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

ББК 32.87
Р12
УДК 534.782.001:621.39

Рабинер Л. Р., Шафер Р. В.

Р12 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: временна́я 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.

ББК 32.87

Р $\frac{30602-185}{046(01)-81}$ 11—81 (С.) 2402040000

6Ф1

Редакция литературы по электросвязи

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 МТИ[1], 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 ТИИЭР[2] 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 Transations». (*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

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 зачительное 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. Патуто, 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
Введение

1.0. Цель книги

Цель книги заключается в том, чтобы показать, как методы цифровой обработки могут быть использованы в задачах речевой связи[1]. 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-

[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 обусловливают 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 сигнала[1] (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 описания[2] 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.

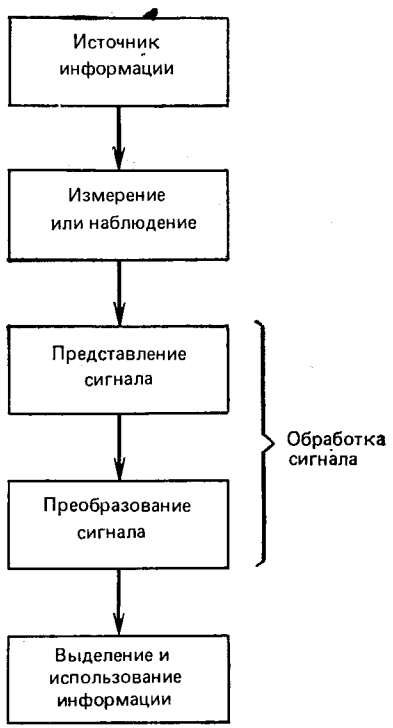


Рис. 1.1. Схема обработки информации

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 быстродействующих 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 реализации 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 частотой[1]. 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 —

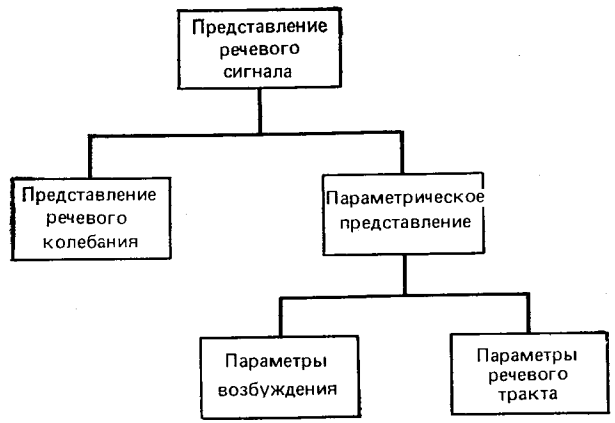


Рис. 1.2. Способы представления речевого сигнала

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 объемом[2]. 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 прикладной 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.

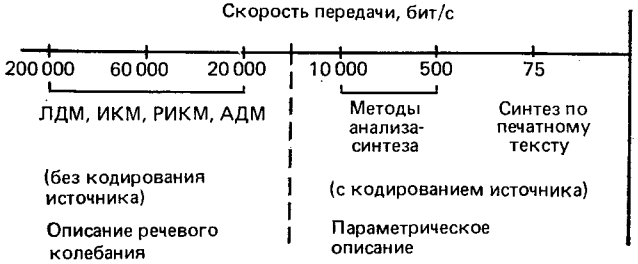


Рис. 1.3. Диапазон скоростей передачи при различном представлении речевого сигнала

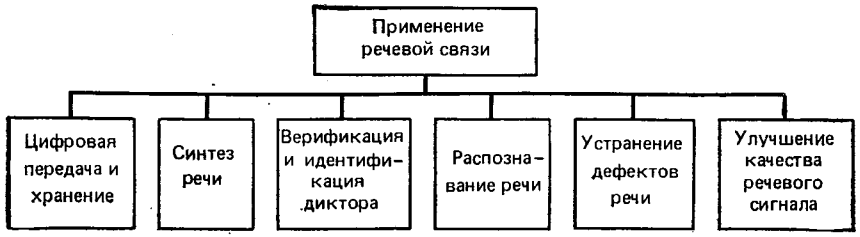


Рис. 1.4. Области применения речевой связи

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 речи[1]. 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 экономического 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 апарат 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

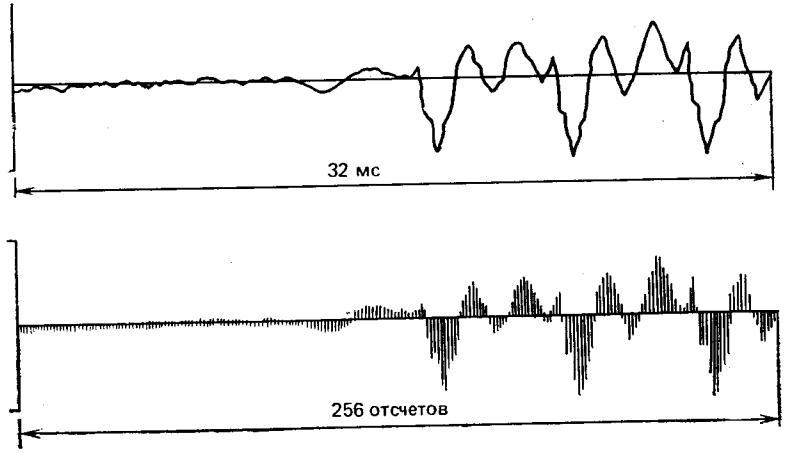


Рис. 2.1. Представление речевого сигнала

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

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

Последовательность единичного скачка имеет вид

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

Экспоненциальная последовательность

$$x(n)=a^n. \tag{2.3}$$

Если a — комплексное число, т. е. $a=re^{i\omega_0}$, то

$$x(n)=r^n e^{i\omega_0 n}=r^n(\cos\omega_0 n+i\sin\omega_0 n). \qquad (2.4)$$

Если $r=1$ и $\omega_0\neq 0$, $x(n)$ — комплексная синусоида; если $\omega_0=0$, $x(n)$ — действительное; если $r<1$ и $\omega_0\neq 0$, то $x(n)$ — экспоненциально-затухающая осциллирующая последовательность. Последовательности этого типа часто используются при представлении линейных систем и моделировании речевых сигналов.

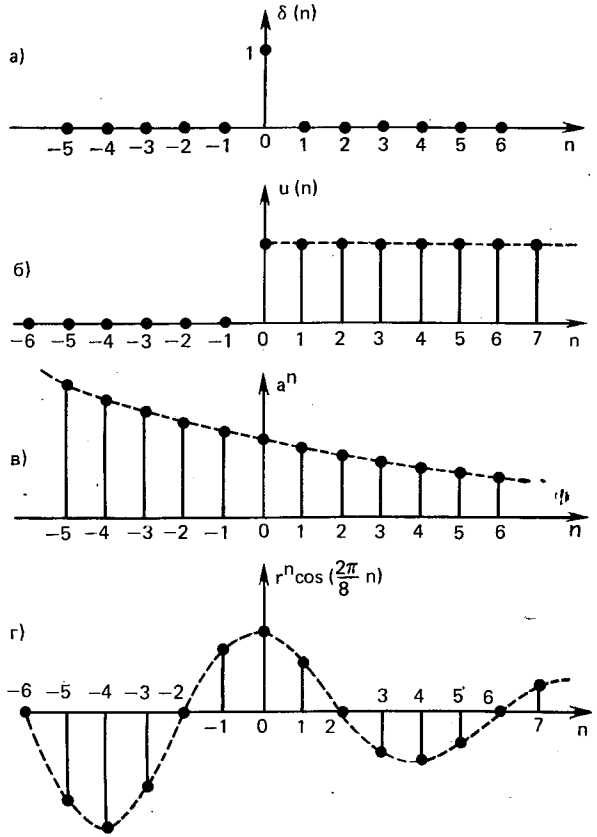


Рис. 2.2. Единичный отсчет ($а$), функция единичного скачка ($б$), действительная экспонента ($в$) и затухающий косинус ($г$)

Обработка сигналов включает преобразование их в форму, удобную для дальнейшего использования. Таким образом, предметом изучения являются дискретные системы или, что то же самое, преобразования входной последовательности в выходную. Подобные преобразования далее изображаются на структурных схемах так, как это показано на рис. 2.3$а$. Многие системы анализа речевых сигналов разработаны для оценивания переменных во времени параметров по последовательности мгновенных значений речевого колебания. Подобные системы имеют многомерный выход, т. е. одномерная последовательность на входе, представляющая собой речевой сигнал, преобразуется в векторную последовательность на выходе, как это изображено на рис. 2.3$б$. В данной книге рассматриваются системы как с одномерным, так и с многомерным выходами.

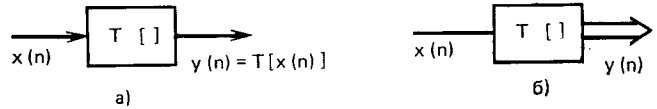


Рис. 2.3. Структурная схема представления сигнала при одномерном входе и выходе ($а$), одномерном входе и многомерном выходе ($б$)

При обработке речевых сигналов особенно широкое применение находят системы, инвариантные к временно́му сдвигу. Такие системы полностью описываются откликом на единичный импульс. Сигнал на выходе системы может быть рассчитан по сигналу на входе и отклику на единичный импульс $h(n)$ с помощью дискретной свертки

$$y(n)=\sum_{k=-\infty}^{\infty}x(k)\,h(n-k)=x(n)*h(n), \qquad (2.5а)$$

где символ * обозначает свертку. Эквивалентное выражение имеет вид

$$y(n)=\sum_{k=-\infty}^{\infty}h(k)\,x(n-k)=h(n)*x(n). \qquad (2.5б)$$

Линейные системы, инвариантные к временно́му сдвигу, применяются при фильтрации сигнала и, что более важно, они полезны как модели речеобразования.

2.2. Описание преобразований сигналов и систем

Анализ сигналов и расчет систем значительно облегчаются при их описании в частотной области. В этой связи полезно кратко остановиться на представлении сигналов и систем в дискретном времени с использованием преобразования Фурье и z-преобразования.

2.2.1. Прямое и обратное z-преобразование

Прямое и обратное z-преобразование последовательности определяется двумя уравнениями:

$$X(z)=\sum_{n=-\infty}^{\infty}x(n)\,z^{-n}; \qquad (2.6а)$$

$$x(n)=\frac{1}{2\pi i}\oint_C X(z)(z)^{n-1}\,dz. \qquad (2.6б)$$

Прямое z-преобразование $x(n)$ определяется уравнением (2.6а). В общем случае $X(z)$ — бесконечный ряд по степеням z^{-1}; последовательность $x(n)$ играет роль коэффициентов ряда. В общем случае подобные степенные ряды сходятся к конечному пределу только для некоторых значений z. Достаточное условие сходимости имеет вид

$$\sum_{n=-\infty}^{\infty}|x(n)||z^{-n}|<\infty. \qquad (2.7)$$

Множество значений, для которых ряды сходятся, образует область на комплексной плоскости, известную как *область сходимости*. В общем случае эта область имеет вид

$$R_1<|z|<R_2. \qquad (2.8)$$

Для того чтобы выявить связь между областью сходимости и структурой последовательности, рассмотрим несколько примеров.

Пример 1. Пусть $x(n)=\delta(n-n_0)$. Тогда подстановка в (2.6а) дает $X(z)=z^{-n_0}$.

Пример 2. Пусть $x(n)=u(n)-u(n-N)$. Тогда $X(z)=\sum_{n=0}^{N-1}(1)z^{-n}=(1-z^{-N})/(1-z^{-1})$.

В обоих случаях $x(n)$ имеет конечную длительность. Таким образом, $X(z)$ есть просто полином переменной z^{-1} и область сходимости представляет собой всю z-плоскость, за исключением точки $z=0$. Все последовательности конечной длительности имеют область сходимости, которая по крайней мере равна $0<|z|<\infty$.

Пример 3. Пусть $x(n)=a^n u(n)$. Тогда

$$X(z)=\sum_{n=0}^{\infty}a^n z^{-n}=\frac{1}{1-az^{-1}},\quad |a|<|z|.$$

В этом случае степенной ряд представляет собой сумму членов геометрической прогрессии, для которой существует удобное замкнутое выражение. Этот результат типичен для последовательностей бесконечной протяженности, отличных от нуля при $n>0$. В общем случае область сходимости имеет вид $|z|>R_1$.

Пример 4. Пусть $x(n)=-b^n u(-n-1)$. Тогда

$$X(z)=\sum_{n=-\infty}^{-1}b^n z^{-n}=\frac{1}{1-bz^{-1}},\quad |z|<|b|.$$

Это типичный результат для бесконечных последовательностей, отличных от нуля при $n<0$, область сходимости которых в общем случае имеет вид $|z|<R_2$. Наиболее общий случай, где $x(n)$ отлично от нуля для $-\infty<n<\infty$, может быть представлен в виде комбинации ситуаций, иллюстрированных в примерах 3 и 4. Для этого случая область сходимости имеет вид $R_1<|z|<R_2$.

Обратное преобразование определяется контурным интегралом (2.6б), где C — замкнутый контур, охватывающий начало координат z-плоскости и расположенный в области сходимости $X(z)$. Для случая действительного z-преобразования удобным способом вычисления обратного преобразования является разложение на простые дроби [1].

Существует много теорем и свойств z-преобразования, полезных при изучении систем в дискретном времени. Хорошее знание и умение применять эти теоремы и свойства на практике являются необходимыми условиями полного понимания материала последующих глав. Перечень основных теорем представлен в табл. 2.1. Можно заметить, что эти теоремы близки к соответствующим теоремам преобразования Лапласа для непрерывных функций. Однако это сходство не должно приводить к ложному выводу, что z-преобразование представляет собой в некотором смысле аппроксимацию преобразования Лапласа. Преобразование Лапласа — это *точное* представление непрерывной функции времени, а z-преобразование — *точное* представление последовательности чисел. Соответствующая связь между непрерывным и дискретным представлениями сигнала может быть установлена на основе теоремы дискретизации, что обсуждается в § 2.4.

Таблица 2.1

Последовательности и их z-преобразования

	Последовательность	z-преобразование
Линейность	$ax_1(n)+bx_2(n)$	$aX_1(z)+bX_2(z)$
Сдвиг	$x(n+n_0)$	$z^{n_0}X(z)$
Экспоненциальное взвешивание	$a^n x(n)$	$X(a^{-1}z)$
Линейное взвешивание	$nx(n)$	$-z\frac{dX(z)}{dz}$
Обращение времени	$x(-n)$	$X(z^{-1})$
Свертка	$x(n)*h(n)$	$X(z)H(z)$
Произведение последовательностей	$x(n)w(n)$	$\frac{1}{2\pi i}\oint_C X(v)W(z/v)v^{-1}dv$

2.2.2. Преобразование Фурье

Описание сигнала в дискретном времени с помощью преобразования Фурье задается в виде

$$X(e^{i\omega})=\sum_{n=-\infty}^{\infty}x(n)e^{-i\omega n}; \qquad (2.9а)$$

$$x(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}X(e^{i\omega})e^{i\omega n}\,d\omega. \qquad (2.9б)$$

Эти уравнения, как нетрудно убедиться, представляют собой частный случай уравнений (2.6). Преобразование Фурье получается путем вычисления z-преобразования на единичной окружности, т. е. подстановкой $z=e^{i\omega}$. Как показано на рис. 2.4, частота ω может быть интерпретирована как угол на z-плоскости. Достаточное условие существования преобразования Фурье можно получить, подставляя $|z|=1$ в (2.7):

$$\sum_{n=-\infty}^{\infty}|x(n)|<\infty. \qquad (2.10)$$

С целью иллюстрации вычисления преобразований Фурье вернемся к примерам из 2.2.1. Преобразование Фурье получается простой заменой. В первых двух примерах в результате действительно получается преобразование Фурье, поскольку единичная окружность входит в область сходимости $X(z)$. Однако в примерах 3 и 4 преобразование Фурье существует только при $|a|<1$ и $|b|>1$ соответственно. Эти условия, конечно, соответствуют затуханию последовательностей, для которых выполнено соотношение (2.10).

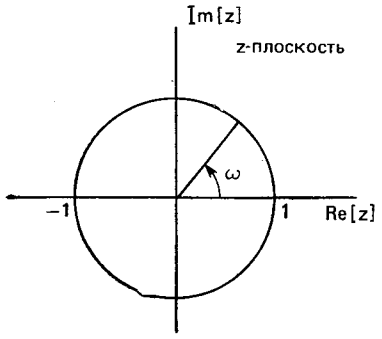


Рис. 2.4. Единичная окружность на z-плоскости

Важная особенность преобразования Фурье последовательности состоит в том, что оно является периодической функцией ω с периодом 2π. Это легко показать, подставив $\omega+2\pi$ в соотношение (2.9а). С другой стороны, поскольку $X(e^{i\omega})$ представляет собой значение $X(z)$ на единичной окружности, оно должно повторяться после каждого полного обхода этой окружности, т. е. когда ω изменится на 2π рад.

Подставляя $z=e^{i\omega}$ в условие каждой теоремы в табл. 2.1, получим соответствующее множество теорем для преобразования Фурье. Конечно, эти результаты справедливы лишь при условии существования преобразований Фурье.

2.2.3. Дискретное преобразование Фурье

Как и в случае аналоговых сигналов, если последовательность периодическая с периодом N, т. е.

$$\tilde{x}(n)=\tilde{x}(n+N), -\infty<n<\infty, \tag{2.11}$$

то $\tilde{x}(n)$ можно представить в виде суммы синусоид, а не в виде интеграла, как это было в (2.9б). Преобразование Фурье для периодической последовательности имеет вид

$$\tilde{X}(k)=\sum_{n=0}^{N-1}\tilde{x}(n)\,e^{-i2\pi kn/N}; \tag{2.12а}$$

$$\tilde{x}(n)=\frac{1}{N}\sum_{k=0}^{N-1}\tilde{X}(k)\,e^{i\,2\pi\,kn/N}. \tag{2.12б}$$

Это *точное* представление периодической последовательности. Однако, основное преимущество данного описания заключается в возможности несколько иной интерпретации уравнений (2.12). Рассмотрим последовательность конечной длины $x(n)$, равную нулю

вне интервала $0\leqslant n\leqslant N-1$. В этом случае z-преобразование имеет вид

$$X(z)=\sum_{n=0}^{N-1}x(n)\,z^{-n}. \tag{2.13}$$

Если записать $X(z)$ в N равноотстоящих точках единичной окружности, т. е. $z_k=e^{i2\pi k/N}$, $k=0, 1, \ldots, N-1$, то получим

$$X(e^{i2\pi k/N})=\sum_{n=0}^{N-1}x(n)\,e^{-i2\pi kn/N}, k=0, 1, \ldots, N-1. \tag{2.14}$$

Если при этом построить периодическую последовательность в виде бесконечного числа повторений сегмента $x(n)$,

$$\tilde{x}(n)=\sum_{r=-\infty}^{\infty}x(n+rN), \tag{2.15}$$

то отсчеты $X(e^{i2\pi k/N})$, как это видно из (2.12а) и (2.14), будут представлять собой коэффициенты Фурье периодической последовательности $\tilde{x}(n)$ в (2.15). Таким образом, последовательность длиной N можно точно описать с помощью дискретного преобразования Фурье (ДПФ) в виде

$$X(k)=\sum_{n=0}^{N-1}x(n)\,e^{i2\pi kn/N}, k=0, 1, \ldots N-1; \tag{2.16а}$$

$$x(n)=\frac{1}{N}\sum_{k=0}^{N-1}X(k)\,e^{i2\pi kn/N}, n=0, 1, \ldots, N-1. \tag{2.16б}$$

Очевидно, что различие между соотношениями (2.16) и (2.12) состоит лишь в небольшом изменении обозначений (опущен символ, означающий периодичность) и в ограничении интервалов $0\leqslant k\leqslant N-1$ и $0\leqslant n\leqslant N-1$. Следует, однако, иметь в виду, что все последовательности при использовании ДПФ ведут себя так, как если бы они были периодическими функциями, т. е. ДПФ является на самом деле представлением периодической функции времени, заданной (2.15). Несколько иной подход при использовании ДПФ заключается в том, что индексы последовательности интерпретируются по модулю N. Это следует из того факта, что если $x(n)$ имеет длину N, то

$$\tilde{x}(n)=\sum_{k=-\infty}^{\infty}x(n+rN)=x(n\text{ по модулю }N)=x((n))_N. \tag{2.17}$$

Введение двойных обозначений позволяет отразить периодичность, присущую представлению с помощью ДПФ. Эта периодичность существенно отражается на свойствах ДПФ. Наиболее важные теоремы о ДПФ представлены в табл. 2.2. Очевидно, что задержка последовательности должна рассматриваться по модулю

Таблица 2.2

Последовательности и их дискретное преобразование Фурье

	Последовательность	N-точечное ДПФ
Линейность	$ax_1(n)+bx_2(n)$	$aX_1(k)+bX_2(k)$
Сдвиг	$x((n+n_0))_N$	$e^{(i\,2\pi/N)\,kn_0}X(k)$
Обращение времени	$x((n))_N$	$X^*(k)$
Свертка	$\sum_{m=0}^{N-1} x(m)\,h((n-m))_N$	$X(k)H(k)$
Умножение последовательностей	$x(n)w(n)$	$\frac{1}{N}\sum_{r=0}^{N-1} X(r)\,W((k-r))_N$

N. Это приводит, например, к некоторым особенностям выполнения дискретной свертки.

Дискретное преобразование Фурье со всеми его особенностями является важным способом описания сигналов по следующим причинам: 1) ДПФ можно рассматривать как дискретизированный вариант z-преобразования (или преобразования Фурье) последовательности конечной длительности; 2) ДПФ очень сходно по своим свойствам (с учетом периодичности) с преобразованием Фурье и z-преобразованием; 3) N значений $X(k)$ можно вычислить с использованием эффективного (время вычисления пропорционально $N\log N$) семейства алгоритмов, известных под названием быстрых преобразований Фурье (БПФ) [1—4].

Дискретное преобразование Фурье широко используется при вычислении корреляционных функций, спектров и при реализации цифровых фильтров [5—6], а также часто используется и при обработке речевых сигналов.

2.3. Основы цифровой фильтрации

Цифровой фильтр представляет собой систему с постоянными параметрами (инвариантную к сдвигу), работающую в дискретном времени. Напомним, что для таких систем сигнал на входе и выходе связан дискретной сверткой (2.5). Соответствующее соотношение между z-преобразованиями, как это видно из табл. 2.1, имеет вид

$$Y(z)=H(z)\,X(z). \tag{2.18}$$

Прямое z-преобразование отклика на единичный импульс $H(z)$ называется *передаточной функцией* системы. Преобразование Фурье отклика на единичный импульс $H(e^{i\omega})$ называется *частотной характеристикой*. Обычно $H(e^{i\omega})$ представляет собой комплексную функцию ω, которую можно записать в виде

$$H(e^{i\omega})=H_r(e^{i\omega})+i\,H_i(e^{i\omega}) \tag{2.19}$$

или через модуль и фазу:

$$H(e^{i\omega})=|H(e^{i\omega})|\,e^{i\arg[H(e^{i\omega})]}. \tag{2.20}$$

Инвариантная к сдвигу линейная система называется *физически реализуемой*, если $h(n)=0$ при $n<0$. Линейная система устойчива, если для любой ограниченной по уровню входной последовательности выходная последовательность также ограничена. Необходимым и достаточным условием устойчивости линейной системы с постоянными параметрами является

$$\sum_{n=-\infty}^{\infty}|h(n)|<\infty. \tag{2.21}$$

Это условие аналогично (2.10) и оказывается достаточным для существования $H(e^{i\omega})$.

Сигналы на входе и выходе линейных инвариантных к сдвигу систем, таких, например, как фильтры, связаны дискретной сверткой (2.5) и, кроме того, разностным уравнением

$$y(n)-\sum_{k=1}^{N} a_k\,y(n-k)=\sum_{r=0}^{M} b_r\,x(n-r). \tag{2.22}$$

Вычисляя z-преобразование от обеих частей, можно получить

$$H(z)=\frac{Y(z)}{X(z)}=\frac{\sum_{r=0}^{M} b_r\,z^{-r}}{1-\sum_{k=1}^{N} a_k\,z^{-k}}. \tag{2.23}$$

Сравнивая (2.22) и (2.23), полезно отметить следующее. Если задано разностное уравнение вида (2.22), то $H(z)$ можно получить непосредственной подстановкой коэффициентов при входном сигнале в числитель передаточной функции к соответствующим степеням z^{-1}, а коэффициенты при выходном сигнале — в знаменатель к соответствующим степеням z^{-1}.

Передаточная функция в общем случае является дробно рациональной. Таким образом, она определяется положением нулей и полюсов на z-плоскости. Это означает, что $H(z)$ можно представить в виде

$$H(z)=\frac{A\prod_{r=1}^{M}(1-c_r\,z^{-1})}{\prod_{k=1}^{N}(1-d_k\,z^{-1})}. \tag{2.24}$$

При рассмотрении z-преобразования отмечалось, что физически реализуемые системы имеют область сходимости вида $|z|>R_1$. Если система, кроме того, еще и устойчива, то R_1 должно быть меньше единицы, таким образом единичная окружность входит в область сходимости. Иначе говоря, для устойчивой системы все полюсы $H(z)$ должны лежать внутри единичной окружности.

Достаточно определить два типа линейных систем с постоянными параметрами. Это системы с конечной импульсной характеристикой (КИХ) и системы с бесконечной импульсной характеристикой (БИХ). Эти два класса обладают отличными друг от друга свойствами, которые будут рассмотрены ниже.

2.3.1. Системы с конечными импульсными характеристиками

Если все коэффициенты a_k в уравнении (2.22) равны нулю, то разностное уравнение принимает вид

$$y(n)=\sum_{r=0}^{M} b_r\, x(n-r). \tag{2.25}$$

Сравнивая (2.25) с (2.5б), можно отметить, что

$$h(n)=\begin{cases} b_n, & 0\leqslant n\leqslant M,\\ 0, & \textit{в противном случае.}\end{cases} \tag{2.26}$$

Системы с КИХ обладают рядом важных свойств. Передаточная функция $H(z)$ таких систем представляет собой полином по степеням z^{-1} и, таким образом, не имеет ненулевых полюсов, а содержит только нули. Системы с КИХ могут обладать строго линейной фазо-частотной характеристикой (ФЧХ). Если $h(n)$ удовлетворяет условию

$$h(n)=\pm h(M-n), \tag{2.27}$$

то $$H(e^{i\omega})=A(e^{i\omega})\,e^{-i\omega(M/2)}, \tag{2.28}$$

где $A(e^{i\omega})$ — действительная или чисто мнимая величина в зависимости от знака в (2.27).

Возможность получения *строго* линейной ФЧХ является очень важным обстоятельством применительно к речевым сигналам в тех случаях, когда требуется сохранить взаимное расположение элементов сигнала. Это свойство систем с КИХ существенно облегчает решение задачи их проектирования, поскольку все внимание можно уделять лишь аппроксимации амплитудно-частотной характеристики (АЧХ). За это достоинство фильтра с линейной ФЧХ приходится расплачиваться необходимостью аппроксимации протяженной импульной реакции в случае фильтров с крутыми АЧХ.

Хорошо разработаны три метода проектирования КИХ-фильтров с линейными ФЧХ: взвешивания [1, 2, 5, 7], частотной выборки [1, 2, 8] и проектирования оптимальных фильтров с минимаксной ошибкой [1, 2, 9—11]. Первый из трех перечисленных методов чисто аналитический, т. е. приводит к замкнутой системе уравнений относительно коэффициентов фильтра. Второй и третий методы являются оптимизационными и используют итеративный (в отличие от замкнутой формы) подход для определения коэффициентов фильтра. Несмотря на простоту метода взвешивания, широкое применение нашли все три метода. Это обусловлено завершенностью глубоких исследований оптимальных КИХ-фильтров и, кроме того, наличием подробно описанных программ, позволяющих пользователю легко рассчитать любой фильтр [2, 10].

При рассмотрении вопросов реализации цифровых фильтров полезно изображать их в виде схем. Разностное уравнение (2.25) изображено на рис. 2.5. Подобные схемы, называемые структурными, описывают в графической форме те операции, которые необходимо проделать над входной последовательностью для получения сигнала на выходе. Основные элементы на схеме отображают устройства суммирования, умножения на постоянные коэффициенты (последние показаны возле стрелок, обозначающих операцию умножения) и хранения последних значений входной последовательности. Эта структурная схема позволяет составить наглядное представление о сложности устройства. Когда система обладает линейной фазо-частотной характеристикой, возможны дополнительные упрощения в технической реализации (см. задачу 2.7).

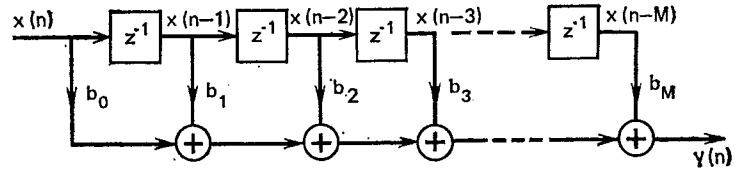


Рис. 2.5. Структурная схема КИХ-фильтра

2.3.2. Системы с бесконечными импульсными характеристиками

Если передаточная функция (2.24) имеет полюсы и нули, то разностное уравнение (2.22) можно переписать в виде

$$y(n)=\sum_{k=1}^{N} a_k\, y(n-k)+\sum_{r=0}^{M} b_r\, x(n-r). \tag{2.29}$$

Это уравнение представляет собой рекуррентную формулу, которая может использоваться для последовательного вычисления текущего значения выходного сигнала по прошлым значениям выходного сигнала. Если в (2.24) $M<N$, то $H(z)$ можно разложить на простые дроби:

$$H(z)=\sum_{k=1}^{N}\frac{A_k}{1-d_k z^{-1}}. \tag{2.30}$$

Для физически реализуемых систем легко показать (см. задачу 2.9), что

$$h(n)=\sum_{k=1}^{N} A_k (d_k)^n u(n). \qquad (2.31)$$

Таким образом, $h(n)$ имеет бесконечную протяженность. Однако вследствие рекуррентности соотношения (2.29) часто оказывается возможным построить БИХ-фильтр, позволяющий получить тот же результат с более высокой эффективностью (т. е. при меньшем объеме вычислений), чем при использовании КИХ-фильтра. Это особенно справедливо для фильтров с крутым срезом.

Существует ряд методов проектирования БИХ-фильтров. Расчет частотно-селективных фильтров (ФНЧ, полосовых и т. д.) часто основывается на известных методах проектирования аналоговых фильтров. В этот класс входят методы Баттерворта (максимально плоской АЧХ), Бесселя (равного времени групповой задержки), Чебышева (равновеликих колебаний в полосе пропускания или затухания) и метод равновеликих колебаний как в полосе пропускания, так и в полосе затухания. Перечисленные методы являются аналитическими и широко применяются при проектировании цифровых КИХ-фильтров [1, 2]. В дополнение к перечисленным, разработан ряд оптимизационных методов расчета фильтров [12].

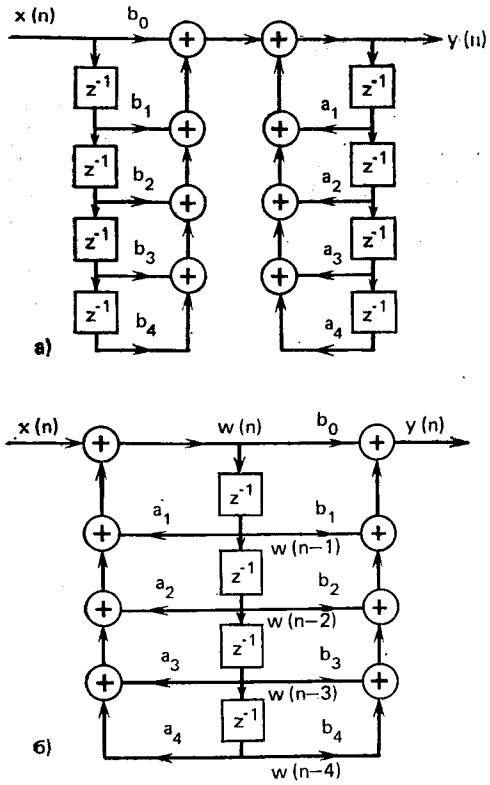


Рис. 2.6. Прямая форма БИХ-фильтра (*а*) и БИХ-фильтра с минимальной памятью (*б*)

Главное отличие БИХ-фильтров от КИХ-фильтров состоит в том, что невозможно спроектировать БИХ-фильтр со строго линейной ФЧХ, в то время как КИХ-фильтр может обладать такой характеристикой [13].

Фильтры с БИХ обладают большой гибкостью при реализации. На рис. 2.6*а* показан фильтр, описываемый уравнением (2.29) для случая $M=N=4$. Этот способ реализации называют прямой формой. Обобщение на случай произвольных M и N очевидно. Уравнение

(2.29) можно преобразовать в ряд различных эквивалентных форм. Особенно полезной среди них является форма, описываемая уравнениями

$$w(n)=\sum_{k=1}^{N} a_k w(n-k)+x(n)\,;\; y(n)=\sum_{r=0}^{M} b_r w(n-r) \qquad (2.32)$$

(см. задачу 2.10). Эта система уравнений может быть реализована так, как это показано на рис. 2.6*б*, что существенно экономит объем памяти, необходимый для хранения значений входной и выходной последовательностей.

Соотношение (2.24) показывает, как можно выразить $H(z)$ через нули и полюсы. Нули и полюсы появляются на z-плоскости комплексно-сопряженными парами, поскольку a_k и b_r действитель-

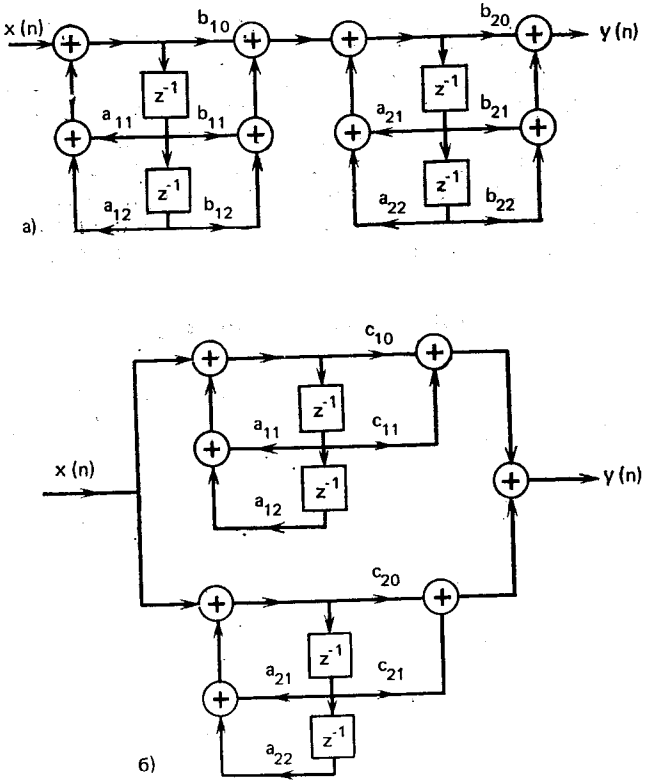


Рис. 2.7. Каскадная (*а*) и параллельная (*б*) формы цифрового фильтра

ы. Группируя комплексно-сопряженные нули и полюсы, $H(z)$ жно выразить в виде произведения элементарных передаточных нкций второго порядка

$$H(z)=A\prod_{k=1}^{K}\left[\frac{1+b_{1k}z^{-1}+b_{2k}z^{-2}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}\right], \qquad (2.33)$$

где K — целая часть $(N+1)/2$. Любая система, таким образом, может быть реализована в виде каскадного соединения систем второго порядка, каждая из которых реализуется в виде рис. 2.6. На рис. 2.7*а* такое соединение изображено для случая $N=M=4$. Обобщение на случай системы произвольного порядка очевидно. Разложение на простые дроби (2.30) предполагает несколько иную реализацию. Объединяя комплексно-сопряженные полюсы, $H(z)$ можно представить в виде

$$H(z)=\sum_{k=1}^{K}\frac{c_{0k}+c_{1k}z^{-1}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}. \qquad (2.34)$$

Это приводит к параллельной форме реализации, которая изображена на рис. 2.7*б* для $N=4$.

Все рассмотренные способы реализации используются и при обработке речи. Каскадная форма часто предпочтительнее, так как наименее чувствительна к шумам округления, точности квантования коэффициентов и нарушениям устойчивости [1, 2]. Все перечисленные выше формы используются при построении синтезаторов речевого сигнала, а прямая форма особенно важна при синтезе по параметрам линейного предсказания (см. гл. 8).

2.4. Дискретизация

Для применения методов цифровой обработки к такому аналоговому сигналу, как речевое колебание, необходимо представить его в виде последовательности чисел. Обычно это осуществляется путем периодической дискретизации аналогового сигнала для получения последовательности его значений

$$x(n)=x_a(nT), \quad -\infty<n<\infty, \qquad (2.35)$$

где n, конечно, принимает только целые значения. На рис. 2.1 показан речевой сигнал и соответствующая последовательность отсчетов, взятых с периодом $T=1/8000$ с.

2.4.1. Теорема дискретизации [1]

Условия, которые должны выполняться для того, чтобы аналоговый сигнал можно было представить последовательностью своих отсчетов единственным образом, хорошо известны и часто формулируются в следующем виде.

Теорема дискретизации: если сигнал $x_a(t)$ имеет преобразование Фурье $X_a(i\Omega)$ такое, что $X_a(i\Omega)=0$ при $\Omega\geqslant 2\pi F_N$, то $x_a(t)$ может быть восстановлен единственным образом по последовательности равноотстоящих отсчетов $x_a(nT)$, $-\infty<n<\infty$, если $1/T>2F_N$.

[1] См. прим. ред. на с. 13.

Данная теорема вытекает из того факта, что если преобразование Фурье сигнала $x_a(t)$ определяется выражением

$$X_a(i\Omega)=\int_{-\infty}^{\infty}x_a(t)e^{-i\Omega t}\,dt \qquad (2.36)$$

и преобразование Фурье последовательности $x(n)$ определено в соответствии с (2.9а), то в частотной области выполняется соотношение [1, 2]

$$X(e^{i\Omega T})=\frac{1}{T}\sum_{k=-\infty}^{\infty}X_a\left(i\Omega+i\frac{2\pi}{T}k\right). \qquad (2.37)$$

Для пояснения соотношения (2.37) предположим, что $X_a(i\Omega)$ имеет вид, показанный на рис. 2.8*а*, т. е. допустим, что $X_a(i\Omega)=0$ для $|\Omega|>\Omega_N=2\pi F_N$. Частоту F_N называют частотой Найквиста. В

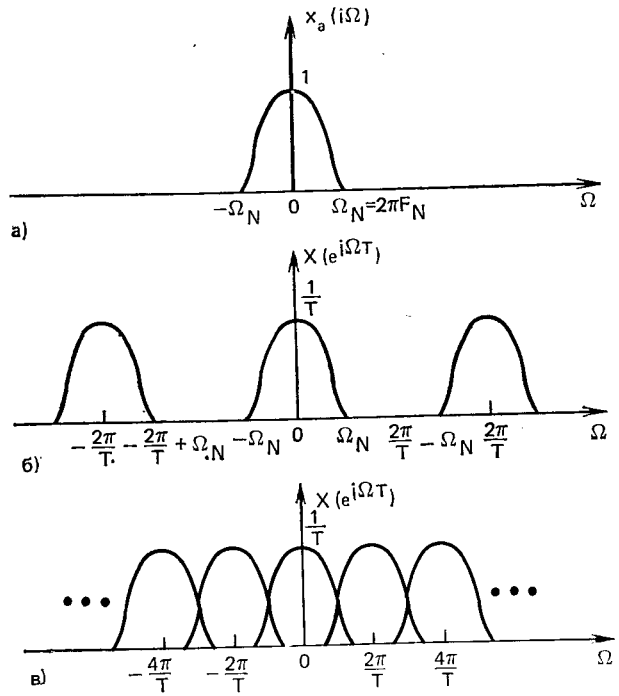


Рис. 2.8. Дискретизация

соответствии с (2.37) $X(e^{i\Omega T})$ представляет собой сумму бесконечного числа спектров $X_a(i\Omega)$, каждый из которых расположен на высших гармониках частоты $2\pi/T$. На рис. 2.8*б* показан случай, когда $1/T>2F_N$. Здесь дополнительные компоненты преобразования Фурье не попадают в основной диапазон $|\Omega|<2\pi F_N$. На рис. 2.8*в* приведен обратный случай, когда $1/T<2F_N$. Здесь спектры,

отстоящие друг от друга на $2\pi/T$, пересекаются. Такая ситуация, при которой смежные спектры перекрываются, называется *наложением частот*. Очевидно, что наложения частот можно избежать только при условии, что преобразование Фурье исходного сигнала ограничено по полосе частот и частота дискретизации, по крайней мере, равна удвоенной частоте Найквиста ($1/T > 2F_N$).

Если $1/T > 2F_N$, то преобразование Фурье последовательности отсчетов пропорционально преобразованию Фурье аналогового сигнала в основной полосе частот:

$$X(e^{i\Omega T}) = X_a(i\Omega)/T, \quad |\Omega| < \pi/T. \tag{2.38}$$

Используя этот результат, можно показать, что исходный сигнал связан с последовательностью отсчетов следующей формулой [1, 2]:

$$x_a(t) = \sum_{n=-\infty}^{\infty} x_a(nT) \left[\frac{\sin[\pi(t-nT)/T]}{\pi(t-nT)/T} \right]. \tag{2.39}$$

Таким образом, по последовательности отсчетов аналогового сигнала, взятых с частотой, равной, по крайней мере, удвоенной частоте Найквиста, можно по (2.39) восстановить исходный аналоговый сигнал. Применяемые на практике цифроаналоговые преобразователи основаны на приближении соотношения (2.39).

Дискретизация предполагается во многих алгоритмах обработки речевых сигналов, предназначенных для оценки таких важных параметров речи, как частоты формант или период основного тона. В этих случаях аналоговая функция, подвергаемая дискретизации, недоступна наблюдению. Однако параметры изменяются во времени медленно, и поэтому их можно оценивать со скоростью порядка 100 отсч./с (т. е. дискретизировать). Полученные отсчеты параметра являются значениями ограниченной по частоте функции, которую можно восстановить в соответствии с (2.39).

2.4.2. Прореживание и интерполяция дискретизированного сигнала

В ряде примеров, рассматриваемых в книге, возникает задача изменения частоты дискретизации сигнала, представленного в дискретном времени. Такая задача появляется, например, когда сигнал, дискретизированный с высокой частотой и представленный в разностной форме с использованием двухуровневого квантователя (дельта-модуляция), преобразуется в многоуровневый сигнал ИКМ с более низкой частотой дискретизации. Другой пример соответствует случаю, в котором параметр речевого сигнала дискретизируют с низкой частотой для более эффективного кодирования, а для восстановления сигнала требуется более высокая частота дискретизации. В первом случае частоту дискретизации следует понизить, а во втором — повысить. Процесс понижения и повышения частоты дискретизации будет далее называться прореживанием и интерполяцией соответственно. В обоих случаях будем предполагать, что имеется последовательность отсчетов $x(n) = x_a(nT)$, где аналоговая функция $x_a(t)$ имеет ограниченное по частоте преобразование Фурье, такое, что $X_a(i\Omega) = 0$, $|\Omega|/2\pi > F_N$. Как было показано выше, если $1/T > 2F_N$, то преобразование Фурье удовлетворяет соотношению

$$X(e^{i\Omega T}) = X_a(i\Omega)/T, \quad |\Omega| < \pi/T. \tag{2.40}$$

Прореживание. Пусть требуется понизить частоту дискретизации в M раз, т. е. необходимо построить новую последовательность, соответствующую отсчетам $x_a(t)$, взятым с периодом $T' = MT$, т. е.

$$y(n) = x_a(nT') = x_a(nTM). \tag{2.41}$$

Легко видеть, что

$$y(n) = x(Mn), \quad -\infty < n < \infty. \tag{2.42}$$

Таким образом, $y(n)$ получается путем сохранения только одного из M отсчетов. Из приведенного выше обсуждения теоремы дискретизации следует, что если $1/T' > 2F_N$, то $y(n)$ также единственным образом описывает исходный аналоговый сигнал. Преобразования Фурье $x(n)$ и $y(n)$ связаны соотношением [14]

$$Y(e^{i\Omega T}) = \frac{1}{M} \sum_{k=0}^{M-1} X\left(e^{\frac{i(\Omega T' - 2\pi k)}{M}}\right). \tag{2.43}$$

Из (2.43) видно, что для устранения наложения между спектрами необходимо, чтобы $1/T' > 2F_N$. Если это условие выполняется, то получаем

$$Y(e^{i\Omega T'}) = \frac{1}{M} X(e^{i\Omega T'/M}) = \frac{1}{M}\frac{1}{T} X_a(i\Omega) = \frac{1}{T'} X_a(i\Omega), \quad -\pi/T' < \Omega < \pi/T'. \tag{2.44}$$

На рис. 2.9 показан пример снижения частоты дискретизации. На рис. 2.9*а* приведено преобразование Фурье исходного аналогового сигнала. На рис. 2.9*б* показано преобразование Фурье для $x(n) = x_a(nT)$, где частота дискретизации ($1/T$) несколько больше частоты Найквиста ($2F_N$). На рис. 2.9*в* представлен случай снижения втрое частоты дискретизации, т. е. $T' = 3T$. Здесь возникает наложение частот, поскольку $1/T' < 2F_N$. Пусть $x(n)$ формируется на выходе ФНЧ с частотой среза $\pi/T' = \pi/(3T)$ и далее получается последовательность $w(n)$. Преобразование Фурье сигнала на выходе ФНЧ показано на рис. 2.9*г*. В данном случае при снижении частоты дискретизации втрое наложения частот не возникает, но последовательность отсчетов $y(n)$ относится уже не к сигналу $x(t)$, а к сигналу $y_a(t)$, полученному на выходе фильтра нижних частот. Структурная схема обобщенной системы прореживания приведена на рис. 2.10.

Интерполяция. Пусть имеется последовательность отсчетов аналогового сигнала $x(n)=x_a(nT)$. Если необходимо повысить частоту дискретизации в L раз, то следует вычислить новую последовательность, соответствующую отсчетам $x_a(t)$, взятым с периодом $T'=T/L$, т. е.

$$y(n)=x_a(nT')=x_a(nT/L). \tag{2.45}$$

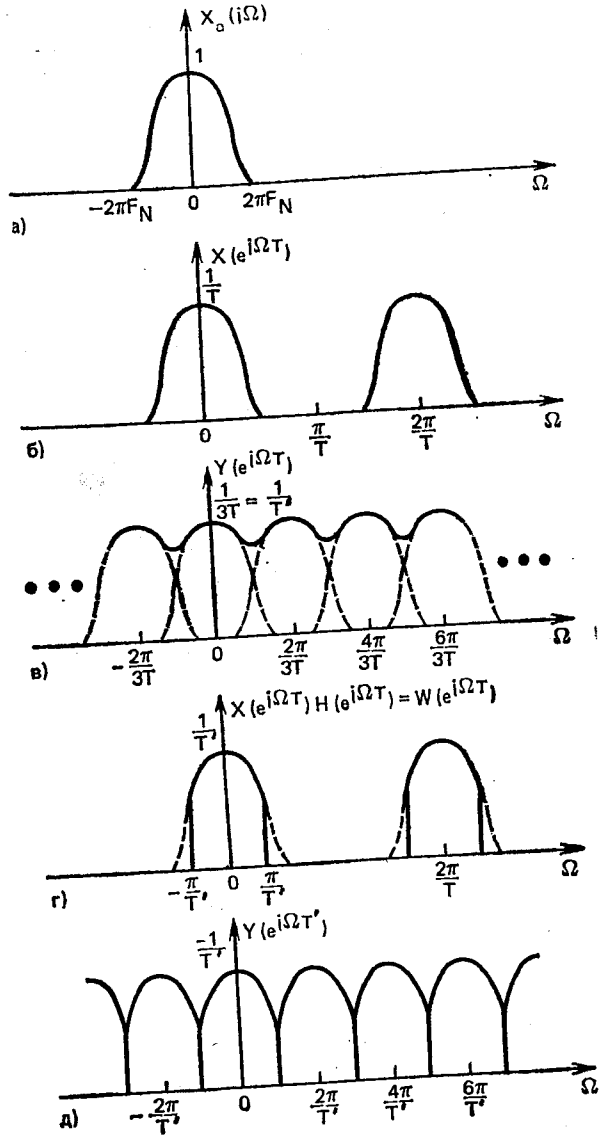


Рис. 2.9. Прореживание

Очевидно, $y(n)=x(n/L)$ для $n=0, \pm L, \pm 2L$, но для других значений недостающие отсчеты необходимо получить с использованием методов интерполяции [14]. Для того чтобы выяснить, как

Рис. 2.10. Структурная схема прореживания

это сделать с помощью цифрового фильтра, рассмотрим последовательность

$$v(n)=\begin{cases} x,\left(\dfrac{n}{L}\right), & n=0,\pm L,\pm 2L,\ \ldots, \\ 0, & \textit{в противном случае.}\end{cases} \tag{2.46}$$

Легко показать [14], что преобразование Фурье $v(n)$ имеет вид

$$V(e^{i\Omega T'})=X(e^{i\Omega T'L})=X(e^{i\Omega T}). \tag{2.47}$$

Таким образом, $V(e^{i\Omega T'})$ оказывается периодической функцией с периодом $2\pi/T=2\pi/LT'$. На рис. 2.11*а* показан $V(e^{i\Omega T'})$ и $(X(e^{i\Omega T}))$ при $T'=T/3$. Для получения последовательности $y(n)=x_a(nT')$ по $v(n)$ необходимо, чтобы

$$Y(e^{i\Omega T'})=X_a(i\Omega)/T', \quad -\pi/T'\leqslant\Omega\leqslant\pi/T'. \tag{2.48}$$

Предположим, что

$$X(e^{i\Omega T})=X_a(i\Omega)/T, \quad -\pi/T\leqslant\Omega\leqslant\pi/T. \tag{2.49}$$

Тогда из рис. 2.11*б* видно, что необходимо с помощью фильтра нижних частот выделить составляющие, расположенные в полосе частот $-\pi/T\leqslant\leqslant\Omega\leqslant\pi/T$, и подавить гармоники, расположенные вокруг частот $\Omega=\pm 2\pi/T$ и $\Omega=4\pi/T$. Более того, для сохранения амплитудных значений, соответствующих интервалу дискретизации T', коэффициент усиления фильтра должен быть равен $L=T/T'$. Таким обра-

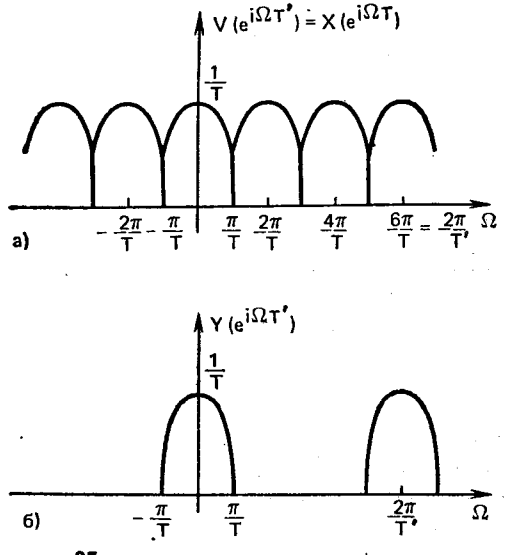


Рис. 2.11. Интерполяция

$$Y(e^{i\Omega T'}) = H(e^{i\Omega T'})\,V(e^{i\Omega T'}) = H(e^{i\Omega T'})\,X(e^{i\Omega T}) =$$
$$= H(e^{i\Omega T'})\,\frac{1}{T}\,X_a(i\,\Omega). \qquad (2.50)$$

Для того чтобы $Y(e^{i\Omega T'}) = (1/T)X_a(i\Omega)$, $\Omega \leqslant \pi/T'$, необходимо выполнение соотношения

$$H(e^{i\Omega T'} = \begin{cases} L, & |\Omega| \leqslant \pi/T, \\ 0, & \textit{в противном случае.} \end{cases} \qquad (2.51)$$

Общая структурная схема процесса интерполяции представлена на рис. 2.12.

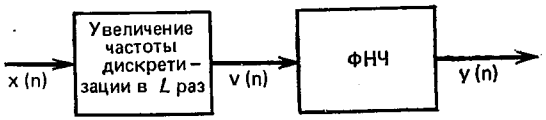


Рис. 2.12. Структурная схема интерполяции

Изменение частоты дискретизации в дробное число раз. Отсчеты, соответствующие периоду дискретизации $T' = MT/L$, можно получить путем комбинации интерполяции с параметром L и последующей процедуры прореживания с параметром M. Соответствующим подбором целых чисел M и L можно получить любое необходимое соотношение между частотами дискретизации. Объединив структурные схемы на рис. 2.9 и 2.11, легко заметить, что вместо двух достаточно иметь один фильтр нижних частот (рис. 2.13).

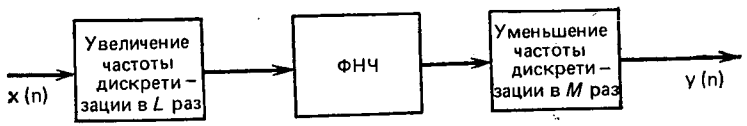


Рис. 2.13. Структурная схема повышения частоты дискретизации

Преимущества КИХ-фильтров. Важным аспектом при использовании методов интерполяции и прореживания является выбор фильтра нижних частот. Значительная экономия в объеме вычислений в таких системах достигается использованием КИХ-фильтров в стандартной прямой форме. Экономия в вычислениях достигается вследствие того, что при прореживании только один из каждых M отсчетов подвергается фильтрации, а при интерполяции каждые $L-1$ из L отсчетов равны нулю и потому не влияют на процесс вычисления. Это обстоятельство трудно использовать в полной мере при применении БИХ-фильтров [14].

Если предположить, что фильтрация будет осуществлена с использованием КИХ-фильтра нижних частот, то для большого изменения частоты дискретизации (т. е. большого M при прореживании и большого L при интерполяции) более целесообразным оказывается уменьшать (или увеличивать) частоту дискретизации с помощью серии последовательных прореживаний. В этом случае

частота дискретизации уменьшается постепенно и на каждом шаге требуется фильтр нижних частот с менее крутым спадом частотной характеристики. Детальное исследование процедуры последовательного прореживания и интерполяции, а также узкополосной фильтрации содержится в работах [15—18].

2.5. Заключение

В этой главе представлен обзор основ теории обработки сигналов в дискретном времени. Введенные здесь понятия дискретной свертки, разностного уравнения, а также описание сигналов и систем в частотной области будут широко использоваться при последующем изложении. Вопросы дискретизации и изменения частоты дискретизации, рассмотренные в § 2.4, также чрезвычайно важны для систем цифровой обработки речевых сигналов.

Задачи

2.1. Рассмотрим последовательность

$$x(n) = \begin{cases} a^n, & n \geqslant n_0 \\ 0, & n < n_0. \end{cases}$$

а) Определить z-преобразование $x(n)$.

б) Определить преобразование Фурье $x(n)$ и указать условия существования преобразования Фурье.

2.2. На входе стационарной линейной системы действует сигнал

$$x(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N-1; \\ 0, & \textit{в противном случае.} \end{cases}$$

Импульсная характеристика системы имеет вид

$$h(n) = \begin{cases} a^n, & n \geqslant 0; \\ 0, & n < 0. \end{cases}$$

а) Используя понятие дискретной свертки, определить сигнал на выходе системы для любого n.

б) Определить z-преобразование выходного сигнала.

2.3. Определить z-преобразование и преобразование Фурье для каждой из следующих последовательностей (все они часто применяются как весовые функции — «окна» — при обработке речевых сигналов).

Экспоненциальное окно

$$w_1(n) = \begin{cases} a^n, & 0 \leqslant n \leqslant N-1; \\ 0, & \textit{в противном случае.} \end{cases}$$

Прямоугольное окно

$$w_2(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N-1; \\ 0, & \textit{в противном случае.} \end{cases}$$

Окно Хемминга

$$w_3(n) = \begin{cases} 0{,}54 - 0{,}46\cos[2\pi n/(N-1)], & 0 \leqslant n \leqslant N-1; \\ 0, & \textit{в противном случае.} \end{cases}$$

Изобразить амплитуды преобразования Фурье для каждого случая. Указание: получить соотношение между $W_3(e^{i\omega})$ и $W_2(e^{i\omega})$.

2.4. Частотная характеристика идеального фильтра нижних частот имеет вид

$$H(e^{i\omega}) = \begin{cases} 1, & |\omega| < \omega_c; \\ 0, & \omega_c < |\omega| \leqslant \pi \end{cases}$$

($H(e^{i\omega})$, разумеется, периодическая с периодом 2π).

а) Определить импульсную характеристику идеального фильтра нижних частот.

б) Изобразить импульсную характеристику для $\omega_c = \pi/4$. Частотная характеристика идеального полосового фильтра имеет вид

$$H(e^{i\omega}) = \begin{cases} 1, & \omega_a < |\omega| < \omega_b; \\ 0, & |\omega| < \omega_a \text{ и } \omega_b < |\omega| \leqslant \pi. \end{cases}$$

в) Определить импульсную характеристику идеального полосового фильтра.

г) Изобразить импульсную характеристику для $\omega_a = \pi/4$ и $\omega_b = 3\pi/4$.

2.5. Частотная характеристика идеальной дифференцирующей цепи имеет вид $H(e^{i\omega}) = i\omega e^{-i\omega\tau}$, $-\pi < \omega < \pi$ (эта характеристика повторяется с периодом 2π). Величина τ представляет собой задержку в числе отсчетов.

а) Изобразить амплитудно-частотную и фазо-частотную характеристики системы.

б) Найти импульсную характеристику системы.

в) Импульсную характеристику данной системы можно ограничить по протяженности величиной N с помощью окна, такого, как в задаче 2.3. В этом случае задержка составит $\tau = (N-1)/2$, так что идеальная импульсная характеристика может быть ограничена симметрично [1]. Если $\tau = (N-1)/2$ и N — нечетное целое, показать, что идеальная импульсная характеристика убывает как $1/n$. Изобразить идеальную импульсную характеристику для случая $N = 11$.

г) Показать, что при четном N, $h(n)$ убывает как $1/n^2$. Изобразить импульсную характеристику для $N = 10$.

2.6. Частотная характеристика идеального преобразователя Гильберта (90-градусного фазовращателя) с задержкой τ есть

$$H(e^{i\omega}) = \begin{cases} -ie^{-i\omega\tau}, & 0 < \omega < \pi; \\ ie^{-i\omega\tau} & -\pi < \omega < 0. \end{cases}$$

Рассчитать и построить импульсную характеристику этой системы.

2.7. Рассмотрим КИХ-фильтр с линейной фазо-частотной характеристикой. Импульсная характеристика такого фильтра обладает следующим свойством:

$$h(n) = \begin{cases} h(N-1-n), & 0 \leqslant n \leqslant N-1; \\ 0, & \textit{в противном случае.} \end{cases}$$

а) Показать, что если N — четное целое, то дискретная свертка для сигнала на выходе может быть представлена как

$$y(n) = \sum_{k=0}^{(N-2)/2} h(k)\,[x(n-k) + x(n-N-1+k)],$$

и если N — нечетное, то

$$y(n) = \sum_{k=0}^{(N-3)/2} h(k)\,[x(n-k) + x(n-N+1+k)] + h((N-1)/2)\,x(n-(N-1)/2).$$

Таким образом, для вычисления каждого отсчета выходного сигнала необходимо выполнить лишь половину умножений по сравнению с общим случаем.

б) Изобразить структурную схему цифрового фильтра для каждого из приведенных выше соотношений.

2.8. Рассмотрим систему первого порядка

$$y(n) = \alpha y(n-1) + x(n).$$

а) Определить передаточную функцию $H(z)$ для этой системы.

б) Определить импульсную характеристику системы.

в) Определить, при каких α система устойчива?

г) Предположим, что сигнал на входе получен дискретизацией с периодом T. Определить такое значение α, чтобы $h(n) < e^{-1}$ для $nT < 2$ мс, т. е. так, чтобы постоянная времени составляла 2 мс.

2.9. а) Показать, что при $M < N$ $H(z)$ можно представить в виде разложения на простые дроби в соответствии с (2.30), где коэффициенты A_m определяются из соотношения $A_m = H(z)(1-d_m z^{-1})|_{z=d_m}$, $m = 1, 2, \ldots, N$.

б) Показать, что z-преобразование последовательности $A_k(d_k)^n u(n)$ имеет вид $A_k/(1-d_k z^{-1})$, $|z| > |d_k|$, и, таким образом, $h(n)$ задается соотношением (2.31).

2.10. Рассмотрим две инвариантные к сдвигу линейные системы в каскадной форме, как это показано на рис. 3.2.1, т. е. выход первой системы представляет собой вход второй.

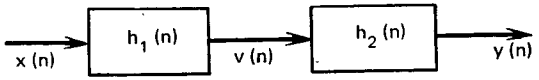


Рис. 3.2.1

а) Показать, что импульсная характеристика общей системы имеет вид $h(n) = h_1(n) * h_2(n)$.

б) Показать, что $h_1(n)h_2(n) = h_2(n)h_1(n)$ и, таким образом, общая импульсная характеристика не зависит от порядка включения каскадов.

в) Рассмотрим передаточную функцию (2.23), записанную в форме

$$H(z) = \left[\sum_{r=0}^{M} b_r z^{-r}\right]\left[\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right] = H_1(z)\,H_2(z),$$

т. е. в виде каскадного соединения двух систем. Записать разностное уравнение для системы в целом.

г) Рассмотрим две части системы в обратном порядке, т. е. $H(z) = H_2(z)H_1(z)$. Показать, что результирующая система описывается (2.33).

2.11. Для разностного уравнения $y(n) = 2\cos(bT)y(n-1) - y(n-2)$ найти начальные условия $y(-1)$ и $y(-2)$ так, чтобы

а) $y(n) = \cos(bT_n)$, $n \geqslant 0$;

б) $y(n) = \sin(bT_n)$, $n \geqslant 0$.

2.12. Рассмотрим систему разностных уравнений:

$$y_1(n) = Ay_1(n-1) + By_2(n-1) + x(n),$$
$$y_2(n) = cy_1(n-1) + Dy_2(n-1).$$

а) Изобразить структурную схему для этой системы.

б) Определить передаточные функции.

в) Для случая $A = D = r\cos\theta$ и $c = -B = r\sin\theta$ определить импульсную характеристику $h_1(n)$ и $h_2(n)$ и результирующую, если система возбуждается сигналом $x(n) = \delta(n)$.

2.13. Реализуемая инвариантная к сдвигу система имеет передаточную функцию вида

$$H(z) = \frac{(1 + 2z^{-1} + z^{-2})(1 + 2z^{-1} + z^{-2})}{\left(1 + \frac{7}{8}z^{-1} + \frac{5}{16}z^{-2}\right)\left(1 + \frac{3}{4}z^{-1} + \frac{7}{8}z^{-2}\right)}.$$

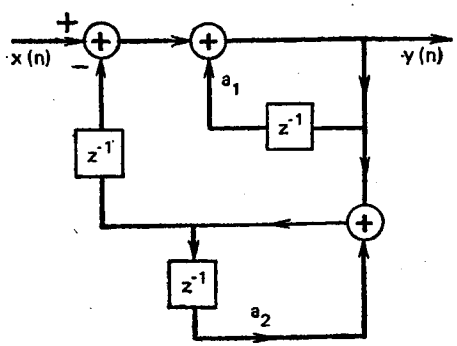


Рис. 3.2.2

а) Изобразить цифровую структурную схему, соответствующую данной системе в каскадной и прямой формах.

б) Определить, устойчива ли эта система? Привести пример.

2.14. Для системы на рис. 3.2.2:

а) Записать разностное уравнение, описывающее эту систему.

б) Определить передаточную функцию для данной системы.

2.15. Определить a_1, a_2 и a_3 через b_1 и b_2 так, чтобы обе системы на рис. 3.2.3 обладали одной и той же передаточной функцией.

2.16. Передаточная функция обычного резонатора имеет вид

$$H(z) = \frac{1 - 2e^{-aT}\cos(bT) + e^{-2aT}}{1 - 2e^{-aT}\cos(bT)\,z^{-1} + e^{-2aT}\,z^{-2}}.$$

а) Найти полюса и нули $H(z)$ и изобразить их на z-плоскости.

б) Найти импульсную характеристику для $T=10^{-4}$; $b=1000$: $a=200$ и изобразить ее на графике.

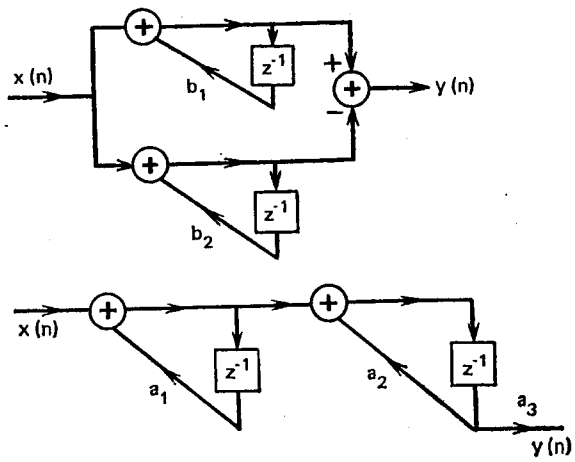


Рис. 3.2.3

2.17. Рассмотрим последовательность конечной длительности $x(n)=\delta(n)+0{,}5\delta(n-5)$.

а) Найти z-преобразование и преобразование Фурье для $x(n)$.

б) Найти N-точечное ДПФ для $x(n)$ при $N=50$; 10; 5.

в) Сравнить ДПФ при $N=5$ и 50.

г) Определить взаимосвязь между ДПФ и преобразованием Фурье.

2.18. Речевой сигнал дискретизирован с частотой 20 000 отсч./с. Для вычисления 1024-точечного ДПФ выделен сегмент, равный 1024 отсчетам.

а) Определить длительность сегмента.

б) Какое частотное разрешение достигается в ДПФ?

в) Как изменятся ответы на вопросы а) и б), если сегмент будет содержать 512 отсчетов (для вычисления ДПФ сегмент дополняется 512 нулями)?

3

Цифровые модели речевых сигналов

3.0. Введение

Для того чтобы научиться применять методы цифровой обработки сигналов в задачах связи, надо хорошо представлять основные положения как теории речеобразования, так и теории цифровой обработки сигналов. В данной главе приведен обзор положений акустической теории речеобразования и показано, как из нее вытекают различные способы представления речи. Особенно большое внимание уделяется моделям в дискретном времени, с помощью которых описывается дискретизированный речевой сигнал. Эти модели служат основой применения методов цифровой обработки.

Назначение данной главы близко к назначению гл. 2, в которой излагаются предварительные сведения об изучаемых вопросах. Более детальное изложение можно найти в [1—5]. Повышенного внимания заслуживают книги Фанта [1] и Фланагана [2]. В [1] подробно рассмотрены положения акустической теории речеобразования и содержится большое количество данных о системах измерения акустических характеристик речи и ее моделях. В книге Фланагана, охватывающей более широкий круг вопросов, дается богатое множество способов физического моделирования процесса речеобразования и указываются пути их применения для представления и обработки речевых сигналов. Эти книги могут быть рекомендованы читателю, желающему глубоко изучить состояние проблемы.

Перед изложением акустической теории речеобразования и ее математических положений полезно познакомиться с различными типами звуков, из которых состоит речь. Поэтому глава начинается с краткого введения в акустическую фонетику, в котором приводятся основные классы фонем английского языка, рассматриваются особенности их произнесения. Далее излагаются основы акустической теории речеобразования. Рассматриваются вопросы распространения звуковых волн в голосовом тракте, электрические аналоги голосового тракта, квазистационарное поведение артикуляторного аппарата при произнесении протяжных звуков речи. Теория позволяет представить речевой сигнал в виде отклика нестационарной линейной системы (голосового тракта), возбуждаемой либо шумом, либо квазипериодической последовательностью импульсов. Такое представление применяется для получения моделей речевого сигнала в дискретном времени. Эти модели, разработанные на основе положений акустической теории, формируются с позиций теории цифровой обработки сигналов и используются далее при изложении основного предмета книги — методов цифровой обработки речевых сигналов.

3.1. Процесс образования речи

Речь состоит из последовательности звуков. Звуки и переходы между ними служат символическим представлением информации. Порядок следования звуков (символов) определяется правилами языка. Изучение этих правил и их роли в общении между людьми составляет предмет *лингвистики*, анализ и классификация самих звуков речи — предмет *фонетики*. В подробном изучении фонетики и лингвистики здесь нет необходимости. Однако при обработке речевых сигналов с целью повышения их информативного содержания либо для выделения содержащейся в сигнале информации полезно располагать как можно большим количеством сведений о структуре сигнала, например о способе кодирования информации в сигнале. Таким образом, прежде чем подробно рассматривать математические модели речеобразования, уместно обсудить основные группы звуков речи. Этим упоминанием о фонетике и лингвистике мы и ограничимся. Укажем, однако, что эта краткость не умаляет важности этих наук, особенно в области распознавания и синтеза речи.

3.1.1. Механизм речеобразования

На рентгеновском снимке (рис. 3.1) показаны наиболее важные органы речеобразующей системы человека [6]. *Голосовой тракт*, который на рисунке обведен пунктиром, начинается с прохода между голосовыми связками, называемого *голосовой щелью*, и заканчивается у губ. Голосовой тракт, таким образом, состоит из *гортани* (от пищевода до рта) и рта, или *ротовой полости*. У взрослого мужчины общая длина голосового тракта составляет примерно 17 см. Площадь поперечного сечения голосового тракта, которая определяется положением языка, губ, челюстей и небной занавески, может изменяться от нуля (тракт полностью перекрыт) до примерно 20 см^2. *Носовая полость* начинается у небной занавески и заканчивается ноздрями. При опущенной небной занавеске носовая полость акустически соединена с голосовым трактом и участвует в образовании носовых звуков речи. При изучении процесса речеобразования полезно изображать основные органы физической системы в таком виде, при котором становится ясной математическая сторона вопроса. На рис. 3.2 показано подробное схематическое изображение речеобразующей системы. Для полноты в диаграмму включены и такие органы, как легкие, бронхи и трахея, расположенные ниже гортани. Совокупность этих органов служит источником энергии для образования речи. Речь представляет собой акустическую волну, которая вначале излучается этой системой при выталкивании воздуха из легких и затем преобразуется в голосовом тракте. В качестве примера на рис. 3.3*а* показано речевое колебание, соответствующее фразе «Should we cha (se)», произнесенной мужским голосом. Основные особенности колебания легко объяснить на основе подробного анализа механизма образования речи.

Звуки речи могут быть разделены на три четко выраженные группы по типу возбуждения. *Вокализованные* звуки образуются

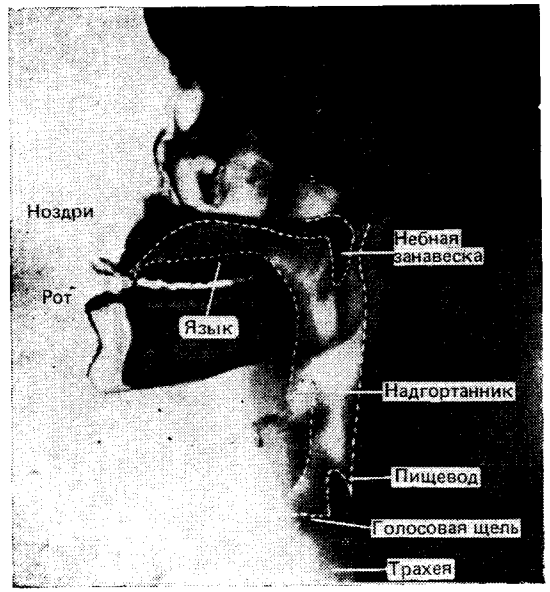


Рис. 3.1. Рентгеновский снимок речеобразующих органов человека [6]

проталкиванием воздуха через голосовую щель, при котором периодически напрягаются и расслабляются голосовые связки и возникает квазипериодическая последовательность импульсов потока воздуха, возбуждающая голосовой тракт. Вокализованные сегменты обозначены на рис. 3.3*а* знаками $|U|$, $|d|$, $|w|$, $|i|$, $|e|$. *Фрикативные* или *невокализованные* звуки генерируются при сужении

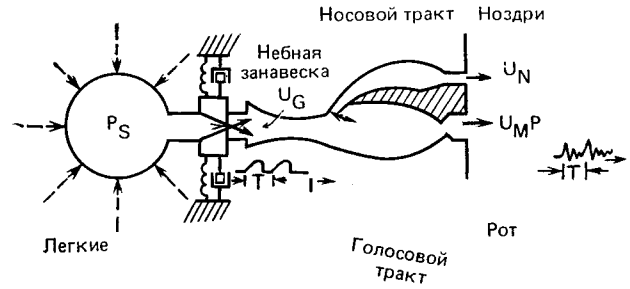


Рис. 3.2. Схематическое изображение речеобразующих органов человека [6]

голосового тракта в каком-либо месте (обычно в конце рта) и проталкивании воздуха через суженное место со скоростью, достаточно высокой для образования турбулентного воздушного потока. Таким образом, формируется источник широкополосного шу-

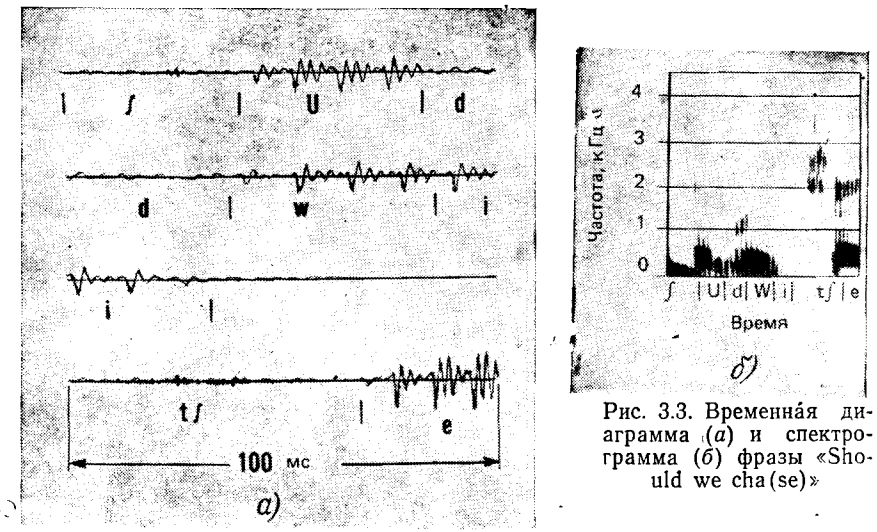


Рис. 3.3. Временна́я диаграмма (*а*) и спектрограмма (*б*) фразы «Should we cha(se)»

ма, возбуждающего голосовой тракт. На рис. 3.3*а* сегмент, обозначенный знаком $|\int|$, соответствует фрикативному звуку «*sh*». При произнесении *взрывных звуков* голосовой тракт полностью закрывается (обычно в начале голосового тракта). За этой смычкой возникает повышенное сжатие воздуха. Затем воздух внезапно высвобождается. Такое явление имеет место при произнесении звука, обозначенного на рис. 3.3*а* символом $|t\int|$. Область малого уровня в конце третьей линии, которая предшествует шумоподобному участку колебания, соответствует периоду полного закрытия голосового тракта. Голосовой тракт и носовая полость показаны на рис. 3.2 в виде труб с переменной по продольной оси площадью поперечного сечения. При прохождении звуковых волн через эти трубы их частотный спектр изменяется в соответствии с частотной избирательностью трубы. Этот эффект похож на резонансные явления, происходящие в трубах органов и духовых музыкальных инструментов. При описании речеобразования резонансные частоты трубы голосового тракта называют *формантными частотами* или просто *формантами*. Формантные частоты зависят от конфигурации и размеров голосового тракта: произвольная форма тракта может быть описана набором формантных частот. Различные звуки образуются путем изменения формы голосового тракта. Таким образом, спектральные свойства речевого сигнала

изменяются во времени в соответствии с изменением формы голосового тракта.

Переменные во времени спектральные характеристики речевого сигнала с помощью звукового спектрографа могут быть высвечены в виде графика. Этот прибор позволяет получить двумерный график, называемый *спектрограммой*, на которой по вертикальной оси отложена частота, а по горизонтальной — время. Плотность зачернения графика пропорциональна энергии сигнала. Таким образом, резонансные частоты голосового тракта имеют вид затемненных областей на спектрограмме. Вокализованным областям сигнала соответствует появление четко выраженной периодичности временно́й зависимости, в то время как невокализованные интервалы выглядят почти сплошными. Спектрограмма фразы рис. 3.3*а* показана на рис. 3.3*б*. На спектрограмме отдельные участки помечены теми же символами, что и на рис. 3.3*а*, так что особенности сигнала во временно́й и частотной областях могут быть сопоставлены.

Звуковой спектрограф весьма долго служил основным инструментом исследования речевого сигнала, и хотя в настоящее время с помощью цифровой обработки можно получить более гибкие устройства визуального изображения (см. гл. 6), основные принципы спектрографа используются широко и в настоящее время. Хорошим пособием по спектрографическому представлению речи, не утратившим своего значения и до сих пор, является книга [8]. Хотя книга написана для обучения «чтению» спектрограмм, она является замечательным введением в акустическую фонетику.

3.1.2. Акустическая фонетика

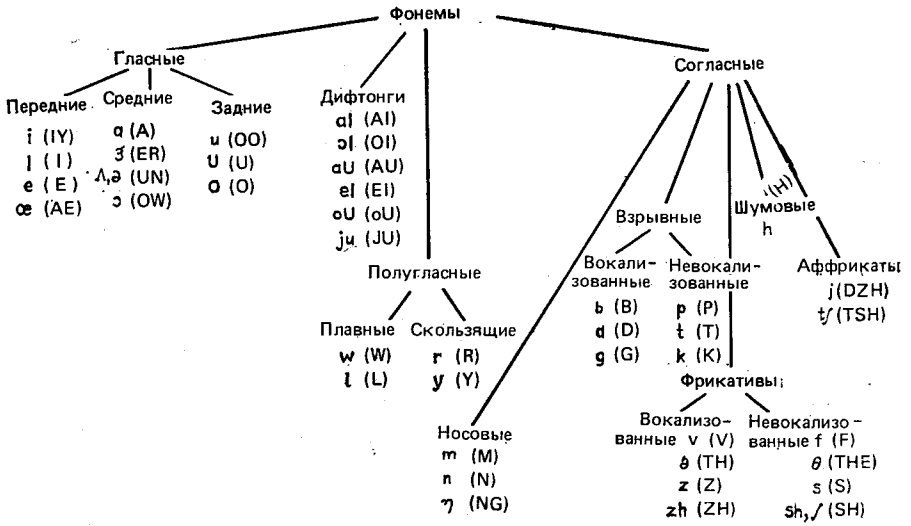
Многие языки, в том числе и английский, можно описать набором отдельных звуков или *фонем*. В частности, в американском произношении английского языка существует 42 фонемы, которые подразделяются на гласные, дифтонги, полугласные и согласные. Изучать фонему можно по-разному. Лингвисты, например, изучают отличительные характеристики фонем [9, 10]. Нам достаточно рассмотреть акустические свойства различных звуков, в том числе место и способ артикуляции, форму акустического колебания, характеристики спектрограмм. В табл. 3.1 приведены различные классы фонем английского языка в его американском произношении[1]. Четыре широких класса звуков образуют гласные, дифтонги, полугласные и согласные. Каждый из классов разбит на подклассы по способу и месту образования звука в голосовом тракте. Каждая фонема табл. 3.1 может быть отнесена к классу протяжных или кратковременных звуков. Протяжные звуки образуются при фиксированной (инвариантной ко времени) форме голосового тракта, который возбуждается соответствующим источником. К этому классу относятся гласные, фрикативные (вокализованные и невокализованные) носовые согласные. Остальные звуки (дифтонги, полугласные, аффрикаты и взрывные согласные) произносятся при изменяющейся форме голосового тракта. Они образуют класс кратковременных звуков[2].

---

[1] В таблице указаны как фонетическое, так и орфографическое представления фонем, которые используются далее во всей книге. (*Прим. ред.*)

[2] Имеется в виду, что при их произнесении решающую роль играет кратковременная динамика артикуляционных движений. (*Прим. ред.*)

Таблица 3.1



Гласные. Гласные образуются при квазипериодическом возбуждении голосового тракта неизменной формы импульсами воздуха, возникающими вследствие колебания голосовых связок. Как будет показано ниже, зависимость площади поперечного сечения голосового тракта от координаты (расстояния) вдоль его продольной оси определяет резонансные частоты тракта (форманты) и характер произносимого звука. Эта зависимость называется *функцией площади поперечного сечения*. Функция площади поперечного сечения для каждой гласной зависит в первую очередь от положения языка; вместе с тем на характер звука оказывают влияние положения челюстей, губ и, в меньшей степени, небной занавески. Например, при произнесении звука $|a|$, как в слове «father», голосовой тракт открыт в начале, а в его конце тело языка образует сужение. Наоборот, при произнесении звука $|i|$, как в слове «eve», язык образует сужение в начале голосового тракта и оставляет его открытым в конце. Таким образом, каждому гласному звуку может быть поставлена в соответствие форма голосового тракта (функция площади поперечного сечения), характерная для его произношения. Очевидно, что это соответствие неоднозначное, так как у разных дикторов голосовые тракты различны. Другим представлением гласного звука является его описание с помощью набора резонансных частот голосового тракта. Это описание также зависит от диктора. Петерсон и Барней [11] провели измерения формантных (резонансных) частот с помощью звукового спектрографа для гласных, произнесенных различными дикторами. Эти результаты приведены на рис. 3.4, где показан график зависимости частоты второй форманты от частоты первой форманты для некоторых гласных, произнесенных взрослыми дик-

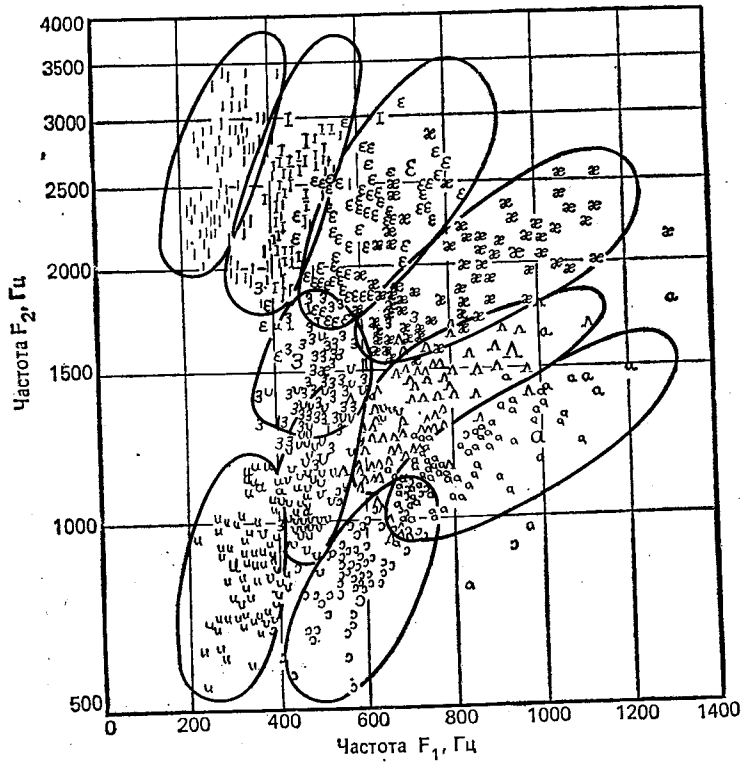


Рис. 3.4. Зависимость частоты второй форманты от частоты первой форманты для гласных, произнесенных разными дикторами [11]

торами и детьми. Эллипсы на рис. 3.4 ограничивают область изменения формантных частот для каждой из гласных. В табл. 3.2 приведены средние значения первых трех формантных частот для гласных, произнесенных мужскими голосами. Хотя существует

Таблица 3.2

Средние значения формантных частот для гласных [11]

Письменный символ	Транскрипция	Типичное слово	F_1	F_2	F_3
IY	*i*	heet	270	2290	3010
I	*ɪ*	bit	390	1990	2550
E	*ɛ*	bet	530	1840	2480
AE	*æ*	bat	660	1720	2410
UH	*ʌ*	but	520	1190	2390
A	*a*	hot	730	1090	2440
OW	*ɔ*	bought	570	840	2410
U	*U*	foot	440	1020	2240
OO	*u*	boot	300	870	2240

большой разброс формантных частот, данные табл. 3.2 являются полезной характеристикой гласных. На рис. 3.5 приведен график зависимости частоты второй форманты от частоты первой форманты для гласных табл. 3.2. В верхнем левом углу так называемого треугольника гласных расположена гласная $|i|$ с низкой частотой

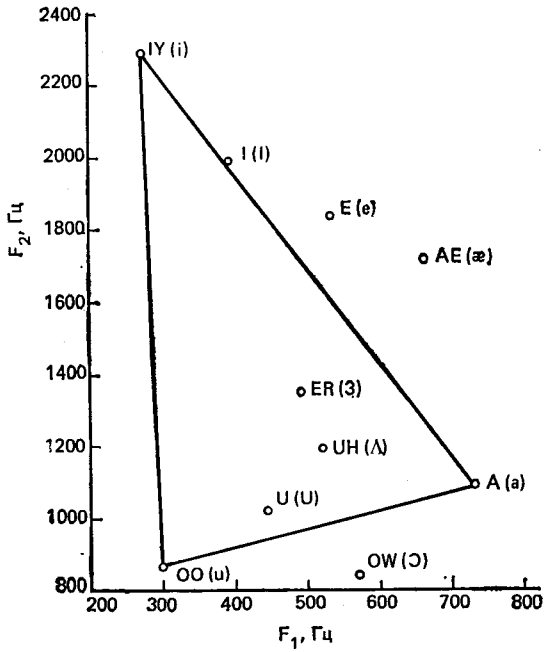


Рис. 3.5. Треугольник гласных

первой форманты и высокой частотой второй форманты. В нижнем левом углу расположена гласная $|u|$ с низкими частотами первой и второй формант. В третьей вершине треугольника находится гласная $|a|$ с высокой частотой первой форманты и низкой частотой второй форманты. Далее будет показано, как форма голосового тракта влияет на частоты формант гласных.

На рис. 3.6 приведены акустические колебания и спектрограммы для всех гласных английского языка. На спектрограммах четко выделяются различные резонансные области, характерные для каждой гласной. Акустические колебания, иллюстрируя периодичность вокализованных звуков, позволяют также путем анализа одного периода выявить грубые спектральные характеристики. Например, акустическое колебание звука $|i|$ состоит из низкочастотного затухающего колебания, на которое накладывается относительно высокочастотная составляющая. Это соответствует низкой частоте первой форманты и высоким частотам второй и третьей формант (см. табл. 3.2). Два резонанса, расположенных на близких частотах, расширяют спектр колебания. Наоборот, в акусти-

ческом колебании гласной $|u|$ энергия высокочастотных составляющих относительно мала, что соответствует низким частотам первой и второй формант. Подобный анализ может быть проведен для всех гласных, акустические колебания которых приведены на рис. 3.6.

Дифтонги. Дифтонгом называется участок речи, соответствующий одному слогу, который начинается с одной гласной и затем постепенно переходит в другую. На основе этого определения в американском произношении можно выделить шесть дифтонгов: $|eI|$ (как в слове «bay»), $|oU|$ (как в слове «boat»), $|aU|$ (как в слове «how»), $|oI|$ (как в слове «boy»), $|aI|$ (как в слове «buy») и $|ju|$ (как в слове «you»).

Дифтонги образуются путем плавного изменения формы голосового тракта. Для иллюстрации этого положения на рис. 3.7 показана временна́я зависимость частоты второй форманты от частоты первой форманты для дифтонгов [12]. Стрелками показаны направления изменения формантных частот во времени. Пунктирными линиями обведены средние значения формант для гласных. Дифтонги можно описать изменением во времени функции площади поперечного сечения голосового тракта от значения, соответствующего первой гласной, до значения, соответствующего второй гласной дифтонга.

Полугласные. Группу звуков, содержащих $|w|$, $|l|$, $|r|$ и $|y|$, описать довольно трудно. Эти звуки называются полугласными, так как по своим свойствам они напоминают гласные звуки. Обычно их характеризуют плавным изменением функции площади поперечного сечения голосового тракта между смежными фонемами. Таким образом, акустические характеристики этих звуков существенно зависят от произносимого текста. Нам удобно рассматривать эти звуки как переходные, сходные с гласными. Их структура близка к структуре гласных и дифтонгов. Пример сигнала, соответствующего полугласному звуку $|w|$, показан на рис. 3.3.

Носовые звуки. Носовые согласные $|m|$, $|n|$ и $|\eta|$ образуются при голосовом возбуждении. В полости рта при этом возникает полная смычка. Небная занавеска опущена, поэтому поток воздуха проходит через носовую полость и излучается через ноздри. Полость рта, которая вначале закрыта, акустически соединена с гортанью. Таким образом, рот служит резонансной полостью, в которой задерживается часть энергии при определенных частотах воздушного потока. Эти резонансные частоты соответствуют антирезонансам или нулям передаточной функции тракта речеобразования [2]. Более того, для носовых согласных и гласных (т. е. гласных, расположенных перед носовыми согласными) характерны менее выраженные резонансы, чем для гласных. Расширение резонансных областей происходит из-за того, что внутренняя поверхность носового тракта напрягается и при этом носовая полость имеет большое отношение площади поверхности к площади поперечного сечения. Вследствие этого потери за счет теплопроводности и вязкости оказываются бо́льшими, чем обычно.

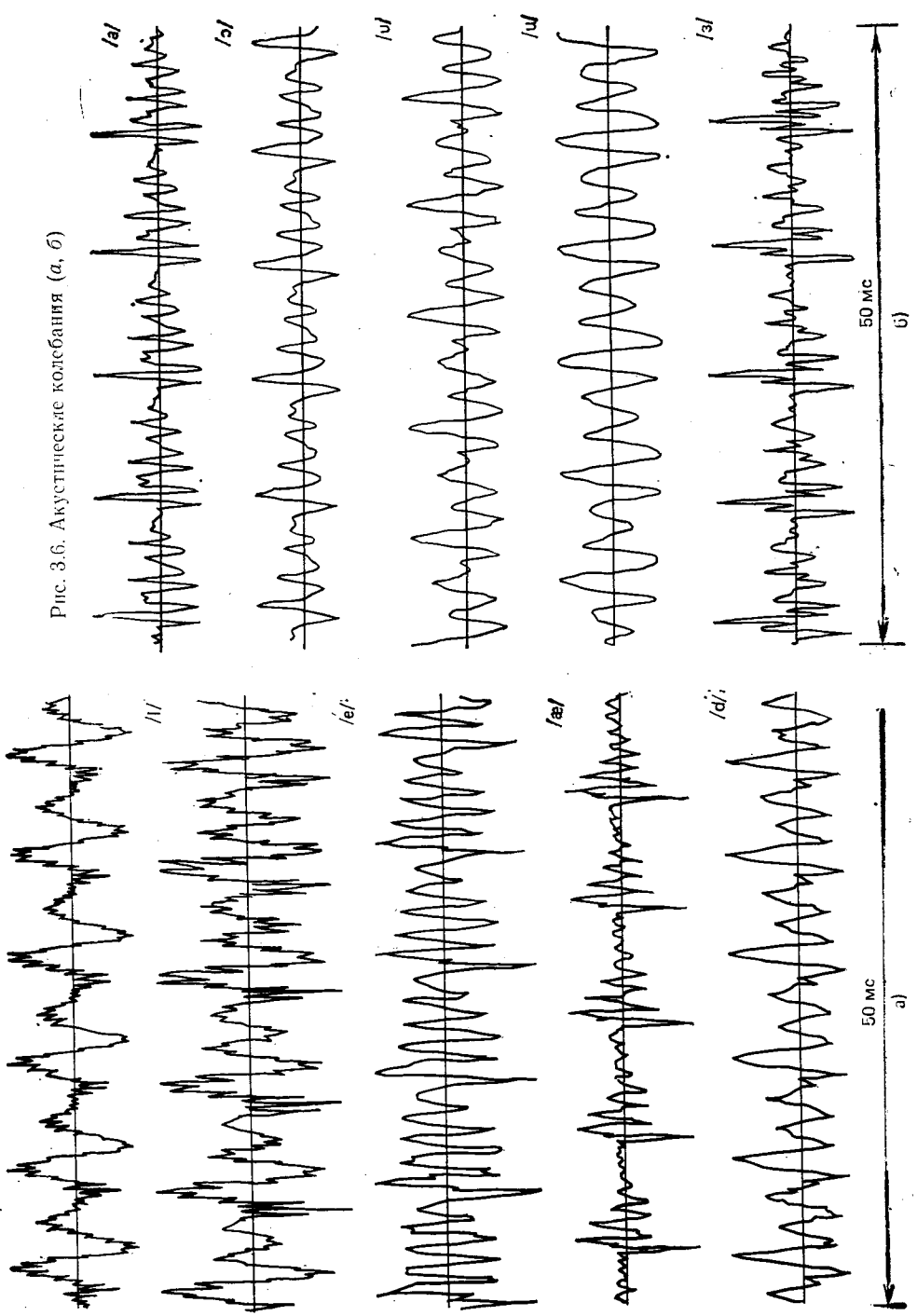


Рис. 3.6. Акустические колебания (*а*, *б*)

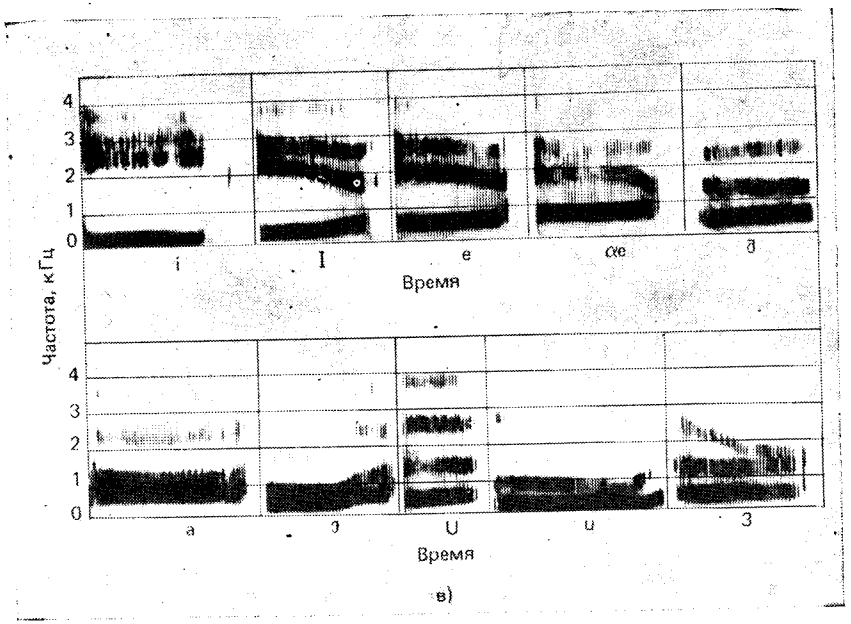


Рис. 3.6.*в*. Спектрограммы в некоторых гласных английского языка (американское произношение)

Три носовых согласных различаются местом расположения полной смычки. При произнесении звука |*m*| смычка образуется между губами, |*n*| — у внутренней стороны зубов и |η| — у небной занавески. На рис. 3.8 приведены типичные колебания и спектрограммы для двух носовых согласных в сочетании гласный — носовой согласный — гласный. Из рисунка видно, что временные колебания согласных |*m*| и |*n*| очень похожи. Спектрограммы иллюстрируют подъем спектра на низких частотах и отсутствие четко выраженных резонансов в диапа-

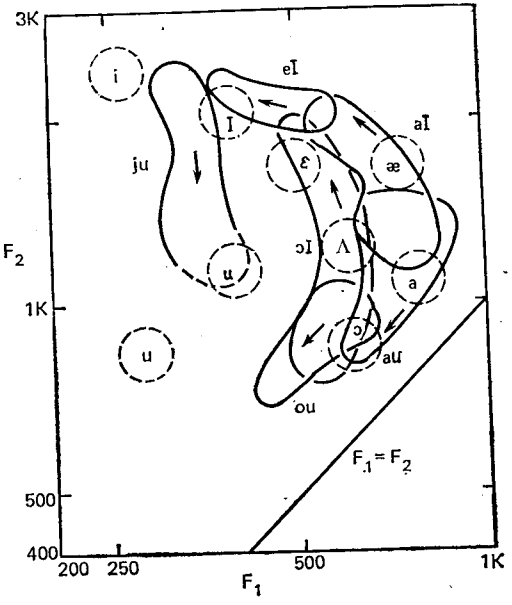
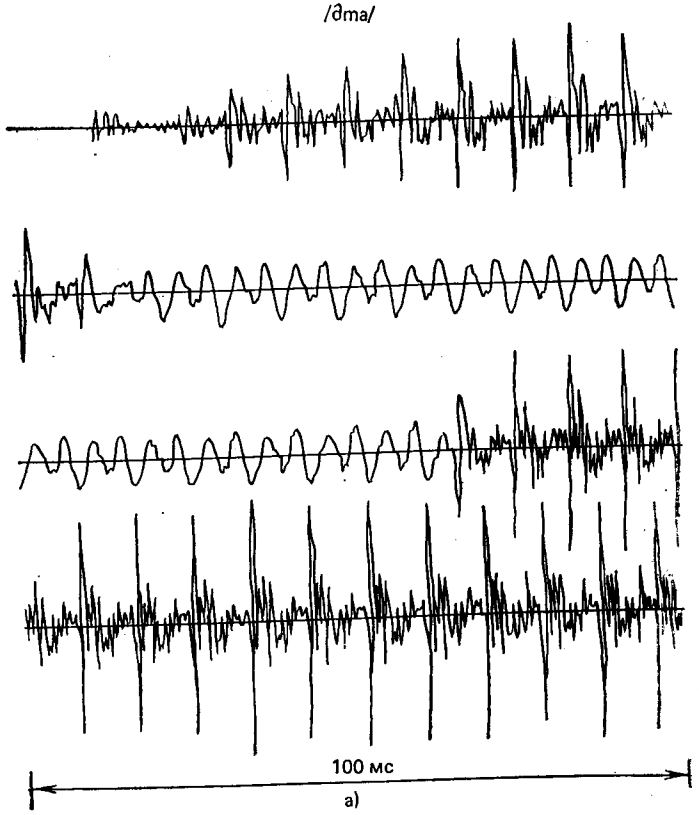
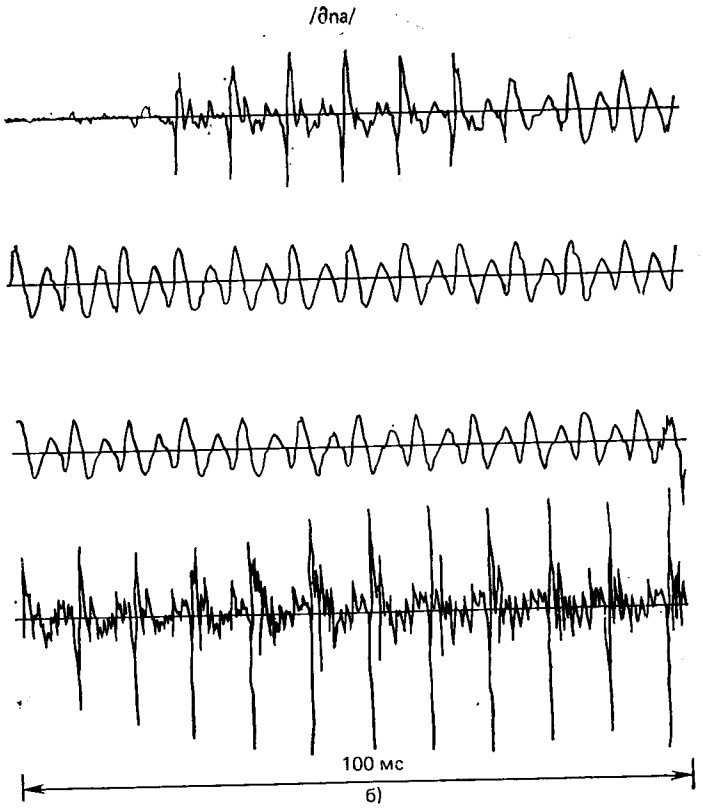


Рис. 3.7. Временные изменения двух первых формант дифтонгов [27]

зоне средних частот. Это происходит вследствие взаимного влияния резонансов и антирезонансов, образующихся за счет взаимодействия полостей носа и рта [13].

Глухие фрикативные звуки. Глухие фрикативные звуки |*f*|, |θ|, |*s*| и |*sh*| образуются путем возбуждения голосового тракта турбулентным воздушным потоком, возникающим в области смычки голосового тракта. Расположение смычки характеризует тип фрикативного звука. При произнесении звука |*f*| смычка возникает около губ, |θ| — около зубов, |*s*| — в середине полости рта и |*sh*| — в конце полости рта. Таким образом, система образования глухих фрикативных звуков содержит источник шума, расположенный в области смычки, которая разделяет голосовой тракт на две полости. Звуковая волна излучается через губы, т. е. через переднюю полость. Другая полость служит, как и в случае произнесения носовых звуков, для задерживания акустического потока, и таким образом в речеобразующем тракте возникают антирезонансы [2, 14]. На рис. 3.9 приведены колебания и спектрограммы фрикативных звуков |*f*|, |*s*| и |*sh*|. Непериодическая структура возбуждения отчетливо видна на временных диаграммах колеба-





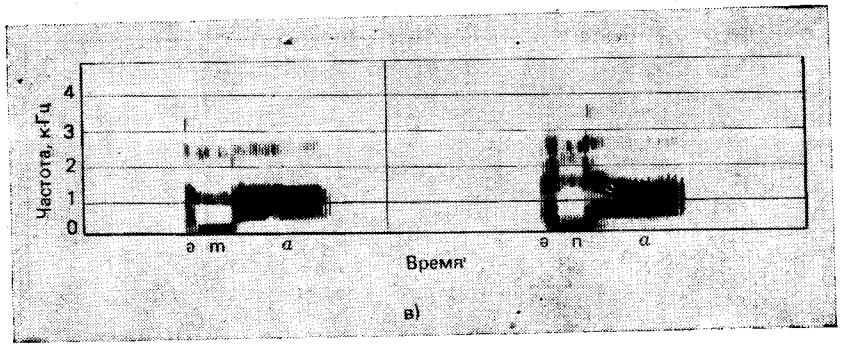


Рис. 3.8. Акустические колебания (*а*, *б*) и спектрограммы (*в*) для сочетаний |*UH—M—A*| и |*UH—N—A*|

ний. Спектральные отличия фрикативных звуков легко определяются по спектрограммам.

Звонкие фрикативные звуки. Звонкие фрикативные звуки |*v*|, |*th*|, |*z*| и |*zh*| являются прототипами глухих звуков |*f*|, |θ|,

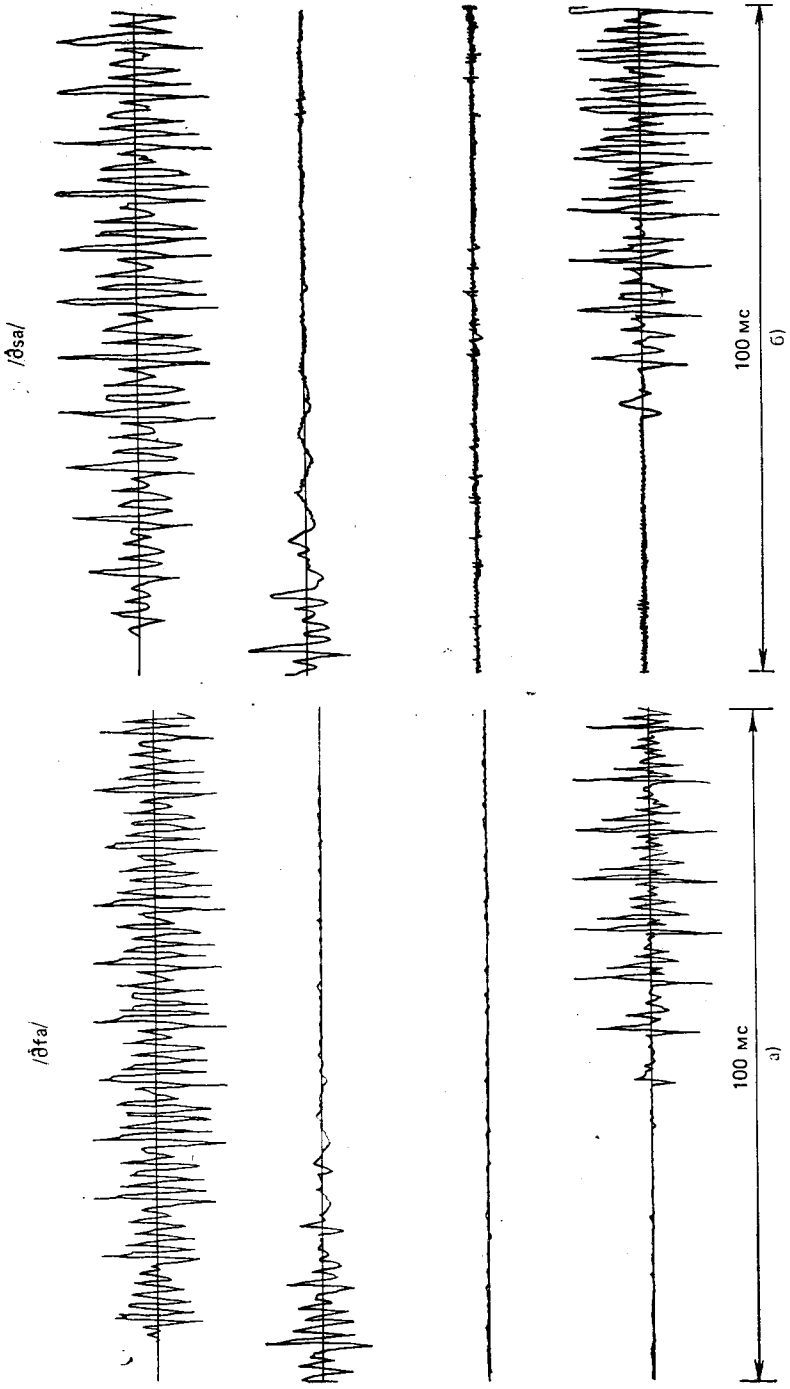


Рис. 3.9. Акустические колебания (*а*, *б*)

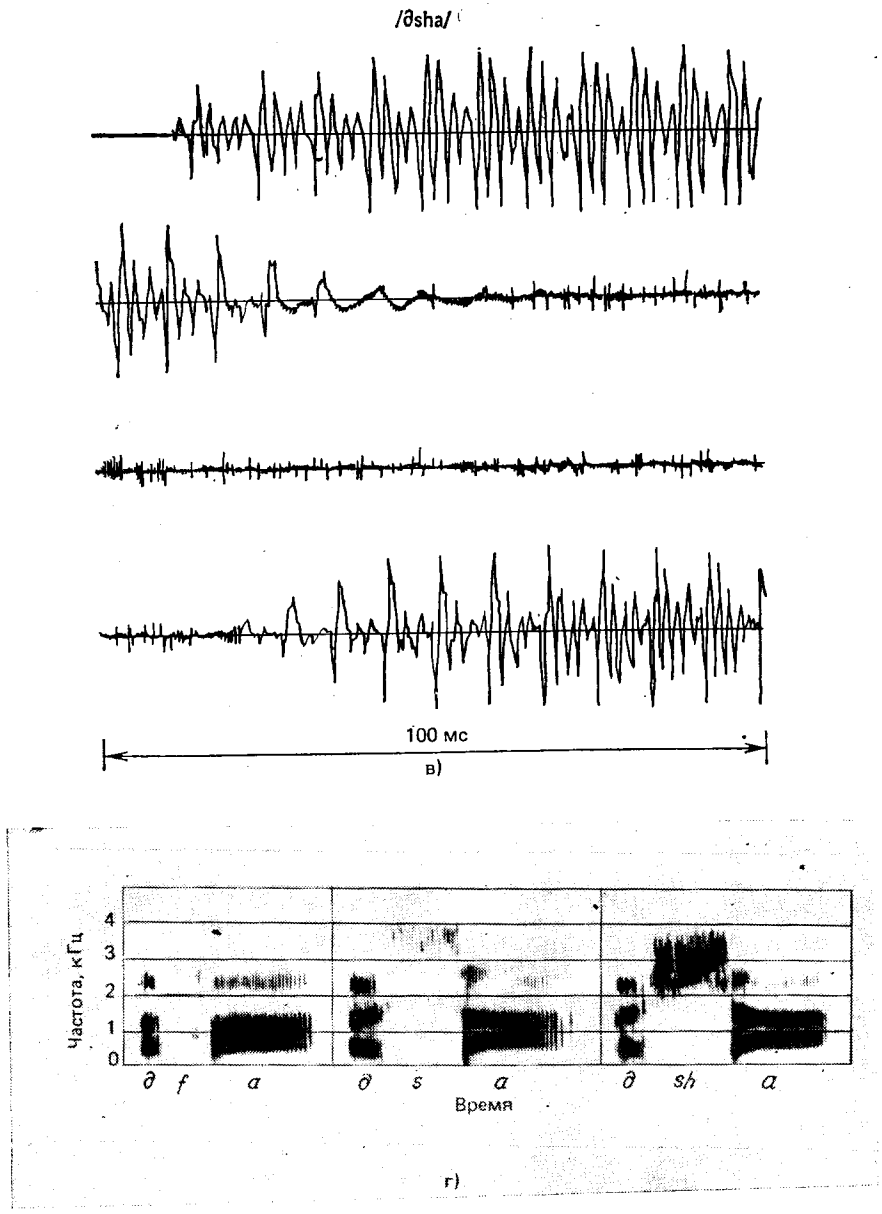
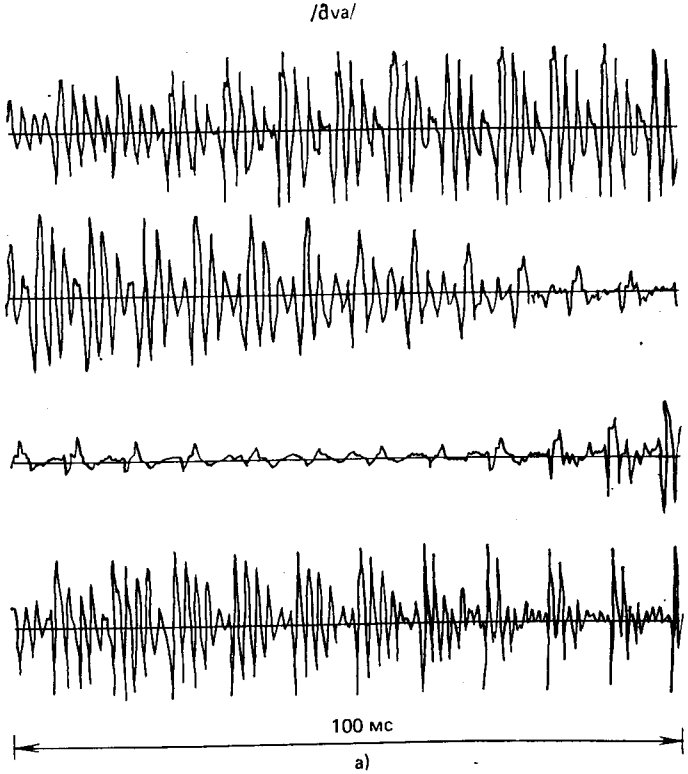


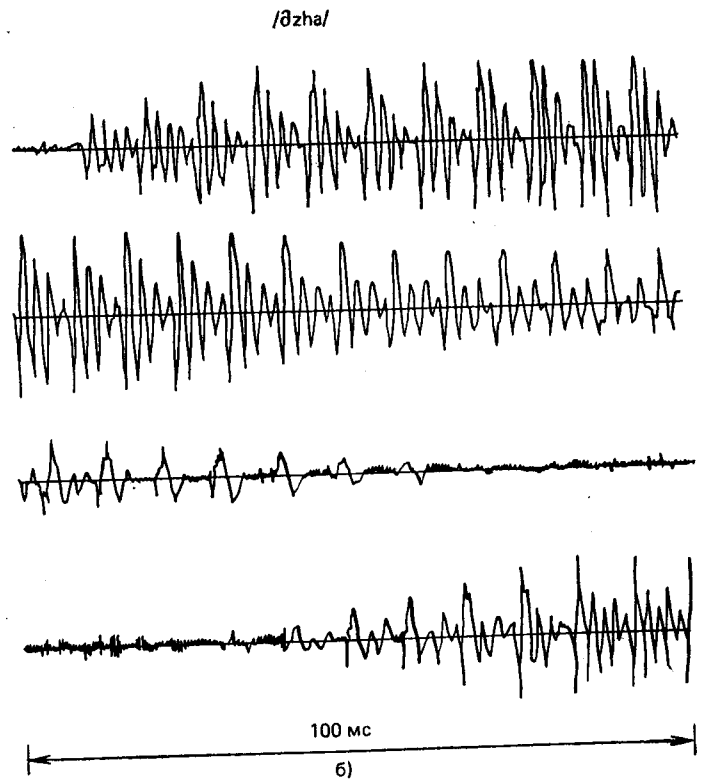
Рис. 3.9. Акустические колебания (*в*) и спектрограммы (*г*) для |*UH—F—A*| и |*UH—SH—A*|

|*s*| и |*sh*| соответственно. Место расположения смычки для этих пар звуков совпадает. Однако звонкие фрикативные отличаются от своих глухих аналогов тем, что при их образовании участвуют

два источника возбуждения. При образовании звонких звуков голосовые связки колеблются и, таким образом, один источник возбуждения находится в гортани. Однако, так как в голосовом тракте образуется смычка, поток воздуха в этой области становится турбулентным. Можно ожидать, что в спектре звонких фрикативных звуков будут две различные составляющие. Эти особенности возбуждения отчетливо видны на рис. 3.10, на котором приведены типичные колебания и спектры для нескольких звонких фрикативных звуков. Сходство структуры звонкого |*v*| и глухого |*f*| также легко установить путем сравнения соответствующих спектрограмм (рис. 3.9 и 3.10). Аналогично можно сравнить и спектрограммы звуков |*sh*| и |*zh*|.

Звонкие взрывные согласные. Звонкие взрывные согласные |*b*|, |*d*| и |*g*| являются переходными непротяжными звуками. При их образовании голосовой тракт смыкается в какой-нибудь области полости рта. За смычкой воздух сжимается и затем внезапно высвобождается. При произнесении звука |*b*| смычка образуется между губами, |*d*| — с внутренней стороны зубов, |*g*| — вблизи небной занавески. В течение периода, когда голосовой тракт полностью закрыт, звуковые волны практически не излучаются через губы. Однако слабые низкочастотные колебания излучаются стенками горла (эту область иногда называют голосовым затвором — «voice bar»). Колебания возникают из-за того, что голосовые связки могут вибрировать даже тогда, когда голосовой тракт перекрыт.





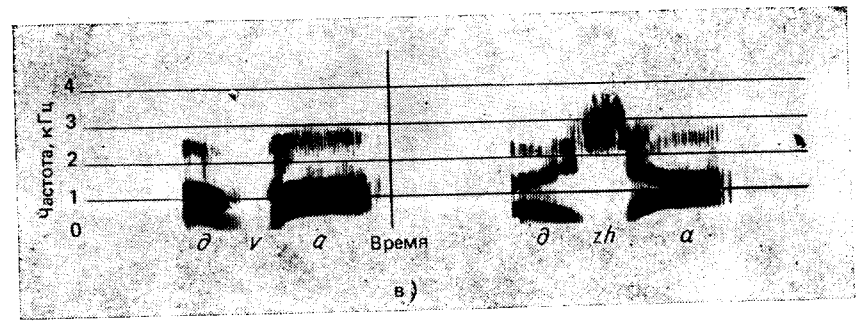


Рис. 3.10. Акустические колебания (*а*, *б*) и спектрограммы (*в*) для |*UH—V—A*| и |*UH—ZH—A*|

Так как структура взрывных звуков изменчива, их свойства существенно зависят от последующего гласного [15]. В этой связи

характер временны́х колебаний несет мало сведений о свойствах этих согласных. На рис. 3.11 показаны временна́я диаграмма колебания и спектрограмма сочетания |*UH—B—A*|. Из временно́й диаграммы, соответствующей звуку |*b*|, видно лишь, что при его произнесении имеет место голосовое возбуждение и в сигнале отсутствуют высокочастотные составляющие.

Глухие взрывные согласные. Глухие взрывные согласные |*p*|, |*t*| и |*k*| подобны своим звонким прототипам |*b*|, |*d*| и |*g*|, но

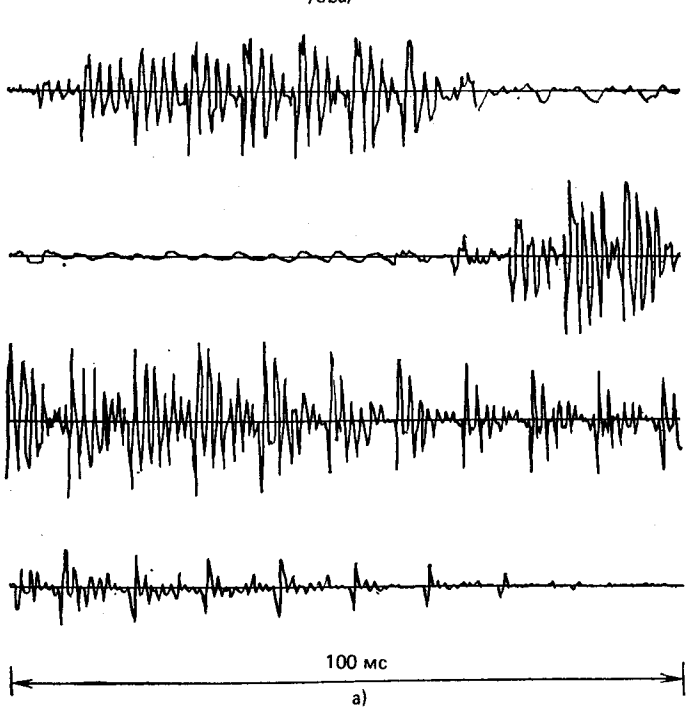


Рис. 3.11. Акустическое колебание (*а*) и спектрограмма (*б*) для |*UH—B—A*|

имеют одно важное отличие. В течение периода полного смыкания голосового тракта голосовые связки не колеблются. После этого периода, когда воздух за смычкой высвобождается, в течение короткого промежутка времени потери на трение возрастают из-за внезапной турбулентности потока воздуха. Далее следует период придыхания (шумовой воздушный поток из голосовой щели возбуждает голосовой тракт). После этого возникает голосовое возбуждение.

На рис. 3.12 приведены временны́е колебания и спектрограммы согласных |*p*| и |*t*|. На рисунке отчетливо виден период смычки (интервал, в течение которого сжимается воздух за смычкой). Видно также, что длительность и частотный состав шума в периоды повышенного трения и придыхания существенно зависят от типа взрывной согласной.

Аффрикаты и звук |*h*|. Остальными согласными американского произношения являются аффрикаты |*t*∫| и |*j*| и фонема |*h*|. Глухая аффриката |*t*∫| является динамичным звуком, который можно представить как сочетание взрывного |*t*| и фрикативного согласного |∫| (см. рис. 3.3*а*). Звонкий звук |*j*| можно представить как сочетание взрывного |*d*| и фрикативного звука |*zh*|. Наконец, фонема |*h*| образуется путем возбуждения голосового тракта турбулентным воздушным потоком, т. е. без участия голосовых связок, но при возникновении шумового потока в голосовой щели[1]. Структура звука |*h*| не зависит от следующей за ним гласной. Поэтому голосовой тракт может перестраиваться для произнесения следующей гласной в процессе произнесения звука |*h*|.

3.2. Акустическая теория речеобразования

В предыдущем параграфе дано качественное описание звуков речи и способов их образования. В настоящем параграфе изучим математическое описание речеобразования, которое служит основой анализа и синтеза речи.

3.2.1. Распространение звуков

Понятие звука почти совпадает с поня́тием колебаний. Звуковые волны возникают за счет колебаний. Они распространяются в воздухе или другой среде с помощью колебаний частиц этой среды. Следовательно, образование и распространение звуков в голосовом тракте подчиняется законам физики. В частности, основные законы сохранения массы, сохранения энергии, сохранения количества движения вместе с законами термодинамики и механики жидкостей применимы к сжимаемому воздушному потоку с низкой вязкостью, который является средой распространения звуков речи. Используя эти основные физические законы, можно соста-

[1] Этот способ возбуждения характерен и для шепота.

вить систему дифференциальных уравнений в частных производных, описывающую движение воздуха в речеобразующей системе [16—20]. Составление и решение этих уравнений весьма затруднительны даже для простых предположений относительно формы голосового тракта и потерь энергии в речеобразующей системе. Полная акустическая теория должна учитывать следующие факторы:

изменение во времени формы голосового тракта;

потери энергии на стенках голосового тракта за счет вязкого трения и теплопроводности;

мягкость стенок голосового тракта;

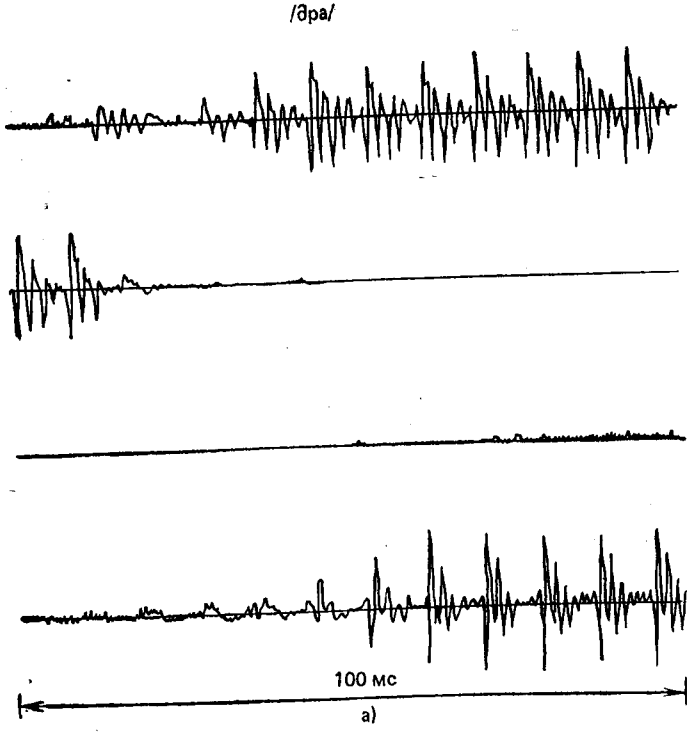
излучение звуковых волн через губы;

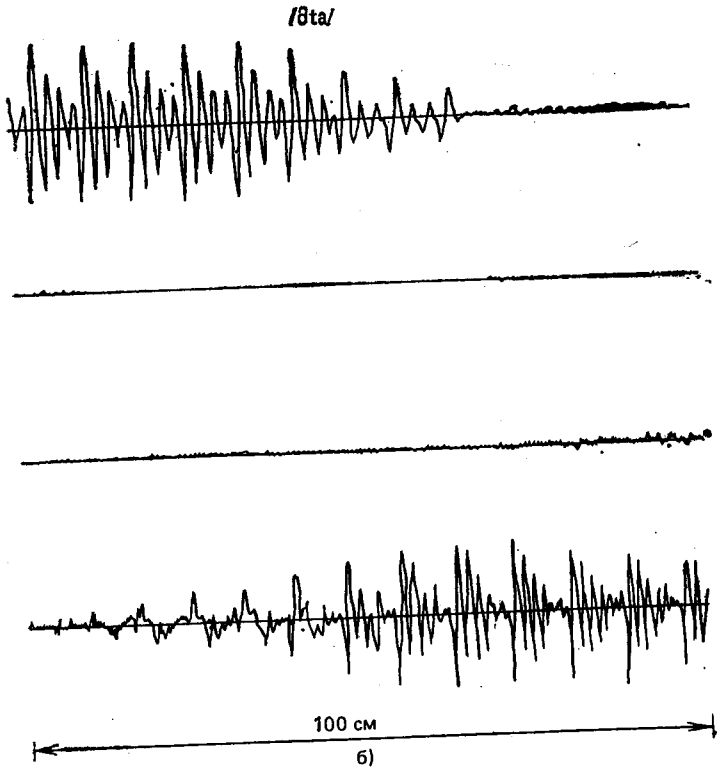
влияние носовой полости;

возбуждение голосового тракта.

Построение акустической теории, охватывающей все эти факторы, выходит за рамки этой главы и, кроме того, создание такой теории пока еще невозможно. Дадим обзор этих явлений и ссылки на соответствующую литературу. Если подходящей литературы по какому-либо вопросу не имеется, придется ограничиться качественным описанием.

Простейшая физическая интерпретация системы речеобразования показана на рис. 3.13*а*. Голосовой тракт здесь представлен в виде неоднородной трубы с переменной во времени площадью по-





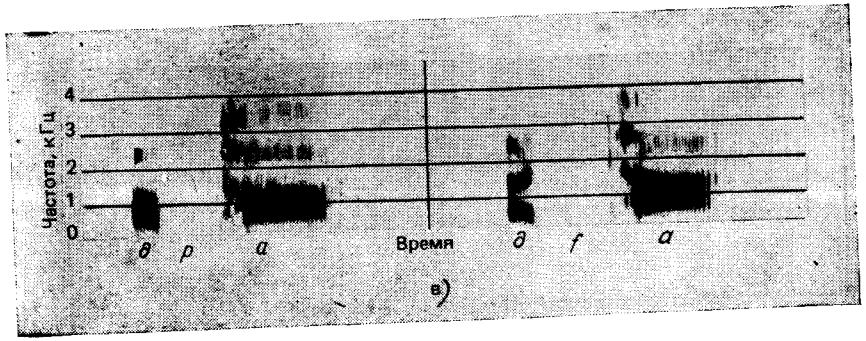


Рис. 3.12. Акустические колебания (*а*, *б*) и спектрограммы (*в*) для |*UH—P—A*| и |*UH—T—A*|

перечного сечения. Для колебаний, длина волны которых превышает размеры голосового тракта (это обычно имеет место на частотах ниже 4000 Гц), можно допустить, что вдоль продольной оси трубы распространяется плоская волна. Дальнейшее упрощение состоит в предположении отсутствия потерь на вязкость и тепло-

проводность как внутри воздушного потока, так и на стенках трубы. На основе законов сохранения массы, количества движения и энергии с учетом перечисленных допущений Портнов [18] показал, что звуковые волны в трубе удовлетворяют следующим уравнениям:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial (u/A)}{\partial t}; \tag{3.1а}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (pA)}{\partial t} + \frac{\partial A}{\partial t}, \tag{3.1б}$$

где $p = p(x, t)$ — звуковое давление как функция x и t; $u = u(x, t)$ — скорость воздушного потока (volume velocity) как функция x и t; ρ — плотность воздуха в трубе; c — скорость распространения звука; $A = A(x, t)$ — «функция площади», т. е. площадь поперечного сечения в направлении, перпендикулярном продольной оси трубы, как функция расстояния вдоль этой оси и времени. Сходная система уравнений была получена Сондхи [20].

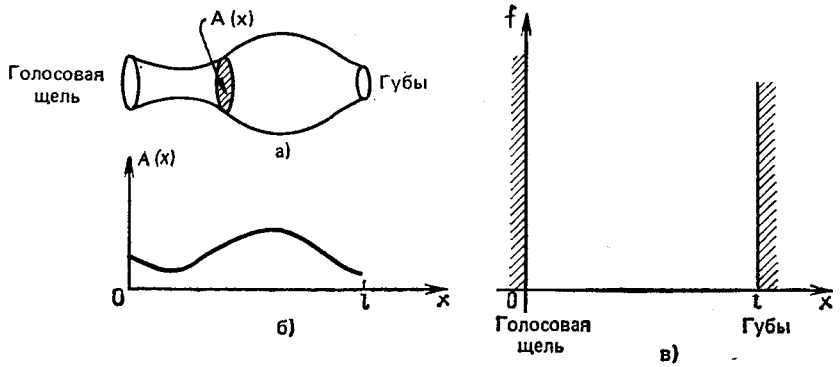


Рис. 3.13. Схематическое изображение голосового тракта (*а*), функция площади его поперечного сечения (*б*) и плоскость x—t для решения волнового уравнения (*в*)

Замкнутое решение уравнений (3.1) получить невозможно даже для простых форм трубы. Однако могут быть получены численные решения. Полное решение дифференциальных уравнений предполагает заданными давление и скорость потока для значений x и t в области голосовой щели и около губ, т. е. для получения решения должны быть заданы граничные условия у обоих концов трубы. Со стороны губ граничные условия должны отображать эффект излучения, а со стороны голосовой щели — характер возбуждения.

Кроме граничных условий необходимо задать функцию площади $A(x, t)$. На рис. 3.13*б* показана функция площади для трубы рис. 3.13*а* в некоторый момент времени. Для протяжных звуков можно предположить, что $A(x, t)$ не изменяется во времени.

Однако это предположение неверно для непротяжных звуков. Подробные измерения $A(x, t)$ весьма затруднительны и могут быть выполнены только для протяжных звуков. Одним из методов проведения таких измерений является рентгеновская киносъемка. Фант [1] и Перкелл [21] провели несколько таких экспериментов, однако подобные измерения могут быть выполнены лишь в ограниченном объеме. Другим методом является вычисление формы голосового тракта по акустическим измерениям. В [22] описан подобный метод, предполагающий возбуждение голосового тракта внешним источником. Оба метода являются полезными для получения сведений о динамике речеобразования. Тем не менее они не могут быть применены для получения описания речевых сигналов, например, в задачах связи. В работе Атала [23] описаны результаты прямого измерения $A(x, t)$ по сигналу речи, произнесенной в нормальных условиях.

Точное решение уравнений (3.1) является весьма сложным, даже если значение $A(x, t)$ точно известно. Вместе с тем для изучения структуры речевого сигнала нет необходимости в точном и общем решениях этих уравнений. Достаточно рассмотреть приближенные решения для простых ситуаций.

3.2.2. Однородная труба без потерь (пример)

Для изучения структуры речевого сигнала весьма полезно рассмотреть простую модель речеобразования, в которой функция площади постоянна по обоим аргументам (однородная труба инвариантная к временно́му сдвигу). Такая форма голосового тракта является празомерной для нейтрального гласного */UH/*. Рассмотрим эту модель, оставив для более позднего анализа исследование сложных ситуаций. На рис. 3.14*а* приведена схема однородной трубы, возбуждаемая идеальным источником. Источник представляет собой поршень, который может передвигаться независимо от давления в трубе. Дальнейшее упрощение состоит в предположении, что конец трубы открыт. Звуковое давление, таким образом, постоянно. Изменяется только скорость воздушного потока. Очевидно, что эти существенные упрощения не имеют места в действительности. Однако на этом примере мы убедимся в том, что основной подход к анализу и особенности полученного решения имеют много общего и с более сложной ситуацией. Покажем также, что более общие модели могут быть получены путем соединения однородных труб.

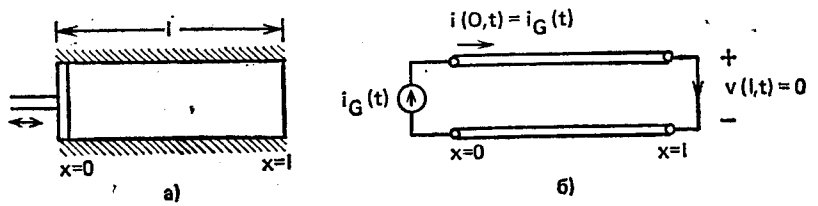


Рис. 3.14. Однородная труба без потерь с идеальной нагрузкой (*а*) и эквивалентная электрическая линия (*б*)

Если $A(x,\ t)=A$, то уравнения в частных производных (3.1) запишутся в виде

$$-\frac{\partial p}{\partial x}=\frac{\rho}{A}\frac{\partial u}{\partial t};\quad -\frac{\partial u}{\partial x}=\frac{A}{\rho c^2}\frac{\partial p}{\partial t}. \qquad (3.2а;\ б)$$

Можно показать (см. задачу 3.3), что решением уравнений (3.2) являются соотношения

$$u(x,\ t)=[u^+(t-x/c)-u^-(t+x/c)]; \qquad (3.3а)$$

$$p(x,\ t)=(\rho c/A)\,[u^+(t-x/c)+u^-(t+x/c)]. \qquad (3.3б)$$

В (3.3) функции $u^+(t-x/c)$ и $u^-(t+x/c)$ могут рассматриваться как волны, распространяющиеся в положительном (прямом) и отрицательном (обратном) направлениях. Взаимосвязь этих волн определяется граничными условиями.

Читатели, знакомые с теорией электрических длинных линий, знают, что для однородной линии без потерь напряжение $v(x,\ t)$ и ток $i(x,\ t)$ удовлетворяют уравнениям

$$-\frac{\partial v}{\partial x}=L\frac{\partial i}{\partial t};\quad -\frac{\partial i}{\partial x}=C\frac{\partial v}{\partial t}, \qquad (3.4а;\ б)$$

где L и C — индуктивность и емкость на единицу длины соответственно. Таким образом, теория однородных электрических длинных линий без потерь [24, 25] применима непосредственно для анализа однородной акустической трубы (табл. 3.3). Из таблицы следует, что однородная акустическая труба эквивалентна однородной линии без потерь, коротко замкнутой на одном конце ($v(l,\ t)=0$) и возбуждаемой источником тока ($i(0,\ t)=i_G(t)$) на другом. Это показано на рис. 3.14*б*.

Таблица 3.3

Акустическая переменная	Электрическая переменная (аналог)
p — давление	U — напряжение
u — скорость потока	i — ток
ρ/A — акустическая индуктивность	L — индуктивность
$A/(\rho c^2)$ — акустическая емкость	C — емкость

Весьма полезным является частотное описание таких линейных систем как длинные линии и линейные цепи. По аналогии можно получить такое описание и для однородной трубы без потерь. Для описания такой трубы в частотной области положим, что граничные условия в точке $x=0$:

$$u(0,\ t)=u_G(t)=u_G(\Omega)\,e^{i\Omega t}. \qquad (3.5)$$

Таким образом, труба возбуждается комплексным экспоненциальным потоком с круговой частотой Ω и комплексной амплитудой $U_G(\Omega)$. Так как уравнения (3.2) линейные, решения $u^+(t-x/c)$ и $u^-(t+x/c)$ будут иметь вид

$$u^+(t-x/c)=K^+e^{i\Omega(t-x/c)}; \qquad (3.6а)$$

$$u^-(t+x/c)=K^-e^{i\Omega(t+x/c)}. \qquad (3.6б)$$

Подставляя эти соотношения в (3.3) и используя граничные условия

$$p(l,\ t)=0 \qquad (3.7)$$

на конце трубы со стороны губ и (3.5) — со стороны голосовой щели, можно

решить (3.2) относительно неизвестных коэффициентов K^+ и K^-. В результате для $p(x,\ t)$ и $u(x,\ t)$ получаем

$$p(x,\ t)=iZ_0\frac{\sin[\Omega(l-x)/c]}{\cos[\Omega l/c]}U_G(\Omega)\,e^{i\Omega t}; \qquad (3.8а)$$

$$u(x,\ t)=\frac{\cos[\Omega(lx)/c]}{\cos[\Omega l/c]}U_G(\Omega)\,e^{i\Omega t}; \qquad (3.8б)$$

по аналогии с длинными линиями называется *характеристическим сопротивлением акустической трубы.*

В дальнейшем будем избегать решения уравнений относительно прямой и отраженной волн, принимая в качестве функций $p(x,\ t)$ и $u(x,\ t)$ при комплексном возбуждении выражения[1]

$$p(x,\ t)=P(x,\ \Omega)\,e^{i\Omega t}; \qquad (3.10а)$$

$$u(x,\ t)=U(x,\ \Omega)\,e^{i\Omega t}. \qquad (3.10б)$$

Подставляя эти соотношения в (3.1), получаем уравнения в полных производных относительно комплексных амплитуд

$$-\frac{dP}{dx}=ZU;\quad -\frac{dU}{dx}=YP, \qquad (3.11а;\ б)$$

где

$$Z=i\Omega\rho/A \qquad (3.12)$$

называется *акустическим сопротивлением на единицу длины*;

$$Y=i\Omega A/\rho c^2 \qquad (3.13)$$

называется *акустической проводимостью на единицу длины.* Дифференциальные уравнения (3.11) имеют следующее решение:

$$P(x,\ \Omega)=Ae^{\gamma x}+Be^{-\gamma x}; \qquad (3.14а)$$

$$U(x,\ \Omega)=Ce^{\gamma x}+De^{-\gamma x}, \qquad (3.14б)$$

где

$$\gamma=\sqrt{ZY}=i\Omega/c. \qquad (3.14в)$$

Здесь неизвестные коэффициенты могут быть найдены из граничных условий

$$P(l,\ \Omega)=0;\quad U(0,\ \Omega)=U_G(\Omega). \qquad (3.15а;\ б)$$

Очевидно, что полученное решение полностью совпадает с (3.8). Уравнения (3.8) описывают взаимосвязь между синусоидальным потоком источника, звуковым давлением и скоростью потока в каждой точке трубы. В частности, если нас интересует взаимосвязь скорости потока около губ со скоростью потока источника, из (3.8б) получаем

$$u(l,\ t)=U(l,\ \Omega)\,e^{i\Omega t}=\frac{1}{\cos(\Omega l/c)}U_G(\Omega)\,e^{i\Omega t}. \qquad (3.16)$$

Отношение

$$\frac{U(l,\ \Omega)}{U_G(\Omega)}=V_a(i\Omega)=\frac{1}{\cos(\Omega l/c)}. \qquad (3.17)$$

является частотной характеристикой, связывающей скорости потоков на входе и выходе. Эта функция изображена на рис. 3.15*а* для $l=17{,}5$ см и $c=35\,000$ см/с.

[1] Далее переменные во временной области будут обозначаться малыми буквами, например $u(x,\ t)$, а соответствующие им переменные в частотной области — большими, например $U(x,\ \Omega)$.

Заменяя Ω на s/i, получаем передаточную функцию в виде преобразования Лапласа (системную функцию)

$$V_a(s) = 2\mathrm{e}^{-sl/c}/(1+\mathrm{e}^{-s2l/c}). \qquad (3.18)$$

Заметим, что $V_a(s)$ имеет бесконечное число полюсов, равномерно распределенных на оси $\mathrm{i}\Omega$ так, что (рис. 3.15*б*)

$$s_n = \pm\,\mathrm{i}\,[(2n+1)\,\pi c/2l], \quad n = 0,\ \pm 1,\ \pm 2, \ldots \qquad (3.19)$$

Полюса системной функции линейной системы, инвариантной к сдвигу, определяют собственные частоты системы. Кроме того, они указывают на расположение резонансных частот системы. Эти резонансные частоты называют формантными, когда говорят о процессе речеобразования. Как мы увидим далее, аналогичные резонансные явления будут проявляться при разных конфигурациях голосового тракта.

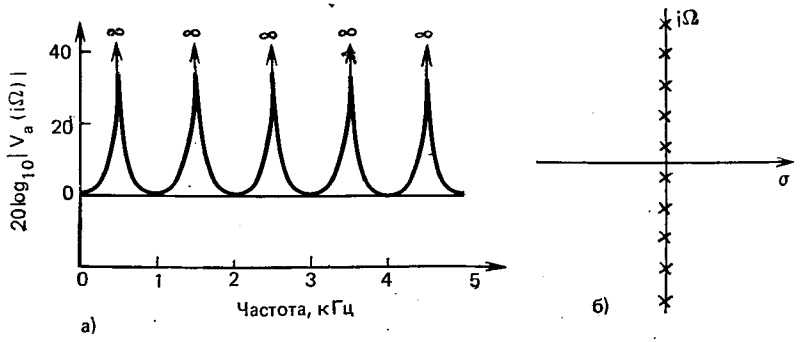


Рис. 3.15. Частотная характеристика (*а*) и расположение полюсов (*б*) для однородной трубы без потерь

Следует подчеркнуть, что частотная характеристика позволяет определить отклик системы не только для синусоидальных воздействий, но и для произвольных сигналов с помощью преобразования Фурье. Действительно, (3.17) можно рассматривать как отношение $V_a(\mathrm{i}\Omega)$ преобразования Фурье скорости потока у губ (на выходе) к преобразованию Фурье скорости потока у голосовой щели (на входе). Таким образом, частотная характеристика является удобным средством описания голосового тракта. После того как мы выяснили способ определения частотной характеристики простейшей акустической модели речеобразования, перейдем к анализу более точных моделей.

3.2.3. Потери в голосовом тракте

Уравнения, определяющие распространение звуковых волн в голосовом тракте, были получены в предположении отсутствия потерь энергии в трубе. В действительности потери энергии будут возникать за счет вязкого трения между потоком воздуха и стенками трубы, теплопроводности и колебаний стенок голосового тракта. Для анализа этих эффектов необходимо использовать основные законы физики и вывести новую систему уравнений распространения звуковых волн в голосовом тракте. Получение таких уравнений особенно затрудняется тем, что потери зависят от частоты потока. Поэтому наиболее удобно формировать и решать эти уравнения в частотной области [2, 18].

Рассмотрим вначале эффект колебаний стенок голосового тракта. При изменении звукового давления внутри голосового тракта на его стенки будет воздействовать сила переменной величины. Если стенки тракта мягкие (не жесткие), то площадь поперечного сечения трубы будет изменяться в зависимости от давления. Предполагая, что стенки реагируют на давление «локально» [17, 18], будем считать, что площадь поперечного сечения $A(x, t)$ является функцией $p(x, t)$. Так как изменения звукового давления очень малы, отклонения площади поперечного сечения от «номинального» значения будут небольшими и можно записать

$$A(x, t) = A_0(x, t) + \delta A(x, t), \qquad (3.20)$$

где $A_0(x, t)$ — номинальное значение площади и $\delta A(x, t)$ — малые отклонения. Это показано на рис. 3.16. С учетом влияния массы и мягкости стенок взаимосвязь между $\delta A(x, t)$ и $p(x, t)$ можно записать в виде следующего дифференциального уравнения:

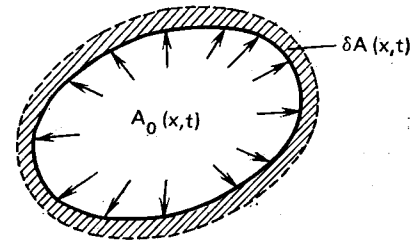


Рис. 3.16. Иллюстрация вибрации стенок голосового тракта

$$m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w(\delta A) = p(x, t), \qquad (3.21)$$

где $m_w(x)$ — масса стенок на единицу длины; $b_w(x)$ — затухание на единицу длины стенок; $k_w(x)$ — жесткость на единицу длины стенок голосового тракта.

Пренебрегая членами второго порядка относительно величин u/A и pA, перепишем (3.1) в виде

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial (u/A_0)}{\partial t}; \qquad (3.22а)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial (p A_0)}{\partial t} + \frac{\partial A_0}{\partial t} + \frac{\partial (\delta A)}{\partial t}. \qquad (3.22б)$$

Итак, распространение звуковых волн в трубе с мягкими стенками, такой как голосовой тракт, описывается системой уравнений (3.20)—(3.22).

Для подробного изучения этого эффекта перейдем в частотную область и, как и раньше, рассмотрим трубу, инвариантную к сдвигу, возбуждаемую потоком с комплексной скоростью. Граничные условия со стороны голосовой щели равны

$$u(0, t) = U_G(\Omega)\,\mathrm{e}^{\mathrm{i}\Omega t}. \qquad (3.23)$$

Так как уравнения (3.21) и (3.22) линейные и инвариантны к сдвигу, скорость потока и давление могут быть представлены в виде

$$p(x, t) = P(x, \Omega)\,\mathrm{e}^{\mathrm{i}\Omega t}; \qquad (3.24а)$$

$$u(x, t) = U(x, \Omega)\, e^{i\Omega t}. \qquad (3.24б)$$

Подставляя (3.24) в (3.21) и (3.22), получаем

$$-\frac{\partial P}{\partial x} = ZU\,; \quad -\frac{\partial U}{\partial x} = YP + Y_w P, \qquad (3,25а);\ (3.25б)$$

где

$$Z(x, \Omega) = i\Omega \frac{\rho}{A_0(x)}\,; \qquad (3.26а)$$

$$Y(x, \Omega) = i\Omega \frac{A_0(x)}{\rho c^2}\,; \qquad (3.26б)$$

$$Y_w(x, \Omega) = \frac{1}{i\Omega\, m_w(x) + b_w(x) + [k_w(x)/i\Omega]}\,. \qquad (3.26в)$$

Заметим, что (3.25) почти совпадает с (3.11). Различие состоит в том, что появился член Y_w (за счет учета проводимости стенок) и акустические сопротивление и проводимость оказываются функциями x. Если труба однородна и $A_0(x)$ постоянна, то (3.12), (3.13) совпадают с (3.26а, 3.26б).

По измерениям, проведенным на теле языка [2], в [18, 19] вычислены параметры (3.26в), после чего дифференциальные уравнения (3.25) были решены для граничных условий $p(l, t) = 0$ около губ. Отношение

$$V_a(i\Omega) = \frac{U(l, \Omega)}{U_G(\Omega)} \qquad (3.27)$$

изображено на рис. 3.17 как функция Ω для однородной трубы длиной 17,5 см [18]. Результат близок к рис. 3.15, но имеет существенное отличие. Видно, что резонансы в этом случае расположены не на оси $i\Omega$ в s-плоскости. Это означает, что частотная характеристика уже не равна бесконечности на частотах 500, 1500, 2500 Гц и т. д., хотя она и имеет максимумы на этих частотах. Центральные частоты и ширина этих резонансных областей[1] приведены в таблице на рис. 3.17. Отметим несколько существенных особенностей этого примера. Во-первых, центральные частоты расположены выше, чем в случае трубы без потерь. Во-вторых, ширина резонансных областей отлична от нуля, так как на резонансной частоте частотная характеристика имеет конечное значение. Влияние мягкости стенок наиболее существенно проявляется на низких частотах. Этого можно было ожидать, так как массивные стенки незначительно отклоняются на высоких частотах.

Результаты этого примера отражают типичные явления, происходящие при вибрации стенок голосового тракта; центральные частоты несколько смещаются в область верхних частот, резонансные области низких частот оказываются более широкими, чем в других диапазонах. Эффекты вязкого трения и теплопроводности оказывают значительно меньшее влияние, чем вибрация стенок. В

[1] Ширина резонансных областей определяется по уровню 0,707.

[2] Фланаган подробно исследовал эти явления и установил, что влияние вязкого трения можно учесть в частотной области путем добавления в (3.25) действительного зависящего от частоты члена в выражение для акустического сопротивления Z:

$$Z(x, \Omega) = \frac{S(x)}{[A_0(x)]^2}\sqrt{\Omega\rho\mu/2} + i\Omega\frac{\rho}{A_0(x)}, \qquad (3.28а)$$

где $S(x)$ — длина окружности сечения трубы; μ — коэффициент трения; ρ — плотность воздуха в трубе.

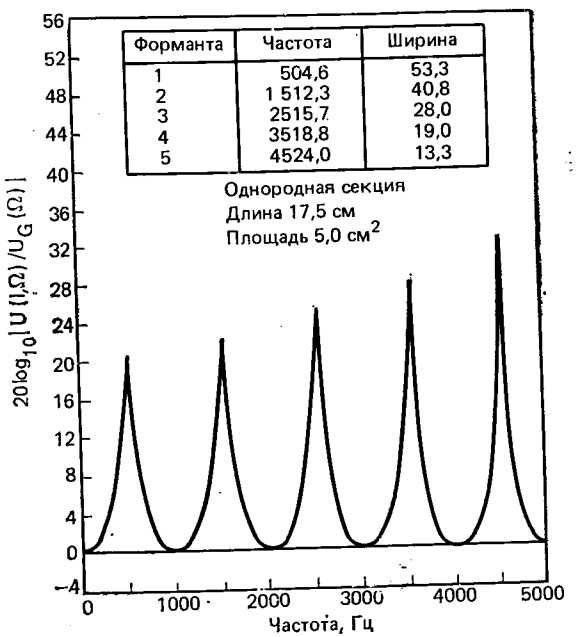


Рис. 3.17. Частотная характеристика однородной трубы с мягкими стенками. Нагрузка — короткое замыкание ($p(l, t) = 0$) [18]

Эффект теплопроводности стенок можно учесть аналогичным образом, добавляя действительный зависящий от частоты член в выражение для акустической проводимости:

$$Y(x, \Omega) = \frac{S(x)(\eta - 1)}{\rho c^2}\sqrt{\frac{\lambda\Omega}{2c_p\rho}} + i\Omega\frac{A_0(x)}{\rho c^2}, \qquad (3.28б)$$

где c_p — удельная теплоемкость при постоянном давлении; η — отношение удельной теплоемкости при постоянном давлении к этой же величине при постоянном объеме (адиабатическая постоянная); λ — коэффициент теплопроводности [2]. Значения этих постоянных в (3.28) получены Фланаганом [2]. Здесь достаточно отметить, что потери на трение пропорциональны действительной части $Z(x, \Omega)$, т. е. $\Omega^{1/2}$. Аналогично потери на теплопро-

водность пропорциональны действительной части $Y(x, \Omega)$, т. е. тоже $\Omega^{1/2}$. Используя (3.28) для $Z(x, \Omega)$ и $Y(x, \Omega)$ и выражение для $Y_w(x, \Omega)$ (3.26в), уравнения (3.25) можно решить численно [18]. Частотная характеристика, полученная для граничных условий $p(l, t)=0$, показана на рис. 3.18.

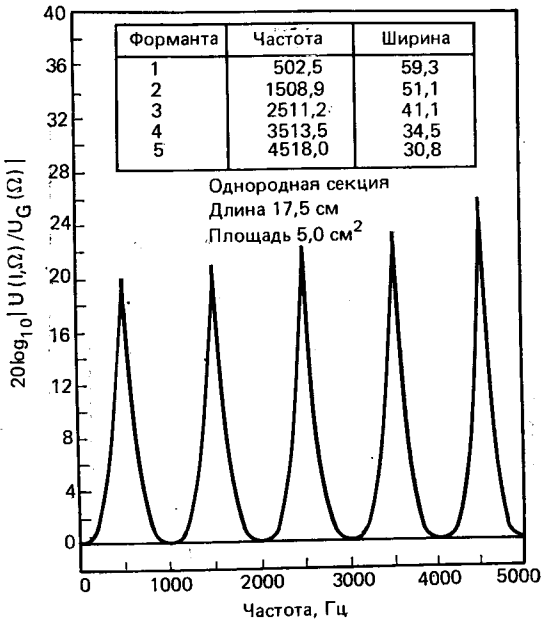


Рис. 3.18. Частотная характеристика однородной трубы с мягкими стенками, потерями на трение и теплопроводность. Нагрузка — короткое замыкание ($p(l, t)=0$) [18]

Как и раньше, значения резонансных частот и ширины резонансных областей приведены в таблице. Сравнивая рис. 3.18 и 3.17, видим, что центральные частоты при потерях на трение и теплопроводность уменьшаются, а ширина резонансных областей увеличивается. Так как потери на трение и теплопроводность возрастают пропорционально $\Omega^{1/2}$, резонансы на высоких частотах расширяются больше, чем в низкочастотной области. Примеры рис. 3.17 и 3.18 отражают типичное влияние потерь в голосовом тракте. Итак, потери на вязкое трение и теплопроводность увеличиваются с возрастанием частоты и оказывают наибольшее влияние на резонансы в высокочастотной области. Потери за счет вибрации стенок в то же время оказывают наибольшее влияние на характеристики в области низких частот. Эффект вибрации стенок приводит к возрастанию резонансных частот, в то время как потери на трение и теплопроводность оказывают обратное влияние. В результате всех этих эффектов резонансы в низкочастотной области несколько смещаются по сравнению с резонансами в трубе

без потерь и с жесткими стенками. Влияние потерь на трение и теплопроводность мало́ по сравнению с эффектом вибрации стенок на частотах ниже 3—4 кГц. Таким образом, уравнения (3.21), (3.22), в которых не учтены эти потери, хорошо описывают распространение звуковых волн в голосовом тракте. Как мы увидим далее, эффект излучения через губы является источником существенных потерь на высоких частотах. Это еще раз подтверждает правомерность пренебрежения потерями на трение и теплопроводность в моделях речеобразования.

3.2.4. Излучение через губы

Мы изучили, каким образом внутренние потери влияют на распространение звуковых волн в голосовом тракте. В рассмотренных примерах граничные условия со стороны губ были всегда равны $p(l, t)=0$. Это соответствует короткому замыканию эквивалентной длинной линии. Получить акустический аналог короткого замыкания очень трудно, так как это предполагает такую конфигурацию трубы, при которой скорость потока может изменяться на конце трубы при неизменном давлении. Однако в реальных условиях труба голосового тракта заканчивается отверстием между губами (или в ноздрях в случае произнесения носовых звуков). Возможная модель такой конфигурации показана на рис. 3.19*а*, где отверстие между губами показано как отверстие в сфере. В этой модели на нижних частотах отверстие может рассматриваться как излучающая поверхность. Звуковые волны отражаются сферой, которой является голова человека.

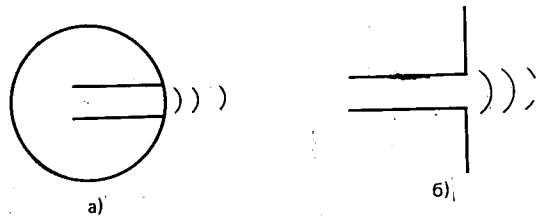


Рис. 3.19. Излучение сферической (*а*) и плоской бесконечной (*б*) поверхностью отражения

Эффект отражения очень сложен и труден для описания. Однако для того, чтобы определить граничные условия около губ, достаточно установить взаимосвязь между давлением и скоростью потока у излучающей поверхности. Это тоже оказывается весьма сложным для конструкции, изображенной на рис. 3.19*а*. Если излучающая поверхность отверстия между губами мала по сравнению с размерами сферы, можно считать отражающую поверхность плоской и бесконечно протяженной (рис. 3.19*б*). В этом случае можно показать, что связь между комплексными амплитудами давления и скорости потока около губ имеет вид

$$P(l, \Omega) = Z_L(\Omega)\, U(l, \Omega), \qquad (3.29а)$$

где нагрузочное сопротивление излучения через губы приближенно равно

$$Z_L(\Omega) = \frac{i\Omega L_r R_r}{R_r + i\Omega L_r}. \qquad (3.29б)$$

Электрический аналог этого нагрузочного сопротивления есть параллельное соединение сопротивления излучения R_r и индуктивности излучения L_r. Значения R_r и L_r, которые хорошо соответствуют предположению о плоской бесконечной поверхности отражения, равны [2]

$$R_r = 128/9\,\pi^2;\ L_r = 8a/3\,\pi c, \qquad (3.30а);\ (3.30б)$$

где a — радиус отверстия и c — скорость звука.

Характер нагрузочного сопротивления излучения влияет на структуру волны в голосовом тракте через граничные условия (3.29). Легко видеть из (3.29б), что на низких частотах $Z_L(\Omega) \approx \approx 0$, т. е. на очень низких частотах сопротивление излучения примерно соответствует короткому замыканию, которое и предполагалось ранее. Аналогично из (3.29) видно, что для диапазона средних частот (когда $\Omega L_r \ll R_r$) $Z_L(\Omega) \approx i\Omega L_r$. На высоких частотах ($\Omega L_r \gg R_r$) $Z_L(\Omega) \approx R_r$. Это отчетливо видно на рис. 3.20, где показаны действительная и мнимая части $Z_L(\Omega)$ как функции от Ω для типичных значений параметров. Энергия излучения пропорциональна реальной части сопротивления излучения. Таким образом, для полной системы речеобразования (голосовой тракт, рас-

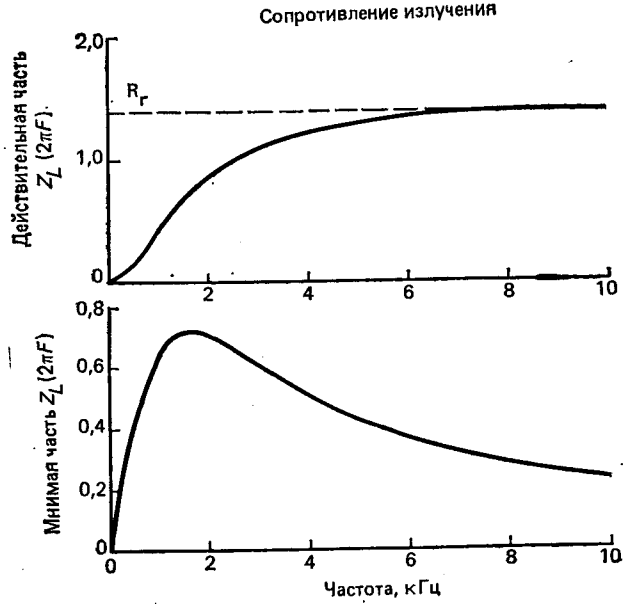


Рис. 3.20. Действительная и мнимая части сопротивления излучения

сматриваемый совместно с излучением) потери на излучение на высоких частотах значительны. Для количественной оценки этого эффекта уравнения (3.25), (3.26в), (3.29) можно решить совместно для случая однородной инвариантной к сдвигу трубы с мягкими стенками при наличии потерь на трение, теплопроводность и излучение с плоской и бесконечно протяженной поверхностью отражения. На рис. 3.21 показана результирующая частотная характеристика

$$V_a(i\Omega) = U(l, \Omega)/U_G(\Omega) \qquad (3.31)$$

для входного сигнала $U(0, t) = U_G(\Omega)e^{i\Omega t}$. Сравнение рис. 3.21, 3.17 и 3.18 показывает, что основным эффектом наличия потерь является расширение резонансных областей, т. е. увеличение зату-

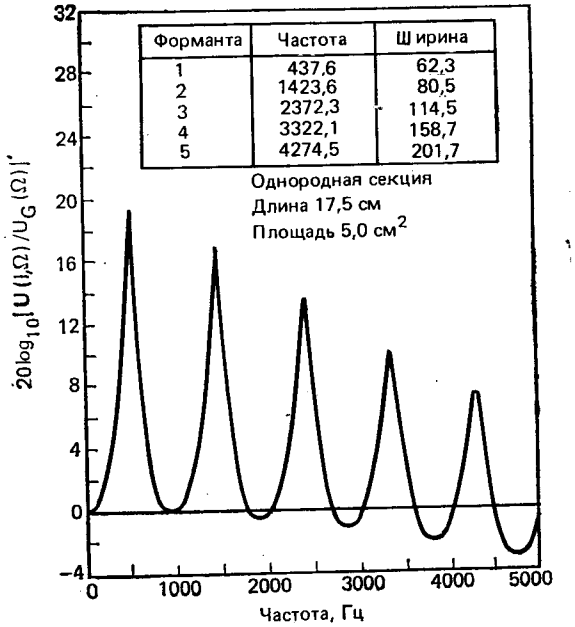


Рис. 3.21. Частотная характеристика однородной трубы с мягкими стенками, потерями на трение и теплопроводностью [18]

хания, и снижение резонансных (формантных) частот. Как и следовало ожидать, расширение резонансных областей происходит в основном на высоких частотах. Ширина первой форманты изменяется за счет потерь на стенках голосового тракта, ширина высокочастотных формант — за счет потерь на излучение, ширина второй и третьей формант формируется в результате влияния всех потерь в голосовом тракте. Частотная характеристика, показанная на рис. 3.21, определена для скорости потока у губ и скорости входного потока у губ. Представляет интерес взаимосвязь между

звуковым давлением у губ и скоростью потока у голосовой щели, особенно в том случае, когда для преобразования акустической волны в электрическое колебание используется микрофон, чувствительный к звуковому давлению. Так как $P(l, \Omega)$ и $U(l, \Omega)$ описываются выражением (3.29а), передаточная функция для звукового давления имеет простой вид:

$$H_a(\Omega) = \frac{P(l,\Omega)}{U_G(\Omega)} = \frac{P(l,\Omega)}{U(l,\Omega)} \frac{U(l,\Omega)}{U_G(\Omega)} = Z_L(\Omega) V_a(\Omega). \qquad (3.32)$$

Из рис. 3.21 видно, что влияние потерь приводит к подъему характеристик на высоких частотах и к их спаду до нуля при $\Omega = 0$. На рис. 3.22 показана величина $20\log_{10}|H_a(\Omega)|$, полученная

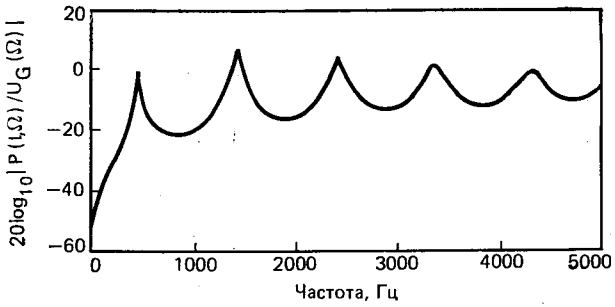


Рис. 3.22. Частотная характеристика, связывающая звуковое давление около губ и скорость потока в голосовой щели, для однородной трубы

с учетом потерь на стенках тракта и потерь на излучение при плоской и бесконечно протяженной отражательной поверхности. Сравнивая рис. 3.21 и 3.22, можно сделать вывод, что характеристики равны нулю при $\Omega = 0$ и имеют подъем на высоких частотах!

3.2.5. Передаточная функция голосового тракта для гласных

Распространение звуковых волн и их излучение при речеобразовании описываются уравнениями, приведенными в 3.2.3 и 3.2.4. Применяя методы численного интегрирования, можно решить эти уравнения в частотной или временной области и найти характеристики голосового тракта. Эти характеристики позволяют изучить подробнее процесс речеобразования и структуру речевого сигнала. В качестве примера в [18] решены уравнения (3.25), (3.26в), (3.28) и (3.29) и найдены частотные характеристики голосового тракта для функций площадей, измеренных в работе Фанта [1]. На рис. 3.23—3.26 показаны функции площади поперечного сечения голосового тракта и частотные характеристики $(U(l, \Omega)/U_G(\Omega))$ для гласных $|a|$, $|e|$, $|i|$ и $|u|$ русского языка. Эти рисунки иллюстрируют влияние потерь, которые изучались в 3.2.3 и 3.2.4. Формантные частоты и ширина формантных областей хорошо совпадают с результатами измерений, проведенных на реальных гласных Петерсоном, Барни [11], Данном [27].

В заключение сделаем следующие выводы.

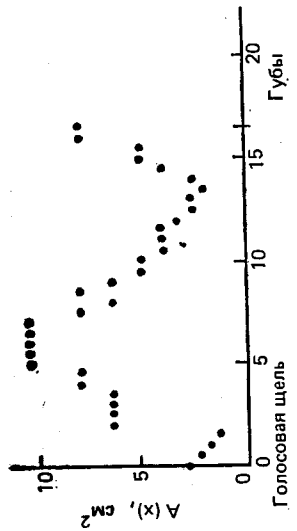


Рис. 3.23. Функция площади [1] и частотная характеристика [18] для гласного /a/ русского языка

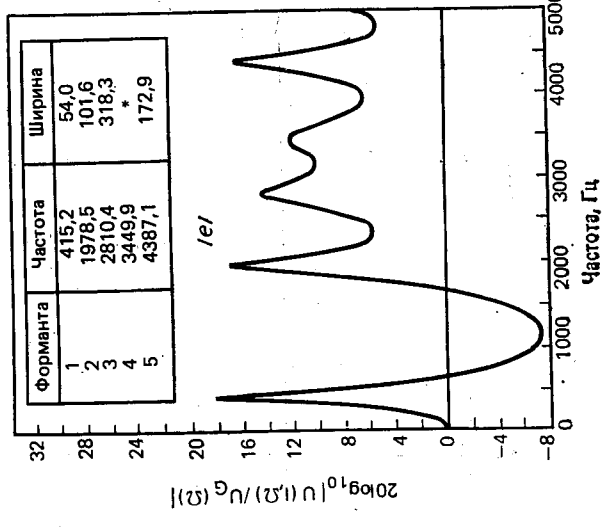
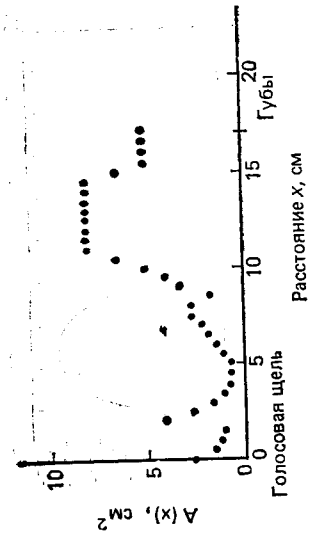
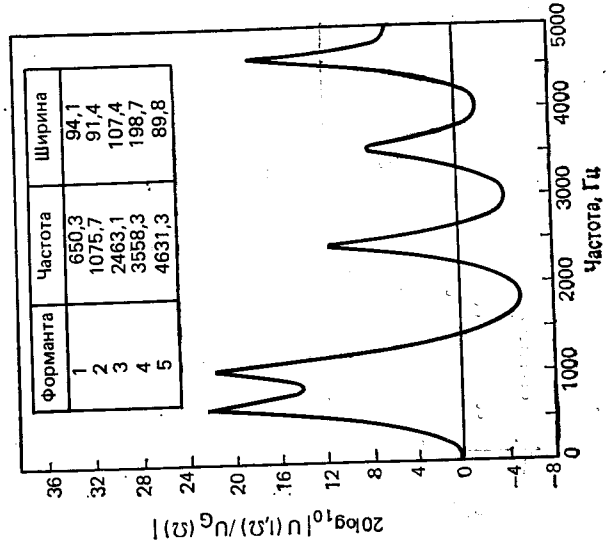


Рис. 3.24. Функция площади [1] и частотная характеристика [18] для гласного /e/ русского языка





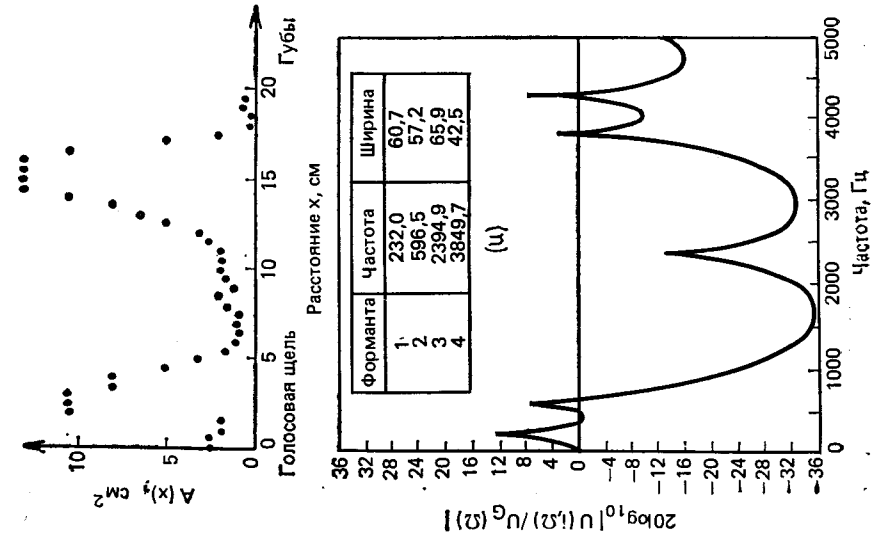


Рис. 3.25. Функция площади [1] и частотная характеристика [18] для гласного /i/ русского языка

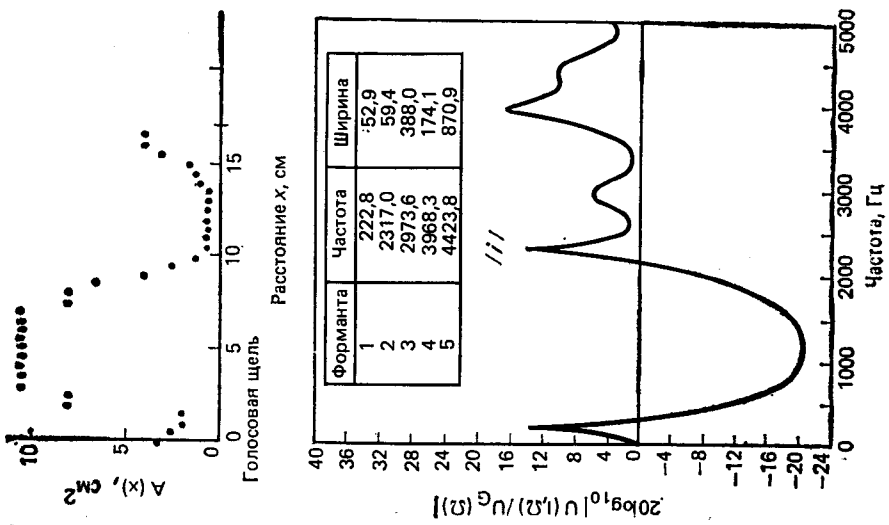


Рис. 3.26. Функция площади [1] и частотная характеристика [18] для гласного /u/ русского языка

1. Система речеобразования описывается набором резонансов (формант), которые определяются в первую очередь функцией площади поперечного сечения голосового тракта. Некоторый сдвиг резонансных частот возникает за счет потерь.

2. Ширина низкочастотных формантных областей (первой и второй) зависит от потерь на стенках голосового тракта[1].

3. Ширина высокочастотных формантных областей зависит в первую очередь от потерь на вязкое трение, теплопроводность и излучение.

3.2.6. Влияние носовой полости

При образовании носовых согласных $|m|$, $|n|$ и $|\eta|$ небная занавеска играет роль «двери» для подключения носового тракта к гортани. Одновременно в ротовой полости формируется полная смычка (перекрытие), например между губами при произнесении звука $|m|$. Такая конфигурация голосового тракта представлена на рис. 3.27*а*, где показаны два ответвления трубы, одно из кото-

Рис. 3.27. Модель образования носовых звуков (*а*) и ее электрический аналог (*б*)

рых полностью перекрыто. В точке соединения звуковое давление такое же, как на входе каждой трубы. Скорость звукового потока в этой точке должна быть непрерывной, т. е. скорость потока на выходе гортани должна быть равна сумме скоростей потоков на входах носовой и ротовой полостей. Соответствующий аналог в виде электрической линии показан на рис. 3.27*б*. Заметим, что непрерывность скоростей потока в месте соединения трех труб соответствует закону Кирхгофа для токов в точке соединения длинных линий.

Излучение звуковых волн при произнесении носовых звуков происходит в основном через ноздри. Таким образом, труба носового тракта имеет сопротивление излучения, зависящее от размера отверстия в ноздрях. Ротовой тракт, который полностью перекрыт, соответствует разомкнутой электрической линии, т. е. ток по линии не протекает. Носовые гласные формируются в подобной же системе, но ротовой тракт при этом имеет ту же форму, что и для обычных гласных. Речевой сигнал определяется взаимодейст-

[1] В 3.2.7 будет показано, что потери, связанные с источником возбуждения, также влияют на низкочастотные форманты.

вием звуковых волн на выходах носовой и ротовой полостей. Математическое описание голосового тракта такой конфигурации сводится к трем системам уравнений в частных производных с граничными условиями, зависящими от формы возбуждения, конфигурации ротовой и носовой полостей и соотношения непрерывности в точке их соединения. В итоге получается весьма сложная система уравнений, которую в принципе можно решить, если располагать измерениями функции площади поперечного сечения для всех трех труб. Передаточная функция этой сложной системы будет обладать свойствами, весьма сходными со свойствами характеристик уже рассмотренных примеров. Так, система будет обладать набором формант, которые зависят от формы и длины трех труб. Существенным отличием этой системы является то, что перекрытая ротовая полость способна задерживать акустический поток при определенных его частотах, не допуская прохождения потока в носовую полость. В эквивалентной электрической линии эти частоты соответствуют частотам, на которых входное сопротивление разомкнутой линии равно нулю. На этих частотах место соединения эквивалентных линий накоротко замкнуто линией, отображающей ротовую полость. В результате в передаточной функции речеобразующей системы для носовых звуков наряду с резонансами появляются антирезонансы (нули). В результате измерений было выяснено, что для носовых звуков формантные области являются более широкими, чем для неносовых звуков. Это объясняется большими потерями за счет вязкого трения и теплопроводности, возникающими вследствие большой площади поверхности носовой полости.

3.2.7. Возбуждение звуков в голосовом тракте

В предыдущих разделах показано, как законы физики могут быть применены для описания распространения и излучения звуковых волн при речеобразовании. Для завершения изучения акустических аспектов речеобразования необходимо рассмотреть механизм возбуждения звуковых волн в речеобразующей системе. Напомним, что в общем обзоре в 3.1.1 выделено три способа возбуждения:

1. Воздушный поток, нагнетаемый из легких, модулируется за счет вибраций голосовых связок. В результате возникает квазипериодический импульсный поток.

2. Воздушный поток из легких становится турбулентным при прохождении сужения голосового тракта. В результате возникает шумоподобное возбуждение.

3. Воздушный поток сжимается легкими перед смычкой голосового тракта. Далее этот воздух внезапно высвобождается при устранении смычки, вызывая шумоподобное возбуждение.

Подробная схема возбуждения звуковых волн включает подглоточную систему (легкие, бронхи, трахею), голосовую щель и голосовой тракт. Безусловно полная модель описывает не только

речеобразование, но и процесс дыхания [2]!. Первая попытка создания физической модели возбуждения звуков в речеобразующей системе сделана в работе Фланагана [2, 28]. В последующих исследованиях разработана более совершенная модель, которая подробно описывает процесс образования вокализованной и невокализованной речи [28—31]. Эта модель основана на классической механике и механике жидкостей, и ее анализ выходит за рамки настоящей книги. Однако даже краткое качественное описание основных принципов возбуждения звуков оказывается весьма полезным для объяснения упрощенных моделей, которые широко используются при обработке речевых сигналов.

Вибрацию голосовых связок при образовании вокализованной речи можно упрощенно представить в виде рис. 3.28. Голосовые

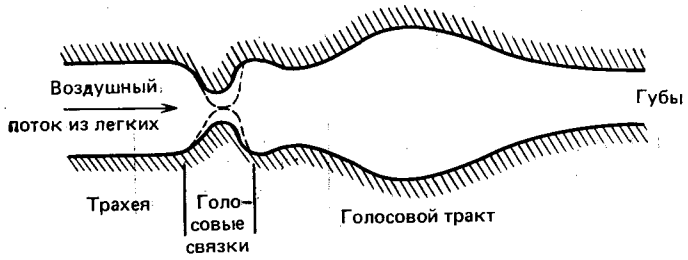


Рис. 3.28. Схематическое изображение речеобразующей системы

связки преграждают путь воздушному потоку из легких в голосовой тракт. Когда звуковое давление в легких возрастает, воздушный поток, нагнетаемый из легких, проходит через отверстие между голосовыми связками. По мере нарастания воздушного потока давление в голосовой щели падает в соответствии с законом Бернулли. Вследствие натяжения голосовых связок и уменьшения давления в голосовой щели связки соединяются, образуя полное перекрытие. На рис. 3.28 это показано пунктирными линиями. В результате давление звукового потока перед связками начинает возрастать. Когда давление повышается до уровня, достаточного чтобы раздвинуть связки, голосовая щель раскрывается и воздушный поток вновь проходит в голосовой тракт. Давление в голосовой щели снова падает, и цикл повторяется. Таким образом возникают условия, при которых голосовые связки начинают вибрировать. Частота вибрации связок зависит от давления потока, нагнетаемого из легких, массы и упругости голосовых связок, а также площади голосовой щели в свободном состоянии. Эти параметры могут быть приняты за основу создания модели голосовых связок. Такие модели должны учитывать и влияние голосового тракта, так как изменение звукового давления в голосовом тракте влияет на давление в голосовой щели. С точки зрения электрических аналогий голосовой тракт играет роль нагрузки генератора звукового возбуждения.

Схематическая диаграмма модели голосовых связок [30] показана на рис. 3.29*а*. Модель описывается системой сложных нелинейных дифференциальных уравнений. Объединение этих уравнений с дифференциальными уравнениями в частных производных,

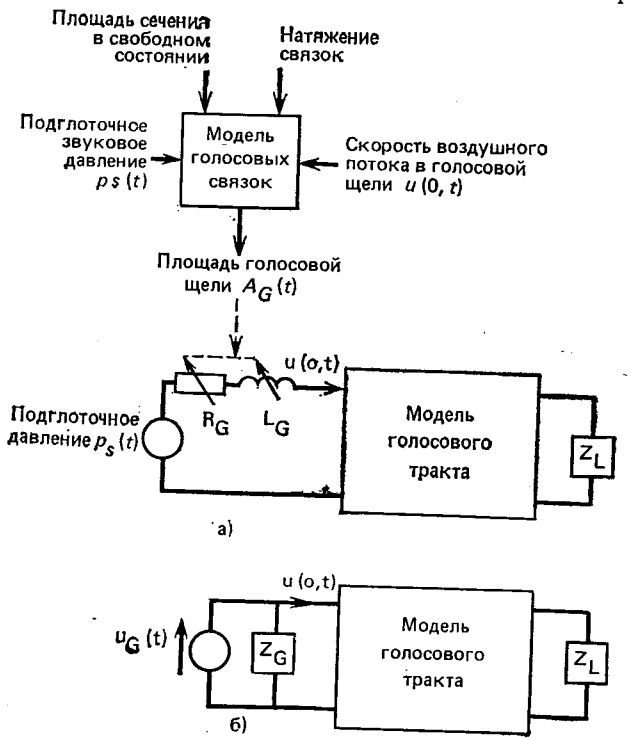


Рис. 3.29. Схематическое изображение модели голосовых связок

которыми описывается голосовой тракт, можно выполнить путем введения переменного во времени акустического сопротивления и индуктивности [30]. Это сопротивление является функцией $1/A_G(t)$. Например, когда $A_G(t)=0$ (голосовая щель закрыта), сопротивление равно бесконечности, а скорость воздушного потока равна нулю. Таким образом, воздушный поток автоматически приобретает импульсный характер. Пример сигналов, формируемых в таких моделях, показан на рис. 3.30 [30]. В верхней части рисунка показана скорость воздушного потока, а в нижней — давление около губ для конфигурации голосового тракта, соответствующей гласной $|a|$. Импульсная структура потока в голосовой щели согласуется с ранее изложенным материалом и с результатами высокоскоростной киносъемки [2]. Естественно также, что затухающие колебания на выходе согласуются с изложенной трактовкой природы распространения звука в голосовом тракте.

Так как площадь голосовой щели является функцией потока в голосовом тракте, система, изображенная на рис. 3.29*а*, в общем случае нелинейна, хотя голосовой тракт и тракт излучения являются линейными. Взаимодействие между голосовым трактом и голосовой щелью невелико, и, как правило, им пренебрегают[1]. При

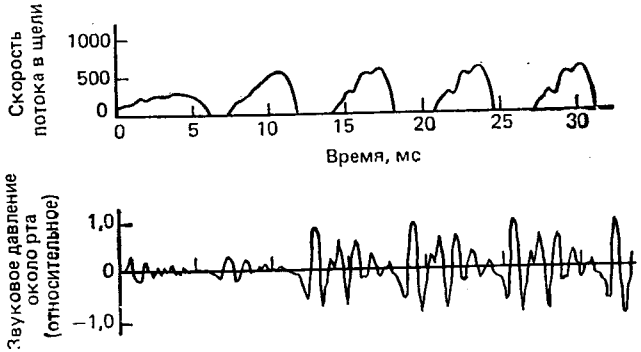


Рис. 3.30. Скорость звукового потока в голосовой щели и звуковое давление около рта для гласного /*a*/ [30]

этом возникает возможность разделения систем возбуждения и преобразования звуковых волн и их линеаризации, как это показано на рис. 3.29*б*. В этом случае $u_G(t)$ является источником скорости звукового потока (источником тока), сигнал которого показан в верхней части рис. 3.30. Акустическое сопротивление голосовой щели Z_G можно получить путем линеаризации соотношений, связывающих давление и скорость воздушного потока в голосовой щели [2]. Это сопротивление равно

$$Z_G(\Omega) = R_G + \mathrm{i}\,\Omega L_G, \tag{3.33}$$

где R_G и L_G постоянны. В этом случае идеальные граничные условия в частотной области $U(0, \Omega)=U_G(\Omega)$ следует заменить соотношением

$$U(0, \Omega) = U_G(\Omega) - P(0, \Omega)/Z_G(\Omega). \tag{3.34}$$

Сопротивление источника оказывает значительное влияние на ширину резонансных областей речеобразующей системы. Наиболее сильно это влияние сказывается на ширине низкочастотной резонансной области. Это происходит потому, что $Z_G(\Omega)$ растет с увеличением частоты так, что на высоких частотах Z_G соответствует разомкнутой цепи и весь поток от источника возбуждения проходит в голосовой тракт. Таким образом, вибрация стенок голосового тракта и потери в голосовой щели влияют на ширину

[1] Строго говоря, такие рассуждения неверны, так как в нелинейной системе малому по величине взаимодействию могут соответствовать существенные изменения выходного сигнала. По-видимому, именно это и проявляется при речеобразовании, что объясняет несовершенство современных систем обработки и передачи речи, в основу которых положено это «мелкое» упрощение. (*Прим. ред.*)

низкочастотных формантных областей, в то время как потери на излучение, трение и теплопроводность влияют на ширину высокочастотных формантных областей.

Механизм образования невокализованных звуков основан на формировании турбулентного воздушного потока. Он формируется в месте сужения голосового тракта, когда скорость потока возрастает до определенного критического уровня [2, 29]. Такое возбуждение можно имитировать путем введения источника случайного нестационарного шума в область сужения. Мощность возбуждения должна зависеть от скорости потока в трубе. Это позволяет учесть потери на трение [2, 29, 31]. При образовании фрикативных звуков параметры голосовых связок принимают такие значения, при которых связки не вибрируют. При образовании вокализованных фрикативных звуков голосовые связки вибрируют и одновременно, когда скорость потока достигает критического значения, в месте сужения голосового тракта возникает турбулентный поток. Обычно это сказывается в моменты пиков скорости импульсного воздушного потока. При произнесении взрывных звуков голосовой тракт перекрывается на период времени, когда перед смычкой воздух, нагнетаемый из легких, сжимается. Голосовые связки в это время неподвижны. Далее воздух за смычкой внезапно высвобождается, поток приобретает большую скорость и, таким образом, возникает турбулентность.

3.2.8. Модели сигнала, основанные на акустической теории

В § 3.2 изложены основные положения акустической теории речеобразования. Модели возбуждения, распространения и излучения звуковых волн описываются сложными уравнениями. Для определения речевого колебания на выходе эти уравнения можно разрешить при соответствующих значениях параметров возбуждения и голосового тракта. Естественно, что такой способ синтеза речи является наиболее эффективным [31]. Однако во многих случаях такой сложный синтез оказывается неприемлемым. В этих ситуациях на основе акустической теории можно получить упрощенные модели синтеза. На рис. 3.31 показана общая структурная

Рис. 3.31. Модель речеобразования

схема, по которой разработано множество моделей, применяемых при обработке речевых сигналов. Основной особенностью этих моделей является то, что источник возбуждения и голосовой тракт рассматриваются как отдельные системы[1]. Голосовой тракт с учетом излучения представлен линейной системой с переменными параметрами. Эта система отображает резонансные явления в голосовом тракте. Генератор возбуждения формирует сигнал либо в виде последовательности импульсов, либо в форме шумоподобного процесса. Параметры источника возбуждения и линейной системы выбираются так, что формируемый на выходе сигнал оказывается речеподобным. Если удается достигнуть этого, то полученная модель может быть использована при обработке речевого сигнала. В последующей части главы изучаются модели такого типа.

[1] См. прим. ред. на с. 81.

3.3. Модели с трубами без потерь

Одна из наиболее распространенных моделей речеобразования основана на предположении, что голосовой тракт можно представить соединением акустических труб без потерь (рис. 3.32). Площади поперечного сечения труб выбираются так, чтобы результирующая оказалась равной функции площади поперечного сечения голосового тракта $A(x)$. Если количество труб велико, а их длина достаточно мала, то можно ожидать, что резонансные частоты такого соединения будут близки к резонансным частотам трубы с непрерывной функцией площади сечения. Поскольку, однако, при таком приближении пренебрегают потерями на трение, теплопроводность и вибрацию стенок, можно ожидать, что ширина резонансных областей будет отличаться от ширины таких областей в сложной модели, учитывающей и потери. Потери в голосовой щели и около губ могут быть учтены и, как будет показано в гл. 8, это позволяет достаточно точно описать резонансные свойства речевого сигнала.

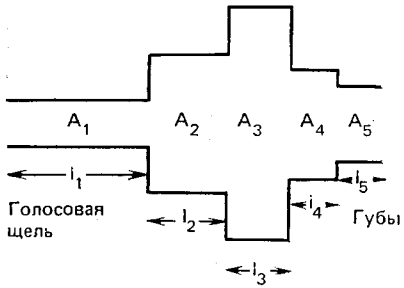


Рис. 3.32. Соединение пяти труб без потерь

Более важным вопросом является то обстоятельство, что модели с трубами без потерь позволяют перейти от непрерывного времени к дискретному. Иначе говоря, предметом последующего изучения является модель рис. 3.32.

3.3.1. Распространение звуковых волн в соединении труб без потерь

Так как потери в трубах (см. рис. 3.32) отсутствуют, распространение звуковых волн в каждой трубе описывается уравнениями (3.2) с соответствующими значениями площади поперечного сечения. Если рассмотреть k-ю трубу с площадью сечения A_k, то звуковое давление и скорость потока в трубе

$$p_k(x,t)=\frac{\rho c}{A_k}\left[u_k^+(t-x/c)+u_k^-(t+x/c)\right]; \qquad (3.35а)$$

$$u_k(x,t)=u_k^+(t-x/c)-u_k^-(t+x/c), \qquad (3.35б)$$

где x — расстояние от левого конца соединения до k-й трубы ($0\leqslant x\leqslant l_k$) и $u^+{}_k(\cdot)$, $u^-{}_k(\cdot)$ — прямая и отраженная волны. Взаимосвязь между волнами в соседних трубах можно установить на ос-

нове следующего соображения. Звуковое давление и скорость потока должны быть непрерывными в каждый момент времени и в любой точке системы. Это позволяет задать граничные условия для обоих концов трубы.

Рассмотрим, в частности, соединение k-й и $(k+1)$-й труб (рис. 3.33). Из условия непрерывности получаем

$$p_k(l_k, t) = p_{k+1}(0, t);\ u_k(l_k, t) = u_{k+1}(0, t). \quad (3.36а);\ (3.36б)$$

Подставляя (3.35) в (3.36), имеем

$$\frac{A_{k+1}}{A_k}\left[u_k^+(t-\tau_k) + u_k^-(t+\tau_k)\right] = u_{k+1}^+(t) + u_{k+1}^-(t); \quad (3.37а)$$

$$u_k^+(t-\tau_k) - u_k^-(t+\tau_k) = u_{k+1}^+ - u_{k+1}^-(t), \quad (3.37б)$$

где $\tau_k = l_k/c$ — время прохождения волны через k-ю трубу. Из рис. 3.33 видно, что часть прямой волны распространяется далее (направо), а часть — отражается. Аналогично часть отраженной волны

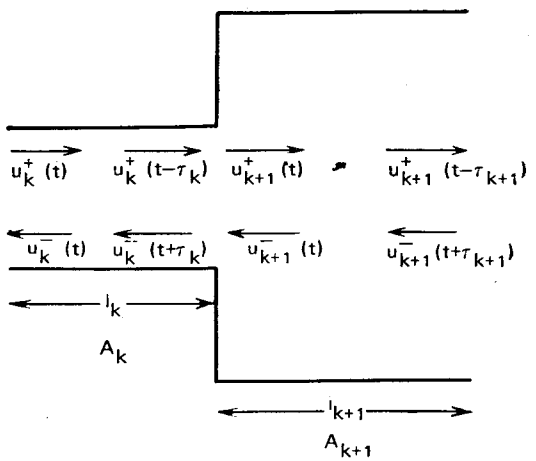


Рис. 3.33. Соединение двух труб без потерь

ны распространяется далее (налево), а другая часть — вновь отражается. Таким образом, если выразить $u^+_{k+1}(t)$ и $u^-_k(t+\tau_k)$ через $u^-_{k+1}(t)$ и $u^+_k(t-\tau_k)$, то можно подробно изучить, как распространяются прямые и отраженные волны через всю систему. Решая (3.37б) для $u^-_k(t+\tau_k)$ и подставляя результат в (3.37а), получаем

$$u_{k+1}^+(t) = \left[\frac{2A_{k+1}}{A_{k+1}+A_k}\right] u_k^+(t-\tau_k) + \left[\frac{A_{k+1}-A_k}{A_{k+1}+A_k}\right] u_{k+1}^-(t). \quad (3.38а)$$

Вычитая (3.37б) из (3.37а), имеем

$$u_k^-(t+\tau_k) = -\left[\frac{A_{k+1}-A_k}{A_{k+1}+A_k}\right] u_k^+(t-\tau_k) + \left[\frac{2A_k}{A_{k+1}+A_k}\right] u_{k+1}^-(t). \quad (3.38б)$$

Из (3.38а) видно, что величина

$$r_k = (A_{k+1}-A_k)/A_{k+1}+A_k. \quad (3.39)$$

определяет величину отраженной волны $u^-_{k+1}(t)$. Поэтому величину r_k называют коэффициентом отражения для k-го соединения труб. Можно показать (см. задачу 3.4), что так как площади поперечного сечения положительны, то

$$-1 \leqslant r_k \leqslant 1. \quad (3.40)$$

На основе определения r_k уравнения (3.38) можно записать так:

$$u_{k+1}^+(t) = (1+r_k)\,u_k^+(t-\tau_k) + r_k\,u_{k+1}^-(t); \quad (3.41а)$$

$$u_k^-(t+\tau_k) = -r_k\,u_k^+(t-\tau_k) + (1-r_k)\,u_{k+1}^-(t). \quad (3.41б)$$

Эти уравнения впервые были использованы для синтеза речи Келли и Лохбаумом [32]. Удобно изображать эти уравнения в виде графа рис. 3.34. На рисунке для отображения операций сложения

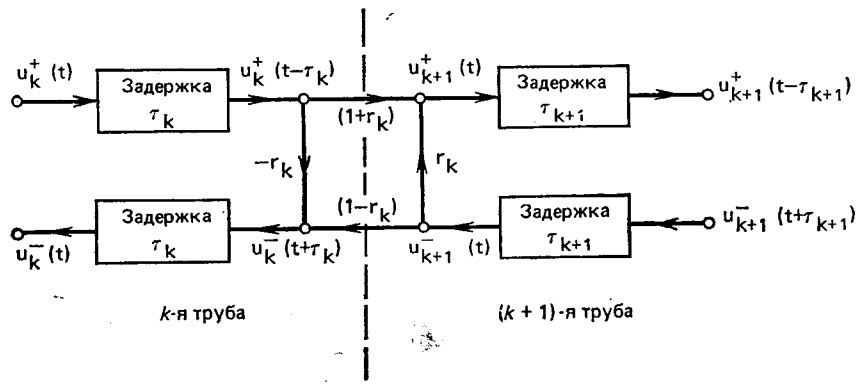


Рис. 3.34. Граф прохождения сигнала через соединение двух труб без потерь

и умножения в (3.41) использовано графическое изображение преобразований сигнала[1]. Любое соединение труб на рис. 3.32 можно представить аналогично рис. 3.34, если анализу подвергаются только значения звукового давления и скорости потока на входе и выходе труб. Такой анализ не вносит ограничений так как нас интересует только соотношение между выходом последней трубы и входом первой трубы. Модель из пяти труб рис. 3.32 будет состоять из пяти элементов задержки в прямом и обратном направлениях и четырех соединений, каждое из которых описывается коэффициентом отражения. Для полного описания распространения волны в такой системе остается ввести граничные условия со стороны губ и голосовой щели.

[1] В качестве введения к использованию сигнальных графов при обработке сигналов рекомендуется [32].

3.3.2. Граничные условия

Пронумеруем N секций модели числами от 1 до N, начиная со стороны голосовой щели. Тогда граничные условия со стороны губ будут определять звуковое давление $p_N(l_N, t)$ и скорость потока $u_N(l_N, t)$ на выходе N-й трубы с учетом излучения. В частотной области справедливо соотношение

$$P_N(l_N, \Omega) = Z_L U_N(l_N, \Omega). \tag{3.42}$$

Если предположить, что Z_L действительная величина, то во временно́й области справедливо соотношение

$$\rho c\,(u_N^+(t-\tau_N) + u_N^-(t+\tau_N))/A_N = Z_L\,(u_N^+(t-\tau_N) - u_N^-(t+\tau_N)). \tag{3.43}$$

(Если Z_L — комплексная величина, то вместо (3.43) получается дифференциальное уравнение, связывающее $p_N(l_N, t)$ и $u_N(l_N, t)$.) Решая уравнение для $u^-_N(t+\tau_N)$, получаем

$$u_N^-(t+\tau_N) = -r_L\, u_N^+(t-\tau_N), \tag{3.44}$$

где коэффициент отражения около губ равен

$$r_L = \left[\frac{\rho c/A_N - Z_L}{\rho c/A_N + Z_L}\right]. \tag{3.45}$$

Скорость выходного потока около губ

$$u_N(l_N, t) = u_N^+(t-\tau_N) - u_N^-(t+\tau_N) = (1+r_L)\,u_N^+(t-\tau_N). \tag{3.46}$$

На выходе трубы справедливы уравнения (3.44), (3.46). Модель окончания трубы изображена на рис. 3.35. Заметим, что если Z_L комплексная величина, то (3.45) остается справедливым, но величина r_L здесь также будет комплексной и для получения соотношения, аналогичного (3.44), надо перейти в частотную область. Во временно́й области $u^-_N(t+\tau_N)$ и $u^+_N(t-\tau_N)$ будут связаны дифференциальным уравнением (см. задачу 3.5).

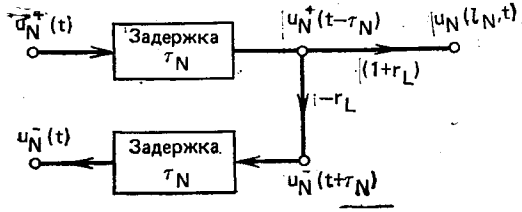


Рис. 3.35. Нагрузка со стороны губ

Соотношения в частотной области в предположении, что источник возбуждения линейно независим от голосового тракта, приведены в § 3.2.7. Применяя это предположение, здесь также можно получить соотношение для давления и скорости входного потока первой трубы

$$U_1(0, \Omega) = U_G(\Omega) - P_1(0, \Omega)/Z_G. \tag{3.47}$$

Снова предположив, что Z_G — действительная величина, имеем

$$u_1^+(t) - u_1^-(t) = u_G(t) - \frac{\rho c}{A_1}\left[\frac{u_1^+(t) + u_1^-(t)}{Z_G}\right]. \tag{3.48}$$

Решая для $u^+_1(t)$, получаем (см. задачу 3.6)

$$u_1^+(t) = \frac{(1+r_G)}{2}\,u_G(t) + r_G\,u_1^-(t), \tag{3.49}$$

где коэффициент отражения у голосовой щели равен

$$r_G = \left[\frac{Z_G - \rho c/A_1}{Z_G + \rho c/A_1}\right]. \tag{3.50}$$

Выражение (3.49) изображено в виде схемы на рис. 3.36. Так же, как и в случае излучения, если Z_G комплексная величина, то (3.50) остается справедливым. Однако r_G будет комплексным и (3.49) следует заменить аналогичным соотношением в частотной области или $u^+(t)$ и $u_G(t)$, $u_1^-(t)$ могут быть связаны дифференциальным уравнением. Обычно для упрощения полагают, что Z_G и Z_L — действительные величины.

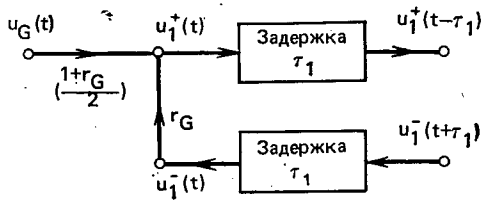


Рис. 3.36. Нагрузка со стороны голосовой щели

В качестве примера на рис. 3.37 приведена полная схема распространения волны в системе из двух труб. Скорость потока около губ обозначена через $u_L(t) = u_2(l_2, t)$. В частотной области

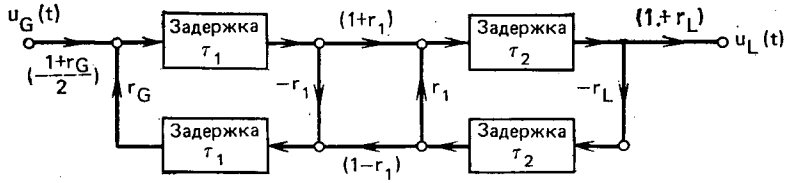


Рис. 3.37. Полный граф двухтрубной модели

можно получить (см. задачу 3.7) выражение для частотной характеристики

$$V_a(\Omega) = \frac{U_L(\Omega)}{U_G(\Omega)} = \frac{0{,}5\,(1+r_G)\,(1+r_L)\,(1+r_1)\,e^{-i\Omega(\tau_1+\tau_2)}}{1 + r_1 r_G e^{-i\Omega 2\tau_1} + r_1 r_L e^{-i\Omega 2\tau_2} + r_L r_G e^{-i\Omega 2(\tau_1+\tau_2)}}. \tag{3.51}$$

Отметим несколько особенностей (3.51). Во-первых, в числителе имеется коэффициент $e^{-i\Omega(\tau_1+\tau_2)}$. Он отображает задержку прохождения потока от голосовой щели до губ. Системную функцию можно найти путем замены $i\Omega$ на s:

$$V_a(s) = \frac{0{,}5\,(1+r_G)\,(1+r_L)\,(1+r_1)\,e^{-s(\tau_1+\tau_2)}}{1 + r_1 r_G e^{-s2\tau_1} + r_1 r_L e^{-s2\tau_2} + r_L r_G e^{-s2(\tau_1+\tau_2)}}. \tag{3.52}$$

Полюса $V_a(s)$ являются комплексными резонансными частотами системы. Видно, что количество полюсов бесконечно, так как зависимость от s экспоненциальная. В работах Фанта [1] и Фланагана [2] показано, что можно получить хорошее соответствие резонансных частот модели реальным формантным частотам гласных, если подобрать длины труб и площади их поперечного сечения (см. также задачу 3.8).

3.3.3. Связь с цифровыми фильтрами

Анализ $V_a(s)$ показывает, что модель из двух труб без потерь имеет свойства, близкие к цифровым фильтрам. Чтобы убедиться в этом, рассмотрим систему, состоящую из N труб без потерь длиной $\Delta x = l/N$, где l — длина голосового тракта. Такая система изображена на рис. 3.38 для $N=7$. Распространение волн в такой системе можно представить схемой, изображенной на рис. 3.34, с задержками на интервал $\tau = \Delta x/c$, равный времени прохождения волны через одну трубу. Изучение удобно начинать с анализа отклика системы на возбуждение в виде единичного импульса $u_G(t) = \delta(t)$. Импульс возбуждения проходит через все трубы, частично отражаясь от мест их соединений. Подробный анализ распространения импульса показывает, что импульсная характеристика системы (скорость потока около губ при импульсном возбуждении со стороны голосовой щели)

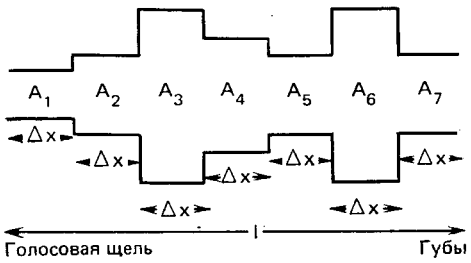


Рис. 3.38. Соединение ($N=7$) труб без потерь одинаковой длины

$$v_a(t) = \alpha_0 \delta(t - N\tau) + \sum_{k=1}^{\infty} \alpha_k \delta(t - N\tau - 2k\tau). \qquad (3.53)$$

Очевидно, что импульс может появиться на выходе только через $N\tau$ с. Импульсы, отраженные от мест соединений, достигнут выхода на 2τ с позже. Величина 2τ — время распространения импульса вдоль одной трубы в прямом и обратном направлениях. Системная функция имеет вид

$$V_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-s(N+2k)\tau} = e^{-sN\tau} \sum_{k=0}^{\infty} \alpha_k e^{-s2\tau k}. \qquad (3.54)$$

Коэффициент $e^{-sN\tau}$ отображает задержку, т. е. время распространения импульса вдоль всех N секций. Выражение

$$\hat{V}_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-sk2\tau} \qquad (3.55)$$

является передаточной функцией линейной системы, импульсная характеристика которой равна $\hat{v}_a(t) = v_a(t+N\tau)$. Эта функция определяет резонансные свойства системы. На рис. 3.39*а* приведена структурная схема модели с трубами без потерь, в которой задержка отделена от системы $\hat{v}_a(t)$. Частотная характеристика $\hat{V}_a(\Omega)$ равна

$$\hat{V}_a(\Omega) = \sum_{k=0}^{\infty} \alpha_k e^{-i\Omega k 2\tau}. \qquad (3.56)$$

Легко показать, что

$$\hat{V}_a(\Omega + 2\pi/2\tau) = \hat{V}_a(\Omega). \qquad (3.57)$$

Очевидно, что эта характеристика напоминает частотную характеристику системы в дискретном времени. Действительно, если спектр входного сигнала сосредоточен в диапазоне ниже $\pi/(2\tau)$, то можно дискретизировать входной сигнал с периодом $T=2\tau$ и затем пропустить сигнал через цифровой фильтр с импульсной характеристикой

$$v(n) = \begin{cases} \alpha_n, & n \geqslant 0, \\ 0, & n < 0. \end{cases} \qquad (3.58)$$

Для периода $T=2\tau$ время задержки на $N\tau$ с соответствует сдвигу на $N/2$. Эквивалентная система в дискретном времени для входного сигнала с ограниченным по частоте спектром показана на рис. 3.39*б*. Отметим, что если N — четное, то $N/2$ — целое число и задержка может быть реализована как простой сдвиг выходной последовательности первой системы. Если N нечетное, то для получения выходного сигнала в схеме рис. 3.39*а* необходимо проводить интерполяцию. Однако задержкой можно пренебречь, так как в большинстве приложений она не имеет существенного значения.

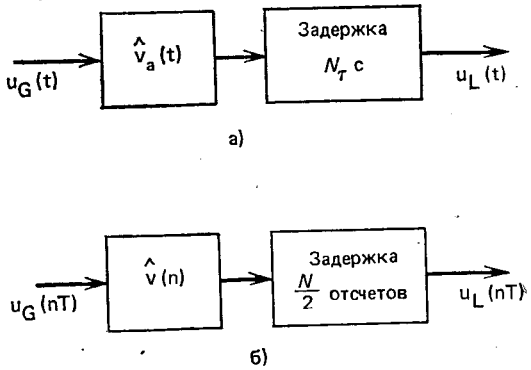


Рис. 3.39. Модель акустической трубы без потерь (*а*) и эквивалентная дискретная система (*б*)

Выполнив замену e^{sT} на z в $\hat{V}_a(s)$, получим z-преобразование $\hat{v}(n)$:

$$\hat{V}(z) = \sum_{k=0}^{\infty} \alpha_k z^{-k}. \qquad (3.59)$$

Граф прохождения сигнала в системе с дискретным временем можно построить по графу аналоговой системы. Каждую переменную аналоговой системы надо заменить на соответствующую последовательность отсчетов. Кроме того, каждую задержку на τ с надо заменить элементом задержки на половину периода дискретизации, так как $\tau = T/2$. Пример такого графа показан на рис. 3.40. Отметим, что задержка отображена на рис. 3.40*б* ветвью с коэффициентом передачи $z^{-1/2}$.

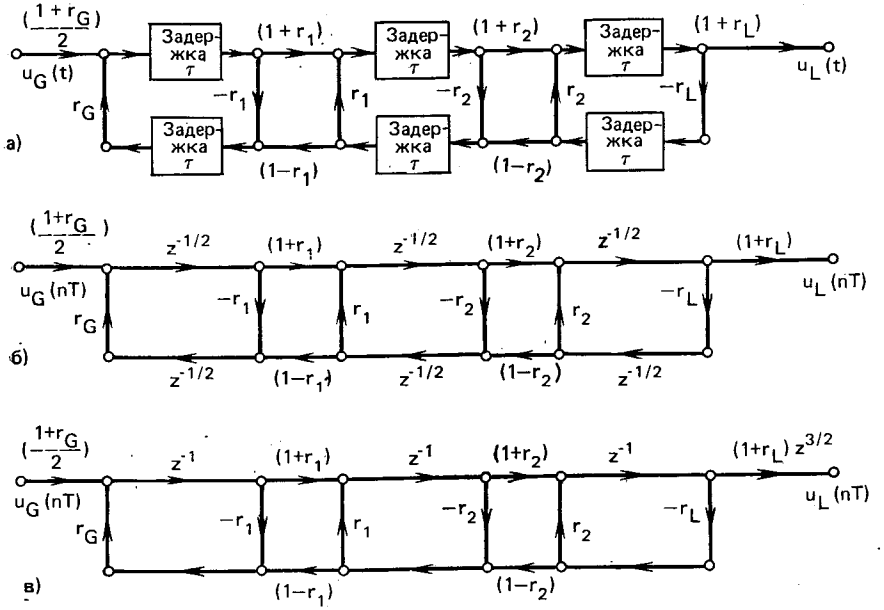


Рис. 3.40. Граф модели голосового тракта (*а*), эквивалентная дискретная система (*б*) и эквивалентная дискретная система с объединенными элементами задержки (*в*)

Задержка на половину периода дискретизации (рис. 3.40*б*) предполагает интерполяцию сигнала между двумя отсчетами. Такая интерполяция не может быть реализована непосредственно. Можно получить более удобную структуру, если использовать то обстоятельство, что схема рис. 3.40*б* имеет лестничную форму с элементами задержки в верхних и нижних ветвях. Сигнал распространяется в прямом направлении (вправо) в верхних ветвях и в обратном направлении (влево) — в нижних ветвях. Из рисунка видно, что задержка внутри замкнутого пути сохранится, если задержку в нижней ветви объединить с задержкой в верхней ветви. Общая задержка прохождения сигнала от входа к выходу при этом изменится, однако это не имеет большого значения на практике, а теоретически это изменение может быть устранено введением корректирующего упреждения[1] $z^{N/2}$. На рис. 3.40*в* показано, как эти рассуждения можно реализовать на примере модели из трех труб. Достоинством такой структуры является то, что для этой системы могут быть записаны разностные уравнения, которые позволяют путем итераций получить выходной сигнал при заданном входном.

Цифровые системы [33], подобные изображенной на рис. 3.40*в*, могут быть использованы для формирования отсчетов синтезированного речевого сигнала по отсчетам сигнала возбуждения [32]. Система такой структуры является довольно сложной. Каждая ветвь, коэффициент передачи которой не равен единице, предполагает выполнение операции умножения. Таким образом, в каждой секции требуется выполнить четыре операции умножения и две — сложения. В общем для модели с N трубами (рис. 3.40*в*) надо выполнять $4N$ операций умножения и $2N$ — сложения. Так как выполнение операции умножения требует много времени, представляет интерес разработка других структур (или другой организации вычислений), которые требуют выполнения меньшего числа операций умножения. Такие структуры могут быть получены в результате подробного анализа схемы рис. 3.41*а*. Разностные уравнения для этой схемы имеют вид

$$u^+(n) = (1+r)\,w^+(n) + r\,u^-(n); \qquad (3.60а)$$

$$w^-(n) = -r\,w^+(n) + (1-r)\,u^-(n). \qquad (3.60б)$$

Эти уравнения можно переписать:

$$u^+(n) = w^+(n) + r\,w^+(n) + r\,u^-(n); \qquad (3.61а)$$

$$w^-(n) = -r\,w^+(n) - r u^-(n) + u^-(n). \qquad (3.61б)$$

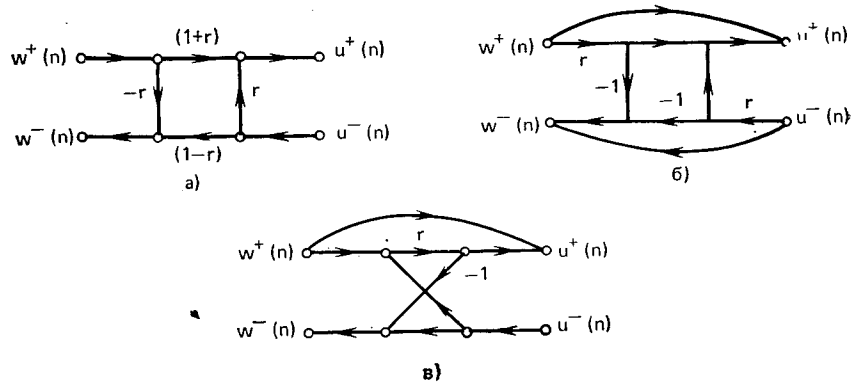


Рис. 3.41. Модель с четырьмя операциями умножения (*а*), двумя операциями умножения (*б*) и одной операцией умножения (*в*)

[1] Все задержки можно объединить в нижних ветвях. Тогда потребуется ввести корректирующую задержку на $N/2$.

Так как члены $rw^+(n)$ и $rw^-(n)$ входят в оба уравнения, два выхода четырех ветвей с операциями умножения можно устранить так, как это сделано на рис. 3.41*б*. В полученной структуре имеется две операции умножения и четыре — сложения. Другую структуру можно получить, группируя слагаемые с коэффициентом r следующим образом:

$$u^+(n) = w^+(n) + r[w^+(n) + u^-(n)]; \tag{3.62а}$$

$$w^-(n) = u^-(n) - r[w^+(n) + u^-(n)]. \tag{3.62б}$$

Так как слагаемое $r[w^+(n)+u^-(n)]$ входит в оба уравнения, здесь требуется выполнить одну операцию умножения и три — сложения (рис. 3.41*в*). Такая форма представления модели с трубами без потерь впервые была получена Итакурой и Саито [34]. При использовании модели с трубами без потерь для синтеза речевых сигналов выбор вычислительной схемы зависит от скорости, с которой могут быть выполнены операции умножения и сложения, а также от простоты выполнения контроля над вычислениями.

3.3.4. Передаточная функция модели с трубами без потерь

Для полного завершения изучения модели с трубами без потерь в дискретном времени полезно получить общее выражение для передаточной функции через коэффициенты отражения. Рассматриваемые далее уравнения впервые были получены Аталом и Ханауером [35], Маркелом, Греем [36] и Вакитой [37] при исследовании анализа речи с помощью линейного предсказания. В гл. 8 будет рассмотрена взаимосвязь модели с трубами без потерь с моделью линейного предсказания. Основными вопросами данного раздела являются вывод общего выражения для передаточной функции модели с трубами без потерь и выявление вытекающих из этой модели разнообразных представлений сигнала.

Передаточная функция определена выражением

$$V(z) = U_L(z)/U_G(z). \tag{3.63}$$

Таким образом, для получения $V(z)$ достаточно выразить $U_G(z)$ через $U_L(z)$ и записать их отношение. С этой целью рассмотрим рис. 3.42, на котором показана одна секция модели трубы без потерь. Уравнения в z-плоскости для этой секции имеют вид

$$U^+_{k+1}(z) = (1+r_k)\, z^{-1/2} U^+_k(z) + r_k U^-_{k+1}(z); \tag{3.64а}$$

$$U^-_k(z) = -r_k z^{-1} U^+_k(z) + (1-r_k)\, z^{-1/2} U^-_{k+1}(z). \tag{3.64б}$$

Решая эти уравнения относительно $U^+_k(z)$, $U^-_k(z)$, получим

$$U^+_k(z) = \frac{z^{1/2}}{1+r_k} U^+_{k+1}(z) - \frac{r_k z^{1/2}}{1+r_k} U^-_{k+1}(z); \tag{3.65а}$$

$$U^-_k(z) = \frac{-r_k z^{-1/2}}{1+r_k} U^+_{k+1}(z) + \frac{z^{-1/2}}{1+r_k} U^-_{k+1}(z). \tag{3.65б}$$

Уравнения (3.65) позволяют выразить $U_G(z)$ через $U_L(z)$ на основе связи между выходом и входом модели с трубами без потерь.

Для получения результата в компактной форме удобно записать граничные условия около губ в том же виде, что и для стыков между трубами. Чтобы сделать это, введем $U_{N+1}(z)$ как z-

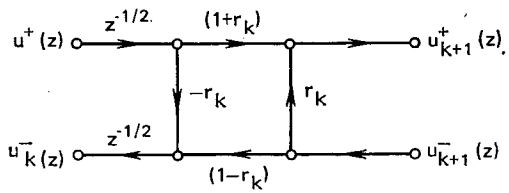


Рис. 3.42. Граф для z-преобразований

преобразования входного сигнала фиктивной $(N+1)$-й трубы, которая имеет бесконечную длину, так что отраженная волна в ней отсутствует. Можно рассуждать и иначе, полагая, что нагрузкой $(N+1)$-й трубы является характеристическое сопротивление. В обоих случаях $U^+_{N+1}(z) = U_L(z)$ и $U^-_{N+1}(z) = 0$. Тогда из (3.39) и (3.45) видно, что если $A_{N+1} = \rho c/Z_L$, то можно положить $r_N = r_L$.

Уравнение (3.65) можно записать в матричной форме:

$$\mathbf{U}_k = \mathbf{Q}_k \mathbf{U}_{k+1}, \tag{3.66}$$

где

$$\mathbf{U}_k = \begin{Vmatrix} U^+_k(z) \\ U^-_k(z) \end{Vmatrix} \tag{3.67}$$

и

$$\mathbf{Q}_k = \begin{Vmatrix} \dfrac{z^{1/2}}{1+r_k} & \dfrac{-r_k z^{1/2}}{1+r_k} \\ \dfrac{-r_k z^{-1/2}}{1+r_k} & \dfrac{z^{-1/2}}{1+r_k} \end{Vmatrix}. \tag{3.68}$$

Итерируя (3.66), можно показать, что входная переменная первой трубы может быть выражена через переменную на выходе с помощью матричного произведения:

$$\mathbf{U}_1 = \mathbf{Q}_1 \mathbf{Q}_2 \ldots \mathbf{Q}_N U_{N+1} = \prod_{k=1}^{N} \mathbf{Q}_k U_{N+1}. \tag{3.69}$$

Из рис. 3.36 видно, что граничные условия со стороны голосовой щели можно записать в виде

$$U_G(z) = \frac{2}{(1+r_G)} U^+_1(z) - \frac{2r_G}{1+r_G} U^-_1(z). \tag{3.70}$$

Это выражение можно переписать:

$$U_G(z) = \left\| \frac{2}{1+r_G}, \; -\frac{2r_G}{1+r_G} \right\| \mathbf{U}_1. \tag{3.71}$$

Так как

$$\mathbf{U}_{N+1}=\left\|\begin{matrix}U_L(z)\\0\end{matrix}\right\|=\left\|\begin{matrix}1\\0\end{matrix}\right\|U_L(z), \tag{3.72}$$

и окончательно можно записать

$$\frac{U_G(z)}{U_L(z)}=\left\|\frac{2}{1+r_G},\ -\frac{2r_G}{1+r_G}\right\|\prod_{k=1}^{N}\mathbf{Q}_k\left\|\begin{matrix}1\\0\end{matrix}\right\|. \tag{3.73}$$

Это решение равно $1/V(z)$.

Для изучения свойств $V(z)$ полезно записать

$$\mathbf{Q}_k=z^{1/2}\left\|\begin{matrix}\dfrac{1}{1+r_k} & \dfrac{-r_k}{1+r_k}\\[2ex] \dfrac{-r_k z^{-1}}{1+r_k} & \dfrac{z^{-1}}{1+r_k}\end{matrix}\right\|=z^{1/2}\hat{\mathbf{Q}}_k. \tag{3.74}$$

Выражение (3.73) можно записать в форме

$$\frac{1}{V(z)}=z^{N/2}\left\|\frac{2}{1+r_G},\ -\frac{2r_G}{1+r_G}\right\|\prod_{k=1}^{N}\hat{\mathbf{Q}}_k\left\|\begin{matrix}1\\0\end{matrix}\right\|. \tag{3.75}$$

Так как элементы $\hat{\mathbf{Q}}_k$ либо постоянные величины, либо пропорциональны z^{-1}, то полное матричное произведение можно представить в виде полинома от z^{-1} степени N. Например, можно показать (см. задачу 3.9), что для $N=2$ справедливо выражение

$$\frac{1}{V(z)}=\frac{2\left(1+r_1r_2z^{-1}+r_1r_Gz^{-1}+r_2r_Gz^{-2}\right)z}{(1+r_G)(1+r_1)(1+r_2)} \tag{3.76}$$

или

$$V(z)=\frac{0{,}5(1+r_G)(1+r_1)(1+r_2)z^{-1}}{1+(r_1r_2+r_1r_G)z^{-1}+r_2r_Gz^{-2}}. \tag{3.77}$$

Из (3.74) и (3.75) видно, что в общем случае передаточная функция модели с трубами без потерь может быть записана как

$$V(z)=\frac{0{,}5(1+r_G)\prod_{k=1}^{N}(1+r_k)z^{-N/2}}{D(z)}, \tag{3.78а}$$

где $D(z)$ — полином переменной z^{-1} — задается матрицей

$$\mathrm{D}(z)=\|1,\ -r_G\|\left\|\begin{matrix}1 & -r_1\\ -r_1z^{-1} & z^{-1}\end{matrix}\right\|\cdots\left\|\begin{matrix}1 & -r_N\\ -r_Nz^{-1} & z^{-1}\end{matrix}\right\|\left\|\begin{matrix}1\\0\end{matrix}\right\|. \tag{3.78б}$$

Из (3.78б) для $D(z)$ можно получить

$$D(z)=1-\sum_{k=1}^{N}\alpha_kz^{-k}. \tag{3.79}$$

Другими словами, передаточная функция модели с трубами без потерь имеет столько элементов задержки, сколько секций в модели. Передаточная функция имеет полюса и не имеет нулей. Эти полюса определяют резонансы или форманты модели.

В специальном случае $r_G=1$ ($Z_G=\infty$) полином $D(z)$ можно найти по рекурсивной формуле, вытекающей из (3.78б). Если вычислять матричное произведение слева, то вначале надо вычислить произведение матрицы-строки размером 1×2 и матрицы размером 2×2, а в конце вычислений умножить результат на матрицу-столбец размером 2×1 в правой части (3.78б). Рекурсивная формула может быть получена путем вычисления первых нескольких матричных произведений. Введем

$$\mathbf{P}_1=\|1,-1\|\left\|\begin{matrix}1 & -r_1\\ -r_1z^{-1} & z^{-1}\end{matrix}\right\|=\|(1+r_1z^{-1}),-(r_1+z^{-1})\|. \tag{3.80}$$

Если далее ввести

$$D_1(z)=1+r_1z^{-1}, \tag{3.81}$$

то легко показать, что

$$\mathbf{P}_1=\|D_1(z),-z^{-1}D_1(z^{-1})\|. \tag{3.82}$$

Аналогично введем матрицу-строку

$$\mathbf{P}_2=\mathbf{P}_1\left\|\begin{matrix}1 & -r_2\\ -r_2z^{-1} & z^{-1}\end{matrix}\right\|. \tag{3.83}$$

Если выполнить умножение, то можно показать, что

$$\mathbf{P}_2=\|D_2(z),\quad -z^{-2}D_2(z^{-1})\|, \tag{3.84}$$

где

$$D_2(z)=D_1(z)+r_2z^{-2}D_1(z^{-1}). \tag{3.85}$$

По индукции получаем

$$\mathbf{P}_k=\mathbf{P}_{k-1}\left\|\begin{matrix}1 & -r_k\\ -r_kz^{-1} & z^{-1}\end{matrix}\right\|=\|D_k(z),\quad -z^{-k}D_k(z^{-1})\|. \tag{3.86}$$

где

$$D_k(z)=D_{k-1}(z)+r_kz^{-k}D_{k-1}(z^{-1}). \tag{3.87}$$

Окончательно

$$D(z)=\mathbf{P}_N\left\|\begin{matrix}1\\0\end{matrix}\right\|=D_N(z). \tag{3.88}$$

Таким образом, нет необходимости вычислять все матричные произведения. Достаточно выполнить вычисления по рекурсивной формуле:

$$D_0(z)=1; \tag{3.89а}$$

$$D_k(z)=D_{k-1}(z)+r_kz^{-k}D_{k-1}(z^{-1}),\quad k=1,\ 2,\ \ldots,N; \tag{3.89б}$$

$$D(z)=D_N(z). \tag{3.89в}$$

Эффективность модели с трубами без потерь можно продемонстрировать путем вычисления передаточной функции для функций площадей поперечного сечения, использованных для получения рис. 3.23—3.26. При выполнении этого расчета нужно учесть нагрузку со стороны губ и количество секций в модели. В наших определениях нагрузка излучения была представлена трубой с площадью сечения A_{N+1}, в которой отсутствует отраженная волна. Значение A_{N+1} выбирается из соображений получения требуемого коэффициента отражения на выходе. Влияние излучения является единственным источником потерь в системе (если $r_G=1$), и, таким образом, можно ожидать, что выбором величины A_{N+1} можно изменять ширину резонансных областей $V(z)$. Например, при $A_{N+1}=\infty$ имеем $r_N=r_L=1$, т. е. коэффициент отражения соответствует акустическому короткому замыканию. Этот случай соответствует полному отсутствию потерь. Обычно A_{N+1} выбирают так, чтобы получить коэффициент отражения со стороны губ, при котором получается требуемая ширина резонансных областей. Пример подобного расчета излагается ниже.

Выбор числа секций зависит от частоты дискретизации речевого сигнала. Пусть частотная характеристика модели с трубами без потерь периодическая. В этом случае модель правильно отображает поведение голосового тракта только в диапазоне частот $|F|<1/(2T)$, где T — период дискретизации. Ранее было установлено, что $T=2\tau$, где τ — время распространения звуковой волны через одну секцию в одном направлении. Если имеется N секций общей длиной l, то $\tau=l/(cN)$. Так как знаменатель передаточной функции является полиномом степени N, существует $N/2$ комплексных полюсов, соответствующих резонансам в полосе частот $|F|<1/(2T)$. Если принять $l=17{,}5$, $c=35\,000$ см/с, то получаем

$$1/2T = 1/4\tau = NC/4l = 1000\,N/2 \text{ Гц}. \qquad (3.90)$$

Это означает, что в полосе частот до 1000 Гц в голосовом тракте длиной 17,5 см будет $N/2$ резонансов (формант). Если $1/T=10\,000$ Гц, то общая полоса частот составляет 5000 Гц. Это означает, что N должно быть равно 10. Просматривая рис. 3.21—3.26, можно сделать вывод о том, что резонансы голосового тракта расположены с плотностью примерно одна форманта на 1000 Гц. Голосовому тракту меньшей длины будет соответствовать меньшая плотность расположения резонансов и наоборот.

На рис. 3.43 приведен пример для $N=10$ и $1/T=10$ кГц. На рис. 3.43*а* показаны значения функций площадей поперечного сечения, взятые из рис. 3.23 и дискретизированные для получения десятисекционной модели гласного /*a*/. На рис. 3.43*б* показаны десять коэффициентов отражения для $A_{11}=30$ см^2. При этом коэффициент отражения около губ был равен $r_N=0{,}714$. Отметим, что наибольшая величина коэффициента отражения соответствует месту наибольшего изменения функции площади. На рис. 3.43*в* показана частотная характеристика для $r_N=1$ и 0,714 (пунктирная линия). Сравнивая пунктирную линию рис. 3.43*в* с рис. 3.23, можно

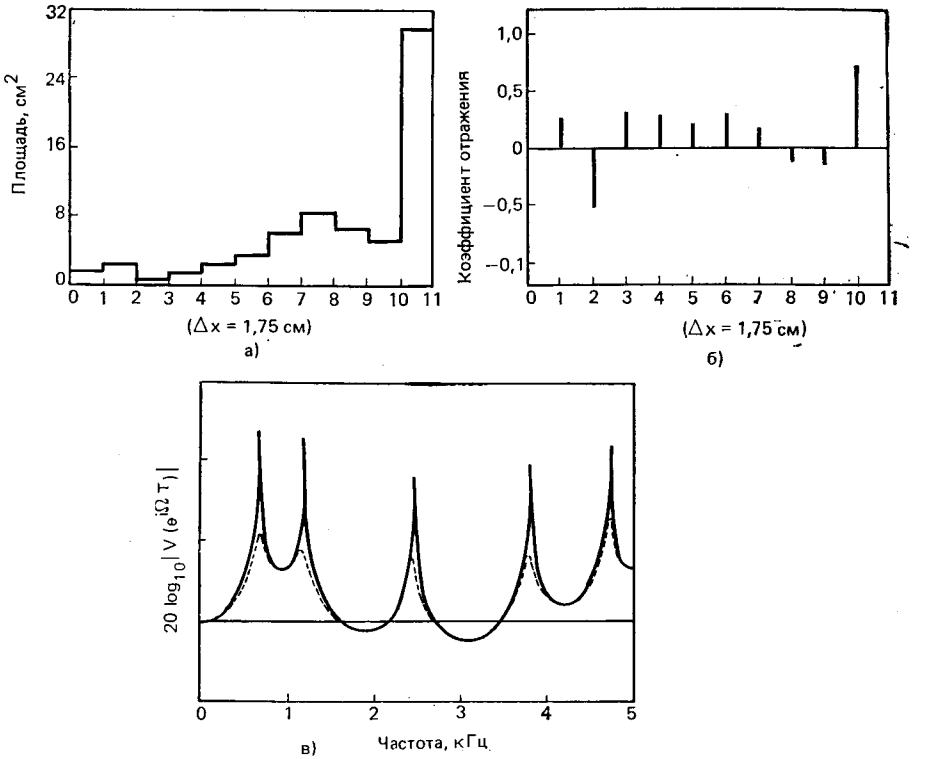


Рис. 3.43. Функция площади десятисекционной трубы без потерь, нагруженной трубой с площадью 30 см^2 (*а*); коэффициенты отражения десятисекционной трубы (*б*); частотная характеристика десятисекционной трубы (*в*). Пунктирная линия соответствует условиям (*б*); сплошная линия соответствует короткому замыканию. Данные (*а*) взяты из [1] для гласного /*a*/ русского языка

сделать вывод, что при наличии потерь на излучение частотная характеристика модели с трубами без потерь очень близка к характеристике сложной модели с потерями.

3.4. Цифровые модели речевых сигналов

В § 3.2 было показано, что можно получить довольно подробное математическое описание акустического процесса речеобразования. Наша цель при изучении этой теории состоит в выяснении основных особенностей речевого сигнала и в установлении того, ка́к эти особенности согласуются с физикой речеобразования. Было рассмотрено три способа возбуждения звуков и выяснено, что каждому способу соответствует свой тип выходного сигнала. Установлено также, что сигнал возбуждения проходит через голосовой тракт и преобразуется в нем в соответствии с резонансами тракта, образуя звуки речи. Это явление составляет предмет дальнейшего изучения.

Из проведенного обстоятельного рассмотрения моделей речеобразования можно сделать важный вывод. Уже ясно, что для общего описания речевых сигналов можно разработать «эквивалентную модель», такую, как показано на рис. 3.31. Здесь линейная система, выходной сигнал которой обладает свойствами, близкими к свойствам речи, управляется множеством параметров примерно так, как это происходит при речеобразовании. Таким образом, выходной сигнал этой модели эквивалентен выходному сигналу физической модели, но ее внутренняя структура не связана с физикой речеобразования. Представляет интерес вопрос о построении подобных эквивалентных моделей в дискретном времени для описания дискретизированных речевых сигналов. Для формирования речеподобного сигнала тип возбуждения и резонансные свойства линейной системы должны изменяться во времени. Характер этих изменений рассматривался в § 3.1. В частности, колебания (см. рис. 3.3*а*) показывают, что свойства сигнала изменяются довольно медленно. Для многих звуков речи можно считать, что тип возбуждения и свойства голосового тракта остаются неизменными в течение 10—20 мс. Таким образом, эквивалентная аналоговая модель должна состоять из линейной системы с медленно изменяющимися во времени параметрами, возбуждаемой сигналом от источника возбуждения. Сигнал возбуждения представляет собой квазипериодическую последовательность импульсов для вокализованной речи и шумоподобный процесс для невокализованной речи.

Примером такой подходящей модели является модель с трубами без потерь в дискретном времени, которая рассматривалась выше. Основные особенности этой модели отображены на рис. 3.44*а*. Пусть система, отображающая голосовой тракт, описывается

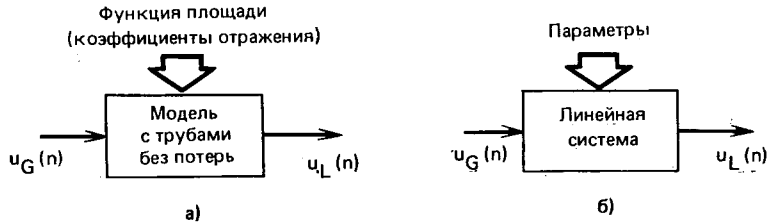


Рис. 3.44. Диаграмма модели с трубами без потерь (*а*) и эквивалентная модель (*б*)

набором площадей или коэффициентами отражения. Системы вида рис. 3.40*в* могут быть применены для формирования речевого сигнала по соответствующему возбуждению. Соотношение между входным и выходным сигналами может быть записано через передаточную функцию

$$V(z) = G \Big/ \left(1 - \sum_{k=1}^{N} \alpha_k z^{-k}\right), \tag{3.91}$$

где G и $\{\alpha_k\}$ зависят от функции площади (постоянная задержка

из (3.78а) здесь опущена). Нас интересует только выходной сигнал, поэтому любая система, которая обладает данной передаточной функцией, способна формировать такой выходной сигнал по заданному воздействию. (Это не вполне справедливо для систем с переменными параметрами, но возможные различия можно свести к минимуму.) Таким образом, эквивалентные модели в дискретном времени в общем виде имеют структуру рис. 3.44*б*. Это указывает на возможность описания голосового тракта с помощью фильтра.

Для получения полной модели речевого сигнала необходимо также учесть изменение звукового возбуждения и эффект излучения через губы. В заключительной части этой главы будут рассмотрены отдельно составляющие такой полной модели.

3.4.1. Голосовой тракт

Резонансы (форманты) речевого сигнала соответствуют полюсам передаточной функции $V(z)$. Полюсная модель дает хорошее описание голосового тракта для большинства звуков речи. Однако акустическая теория показывает, что для носовых и фрикативных звуков надо учитывать и резонансы и антирезонансы (полюса и нули). В этом случае в передаточную функцию следует ввести нули или, следуя работе Атала [35], полагать, что наличие нулей можно учесть, увеличивая количество полюсов (см. задачу 3.10). Последнее соображение применяется наиболее часто.

Так как коэффициенты знаменателя $V(z)$ в (3.91) действительные, корни полинома в знаменателе могут быть действительными или комплексно-сопряженными. Типичная комплексная резонансная частота голосового тракта имеет вид

$$s_k, s_k^* = -\sigma_k \pm \mathrm{i}\, 2\pi F_k. \tag{3.92}$$

Соответствующая пара комплексно-сопряженных полюсов в модели с дискретным временем может быть записана так:

$$z_k,\ z_k^* = \mathrm{e}^{-\sigma_k T}\, \mathrm{e}^{\pm \mathrm{i}\, 2\pi F_k T} = \mathrm{e}^{-\sigma_k T} \cos(2\pi F_k T) \pm \mathrm{i}\, \mathrm{e}^{-\sigma_k T} \sin(2\pi F_k T). \tag{3.93}$$

Ширина резонансной области голосового тракта равна примерно $2\sigma_k$, а центральная частота — $2\pi F_k$ [26]. В z-плоскости длина линии от начала координат до координаты полюса определяет ширину резонансной области:

$$|z_k| = \mathrm{e}^{-\sigma_k T}, \tag{3.94а}$$

а угол наклона линии определяется соотношением

$$\theta_k = 2\pi F_k T. \tag{3.94б}$$

Таким образом, если знаменатель $V(z)$ факторизован, то значения формантных частот и ширины формантных областей могут быть найдены по (3.94). Как показано на рис. 3.45, комплексные частоты голосового тракта расположены в левой полуплоскости s,

так как голосовой тракт является устойчивой системой. Следовательно, $\sigma_k > 0$ и $|z_k| < 1$. Это означает, что все полюса модели в дискретном времени располагаются внутри единичного круга. На рис. 3.45 показаны типичные комплексные резонансные частоты в s- и z-плоскостях.

В § 3.3 было показано, как для модели с трубами без потерь получить передаточную функцию (3.91). Можно показать [35, 36], что если площади поперечного сечения труб положительные, все полюса $V(z)$ будут находиться внутри единичного круга. Обратно, по $V(z)$ из (3.91) можно определить модель с трубами без потерь [35, 36]. Один из путей реализации требуемой передаточной функции состоит в использовании лестничной структуры рис. 3.40*в*, одну секцию которой можно представить в виде рис. 3.41. Другим способом реализации передаточной функции является применение описанных в гл. 2 структур, используемых в обычных цифровых фильтрах. Например, можно применить прямую форму реализа-

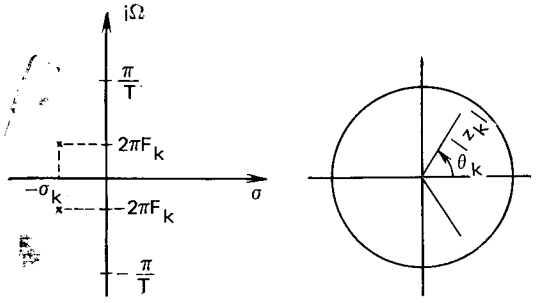


Рис. 3.45. Резонансы голосового тракта: *а*) в s-плоскости; *б*) в z-плоскости

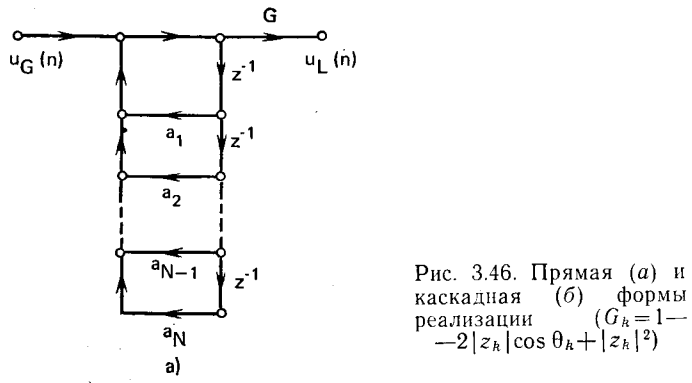


Рис. 3.46. Прямая (*а*) и каскадная (*б*) формы реализации ($G_k = 1 - 2|z_k|\cos\theta_k + |z_k|^2$)

ции (рис. 3.46*а*); можно использовать и каскадное соединение систем второго порядка (резонаторов), для которых

$$V(z) = \prod_{k=1}^{M} V_k(z), \tag{3.95}$$

где M равно $((N+1)/2)$, округленному до большего целого, и

$$V_k(z) = \frac{(1 - 2|z_k|\cos(2\pi F_k T) + |z_k|^2)}{(1 - 2|z_k|\cos(2\pi F_k T)\, z^{-1} + |z_k|^2 z^{-2}}. \tag{3.96}$$

Числитель $V_k(z)$ следует выбирать так, чтобы коэффициент усиления фильтра совпадал с коэффициентом усиления модели с трубами без потерь. Заметим, что на нулевой частоте ($z=1$) $V_k(1) = 1$. Каскадная модель показана на рис. 3.46*б*. В задаче 3.11 указан путь уменьшения числа операций умножения в такой модели. Другой способ реализации $V(z)$ основан на ее разложении на простые дроби и построении параллельной структуры. Этот способ рассматривается в задаче 3.12.

Интересно отметить, что каскадные и параллельные структуры вначале использовались как аналоговые модели. Они обладают существенным ограничением: аналоговые системы второго порядка (резонаторы) имеют частотные характеристики, быстро убывающие с ростом частоты. Это потребовало введения звеньев «коррекции высокочастотных полюсов», соединенных последовательно с формантными резонаторами для обеспечения необходимого наклона частотной характеристики на высоких частотах. В результате цифрового моделирования Голд и Рабинер [38] показали, что цифровые резонаторы, благодаря присущим им свойствам периодичности частотных характеристик, обеспечивают требуемый наклон характеристики на высоких частотах. В этом можно убедиться на примере модели с трубами без потерь. Таким образом, в цифровых моделях нет необходимости вводить цепи коррекции.

3.4.2. Излучение

Ранее рассматривалась передаточная функция $V(z)$, описывающая взаимосвязь скорости потока источника и скорости выходного потока около губ. Если нужно получить модель, описывающую характер звукового давления около губ (как это обычно и требуется), необходимо учитывать эффект излучения. В 3.2.4 показано, что в аналоговых моделях звуковое давление и скорость потока связаны уравнением (3.29). Аналогичное соотношение в z-плоскости имеет вид

$$P_L(z) = R(z)\, U_L(z). \tag{3.97}$$

Из 3.2.4 и рис. 3.20 можно сделать вывод, что звуковое давление связано со скоростью воздушного потока операцией высокочастотной фильтрации. В частности, можно считать, что на низких частотах звуковое давление примерно равно производной скорости

потока. Для представления этой взаимосвязи в дискретном времени необходимо использовать методы цифровой техники, позволяющие избежать явления «наложения частот». Например, метод билинейного преобразования (обычный метод проектирования цифровых фильтров) обеспечивает достаточно хорошее описание эффекта излучения (см. задачу 3.13) при

$$R(z) = R_0(1 - z^{-1}), \tag{3.98}$$

т. е. при вычислении разности первого порядка (более точное приближение рассматривается в задаче 3.13). «Грубое» дифференцирование путем вычисления разности первого порядка соответствует обычно применяемому «грубому» описанию связи давления со скоростью потока посредством операции дифференцирования на низких частотах.

Эта модель излучения может быть включена последовательно с моделью голосового тракта (рис. 3.47). Передаточную функцию

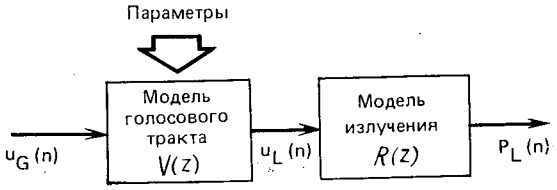


Рис. 3.47. Эквивалентная модель речеобразования с учетом излучения

$V(z)$ можно реализовать любым подходящим способом с параметрами, соответствующими требуемой конфигурации голосового тракта: функции площади сечения модели с трубами без потерь или формантным областям каскадной модели.

3.4.3. Возбуждение

Для завершения построения эквивалентной модели необходимо изучить способы описания возбуждения голосового тракта. Будем считать, что большинство звуков речи можно отнести либо к вокализованным, либо к невокализованным. В первом случае источник возбуждения должен формировать квазипериодическую последовательность импульсов, а во втором — шумоподобное случайное колебание.

Для синтеза вокализованной речи сигнал возбуждения должен иметь вид, изображенный на рис. 3.30. Один из способов получения такого сигнала показан на рис. 3.48. Генератор последовательности импульсов формирует единичные импульсы, повторяющиеся через период основного тона. Этот сигнал поступает на линейную систему, импульсная характеристика которой $g(n)$ соответствует форме колебания в голосовой щели. Коэффициент усиления A_v определяет интенсивность голосового возбуждения.

Форма $g(n)$ несущественна и требуется только, чтобы преобразование Фурье от $g(n)$ обладало «правильными» свойствами. Розенберг [39], исследуя влияние формы импульса возбуждения на

качество восприятия речевого сигнала, выяснил, что импульс голосового возбуждения может иметь следующий вид:

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/N_1)], & 0 \leqslant n \leqslant N_1; \\ \cos(\pi(n - N_1)/2N_2), & N_1 \leqslant n \leqslant N_1 + N_2; \\ 0, & \textit{в противном случае.} \end{cases} \tag{3.99}$$

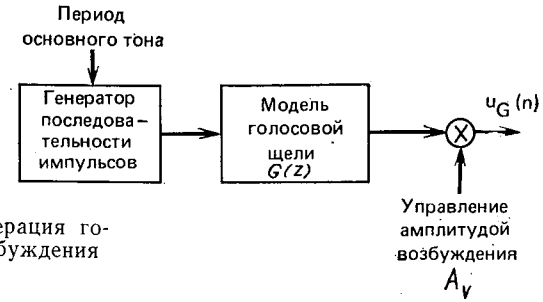


Рис. 3.48. Генерация голосового возбуждения

Этот импульс очень напоминает импульсы, изображенные на рис. 3.30. На рис. 3.49 показаны импульсное колебание и его амплитудный спектр для типичных значений N_1 и N_2. Из рисунка видно, что, как и следовало ожидать, спектр импульсного возбуждения сосредоточен в низкочастотном диапазоне.

Так как $g(n)$ в (3.99) имеет конечную длительность, z-преобразование $G(z)$ имеет только нули. Однако в большинстве случаев требуется располагать полюсной моделью, поэтому обычно применяют двухполюсную функцию $G(z)$ [36].

Для невокализованных звуков модель возбуждения гораздо проще. Здесь достаточно располагать источником шума и изменять коэффициент усиления для получения требуемой мощности возбуждения. Для моделей в дискретном времени в качестве такого источника мо-

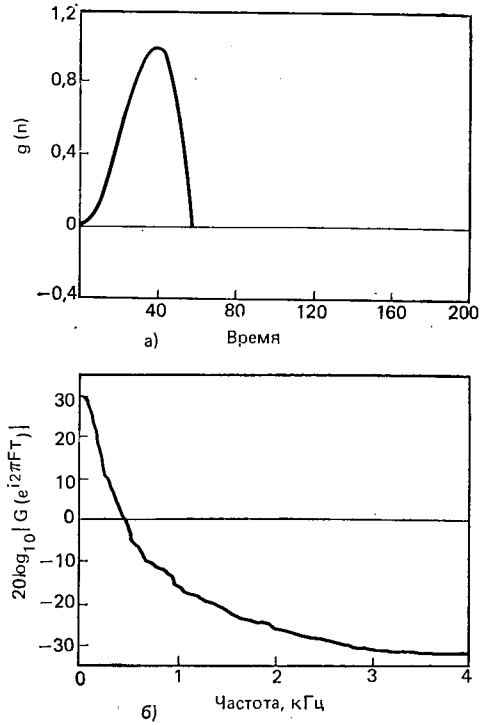


Рис. 3.49. Аппроксимация импульса голосового возбуждения Розенберга (*а*) и спектр (*б*)

жет быть использован генератор случайных чисел, формирующих последовательность с равномерным спектром, функция распределения шумового возбуждения при этом несущественна.

3.4.4. Полная модель

Объединим все компоненты модели (рис. 3.50). Здесь переключением источников возбуждения можно изменять характер сигнала возбуждения. Голосовой тракт можно представить различным

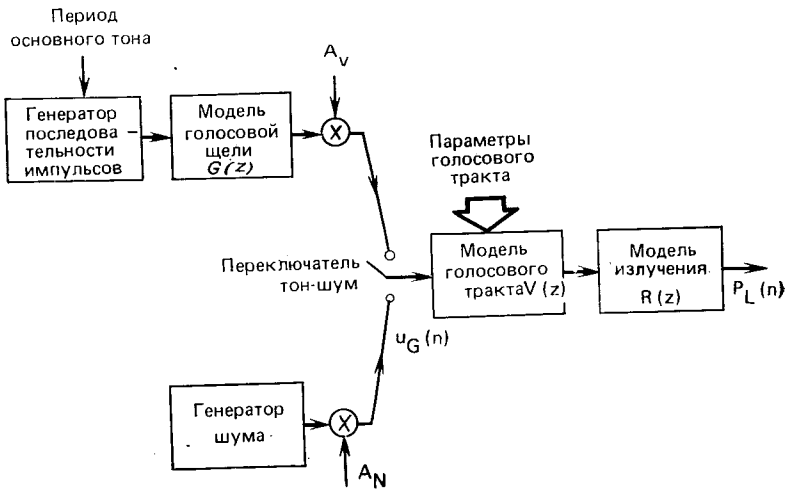


Рис. 3.50. Общая дискретная модель речеобразования

образом. В некоторых случаях удобно объединить модели голосового тракта, возбуждения и излучения в одну систему. В частности, далее будет показано, что в случае анализа на основе линейного предсказания модели голосового возбуждения, излучения и голосового тракта удобно объединить вместе, записав общую передаточную функцию в виде

$$H(z) = G(z)\,V(z)\,R(z). \qquad (3.100)$$

Эта функция имеет только полюса. Другими словами, схема, изображенная на рис. 3.50, дает лишь общее представление о речеобразовании. Существует много разновидностей этой модели.

Важным вопросом является выяснение ограничений этой модели. Очевидно, что модель весьма далека от тех дифференциальных уравнений в частных производных, с которых было начато изучение. Можно выделить несколько ограничений. Первое состоит в характере изменения параметров. Для протяжных звуков, таких как гласные, параметры изменяются довольно медленно и в этом случае модель оказывается достаточно точной. При произнесении кратковременных, например, взрывных звуков модель уже не является адекватной. Следует подчеркнуть, что использование понятий «передаточная функция» и «частотная характеристика» предполагает «кратковременный» анализ сигнала. Таким образом, предполагается, что параметры модели постоянны на интервалах 10—20 мс. Передаточная функция $V(z)$ хорошо отображает структуру звуков, для которых параметры медленно изменяются во времени. В последующих главах это обстоятельство будет часто упоминаться. Второе ограничение состоит в отсутствии нулей передаточной функции, необходимых для точного описания носовых и фрикативных звуков. Это ограничение имеет большее значение для носовых звуков и несколько меньшее для фрикативных. При необходимости нули можно ввести в передаточную функцию модели. Третье ограничение состоит в упрощенном дихотомическом разделении типов возбуждения: вокализованное или невокализованное; такое разделение не соответствует вокализованным фрикативным звукам. Устранить это ограничение путем простого сложения сигналов возбуждения двух типов не удается, так как для фрикативных звуков импульсы основного тона коррелированы с шумовым возбуждением. Смешанная модель возбуждения для вокализованных фрикативных звуков разработана в [40] и может быть использована там, где это необходимо. Наконец, еще одним недостатком модели, изображенной на рис. 3.50, является то, что импульсы голосового возбуждения повторяются с периодом, кратным интервалу дискретизации T. В [41] рассмотрены пути устранения этого ограничения для ситуаций, в который требуется точное управление основным тоном.

3.5. Заключение

В данной главе изложены три основных вопроса: звуки речи, физика речеобразования и дискретные модели речеобразования. Обзор акустической фонетики и теории речеобразования был довольно обстоятельным, но далеко не полным. Основная задача заключалась в изучении таких характерных особенностей речи, располагая которыми можно осознанно вводить модели, полезные для обработки речевых сигналов.

Модели, рассмотренные в § 3.3 и 3.4, являются основой всего последующего содержания книги. Эти модели будут использоваться в двух направлениях. Первое направление называется анализом речи, второе — синтезом. В первом случае предмет изучения охватывает методы оценивания параметров модели в предположении, что выходным сигналом модели является речь. При синтезе речи модели используются для формирования речевого сигнала путем управления соответствующими параметрами модели. Такие задачи встречаются во многих приложениях. Основой изучения являются модели данной главы. Располагая обзорами теории цифровой обработки сигналов (гл. 2) и акустической теории речеобразования (гл. 3), можно приступить к изучению способов применения методов цифровой обработки к речевым сигналам.

Задачи

3.1. На рис. 3.3.1 приведена временна́я диаграмма речевого сигнала длительностью 500 мс (100 мс на каждом отрезке).

а) Указать области вокализованной, невокализованной речи и пауз (шума).

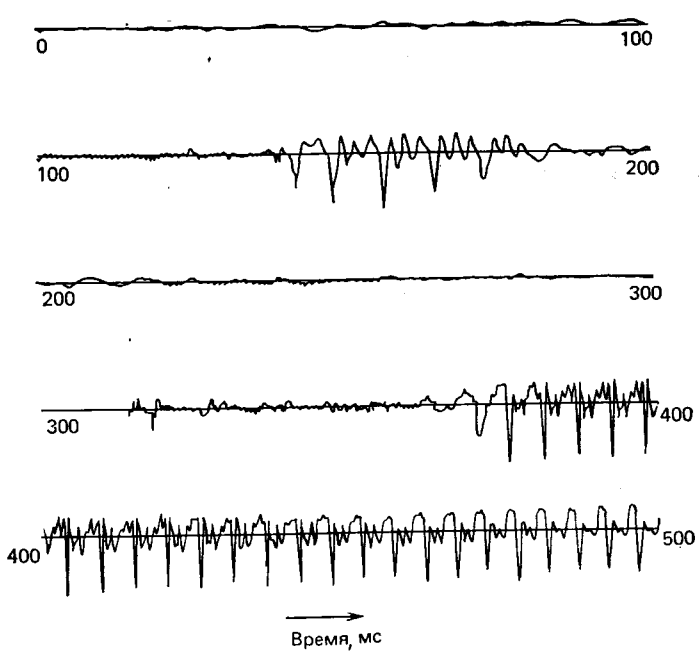


Рис. 3.3.1

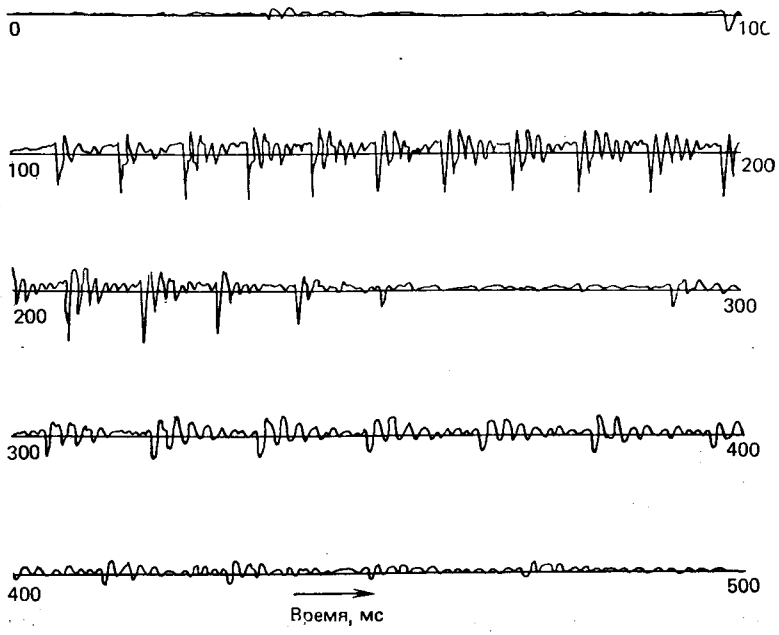


Рис. 3.3.2

б) Для участков вокализованной речи оценить период основного тона и изобразить зависимость периода от времени (для невокализованных участков и пауз положить, что период равен нулю).

3.2. Колебание рис. 3.3.2 соответствует слову cattle. Каждый отрезок соответствует 100 мс.

а) Отметить границы между фонемами, т. е. указать временные границы между звуками $|c|a|tt|le|$.

б) Отметить моменты, где частота основного тона (*i*) — максимальна, (*ii*) — минимальна. Определить частоту основного тона в эти моменты времени.

в) Является ли диктор мужчиной, женщиной или ребенком, и почему?

3.3. Показать что (3.3) есть решение дифференциальных уравнений в частных производных (3.2). Подставить для этого (3.3) в (3.2).

3.4. Коэффициенты отражения в месте соединения двух труб без потерь с площадями сечения A_k и A_{k+1} равны

$$r_k = \frac{A_{k+1}/A_k - 1}{A_{k+1}/A_k + 1} \quad \text{или} \quad r_k = \frac{1 - A_k/A_{k+1}}{1 + A_k/A_{k+1}}.$$

Показать, что, если A_k и A_{k+1} положительны, то $-1 \leqslant r_k \leqslant 1$.

3.5. При анализе влияния нагрузочного сопротивления излучения в модели с трубами без потерь предполагалось, что Z_L — действительная постоянная величина. Более точным является соотношение (3.29б).

а) Задавая граничные условия $P_N(l_N, \Omega) = Z_L U_N(l_N, \Omega)$, найти соотношения между преобразованиями Фурье $u^-_N(t+\tau_N)$ и $u^+_N(t-\tau_N)$.

б) Используя соотношение в частотной области, полученное в а) и (3.29б), показать, что $u^-_N(t+\tau_N)$ и $u^+_N(t-\tau_N)$ удовлетворяют обычному дифференциальному уравнению:

$$L_r\left(R_r + \frac{\rho c}{A_N}\right)\frac{du^-_N(t+\tau_N)}{dt} + \frac{\rho c}{A_N} R_r u^-_N(t+\tau_N) =$$

$$= L_r\left(R_r - \frac{\rho c}{A_N}\right)\frac{du^+_N(t-\tau_N)}{dt} - \frac{\rho c}{A_N} R_r u^+_N(t-\tau_N).$$

3.6. Подставляя (3.50) в (3.49), показать, что (3.48) и (3.49) эквивалентны.

3.7. Пусть имеется модель из двух труб (см. рис. 3.37). Написать для этой модели уравнения в частотной области и показать, что передаточная функция, связывающая скорости входного и выходного потоков, определяется выражением (3.51).

3.8. Рассмотрим идеальную двухсекционную модель без потерь для образования гласных (рис. 3.3.3). Предположим, что и со стороны голосовой щели, и со стороны губ потери отсутствуют. Для этих условий системная функция модели может быть получена из (3.52) подстановкой $r_G = r_L = 1$ и $r_1 = (A_2 - A_1)/(A_2 + A_1)$.

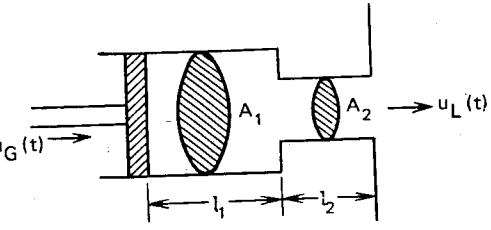


Рис. 3.3.3

а) Показать, что полюса системы на оси $i\Omega$ располагаются на частотах Ω, удовлетворяющих уравнению $\cos\Omega(\tau_1 + \tau_2) + r_1 \cos\Omega(\tau_2 - \tau_1) = 0$ или $(A_1/A_2)\tan(\Omega\tau_2) = \cot(\Omega\tau_1)$, где $\tau_1 = l_1/c$; $\tau_2 = l_2/c$ и c — скорость звука.

б) Значения частоты Ω, найденные решением уравнений а), являются формантными частотами модели с трубами без потерь. Путем соответствующего выбора параметров l_1, l_2, A_1 и A_2 можно аппроксимировать конфигурацию голосового тракта для гласных и, решая приведенные выше уравнения, получить формантные частоты модели. В табл. 3.3.1 даны параметры для нескольких гласных из [2]. Найти формантные частоты для каждого случая (не-

Таблица 3.3.1

Гласный звук	l_1, см	A_1, см²	l_2, см	A_2, см²
| i |	9	8	6	1
| ae |	4	1	13	8
| a |	9	1	8	7
| Λ |	17	6	0	6

линейные уравнения можно решить графически или путем итераций). Принять $c=35000$ см/с.

3.9. Подставляя $\hat{\mathbf{Q}}_1$ и $\hat{\mathbf{Q}}_2$ в (3.75), показать, что передаточная функция дискретной модели голосового тракта с двумя трубами равна выражению (3.77).

3.10. Показать, что если $|a|<1$, то

$$1-az^{-1}=1\Big/\sum_{n=0}^{\infty}a^n z^{-n},$$

и, таким образом, наличие нуля в передаточной функции можно заменить бесконечным количеством полюсов.

3.11. Передаточная функция цифрового формантного резонатора

$$V_k(z)=\frac{1-2|z_k|\cos\theta_k+|z_k|^2}{1-2|z_k|\cos\theta_k\,z^{-1}+|z_k|^2z^{-2}},$$

где $|z_k|=e^{-\sigma_k T}$ и $\theta^k=2\pi F_k T$.

а) Изобразить полюса $V_k(z)$ на z-плоскости. Отметить соответствующие полюса аналоговой системы в s-плоскости.

б) Написать разностные уравнения, связывающие выходной сигнал $y_k(n)$, $V_k(z)$ и входной сигнал $x_k(n)$.

в) Изобразить схему цифровой системы формантного резонатора с тремя операциями умножения.

г) Используя разностные уравнения п. б), изобразить схему цифрового формантного резонатора с двумя операциями умножения.

3.12. Пусть имеется системная функция дискретной модели голосового тракта

$$V(z)=G\Big/\prod_{k=1}^{N}(1-z_k z^{-1}).$$

а) Показать, что $V(z)$ можно разложить на простые дроби:

$$V(z)=\sum_{k=1}^{M}\left[\frac{G_k}{1-z_k z^{-1}}+\frac{G_k^*}{1-z_k^* z^{-1}}\right],$$

где M равно $(N+1)/2$ с округлением до большего целого. Предполагается, что все полюса $V(z)$ — комплексные. Получить выражение для G_k.

б) Объединяя слагаемые в а), показать, что

$$V(z)=\sum_{k=1}^{M}\frac{B_k-C_k z^{-1}}{1-2|z_k|\cos\theta_k\,z^{-1}+|z_k|^2z^{-2}}.$$

Это выражение задает параллельную форму реализации $V(z)$, где $z_k=|z_k|e^{i\theta_k}$ Получить формулы для B_k и C_k через G_k и z_k.

в) Изобразить схему параллельной реализации $\mathbf{V}(z)$ при $M=3$.

г) Пусть задана полюсная система с передаточной функцией $V(z)$. Какая форма реализации потребует выполнения большего числа операций умножения — параллельная или каскадная (см. задачу 3.11)?

3.13. Взаимосвязь между звуковым давлением и скоростью потока около губ задается соотношением $P(l, s)=Z_L(s)U(l, s)$, где $P(l, s)$ и $U(l, s)$ — преобразования Лапласа $p(l, t)$ и $u(l, t)$ соответственно, $Z_L(s)=sR_rL_r/(R_r+sL_r)$, где $R_r=128/9\pi^2$ и $L_r=8a/3\pi c$. Здесь c — скорость звука; a — радиус отверстия между губами. В дискретной модели аналогичное соотношение имеет вид (3.97): $P_L(z)=R(z)U_L(z)$, где $P_L(z)$ и $U_L(z)$ есть z-преобразования $p_L(n)$ и $u_L(n)$ — дискретные аналоги звукового давления и скорости потока, спектр которых ограничен по частоте. Один из способов получения состоит в использовании билинейного преобразования [33]

$$R(z)=Z_L(s)\Big|_{s=\frac{2}{T}\left[\frac{1-z^{-1}}{1+z^{-1}}\right]}.$$

а) Для заданного $Z_L(s)$ определить $R(z)$.

б) Записать разностное уравнение для $p_L(n)$ и $u_L(n)$.

в) Найти полюса и нули $R(z)$.

г) Если $c=35\,000$ см/с, $T=10^{-4}$ с$^{-1}$, 0,5 см$<a<$1,3 см, то где располагаются полюса?

д) Простым приближением $R(z)$ является соотношение $\tilde{R}(z)=R_0(1-z^{-1})$. Для $a=1$ см и $T=10^{-4}$ найти R_0, при котором $\tilde{R}(-1)=Z_L(\infty)=R(-1)$.

ж) Нарисовать частотные зависимости $Z_L(\Omega)$, $R(e^{i\Omega T})$ и $\tilde{R}(e^{i\Omega T})$ как функции Ω для $a=1$ см и $T=10^{-4}$, $0\leqslant\Omega\leqslant\pi/T$.

3.14. Простое приближение импульса голосовой щели изображено на рис. 3.3.4*а*.

а) Найти z-преобразование $G_1(z)$ этой последовательности. (Указание: $g_1(n)$ можно записать как свертку двух последовательностей $p(n)$:

$$p(n)=\begin{cases}1, & 0\leqslant n\leqslant N-1;\\ 0, & \textit{в противном случае.}\end{cases}$$

б) Показать полюса и нули $G_1(z)$ на z-плоскости при $N=10$.

в) Изобразить амплитудный спектр $g_1(n)$ как функцию ω. Рассмотрим приближение импульса голосовой щели рис. 3.3.4*б*.

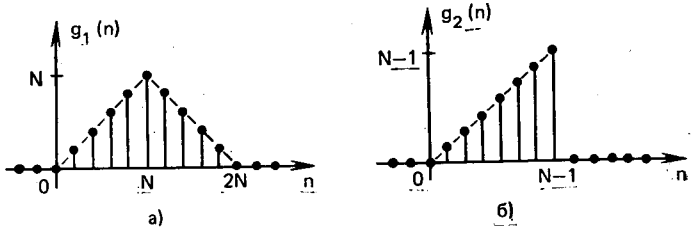


Рис. 3.3.4

г) Показать, что z-преобразование $G_2(z)$ равно

$$G_2(z)=z^{-1}\sum_{n=0}^{N-2}(n+1)z^{-n}=z^{-1}\left[\frac{1-Nz^{-(N-1)}+(N-1)z^{-N}}{(1-z^{-1})^2}\right].$$

(Указание: z-преобразование $nx(n)$ равно $-z\dfrac{dX(z)}{dz}$)

д) Показать, что в общем случае $G_2(z)$ должно иметь по крайней мере один нуль вне единичного круга. Найти нули $G_2(z)$ для $N=4$.

3.15. Наиболее часто импульс в голосовой щели формируется в виде

$$g(n) = \begin{cases} n\,a^n, & n \geqslant 0; \\ 0, & n < 0. \end{cases}$$

а) Найти z-преобразование $g(n)$.
б) Изобразить преобразование Фурье $G(e^{i\omega})$ как функцию ω.
в) Показать, как можно выбрать a, чтобы выполнялось соотношение: $20\log_{10}|G(e^{i0})| - 20\log_{10}|G(e^{i\pi})| = 60$ дБ.

4

Методы обработки речевых сигналов во временной области

4.0. Введение

В гл. 2 и 3 были изложены наиболее эффективные методы цифровой обработки, а также основные свойства речевых сигналов. Рассмотрим теперь применение методов цифровой обработки речевых сигналов. Основной целью обработки речевых сигналов является получение наиболее удобного и компактного представления содержащейся в них информации. Точность представления определяется той информацией, которую необходимо сохранить или выделить. Например, цифровая обработка может применяться для выяснения, является ли данное колебание речевым сигналом. Сходная, но несколько более сложная задача состоит в том, чтобы классифицировать колебания на вокализованную речь, невокализованную речь и паузу (шум). В этих случаях целесообразно использовать такие характеристики сигнала, в которых признаки классификации представлены с максимальной точностью. В других задачах (например, при цифровой передаче) может потребоваться точное восстановление речевого сигнала по его сокращенному представлению. В этой главе рассматриваются методы обработки речевого колебания *во временной области*. В гл. 6—8, напротив, излагаются методы обработки спектрального представления сигнала в частотной области[1].

Примерами временных характеристик речевого сигнала могут служить среднее число переходов через нулевой уровень, энергия сигнала и его корреляционная функция. Эти характеристики часто используются, так как их измерение не требует сложных устройств, а располагая их значениями, можно получить представление о некоторых особенностях сигнала.

В начале главы обсуждаются принципы обработки во временной области; далее приводится несколько примеров такой обработки. В заключение рассматриваются такие задачи, как разделение сигнала на вокализованные и невокализованные сегменты, выделение основного тона, измерение функции кратковременной мощности. Существует множество других задач, которые можно было бы здесь привести. Однако наша цель состоит не в составлении исчерпывающего обзора задач, а в иллюстрации эффективности методов обработки во временной области.

4.1. Текущая обработка речевых сигналов

На рис. 4.1 показана последовательность отсчетов (с частотой 8000 отсч./с), представляющая типичный речевой сигнал. Из рисунка видно, что свойства речевого сигнала изменяются во времени, например характер возбуждения на вокализованных и невокализованных участках, пиковая амплитуда, период основного тона на вокализованных сегментах. Тот факт, что эти изменения видны на осциллограмме речевого сигнала, означает, что методы его обработки во временной области должны обеспечивать хорошее описание таких текущих характеристик сигнала, как мощность, характер возбуждения, основной тон, и, возможно, даже таких параметров голосового тракта, как форматные частоты.

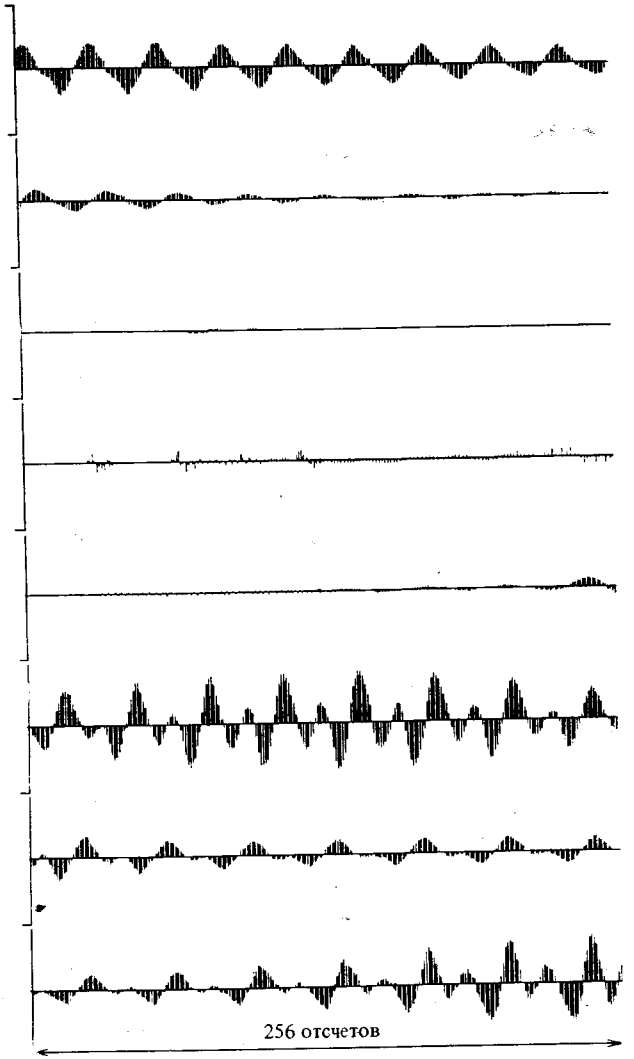


Рис. 4.1. Отсчеты типичного речевого сигнала (частота дискретизации 8 кГц)

[1] Во всех случаях предполагается, что сигнал ограничен по частоте и дискретизирован, по крайней мере, с частотой Найквиста. Предполагается также, что отсчеты квантованы с пренебрежимо малой ошибкой (см. гл. 5, где обсуждаются эффекты квантования).

В основе большинства методов обработки речи лежит предположение о том, что свойства речевого сигнала с течением времени медленно изменяются. Это предположение приводит к методам кратковременного анализа, в которых сегменты речевого сигнала выделяются и обрабатываются так, как если бы они были короткими участками отдельных звуков с отличающимися свойствами. Процедура повторяется так часто, как это требуется. *Сегменты*, которые иногда называют *интервалами* (кадрами) *анализа*, обычно пересекаются. Результатом обработки на каждом интервале является число или совокупность чисел. Следовательно, подобная обработка приводит к новой, зависящей от времени последовательности, которая может служить характеристикой речевого сигнала.

Большинство методов кратковременного анализа, излагаемых в главе, в том числе и кратковременный Фурье-анализ (см. гл. 6), могут быть описаны выражением

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)]\, w(n-m). \tag{4.1}$$

Речевой сигнал (возможно, после ограничения частотного диапазона в линейном фильтре) подвергается преобразованию $T[\cdot]$, линейному или нелинейному, которое может зависеть от некоторого управляющего параметра или их совокупности. Результирующая последовательность умножается затем на последовательность значений временного окна (весовой функции), расположенную во времени в соответствии с индексом n. Результаты затем суммируются по всем ненулевым значениям. Обычно, хотя и не всегда, последовательность значений временного окна имеет конечную протяженность. Значение Q_n представляет собой, таким образом, «взвешенное» среднее значение последовательности $T[x(m)]$.

Простым примером, иллюстрирующим изложенное, может служить измерение кратковременной энергии сигнала. Полная энергия сигнала в дискретном времени определяется как

$$E = \sum_{m=-\infty}^{\infty} x^2(m). \tag{4.2}$$

Вычисление этой величины не имеет особого смысла при обработке речевых сигналов, поскольку она не содержит информации о свойствах сигнала, изменяющихся во времени. Кратковременная энергия определяется выражением

$$E_n = \sum_{m=n-N+1}^{n} x^2(m). \tag{4.3}$$

Таким образом, кратковременная энергия в момент n есть просто сумма квадратов N отсчетов от $n-N+1$ до n. Из (4.1) видно, что в (4.3) $T[\cdot]$ есть просто операция возведения в квадрат, а

$$w(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N-1, \\ 0, & \textit{в противном случае}. \end{cases} \tag{4.4}$$

Вычисление кратковременной энергии иллюстрирует рис. 4.2. Отметим, что окно «скользит» вдоль последовательности квадратов значений сигнала, в общем случае вдоль последовательности $T[x(m)]$, ограничивая длительность интервала, используемого в вычислениях.

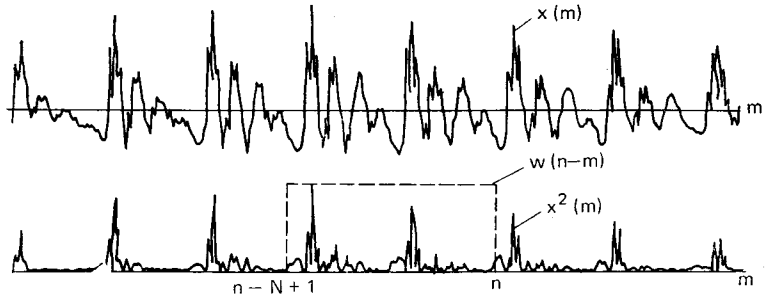


Рис. 4.2. Иллюстрация вычисления функции кратковременной энергии

Более подробно вычисление кратковременной энергии будет обсуждаться в следующем параграфе. Здесь же уместно отметить одно важное свойство преобразования (4.1). Выражение (4.1) описывает дискретную свертку окна $w(n)$ с последовательностью $T[x(n)]$. Таким образом, последовательность Q_n[1] может быть интерпретирована как выходной сигнал линейной инвариантой к сдвигу системы с импульсной характеристикой $h(n) = w(n)$. Такая система изображена на рис. 4.3. Важность этого подхода станет яснее в процессе изучения материала этой главы и гл. 6.

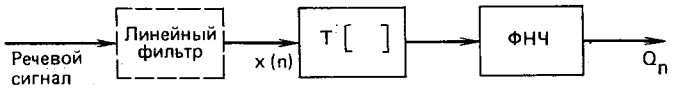


Рис. 4.3. Общее представление принципа кратковременного анализа

4.2. Кратковременная энергия и кратковременное среднее значение сигнала [2]

Как отмечалось выше, амплитуда речевого сигнала существенно изменяется во времени. В частности, амплитуда невокализованных сегментов речевого сигнала значительно меньше амплитуды вокализованных сегментов. Подобные изменения амплитуды хорошо описываются с помощью функции кратковременной энергии

[1] Нижний индекс используется для кратковременных характеристик. Сейчас это не должно вызывать больших затруднений, а в дальнейшем позволит получить простые и ясные обозначения.

[2] Далее везде, где это не вызывает недоразумений, слово «кратковременная» будет опускаться. (*Прим. ред.*)

сигнала. В общем случае определить функцию энергии можно как

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)\, w(n-m)]^2. \qquad (4.5)$$

Это выражение может быть переписано в виде

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m)\, h(n-m), \qquad (4.6)$$

где

$$h(n) = w^2(n). \qquad (4.7)$$

Уравнение (4.6) можно интерпретировать в соответствии с рис. 4.4*а*. Сигнал $x^2(n)$ в этом случае фильтруется с помощью линейной системы с импульсной характеристикой $h(n)$.

Выбор импульсной характеристики $h(n)$ или окна составляет основу описания сигнала с помощью функции энергии. Чтобы понять, как влияет выбор окна на функцию кратковременной энергии сигнала, предположим, что $h(n)$ в (4.6) является достаточно длительной и имеет постоянную амплитуду; значение E_n будет при этом изменяться во времени незначительно. Такое окно эквивалентно фильтру нижних частот с узкой полосой пропускания. Полоса фильтра нижних частот не должна быть столь узкой, чтобы выходной сигнал оказался постоянным, иначе говоря, полосу следует выбрать так, чтобы функция энергии отражала изменения амплитуды речевого сигнала. Описанная ситуация выражает противоречие, которое нередко возникает при изучении кратковременных характеристик речевых сигналов. Суть его состоит в том, что для описания быстрых изменений амплитуды желательно иметь узкое окно (короткую импульсную характеристику), однако слишком малая ширина окна может привести к недостаточному усреднению и, следовательно, к недостаточному сглаживанию функции энергии.

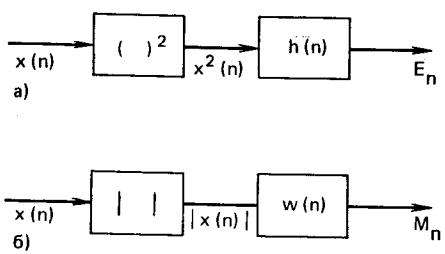


Рис. 4.4. Структурная схема для функции: *а*) кратковременной энергии; *б*) кратковременного среднего значения

Влияние вида окна на вычисление изменяющейся во времени энергии сигнала можно проиллюстрировать на примере использования двух наиболее распространенных окон: прямоугольного

$$h(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N-1, \\ 0, & \textit{в противном случае.} \end{cases} \qquad (4.8)$$

и окна Хемминга

$$h(n) = \begin{cases} 0{,}54 - 0{,}46 \cos(2\pi n/(N-1)), & 0 \leqslant n \leqslant N-1, \\ 0, & \textit{в противном случае.} \end{cases} \qquad (4.9)$$

Прямоугольное окно, как это видно из (4.3), соответствует случаю, когда всем отсчетам на интервале от $(n-N+1)$ до n приписывается одинаковый вес. Частотная характеристика прямоугольного окна (с импульсной характеристикой (4.8)), как легко показать (см. задачу 4.1), равна

$$H(e^{i\Omega T}) = \frac{\sin(\Omega N T/2)}{\sin(\Omega T/2)}\, e^{-i\Omega T (N-1)/2}. \qquad (4.10)$$

Для окна с шириной 51 отсчет ($N=51$) логарифм амплитудно-частотной характеристики представлен на рис. 4.5*а*. Отметим, что

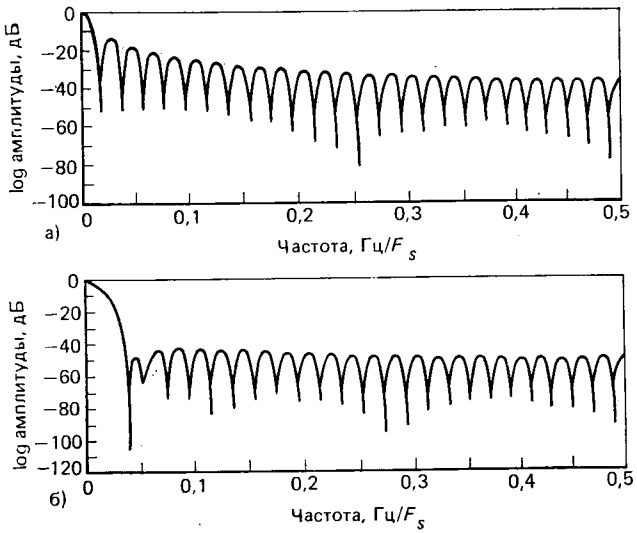


Рис. 4.5. Преобразование Фурье для:
а) прямоугольного окна; *б*) окна Хемминга

первое нулевое значение амплитудно-частотной характеристики (4.10) соответствует частоте

$$F = F_s/N, \qquad (4.11)$$

где $F_s = 1/T$ — частота дискретизации. Это номинальная частота среза фильтра нижних частот, соответствующего прямоугольному окну. Амплитудно-частотная характеристика окна Хемминга при $N=51$ показана на рис. 4.5*б*. Полоса пропускания фильтра с окном Хемминга при одинаковой ширине примерно вдвое превосходит полосу фильтра с прямоугольным окном. Очевидно также, что окно Хемминга обеспечивает большее затухание вне полосы пропускания по сравнению с прямоугольным окном. Затухание, вносимое вне полосы, несущественно зависит от ширины каждого из

окон. Это означает, что увеличение ширины приведет просто к сужению полосы[1]. Если N мало (порядка периода основного тона или менее), то E_n будет изменяться очень быстро, в соответствии с тонкой структурой речевого колебания. Если N велико (порядка нескольких периодов основного тона), то E_n будет изменяться медленно и не будет адекватно описывать изменяющиеся особенности речевого сигнала. Это, к сожалению, означает, что не существует единственного значения N, которое в полной мере удовлетворяло бы перечисленным требованиям, так как период основного тона изменяется от 10 отсчетов (при частоте дискретизации 10 кГц) для высоких женских и детских голосов до 250 отсчетов для очень низких мужских голосов. На практике N выбирают равным 100—200 отсчетов при частоте дискретизации 10 кГц (т. е. длительность порядка 10—20 мс).

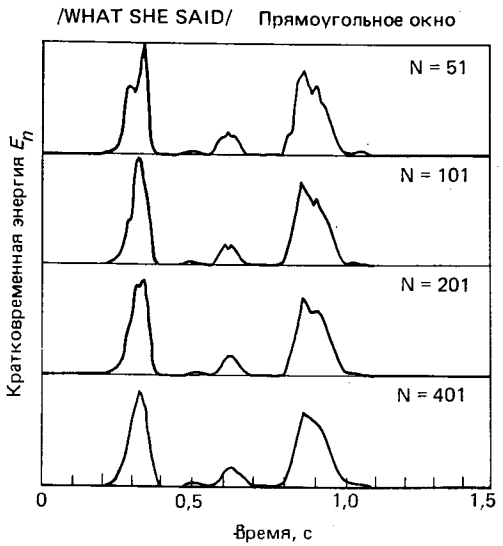


Рис. 4.6. Функции кратковременной энергии для прямоугольных окон различной длительности

На рис. 4.6 и 4.7 показано влияние длительности окна (для прямоугольного окна и окна Хемминга) на энергию, вычисленную по фразе «what she said», произнесенной женским голосом. Легко видеть, что при увеличении N траектория энергии становится более гладкой при использовании любого временного окна.

Основное назначение E_n состоит в том, что эта величина позволяет отличить вокализованные речевые сегменты от невокализованных. Как видно из рис. 4.6 и 4.7, значения E_n для невокализованных сегментов значительно меньше, чем для вокализованных. Функция кратковременной энергии может быть использована для приближенного определения момента перехода от вокализованного сегмента к невокализованному и наоборот, а в случае высококачественного речевого сигнала (с большим отношением сигнала к шуму) функцию энергии можно использовать и для отделения речи от пауз.

Одним из недостатков функции кратковременной энергии, определяемой выражением (4.6), является ее чувствительность к большим уровням сигнала (поскольку в (4.6) каждый отсчет возводится в квадрат). Вследствие этого значительно искажается соотношение между значениями последовательности $x(n)$. Простым способом устранения этого недостатка является переход к определению функции среднего значения в виде

$$M_n = \sum_{m=-\infty}^{\infty} |x(m)|\, w(n-m), \qquad (4.12)$$

где вместо суммы квадратов вычисляется взвешенная сумма абсолютных значений. На рис. 4.4*б* показано, как соотношение (4.12) может быть представлено посредством линейной фильтрации последовательности $|x(n)|$. Исключение операции возведения в квадрат упрощает арифметические вычисления.

На рис. 4.8 и 4.9 показаны траектории среднего значения, соответствующие рис. 4.6 и 4.7. Различия заметны практически лишь на невокализованных сегментах. При вычислении среднего значения по (4.12) динамический диапазон (отношение максимального значения к минимальному) определяется примерно как квадратный корень из динамического диапазона при обычном вычислении энергии. Таким образом, в данном случае различия в уровнях между вокализованной и невокализованной речью выражены не столь ярко, как при использовании функций энергии.

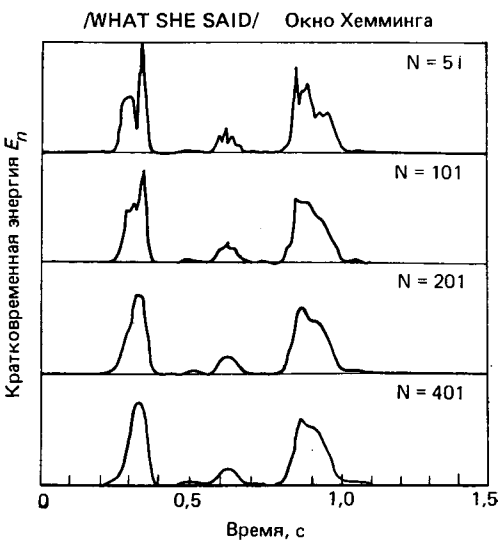


Рис. 4.7. Функции кратковременной энергии для окон Хемминга различной длительности

Поскольку полоса частот при определении как функции энергии, так и среднего значения приближенно совпадает с полосой пропускания используемого фильтра нижних частот, то нет необходимости дискретизировать эти функции столь же часто, как исходный речевой сигнал. Например, для окна длительностью 20 мс достаточна частота дискретизации около 100 Гц. Это означает, что значительная часть информации теряется при использовании подобных кратковременных представлений. Очевидно также, что информация, относящаяся к динамике амплитуд речевого сигнала, сохраняется в весьма удобной форме.

Завершая рассмотрение свойств функций энергии и среднего значения, следует отметить, что используемое окно не обязательно должно быть прямоугольным, или окном Хемминга, или какой-ли-

[1] Здесь нет необходимости в подробном изложении свойств временных окон. Оно содержится в гл. 6.

бо функцией, обычно применяемой в качестве окна при спектральном анализе и при цифровой фильтрации сигналов. Необходимо лишь, чтобы применяемый фильтр обеспечивал адекватное сглаживание. Таким образом, можно использовать фильтр нижних частот, синтезированный любым стандартным способом [1, 2]. Кроме того, фильтр может быть как КИХ-, так и БИХ-фильтром. Импульсная характеристика должна быть всегда положительной, поскольку это гарантирует, что энергия и среднее значение окажутся больше нуля. Фильтры с КИХ (с прямоугольной импульсной характеристикой и характеристикой окна Хемминга) обладают тем преимуществом, что сигнал на их выходе может быть рассчитан сразу для пониженной частоты дискретизации путем сдвига окна более чем на один отсчет входного сигнала. Например, если речевой сигнал дискретизирован с частотой 10 кГц и применяемое окно имеет длительность 20 мс (200 отсчетов), то функция энергии может быть определена при частоте дискретизации около 100 Гц, т. е. 1 раз на каждые 100 отсчетов входного сигнала.

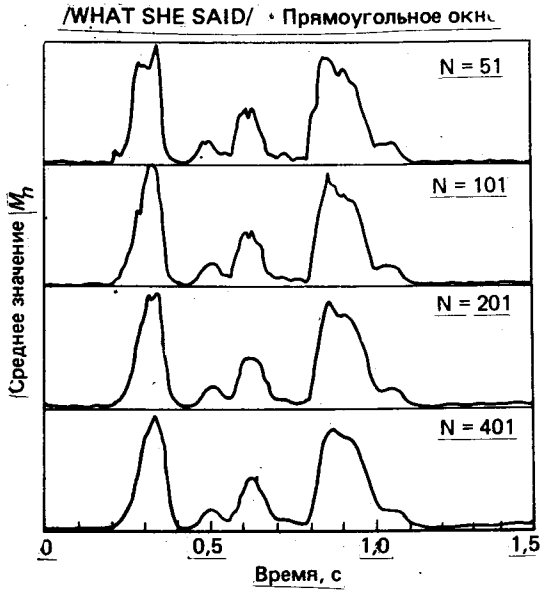


Рис. 4.8. Функции среднего значения для прямоугольных окон различной длительности

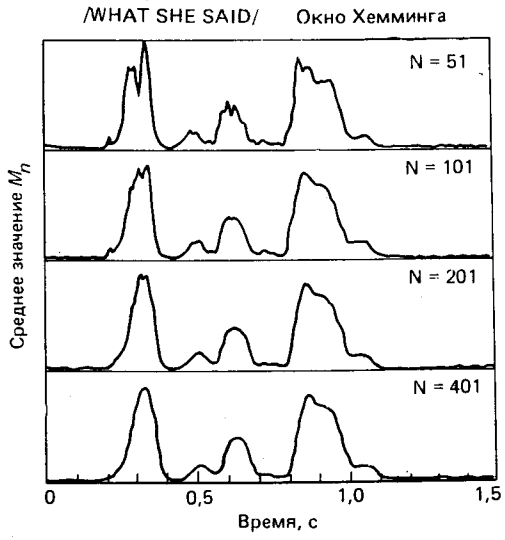


Рис. 4.9. Функции среднего значения для окна Хемминга различной длительности

Совершенно не обязательно использовать окна конечной длительности. Хотя на первый взгляд это кажется необычным, для фильтрации можно использовать фильтр с бесконечно протяженной импульсной характеристикой, если ее z-преобразование представляет собой

рациональную функцию. Примером может служить окно следующего вида:

$$h(n)=\begin{cases} a^n, & n \geqslant 0, \\ 0, & n<0. \end{cases} \tag{4.13}$$

Значение $0<a<1$ позволяет выбирать эффективную длительность окна. Соответствующее z-преобразование окна имеет вид

$$H(z)=1/(1-a\,z^{-1}), \quad |z|>|a|, \tag{4.14}$$

откуда легко видеть, что передаточная функция $H(e^{-i\Omega T})$, как и требуется, сосредоточена в области нижних частот. Подобный фильтр описывается простейшим разностным уравнением, т. е. функция энергии должна удовлетворяться рекуррентному соотношению:

$$E_n=a\,E_{n-1}+x^2(n), \tag{4.15}$$

а среднее значение — рекуррентному соотношению

$$M_n=a\,M_{n-1}+|x(n)|. \tag{4.16}$$

Использование (4.15) и (4.16) приводит к тому, что функции энергии и среднего значения надо вычислять для каждого отсчета входного сигнала, даже если требуется значительно меньшая частота дискретизации. Однако иногда полученные рекуррентные уравнения весьма полезны, например, при кодировании формы речевого сигнала, как это описано в гл. 5. Но если частота дискретизации достаточно снижена, то нерекуррентные методы требуют меньшего объема вычислений (см. задачу 4.4). Другой вопрос, который представляет интерес, относится к определению задержки, связанной с обработкой в фильтре нижних частот. Временные окна (4.8) и (4.9) определены таким образом, что они соответствуют фильтрам, в которых не нарушается принцип причинности (они реализуемы). В силу симметрии импульсной характеристики фильтры имеют абсолютно линейную фазо-частотную характеристику и вносят задержку на $(N-1)/2$ отсчетов. Поэтому исходная функция энергии может быть уточнена с учетом вносимой задержки. При рекуррентной обработке фазо-частотная характеристика нелинейна, поэтому задержку нельзя скомпенсировать полностью.

4.3. Кратковременная функция среднего числа переходов через нуль

При обработке сигналов в дискретном времени считают, что если два последовательных отсчета имеют различные знаки, то произошел переход через нуль. Частота появления нулей в сигнале может служить простейшей характеристикой его спектральных свойств. Это наиболее справедливо для узкополосных сигналов. Например, синусоидальный сигнал с частотой F_0, подвергнутый дискретизации с частотой F_s, имеет F_s/F_0 отсчетов за период. Каждый

период содержит два перехода через нуль, таким образом, среднее число нулевых переходов за большой интервал времени.

$$z = 2F_0/F_s. \tag{4.17}$$

Среднее число нулевых переходов можно принять в качестве подходящей оценки частоты синусоидального колебания.

Речевой сигнал является широкополосным и, следовательно, интерпретация среднего числа переходов через нуль менее очевидна. Однако можно получить грубые оценки спектральных свойств сигнала, основанные на использовании функции среднего числа переходов через нуль для речевого сигнала; рассмотрим способ вычисления этой величины. Определим среднее число переходов через нуль:

$$Z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]|\, w(n-m), \tag{4.18}$$

где

$$\operatorname{sgn}[x(n)] = \begin{cases} 1, & x(n) \geqslant 0, \\ -1, & x(n) < 0 \end{cases} \tag{4.19}$$

и

$$w(n) = \begin{cases} 1/2N, & 0 \leqslant n \leqslant N-1, \\ 0, & \textit{в противном случае.} \end{cases} \tag{4.20}$$

Операции, входящие в (4.18), представлены в виде структурной схемы на рис. 4.10. Такое представление показывает, что функция среднего числа переходов через нуль имеет те же общие свойства,

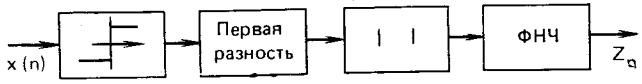


Рис. 4.10. Структурная схема вычисления кратковременной функции среднего нулевых пересечений

что и функции энергии и среднего значения. Может показаться, однако, что вычисления по (4.18) и в соответствии с рис. 4.10 являются более сложными, чем это есть на самом деле. Все, что в действительности требуется, это проверить пары отсчетов с целью определения нулевых пересечений, а затем вычислить среднее по всем N последовательным отсчетам (деление на N, конечно, не обязательно). Как и ранее, может быть вычислено взвешенное среднее и при использовании симметричных окон конечной длительности задержка может быть скомпенсирована точно. Могут быть получены и рекуррентные уравнения, сходные с (4.15) и (4.16) (см. задачу 4.5).

Рассмотрим теперь применение функции среднего числа переходов через нуль для обработки речевых сигналов. Модель речеобразования предполагает, что энергия вокализованных сегментов речевого сигнала концентрируется на частотах ниже 3 кГц, что обусловлено убывающим спектром сигнала возбуждения, тогда как для невокализованных сегментов бóльшая часть энергии лежит в области высоких частот. Поскольку высокие частоты приводят к большому числу переходов через нуль, а низкие — к малому, то существует жесткая связь между числом нулевых пересечений и распределением энергии по частотам. Разумно предположить, что большому числу нулевых пересечений соответствуют невокализованные сегменты, а малому числу — вокализованные сегменты речи. Это, однако, очень расплывчатое утверждение, поскольку мы не определили, что означает «много» или «мало», и количественно определить эти понятия в действительности трудно. На рис. 4.11 представлены гистограммы среднего числа нулевых пересечений (усреднение за 10 мс) как для вокализованных, так и для невокализованных сегментов речевого сигнала. Отметим, что гауссовская кривая хорошо согласуется с приведенными гистограммами. Среднее число пересечений составляет 49 для вокализованных и 14 для невокализованных сегментов длительностью 10 мс. Поскольку оба распределения перекрываются, нельзя вынести однозначное решение о принадлежности сегмента к вокализованным или невокализованным отрезкам только по среднему числу переходов через нуль. Тем не менее, подобное представление весьма полезно при осуществлении такой классификации.

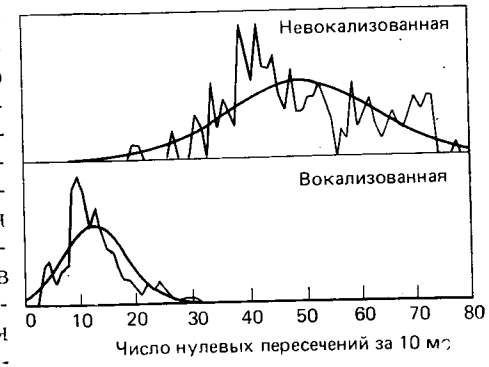


Рис. 4.11. Распределение нулевых пересечений для вокализованной и невокализованной речи

Некоторые результаты измерения среднего числа переходов через нуль представлены на рис. 4.12. В приведенных примерах длительность окна составляла 15 мс (150 отсчетов при частоте дискретизации 10 кГц). Результат вычислялся 100 раз в секунду (окно перемещалось с шагом в 100 отсчетов). Отметим, что так же, как и в случае функций энергии и среднего, функцию среднего числа переходов через нуль можно дискретизировать с очень низкой частотой. Хотя среднее число переходов через нуль изменяется значительно, вокализованные и невокализованные сегменты на рис. 4.12 просматриваются очень четко.

При использовании описания сигнала средним числом переходов через нуль следует иметь в виду ряд практических соображений. Хотя в основу алгоритма вычисления нулевых переходов положено сравнение знаков соседних отсчетов, тем не менее при дискретизации сигнала следует предпринимать специальные меры.

Очевидно, что число нулевых переходов зависит от уровня шума при аналого-цифровом преобразовании, интенсивности фона переменного тока и других шумов, которые могут присутствовать в цифровой системе. Таким образом, с целью уменьшения влияния

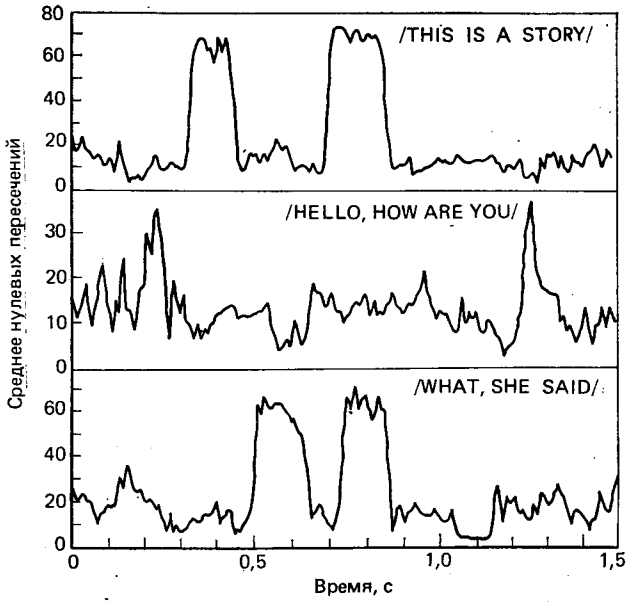


Рис. 4.12. Функция среднего нулевых пересечений для трех различных фраз

этих факторов следует проявлять особую осторожность при аналоговой обработке сигнала, предшествующей дискретизации. Например, часто оказывается более целесообразным использовать полосовой фильтр вместо фильтра нижних частот для уменьшения эффекта наложения при аналого-цифровом преобразовании и устранения фона переменного тока из сигнала. Кроме того, при измерении числа переходов через нуль следует учитывать соотношение между периодом дискретизации и интервалом усреднения N. Период дискретизации определяет точность выделения нулевых пересечений по времени (и по частоте), т. е. чтобы добиться высокой точности, нужна большая частота дискретизации. Вместе с тем от каждого отсчета требуется информация объемом лишь 1 бит (информация только о знаке сигнала).

Вследствие практической ограниченности этого метода было предложено множество сходных представлений сигнала. В каждом из них содержатся дополнительные особенности, направленные на снижение чувствительности оценок к шуму, но все они имеют и свои собственные ограничения. Наиболее заметным среди них является представление сигнала, исследованное Бейкером

[3]. Представление основано на интервалах времени между положительными переходами через нуль (снизу вверх). Бейкер применил это описание для фонетической классификации звуков речи [3].

Другое применение анализа переходов через нуль состоит в получении промежуточного представления речевого сигнала в частотной области. Метод включает фильтрацию речевого сигнала в нескольких смежных частотных диапазонах. Затем по сигналам на выходе фильтров измеряют кратковременную энергию и среднее число переходов через нуль. Совместное использование этих характеристик дает грубое описание спектральных свойств сигнала. Этот подход, предложенный Рэдди и исследованный Вайсенсом [4] и Эрманом [5], положен в основу систем распознавания речи.

4.4. Разделение речи и пауз на основе функций кратковременной энергии и среднего числа переходов через нуль

Задача определения моментов начала и окончания фразы при наличии шума является одной из важных задач в области обработки речи. В частности, при автоматическом распознавании слов важно точно определить моменты начала и окончания слова. Методы обнаружения моментов начала и окончания фразы можно использовать для уменьшения числа арифметических операций, если обрабатывать только те сегменты, в которых имеется речевой сигнал, например, в системах, работающих не в реальном масштабе времени.

Проблема отделения речи от окружающего шума очень сложна, за исключением случаев очень большого отношения сигнал/шум, т. е. в случае высококачественных записей, выполненных в заглушенной камере или звуконепроницаемой комнате. В этих случаях энергия даже наиболее слабых звуков речи (фрикативных согласных) превышает энергию шума и, таким образом, достаточно лишь измерить энергию сигнала. Но подобные условия записи, как правило, не встречаются в реальных ситуациях.

Рассматриваемый ниже алгоритм основан на измерении двух простых характеристик — энергии и числа переходов через нуль. На примере простых ситуаций иллюстрируются трудности, возникающие при обнаружении моментов начала и окончания фразы. На рис. 4.13 представлено колебание (начало слова eight), в котором шум, как это видно из рисунка, легко отделяется от речевого сигнала. В этом случае значительное различие энергий сигнала и шума достаточно для определения момента начала фразы. На рис. 4.14 изображен другой случай (начало слова six), в котором также легко определить начало речевого сигнала. Здесь спектральный состав речи существенно отличается от спектрального состава окружающего шума, что видно по резкому увеличению числа переходов через нуль в сигнале. Следует отметить, что в данном слу-

чае энергия речевого сигнала в начале фразы сравнима с энергией шума.

На рис. 4.15 представлен случай, в котором чрезвычайно трудно выделить начало речевого сигнала. На данном рисунке изображено колебание в начале слова (four). Поскольку это слово начинается со слабого (с малой энергией) фрикативного согласного, очень трудно определить момент его начала. Хотя точка *B* могла бы служить началом слова, в действительности оно начинается в точке *A*. В общем случае очень трудно определить начало или концы слов, в которых: 1) слабые фрикативные согласные (|*f*|, |*th*|, |*h*|) в начале или конце; 2) слабые глухие взрывные звуки (|*p*|, |*t*|, |*k*|) в начале или в конце; 3) носовые звуки в конце; 4) вокализованные фрикативные звуки, которые переходят в невокализованные в конце слова; 5) протяженные гласные звуки в конце слова.

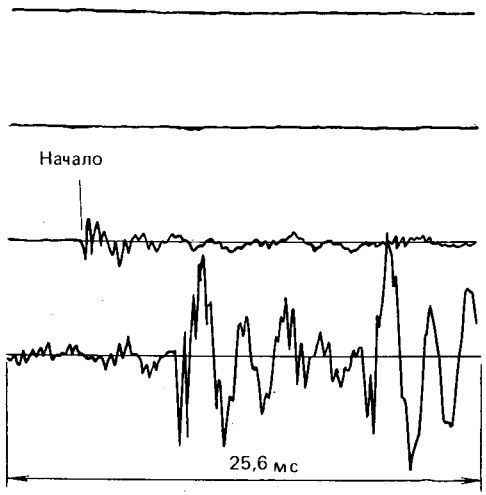


Рис. 4.13. Временна́я диаграмма начала слова */eight/* [6]

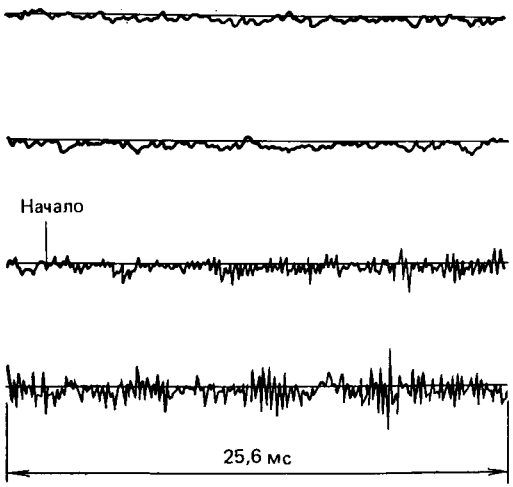


Рис. 4.14. Временна́я диаграмма начала слова */six/* [6]

Несмотря на сложность поставленной выше задачи, энергия и переходы через нуль могут быть совместно использованы в качестве основы для разработки алгоритмов выделения момента начала и конца фразы. Один из таких алгоритмов изучен Рабинером и Самбуром [6] при построении системы распознавания изолированных слов. В этой системе диктор произносит слово в течение интервала времени, установленного для регистрации сигнала. Сигнал далее на всем интервале дискретизируется и заносится в память для последующей обработки. Цель алгоритма состоит в определении начала и конца слова с тем, чтобы при распознавании исключить сегменты, содержащие только шум.

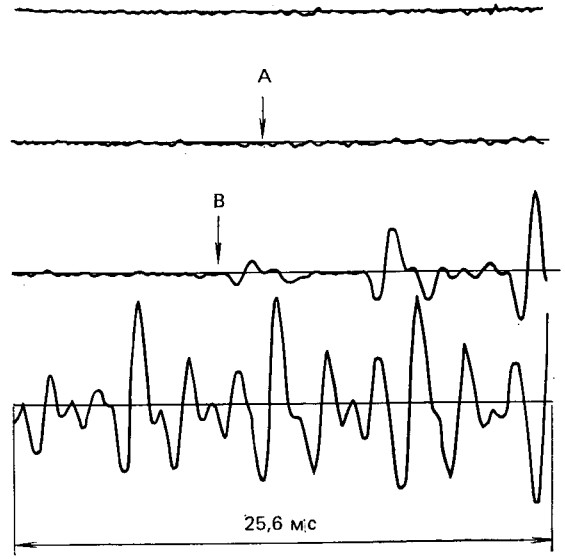


Рис. 4.15. Временна́я диаграмма начала слова */four/* [6]

Алгоритм можно пояснить с помощью рис. 4.16. В качестве основных параметров используются число переходов через нуль в течение 10 мс (4.18) и функция среднего значения (4.12), вычисленные с использованием окна длительностью 10 мс. Обе функции вычисляются на всем интервале с частотой 100 Гц. Предполагается,

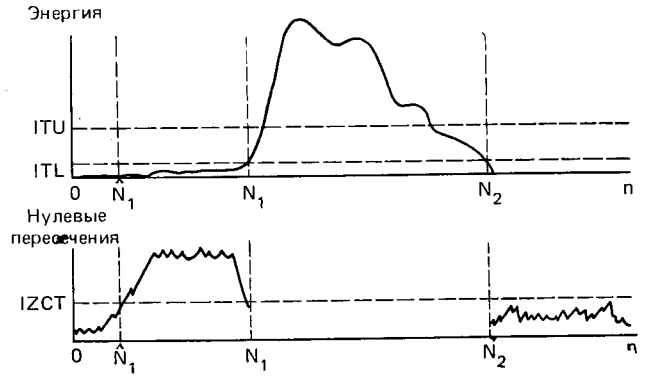


Рис. 4.16. Типичный пример измерений среднего значения и нулевых пересечений для слов с фрикативным звуком в начале [6]

что первые 100 мс не содержат речевого сигнала. По этому участку вычисляется среднее значение и дисперсия каждой из величин (4.12), (4.18) для определения статистических характеристик шума. Затем, с учетом этих характеристик и максимального среднего значения на интервале вычисляются пороги для среднего числа нулевых переходов и энергии сигнала [6]. Определяется фрагмент колебания, на котором траектория среднего значения превышает верхний порог (*ITU* на рис. 4.16). Предполагается, что начало и конец слова лежат вне этого фрагмента. Затем, двигаясь в обратном направлении по оси времени от момента, где M_n впервые превысила порог *ITU*, определяют момент, в котором M_n впервые оказалась меньше нижнего порога *ITL* (точка N_1). Этот момент выбирается в качестве предполагаемого начала. Сходным образом определяется и предполагаемое окончание слова N_2.

Данный двухпороговый алгоритм гарантирует, что провалы в траектории среднего значения не приведут к ложному выделению моментов начала и конца слова. На этом этапе главное — получить данные о том, что начало и конец слова расположены вне интервала от N_1 до N_2. Следующий шаг состоит в перемещении влево от N_1 (вправо от N_2) и сравнении числа переходов через нуль с порогом (*IZCT* на рис. 4.16), вычисленным по начальному участку. Это перемещение не должно превышать 25 интервалов слева от N_1 (справа от N_2). Если число переходов через нуль превышает порог 3 или более раз, начало слова переносится туда, где кривая числа нулевых пересечений впервые превысила порог. В противном случае N_1 считается началом слова. Аналогично поступают и с N_2. На рис. 4.17 показан пример работы алгоритма на типичных изолированных словах. На рисунке представлены восемь функций среднего значения для восьми различных слов двух различных дикторов. Некоторые слова записаны в машинном зале, а другие представляют собой магнитную запись в звуконепроницаемой комнате.

На каждом рисунке помечены начало и конец слова, как они были определены с помощью алгоритма. Например, на рис. 4.17*а* (слово |*nine*|) контроль среднего значения оказался достаточным для определения границ слова. А на примере 4.17*б* (слово |*replace*|) для определения конца слова использована функция числа нулевых пересечений, так как здесь расположен фрикативный звук |*s*|. Несмотря на то что звук |*s*| в конце слова имеет большое среднее значение, конец слова не может быть точно выделен по этому значению из-за высокого порога. В этом случае момент окончания слова уточнен по кривой числа нулевых пересечений. На рис. 4.17*в* конечное |*t*| в слове |*delete*| легко выделяется из-за значительного числа нулевых пересечений на участке длительностью 70 мс, расположенном после смычки и соответствующем взрывному согласному |*t*|. Таким образом, хотя среднее значение и число переходов через нуль малы на интервале 50 мс, соответствующем смычке, алгоритм позволяет правильно определить конец слова за счет большой интенсивности взрывного звука. В то же

время, если интенсивность взрывного звука будет мала, то конец слова будет отождествлен с началом смычки.

На рис. 4.17*г* представлен пример, в котором среднее значение шума было значительным в двух местах до начала слова |*substract*|, однако алгоритм устранил эти моменты из рассмотрения ввиду малого числа переходов через нуль. В этом примере относительно слабый взрывной звук (|*t*|) был правильно идентифициро-

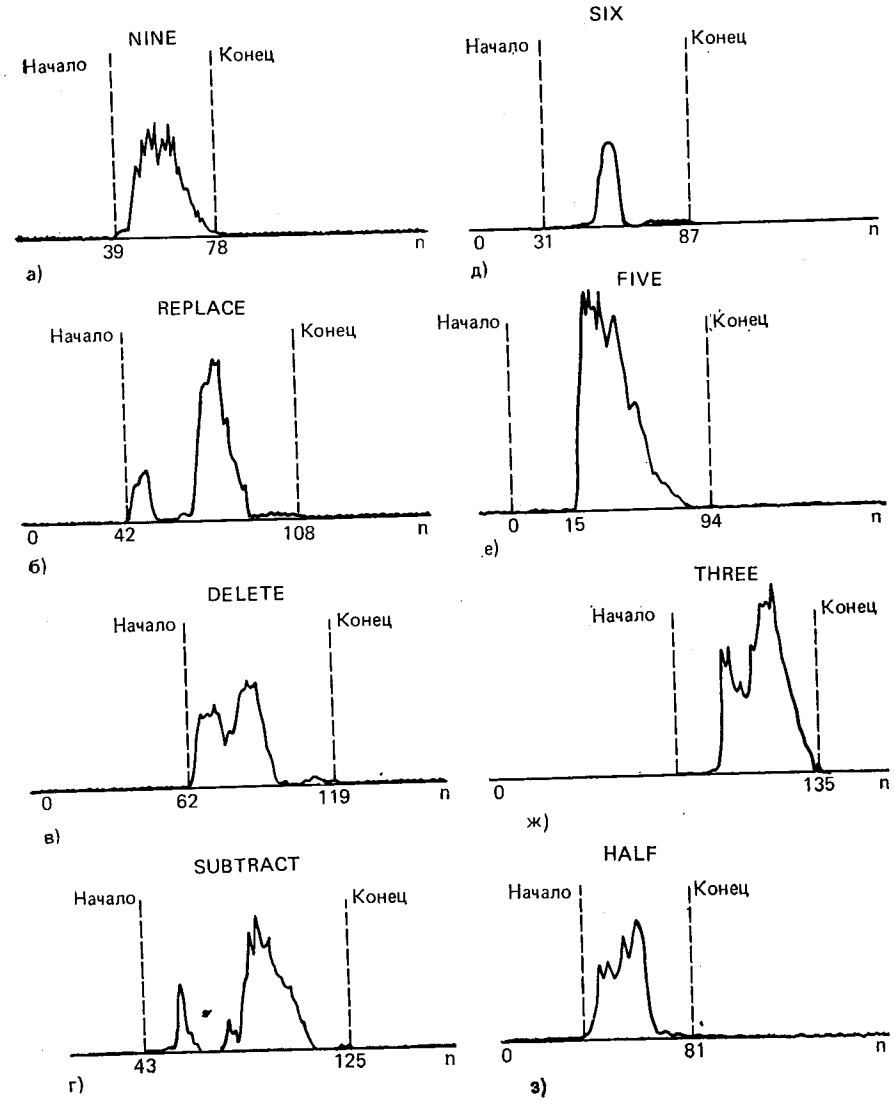


Рис. 4.17. Последовательности среднего значения, иллюстрирующие работу алгоритма выделения конца и начала слова на разных сигналах [6]

ван, как конец слова. На рис. 4.17*д* — показаны слова с фрикативными согласными в начале либо в конце. Во всех случаях алгоритм позволил установить граничные моменты слов так, что значительная часть невокализованного сегмента оказывалась в пределах этих границ.

Это применение функций числа переходов через нулевой уровень и среднего значения показывает, насколько полезны такие характеристики при решении практических задач. Большая практическая ценность рассмотренных способов обусловлена их простотой. Подобные примеры обработки будут встречаться и в последующих параграфах этой главы.

4.5. Оценивание периода основного тона на основе параллельной обработки

Оценивание периода (или частоты) основного тона является одной из наиболее важных задач в обработке речи. Выделители основного тона используются в вокодерах [8], системах распознавания и верификации дикторов [9, 10], в устройствах, предназначенных для глухих [11]. Поскольку задача очень важна, предложен ряд способов ее решения [12—19]. Все они обладают ограничениями и можно с уверенностью сказать, что в настоящее время отсутствует метод выделения основного тона, обеспечивающий удовлетворительные результаты для различных дикторов, в разных областях применения и условиях эксплуатации.

В этом параграфе рассмотрим только один метод выделения основного тона, предложенный Голдом и затем усовершенствованный Голдом и Рабинером [14]. Причины выбора именно этого метода выделения основного тона состоят в следующем: метод с успехом применялся в ряде приложений, основан исключительно на обработке во временной области, требует малых затрат времени при моделировании на универсальной ЭВМ и просто реализуем в спецвычислителе, а также хорошо иллюстрирует принцип параллельной обработки.

Основные положения этого метода:

1. По речевому сигналу формируется несколько импульсных последовательностей, которые сохраняют периодичность входного сигнала и не содержат других его особенностей, бесполезных с точки зрения выделения основного тона.

2. Обработка предполагает использование набора простых выделителей основного тона для каждой последовательности.

3. Оценки основного тона каждой последовательности подвергаются логической обработке для получения результирующей оценки периода основного тона речевого сигнала.

На рис. 4.18 изображена схема, предложенная Голдом и Рабинером [14]. Речевой сигнал дискретизируется с частотой 10 кГц, что позволяет оценить период с точностью $T = 10^{-4}$ с. Далее сигнал сглаживается в фильтре нижних частот с частотой среза около 900 Гц. Можно применить и полосовой фильтр с полосой от 100 до

900 Гц для устранения фона переменного тока питания (фильтр может быть выполнен в виде аналогового устройства до дискретизации или в виде цифрового — после дискретизации).

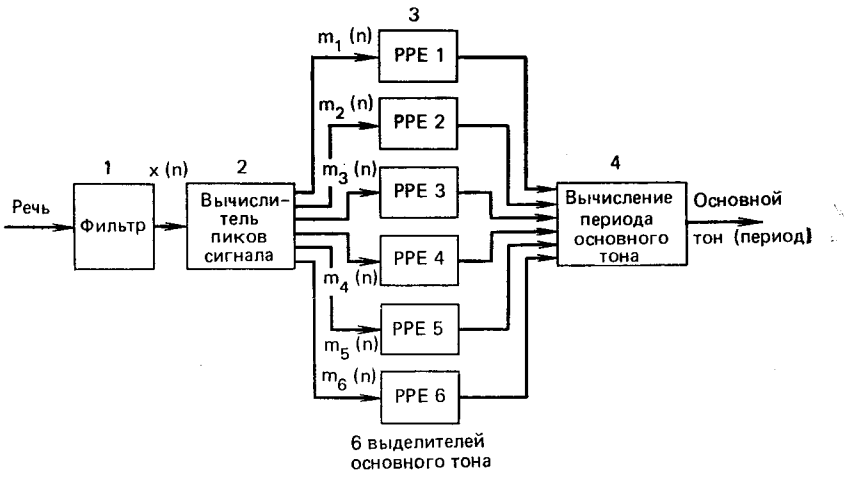


Рис. 4.18. Структурная схема выделения основного тона параллельной обработкой сигнала во временной области

Вслед за фильтрацией определяются локальные максимумы и минимумы в сигнале и по их амплитуде и положению из отфильтрованного сигнала формируется несколько (на рис. 4.18 — шесть) импульсных последовательностей. Каждая импульсная последовательность состоит из положительных импульсов, возникающих в месте расположения максимума или минимума сигнала. Эти шесть последовательностей в [14] имеют следующий вид.

1. $m_1(n)$: импульс, равный по амплитуде максимальному значению сигнала и формирующийся в месте расположения максимума.

2. $m_2(n)$: импульс, равный по амплитуде разности между максимумом и предшествующим минимумом и формирующийся в точке каждого максимума.

3. $m_3(n)$: импульс, равный по амплитуде разности между текущим максимумом и предшествующим максимумом и возникающий в точке каждого максимума (если эта разность отрицательна, то импульс обращается в нуль).

4. $m_4(n)$: импульс, равный по амплитуде минимальному отрицательному значению, взятому со знаком «минус», и возникающий в точке каждого минимума.

5. $m_5(n)$: импульс, равный сумме значений сигнала в точке минимума, взятого со знаком «минус», и сигнала в точке предшествующего максимума; формируется в точке каждого минимума.

6. $m_6(n)$: импульс, равный сумме минимального значения сигнала, взятого со знаком «минус», и предшествующего минимально-

го значения (если эта разность отрицательная, то импульс обращается в нуль).

На рис. 4.19 и 4.20 показаны два примера — синусоидальный сигнал и сумма синусоидального сигнала и его второй гармоники вместе с импульсными последовательностями, определенными выше. Очевидно, что импульсные последовательности имеют тот же период, что и исходный сигнал, хотя $m_5(n)$ на рис. 4.20 ближе к периодической последовательности с периодом, равным половине основного периода. Целью формирования импульсных последовательностей является упрощение текущего оценивания периода основного тона. Работа простейшего устройства оценивания иллюстрируется рис. 4.21. Каждая импульсная последовательность обрабатывается нелинейной системой с переменными параметрами (названной в [13] выделителем основного тона с экспоненциальной границей раздела). Когда на входе появляется импульс достаточно большой амплитуды, на выходе цепи устанавливается постоянный сигнал, равный этой амплитуде. Этот сигнал поддерживается неизменным в течение фиксированного интервала времени $\tau(n)$. В конце интервала выходной сигнал начинает уменьшаться по экспоненте. Когда входной импульс превысит уровень экспоненциально затухающего сигнала на выходе, процесс повторяется. Скорость затухания и длительность интервала зависят от последней оценки основного тона [14]. В результате такого своеобразного сглаживания получается квазипериодическая импульсная последовательность, показанная на рис. 4.21. Длительность каждого импульса представляет собой оценку периода основного тона, который обновляется с частотой 100 Гц.

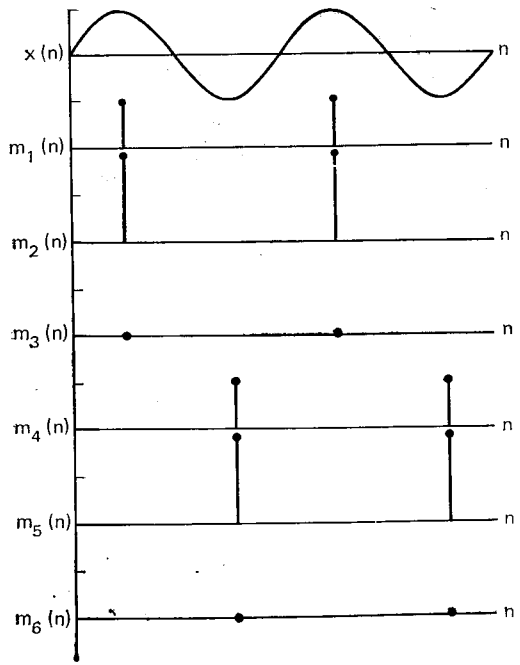


Рис. 4.19. Входной сигнал (синусоида) и соответствующая последовательность импульсов, сформированная по пикам и провалам.

Этот метод применяется в каждом из шести каналов, и таким образом получается шесть оценок периода основного тона. Полученные текущие оценки рассматриваются совместно с двумя последними оценками в каждом из шести каналов. Все оценки затем сравниваются и за оценку периода основного тона принимается та,

которая чаще всего встречается в данном множестве. Метод дает очень хорошие результаты при оценивании периода основного тона на вокализованных сегментах речевого сигнала. Для невокализованных сегментов возникает значительный разброс в оценках периода в каждом из шести каналов. Если такой разброс обнаруживается, то речь классифицируется как невокализованная. Весь процесс повторяется периодически для получения периода основного тона как функции времени и разделения всего сигнала на вокализованные и невокализованные участки.

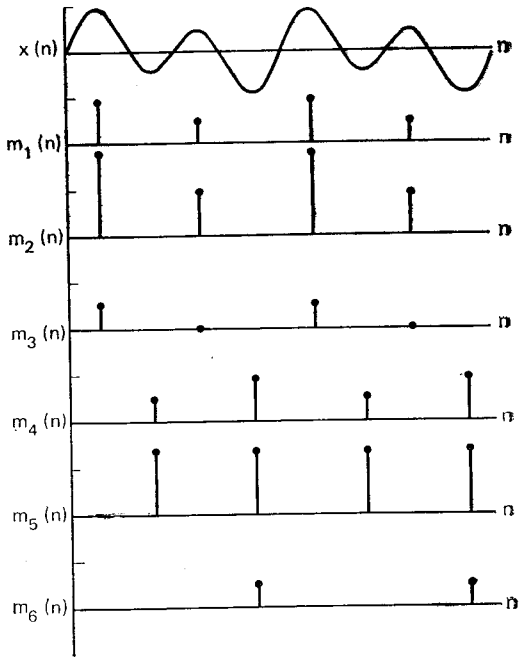


Рис. 4.20. Входной сигнал (ослабленная основная гармоника в сумме со второй гармоникой) и соответствующая последовательность импульсов, сформированная по пикам и провалам

Хотя описанный способ может показаться чрезмерно хитроумным, такая схема выделения периода основного тона может быть рекомендована как для целей технической реализации, так и для моделирования в универсальной ЭВМ. Действительно, при использовании современных ЭВМ метод позволяет обрабатывать сигнал почти в реальном масштабе времени (с коэффициентом трансформации времени, равным 2).

Работу схемы выделения периода основного тона иллюстрирует рис. 4.22, где приведен результат обработки синтезированного речевого сигнала. Преимущество использования синтетической ре-

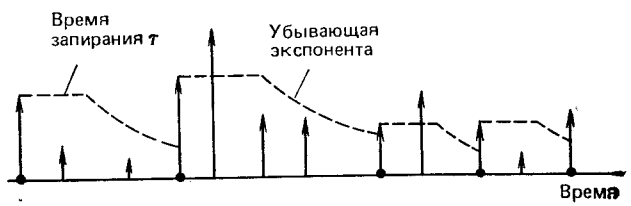


Рис. 4.21. Иллюстрация работы отдельного устройства оценивания периода основного тона во временной области [14]

чи состоит в том, что истинный период основного тона известен точно (поскольку он задается при синтезе). Это позволяет установить точность алгоритма. Недостаток синтетической речи заключается в том, что она формируется с помощью простой модели и по-

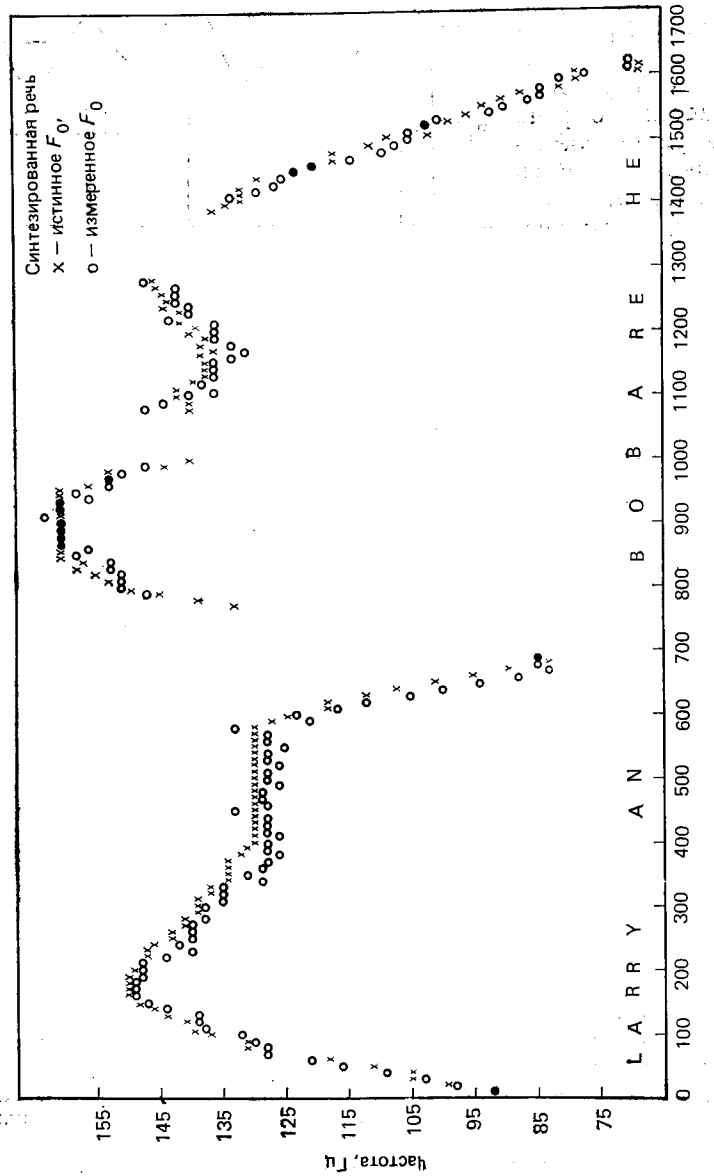


Рис. 4.22. Истинное и измеренное значение частоты основного тона для синтезированной фразы [14]

этому не может отражать всех свойств натурального речевого сигнала. Во всяком случае, эксперимент на синтезированном сигнале показал, что метод позволяет следить за периодом основного тона с точностью до двух интервалов дискретизации. Более того, было обнаружено, что в начале вокализованных сегментов (т. е. первые 10—30 мс) речь часто классифицируется как невокализованная. Этот результат обусловлен тем, что для устойчивого выделения требуется примерно три периода. Таким образом, возникает задержка приблизительно на два периода основного тона. В проведенном недавно сравнительном анализе алгоритмов выделения основного тона на реальном речевом сигнале и для различных условий оценивания этот метод дал хорошие результаты по сравнению с другими известными алгоритмами [12].

В заключение подчеркнем несколько основных положений метода. Во-первых, речевой сигнал обрабатывается с целью получения совокупности импульсных последовательностей, сохраняющих только периодичность сигнала (или фиксирующих ее отсутствие). Из-за такого упрощения структуры речевого сигнала для получения хорошей оценки основного тона оказывается возможным использовать простейшее устройство оценивания. И, во-вторых, несколько оценок основного тона рассматриваются в совокупности для повышения качества выделения. Таким образом, простота обработки сигнала достигнута ценой увеличения сложности логической части алгоритма. Поскольку логические операции осуществляются значительно реже (100 раз в секунду), чем дискретизация сигнала, скорость обработки оказывается высокой. Аналогичный подход был использован Барнвеллом [15] при разработке выделителя основного тона с помощью четырех простых схем выделения по нулевым пересечениям с последующей совместной обработкой решений для получения надежной оценки.

4.6. Кратковременная автокорреляционная функция

Автокорреляционная функция детерминированного сигнала в дискретном времени определяется выражением

$$\varphi(k) = \sum_{m=-\infty}^{\infty} x(m)\, x(m+k). \tag{4.21}$$

Если сигнал случайный или периодический, то $\varphi(k)$ целесообразно определить по-другому:

$$\varphi(k) = \lim_{N\to\infty} \frac{1}{(2N+1)} \sum_{m=-N}^{N} x(m)\, x(m+k). \tag{4.22}$$

Во всех случаях представление сигнала с помощью автокорреляционной функции позволяет отразить определенные свойства сиг-

нала. Например, если сигнал имеет период в P отсчетов, то легко показать, что

$$\varphi(k)=\varphi(k+P), \tag{4.23}$$

т. е. автокорреляционная функция периодического сигнала тоже периодическая. Автокорреляционная функция обладает и рядом других важных свойств: 1) она является четной функцией; 2) достигает максимального значения при $k=0$, т. е. $|\varphi(k)|\leqslant\varphi(0)$ для всех k; 3) величина $\varphi(0)$ равна полной энергии для детерминированного сигнала и средней мощности для случайного или периодического сигнала.

Таким образом, по автокорреляционной функции можно определить энергию сигнала и, кроме того, его периодические свойства. Если рассмотреть (4.23) и свойства 1 и 2, то можно отметить, что для периодического сигнала автокорреляционная функция достигает максимального значения в точках $0, \pm P, \pm 2P, \ldots$, т. е. при любом временно́м расположении сигнала его период можно оценить путем определения местоположения первого максимума автокорреляционной функции. Это свойство автокорреляционной функции позволяет использовать ее для оценки периодичности в любом сигнале, в том числе и речевом. Более того (см. гл. 8), автокорреляционная функция содержит значительно больше информации о тонкой временно́й структуре сигнала. Таким образом, весьма важно рассмотреть, как следует изменить приведенное определение для описания сигнала с помощью кратковременной автокорреляционной функции.

Используя развитый выше подход к определению кратковременных характеристик, определим корреляционную функцию в виде

$$R_n(k)=\sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,x(m+k)\,w(n-k-m). \tag{4.24}$$

Это уравнение можно интерпретировать следующим образом: сначала выделяется сегмент речевого сигнала с помощью функции временного окна, затем к взвешенному таким образом речевому сигналу применяется преобразование (4.21). Легко установить, что

$$R_n(-k)=R_n(k). \tag{4.25}$$

Используя это соотношение, можно выразить $R_n(k)$ в виде (4.10). Сначала заметим, что

$$R_n(k)=R_n(-k)=\sum_{m=-\infty}^{\infty} x(m)\,x(m-k)\,[w(n-m)\,w(n+k-m)]. \tag{4.26}$$

Если обозначить

$$h_k(n)=w(n)\,w(n+k), \tag{4.27}$$

то (4.26) можно переписать в виде

$$R_n(k)=\sum_{m=-\infty}^{\infty} x(m)\,x(m-k)\,h_k(n-m). \tag{4.28}$$

Таким образом, значение автокорреляционной функции при задержке k и в момент n получается путем фильтрации последовательности $x(n)x(n-k)$ в фильтре с импульсной характеристикой $h_k(n)$. Это изображено на рис. 4.23.

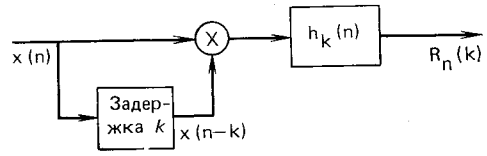


Рис. 4.23. Структурная схема вычисления кратковременной автокорреляции

Вычисление автокорреляционной функции обычно осуществляют, переписав соотношение (4.24) в виде

$$R_n(k)=\sum_{m=-\infty}^{\infty}[x(n+m)\,w'(m)]\,[x(n+m+k)\,w'(k+m)], \tag{4.29}$$

где $w'(n)=w(-n)$. Уравнение (4.29) показывает, что начало отсчета времени во входной последовательности действительно сдвигается к n-му отсчету, после чего она умножается на временно́е окно w' для выделения короткого сегмента речи. Если окно имеет конечную длительность, как в (4.8) и (4.9), то результирующая последовательность также будет иметь конечную длительность и (4.29) запишется в виде

$$R_n(k)=\sum_{m=0}^{N-1-k}[x(n+m)\,w'(m)]\,[x(n+m+k)\,w'(k+m)]. \tag{4.30}$$

Отметим, что при использовании в качестве w' прямоугольного окна или окна Хемминга уравнение (4.30) соответствует нереализуемому фильтру в (4.28). Для окон конечной длительности это, однако, не приводит к затруднениям, поскольку всегда, даже при обработке в реальном масштабе времени, может быть введена необходимая задержка.

Для вычисления автокорреляционной функции при задержке k в соответствии с (4.30) требуется N умножений для вычисления $x(n+m)w'(m)$ и $(N-k)$ умножений и сложений для получения суммы задержанных произведений. Таким образом, для вычисления совокупности значений корреляционной функции, как это нужно при выделении периодичности, требуется выполнять большой объем вычислений. Его можно сократить путем использования специальных свойств уравнения (4.30). Несколько таких примеров описано в приложении.

В отличие от (4.30) соотношение (4.28) можно использовать и при вычислении совокупности значений корреляционной функции,

так как при правильном выборе окна $R_n(k)$ можно вычислять рекуррентно (см. задачу 4.7).

На рис. 4.24 представлены три примера автокорреляционных функций, вычисленных по речевому сигналу, дискретизированному

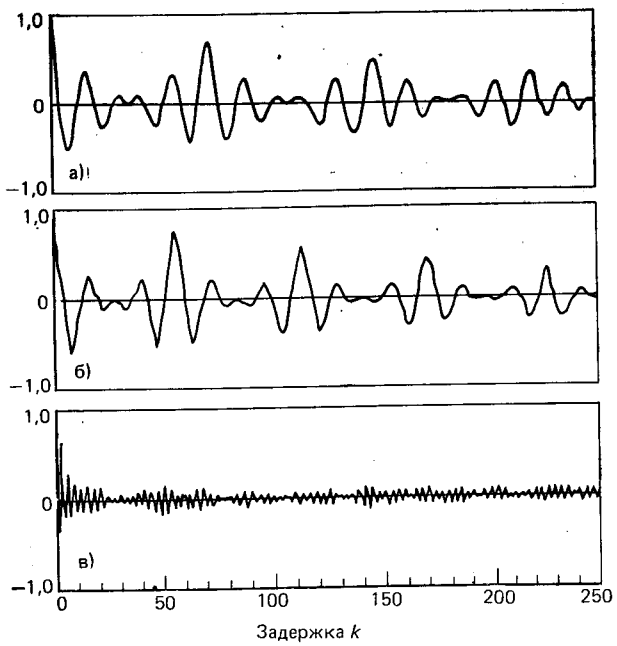


Рис. 4.24. Автокорреляционная функция вокализованной речи (*а*, *б*) и невокализованной речи (*в*), полученная с прямоугольным окном длительностью $N=401$

с частотой 10 кГц с использованием (4.30) при $N=401$. Как видно из рисунка, автокорреляционные функции вычислены для задержек $0 \leqslant k \leqslant 250$. Первые два примера соответствуют сегментам вокализованной, а последний — невокализованной речи[1]. Для первого сегмента пики корреляции возникают при задержках, кратных 72, что указывает на наличие периода, равного 7,2 мс, или частоты основного тона, равной примерно 140 Гц. Заметим, что даже очень короткий сегмент речи отличается от сегмента строго периодического сигнала. В течение интервала длительностью 401 отсчет и «период» сигнала и его форма изменяются. Это одна из причин, по которой пики уменьшаются по амплитуде с ростом задержки. Для другого вокализованного сегмента (взятого в другом месте фразы) видна сходная периодичность, только теперь пики возникают при задержках, кратных 58, что показывает наличие основного тона с периодом 5,8 мс. Наконец, в автокорреляционной

[1] Здесь и на последующих рисунках автокорреляционная функция нормирована таким образом, что $R_n(0)=1$.

функции для невокализованной речи отсутствует ярко выраженная периодичность, что говорит о непериодическом характере сигнала в данном случае. Видно, что автокорреляционная функция невокализованной речи представляет собой шумоподобное колебание, напоминающее речевой сигнал, по которому оно вычислено.

На рис. 4.25 приведены примеры применения временно́го окна Хемминга. Сравнивая эти результаты с соответствующими результатами на рис. 4.24, отметим, что прямоугольное окно значительно сильнее выявляет периодичность в сигнале, чем окно Хемминга. Этот результат не покажется неожиданным, если учесть, что окно Хемминга вносит затухание на концах обрабатываемого сегмента речи.

Примеры на рис. 4.24 и 4.25 рассчитаны для $N=401$. Важным вопросом является выбор N для надежного обнаружения периодичности. Здесь мы вновь сталкиваемся с противоречивыми требо-

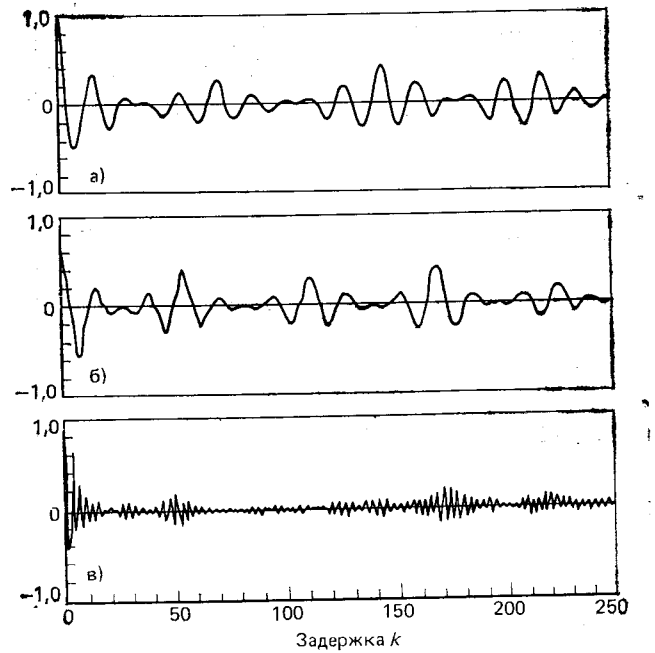


Рис. 4.25. Автокорреляционная функция вокализованной речи (*а*, *б*) и невокализованной речи (*в*), полученная с окном Хемминга длительностью $N=401$

ваниями. Вследствие изменения свойств речевого сигнала N следовало бы выбирать как можно меньше. С другой стороны очевидно, что для измерения периодичности по автокорреляционной функции окно должно иметь длительность, равную по крайней мере двум периодам основного тона. Фактически из-за конечной длительности взвешенного речевого сигнала, используемого при

вычислении $R_n(k)$, по мере увеличения k используется все меньше и меньше данных [см. верхний предел в сумме (4.30)]. Это приводит к уменьшению амплитуды пиков автокорреляционной функции при увеличении k, что видно в случае периодической импульсной последовательности (см. задачу 4.8) и легко может быть продемонстрировано на речевом сигнале. На рис. 4.26 показано влияние прямоугольного окна различной длительности на вид корреляционной функции. Пунктиром изображена функция

$$R(k) = 1 - k/N, \quad |k| < N, \tag{4.31}$$

которая представляет собой автокорреляционную функцию прямоугольного временно́го окна. Очевидно, что эта пунктирная линия является границей значений амплитуды автокорреляционной функции сигнала. В задаче 4.8 показано, что для периодической последовательности максимумы будут лежать точно на этой линии. В

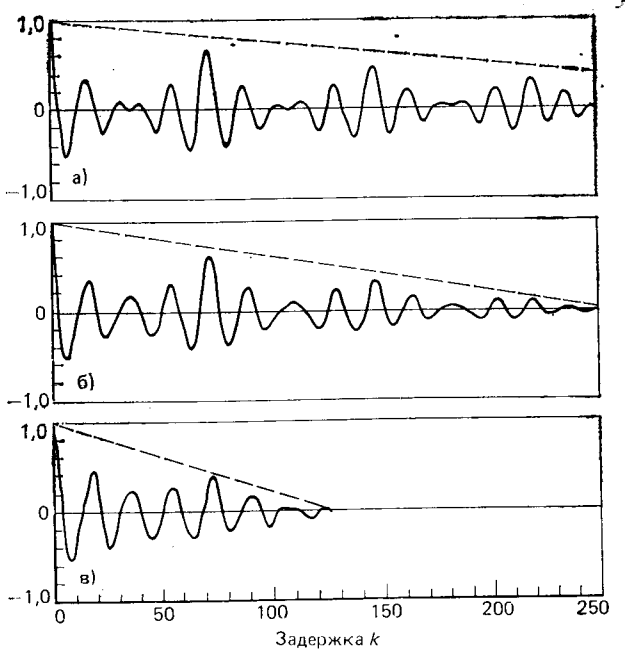


Рис. 4.26. Автокорреляционная функция вокализованной речи, полученная с прямоугольным окном при $N=401$ (*а*), 251 (*б*) и 125 (*в*)

рассматриваемом примере максимумы более отдалены от нее при $N=401$, чем в двух других случаях. Это происходит из-за того, что форма сигнала и период основного тона больше изменяются на интервале, равном 401 отчету, чем на более коротких интервалах. В этом состоят причины общего затухания амплитуды корреляционной функции.

Рисунок 4.26*в* соответствует окну длительностью 72 отсчета. Поскольку период основного тона в этом примере того же порядка, окно не охватывает даже двух полных периодов. Это случай, которого следует избегать. Но сделать это очень трудно из-за широкого диапазона значений периода основного тона. Один способ состоит в выборе столь длительного окна, чтобы оно охватывало даже наибольший период основного тона. Но это, очевидно, приведет к нежелательному усреднению большого числа периодов, когда период основного тона мал. Другой подход состоит в изменении длительности окна в соответствии с ожидаемым периодом основного тона. Еще один подход, позволяющий использовать короткие окна, состоит в модификации определения автокорреляционной функции.

Модифицированная кратковременная корреляционная функция определяется выражением

$$\hat{R}_n(k) = \sum_{m=-\infty}^{\infty} x(m)\, w_1(n-m)\, x(m+k)\, w_2(n-m-k). \tag{4.32}$$

Это выражение можно переписать в виде

$$\hat{R}_n(k) = \sum_{m=-\infty}^{\infty} x(n+m)\, \hat{w}_1(m)\, x(n+m+k)\, \hat{w}_2(m+k), \tag{4.33}$$

где

$$\hat{w}_1(m) = w_1(-m), \tag{4.34а}$$

$$\hat{w}_2(m) = w_2(-m). \tag{4.34б}$$

Для того чтобы исключить затухание, обусловленное переменным верхним пределом в (4.30), выберем окно $\hat{w}_2$ таким, чтобы оно включало отсчеты вне ненулевого интервала окна w_1, т. е. определим временны́е окна в виде

$$\hat{w}_1(m) = \begin{cases} 1, & 0 \leqslant m \leqslant N-1 \\ 0, & \text{в противном случае} \end{cases} \tag{4.35а}$$

и

$$\hat{w}_2(m) = \begin{cases} 1, & 0 \leqslant m \leqslant N-1+K \\ 0, & \text{в противном случае} \end{cases} \tag{4.35б}$$

где K — наибольшая требуемая задержка. Таким образом, уравнение (4.33) можно переписать в виде

$$\hat{R}_n(k) = \sum_{m=0}^{N-1} x(n+m)\, x(n+m+k) \quad 0 \leqslant k \leqslant K, \tag{4.36}$$

т. е. усреднение осуществляется по всем N отсчетам, включая отсчеты вне интервала от n до $n+N-1$. Различия в исходных данных для (4.30) и (4.36) изображены на рис. 4.27. На рис. 4.27*а* изображен исходный речевой сигнал, а на рис. 4.27*б* N отсчетов, выделенных с помощью прямоугольного временно́го окна. В слу-

чае прямоугольного окна этот сегмент будет использован в обоих сомножителях в (4.30) и будет сомножителем $x(n+m)w_1(m)$ в (4.36). На рис. 4.27*в* изображен другой сомножитель в (4.36), включающий K дополнительных отсчетов.

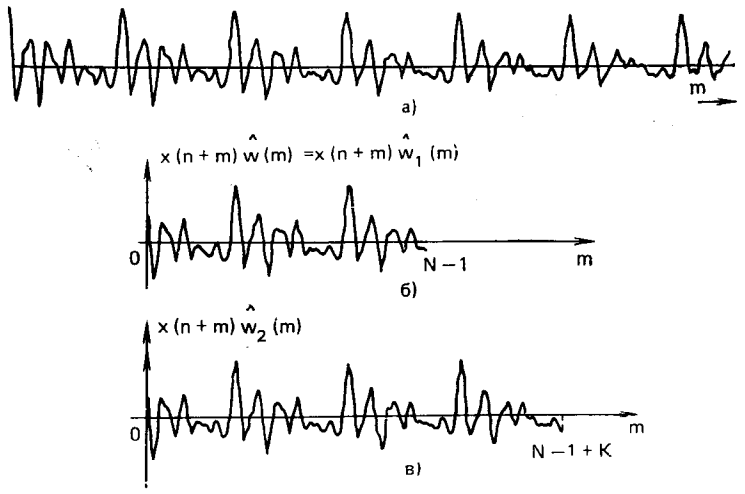


Рис. 4.27. Отсчеты, применяемые для вычисления кратковременной автокорреляционной функции

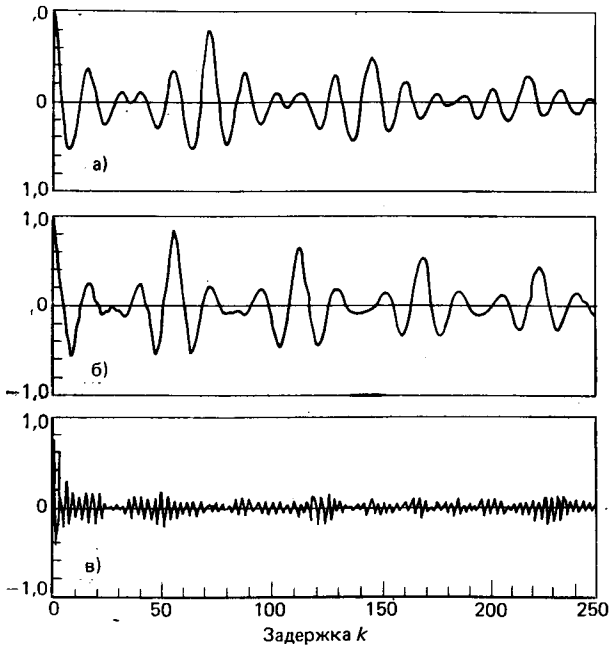


Рис. 4.28. Модифицированная автокорреляционная функция вокализованной речи для сегментов рис. 4.24 при $N=401$

Выражение (4.36) далее называется *модифицированной* кратковременной корреляционной функцией. Однако, строго говоря, это *взаимная* корреляционная функция двух сегментов речи $x(n+m)\hat{w}_1(m)$ и $x(n+m)\hat{w}_2(m)$. Таким образом, $\hat{R}_n(k)$ имеет свойства взаимной корреляционной функции, а не автокорреляционной, например, $\hat{R}_n(-k)\neq\hat{R}_n(k)$. Тем не менее $\hat{R}_n(k)$ имеет пики при задержках, кратных периоду сигнала, и эти пики не затухают с ростом k. На рис. 4.28 представлены модифицированные автокорреляционные функции, соответствующие примерам рис. 4.24. Поскольку при $N=401$ эффект изменения формы сигнала преобладает над краевым эффектом на рис. 4.24, то оба рисунка выглядят почти одинаково. Сравнение рис. 4.29 и 4.26 показывает, что различия более заметны для малых значений N. Очевидно, что пики

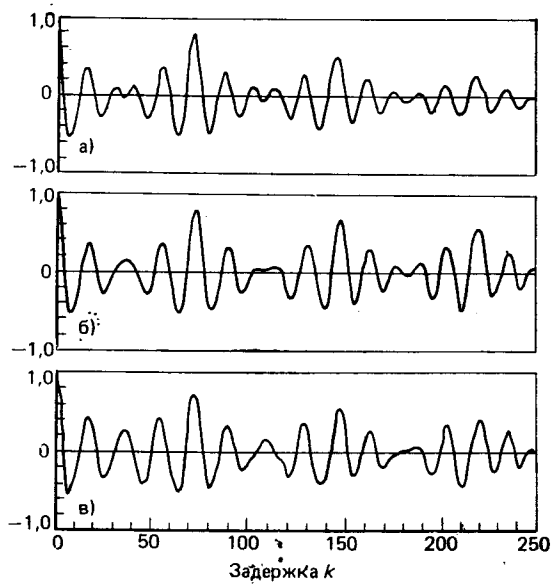


Рис. 4.29. Модифицированная автокорреляционная функция вокализованной речи при $N=401$ (*а*), 251 (*б*) и 125 (*в*) (см. рис. 4.26)

на рис. 4.29 меньше, чем при $k=0$, только вследствие изменения периодичности на интервале от n до $n+N-1+K$, который используется в уравнении (4.36). В задаче 4.8 показано, что для строго периодической последовательности все пики будут иметь одинаковые амплитуды.

4.7. Кратковременная функция среднего значения разности

Как отмечалось выше, вычисление автокорреляционной функции требует выполнения большого числа арифметических операций, даже при использовании упрощений, изложенных в приложе-

нии. Метод, исключающий необходимость умножений, основан на том, что для строго периодической функции с периодом P последовательность

$$d(n)=x(n)-x(n-k) \tag{4.37}$$

будет равна нулю при $k=0\pm P,\ \pm 2P, \ldots$ Для сегментов вокализованного речевого сигнала естественно ожидать, что последовательность $d(n)$ будет близка к нулю (но не равна ему) при k, кратном периоду основного тона. Кратковременное среднее значение величины $d(n)$ как функция k будет мало, если k близко к периоду основного тона. Кратковременная функция среднего значения разности (КФСР) определяется как

$$\gamma_n(k)=\sum_{m=-\infty}^{\infty}|x(n+m)\,w_1(m)-x(n+m-k)\,w_2(m-k)|. \tag{4.38}$$

Очевидно, что если $x(n)$ близка к периодической функции на интервале, выделенном с помощью временно́го окна, то $\gamma_n(k)$ будет иметь глубокие провалы при $k=P,\ 2P\ldots$ Отметим, что целесообразно применять прямоугольные окна. Если оба окна имеют одинаковую длительность, то получается функция, сходная с автокорреляционной функцией (4.30). Если длительность $w_2(n)$ превышает длительность $w_1(n)$, то ситуация аналогична вычислению модифицированной автокорреляции (4.36). Можно показать [16], что

$$\gamma_n(k)\approx\sqrt{2}\,\beta(k)\,[\hat{R}_n(0)-\hat{R}_n(k)]^{1/2}, \tag{4.39}$$

причем $\beta(k)$ в (4.39) изменяется в пределах от 0,6 до 1,0 на различных сегментах речи, но слабо зависит от k.

На рис. 4.30 представлена КФСР для речевых сегментов рис. 4.24 и 4.28 при окне той же протяженностью. Легко видеть, что КФСР действительно имеет вид (4.39); функция содержит глубокие провалы в точках, кратных периоду основного тона, для вокализованной речи, и не имеет заметных провалов для невокализованной речи.

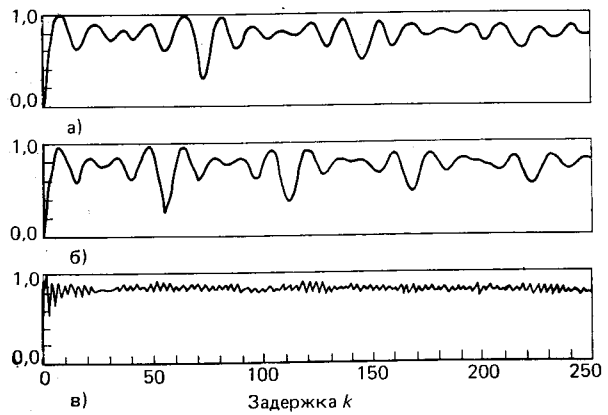


Рис. 4.30. Функция КФСР (нормированная) для сегментов речи рис. 4.24, 4.28

Для получения КФСР требуется выполнить операции сложения, вычитания и вычисления модуля, в то время как для вычисления автокорреляционной функции требуется выполнить операции сложения и умножения. При использовании системы счисления с плавающей запятой, где на операцию сложения и умножения требуется приблизительно одно и то же время, оба метода при одинаковых окнах сравнимы по быстродействию. Однако с точки зрения технической реализации, когда применяется система счисления с фиксированной запятой, КФСР имеет некоторые преимущества. В этом случае на выполнение операции умножения требуется значительно больше времени, чем операции сложения. Кроме того, для вычисления суммы произведений требуется либо масштабирование, либо удвоенная точность вычислений. С этой точки зрения именно КФСР использовалась в цифровых системах обработки речи, функционирующих в реальном масштабе времени.

4.8. Оценивание периода основного тона по автокорреляционной функции

Как показано в § 4.6, кратковременная автокорреляционная функция является удобной характеристикой сигнала, по которой можно производить текущее оценивание периода основного тона. В данном параграфе рассматривается несколько особенностей применения автокорреляционных выделителей основного тона.

Одно из основных ограничений применения автокорреляционной функции состоит в том, что она «содержит» излишне много сведений о сигнале. (В гл. 8 показано, что по 10—12 значениям автокорреляционной функции можно достаточно точно оценить передаточную функцию голосового тракта.) В результате (см. рис. 4.26) автокорреляционная функция имеет много побочных пиков. Большинство этих пиков обусловлено откликом голосового тракта, состоящего из затухающих колебаний, которые определяют форму речевого колебания на каждом периоде основного тона. На рис. 4.26*а* и *б* пик, соответствующий периоду основного тона, имеет наибольшую амплитуду, однако на рис. 4.26*в* пик в точке $k=15$ фактически больше, чем в точке $k=72$. Так получилось из-за малой протяженности окна по сравнению с периодом основного тона. Быстрое изменение форматных частот также может привести к подобному эффекту. В том случае, если побочные пики автокорреляционной функции превышают пик основного тона, простая процедура выделения наибольшего пика приведет к ошибкам.

Для устранения этих затруднений целесообразно так обработать речевой сигнал, чтобы подчеркнуть его периодичность и устранить несущественные в данном случае особенности его тонкой структуры. Этот подход, введенный в § 4.5, допускает использова-

ние очень простого выделителя основного тона. Методы, которые используют этот вид обработки сигнала, иногда называют методами «выравнивания спектра», поскольку они основаны на устранении влияния передаточной функции речевого тракта. При этом каждая гармоника обработанного сигнала имеет одинаковую амплитуду. Предложен ряд методов спектрального сглаживания [17], однако в данном случае наиболее удобным оказался метод центрального ограничения.

В методе, предложенном Сондхи [17], центральное ограничение речи осуществляется посредством нелинейного преобразования:

$$y(n) = C[x(n)], \tag{4.40}$$

где функция $C[\cdot]$ изображена на рис. 4.31. Рисунок 4.32 иллюстрирует обработку речевого сигнала с использованием центрального ограничения. Сегмент речевого сигнала, используемый при

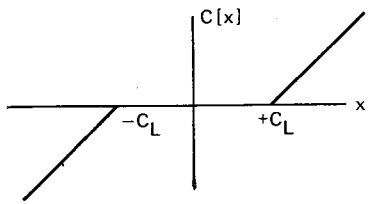


Рис. 4.31. Центральное ограничение

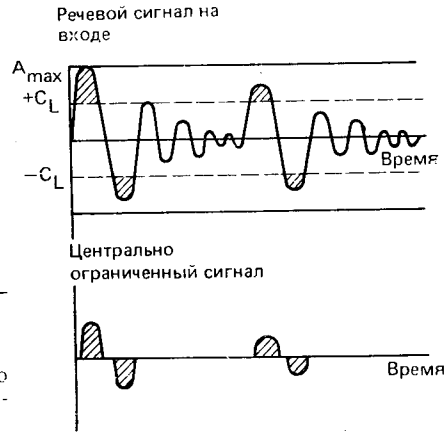


Рис. 4.32. Пример обработки речевого сигнала с помощью центрального ограничения [17]

вычислении автокорреляционной функции, показан сверху на рис. 4.32. Для этого сегмента определяется максимальная амплитуда A_{max} и уровень ограничения C_L устанавливается как некоторый процент от A_{max} (в [17] — 30%). Как видно из рис. 4.31, для отсчетов, превышающих C_L, выходной сигнал центрального ограничителя равен разности входного сигнала и уровня ограничения. Для отсчетов, меньших уровня ограничения, сигнал на выходе равен нулю. Нижняя кривая на рис. 4.32 — сигнал на выходе ограничителя. В отличие от метода, изложенного в § 4.5, где пики сигнала отображались импульсами, в данном случае преобразованный сигнал состоит из части входного сигнала, превысившей порог ограничения.

Рисунок 4.33 [18] иллюстрирует процесс вычисления автокорреляционной функции с помощью центрального ограничения. На рис. 4.33*а* показан вокализованный сегмент сигнала длиной 300 отсчетов ($F_s = 10$ кГц). На автокорреляционной функции, показан-

ной справа, имеется четкий пик, соответствующий основному тону. Однако имеется и много побочных пиков, обусловленных затухающими колебаниями в голосовом тракте. На рис. 4.33*в* изображен сигнал после центрального ограничения, где уровень ограничения установлен так, как это показано на рис. 4.33*в* (в данном случае

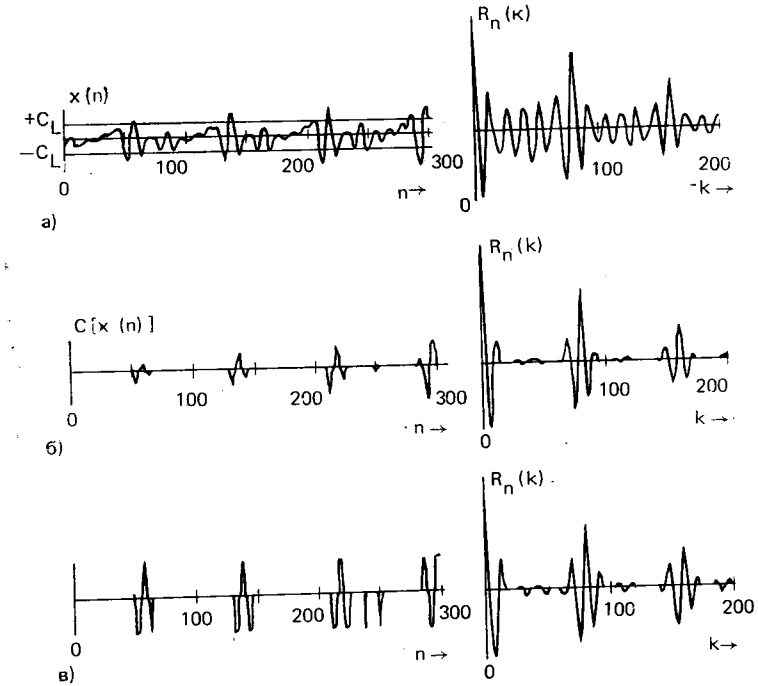


Рис. 4.33. Временна́я диаграмма речевого сигнала и корреляционная функция: *а*) без ограничения; *б*) с центральным ограничением; *в*) с трехуровневым центральным ограничением (Функции корреляции нормированы) [18]

он составляет 68% от максимальной амплитуды на первых 100 отсчетах). Отметим, что все, что остается при этом от речевого сигнала, представляет собой несколько импульсов, расположенных там же, где и исходные импульсы основного тона. Автокорреляционная функция имеет теперь значительно меньше побочных пиков, что уменьшает вероятность ошибки.

Рассмотрим эффект ограничения уровня. При высоком уровне ограничения количество отсчетов, превышающих его, будет небольшим. Это приведет к малому количеству посторонних пиков в автокорреляционной функции, что иллюстрируется рис. 4.34, на котором изображена автокорреляционная функция сегмента речи рис. 4.26*а* для уменьшающихся уровней ограничения. Очевидно, что по мере уменьшения уровня ограничения через ограничитель проходит больше пиков, что приводит к усложнению формы автокорреляционной функции (случай, когда уровень ограничения ра-

вен нулю, соответствует рис. 4.26*а*). Из данного рисунка можно сделать вывод, что наиболее надежное оценивание периода основного тона достигается при наиболее высоком уровне ограничения. Использование слишком высокого уровня может, однако, также привести к трудностям. Может случиться так, что амплитуда речевого сигнала будет заметно изменяться на протяжении сегмента.

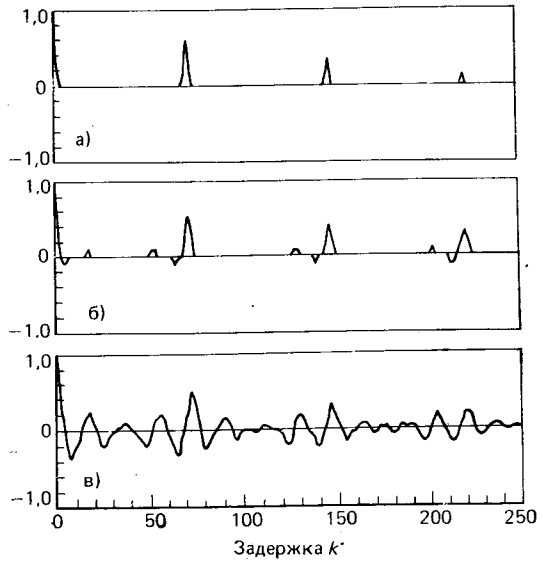


Рис. 4.34. Автокорреляционная функция на выходе центрального ограничителя при $N=401$; *а*) $C_L=80\%$ от максимума; *б*) 64%; *в*) 48% (Сегмент речи соответствует рис. 4.26*а*)

Тогда, если уровень ограничения выставлен по наибольшей амплитуде сигнала, бо́льшая часть колебания окажется ниже уровня и будет потеряна. Поэтому Сондхи выбрал величину уровня ограничения как 30% от максимального значения. В методах, позволяющих использовать более высокие уровни (от 60 до 80%), определяется максимальное значение сигнала на первой и последней трети обрабатываемого сегмента и затем в качестве уровня ограничения выбирается меньшее из этих значений. Этот способ иллюстрируется рис. 4.33*б*.

Решение задачи устранения побочных максимумов корреляционной функции может быть существенно облегчено применением центрального ограничения перед вычислением автокорреляционной функции. Однако другая трудность при автокорреляционном спи-сании сигнала (которая не устраняется даже при центральном ограничении) состоит в большом объеме вычислений. Простая модификация функции центрального ограничения приводит к значительному упрощению автокорреляционной функции, практически без потерь точности оценивания основного тона. Модифицирован-

ная функция представлена на рис. 4.35. Выходной сигнал ограничителя, как это следует из рис. 4.35, равен $+1$, если $x(n)>C_L$, и равен -1, если $x(n)<-C_L$. Во всех других случаях он равен нулю. Устройство, описываемое данной функцией, будем далее называть трехуровневым центральным ограничителем. На рис. 4.33*в* по-

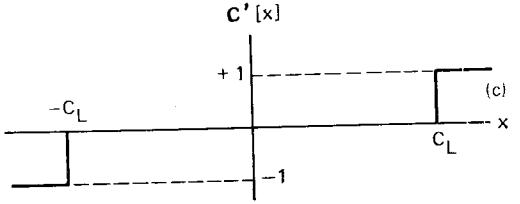


Рис. 4.35. Трехуровневое центральное ограничение

казан сигнал на выходе трехуровневого центрального ограничителя в случае, когда на вход поступает сигнал рис. 4.33*а*. Несмотря на то что при этом все пики, превысившие уровень ограничения, имеют одну и ту же амплитуду, корреляционная функция сигнала практически не отличается от изображенной на рис. 4.33*б*. Таким образом, и в этом случае подавляется большинство побочных пиков, что позволяет точно оценить период основного тона.

Вычисление автокорреляционной функции сигнала на выходе трехуровневого ограничителя отличается простотой. Если обозначить выходной сигнал через $y(n)$, то слагаемое $y(n+m)y(n+m+k)$ в автокорреляционной функции

$$R_n(k)=\sum_{m=0}^{N-k-1} y(n+m)\,y(n+m+k) \tag{4.41}$$

может принимать только три различных значения:

$$y(n+m)\,y(n+m+k)=\begin{cases}0, & y(n+m)=0 \text{ или } y(n+m+k)=0\\ +1, & y(n+m)=y(n+m+k),\\ -1, & y(n+m)\neq y(n+m+k).\end{cases} \tag{4.42}$$

Таким образом, при технической реализации алгоритма вычислений требуется выполнить лишь логические операции и накопление значений автокорреляционной функции в реверсивном счетчике для каждого k.

Обсуждая дальнейшие детали технической реализации, отметим, что целесообразно использовать модифицированную автокорреляционную функцию, определенную выражением (4.36), как в случае трехуровневого ограничения, так и при другом центральном ограничении, поскольку в этом случае пики автокорреляционной функции не уменьшаются при увеличении задержки. При вычислении КФСР по (4.38) сигнал также может быть подвергнут одному из видов ограничения. В действительности существует множество комбинаций, которые можно использовать в разных ситуациях [18].

Алгоритмов оценивания тона по кратковременной корреляционной функции уже предложено много и несомненно, что будет пред-

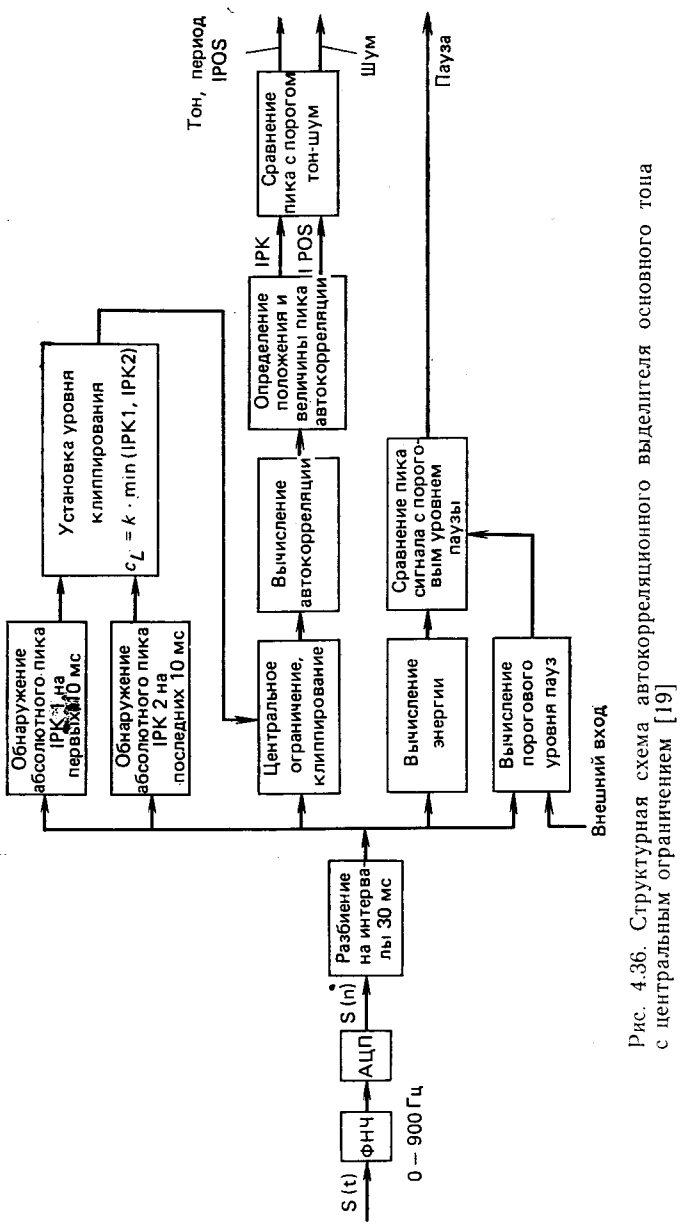


Рис. 4.36. Структурная схема автокорреляционного выделителя основного тона с центральным ограничением [19]

ложено еще больше. Мы завершим этот параграф рассмотрением одного способа, который был реализован в цифровой форме [19]. Подробный алгоритм представлен на рис. 4.36, а его краткое описание приводится ниже.

1. Речевой сигнал пропускается через фильтр нижних частот с частотой среза 900 Гц и дискретизируется с частотой 10 кГц.

2. Выделяются сегменты длительностью 30 мс (300 отсчетов) через каждые 10 мс. Таким образом, соседние сегменты имеют перекрытие 20 мс.

3. Вычисляется среднее значение (4.12) с прямоугольным окном протяженностью 100 отсчетов. Максимальное значение сигнала на каждом сегменте сравнивается с порогом, вычисленным по сегменту шумового фона длительностью 50 мс. Если пиковый уровень сигнала превосходит порог, то принимается решение о наличии речевого сигнала на сегменте и продолжается его обработка, в противном случае сегмент классифицируется как пауза и дальнейшая обработка не производится.

4. Уровень ограничения составляет фиксированный процент (68%) минимального значения двух максимумов сигнала, рассчитанных по первым и последним 100 отсчетам.

5. Речевой сигнал преобразуется в трехуровневом ограничителе и далее вычисляется автокорреляционная функция в области предполагаемого значения периода основного тона.

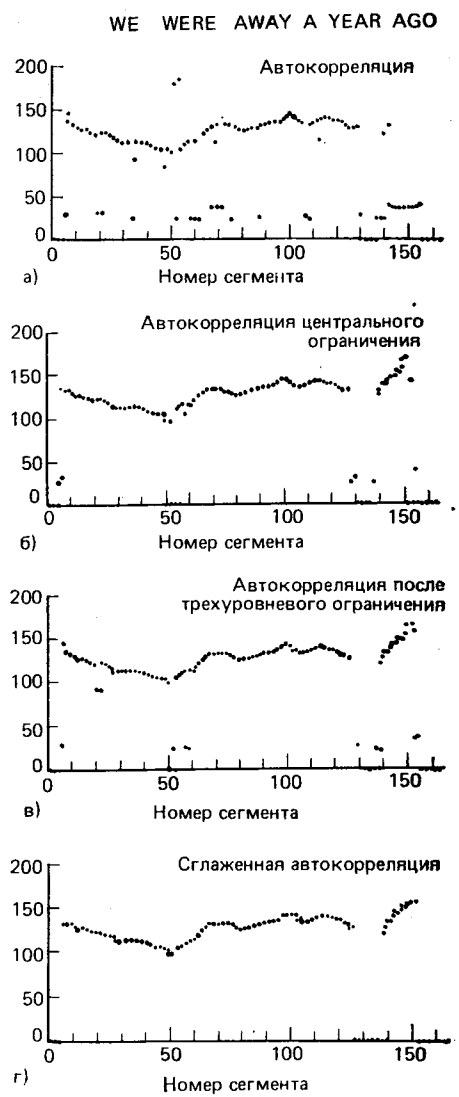


Рис. 4.37. Выходной сигнал автокорреляционного выделителя основного тона: *а*) без ограничения; *б*) с центральным ограничением; *в*) с трехуровневым центральным ограничением (см. рис. 4.35); *г*) с нелинейным сглаживанием сигнала (*в*) [18]

6. Определяется максимальный пик автокорреляционной функции и сравнивается с порогом, который составляет примерно 30% от $R_n(0)$. Если этот пик оказался меньше порога, то сегмент классифицируется как невокализованный, а в противном случае — как вокализованный с периодом основного тона, соответствующим положению этого пика. Выше приведен алгоритм, реализованный в цифровом виде [19]. Однако имеется возможность значительно изменить вычислительный процесс. Например, на шагах 4 и 5 алгоритма можно использовать центральное ограничение и обычную процедуру вычисления автокорреляционной функции или вообще устранить всякое ограничение. Другая возможность заключается в использовании КФСР (и поиске минимумов, а не максимумов) как с каким-либо ограничением, так и без него.

На рис. 4.37 изображена кривая основного тона для трех вариантов алгоритма. На рис. 4.37*а* приведена траектория основного тона для случая вычисления автокорреляционной функции без какого-либо ограничения. Следует отметить наличие значительного разброса оценок вследствие ошибок в оценивании, обусловленных, очевидно, тем, что побочные пики оказываются больше, чем пики в области основного тона. Поскольку нормированный период основного тона в среднем расположен между 100 и 150, уменьшение автокорреляционной функции вызывает значительное затухание пика в области основного тона. Поэтому побочные пики автокорреляционной функции оказываются больше основного. На рис. 4.37*б* и *в* иллюстрируется применение центрального ограничения и трехуровневого центрального ограничения при вычислении автокорреляционной функции. Видно, что большинство ошибок здесь устранено и, более того, между этими двумя результатами нет существенного различия. Небольшое количество ошибок остается в обоих случаях. Эти ошибки можно эффективно устранить путем применения нелинейного сглаживания с помощью метода, излагаемого ниже. Пример показан на рис. 4.37*г*.

4.9. Медианное сглаживание и обработка речи

Часто для сглаживания шумоподобной компоненты в сигнале используются линейные фильтры. Однако для некоторых приложений, вследствие особенностей обрабатываемых данных, линейное сглаживание не является адекватным. Примером может служить траектория основного тона на рис. 4.37*в*, содержащая очевидные ошибки, которые следует устранить, переместив ошибочные точки на основную траекторию. Обычный линейный фильтр не только не устранит этих ошибок, но и внесет искажения на резких переходах от вокализованного сегмента к невокализованному. Последнему соответствует нулевое значение периода основного тона. В таких случаях требуется нелинейный алгоритм обработки, устраняющий большие ошибки и не приводящий к значительным искажениям. Хотя идеального алгоритма такого типа не существует, можно показать, что комбинация алгоритма вычисления медианы и линейного сглаживания (впервые предложенная Тьюки [20]) обладает требуемыми свойствами [21].

Сущность линейного сглаживания состоит в разделении таких сигналов, спектры которых почти не пересекаются. Для нелинейного сглаживания целесообразно ввести разделение с учетом того, является ли сигнал «гладким» или шумоподобным. Таким образом, сигнал можно представить в виде

$$x(n) = S[x(n)] + R[x(n)], \tag{4.43}$$

где $S[x]$ — гладкая компонента, а $R[x]$ — шумоподобная компонента сигнала x. Нелинейностью, позволяющей разделить эти компоненты, является текущая медиана сигнала. Выходным сигналом устройства медианного сглаживания $M_L[x(n)]$ является медиана L отсчетов $x(n), \ldots, x(n-L+1)$. Текущие медианы протяженностью L обладают рядом полезных для сглаживания свойств:

1. $M_L[\alpha x(n)] = \alpha M_L[x(n)]$.
2. Медианы не «смазывают» основных разрывов в сигнале, если сигнал не имеет других разрывов среди $L/2$ отсчетов.
3. Медианы приблизительно повторяют полиномиальный тренд низкого порядка. Следует отметить, что алгоритмы медианного сглаживания, как и другие нелинейные алгоритмы обработки, не обладают принципом суперпозиции, т. е.

$$M_L[\alpha x_1(n) + \beta x_2(n)] \neq \alpha M_L[x_1(n)] + \beta M_L[x_2(n)]. \tag{4.44}$$

Хотя в общем случае медианы сохраняют резкие разрывы в сигнале, применение такой обработки позволяет значительно сгладить нежелательные шумоподобные компоненты сигнала. Хорошие результаты дает совместное использование линейных методов и методов медианного сглаживания. Поскольку текущие медианы обеспечивают некоторое сглаживание сигнала, линейная система может быть низкого порядка. Например, вполне подходит фильтр с импульсной характеристикой

$$h(n) = \begin{cases} 1/4 & n = 0, \\ 1/2 & n = 1, \\ 1/4 & n = 2. \end{cases} \tag{4.45}$$

На рис. 4.38*а* приведена структурная схема системы совместного сглаживания. Сигнал на выходе устройства приблизительно равен

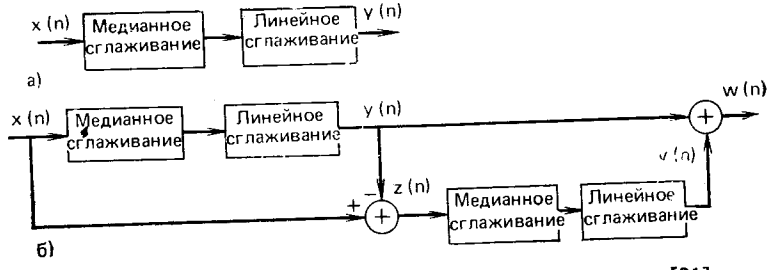


Рис. 4.38. Структурная схема нелинейного сглаживания [21]

$S[x(n)]$. Так как приближение не идеальное, далее применяется повторное сглаживание рис. 4.38*б*. Поскольку

$$y(n) = S[x(n)], \tag{4.46}$$

то

$$z(n) = x(n) - y(n) = R[x(n)]. \tag{4.47}$$

Повторное нелинейное сглаживание $z(n)$ дает корректирующий сигнал, который прибавляется к $y(n)$ для получения $w(n)$, более точно описывающей $S[x(n)]$. Сигнал $w(n)$ удовлетворяет соотношению

$$w(n) = S[x(n)] + S[R[x(n)]]. \tag{4.48}$$

Если $z(n) = R[x(n)]$ точно, т. е. устройство нелинейного сглаживания идеально, то $S[R[x(n)]]$ равно нулю и корректирующее слагаемое будет отсутствовать.

При использовании устройства нелинейного сглаживания с алгоритмом рис. 4.38 следует учитывать задержки в каждой его ветви. Медианное сглаживание вносит задержку на $(L-1)/2$ отсчетов, а линейное — в соответствии с импульсной характеристикой фильтра. Например, устройство медианного сглаживания на пять отсчетов вносит задержку в два отсчета, а окно Хемминга на три отсчета — задержку на один отсчет. На рис. 4.39 изображен реализуемый вариант устройства рис. 4.38*б*.

Наконец, при технической реализации устройства медианного сглаживания (рис. 4.39) надо установить граничные условия, т. е.

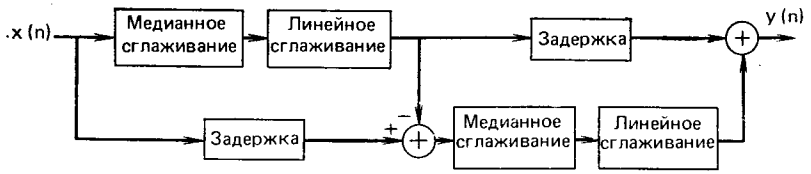


Рис. 4.39. Система нелинейного сглаживания с компенсирующей задержкой [21]

определить, как вычислять текущую медиану в начале и конце сегмента сигнала. Хотя возможны различные подходы к выбору граничных условий, для речевого сигнала целесообразно доопределить сигнал за пределы сегмента, полагая его периодически повторяющимся.

На рис. 4.40 показаны результаты применения различных устройств сглаживания к функции среднего числа нулевых пересечений. Исходная функция (рис. 4.40*а*) имеет шумоподобную компоненту из-за малого времени усреднения. На рис. 4.40*г* показан выходной сигнал устройства медианного сглаживания (последовательное пятиточечное и трехточечное сглаживание). Можно заметить, что сигнал имеет квазипрямоугольную форму, что обусловлено наличием высокочастотных компонент в сглаживаемом сигнале.

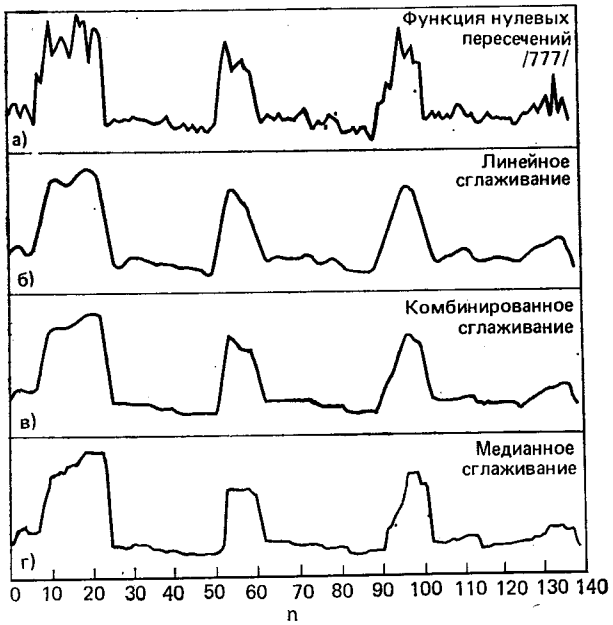


Рис. 4.40. Пример нелинейного сглаживания функции нулевых пересечений [21]

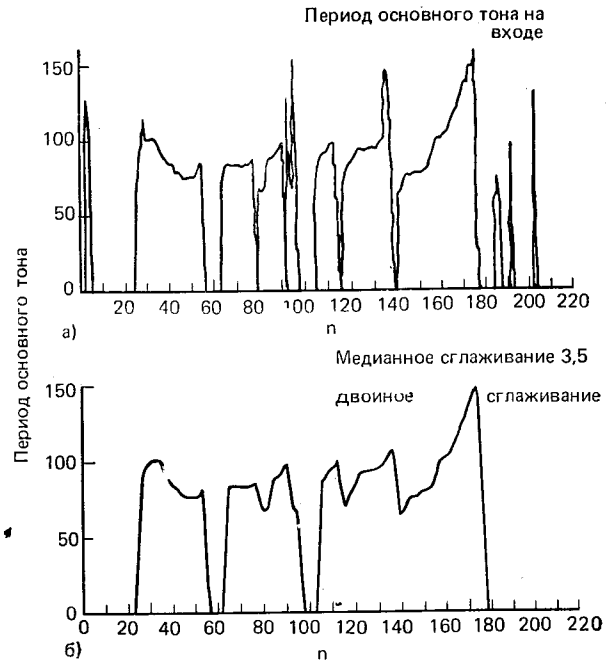


Рис. 4.41. Пример нелинейного сглаживания траектории основного тона [21]

Входной сигнал устроиства линейного сглаживания (КИХ-фильтр нижних частот 19-го порядка) изображен на рис. 4.40*б* и иллюстрирует «смазывание» сигнала при возникновении в нем резких переходов. На рис. 4.40*в* показан сигнал на выходе устройства при совместном использовании медианного сглаживания, описанного выше, и линейного сглаживания с трехточечным временны́м окном типа (4.45). В данном случае оценка хорошо следует за изменениями входного сигнала и бо́льшая часть шума устранена.

На рис. 4.41 показан пример использования нелинейного сглаживания для обработки траектории основного тона, содержащей ряд очевидных ошибок оценивания. Полезным свойством медианного сглаживания является возможность устранения отдельных ошибок в исходном сигнале совместно со сглаживанием траектории. Как это видно из рис. 4.41, применение совместного сглаживания позволяет устранить значительные ошибки при оценивании и адекватно сгладить сигнал, сохранив резкие переходы тон/шум.

4.10. Заключение

В гл. 4 рассмотрены некоторые характеристики речевого сигнала, полученные путем обработки во временно́й области. Мы подробно изучили методы обработки, так как они широко применяются при анализе речи, и надеемся, что подробное изложение будет способствовать их эффективному использованию. В данную главу включены также несколько примеров устройств обработки речевых сигналов на основе совместного использования функций кратковременной энергии, переходов через нуль и автокорреляционной функции речевого сигнала. Это сделано с тем, чтобы показать, как используются основные методы обработки сигналов при построении систем обработки речи.

ПРИЛОЖЕНИЕ. СОКРАЩЕНИЕ ОБЪЕМА ВЫЧИСЛЕНИЙ ПРИ РАСЧЕТЕ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ

Вычисление K значений автокорреляционной функции по N-точечному окну требует по крайней мере KN операций умножения и сложения. Поскольку для многих практических приложений как K, так и N велики (например, $K=250$ и $N=401$), для уменьшения объема вычислений желательно использовать некоторые свойства автокорреляционной функции. Рассмотрим три метода, позволяющие сократить число арифметических операций при вычислении автокорреляционной функции.

Первый метод предложенный Бланкеншипом [22], основан на том, что при $m\neq 0$ большинство входных отсчетов используется при умножении дважды, т. е. для модифицированной автокорреляционной функции при $k=1$ получаем

$$\begin{aligned}\hat{R}_n(1) = &\sum_{m=0}^{N-1} x(m+n)\,x(m+n+1) = \\ &= x(n)\,x(n+1) + x(n+1)\,x(n+2) + \ldots + \\ &+ x(n+N-1)\,x(n+N) = \\ &= x(n+1)\,[x(n) + x(n+2)] + x(n+3)\,[x(n+2) + \\ &+ x(n+4)] + \ldots\end{aligned} \tag{П.1}$$

Таким образом, когда $k\neq 0$, можно, используя выражение (4.36), сократить число умножений без увеличения числа сложений. Формально автокорреляционная функция может быть записана в виде

$$\hat{R}(k) = B(k) + C(k) \tag{П.2}$$

(для упрощения индекс n опущен), где

$$\underset{\text{четная компонента}}{N = 2qk} \qquad \underset{\text{произвольная компонента}}{+ak} \qquad + b \tag{П.3}$$

Здесь $a=0$ или 1 и b находится в области

$$0 \leqslant b < k. \tag{П.4}$$

В уравнении (П2)

$$B(k) = \sum_{j=0}^{q-1} \sum_{i=1}^{k} x(2jk+i+k)\,[x(2jk+i) + x(2jk+i+2k)], \tag{П.5}$$

и если $a=0$, то

$$C(k) = \sum_{i=1}^{b} x(2qk+i)\,x(2qk+i+k) \tag{П.6}$$

или, если $a=1$, то

$$\begin{aligned}C(k) = &\sum_{i=1}^{b} x(2qk+i+k)\,[x(2qk+i) + x(2qk+i+2k)] + \\ &+ \sum_{i=b+1}^{k} x(2qk+i)\,x(2qk+i+k).\end{aligned} \tag{П.7}$$

Например, рассмотрим $N=60$ с $k=6$, 7 и 8. Величины q, a и b в (П3) равны:

k	q	a	b
6	5	0	0
7	4	0	4
8	3	1	4

Поскольку значения q, a и b получены, то можно использовать уравнение (П2) и (П5)—(П7) для вычисления $\hat{R}(k)$. Легко показать, что число умножений, необходимых для вычисления $\hat{R}(k)$, удовлетворяет неравенству.

$$N_M < \frac{1}{2}(N+k). \tag{П.8}$$

Таким образом, при $k \ll N$ этот метод дает примерно двукратное сокращение числа умножений. Если вычислять небольшое число значений автокорреляционной функции (например, при линейном предсказании методами гл. 8), этот метод очень эффективен. Например, Бланкеншип показал, что если $K=12$ и $N=128$, то при прямом вычислении автокорреляционной функции потребуется 1664 умножения ($N_1(K+1)$), в то время как при использовании модифицированной процедуры требуется лишь 912 умножений, т. е. в 1,825 раз меньше. Если требуется вычислить много значений автокорреляционной функции, как в примерах § 4.6, этот метод дает незначительную экономию времени.

Модификация алгоритма предложена Кендаллом [23] в случае определения автокорреляционной функции в соответствии с (4.30). Обозначая взвешенный речевой сигнал $x(n)w(n)$ через $\hat{x}(n)$ и опуская индекс n, перепишем (4.30):

$$R(k) = \sum_{m=0}^{N-1-k} \hat{x}(m)\,\hat{x}(m+k), \tag{П.9}$$

что можно представить, полагая N четным, в виде

$$R(k)=\sum_{m=0}^{(N-k)/2-1}[\hat{x}(2m)+\hat{x}(2m+k+1)][\hat{x}(2m+1)+\hat{x}(2m+k)]-$$
$$-A(k)-B(k) \qquad (\text{П.10})$$

при четном k,

$$R(k)=\sum_{m=0}^{(N-k-1)/2-1}[\hat{x}(2m)+\hat{x}(2m+k+1)][\hat{x}(2m+1)+\hat{x}(2m+k)]-$$
$$-A(k)-B(k)+\hat{x}(N-1-k)\,\hat{x}(N-1), \qquad (\text{П.11})$$

при нечетном k, где $A(k)$ и $B(k)$ получаются с помощью рекурсивных соотношений

$$A(k)=A(k+2)+\hat{x}(N-2-k)\,\hat{x}(N-1-k) \qquad (\text{П.12})$$

при четном k с начальным условием $A(N)=0$,

$$A(k)=A(k+1) \qquad (\text{П.13})$$

при нечетном k и

$$B(k)=B(k+2)+\hat{x}(k)\,\hat{x}(k+1) \qquad (\text{П.14})$$

при четном k с начальным условием $B(N)=0$ и

$$B(k)=B(k+2)+\hat{x}(k)\,\hat{x}(k+1) \qquad (\text{П.15})$$

при нечетном k с начальным условием $B(N-1)=0$. **Уравнения (П10) и (П11)** показывают, что число умножений, необходимое для вычисления $R(k)$, примерно равно $(N-k-1)/2$, т. е. составляет половину от обычно требуемого числа, однако число сложений увеличилось примерно на 50%. Легко видеть, что число умножений в данном случае сократилось для любых k, а не только для $k \ll N$, как в предыдущем случае.

Третий метод ускорения вычислений автокорреляционной функции заключается в применении быстрого преобразования Фурье (БПФ), в котором автокорреляционная функция определяется как обратное преобразование Фурье спектральной плотности мощности последовательности сигнала (т. е. квадрата модуля спектра сигнала) [1, 2, 24]. При этом методе требуется двукратное применение БПФ и операция возведения в квадрат. Для того чтобы избежать наложения при вычислении автокорреляционной функции, применяется 2 N-точечное дискретное преобразование Фурье (вычисляемое с помощью БПФ), в котором N-точечная последовательность заполняется нулями. Процесс вычисления квадрата модуля спектра требует примерно $2N$ умножений, а для получения $2N$ точек БПФ необходимо осуществить $2N\log_2(2N)$ умножений для вычисления всех N значений автокорреляционной функции. Таким образом, использование БПФ для вычисления автокорреляции предполагает выполнение следующего количества операций умножения:

$$N_F=2\cdot 2N\log_2(2N)+2N. \qquad (\text{П.16})$$

Кендалл [23] показал, что прямое модифицированное вычисление автокорреляционной функции при $N \leqslant 256$ более эффективно, чем применение БПФ с точки зрения количества операций умножения. Если при вычислении эффективности учитывать и операции сложения, то прямой метод более эффективен при $N \leqslant 128$.

Задачи

4.1. Прямоугольное окно определяется выражением

$$w_R(n)=\begin{cases}1, & 0\leqslant n\leqslant N-1;\\ 0, & \textit{в противном случае.}\end{cases}$$

Окно Хемминга определяется выражением

$$w_H(n)=\begin{cases}0{,}54-0{,}46\cos[2\pi n/(N-1)], & 0\leqslant n\leqslant N-1;\\ 0, & \textit{в противном случае.}\end{cases}$$

а) Показать, что преобразование Фурье прямоугольного окна имеет вид

$$w_R(e^{i\omega})=\frac{\sin(\omega N/2)}{\sin(\omega/2)}e^{-i\omega(N-1)/2}.$$

б) Изобразить $W_R(e^{i\omega})$ как функцию ω, опуская множитель $e^{-i\omega(N-1)/2}$ с линейной фазой.

в) Выразить $w_H(n)$ через $w_R(n)$ и таким образом получить выражение $W_H(e^{i\omega})$ через $W_R(e^{i\omega})$.

г) Изобразить отдельные члены в $W_H(e^{i\omega})$ (опуская общий для них множитель с линейной фазой.) Рисунок должен иллюстрировать изменения в частотной характеристике окна Хемминга, приводящие к улучшению подавления высших частот.

4.2. Кратковременная энергия последовательности определяется в виде

$$E_n=\sum_{m=-\infty}^{\infty}[x(m)\,w(n-m)]^2.$$

В случае выбора

$$w(m)=\begin{cases}a^m, & m\geqslant 0;\\ 0, & m<0\end{cases}$$

можно получить рекуррентную формулу для E_n.

а) Вывести рекуррентное разностное уравнение для E_n через E_{n-1} и входной сигнал $x(n)$.

б) Изобразить схему цифрового фильтра, соответствующую полученному уравнению.

в) Показать, каким основным свойством должна обладать последовательность $h(m)=w^2(m)$, чтобы можно было получить рекуррентное уравнение.

4.3. Кратковременная энергия сигнала определяется выражением

$$E_n=\sum_{m=-N}^{N}h(m)\,x^2(n-m).$$

Предположим, что мы хотим вычислять E_n по каждому отсчету входного сигнала.

а) Пусть $h(m)$ имеет вид:

$$h(m)=\begin{cases}a^{|m|}, & |m|\leqslant N;\\ 0, & |m|>N.\end{cases}$$

Получить рекуррентное (разностное) уравнение для E_n.

б) Какой выигрыш по числу умножений достигается применением рекуррентных соотношений вместо непосредственного вычисления E_n?

в) Изобразить схему цифрового фильтра, соответствующего рекуррентной формуле для E_n. (Поскольку $h(m)$ нереализуема, следует предусмотреть соответствующую задержку.)

4.4. Предположим, что среднее значение модуля оценивается для каждых L отсчетов входного сигнала. Один из способов состоит в использовании окна конечной длительности

$$M_n=\sum_{m=n-N+1}^{n}|x(m)|\,w(n-m).$$

В этом случае M_n вычисляется только один раз для каждых входных отсчетов. Другой подход состоит в применении окна, для которого сложно получить ре-

куррентную формулу, например $M_n = aM_{n-1} + |x(n)|$. В этом случае M_n необходимо вычислять по каждому отсчету, даже если требуется получить его только для каждых L отсчетов.

а) Определить, сколько операций умножения и сложения требуется для вычисления среднего значения модуля для каждых L отсчетов с использованием окна конечной длительности.

б) Повторить вычисления п. а) для рекуррентного случая.

в) В каких случаях окно конечной длительности оказывается более эффективным с этой точки зрения.

4.5. Среднее число переходов через нуль определяется уравнениями (4.18) и (4.20): $z_n = \frac{1}{2N} \sum_{m=n-N+1}^{n}{}' |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]|$. Показать, что z_n можно представить в виде

$$z_n = z_{n-1} + \frac{1}{2N} \{|\operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n-1)]| - |\operatorname{sgn}[x(n-N)] - \operatorname{sgn}[x(n-N-1)]|\}.$$

4.6. Чтобы показать, как с помощью параллельной обработки добиться высокой точности выделения основного тона, рассмотрите следующий идеализированный случай. Предположим, что имеется семь выделителей основного тона; вероятность правильного оценивания периода равна p, а вероятность ошибки $1-p$. Эти оценки обрабатываются таким образом, что полная ошибка возникает тогда, когда период ошибочно оценен в четырех или более выделителях основного тона.

а) Получить явное выражение для вероятности ошибки при параллельном оценивании основного тона. (Указание: рассмотреть результат работы каждого выделителя в рамках схемы Бернулли с вероятностью ошибки $1-p$ и вероятностью правильного решения p.)

б) Изобразить графически зависимость полной вероятности ошибки от p.

в) Определить, при каком значении p полная вероятность ошибки менее 0,05?

4.7. В соответствии с (4.24) кратковременная автокорреляционная функция определяется выражением

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)\, w(n-m)\, x(m+k)\, w(n-k-m).$$

а) Показать, что $R_n(k) = R_n(-k)$, т. е. что $R_n(k)$ — четная функция k.

б) Показать, что $R_n(k)$ можно представить в виде

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)\, x(m-k)\, h_k(n-m),$$

где $h_k(n) = w(n)\, w(n+k)$.

в) Пусть

$$w(n) = \begin{cases} a^n, & n \geqslant 0; \\ 0, & n < 0. \end{cases}$$

Определить импульсную характеристику $h_k(n)$ для вычисления k-го значения корреляционной функции.

г) Определить z-преобразование $h_k(n)$ в п. в) и из него получить рекуррентное уравнение для $R_n(k)$. Изобразить схему цифрового фильтра для вычисления $R_n(k)$ как функции n при временно́м окне, приведенном в п.в).

д) Определить то же, что в п.в) и г) для случая

$$w(n) = \begin{cases} na^n, & n \geqslant 0; \\ 0, & n < 0. \end{cases}$$

4.8. Пусть дана периодическая импульсная последовательность

$$x(m) = \sum_{r=-\infty}^{\infty} \delta(m - rP).$$

а) Используя уравнение (4.30) с прямоугольным окном, длина которого N удовлетворяет соотношению $QP < N-1 < (Q+1)P$, где Q — целое, определить и изобразить $R_n(k)$ для $0 \leqslant k \leqslant N-1$.

б) Как изменится результат п.а), если в качестве окна использовать окно Хемминга той же длины?

в) Определить и построить кратковременную модифицированную корреляционную функцию $\hat{R}_n(k)$, задаваемую для тех же значений N.

4.9. Корреляционная функция случайного или периодического сигнала определяется выражением

$$\Phi(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{m=-N}^{N} x(m)\, x(m+k).$$

Кратковременная автокорреляционная функция определяется выражением

$$R_n(k) = \sum_{m=0}^{N-|k|-1} x(n+m)\, w'(m)\, x(n+m+k)\, w'(m+k).$$

Модифицированная кратковременная автокорреляционная функция равна

$$\hat{R}_n(k) = \sum_{m=0}^{N-1} x(n+m)\, x(n+m+k).$$

Показать, являются ли справедливыми следующие соотношения:

Если $x(n) = x(n+P)$, $-\infty < n < \infty$, то

а)
$$\Phi(k) = \Phi(k+P), \qquad -\infty < k < \infty;$$
$$R_n(k) = R_n(k+P), \qquad -(N-1) \leqslant k \leqslant N-1$$
$$\hat{R}_n(k) + \hat{R}_n(k+P), \qquad -(N-1) \leqslant k \leqslant N-1.$$

б)
$$\Phi(-k) = \Phi(k), \qquad -\infty < k < \infty;$$
$$R_n(-k) = R_n(k). \qquad -(N-1) \leqslant k \leqslant N-1;$$
$$\hat{R}_n(-k) = \hat{R}_n(k), \qquad -(N-1) \leqslant k \leqslant N-1.$$

в)
$$\Phi(k) \leqslant \Phi(0), \qquad -\infty < k < \infty;$$
$$R_n(k) \leqslant R_n(0), \qquad -(N-1) \leqslant k \leqslant N-1;$$
$$\hat{R}_n(k) \leqslant \hat{R}_n(0), \qquad -(N-1) \leqslant k \leqslant N-1.$$

г) $\Phi(0)$ равна мощности сигнала; $R_n(0)$ равна кратковременной энергии сигнала; $\hat{R}_n(0)$ равна кратковременной энергии сигнала.

4.10. Рассмотрим сигнал $x(n) = \cos \omega_0 n$, $-\infty < n < \infty$.

а) Определить корреляционную функцию для $x(n)$ по (4.21).

б) Изобразить $\Phi(k)$ как функцию от k.

г) Рассчитать и построить автокорреляционную функцию сигнала

$$y(n) = \begin{cases} 1, & x(n) \geqslant 0; \\ 0, & x(n) < 0. \end{cases}$$

4.11. Кратковременная функция среднего разности значений (КФСР) сигнала $x(n)$ определяется выражением [см. (4.38)]

$$\gamma_n(k) = \frac{1}{N}\sum_{m=0}^{N-1} |x(n+m) - x(n+m-k)|.$$

а) Используя неравенство [16]

$$\frac{1}{N}\sum_{m=0}^{N-2} |x(m)| \leqslant \left[\frac{1}{N}\sum_{m=0}^{N-1} |x(m)|^2\right]^{1/2},$$

показать, что $\gamma_n(k) \leqslant [2(\hat{R}_n(0) - \hat{R}_n(k))]^{1/2}$. Этот результат приводит к равенству (4.39).

б) Изобразить $\gamma_n(k)$ и величину $[2(\hat{R}_n(0) - \hat{R}_n(k))]^{1/2}$ при $0 \leqslant k \leqslant 200$ для сигнала $x(n) = \cos(\omega_0 n)$ с $N = 200$, $\omega_0 = 200\pi/(10\,000)$.

4.12. Рассмотрим входной сигнал $x(n) = A\cos(\omega_0 n)$ трехуровневого центрального ограничителя вида

$$y(n) = \begin{cases} 1, & x(n) > C_L; \\ 0, & |x(n)| \leqslant C_L; \\ -1, & x(n) < -C_L. \end{cases}$$

а) Изобразить $y(n)$ как функцию n для $C_L = 0{,}5A$, $0{,}75A$ и A.

б) Изобразить автокорреляционные функции для $y(n)$ и значений C_L из п.а).

в) Обсудить влияние взаимного расположения C_L и A. Пусть A изменяется во времени, удовлетворяя неравенству $0 < A(n) \leqslant A_{max}$. Рассмотреть ситуацию, которая может возникнуть при C_L, близком к A_{max}.

5
Цифровое представление речевых сигналов

5.0. Введение

«Если бы я смог заставить поток электричества изменяться по интенсивности точно в соответствии с изменением плотности воздуха во время распространения в нем звука, я бы смог передавать по телеграфу любые звуки, даже звуки речи» — А. Г. Белл [1]. Эта простая идея, имеющая столь важное значение для истории связи, кажется сегодня очевидной. Принцип, изложенный в открытии Белла, положен в основу множества устройств и систем, предназначенных для записи, передачи или обработки речевых сигналов и в которых речевой сигнал отражает колебания плотности звуковых (речевых) волн. Это относится и к цифровым системам, в которых речевой сигнал представлен последовательностью своих мгновенных значений.

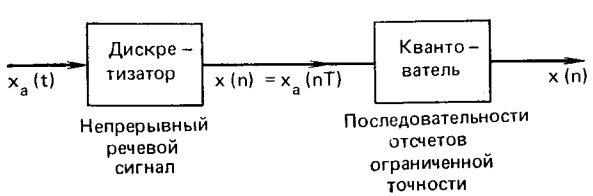


Рис. 5.1. Общая схема цифрового представления

Общая схема цифрового представления речевого сигнала изображена на рис. 5.1. Из рисунка следует, что речевое колебание как непрерывная функция времени подвергается дискретизации, чаще всего периодической, в результате которой образуется последовательность отсчетов $x_a(nT)$. Эти отсчеты могут в общем случае принимать непрерывное множество значений. Поэтому для получения цифрового, т. е. дискретного по амплитуде и по времени, представления необходимо проквантовать каждый отсчет до конечного множества значений.

Мы увидим далее, что рис. 5.1 достаточно полно отражает процесс формирования цифрового представления речевого сигнала. Может быть не во всех случаях можно разделить эту процедуру на два отдельных этапа, но основные операции — дискретизация и квантование — свойственны всем методам, приведенным в данной главе.

В начале главы изложены вопросы дискретизации применительно к речевым сигналам. Далее излагаются методы квантования отсчетов речевого колебания.

5.1. Дискретизация речевых сигналов

Теорема дискретизации уже обсуждалась в гл. 2. Последовательность отсчетов сигнала, как показано в гл. 2, единственным образом описывает аналоговый сигнал, если он ограничен по полосе частот и частота дискретизации по крайней мере вдвое больше наивысшей частоты спектра сигнала. Поскольку нас интересует цифровое представление речевых сигналов, изучим спектральные свойства речи. В соответствии с изложенным в гл. 3 описанием гласных и фрикативных звуков речевой сигнал не ограничен по полосе частот, хотя его спектр быстро спадает в области высоких частот. На рис. 5.2 изображены спектры типичных звуков речи.

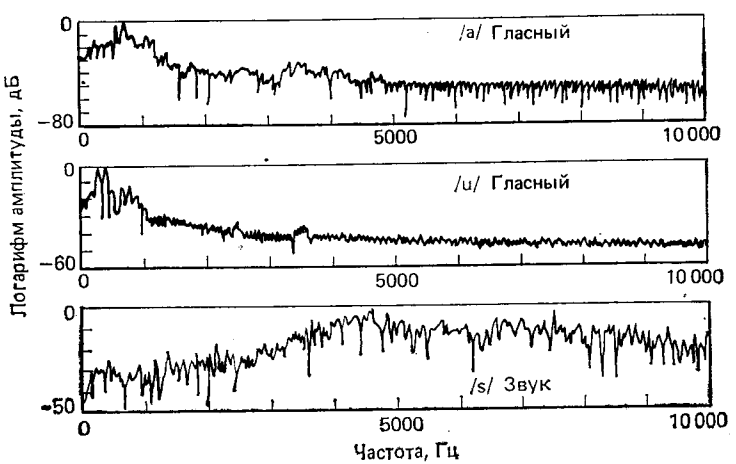


Рис. 5.2. Спектры вокализованных звуков /a/ и /u/ и невокализованного /s/ при частоте дискретизации 20 кГц

Видно, что для вокализованных звуков наивысшая частота, ниже которой максимумы спектра меньше уровня 40 дБ, составляет около 4 кГц. С другой стороны, для невокализованных звуков

спектр не затухает даже на частотах выше 8 кГц. Таким образом, для точного воспроизведения всех звуков речи требуется частота дискретизации около 20 кГц. В большинстве приложений такая частота дискретизации, однако, не требуется. Например, если дискретизация предшествует оцениванию трех первых формантных частот вокализованной речи, то достаточно располагать частью спектра до частоты около 3,5 кГц. Таким образом, если перед дискретизацией речевой сигнал пропускается через фильтр нижних частот так, что частота Найквиста равна 4 кГц, то частота дискретизации должна составлять 8 кГц. Другой пример. Рассмотрим речевой сигнал, который требуется передать по телефонному каналу. На рис. 5.3 приведена типичная частотная характеристика телефонного канала. Очевидно, что телефонный канал ограничивает полосу частот сигнала и частота Найквиста для «телефонной речи» составляет примерно 4 кГц.

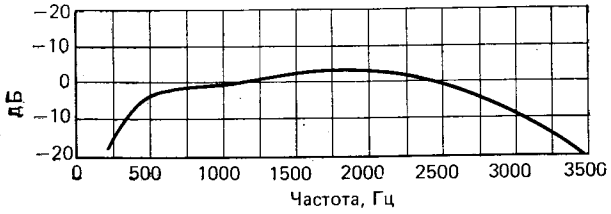


Рис. 5.3. Типичная частотная характеристика тракта телефонной связи (по BTL, Transmission Systems for communication, стр. 73)

Важная особенность, которую часто не замечают при обсуждении дискретизации, состоит в том, что даже если сигнал имеет ограниченный по частоте спектр, он может быть искажен широкополосным случайным шумом перед аналого-цифровым преобразованием. В таких случаях смесь сигнала и шума должна быть пропущена через фильтр с частотой среза, близкой к частоте Найквиста, что позволит избежать эффекта наложения частот при цифровом представлении.

5.2. Обзор статистических моделей речевых сигналов

При рассмотрении цифровых методов представления часто достаточно предполагать, что речевой сигнал является эргодическим случайным процессом. Хотя это является большим упрощением, далее будет показано, что статистическая точка зрения приводит к полезным результатам, тем самым подтверждая целесообразность подобной модели.

Если предположить, что сигнал $x_a(t)$ представляет собой непрерывный случайный процесс, то периодическая последовательность отсчетов этого сигнала может рассматриваться как случайный процесс с дискретным временем. В ряде случаев при анализе систем связи адекватными характеристиками аналогового сигнала являются одномерная функция плотности вероятности и автокорреляционная функция, определенная выражением

$$\varphi_a(\tau) = E[x_a(t)\, x_a(t+\tau)], \tag{5.1}$$

где $E[\cdot]$ означает усреднение по ансамблю величины, стоящей в квадратных скобках. Непрерывная спектральная плотность мощности представляет собой преобразование Фурье от $\varphi_a(\tau)$:

$$\Phi_a(\Omega) = \int_{-\infty}^{\infty} \varphi_a(\tau)\, e^{-i\Omega\tau}\, d\tau. \tag{5.2}$$

Сигнал с дискретным временем, полученный из непрерывного сигнала, имеет автокорреляционную функцию

$$\varphi(m) = E[x(n)\, x(n+m)] = E[x_a(nT)\, x_a(nT+mT)] = \varphi_a(mT). \tag{5.3}$$

Это просто дискретизированная функция $\varphi_a(\tau)$, поэтому спектральная плотность мощности равна

$$\begin{aligned} \Phi\left(e^{i\Omega T}\right) &= \sum_{m=-\infty}^{\infty} \varphi(m)\, e^{-i\Omega T m} = \\ &= \frac{1}{T} \sum_{k=-\infty}^{\infty} \Phi_a\left(\Omega + \frac{2\pi}{T} k\right). \end{aligned} \tag{5.4}$$

Из (5.4) следует, что спектральная плотность дискретизированного сигнала представляет собой периодическую последовательность, каждый член которой повторяет спектр аналогового сигнала.

Функция плотности вероятности величины $x(n)$ такая же, как и величины $x_a(t)$, так как $x(n) = x_a(nT)$. Это означает, в свою очередь, что и среднее и дисперсия непрерывного сигнала и сигнала с дискретным временем одинаковы.

Для использования статистических понятий при описании речевых сигналов необходимо оценить функцию плотности вероятности и корреляционную функцию (или спектральную плотность мощности) речевого колебания. Функция плотности вероятности оценивается путем определения гистограммы по большому числу отсчетов, т. е. в течение большого отрезка времени. Давенпорт [2] провел обширные исследования такого рода, а позже Паез и Глиссон [3], используя сходные измерения, показали, что хорошей аппроксимацией для экспериментальной функции плотности вероятности может служить гамма-распределение

$$p(x) = \left(\frac{\sqrt{3}}{8\pi\sigma_x |x|}\right)^{1/2} e^{\frac{-\sqrt{3}|x|}{2\sigma_x}}. \tag{5.5}$$

Более простой аппроксимацией является функция плотности вероятности Лапласа

$$p(x) = \frac{1}{\sqrt{2}\sigma_x} e^{\frac{-\sqrt{2}|x|}{\sigma_x}}. \tag{5.6}$$

На рис. 5.4 показана экспериментальная функция плотности вероятности совместно с функцией плотности вероятности Лапласа и

гамма-распределением. Все функции нормализованы таким образом, что среднее значение равно нулю, а дисперсия — единице. Хотя обе функции плотности вероятностей хорошо аппроксимируют экспериментальный результат, гамма-распределение, очевидно, обеспечивает лучшую аппроксимацию.

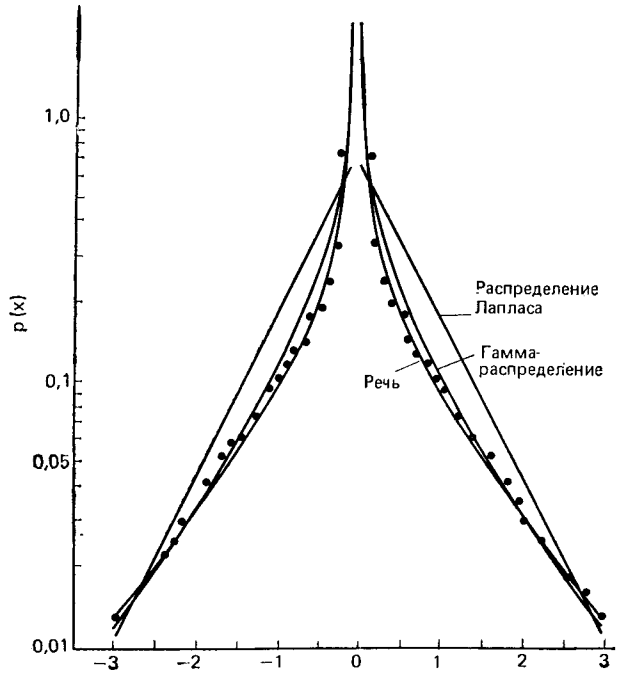


Рис. 5.4. Функции плотности вероятности речи, гамма-распределения и распределения Лапласа [3]

Автокорреляционная функция и спектральная плотность мощности речевого сигнала могут быть получены с использованием стандартных методов анализа временны́х рядов. Оценка автокорреляционной функции эргодического случайного процесса может быть получена путем усреднения за большой отрезок времени. Например, для получения усреднения за большой интервал времени достаточно немного изменить определение кратковременной автокорреляционной функции (4.30):

$$\hat{\varphi}(m) = \frac{1}{L} \sum_{n=0}^{L-1-m} x(n)\, x(n+m), \quad 0 \leqslant |m| \leqslant L-1, \tag{5.7}$$

где L — большое целое число. Пример такой оценки показан на рис. 5.5 при частоте дискретизации 8 кГц [4]. Верхняя кривая вычислена по сигналу, пропущенному через фильтр нижних частот, нижняя — через полосовой фильтр. Заштрихованные области

вокруг каждой кривой показывают изменения в корреляции, возникающие для различных дикторов. Корреляция весьма велика между соседними отсчетами и быстро убывает при увеличении расстояния между ними. Видно, что речевой сигнал на выходе фильтра нижних частот более коррелирован, чем на выходе полосового фильтра.

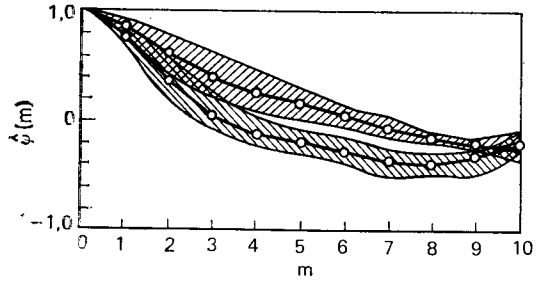


Рис. 5.5. Автокорреляционная функция речевых сигналов: верхняя кривая — для низкочастотной составляющей речи, нижняя — для высокочастотной [4]

Спектральную плотность мощности можно оценить различными путями. Для речевого сигнала один из наиболее ранних результатов был получен путем измерения сигнала на выходе гребенки полосовых фильтров [5]. На рис. 5.6 показан пример, в котором

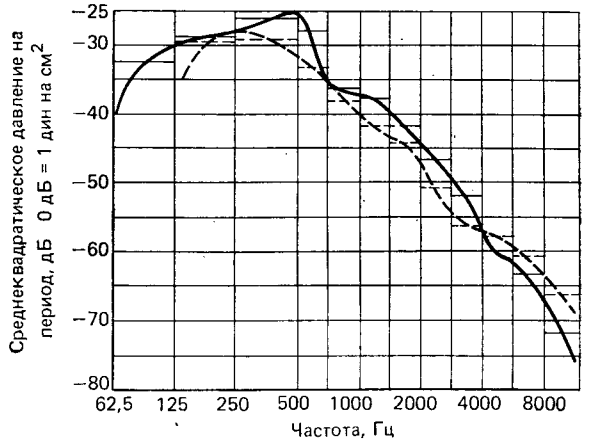


Рис. 5.6. Усредненная спектральная плотность мощности непрерывного речевого сигнала [5]: ——шесть дикторов-мужчин; — — — пять дикторов-женщин

мощность усреднялась за минуту непрерывной речи. Этот рисунок показывает, что усредненная спектральная плотность мощности имеет максимум в диапазоне 250—500 Гц и затухает примерно на 8—10 дБ на октаву. Другой подход к оценке усредненной спектральной плотности состоит в оценивании $\hat{\varphi}(m)$ соответственно (5.7) и последующем вычислении

$$\hat{\Phi}(e^{i\Omega T}) = \sum_{m=-M}^{M} w(m)\, \hat{\varphi}(m)\, e^{-i\Omega m T} \tag{5.8}$$

для дискретной последовательности $\Omega_k = 2\pi k/T$ при $k=0, 1, \ldots, \ldots, N-1$, используя дискретное преобразование Фурье [6], где $w(m)$ окно (взвешивающая функция) для автокорреляционной функции. В качестве примера применения этого метода к оценке спектральной плотности речи на рис. 5.7 приведены соответствующие результаты при использовании окна Хемминга [7]. Еще один

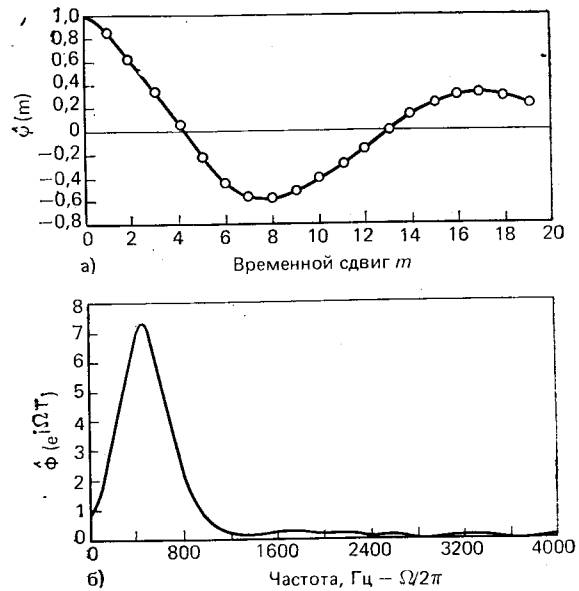


Рис. 5.7. Автокорреляционная функция (*а*) и спектральная плотность мощности (*б*) речевого сигнала [7]

подход состоит в вычислении передаточной функции цифрового фильтра, на входе которого действует белый шум, а сигнал на выходе имеет те же спектральные свойства, что и данный сигнал (см. гл. 8).

5.3. Квантование мгновенных значений

Как уже отмечалось, операции дискретизации и квантования удобно рассматривать отдельно, хотя часто разделить их затруднительно. Предположим, что речевой сигнал пропущен через фильтр нижних частот и в результате дискретизации получена последовательность непрерывных величин $\{x(n)\}$. В большинстве случаев в данной главе последовательность $\{x(n)\}$ рассматривается как случайный процесс в дискретном времени. Для того чтобы передать эту последовательность отсчетов по цифровому каналу связи, зарегистрировать ее в цифровом блоке памяти или использовать ее как входной сигнал некоторого алгоритма цифровой обработки,

каждый отсчет необходимо проквантовать до конечного множества значений, которые можно описать конечным множеством символов. Этот процесс квантования и кодирования изображен на рис. 5.8. Так же, как полезно разделять операции дискретизации и квантования, целесообразно разделить процесс представления последовательности $\{x(n)\}$ множеством символов на два этапа: квантование, результатом которого является последовательность величин $\{\hat{x}(n)\} = \{Q[x(n)]\}$, и кодирование, при котором каждой квантованной величине ставится в соответствие кодовое слово $c(n)$. Этот процесс изображен на рис. 5.8*а*. (Величина Δ на рисунке означает шаг квантования в квантователе.) Аналогично определим декодер как устройство, которое последовательности кодовых

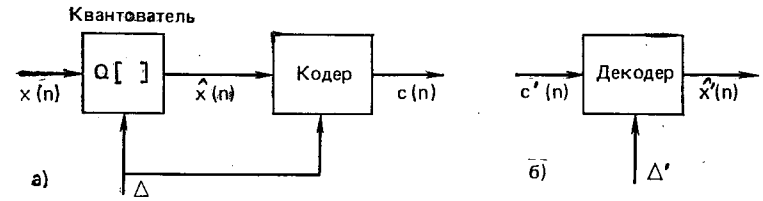


Рис. 5.8. Квантование и кодирование:
а) кодер; *б*) декодер

слов $\{c'(n)\}$ ставит в соответствие последовательность квантованных отсчетов $\{\hat{x}'(n)\}$, как это показано на рис. 5.8*б*. Если последовательность кодовых слов $c'(n)$ точно совпадает с последовательностью кодовых слов $c(n)$, т. е. ошибки отсутствуют, то сигнал на выходе идеального декодера точно совпадает с последовательностью квантованных отсчетов входного сигнала, т. е. $x'(n) = \hat{x}(n)$.

В большинстве случаев целесообразно для кодирования квантованных отсчетов использовать двоичную последовательность. С помощью B-разрядного двоичного кодового слова можно представить 2^B различных уровней квантования. Информационный объем цифрового представления, который надо знать при передаче или хранении сигнала, можно подсчитать:

$$I = B \cdot F_s = \text{скорость, бит/с}, \qquad (5.9)$$

где F_s — частота дискретизации (т. е. отсч./с); B — число бит на отсчет сигнала. В общем случае желательно выбирать скорость передачи наиболее низкой, при которой еще сохраняется требуемое качество восприятия сигнала. Для данной полосы частот речевого сигнала минимальная частота дискретизации определяется теоремой о дискретизации. Таким образом, единственный путь уменьшения скорости передачи состоит в сокращении числа двоичных единиц на отсчет сигнала. С этой целью продолжим обсуждение различных способов квантования сигнала.

В общем случае целесообразно предполагать, что отсчеты сигнала будут попадать в конечный интервал значений, при котором

$$|x(n)| \leqslant X_{max}. \qquad (5.10)$$

Для удобства следует предположить, что величина X_{max} бесконечно велика, что соответствует, например, функциям плотности вероятности гамма-распределения или Лапласа. Однако следует иметь в виду, что предположение о конечности диапазона значений в большей мере отвечает реальной ситуации. Даже если для описания сигнала используется функция плотности вероятности Лапласа, то легко показать (см. задачу 5.2), что только 0,35% отсчетов сигнала окажется вне диапазона

$$-4\sigma_x \leqslant x(n) \leqslant 4\sigma_x. \qquad (5.11)$$

Таким образом, целесообразно считать, что полный размах сигнала пропорционален среднему квадратическому отклонению.

Диапазон изменения входного сигнала делится на интервалы, и операция квантования сводится к тому, что всем отсчетам входного сигнала, попавшим в некоторый интервал, приписывается одно и то же заданное значение. Этот процесс иллюстрирует рис. 5.9 для восьмиуровневого квантователя. Например, для всех значений входного сигнала $x(n)$, расположенных между x_1 и x_2, значение сигнала на выходе будет $\hat{x}(n) = Q[x(n)] = \hat{x}_2$. Каждому уровню поставлено в соответствие трехразрядное слово, которым кодируется значение соответствующего уровня. Например, если отсчет попадает в интервал между x_1 и x_2 (рис. 5.9), то на выходе кодера появится слово 101. Конкретный способ кодирования уровней на рис. 5.9 произволен. В принципе можно использовать все восемь способов обозначения уровней, однако часто имеются причины выбирать вполне определенный способ кодирования.

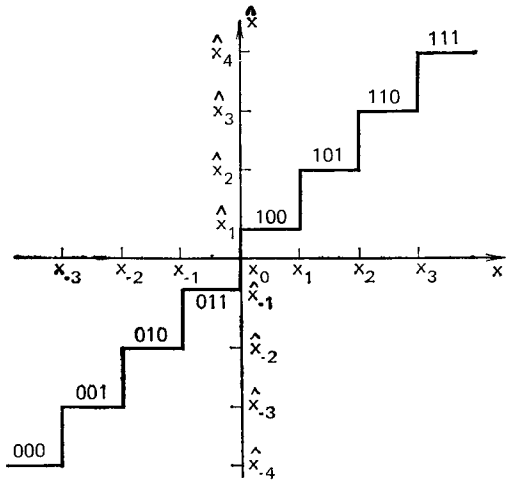


Рис. 5.9. Характеристика трехразрядного квантователя

5.3.1. Равномерное квантование

Интервалы и уровни квантования можно выбирать по-разному в зависимости от предлагаемого использования цифрового представления. Когда цифровое представление сигнала предназначено для обработки в некоторой системе, уровни и интервалы квантования выбирают обычно равномерно. Таким образом, для равномерного квантователя (см. рис. 5.9) получаем

$$x_i - x_{i-1} = \Delta \qquad (5.12)$$

и

$$\hat{x}_i - \hat{x}_{i-1} = \Delta, \qquad (5.13)$$

где Δ — шаг квантования. Для случая восьми уровней квантования на рис. 5.10 приведены характеристики двух обычно используемых квантователей. На рис. 5.10*а* изображен случай, когда на-

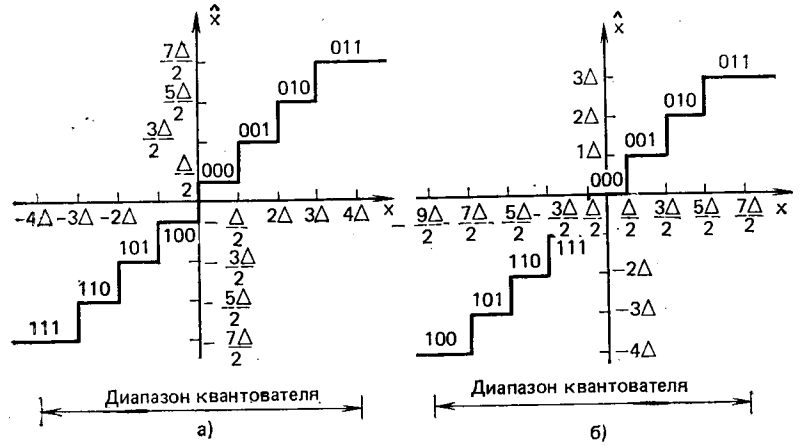


Рис. 5.10. Характеристики равномерных квантователей: *а*) с усечением; *б*) с округлением

чало отсчета приходится на середину вертикального участка ступенчатой функции. Этот класс квантователей называется квантователями с усечением. Аналогично на рис. 5.10*б* показан пример квантователей с округлением. Для случая, когда число уровней равно степени 2, как это обычно и бывает при использовании методов двоичного кодирования, квантователь с усечением имеет одинаковое количество положительных и отрицательных уровней, расположенных симметрично относительно начала координат. В отличие от этого квантователь с округлением имеет на один отрицательный уровень больше, но при этом обладает нулевым уровнем, который отсутствует у квантователя с усечением. Кодовые слова распределены по уровням на рис. 5.10 аналогично тому, как это сделано на рис. 5.9. Но в данном случае они обозначают непосредственно номер уровня в двоичной системе счисления. Например, если интерпретировать кодовые слова на рис. 5.10*а* как представление значения сигнала со знаком, полагая, что знаковый разряд — крайний слева, то уровни квантования связаны с кодовыми словами соотношением

$$\hat{x}(n) = (\Delta/2)\,\mathrm{sign}\,(c(n)) + \Delta\, c(n), \qquad (5.14)$$

где $\mathrm{sign}(c(n)) = +1$, если первый разряд $c(n) = 0$, и $\mathrm{sign}(c(n)) = -1$, если знаковый разряд $c(n) = 1$. Аналогично можно предста-

вить и кодовые слова на рис. 5.10*б*, но в этом случае соотношение между уровнями имеет вид

$$\hat{x}(n) = \Delta c(n). \tag{5.15}$$

Последний способ отображения уровней кодовыми словами используется обычно, когда последовательность отсчетов обрабатывается с представлением чисел в дополнительном коде (как в большинстве микропроцессоров), так как кодовые слова в этом случае являются просто численным значением отсчета сигнала.

Для описания равномерных квантователей (таких, как показанные на рис. 5.10) достаточно задать два параметра: число уровней и шаг квантования Δ. Число уровней выбирается обычно в виде 2^B с тем, чтобы использовать все B-разрядные кодовые слова. Параметры Δ и B выбираются таким образом, чтобы охватить весь диапазон сигнала. Если предположить, что $|x(n)| \leqslant X_{max}$, то (полагая симметричной функцию плотности вероятности $x(n)$) имеем

$$2X_{max} = \Delta \cdot 2^B. \tag{5.16}$$

При изучении эффектов квантования полезно представить квантованный сигнал в виде

$$\hat{x}(n) = x(n) + e(n), \tag{5.17}$$

где $x(n)$ — непрерывный отсчет, а $e(n)$ — ошибка, или шум квантования. Из рис. 5.10*а* и *б* легко установить, что если Δ и B выбирать в соответствии с (5.16), то

$$-\Delta/2 \leqslant e(n) \leqslant \Delta/2. \tag{5.18}$$

Если выбирать в качестве примера размах сигнала в $8\sigma_x$ и предположить, что сигнал имеет распределение Лапласа, то только 0,35% отсчетов окажутся вне диапазона квантователя. Квантование этих отсчетов будет сопровождаться ошибкой, большей $\pm\Delta/2$, то их число крайне мало, поэтому целесообразно выбирать диапазон квантования около $8\sigma_x$ и пренебречь при теоретическом анализе редкими большими ошибками [8].

Очевидно, что нам известен только $\hat{x}(n)$, а $x(n)$ и $e(n)$ неизвестны. Для изучения эффектов квантования удобно и полезно предположить простую статистическую модель шума квантования. Эта модель основана на следующих предположениях.

1. Шум квантования является стационарным белым шумом, т. е.

$$E[x(n)\,e(n+m)] = \begin{cases} \sigma_e^2, & m = 0, \\ 0, & \textit{в противном случае} \end{cases} \tag{5.19}$$

2. Шум квантования не коррелирован с входным сигналом, т. е.

$$E[x(n)\,e(n+m)] = 0, \quad \textit{для всех } m \tag{5.20}$$

3. Распределение шума равномерно в любом интервале квантования и поскольку все интервалы равны между собой, то

$$p_e(e) = \begin{cases} \dfrac{1}{\Delta}, & -\dfrac{\Delta}{2} \leqslant e \leqslant \dfrac{\Delta}{2}, \\ 0, & \textit{в противном случае} \end{cases} \tag{5.21}$$

Очевидно, что эти предположения не выполняются для некоторых сигналов. Они нарушаются, например, если входной сигнал постоянен для всех n. Речевой сигнал, однако, является достаточно сложным и быстроизменяющимся в пределах любого уровня квантования, и если шаг Δ достаточно мал, то вероятность попадания двух последовательных отсчетов в различные далеко отстоящие интервалы достаточно велика. Справедливость сделанных выше предположений подтверждается экспериментом [9].

Пример, иллюстрирующий правомерность предположений, приведен на рис. 5.11 [6]. На рис. 5.11*а* изображено 400 последовательных отсчетов сигнала, которые были подвергнуты операции

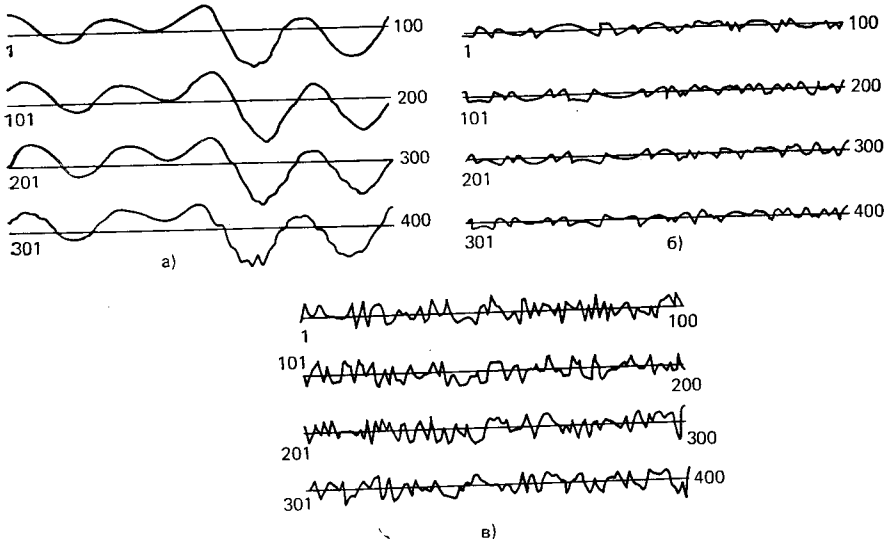


Рис. 5.11. Временная диаграмма речи (*а*); шум квантования для трехразрядного квантователя (*б*); шум квантования для восьмиразрядного квантователя (*в*) (увеличено в 66 раз по сравнению с (*а*) и (*б*))

квантования в трех- и восьмиразрядном квантователях[1] (рис. 5.10*б*). Ошибки квантования изображены на рис. 5.11*б* и *в* соответственно. В случае трехразрядного квантования видна корреляция сигнала и ошибки, в то время как во втором случае такая корреляция не наблюдается. Для подтверждения этого факта на

[1] Термин восьмиразрядный (8-bit) квантователь означает 2^8-уровневый квантователь. (*Прим. ред.*)

рис. 5.12*а* и *в* представлены корреляционные функции для первого и второго случаев соответственно. Очевидно, что корреляционная функция рис. 5.12*в* более согласуется с предложением, что $\varphi(m)=\sigma^2_e\delta(m)$, поскольку на рис. 5.12*а* видна значительная корреляция при $m>0$. Этот же эффект наблюдается и на спектральной плотности мощности (рис. 5.12*б* и *г*). Спектр погрешности

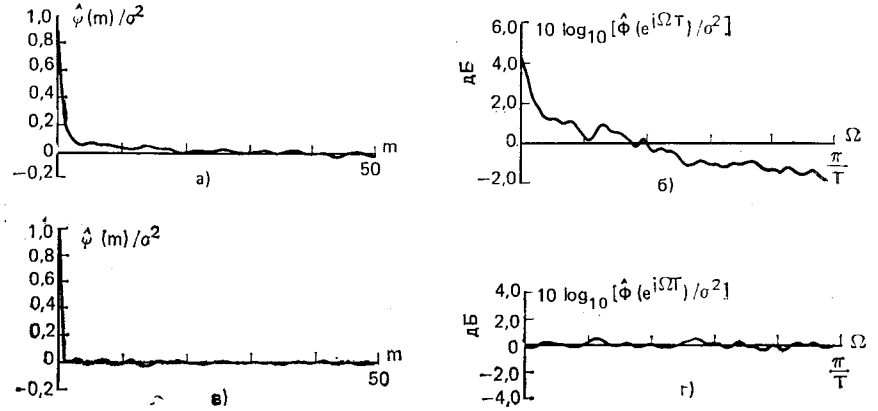


Рис. 5.12. Нормализованная оценка автокорреляции для трехразрядного квантователя (*а*); спектр мощности для трехразрядного квантователя (*б*), нормализованная оценка автокорреляции для восьмиразрядного квантователя (*в*); спектр мощности для восьмиразрядного квантователя (*г*)

при трехразрядном квантовании уменьшается в области высших частот (как и спектр речи), а для восьмиразрядного квантователя спектр равномерен. Заметим однако, что даже для трехразрядного квантователя спектр затухает со скоростью 6 дБ на октаву.

В рамках введенной статистической модели можно связать мощность шума и сигнала с параметрами квантователя. Для этой цели удобно вычислить отношение сигнал/шум квантования, определяемое выражением[1]

$$SNR=\frac{\sigma_x^2}{\sigma_e^2}=\frac{E\,[x^2(n)]}{E\,[e^2(n)]}=\frac{\sum\limits_n x^2(n)}{\sum\limits_n e^2(n)}. \qquad (5.22)$$

Если предположить, что диапазон непрерывных значений равен $2X_{max}$, то для B-разрядного квантователя получаем

$$\Delta=2X_{max}/2^B. \qquad (5.23)$$

Если предположить равномерное распределение шума, получим (см. задачу 5.1)

$$\sigma_e^2=\Delta^2/12=X_{max}^2/(3)\,2^{2B}. \qquad (5.24)$$

[1] Предполагается, что сигнал имеет нулевое среднее значение. В противном случае оно вычитается из сигнала перед вычислением отношения сигнал/шум.

Подставляя (5.24) в (5.22), имеем

$$SNR=(3)\,2^{2B}/[X_{max}/\sigma_x]^2 \qquad (5.25)$$

или, выражая отношение сигнал/шум в децибелах,

$$SNR=10\log_{10}[\sigma_x^2/\sigma_e^2]=6B+4{,}77-20\log_{10}[X_{max}/\sigma_x]. \qquad (5.26)$$

Предполагая диапазон квантования $X_{max}=4\sigma_x$, из (5.26) получим

$$SNR=6B-7{,}2. \qquad (5.27)$$

Соотношение (5.27), из которого следует, что каждое добавление одного разряда в кодовом слове улучшает отношение сигнал/шум на 6 дБ, справедливо при следующих предположениях: 1) входной сигнал изменяется таким образом, что справедлива описанная ранее статистическая модель шума квантования; 2) шаг квантования мал настолько, что шум белый и не коррелирован с сигналом; 3) диапазон квантования установлен таким образом, что он превышает размах сигнала. Следовательно, диапазон квантования используется полностью, и в тоже время количество отсчетов, не попадающих в него, достаточно мало.

Для речевых сигналов первые два предположения выполняются, если количество уровней квантования больше, чем 2^6. Однако третье предположение менее справедливо, поскольку энергия сигнала может изменяться более чем на 40 дБ в зависимости от диктора и условий передачи. Даже для данного диктора амплитуда речевого сигнала существенно меняется при переходе от вокализованной речи к невокализованной и на протяжении вокализованных сегментов. Соотношение (5.27) предполагает полное использование диапазона квантования, и если размах сигнала очень мал, то это эквивалентно использованию лишь нескольких уровней квантования квантователя. Например, из (5.26) следует, что в случае если дисперсия входного сигнала составляет лишь половину той, на которую рассчитан квантователь, то это приведет к ухудшению отношения сигнал/шум на 6 дБ. В то же время из кратковременного анализа сигнала известно, что дисперсия на невокализованных сегментах может быть на 20—30 дБ меньше, чем дисперсия на вокализованных, т. е. кратковременное отношение сигнал/шум на невокализованных сегментах может быть значительно меньше, чем на вокализованных.

Для поддержания ошибки квантования на приемлемом уровне необходимо выбирать значительно больше уровней квантования, чем это следует из предварительного анализа в рамках предположения о стационарности сигнала. Например, использование соотношения (5.27) позволяет сделать вывод, что значение $B=7$ обеспечит отношение сигнал/шум, равное 36 дБ, т. е. хорошее качество связи. Однако известно, что необходимо около 11 разрядов квантователя для получения высококачественного речевого сигнала при равномерном квантовании.

Таким образом, желательно иметь устройство квантования, при котором отношение сигнал/шум не зависит от уровня сигнала,

т. е. вместо постоянной не зависящей от уровня сигнала ошибки (как это имеет место при равномерном квантовании) хотелось бы получить постоянную относительную ошибку. Это достигается путем использования неравномерного распределения уровней квантования.

5.3.2. Мгновенное компандирование

Для того чтобы относительная ошибка была постоянной, уровни квантования должны быть распределены логарифмически. С другой стороны, вместо квантования исходного сигнала для достижения постоянной ошибки можно квантовать его логарифм. Этот процесс изображен на рис. 5.13, где входной сигнал компрес-

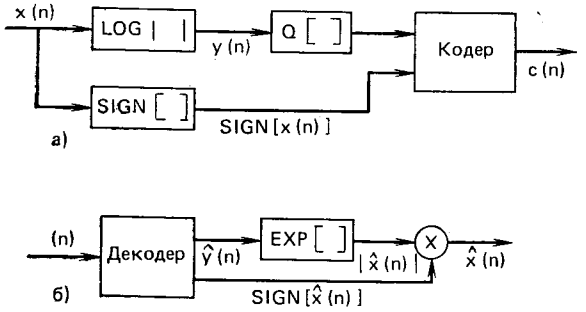


Рис. 5.13. Структурная схема системы логарифмического кодирования

сируется перед квантованием с помощью логарифмического преобразования, а выходной сигнал после декодирования экспандируется с помощью экспоненциального преобразования. Убедимся, что это приводит к требуемой нечувствительности ошибки квантования от значения сигнала. Предположим, что

$$y(n) = \ln |x(n)|. \tag{5.28}$$

Обратное преобразование равно

$$x(n) = \exp[y(n)] \operatorname{sign}[x(n)], \tag{5.29}$$

где $\operatorname{sign}[x(n)] = +1$, если $x(n)$ положительно, и $\operatorname{sign}[x(n)] = -1$, если $x(n)$ отрицательно. Теперь квантованный логарифм имеет вид

$$\hat{y}(n) = Q[\log |x(n)|] = \log |x(n)| + \varepsilon(n), \tag{5.30}$$

где, как и ранее, предполагается, что $\varepsilon(n)$ не зависит от $\log|x(n)|$. Применяя к квантованной величине обратное преобразование, получаем

$$\hat{x}(n) = \exp[\hat{y}(n)] \operatorname{sign}[x(n)] = |x(n)| \operatorname{sign}[x(n)] \exp[\varepsilon(n)] = x(n) \exp[\varepsilon(n)]. \tag{5.31}$$

Если $\varepsilon(n)$ мало, можно аппроксимировать это уравнение в виде

$$\hat{x}(n) \approx x(n)[1 + \varepsilon(n)] = x(n) + \varepsilon(n) x(n) = x(n) + f(n), \tag{5.32}$$

где $f(n) = x(n)\varepsilon(n)$. Таким образом, поскольку $x(n)$ и $\varepsilon(n)$ предполагаются некоррелированными, получаем

$$\sigma_f^2 = \sigma_x^2 \sigma_\varepsilon^2; \quad SNR = \sigma_x^2/\sigma_f^2 = 1/\sigma_\varepsilon^2. \tag{5.33); (5.34}$$

Следовательно, отношение сигнал/шум не зависит от мощности сигнала, а зависит только от шага квантования. Квантователь такого типа не имеет практического значения, поскольку динамический диапазон (отношение максимального значения к минимальному) бесконечен и, таким образом, требуется бесконечное число уровней квантования. Выполненный анализ, возможно и лишенный практического смысла, позволяет, однако, сделать вывод о том, что характеристика компрессора может быть близкой к логарифмической. Использование системы компрессор—экспандер для

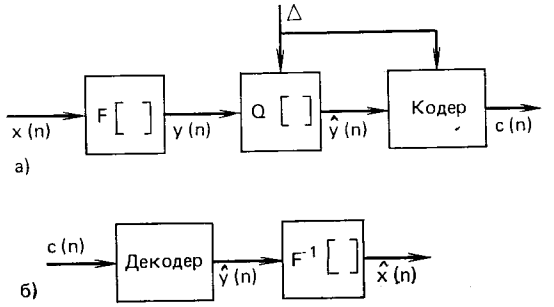


Рис. 5.14. Структурная схема системы компрессор-экспандер для квантования

квантования показано на рис. 5.14. Смит [10] исследовал характеристики компрессора, названные μ-законом компандирования. В этом случае

$$y(n) = F[x(n)] = X_{max} \frac{\log\left[1 + \mu \dfrac{|x(n)|}{X_{max}}\right]}{\log[1 + \mu]} \operatorname{sign}[x(n)]. \tag{5.35}$$

На рис. 5.15 представлено семейство зависимостей $y(n)$ от $x(n)$ для различных значений μ. Очевидно, что использование функции (5.35) решает проблему малых амплитуд, поскольку $y(n) = 0$, если $|x(n)| = 0$. При $\mu = 0$ уравнение (5.35) превращается в равенство

$$y(n) = x(n), \tag{5.36}$$

т. е. уровни квантования располагаются равномерно. Однако для больших μ и больших $|x(n)|$ получаем

$$|y(n)| \approx X_{max} \log |x(n)/X_{max}|. \tag{5.37}$$

Таким образом, за исключением очень малых амплитуд, кривые, соответствующие μ-закону, позволяют получить постоянный процент дисперсии шума от дисперсии сигнала. На рис. 5.16 показано

распределение уровней квантования при $\mu=40$ и восьми уровнях (характеристика квантователя антисимметрична относительно начала координат).

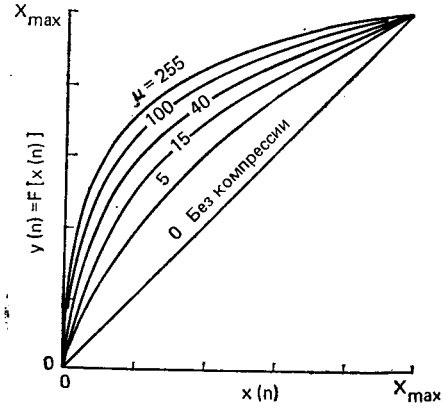


Рис. 5.15. Характеристика компрессии по μ-закону [10]

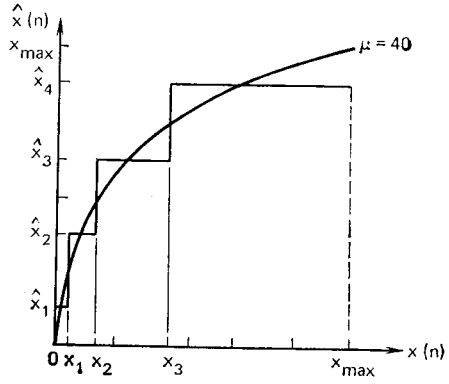


Рис. 5.16. Распределение уровней квантования для μ-закона в трехразрядном квантователе при $\mu=40$

Используя предположения такого же типа, как и в случае анализа равномерного квантователя, Смит [10] получил формулу для отношения сигнал/шум при использовании μ-закона в виде

$$SNR=6B+4{,}77-20\log_{10}[\ln(1+\mu)]-$$
$$-10\log_{10}\left[1+\left(\frac{X_{max}}{\mu\sigma_x}\right)^2+\sqrt{2}\left(\frac{X_{max}}{\mu\sigma_x}\right)\right], \qquad (5.38)$$

Сравнивая это уравнение с (5.26), можно отметить, что в данном случае отношение сигнал/шум значительно меньше зависит от величины (X_{max}/σ_x). Эта зависимость уменьшается при возрастании μ, т. е. хотя член $20\log_{10}[\ln(1+\mu)]$ уменьшает отношение сигнал/шум, второе слагаемое с ростом μ возрастает. На рис. 5.17 и 5.18 графически представлены соотношения (5.26) и (5.38) как функции величины X_{max}/σ_x при $\mu=100$ и 500 соответственно. Величина X_{max} является параметром устройства квантования. Она определяет порог «переполнения», т. е. значение, выше которого все отсчеты ограничиваются. Величина σ_x является параметром сигнала, определяющим «среднюю» амплитуду сигнала. Величина X_{max}/σ_x показывает, насколько диапазон сигнала согласован с диапазоном квантователя. Пунктирные кривые на рис. 5.17 иллюстрируют зависимость отношения сигнал/шум в децибелах от X_{max}/σ_x. При заданном значении X_{max} уменьшение вдвое величины σ_x приводит к потере в отношении сигнал/шум на 6 дБ. Для заданного значения X_{max}/σ_x отношение сигнал/шум возрастает на 6 дБ при добавлении одного разряда квантователя. Это справедливо как для равномерного квантователя, так и при использовании μ-закона.

Поясним правомерность соотношений (5.26) и (5.38) и кривых рис. 5.17 и 5.18. Одно из предположений, сделанных при выводе этих уравнений, состоит в том, что количество переполнений в квантователе пренебрежимо мало, т. е. вероятность того, что некоторый отсчет превысит X_{max}, пренебрежимо мала. Это предположение, очевидно, нарушается, если дисперсия σ_x приблизительно равна X_{max}, т. е. $X_{max}/\sigma_x\approx 1$. Экспериментальные кривые отношения сигнал/шум подтверждают сделанное утверждение. Соотношения (5.26) и (5.38) хорошо описывают отношение сигнал/шум при $(X_{max}/\sigma_x)>8$ [10].

Важное свойство μ-закона, иллюстрируемое этими кривыми, состоит в том, что отношение сигнал/шум более или менее постоянно в широком диапазоне. Например, из рис. 5.17 видно, что при $\mu=100$ отношение сигнал/шум уменьшается всего на 2 дБ при

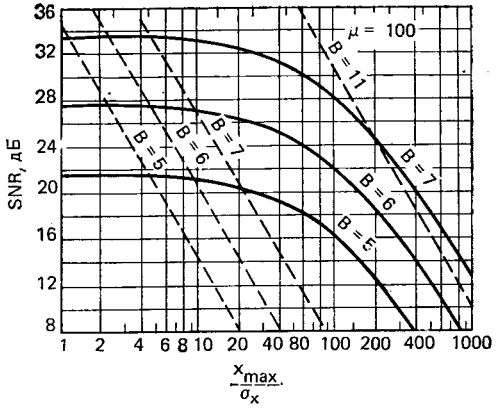


Рис. 5.17. *SNR* для μ-закона и равномерного квантователя как функция X_{max}/σ_x при $\mu=100$ и различном числе разрядов B [10]

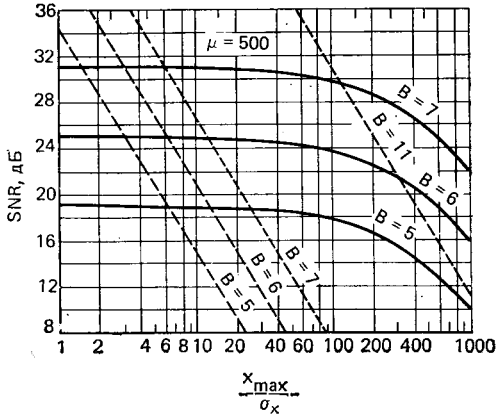


Рис. 5.18. *SNR* для μ-закона и равномерного квантователя при $\mu=500$, $B=5$; 6; 7; 11 [10]

$$8<X_{max}/\sigma_x<30, \qquad (5.39)$$

а из рис. 5.18 видно, что при $\mu=500$ отношение сигнал/шум — менее чем на 2 дБ в диапазоне

$$8<X_{max}/\sigma_x<150. \qquad (5.40)$$

Однако сравнение рис. 5.17 и 5.18 показывает, что максимальное отношение сигнал/шум во втором случае уменьшается на 2,6 дБ. Таким образом, используя большие значения коэффициента компрессии, мы получаем выигрыш в динамическом диапазоне ценой проигрыша в отношении сигнал/шум.

Как следует из рис. 5.17 и 5.18, при $B=7$ отношение сигнал/шум, равное 34 дБ, достигается в широком диапазоне уровней

входного сигнала. Поэтому семиразрядная ИКМ с компрессией используется как стандарт для получения речевого сигнала с хорошим качеством. При равномерном квантовании для получения такого же динамического диапазона требуется 11 разрядов. Как следует из рис. 5.18, 11-разрядное равномерное квантование оказывается лучше семиразрядного квантования для $\mu=500$ ($\sigma_x > 0{,}01X_{max}$). Таким образом, можно сказать, что 11-разрядное равномерное квантование будет таким же или лучшим, чем семиразрядное квантование при $\mu=500$ для уровней входного сигнала, составляющих, по крайней мере, 1% максимального уровня квантования.

5.3.3. Оптимальное квантование

Квантование по μ-закону позволяет получить постоянное отношение сигнал/шум в широком диапазоне дисперсий входного сигнала. Как отмечалось выше, это достигается ценой ухудшения отношения сигнал/шум по сравнению с тем, которое можно получить, если диапазон квантования согласован с дисперсией сигнала. В тех случаях, когда дисперсия сигнала известна, можно так выбирать уровни квантования, чтобы минимизировать мощность шума и, таким образом, максимизировать отношение сигнал/шум квантования. Эта задача изучалась Максом [11] и позже Паезом и Глиссоном [3]. Дисперсия шума квантования имеет вид

$$\sigma_e^2 = E\,[e^2(n)] = E\,[(\hat{x}(n) - x(n))]^2, \tag{5.41}$$

где $\hat{x}(n) = Q[x(n)]$. На основе рис. 5.9 предположим, что имеется M уровней квантования, которые можно обозначить через $\{\hat{x}_{-M/2}, \hat{x}_{-M/2+1}, \ldots, \hat{x}_{-1}, \hat{x}_1, \ldots, x_{M/2}\}$, полагая, что M четное. Уровень квантования, соответствующий интервалу от x_{j-1} до x_j, обозначен как $\hat{x}_j$. Для симметричной функции плотности вероятности с нулевым средним удобно обозначить центральную граничную точку $x_0=0$, и если функция плотности вероятности не обращается в нуль при больших значениях, как, например, функция Лапласа, то максимальное значение границ внешнего интервала квантования будет $\pm\infty$, т. е. $x_{\pm M/2} = \pm\infty$. В этих предположениях можно записать

$$\sigma_e^2 = \int e^2 p_e(e)\, de. \tag{5.42}$$

На рис. 5.19 показана зависимость e от x. Как видно из рисунка, вклад каждого интервала квантования в функцию плотности вероятности погрешности оценить сложно. Поскольку

$$e = \hat{x} - x, \tag{5.43}$$

Можно выполнить замену переменной в (5.42)

$$p_e(e) = p_e(\hat{x} - x) = p_{x/\hat{x}}(x/\hat{x}) \triangleq p_x(x), \tag{5.44}$$

что приводит к соотношению

$$\sigma_e^2 = \sum_{i=-M/2+1}^{M/2} \int_{x_{i-1}}^{x_i} (\hat{x}_i - x)^2 p(x)\, dx \tag{5.45}$$

(отметим, что в данную формулу для дисперсии шума входят и ошибки за счет ограничения или переполнений). Если $p(x) = p(-x)$, то характеристика квантования будет антисимметричной, т. е. $\hat{x}_i = -\hat{x}_{-i}$ и $x_i = -x_{-i}$. Таким образом,

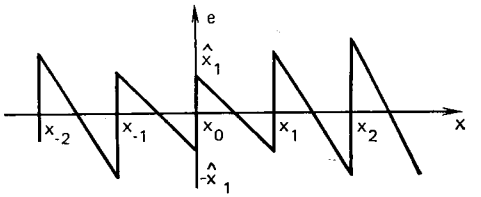


Рис. 5.19. Ошибка квантования как функция уровня x неравномерного квантователя

$$\sigma_e^2 = 2\sum_{i=1}^{M/2} \int_{x_{i-1}}^{x_i} (\hat{x}_i - x)^2 p(x)\, dx. \tag{5.46}$$

Теперь требуется выбрать множество таких параметров $\{x_i\}$ и $\{\hat{x}_i\}$, чтобы минимизировать σ_e^2. Для решения этой задачи продифференцируем σ_e^2 по этим параметрам и приравняем производные к нулю. Это приводит к системе уравнений

$$\int_{x_{i-1}}^{x_i} (\hat{x}_i - x)\, p(x)\, dx = 0, \quad i = 1, 2, \ldots, M/2; \tag{5.47а}$$

$$x_i = (\hat{x}_i + \hat{x}_{i+1})/2, \qquad i = 1, 2, \ldots, (M/2) - 1 \tag{5.47б}$$

в предположении

$$x_0 = 0; \tag{5.48а}$$

$$x_{\pm M/2} = \pm\infty. \tag{5.48б}$$

Уравнения (5.47) показывают, что оптимальные пороги равны полусумме соответствующих уровней квантования. Уравнение (5.47а) означает, что соответствующие уровни есть средние значения функции плотности вероятности на интервале от x_{i-1} до x_i. Два этих множества уравнений следует решать совместно относительно $M-1$ неизвестных параметров квантователя. Поскольку данные уравнения нелинейны, замкнутое решение можно получить лишь в специальных случаях. Для других случаев следует использовать итеративные методы решения. Такой метод последовательных приближений дан Максом [1]. Паез и Глиссон использовали эту процедуру для получения оптимальных порогов в случае распределения Лапласа и гамма-распределения [3].

В общем случае решение уравнений (5.47) приводит к неравномерному распределению уровней квантования. Только при равномерном распределении сигнала оптимальное решение будет равномерным, т. е.

$$\hat{x}_i - \hat{x}_{i-1} = x_i - x_{i-1} = \Delta. \tag{5.49}$$

Полагая квантователь равномерным, можно определить значение оптимального шага квантования Δ, при котором минимальна мощность шума квантования и максимально отношение сигнал/(шум квантования). В этом случае

$$x_i = \Delta \cdot i; \tag{5.50}$$

$$\hat{x}_i = (2i-1)\,\Delta/2 \tag{5.51}$$

и Δ удовлетворяет уравнению

$$\sum_{i=1}^{M/2-1} (2_i - 1) \int_{(i-1)\,\Delta}^{i\,\Delta} \left[\left(\frac{2i-1}{2}\right)\Delta - x\right] p(x)\,dx + \\ + (M-1) \int_{(M/2-1)\,\Delta}^{\infty} \left[\left(\frac{M-1}{2}\right)\Delta - x\right] p(x)\,dx = 0. \tag{5.52}$$

Таблица 5.1

Оптимальный квантователь для сигналов с распределением Лапласа ($m_x=0$, $\sigma_x=1$) [3]

N	2		4		8		16		32	
i	x_j	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$
1	∞	0,707	1,102	0,395	0,504	0,222	0,266	0,126	0,147	0,072
2			∞	1,810	1,181	0,785	0,566	0,407	0,302	0,222
3					2,285	1,576	0,910	0,726	0,467	0,382
4					∞	2,994	1,317	1,095	0,642	0,551
5							1,821	1,540	0,829	0,732
6							2,499	2,103	1,031	0,926
7							3,605	2,895	1,250	1,136
8							∞	4,316	1,490	1,365
9									1,756	1,616
10									2,055	1,896
11									2,398	2,214
12									2,804	2,583
13									3,305	3,025
14									3,978	3,586
15									5,069	4,371
16									∞	5,768
MSE	0,5		0,1765		0,0548		0,0154		0,00414	
SNR, дБ	3,01		7,53		12,61		18,12		23,83	

Если $p(x)$ известна или задана (например, функция плотности вероятности Лапласа), то интегралы могут быть вычислены. Тогда получается простое уравнение, которое можно решить на ЭВМ, пользуясь итеративными методами и изменяя Δ до получения оптимального значения.

В табл. 5.1 и 5.2 содержатся оптимальные уровни квантования для гамма-распределения и распределения Лапласа [3]. (Результаты получены в предположении единичной дисперсии. Если в действительности дисперсия σ^2_x, то каждое число в таблице следует

Таблица 5.2

Оптимальный квантователь для сигналов с гамма-распределением ($m_x=0$, $\sigma^2_x=1$) [3]

N	2		4		8		16		32	
i	x_j	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$
1	∞	0,577	1,205	0,302	0,504	0,149	0,229	0,072	0,101	0,033
2			∞	2,108	1,401	0,859	0,588	0,386	2,252	0,169
3					2,872	1,944	1,045	0,791	0,429	0,334
4					∞	3,799	1,623	1,300	0,630	0,523
5							2,372	1,945	0,857	0,737
6							3,407	3,798	1,111	0,976
7							5,050	4,015	1,397	1,245
8							∞	6,085	1,720	1,548
9									2,089	1,892
10									2,517	2,287
11									3,022	2,747
12									3,633	3,296
13									4,404	3,970
14									5,444	4,838
15									7,046	6,050
16									∞	8,043
MSE	0,6680		0,2326		0,0712		0,0196		0,0052	
SNR, дБ	1,77		6,33		11,47		17,07		22,83	

умножить на σ_x.) На рис. 5.20 изображена схема трехразрядного (восьмиуровневого) квантователя для функции плотности вероятности Лапласа. Легко видеть, что шаг квантования возрастает при убывании значения функции плотности вероятности. Это согласуется с интуитивным представлением о том, что большие значения ошибки квантования должны соответствовать наиболее редко возникающим отсчетам. Сравнение рис. 5.16 и 5.20 позволяет отметить сходство оптимального квантователя с квантователем на основе μ-закона. Таким образом, можно ожидать, что неравномерный квантователь обладает улучшенным динамическим диапазоном. Это подтверждается результатами работы [3].

На рис. 5.21 показана зависимость шага квантования оптимального равномерного квантователя для гамма-распределения,

распределений Лапласа [3] и Гаусса [11] от числа уровней. Как это и можно было ожидать, шаг квантования убывает приблизительно экспоненциально с увеличением разрядности кодовых слов. Некоторые различия в форме кривых объясняются различиями в

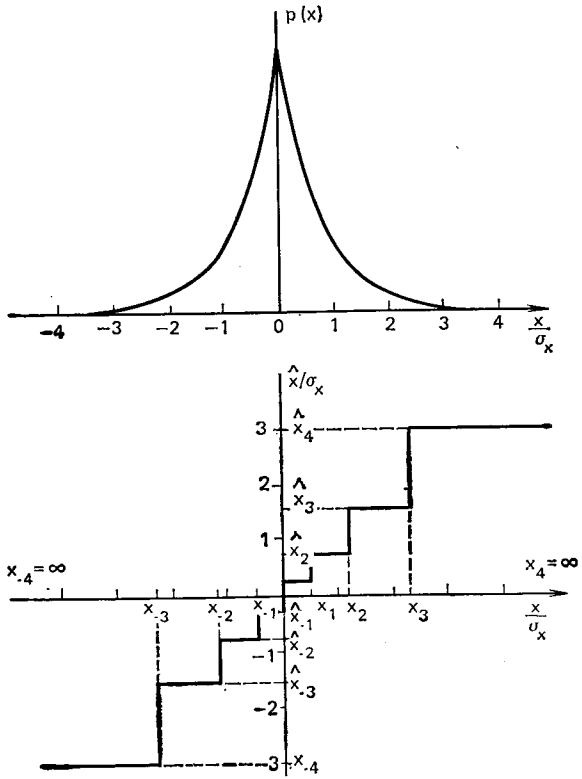


Рис. 5.20. Функция плотности вероятности и характеристика квантователя для распределения Лапласа и неравномерного квантователя с тремя разрядами

виде функции плотности вероятности. Хотя оптимальное квантование дает минимальные погрешности при точном соответствии дисперсии и распределении сигнала, нестационарный характер речевого сигнала в системах связи приводит к неудовлетворительным результатам и в этом случае. Простым подтверждением этого обстоятельства является ситуация, возникающая в системах передачи в период молчания, т. е. при незанятом канале передачи. При этом сигнал на входе квантователя весьма мал (предполагается малый шум), что приводит к колебаниям сигнала на выходе квантователя между наименьшими уровнями квантования.

Для симметричного квантователя (см. рис. 5.10*а*), если внутренние уровни квантования больше, чем мгновенные значения шума, мощность шума на выходе будет больше мощности шума на

входе. Поэтому применение оптимального квантователя с малым числом уровней квантования не имеет практического смысла. В табл. 5.3 для сравнения представлены значения наименьшего уровня квантования [3] для ряда оптимальных равномерных и неравномерных квантователей и квантователя, построенного по μ-закону при μ=100. Легко видеть, что квантователь, построенный по μ-закону, приводит к меньшему шуму незанятого канала по сравнению с любым оптимальным квантователем. А для больших значений μ наименьший уровень квантования лежит еще ниже (так при μ=255 минимальный уровень составляет 0,031). Поэтому μ-квантователь чаще используется на практике, несмотря на несколько меньшее отношение сигнал/шум по сравнению с оптимальным.

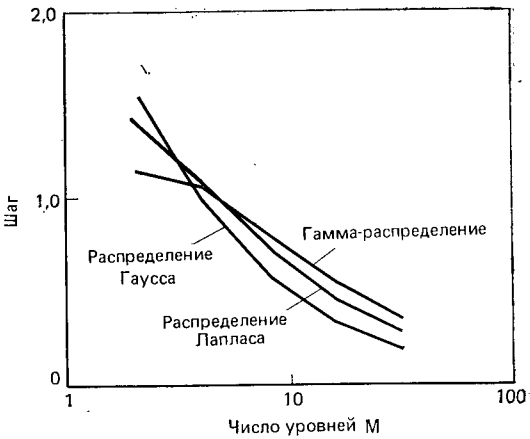


Рис. 5.21. Оптимальные размеры шага равномерного квантователя при лапласовском, гамма- и гауссовском распределениях [11]

Таблица 5.3

Отношение сигнал/шум для восьмиуровневого квантователя [12]

Закон	SNR, дБ	Наименьший уровень (δ_x=1)
μ-закон ($x_{max}=8\sigma_x=\mu=100$)	9,5	0,062
Гаусса	14,6/14,3*	0,245/0,293
Лапласа	12,6/11,4	0,222/0,366
Гамма	11,5/11,5	0,149/0,398
Речь	12,1/8,4	0,124/0,398

* В числителе — для неравномерного, в знаменателе — для равномерного квантователя.

5.4. Адаптивное квантование

Как следует из результатов предыдущего параграфа, при квантовании речевого сигнала возникают серьезные трудности. С одной стороны, шаг квантования следует выбирать достаточно большим для согласования диапазона квантования с размахом сигнала. С другой стороны, шаг квантования следует сделать малым для уменьшения шума квантования. Это еще более усложняется нестационарным характером речевого сигнала и его последующей передачей по каналу связи. Амплитуда речевого сигнала может

изменяться в широких пределах в зависимости от диктора, условий передачи, а также внутри фразы при переходе от вокализованного к невокализованному сегменту. Как уже отмечалось, один из методов учета этих флуктуаций состоит в применении неравномерного квантования. Другой метод состоит в адаптации свойств квантователя к уровню входного сигнала. В этом параграфе обсуждаются основные принципы адаптивного квантования, а в последующих приводятся примеры применения методов адаптивного квантования совместно с линейным предсказанием речи. Если адаптивное квантование применяется непосредственно к отсчетам входного сигнала, то такой метод обработки называется адаптивной ИКМ или сокращенно АИКМ.

Основная идея адаптивного квантования состоит в том, что шаг квантования (или, в общем случае, интервалы и уровни квантования) изменяется таким образом, чтобы соответствовать изменяющейся дисперсии входного сигнала (рис. 5.22*а*). Другой метод адаптации (рис. 5.22*б*) соответствует случаю, когда характе-

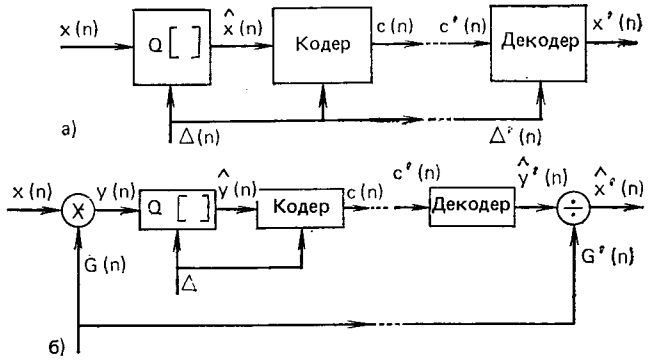


Рис. 5.22. Структурная схема адаптивного квантователя; *а*) с переменным шагом; *б*) с переменным усилением

ристики квантователя не изменяются, а постоянный уровень дисперсии поддерживается за счет переменного коэффициента усиления. В первом случае шаг квантования должен увеличиваться или уменьшаться при увеличении или уменьшении дисперсии входного сигнала соответственно. В случае неравномерного квантования это приводит к соответствующему масштабированию интервалов и уровней квантования. Во втором подходе, применимом в равной мере как к равномерному, так и к неравномерному квантователям, коэффициент усиления изменяется обратно пропорционально дисперсии входного сигнала так, чтобы поддерживать ее постоянной. В обоих случаях необходимо оценивать изменяющиеся во времени характеристики сигнала.

При изучении текущих свойств сигнала полезно выяснить, как быстро происходят их изменения. Значения сигнала изменяются от отсчета к отсчету или на малом количестве отсчетов и могут

рассматриваться как *мгновенные*. Максимальное значение амплитуды сигнала на невокализованном или вокализованном сегменте речи остаются относительно постоянными в течение длительного интервала времени. Такие изменения называются *слоговыми*, и это означает, что они проявляются на интервалах времени, сравнимых по протяженности с длительностью одного слога. При рассмотрении методов квантования целесообразно их классифицировать в соответствии с тем, медленно или быстро происходит адаптация, т. е. является она слоговой или мгновенной.

Имеется два класса схем адаптивного квантования. В первом амплитуда или дисперсия входного сигнала оценивается непосредственно по этому сигналу. Такие схемы называются квантователями с адаптацией по входу. В схемах второго класса шаг квантования подстраивается по выходному сигналу $\hat{x}(n)$ или, что то же самое, по выходной последовательности кодовых слов $c(n)$. Это квантователи с адаптацией по *выходу*. В обоих случаях адаптация может быть как слоговой, так и мгновенной.

5.4.1. Адаптация по входному сигналу

На рис. 5.23 показана в общем виде схема квантователя с адаптацией по входному сигналу. Для простоты будем полагать, что квантователь равномерный и, таким образом, достаточно изменять только шаг квантования. Полученные результаты затем

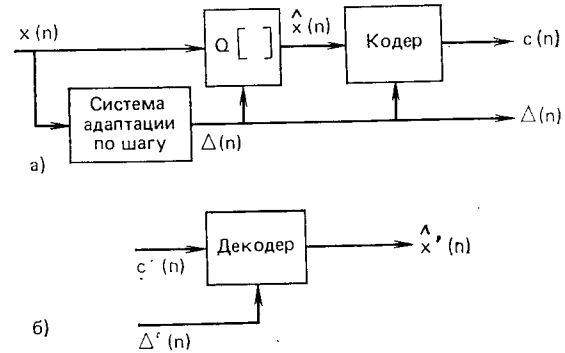


Рис. 5.23. Общая схема адаптивного по входу квантователя: *а*) кодер; *б*) декодер

легко можно обобщить на случай неравномерного квантования. Шаг квантования, используемый при квантовании $x(n)$ (рис. 5.23*а*), должен быть известен на приемной стороне (рис. 5.23*б*). Таким образом, отсчет описывается кодовым словом и шагом квантования. Если $c'(n)=c(n)$ и $\Delta'(n)=\Delta(n)$, то $\hat{x}'(n)=\hat{x}(n)$, однако если $c'(n)\neq c(n)$ или $\Delta'(n)\neq\Delta(n)$, например имеются ошибки при передаче, то $\hat{x}(n)\neq\hat{x}(n)$. Конкретное влияние ошибок определяется методом адаптации. На рис. 5.24 изображена общая схема квантователя с адаптацией по входу на основе усилителя с

переменным коэффициентом усиления. В этом случае квантованный сигнал описывается совместно кодовым словом и коэффициентом усиления.

Для того чтобы разобраться в работе схемы квантователя с адаптацией по входу, полезно рассмотреть ряд примеров. В боль-

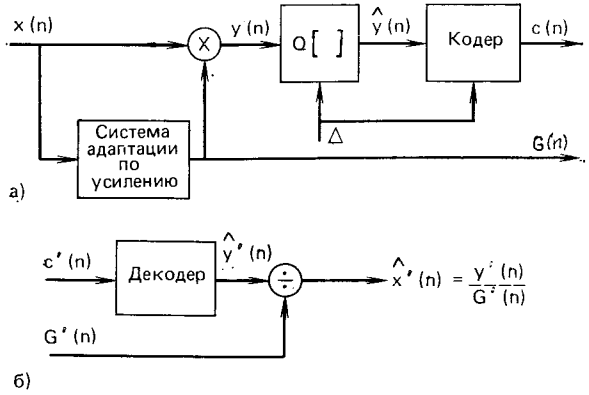


Рис. 5.24. Общий квантователь с адаптацией по входу и с переменным усилением:
а) кодер; *б*) декодер

шинстве систем такого рода используется оценка дисперсии входного сигнала. В этом случае шаг или уровни квантования устанавливаются пропорционально среднему квадратическому отклонению сигнала, а коэффициент усиления — обратно пропорционально.

Общий подход состоит в предложении, что дисперсия пропорциональна кратковременной энергии, которая, как отмечалось ранее, представляет собой сигнал на выходе фильтра нижних частот, на входе которого действует сигнал $x^2(n)$, т. е.

$$\sigma^2(n) = \sum_{n=-\infty}^{\infty} x^2(m)\, h(n-m), \tag{5.53}$$

где $h(n)$ — импульсная характеристика фильтра нижних частот. [Для стационарного сигнала легко показать (см. задачу 5.7), что математическое ожидание $\sigma^2(n)$ пропорционально σ^2_x.] Например,

$$h(n) = \begin{cases} \alpha^{n-1}, & n \geqslant 1; \\ 0, & \textit{в противном случае,} \end{cases} \tag{5.54}$$

Воспользовавшись (5.54) и (5.53), получаем

$$\sigma^2(n) = \sum_{m=-\infty}^{n-1} x^2(m)\, \alpha^{n-m-1}. \tag{5.55}$$

Можно показать, что $\sigma^2(n)$ удовлетворяет также уравнению

$$\sigma^2(n) = \alpha\sigma^2(n-1) + x^2(n-1) \tag{5.56}$$

(для устойчивости потребуем, чтобы $0<\alpha<1$). Шаг квантования для схемы рис. 5.23*а* теперь будет равен

$$\Delta(n) = \Delta_0 \sigma(n) \tag{5.57}$$

и переменный параметр адаптации[1]

$$G(n) = G_0/\sigma(n). \tag{5.58}$$

Параметр α определяет протяженность интервала времени, на котором сигнал вносит основной вклад в оценку дисперсии. На рис. 5.25 изображен пример квантователя в системе разностной ИКМ [13]. На рис. 5.25*а* показана траектория оценки среднего

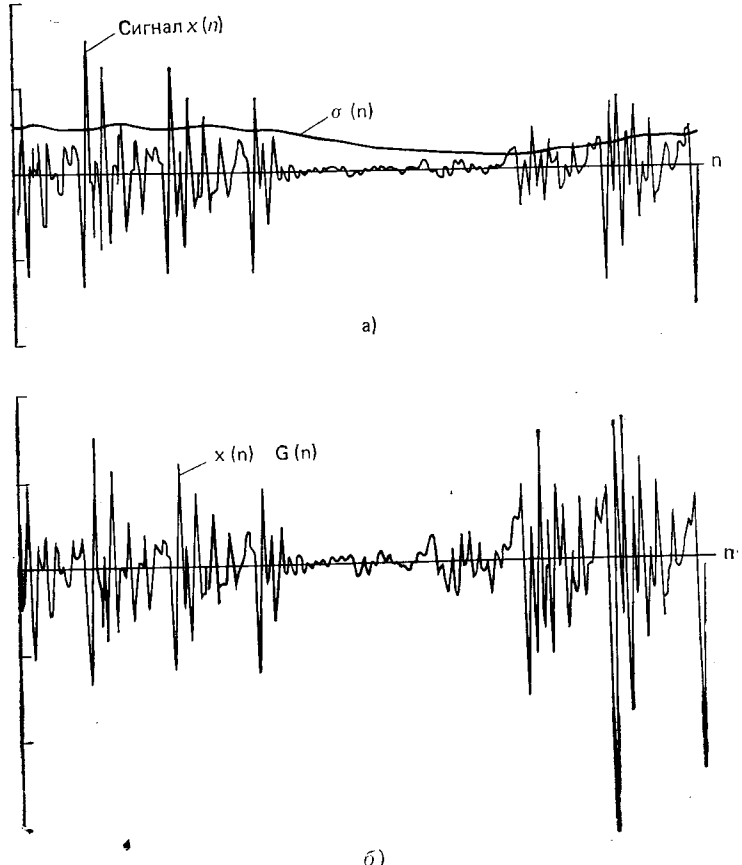


Рис. 5.25. Пример оценки (5.56):
а) временна́я диаграмма сигнала $x(n)$ и $\sigma(n)$ для $\alpha=0{,}99$; *б*) произведение переменного коэффициента усиления и сигнала [13]

[1] Предполагается, что константы Δ_0 и G_0 входят в коэффициент усиления фильтра.

квадратического отклонения совместно с входным сигналом $\alpha = 0{,}99$. На рис. 5.25*б* показана последовательность $y(n) = x(n)G(n)$. При таком выборе параметра α глубокий провал в амплитуде сигнала не в полной мере компенсируется изменением коэффициента усиления. На рис. 5.26 представлены аналогичные

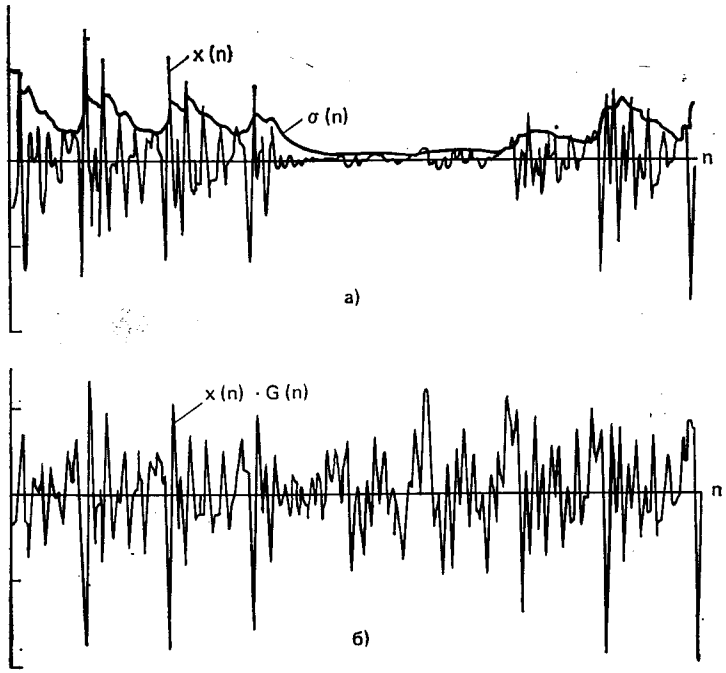


Рис. 5.26. Оценка (5.56):
а) $x(n)$ и $\sigma(n)$ для $\alpha = 0{,}9$; *б*) произведение $x(n)G(n)$

результаты при $\alpha = 0{,}90$. В этом случае система быстрее реагирует на изменение амплитуды входного сигнала. Таким образом, в этом случае дисперсия $y(n) = G(n)x(n)$ остается постоянной, несмотря на внезапное уменьшение амплитуды входного сигнала. В первом случае при $\alpha = 0{,}99$ постоянная времени (число отсчетов до затухания (e^{-1})) составляет около 100 (или 12,5 мс при частоте дискретизации 8 кГц). Во втором случае при $\alpha = 0{,}90$ постоянная времени равна 9 (или около 1 мс при частоте дискретизации 8 кГц). Таким образом, можно считать, что при $\alpha = 0{,}99$ имеется слоговая, а при $\alpha = 0{,}90$ — мгновенная адаптация.

Как следует из рис. 5.25*а* и 5.26*а*, оценка среднего квадратического отклонения или обратная ей величина $G(n)$ представляет собой медленно меняющуюся функцию времени по сравнению с исходным сигналом. Частота дискретизации коэффициента усиления (или шага квантования) определяется шириной полосы пропускания фильтра нижних частот. Так, для случая рис. 5.25 и 5.26 частоты, на которых коэффициент усиления фильтра уменьшается на 3 дБ, равен соответственно 13 и 135 Гц при частоте дискретизации 8 кГц. Важно выбрать наиболее низкую частоту дискретизации, поскольку общая скорость передачи информации складывается из скорости передачи выходного сигнала квантователя и скорости передачи коэффициента усиления. Коэффициент усиления (или шаг квантования) в схемах рис. 5.24 и 5.23 перед передачей следует подвергнуть дискретизации и квантованию.

Перед квантованием следует ограничить диапазон изменения коэффициента усиления или шага квантования. Определим пределы G и Δ интервалом

$$G_{min} \leqslant G(n) \leqslant G_{max} \tag{5.59}$$

или

$$\Delta_{min} \leqslant \Delta(n) \leqslant \Delta_{max}. \tag{5.60}$$

Отношение этих пределов определяет динамический диапазон системы. Для получения примерно постоянного отношения сигнал/шум в диапазоне 40 дБ требуется, чтобы G_{max}/G_{min} или $\Delta_{max}/\Delta_{min} = 100$.

Пример улучшения отношения сигнал/шум при адаптации дан в исследовании, проведенном Ноллом [12][1]. Он рассмотрел алгоритмы с адаптацией по входу и оценкой дисперсии в виде

$$\sigma^2(n) = \frac{1}{M} \sum_{m=n}^{n+M-1} x^2(m). \tag{5.61}$$

Коэффициент усиления (или шаг квантования) передавался через каждые M отсчетов. В данном случае в системе используется буфер объемом M ячеек, и коэффициент усиления, как и шаг квантования, определяется по всем отсчетам, а не только по последнему отсчету, как в предыдущем случае.

В табл. 5.4 приведены результаты сравнения различных трехразрядных квантователей в случае речевого сигнала с известной

Таблица 5.4

Отношение сигнал/шум для адаптивного восьмиуровневого квантователя с адаптацией по входу

Закон	Неадаптивный SNR, дБ	Адаптивный (M=128) SNR, дБ	Адаптивный (M=1024) SNR, дБ
μ-закон ($\mu = 100$, $X_{max} = 8\sigma_x$)	9,5	—	—
Гаусса	7,3/6,7*	15,0/14,7	12,1/11,3
Лапласа	9,9/7,4	13,3/13,4	12,8/11,5

* В числителе — для неравномерного, в знаменателе — для равномерного квантователя.

[1] Эти методы исследовались также в [14]. Для обозначения процесса вычисления коэффициента или шага квантования по M отсчетам здесь введен термин «блоковое компандирование».

дисперсией[1]. В первой колонке представлены различные типы квантователей. Во второй колонке — отношение сигнал/шум без адаптации. В третьей и четвертой колонках — отношение сигнал/шум при адаптации на основе оценки дисперсии (5.61) с $M=128$ и 1024 соответственно. Для данного речевого сигнала адаптивные квантователи позволяют получить выигрыш в отношении сигнал/шум на 5,6 дБ. Сходных результатов можно ожидать и на других фразах речевого сигнала. Таким образом, адаптивный квантователь обладает преимуществом по сравнению с неадаптивным неравномерным квантователем. Дополнительное преимущество адаптивных квантователей, не отраженное в табл. 5.4, заключается в том, что соответствующим выбором Δ_{min} и Δ_{max} можно увеличить отношение сигнал/шум при сохранении малых шумов незанятого канала и широкого динамического диапазона. Это справедливо и в общем случае для большинства правильно спроектированных адаптивных схем. Таким образом, адаптивное квантование позволяет достигнуть лучших результатов по сравнению с мгновенным компандированием и квантованием по минимуму дисперсии ошибки.

5.4.2. Адаптация по выходному сигналу

Квантователь второго типа показан на рис. 5.27 и 5.28, где отмечено, что дисперсия входного сигнала оценивается по выходному квантованному сигналу или по последовательности кодовых слов. Как и в случае адаптации по входному сигналу, в данном случае шаг квантования и коэффициент усиления прямо и обратно пропорциональны дисперсии входного сигнала в соответствии

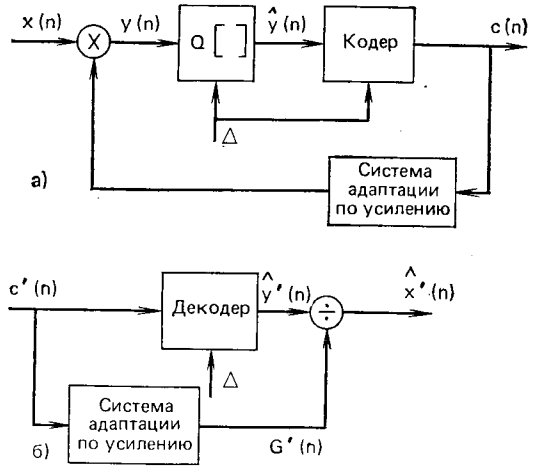


Рис. 5.27. Общая схема адаптации по выходу с переменным усилением:
а) кодер; *б*) декодер

[1] Результаты получены на реальном речевом сигнале.

с (5.57) и (5.58). Такие схемы обладают важным преимуществом, состоящим в том, что шаг квантования или коэффициент усиления не требуется хранить или передавать по каналу связи, поскольку они получены по последовательности кодовых слов. Недостатком подобных квантователей является высокая чувствительность к ошибкам в кодовых словах, ибо эти ошибки приводят не только к неправильной установке уровня квантования, но и к ошибкам в шаге квантования.

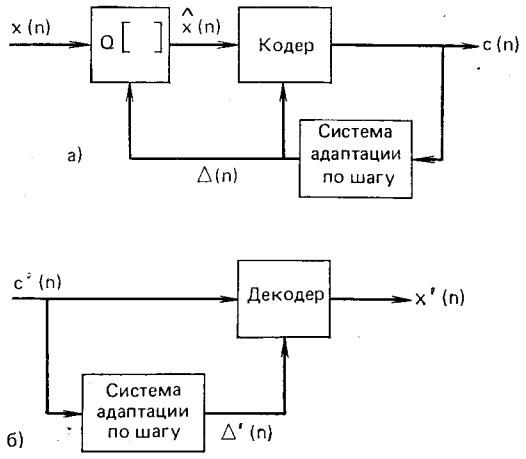


Рис. 5.28. Общая схема адаптации по выходу с переменным шагом:
а) кодер; *б*) декодер

Один из простых подходов состоит в использовании уравнения (5.53) непосредственно для выходного сигнала квантователя:

$$\sigma^2(n) = \sum_{m=-\infty}^{\infty} \hat{x}^2(m)\, h(n-m). \tag{5.62}$$

В данном случае, однако, нельзя применить буфер для использования нереализуемого фильтра, т. е. оценка дисперсии должна быть основана только на последнем отсчете $\hat{x}(n)$, поскольку текущее значение $\hat{x}(n)$ можно получить только после квантования, которое осуществляется после оценки дисперсии. Можно, например, использовать фильтр с импульсной характеристикой

$$h(n) = \begin{cases} \alpha^{n-1}, & n \geqslant 1; \\ 0, & \textit{в противном случае} \end{cases} \tag{5.63}$$

как в уравнении (5.55). Фильтр может также иметь импульсную характеристику, равную

$$h(n) = \begin{cases} 1/M, & 1 \leqslant m \leqslant M; \\ 0, & \textit{в противном случае,} \end{cases} \tag{5.64}$$

так что

$$\sigma^2(n) = \frac{1}{M} \sum_{m=n-M}^{n-1} x^2(m). \tag{5.65}$$

Такой квантователь исследован Ноллом [12], который обнаружил, что удовлетворительные результаты по настройке констант Δ_0 или G_0 в (5.57) и (5.58) (отношение сигнал/шум квантования порядка 12 дБ при трехразрядном квантователе) достигаются при ис-

пользовании окна протяженностью всего в два отсчета. Бо́льшие значения M приводят лишь к незначительному улучшению результатов.

Несколько иной подход (рис. 5.28) подробно исследован Джаянтом [15]. Здесь шаг квантования в адаптивном квантователе изменяется в соответствии с уравнением

$$\Delta(n) = P\,\Delta(n-1), \tag{5.66}$$

где множитель P зависит лишь от значения предшествующего кодового слова $|c(n-1)|$. Это показано на рис. 5.29 для трехразряд-

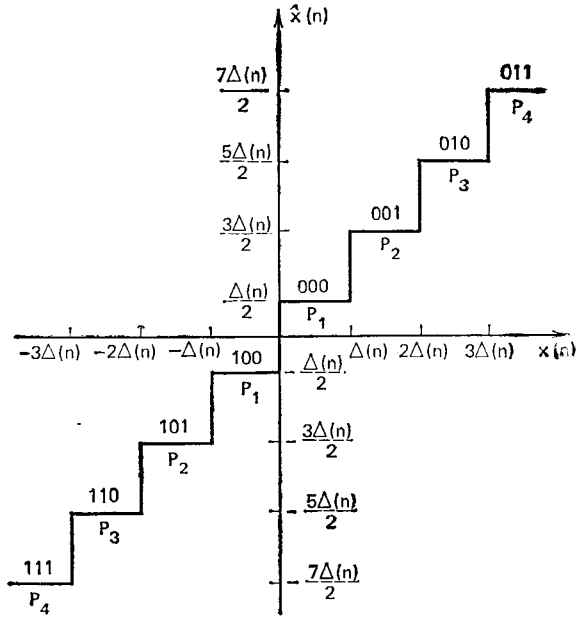


Рис. 5.29. Характеристика трехразрядного адаптивного квантователя

ного равномерного квантователя, где старший разряд — знаковый, а остаток кодового слова — величина отсчета. При этом

$$\hat{x}(n) = \frac{\Delta(n)\,\mathrm{sign}\,(c(n))}{2} + \Delta(n)\,c(n), \tag{5.67}$$

где $\Delta(n)$ удовлетворяет уравнению (5.66). Поскольку $\Delta(n)$ зависит от предшествующего значения шага квантования и предшествующего значения кодового слова, последовательность кодовых слов полностью описывает сигнал. На практике необходимо ввести ограничения

$$\Delta_{min} \leqslant \Delta(n) \leqslant \Delta_{max}. \tag{5.68}$$

Отношение $\Delta_{max}/\Delta_{min}$ определяет динамический диапазон шага квантования.

Способ изменения значения P в уравнении (5.66) на основе кодового слова $|c(n-1)|$ очевиден. Если предшествующее кодовое слово соответствует наибольшему положительному или наибольшему отрицательному уровню, то естественно предположить, что квантователь переполнился и шаг квантования слишком мал. В этом случае множитель должен быть больше единицы. Наоборот, в том случае, если кодовое слово соответствует наименьшему положительному или наименьшему отрицательному уровню, целесообразно уменьшить шаг квантования путем умножения его на число, меньшее единицы. Разработка подобного квантователя включает выбор множителя, соответствующего каждому из 2^b кодовых слов B-разрядного квантователя. Джаянт [15] решил эту проблему путем определения набора множителей, которые минимизируют мощность шума квантования. Были получены теоретические результаты для сигнала, распределенного по закону Гаусса, и эмпирические результаты для речевого сигнала на основе итеративных вычислений. Результаты, полученные Джаянтом, приведены на рис. 5.30, где показана приближенная зависимость множителей от величины

$$Q = [1 + 2\,|\,c(n-1)\,|]/(2^B - 1). \tag{5.69}$$

Заштрихованная область соответствует ожидаемому изменению множителей при изменении статистики входного сигнала или B. Конкретные значения множителей будут лежать внутри заштрихованной области. Важно подчеркнуть, что множители будут такими, чтобы нарастание происходило более интенсивно, чем убывание. В табл. 5.5 приведены значения множителей для $B=2, 3, 4, 5$.

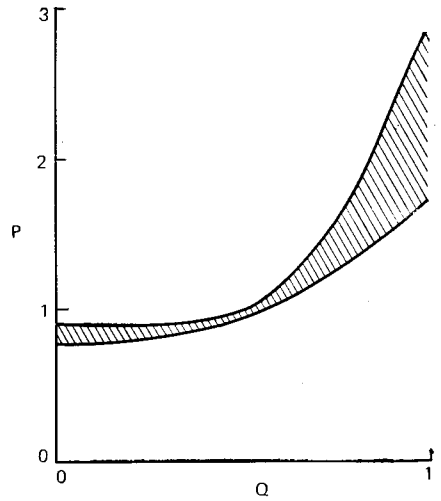


Рис. 5.30. Общая форма функции кратности речевого квантователя для $B>2$ [15]

Увеличение отношения сигнал/шум, достигаемое при использовании метода адаптивного квантования, иллюстрирует табл. 5.6. Множители из табл. 5.5 обеспечивают $\Delta_{max}/\Delta_{min}=100$. Как следует из табл. 5.6, достигается выигрыш около 4,7 дБ по отношению к квантованию на основе μ-закона. Улучшение на 2—4 дБ достигается также по отношению к неадаптивному случаю. В другом исследовании Нолл [7] установил, что отношения сигнал/шум для трехразрядного квантователя, построенного по μ-закону, и адаптивного квантователя составляют 9,4 и 14,1 дБ соответственно. В этом эксперименте, в отличие от эксперимента Джаянта, множители были равны {0,8; 0,8; 1,3; 1,9}. Тот факт, что столь различ-

Таблица 5.5

Множители изменения шага в методах адаптивного квантования

B	Тип кодера	
	ИКМ	РИКМ
2	0,6; 2,2	0,8; 1,6
3	0,85; 1; 1; 1,5	0,9; 0,9; 1,25; 1,75
4	0,8; 0,8; 0,8; 0,8 1,2; 1,6; 2,0; 2,4	0,9; 0,9; 0,9; 0,9 1,2; 1,6; 2,0; 2,4
5	0,85; 0,85; 0,85; 0,85 0,85; 0,85; 0,85; 0,85 1,2; 1,4; 1,6; 1,8; 2,0; 2,2; 2,4; 2,6	0,9; 0,9; 0,9; 0,9; 0,95; 0,95; 0,95; 0,95; 1,2; 1,5; 1,8; 2,1; 2,4; 2,7; 3,0; 3,3

ные множители в разных экспериментах дают сравнимые результаты, означает, что значения множителей не столь существенны.

Таблица 5.6

Выигрыш в отношении сигнал/шум, дБ, при использовании оптимальных множителей изменения шага для адаптивного квантования [15]

B	Логарифмическая ИКМ с μ-законом (μ=100) квантования	Адаптивная ИКМ с равномерным квантованием
2	3	9
3	8	15
4	15	19

5.4.3. Общие замечания

Как следует из результатов данного параграфа, имеется практически неограниченное множество методов адаптивного квантования. Большинство этих методов приводит к выигрышу в отношении сигнал/шум по сравнению с квантованием по μ-закону и обеспечивает такой же динамический диапазон. Выбирая малым Δ_{min}, можно уменьшить шум незанятого канала. Таким образом, адаптивное квантование обладает рядом полезных свойств. Однако не следует ожидать, что дальнейшее совершенствование методов адаптации даст существенный выигрыш в скорости передачи, поскольку эти методы используют лишь наши знания о распределении мгновенных значений речевого сигнала. Поэтому в следующем параграфе рассмотрим использование корреляции между соседними отсчетами на основе методов разностного квантования.

5.5. Общая теория разностного квантования

Анализ рис. 5.7 позволяет сделать вывод, что между соседними отсчетами сигнала имеется значительная корреляция, которая слабо убывает по мере увеличения интервала между отсчетами.

Это означает, что, вообще говоря, сигнал изменяется медленно и разность между соседними отсчетами будет иметь меньшую дисперсию, чем исходный сигнал, в чем можно легко убедиться (см. задачу 5.10). Такие рассуждения позволяют ввести общие методы разностного квантования, показанные на рис. 5.31 [16, 17]. Здесь на входе квантователя действует сигнал

$$d(n) = x(n) - \tilde{x}(n), \tag{5.70}$$

который представляет собой разность сигнала $x(n)$ и оценки предсказанного значения входного сигнала, которое обозначено через $\tilde{x}(n)$. Предсказанное значение представляет собой выходной сигнал предсказателя P; входным сигналом, как будет ясно из дальнейшего, является квантованный входной сигнал $x(n)$. Разностный сигнал можно также назвать погрешностью предсказания, поскольку это — величина, на которую предсказанное значение отличается от точного значения входного сигнала. Оставив пока в стороне вопрос о вычислении предсказанного значения, отметим, что вместо входного сигнала квантованию подвергается разностный сигнал. Квантователь может быть адаптивный или неадаптивный, равномерный или неравномерный, но во всех случаях его параметры должны соответствовать дисперсии погрешности предсказания. Квантованная погрешность может быть представлена в виде

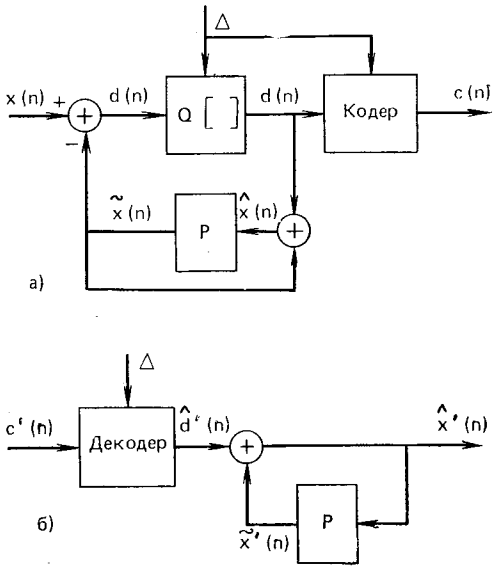


Рис. 5.31. Общая схема разностного квантования:
а) кодер; *б*) декодер

$$\hat{d}(n) = d(n) + e(n), \tag{5.71}$$

где $e(n)$ — ошибка квантования. В соответствии с рис. 5.31*а* погрешность с выхода квантователя складывается с предсказанным значением для получения квантованного сигнала:

$$\hat{x}(n) = \tilde{x}(n) + \hat{d}(n). \tag{5.72}$$

Подставляя (5.70) и (5.71) в (5.72), видим, что

$$\hat{x}(n) = x(n) + e(n), \tag{5.73}$$

Таким образом, независимо от свойств устройства, обозначенного через p, квантованный речевой сигнал отличается от входного только на величину шума квантования разностного сигнала. Таким образом, если предсказатель хороший, дисперсия $d(n)$ будет меньше, чем дисперсия $x(n)$, так что квантователь с заданным количеством уровней даст меньшую погрешность при квантовании разности, чем при квантовании исходного сигнала.

Следует отметить, что для передачи по каналу связи или занесения в запоминающее устройство используется квантованный разностный сигнал. Схема, восстанавливающая входной сигнал по последовательности кодовых слов, неявно имеется и на рис. 5.31*а*. В явном виде она показана на рис. 5.31*б* и содержит декодер, восстанавливающий разностный сигнал, и предсказатель (такой же, как и на передающей стороне). Очевидно, что при совпадении $c'(n)$ и $c(n)$ сигнал $\hat{x}'(n)=\hat{x}(n)$ и отличается от $x(n)$ лишь ошибкой квантования, вносимой квантованием сигнала $d(n)$.

Отношение сигнал/(шум квантования) для системы рис. 5.31 равно

$$SNR = E[x^2(n)]/E[e^2(n)] = \sigma_x^2/\sigma_e^2. \tag{5.74}$$

Выражение (5.74) можно переписать в виде

$$SNR = (\sigma_x^2/\sigma_d^2)(\sigma_d^2/\sigma_e^2) = G_p \cdot SNR_Q, \tag{5.75}$$

где

$$SNR_Q = \sigma_d^2/\sigma_e^2 \tag{5.76}$$

представляет собой отношение сигнал/(шум квантования), а величина

$$G_p = \sigma_x^2/\sigma_d^2 \tag{5.77}$$

определяется как коэффициент усиления, обусловленный разностным кодированием[1].

Величина SNR_Q зависит от конкретно используемого квантователя и известных свойств $d(n)$, SNR_Q и может быть максимизирован на основе методов предыдущего параграфа. Величина G_p определяет выигрыш в отношении сигнал/шум при использовании разностного представления. Очевидно, что необходимо увеличить p выбором схемы предсказания. Для данного сигнала σ^2_x — величина фиксированная, следовательно, G_p можно максимизировать за счет минимизации знаменателя в (5.77), т. е. минимизацией дисперсии погрешности предсказания.

Для решения поставленной задачи следует определить тип предсказателя p. Один из методов, хорошо согласующийся с предшествующими рассуждениями о моделях сигнала и приводящий к несложным математическим выкладкам, состоит в использовании

[1] Величина G_p является коэффициентом усиления системы, обратной предсказателю, т. е. авторегрессионной модели сигнала. Этот коэффициент характеризует эффективность предсказания. (*Прим. ред.*)

линейного предсказания, т. е. $\tilde{x}(n)$ представляется в виде линейной комбинации предшествующих значений сигнала

$$\tilde{x}(n) = \sum_{k=1}^{p} \alpha_k \hat{x}(n-k). \tag{5.78}$$

Предсказанное значение является, таким образом, выходным сигналом фильтра с передаточной функцией вида

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k}, \tag{5.79}$$

на вход которого поступает восстановленный сигнал $\hat{x}(n)$. Попутно отметим, что восстановленный сигнал представляет собой выходной сигнал фильтра с передаточной функцией

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}, \tag{5.80}$$

на вход которого поступает квантованная погрешность. Дисперсия погрешности предсказания на рис. 5.31

$$\sigma_d^2 = E[d^2(n)] = E[(x(n)-\tilde{x}(n))^2] = E\left[x(n) - \sum_{k=1}^{p} \alpha_k \hat{x}(n-k)\right]^2 =$$

$$= E\left[x(n) - \sum_{k=1}^{p} \alpha_k x(n-k) - \sum_{k=1}^{p} \alpha_k e(n-k)\right]^2. \tag{5.81}$$

Для вычисления множества параметров предсказания, которые минимизируют σ^2_d, продифференцируем σ^2_d по каждому параметру и приравняем производную к нулю. Это приводит к системе уравнений

$$\frac{\partial \sigma_d^2}{\partial \alpha_j} = -2E\left[\left[x(n) - \sum_{k=1}^{p} \alpha_k (x(n-k) + e(n-k))\right][x(n-j) + e(n-j)]\right] = 0, \quad 1 \leqslant j \leqslant p. \tag{5.82}$$

Уравнения (5.82) можно переписать в более компактной форме:

$$E[(x(n)-\tilde{x}(n))\hat{x}(n-j)] = E[d(n)\hat{x}(n-j)] = 0, \quad 1 \leqslant j \leqslant p. \tag{5.83}$$

Если коэффициенты предсказания таковы, что σ^2_d минимально, то погрешность предсказания не коррелирована с последними значениями сигнала на входе предсказателя, т. е. ортогональна $\hat{x}(n-j)$ при $1 \leqslant j \leqslant p$.

Уравнения (5.82) можно представить в виде

$$E[x(n-j)x(n)] + E[e(n-j)x(n)] = \sum_{k=1}^{p} \alpha_k E[x(n-j)x(n-k)] +$$

$$+\sum_{k=1}^{p}\alpha_k E[e(n-j)x(n-k)]+\sum_{k=1}^{p}\alpha_k E[x(n-j)e(n-k)]+$$

$$+\sum_{k=1}^{p}\alpha_k E[e(n-j)e(n-k)], \tag{5.84}$$

где $1\leqslant j\leqslant p$. Предполагая, что квантование достаточно точное, допустим, что $e(n)$ не коррелирована с $x(n)$ и $e(n)$ — стационарная последовательность. Тогда

$$E[x(n-j)e(n-k)]=0 \text{ для всех } n,\ j \text{ и } k \tag{5.85}$$

и

$$E[e(n-j)e(n-k)]=\sigma_e^2\delta(j-k). \tag{5.86}$$

Используя эти предположения, уравнение (5.84) можно упростить:

$$\varphi(j)=\sum_{k=1}^{p}\alpha_k[\varphi(j-k)+\sigma_e^2\delta(j-k)], \tag{5.87}$$
$$1\leqslant j\leqslant p,$$

где $\varphi(j)$ — автокорреляционная функция $x(n)$. Если разделить обе части этого уравнения на σ^2_x и обозначить нормированную корреляционную функцию через

$$\rho(j)=\varphi(j)/\sigma_x^2, \tag{5.88}$$

то (5.87) перепишется в матричной форме в виде

$$\boldsymbol{\rho}=\mathbf{C}\boldsymbol{\alpha}, \tag{5.89а}$$

где

$$\mathbf{C}=\begin{Vmatrix} \left(1+\frac{1}{SNR}\right) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \left(1+\frac{1}{SNR}\right) & \cdots & \rho(p-2) \\ \vdots & \vdots & & \\ \rho(p-1) & \rho(p-2) & \cdots & \left(1+\frac{1}{SNR}\right) \end{Vmatrix}, \tag{5.89б}$$

$$\boldsymbol{\rho}=\begin{Vmatrix}\rho(1)\\ \rho(2)\\ \vdots\\ \rho(p)\end{Vmatrix}, \qquad \boldsymbol{\alpha}=\begin{Vmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_p\end{Vmatrix} \tag{5.89в), (5.89г)}$$

и $SNR=\sigma^2_x/\sigma^2_e$. Таким образом, вектор оптимальных коэффициентов предсказания получается как решение матричного уравнения (5.89а):

$$\boldsymbol{\alpha}=\mathbf{C}^{-1}\boldsymbol{\rho}. \tag{5.90}$$

В общем случае матрица $\mathbf{C}^{-1}$ может быть вычислена различными численными методами, в том числе и методом, основанным на том обстоятельстве, что $\mathbf{C}$ — теплицева матрица (см. гл. 8). Однако в общем случае уравнения (5.89а) неразрешимы, поскольку матрица содержит члены, которые зависят от отношения сигнал/шум [см. (5.89в)], а отношение сигнал/шум, в свою очередь, определяется коэффициентами линейного предсказания (5.89а). Один из выходов из этого положения заключается в том, чтобы пренебречь членом $1/SNR$ в (5.89). При $p=1$ такое упрощение не требуется, поскольку уравнение (5.90) можно решить непосредственно:

$$\alpha_1=\frac{\rho(1)}{1+(1/SNR)}. \tag{5.91}$$

Уравнение (5.91) показывает, что $\alpha_1<\rho(1)$.

Несмотря на трудности прямого решения уравнений относительно коэффициентов предсказания, можно получить выражение для оптимального G_p через α'_i. Для этого определим σ^2_d, переписав (5.81) в виде

$$\sigma_d^2=E[(x(n)-\tilde{x}(n))(x(n)-\tilde{x}(n))]=E[(x(n)-\tilde{x}(n))x(n)]-$$
$$-E[(x(n)-\tilde{x}(n))\tilde{x}(n)]. \tag{5.92}$$

Используя (5.83), можно показать, что для оптимальных коэффициентов предсказания второе слагаемое в последнем уравнении обращается в нуль, т. е. предсказанное значение не коррелировано с погрешностью предсказания (см. задачу 5.12). Таким образом, можно записать

$$\sigma_d^2=E[(x(n)-\tilde{x}(n))x(n)]=E[x^2(n)]-E\sum_{k=1}^{p}\alpha_k(x(n-k)+$$
$$+e(n-k))x(n)\Big]. \tag{5.93}$$

Используя предположение о некоррелированности сигнала и шума, получаем

$$\sigma_d^2=\sigma_x^2-\sum_{k=1}^{p}\alpha_k\varphi(k)=\sigma_x^2\left[1-\sum_{k=1}^{p}\alpha_k\rho(k)\right]. \tag{5.94}$$

Учитывая (5.77), имеем

$$(G_p)_{opt}=\left[1-\sum_{k=1}^{p}\alpha_k\rho(k)\right]^{-1} \tag{5.95}$$

где α_k удовлетворяют уравнению (5.89а). При $p=1$ можно оценить эффект влияния субоптимального значения α_1 на величину $G_p=\sigma^2_x/\sigma^2_d$. Из (5.95) получим

$$(G_p)_{opt}=[1-\alpha_1\rho(1)]^{-1}. \tag{5.96}$$

Выбирая произвольное значение α_1 и повторяя дифференцирование, приводящее к (5.94), получим

$$\sigma_d^2 = \sigma_x^2 [1 - 2\alpha_1 \rho(1) + \alpha_1^2] + \alpha_1^2 \sigma_e^2 \tag{5.97}$$

или

$$(G_p)_{\text{arb}} = \left[1 - 2\alpha_1 \rho(1) + \alpha_1^2 \left(1 + \frac{1}{SNR}\right)\right]^{-1}. \tag{5.98}$$

Член α^2_1/SNR показывает увеличение дисперсии $d(n)$ из-за шума $e(n)$ в петле обратной связи. Легко показать (см. задачу 5.13), что (5.98) можно переписать в виде

$$(G_p)_{\text{arb}} = \frac{1 - (\alpha_1^2 \mid SNR_Q)}{1 - 2\alpha_1 \rho(1) + \alpha_1^2} \tag{5.99}$$

при любом α_1, включая оптимальное. Например, если $\alpha_1 = \rho(1)$ (что соответствует субоптимальному случаю в соответствии с (5.91)), получаем

$$(G_p)_{\text{subopt}} = \frac{1 - (\rho^2(1)/SNR_Q)}{1 - \rho^2(1)} = \left[\frac{1}{1 - \rho^2(1)}\right]\left[1 - \frac{\rho^2(1)}{SNR_Q}\right]. \tag{5.100}$$

Таким образом, коэффициент, рассчитанный при отсутствии квантователя, уменьшается за счет второго сомножителя в (5.100), обусловленного шумом квантования в петле обратной связи.

Для получения оптимального коэффициента усиления можно продифференцировать уравнение (5.99) по α_1. Тогда получим уравнение[1]

$$\frac{d(G_p)}{d\alpha_1} = 0, \tag{5.101}$$

которое можно решить при оптимальном α_1.

Предположим, что можно пренебречь слагаемым $1/SNR$ в (5.89). Тогда для предсказания первого порядка из (5.91) следует, что $\alpha_1 = \rho(1)$ и коэффициент имеет вид

$$(G_p)_{\text{opt}} = [1 - \rho^2(1)]^{-1}. \tag{5.102}$$

Таким образом, пока $\rho(1) \neq 0$, отношение сигнал/шум будет увеличиваться за счет предсказания. Ранее приводилась типичная корреляционная функция (см. рис. 5.5) для речевого сигнала на выходе полосового фильтра и фильтра нижних частот при частоте дискретизации 8 кГц [4]. Заштрихованная область показывает изменения $\rho(n)$ корреляции для четырех различных дикторов, а центральная кривая — среднее значение корреляционной функции по всем четырем дикторам. Из кривых видно, что при частоте дискретизации, соответствующей частоте Найквиста, справедливо неравенство

$$\rho(1) > 0{,}80. \tag{5.103}$$

[1] Авторы выражают благодарность профессору П. Ноллу за полезные замечания по изложенному анализу.

Это означает, что

$$(G_p)_{\text{opt}} > 2{,}77 \text{ (или } 4{,}43 \text{ дБ)}. \tag{5.104}$$

Нолл [4] использовал данные рис. 5.5 для вычисления оптимального значения G_p как функции от p для сегментов речи длительностью 55 с, полученных как на выходе фильтра нижних частот, так и на выходе полосового фильтра. Результаты приведены на рис. 5.32[1]. Заштрихованная область показывает разброс результатов для четырех дикторов, центральная кривая получена усреднением по всем дикторам. Видно, что даже при простом предсказателе можно получить выигрыш 6 дБ. Это эквивалентно добавлению еще одного разряда в квантователь, однако, поскольку на самом деле ничего не добавляется, скорость остается неизменной. Отметим также, что коэффициент не достигает 12 дБ ни при каком p, т. е. невозможно получить выигрыш, эквивалентный добавлению

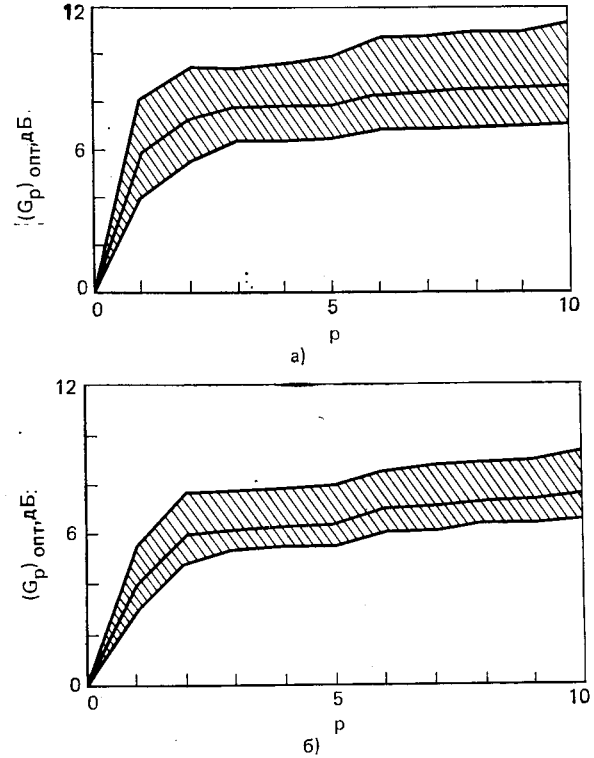


Рис. 5.32. Зависимости оптимального по критерию *SNR* усиления G от числа коэффициентов предсказания:
а) низкочастотная фильтрация речи; *б*) высокочастотная фильтрация речи [7]

[1] Как и ранее, частота дискретизации 8 кГц, т. е. n надо умножить на 125 мкс.

двух разрядов. С другой стороны, применение разностного квантования позволяет понизить скорость передачи при том же отношении сигнал/шум. Это достигается, конечно, ценой усложнения квантователя.

Несколько основных результатов применения схем разностного квантования вытекают из рис. 5.5 и 5.32. Во-первых, разностное квантование обеспечивает выигрыш по сравнению с непосредственным квантованием. Во-вторых, величина выигрыша зависит от величины корреляции. В-третьих, один и тот же предсказатель не может быть оптимальным для различных дикторов и различного речевого сигнала. Эти обстоятельства стимулируют разработку усовершенствованных схем, в которых сохраняется основная структура, изображенная на рис. 5.31. В них применяются различные адаптивные или неадаптивные квантователи или предсказатели для достижения лучшего качества или меньшей скорости передачи речевого сигнала. Далее рассматриваются примеры, позволяющие судить о возможностях таких систем.

5.6. Дельта-модуляция

Примером простого применения разностного квантования является дельта-модуляция (ДМ) [18—24]. В системах такого типа частота дискретизации выбирается во много раз больше, чем частота Найквиста. В результате соседние отсчеты оказываются в большой степени коррелированными. Это следует из результатов § 5.2, где показано, что автокорреляционная функция последовательности отсчетов представляет собой дискретизированную автокорреляционную функцию непрерывного сигнала:

$$\varphi(m) = \varphi_a(mT). \tag{5.105}$$

Используя свойства автокорреляционной функции, естественно предположить, что она возрастает при $T \to 0$. Действительно, можно считать, что, за исключением случая некоррелированного сигнала, имеет место соотношение

$$\varphi(1) \to \sigma_x^2 \quad \text{при} \quad T \to 0. \tag{5.106}$$

Большая корреляция между отсчетами означает, что при уменьшении T можно более точно предсказать текущий отсчет по предшествующим и, следовательно, уменьшить дисперсию погрешности предсказания. Поэтому более «грубый» квантователь может дать хорошие результаты. В действительности в системе с дельта-модуляцией используется простой одноразрядный (двухуровневый) квантователь. Таким образом, скорость передачи при использовании ДМ численно равна частоте дискретизации.

5.6.1. Линейная дельта-модуляция

Схема простейшей системы с дельта-модуляцией приведена на рис. 5.33. В этом случае квантователь имеет только два уровня и шаг квантования фиксирован. Положительный уровень квантования соответствует $c(n) = 0$, а отрицательный $c(n) = 1$. Таким образом,

$$\hat{d}(n) = \begin{cases} \Delta, & c(n) = 0; \\ -\Delta, & c(n) = 1. \end{cases} \tag{5.107}$$

На рис. 5.33 показан простой одношаговый предсказатель первого порядка, для которого оптимальный коэффициент усиления

$$(G_p)_{\text{opt}} = [1 - \rho^2(1)]^{-1}. \tag{5.108}$$

Таким образом, если $\rho(1) \to 1$, то $(G_p)_{\text{opt}} \to \infty$. Этот результат носит качественный характер, ибо при использовании «грубого» двухуровневого квантователя предположения, при которых получено выражение для $(G_p)_{\text{opt}}$, не справедливы.

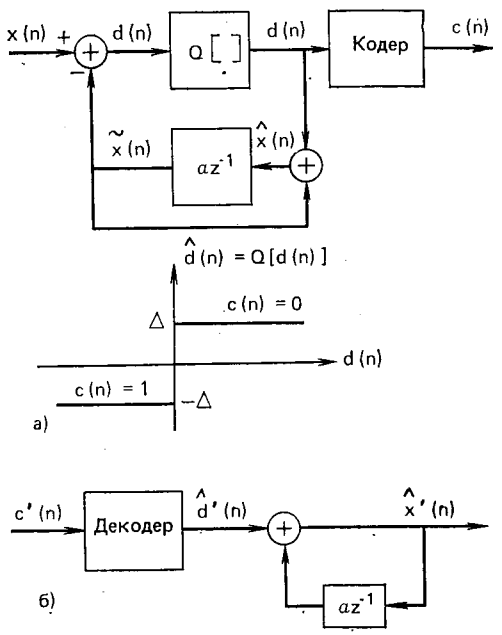


Рис. 5.33. Структурная схема системы с дельта-модуляцией:
а) кодер; *б*) декодер

Эффект квантования можно увидеть из рис. 5.34*а*, где показаны аналоговый сигнал $x_a(t)$, результирующие отсчеты $x(n)$ $\tilde{x}(n)$ и $\hat{x}(n)$ при данном периоде дискретизации в предположении, что α (множитель в петле обратной связи) равен единице. Из рис. 5.33*а* видно, что $\hat{x}(n)$ в общем случае удовлетворяет уравнению

$$\hat{x}(n) = \alpha \hat{x}(n-1) + \hat{d}(n). \tag{5.109}$$

При $\alpha \approx 1$ уравнение описывает дискретный аналог интегратора в том смысле, что осуществляется накопление положительных и отрицательных приращений величины Δ. Отметим, что входной сигнал квантователя имеет вид

$$d(n) = x(n) - \hat{x}(n-1) = x(n) - x(n-1) - e(n-1). \tag{5.110}$$

Таким образом, пренебрегая ошибкой квантования $\hat{x}(n-1)$, значение $d(n)$ можно представить как разность первого порядка сигнала $x(n)$. Разность может рассматриваться как аппроксимация производной входного сигнала, а ее вычисление — как операция, обратная цифровому интегрированию. Если крутизна входного сигнала максимальна, то очевидно, что для того, чтобы последова-

тельность отсчетов $\{\hat{x}(n)\}$ возрастала так же быстро, как и последовательность $\{x(n)\}$ в области максимальной крутизны, необходимо потребовать выполнения неравенства

$$\frac{\Delta}{T} \geqslant \max \left| \frac{dx_a(t)}{dt} \right|. \qquad (5.111)$$

Иначе восстановленный сигнал будет «отставать» от исходного, как это показано в левой части рис. 5.34*а*. Если условие (5.111) выполняется, то такие искажения, называемые *перегрузкой по*

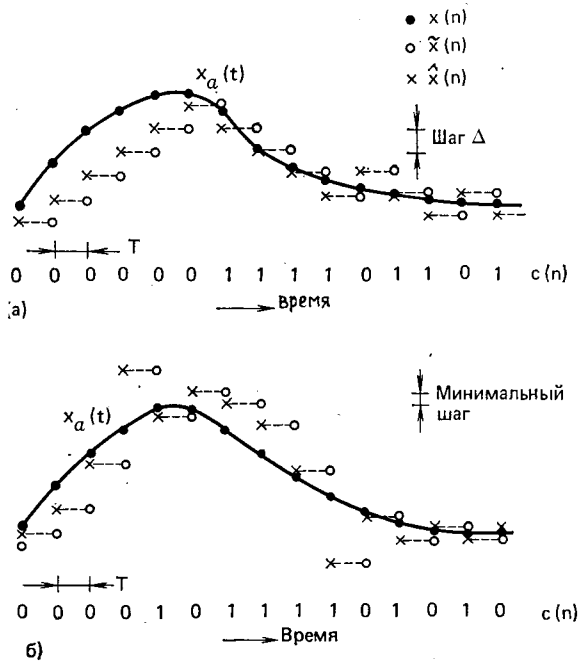


Рис. 5.34. Иллюстрация дельта-модуляции:
а) фиксированный шаг; *б*) адаптивный шаг

крутизне, отсутствуют. Отметим, что поскольку максимальная крутизна $\hat{x}(n)$ ограничивается шагом квантования, то возрастание или убывание последовательности $\hat{x}(n)$ происходит по соответствующей ступенчатой линии. По этой причине фиксированную (неадаптивную) дельта-модуляцию иногда называют линейной (ЛДМ).

Шаг квантования определяет также и максимальную ошибку, когда крутизна мала. Например, если сигнал на входе равен нулю (канал не занят), сигнал на выходе квантователя представляет собой переменную последовательность нулей и единиц, что приводит к флуктуации восстановленного сигнала вокруг нулевого или иного постоянного уровня с размахом Δ. Этот тип ошибок квантования, изображенный в правой части рис. 5.34*а*, называется *шумом дробления*.

Ранее было показано, что для получения большого динамического диапазона необходимо иметь большой шаг квантования; в то же время для точного описания малых сигналов шаг квантования должен быть малым. В данном случае это относится к динамическому диапазону и амплитуде разностного сигнала (или производной аналогового сигнала). Интуитивно ясно, что выбор шага квантования, минимизирующего среднее квадратическое значение шума квантования, приведет к компромиссу между перегрузкой по крутизне и шумом дробления.

На рис. 5.35, заимствованном из подробного исследования дельта-модуляции, проведенного Абатом [21], изображена зави-

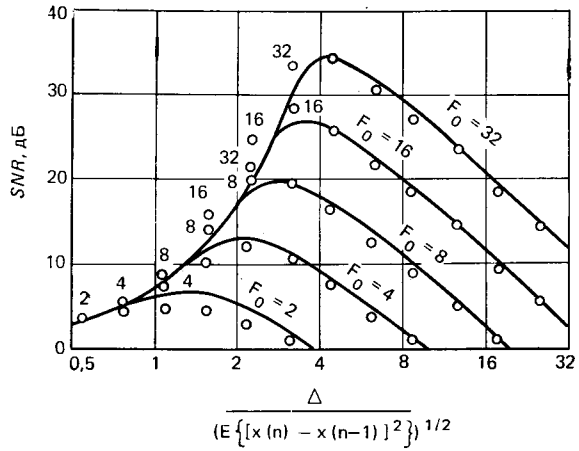


Рис. 5.35. *SNR* для дельта-модуляторов как функция нормированного шага [21]

симость отношения сигнал/шум от $\Delta/(E[(x(n)-x(n-1))^2])^{1/2}$ и параметра $F_0 = F_s/2F_N$, где F_s — частота дискретизации, а F_N — частота Найквиста. Отметим, что

$$\textit{Скорость передачи} = F_s\,(1 \text{ бит}) = F_s = 2F_N F_0. \qquad (5.112)$$

Таким образом, параметр F_0 играет здесь ту же роль, что и количество двоичных единиц на отсчет в многоуровневом квантователе при дискретизации с удвоенной частотой Найквиста. Кривые получены для сигнала с равномерным спектром в полосе и гауссовским распределением. Для речевого сигнала отношение сигнал/шум несколько больше вследствие большей корреляции, хотя форма кривых такая же. Из рис. 5.35 видно, что для данного значения F_0 отношение сигнал/шум достигает максимума при некотором Δ. Значения F_0, лежащие левее этого максимума, соответствуют перегрузке по крутизне, а правее — шуму дробления. Абат [21] вывел эмпирическую формулу

$$\Delta_{\text{opt}} = \{E[(x(n)-x(n-1))^2]\}^{1/2} \ln(2F_0) \qquad (5.113)$$

для оптимального шага квантования, т. е. для расположения максимумов кривой при заданной F_0. Из рис. 5.35 следует, что оптимальное значение сигнал/шум увеличивается на 9 дБ при удвоении F_0. Поскольку удвоение эквивалентно удвоению F_s, можно сказать, что при удвоении скорости передачи отношение сигнал/шум возрастает на 9 дБ. Таким образом, в отличие от ИКМ, где при удвоении количества двоичных единиц на отсчет достигается увеличение отношения сигнал/шум, равное 6 дБ на каждый добавленный бит, здесь увеличение происходит значительно быстрее.

Другая важная особенность кривых рис. 5.35 состоит в том, что они весьма «остроконечны», т. е. отношение сигнал/шум в большой степени зависит от уровня входного сигнала (отметим, что $E[(x(n)-x(n-1))^2]=2\sigma_x^2(1-\rho(1))$. Таким образом, для получения отношения сигнал/шум, равного 35 дБ при частоте Найквиста 3 кГц, требуется скорость передачи 200 кбит/с. Однако даже на этой скорости при постоянном шаге квантования требуемое качество может быть достигнуто для весьма узкого диапазона уровней входного сигнала. Для получения хорошего качества восстановленного речевого сигнала, сравнимого с качеством, достигаемым в семиразрядной логарифмической ИКМ, требуется значительно бо́льшая скорость.

Основное достоинство ЛДМ состоит в ее простоте. Система может быть реализована на простом аналоговом или цифровом интеграторе и, поскольку используется только одноразрядный код, не требовать никакой синхронности по кодовым словам между передатчиком и приемником. Ограничения линейной дельта-модуляции состоят, главным образом, в весьма грубом квантовании погрешности предсказания. С учетом предыдущего обсуждения адаптивных методов квантования естественно предположить, что использование этого подхода в данном случае может существенно улучшить характеристики дельта-модулятора. Наибольший интерес представляют простые адаптивные схемы, которые улучшают характеристики, но не приводят к существенному усложнению системы передачи.

5.6.2. Адаптивная дельта-модуляция

Известен ряд методов адаптивной дельта-модуляции (АДМ). Большинство этих методов основано на адаптации по выходу, когда шаг квантования перестраивается по выходной последовательности кодовых слов. Общий вид системы показан на рис. 5.36. Подобные схемы обладают тем преимуществом, что не требуют синхронизации по кодовым словам, поскольку при отсутствии ошибок шаг квантования как передатчика, так и приемника перестраивается в одной и той же кодовой последовательности.

В данном подпараграфе дается иллюстрация применения адаптивного квантования в дельта-модуляции на примере двух специальных алгоритмов адаптации. Имеется и множество других возможностей, которые можно найти в литературе [20—24].

Первая из рассматриваемых систем была подробно исследована Н. С. Джаянтом [22]. Алгоритм Джаянта для адаптивной

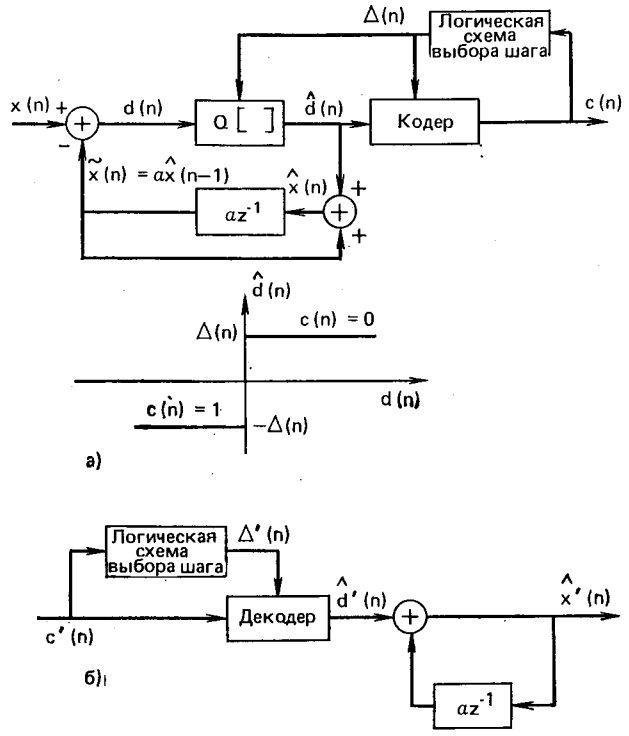


Рис. 5.36. Дельта-модулятор с адаптацией по шагу:
а) кодер; *б*) декодер

дельта-модуляции представляет собой модификацию алгоритма, рассмотренного в 5.4.2. Как и в случае многоуровневого квантователя, шаг квантования подчиняется следующему правилу:

$$\Delta(n)=M\Delta(n-1); \tag{5.114а}$$

$$\Delta_{min}\leqslant\Delta(n)\leqslant\Delta_{max}. \tag{5.114б}$$

В том случае множитель является функцией текущего и предшествующего кодовых слов $c(n)$ и $c(n-1)$. Это возможно, поскольку $c(n)$ зависит только от знака $d(n)$, который задается соотношением

$$d(n)=x(n)-\alpha\hat{x}(n-1). \tag{5.115}$$

Таким образом, знак $d(n)$ определяется перед получением квантованного значения $\hat{d}(n)$, которое возникает после вычисления

$\Delta(n)$ в соответствии с (5.114). Алгоритм выбора множителя шага квантования в уравнении (5.114а) имеет вид

$$\left.\begin{aligned} M=P>1, \quad & c(n)=c(n-1); \\ M=Q<1, \quad & c(n)\neq c(n-1). \end{aligned}\right\} \qquad (5.116)$$

Выбор такого метода адаптации объясняется видом последовательности кодовых слов, наблюдаемой в линейной дельта-модуляции. Например, из рис. 5.34*а* видно, что период перегрузки по крутизне соответствует отрезкам последовательности, состоящим только из нулей или только из единиц. Период шума дробления соответствует последовательности из чередующихся нулей и единиц вида 010101 ... На рис. 5.34*б* показано, как будет квантован сигнал, изображенный на рис. 5.34*а* с использованием адаптивного дельта-модулятора, описанного соотношениями (5.114) и (5.116).

Для удобства параметры системы в этом случае были приняты следующими: $P=2$, $Q=1/2$, $a=1$; минимальный шаг квантования показан на рисунке. Можно отметить, что начальный участок области большой положительной крутизны порождает последовательность нулей, но в этом случае шаг квантования увеличивается экспоненциально, и это позволяет следить за увеличением крутизны входного сигнала. Области дробления в правой части рисунка вновь соответствует чередующаяся последовательность из нулей и единиц, но в этом случае шаг квантования быстро уменьшается до минимального (Δ_{min}) и остается таковым до тех пор, пока крутизна мала. Поскольку минимальный шаг квантования может быть сделан значительно меньше, чем тот, который необходим для оптимальной работы линейного дельта-модулятора, шум дробления может быть существенно уменьшен. Аналогично максимальный шаг квантования можно сделать бо́льшим, чем максимальная крутизна входного сигнала, что приведет к уменьшению шума перегрузки по крутизне.

Параметрами этой системы адаптивной дельта-модуляции являются: P, Q, Δ_{min} и Δ_{max}. Границы шага квантования следует выбирать таким образом, чтобы обеспечить необходимый динамический диапазон входного сигнала. Отношение $\Delta_{max}/\Delta_{min}$ должно быть достаточно большим, чтобы обеспечить большую величину отношения сигнал/шум в требуемом диапазоне уровней входного сигнала. Минимальный шаг квантования должен быть настолько малым, чтобы минимизировать шум незанятого канала. Джаянт [22] показал, что P и Q должны удовлетворять соотношению

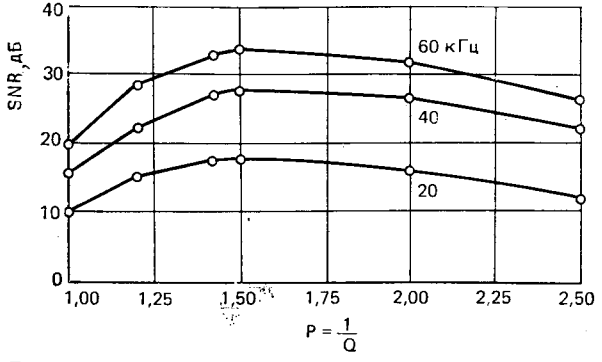


Рис. 5.37. *SNR* для адаптивного дельта-модулятора как функция P [22]

$$PQ \leqslant 1 \qquad (5.117)$$

для устойчивости системы, т. е. для поддержания шага квантования таким, чтобы он соответствовал уровню входного сигнала. На рис. 5.37 представлены результаты моделирования на речевом сигнале с $PQ=1$ для трех различных частот дискретизации. Видно, что значение отношения сигнал/шум достигает максимума при $P=1{,}5$, однако во всех трех случаях отношение сигнал/шум мало меняется при изменении P в пределах

$$1{,}25<P<2. \qquad (5.118)$$

На рис. 5.38 для сравнения систем АДМ, ЛДМ и логарифмической ИКМ приведены зависимости отношения сигнал/шум от скорости передачи для всех трех случаев. Представленные на рисунке результаты для ЛДМ соответствуют случаю $P=1/Q$ при дополнительном условии $P=1=1/Q$. Результаты для АДМ получены при $P=1{,}5$. В случае логарифмической ИКМ зависимость отношения сигнал/шум от скорости передачи вычислена в соответствии с соотношением (5.38) в предположении частоты дискретизации Найквиста ($2F_N=6{,}6$ кГц) при $\mu=100$.

Рисунок 5.38 показывает, что при АДМ отношение сигнал/шум на 8 дБ выше, чем при ЛДМ (скорость передачи 20 кбит/с). Этот выигрыш достигает 14 дБ при скорости 60 кбит/с. С удвоением частоты дискретизации (скорости передачи) отношение сигнал/шум увеличивается на 6 дБ при ЛДМ, и на 10 дБ при АДМ. Сравнивая АДМ и логарифмическую ИКМ, отметим, что при скоростях меньше 40 кбит/с АДМ имеет лучшие характеристики, чем логарифмическая ИКМ. Для больших скоростей логарифмическая ИКМ имеет лучшее отношение сигнал/шум. Например, как следует из рис. 5.38, система с АДМ требует скорости, приблизительно равной 60 кбит/с для достижения того же качества, что и при се-

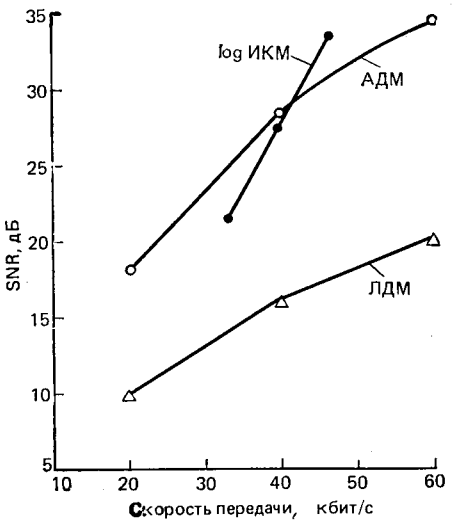


Рис. 5.38. Зависимость *SNR* от скорости передачи для трех схем кодирования с частотой дискретизации 6,6 кГц

миразрядной логарифмической ИКМ со скоростью передачи около 46 кбит/с.

Улучшение качества системы АДМ достигнуто путем ее незначительного усложнения. Поскольку адаптация осуществляется по выходному потоку двоичных символов, система АДМ сохраняет основное преимущество систем с дельта-модуляцией, т. е. не требует синхронизации по кодовым словам. Таким образом, во многих случаях целесообразно использовать АДМ вместо логарифмической ИКМ даже за счет незначительного увеличения скорости передачи.

Другим примером адаптивного квантования с дельта-модуляцией является дельта-модуляция с изменяющейся крутизной (ИКДМ). Эта система (впервые предложенная Грифкесом [23]) не отличается от системы, изображенной на рис. 5.35, но шаг квантования в этом случае изменяется в соответствии с уравнениями

$$\Delta(n)=\begin{cases}\beta\Delta(n-1)+D_2, & c(n)=c(n-1)=c(n-2); \quad (5.119а)\\ \beta\Delta(n-1)+D_1, & \textit{в противном случае,} \quad (5.119б)\end{cases}$$

где $0<\beta<1$ и $D_2\gg D_1>0$. В этом случае минимальный и максимальный шаги квантования определяются рекуррентным соотношением $\Delta(n)$ (см. задачу 5.14).

Суть метода, как и раньше, состоит в том, чтобы увеличить шаг квантования при возникновении последовательности двоичных символов, свидетельствующей о перегрузке по крутизне. В случае возникновения трех последовательных символов единиц или нулей к шагу квантования добавляется приращение D_2. В отсутствие трех последовательных одинаковых символов шаг квантования уменьшается (так как $\beta<1$), пока не достигнет Δ_{min}. Таким образом, шаг квантования увеличивается при перегрузке по крутизне и уменьшается при ее отсутствии. Величины Δ_{min} и Δ_{max} вновь выбираются из условия обеспечения требуемого динамического диапазона и малого шума дробления в условиях незанятого канала. Параметр β определяет скорость адаптации. Если β близко к единице, то скорость нарастания и уменьшения $\Delta(n)$ мала. С другой стороны, если β намного меньше единицы, то адаптация происходит быстрее. Таким образом, адаптация в данном случае может быть как слоговой, так и мгновенной.

Такая система может быть использована в случае, когда необходима малая чувствительность к ошибкам в канале и понижены требования к качеству речевого сигнала по сравнению с коммерческими каналами связи. При этом используется слоговая адаптация. Кроме того, коэффициент предсказания α устанавливается значительно меньшим единицы, в результате чего влияние ошибок в канале существенно ослабляется. За нечувствительность к ошибкам в канале приходится расплачиваться понижением качества восприятия речи при их отсутствии. Основное достоинство системы АДМ в данном случае состоит в том, что она обладает достаточной гибкостью, позволяющей осуществлять обмен между качеством передачи и помехоустойчивостью.

5.6.3. Предсказание высокого порядка в дельта-модуляции

Для простоты в большинстве систем ЛДМ и АДМ используются предсказатели первого порядка вида

$$\tilde{x}(n)=\alpha\hat{x}(n-1), \quad (5.120)$$

как показано на рис. 5.36. В этом случае восстановленный сигнал удовлетворяет разностному уравнению

$$\hat{x}(n)=\alpha\hat{x}(n-1)+\hat{d}(n), \quad (5.121)$$

которое определяется передаточной функцией

$$H_1(z)=(1-\alpha z^{-1})^{-1}. \quad (5.122)$$

Ранее отмечалось, что (5.121) соответствует цифровому интегратору (если $\alpha=1$). Когда $\alpha<1$, такое устройство иногда называют квазиинтегратором.

Результаты рис. 5.32 показывают[1], что в системе с дельта-модуляцией можно получить большее отношение сигнал/шум при использовании предсказателя второго порядка, для которого

$$\tilde{x}(n)=\alpha_1\hat{x}(n-1)+\alpha_2\hat{x}(n-2). \quad (5.123)$$

В этом случае

$$\hat{x}(n)=\alpha_1\hat{x}(n-1)+\alpha_2\hat{x}(n-2)+\hat{d}(n), \quad (5.124)$$

что соответствует передаточной функции

$$H_2(z)=(1-\alpha_1 z^{-1}-\alpha_2 z^{-2})^{-1}. \quad (5.125)$$

В [25] показано, что предсказатель второго порядка дает выигрыш по сравнению с предсказателем первого порядка, когда оба полюса $H_2(z)$ действительны:

$$H_2(z)=\frac{1}{(1-az^{-1})(1-bz^{-1})}, \quad 0<a,\ b<1. \quad (5.126)$$

Такую систему часто называют системой с двойным интегрированием. Увеличение отношения сигнал/шум по сравнению с системой с одним интегратором может достигать 4 дБ в зависимости от диктора и речевого сигнала [25].

К сожалению, использование предсказателей более высокого порядка в системах с АДМ не является таким же простым делом, как замена предсказателя первого порядка предсказателем второго порядка, так как алгоритм адаптивного квантования связан с алгоритмом предсказания. Например, случаю незанятого канала будут соответствовать различные последовательности двоичных символов в зависимости от порядка предсказателя. Для предсказателя второго порядка эта последовательность может быть

[1] Для более точного анализа необходимо знать значения автокорреляционной функции речевого сигнала на задержках, меньших чем 125 мкс (соответствующих высшей частоте дискретизации при дельта-модуляции), для вычисления коэффициента усиления.

010101 ... или 00110011 ... в зависимости от выбора α_1 и α_2 и последнего состояния системы перед тем, как сигнал на входе стал равен нулю. Это требует использовать алгоритм адаптации, основанный более чем на двух последовательных двоичных символах в случае, если шаг квантования достиг своего минимального значения для незанятого канала.

Построение систем с АДМ с предсказателем высокого порядка в настоящее время подробно не исследовано. Вопрос о целесообразности усложнения предсказателя и квантователя зависит от величины выигрыша в качестве передачи, которое может быть при этом получено. Использование многоуровневых квантователей, подобных рассмотренным в § 5.4, до некоторой степени упрощает разработку систем, но предполагает разделение двоичного потока на кодовые слова. Ниже рассматриваются методы разностного квантования с использованием многоуровневых квантователей.

5.7. Разностная ИКМ

Системы, аналогичные изображенной на рис. 5.31, будут далее называться системами с разностной ИКМ (РИКМ). Дельта-модулятор также можно называть одноразрядной системой с ИКМ. В общем случае, однако, термин «разностная ИКМ» применяется по отношению к системам, в которых квантователь имеет более двух уровней квантования.

Как следует из рис. 5.32, системы с РИКМ обеспечивают выигрыш от 4 до 11 дБ по сравнению с прямым квантованием (ИКМ). Наибольший выигрыш достигается при переходе от системы без предсказания к предсказателю первого порядка, несколько меньший — при увеличении порядка от одного до 4—5, после чего выигрыш незначителен. Как указывалось в § 5.5, это увеличение отношения сигнал/шум означает, что системы с РИКМ могут обеспечивать данное отношение сигнал/шум при разрядности, меньшей на единицу, чем это требовалось бы при прямом квантовании речевого сигнала. Следовательно, методы, изложенные в § 5.3 и 5.4, могут быть использованы для оценки качества, которое может быть достигнуто при применении обычного квантователя в разностной схеме. Например, для системы с разностной ИКМ и равномерным неадаптивным квантователем отношение сигнал/шум будет приблизительно на 6 дБ больше, чем для такого же квантователя при прямом квантовании входного сигнала. Разностная ИКМ будет обладать теми же свойствами, что и обычная ИКМ, т. е. отношение сигнал/шум будет увеличиваться на 6 дБ для каждого дополнительного разряда кодового слова, и будет зависеть также от уровня входного сигнала. Аналогично использование квантователя по μ-закону в разностной схеме увеличит отношение сигнал/шум на 6 дБ, и в то же время ее характеристики будут нечувствительны к уровню входного сигнала.

На рис. 5.32 показаны изменения коэффициента усиления в зависимости от диктора и ширины полосы частот сигнала. Значительный разброс возникает при обработке различных фраз речевого сигнала, что является следствием нестационарности речи. Не существует единственного множества коэффициентов предсказателя, которые были бы оптимальными для различного речевого материала и различных дикторов.

Этот разброс совместно с изменениями уровня сигнала, которые характерны для систем связи, приводит к необходимости использования адаптивных предсказателей и адаптивных квантователей для получения наилучших характеристик при различных дикторах и в различных условиях. Такие системы называются системами адаптивной разностной ИКМ (АРИКМ).

5.7.1. АРИКМ с адаптивным квантованием

Результаты анализа адаптивного квантования, изложенные в § 5.4, можно применить и в случае РИКМ. Имеется два основных способа управления адаптивным квантованием.

На рис. 5.39 показано применение квантователей с адаптацией по входу в системе АРИКМ [7]. Здесь шаг квантования пропорционален дисперсии сигнала на его входе. Однако, поскольку раз-

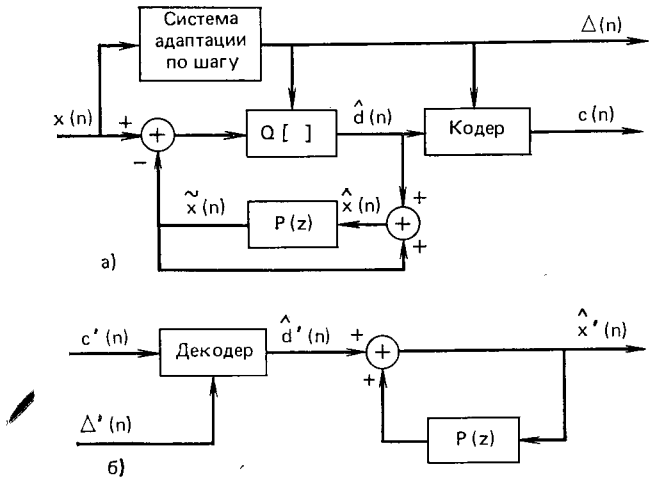


Рис. 5.39. Система АРИКМ с адаптивным по входу квантователем:
а) кодер; *б*) декодер

ностный сигнал $d(n)$ пропорционален входному сигналу, целесообразно управлять шагом квантования или по $d(n)$, или, как это показано на рис. 5.39, по входному сигналу $x(n)$. В 5.4.1 дано несколько алгоритмов управления шагом квантования. Как следует из результатов этого раздела, адаптивное квантование может обеспечить выигрыш около 5 дБ по сравнению со стандартной не-

адаптивной ИКМ μ-законом квантования. Этот выигрыш совместно с дополнительным увеличением отношения сигнал/шум 6 дБ, которое можно получить при применении разностной схемы с неадаптивным квантованием, означает, что АРИКМ с адаптацией по входу позволит получить отношение сигнал/шум на 10—11 дБ больше, чем при использовании неадаптивного квантователя с тем же числом уровней.

На рис. 5.40 показано использование квантователя с адаптацией по выходу в системе АРИКМ [26]. Если, например, адаптация осуществляется в соответствии с уравнениями (5.66)—(5.68),

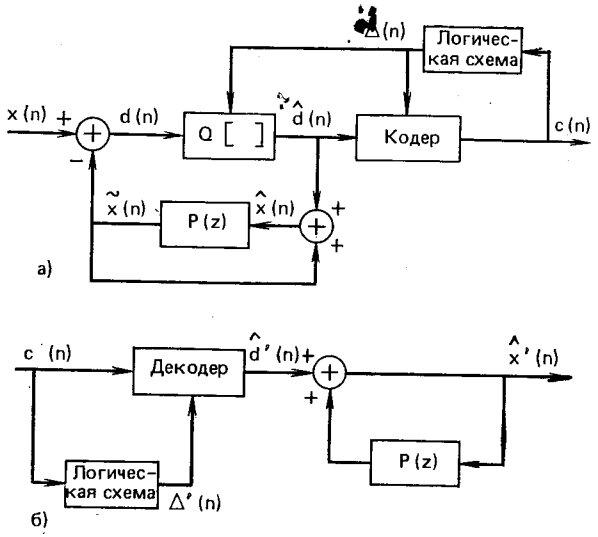


Рис. 5.40. Система АРИКМ с адаптивным по выходу квантователем:
а) кодер; *б*) декодер

можно ожидать выигрыш 4—6 дБ по сравнению с неадаптивным квантованием по μ-закону с тем же числом уровней. Таким образом, адаптация как по выходу, так и по входу позволит достигнуть выигрыша, равного 10—12 дБ по сравнению с неадаптивным квантованием с тем же числом уровней. Кроме того, адаптивный квантователь позволяет расширить динамический диапазон. Дополнительным преимуществом адаптации по выходу является то, что не требуется передавать дополнительную информацию о шаге квантования. Это, однако, делает восстановленный сигнал более чувствительным к ошибкам в канале связи. При адаптации по входу кодовые слова и шаг квантования представляют собой описание сигнала. Хотя это увеличивает сложность представления, однако появляется возможность передачи шага квантования с защитой его от ошибок, что позволяет существенно улучшить качество восстановленного сигнала [27, 28].

5.7.2. АРИКМ с адаптивным предсказанием

Выше рассматривались системы с неадаптивным предсказателем и было выяснено, что даже при использовании предсказателей высокого порядка можно ожидать, что разностное квантование при благоприятных условиях даст выигрыш около 10—12 дБ. Кроме того, величина выигрыша зависит от диктора и речевого материала. Учитывая нестандартность речевого сигнала, естественно рассмотреть адаптивный предсказатель, который, как и адаптивный квантователь, следит за текущими изменениями в речевом сигнале [29]. Система АРИКМ, содержащая адаптивный квантователь и адаптивный предсказатель, изображена на рис. 5.41.

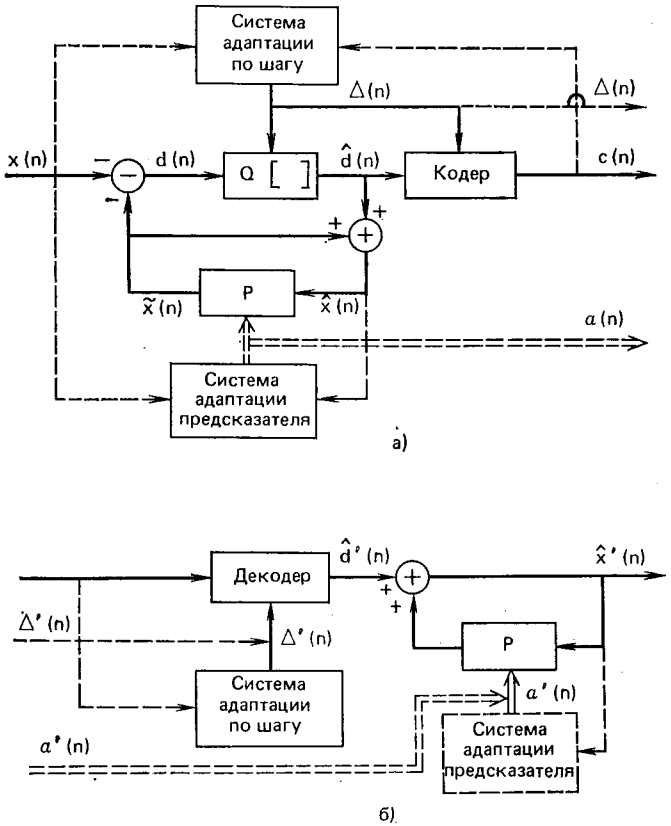


Рис. 5.41. Система АРИКМ с адаптивным квантователем и адаптивным предсказанием:
а) кодер; *б*) декодер

Пунктирные линии показывают, что адаптация предсказателя и квантователя может осуществляться по входному и выходному сигналам. В первом случае к последовательности кодовых слов $c(n)$ для полного описания речевого сигнала необходимо доба-

вить $\Delta(n)$ или коэффициенты предсказания $\alpha(n)=\{\alpha_k(n)\}$, (или и то и другое).

Предполагается, что коэффициенты предсказания зависят от времени так, что предсказанное значение имеет вид

$$\tilde{x}(n)=\sum_{k=1}^{p}\alpha_k(n)\,\hat{x}(n-k). \qquad (5.127)$$

При адаптации коэффициентов предсказателя $\alpha(n)$ обычно предполагается, что свойства речевого сигнала не меняются в течение короткого интервала времени. Коэффициенты предсказания выбираются, следовательно, так, чтобы минимизировать средний квадрат погрешности предсказания на коротком интервале времени. При адаптации по входу предсказатель адаптируется по измерениям входного сигнала (это справедливо, если в соотношениях § 5.5 пренебречь членом $1/SNR$). Применяя те же выкладки, которые использовались при выводе (5.87) и (5.89), и пренебрегая влиянием ошибок квантования, можно показать, что оптимальные коэффициенты предсказателя удовлетворяют уравнениям

$$R_n(j)=\sum_{k=1}^{p}\alpha_k(n)\,R_n(j-k), \quad j=1,\,2,\,\ldots,\,p, \qquad (5.128)$$

где $R_n(j)$ — кратковременная автокорреляционная функция [соотношение (4.24)]

$$R_n(j)=\sum_{m=-\infty}^{\infty}x(m)\,w(n-m)\,x(j+m)\,w(n-m-j), \quad 0\leqslant j\leqslant p \qquad (5.129)$$

и $w(n-m)$ — взвешивающая функция (временно́е окно). Можно использовать прямоугольное окно или другие подходящие окна (например, окно Хемминга длиной N). Поскольку параметры речи меняются относительно медленно, целесообразно подстраивать параметры предсказателя $\alpha(n)$ периодически. Например, новую оценку можно вычислять через каждые 10—20 мс, полагая, что на этих интервалах она остается постоянной. Длительность окна может быть равна этому интервалу или быть несколько больше. В последнем случае соседние сегменты речевого сигнала будут пересекаться. Как следует из (5.129), при вычислении оценок корреляционной функции в (5.128) предполагается, что перед вычислением $R_n(j)$ необходимо записать N отсчетов $x(n)$ в буфер. Множество параметров $\alpha(n)$, удовлетворяющих (5.128), используется в схеме рис. 5.41*а* для того, чтобы квантовать входной сигнал в течение интервала N отсчетов, начиная с n. Таким образом, для восстановления входного сигнала по последовательности кодовых слов необходимо знать коэффициенты предсказания (и возможно шаг квантования), как это показано на рис. 5.41*б*. Особенности вычисления изменяющихся во времени параметров предсказания изучаются в гл. 8.

Для оценки преимуществ адаптивного предсказания Нолл [7] исследовал зависимость коэффициента качества предсказания G_p от порядка предсказателя для фиксированного и адаптивного случаев. На рис. 5.42 приведена зависимость величины[1]

$$10\log_{10}[G_p]=10\log_{10}\left[\frac{E[x^2(n)]}{E[d^2(n)]}\right] \qquad (5.130)$$

от порядка предсказателя p для случаев адаптивного и неадаптивного предсказания. Нижняя кривая, полученная вычислением автокорреляционной функции с усреднением по всей фразе с после-

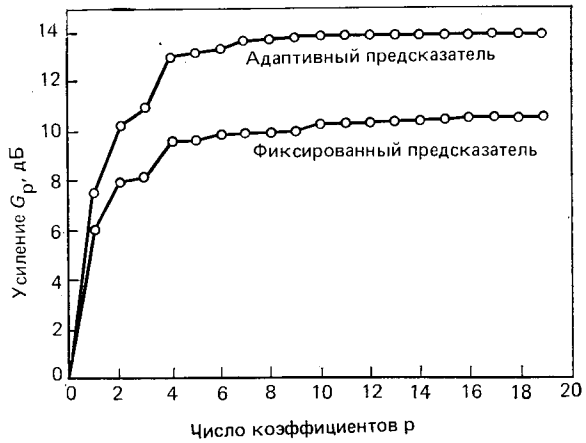


Рис. 5.42. Зависимость показателя качества предсказания от числа коэффициентов предсказания для женского голоса (полоса частот 0—32 000 Гц) [7]

дующим решением системы уравнений (5.89), показывает, что максимальное значение коэффициента качества предсказания составляет около 10,5 дБ. Верхняя кривая получена путем определения такой длины окна L и таких значений коэффициентов предсказания $\alpha(n)$, которые максимизируют G_p на всей входной фразе при данном значении p. Эта максимальная величина изображена для каждого p. Максимальное значение коэффициента качества предсказания при этом составляет около 14 дБ. Таким образом, Нолл [7] предположил, что приемлемая верхняя граница выигрыша АРИКМ систем с неадаптивным и адаптивным предсказателями составляет 10,5 и 14 дБ соответственно. Кривые рис. 5.42, однако, не отражают того обстоятельства, что неадаптивный оптимальный предсказатель весьма чувствителен как к диктору, так и к речевому материалу, в отличие от адаптивного предсказателя, чувствительность которого заметно меньше.

При адаптивном предсказании устраняется избыточность речевого сигнала. Если возможно качественное предсказание, то погрешность предсказания $d(n)$ будет совершенно некоррелированной (белым шумом). Можно сказать, что избыточность устраняет-

[1] На данном рисунке представлены результаты, полученные для одного диктора. Кроме того, ошибки в петле обратной связи не учитывались.

ся в строгом соответствии с моделью речевого сигнала. Как следует из рис. 5.42, коэффициент усиления незначительно увеличивается при увеличении порядка предсказания от 4 до 5. Такое предсказание, однако, не учитывает важного источника избыточности в речевом сигнале, а именно корреляцию вследствие квазипериодического характера вокализованной речи. Один из подходов к учету этой корреляции предложен Аталом и Шредером [29], которые использовали предсказатель вида

$$\tilde{x}(n)=\beta\hat{x}(n-M)+\sum_{k=1}^{p}\alpha_k[\hat{x}(n-k)-\beta\hat{x}(n-k-M)], \qquad (5.131)$$

где параметры предсказания β, M и $\{\alpha_k\}$ адаптируются на интервале, равном N отсчетов. Пренебрегая эффектами погрешности квантования в $\hat{x}(n)$, можем записать ошибку предсказания в виде

$$d(n)=x(n)-\tilde{x}(n)=x(n)-\beta x(n-M)-$$

$$-\sum_{k=1}^{p}\alpha_k[x(n-k)-\beta x(n-k-M)], \qquad (5.132)$$

которая может быть выражена как

$$d(n)=v(n)-\sum_{k=1}^{p}\alpha_k v(n-k). \qquad (5.133)$$

Здесь

$$v(n)=x(n)-\beta x(n-M). \qquad (5.134)$$

Непосредственное вычисление значений β, M и $\{\alpha_k\}$, которые минимизируют $d(n)$, затруднительно. Поэтому Атал и Шредер [29] предложили субоптимальное решение, в котором сначала минимизируется дисперсия $v(n)$, а затем — дисперсия $d(n)$ при заданных β и M. Таким образом, в данном случае автокорреляционная функция вычисляется так же, как и раньше, но для задержек, попадающих в область периода основного тона. Коэффициент предсказания β выбирается совпадающим со значением пика нормированной автокорреляционной функции, а M представляет собой положение этого пика в $R_n(j)$. Таким образом, значение β учитывает изменение амплитуды между последовательными периодами, в то время как M — это период основного тона (в числе отсчетов). По известным M и β вычисляется последовательность $v(n)$ и определяется ее корреляционная функция для $j=0, 1, \ldots, p$, по которой можно определить коэффициенты предсказания из уравнения (5.128), где $R_n(j)$ является кратковременной автокорреляционной функцией последовательности $v(n)$.

Для представления речевого сигнала на основе подобного метода необходимо передавать или хранить квантованный разностный сигнал, шаг квантования (если осуществляется адаптация по входу) и коэффициенты предсказания (квантованные). В работе Атала и Шредера использовался одноразрядный квантователь для разностного сигнала и шаг квантования изменялся каждые 5 мс (33 отсчета на частоте 6,67 кГц) для минимизации ошибки квантования. Кроме того, через каждые 5 мс оценивались и параметры предсказателя. Хотя в работе не приводятся данные по отношению сигнал/шум, предполагается, что можно получить высокое качество восстановленного сигнала на скоростях порядка 10 кбит/с. Используя соответствующее квантование параметров и коэффициентов, Джаянт [8] утверждает, что можно получить выигрыш, равный 20 дБ, по сравнению с ИКМ.

К сожалению, до настоящего времени не проведены исследования потенциальных возможностей адаптивного предсказания, включая параметры основного тона. Однако, несомненно, что методы, подобные приведенным выше являются наиболее сложными среди цифровых методов кодирования речевого колебания. Антиподом АРИКМ может служить линейная дельта-модуляция с ее простым процессом квантования и однородным потоком однозарядных двоичных кодовых слов. Выбор схемы квантования зависит от различных факторов, в том числе от требуемой скорости передачи, качества передачи, сложности кодера и цифрового представления. В следующем параграфе собраны результаты нескольких сравнительных исследований, позволяющие классифицировать различные методы квантования. Но перед этим кратко рассмотрим вопросы управления по выходу в адаптивном предсказании.

Один из подходов основан на вычислении корреляционной функции квантованного сигнала. Таким образом, в (5.128) $R_n(j)$ заменяется на

$$\sum_{m=-\infty}^{\infty}\hat{x}(m)w(n-m)\hat{x}(m+j)w(n-m-j), \quad 0\leqslant j\leqslant p. \qquad (5.135)$$

В этом случае окно должно иметь вид

$$w(m)=\begin{cases}1, & 0\leqslant m\leqslant N-1;\\ 0, & \textit{в противном случае},\end{cases} \qquad (5.136)$$

т. е. оценки коэффициентов предсказания должны основываться на прошлых квантованных отсчетах, а не на будущих, которые не могут быть получены до коэффициентов предсказания. Как и в случае адаптации квантователя, в данном случае по каналу связи передаются только последовательность кодовых слов. Но предсказание с управлением по выходу не получило широкого распространения из-за присущей ему чувствительности к ошибкам и худших характеристик, обусловленных использованием для управления искаженного входного сигнала. Интересный подход к управлению по выходу рассмотрен Строхом [30], исследовавшим градиентные методы для подстройки коэффициентов предсказателя.

5.8. Сравнение систем

При сравнении цифровых систем кодирования речевого колебания в качестве критерия достаточно использовать отношение сигнал/шум. Однако в конечном счете качество систем передачи речевого сигнала необходимо оценивать по слуховому восприятию. Вопрос о том, насколько хорошо звучит кодированный сигнал по сравнению с исходным, является вопросом большой важности. К сожалению, субъективное качество восприятия, как правило, не поддается количественным оценкам и законченные результаты в этом направлении отсутствуют. Поэтому в данном параграфе содержится обзор и сравнение ряда систем кодирования речевого сигнала по отношению сигнал/шум квантования, а также приводится ряд субъективных оценок, позволяющих сопоставить полученные результаты.

Нолл [7] приводит подробное сравнительное исследование схем кодирования речи. Он рассматривал следующие системы:

1) неадаптивную ИКМ с логарифмическим компандированием и с $\mu=100$, $X_{max}=8\sigma_x$ (ИКМ);

2) адаптивную ИКМ (оптимальный гауссовский квантователь) с управлением по входному сигналу (ИКМ-АК$_{вх}$);

3) разностную ИКМ с предсказателем первого порядка (неадаптивным) и адаптивным «гауссовским квантователем» с управлением по выходному сигналу (РИКМ-АК$_{вых}$);

4) адаптивную разностную ИКМ с адаптивным предсказателем первого порядка и адаптивным гауссовским квантователем, управляемыми по входному сигналу (АРИКМ1-АК$_{вх}$) (протяженность окна 32);

5) адаптивную разностную ИКМ с предсказателем четвертого порядка и адаптивным «квантователем Лапласа», адаптирующимися по входному сигналу (АРИКМ4-АК$_{вых}$) (протяженность окна 128);

6) адаптивную разностную ИКМ с адаптивным предсказателем 12-го порядка и адаптивным «гамма-квантователем» с адаптацией по входному сигналу (АРИКМ12-АК$_{вх}$) (протяженность окна 256).

Во всех системах частота дискретизации составляла 8 кГц, а длина кодового слова изменялась от 2 до 5 бит на отсчет, т. е. скорость изменялась от 16 до 40 кбит/с. Отношение сигнал/(шум квантования) для всех систем приведено на рис. 5.43. Представленные кривые позволяют отметить ряд интересных особенностей. Во-первых, нижняя кривая соответствует квантователю с кодовыми словами длиной 2 бит. По мере увеличения длины кодового слова кривые поднимаются вверх и отстоят одна от другой на 6 дБ. Заметим также, что добавление как неадаптивного предсказателя, так и адаптивного квантователя позволяет получить выигрыш в отношении сигнал/шум, но адаптация простого предсказателя практически не приводит к выигрышу. Однако видно, что

адаптивные предсказатели более высокого порядка приводят к существенному выигрышу в отношении сигнал/шум.

При телефонной передаче обычно предполагают, что приемлемое качество речи обеспечивается квантователем по μ-закону при 6--7 бит на отсчет. Из (5.38) видно, что семиразрядный квантователь при $\mu=100$ дает отношение сигнал/шум около 33 дБ. На основе рис. 5.43 можно сделать вывод, что такое же качество обеспечивает пятиразрядный адаптивный квантователь и адаптивный предсказатель. На практике субъективное качество речевого сигнала в системах с АРИКМ оказывается лучше, чем с ИКМ при том же отношении сигнал/шум. Исследуя системы с АРИКМ и адаптацией квантователя по выходу и неадаптивным предсказателем, в [26] установлено, что слушатели отдают предпочтение АРИКМ сигналу перед сигналом логарифмической ИКМ с большим отношением сигнал/шум. Результаты теста представлены в табл. 5.7, где ИКМ — это система 1 в исследовании Нолла, а АРИКМ — это система 3 (см. выше). Из представленных результатов следует, что четырехразрядная АРИКМ оказывается лучше шестиразрядной ИКМ. Этот результат не будет казаться удивительным, если вспомнить, что выигрыш в отношении сигнал/шум при использовании АРИКМ с неадаптивным предсказателем и адаптивным квантователем составляет 10—12 дБ или, грубо говоря, двухразрядной, но фактически четырехразрядной АРИКМ отдано предпочтение перед шестиразрядной ИКМ, несмотря на несколько меньшее отношение сигнал/шум.

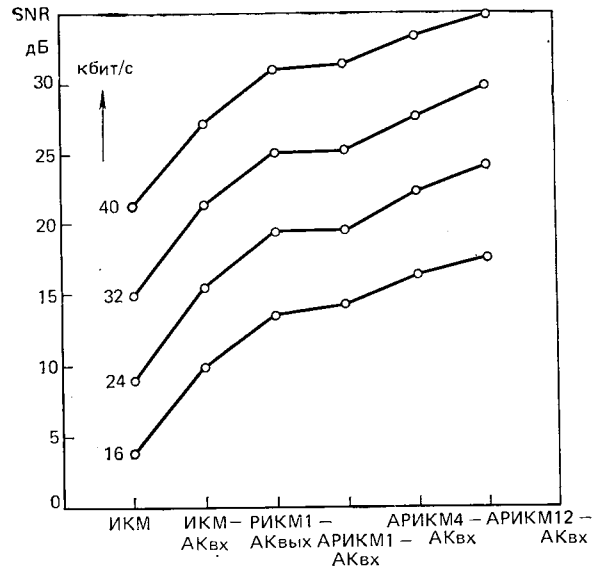


Рис. 5.43. Отношение сигнал/шум для квантования от 2 до 5 бит на отсчет (16—40 кбит/с): АК$_{вх}$ — адаптивный квантователь с адаптацией по входу; АК$_{вых}$ — адаптивный квантователь с адаптацией по выходу; АРИКМ$_r$ — система АРИКМ с порядком предсказателя r [7]

Исследуя адаптивное предсказание, Атал и Шредер [29] установили, что их система АРИКМ с двухуровневым адаптивным

Таблица 5.7

Сравнение объективного и субъективного качества системы АРИКМ и лог-ИКМ

Объективная оценка (по SNR)	Субъективная оценка (по предпочтению)
7 бит ИКМ	7 бит ИКМ (высокое качество)
6 бит ИКМ	4 бит АРИКМ
4 бит АРИКМ	6 бит ИКМ
5 бит ИКМ	3 бит АРИКМ
3 бит АРИКМ	5 бит ИКМ
4 бит ИКМ	4 бит ИКМ (низкое качество)

квантователем и сложным адаптивным предсказателем позволяет получить сигнал, не намного худший, чем при шестиразрядной логарифмической адаптивной ИКМ. Оцененная скорость передачи в этой системе составляла 10 кбит/с, а для ИКМ в тех же условиях, т. е. при частоте дискретизации 6,67 кГц, требуется скорость 40 кбит/с. В этом случае особенно явно сказывается различие между субъективным качеством и тем, которое можно ожидать из анализа величины отношения сигнал/шум.

Дать точное объяснение этому эффекту затруднительно, но можно предположить, что такое различие возникает вследствие влияния двух факторов: лучших характеристик незанятого канала для адаптивного квантователя и большей корреляции между шумом квантования и сигналом [7].

5.9. Преобразования способов кодирования

Из материала данной главы достаточно ясно, что существует множество возможностей для квантования речевого сигнала. Эти методы различаются по своей сложности — от чрезвычайно простой в технической реализации линейной дельта-модуляции, требующей больших скоростей передачи, до разнообразных методов адаптивной разностной ИКМ, обеспечивающих хорошее качество на малых скоростях передачи, но при большей сложности алгоритмов обработки. В результате значительный интерес представляет разработка методов прямого преобразования одного цифрового представления в другое, минуя аналоговое представление сигнала. Эта задача важна по следующим причинам.

1. В больших системах связи весьма вероятно возникновение ситуаций, в которых на районных сетях более важно иметь дешевое оконечное оборудование, а не низкую скорость передачи информации. В других случаях, например при междугородной связи или при хранении речевого сигнала в цифровом блоке памяти, важнее получить сокращенное описание сигнала. Для объединения различных частей системы связи с различными скоростями чрезвычайно важно располагать возможностью прямого преобразования цифровых представлений, используемых в каждой из

подсистем, что позволит избежать дополнительных потерь качества.

2. Реализация низкоскоростных представлений сигнала существенно упрощается при применении простых методов цифрового преобразования. Например, может оказаться полезным использовать линейную дельта-модуляцию, скажем, для аналого-цифрового преобразования, чтобы далее представить полученный сигнал в более компактной форме, такой, как ИКМ или АРИКМ.

3. При обработке речевого сигнала требуется представить его в цифровой форме с помощью линейной ИКМ для того, чтобы далее, например при цифровой фильтрации, можно было использовать отдельные отсчеты сигнала.

5.9.1. Преобразование ЛДМ в ИКМ

Для получения высококачественного цифрового представления с помощью линейной дельта-модуляции требуется очень большая частота дискретизации и простой двухуровневый квантователь. Такие системы могут быть просто реализованы на основе совместного использования аналоговых и цифровых компонент. Действительно, весь кодер ЛДМ можно легко реализовать как простую интегрирующую цепь [31]. На рис. 5.44 показаны обычный анало-

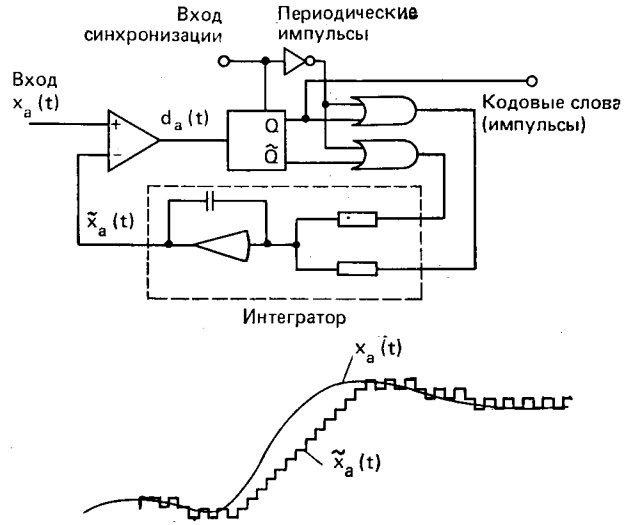


Рис. 5.44. Схема линейного дельта-модулятора [31]

говый компаратор, формирующий разностный сигнал, триггер, определяющий знак этого сигнала, и интегратор, восстанавливающий предсказанное значение для сравнения его с входным сигналом. Это простое объединение аналоговых и цифровых цепей — все, что необходимо для реализации линейного дельта-модулято-

ра. Данное устройство реализовано в виде интегральной схемы, работающей на частотах до 17 МГц [31].

Простота схемы совместно с возможностью работы на весьма высоких скоростях делает это устройство весьма удобным для использования при аналого-цифровых преобразованиях. Простота достигается, конечно, за счет чрезвычайно высокой скорости, которая необходима для получения хорошего качества передачи.

Скорость передачи, однако, можно снизить за счет использования преобразования линейной дельта-модуляции в более эффективные виды цифрового представления, такие, как ИКМ или АРИКМ. Наиболее важным является преобразование ЛДМ в линейную ИКМ, поскольку ИКМ представление требуется при любой цифровой обработке отсчетов аналогового сигнала.

Процесс преобразования ЛДМ в ИКМ включает в себя, во-первых, получение ИКМ представления и, во-вторых, снижение частоты дискретизации до частоты Найквиста. Первый шаг заключается в декодировании последовательности нулей и единиц в последовательность величин $\pm\Delta$, а затем в цифровом интегрировании положительных и отрицательных приращений для получения квантованных отсчетов $x_a(t)$ с частотой дискретизации, соответствующей ЛДМ. Полученная последовательность содержит шумы квантования в полосе $|\Omega| \leqslant \pi/T$, где T — период дискретизации в ЛДМ, хотя спектр входного сигнала ограничен по полосе значительно более низкой частотой. Таким образом, перед уменьшением частоты дискретизации необходимо устранить шум квантования в полосе от максимальной частоты спектра сигнала до половины частоты дискретизации ЛДМ. Как обсуждалось в 2.4.2, это можно сделать весьма эффективно при использовании КИХ-фильтра нижних частот с частотой среза, равной частоте Найквиста обрабатываемого сигнала [32]. Выходной сигнал фильтра вычисляется для каждых M отсчетов, где M — отношение частоты ЛДМ дискретизации к частоте ИКМ дискретизации. Таким образом, ЛДМ-ИКМ-преобразователь содержит интегратор, выход которого фильтру-

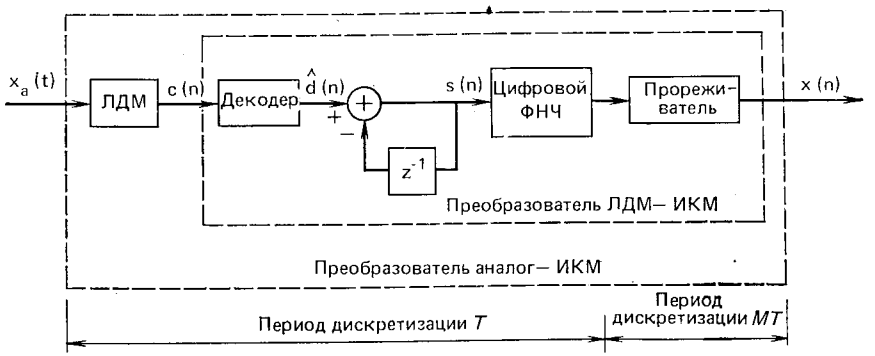


Рис. 5.45. Преобразователь аналог — ИКМ в системе ЛДМ и преобразователь ЛДМ—ИКМ

ется и подвергается дискретизации. Как следует из рис. 5.45, преобразователь ЛДМ-ИКМ совместно с кодером ЛДМ представляет собой устройство аналого-цифрового преобразования, почти полностью выполненное в цифровой форме.

5.9.2. Преобразование ИКМ—АРИКМ

Другой пример преобразования — это преобразование линейной ИКМ в АРИКМ [33]. Интерес к этому виду преобразования определяется желанием получить более эффективное представление сигнала, чем то, которое дает линейная ИКМ. На рис. 5.46 показана структурная схема данного преобразования. Очевидно,

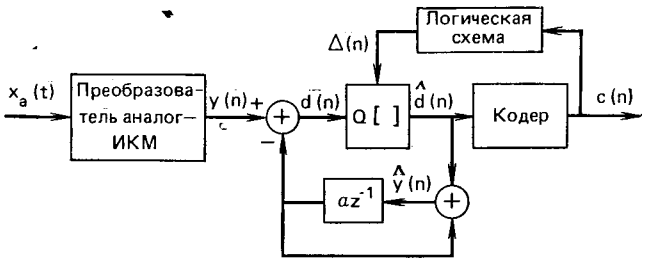


Рис. 5.46. Преобразователь ИКМ—АРИКМ

что основной принцип состоит в реализации обычного алгоритма АРИКМ, применяемого к последовательности на выходе кодера линейной ИКМ. Это достигается путем реализации операций, изображенных на рис. 5.46. Важный вопрос, возникающий при любых цифровых преобразованиях, состоит в ухудшении качества, обусловленном последующей цифровой обработкой. Очевидно, что входной ИКМ сигнал будет содержать погрешность, которую можно охарактеризовать отношением сигнал/шум SNR_1. Алгоритм АРИКМ внесет дополнительные искажения. Если предположить, что эти искажения не зависят от обрабатываемого сигнала и ошибок, вносимых ИКМ, то можно считать, что общие искажения приближенно равны (см. задачу 5.17)

$$SNR = SNR_1/(1 + SNR_1/SNR_2), \qquad (5.137)$$

где SNR_2 — отношение сигнал/шум системы с АРИКМ. Из данного соотношения видно, что полное SNR не может быть больше, чем SNR_1. Однако если SNR_2 порядка SNR_1, то потери малы. Действительно, если $SNR_1 = SNR_2$, они составляют около 3 дБ.

Кодопреобразование можно рассматривать как точный и эффективный метод применения обычных алгоритмов кодирования речи. Так изначально получают линейную ИКМ, например так, как это описано в 5.9.1, а затем осуществляют цифровую обработку с использованием различных методов кодирования сигнала.

5.10. Заключение

В данной главе подробно изложены методы цифрового представления речевого колебания. Показано, что существуют несколько подходов к решению этой задачи. Главное внимание уделено основным принципам, а не описанию всех систем такого типа, предложенных к настоящему времени. Дальнейшее изучение этой темы можно провести по обзорной статье Джаянта [8] и по сборнику статей [34] под его редакцией.

Задачи

5.1. Функция плотности вероятности равномерного закона распределения

$$p(x)=\begin{cases}1/\Delta, & |x|<\Delta/2;\\ 0, & \textit{в противном случае.}\end{cases}$$

Определить среднее значение и дисперсию равномерного распределения.

5.2. Рассмотреть функцию плотности вероятности распределения Лапласа $p(x)=\frac{1}{\sqrt{2}\sigma_x}e^{-\sqrt{2}|x|/\sigma_x}$ и определить вероятность того, что $|x|>4\sigma_x$.

5.3. Пусть $x(n)$ — сигнал на входе линейной системы, инвариантной к сдвигу, — представляет собой стационарный белый шум с нулевым средним значением и единичной дисперсией. Показать, что автокорреляционная функция процесса на выходе имеет вид $\varphi(m)=\sigma^2_x\sum\limits_{k=-\infty}^{\infty}h(k)h(k+m)$, где σ^2_x — дисперсия входного сигнала и $h(n)$ — импульсная характеристика линейной системы.

5.4. Рассмотреть вопросы разработки высококачественной цифровой акустической системы. Отношение сигнал/шум, равное 60 дБ, необходимо обеспечить для пиковых уровней сигнала в диапазоне от 1 до 100. Полезная полоса частот сигнала должна быть не менее 8 кГц.

Требуется:

а) Изобразить основные узлы аналого-цифрового и цифро-аналогового преобразователей.

б) Определить количество разрядов при аналого-цифровом преобразовании.

в) Что является главным условием при выборе частоты дискретизации? Какого типа фильтры необходимо использовать перед аналого-цифровым и после цифро-аналогового преобразований? Оценить наименьшую частоту дискретизации, которая возможна в практическом случае.

г) Как изменяются предъявляемые к системе требования и выбранные параметры, если необходимо обеспечить лишь «телефонное» качество речевого сигнала?

5.5. Речевой сигнал ограничен по полосе идеальным фильтром нижних частот, дискретизирован с частотой Найквиста, квантован в B-разрядном квантователе и затем преобразован обратно в аналоговую форму в идеальном цифро-аналоговом преобразователе (рис. 3.5.1.*а*). Определим $y(n)=x(n)+e_1(n)$, где $e_1(n)$ — погрешность квантования. Предположим, что шаг квантования $\Delta=8\sigma_x/2^B$ и B достаточно велики, чтобы выполнялись следующие предположения: $e_1(n)$ — стационарная последовательность; $e_1(n)$ — не коррелирована с $x(n)$; $e_1(n)$ — равномерно распределенная последовательность белого шума. Как было показано, отношение сигнал/(шум квантования) в этих условиях равно $SNR_1=\sigma^2_x/\sigma^2_e=(12/64)\cdot 2^{2B}$. Предположим, что аналоговый сигнал дискретизирован с частотой Найквиста и квантован в B-разрядном квантователе (рис. 3.5.1*б*) (предположим также, что $0<\varepsilon<T$, т. е. устройства работают не абсолютно синхронно во времени). При этом $w(n)=y'(n)+e_2(n)$, где $e_2(n)$ обладает теми же свойствами, что и $e_1(n)$.

а) Показать, что полное отношение сигнал/шум равно $SNR_2=SNR_1/2$.

б) Обобщить результат п.а) для N-кратного аналого-цифрового преобразования и обратно.

5.6. Хотя обычно предполагают, что погрешность квантования не зависит от сигнала, это предположение нарушается при малом числе уровней квантования.

а) Показать, что $e(n)=x(n)-\hat{x}(n)$ не является статистически независимым от $x(n)$ ($\hat{x}(n)$ — квантованный сигнал). Указание: представить $\hat{x}(n)$ как $\hat{x}(n)=[x(n)/\Delta]\Delta+(\Delta/2)$, где $[\cdot]$ означает целую часть, т. е. наибольшее целое число,

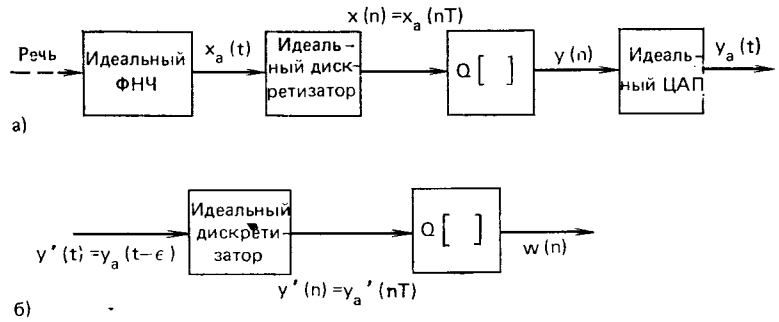


Рис. 3.5.1

меньшее или равное величине, указанной в скобках. Процесс представить в виде $x(n)=[x(n)/\Delta]\Delta+x_f(n)=x_i(n)+x_f(n)$, где $x_i(n)$ — целая часть $x(n)$, а $x_f(n)$ остаток от $x(n)$. Таким образом, $e(n)$ можно представить как функцию от $x(n)$. Показать, что они не могут быть статистически независимыми.

б) При каких условиях справедливо приближение, что $x(n)$ и $e(n)$ независимы?

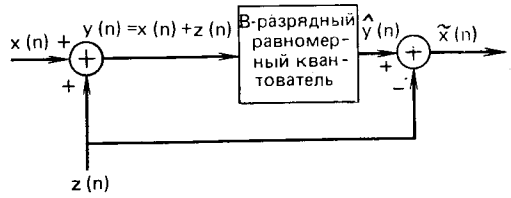


Рис. 3.5.2

в) На рис. 3.5.2 представлен метод, позволяющий сделать $e(n)$ и $x(n)$ статистически независимыми даже при малом числе уровней квантования. В этом случае $z(n)$ представляет собой псевдослучайную реализацию типа белого шума с равномерным распределением и функцией плотности вероятности, равной $p(z)=1/\Delta$, $-\Delta/2\leqslant z\leqslant\Delta/2$. Показать, что при этом погрешность квантования статистически не зависит от сигнала при любых значениях B (последовательность шума называют возмущающей). Указание: рассмотреть последовательность $e(n)$ для $y(n)$.

г) Показать что дисперсия ошибки при возмущении возрастает и $\sigma^2_{e_1}>\sigma^2_e$, где $e_1(n)=x(n)-\hat{y}(n)$ и $e(n)=x(n)-\hat{x}(n)$.

д) Показать, что путем простого вычитания возмущающего шума из выходной последовательности квантователя можно получить такую же ошибку квантования $e_2(n)=x(n)-(\hat{y}(n)-z(n))$, как и при отсутствии возмущений $\sigma^2_{e_2}=\sigma^2_e$.

5.7. Обычно дисперсию сигнала оценивают, полагая, что она пропорциональна кратковременной энергии, равной $\sigma^2(n)=\sum\limits_{m=-\infty}^{\infty}x^2(m)h(n-m)$.

а) Показать, что если $x(n)$ — стационарный случайный процесс с нулевым средним, то $E[\sigma^2(n)]$ пропорциональна σ^2_x.

б) Для

$$h(n)=\begin{cases}\alpha^n, & n\geqslant 0,\quad (|\alpha|<1);\\ 0, & n<0\end{cases}$$

и для

$$E[x^2(m)\,x^2(l)] = \begin{cases} B, & m = l; \\ 0, & m \neq l \end{cases}$$

определить дисперсию $\sigma^2(n)$ как функцию B и α.

в) Объяснить поведение дисперсии в п.б), если α изменяется от 0 до 1.

5.8. Рассмотреть адаптивный квантователь, показанный на рис. 3.5.3*а*. Характеристика двухразрядного квантователя и соответствие кодовых слов представлена на рис. 3.5.3*б*. Предположим, что шаг квантования адаптируется в соответствии с правилом $\Delta(n) = M\Delta(n-1)$, где M зависит от предшествующего кодового слова $c(n-1)$ и $\Delta_{min} \leqslant \Delta(n) \leqslant \Delta_{max}$. Далее предположим, что

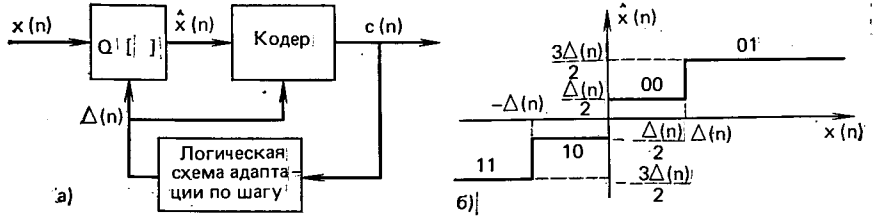


Рис. 3.5.3

$$M = \begin{cases} P, & c(n-1) = 01 \text{ или } 11; \\ 1/P, & c(n-1) = 00 \text{ или } 10. \end{cases}$$

а) Изобразить структурную схему блока адаптации шага квантования.

б) Пусть

$$x(n) = \begin{cases} 0, & n < 5; \\ 20, & 5 \leqslant n \leqslant 13; \\ 0, & 13 < n. \end{cases}$$

Предположим, что $\Delta_{min} = 2$ и $\Delta_{max} = 30$, и $P = 2$. Рассчитать таблицу значений $x(n)$, $\Delta(n)$, $c(n)$ и $\hat{x}(n)$ для $0 \leqslant n \leqslant 25$ (пусть $n = 0$, $\Delta(n) = \Delta_{min} = 2$ и $c(n) = 00$).

г) Построить в одном масштабе последовательности $x(n)$ и $\hat{x}(n)$.

5.9. Рассмотрим систему двухразрядного адаптивного квантования из задачи 5.8. Алгоритм адаптации шага квантования зададим выражением

$$\Delta(n) = \begin{cases} \beta\Delta(n-1) + D, & \text{если } \sum_{k=1}^{M} LSB[c(n-k)] \geqslant 2, \\ \beta\Delta(n-1) - \textit{в противном случае}, \end{cases}$$

где $LSB[c(n-k)]$ означает последний значащий разряд в кодовом слове $c(n-k)$.

а) Изобразить структурную схему устройства адаптации шага.

б) Определить максимально возможный шаг квантования и выразить его через β и Δ. (Указание: рассмотреть отклик на функцию скачка первого уравнения задачи.)

в) Предположим, что

$$x(n) = \begin{cases} 0, & n < 5; \\ 20, & 5 \leqslant n \leqslant 13; \\ 0, & 13 < n. \end{cases}$$

Предположим также, что $M = 1$, $\beta = 0{,}8$ и $D = 6$. Рассчитать таблицу значений

$x(n)$, $\Delta(n)$, $c(n)$ и $\hat{x}(n)$ для $0 \leqslant n \leqslant 25$ (при этом пусть $n = 0$, $\Delta(n) = 0$ и $c(n) = 00$). Изобразить $x(n)$ и $\hat{x}(n)$ в одной координатной системе.

г) Определить значение β, при котором постоянная времени шага адаптации составляет 10 мс.

5.10. Рассмотрим предсказатель первого порядка $\tilde{x}(n) = \alpha x(n-1)$, где $x(n)$ — стационарный случайный процесс с нулевым средним.

а) Показать, что погрешность предсказания имеет дисперсию $\sigma^2_d = \sigma^2_x(1 + \alpha^2 - 2\alpha\varphi(1)/\sigma^2_x)$.

б) Показать, что σ^2_d минимальная при $\alpha = \varphi(1)/\sigma^2_x = \rho(1)$.

в) Показать, что минимальная дисперсия погрешности равна $\sigma^2_d = \sigma^2_x(1 - \rho^2(1))$.

г) При каких условиях справедливо соотношение $\sigma^2_d < \sigma^2_x$?

5.11. Дана последовательность $x(n)$ с автокорреляционной функцией вида $\varphi(m)$. Показать, что разностный сигнал $d(n) = x(n) - x(n-n_0)$ имеет тем меньшую дисперсию, по сравнению с исходным, чем больше корреляция между $x(n)$ и $x(n-n_0)$. (Предположить, что среднее значение $x(n)$ равно нулю.)

а) Определить условие на $\varphi(n_0)$, при котором $\sigma^2_d \leqslant \sigma^2_x$.

б) Пусть $d(n)$ формируется в виде $d(n) = x(n) - \alpha x(n-n_0)$, где $\alpha = \varphi(n_0)/\varphi(0)$.

Установите ограничения на $\varphi(n_0)$, при которых справедливо неравенство $\sigma^2_d \leqslant \sigma^2_x$.

5.12. Используя соотношения (5.78) и (5.83), доказать, что для оптимальных коэффициентов предсказания имеет место равенство $E[(x(n) - \tilde{x}(n))\tilde{x}(n)] = E[d(n)\tilde{x}(n)] = 0$, т. е. оптимальная погрешность предсказания не коррелирована с сигналом.

5.13. Рассмотрим разностный сигнал $d(n) = x(n) - \alpha_1\hat{x}(n-1)$, где $\hat{x}(n)$ — квантованный сигнал в разностном кодере.

а) Показать, что $\sigma^2_d = \sigma^2_x[1 - \alpha_1\rho(1) + \alpha^2_1] + \alpha^2_1\sigma^2_e$.

б) Используя результат а), показать, что $G_p = \dfrac{\sigma^2_x}{\sigma^2_d} = \dfrac{1 - \alpha_1/(SNR_Q)}{1 - 2\alpha_1\rho(1) + \alpha^2_1}$, где $SNR_Q = \sigma^2_d/\sigma^2_e$.

5.14. Для системы с дельта-модуляцией с переменной крутизной алгоритм адаптации шага квантования имеет вид

$$\Delta(n) = \begin{cases} \beta\Delta(n-1) + D_2, & c(n) = c(n-1) = c(n-2); \\ \beta\Delta(n-1) + D_1, \end{cases}$$

где $0 < \beta < 1$ и $0 < D_1 \ll D_2$.

а) Максимальный шаг квантования возникает при поступлении на вход фильтра сигнала в течение длительного времени D_2, что соответствует длительному периоду перегрузки по крутизне. Определить Δ_{max} через D_2 и β.

б) Минимальный шаг достигается, если в течение длительного времени не возникает последовательности: $c(n) = c(n-1) = c(n-2)$.

Определить Δ_{min} через D_1 и β.

5.15. Рассмотрим адаптивный дельта-модулятор (рис. 3.5.4*а*). Двухуровневый квантователь представлен на рис. 3.5.4*б*. Шаг квантования адаптируется по правилу $\Delta(n) = M\Delta(n-1)$, где $\Delta_{min} \leqslant \Delta(n) \leqslant \Delta_{max}$ и множители шага равны

$$M = \begin{cases} P, & \text{если } c(n) = c(n-1); \\ 1/P, & \text{если } c(n) \neq c(n-1). \end{cases}$$

а) Изобразить структурную схему алгоритма адаптации.

б) Пусть

$$x(n) = \begin{cases} 0, & n < 5; \\ 20, & 5 \leqslant n \leqslant 13; \\ 0, & 13 < n. \end{cases}$$

Предположим, что $\Delta_{min} = 1$, $\Delta_{max} = 15$, $\alpha = 1$, и $P = 2$. Рассчитать таблицу значений $x(n)$, $\tilde{x}(n)$, $d(n)$, $\Delta(n)$, $\tilde{a}(n)$ и $\hat{x}(n)$ для $0 \leqslant n \leqslant 25$.

Предположим, что при $n=0$, $x(0)=0$, $\tilde{x}(0)=1$, $d(0)=1$, $\Delta(0)=\Delta_{min}=1$ и $\hat{d}(0)=1$. Изобразить $x(n)$ и $\hat{x}(n)$ для $0\leqslant n\leqslant 25$.

5.16. Рассмотрим два кодера, представленных на рис. 3.5.5*а* и *б*. В каждом кодере используется двухразрядный квантователь с характеристикой рис. 3.5.5*в*.

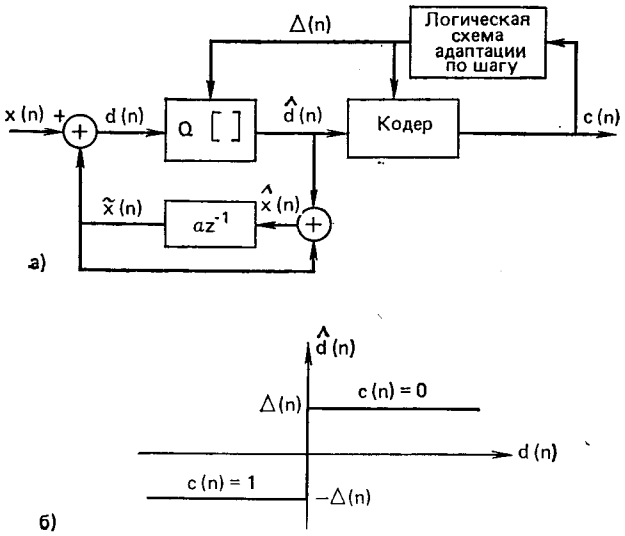


Рис. 3.5.4

Рассмотрим случай свободного канала, т. е. случай, когда есть шум малого уровня. Для простоты положим $x(n)=0{,}1\cos(\pi n/4)$.

а) Для $0\leqslant n\leqslant 20$ построить $\hat{x}(n)$ для обоих кодеров.

б) Для какого кодера шум незанятого канала будет более неприятным и почему?

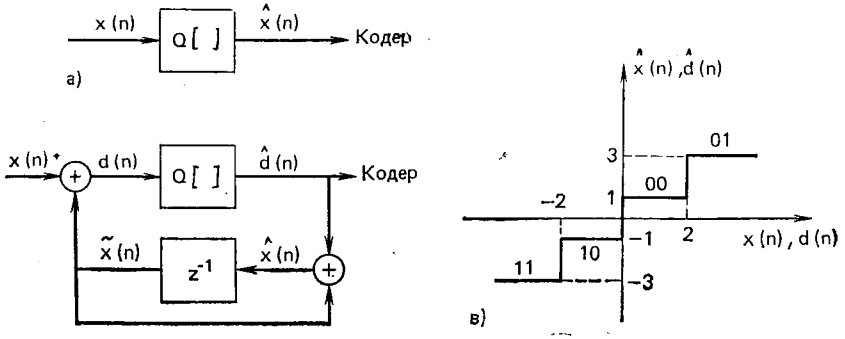


Рис. 3.5.5

5.17. Рассмотрим преобразование ИКМ-АРИКМ (см. рис. 5.46). Сигнал ИКМ $y(n)$ можно представить в виде $y(n)=x(n)+e_1(n)$, где $x(n)=x_a(nT)$ и $e_1(n)$ — погрешность квантования при ИКМ представлении. Квантованный АРИКМ сигнал можно представить в виде $\hat{y}(n)=y(n)+e_2(n)$, где $e_2(n)$ — погрешность квантования АРИКМ.

а) Полагая погрешности квантования некоррелированными, показать, что полное отношение сигнал/шум $SNR=\sigma^2_x/(\sigma^2_{e_1}+\sigma_{e_2})$.

б) Показать, что отношение сигнал/шум можно записать в форме $SNR=SNR_1/[1+(1+SNR_1)/SNR_2]$, где $SNR_1=\sigma^2_x/\sigma^2_{e_1}$ и $SNR_2=\sigma^2_y/\sigma^2_{e_2}$.

6
Кратковременный анализ Фурье

6.0. Введение

Представление сигналов или других функций с помощью сумм синусоид и комплексных экспонент часто приводит к удобным решениям задач науки и техники и помогает понять физику явления глубже, чем это возможно другими методами. Такое представление, часто называемое Фурье-представлением, полезно при обработке сигналов по двум причинам. Во-первых, в линейных системах легко определить отклик на суперпозицию синусоид или комплексных экспонент. Во-вторых, Фурье-представление часто выявляет такие свойства сигнала, которые в первоначальном виде скрыты или по крайней мере не очевидны.

Исследования и техника передачи речи представляют собой области, в которых по традиции преобразование Фурье играло ведущую роль. Чтобы понять, почему это так, полезно вспомнить, что модель образования стационарного речевого сигнала состоит просто из линейной системы, возбуждаемой либо периодически, либо случайно. Спектр на выходе такой модели равен произведению частотной характеристики голосового тракта и спектра возбуждения. Следует ожидать поэтому, что спектр на выходе должен отражать свойства как возбуждения, так и частотной характеристики голосового тракта. Однако речевой сигнал гораздо сложнее, чем просто продолжительный гласный или фрикативный звук. Поэтому стандартное Фурье-представление, вполне пригодное для периодических, импульсных или стационарных случайных сигналов, неприменимо к речевому сигналу, характеристики которого значительно меняются во времени. Однако мы уже имели возможность убедиться в том, что принцип кратковременного анализа оказывается полезным при обработке речи. Например, такие изменяющиеся во времени характеристики, как энергия, переходы через нуль и корреляция, можно считать постоянными на интервалах времени около 10—30 мс. В этой главе будет показано, что аналогичным образом и спектральные характеристики речи можно считать изменяющимися во времени сравнительно медленно.

Для изучения спектральных характеристик речевого сигнала удобно ввести формализм, связанный с понятием кратковременного преобразования Фурье. Введем кратковременное преобразование Фурье и процедуру синтеза, основанную на нем. При этом удобно рассматривать анализ Фурье, как преобразование сигнала в гребенке фильтров. Такой подход позволяет лучше понять как теоретические, так и практические (вычислительные) аспекты кратковременного анализа. Будут рассмотрены также и другие вычислительные методы, основанные на быстром алгоритме вычисления дискретного преобразования Фурье (алгоритмы БПФ). И, наконец, подробно изучив теорию и способ вычисления кратковременного преобразования Фурье, рассмотрим его применение в задачах анализа — синтеза речи (вокодеры), визуального отображения спектра и в таких важных задачах обработки речи, как формантный анализ и выделение основного тона.

---

[1] Перевод данной главы выполнен при участии канд. физ.-мат. наук А. Ю. Шевердяева. (*Прим. ред.*)

6.1. Определения и свойства

Потребность в спектральном представлении, отображающем меняющиеся во времени свойства речевых сигналов, побуждает нас ввести представление Фурье, зависящее от времени. Подходящим определением зависящего от времени преобразования Фурье будет следующее:

$$X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} w(n-m)\,x(m)\,e^{-i\omega m}. \tag{6.1}$$

В (6.1) $w(n-m)$ представляет собой действительную последовательность временно́го окна. Этой последовательностью выделяется часть входного сигнала в момент времени n. Ясно, что зависящее от времени преобразование Фурье представляет собой функцию двух переменных: времени n, которое предполагается дискретным, и частоты ω, предполагаемой здесь непрерывной. Другая форма для (6.1) получается при замене переменной суммирования, что дает

$$X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} w(m)\,x(n-m)\,e^{-i\omega(n-m)} =$$

$$= e^{-i\omega n} \sum_{m=-\infty}^{\infty} x(n-m)\,w(m)\,e^{i\omega m}. \tag{6.2}$$

Определив

$$\tilde{X}_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(n-m)\,w(m)\,(e^{i\omega m}), \tag{6.3}$$

получим для $X_n(e^{i\omega})$ выражение

$$X_n(e^{i\omega}) = e^{-i\omega n}\,\tilde{X}_n(e^{i\omega}). \tag{6.4}$$

Приведенные соотношения допускают две различные интерпретации. Во-первых, зафиксировав n, видим, что $X_n(e^{i\omega})$ представляет собой обычное преобразование Фурье для последовательности $w(n-m)x(m)$, $-\infty<m<\infty$. Поэтому для фиксированного n $X_n(e^{i\omega})$ обладает свойствами обычного преобразования Фурье. Вторая интерпретация получается, если зафиксировать ω и рассматривать $X_n(e^{i\omega})$ как функцию времени n. При таком подходе видно, что и (6.1) и (6.3) представляют собой свертки. Это позволяет рассматривать зависящее от времени представление Фурье с помощью линейной фильтрации. Как мы увидим ниже, обе интерпретации позволяют глубже понять существо дела, поэтому представляется целесообразным подробно изучить преобразование Фурье с обеих точек зрения.

6.1.1. Интерпретация преобразования Фурье

Рассмотрим $X_n(e^{i\omega})$ как обычное преобразование Фурье последовательности $w(n-m)x(m)$, $-\infty<m<\infty$ при фиксированном n. Зависящее от времени преобразование Фурье представляет собой функцию индекса времени n, принимающего все целые значения, так что окно $w(n-m)$ «скользит» вдоль последовательности $x(m)$. Это иллюстрируется на рис. 6.1, где $x(m)$ и $w(n-m)$ показаны как функции m для нескольких значений n. (И сигнал, и окно изображены для удобства непрерывными функциями, хотя определены они только для целых значений m и $n-m$.)

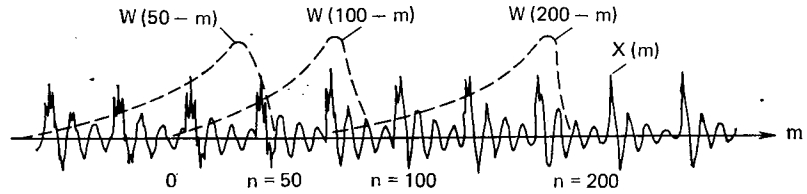


Рис. 6.1. $x(m)$ и $w(n-m)$ для нескольких значений n

Условия существования зависящего от времени преобразования Фурье легко получить, вспомнив, что достаточным условием существования обычного преобразования Фурье служит абсолютная суммируемость последовательности. Потребуем, чтобы последовательность $x(m)w(n-m)$ была абсолютно суммируемой при всех значениях n. Очевидно, что это выполняется, если, как это часто бывает, $w(n-m)$ имеет конечную длительность.

Как и в случае обычного преобразования Фурье сигналов с дискретным временем, зависящее от времени преобразование Фурье периодично по ω с периодом 2π, в чем легко убедиться, подставив $\omega+2\pi$ в (6.1). Отметим также, что это преобразование Фурье можно выразить как функцию частоты различными способами. Если, например, $\omega=\Omega T$, где T — период дискретизации, приводящей к последовательности $x(m)$, то Ω будет непрерывной частотой в радианах. При подстановке $\omega=2\pi f$ или $\omega=2\pi FT$ можно выразить преобразование Фурье как функцию нормированной частоты f и соответственно как функцию обычной непрерывной частоты F (в герцах). Нам представится возможность воспользоваться разными частотными переменными в формулах и на рисунках; по мере знакомства с подобными простыми соотношениями возможностей для путаницы будет все меньше. То обстоятельство, что $X_n(e^{i\omega})$ обладает при заданном n свойствами обычного преобразования Фурье, позволяет легко доказать, что входная последовательность $x(m)$ может быть полностью восстановлена по ее зависящему от времени преобразованию Фурье. Вспомним, что, как было отмечено выше, $X_n(e^{i\omega})$ представляет собой обычное преобразование

Фурье для $w(n-m)x(m)$:

$$w(n-m)x(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}X_n(e^{i\omega})e^{i\omega m}d\omega. \tag{6.5}$$

Заметим, что интегрировать в (6.5) можно по любому интервалу длиной 2π (например, от 0 до 2π), поскольку интегрируемая функция периодична с периодом 2π. Если теперь $w(0)\neq 0$, из (6.5) при $m=n$ можно получить

$$x(n)=\frac{1}{2\pi w(0)}\int_{-\pi}^{\pi}X_n(e^{i\omega})e^{i\omega n}d\omega. \tag{6.6}$$

Следовательно, при слабом ограничении на $w(0)$ и если $X_n(e^{i\omega})$ известна при ω в интервале, перекрывающем полный период, последовательность $x(n)$ точно восстановима по значениям $X_n(e^{i\omega})$. Это важный теоретический результат. Он важен и в приложениях, если на окно наложены некоторые дополнительные ограничения.

Связь с кратковременной автокорреляционной функцией, определенной в гл. 4, представляет собой еще одно важное свойство $X_n(e^{i\omega})$. Рассматривая $X_n(e^{i\omega})$ как обычное преобразование Фурье для $w(n-m)x(m)$ при каждом n, легко понять, что

$$S_n(e^{i\omega})=|X_n(e^{i\omega})|^2=X_n(e^{i\omega})X_n^*(e^{i\omega}) \tag{6.7}$$

есть преобразование Фурье для

$$R_n(k)=\sum_{m=-\infty}^{\infty}w(n-m)x(m)w(n-k-m)x(m+k). \tag{6.8}$$

Таким образом, соотношения (6.7) и (6.8) связывают кратковременное спектральное представление с кратковременной корреляцией, введенной в гл. 4.

Кратковременное преобразование Фурье[1] можно представить многими способами. Особенно простой вид оно принимает, если выражено через действительную и мнимую части[2]:

$$X_n(e^{i\omega})=a_n(\omega)-ib_n(\omega). \tag{6.9}$$

Можно показать, что $x(m)$ и $w(n-m)$ удовлетворяют определенным условиям симметрии и периодичности, когда и $a_n(\omega)$ и $b_n(\omega)$ действительны (см. задачу 6.1). Другое представление $X_n(e^{i\omega})$ через амплитуду и фазу:

$$X_n(e^{i\omega})=|X_n(e^{i\omega})|e^{i\theta_n(\omega)}. \tag{6.10}$$

[1] Авторы для преобразования (6.1) применяют три термина: short-time, time-dependent, time-varying. С целью устранения возможной путаницы далее сохраняется одно название — кратковременное преобразование Фурье (*Прим. ред.*)

[2] Обратите внимание на то, что $a_n(\omega)$ есть действительная часть $X_n(e^{i\omega})$, а $b_n(\omega)$ — мнимая, взятая со знаком «минус». Последнее сделано для удобства.

Величины $|X_n(e^{i\omega})|$ и $\theta_n(\omega)$ легко связать с $a_n(\omega)$ и $b_n(\omega)$ (см. задачу 6.3). Другие задачи в конце этой главы указывают на дополнительные свойства функций $a_n(\omega)$, $b_n(\omega)$ и $X_n(e^{i\omega})$.

До сих пор единственная роль окна $w(n-m)$ состояла в том, чтобы выделить для анализа часть последовательности $x(m)$. Форма временно́го окна оказывает существенное влияние на характер кратковременного преобразования Фурье. Рассмотрим это влияние последовательности $w(n-m)$. Если представить себе $X_n(e^{i\omega})$ как обычное преобразование Фурье последовательности $w(n-m)x(m)$ и если предположить, что обычные преобразования Фурье

$$X(e^{i\omega})=\sum_{m=-\infty}^{\infty}x(m)e^{-i\omega m} \tag{6.11}$$

и

$$W(e^{i\omega})=\sum_{m=-\infty}^{\infty}w(m)e^{-i\omega m} \tag{6.12}$$

существуют, то обычное преобразование Фурье для $w(n-m)x(m)$ (при фиксированном n) представляет собой свертку преобразования $w(n-m)$ и $x(m)$. Поскольку при фиксированном n преобразование Фурье $w(n-m)$ есть $W(e^{i\omega})e^{-i\omega n}$, имеем

$$X_n(e^{i\omega})=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{-i\theta})e^{-i\theta n}X(e^{i(\omega-\theta)})d\theta. \tag{6.13}$$

Заменив в (6.13) θ на $-\theta$, получим

$$X_n(e^{i\omega})=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{i\theta})e^{i\theta n}X(e^{i(\omega+\theta)})d\theta. \tag{6.14}$$

Таким образом, преобразование Фурье последовательности $x(m)$, $-\infty<m<\infty$, свертывается с преобразованием Фурье сдвинутой последовательности окна. Этот результат нуждается в оговорке, поскольку, строго говоря, обычного преобразования Фурье речевого сигнала не существует. Соотношение (6.14) станет полезным, если вспомнить, что смысл использования окна состоит в том, чтобы выделить некоторый конечный сегмент речевого сигнала в окрестности отсчета n и устранить остальную часть сигнала. И действительно, типичное окно таково, что $w(n-m)=0$ для m, не принадлежащих конечному интервалу вокруг n. Имея в виду конечный результат, вполне оправдано считать, что свойства последовательности определяются той ее частью, которая попадает в окно. Если, например, речевой сигнал в пределах окна соответствует гласному или другому вокализованному звуку, то можно рассматривать результирующую последовательность $x(m)w(n-m)$, как соответствующую периодически продолженному вокализованному звуку. Аналогичным образом, если речь в пределах окна не вокализована, можно считать, что характеристики невокализованного

сигнала сохраняются и вне окна. Равным образом можно считать, что сигнал вне окна равен нулю. Эта интерпретация пригодится при анализе переходных звуков, таких, как взрывные.

Следовательно, (6.14) имеет смысл как в случае, когда мы считаем, что $X(e^{i\theta})$ представляет собой преобразование Фурье сигнала, свойства которого сохраняются и вне окна, так и в случае, когда $X(e^{i\theta})$ соответствует сигналу, обращающемуся вне окна в нуль. Таким образом, кратковременное преобразование Фурье можно интерпретировать как сглаженное преобразование Фурье для части сигнала, попавшего в окно.

В этом смысле становятся важными свойства преобразования Фурье окна $W(e^{i\theta})$. Из (6.14) ясно, что для достоверного отражения свойств $X(e^{i\omega})$ в $X_n(e^{i\omega})$, необходимо чтобы функция $W(e^{i\theta})$ носила импульсный характер по сравнению с $X(e^{i\omega})$. В гл. 4 уже обсуждались свойства прямоугольного окна и окна Хемминга. Было показано, что ширина главного лепестка $W(e^{i\theta})$ обратно пропорциональна ширине окна, в то время как уровень боковых лепестков от ширины окна по существу не зависит.

Эффекты, связанные с использованием окон при спектральном анализе речи, иллюстрируются рис. 6.2—6.5. Часть *а*) каждого из этих рисунков показывает сигнал $x(m)w(n—m)$, взвешенный окном Хемминга, часть *б*) представляет собой логарифм амплитуды $X_n(e^{i\omega})$ (в децибелах), часть *в*) — сигнал, взвешенный прямоугольным окном, и часть *г*) — логарифм амплитуды соответствующего спектра. На рис. 6.2 представлены результаты для окна длительностью 500 отсчетов (50 мс при частоте дискретизации 10 кГц) и сегмента вокализованной речи. Отчетливо прослеживается периодичность сигнала на рис. 6.2*а* (временна́я диаграмма) и 6.2*б*, где основная частота и ее гармоники проявляются в кратковременном преобразовании Фурье как узкие пики, разнесенные равномерно по частоте. Из рис. 6.2*б* видно также, что спектр состоит из пика первой форманты в районе 300—400 Гц, и широкого пика около 2200 Гц, соответствующего второй и третьей формантам. Заметен также пик четвертой форманты — около 3800 Гц. И, наконец, видно, что спектр имеет тенденцию спадать на высоких частотах, что характерно для спектра импульсов возбуждения. Сравнение спектров на рис. 6.2*б* (окно Хемминга) и 6.2*г* (прямоугольное окно), выявляет их большое сходство по гармоникам основного тона, формантной структуре и форме спектра в целом. Видны и различия в спектрах, из них наиболее отчетливо — большая острота гармоник основного тона на рис. 6.2*г*, что вызвано лучшим частотным разрешением прямоугольного окна в сравнении с окном Хемминга той же ширины. Другое различие в спектрах вызвано тем, что сравнительно большие боковые лепестки прямоугольного окна дают «рваный», или «зашумленный» спектр. Этот эффект связан с тем, что боковые лепестки смежных гармоник взаимодействуют в интервалах между гармониками , иногда усиливаясь, иногда уничтожая друг друга, в целом производя впечатление довольно беспорядочного характера спектра между гармониками. Нежела-

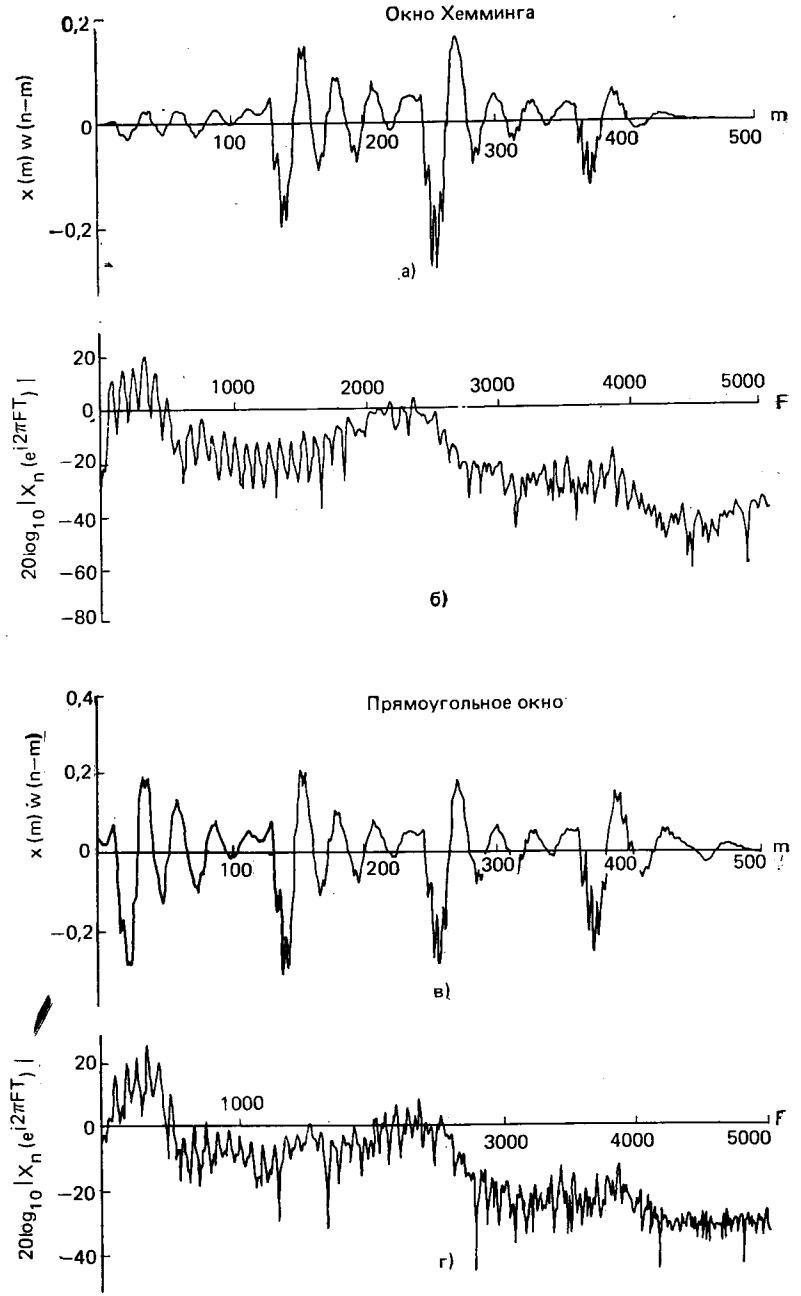


Рис. 6.2. Спектральный анализ вокализованной речи с окном длительностью 50 мс:
а, *б*) — окно Хемминга; *в*, *г*) — прямоугольное окно (на рис. *а* и *в* показан сигнал, а на рис. *б* и *г* — спектры)

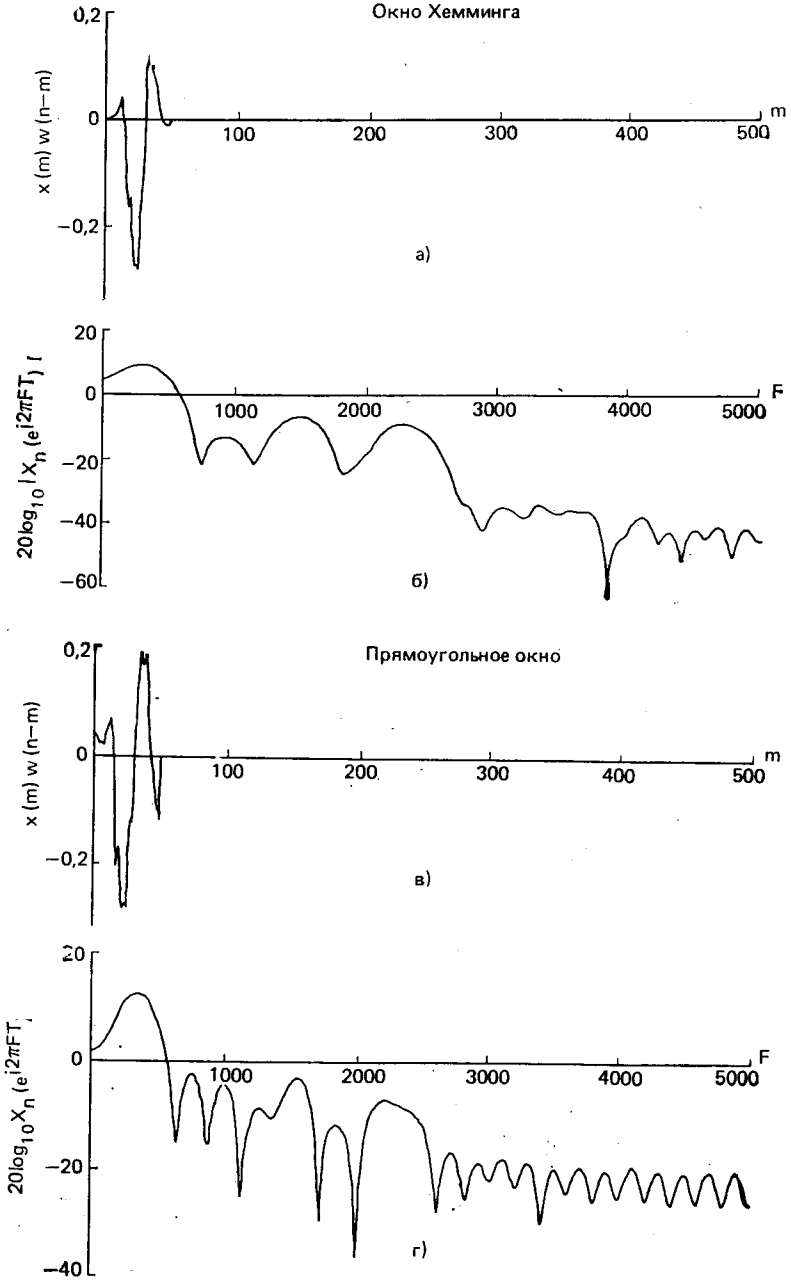


Рис. 6.3. Спектральный анализ вокализованной речи с окном длительностью 5 мс:

а, б) — окно Хемминга; *в, г*) — прямоугольное окно (на рис. *а* и *в* показан сигнал, а на рис. *б* и *г* — спектры)

тельное «рассеивание» спектра между смежными гармониками сводит на нет выгоды, связанные с узостью основного лепестка прямоугольного окна. Поэтому такие окна редко используются в спектральном анализе речи.

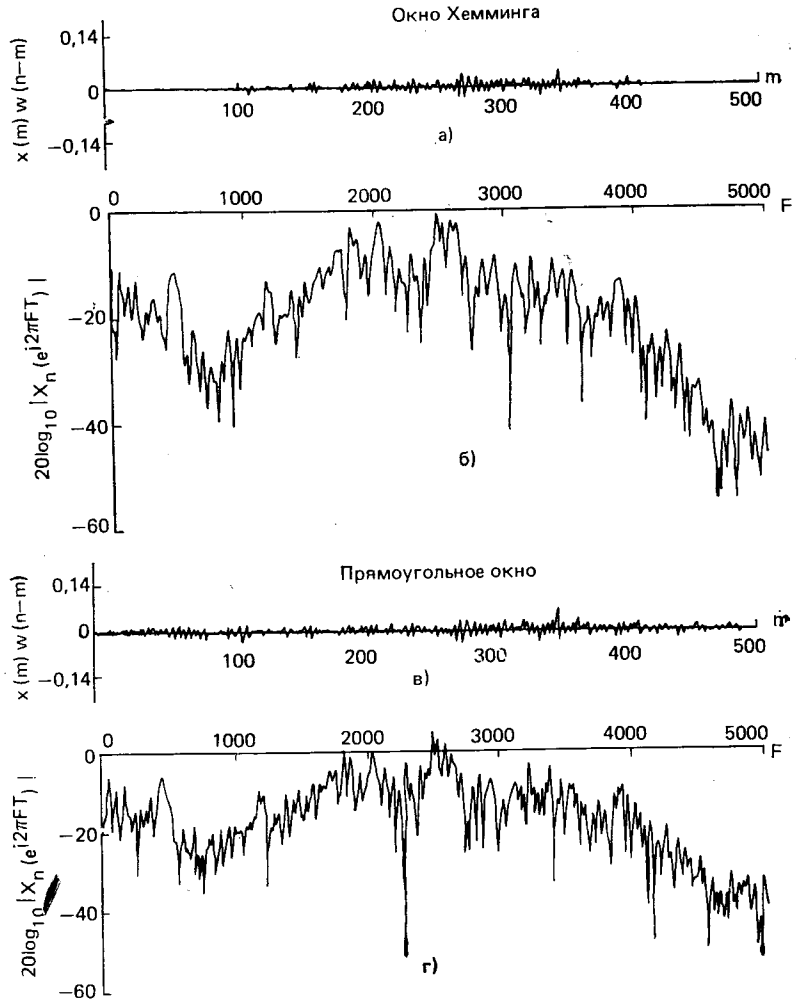


Рис. 6.4. Спектральный анализ невокализованной речи с окном длительностью 50 мс:

а, б) — окно Хемминга; *в, г*) — прямоугольное окно (на рис. *а* и *в* показан сигнал, а на рис. *б* и *г* — спектры)

На рис. 6.3 аналогичным образом проведено сравнение для сегмента вокализованной речи длительностью 50 отсчетов (5 мс). Для столь узкого окна ни временна́я последовательность $x(m)\times$ $\times w(n-m)$ (рис. 6.3*а*, *в*), ни спектр сигнала (рис. 6.3*б*, *г*) не вы-

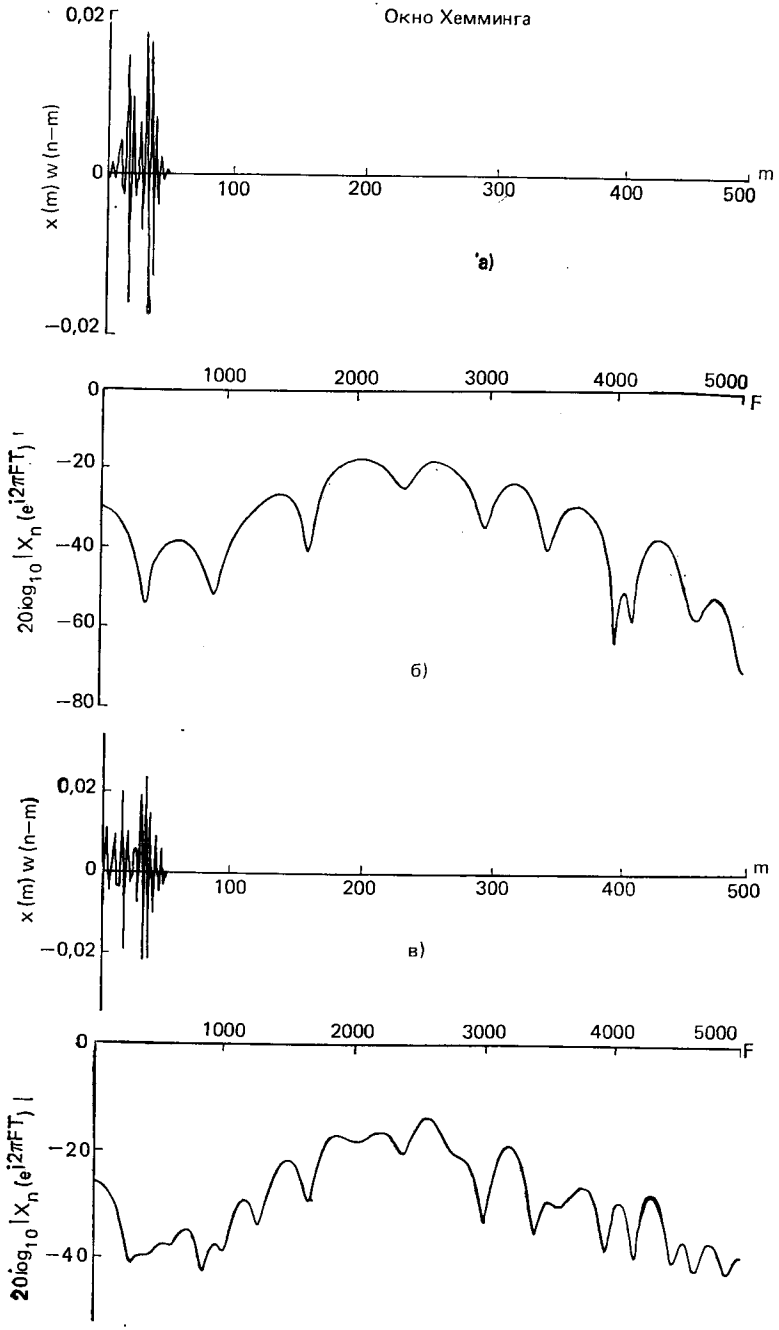


Рис. 6.5. Спектральный анализ невокализованной речи с окном длительностью 5 мс:
а, *б*) — окно Хемминга; *в*, *г*) — прямоугольное окно (на рис. *а* и *в* показан сигнал, а на рис. *б* и *г* — спектры)

являют периодичности сигнала. В сравнении с рис. 6.2 спектр на рис. 6.3 имеет только несколько довольно широких пиков вблизи 400, 1400 и 2200 Гц, соответствующих первым трем формантам отрезка речи, попавшего в окно. Сравнение спектров снова выявляет большее частотное разрешение прямоугольного окна.

Рисунки 6.4 и 6.5 иллюстрируют влияние окна для сегментов невокализованной речи (соответствующих фрикативу /*sh*/) длиной 500 отсчетов (рис. 6.4) и длиной 50 отсчетов (рис. 6.5). На этих рисунках видно, что спектр имеет медленно меняющийся тренд по частоте, на который наложена последовательность острых пиков и провалов. «Рваный» характер спектра (для обоих окон) связан со случайной природой невокализованной речи. И, наконец, видно, что окно Хемминга дает несколько более гладкий спектр в сравнении с прямоугольным окном.

Примеры, приведенные на рис. 6.2 и 6.5, хорошо иллюстрируют зависимость между длительностью окна и свойствами кратковременного преобразования Фурье: разрешение по частоте обратно пропорционально ширине окна. Вспомнив, что задача окна состоит в том, чтобы выделить подлежащий анализу интервал времени, сохранив, однако, характеристики колебания без существенных изменений, мы убеждаемся в необходимости компромисса. На рис. 6.2, например, видно, что частоты формант явно меняются на интервале, равном 50 мс. Необходим более короткий интервал анализа, чтобы выявить эти изменения. Окна шириной 5 мс, размещенные в начале и в конце 50-миллисекундного интервала, явно дадут другие кратковременные преобразования Фурье. Следовательно, хорошее разрешение по времени требует узкого окна, а хорошее разрешение по частоте — широкого. Позже, изучая приложения, мы приведем примеры использования обоих типов окон.

Итак, интерпретация кратковременного преобразования Фурье как обычного преобразования Фурье выделенного окном сегмента речевого сигнала позволяет глубже понять свойства кратковременного представления Фурье и роль самого окна.

6.1.2. Интерпретация посредством линейной фильтрации

Как ясно из (6.1), для каждого ω $X_n(e^{i\omega})$ представляет собой свертку последовательности $w(n)$ с последовательностью $x(n)e^{-i\omega n}$. Поэтому для фиксированных ω можно представлять себе $X_n(e^{i\omega})$ как выход системы, изображенной на рис. 6.6*а*, где $w(n)$ играет роль импульсной характеристики линейной системы, инвариантной относительно сдвига. На рис. 6.6*а* вход и выход линейной системы комплексные. Выразим $X_n(e^{i\omega})$ в виде

$$X_n(e^{i\omega}) = a_n(\omega) - i\, b_n(\omega). \qquad (6.15)$$

Операции, необходимые для расчета $a_n(\omega)$ и $b_n(\omega)$, показаны на рис. 6.6*б*, где имеются только действительные последовательности.

Чтобы понять, как в системе на рис. 6.6*а* формируется кратковременное преобразование Фурье на частоте ω, полезно снова

предположить, что существует обычное преобразование Фурье $x(n)$. Обозначим преобразование Фурье $x(n)$ через $X(e^{i\theta})$, чтобы избежать путаницы с частотными переменными. (Напомним, что теперь ω — фиксированное значение частоты в радианах.) В результате модуляции преобразование Фурье входа линейного фильтра равно $X(e^{i(\theta+\omega)})$. Поэтому спектр $x(n)$ на частоте ω сдвигается

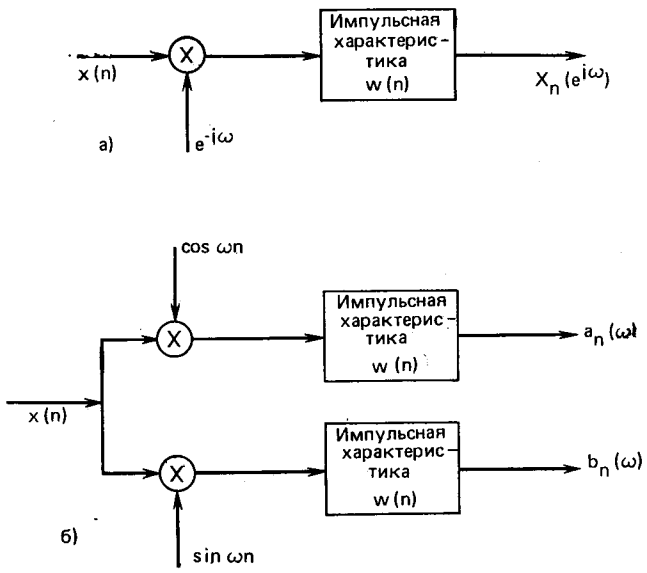


Рис. 6.6. Интерпретация кратковременного анализа как линейной фильтрации: *а*) — комплексные операции; *б*) — действительные операции

к нулю. Поскольку преобразование Фурье последовательности, поступающей с выхода фильтра, есть $X(e^{i(\theta+\omega)})W(e^{i\theta})$, то в случае фильтра нижних частот с очень узкой полосой пропускания выходная последовательность фильтра существенным образом зависит от $X(e^{i\omega})$. Поэтому, как следует из предшествующей интерпретации, $W(e^{i\theta})$ должна быть отличной от нуля только в очень узком интервале вокруг нулевой частоты и быть по возможности меньшей вне этого интервала. Отметим интересный факт: правая часть (6.14) представляет собой преобразование Фурье $W(e^{i\theta})X(e^{i(\theta+\omega)})$.

Еще одна интерпретация для $X_n(e^{i\omega})$ через линейные фильтры получается из рассмотрения (6.2). Как видно из рис. 6.7*а*, $X_n(e^{i\omega})$ можно представить себе как результат модуляции $e^{-i\omega n}$ выходным сигналом комплексного полосового фильтра с импульсной характеристикой $w(n)e^{i\omega n}$. Если преобразование Фурье $W(e^{i\theta})$ имеет вид характеристики фильтра нижних частот, то фильтр, изображенный на рис. 6.7*а*, будет полосовым фильтром с центральной частотой ω. На рис. 6.7*б* приведена та же система, что и на рис. 6.7*а*, но используются только действительные величины.

Сравнение рис. 6.6*б* и 6.7*б* показывает, что в случае, когда необходимы $a_n(\omega)$ и $b_n(\omega)$, реализация системы рис. 6.6*б* проще. Если, однако, требуется получить только $|X_n(e^{i\omega})|$, то проще реализация с полосовым фильтром. Это становится понятным, если заметить, что из (6.4) и (6.9) получается

$$|X_n(e^{i\omega})| = [a_n^2(\omega) + b_n^2(\omega)]^{1/2} = \tag{6.16а}$$

$$= |\tilde{X}(e^{i\omega})| = [\tilde{a}_n^2(\omega) + \tilde{b}_n^2(\omega)]^{1/2}. \tag{6.16б}$$

Рисунок 6.8 иллюстрирует выражение (6.16а), а рис. 6.8*б* — (6.16б). Система, изображенная на рис. 6.8*б*, будет, вообще говоря, проще.

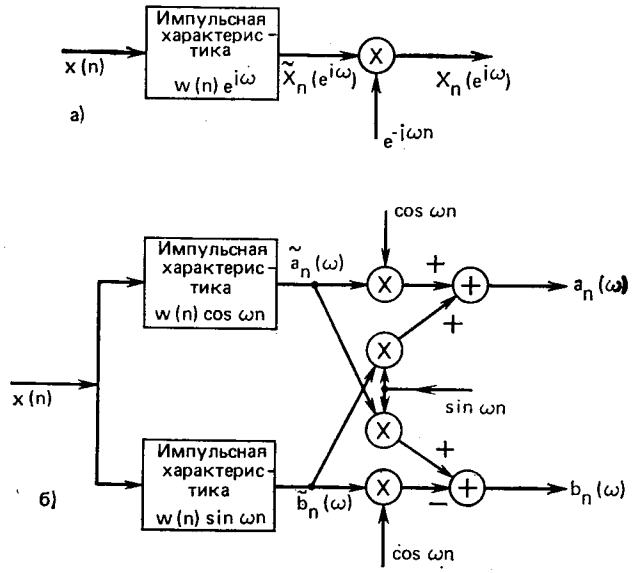


Рис. 6.7. Другая интерпретация кратковременного спектрального анализа как линейной фильтрации:
а) комплексные операции; *б*) действительные операции

Рассматривая $X_n(e^{i\omega})$ при фиксированном ω как выход системы, изображенной на рис. 6.6 и 6.7, можно воспользоваться нашими знаниями линейных систем, чтобы уяснить свойства кратковременного представления Фурье. Полезно, например, вспомнить, что импульсная характеристика линейной системы с дискретным временем, инвариантной относительно сдвигом, может быть либо конечной (КИХ), либо бесконечной (БИХ) длительности. Аналогичным образом можно определить два класса окон для кратковременного анализа Фурье. Вспомним также, что линейная инвариантная к сдвигу система может быть или не быть физически реализуемой в зависимости от того, равна ли нулю ее импульсная характеристика при $n<0$. Точно так же можно классифицировать окна на физиче-

ски реализуемые или нереализуемые. Для физически реализуемого окна

$$w(n)=0, \quad n<0, \tag{6.17а}$$

или, что эквивалентно, $$w(n-m)=0, \quad n<m. \tag{6.17б}$$

Окно Хемминга и прямоугольное окно являются окнами конечной ширины. При подходящем выборе начала отсчета во времени

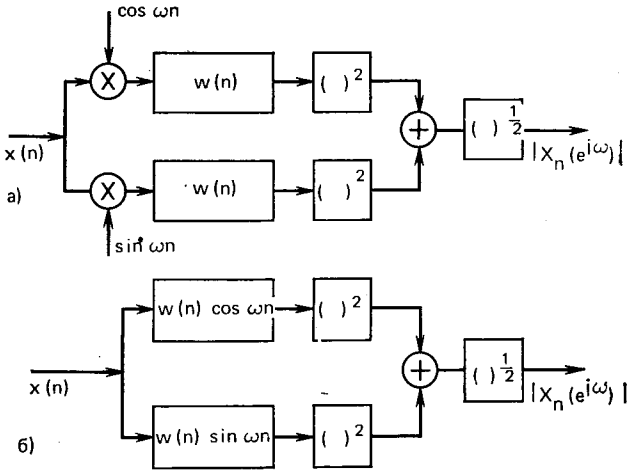


Рис. 6.8. Две схемы вычисления амплитудного кратковременного спектра: *а*) с фильтрами низкой частоты; *б*) с полосовыми фильтрами

их можно задать и как физически реализуемые. Как мы увидим ниже, такие окна пригодны как для реализации, основанной на схемах, изображенных на рис. 6.6 и 6.7, так и при реализации, основанной на дискретном преобразовании Фурье. Окна бесконечной ширины также оказываются полезными, особенно если $X_n(e^{i\omega})$ вычисляется с помощью линейной фильтрации (рис. 6.6 и 6.7). В подобных случаях можно получить рекуррентные формулы для функции $X_n(e^{i\omega})$ через ее значения в предыдущие моменты времени (см. задачу 6.6).

6.1.3. Частоты дискретизации $X_n(e^{i\omega})$ по времени и частоте [1]

Кратковременное преобразование Фурье есть двумерное представление одномерного сигнала $x(n)$. Иначе говоря, $X_n(e^{i\omega})$ представляет собой функцию и времени n и круговой частоты ω. При цифровой реализации системы кратковременного анализа Фурье одним из основных оказывается вопрос о выборе частот дискретизации $X_n(e^{i\omega})$ по времени и частоте так, чтобы избежать наложения и получить представление для $X_n(e^{i\omega})$, по которому можно было бы точно восстановить $x(n)$.

[1] Содержание оставшейся части § 6.1 основано на работах [1—4].

Вопрос этот ни в коей мере не тривиален и требует тщательного учета всех факторов, связанных с вычислением $X_n(e^{i\omega})$, для того, чтобы получить правильные значения частот дискретизация по времени и частоте. Как будет показано ниже, выбор частот дискретизации усложняется тем, что и в частотной и во временно́й областях можно использовать частоты дискретизации, меньшие теоретически минимальной, и все же точно восстановить $x(n)$ по кратковременному представлению (при наличии наложения частот).

Такого рода представления (с пониженной частотой дискретизации) в действительности весьма полезны для приложений, в которых важен лишь кратковременный анализ Фурье (например, для оценки спектра, формантного анализа и анализа основного тона, для получения спектрограмм речи в дискретном времени и т. п.), а также применительно к вокодерам, в которых важнее всего минимизация общей скорости передачи в системе. В тех случаях, когда требуется получить кратковременное преобразование Фурье и далее произвести некоторые операции над ним (например, линейно или нелинейно отфильтровать), а затем синтезировать сигнал, важно, чтобы наложения не возникали ни в частотной, ни во временно́й областях.

Рассмотрим требования к частоте дискретизации, начиная с временно́й области. В этом случае наша интуиция может опираться на приведенную в предыдущем разделе интерпретацию посредством линейной фильтрации. Там было показано, что для фиксированного значения ω $X_n(e^{i\omega})$ будет выходом фильтра с импульсной характеристикой $w(n)$. Обозначим преобразование Фурье $w(n)$ через $W(e^{i\omega})$. Для большинства разумно выбранных окон функция $W(e^{i\omega})$ обладает свойствами частотной характеристики фильтра нижних частот. Обозначим эффективную полосу анализирующего фильтра через B Гц[1]. Таким образом, у $X_n(e^{i\omega})$ такая же полоса, как и у окна, а следовательно, в соответствии с теоремой отсчетов $X_n(e^{i\omega})$ должна быть дискретизирована, по крайней мере, с частотой 2 B отсч./с для того, чтобы избежать наложений. В качестве примера рассмотрим окно Хемминга:

$$w(n)=\begin{cases} 0{,}54-0{,}46\cos(2\pi n/(L-1)), & 0\leqslant n\leqslant L-1; \\ 0, & \textit{в противном случае,} \end{cases} \tag{6.18}$$

Ширина полосы $W(e^{i\omega})$, выраженная через непрерывную частоту, приближенно равна

$$B=2F_s/L, \tag{6.19}$$

где F_s — частота дискретизации сигнала $x(n)$ и, следовательно, требуемая частота дискретизации $X_n(e^{i\omega})$ во временной области равна $2B=4F_s/L$ отсч./с. Таким образом, для $L=100$, $F_s=$

[1] Отметим, что здесь может возникнуть путаница с частотными переменными. Напомним, что когда мы рассматриваем $X_n(e^{i\omega})$ как функцию времени, ω — число (фиксировано). Однако эта переменная ω используется также для обозначения переменной частоты в $X_n(e^{i\omega})$.

$=10\,000$ Гц получим $B=200$ Гц, и $X_n(e^{i\omega})$ должна вычисляться 400 раз/с, т. е. через каждые 25 отсчетов.

Ввиду того что функция $X_n(e^{i\omega})$ периодична по ω с периодом 2π, дискретизация необходима только в интервале длиной 2π. Воспользуемся интерпретацией $X_n(e^{i\omega})$ через преобразование Фурье, чтобы найти подходящий конечный набор частот $\omega_k=2\pi k/N$; $k=0, 1, \ldots, N-1$, в которых необходимо задать $X_n(e^{i\omega})$ для однозначного восстановления $x(n)$. Если окно ограничено во времени и если $X_n(e^{i\omega})$ рассматривать как преобразование Фурье, то и обратное преобразование окажется ограниченным во времени. В таком случае теорема отсчетов требует, чтобы мы дискретизовали $X_n(e^{i\omega})$ в частотной области с частотой, по крайней мере вдвое превышающей «ширину временно́й полосы». Поскольку обратным преобразованием Фурье $X_n(e^{i\omega})$ будет сигнал $x(m)w(n-m)$ и этот сигнал имеет продолжительность L отсчетов (опять из-за конечной ширины окна), то по теореме отсчетов $X_n(e^{i\omega})$ необходимо дискретизовать (по частоте) на частотах

$$\omega_k=2\pi k/L, \quad k=0, 1, \ldots, L-1, \tag{6.20}$$

чтобы точно восстановить $x(n)$ по $X_n(e^{i\omega_k})$ (см. задачу 6.8). Таким образом, в примере с окном Хемминга шириной $L=100$ отсчетов требуется, чтобы $X_n(e^{i\omega})$ вычислялось по крайней мере для 100 равномерно распределенных на единичной окружности частот.

Основываясь на приведенном обсуждении, можно определить полное число отсчетов, подлежащее вычислению каждую секунду для того, чтобы избежать наложений и получить представление исходного сигнала. Минимальная частота дискретизации $X_n(e^{i\omega_k})$ во временно́й области равна $2B$, где B — ширина полосы частот окна, а минимальное число отсчетов в частотной области равно L — ширине окна во времени. Таким образом, «полная частота дискретизации» для $X_n(e^{i\omega_k})$ равна

$$SR=2BL \text{ отсч./с.} \tag{6.21}$$

Для большинства практически используемых окон B можно представить как кратное от (F_s/L), где F_s — частота дискретизации $x(n)$, т. е.

$$B=CF_s/L \text{ Гц.} \tag{6.22}$$

Здесь C — константа пропорциональности. Следовательно, (6.21) можно переписать в виде

$$SR=2CF_s \text{ отсч./с.} \tag{6.23}$$

Соотношение SR и F_s поэтому равно

$$SR/F_s=2C. \tag{6.24}$$

Величина $2C$ служит коэффициентом «излишка дискретизации» кратковременного анализа в сравнении с обычным дискретизированным представлением $x(n)$.

Из примеров видно, что $2C=4$, если $w(n)$ — окно Хемминга, и $2C=2$, если $w(n)$ — прямоугольное окно (и если ширина полосы

определена по первому нулю $W(e^{i\omega})$). Следовательно, для кратковременного спектрального представления $x(n)$ требуется примерно в 2—4 раза больше отсчетов, чем для описания самого сигнала. Взамен, однако, получено весьма гибкое представление сигнала, позволяющее производить целый ряд манипуляций как в частотной, так и во временно́й областях. Хотя рассчитанные частоты дискретизации и представляют собой теоретически минимальные, все же существуют специальные случаи, в которых можно дискретизовать $X_n(e^{i\omega_k})$ с пониженной частотой во временно́й или частотной области и для которых сохраняется возможность точного восстановления $x(n)$ без ошибок наложения. Такие случаи важны при реализации систем с минимальной памятью (скоростью передачи), например систем анализа — синтеза, визуального отображения спектра и т. п. Далее в этой главе мы рассмотрим, как спроектировать и реализовать такие системы. Вначале, однако, опишем два различных способа восстановления $x(n)$ по дискретизированной $X_n(e^{i\omega_k})$, а затем влияние модификаций $X_n(e^{i\omega_k})$ на результирующий сигнал.

6.1.4. Кратковременный синтез методом суммирования выходов гребенки фильтров

Первый метод синтеза связан с интерпретацией кратковременного анализа как преобразования в гребенке фильтров. Ранее было показано, что для любой частоты ω_k $X_n(e^{i\omega_k})$ есть узкополосное представление сигнала в полосе с центральной частотой ω_k. Из (6.1) и (6.2) можно получить для $X_n(e^{i\omega_k})$ выражения

$$X_n(e^{i\omega_k})=\sum_{m=-\infty}^{\infty} w_k(n-m)\,x(m)\,e^{-i\omega_k m} \tag{6.25}$$

или

$$X_n(e^{i\omega_k})=e^{-i\omega_k n}\sum_{m=-\infty}^{\infty} x(n-m)\,w_k(m)\,e^{i\omega_k m}, \tag{6.26}$$

где $w_k(m)$ — окно, используемое на частоте ω_k. Определив

$$h_k(n)=w_k(n)\,e^{i\omega_k n}, \tag{6.27}$$

перепишем (6.26):

$$X_n(e^{i\omega_k})=e^{-i\omega_k n}\sum_{m=-\infty}^{\infty} x(n-m)\,h_k(m). \tag{6.28}$$

Поскольку окно $w_k(n)$ обладает свойствами фильтра нижних частот, (6.28) можно интерпретировать как полосовую фильтрацию с импульсной характеристикой $h_k(n)$ и последующей модуляцией комплексной экспонентой $e^{-i\omega_k n}$, как это показано на рис. 6.7. Определив

$$y_k(n)=X_n(e^{i\omega_k})\,e^{i\omega_k n}, \tag{6.29}$$

получим из (6.28) следующее выражение:

$$y_k(n) = \sum_{m=-\infty}^{\infty} x(n-m)\, h_k(m). \qquad (6.30)$$

Таким образом, $y_k(n)$ есть просто выходная последовательность полосового фильтра с импульсной характеристикой $h_k(n)$ (6.27). Операции, определенные (6.28) и (6.29), изображены на рис. 6.9*а*. Поскольку (6.25) и (6.28) эквивалентны, то в (6.29) можно использовать любое из выражений для $X_n(e^{i\omega_k})$. В обоих случаях система, связывающая $x(n)$ с $y_k(n)$, представляет собой полосовой фильтр с импульсной характеристикой $h_k(n)$. Это показано на рис. 6.9, причем рис. 6.9*а* отображает (6.28) и (6.29), а рис. 6.9*б*—(6.25) и (6.29). На рис. 6.9*в* показан эквивалентный полосовой фильтр.

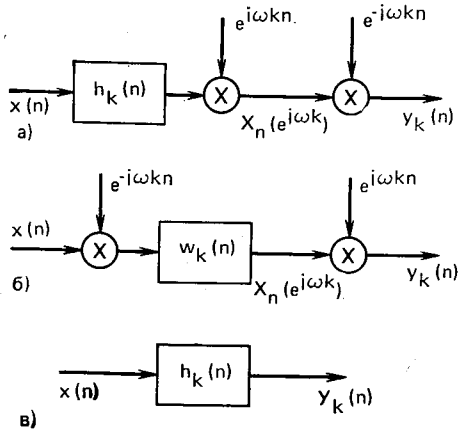


Рис. 6.9. Метод построения одного канала синтеза в рамках линейной фильтрации

Результаты, объединенные в схеме рис. 6.9, подсказывают практический способ восстановления входного сигнала по его кратковременному преобразованию Фурье. Возьмем набор из N частот $\{\omega_k\}$, $k=0, 1, \ldots, N-1$, и положим, что нам известны $X_n(e^{i\omega_k})$ для каждой из частот. Рассмотрим теперь N полосовых фильтров вида (6.27). Допустим, например, что $w_k(n)$ представляет собой импульсную характеристику идеального фильтра нижних частот с частотой среза ω_{pk}. Частотная характеристика $W_k(e^{i\omega})$ такого фильтра приведена на рис. 6.10*а*[1]. Частотной характеристикой соответствующего комплексного полосового фильтра с импульсной характеристикой $h_k(n) = w_k(n)\, e^{i\omega_k n}$ будет тогда

$$H_k(e^{i\omega}) = W_k\left(e^{i(\omega-\omega_k)}\right), \qquad (6.31)$$

как это показано на рис. 6.10*б*. Заметим, что центральная частота равна ω_k, а ширина полосы равна $2\omega_{pk}$.

Рассмотрим набор из N полосовых фильтров с центральными частотами, равномерно разнесенными так, что перекрывается весь диапазон основных частот:

$$\omega_k = 2\pi k/N, \quad k=0,\ 1, \ldots,\ N-1. \qquad (6.32)$$

Допустим также, что окно одинаково для всех каналов:

$$w_k(n) = w(n), \quad k=0,1,\ldots,\ N-1. \qquad (6.33)$$

[1] Здесь ω — частотная переменная.

Если теперь рассмотреть все полосовые фильтры вместе, считая, что вход одинаков, а выходы суммируются, как это показано на рис. 6.11, то общей частотной характеристикой, связывающей $y(n)$ с $x(n)$, будет

$$\tilde{H}(e^{i\omega}) = \sum_{k=0}^{N-1} H_k(e^{i\omega}) = \sum_{k=0}^{N-1} W\left(e^{i(\omega-\omega_k)}\right). \qquad (6.34)$$

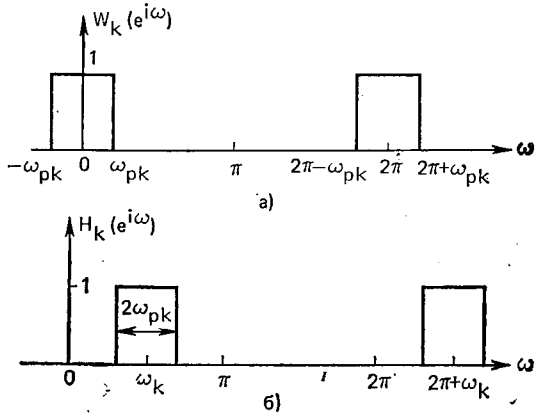


Рис. 6.10 Частотные характеристики:
а) идеального ФНЧ; *б*) идеального полосового фильтра

Если $W(e^{i\omega_k})$ должным образом дискретизована по частоте (т. е. если $N \geqslant L$, где L — длительность окна), то, как можно показать,

$$\frac{1}{N}\sum_{k=0}^{N-1} W\left(e^{i(\omega-\omega_k)}\right) = w(0) \qquad (6.35)$$

для всех ω.

Равенство (6.35) выводится следующим образом. Обратным преобразованием Фурье для $W(e^{i\omega})$ служит окно $w(n)$. Если $W(e^{i\omega})$ дискретизована по частоте с N равномерно разнесенными частотами, то обратное преобразование Фурье дискретизованной $W(e^{i\omega k})$ имеет вид

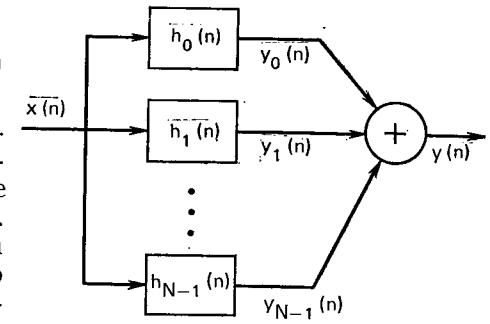


Рис. 6.11. Эквивалентная линейная система, связывающая $y_k(n)$ и $y(n)$ с $x(n)$

$$\frac{1}{N}\sum_{k=0}^{N-1} W\left(e^{i\omega_k}\right) e^{i\omega_k n} = \sum_{r=-\infty}^{\infty} w(n+rN), \qquad (6.36)$$

т. е. получено представление $w(n)$ с наложениями (см. задачу 6.8). Если $w(n)$ имеет продолжительность в L отсчетов, то

$$w(n)=0, \quad n<0, \quad n\geqslant L. \tag{6.37}$$

и наложений за счет дискретизации $W(e^{i\omega})$ по частоте нет. В этом случае, вычислив (6.36) для $n=0$, получим

$$\frac{1}{N}\sum_{k=0}^{N-1} W\left(e^{i\omega_k}\right)=w(0). \tag{6.38}$$

Теперь легко получить (6.35), заметив, что $W(e^{i(\omega-\omega_k)})$ представляет собой равномерно дискретизованную $W(e^{i\omega})$, вычисленную в точке $\omega-\omega_k$ вместо ω_k. В соответствии с теоремой отсчетов годится любой набор из N равномерно разнесенных отсчетов. Поэтому (6.35) следует из (6.38) и теоремы отсчетов.

Из (6.38) и (6.35) видно, что для системы в целом импульсная характеристика

$$\tilde{h}(n)=\sum_{k=0}^{N-1} h_k(n)=\sum_{k=0}^{N-1} w(n)\,e^{i\omega_k n}=N\,w(0)\,\delta(n) \tag{6.39}$$

просто совпадает с масштабированным единичным отсчетом $Nw(0)\delta(n)$, а следовательно, общим выходом $y(n)$ будет $Nw(0)\times$ $\times x(n)$.

Таким образом, в методе суммирования выходов гребенки фильтров восстановленный сигнал формируется как

$$y(n)=\sum_{k=0}^{N-1} y_k(n)=\sum_{k=0}^{N-1} X_n\left(e^{i\omega_k}\right)e^{i\omega_k n}, \tag{6.40}$$

и мы показали, что, если $X_n(e^{i\omega_k})$ дискретизована должным образом по частоте, $y(n)=Nw(0)x(n)$ независимо от конкретной формы окна. Операции анализа и синтеза, подразумеваемые в (6.40), изображены на рис. 6.12, причем фильтры на рисунке — полосовые.

Мы только что установили весьма важный результат: при условии, что $w(n)$ имеет конечную длительность L, последовательность $x(n)$ можно точно восстановить по кратковременному преобразованию Фурье, дискретизованному и по времени, и по частоте. Можно также показать, что если $W(e^{i\omega})$ строго ограничена по частоте, то аналогично $x(n)$ может быть точно восстановлена по $X_n(e^{i\omega_k})$. В действительности, существует много способов точного восстановления $x(n)$ по кратковременному преобразованию.

Для того чтобы избежать наложений, нужно вычислить $X_n(e^{i\omega_k})$, по крайней мере на L равномерно разнесенных частотах, где L — длительность окна. Ширина полосы окна длительностью в L отсчетов лежит, вообще говоря, между $2\pi/L$ (для прямоугольного окна) и $4\pi/L$ (для окна Хемминга). Поскольку частоты анализа равны $2\pi k/L$, эффективные полосы полосовых фильтров пере-

крываются. Как указывалось ранее, существует способ, при котором можно вычислять $X_n(e^{i\omega_k})$ в неперекрывающихся полосах и тем не менее восстановить $x(n)$ точно (по крайней мере теоретически).

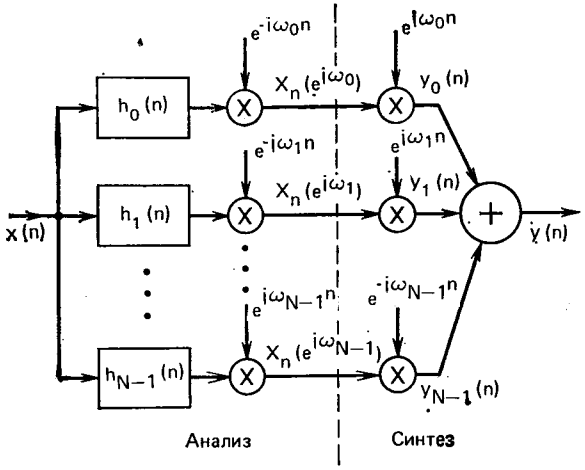


Рис. 6.12. Операции анализа и синтеза для процедуры кратковременного спектрального анализа

Чтобы показать это, допустим, что ширина окна для всех полос равна L отсчетам и что такое одно и то же окно используется для N равноразнесенных полос частот с частотами анализа:

$$\omega_k=2\pi k/N, \quad k=0,\ 1,\ldots,\ N-1, \tag{6.41а}$$

где N может быть меньше L. Допустим также, что $w(n)$ представляет собой идеальный фильтр нижних частот с частотой среза

$$\omega_p=\pi/N. \tag{6.41б}$$

Эта ситуация показана на рис. 6.13, где приведена общая характеристика для $N=6$ равноразнесенных идеальных фильтров. В этом случае (6.39) переходит в

$$\tilde{h}(n)=\sum_{k=0}^{N-1} w(n)\,e^{i\omega_k n}=w(n)\sum_{k=0}^{N-1} e^{i\omega_k n}. \tag{6.42}$$

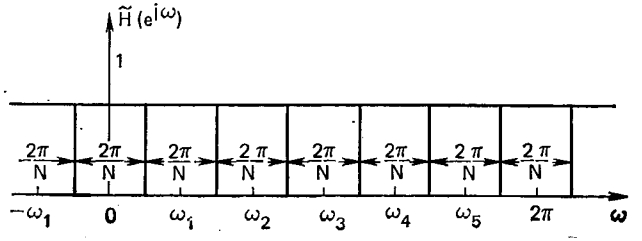


Рис. 6.13. Общая частотная характеристика для шести равноразнесенных идеальных фильтров

Определив

$$p(n)=\sum_{k=0}^{N-1} e^{i\omega_k n}=\sum_{k=0}^{N-1} e^{i 2\pi kn/N}, \tag{6.43}$$

перепишем $\tilde{h}(n)$ в виде

$$\tilde{h}(n)=w(n)\,p(n). \tag{6.44}$$

Легко видеть, что последовательность $p(n)$ периодична с периодом N. В действительности можно показать (см. задачу 6.7), что $p(n)$ есть периодическая последовательность импульсов с амплитудой N:

$$p(n)=N\sum_{r=-\infty}^{\infty}\delta(n-rN). \tag{6.45}$$

Поэтому для $\tilde{h}(n)$ имеем

$$\tilde{h}(n)=N\sum_{r=-\infty}^{\infty} w(rN)\,\delta(n-rN). \tag{6.46}$$

Следовательно, общая импульсная характеристика представляет собой просто последовательность окна, дискретизованную через интервалы длиной в N отсчетов. Это показано на рис. 6.14: на рис. 6.14*а* — последовательность $p(n)$, на рис. 6.14*б* — импульсная характеристика идеального фильтра нижних частот с частотой среза π/N, т. е.

$$w(n)=\sin(\pi n/N)/\pi n. \tag{6.47}$$

Сравнивая рис. 6.14*а* и *б* можно понять, что произведение $\tilde{h}(n)=p(n)w(n)$ равно нулю всюду, кроме точки $n=0$, где оно равно единице. Поэтому общая импульсная характеристика равна

$$\tilde{h}(n)=\delta(n). \tag{6.48}$$

И хотя при этом фильтр нижних частот предполагался идеальным, характер взаимодействия $p(n)$ и $w(n)$, при котором образуется общая характеристика, подсказывает множество возможностей для выбора $w(n)$ так, чтобы можно было восстановить сигнал по дискретизованному кратковременному преобразованию. Заметим, что если $w(n)$ имеет конечную ширину $L<N$ и физически реализуемо, то общая импульсная характеристика будет такой, как в (6.39), что и подтверждает наши рассуждения в предыдущем разделе. На рис. 6.14*в* приведен пример для этого случая. С другой стороны, можно использовать физически реализуемое окно, с шириной, большей N, если при этом $w(n)$ обладает следующими свойствами:

$$w(n)=\begin{cases}1/N, & n=r_0N;\\ 0, & n=rN\begin{cases}r\neq r_0\\ r=0,\ \pm1,\ \pm2,\ldots\end{cases}\end{cases} \tag{6.49а}$$

Тогда

$$\tilde{h}(n)=p(n)\,w(n)=\delta(n-r_0N). \tag{6.49б}$$

На рис. 6.14 приведен пример окна конечной длительности, для которого $r_0=2$. На самом деле ясно, что для того, чтобы можно было по $X_n(e^{i\omega})$ восстановить возможно с задержкой $x(n)$, вовсе не обязательно, чтобы $w(n)$ было ограничено во времени или по частоте. Все, что необходимо — это, чтобы для $w(n)$ (6.49а) выполнялось. На рис. 6.14*д* приведен пример окна бесконечной длительности с подходящими свойствами.

Смысл (6.49б) заключается в том, что общая частотная характеристика системы анализ — синтез (рис 6.12), должна иметь

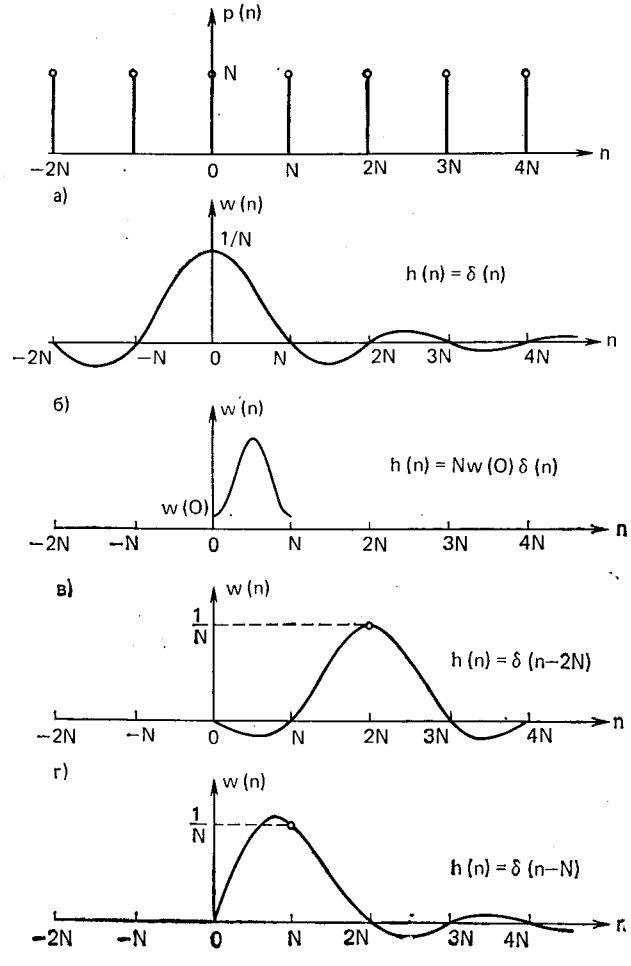


Рис. 6.14. Типичные последовательности $p(n)$ и $w(n)$ гребенки фильтров

плоскую амплитудно-частотную характеристику и линейную фазочастотную, соответствующую задержке на r_0N отсчетов, т. е.

$$\tilde{H}(e^{i\omega}) = e^{-i\omega r_0 N}. \qquad (6.50)$$

Отсюда следует, что выходом системы анализ — синтез служит

$$y(n) = x(n - r_0 N). \qquad (6.51)$$

Таким образом, если не считать задержки в r_0 отсчетов, выход системы кратковременного анализа и синтеза Фурье представляет собой точную копию входной последовательности. Мы показали, следовательно, что возможно точное восстановление входного сигнала при числе частотных каналов, меньшем, чем необходимо по теореме отсчетов, и с помощью физически реализуемого окна, позволяющего реализовать анализ на полосовых фильтрах или фильтрах нижних частот. Возникает поэтому практически важный вопрос о том, насколько точно проектируемый фильтр будет аппроксимировать характеристики, приведенные на рис. 6.14. Этот вопрос рассмотрен в § 6.2.

Прежде чем перейти к другому методу синтеза по кратковременному спектру, нам следует обсудить способ практической реализации (6.40) (уравнение синтеза), поскольку было показано, что достаточно вычислить $X_n(e^{i\omega_k})$ с частотой, определяемой шириной полосы окна. Допустим, что сигнал вычисляется в k-м канале только для каждого D_k-го отсчета входного сигнала. (Для равноразнесенных каналов $D_k = D$ независимо от k.) Предположив, что $X_n(e^{i\omega_k})$ вычисляется с частотой отсчетов входного сигнала (хотя фактически этого может и не быть), можно переделать рис. 6.9*а* и *б*, включив прореживатель на входе анализатора и интерполятор на выходе синтезатора, так, как это показано на рис. 6.15. Этим отображается утверждение о том, что $X_n(e^{i\omega_k})$ можно дискретизовать с частотой F_s/D_k. Как уже отмечалось в гл. 2, прореживание заключается просто в том, что выкидываются $D_k - 1$

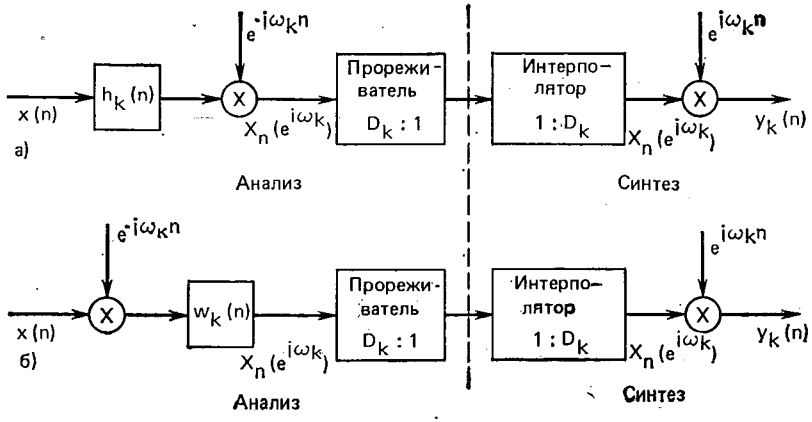


Рис. 6.15. Реализация одного канала кратковременного спектрального анализа с гребенкой фильтров

отсчетов из каждых D_k, или, что то же самое, $X_n(e^{i\omega_k})$ вычисляется через каждые D_k отсчетов. Интерполяция реализуется заполнением нулями $D_k - 1$ отсчетов между значениями $X_n(e^{i\omega_k})$, полученными с уменьшенной частотой дискретизации и последующей фильтрацией в фильтре нижних частот.

6.1.5. Кратковременный синтез методом суммирования с наложением

Альтернативный метод восстановления $x(n)$ по кратковременному спектру основан на интерпретации кратковременного спектра посредством обычного преобразования Фурье. Поскольку $X_n(e^{i\omega_k})$ можно рассматривать как обычное дискретное преобразование Фурье последовательности

$$y_n(m) = x(m)\, w(n - m), \qquad (6.52)$$

то, следовательно, можно восстановить $x(m)$, вычислив обратное дискретное преобразование Фурье от $X_n(e^{i\omega_k})$ и разделив затем на окно (в предположении, что оно не равно нулю при всех m). Этим способом можно вычислить L значений сигнала $x(m)$, где L — длительность окна. Затем окно сдвигается на L отсчетов и процесс повторяется. Из рассуждений, приведенных в 6.1.3, можно понять, что в этой процедуре используется представление $X_n(e^{i\omega_k})$ с пониженной частотой дискретизации и, следовательно, оно весьма чувствительно к ошибкам наложения. Поэтому, несмотря на то, что такая процедура обоснована, она не нашла сколько-нибудь широкого применения там, где важно восстановить исходный (или преобразованный) сигнал. В этом разделе приведена более устойчивая процедура, сходная с методом вычисления периодической свертки на основе дискретного преобразования Фурье.

Допустим, что кратковременное преобразование дискретизируется с периодом, равным R отсчетов, во временной области, т. е. пусть $Y_r(e^{i\omega \hat{k}}) = X_{rR}(e^{i\omega \hat{k}})$ где r — целое и $0 \leqslant k \leqslant N-1$. Метод суммирования с наложением основан на соотношении

$$y(n) = \sum_{r=-\infty}^{\infty} \left[\frac{1}{N} \sum_{k=0}^{N-1} Y_r(e^{i\omega_k}) e^{i\omega_k n} \right]. \qquad (6.53)$$

Иначе говоря, сигнал получается вычислением для каждого значения r обратного преобразования $Y_r(e^{i\omega_k})$, что дает последовательность

$$y_k(m) = x(m)\, w(rR - m), \quad -\infty < m < \infty. \qquad (6.54)$$

Затем получают сигнал в момент n, суммируя значения всех последовательностей $y_r(m)$, перекрывающихся в этот момент. Таким образом,

$$y(n) = \sum_{r=-\infty}^{\infty} y_r(n) = x(n) \sum_{r=-\infty}^{\infty} w(rR - n). \qquad (6.55)$$

Легко показать (см. задачу 6.8), что если преобразование Фурье от $w(n)$ ограничено по частоте и если $X_n(e^{i\omega_k})$ дискретизирована надлежащим образом по частоте, т. е. если R достаточно мала, чтобы избежать наложения[1] во времени, то

$$\sum_{r=-\infty}^{\infty} w(rR-n) \approx W(e^{i0})/R \qquad (6.56)$$

независимо от n.

Поэтому (6.55) переходит в

$$y(n) = x(n)\, W(e^{i0})/R. \qquad (6.57)$$

Это показывает, что правило синтеза (6.53) приводит к точному восстановлению $x(n)$ с точностью до множителя при суммировании перекрывающихся во времени сегментов сигнала.

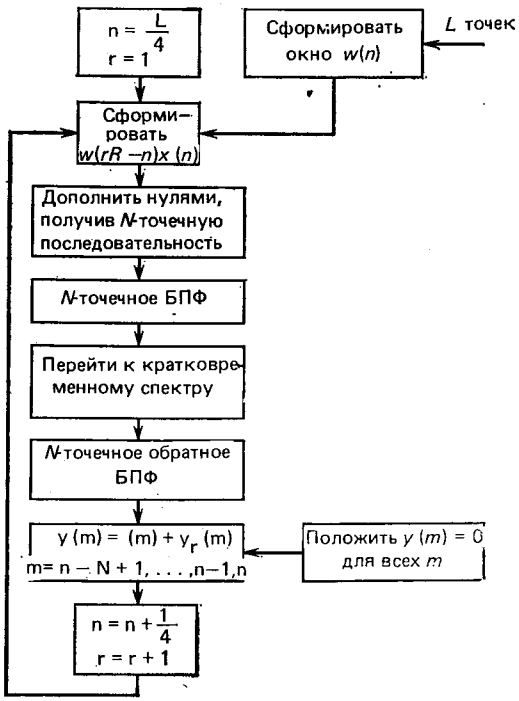


Рис. 6.16. Схема метода суммирования с наложением

На рис. 6.16 и 6.17 показано, как метод суммирования с наложением реализуется для L-точечного окна Хемминга с $R=L/4$. На рис. 6.16 приведена блок-схема алгоритма в предположении, что $x(n)=0$ при $n<0$. Поскольку для окна Хемминга требуется перекрытие по времени 4 : 1, то для того, чтобы получить правильные начальные условия, начало первого анализируемого сегмента принято равным $L/4$ (рис. 6.17). Используемое окно (предполагаемое отличным от нуля при $0\leqslant n\leqslant L-1$) дает сигнал $y_r(m) = w(rR-m)x(m)$, отличный от нуля для $rR-L+1\leqslant m\leqslant rR$. Последовательность длиной L дополняется нулями для получения нужного изменения кратковременного спектра (см. следующий раздел), и затем к результирующей последовательности применяется N-точечное БПФ, дающее $Y_r(e^{i\omega_k})$.

Для восстановления сигнала в момент n можно воспользоваться выражением (6.53). На рис. 6.17 показаны операции, необходимые для вычисления (6.53) при $0\leqslant n\leqslant R-1$. Заметим, что $y(n)$ представляется суммой четырех членов

$$y(n) = x(n)w(R-n) + x(n)w(2R-n) + x(n)w(3R-n) + x(n)w(4R-n). \qquad (6.58)$$

При $R\leqslant n\leqslant 2R-1$ член $x(n)w(R-n)$ следует заменить членом $x(n)w(5R-n)$ и т. д.

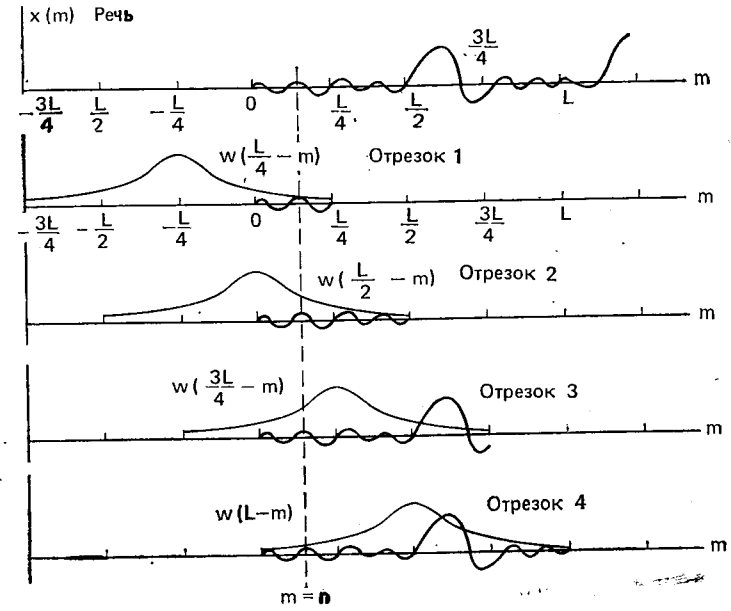


Рис. 6.17. Восстановление $w(n)$ с L-точечным окном Хемминга

Метод суммирования выходов гребенки фильтров и метод суммирования с наложением по сути дела дуальны: один связан с дискретизацией по частоте, а другой — с дискретизацией во времени. Метод суммирования выходов гребенки фильтров требует, чтобы при дискретизации по частоте преобразование окна удовлетворяло соотношению

$$\sum_{k=0}^{N-1} W\left(e^{i(\omega-\omega_k)}\right) = w(0), \qquad (6.59а)$$

в то время как в методе суммирования с наложением требуется, чтобы при дискретизации по времени окно удовлетворяло равенству

$$\sum_{r=-\infty}^{\infty} w(rR-n) = W(e^{i0})/R. \qquad (6.59б)$$

Дуальность (6.59а и б) очевидна.

[1] Для L-точечного окна Хемминга $R\leqslant L/4$.

Для того чтобы сопоставить эти два метода восстановления сигнала по кратковременному преобразованию Фурье, ниже рассматривается влияние изменений кратковременного спектра на результат синтеза.

6.1.6. Влияние преобразований кратковременного спектра на синтез

Итак, имеется два метода восстановления сигнала по его кратковременному спектру. Оба метода позволяют точно восстановить исходный сигнал с точностью до множителя, если кратковременный спектр должным образом дискретизован как по частоте, так и по времени. Иногда требуется изменить кратковременный спектр для того, чтобы произвести постоянную или динамичную (т. е. переменную во времени) фильтрацию анализируемого сигнала. Рассмотрим влияние фиксированных или переменных во времени преобразований кратковременного спектра на результат синтеза.

Метод суммирования выходов гребенки фильтров (СГФ). Представим фиксированное преобразование кратковременного спектра в виде

$$\hat{X}_n(e^{i\omega_k}) = X_n(e^{i\omega_k})\,P(e^{i\omega_k}), \tag{6.60}$$

где $P(e^{i\omega_k})$ — весовая частотная функция кратковременного спектра. Предположим, что существует обратное дискретное преобразование Фурье от $P(e^{i\omega_k})$, и обозначим эту последовательность через $p(n)$:

$$p(n) = \frac{1}{N}\sum_{k=0}^{N-1} P\left(e^{i\omega_k}\right) e^{i\omega_k n}, \tag{6.61}$$

где N — количество частот, на которых вычисляется $P(e^{i\omega_k})$. Сигнал, восстановленный методом СГФ, получается подстановкой (6.60) в (6.40):

$$\hat{y}(n) = \sum_{k=0}^{N-1} X_n\left(e^{i\omega_k}\right) P\left(e^{i\omega_k}\right) e^{i\omega_k n} = \sum_{k=0}^{N-1}\left[\sum_{m=-\infty}^{\infty} w(n-m)\,x(m)\,e^{-i\omega_k m}\right] P\left(e^{i\omega_k}\right) e^{i\omega_k n} = \sum_{m=-\infty}^{\infty} w(n-m)\,x(m) \times$$

$$\times \sum_{k=0}^{N-1} P\left(e^{i\omega_k}\right) e^{i\omega_k (n-m)} = \sum_{m=-\infty}^{\infty} w(n-m)\,x(m)\,N\,p(n-m) =$$

$$= N\,x(n) * [w(n)\,p(n)]. \tag{6.62}$$

Следовательно, эффект постоянного во времени преобразования с $P(e^{i\omega_k})$ сводится к свертке сигнала $x(n)$ с произведением окна $w(n)$ и периодической последовательности $p(n)$. Преобразование кратковременного спектра вида (6.60) применяется тогда, когда

требуется линейно отфильтровать $x(n)$. Можно, например, потребовать, чтобы

$$w(n)\,p(n) = h_p(n) \tag{6.63}$$

была импульсной характеристикой линейного фильтра. Последовательность $p(n)$ представляет собой периодическую последовательность, и если длительность $w(n)$ не превосходит N, то структура $h_p(n)$ окажется повторяющейся. В следующем разделе будет показано, что преобразования такого вида возникают при построении гребенок БИХ-фильтров. Таким образом, в методе суммирования выходов гребенки фильтров окно существенно влияет на фиксированное преобразование спектра. Справедливо лишь приближенное равенство

$$h_p(n) \approx p(n), \quad 0 \leqslant n \leqslant N-1, \tag{6.64}$$

да и то только в случае, когда $p(n)$ сильно сконцентрирована или когда используется прямоугольное окно.

Для переменного во времени преобразования запишем $\hat{X}_n(e^{i\omega_k})$ в виде

$$\hat{X}_n(e^{i\omega_k}) = X_n(e^{i\omega_k})\,P_n(e^{i\omega_k}) \tag{6.65}$$

и определим переменную во времени импульсную характеристику $p_n(m)$ так:

$$p_n(m) = \frac{1}{N}\sum_{k=0}^{N-1} P_n\left(e^{i\omega_k}\right) e^{i\omega_k m}. \tag{6.66}$$

Поступая как и раньше, получим

$$\hat{y}(n) = \sum_{k=0}^{N-1} X_n\left(e^{i\omega_k}\right) P_n\left(e^{i\omega_k}\right) e^{i\omega_k n} = \sum_{k=0}^{N-1} e^{-i\omega_k n} \sum_{m=-\infty}^{\infty} x(n-m)\,w(m)\,e^{i\omega_k m} P_n\left(e^{i\omega_k}\right) e^{i\omega_k n} = \sum_{m=-\infty}^{\infty} x(n-m)\,w(m) \times$$

$$\times \sum_{k=0}^{N-1} P_n\left(e^{i\omega_k}\right) e^{i\omega_k m} = \sum_{m=-\infty}^{\infty} x(n-m)\,w(m)\,N\,p_n(m) = N\sum_{m=-\infty}^{\infty} x(n-m)\,[p_n(m)\,w(m)]. \tag{6.67}$$

Равенства (6.67) снова показывают, что в методе СГФ временна́я характеристика преобразования спектра взвешивается окном до свертывания с $x(n)$.

Итак, в методе суммирования выходов гребенки фильтров влияние преобразования спектра (постоянного или переменного во времени) сводится к свертыванию исходного сигнала с взвешенной временно́й характеристикой преобразования.

Метод суммирования с наложением (СН). Воспользовавшись (6.60), с помощью (6.53) можно выразить восстановленный сигнал в виде

$$\hat{y}(n)=\sum_{r=-\infty}^{\infty}\frac{1}{N}\sum_{k=0}^{N-1}Y_r\left(e^{i\omega_k}\right)P\left(e^{i\omega_k}\right)e^{i\omega_k n}=\frac{1}{N}\sum_{r=-\infty}^{\infty}\sum_{k=0}^{N-1}\sum_{l=-\infty}^{\infty}x(l)\,w\times$$

$$\times(rR-l)\,e^{-i\omega_k l}P\left(e^{i\omega_k}\right)e^{i\omega_k n}=\frac{1}{N}\sum_{l=-\infty}^{\infty}x(l)\left[\sum_{k=0}^{N-1}P\left(e^{i\omega_k}\right)e^{i\omega_k(n-l)}\right]\times$$

$$\times\left[\sum_{r=-\infty}^{\infty}w(rR-l)\right]=\sum_{l=-\infty}^{\infty}x(l)\,p(n-l)\,W(e^{i0})/R \qquad (6.68)$$

или

$$\hat{y}(n)=(1/R)\,W(e^{i0})\,[x(n)*p(n)]. \qquad (6.69)$$

Соотношение (6.69) показывает, что $\hat{y}(n)$ представляет собой свертку исходного сигнала с временно́й характеристикой преобразования спектра, т. е. в этом методе окно не изменяет последовательности $p(n)$[1]. (Читателю следует отдавать себе отчет в том, что схему, изображенную на рис. 6.16, следует видоизменить должным образом — дополнить нулями сигнал, — чтобы избежать наложений при реализации операций анализа и синтеза посредством БПФ).

Для случая переменного преобразования получим

$$\hat{y}(n)=\sum_{r=-\infty}^{\infty}\frac{1}{N}\left[\sum_{k=0}^{N-1}Y_r\left(e^{i\omega_k}\right)P_r\left(e^{i\omega_k}\right)\right]e^{i\omega_k n}, \qquad (6.70)$$

что можно преобразовать к виду

$$\hat{y}(n)=\frac{1}{N}\sum_{l=-\infty}^{\infty}x(l)\sum_{r=-\infty}^{\infty}w(rR-l)\left[\sum_{k=0}^{N-1}P_r\left(e^{i\omega_k}\right)e^{i\omega_k(n-l)}\right]. \qquad (6.71)$$

Из (6.66) получим

$$\hat{y}(n)=\sum_{l=-\infty}^{\infty}x(l)\sum_{r=-\infty}^{\infty}w(rR-l)\,p_r(n-l). \qquad (6.72)$$

Положив $q=n-l$ или $l=n-q$, получим для (6.72)

$$\hat{y}(n)=\sum_{q=-\infty}^{\infty}x(n-q)\sum_{r=-\infty}^{\infty}p_r(q)\,w(rR-n+q). \qquad (6.73)$$

Если определить $\hat{p}$ как

$$\hat{p}(n-q,q)=\hat{p}(m,q)=\sum_{r=-\infty}^{\infty}p_r(q)\,w(rR-m), \qquad (6.74)$$

[1] Из-за вычисления выхода блоками по N отсчетов в (6.68), $p(n)$ не периодична и имеет максимальную длительность, равную N отсчетам.

соотношение (6.72) перейдет в

$$\hat{y}(n)=\sum_{q=-\infty}^{\infty}x(n-q)\,\hat{p}(n-q,q). \qquad (6.75)$$

Интерпретация (6.74) такова: для q-го значения $\hat{p}(m,q)$ представляет свертку $p_r(q)$ и $w(r)$. Поэтому каждый из коэффициентов временно́й характеристики преобразования сглаживается окном (фильтруется фильтром нижних частот). Следовательно, в методе суммирования с наложением любое преобразование ограничивается по частоте окном, но соответствует обычной свертке. Это прямо противоположно методу суммирования выходов гребенки фильтров, в котором преобразование ограничивается во времени окном и может мгновенно меняться.

6.1.7. Аддитивное преобразование

Мы рассмотрели влияние неслучайного мультипликативного преобразования кратковременного спектра. Важно также понимать влияние аддитивного, не зависящего от сигнала (случайного) изменения кратковременного спектра, такого, какое может возникнуть при реализации анализа с конечной точностью вычисления (шумы округления) или при квантовании спектра, как это происходит в вокодерах. Запишем такое аддитивное преобразование спектра в виде

$$\hat{X}_n\left(e^{i\omega_k}\right)=X_n\left(e^{i\omega_k}\right)+E_n\left(e^{i\omega_k}\right), \qquad (6.76)$$

где шумовая последовательность $E_n(e^{i\omega_k})$ определена так:

$$e(n)=\sum_{k=0}^{N-1}E_n\left(e^{i\omega_k}\right)e^{i\omega_k n}. \qquad (6.77)$$

В случае, когда $e(n)$ представляет собой случайный шум, требуется статистическая модель для $e(n)$ и $E(e^{i\omega_k})$. Приводимые ниже результаты не зависят от природы модели. В методе СГФ эффект аддитивного преобразования (6.76) сводится к

$$\hat{y}(n)=\sum_{k=0}^{N-1}\left[X_n\left(e^{i\omega_k}\right)+E_n\left(e^{i\omega_k}\right)\right]e^{i\omega_k n}, \qquad (6.78)$$

что, ввиду линейности, можно записать следующим образом:

$$\hat{y}(n)=y(n)+\sum_{k=0}^{N-1}E_n\left(e^{i\omega_k}\right)e^{i\omega_k n} \qquad (6.79)$$

или

$$\hat{y}(n)=y(n)+e(n). \qquad (6.80)$$

Следовательно, аддитивное преобразование спектра приводит к появлению аддитивной компоненты в восстановленном сигнале. Следует заметить, что «анализирующее» окно не оказывает прямого влияния на аддитивные члены при синтезе.

Для СН метода аддитивное преобразование (6.76) сводится к

$$\hat{y}(n)=\sum_{r=-\infty}^{\infty}\frac{1}{N}\sum_{k=0}^{N-1}\left(Y_r\left(e^{i\omega_k}\right)+E_r\left(e^{i\omega_k}\right)\right)e^{i\omega_k n}, \qquad (6.81)$$

что можно переписать в виде

$$\hat{y}(n)=y(n)=\sum_{r=-\infty}^{\infty}\left[\frac{1}{N}\sum_{k=0}^{N-1}E_r\left(e^{i\omega_k}\right)e^{i\omega_k n}\right]=y(n)+\sum_{r=-\infty}^{\infty}e_r(n). \qquad (6.82)$$

Следовательно, результат синтеза содержит бóльшую аддитивную (шумовую) компоненту в случае СН метода по сравнению с методом СГФ из-за перекрытий между сегментами анализа. Для окна Хемминга с перекрытием 4 : 1 аддитивная добавка будет примерно в 4 раза больше при синтезе методом СН в сравнении с методом СГФ. Таким образом, метод СН оказывается более чувствительным к ошибкам вычислений, чем СГФ, и, следовательно, менее полезен для применения в вокодерах и т. п.

6.1.8. Обзор методов кратковременного анализа и синтеза речи

В этой главе было показано, что полезное определение кратковременного преобразования Фурье имеет вид

$$X_n(e^{i\omega})=\sum_{m=-\infty}^{\infty}w(n-m)\,x(m)\,e^{-i\omega m},$$

где $w(n)$ — анализирующее окно, выделяющее часть входного сигнала в момент n. Показано также, что $X_n(e^{i\omega})$ можно интерпретировать посредством линейной фильтрации, как выход полосового фильтра с близкой к нулю полосой пропускания или, равным образом, как обычное преобразование Фурье последовательности $w(n-m)\,x(m)$.

Показано также, что можно определить частоты дискретизации по времени и частоте, основываясь на теореме отсчетов и рассматривая представление окна в частотной и временно́й областях. Требуемые частоты надлежащим образом дискретизированного кратковременного спектрального представления оказались в 2—4 раза выше, чем в эквивалентном представлении самого сигнала во временно́й области.

На основе двух интерпретаций кратковременного анализа были введены две различные процедуры синтеза. В первом методе, названном суммированием выходов гребенки фильтров, сигнал синтезируется так:

$$y(n)=\sum_{k=0}^{N-1}X_n\left(e^{i\omega_k}\right)e^{i\omega_k n},$$

т. е. выходной сигнал представляет собой сумму сигналов гребенки фильтров, смещенных к центральным частотам полос.

Второй метод синтеза, названный методом суммирования с наложением, приводит к следующей процедуре синтеза:

$$y(n)=\sum_{r=-\infty}^{\infty}\frac{1}{N}\sum_{k=0}^{N-1}Y_r\left(e^{i\omega_k}\right)e^{i\omega_k n},$$

где $Y_r(e^{i\omega_k})=X_{rR}(e^{i\omega_k})$, т. е. взвешенные окном сегменты, разнесенные по времени на R отсчетов, суммируются, образуя восстановленный сигнал. Эти два метода синтеза обладают двойственными свойствами, дуальными как по отношению к синтезу, так и по реакции на преобразование кратковременного спектра.

На этом мы заканчиваем формальное обсуждение общих свойств методов кратковременного анализа и синтеза речи. В последующих параграфах внимание сконцентрировано на методах проектирования цифровых гребенок фильтров для случая пониженной частоты дискретизации, например для вокодера, и обсуждаются некоторые из многочисленных приложений теории кратковременного анализа и синтеза к обработке речи.

6.2. Проектирование гребенок цифровых фильтров

Ниже будут рассмотрены некоторые практические методы проектирования цифровых фильтров, используемых в методе суммирования выходов гребенки фильтров. Нашей целью будет проектирование гребенки фильтров с общей частотной характеристикой, хорошо аппроксимирующей идеальную — с плоской амплитудно-частотной и линейной фазо-частотной частями. Начнем с некоторых подробностей, общих для проектирования гребенок фильтров, независимо от типа используемых фильтров. Затем приведем примеры использования БИХ- и КИХ-фильтров.

6.2.1. Соображения практического характера

Из ранее изложенного теоретического анализа следует, что достигнуть точного воспроизведения входного сигнала на выходе гребенки равномерно разнесенных фильтров можно, когда нули фильтра нижних частот (или анализирующего окна $w(n)$) равномерно разнесены с интервалом, равным N отсчетам. Казалось бы, что при проектировании такой гребенки нужно сначала выбрать количество фильтров, а следовательно, разнесение по частоте, а затем спроектировать фильтр нижних частот с надлежащим разрешением по частоте и разносом нулей во временно́й области. К сожалению, возникает ряд соображений практического характера, усложняющих эту процедуру. Во-первых, часто желательно использовать неравномерное разнесение фильтров, поэтому данное выше доказательство того, что можно получить безупречную общую характеристику, не дает каких-либо указаний по поводу практического решения. Во-вторых, некоторая часть спектра на практике часто анализу не подвергается. Следовательно, необходимо рассмотреть, как повлияет отсутствие фильтра на общую характеристику. И, наконец, большинство процедур проектирования фильтров нижних частот не допускает введения одновременных ограни-

чений на частотную и временну́ю характеристики. Поэтому может оказаться невозможным получить и требуемое разрешение по частоте и необходимое распределение нулей импульсной характеристики.

Для того чтобы представить в простом виде некоторые из таких специальных соображений, а также предвосхищая разрешение некоторых из указанных трудностей, полезно рассмотреть несколько модифицированную структуру гребенки фильтров[1], в которой в каждом из каналов (см. рис. 6.11 или равным образом рис. 6.12) комплексный сигнал умножается на комплексную величину, обозначенную через $P_k = |P_k| e^{i\Phi_k}$. Для одного канала этот случай показан на рис. 6.18. Заметим, что поскольку каждый из каналов представляет собой линейную систему, умножение можно отнести как ко входу, так и к выходу. На рис. 6.18*б* и *в* комплексная кон-

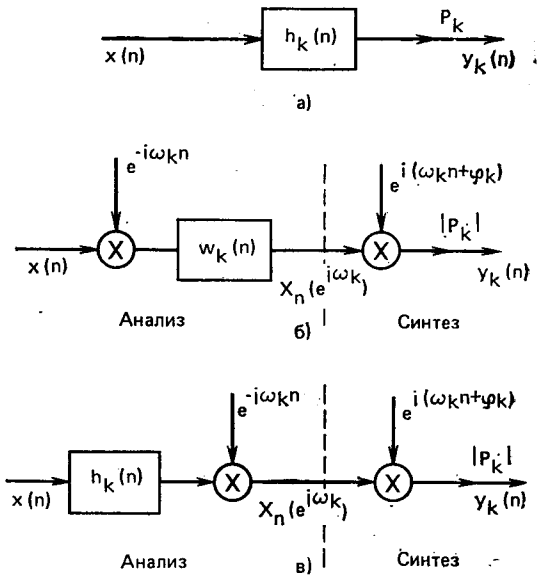


Рис. 6.18. Канал модифицированной гребенки фильтров:
а) эквивалентный полосовой фильтр для процедуры анализ — синтез; *б*) реализация с модулятором и фильтром нижних частот; *в*) реализация с модулятором и полосовым фильтром

станта включена в часть канала, относящуюся к синтезу. В другом случае, когда умножение выполняется на входе, выходом блока анализа будет $P_k X_n(e^{i\omega_k})$. В каждом из этих случаев, рассмотрев общий выход системы N комплексных каналов, получим

$$y(n) = \sum_{k=0}^{N-1} P_k y_k(n) = \sum_{k=0}^{N-1} P_k X_n\left(e^{i\omega_k}\right) e^{i\omega_k n}. \qquad (6.83)$$

[1]) Эта модификация относится к преобразованиям общего вида, рассмотренным в 6.1.6.

Общая импульсная характеристика системы имеет вид

$$\tilde{h}(n) = \sum_{k=0}^{N-1} P_k h_k(n) = \sum_{k=0}^{N-1} |P_k| w_k(n) e^{i(\omega_k n+\Phi_k)}. \qquad (6.84)$$

Таким образом, в терминах кратковременного анализа — синтеза Фурье $X_n(e^{i\omega_k})$ взвешивается комплексной последовательностью $\{P_k\}$, $k=0, 1, \ldots, N-1$. В случае гребенки фильтров комплексные константы P_k позволяют изменять усиление и фазу отдельных фильтров гребенки.

Первый шаг в проектировании системы с гребенкой фильтров (или системы кратковременного анализа — синтеза Фурье) состоит в выборе частот анализа $\{\omega_k\}$ для $0 \leqslant k \leqslant N-1$. При их выборе обычно руководствуются требованиями разрешения по частоте. Например, если требуется, чтобы основная частота и ее гармоники оказались разделенными, понадобятся близкие частоты анализа и полосовые или фильтры нижних частот с достаточно узкими полосами. Во многих случаях требуются равномерно разнесенные частоты анализа и фильтры с одинаковой шириной полосы. Однако иногда приходится использовать неравномерное распределение частот анализа. Так будет, например, в случае системы обработки речи, в которой используется ухудшение чувствительности слуха на высоких частотах. Обычно выбирают симметричный набор частот в диапазоне $0 \leqslant \omega \leqslant 2\pi$, т. е. так, что $\omega_{N-k} = 2\pi - \omega_k$. Это показано на рис. 6.19 для четных и нечетных N. Заметим, что для четных N имеется канал с центральной частотой $\omega = \pi$. Если к тому же $w_k(n) = w_{N-k}(n)$, то, как и ранее,

$$X_n(e^{i\omega_k}) = X_n^*(e^{i(2\pi-\omega_k)}) = X_n^*(e^{-i\omega_k}), \qquad (6.85)$$

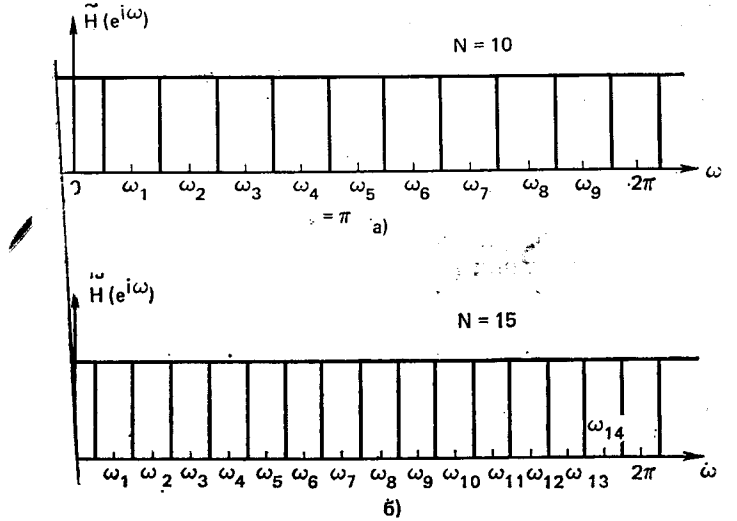


Рис. 6.19. Распределение частот анализа для гребенки фильтров с четным (*а*) и нечетным (*б*) N

что приводит к упрощению гребенки фильтров, поскольку здесь необходимы только действительные полосовые фильтры.

Допустим, что $\omega_k = 2\pi - \omega_{N-k}$, $P_k = P^*_{N-k}$ и $\omega_k(n) = w_{N-k}(n)$. Тогда из (6.84) можно показать, что для четных N

$$\widetilde{h}(n) = P_0 w_0(n) + \sum_{k=1}^{\frac{N}{2}-1} 2|P_k| w_k(n) \cos(\omega_k n + \Phi_k) + P_{N/2} w_{N/2} \times \times (n)(-1)^n, \quad (6.86)$$

а для N нечетных

$$\widetilde{h}(n) = P_0 w_0(n) + \sum_{k=1}^{\frac{N-1}{2}} 2|P_k| w_k(n) \cos(\omega_k n + \Phi_k). \quad (6.87)$$

Поэтому можно ограничиться гребенкой, состоящей из фильтров нижних частот $P_0 w_0(n)$ и набора полосовых фильтров с действительными импульсными характеристиками:

$$h_k(n) = 2|P_k| w_k(n) \cos(\omega_k n + \Phi_k). \quad (6.88)$$

Если N четное, то потребуется дополнительный фильтр высоких частот с центральной частотой $\omega = \pi$; тогда окажется перекрытым весь диапазон частот $0 \leqslant \omega < 2\pi$. Импульсная характеристика этого фильтра должна быть равной $P_{N/2} w_{N/2}(n)(-1)^n$.

За исключением того, что частоты анализа должны быть размещены симметрично на интервале $0 \leqslant \omega < 2\pi$, нет никаких других ограничений на частоты $\{\omega_k\}$. Коль скоро эти частоты выбраны, нам нужно найти соответствующий набор фильтров нижних частот или анализирующих окон $\{w_k(n)\}$ с требуемым частотным разрешением и нужной общей характеристикой.

Часто необходимо вычислять $X_n(e^{i\omega_k})$ только на частотах некоторого поддиапазона основной полосы $0 \leqslant \omega < 2\pi$. Например, часто не проводят анализа на частоте $\omega_k = 0$, поскольку эта часть речевого спектра не представляет интереса в большинстве систем обработки речи. Аналогичным образом опускают анализ на высоких частотах (ω_k, близких к π), поскольку в процессе ограничения спектра входного сигнала, предшествующего дискретизации, эта часть спектра сильно ослаблена и поэтому несет мало достоверной информации.

Чтобы оценить зависимость общей характеристики от пропуска каналов, вернемся к случаю разноразнесенных частот анализа. Предположив, что имеются идентичные анализирующие окна, получим из (6.84)

$$\widetilde{h}(n) = w(n) \sum_{k=0}^{N-1} P_k e^{i2\pi kn/N}. \quad (6.89)$$

Следовательно, если определить

$$p(n) = \sum_{k=0}^{N-1} P_k e^{i2\pi kn/N}, \quad (6.90)$$

то

$$\widetilde{h}(n) = w(n)\, p(n) \quad (6.91)$$

в соответствии с (6.89). Видно, что, как и раньше, последовательность $p(n)$ периодична с периодом N. В действительности комплексные коэффициенты усиления в каналах играют роль коэффициентов ряда Фурье. Заметим, что в случае, когда $P_k = 1$, $0 \leqslant k \leqslant N-1$, все каналы оказываются включенными и (6.89) становится идентичным (6.42). Эффект пропуска каналов весьма удобно наблюдать, положив соответствующие P_k равными нулю. Для того чтобы исключить, например, нулевую частоту, можно положить $P_0 = 0$. Чтобы, кроме того, исключить каналы с частотами выше $\omega_M = 2\pi M/N$, положим $P_k = 0$ для $k > M$. В этом случае

$$p(n) = \sum_{k=1}^{M} e^{i\frac{2\pi}{N}kn} + \sum_{k=N-M}^{N-1} e^{i\frac{2\pi}{N}kn} = \sum_{k=1}^{M} \left[e^{i\frac{2\pi}{N}kn} + e^{-i\frac{2\pi}{N}kn} \right]. \quad (6.92)$$

Это можно выразить в более компактной форме

$$p(n) = \frac{\sin\left[\frac{\pi}{N}(2M+1)n\right]}{\sin\left[\frac{\pi}{N}n\right]} - 1. \quad (6.93)$$

В качестве примера на рис. 6.20 изображена последовательность (6.93) для случая $N = 15$ и $M = 2$. Ясно, что здесь $p(n)$ периодична

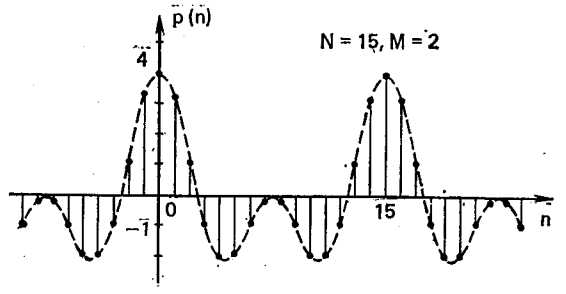


Рис. 6.20. График $p(n)$ для гребенки фильтров с $N = 15$, $M = 2$

с периодом 15, но вместо одного единичного импульса через каждые 15 отсчетов $p(n)$ теперь содержит импульсы с амплитудой и шириной, зависящими от N и M. Можно понять, что если канал на нулевой частоте будет включен, то член -1 в правой части (6.93) исчезнет. Кроме того, для нечетного N при наличии всех каналов будет $M = (N-1)/2$, так что $p(n)$ можно выразить в виде

$$p(n) = \frac{\sin(\pi n)}{\sin(\pi n/N)} = N \sum_{r=-\infty}^{\infty} \delta(n - rN). \quad (6.94)$$

Таким образом, только в случае, когда имеются все каналы, последовательность $p(n)$ может быть такой, что окна́ $w(n)$ с разнесенными на интервалы в N отсчетов нулями достаточно для выполнения равенства

$$\widetilde{h}(n)=\delta(n-r_0 N). \tag{6.95}$$

Это, разумеется, оправдано, поскольку присутствует не весь частотный спектр. Разумно предположить, что в области включенных фильтров общая амплитудно-частотная характеристика будет плоской, а общая фазо-частотная — линейной. Выбрасывание канала на нулевой частоте и каналов на высших частотах эквивалентно полосовой фильтрации. Это подтверждается на последующих примерах.

Многие из стандартных методов проектирования фильтров не допускают введения одновременных ограничений на частотную и временну́ю характеристики. Следовательно, может оказаться невозможным получить фильтр нижних частот с импульсной характеристикой, обращающейся в нуль через каждые N отсчетов. Чтобы уяснить влияние этого обстоятельства, рассмотрим рис. 6.21.

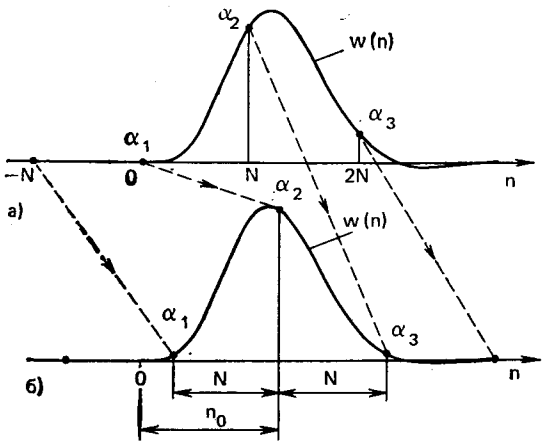


Рис. 6.21. Иллюстрация подстройки параметра n_0:
а) общая импульсная характеристика для $n_0=0$; *б*) n_0 выбрано так, чтобы минимизировать амплитудную и фазовую ошибки (пунктиром показано перемещение отдельных импульсов) [1]

Для удобства здесь принято, что задействованы все каналы так, что $p(n)$ представляет собой последовательность единичных импульсов с периодом N. Огибающая последовательности $w(n)$ показана как непрерывная кривая. Произведение $p(n)$ и $w(n)$ представлено отсчетами, обозначенными через α_1, α_2, α_3 и т. д. Ясно, что в этом случае для общей характеристики имеем приближенно

$$\widetilde{h}(n)=\alpha_2\delta(n-N)+\alpha_3\delta(n-2N), \tag{6.96}$$

если пренебречь импульсами в моменты $3N$, $4N$ и т. д. Общая частотная характеристика будет иметь вид

$$\widetilde{H}(e^{i\omega})=\alpha_2 e^{-i\omega N}+\alpha_3 e^{-i\omega 2N}=\alpha_2 e^{-i\omega N}\left(1+\frac{\alpha_3}{\alpha_2}e^{-i\omega N}\right)=$$
$$=\alpha_2 e^{-i\omega N}\widetilde{G}(e^{i\omega}), \tag{6.97}$$

где

$$\widetilde{G}(e^{i\omega})=1+(\alpha_3/\alpha_2)e^{-i\omega N} \tag{6.98}$$

задает отклонение общей частотной характеристики от $\alpha_2 e^{-i\omega N}$. Последняя соответствует точному воспроизведению входного сигнала с задержкой в N отсчетов и при масштабировании коэффитиентом α_2. Амплитуда и фаза $\widetilde{G}(e^{i\omega})$ задаются соответственно

$$\left|\widetilde{G}(e^{i\omega})\right|=\left[1+\left(\frac{\alpha_3}{\alpha_2}\right)^2+2\left(\frac{\alpha_3}{\alpha_2}\right)\cos\omega N\right]^{1/2} \tag{6.99}$$

$$\arg[\widetilde{G}(e^{i\omega})]=\mathrm{tg}^{-1}\left[\frac{-(\alpha_3/\alpha_2)\sin\omega N}{1+(\alpha_3/\alpha_2)\cos\omega N}\right]. \tag{6.100}$$

Набросок этих функций для $N=4$ приведен на рис. 6.22. Можно видеть, что модуль общей характеристики имеет множитель ошибки, который, вообще говоря, осциллирует с периодом $2\pi/N$. Это вызвано разносом между фильтрами. Действительно, пики $|\widetilde{G}(e^{i\omega})|$ приходятся на частоты анализа $\omega_k=(2\pi/N)k$, а провалы лежат строго посередине между ними, т. е. в «перекрестных» точках фильтров. Величина пульсаций общей характеристики зависит от соотношения между α_2 и α_3. И для фазы видно уклонение от линейной кривой; оно также периодично с периодом $2\pi/N$. Отметим, что амплитудная и фазовая ошибки исчезают при $\alpha_3=0$ и максимальны при $\alpha_3=\alpha_2$.

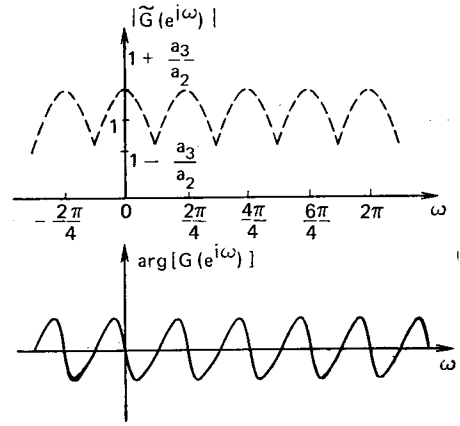


Рис. 6.22. Амплитуда и фаза гребенки фильтров

На рис. 6.21 предлагается несколько подходов к уменьшению этих погрешностей. Одна из возможностей состоит в использовании более узкого окна, что позволяет добиться уменьшения α_3 путем снижения $w(n)$ при $n=2N$. Неприятным эффектом оказывается расширение преобразования Фурье $w(n)$ и, следовательно, сокращение частотного разрешения в отдельных каналах. Вторая возможность заключается в том, чтобы увеличить N, сохранив $w(n)$ без изменений. Третья возможность показана на рис. 6.21*б*. Если сдвинуть бесконечную последовательность импульсов относительно $w(n)$, то получим последовательность из трех импульсов вместо двух:

$$\widetilde{h}(n)=\alpha_1'\delta(n-n_0)+\alpha_2'\delta(n-N-n_0)+\alpha_3'\delta(n-2N-n_0) \tag{6.101}$$

Тогда общей частотной характеристикой будет функция

$$\widetilde{H}(e^{i\omega}) = e^{-i\omega(n_0+N)}[\alpha_1' e^{i\omega N} + \alpha_2' + \alpha_3' e^{-i\omega N}]. \quad (6.102)$$

Если определить, как и раньше,

$$\widetilde{G}'(e^{i\omega}) = \frac{\alpha_1'}{\alpha_2'} e^{i\omega N} + 1 + \frac{\alpha_3'}{\alpha_2'} e^{-i\omega N}, \quad (6.103)$$

то

$$\widetilde{H}(e^{i\omega}) = \alpha_2' e^{-i\omega(n_0+N)} \widetilde{G}'(e^{i\omega}). \quad (6.104)$$

Ясно, что для $\alpha'_1 = \alpha'_3$

$$\widetilde{G}'(e^{i\omega}) = (2\alpha_3'/\alpha_2') \cos \omega N + 1, \quad (6.105)$$

т. е. фазовая ошибка отсутствует. Амплитудная ошибка отложена для $N=4$ (рис. 6.23). В первом случае максимум ошибки равен

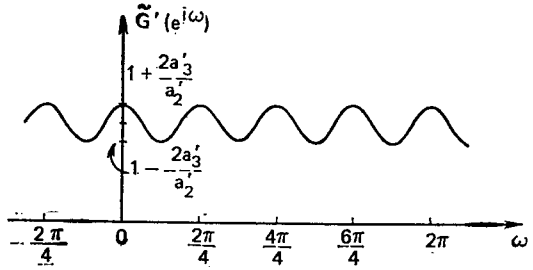


Рис. 6.23. Амплитуда гребенки фильтров, когда $p(n)$ сдвинута относительно $w(n)$

$2\alpha_3/\alpha_2$. Во втором — $4\alpha'_3/\alpha'_2$. Следовательно, в дополнение к тому, что при $\alpha'_1 = \alpha'_3$ отсутствует фазовая ошибка, при

$$2\alpha_3'/\alpha_2' < \alpha_3/\alpha_2 \quad (6.106)$$

и амплитудная ошибка будет меньше.

Механизм сдвига $p(n)$ относительно $w(n)$ подсказывается (6.90). Чтобы сдвинуть бесконечную последовательность единичных импульсов на n_0, положим

$$P_k = e^{-i2\pi k n_0/N}, \quad 0 \leqslant k \leqslant N-1, \quad (6.107)$$

тогда

$$p(n) = \sum_{k=0}^{N-1} e^{i2\pi k(n-n_0)/N} = N \sum_{r=-\infty}^{\infty} \delta(n - n_0 - rN). \quad (6.108)$$

Если, положив некоторые из P_k равными нулю, мы пропускаем соответствующие каналы, то в результате вместо каждого импульса появится сдвинутая последовательность единичных импульсов. Подбирая фазы каналов в соответствии с (6.107), можно переместить $p(n)$ относительно $w(n)$. Регулировку фазы можно производить так, как это показано на рис. 6.18. Рассмотрим далее некоторые примеры гребенок БИХ- и КИХ-фильтров.

6.2.2. Проектирование гребенок с БИХ-фильтрами

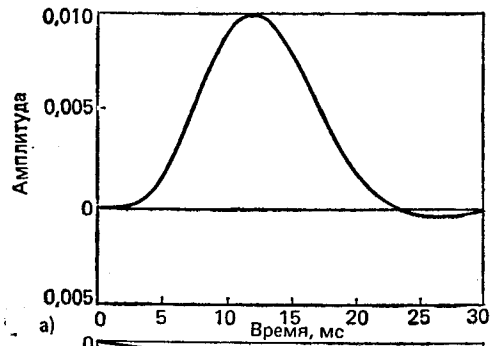
Проектирование гребенок БИХ-фильтров рассмотрено в [1], где показано, что общий принцип подбора фазы в отдельных каналах в соответствии с (6.107) может оказаться полезным для оптимизации.

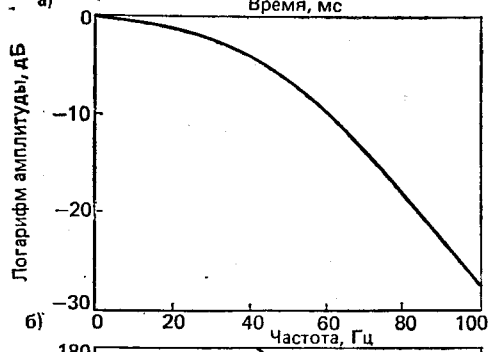
Следующий пример также взят из [1]. Предполагается, что входная последовательность получается дискретизацией с частотой 10 000 отсч./с. Нужно спроектировать гребенку фильтров с равномерным разнесением в 100 Гц (по аналоговой частоте). Отсюда следует, что $N = 10\,000/100 = 100$ и что частоты анализа равны $\omega_k = 2\pi k/100$, $k = 0, 1, \ldots, M$. Область анализа предполагается лежащей между 100 и 3000 Гц. Это означает, что потребуется 30 каналов, т. е. $M = 30$. Плоской общей амплитудно-частотной и линейной фазо-частотной характеристик проще всего добиться, когда отдельные фильтры обладают такими характеристиками. По этой причине используются фильтры Бесселя с максимально-плоской характеристикой. Для этого примера цифровой фильтр нижних частот был получен из фильтра Бесселя шестого порядка методом инвариантной импульсной характеристики. На рис. 6.24 показаны свойства отклика базового для анализа фильтра. Отметим, что 30 мс соответствуют 300 отсчетам при частоте дискретизации 10 кГц. Отметим также, что фаза на рис. 6.24*в* вполне линейна. Номинальная частота среза фильтра равна 60 Гц. Этот фильтр использовался в 30-канальной гребенке фильтров с разнесением между каналами в 100 Гц. Канал на нулевой частоте был опущен. Из (6.93) при $M = 30$ и $N = 100$ имеем

$$p(n) = \frac{\sin[0{,}61\,\pi n]}{\sin[0{,}01\,\pi n]} - 1. \quad (6.109)$$

Отметим, что $p(n)$ периодична с периодом $N = 100$ отсчетов. Результирующая общая характеристика показана на рис. 6.25. На общей импульсной характеристике видно расширение импульсов, что и следовало ожидать, поскольку все каналы участвуют в формировании отклика. Видно также, что из-за того, что продолжительность $w(n)$ превышает $2N = 200$ отсчетов, у $\tilde{h}(n)$ появляются большие пики. Это приводит к значительным пульсациям в общих амплитудно-частотной и фазо-частотной характеристиках — около 4 дБ пульсаций по амплитуде и 25° для максимума фазовой ошибки.

Для того чтобы улучшить характеристики, не меняя $w(n)$, была введена задержка, достаточная для выравнивания амплитуды первого и третьего пиков. Это показано на рис. 6.26. Пик функции $p(n)$ при $n=0$ сдвинулся на 129 отсчетов вправо. На рис. 6.26*б* и *в* видно, что результирующие амплитудно-частотная и фазо-частотная характеристики существенно улучшились. Теперь значение пульсации по амплитуде стало около 0,8 дБ, а максимальная ошибка по фазе — всего лишь 0,6°.





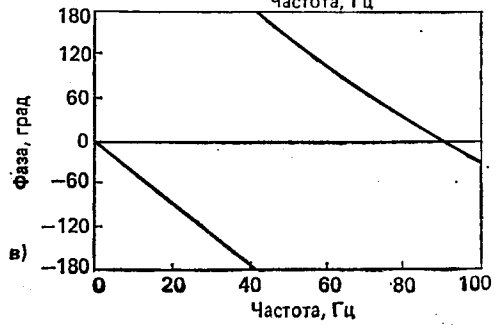
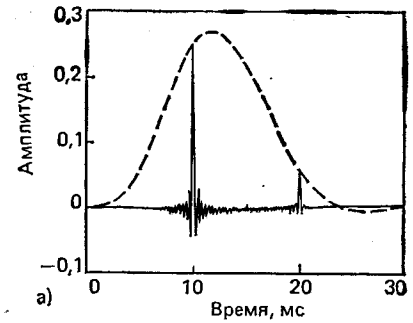
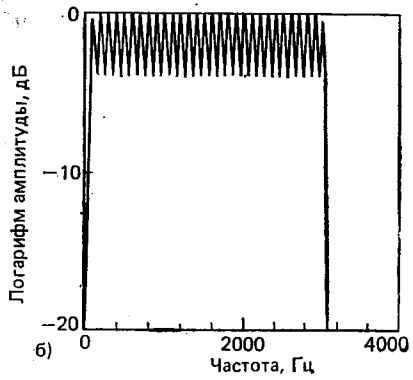


Рис. 6.24. Характеристики фильтра Бесселя 6-го порядка:
а) импульсная характеристика; *б*) амплитудно-частотная характеристика; *в*) фазо-частотная характеристика [1]





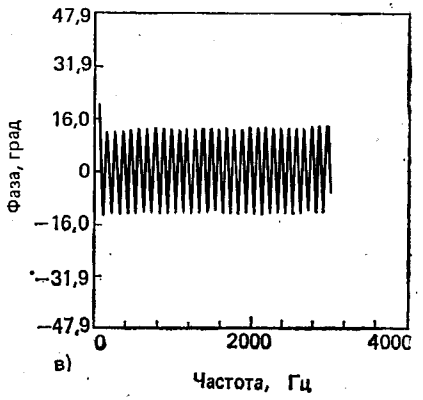


Рис. 6.25. Характеристики 30-канальной гребенки фильтров:
а) импульсная характеристика (пунктир — импульсная характеристика фильтра-прототипа нижних частот); *б*) общая амплитудно-частотная характеристика; *в*) общая фазо-частотная характеристика за вычетом задержки 10 мс [1]

Приведенный пример иллюстрирует следующую процедуру проб и ошибок, которая позволяет получать гребенки фильтров с хорошими общими характеристиками:

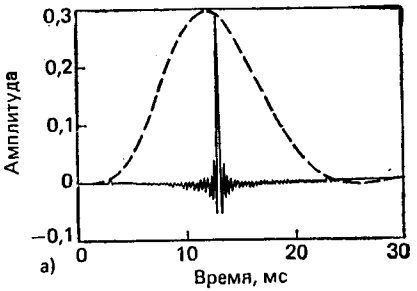
1. Определить разнос и количество фильтров, требуемых для перекрытия полосы анализа.

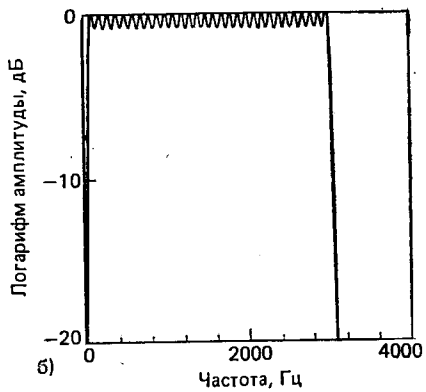
2. Спроектировать фильтр с нужной избирательностью в каждом канале. Здесь будет получено окно $w(n)$.

3. Вычислить $w(n)$ и определить n_0, при котором $\alpha'_1 = \alpha'_3$ (см. рис. 6.21).

4. Вычислить характеристику. Если она не удовлетворяет вас, изменить разнос, полосу частот и повторить расчет.

Эта процедура предназначена для равноразнесенных частот анализа. В [1] рассматривается также подход к проектированию неравноразнесенных гребенок, которые разбиваются на несколько подгребенок. Каждая из последних набрана из равноразнесенных фильтров. Этот подход довольно эффективен, однако лучшие результаты можно получить с помощью КИХ-фильтров. Поэтому вопросы проектирования гребенок с БИХ-фильтрами более рассматриваться не будут.





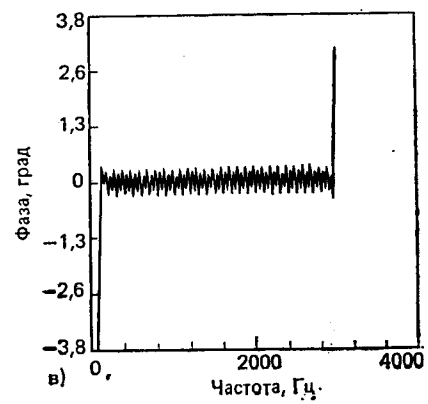


Рис. 6.26. Характеристики 30-канальной гребенки фильтров:
а) импульсная характеристика для $n_0 = 129$; *б*) общая амплитудно-частотная характеристика; *в*) общая фазо-частотная характеристика за вычетом задержки 12,9 мс [1]

6.2.3. Проектирование гребенок с КИХ-фильтрами

Цифровые КИХ-фильтры весьма удобны для проектирования гребенок фильтров анализаторов речи. Во-первых, такие фильтры легко спроектировать со строго линейной фазо-частотной характеристикой, просто наложив ограничение

$$w(n) = w(L-1-n),\ 0 \leqslant n \leqslant L-1 \qquad (6.110)$$

для каждого фильтра[1], где $w(n)$ — импульсная характеристика фильтра, а L — ее длина в отсчетах. Это означает, что требование линейности фазы общей характеристики гребенки фильтров выполняется тривиально, если отдельные фильтры обладают одинаковыми линейными фазо-частотными характеристиками. Можно поэтому сконцентрировать внимание на достижении нужной частотной избирательности каждого из фильтров и требуемой плоской характеристики всей гребенки фильтров. Второе достоинство КИХ-фильтров состоит в том, что существует множество методов проектирования — от прямого метода взвешивания до метода последовательных приближений, допускающих гибкий подход к реализации сложных требований проектирования.

Представляется, что метод взвешивания имеет ряд преимуществ при проектировании полосовых фильтров или фильтров нижних частот класса КИХ, предназначенных для использования в гребенках. Этот метод иллюстрируется на рис. 6.27. Прежде чем

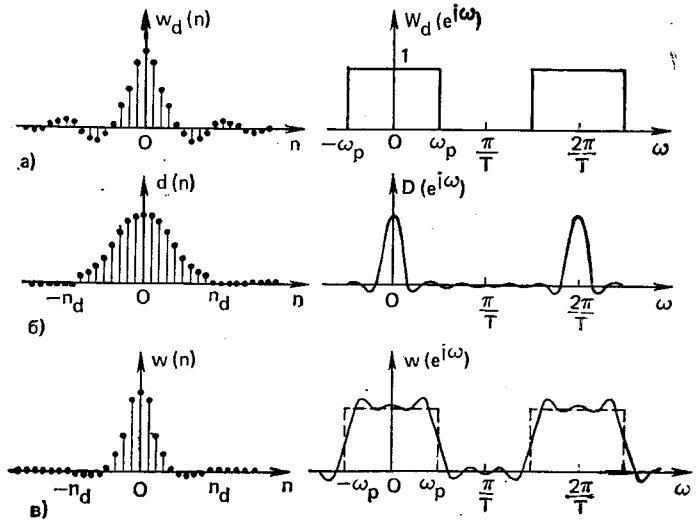


Рис. 6.27. Конструктивное окно для идеального фильтра нижних частот [2]

перейти к подробностям проектирования гребенки, рассмотрим метод взвешивания. Во-первых, идеальный фильтр нижних частот вида

$$W_d(e^{i\omega}) = \begin{cases} e^{-i\omega n_d}, & |\omega| \leqslant \omega_p; \\ 0, & \textit{в противном случае} \end{cases} \qquad (6.111)$$

[1]) Для простоты предполагается, что импульсная характеристика каждого полосового фильтра имеет длительность L, хотя такое ограничение снимается тривиально, если добавить в каждый канал соответствующую задержку.

задается частотой среза ω_p. Отметим, что для простоты на рисунке опущен член $e^{-i\omega n_d}$ с линейной фазой, соответствующий задержке на n_d отсчетов. Требуемое значение n_d равно $(L-1)/2$. Это означает, что при четном L задержка соответствует нецелому числу отсчетов. Следовательно, идеальная импульсная характеристика задается выражением

$$w_d(n) = \frac{1}{2\pi}\int_{-\omega_p}^{\omega_p} e^{-i\omega n_d} e^{i\omega n}\, d\omega = \frac{\sin[\omega_p(n-n_d)]}{\pi[n-n_d]} \text{ для всех } n. \qquad (6.112)$$

Эта характеристика имеет бесконечную длительность, и ее необходимо усечь, чтобы получить КИХ-фильтр. Это достигается вводом

$$w(n) = d(n-n_d)\, w_d(n), \qquad (6.113)$$

где $d(n)$ — конструктивное окно фильтра, а $w(n)$ — импульсная характеристика результирующего фильтра нижних частот[1]. Ширина окна, обозначенная через L, может быть либо четным ($L=2q$), либо нечетным ($L=2q+1$) целым числом. На рис. 6.27 приведен случай нечетного L.

В частотной области результатом умножения импульсной характеристики фильтра нижних частот на конструктивное окно будет свертка идеальной частотной характеристики с преобразованием Фурье $D(e^{i\omega})$ конструктивного окна:

$$W(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_d(e^{i\theta})\, D(e^{i(\omega-\theta)})\, d\theta. \qquad (6.114)$$

Результат этой свертки приведен на рис. 6.27*в*. Основные эффекты сводятся к появлению мягкого перехода от полосы пропускания к полосе задерживания и появлению пульсаций в полосе пропускания и полосе задерживания. Свойства такой аппроксимации отражены на рис. 6.28. Когда ω_p больше ширины основного лепестка $D(e^{i\omega})$, справедливо следующее:

1. Переходная область $\Delta\omega$ обратно пропорциональна L.

2. Функция $W(e^{i\omega})$ весьма близка к нечетной в окрестности точки $(\omega_p = 0{,}5)$.

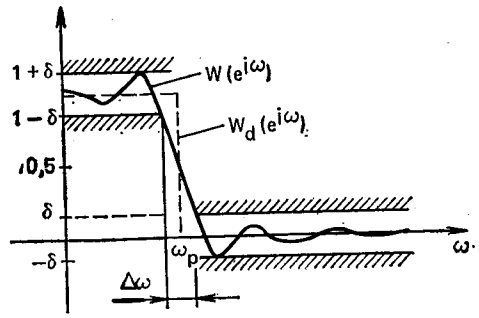


Рис. 6.28. Характеристика ФНЧ, спроектированного методом взвешивания [2]

[1] Здесь важно не запутаться в терминологии. Мы назвали $d(n)$ конструктивным окном фильтра, а $w(n)$ — импульсной характеристикой фильтра нижних частот; $w(n)$ представляет собой также и окно кратковременного анализа Фурье.

3. Максимумы погрешностей аппроксимации в полосе пропускания и в полосе задерживания весьма близки.

4. Погрешность аппроксимации максимальна в окрестности ω_p и убывает с изменением ω в обе стороны от ω_p.

Указанные выше свойства метода взвешивания выполняются для всех обычно применяемых окон. Однако Кайзер предложил семейство весьма гибких окон, близких к оптимальным, если они используются при проектировании фильтров [5]. Окно Кайзера равно:

$$d(n)=\begin{cases}\dfrac{I_0[\alpha\sqrt{1-(n/n_d)^2}]}{I_0(\alpha)}, & |n|\leqslant n_d,\\ 0, & \textit{в противном случае.}\end{cases} \qquad (6.115)$$

Здесь $n_d=(L-1)/2$, а $I_0[x]$ — модифицированная функция Бесселя нулевого порядка первого рода. Подбирая параметр α, можно добиться компромисса между шириной переходной области и пиковой погрешностью аппроксимации. Больше того, Кайзер формализовал процедуру выбора окон, дав эмпирическую формулу

$$L=\frac{-20\log_{10}\delta-7{,}95}{14{,}36\,\Delta f}+1, \qquad (6.116а)$$

где L — порядок фильтра; δ — пиковая погрешность аппроксимации; Δf — нормированная ширина переходной области:

$$\Delta f=\Delta\omega/2\pi. \qquad (6.116б)$$

При использовании этой формулы задают δ и Δf, чем добиваются нужной частотной избирательности. Затем используют (6.116а) для вычисления L, а параметр α можно подсчитать из уравнения [5]:

$$\left.\begin{aligned}&\alpha=0{,}1102\,(-20\log_{10}\delta-8{,}7), \quad \textit{для}\ -20\log_{10}\delta>50;\\ &\alpha=0{,}5842\,(-20\log_{10}\delta-\\ &-21)^{0{,}4}+0{,}07886\times\\ &\times(-20\log_{10}\delta-21), \qquad \textit{для}\ 21<-20\log_{10}\delta<50.\end{aligned}\right\} \qquad (6.117)$$

На практике выбор δ и Δf зависит от требований к полосовым фильтрам, из которых набрана гребенка.

Гребенка фильтров составлена из набора полосовых фильтров с импульсными характеристиками вида

$$h_k(n)=P_k w_k(n)\,e^{i\omega_k n},\ 0\leqslant k\leqslant N-1, \qquad (6.118)$$

где $w_k(n)$ — импульсная характеристика фильтра нижних частот. На рис. 6.29 приведены три идеальных полосовых фильтра с абсолютно плоской характеристикой в области $\omega_{min}\leqslant\omega\leqslant\omega_{max}$. Проектируя фильтры нижних частот, мы выбираем набор частот анализа $\{\omega_k\}$ и набор частот среза ω_{pk} так, чтобы точно перекрыть полосу, как это показано на рис. 6.29. Идеальные полосовые фильтры затем аппроксимируются методом взвешивания.

Выбор пиковой погрешности аппроксимации связан с необходимым для данного применения затуханием в полосе задерживания. Типичное значение $-20\log_{10}\delta$ оказывается, как правило, между 40 и 60 дБ. Значение α вычисляется по (6.117). И, наконец, нормированное значение ширины переходной области Δf задается для

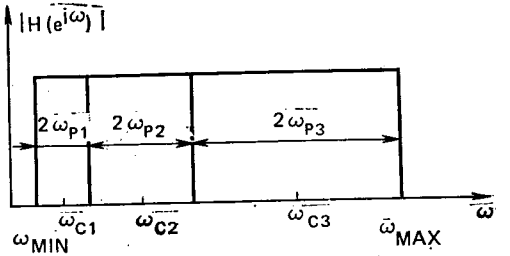


Рис. 6.29. Общая характеристика трех идеальных полосовых фильтров с разными полосами пропускания

того, чтобы по (6.116б) вычислить L. Выбор Δf или $\Delta\omega$ опять-таки определяется нужной частотной избирательностью отдельных фильтров. Ясно, что ширина переходной области $\Delta\omega_k$ не должна превышать $2\omega_{pk}$.

При разработке гребенки фильтров преполагается, что $\Delta\omega$ одинакова для всех фильтров, следовательно, можно воспользоваться приведенным выше свойством 2. Если переходная область у всех фильтров одна и та же и если, кроме того, переходы симметричны в окрестности точки переходов, то можно ожидать, что сумма частотных характеристик окажется близкой к единице. Отсюда вытекает, что $d(n)$ должна быть одинаковой для всех фильтров гребенки.

Чтобы уяснить эффект одинакового выбора $d(n)$ для всех фильтров, рассмотрим общую частотную характеристику, которая, как следует из (6.118), равна

$$\widetilde{H}(e^{i\omega})=\sum_{k=0}^{N-1}P_k W_k(e^{i(\omega-\omega_k)}). \qquad (6.119)$$

Теперь при одинаковых на всех частотах анализа конструктивных окнах можно записать

$$W_k(e^{i(\omega-\omega_k)})=\frac{1}{2\pi}\int_{-\pi}^{\pi}W_{dk}(e^{i(\theta-\omega_k)})\,D(e^{i(\omega-\theta)})\,d\theta. \qquad (6.120)$$

Подставив (6.120) в (6.119) и изменив порядок суммирования и интегрирования, получим

$$\widetilde{H}(e^{i\omega})=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{k=0}^{N-1}P_k W_{dk}(e^{i(\theta-\omega_k)})\right]D(e^{i(\omega-\theta)})\,d\theta. \qquad (6.121)$$

Определив

$$\widetilde{H}_d(e^{i\omega})=\sum_{k=0}^{N-1}P_k W_{dk}(e^{i(\omega-\omega_k)}), \qquad (6.122)$$

можно переписать (6.121) в виде

$$\widetilde{H}(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widetilde{H}_d(e^{i\theta}) D(e^{i(\omega-\theta)})\, d\theta. \qquad (6.123)$$

Видно, что функция $\widetilde{H}_d(e^{i\omega})$ представляет собой требуемую общую характеристику. Если, например, $P_k=1$ для $0\leqslant k\leqslant N-1$, ширина полос и центральные частоты выбраны так, чтобы перекрыть всю полосу частот $-\pi\leqslant\omega\leqslant\pi$, то

$$\widetilde{H}_d(e^{i\omega}) = e^{i\omega n_d}, \quad -\pi \leqslant \omega \leqslant \pi. \qquad (6.124)$$

Подставив этот результат в (6.123), получим

$$\widetilde{H}(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\theta n_d} D(e^{i(\omega-\theta)})\, d\theta = d(n_d)\, e^{-i\omega n_d}. \qquad (6.125)$$

Отсюда следует, что общая импульсная характеристика имеет вид

$$\widetilde{h}(n) = d(n_d)\,\delta(n-n_d). \qquad (6.126)$$

Следовательно, если имеется достаточное количество каналов для линейности фазо-частотной характеристики и если все фильтры проектируются с одинаковым окном, общая характеристика оказывается также идеальной. Иначе говоря, независимо от распределения центральных частот и полос общая характеристика оказывается идеальной при любом конструктивном окне. Итак, при использовании КИХ-фильтров теоретически можно получить точное воспроизведение входного сигнала при произвольном распределении центральных частот и полос.

Эффект пропуска части частотного диапазона $-\pi\leqslant\omega\leqslant\pi$ можно выяснить, заметив, что (6.123) справедливо независимо от выбора P_k в (6.122). Следовательно, если из анализа исключаются как нижние, так и высшие частотные диапазоны (см. рис. 6.29), то P_0 и $P_k=0$ для $k>M$, где M — количество каналов. Общая характеристика будет иметь вид

$$\widetilde{H}_d(e^{i\omega}) = e^{-i\omega n_d}, \ \omega_{min} \leqslant |\omega| \leqslant \omega_{max}. \qquad (6.127)$$

Таким образом, для заданного конструктивного окна $d(n)$, заданных ω_{min} и ω_{max} общая характеристика будет характеристикой полосового фильтра с переходной областью и пульсациями (в полосе пропускания и полосе задерживания), идентичными с каждым из фильтров каналов. Это происходит потому, что одно и то же конструктивное окно перемножается с каждой из отдельных импульсных характеристик. Следовательно, опять общая характеристика не зависит от числа и распределения отдельных полосовых фильтров.

Проектирование гребенок фильтров в соответствии с указанными принципами иллюстрируют следующие примеры.

Пример 1. Допустим, что входная частота дискретизации равна 9,6 кГц и требуется спроектировать гребенку из 15 равноразнесенных фильтров, перекрывающую область от 200 до 3200 Гц. Частотой среза для всех фильтров нижних частот будет

$$F_p = \omega_p/2\pi T = (3200-200)/2\,(15) = 100. \qquad (6.128)$$

а центральные частоты равны

$$F = \omega_k/2\pi T = 200k + 100, \quad k = 1,2,\ldots,15. \qquad (6.129)$$

При таком выборе центральных частот и полос 15 идеальных фильтров перекрывают интервал от 200 до 3200 Гц. Предположив, что требуется затухание на 60 дБ вне переходных областей каждого канала, найдем из (6.117), что $\alpha=5{,}65326$. Поскольку для фильтров-прототипов частота среза равна 100 Гц, разумно выбрать в качестве самой широкой из областей перехода значение 200 Гц. Воспользовавшись этим значением и тем, что $-20\log_{10}\delta=60$, получим из (6.116а) $L=175$ нижнюю оценку для L. Отметим, что, если бы были допустимы меньшие затухания, можно было бы получить меньшее L при том же Δf. Характеристики гребенки, спроектированной с указанными параметрами, приведены на рис. 6.30. В верхней части рисунка показаны характеристики отдельных

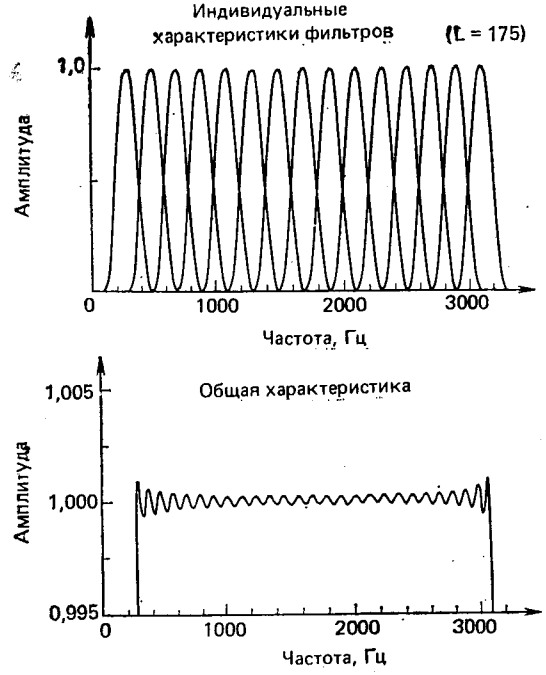


Рис. 6.30. Индивидуальная и общая характеристики гребенки из 15 полосовых фильтров с одинаковыми полосами пропускания при $L=175$ [2]

полосовых фильтров. Спад в переходной области каждого фильтра компенсируется подъемом следующего. Отметим, что смежные каналы пересекаются на уровне 0,5. В нижней части рисунка — общая амплитудно-частотная характеристика гребенки. Фазо-частотная характеристика линейна с задержкой $n_d=(175-1)/2=87$ отсчетов. Ясно, что фильтры хорошо согласуются на краях частотного диапазона. Действительно, отклонение от единицы не превышает или равно пиковой погрешности аппроксимации $\delta=0{,}001$, выбранной при проектировании фильтров нижних частот.

Пример 2. Неравномерное разнесение фильтров используется в тех случаях, когда хотят воспользоваться ухудшением частотного разрешения слуха с ростом частоты. Предположим, что требуется перекрыть тот же самый диапазон от 200 до 3200 Гц, применяя четырехоктавные фильтры, т. е. каждый следующий фильтр имеет полосу вдвое шире предыдущей. Отсюда следует, что диапазон частот от 200 до 3200 Гц следует разбить на четыре полосы шириной в 200, 400, 800 и 1600 Гц с центральными частотами 300, 600, 1200 и 2400 Гц соответственно. Допустив опять, что требуемое затухание составляет 60 дБ, видим, что наименьшая частота среза равна 100 Гц, так что наименьшая переходная область равна 200 Гц. Гребенка фильтров, соответствующая этим параметрам, иллюстрируется рис. 6.31. Дополняющее соотношение между спадами и нараста-

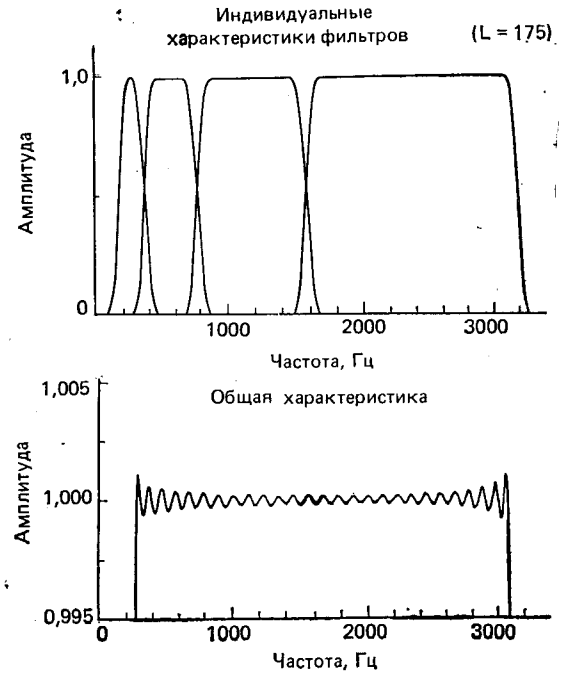


Рис. 6.31. Индивидуальная и общая характеристики гребенки из четырех полосовых фильтров с разными полосами пропускания при $L=175$ [2]

ниями переходов смежных каналов видно в верхней части рисунка. Понятно, что, поскольку L и α одинаковы для каждого из фильтров нижних частот — прототипов, форма кривых в переходной области не зависит от ширины полосы. В нижней части рис. 6.31 приведена общая характеристика, причем отклонение от единицы опять не превосходит 0,001. Как и в примере 1, фаза линейна и соответствует задержке, равной 87 отсчетам. Сравнение рис. 6.30 и 6.31 подтверждает, что в обоих случаях получается одинаковая общая характеристика.

Пример 3. Положим все параметры такими, как и в примере 2, за исключением того, что потребуем более узких переходных областей. Это означает, что необходимо большее значение L. В действительности (6.116а) показывает, что L и Δf, грубо говоря, обратно пропорциональны. На рис. 6.32 показаны характеристики соответствующей гребенки фильтров с параметрами, взятыми из примера 2 с $L=301$ и $\Delta f=0{,}012082$; ширина области перехода равна 116 Гц. В

верхней части рисунка видны более крутые переходы, а из нижней части ясно, что общая характеристика осталась плоской. Задержка в этом случае равна 150 отсчетам.

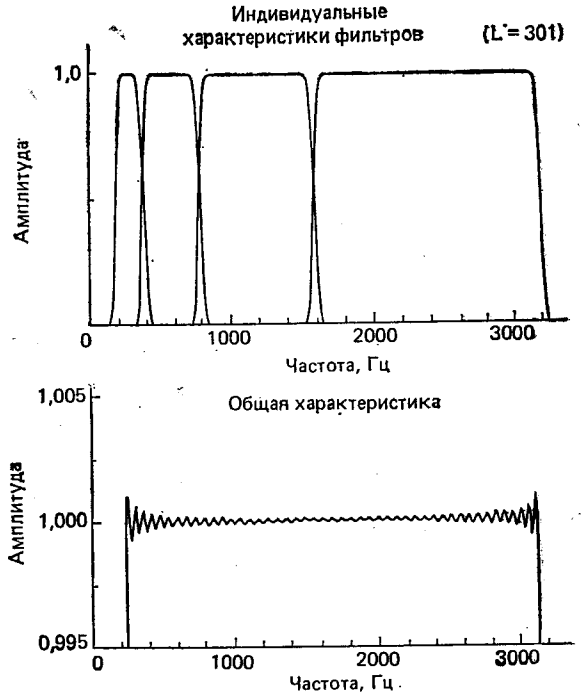


Рис. 6.32. Индивидуальная и общая характеристики гребенки из четырех полосовых фильтров с разными полосами пропускания при $L=301$ [2]

6.3. Реализация метода суммирования выходов гребенки фильтров с помощью БПФ

В предыдущем разделе показано, что можно спроектировать гребенку физически реализуемых фильтров, выход которой совпадает со входом с точностью до задержки и масштабного множителя. В частности, для этой цели особенно хорошо подходят фильтры с конечной импульсной характеристикой. Поскольку кратковременные анализ и синтез Фурье эквивалентны такой гребенке фильтров, то при проектировании систем анализа — синтеза можно эффективно использовать анализирующие окна конечной длительности. Одним из основных недостатков систем на КИХ-фильтрах является большое число вычислений при их реализации. В частном случае кратковременного анализа Фурье имеется удачное обстоятельство — существует ряд методов, позволяющих снизить объем вычислений по сравнению с прямой реализацией.

6.3.1. Методы анализа

Рассмотрим систему кратковременного анализа—синтеза Фурье с равноразнесенными частотами анализа $\omega_k=2\pi k/N$, $0\leqslant k\leqslant N-1$. В 6.1.3 было показано, что нет необходимости вычислять $X_n(e^{i\omega_k})$ с частотой дискретизации на входе, поскольку каж-

дая из последовательностей представляет собой по существу выходную последовательность фильтра нижних частот с цифровой частотой среза π/N. Следовательно, входную последовательность можно вычислять всего 1 раз через каждые N последовательных отсчетов на входе. В таких случаях особенно подходят КИХ-системы, так как в них можно ограничиться вычислением только нужных выходных отсчетов, не вычисляя промежуточных $N-1$ отсчетов. В БИХ-системах приходится вычислять на выходе все значения из-за присущей им рекурсивной природы реализации.

Дополнительное увеличение эффективности вычислений можно получить за счет применения метода быстрого преобразования Фурье (БПФ) [6]. Чтобы уяснить это, выразим кратковременное преобразование Фурье в виде

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=\sum_{m=-\infty}^{\infty}x(m)w(n-m)e^{-i\frac{2\pi}{N}km},\ 0\leqslant k\leqslant N-1. \quad (6.130)$$

Видно, что если бы пределами суммирования были 0 и $N-1$, то отношение (6.130) приняло бы форму дискретного преобразования Фурье (ДПФ). В случае конечной длительности $w(m)$ выражение (6.130) можно привести к виду ДПФ и, следовательно, можно для вычисления $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $0\leqslant k\leqslant N-1$ воспользоваться алгоритмом БПФ. При замене переменных суммирования (6.130) превращается в

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{m=-\infty}^{\infty}x_n(m)e^{-i\frac{2\pi}{N}km}, \quad (6.131)$$

где

$$x_n(m)=x(n+m)w(-m),\ -\infty<m<\infty. \quad (6.132)$$

Иначе говоря, последовательность $x_n(m)$ получается при переносе начала последовательности $x(m)w(n-m)$ в отсчет n, что концентрирует внимание на членах последовательности в окрестности момента времени, для которого необходимо вычислить $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Далее подстановками $m=Nr+q$, $-\infty<r<\infty$ и $0\leqslant q\leqslant N-1$ можно представить (6.131) двойной суммой:

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{r=-\infty}^{\infty}\left(\sum_{q=0}^{N-1}x_n(Nr+q)\right)e^{-i\frac{2\pi}{N}k(Nr+q)}. \quad (6.133)$$

Поскольку $e^{-i2\pi kr}=1$, можно поменять порядок суммирования и получить

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{q=0}^{N-1}\left(\sum_{r=-\infty}^{\infty}x_n(Nr+q)\right)e^{-i\frac{2\pi}{N}kq}. \quad (6.134)$$

Определив конечную последовательность

$$u_n(q)=\sum_{r=-\infty}^{\infty}x_n(Nr+q),\ 0\leqslant q\leqslant N-1, \quad (6.135)$$

видим, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой просто произведение $e^{-i\frac{2\pi}{N}kn}$ на N-точечное ДПФ последовательности $u_n(q)$. Или, по-другому, $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой N-точечное ДПФ последовательности $u_n(q)$ после циклического сдвига на n по модулю N. Иначе говоря,

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=\sum_{m=0}^{N-1}u_n((m-n))_N e^{-i\frac{2\pi}{N}km}, \quad (6.136)$$

где обозначение $((m-n))_N$ указывает на то, что целое в двойных скобках следует рассматривать по модулю N. Мы сумели, таким образом, привести $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ к виду N-точечного ДПФ последовательности конечной длины, полученной по взвешенной окном входной последовательности. Итак, процедура для вычисления $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $0\leqslant k\leqslant N-1$ состоит в следующем:

1. Сформировать последовательность $x_n(m)$, такую, как в (6.132), умножая $x(m+n)$ на обращенную во времени последовательность окна $w(-m)$. На рис. 6.33 показаны $x(m+n)$ и три специальных случая $w(-m)$.

2. Разбить результирующую последовательность на сегменты по N отсчетов и сложить эти сегменты вместе в соответствии с (6.135), что даст последовательность конечной длины $u_n(q)$, $0\leqslant q\leqslant N-1$.

3. Циклически сдвинуть $u_n(q)$ на n по модулю N, что даст $u_n((m-n))_N$, $0\leqslant m\leqslant N-1$.

4. Вычислить N-точечное ДПФ от $u_n((m-n))_N$, что даст $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ $0\leqslant k\leqslant N-1$. Эту процедуру придется повторить для каждого n, при котором необходимо значение $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Ясно, однако, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ можно наращивать любым требуемым способом. Можно, например, вычислять $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $n=0, \pm R, \pm 2R, \ldots$, т. е. с интервалом R отсчетов входного сигнала. Это оправдано тем, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой выход фильтра нижних частот с номинальной частотой среза π/N радиан. Следо-

вательно, «отсчетов» $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ будет достаточно для восстановления входного сигнала, если только $R \leqslant N$.

Этот метод даст значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ для всех k. В общем случае можно ограничиться вычислениями, самое большее для половины каналов. Это связано с сопряженной симметрией $X_n(e^{i\omega})$. Часто каналы на очень низких и очень высоких частотах не реализуют. Возникает, следовательно, вопрос: будет ли метод ДПФ эффективнее непосредственной реализации. Чтобы сравнить их, предположим, что нам требуются значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ только при $1 \leqslant k \leqslant M$. Допустим, кроме того, что длительность окна равна L. Тогда для вычисления полного набора значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ потребуется $4LM$ умножений и примерно $2LM$ сложений, если используется метод, иллюстрированный рис. 6.12. Допустив, что применен обычный комплексный алгоритм БПФ (с N, равным целой степени 2)[1], можно показать, что метод БПФ потребует примерно $L+2N\log_2 N$ умножений и $L+2N\log_2 N$ сложений, чтобы получить

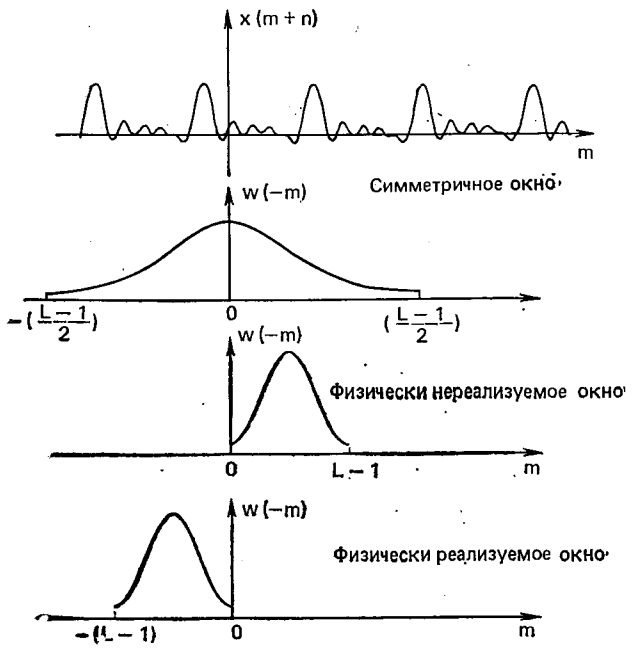


Рис. 6.33. Графики последовательностей $x(m+n)$ и $w(-m)$

[1] Здесь не используется то, что $u_n((m-n))_N$ — действительно. Можно было бы воспользоваться этим и уменьшить число вычислений еще в 2 раза.

все N значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Если за меру сравнения принять число умножений действительных величин, то легко показать, что при $L=N$ метод БПФ потребует меньшего числа вычислений, если только не выполняется неравенство

$$M \leqslant \log_2 N/2. \tag{6.137}$$

Пусть, например, $N=128=2^7$. Видим, что БПФ эффективнее прямого метода, если только не выполняется неравенство $M \leqslant 3{,}5$, т. е. кроме случаев, когда число каналов меньше четырех. Следовательно, во всех приложениях, где необходимо тонкое разрешение по частоте, почти наверняка метод БПФ будет самым эффективным (заметим, что при $L>N$ сравнение еще больше в пользу БПФ).

6.3.2. Методы синтеза

Предыдущее обсуждение методов анализа показало, что, используя алгоритм быстрого преобразования Фурье, можно определить все N равноразнесенные значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при меньшем объеме вычислений, чем требуется для вычисления M каналов с непосредственной реализацией. Реорганизовав вычисления, необходимые в ходе синтеза, можно получить аналогичный выигрыш, восстанавливая $x(n)$ по значениям $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ через каждые R отсчетов $x(n)$, где $R \leqslant N$ [7].

Из (6.83) при $\omega_k = 2\pi k/N$ получим для выхода системы синтеза

$$y(n) = \sum_{k=0}^{N-1} Y_n(k)\, e^{i\frac{2\pi}{N}kn}, \tag{6.138}$$

где

$$Y_n(k) = P_k X_n\left(e^{i\frac{2\pi}{N}k}\right), \quad 0 \leqslant k \leqslant N-1. \tag{6.139}$$

Вспомним, что весовые комплексные коэффициенты P_k позволяют подобрать модуль и фазу каналов. Если имеются значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ только при целых кратных R, то промежуточные значения можно получить с помощью интерполяции. Для этого полезно определить последовательность

$$V_n(k) = \begin{cases} P_k X_n\left(e^{i\frac{2\pi}{N}k}\right), & n = 0, \pm R, \pm 2R, \ldots \\ 0, & \textit{в противном случае.} \end{cases} \tag{6.140}$$

Для каждого значения k имеется последовательность указанного вида. Для каждого k промежуточные значения заполняются теперь за счет обработки последовательности $V_n(k)$ фильтром нижн-

них частот с частотой среза π/N радиан. Если обозначить отклик этого фильтра на единичный импульс через $h(n)$ и допустить, что он симметричен, а полная длина равна $2RQ-1$, то для каждого k из интервала $0 \leqslant k \leqslant N-1$ получим

$$Y_n(k) = \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, V_m(k), \quad -\infty < n < \infty. \tag{6.141}$$

Соотношение (6.141) вместе с (6.138) описывает операции, необходимые при вычислении выхода синтеза в случае, если имеется кратковременное преобразование Фурье, заданное через интервалы в R отсчетов. Этот процесс показан на рис. 6.34, на котором

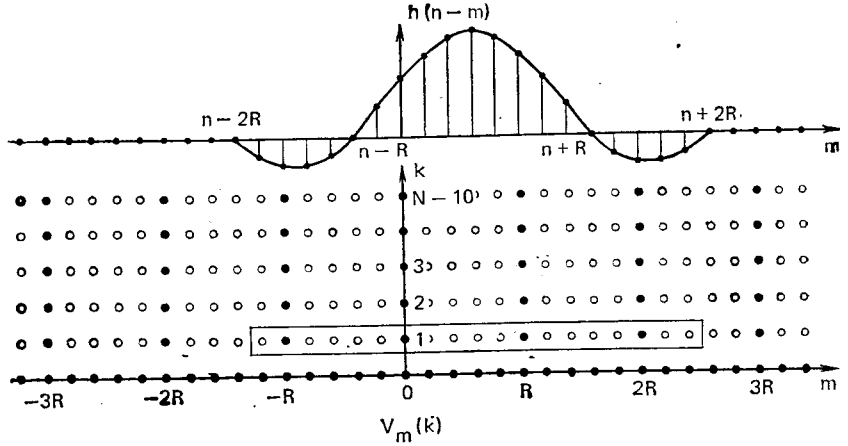


Рис. 6.34. Отсчеты, входящие в вычисление $Y_n(k)$

$V_m(k)$ приведена как функция m и k (напомним, что m — индекс времени, а k — индекс частоты). Точками отмечены координаты, в которых $V_m(k)$ отлична от нуля, т. е. точки, в которых известно $X_m\left(e^{i\frac{2\pi}{N}k}\right)$. Кружками отмечены точки, в которых $V_m(k)$ равна нулю и в которых требуется интерполировать значения $Y_n(k)$. Импульсная характеристика интерполирующего фильтра (для $Q=2$) показана в момент n. Сигнал каждого из каналов интерполируется свертыванием с импульсной характеристикой интерполирующего фильтра. В качестве примера отсчеты, связанные с вычислением $Y_3(1)$, помещены в рамку. В общем случае рамка, указывающая на отсчеты, связанные с вычислением $Y_n(k)$, будет скользить вдоль k-й строки (рис. 6.34), причем центр рамки совпадает с координатой n. Отметим, что каждое из интерполируемых значений зависит от $2Q$ известных значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Предположив, что в процессе синтеза доступны M каналов, легко показать, что потребуется $2(Q+1)M$ умножений и $2QM$ сложений для вычисления каждого из значений выходной последовательности.

Чтобы понять, как можно вычислить выходной сигнал более эффективно, подставим (6.141) в (6.138), что даст

$$y(n) = \sum_{k=0}^{N-1} \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, Y_m(k)\, e^{i\frac{2\pi}{N}kn}. \tag{6.142}$$

Изменив порядок суммирования, получим

$$y(n) = \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, v_m(n), \tag{6.143}$$

где

$$v_m(r) = \sum_{k=0}^{N-1} V_m(k)\, e^{i\frac{2\pi}{N}kr}. \tag{6.144}$$

Воспользовавшись (6.140), видим, что

$$v_m(r) = \begin{cases} \sum_{k=0}^{N-1} P_k X_m\left(e^{i\frac{2\pi}{N}k}\right) e^{i\frac{2\pi}{N}kr}, & m = 0, \pm R, \pm 2R, \ \ldots \\ 0, & \textit{в противном случае.} \end{cases} \tag{6.145}$$

Следовательно, вместо того чтобы интерполировать кратковременное преобразование Фурье, а затем вычислять (6.138), можно вычислять $v_m(r)$ во все те моменты, в которых известна $X_n\left(e^{i\frac{2\pi}{N}k}\right)$, т. е. $m=0, \pm R, \ldots$, а затем интерполировать $v_m(r)$, как в (6.143).

Можно видеть, что $v_m(r)$ имеет вид обратного дискретного преобразования Фурье и, следовательно, $v_m(r)$ периодична по r с периодом N. Поэтому в (6.143) переменную n в $v_m(n)$ надо интерпретировать по модулю N. Такой процесс интерполяции приведен на рис. 6.35. Жирными точками в двумерной сети представлены точки, в которых $v_m(r)$ отлична от нуля. Оставшиеся точки на рис. 6.35 можно интерполировать так, как это было описано при интерполяции $Y_n(k)$. Однако нет необходимости интерполировать все эти точки, поскольку требуется получить только значения одномерной последовательности $y(n)$. Из (6.143) и периодического характера $v_m(r)$ видно, что $y(n)$ равна значениям интерполируемой последовательности вдоль «пилы» (см. рис. 6.35).

Реализуя таким способом систему синтеза, можно вычислить N-точечную последовательность $v_m(r)$, $0 \leqslant r \leqslant N-1$ для каждого значения m, в котором известна $X_n\left(e^{i\frac{2\pi}{N}k}\right)$, используя при этом алгоритм быстрого преобразования Фурье для выполнения вычислений ДПФ в (6.145). Чтобы исключить канал, достаточно просто приравнять значение соответствующего ему сигнала нулю до вычисления обратного дискретного преобразования Фурье. Аналогичным образом, если требуется реализовать линейный фазовый

сдвиг выбором $P = e^{-i\frac{2\pi}{N}kn_0}$, то, как нетрудно показать, результат сведется к простому циклическому сдвигу последовательности $v_m(r)$. Можно, следовательно, избежать умножения на $e^{-i\frac{2\pi}{N}kn_0}$, выполнив обратное дискретное преобразование Фурье непосредственно над $X_m\left(e^{i\frac{2\pi}{N}k}\right)$ и сдвинув затем циклически результат на n_0 отсчетов. Коль скоро получены последовательности $v_m(r)$, выход можно вычислить, интерполируя $v_m(r)$, как в (6.143). Для каждого значения $y(n)$ необходимы $2Q$ значения $v_m(r)$. Отсчеты, необходимые для двух различных значений n, показаны заключенными в рамки на рис. 6.35. Для R последовательных значений $y(n)$ значения $v_m(r)$ получаются из одних и тех же $2Q$ столбцов. Поэтому выход удобно вычислять блоками по R отсчетов.

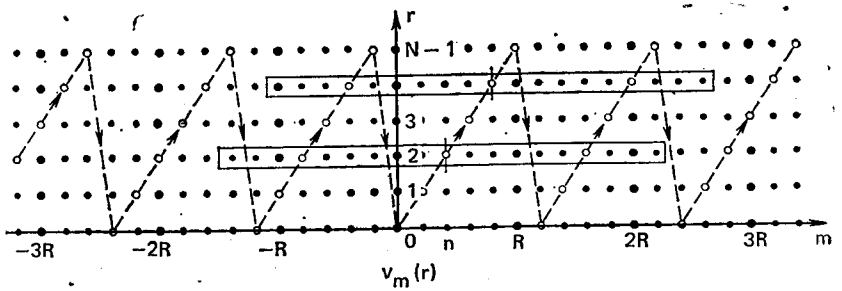


Рис. 6.35. Процесс интерполяции $v_m(r)$ по [7]

Объем вычислений, требуемых при реализации кратковременного синтеза Фурье указанным выше способом, можно снова оценить, предположив, что N — степень двойки и что применен алгоритм БПФ для вычислений обратных преобразований в (6.145). При таком допущении синтез требует $(2QR+2N\log_2 N)$ умножений действительных величин и $(2QR-1+2N\log_2 N)$ сложений действительных величин для вычисления группы из R последовательных значений выхода $y(n)$. Прямой метод синтеза требует $2(Q+1)MR$ умножений и $2QMR$ сложений для вычисления R последовательных отсчетов выхода. Рассмотрев случай, когда прямой метод требует меньшего, чем метод БПФ, числа умножений, найдем

$$M < \frac{Q + (N/R)\log_2 N}{Q+1}. \qquad (6.146)$$

Для типичных значений $N=128$, $Q=2$ (интерполяция через четыре отсчета, как на рис. 6.35) и при $R=N$ (наименьшая возможная частота дискретизации $X_n\left(e^{i\frac{2\pi}{N}k}\right)$) видим, что прямой метод эффективнее только в том случае, если $M<3$. Следовательно, для большинства приложений БПФ дает значительное увеличение эффективности вычислений операций синтеза.

6.4. Спектрографическое отображение

Идеи кратковременного преобразования Фурье появились задолго до появления методов цифровой обработки сигналов. Исследователи речи полагались на спектральные методы анализа начиная с 30-х годов нашего века. Одним из устройств, использующих кратковременное представление Фурье, был звуковой спектрограф — устройство, ставшее важным инструментом в почти каждой фазе исследования речи. В этом приборе короткие (2 с) отрезки речи многократно модулируют сигнал генератора переменной частоты. Модулированный сигнал поступает на полосовой фильтр. Средняя энергия на выходе полосового фильтра для заданных частоты и времени представляет собой грубое приближение кратковременного преобразования Фурье. Эта энергия регистрируется остроумной электромеханической системой на электрочувствительной бумаге. Результат, называемый спектрограммой, представляет собой двумерное представление зависящего от времени спектра, при этом по вертикали откладывается частота, а по горизонтали — время. Амплитуда спектра представляется затемнением отметок на бумаге. Если полосовые фильтры имеют большую ширину полосы (300 Гц), то на спектрограмме получается хорошее разрешение по времени и плохое по частоте. При узкой полосе (45 Гц) спектрограмма имеет хорошее разрешение по частоте и плохое по времени.

На рис. 6.36*а* приведена широкополосная спектрограмма фразы «Every salt breeze comes from the sea». Этот пример иллюстрирует ряд характерных свойств широкополосного кратковременного спектра. Во-первых, видно, что в фиксированный момент времени спектр меняется с частотой, как это показано на рис. 6.3, 6.5, т. е. спектр состоит из ряда широких пиков, соответствующих частотам формант. На спектрограмме четко отражены изменения во времени частот формант. Кроме того, широкополосная спектрограмма имеет линейчатый характер в областях вокализованной речи. Она возникает из-за того, что импульсная характеристика (т. е. анализирующее окно спектра) имеет длительность, примерно совпадающую с периодом основного тона. Поэтому энергия на выходе фильтра максимальна, когда пик импульсной характеристики совпадает с максимумом каждого периода основного тона. В другие моменты энергия на выходе значительно меньше. Для невокализованной речи, которая непериодична, линейчатость исчезает и спектр оказывается более «рваным».

На рис. 6.36*б* показана узкополосная спектрограмма той же фразы. В этом случае у фильтра полоса пропускания выбрана так, что в вокализованной области разделяются отдельные гармоники. Хотя по-прежнему видны частоты формант, сечение в фиксированный момент времени напоминает спектр рис. 6.2 и 6.4. В вокализованной области уже нет линейчатости, поскольку импульсная характеристика при узкой полосе захватывает несколько периодов основного тона; теперь, однако, четко видны основная

частота и ее гармоники. Невокализованные области выделяются отсутствием периодичности по частоте.

На широкополосных и узкополосных спектрограммах отображается существенная часть информации о свойствах речи. Когда

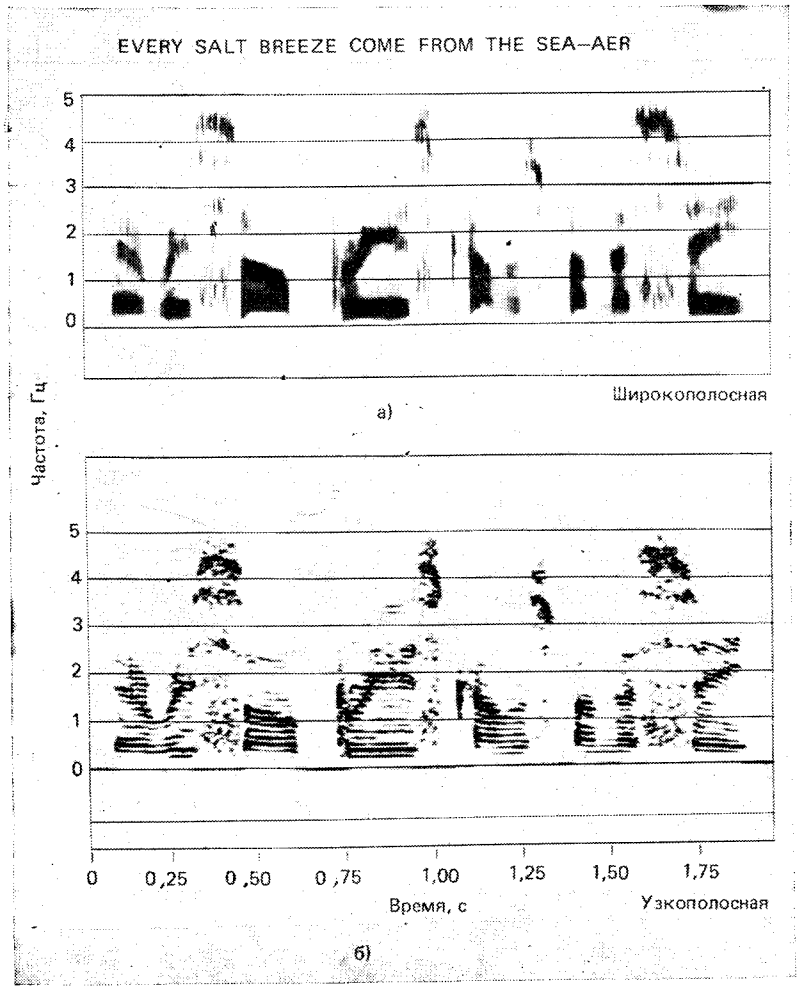


Рис. 6.36. Широкополосная и узкополосная спектрограммы предложения

аппаратура для такого отображения кратковременного представления Фурье появилась впервые, надеялись даже, что получен новый «язык» для общения с глухими. И хотя этим надеждам не суждено было воплотиться в жизнь, последующие исследования привели к написанию книги «Visible Speech» [8], которая и по сей день служит источником информации о спектральных и временных свойствах речи. За годы, прошедшие с появления этой работы, многие исследователи речи вручную определяли по спектрограммам такие параметры речи, как частоты формант и основная частота.

Другим следствием изобретения спектрографа стало то обстоятельство, что по детальному анализу спектрограммы или «отпечатка голоса» произнесенной фразы, как оказалось, можно установить личность говорящего. И хотя вопрос надежности такой идентификации по спектрограмме остается открытым, она получила некоторое признание в юрисдикции.

Звуковой спектрограф долгое время оставался основным инструментом анализа в исследованиях речи. С появлением вычислительных машин, доступных при исследованиях речи, это изменилось. В предыдущих разделах этой главы показаны способы реализации кратковременного представления Фурье. Они гораздо более хитроумные, чем способы, реализуемые на аналоговой аппаратуре. Эти представления могут быть реализованы и в цифровой специализированной аппаратуре, и как программы для универсальной вычислительной машины. Воспользовавшись, например, методами, изложенными в § 6.3, можно получить $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ — комплексное двумерное представление речевого сигнала с дискретным временем и частотой и, кроме того, периодическое по частоте. Возникает, следовательно, задача визуального отображения такого представления. Часто ограничиваются только отображением $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$. Поскольку последовательность $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$ четна и периодична по k с периодом N, на практике необходимо отображать значения только из интервала $0 \leqslant k \leqslant N/2$.

В случаях, когда в вычислительной машине имеется осциллоскоп или графопостроитель, кратковременное преобразование Фурье можно отображать просто в виде последовательности графиков $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$ как функцию k при фиксированных n. Обычно значения n разносятся на интервалы, соответствующие частоте Найквиста в спектральных каналах. Для узкополосного анализа, например, разнос по времени может быть порядка 10—20 мс. На рис. 6.37 [11] приведена последовательность узкополосного спектра, вычисленного через интервалы 20 мс. Видно, что весь сегмент речи вокализован.

Альтернативой отображения спектра как сечений поверхности, определяемой $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$, может служить отображение этой поверхности в перспективе (рис. 6.38) [12].

Ясно, что такой график менее полезен при количественных измерениях. Однако он, как и спектрограмма, имеет преимущество отображения всей фразы в компактной форме.

Поскольку полезность и повсеместная принятость спектрограмм как основного инструмента уже продемонстрирована, постольку

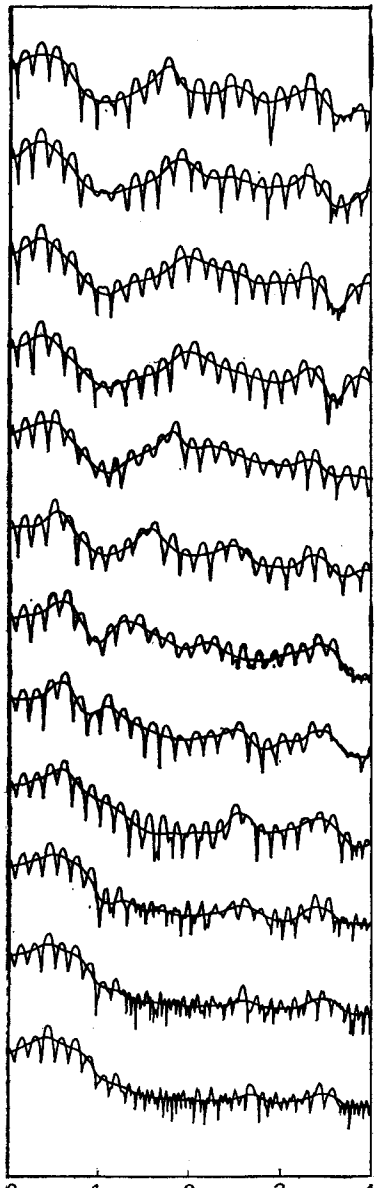


Рис. 6.37. Набор узкополосных спектров сегмента вокализованной речи [11]

по-видимому, цифровые спектрограммы полезнее любых других способов отображения. Если имеется телевизионный или электронно-лучевой дисплей для отображения дискретизованных изображений, то значение $|X_n(e^{i\frac{2\pi}{N}k})|$ из подходящего по размерам интервала можно рассматривать как дискретизованное изображение[1]. Некоторые исследователи рассмотрели такие возможности и нашли, что можно добиться полного сходства с аналоговыми спектрограммами. Действительно, поскольку электрочувствительная бумага имеет диапазон черного, равный 12 дБ [13], достаточно довольно грубого квантования значений $|X_n(e^{i\frac{2\pi}{N}k})|$, чтобы имитировать спектрограмму. У большинства цифровых отображающих устройств динамический диапазон гораздо шире. Следовательно, они могут дать гораздо больше спектральной информации по сравнению с аналоговыми системами.

Другое достоинство цифровых спектрограмм заключается в удобстве формирования спектра разнообразными способами, что увеличивает полезность дисплея. Примером может служить подчеркивание высоких частот, компенсирующее естественный спад спектра речи (это используется и в аналоговых спектрографах). Простой способ добиться подчеркивания высоких частот заключается в вычислении спектра первой разности входного сигнала (см. задачу 6.11). Другой, более гибкий способ, заключается в непосредственном формировании требуемого спектра до его отображения.

[1] Обычно этого добиваются, используя дополнительную память, которая служит для обновления изображения. Л. Р. Моррис изучил, однако, методы отображения спектрограмм, использующие только память и выходные устройства стандартного миникомпьютера (IEEE Tr., ASSP, June, 1975).

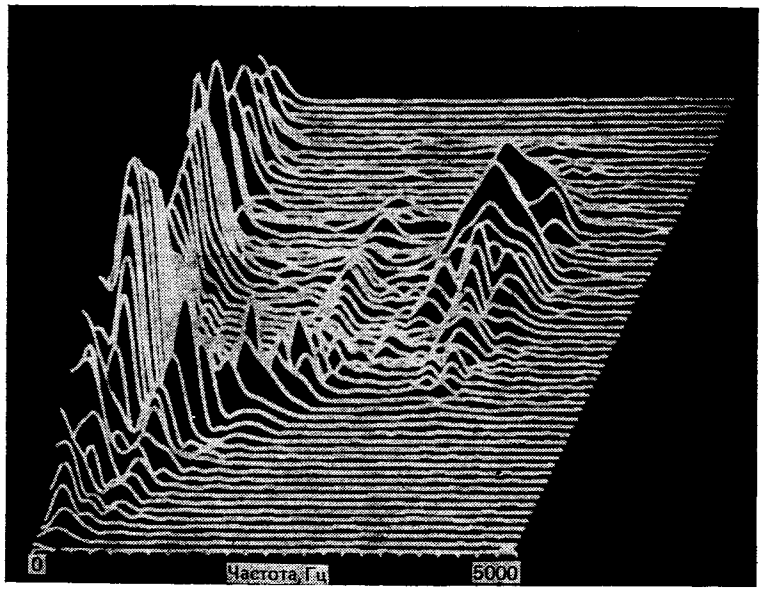


Рис. 6.38. Спектрограмма слова «read», вычисленная по сегментам речи длительностью 8 мс [12]

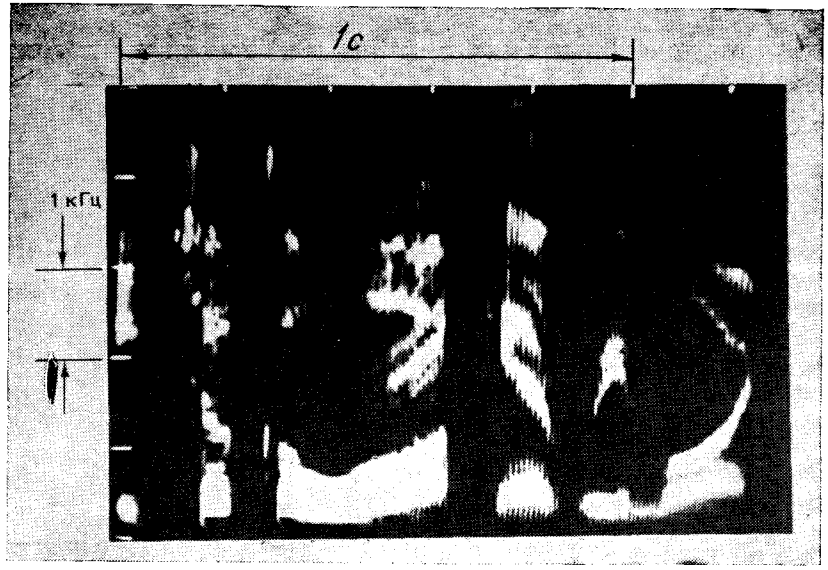


Рис. 6.39. Пример спектрограммы, полученной применением кратковременного анализа и графической системы ЭВМ [14]

Этот подход использован в [14] для создания с помощью ЭВМ спектрограмм, аналогичных показанным на рис. 6.39. Там же показано, что имеются широкие возможности отображения спектральной информации с преобразованием. Например, частотная и временна́я шкалы могут быть сжаты или растянуты по желанию.

Существует еще один подход к созданию спектрограмм с помощью ЭВМ, применяемый в тех случаях, когда нет другой возможности наглядно отобразить выходной сигнал. Если имеется печатающее устройство, допускающее повторную печать, можно получить шкалу черного, сравнимую со шкалой аналоговой спектрограммы, задавая каждый уровень набором накладываемых друг на друга печатных символов (рис. 6.40). Детали получения процедуры таких оттисков приведены в [15].

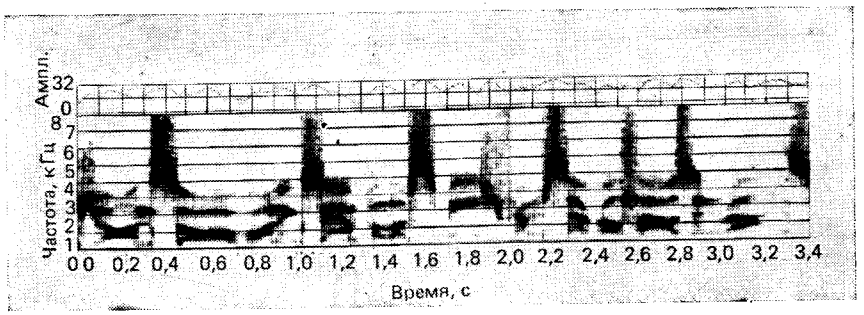


Рис. 6.40. 800-точечная ДПФ спектрограмма [15]

6.5. Выделение основного тона

При узкополосном кратковременном преобразовании Фурье возбуждение вокализованной речи проявляется узкими пиками на частотах, кратных основной частоте. Этот факт положен в основу ряда схем выделения основного тона. Рассмотрим выделитель основного тона, основанный на кратковременном спектральном анализе. Этот пример иллюстрирует как основные концепции использования кратковременного спектра для выделения основного тона, так и гибкость методов цифровой обработки. Внимательному читателю станет ясно, что существует еще много возможностей для использования кратковременного представления Фурье в задаче определения параметров возбуждения (другой пример предлагается в задаче 6.14).

Один из подходов связан с вычислением *произведения гармоник спектра*, определенного в [16] как

$$P_n(e^{i\omega}) = \prod_{r=1}^{K} |X_n(e^{i\omega r})|^2. \qquad (6.147)$$

Взяв логарифм, получим *логарифмическое произведение гармоник спектра:*

$$\hat{P}_n(e^{i\omega}) = 2\sum_{r=1}^{K} \log|X_n(e^{i\omega r})|. \qquad (6.148)$$

Видно, что $\hat{P}_n(e^{i\omega})$ представляет собой сумму K сжатых по частоте $\log|X_n(e^{i\omega})|$. Введение функции (6.148) мотивируется тем, что в вокализованной речи сжатие частотной шкалы в целые числа раз должно привести к совпадению гармоник основной частоты с ней самой. И на промежуточных частотах некоторые из сжатых по частоте гармоник будут совпадать, но всегда усиливаться они будут только на основной частоте. Схематически это приведено на рис. 6.41. Для непрерывной функции $|X_n(e^{i2\pi FTr})|$ пик на частоте F_0 становится все острее с ростом r. Следовательно, в сумме (6.148) острый пик будет на частоте F_0, возможно также появление меньших пиков на других частотах. Было найдено, что этот метод особенно устойчив к аддитивным шумам, поскольку вклад шумов в $X_n(e^{i\omega})$ не имеет коррелированной структуры, если рассматривается как функция частоты. Поэтому и в (6.148) шумовые компоненты в $X_n(e^{i\omega r})$ имеют тенденцию складываться некоррелированно. По той же причине невокализованная речь не даст явного пика в $\hat{P}_n(e^{i\omega})$. Другая важная особенность логарифмического произведения гармоник заключается в том, что на основной частоте пик вовсе не «обязан» явно присутствовать в $|X_n(e^{i\omega})|$, если он имеется в $\hat{P}_n(e^{i\omega})$. Этот метод, следовательно, привлекателен для работы с речью, пропущенной через фильтр высоких частот, как и получается в телефонной линии.

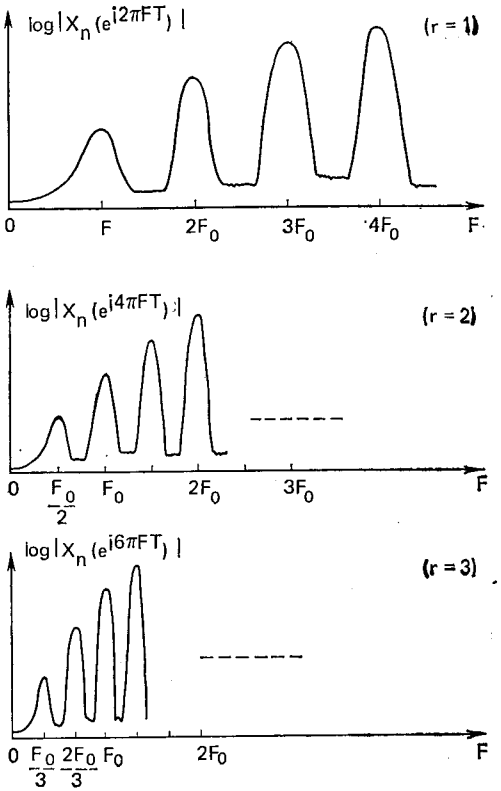


Рис. 6.41. Характер членов в логарифме произведения гармоник спектра

Пример использования метода приведен на рис. 6.42 [16]. Входная речь дискретизировалась с частотой 10 кГц и через каждые 10 мс сигнал умножался на окно Хемминга (400 отсчетов).

Затем вычислялись значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ с помощью алгоритма БПФ с $N=2048$. На рис. 6.42*а* и *б* показана последовательность произведения гармоник и ее логарифма соответственно для (6.147) и (6.148) и при $K=5$. На рис. 6.42*в* и *г* приведены результаты вы-

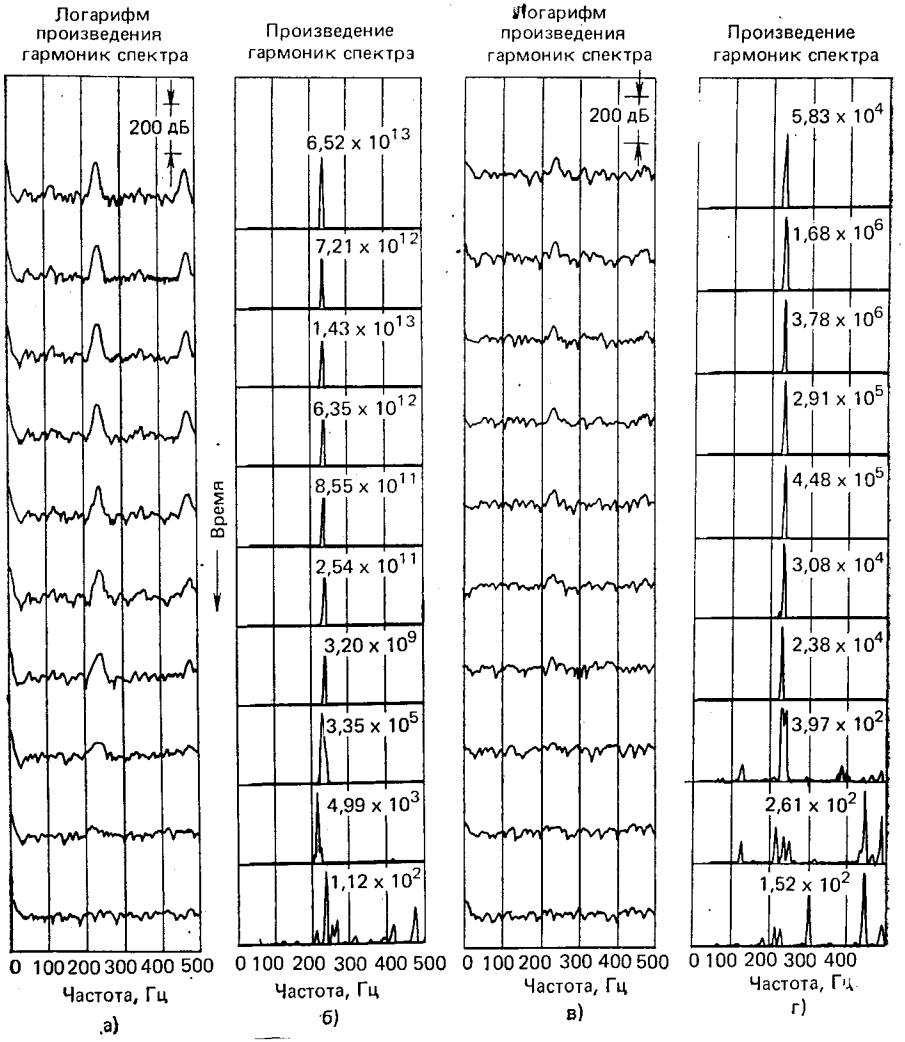


Рис. 6.42. Логарифм произведения гармоник спектра и произведение гармоник: *а*, *б*) — без шума; *в*, *г*) — отношение сигнал/шум равно 0 дБ [17]

числений с теми же параметрами, но во входной сигнал добавлен шум с отношением сигнал/шум, равным 0 дБ. Замечательна четкость, с которой проявляется основная частота. Из этих рисунков ясно, что на основе произведения гармоник можно получить простой алгоритм выделения основного тона. То, что такой алгоритм обладает великолепной устойчивостью к шуму, было проверено в [17].

6.6. Анализ через синтез

В § 6.4 и 6.5 показано, что основные параметры речи ясно проявляются в кратковременном преобразовании Фурье. В этом разделе рассмотрим метод, называемый анализом через синтез. Он оказался полезным при оценке частот формант и для оценивания глоттальных колебаний[1] вокализованной речи.

Основная идея анализа через синтез следующая. Предположим, во-первых, что имеется речевое колебание во времени или какое-либо другое представление речевого сигнала, например кратковременное преобразование Фурье. Допустим затем, что предлагается некоторая модель речеобразования. Эта модель (например, эквивалентная модель голосового тракта) имеет ряд параметров, изменяя которые можно получить различные звуки речи. Можно получить представление для модели, такое же, как и представление речевого сигнала. Если, например, речевой сигнал представлен кратковременным преобразованием Фурье, то можно также получить и кратковременное преобразование Фурье для модели. Меняя затем параметры модели некоторым систематическим образом, можно, например, найти такие значения параметров, при которых модель согласуется с речевым сигналом с минимальной ошибкой. Когда достигнуто такое согласование, считают, что параметры модели представляют собой параметры речи. Это весьма общий подход, не привязанный к кратковременному преобразованию Фурье. Принцип, однако, впервые использовался для анализа речи [18], затем был применен во временной области [19] и для кепструма [20], [21].

Одним из самых ранних описаний применения принципа анализа через синтез к речи было сделано группой в МТИ [22]. В этой работе кратковременное представление Фурье было получено с помощью гребенки аналоговых фильтров. Сигналы на выходах фильтров дискретизировались и подавались на ЭВМ. Результирующее грубое спектральное представление затем преобразовывалось по итеративной процедуре для подбора параметров в модели речеобразования. Параметры включали спектральные компоненты передаточной функции голосового тракта, форму глоттальных колебаний и сопротивление излучения. И хотя алгоритм подбора параметров модели не был полностью автоматическим, работа показала применимость метода анализа через синтез, дав отличные результаты при оценивании формант вокализованной речи [22].

Самым серьезным ограничением в схеме, описанной Беллом и другими [18], было использование аналоговой гребенки фильтров для анализа Фурье. Это ограничение снято в схеме, предложенной Мэтьюзом, Миллером и Дэйвидом [23]. Они начинали с отсчетов речевого сигнала и реализовали анализ Фурье на цифровой машине. Их подход привел к появлению еще одной новой концепции в спектральном анализе речи: понятию анализа речи, синхронного с основным тоном. И хотя работа появилась до того, как были развиты теория и практика применения дискретного анализа Фурье, воспользуемся преимуществами такого анализа для объяснения этого подхода.

6.6.1. Спектральный анализ, синхронный с основным тоном

В нашей модели короткий сегмент вокализованной речи совпадает с сегментом такой же длины периодической последовательности

$$\tilde{x}(n)=\sum_{m=-\infty}^{\infty} h_v(n+mN_p), \qquad (6.149)$$

[1] Имеется в виду форма колебания воздушного потока в голосовой щели. (*Прим. ред.*)

где $h_v(n)$ представляет свертку импульсной характеристики голосового тракта $v(n)$ с глоттальным импульсом $g(n)$ и с импульсной характеристикой сопротивления излучения $r(n)$. Иначе

$$h_v(n) = r(n) * v(n) * g(n). \tag{6.150}$$

Величина N_p есть период основного тона в отсчетах. Эффекты излучения, как правило, проявляющиеся как дифференцирование на нижних частотах, моделируются адекватно для большинства приложений просто вычислением первой разности, которой соответствует следующее z-преобразование:

$$R(z) = 1 - z^{-1}. \tag{6.151}$$

Голосовой тракт характеризуется передаточной функцией вида

$$V(z) = \frac{A}{\prod_{k=1}^{M} (1 - 2e^{-\sigma_k T}\cos(2\pi F_k T)\, z^{-1} + e^{-2\sigma_k T} z^{-2})}, \tag{6.152}$$

где число полюсов зависит от частоты дискретизации входной информации. И, наконец, глоттальный импульс имеет конечную продолжительность, что приводит к тому, что z-преобразование $g(n)$ есть полином по z вида

$$G(z) = \sum_{n=0}^{N_g} g(n) z^{-n} = B \sum_{n=1}^{N_g} (1 - z_n z^{-1}), \tag{6.153}$$

причем N_g меньше, чем N_p. Из (6.150) видно, что z-преобразование $h_v(n)$ есть

$$H_v(z) = R(z) V(z) G(z) \tag{6.154}$$

и что соответствующее преобразование Фурье имеет вид

$$H_v(e^{i\omega}) = R(e^{i\omega}) V(e^{i\omega}) G(e^{i\omega}). \tag{6.155}$$

Преобразование Фурье периодического сигнала $\tilde{x}(n)$ будет состоять из очень острых спектральных линий на частотах, кратных основной частоте.

Периодический сигнал $\tilde{x}(n)$ можно представить рядом Фурье:

$$\tilde{x}(n) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} \tilde{X}(k) e^{i\frac{2\pi}{N_p} kn}, \tag{6.156}$$

где

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right). \tag{6.157}$$

Подставив (6.156) и (6.157) в (6.1), легко показать, что

$$\tilde{X}_n(e^{i\omega}) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} H_v\left(e^{i\frac{2\pi}{N_p} k}\right) W_n\left(e^{i\left(\omega - \frac{2\pi k}{N_p}\right)}\right), \tag{6.158}$$

где $W_n(e^{i\omega})$ — преобразование Фурье анализирующего окна $w(n-m)$. Характер кратковременного преобразования Фурье сильно зависит от длины и формы анализирующего окна. Вспомнив рис. 6.2 и 6.3, мы оказываемся с точки зрения оценивания параметров модели (отличных от основного тона) перед дилеммой. При узкополосном анализе (т. е. при широком анализирующем окне) информация об огибающей спектра скрывается за пиками основного тона; напротив, при широкополосном анализе (т. е. при узком анализирующем окне) пики формант окажутся сглаженными сверткой с преобразованием Фурье окна. Более того, из (6.158) следует, что, хотя $\tilde{x}(n)$ и периодична, $\tilde{X}_n(e^{i\omega})$ представляет собой функцию положения окна. В подходе, предложенном Мэтьюзом и другими, используется то, что коэффициенты ряда Фурье периодического сигнала, такого, как в (6.149), равны просто (6.157) и могут быть вычислены по одному периоду $\tilde{x}(n)$. Имеем

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right) = \sum_{n=0}^{N_p-1} \tilde{x}(n)\, e^{-i\frac{2\pi}{N_p} kn}, \quad 0 \leqslant k \leqslant N_p - 1. \tag{6.159}$$

Следовательно, выделив один период периодического сигнала, можно вычислить отсчеты $H_v(e^{i\omega})$ в N_p разноразнесенных точках. В случае, когда используется один период вокализованной речи вместо $\tilde{x}(n)$ в (6.159), результирующее кратковременное преобразование Фурье называют кратковременным преобразованием Фурье, синхронизированным сигналом основного тона. В общем этот подход к анализу вокализованной речи называют анализом, синхронизированным сигналом основного тона.

Этот подход полностью совместим с нашими рассуждениями о кратковременном анализе Фурье, правда, несколько измененными. Во-первых, моменты, в которых вычисляются значения $X_n(e^{i\omega})$, зависят от периода основного тона речи. Поскольку основной тон меняется со временем, нам необходима теперь неравномерная во времени дискретизация. Поскольку, кроме того, число получаемых значений частот зависит от периода основного тона, частота дискретизации в частотной области также оказывается зависящей от времени. Используемое в этом случае окно обычно выбирают прямоугольным, т. е. выделяется один период речевого колебания, затем он преобразуется с помощью (6.159). Как показано в задаче 6.15, это согласуется с (6.158), поскольку для прямоугольного окна шириной N_p нули $W(e^{i\omega})$ разнесены на интервалы, кратные $2\pi/N_p$. Следовательно, когда (6.158) вычисляется на частотах $\omega_k = 2\pi k/N_p$,

$$\tilde{X}_n\left(e^{i\frac{2\pi}{N_p} k}\right) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right), \quad 0 \leqslant k \leqslant N_p - 1. \tag{6.160}$$

В рассматриваемом подходе не возникает трудностей, связанных с основным тоном в частотной области, так как они устраняются во временно́й. Следовательно, можно избежать временно́й

неопределенности узкополосного спектра, а смазывания в частотной области, присущего широкополосному анализу, можно избежать аккуратным оцениванием спектра только по N_p отсчетам.

6.6.2. Анализ полюсов и нулей модели с помощью анализа через синтез

Воспользовавшись спектром, синхронным с основным тоном, Мэтьюз и другие предложили итеративную процедуру вычисления параметров речи. Они пользовались эквивалентной моделью для передаточных функций сопротивления излучения, голосового тракта и глоттальных импульсов. Это привело к поправочному множителю с числом полюсов, большим, чем, вероятно, было бы нужно, если бы они пользовались соотношениями (6.151)—(6.154). Основной подход остается тем же независимо от конкретной функциональной формы модели речи, так что мы можем продолжить наше изложение, применив цифровую модель (по поводу используемых функций см. [23]).

Параметры $H_v(e^{i\omega})$ можно определить с помощью некоторой итеративной процедуры аппроксимации. Мэтьюз и другие ввели

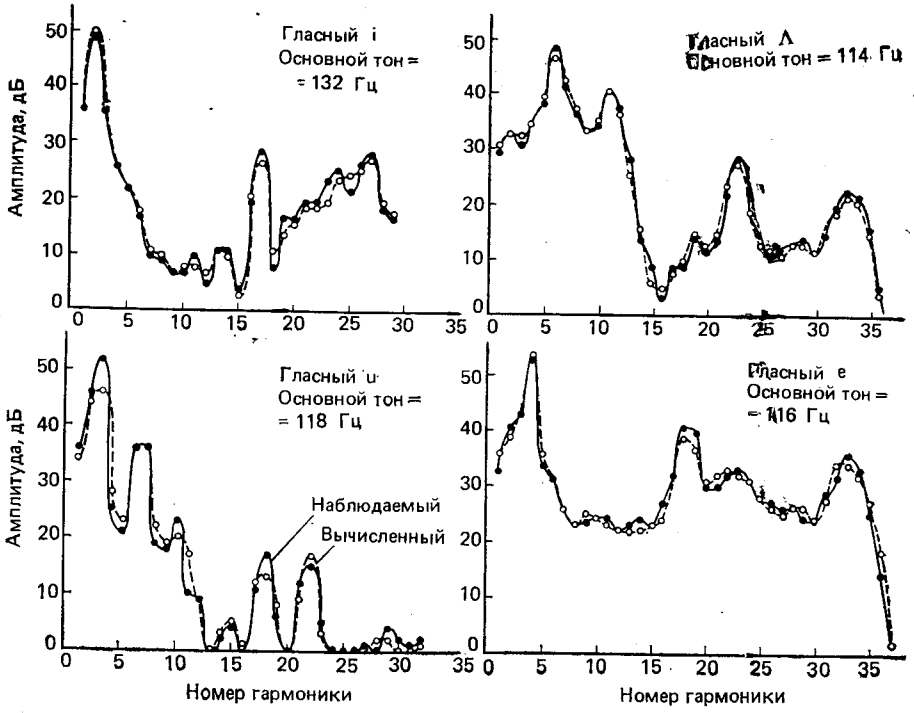


Рис. 6.43. Согласование спектра, синхронное с основным тоном: — наблюдаемый спектр;
— — — согласованный спектр [23]

ряд параметров, вычислили их значения на частотах $2\pi k/N_p$, а затем оценили функцию ошибки в виде

$$E=\sum_k Q(k)\left[\log\left|H_v\left(e^{i\frac{2\pi}{N_p}k}\right)\right|-\log\left|X_n\left(e^{i\frac{2\pi}{N_p}k}\right)\right|\right]^2, \quad (6.161)$$

где $Q(k)$ — весовая функция спектра, а $X_n\left(e^{i\frac{2\pi}{N_p}k}\right)$ — спектр речевого сигнала, синхронный с основным тоном. Параметры подбирались систематическим образом, так чтобы минимизировать функцию ошибки. Правила подбора полюсов $V(z)$ и нулей $G(z)$ рассмотрены в [23]. На рис. 6.43 показаны некоторые примеры согласования спектров, описанные в [23]. Когда ошибка минимизирована, значения полюсов $V(z)$ принимаются за оценки частот формант. Положение нулей дает информацию о глоттальных колебаниях.

6.6.3. Оценивание глоттальных колебаний, синхронное с основным тоном

Работа Мэтьюза, Миллера и Дэйвида связана, прежде всего, с распределением нулей. Были предприняты попытки связать разнесение нулей с формой глоттального импульса. В более поздней (неопубликованной) работе Миллера и Мэтьюза метод модифицирован так, чтобы получить оценки глоттального импульса. В этом случае модель имела вид

$$H_v(z)=R(z)\,G_f(z)\,V(z), \quad (6.162)$$

причем вклад в спектр глоттального колебания первоначально моделировался фиксированной передаточной функцией

$$G_f(z)=\frac{1}{(1-az^{-1})(1-bz^{-1})}. \quad (6.163)$$

Параметры $V(z)$ снова варьировались так, чтобы получить минимум критерия ошибки. Результирующее распределение полюсов служило оценкой частот формант. Чтобы получить форму глоттального колебания на анализируемом периоде речи, Миллер и Мэтьюз вычисляли величину

$$\tilde{G}(k)=\frac{X_n\left(e^{i\frac{2\pi}{N_p}k}\right)}{R\left(e^{i\frac{2\pi}{N_p}k}\right)V\left(e^{i\frac{2\pi}{N_p}k}\right)}, \quad 0\leqslant k\leqslant N_p-1. \quad (6.164)$$

Значения $\tilde{G}(k)$ при $0\leqslant k\leqslant N_p-1$ использовались в качестве коэффициентов Фурье глоттального импульса $g(n)$, который вычислялся с помощью обратного ДПФ:

$$\tilde{g}(n)=\frac{1}{N_p}\sum_{k=0}^{N_p-1}\tilde{G}(k)\,e^{i\frac{2\pi}{N_p}kn}. \quad (6.165)$$

Это выполнимо (поскольку $\tilde{g}(n)$ есть импульс конечной длительности) даже несмотря на то, что в общем случае наличия N_p отсчетов $H_v(e^{i\omega})$, получаемых синхронно с основным тоном, недостаточно для полного задания последовательности $h_v(n)$, которая, вообще говоря, длиннее чем N_p. Таким образом, с помощью модели речеобразования можно выделить из свертки компоненту конечной продолжительности. Этот метод с успехом применялся Розенбергом [24], изучавшим влияние формы глоттального импульса на качество звучания гласной. На рис. 6.44 показан пример речевого колебания и соответствующего глоттального колебания, выделенного указанной выше процедурой.

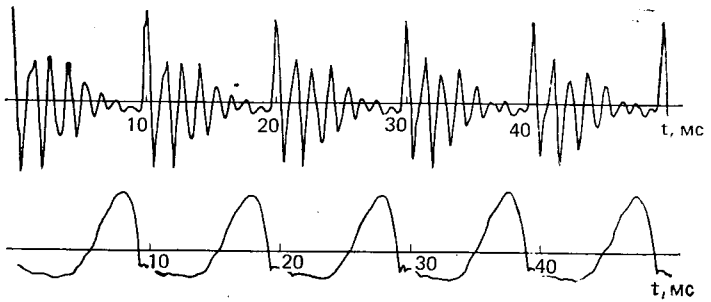


Рис. 6.44. Речь (вверху) и анализируемое колебание возбуждения (внизу) для гласной в слове «hod» [24]

Этот метод оценки глоттального импульса чувствителен к типу выбранной модели. В случаях, когда модель хорошо увязывается с речевым сигналом, как в случае установившихся гласных, получаются отличные результаты. В других ситуациях необходима более сложная модель. Другим фактором, влиявшим на результаты, был способ выделения из речевого колебания периода основного тона. Требуется большая осторожность при определении начала и конца периода. Речевой сигнал интерполировался с большей частотой дискретизации, чтобы облегчить поиск точного положения начала каждого цикла. Не удивительно, что это потребовалось, так как весьма мало вероятно, чтобы момент открытия (и смыкания) голосовой щели совпадал с моментом отсчета.

6.7. Системы анализа—синтеза

До сих пор в этой главе рассматривалась основная теория кратковременного анализа и синтеза Фурье. Было показано, что кратковременное преобразование Фурье может служить основой множества схем оценки параметров модели речеобразования. Однако мы еще не обсудили практического применения того факта, что речевой сигнал может быть точно восстановлен по его кратковременному Фурье-представлению. Этот факт лежит в основе схем кодирования речи, называемых вокодерами. Основная задача вокодеров состоит в цифровом представлении речи с гораздо меньшей скоростью, чем это возможно для схем непосредственного кодирования колебаний. В других случаях использование вокодеров позволяет удалять аддитивный шум и эффект реверберации, а также изменять основные параметры речи с преобразованием масштабов по времени и частоте.

В этом параграфе будут рассмотрены некоторые схемы кодирования речи, основанные на теоретических принципах § 6.1—6.3. Начнем с системы прямой реализации кратковременного анализа и синтеза Фурье, а затем обсудим такие системы, как полосный вокодер. Это не совпадает с историей развития вокодеров, однако при таком подходе станут яснее причины ухудшения качества, связанные с упрощением реализации.

6.7.1. Цифровое кодирование кратковременного преобразования Фурье

В § 6.1 и 6.2 было показано, что речь (практически, любой сигнал) можно точно представить набором полосовых каналов (рис. 6.45*а*). Центральные частоты ω_k и анализирующие окна

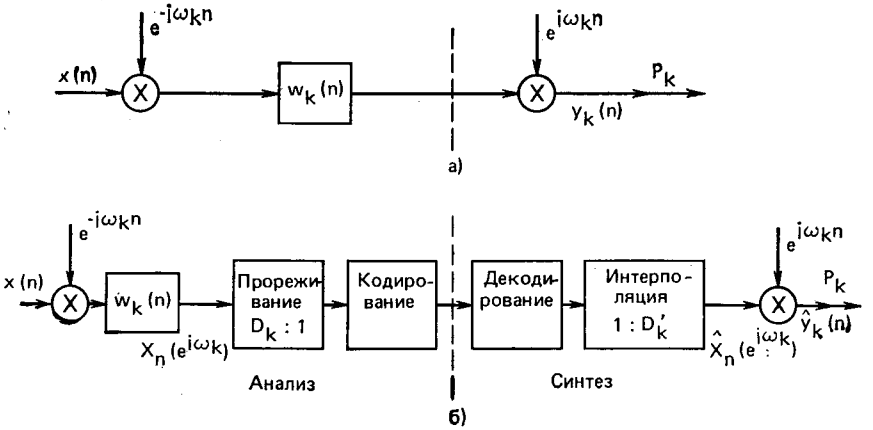


Рис. 6.45. Структурная схема кодирования одного канала процедуры анализ — синтез

$w_k(n)$ выбраны так, чтобы перекрыть нужную полосу частот, а комплексные константы $P_k = |P_k| e^{i\Phi_k}$ — чтобы общая характеристика суммы всех каналов была возможно ближе к идеальной со строго плоской амплитудно-частотной и линейной фазо-частотной характеристиками. В 6.13 было показано, что, поскольку $w_k(n)$ соответствует импульсной характеристике фильтра нижних частот, кратковременное преобразование Фурье на частотах ω_k может быть дискретизировано с частотой, меньшей, чем входной сигнал. Действительно, полное число необходимых отсчетов в секунду для $X_n(e^{i\omega_k})$ можно сделать равным частоте дискретизации входного сигнала. Поэтому чтобы снизить скорость вычислений, необходимую при реализации анализа, сигналы каналов дискретизируются с гораздо меньшей частотой, квантуются и кодируются для последующей передачи или хранения. Для одного канала это показано на рис. 6.45*б*. Операции анализа, показанные слева от пунктирной линии, выполняются модулятором, за которым следуют фильтр нижних частот, прореживатель, кодирую-

щее устройство. Когда $w_k(n)$ представляет собой последовательность конечной длины, операция прореживания просто совмещается с линейной фильтрацией, т. е. выход просто вычисляется для каждых D_k отсчетов на входе. Кодирование включает в себя квантование и собственно кодирование (см. гл. 5). При синтезе цифровое представление сначала декодируется, а затем вычисляется квантованная версия $X_n(e^{i\omega_k})$ с помощью интерполяции. Если опущены высокочастотные каналы, то можно при синтезе выходного колебания использовать частоту дискретизации, меньшую, чем исходная частота дискретизации входа. Следовательно, коэффициент интерполяции D_k может быть меньше коэффициента прореживания D_k. Квантованный канальный сигнал кратковременного преобразования Фурье $\hat{X}_n(e^{i\omega_k})$ модулирует комплексную синусоиду, образуя сигнал $\hat{y}_k(n)$, который складывается затем с другими:

$$\hat{y}(n)=\sum_{k=0}^{N-1} P_k\,\hat{y}_k(n). \qquad (6.166)$$

Чтобы иллюстрировать практические соображения по поводу такого кодирования речи, рассмотрим пример из [6]. Вычисление $X_n(e^{i\omega_k})$ реализуется с помощью БПФ (см. § 6.3). Поскольку программа БПФ требует, чтобы N было степенью двойки, вход дискретизировался с довольно необычной частотой 12195 отсч./с. Значение N равнялось 128, так что частотами анализа были $\omega_k=2\pi k/128$, что соответствует аналоговым частотам $F_k=(95,273k)$ Гц. Анализирующее окно $w(n)$ (одинаковое для всех каналов) имело импульсную характеристику КИХ-фильтров длиной 731 отсчет (с линейной фазо-частотной характеристикой). Оно было спроектировано методом частотной выборки [25]. Частотная характеристика фильтра приведена на рис. 6.46. Заметим,

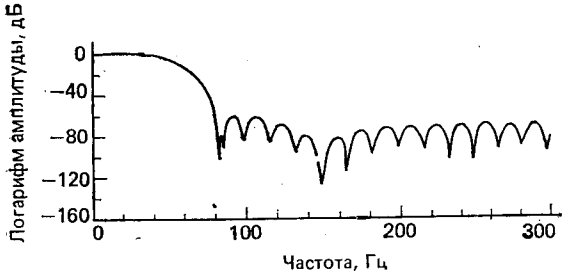


Рис. 6.46. Частотная характеристика анализирующего окна [6]

что выше 80 Гц затухание фильтра составляло не меньше 60 дБ. При подборе надлежащих комплексных констант P_k оказалось возможным получить общую характеристику, приведенную на рис. 6.47. Отметим, что $P_0=0$ и $P_k=0$ при $28<k<100$. Поскольку полоса перекрываемых синтезатором частот содержит частоты только до 2690 Гц, на выходе использовалась частота дискретизации 10004 отсч./с. (На выходе можно было бы использовать еще меньшую частоту дискретизации — примерно 6000 отсч./с,

если бы применялись соответственно более избирательные аналоговые фильтры.)

Результат проектирования показан на спектрограммах (рис. 6.48). На рис. 6.48*а* приведена фраза, поступающая на вход, а на рис. 6.48*б* — выходной сигнал системы анализа — синтеза, состоящей из 28 каналов (см. рис. 6.18*б*). Комплексные константы равны единице при $1\leqslant k\leqslant 28$ и $100\leqslant k\leqslant 127$ и нулю во всех остальных случаях, т. е. никакой специальной фазовой компенсации не применялось. Канальные сигналы дискретизировались с частотой Найквиста (т. е. 160 раз в секунду), что обеспечило точное восстановление на стадии синтеза. Квантование не проводилось, т. е. $X_n(e^{i\omega_k})$ представлялась с точностью 16 разрядов. Сравнение широкополосных спектрограмм (которые обладают хорошим разрешением по времени) выявляет эффект, согласующийся с видом импульсной характеристики рис. 6.25. Нечеткость спектрограммы, изображенной на рис. 6.48*б*, связана с задержкой энергии сигнала эха; искажения такого рода воспринимаются как реверберация. Прослушивание отчетливо выявило эффект «пустой бочки» в звучании сравниваемых фраз, соответствующих рис. 6.48*а* и *б*. Когда же фразы были должным образом подобраны (см. 6.2.1 и рис. 6.47), спектрограмма на выходе (рис. 6.48*в*) неотличима от спектрограммы на входе (рис. 6.48*а*); соответственно речевые сигналы на входе и выходе оказались неразличимыми при восприятии.

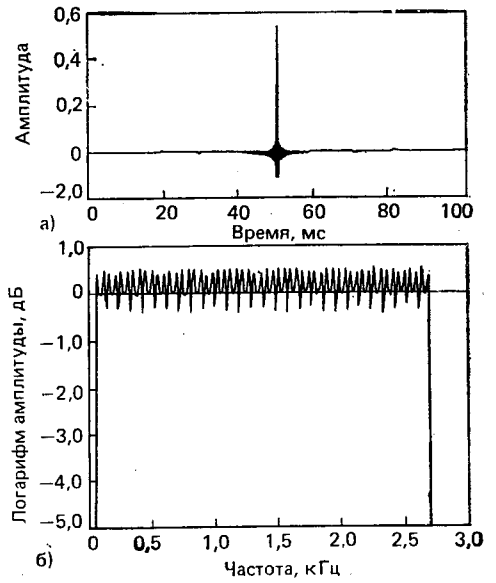


Рис. 6.47. Временна́я и частотная характеристики гребенки фильтров [6]

Уведившись в том, что можно действительно восстановить речевой сигнал по его кратковременному преобразованию Фурье, обратимся к способам кодирования канальных сигналов для цифровой передачи или хранения. В гл. 5 было установлено, что при кодировании любого колебания имеются два основных параметра: частота дискретизации и число бит, требуемых на отсчетах. Произведение этих двух величин дает скорость, которую, как правило, минимизируют, применяя минимальную допустимую частоту дискретизации и минимальное число бит на отсчет. В нашем случае общая информационная скорость представляет собой сумму скоростей для каждого из канальных сигналов в каналах.

Прежде чем перейти к квантованию, полезно рассмотреть эффекты снижения частоты дискретизации канальных сигналов. Вспомним, что действительная и мнимая части $X_n(e^{i\omega_k})$ представляют собой выходы фильтров нижних частот. Следовательно, в

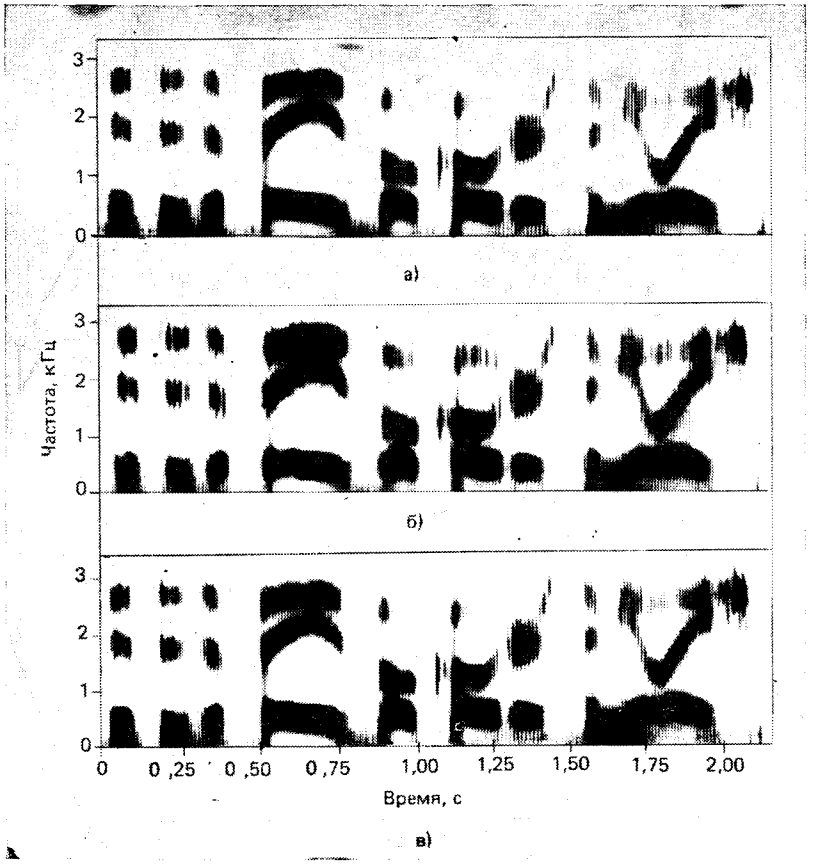


Рис. 6.48. Иллюстрация операции без квантования при $1/T_1 = 160$ Гц; *а*) речь на входе; *б*) речь на выходе без подстройки фазы; *в*) речь на выходе при наилучшей подстройке фазы [6]

рассматриваемом примере (частотная характеристика которого приведена на рис. 6.46) возникнут лишь пренебрежимо малые наложения при частотах дискретизации не меньше 160 Гц, поскольку характеристика фильтра имеет затухание по крайней мере 60 дБ на частотах выше 80 Гц. Если использовать меньшую частоту дискретизации без соответствующего сокращения полосы фильтров, то возникнут наложения во временно́й области. Если уменьшить полосу, не уменьшая разноса каналов, то синтезированная речь окажется более реверберирующей, поскольку характеристики отдельных каналов не будут перекрываться и в спектре синтезированной речи появятся «дыры». Уменьшить разнос каналов без увеличения их числа невозможно. Поэтому, пытаясь сни-

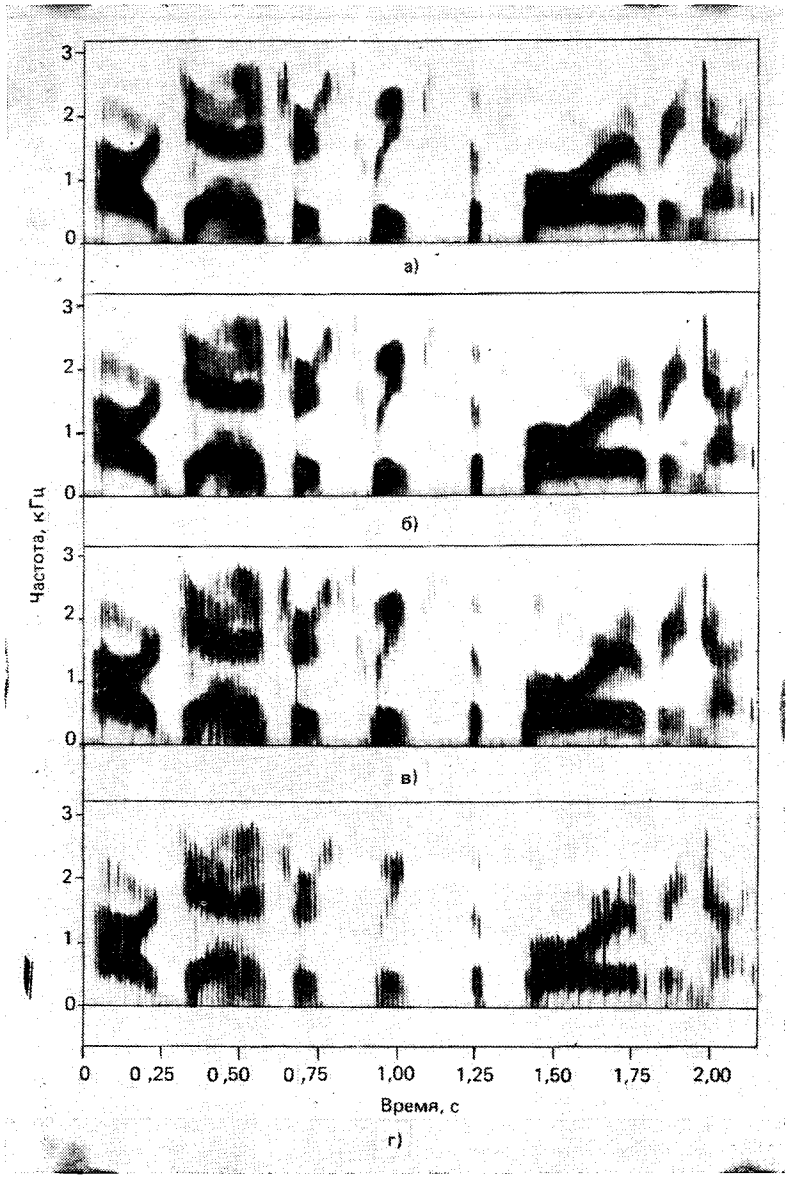


Рис. 6.49. Иллюстрация эффекта наложения при ИКМ (нижняя частота среза 80 Гц, без квантования) при $1/T_1 = 160$ Гц (*а*), 100 Гц (*б*), 80 Гц (*в*) и 60 Гц (*г*) [6]

зить информационную скорость уменьшением частоты дискретизации канальных сигналов, мы должны быть готовы к тому, чтобы смириться либо с искажениями из-за наложений во временно́й области, возникающими вследствие понижения частоты дискретизации, либо с повышенной реверберацией, вызванной сужением полос фильтров.

Оба эти эффекта показаны на рис. 6.49 и 6.50. Первый из них иллюстрирует возникновение эффекта наложения из-за слишком

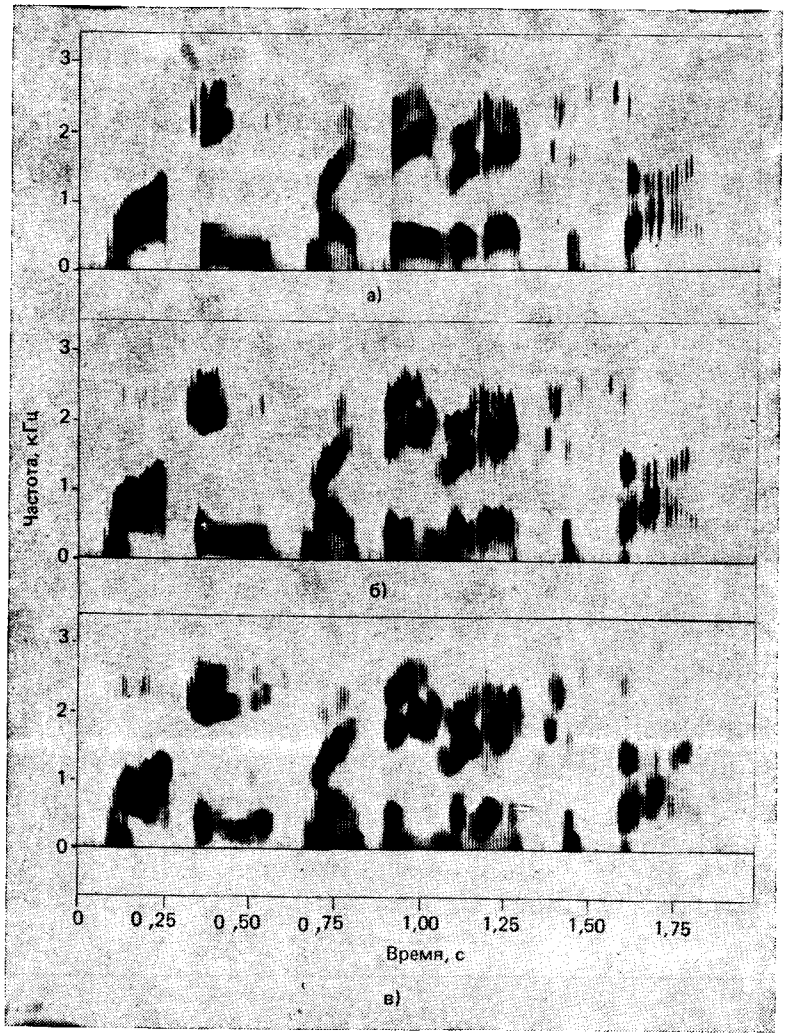


Рис. 6.50. Иллюстрация эффекта узкополосных анализирующих фильтров в ИКМ без квантования при $1/T_1 = 160$ Гц и нижней частоте среза 80 Гц (*а*), 53 Гц (*б*) и 11 Гц (*в*) [6]

малой частоты дискретизации кратковременного спектра. В случаях рис. 6.49*в* и *г* наблюдаются значительные искажения, которые менее заметны в случае рис. 6.49*б*. Сравнив рис. 6.49*г* со спектрограммой исходной речевой фразы (рис. 6.48*а*), видим, что в случаях серьезных искажений из-за наложений во временно́й области основной тон сильно искажается, тогда как частоты формант остаются почти без изменений. Рисунок 6.50 является иллюстрацией эффекта сужения полосы анализирующего фильтра при постоянном разнесении частот каналов. Во всех случаях частота дискретизации канальных сигналов равнялась 160 Гц. В случае, соответствующем рис. 6.50*а*, фильтр был таким же, как и при получении спектра, изображенного на рис. 6.46, т. е. выше частоты 80 Гц затухание фильтра было не меньше 60 дБ. На рис. 6.50*б* соответствующая частота среза равнялась 53 Гц, а на рис. 6.50*в* — 36 Гц. Как и ожидалось, спектрограммы на рис. 6.50*б* и *в* демонстрируют существенное ухудшение качества, что можно отнести за счет реверберации, вызванной расширением эффективной ширины окна. Кроме этого, можно видеть, что, хотя основной тон сигнала остался неизменным, траектории формант сильно пострадали из-за реверберации, вызванной узкополосными фильтрами. По разборчивости искажения, вызванные наложениями во временно́й области, предпочтительнее реверберции, вызванной сужением полос фильтров.

Для того чтобы определить информационную скорость, требуемую для представления речи с помощью кратковременного преобразования Фурье, необходимо выбрать схему квантования. Для квантования действительной и мнимой частей комплексных сигналов можно применить бо́льшую часть схем гл. 5. Два примера, рассматриваемые в [6], используют адаптивную дельта-модуляцию и ИКМ. В системе адаптивной дельта-модуляции [26] 28 каналов кодировались битом на отсчет. Полная скорость системы оказалась в 56 раз больше скорости (частоты) дискретизации канальных сигналов, поскольку нужно было кодировать и действительную и мнимую части канальных сигналов. Так как в системе адаптивной дельта-модуляции частота дискретизации в 5—10 раз превышает частоту Найквиста, можно ожидать, что для достижения хороших результатов потребуется скорость 20—30 кбит/с. На рис. 6.51 приведены примеры кодирования с помощью адаптивной дельта-модуляции для ряда скоростей. На рис. 6.51*а* полная скорость 28 кбит/с соответствует частоте дискретизации 500 отсч./с, на 6.51*б* полная скорость 21 кбит/с соответствует частоте дискретизации 375 отсч./с и на рис. 6.51*в* полная скорость 14 кбит/с соответствует частоте дискретизации 250 отсч./с. Из рис. 6.51 видно, что хорошее качество передачи достигается при скорости 28 кбит/с и что оно быстро ухудшается при меньших скоростях. Альтернативой кодированию АДМ может служить кодирование АИКМ, которое использовалось Крошьером [27] при неравномерном анализе и позволяло осуществлять пе-

редачу с хорошим качеством по четырем-пяти каналам при скорости около 16 кбит/с.

Как пример кодирования ИКМ, та же система с 28 каналами использовалась с частотой дискретизации канальных сигналов 100 отсч./с, т. е. допускался небольшой уровень искажений из-за

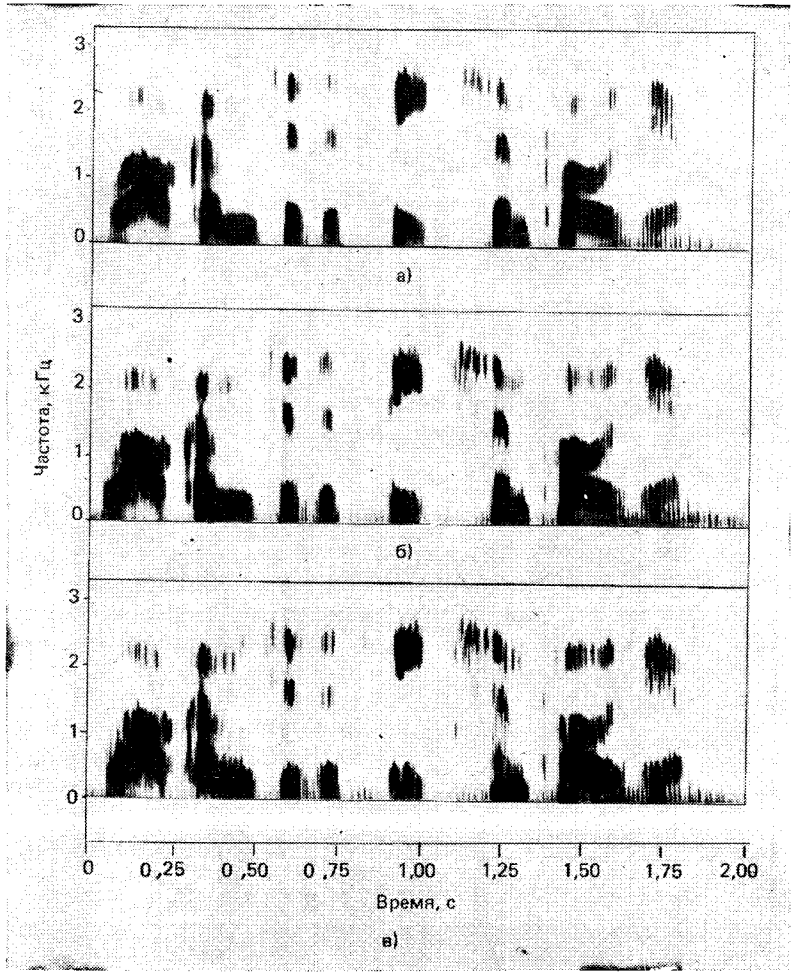


Рис. 6.51. Кодирование параметров спектра адаптивной дельта-модуляцией: *а*) 28 кбит/с, $1/T_1=500$ Гц; *б*) 21 кбит/с, $1/T_1=375$ Гц; *в*) 14 кбит/с, $1/T_1=250$ Гц [6]

наложения с тем, чтобы понизить частоту дискретизации. Квантование логарифмов модуля и фазы кратковременного преобразования Фурье было признано предпочтительным по сравнению с кодированием действительной и мнимой частей комплексных ка-

нальных сигналов. Для того чтобы воспользоваться слабой чувствительностью слуха на высоких частотах, сигналы низкочастотных каналов отображались точнее, чем сигналы высокочастотных каналов. При этом скорость передачи составляла 16 кбит/с; в каналах 1—10 использовалось 3 бит на логарифм амплитуды и 4 на фазу, а в каналах 11—28 соответственно 2 и 3 бит. На рис. 6.52*а* показана широкополосная спектрограмма входного речевого сигнала, на рис. 6.52*б* — результат кодирования со скоростью 16 кбит/с.

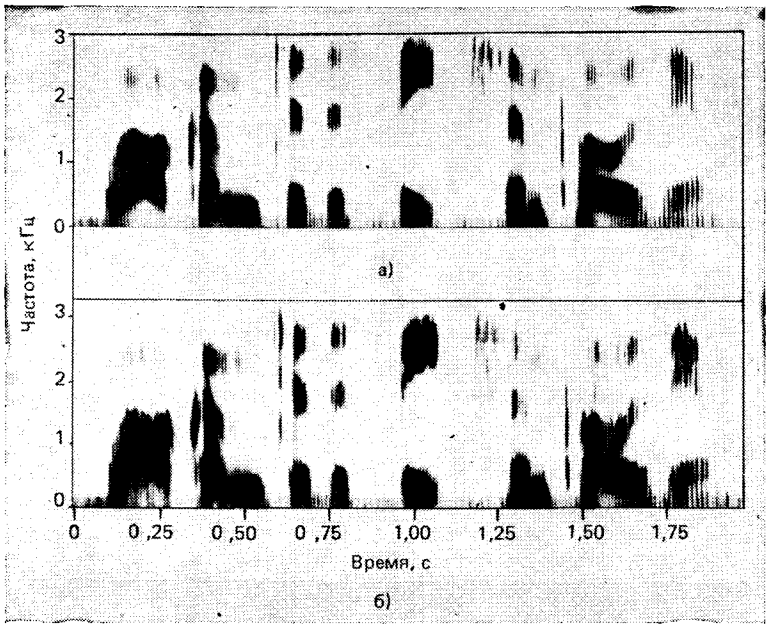


Рис. 6.52. Операции с квантованием:
а) речь на входе; *б*) $1/T_1=100$ Гц (полная скорость 16 кбит/с) [6]

Информационные скорости при использовании цифрового кодирования кратковременного преобразования Фурье сравнительно велики и сравнимы со скоростями непосредственного кодирования речевых колебаний методами адаптивного квантования. Сложность кратковременного преобразования Фурье, конечно, гораздо выше, чем в большинстве систем кодирования непрерывных сигналов. Основным преимуществом представления с помощью кратковременного преобразования Фурье является наличие гибкости, с которой можно манипулировать параметрами речевого сигнала. Это станет очевидным из последующих рассуждений.

Соберем воедино все, что уже изложено относительно схем кодирования речи, основанных на кратковременном преобразовании Фурье. Во-первых, при дискретизации канальных сигналов

с достаточно большой частотой и без квантования отсчетов (достаточно 12 бит/отсч.) можно достичь безукоризненного на слух воспроизведения речи. Скорость, необходимая для передачи такого представления, однако, довольно велика. Действительно, в примере, который был приведен ранее, высококачественное воспроизведение сигнала с полосой частот 3 кГц требовало гораздо меньшей скорости (около 100 бит/с). Скорость передачи можно снизить двумя способами. Во-первых, допустимо более грубо квантовать канальные сигналы и снижать частоту их дискретизации. В этом случае удается достигнуть скорости передачи 16 кбит/с при незначительном ухудшении качества восприятия. Во-вторых, можно использовать свойства речи, удалив часть избыточности сигнала. Ухудшение качества восприятия речи получается, кроме всего прочего, из-за того, что для упрощения реализации системы анализа—синтеза вводится ряд приближений. Такого рода ухудшение воспринимается как изменение разборчивости, отличное от искажений в системах прямого кодирования, представимых, как правило, аддитивным (возможно коррелированным с сигналом) шумом. Следовательно, измерения отношения сигнал/шум (см. гл. 5) не имеют смысла в системах типа вокодеров. По этой причине приходится описывать качество восприятия речи в вокодерах, сравнивая спектрограммы, или по субъективным оценкам искажений, воспринимаемых слушателями.

6.7.2. Фазовый вокодер [1]

Фазовый вокодер представляет собой интересный новый подход к анализу, основанному на кратковременном спектре [28]. Чтобы понять, как работает эта система, рассмотрим отклик в одном канале. Для этого удобно представить систему, изображенную на рис. 6.45*а*, через действительные операции, как это сделано на рис. 6.53. Вспомнив, что обычно выбираются $\omega_{N-k}=2\pi-\omega_k$ и $P_k=P^*_{N-k}$, видим, что мнимые части сокращаются и остаются только действительные, которые легко представить в виде

$$R_e[P_k y_k(n)] = |P_k|\left|X_n\left(e^{i\omega_k}\right)\right|\cos[\omega_k n+\theta_n(\omega_k)+\gamma_k]. \qquad (6.167)$$

Следовательно, общий сигнал на выходе образуется как сумма сигналов. Такие сигналы можно интерпретировать как дискретные косинусоидальные колебания, модулированные по амплитуде и по частоте кратковременным преобразованием Фурье сигналов в канале. Значение $|P_k|$ равно, вообще говоря, единице или нулю, в зависимости от того, участвует ли в суммировании соответствующий канал. Фазовые константы введены для того, чтобы добиться максимально-плоской общей характеристики.

[1] Фазовый вокодер был предложен и исследован Фланаганом и Голденом [28]. Результаты настоящего раздела основаны на этой работе.

Вводя понятие мгновенной частоты, получим полезную интерпретацию (6.167). Удобно рассмотреть непрерывное зависящее от времени преобразование Фурье:

$$X_a(t,\ \Omega_k) = |X_a(t,\ \Omega_k)|\, e^{i\theta_a(t,\ \Omega_k)} = \qquad (6.168а)$$

$$= a_a(t,\ \Omega_k) - i\, b_a(t,\ \Omega_k), \qquad (6.168б)$$

где

$$|X_a(t,\ \Omega_k)| = [a_a^2(t,\ \Omega_k) + b_a^2(t,\ \Omega_k)]^{1/2} \qquad (6.169а)$$

и

$$\theta_a(t,\ \Omega_k) = -\mathrm{tg}^{-1}\left[\frac{b_a(t,\ \Omega_k)}{a_a(t,\ \Omega_k)}\right]. \qquad (6.169б)$$

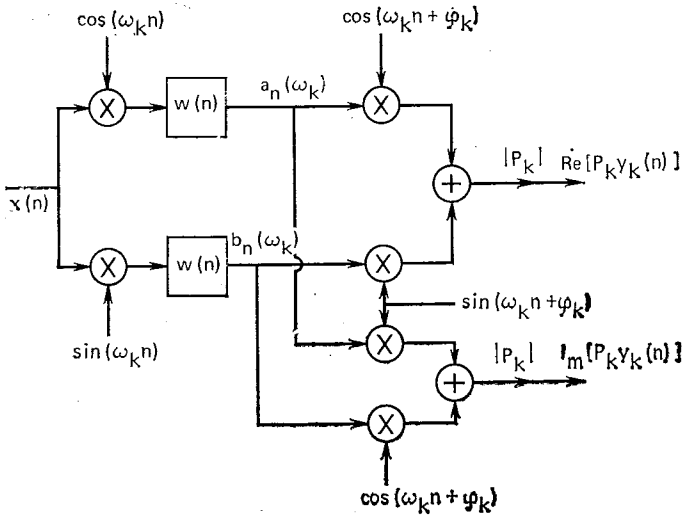


Рис. 6.53. Реализация одного канала фазового вокодера

Это зависящее от времени преобразование Фурье с непрерывным временем можно определить как

$$X_a(t,\ \Omega_k) = \int_{-\infty}^{\infty} x_a(\tau)\, w_a(t-\tau)\, e^{-i\Omega_k \tau}\, d\tau, \qquad (6.170)$$

где $x_a(\tau)$ — речевой сигнал с непрерывным временем, а $w_a(\tau)$ — анализирующее окно с непрерывным временем, или, что то же, импульсная характеристика аналогового фильтра нижних частот. Величина

$$\dot{\theta}_a(t,\ \Omega_k) = \frac{d\,\theta_a(t,\ \Omega_k)}{dt}, \qquad (6.171)$$

называемая фазовой производной, представляет собой мгновенную девиацию частоты от центральной частоты Ω_k в k-м канале.

Производную фазы можно выразить через $a_a(t, \Omega_k)$ и $b_a(t, \Omega_k)$:

$$\dot{\theta}_a(t,\ \Omega_k) = \frac{b_a(t,\ \Omega_k)\,\dot{a}_a(t,\ \Omega_k) - a_a(t,\ \Omega_k)\,\dot{b}_a(t,\ \Omega_k)}{a_a^2(t,\ \Omega_k) + b_a^2(t,\ \Omega_k)}, \qquad (6.172)$$

где точки сверху означают дифференцирование по времени. В случае обработки сигнала с дискретным временем предполагается, что $x_a(t)$ и $X_a(t, \Omega_k)$ ограничены по частоте и что $X_n(e^{i\omega_k})$ представляет собой дискретизованное кратковременное преобразование Фурье с непрерывным временем:

$$X_n\left(e^{i\omega_k}\right) = X_a(nT,\ \omega_k/T). \qquad (6.173)$$

Аналогичным образом, фазовая производная $X_n(e^{i\omega_k})$ определяется как дискретизованная версия $\dot{\theta}_a(t, \Omega_k)$:

$$\dot{\theta}_n(\omega_k) = \frac{b_n(\omega_k)\,\dot{a}_n(\omega_k) - a_n(\omega_k)\,\dot{b}_n(\omega_k)}{a_n^2(\omega_k) + b_n^2(\omega_k)}. \qquad (6.174)$$

В этом случае $\dot{a}_n(\omega_k)$ и $\dot{b}_n(\omega_k)$ предполагаются последовательностями, полученными при дискретизации соответствующих производных с непрерывным временем и ограниченных по частоте. Эти сигналы производных можно получить цифровой фильтрацией последовательностей $a_n(\omega_k)$ и $b_n(\omega_k)$ (см. задачу 6.16).

Чтобы понять, почему представляют интерес сигналы производной фазы, рассмотрим случай, когда центральные частоты каналов мало разнесены. В частности, возьмем случай, когда основной тон постоянен и всего одна гармоника основной частоты попадает в полосу k-го канала. Легко видеть, что $|X_n(e^{i\omega_k})|$ отражает медленно меняющуюся амплитудно-частотную характеристику голосового тракта на частоте вблизи ω_k. Производная фазы будет константой, равной отклонению частоты гармоники от центральной частоты. Если теперь характеристика голосового тракта и основной тон медленно меняются, так, как это бывает при обычной речи, допустимо предположить, что медленно меняются и амплитудный спектр, и производная фазы. Действительно, для восприятия эффекты наложения при дискретизации переменного амплитудного спектра и производной фазы окажутся менее заметными, чем аналогичные эффекты при дискретизации действительной и мнимой частей кратковременного преобразования Фурье [28].

При синтезе $\theta_a(t, \Omega_k)$ получается из $\dot{\theta}_a(t, \Omega_k)$ интегрированием

$$\theta_a(t,\ \Omega_k) = \int_{t_0}^{t} \dot{\theta}_a(\tau,\ \Omega_k)\,d\tau + \theta_a(t_0,\ \Omega_k). \qquad (6.175)$$

Из приведенного равенства следует, что $\theta_n(\omega_k)$, представляющее собой дискретизованную версию $\theta_a(t, \Omega_k)$, окажется более сглаженной, чем $\dot{\theta}_n(\omega_k)$. Поэтому можно предположить, что $\theta_n(\omega_k)$ допустимо дискретизовать с еще меньшей частотой, чем $\dot{\theta}_n(\omega_k)$. Однако мы пренебрегаем тем, что величина $\theta_n(\omega_k)$ не ограничена и, следовательно, непригодна для квантования. (В этом нетрудно убедиться, рассмотрев случай постоянного основного тона.) Ограниченную фазу можно получить, вычисляя главное значение, т. е. ограничив $\theta_n(\omega_k)$ значениями в интервале от 0 до 2π или от $-\pi$ до π. Главное значение фазы окажется, к сожалению, «разрывным» (т. е. главное значение $\theta_a(t, \Omega_k)$ будет разрывной функцией t), а следовательно, не будет сигналом, спектр которого ограничен по частоте в фильтре нижних частот. Разрывность главного значения фазы не означает того, что фазу нельзя квантовать, поскольку единственное, что требуется, это восстановить при подходящей частоте дискретизации соответствующие действительную и мнимую части $X_n(e^{i\omega_k})$. Поэтому частота дискретизации для $\theta_n(\omega_k)$ должна быть не меньше частот, необходимых для $a_n(\omega_k)$ и $b_n(\omega_k)$. В действительности именно главное значение фазы квантовалось в 6.7.1.

Применение производной фазы в системе анализа—синтеза помимо достоинства — гладкости — обладает и рядом недостатков, что видно из (6.175), так как для восстановления $\theta_n(\omega_k)$ по $\dot{\theta}_n(\omega_k)$ необходимы начальные условия. Обычно подобные начальные условия нам неизвестны, а положив произвольно начальную фазу равной нулю, мы получим систематические ошибки по фазе φ_k. Это может привести к существенному отклонению общей характеристики системы анализа—синтеза от идеальной (с плоской амплитудно-частотной и линейной фазо-частотной характеристиками), а синтезированная речь будет звучать с сильными искажениями типа реверберации.

На рис. 6.54 изображен анализатор вокодера, основанный на использовании амплитудного спектра и производной фазы. Здесь

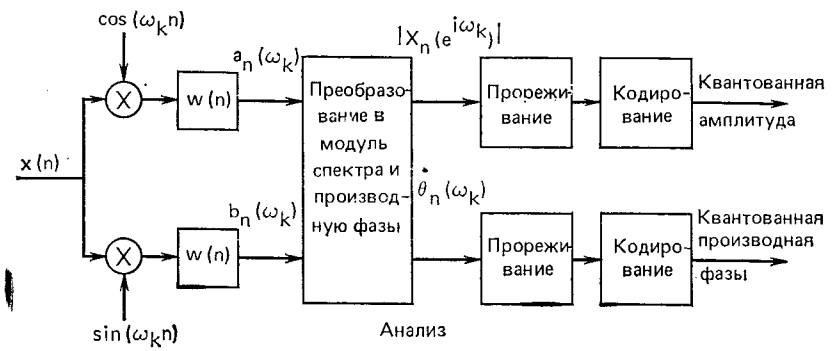


Рис. 6.54. Отдельный канал анализатора фазового вокодера

приведен один канал анализатора с частотой $0 < \omega_k < \pi$. Все остальные каналы строятся точно так же, если не считать возможных различий в прореживании и интерполяции. Операции, требуемые для преобразования $a_n(\omega_k)$ и $b_n(\omega_k)$ в $|X_n(e)|$ и $\theta_n(\omega_k)$, приведены на рис. 6.55. Один из подходов к синтезу по амплитуд-

ному спектру и производной фазы приведен на рис. 6.56. Операции, требуемые для преобразования сигналов амплитудного спектра и производной фазы в действительную и мнимую части, показаны на рис. 6.57. Видно, что сигнал производной фазы нуж-

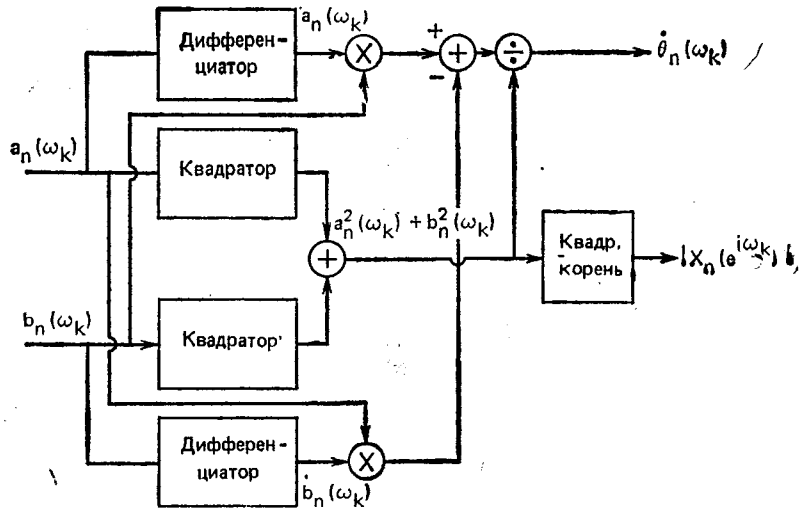


Рис. 6.55. Переход от a и b к $\dot{\theta}$ и $|X(e^{i\omega})|$

но проинтегрировать, чтобы получить сигнал фазы. Косинус и синус фазового угла умножаются затем на значение амплитудного спектра, что и дает действительную и мнимую части. На

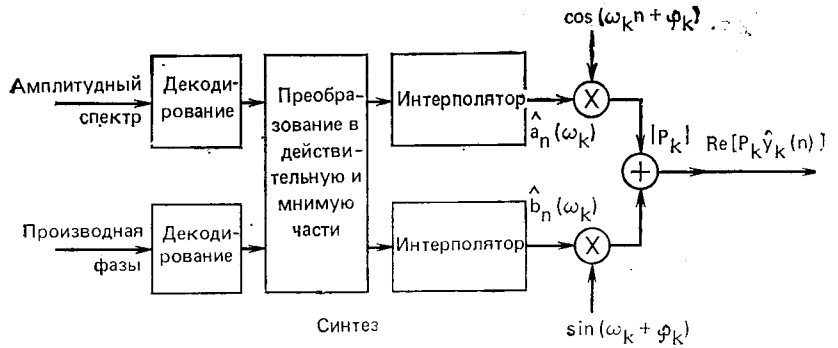


Рис. 6.56. Канал синтезатора фазового вокодера

рис. 6.58 приведен другой подход к синтезу, он позволяет избежать преобразований. В этом случае интерпретируются сигналы амплитудного спектра и производной фазы; результирующие последовательности амплитудного спектра и фазы используются для модуляции синусоиды по амплитуде и фазе. Следовательно, вместо преобразователя амплитудного спектра и производной фазы в

действительную и мнимую части требуется фазовый модулятор. Ясно, что, если реализация такого цифрового фазового модулятора не слишком сложна, схема рис. 6.58 значительно проще схемы рис. 6.56 и 6.57.

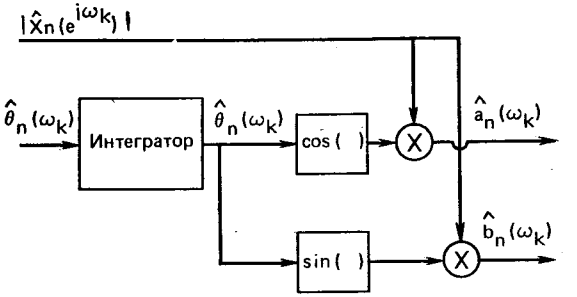


Рис. 6.57. Переход от $|X(e^{i\omega})|$ и θ к a и b

Тщательное исследование методов дискретизации и квантования амплитудного спектра и производной фазы в фазовом вокодере было проведено Карлсом [29]. В этом исследовании был реализован вокодер с 28 каналами и разнесением между ними в

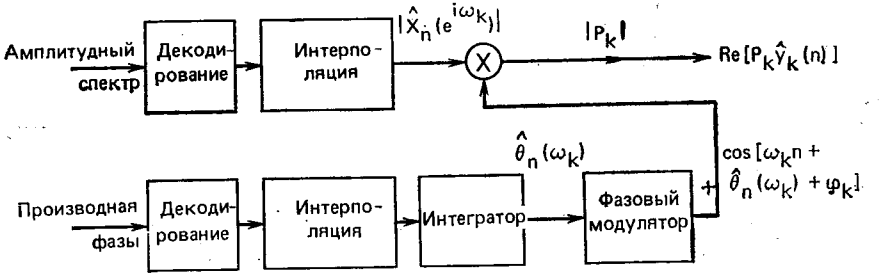


Рис. 6.58. Другая форма синтеза

100 Гц. Для производной фазы использовались линейные квантователи, а для амплитудного спектра — логарифмические. Распределение бит по каналам было неравномерным — больше бит отводилось под низкочастотные каналы и меньше — под высокочастотные. Кроме того, большее число бит отводилось под производную фазы в сравнении с амплитудным спектром. Полученная скорость составляла 7,2 кбит/с при дискретизации сигналов амплитудного спектра и производной фазы с частотой всего 60 раз/с, при этом на сигналы амплитудного спектра отводилось 2 бита в низкочастотных каналах и 1 бит — в высокочастотных каналах, и на сигналы производной фазы: 3 бита — в низкочастотных каналах и 2 бита — в высокочастотных. Неофициальные тесты показали, что такое кодирование речи обеспечивает качество восприятия, сравнимое с качеством восприятия в системе логарифмической ИКМ со скоростью передачи в 2—3 раза выше.

Отметим характерную особенность фазовых вокодеров и вокодеров вообще — большую гибкость в манипуляции параметрами речевого сигнала. По сравнению с представлением в виде колебаний, где изменения в речевом сигнале отображаются некоторой последовательностью чисел, в вокодере сигнал представляется параметрами, теснее связанными с параметрами речеобразования. Например, как уже отмечалось, в случае фазового вокодера можно считать, что амплитуда комплексного канального сигнала несет главным образом сведения о передаточной характеристике голосового тракта, тогда как сигнал производной фазы — сведения о возбуждении. На рис. 6.58 показан простой способ, которым с помощью фазового вокодера можно изменить основные параметры речи. Допустим, что сигнал производной фазы установлен равным нулю, так что выходной сигнал образован произведением модуля кратковременного преобразования Фурье и косинуса фиксированной частоты ω_k. В случае равноразнесенных каналов общий выходной сигнал будет похожим на периодический сигнал с основной частотой, численно равной интервалу частот, на который разнесены каналы. Выходной сигнал не будет строго периодическим, поскольку амплитудный спектр медленно меняется со временем. При таком синтезе выходной сигнал будет иметь монотонное звучание. Если же сигнал производной фазы изменять хаотически, можно ожидать, что синтезированная речь будет похожа на шепот.

Другим более полезным использованием гибкости, присущей системам фазового вокодера, является преобразование временно́го и частотного масштабов сигнала, как это описано в [28]. Обращаясь опять к рис. 6.58, вспомним, что мгновенная частота косинуса равна $[\omega_k+\dot{\theta}_n(\omega_k)]$, следовательно, сигнал с уменьшенной частотой можно получить, просто разделив ω_k и $\dot{\theta}_n(\omega_k)$ на константу q. Если синтезировать каждый канал таким образом, то в результате получится сигнал, сжатый по частоте в q раз. Частотный масштаб результирующего сигнала можно восстановить, записав сигнал на одной скорости и воспроизводя его со скоростью, большей в q раз. Другой способ состоит в использовании преобразователя код—аналог с частотой синхронизации, большей в q раз частоты дискретизации на входе. В любом случае сжатие временно́го масштаба компенсирует сжатие частотного масштаба при синтезе. В результате получается сигнал с обычным частотным масштабом, но со сжатым временны́м. Аналогичными операциями можно растянуть масштаб времени. В этом случае центральная частота ω_k и фазовый угол умножаются на множитель q, а полученный растянутый масштаб частот восстанавливается воспроизведением записанного сигнала с меньшей скоростью. В результате получается растянутый во времени сигнал с обычным частотным масштабом. На рис. 6.59 [28] показано, как описанным способом образуется речь, сжатая и растянутая во времени в $q=2$ раз.

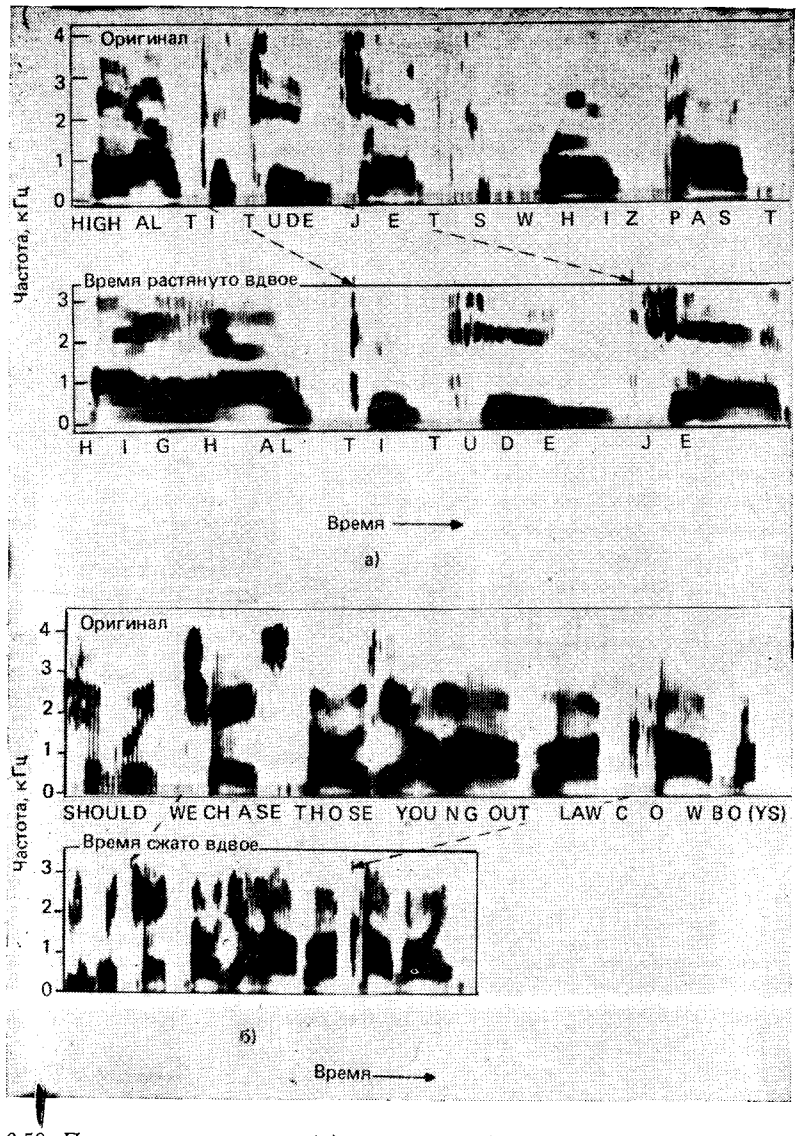


Рис. 6.59. Примеры растяжения (*а*) и сжатия (*б*) временно́го масштаба в фазовом вокодере [28]

6.7.3. Полосный вокодер

Полосный вокодер, изобретенный Дадли [30], представляет собой самое первое устройство для кодирования речи. Он во многих отношениях аналогичен системам, которые уже рассмотрены в этом параграфе. Основные отличия состоят в том, что полосный вокодер теснее связывает схемы анализа и синтеза с моделью

речи, а также в том, что в схемы кратковременного анализа и синтеза Фурье введен для упрощений ряд приближений. Чтобы понять, как связан полосный вокодер с рассмотренными представлениями кратковременного преобразования Фурье, вернемся к (6.167). Вспомним, что это выражение отражает вклад k-го канала в общий выход. Мы интерпретировали это выражение как представление для косинуса с номинальной центральной частотой ω_k, модулированного по фазе и амплитуде, причем по амплитуде — амплитудой (модулем) кратковременного преобразования Фурье, а по фазе — сигналом, соответствующим фазовому углу (аргументу) кратковременного преобразования Фурье. Каждый канал анализа можно представлять себе как полосовой фильтр с центральной частотой ω_k. Отсюда можно предположить, что амплитуду кратковременного преобразования Фурье можно получить детектором огибающей на выходе полосового фильтра с центральной частотой ω_k. Это показано на рис. 6.60, где за полосовым фильтром с импульсной характеристикой $w(n)\cos(\omega_k n)$ следует двухполупериодный выпрямитель (блок амплитуды) и снова фильтр нижних частот.

Рис. 6.60. Метод аппроксимации кратковременного спектра

Двухполупериодный выпрямитель и фильтр нижних частот служат детектором огибающей. Такая система представляет собой базовый блок полосного вокодера. Анализатор состоит из набора таких каналов, причем частоты анализа распределены по частотному диапазону речи. Однако сигнал речи не может быть представлен только амплитудным спектром — сигнал производной фазы содержит информацию о возбуждении. Если приравнять сигнал производной фазы нулю, то результирующая речь окажется полностью вокализованной и монотонной. Для того чтобы отразить должным образом возбуждение, полосной вокодер имеет дополнительное анализирующее устройство определяющее тип возбуждения — вокализованное или невокализованное, и, если вокализованное, основную частоту речевого сигнала. Эти параметры дискретизуются и квантуются для передачи или хранения в цифровой системе. Полностью анализатор полосного вокодера приведен на рис. 6.61. В системе нужно предусмотреть значительные изменения для синтеза сигнала полосного вокодера (рас. 6.62).

Основной принцип синтеза в полосном вокодере можно сформулировать просто. Сигналы каналов отражают амплитуду вклада каждого из каналов, тогда как сигнал возбуждения управляет структурой в заданном канале. Сигнал тон/шум задает нужный тип возбуждения — случайный шум для невокализованной речи

или периодические импульсы в случае вокализованной речи. Основная частота импульсного генератора управляется сигналом основного тона. Следовательно, общий выходной спектр строится по отдельным сегментам, в которых амплитуда в заданной полосе

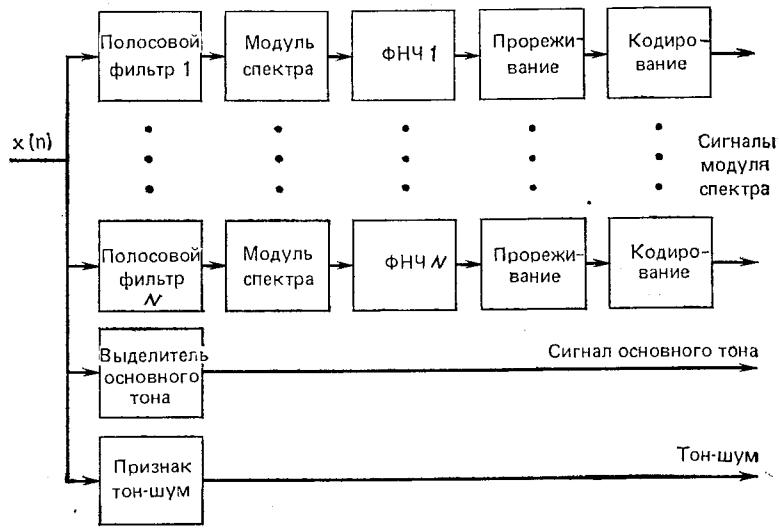


Рис. 6.61. Структурная схема анализатора полосного вокодера

частот, грубо говоря, постоянна. Действительно, амплитуда в каждой полосе частот сохраняет форму, обусловленную частотно-избирательными свойствами используемых при синтезе полосовых

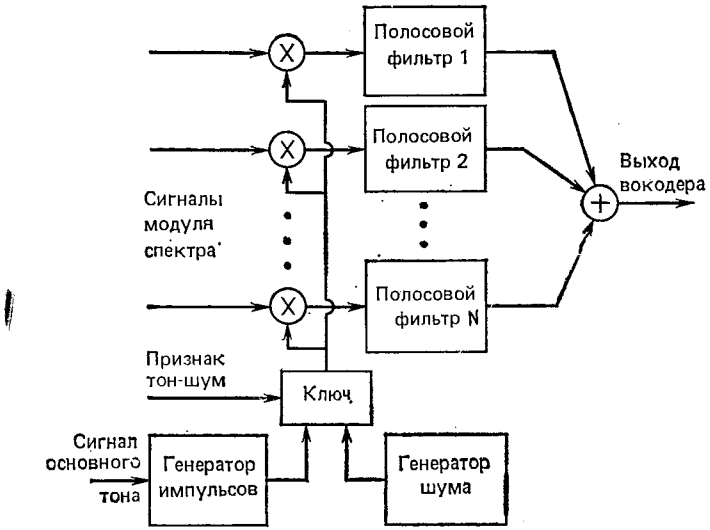


Рис. 6.62. Структурная схема синтезатора полосного вокодера

фильтров. Когда возбуждение вокализованно, выходной сигнал образуется из смежных полос частот, причем тонкая структура спектра характеризуется периодичностью.

Для невокализованного возбуждения спектр непрерывно меняется в каждой из полос частот. Результатом будет речь с сильной реверберацией, что вызвано полным отсутствием контроля над интерференцией смежных частотных полос. На рис. 6.63 [31]

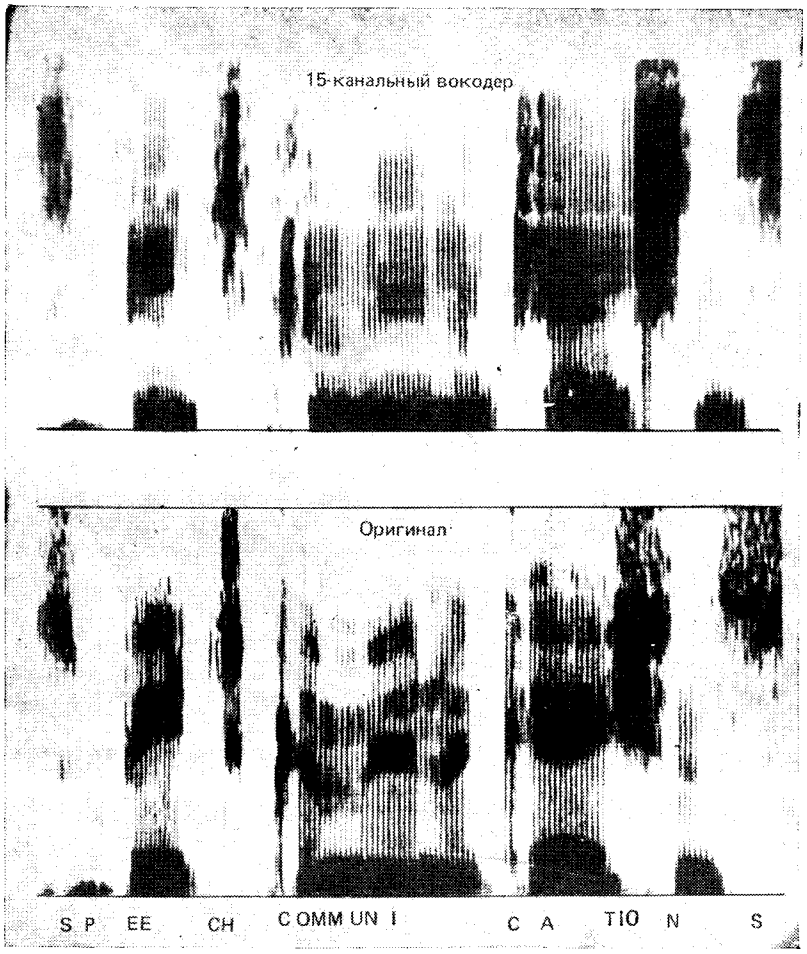


Рис. 6.63. Пример сигнала в 15-канальном вокодере [31]

сравниваются спектрограммы речевого сигнала на входе 15-канального полосного вокодера со спектрограммой соответствующего выхода. Видно, что из-за грубого разнесения каналов формантная структура сильно квантованна, причем частоты формант в

некоторых случаях весьма значительно изменены. Устройства, изображенные на рис. 6.61 и 6.62, обеспечивают значительное снижение скорости передачи информации с сопутствующим ростом искажений.

Полосные вокодеры обычно работают со скоростями 1200—9600 бит/с, причем около 600 бит/с отводится под информацию об основном тоне и типе возбуждения, а оставшиеся — под канальные сигналы. Полосный вокодер еще больше, чем фазовый, допускает модификацию речевого сигнала, поскольку информация о возбуждении и голосовом тракте представляется порознь. Легко понять, например, как можно изменить информацию об основном тоне независимо от информации о голосовом тракте. Если, например, импульсный генератор всегда выдает одну и ту же основную частоту, т. е. информация об основном тоне не используется вовсе, результатом будет монотонная речь. Если возбуждение импульсным генератором не используется, а вместо этого возбуждение всегда представляет собой случайный шум, результатом будет шепот. Используя полосный вокодер, можно также получить независимые изменения временно́й и частотной шкал. Это достигается простым масштабированием центральных частот полосовых фильтров и периода основного тона.

Выделение информации об основном тоне и типе возбуждения обеспечивает основной вклад в снижение скорости, достигаемое в полосном вокодере. Это, однако, является и слабым местом таких систем, поскольку выделение основного тона представляет собой трудную задачу. Следовательно, фазовый вокодер или, точнее, представления, вытекающие из теории, изложенной в предыдущих параграфах этой главы, обладают тем преимуществом, что не требуют отслеживания основного тона.

Полосный вокодер был предметом интенсивных исследований как вследствие традиции, так и вследствие большого числа факторов, влияющих на качество его работы (например, число и тип фильтров, их разнос и др.). Результатом этих исследований оказалось большое число остроумных решений реализации полосных вокодеров [31—34].

6.8. Заключение

В этой главе проведен анализ кратковременного преобразования Фурье применительно к речевым сигналам. Показано, как такое представление может быть эффективно использовано для оценки основных параметров речи, таких, как период основного тона и частота формант. Рассмотрено также применение кратковременного преобразования Фурье при проектировании фазовых и полосных вокодеров.

Задачи

6.1. Пусть кратковременное преобразование Фурье задано в виде $X_n(e^{i\omega}) = a_n(\omega) - ib_n(\omega) = |X_n(e^{i\omega})|e^{i\theta_n(\omega)}$. Доказать, что если $x(n)$ действительна, то:

а) $a_n(\omega) = a_n(2\pi - \omega) = a_n(-\omega)$;

б) $b_n(\omega) = -b_n(2\pi - \omega) = -b_n(-\omega)$;

в) $|X_n(e^{i\omega})| = |X_n(e^{i(2\pi-\omega)})| = |X(e^{-i\omega})|$;
г) $\theta_n(\omega) = -\theta_n(2\pi-\omega) = -\theta_n(-\omega)$.

6.2. Пусть кратковременное преобразование Фурье сигнала задано соотношением $X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m}$.

Показать, что справедливы следующие свойства:
а) линейность[1] — если $v(n) = x(n) + y(n)$, то $V_n(e^{i\omega}) = X_n(e^{i\omega}) + Y_n(e^{i\omega})$;
б) свойство сдвига — если $v(n) = x(n-n_0)$, то $V_n(e^{i\omega}) = X_{n-n_0}(e^{i\omega})\,e^{-i\omega n_0}$;
в) масштабирование — если $v(n) = \alpha x(n)$, то $V_n(e^{i\omega}) = \alpha X_n(e^{i\omega})$;
г) экспоненциальное взвешивание — если $v(n) = a^n x(n)$, то $V_n(e^{i\omega}) = X_n(a^{-1}e^{i\omega})$;
д) сопряженная симметрия — если $x(n)$ действительна, то $X_n(e^{i\omega}) = X^*_n(e^{-i\omega})$.

6.3. По определению $X_n(e^{i\omega}) = a_n(\omega) - ib_n(\omega) = |X_n(e^{i\omega})|\,e^{i\theta_n(\omega)}$:
а) Выразить $|X_n(e^{i\omega})|$ и $\theta_n(\omega)$ через $a_n(\omega)$ и $b_n(\omega)$.
б) Выразить $a_n(\omega)$ и $b_n(\omega)$ через $|X_n(e^{i\omega})|$ и $\theta_n(\omega)$.

6.4. Пусть $x(n)$ и $w(n)$ имеют в качестве обычного преобразования Фурье $X(e^{i\omega})$ и $W(e^{i\omega})$ соответственно. Показать, что кратковременное преобразование Фурье $X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m}$ можно представить в виде

$$X_n(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{i\theta})\,e^{i\theta n}\,X(e^{i(\omega+\theta)})\,d\theta,$$

т. е. $X_n(e^{i\omega})$ представляет собой сглаженную оценку спектра $X(e^{i\omega})$ на частоте ω.

6.5. Определим кратковременную спектральную плотность мощности сигнала посредством кратковременного преобразования Фурье:

$$S_n(e^{i\omega}) = |X_n(e^{i\omega})|^2$$

и зададим кратковременную автокорреляционную функцию сигнала

$$R_n(k) = \sum_{m=-\infty}^{\infty} w(n-m)\,x(m)\,w(n-k-m)\,x(m+k).$$

Показать, что если

$$X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m},$$

то $R_n(k)$ и $S_n(e^{i\omega})$ связаны преобразованием Фурье, т. е. $S_n(e^{i\omega})$ есть преобразование Фурье от $R_n(k)$ и наоборот.

6.6. Допустим, что используемая для кратковременного анализа Фурье последовательность окна $w(n)$ физически реализуема и имеет рациональное z-преобразование вида

$$W(z) = \sum_{r=0}^{Nz} b_r z^{-r} / \left(1 - \sum_{k=1}^{Np} a_k z^{-k}\right).$$

[1] Начинающему полезно иметь в виду, что в математических работах (а также в работах, претендующих на «совместимость» терминологии) свойство а) принято называть аддитивностью, а свойство в) — однородностью. При этом линейность определяют как аддитивность и однородность. Постарайтесь понять, почему все же свойство а) названо здесь линейностью (совет: покажите, что из а) следует однородность для рациональных α). (Прим. перев.)

а) Какими свойствами должна обладать последовательность $W(z)$ или, что то же, $W(e^{i\omega})$, чтобы она годилась для указанного применения?
б) Получить рекуррентную формулу для $X_n(e^{i\omega})$ через сигнал $x(n)$ и предыдущие значения $X_n(e^{i\omega})$.
в) Рассмотрим случай $W(z) = 1/(1-az^{-1})$. Как следует выбрать a, чтобы получить разрешение по частоте около 100 Гц при частоте дискретизации 10 кГц?
г) Использование требуемого в в) значения a может привести к трудностям, если зависящий от времени анализ Фурье узкополосен и реализуется рекурсивно. Рассмотреть природу этих трудностей.

6.7. Доказать, что

$$\sum_{k=0}^{N-1} e^{i\frac{2\pi}{N}kn} = \begin{cases} N\sum_{r=-\infty}^{\infty}\delta(n-rN); \\ N, \quad n = rN, \quad r = 0, \pm 1, \ldots; \\ 0, \quad \textit{в противном случае}. \end{cases}$$

При доказательстве воспользуйтесь тождеством $\sum_{k=0}^{N-1}\alpha^k = \frac{1-\alpha^N}{1-\alpha}$.

6.8. Реализуя зависящее от времени Фурье-представление, можно применить дискретизацию как по времени, так и по частоте. В этой задаче исследуем эффекты обоих типов дискретизации. Рассмотрим последовательность $x(n)$ с обычным преобразованием Фурье

$$X(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,e^{-i\omega m}.$$

а) Для периодической функции $X(e^{i\omega})$, дискретизируемой на частоте $\omega_k = 2\pi k/N$, $k = 0, 1, \ldots, N-1$, имеем

$$\tilde{X}(k) = \sum_{m=-\infty}^{\infty} x(m)\,e^{-i\frac{2\pi}{N}km}.$$

Такие отсчеты можно представлять себе как дискретное преобразование Фурье последовательности $\tilde{x}(n)$, задаваемой соотношением

$$\tilde{x}(n) = \frac{1}{N}\sum_{k=0}^{N-1}\tilde{X}(k)\,e^{i\frac{2\pi}{N}kn}.$$

Показать, что $\tilde{x}(n) = \sum_{r=-\infty}^{\infty} x(n+rN)$.

б) При каких условиях на $x(n)$ при дискретизации $X(e^{i\omega})$ не возникают искажения из-за «наложений» во временно́й области.
в) Рассмотрим теперь «дискретизацию» последовательности $x(n)$, т. е. сформируем новую последовательность $y(n) = x(nM)$, состоящую из M-х отсчетов $x(n)$. Показать, что

$$Y(e^{i\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{i(\omega-2\pi k)/M})$$

есть преобразование Фурье от $y(n)$. При доказательстве вы, возможно, захотите начать с рассмотрения последовательности $v(n) = x(n)p(n)$, где $p(n) = \sum_{r=-\infty}^{\infty}\delta(n+rM)$. После этого заметьте, что $y(n) = v(nM) = x(nM)$.
г) Какие ограничения надо наложить на $X(e^{i\omega})$, чтобы при дискретизации $x(n)$ не возникало наложений в частотной области?

6.9. Рассмотрим окно $w(n)$ с преобразованием Фурье $W(e^{i\Omega T})$, ограниченным по частоте интервалом $0 \leqslant \Omega \leqslant \Omega_c$. Мы хотим показать, что

$$\sum_{r=-\infty}^{\infty} w(rR-n) = W(e^{i0})/R$$

не зависимо от n для достаточно малого (ненулевого) целого R.

а) Пусть $w(r)=w(rR-n)$. Получить выражение для $W(e^{i\Omega T'})$ через R и $W(e^{i\Omega T})$, где T — период дискретизации $w(n)$, а $T'=RT$ — период дискретизации $\hat{w}(r)$ (указание: вспомните задачу прореживания сигнала в отношении $R:1$ или задачу 6.8в).

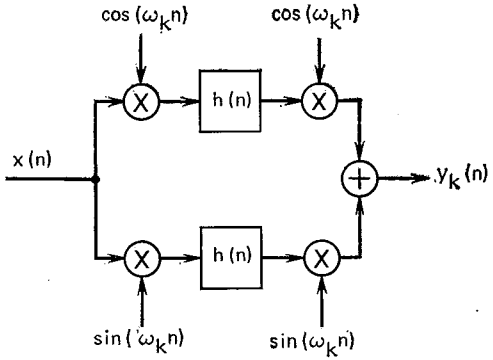


Рис. 3.6.1

б) Допустим, что $W(e^{i\Omega T})=0$ при $|\Omega|>\Omega_c$, получить выражение для максимального значения R (как функции от Ω_c) такого, что $\hat{W}(e^{i0})=W(e^{i0})/R$.

в) Вспомнив соотношение $\sum_{r=-\infty}^{\infty} \hat{w}(r)e^{-i\Omega T'r}=\hat{W}(e^{i\Omega T'})$, покажите, что если выполнены условия пункта б), то справедливо и соотношение, приведенное в формулировке задачи.

6.10. а) Показать, что импульсная характеристика системы, изображенной на рис. 3.6.1, имеет вид $h_k(n)=h(n)\cos(\omega_k n)$.

б) Найти частотную характеристику системы, изображенной на рис. 3.6.1.

6.11. Подчеркивания высокочастотной части спектра часто добиваются переходом к вычислению первой разности. В этой задаче исследуем влияние эффекта от такой операции на кратковременное преобразование Фурье.

а) Пусть $y(n)=x(n)-x(n-1)$. Показать, что $Y_n(e^{i\omega})=X_n(e^{i\omega})-e^{-i\omega}\times X_{n-1}(e^{i\omega})$.

б) При каких условиях можно считать, что $Y_n(e^{i\omega})\approx(1-e^{-i\omega})X_n(e^{i\omega})$? Вообще говоря, можно считать $x(n)$ линейно отфильтрованной: $y(n)=\sum_{k=0}^{N-1} h(k)x(n-k)$.

в) Показать, что $Y_n(e^{i\omega})$ связано с $X_n(e^{i\omega})$ соотношением $Y_n(e^{i\omega})=X_n(e^{i\omega})*h_\omega(n)$. Выразить $h_\omega(n)$ через $h(n)$.

г) Оправдано ли считать, что $Y_n(e^{i\omega})=H(e^{i\omega})X_n(e^{i\omega})$.

6.12. Гребенка из N фильтров характеризуется следующими свойствами: полосы расположены симметрично вокруг $\omega=\pi$, т. е. $\omega_k=2\pi-\omega_{N-k}$, $P_k=P^*_{N-k}$, $w_k(n)=w_{N-k}(n)$; существует канал, для которого $\omega_k=0$. Для четных и нечетных N:

а) прикинуть, как расположены N полос фильтров;

б) выразить общую импульсную характеристику гребенки фильтров через $w_k(n)$, ω_k, P_k, N.

6.13. Чтобы проиллюстрировать эффект реверберации, возникающий в гребенках БИХ-фильтров, рассмотрим общую импульсную характеристику $h(n)=\alpha_1\delta(n)+\alpha_2\delta(n-N)+\alpha_3\delta(n-2N)$, где представлены эхо, разнесенные на N отсчетов.

а) Определить функцию $H(e^{i\omega})$ системы и показать, что ее квадрат можно записать в следующем виде: $|H(e^{i\omega})|^2=(\alpha_2+(\alpha_1+\alpha_3)\cos(\omega N))^2+(\alpha_1-\alpha_3)^2\times\sin^2(\omega N)$.

б) Показать, что фазовую характеристику можно записать в виде

$$\theta(\omega) = -\omega N + \operatorname{tg}^{-1}\left[\frac{(\alpha_1-\alpha_3)\sin(\omega N)}{\alpha_2+(\alpha_1+\alpha_3)\cos(\omega N)}\right].$$

в) Определить, где расположены максимумы и минимумы амплитуды, для чего продифференцировать $|H(e^{i\omega})|^2$ по ω и положить результат равным нулю. Показать, что в случае $|\alpha_1+\alpha_3|\ll|\alpha_2|$ максимумы и минимумы расположены в точках $\omega=\pm k\pi/N$, $k=0, 1, 2, \ldots$

г) Воспользовавшись результатами п. в), показать, что максимальную амплитудную пульсацию (в децибелах) можно записать как

$$R_A = 20\log_{10}\left[\frac{|\alpha_2+\alpha_1+\alpha_3|}{|\alpha_2-\alpha_1-\alpha_3|}\right].$$

д) Найти R_A для

$$\alpha_1=0{,}1;\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}2;$$
$$\alpha_1=0{,}15,\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}15;$$
$$\alpha_1=0{,}1,\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}1$$

е) Дифференцируя $\theta(\omega)$ по ω, можно показать, что максимумы и минимумы θ расположены в тех значениях ω, для которых $\cos(\omega N)=-[(\alpha_1+\alpha_3)/\alpha_2]$. Показать, что максимальная пульсация фазы задается равенством

$$R_p = 2\operatorname{tg}^{-1}\left[\frac{\alpha_1-\alpha_3}{(\alpha_2^2-(\alpha_1+\alpha_3)^2)^{1/2}}\right].$$

ж) Найти R_p для случаев, перечисленных в п. д.). Определить, как влияют изменения α_1 и α_3 на R_A и R_p.

6.14. Предлагается цифровое устройство выделения основного тона, состоящее из гребенки цифровых полосовых фильтров с нижними частотами среза $F_k=2^{k-1}F_1$, $k=1, 2, \ldots, M$, и верхними частотами среза $F_{k+1}=2^kF_1$, $k=1, 2, \ldots, \ldots, M$. При таком выборе частот среза гребенка фильтров обладает следующим свойством: если вход периодичен с основной частотой F_0 такой, что $F_k<F_0<F_{k+1}$, то энергия на выходе фильтров в полосах от 1 до $k-1$ будет мала, выходной сигнал в полосе k будет содержать основную частоту, а в полосы от $k+1$ до M попадут гармоники. Поэтому, если поместить за каждым фильтром выделитель чистого тона, получится хороший индикатор наличия основного тона.

а) Определить такие F_1 и M, чтобы указанный метод использовался для частот основного тона 50—800 Гц.

б) Сделать набросок необходимой частотной характеристики для каждого из M полосовых фильтров.

в) Что Вы можете предложить для реализации обнаружителя тона, необходимого на выходе каждого из фильтров?

г) Какие трудности можно предвидеть при реализации описываемого метода неидеальными полосовыми фильтрами?

д) Что произойдет, если на вход поступает речь, ограниченная полосой 300—3000 Гц, т. е. входной сигнал телефонной линии? Можно ли предложить какие-либо усовершенствования в этом случае?

6.15. Рассмотрим периодическую последовательность

$$\tilde{x}(n) = \sum_{r=-\infty}^{\infty} h_v(n+rN_p),$$

представляющую сегмент вокализованной речи.

а) Показать, что $\tilde{x}(n)$ можно разложить в ряд Фурье:

$$\tilde{x}(n) = \frac{1}{N_p}\sum_{k=0}^{N_p-1} \tilde{X}(k)\, e^{i\frac{2\pi}{N_p}kn},$$

где коэффициенты Фурье $\tilde{X}(k)$ представляют собой отсчеты преобразования Фурье вокализованной речи, т. е.

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right) \quad \text{(см. задачу 6.8).}$$

б) Показать, что кратковременное преобразование Фурье от $\tilde{x}(n)$ можно представить в виде

$$\tilde{X}(e^{i\omega}) = \frac{1}{N_p}\sum_{k=0}^{N_p-1} H_v\left(e^{i\frac{2\pi}{N_p}k}\right) W_n(e^{i(\omega-2\pi k/N_p)}),$$

где $W_n(e^{i\omega})$ — преобразование Фурье от $w(n-m)$.

в) Сколько различных значений принимает $X_n(e^{i\omega})$ при фиксированном ω? Для прямоугольного окна

$$w(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N_p - 1; \\ 0, & \textit{в противном случае} \end{cases}$$

найти функцию $W_n(e^{i\omega})$.

г) Для каких значений N_p справедливо равенство

$$\tilde{X}_n\left(e^{i\frac{2\pi}{N_p}k}\right) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right)$$

для прямоугольного окна шириной n?

6.16. Займемся анализом и синтезом сигнала $x(n)=\cos(\omega_0 n)$. Схема для анализа приведена для k-го канала на рис. 3.6.2*а*.

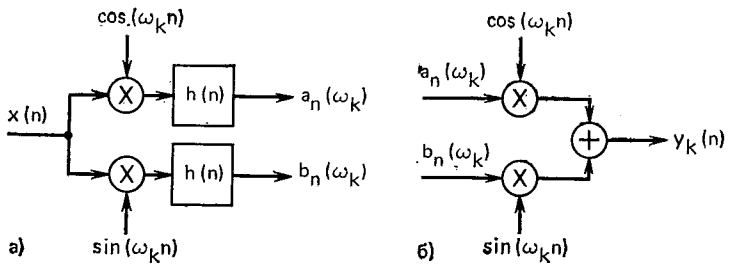


Рис. 3.6.2

а) Определить $a_n(\omega_k)$ и $b_n(\omega_k)$ для заданного входного сигнала.

б) Предположив, что $h(n)$ соответствует узкополосному фильтру нижних частот, упростить полученные выражения для $a_n(\omega_k)$ и $b_n(\omega_k)$ в предположении, что $(\omega_0-\omega_k)$ спадает в полосе фильтра и что $H(e^{i\omega})\approx 1$ для этих частот.

в) Сигналы $a_n(\omega_k)$ и $b_n(\omega_k)$ дают в комбинации мгновенное значение $M_n(\omega_k)$ и производную фазы $\dot{\varphi}_n(\omega_k)$. Определить $M_n(\omega_k)$ и $\dot{\varphi}_n(\omega_k)$ для нашего примера.

г) Показать, что для схемы синтеза, показанной на рис. 3.6.2*б*, выходной сигнал по существу совпадает с входным.

д) Производная фазы $\varphi_n=(\omega_k)$ вычисляется по формуле

$$\dot{\varphi}_n(\omega_k) = \frac{b_n(\omega_k)\dot{a}_n(\omega_k) - a_n(\omega_k)\dot{b}_n(\omega_k)}{[a_n(\omega_k)]^2 + [b_n(\omega_k)]^2}.$$

Найти $\dot{\varphi}_n(\omega_k)$ для нашего примера и сравнить результат с результатом в).

ж) Предположить, что производная фазы из п.д.) вычисляется с помощью первой разности, т. е. $\varphi_n(\omega_k)\approx\frac{1}{T}(a_n(\omega_k)-a_{n-1}(\omega_k))$, где T — период дискретизации во временной области. Найти $\varphi_n(\omega_k)$ по этой формуле и сравнить результат с полученным в п.в). При каких условиях они близки?

7
Гомоморфная обработка речи

7.0. Введение

Одно из основных предположений, сделанных в этой книге, состоит в том, что речевой сигнал трактуется как сигнал на выходе линейной системы с медленно изменяющимися параметрами. Это предположение позволяет считать, что на коротких сегментах речевой сигнал можно рассматривать как сигнал на выходе линейной системы с постоянными параметрами, возбуждаемой либо последовательностью импульсов, либо случайным шумом. Как уже отмечалось, проблема анализа речевого сигнала сводится к измерению параметров модели и оценке изменения этих параметров с течением времени. Поскольку сигнал возбуждения и импульсная характеристика фильтра взаимодействуют через операцию свертки, задача анализа речи может рассматриваться как задача разделения компонент, участвующих в операции свертки. Такая задача иногда называется задачей обратной свертки[1]. В гл. 6 был рассмотрен метод ее решения на основе представления речевого сигнала в виде переменного во времени преобразования Фурье. В данной главе на основе использования теории, изложенной в гл. 6, развивается другой подход к задаче, названный гомоморфной фильтрацией. После краткого введения в общую теорию гомоморфных систем будут рассмотрены различные способы применения методов гомоморфной обратной свертки в области анализа речевых сигналов.

7.1. Гомоморфные относительно свертки системы

Гомоморфные относительно свертки системы удовлетворяют обобщенному принципу суперпозиции. Принцип суперпозиции, если его записать для обычных линейных систем, имеет вид

$$\begin{aligned} L[x(n)] &= L[x_1(n)+x_2(n)] = \\ &= L[x_1(n)]+L[x_2(n)] = \\ &= y_1(n)+y_2(n) = y(n) \end{aligned} \tag{7.1а}$$

и

$$L[ax(n)] = aL[x(n)] = ay(n), \tag{7.1б}$$

где L — линейный оператор. Принцип суперпозиции устанавливает, что если сигнал на входе является линейной комбинацией элементарных сигналов, то и сигнал на выходе будет представлен в виде линейной комбинации соответствующих сигналов. Этот принцип иллюстрируется на рис. 7.1, где символ «+» на входе и выходе

[1] Операцию, обратную свертке (deconvolution), в переводной литературе также называют «разверткой». (*Прим. ред.*)

означает, что аддитивная комбинация сигналов на входе приводит к аддитивной комбинации выходных сигналов.

Как показано в гл. 2, прямым следствием принципа суперпозиции является тот факт, что сигнал на выходе линейной системы может быть представлен в виде дискретной свертки:

$$y(n)=\sum_{k=-\infty}^{\infty} h(n-k)\,(x(k)=h(n)*x(n). \tag{7.2}$$

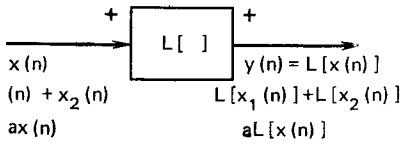


Рис. 7.1. Представление системы, в которой выполняется принцип суперпозиции

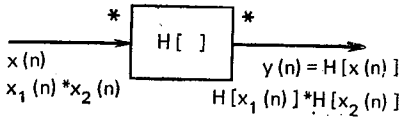


Рис. 7.2. Представление системы, гомоморфной относительно свертки

Символ « * » здесь и далее означает свертку в дискретном времени. По аналогии с принципом суперпозиции для обычных линейных систем определим класс систем, удовлетворяющих обобщенному принципу суперпозиции, в котором сложение заменяется сверткой (легко показать, что свертка обладает такими же алгебраическими свойствами, как и сложение [1]), т. е.

$$\begin{aligned} H[x(n)] &= H[x_1(n)*x_2(n)]= \\ &= H[x_1(n)]*H[x_2(n)]= \\ &= y_1(n)*y_2(n)=y(n). \end{aligned} \tag{7.3}$$

В общем случае возможно сформулировать и уравнение, аналогичное (7.1б), в котором выражено свойство скалярного умножения [2], однако обобщенное скалярное умножение далее не используется. Системы, обладающие свойством (7.3), названы гомоморфными относительно свертки системами. Эта терминология объясняется тем [3], что данные преобразования оказываются гомоморфными преобразованиями линейного векторного пространства. При изображении таких систем (рис. 7.2) операцию свертки представляют в явном виде на входе и выходе системы. Гомоморфный фильтр является гомоморфной системой, обладающей тем свойством, что одна компонента (выделяемая) проходит через эту систему без изменений, а другая — устраняется. В соотношении (7.3), например, если $x_1(n)$ — нежелательная компонента, то необходимо потребовать, чтобы выход, соответствующий $x_1(n)$, представлял собой единичный отсчет, в то время как выход, соответствующий $x_2(n)$, близко совпадал бы с $x_2(n)$. Это полностью аналогично ситуации в линейных системах, где ставится задача выделения сигнала из смеси его с аддитивным шумом.

Важным аспектом теории гомоморфных систем является то, что любая из них может быть представлена в виде каскадного

соединения трех гомоморфных систем, как это изображено на рис. 7.3 для случая систем, гомоморфных относительно свертки [3]. Первый блок преобразует компоненты на входе, представленные в виде свертки, в аддитивную сумму на выходе. Второй блок — обычная линейная система, удовлетворяющая принципам

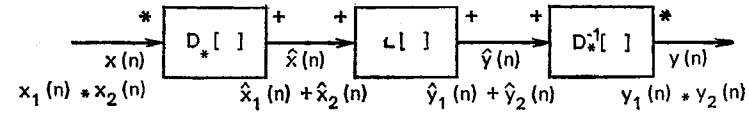


Рис. 7.3. Каноническая форма системы, гомоморфной относительно свертки

суперпозиции в соответствии с (7.1). Третий блок является обратным первому, т. е. преобразует сигналы, представленные в виде суммы, в сигналы, представленные в виде свертки. Важность такого канонического представления заключается в том, что разработка гомоморфной системы сводится к разработке линейной системы. Блок$_*$[ ], называемый характеристическим блоком гомоморфной относительно свертки системы, фиксирован при каноническом представлении, приведенном на рис. 7.3. Очевидно, что обратное преобразование также фиксировано. Характеристическая система для гомоморфной обратной свертки подчиняется обобщенному принципу суперпозиции, в котором операция на входе — свертка, а на выходе — обычное сложение. Свойства характеристической системы определяются выражением

$$\begin{aligned} D_*[x(n)] &= D_*[x_1(n)*x_2(n)]= \\ &= D_*[x_1(n)]+D_*[x_2(n)]= \\ &= \hat{x}_1(n)+\hat{x}_2(n)=\hat{x}(n). \end{aligned} \tag{7.4}$$

Аналогично обратная характеристическая система удовлетворяет соотношению

$$\begin{aligned} D_*^{-1}[\hat{y}(n)] &= D_*^{-1}[\hat{y}_1(n)+\hat{y}_2(n)], = \\ &= D_*^{-1}[\hat{y}_1(n)]*D_*^{-1}[\hat{y}_2(n)]= \\ &= y_1(n)*y_2(n)=y(n). \end{aligned} \tag{7.5}$$

Математическое описание характеристической системы определяется требованиями к выходному сигналу. Если на входе имеется сигнал свертки, то

$$x(n)=x_1(n)*x_2(n) \tag{7.6}$$

и z-преобразование входного сигнала имеет вид

$$X(z)=X_1(z)\,X_2(z). \tag{7.7}$$

Из (7.4) очевидно, что z-преобразование сигнала на выходе системы должно представлять собой сумму z-преобразований компонент. Таким образом, в частотной области характеристическая

система для свертки должна обладать следующим свойством: если на входе имеется произведение компонент, то на выходе должна возникнуть их сумма. Один из подходов к синтезу такой системы представлен на рис. 7.4. Этот подход основан на том,

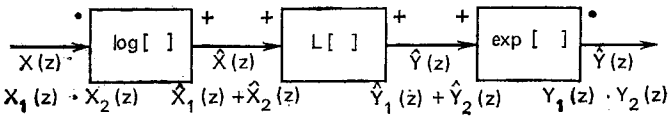


Рис. 7.4. Представление системы, гомоморфной относительно свертки в частотной области

что логарифм произведения равен сумме логарифмов сомножителей, т. е.

$$\hat{X}(z) = \log[X(z)] = \log[X_1(z)\,X_2(z)] = \\ = \log[X_1(z)] + \log[X_2(z)]. \tag{7.8}$$

Если необходимо представлять сигналы во временно́й, а не в частотной области, то характеристическая система примет вид, представленный на рис. 7.5. Аналогичное обратное преобразование показано на рис. 7.6.

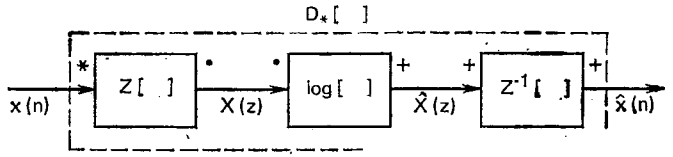


Рис. 7.5. Представление характеристической системы, гомоморфной относительно свертки

Представление прямой и обратной характеристических систем зависит от справедливости соотношения (7.8). Таким образом, логарифм должен быть определен так, чтобы логарифм произведения равнялся сумме логарифмов сомножителей. Это тривиально для действительных положительных величин. Однако в общем случае z-преобразование имеет комплексный характер и вопрос единственности логарифма комплексной случайной величины чрезвычайно важен. С точки зрения вычислений целесообразно рас-

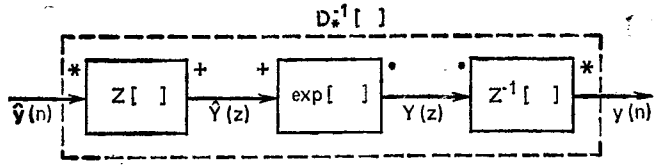


Рис. 7.6. Представление характеристической системы обратной гомоморфной системе

смотреть случай, когда (7.8) справедливо на единичной окружности, т. е. для $z=e^{i\omega}$. Детальное обсуждение проблемы единственности дано в [2]. Для решаемых здесь задач вполне подходит определение логарифма в виде

$$\hat{X}(e^{i\omega}) = \log|X(e^{i\omega})| + i\arg[X(e^{i\omega})]. \tag{7.9}$$

В этом соотношении действительная часть $\log|X(e^{i\omega})|$ не вызывает трудностей. Проблема единственности возникает при определении мнимой части (т. е. $\arg[X(e^{i\omega})]$), которая представляет собой фазовый угол z-преобразования, вычисленного на единичной окружности. В [2] показано, что одним из подходов к решению проблемы единственности является предположение, что фазовый угол представляет собой непрерывную нечетную функцию. В этих условиях уравнение (7.8) справедливо.

С учетом возможности вычисления комплексного логарифма, удовлетворяющего (7.8), обратное преобразование комплексного логарифма преобразования Фурье входного сигнала, являющееся выходом характеристической системы для свертки, имеет вид

$$\hat{x}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{i\omega})\,e^{i\omega n}\,d\omega. \tag{7.10}$$

Выход характеристической системы назван «комплексным кепстром» (термин «кепстр» введен Богертом и др. [4] и является в настоящее время общепринятым для обозначения обратного преобразования Фурье логарифма спектра мощности сигнала; термин «комплексный кепстр» означает, что применяется комплексный логарифм). Термин «кепстр» далее используется для величины

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{i\omega})|\,e^{i\omega n}\,d\omega. \tag{7.11}$$

[Можно показать, что последовательность $c(n)$ представляет собой четную часть комплексного кепстра $\hat{x}(n)$ (см. задачу 7.1).]

Выше была определена характеристическая система для гомоморфной свертки и, таким образом, определена каноническая форма всех гомоморфных систем относительно свертки. Все системы этого класса отличаются только линейной частью. Выбор линейной системы определяется свойствами входного сигнала. Следовательно, для правильного построения линейной системы необходимо прежде всего определить вид и структуру сигнала на выходе характеристической системы, т. е. рассмотреть свойства комплексного кепстра для типичных входных сигналов.

7.1.1. Свойства комплексного кепстра

Для определения свойств комплексного кепстра достаточно рассмотреть случай рационального z-преобразования. Наиболее общая форма преобразования имеет вид

$$X(z)=\frac{Az^r\prod_{k=1}^{M_i}(1-a_k z^{-1})\prod_{k=1}^{M_0}(1-b_k z)}{\prod_{k=1}^{N_i}(1-c_k z^{-1})\prod_{k=1}^{N_0}(1-d_k z)}, \qquad (7.12)$$

где модули величин a_k, b_k, c_k и d_k меньше единицы. Таким образом, сомножители $(1-a_k z^{-1})$ и $(1-c_k z^{-1})$ соответствуют нулям и полюсам внутри единичной окружности, а $(1-b_k z)$ и $(1-d_k z)$ — нулям и полюсам вне единичной окружности. Параметр z^r означает соответствующую задержку во временно́й области. В соответствии с предложением уравнения (7.8) комплексный логарифм $X(z)$ имеет вид

$$\hat{X}(z)=\log[A]+\log[z^r]+\sum_{k=1}^{M_i}\log(1-a_k z^{-1})+\sum_{k=1}^{M_0}\log(1-b_k z)-$$
$$-\sum_{k=1}^{N_i}\log(1-c_k z^{-1})-\sum_{k=1}^{N_0}\log(1-d_k z). \qquad (7.13)$$

Когда (7.13) вычисляется на единичной окружности, легко видеть, что член $\log[e^{i\omega r}]$ вносит вклад только в минимальную часть комплексного логарифма. Поскольку этот член несет информацию только о взаимном расположении во временно́й области, то при вычислении комплексного кепстра он обычно опускается [2]. Таким образом, при обсуждении свойств комплексного кепстра далее этот член не рассматривается. Используя то обстоятельство, что логарифм можно разложить в степенно́й ряд, относительно несложно показать, что комплексный кепстр имеет вид

$$\hat{x}(n)=\begin{cases}\log[A], & n=0,\\ \sum_{k=1}^{N_i}\frac{c_k^n}{n}-\sum_{k=1}^{M_i}\frac{a_k^n}{n}, & n>0,\\ \sum_{k=1}^{M_0}\frac{b_k^{-n}}{n}-\sum_{k=1}^{N_0}\frac{d_k^{-n}}{n}, & n<0.\end{cases} \qquad (7.14)$$

Уравнения (7.14) позволяют выявить ряд важных свойств комплексного кепстра. Прежде всего, комплексный кепстр в общем случае отличен от нуля и бесконечен как для положительных, так и для отрицательных значений n, даже если $x(n)$ удовлетворяет принципу причинности, устойчив и имеет конечную протяженность. Далее видно, что комплексный кепстр является затухающей последовательностью, ограниченной сверху:

$$|\hat{x}(n)|<\beta\alpha^{|n|}/|n|, \quad |n|\to\infty, \qquad (7.15)$$

где α — максимальное абсолютное значение величин a_k, b_k, c_k, d_k; β — постоянный сомножитель.

Если $X(z)$ не содержит нулей и полюсов вне единичной окружности (т. е. $b_k=d_k=0$), то

$$\hat{x}(n)=0, \quad n<0. \qquad (7.16)$$

Такие сигналы называются минимально-фазовыми [5]. Общий результат для последовательности (7.16) состоит в том, что такая последовательность полностью определяется действительной частью преобразования Фурье. Таким образом, для минимально-фазовых систем комплексный кепстр определяется лишь логарифмом модуля преобразования Фурье. Это можно легко показать, если вспомнить, что действительная часть преобразования Фурье представляет собой преобразование Фурье от четной части последовательности, т. е. если $\log|X(e^{i\omega})|$ — преобразование Фурье кепстра, то

$$c(n)=[\hat{x}(n)+\hat{x}(-n)]/2. \qquad (7.17)$$

Используя (7.16) и (7.17) легко показать, что

$$\hat{x}(n)=\begin{cases}0, & n<0,\\ c(n), & n=0,\\ 2c(n), & n>0.\end{cases} \qquad (7.18)$$

Таким образом, для минимально-фазовых последовательностей комплексный кепстр можно получить путем вычисления кепстра и последующего использования (7.18). Другой важный результат для минимально-фазовых систем заключается в том, что комплексный кепстр можно вычислить рекуррентно по входному сигналу [1, 2, 5]. Рекуррентная формула имеет вид

$$\hat{x}(n)=\begin{cases}0, & n<0,\\ \log[x(0)], & n=0,\\ \frac{x(n)}{x(0)}-\sum_{k=0}^{n-1}\left(\frac{k}{n}\right)\hat{x}(k)\frac{x(n-k)}{x(0)}, & n>0.\end{cases} \qquad (7.19)$$

Аналогичные результаты можно получить и тогда, когда $X(z)$ не содержит полюсов и нулей, лежащих внутри единичной окружности. Такие сигналы называют максимально-фазовыми. Для этого случая, как это видно из (7.14),

$$\hat{x}(n)=0, \quad n>0. \qquad (7.20)$$

Совместное использование (7.16) и (7.17) дает

$$\hat{x}(n)=\begin{cases}0, & n>0,\\ c(n), & n=0,\\ 2c(n), & n<0.\end{cases} \qquad (7.21)$$

Как и в случае минимально-фазовых последовательностей, здесь также можно получить рекуррентное соотношение для кепстра:

$$\hat{x}(n)=\begin{cases}\dfrac{x(n)}{x(0)}-\sum\limits_{k=n+1}^{0}\left(\dfrac{k}{n}\right)\hat{x}(k)\dfrac{x(n-k)}{x(0)}, & n<0,\\ \log[x(0)], & n=0,\\ 0, & n>0.\end{cases} \tag{7.22}$$

Важным специальным случаем является случай входного сигнала вида

$$p(n)=\sum_{r=0}^{M}\alpha_r\,\delta(n-rN_p), \tag{7.23}$$

т. е. последовательности импульсов. Преобразование $P(z)$ имеет вид

$$P(z)=\sum_{r=0}^{M}\alpha_r\,z^{-rN_p}. \tag{7.24}$$

Из (7.24) видно, что $P(z)$ представляет собой полином по степеням z^{-N_p}, а не z^{-1}, как это было ранее. Этот полином можно представить как результат произведения $(1-az^{-N_p})$ и $(1-bz^{-N_p})$. Легко видеть, что комплексный кепстр отличен от нуля только для целых значений аргумента, кратных N_p. Например, предположим, что

$$p(n)=\delta(n)+\alpha\,\delta(n-N_p), \tag{7.25}$$

где $0<\alpha<1$. Тогда

$$P(z)=1+\alpha\,z^{-N_p} \tag{7.26}$$

и

$$\hat{P}(z)=\log(1+\alpha\,z^{-N_p})=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\alpha^n}{n}\,z^{-nN_p}. \tag{7.27}$$

Таким образом, $\hat{p}(n)$ представляет собой импульсную последовательность с периодом

$$p(n)=\sum_{r=1}^{\infty}(-1)^{r+1}\frac{\alpha^r}{r}\,\delta(n-rN_p). \tag{7.28}$$

Как будет видно из результатов § 7.2, тот факт, что комплексный кепстр периодической последовательности импульсов также представляет собой периодическую последовательность импульсов, является чрезвычайно важным для анализа речевых сигналов. Однако перед детальным рассмотрением методов гомоморфной обработки речевых сигналов кратко рассмотрим вопросы применения гомоморфных фильтров для обработки сигналов, подвергнутых операции свертки.

7.1.2. Вычислительные аспекты

Математическое описание характеристической системы и обратного преобразования, представленных на рис. 7.5 и 7.6 соответственно, предполагает применение гомоморфной обработки для сигналов, подвергнутых операции свертки. Если ограничиться рассмотрением абсолютно суммируемых сигналов, то область сходимости z-преобразования будет охватывать единичную окружность, т. е. входная последовательность в этом случае будет иметь преобразование Фурье. В этом случае целесообразно заменить z-преобразование (рис. 7.5 и 7.6) преобразованием Фурье. Другими словами, для важного специального случая последовательностей конечной длины математическое представление характеристической системы относительно свертки имеет вид:

$$X(e^{i\omega})=\sum_{n=0}^{N-1}x(n)\,e^{-i\omega n}; \tag{7.29а}$$

$$\hat{X}(e^{i\omega})=\log[X(e^{i\omega})]=\log|X(e^{i\omega})|+i\arg[X(e^{i\omega})]; \tag{7.29б}$$

$$\hat{x}(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{X}(e^{i\omega})\,e^{i\omega n}\,d\omega. \tag{7.29в}$$

Уравнение (7.29а) представляет собой преобразование Фурье входной последовательности, соотношение (7.29б) — комплексный логарифм спектра входного сигнала, а уравнение (7.29в) — обратное преобразование Фурье логарифма спектра входного сигнала. Как уже отмечалось, возникают вопросы единственности этого множества уравнений. Для определения комплексного кепстра необходимо однозначно определить логарифм преобразования Фурье. Полезно потребовать, чтобы комплексный кепстр действительной последовательности также являлся действительной последовательностью. Напомним, что для действительных последовательностей действительная часть преобразования Фурье является четной функцией, а мнимая часть — нечетной. Таким образом, если комплексный кепстр должен быть действительной функцией, то логарифм модуля должен быть четной функцией, а фазу следует определить как нечетную функцию ω. Далее можно показать, что достаточным условием единственности комплексного логарифма является требование, чтобы фаза вычислялась как периодическая функция ω с периодом 2π [1, 2] (эти условия непрерывности необходимы также для существования преобразования Фурье от $\hat{X}(e^{i\omega})$). Алгоритм вычисления фазы разработан и подробно описан в [2, 6].

Соотношения (7.29) записаны в форме, затрудняющей их непосредственное применение, поскольку эти соотношения требуют вычисления интегралов. Однако можно аппроксимировать (7.29) с использованием дискретного преобразования Фурье. Дискретное преобразование Фурье (ДПФ) последовательности конечной

длительности идентично дискретизированному преобразованию Фурье для той же последовательности [5]. Таким образом, алгоритм быстрого преобразования Фурье позволяет быстро вычислить ДПФ [5]. В предложенном подходе вычисления кепстра следует заменить все преобразования Фурье соответствующими дискретными преобразованиями Фурье. Результирующие уравнения имеют вид:

$$X_p(k) = \sum_{n=0}^{N-1} x(n)\, e^{-i\frac{2\pi}{N}kn}, \quad N \leqslant k \leqslant N-1; \tag{7.30а}$$

$$\hat{X}_p(k) = \log [X_p(k)], \qquad 0 \leqslant k \leqslant N-1; \tag{7.30б}$$

$$\hat{x}_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_p(k)\, e^{i\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1. \tag{7.30в}$$

Уравнение (7.30а) представляет собой обратное дискретное преобразование Фурье логарифма ДПФ последовательности конечной длительности. Индекс p указывает на то, что полученная последовательность не является точно эквивалентной комплексному кепстру, определяемому уравнением (7.29). Это обусловлено тем обстоятельством, что комплексный логарифм, используемый при вычислении ДПФ, является дискретным отображением, и, таким образом, результирующее обратное преобразование представляет собой отображение комплексного спектра, искаженного вследствие эффекта наложения частот [1, 2, 5]. Следовательно, комплексный кепстр, полученный с использованием (7.30), связан с действительным комплексным кепстром соотношением

$$\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN). \tag{7.31}$$

Вычислительные операции, необходимые для построения характеристической системы относительно свертки, представлены на рис. 7.7*а*.

Комплексный кепстр, как это было показано выше, основан на вычислении комплексного логарифма, а кепстр в его традицион-

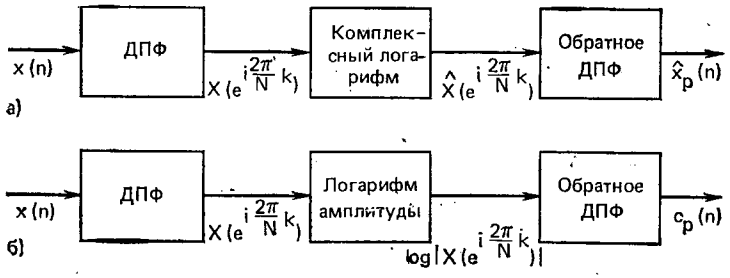


Рис. 7.7. Реализация системы вычисления:
а) комплексного спектра; *б*) кепстра

ном определении основан только на логарифме модуля преобразования Фурье, т. е. кепстр определен в виде

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |X(e^{i\omega})|\, e^{i\omega n}\, d\omega, \qquad -\infty < n < \infty. \tag{7.32}$$

Аппроксимация кепстра получается путем вычисления обратного ДПФ логарифма модуля ДПФ входной последовательности:

$$c_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \log |X_p(k)|\, e^{i\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1. \tag{7.33}$$

Как и ранее, кепстр, полученный с использованием ДПФ, связан с действительным кепстром соотношением

$$c_p(n) = \sum_{r=-\infty}^{\infty} c(n+rN). \tag{7.34}$$

На рис. 7.7*б* показано, как с помощью ДПФ и обратного ДПФ осуществить вычисления, приводящие к (7.34).

Вследствие эффекта наложения частот, присущего дискретному преобразованию Фурье при вычислении кепстра, часто требуется использовать как можно большее значение N. Как показано в [1, 2, 5, 6], большое значение N (т. е. большая частота дискретизации преобразования Фурье) необходимо также и для точного вычисления комплексного логарифма. Однако существование быстрого преобразования Фурье (БПФ) делает возможным использование $N = 512$ или более.

Недавно был предложен иной подход к вычислению кепстра [7] последовательности конечной длительности без нежелательных искажений за счет эффекта наложения частот. Основная идея заключается в непосредственном использовании (7.14) для определения комплексного кепстра через положение нулей (корней) полинома z-преобразования. Этот метод предполагает точное и эффективное вычисление корней полинома сравнительно высокой степени (в задачах обработки речи иногда степень полинома достигает 500). Однако если корни определены с достаточной точностью, то теоретически комплексный кепстр свободен от эффекта наложения частот, присущего вычислительным методам, основанным на реализациях конечной длительности. В [7] приводятся хорошие результаты, полученные для тестовых случаев.

В данном параграфе обсуждались математические и вычислительные аспекты гомоморфных относительно свертки систем. Однако здесь не содержится обсуждение различных частных вопросов и отдельных тонкостей таких систем, поскольку они достаточно полно отражены в [1, 2, 5—7]. Ниже рассмотрено применение гомоморфных относительно свертки систем при анализе речевых сигналов.

7.2. Комплексный кепстр речи

Модели сигналов, с одной стороны, и методы анализа во временной области — с другой, можно объединить и эффективно использовать в теории гомоморфной фильтрации речи. Вспомним, что модель речеобразования обязательно состоит из линейной системы с медленно изменяющимися во времени параметрами и сигнала возбуждения в виде последовательности импульсов или белого шума. Поэтому короткий сегмент вокализованного речевого сигнала целесообразно рассматривать как результат воздействия сигнала возбуждения в виде последовательности импульсов на линейную систему с постоянными параметрами. Аналогично короткий сегмент невокализованного сигнала можно представить как результат возбуждения линейной системы с постоянными параметрами случайным шумом. Короткий сегмент вокализованной речи можно представить в виде

$$s(n) = p(n) * g(n) * v(n) * r(n) = p(n) * h_v(n) = \sum_{r=-\infty}^{\infty} h_v(n - r N_p), \quad (7.35)$$

где $p(n)$ — периодическая импульсная последовательность с периодом N_p отсчетов и $h_v(n)$ — импульсная характеристика линейной системы, отражающая эффект формы источника возбуждения $g(n)$, импульсную характеристику речевого тракта $v(n)$ и импульсную характеристику излучения $r(n)$. Аналогично для невокализованного сегмента сигнала получаем

$$s(n) = u(n) * v(n) * r(n) = u(n) * h_u(n), \quad (7.36)$$

где $u(n)$ — сигнал возбуждения в виде случайного шума; $h_u(n)$ — импульсная реакция системы, объединяющая воздействие речевого тракта и излучения. Для случая вокализованной речи передаточная функция линейной системы имеет вид

$$H_v(z) = G(z) V(z) R(z). \quad (7.37)$$

Для невокализованной речи получаем

$$H_u(z) = V(z) R(z). \quad (7.38)$$

Кратко рассмотрим природу различных компонент в (7.37) и (7.38). Из результатов, приведенных в гл. 3, следует, что передаточная функция речевого тракта имеет вид

$$V(z) = \frac{A z^{-M} \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_o} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})}. \quad (7.39)$$

Для вокализованной речи кроме носовых звуков адекватная модель содержит только полюсы, т. е. $a_k = 0$, $b_k = 0$ для всех k. Для носовых

звуков и невокализованной речи необходимо рассматривать как полюсы, так и нули. Некоторые нули передаточной функции могут лежать вне единичного круга. Для устойчивости системы все ее полюсы должны располагаться внутри единичного круга. Таким образом, поскольку $v(n)$ действительно, полюсы и нули могут возникать лишь в виде комплексно-сопряженных пар. Эффект излучения, результатом которого, как это показано в гл. 3, является подъем в области высоких частот, может быть приближенно смоделирован как

$$R(z) \approx 1 - z^{-1}. \quad (7.40)$$

Наконец, для вокализованной речи импульс возбуждения имеет конечную протяженность. Таким образом, $G(z)$ будет иметь вид

$$G(z) = B \prod_{k=1}^{L_i} (1 - \alpha_k z^{-1}) \prod_{k=1}^{L_o} (1 - \beta_k z), \quad (7.41)$$

где нули α_k и β_k могут быть как внутри, так и вне единичной окружности.

Используя рассмотренные выше модели и результаты 7.1.2, приступим к анализу комплексного кепстра короткого сегмента речевого сигнала (детальное исследование содержится в [8]). Для вокализованного речевого сигнала полный вклад речевого тракта, источника возбуждения и излучения в общем случае может оказаться неминимально-фазовым, что приводит к ненулевым значениям кепстра в области отрицательного времени. Из (7.14) следует, что комплексный кепстр быстро затухает с ростом n. Кроме того, отметим, что вклад в комплексный кепстр от периодического возбуждения проявится в наличии импульсов в точках, кратных периоду возбуждения. Пример анализа (рис. 7.8) иллюстрирует основные особенности вокализованного речевого сигнала. На рис. 7.8*а* показан сегмент вокализованного сигнала, взвешенный с окном Хемминга. На рис. 7.8*б* представлен логарифм модуля дискретного преобразования Фурье. В этой функции имеется периодическая компонента, обусловленная периодическим характером входного сигнала. На рис. 7.8*в* представлен разрывной характер главного значения фазы, а на рис. 7.8*г* — фазовая кривая, лишенная разрывов. Результат преобразования Фурье в комплексный кепстр кривых рис. 7.8*б* и 7.8*г* представлен на рис. 7.8*д*. Отметим наличие пиков в положительном и отрицательном времени и быстрое затухание компонент в области малых времен, что обусловлено совместным воздействием речевого тракта, источника возбуждения и излучением. Кепстр, являющийся обратным преобразованием Фурье логарифма амплитуды модуля спектра, показан на рис. 7.8*е*. В данном случае сохранены все основные особенности комплексного кепстра, как это и предполагалось, поскольку он является четной частью комплексного кепстра.

Последовательность графиков рис. 7.8 показывает, как можно использовать гомоморфную фильтрацию для анализа речевого

сигнала. Прежде всего отметим, что импульс в кепстре, обусловленный квазипериодическим возбуждением, оказывается отделенным от остальных компонент. Это приводит к соответствующей системе гомоморфной фильтрации речевого сигнала, представленной на рис. 7.9. Сегмент речевого сигнала взвешивается с некоторым окном, т. е. кепстр вычисляется так, как это обсуждалось в 7.1.3, и требуемые компоненты кепстра выделяются с использованием «окна по кепстру» $l(n)$. Этот вид фильтрации иногда называют «частотно-инвариантной линейной фильтрацией». В результате взвешенный комплексный кепстр подвергается обратному преобразованию для получения требуемых компонент. Это показано на рис. 7.10. На рис. 7.10*а* и *б* показаны логарифм модуля и фаза, полученные в процессе использования процедуры обратного преобразования в случае, когда

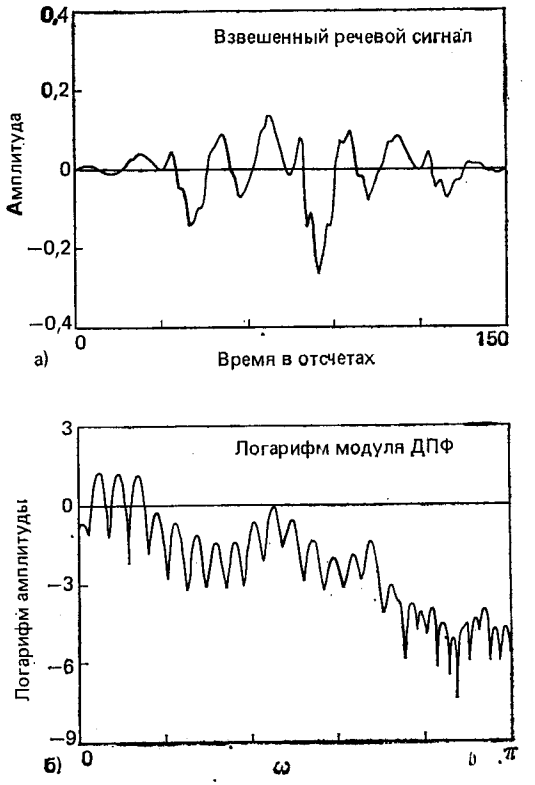
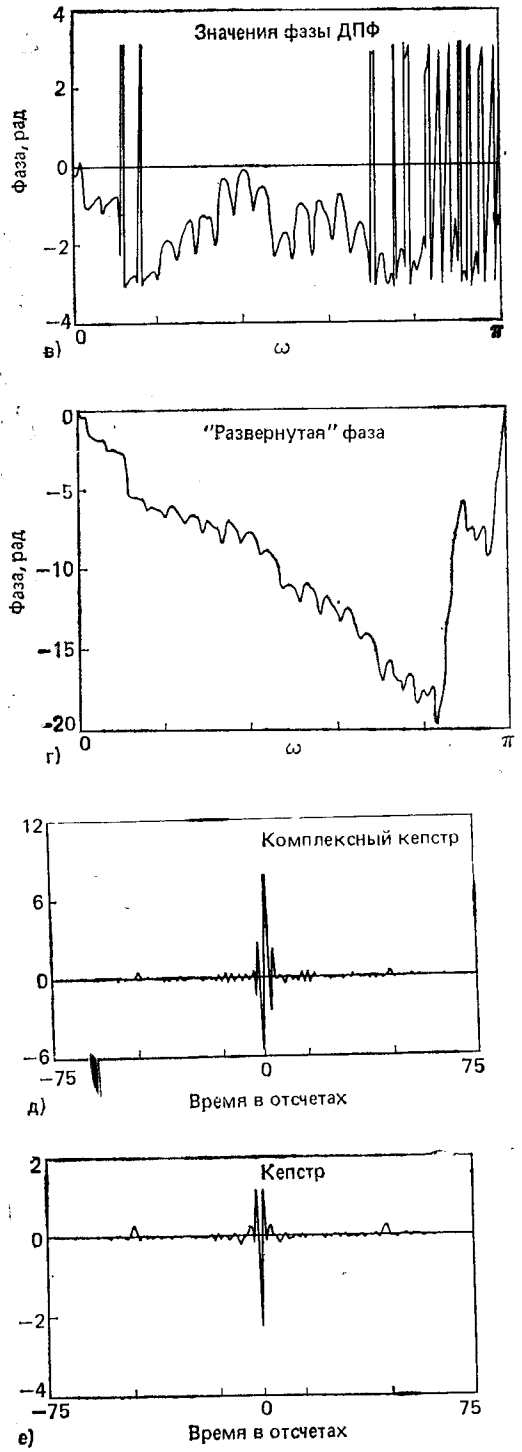


Рис. 7.8. Гомоморфный анализ вокализованной речи:
а) взвешенный речевой сигнал; *б*) логарифм модуля кратковременного преобразования Фурье; *в*) значения фазы; *г*) «развернутая» фаза; *д*) комплексный кепстр; *е*) кепстр

$$l(n)=\begin{cases}1, & |n|<n_0,\\ 0, & |n|\geqslant n_0,\end{cases} \tag{7.42}$$

где n_0 выбрано меньшим, чем период основного тона N_0. Соответствующий выходной сигнал показан на рис. 7.10*в*. (Заметим, что постоянный фазовый сдвиг π радиан был устранен при вычислении кепстра.) Этот сигнал аппроксимирует импульсную реакцию $h_v(n)$, определяемую (7.35). Если выбрать $l(n)$ таким образом, чтобы восстановить компоненты возбуждения, т. е.

$$l(n)=\begin{cases}0, & |n|<n_0,\\ 1, & |n|\geqslant n_0,\end{cases} \tag{7.43}$$



получим результаты (рис. 7.10*г*, *д*, *е*), вычисленные для логарифма модуля, фазы и выходного сигнала. Выходной сигнал аппроксимирует импульсную последовательность возбуждения, амплитуды которой затухают в соответствии с весами окна Хемминга, примененного при взвешивании входного сигнала.

Для полноты иллюстраций применения гомоморфного анализа к обработке речевого сигнала рассмотрим случай анализа невокализованного сегмента речевого сигнала, показанного на рис. 7.11. На рис. 7.11*а* представлен речевой сигнал, взвешенный с окном Хемминга. На рис. 7.11*б* изображен логарифм модуля спектра, а на рис. 7.11*в* — кепстр. Отметим случайные флуктуации в логарифме модуля спектра. Это связано с тем, что возбуждение в данной ситуации случайно и преобразование Фурье короткого сегмента содержит случайную компоненту. В этом случае результаты малочувствительны к вычислению фазы. Сказанное подтверждается рис. 7.11*в*, на котором отсутствуют пики возбуждения, возникавшие в случае вокализованного сигнала, однако область малых времен в кепстре содержит информацию о $H_u(e^{i\omega})$. Это видно из рис. 7.11*г*, где показан ло-

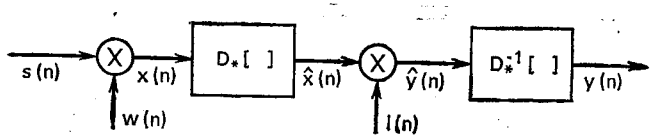
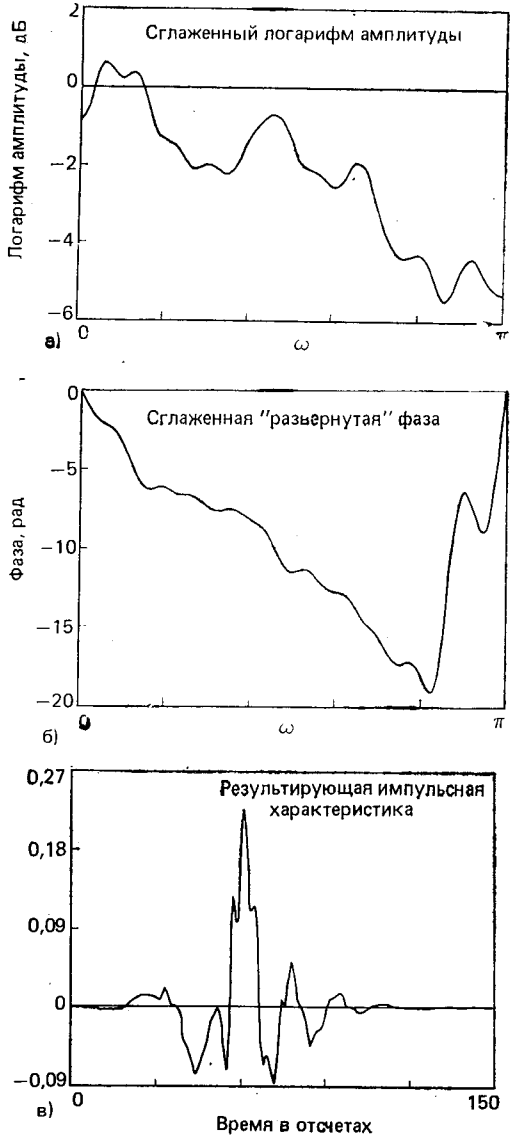


Рис. 7.9. Реализация системы гомоморфной фильтрации речи

тарифм модуля, полученный применением кепстрального окна (7.42) к кепстру 7.11*в*.

Предшествующее обсуждение и примеры показывают, что с помощью гомоморфной фильтрации можно выделить ряд важных компонент речевого сигнала. Тем не менее этот подход не используется достаточно широко, поскольку в ряде речевых приложений полное разложение сигнала не требуется. Чаще сталкиваются с необходимостью оценки таких параметров, как период основного тона и частоты формант. Для этих целей кепстральный анализ весьма эффективен. Таким образом, для большинства задач обработки речи можно избежать обременительных вычислений фазы. Отметим, например, сравнивая 7.8*е* и 7.11*в*, что кепстр позволяет отделять вокализованную речь от невокализованной и, кроме того, период основного тона для вокализованного речевого сигнала хорошо просматривается на кепстральных диаграммах. Частоты формант также можно определить с использованием логарифма модуля передаточной функции речевого тракта, которая вычисляется по кепстру с помощью кепст-



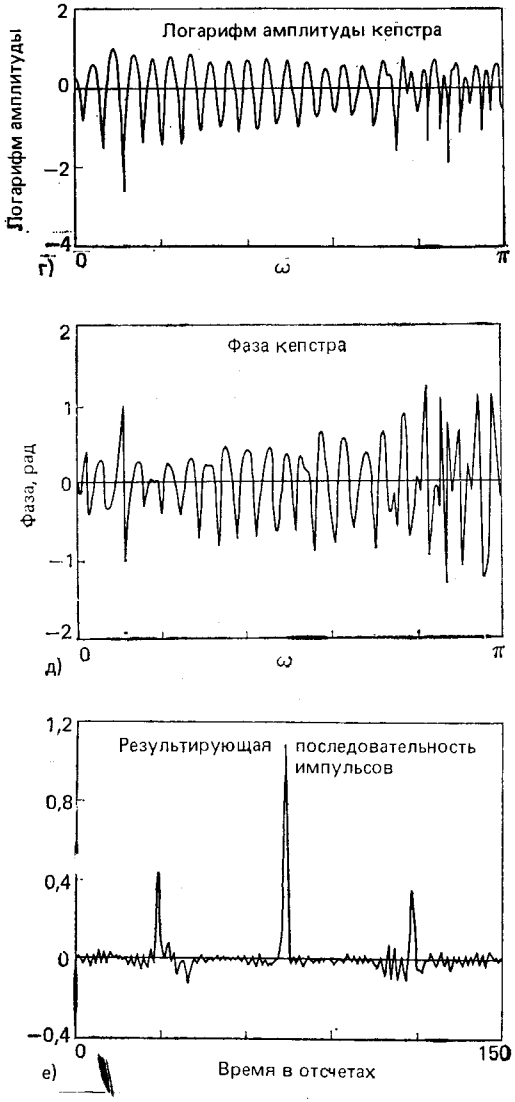


Рис. 7.10. Гомоморфная фильтрация вокализованной речи:
а) и *б*) оценка амплитуды и фазы $H_v(e^{j\omega})$; *в*) оценка импульсной характеристики; *г*) и *д*) оценка амплитуды и фазы $P(e^{j\omega})$; *е*) оценка $p(n)$

рального окна (7.42). В последующих разделах главы кепстральный подход применяется для оценивания формантных частот и периодов основного тона, а также для построения вокодерной системы передачи речевого сигнала.

7.3. Оценивание основного тона

Рисунки 7.8*е* и 7.11*в* позволяют сделать вывод о возможности оценивания периода основного тона с использованием гомоморфной обработки. Было отмечено, что для вокализованного сегмента речи пик в кепстре возникает при задержке, соответствующей периоду основного тона. Для невокализованного сегмента такие пики в кепстре не возникают. Это свойство кепстра может быть использовано для классификации вокализованный/невокализованный и для периода основного тона на вокализованной речи.

Метод оценивания основного тона на основе кепстрального анализа, разработанный для нереального масштаба времени, чрезвычайно прост. Кепстр, полученный в соответствии с результатами 7.1.3, исследуется с целью отыскания пика в области возможных значений основного тона. Если пик в кепстре превышает порог, то сегмент классифицируется как вокализованный, а координата пика дает хорошую оценку периоду основного

тона. Если максимум кепстра не превышает порога, то сегмент классифицируется как невокализованный. Изменение во времени типа возбуждения и периода основного тона можно оценить с использованием зависящего от времени кепстра, что достигается на основе вычисления зависящего от времени преобразования Фурье. Обычно кепстр вычисляется 1 раз через каждые 10—20 мс, поскольку в нормальной речи параметры возбуждения не изменяются быстрее.

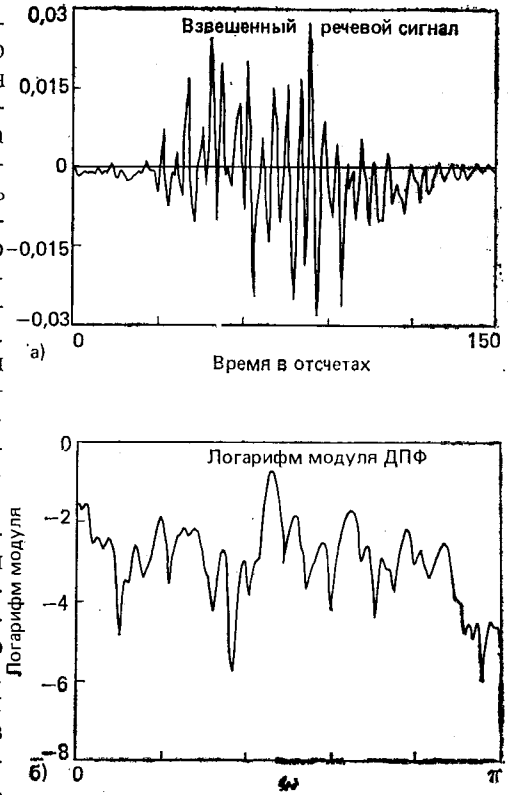
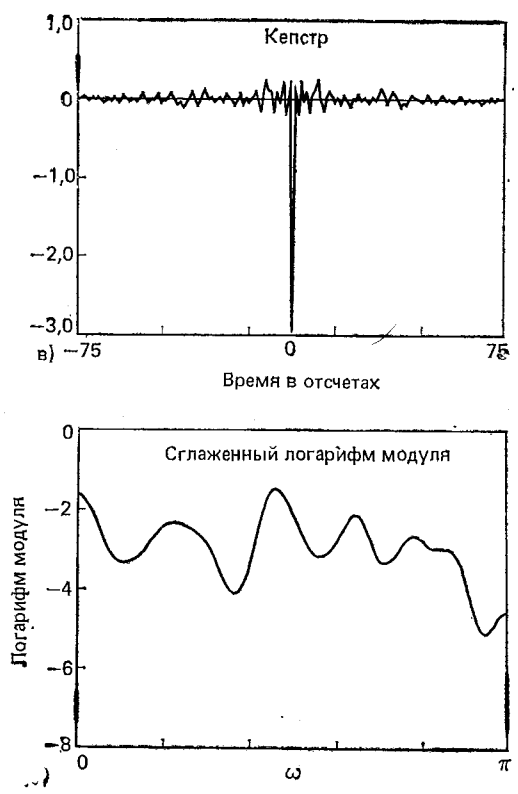


Рис. 7.11. Гомоморфный анализ невокализованной речи:
а) взвешенный речевой сигнал; *б*) логарифм модуля кратковременного преобразования Фурье; *в*) кепстр; *г*) оценка $H_u(e^{j\omega})$

На рис. 7.12 и 7.13 показан пример, полученный Ноллом [9], который первым описал процедуру оценивания периода основного тона на основе кепстра. На рис. 7.12 показана серия логарифмических спектров и соответствующих им кепстров для мужского голоса. Кепстры, изображенные на данном рисунке, представляют собой квадратный корень из $c(n)$, как они были определены выше. В этом примере частота дискретизации на входе составляла 10 кГц. Окно Хемминга протяженностью 40 мс (400 отсчетов) перемещалось с шагом 10 мс, т. е. логарифмические спектры слева и кепстры справа вычислялись через каждые 10 мс. Как следует из рис. 7.12, первые семь 40-миллисекундных интервалов соответствуют невокализованному сигналу, а остальные интервалы — вокализованной речи, причем период основного тона возрастает с течением времени, т. е. частота основного тона падает. На рис. 7.13 показан пример анализа женского голоса. В этом случае речевой сигнал, соответствующий последовательности кепстров и спектров, оказывается вокализованным в начале и невокализованным в конце. Легко видеть, что в конце вокализованного сегмента период основного тона удваивается, что нередко бывает по окончании вокализованного сегмента. Сравнение рис. 7.12 и 7.13 показывает, что для женского голоса частота основного тона гораздо выше, чем для мужского.



Эти два примера, хорошо иллюстрирующие результаты анализа основного тона для речевых сигналов, могут привести к предположению о том, что на основе гомоморфного анализа можно построить очень простой и эффективный алгоритм выделения основного тона и классификации речи на вокализованную/невокализованную. К сожалению, как это зачастую бывает при анализе речи, имеется ряд практических вопросов и трудностей, которые должны быть решены при разработке алгоритма на основе кепстра. Нолл [9] описал одну из возможных схем анализа речи на основе кепстра. Но имеется и ряд других схем, которые успешно используются для этих целей. Вместо того чтобы описывать здесь детально каждую из известных схем, целесообразнее остановиться на тех неизбежных трудностях, которые возникают при построении кепстральных анализаторов основного тона.

Во-первых, наличие выброса в кепстре в диапазоне 3—20 мс очень точно указывает на то, что данный сегмент является вокализованным. Однако отсутствие пика или наличие слабого пика не означает, что данный сегмент является невокализованным. Амплитуда или даже просто существование пика в кепстре зависит от целого ряда факторов, включая длину окна, используемого для взвешивания входного сигнала, и формантной структуры самого сигнала. Легко показать (см. задачу 7.10), что наибольшая амплитуда пика в кепстре равна единице. Это достигается только в случае абсолютного совпадения периодов основного тона. Это, конечно, совершенно не достижимо в реальном случае, даже если использовать прямоугольное временно́е окно, включающее целое число периодов. Прямоугольные временны́е окна применяются весьма редко вследствие худших результатов, даваемых ими при оценивании спектра. В случае, например, окна Хемминга очевид-

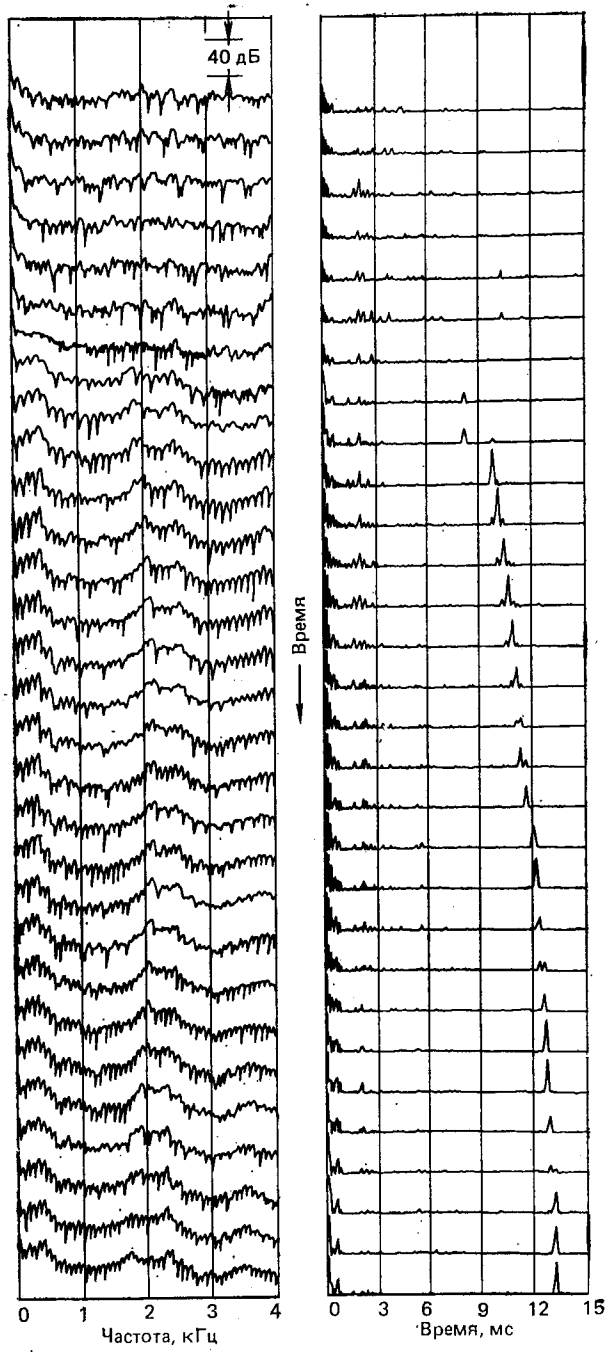


Рис. 7.12. Набор логарифмов спектра и кепстров для мужского голоса [9]

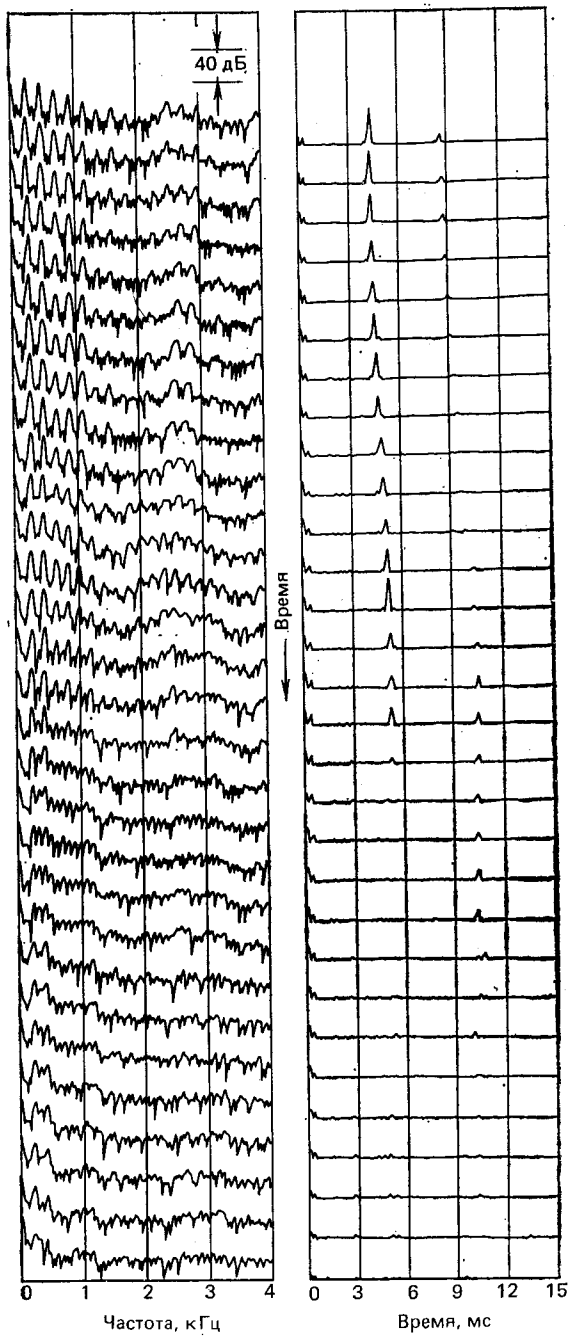


Рис. 7.13. Набор логарифмов спектра и кепстров для женского голоса

но, что как протяженность окна, так и его относительное расположение по отношению к речевому сигналу будут оказывать значительное влияние на величину наибольшего пика в кепстре.

Как крайний случай предположим, что окно имеет протяженность менее двух периодов основного тона. Очевидно, что при этом трудно ожидать точного оценивания периодичности по спектру или кепстру сигнала. Таким образом, протяженность окна может оказаться такой, что с учетом уменьшения амплитуды данных к границам выборки, по крайней мере, два периода основного тона пропадут во взвешенных данных. Для мужской речи с низкой частотой основного тона требуется окно порядка 40 мс. Для голосов с более высокой частотой основного тона требуются пропорционально меньшие окна. Желательно, конечно, выбирать окно настолько малым, насколько это возможно, чтобы избежать значительных изменений параметров сигнала на протяжении используемого сегмента. Чем длиннее окно, тем значительнее изменения параметров в пределах окна и тем больше отклонение от принятой модели анализа. Один из способов выбора окна, при котором оно было и не слишком длинным и не слишком коротким, состоит в адаптации длины окна с учетом предшествующих (или возможно среднего значения) оценок периодов основного тона [10, 11].

Другая причина, по которой сигнал может сильно отличаться от описываемого моделью, заключается в чрезмерном ограничении полосы. Ярким примером подобной неадекватности может служить синусоидальный сигнал. В логарифме спектра такой сигнал даст только один пик. Поскольку в спектре нет периодических колебаний, в кепстре не будет пиков. В речевом сигнале вокализованные сегменты обычно очень узкополосны с плохо выраженной гармонической структурой на частотах выше нескольких сотен герц. В этом случае пики в кепстре отсутствуют. К счастью, для любых значений периода, за исключением самых коротких, область, в которой возникают пики в кепстре, не содержит других компонент, кроме основного тона. Таким образом, для определения положения импульса основного тона можно использовать достаточно низкий порог (порядка 0,1).

При правильно подобранной протяженности окна на входе положение и амплитуда импульса кепстра обеспечивают в большинстве случаев хорошую оценку периода основного тона и классификации тон/шум. В тех случаях, когда кепстральный анализ не позволяет точно ответить на вопрос о наличии импульсов основного тона и значении периода, для вынесения окончательного решения можно привлечь дополнительную информацию о виде функции среднего числа переходов через нуль, энергии сигнала или улучшить оценки с помощью сглаживания [11]. Дополнительная логика при реализации устройств оценивания на основе кепстра требует усложнения аппаратуры. Эта часть общей схемы выделения основного тона вносит незначительную долю в общие вычислительные затраты, но вместе с тем оказывается весьма полезной.

7.4. Оценивание формант

На основе примеров § 7.2 можно сделать вывод, что часть кепстра в области малых времен в основном содержит информацию о речевом тракте, источнике возбуждения и излучении, в то время как в области больших времен заключена информация о сигнале возбуждения. Это свойство использовалось для классификации сегментов и оценивания периода основного тона путем использования части кепстра в области больших времен. Примеры § 7.2 указывают также и на метод получения отклика речевого тракта на основе кепстра. Действительно, отметим, что «сглаженные» логарифмы модуля (см. рис. 7.10*а* и *г*) получаются путем взвешивания кепстра. Эти сглаженные спектры отражают резонансную структуру речевого сигнала, т. е. пики в спектре соответствуют формантным частотам. Это означает, что последние можно оценить по положению максимумов в «кепстрально сглаженном» логарифмическом спектре.

Рассмотрим модель речеобразования, представленную на рис. 7.14. Эта чрезвычайно упрощенная модель описывает вокализо-

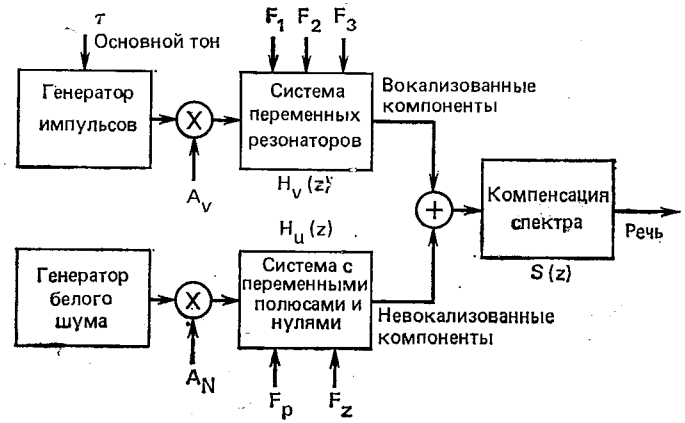


Рис. 7.14. Цифровая модель речеобразования

ванный речевой сигнал с помощью периода и амплитуды основного тона и трех первых формант, а невокализованную речь с помощью амплитудного значения и положения единственного нуля и полюса. Для компенсации свойств в области высоких частот используется дополнительный неперестраиваемый фильтр. Все перечисленные параметры, конечно, изменяются с течением времени. Метод оценивания этих параметров основан на вычислении кепстрально сглаженного логарифма спектра через каждые 10—20 мс. По кепстру производится анализ на вокализованность сегмента и определяется положение максимумов. Если сегмент вокализован, то по кепстру определяются период основного тона и первые три формантные частоты, которые вычисляются по систе-

ме логических правил, учитывающих применяемую модель [11, 12]. В случае невокализованного сегмента полюс определяется в точке максимума спектральной плотности, а нуль — в том месте, где сохраняется относительная амплитуда между максимумом и минимумом [12].

Иллюстрация использования метода оценивания периода основного тона и формантных частот для вокализованной речи представлена на рис. 7.15. Слева показана последовательность кепстров, вычисленных через каждые 20 мс, справа — последова-

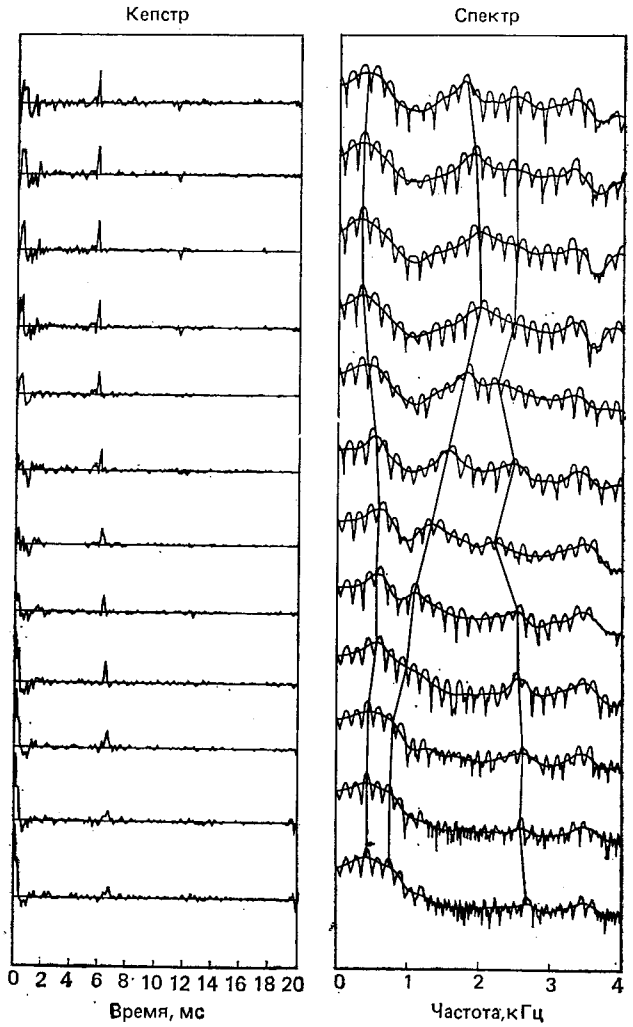


Рис. 7.15. Автоматическое оценивание формантных траекторий по кепстрально сглаженному логарифму спектра

тельность логарифмов спектров с соответствующими сглаженными спектральными оценками, полученными на основе кепстра. Линиями соединены максимумы, выделенные с использованием алгоритма [11] для трех первых формант. Из рис. 7.15 видно, что две первые форманты сблизились настолько, что они не являются уже двумя отдельными максимумами. Эту ситуацию можно обнаружить и отделить пики, если вычислить z-преобразование $H_v(z)$ по контуру, проходящему вблизи полюсов. Вычисление производится с помощью алгоритма спектрального анализа, названного острым z-преобразованием (CZT) [13]. Пример улучшенного разделения показан на рис. 7.16.

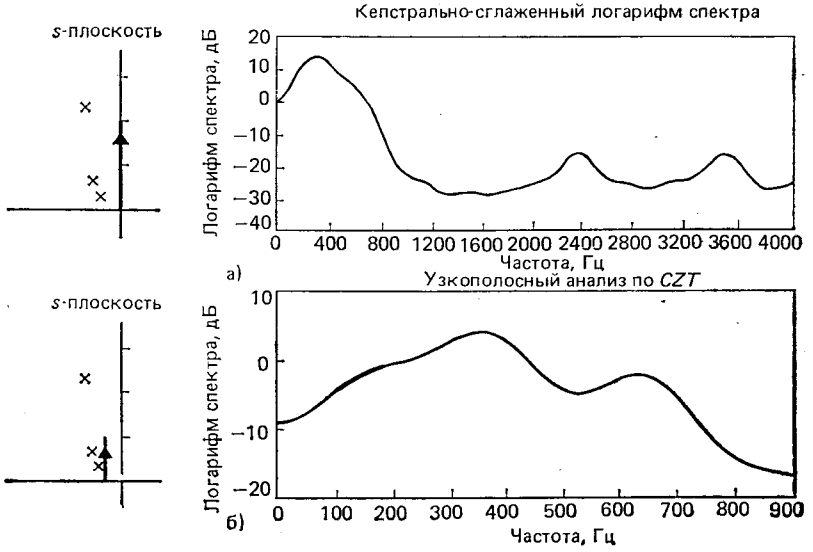


Рис. 7.16. Улучшение разрешения по частоте с помощью алгоритма CZT [11]

Другой подход к оцениванию формант с использованием кепстра предложил Оливье [14], который применил итеративную процедуру, напоминающую метод анализа через синтез, для определения положения полюсов передаточной функции, подгоняемой к сглаженному спектру по критерию минимума среднего квадрата ошибки.

Речевой сигнал можно синтезировать по формантам и периоду основного тона, как это было описано выше, путем простого управления моделью (см. рис. 7.14) с оцененными параметрами. В этом случае для вокализованной речи передаточная функция модели имеет вид

$$V(z)=\prod_{k=1}^{4}\frac{1-2e^{-\alpha_k T}\cos(2\pi F_k T)+e^{-2\alpha_k T}}{1-2e^{-\alpha_k T}\cos(2\pi F_k T)z^{-1}+e^{-2\alpha_k T}z^{-2}}. \qquad (7.44)$$

Это соотношение описывает каскадно соединенные цифровые ре-

зонаторы с единичным коэффициентом усиления на нулевой частоте так, что амплитуда речевого сигнала определяется амплитудой управления A_v. Первые три формантные частоты F_1, F_2 и F_3 изменяются во времени, в то время как F_4 зафиксирована на частоте 4000 Гц и $T=0{,}0001$ с (т. е. частота дискретизации 10 кГц). Полосы формант α_k также зафиксированы на уровне средних для речевого сигнала значений. Неадаптивный фильтр для компенсации влияния источника возбуждения и излучения имеет следующую передаточную функцию:

$$S(z)=\frac{(1-e^{-aT})(1+e^{-bT})}{(1-e^{-aT}z^{-1})(1+e^{-bT}z^{-1})}, \qquad (7.45)$$

где коэффициенты a и b выбраны так, чтобы обеспечивать хорошее приближение спектра. Целесообразно выбрать a и b равными 400π и 5000π соответственно. Более точные значения коэффициентов для данного диктора могут быть получены с использованием спектра, усредненного за большой интервал времени.

Для невокализованного речевого сигнала влияние речевого тракта моделировалось линейной системой с передаточной функцией

$$V(z)=\frac{(1-2e^{-\beta T}\cos(2\pi F_p T)+e^{-2\beta T})(1-2e^{-\beta T}\cos(2\pi F_z T)z^{-1}+\rightarrow}{(1-2e^{-\beta T}\cos(2\pi F_p T)\ z^{-1}+e^{-2\beta T}\ z^{-2})(1-2e^{-\beta T}\times\rightarrow}$$
$$\frac{\rightarrow+e^{-2\beta T}\ z^{-2})}{\rightarrow\times\cos(2\pi F_z T)+e^{-2\beta T})},$$

где F_p взята как максимальное значение сглаженного логарифма спектра на частотах выше 1000 Гц, а F_z удовлетворяет эмпирической формуле

$$F_z=(0{,}0065F_p+4{,}5-\Delta)(0{,}014F_p+28). \qquad (7.46)$$

Здесь

$$\Delta=20\log_{10}|H[e^{i2\pi F_p T}]|-20\log_{10}|H(e^{i0})| \qquad (7.47)$$

обеспечивает сохранение соответствующего соотношения амплитуд [12]. То, что такая относительно простая модель отражает все основные свойства спектра речевого сигнала, иллюстрируется на рис. 7.17 и 7.18, где сравниваются сглаженные логарифмы спектров и результаты, даваемые моделью, определяемой рис. 7.14 и соотношениями (7.44)—(7.47) как для вокализованных, так и для невокализованных речевых сигналов соответственно. В качестве примера речевого сигнала, синтезированного с использованием описанной модели, может служить сигнал, представленный на рис. 7.18. В верхней части рисунка изображены траектории параметров, построенные по речевому сигналу, спектрограмма которого представлена на рис. 7.19*б*. На рис. 7.19*в* показана спектрограмма синтезированного сигнала, полученного с использованием модели рис. 7.14 на основе параметров рис. 7.19*а*. Очевидно, что в синтезированном сигнале хорошо сохранились основные

черты исходной речи. Фактически, несмотря на чрезвычайно грубый способ описания, синтетическая речь очень разборчива и сохраняет многие черты индивидуальности диктора. В действительности период основного тона и формантные частоты оцениваются

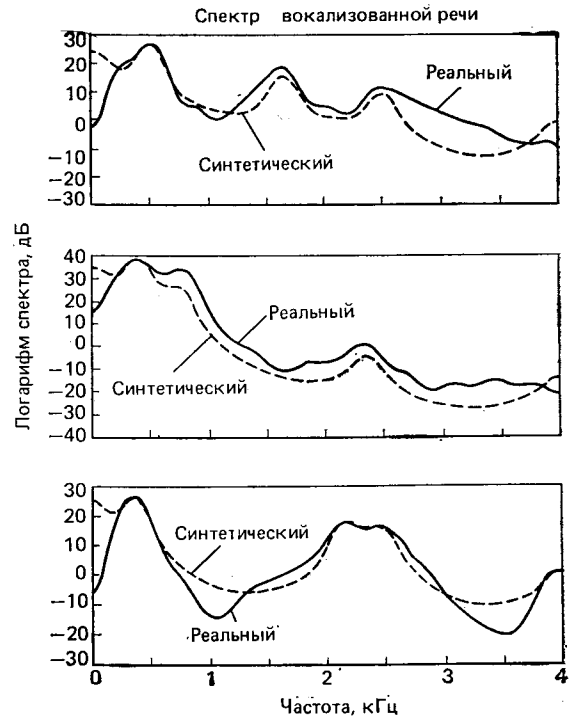


Рис. 7.17. Сравнение кепстрально сглаженного спектра и спектра модели для вокализованного сигнала

с использованием алгоритма, основанного на результатах обширных исследований в области верификации диктора (см. гл. 8 и § 9.2).

Ценное свойство рассмотренного представления заключается в возможности получения очень малых требуемых скоростей передачи. Полная система анализа—синтеза, основанная на этом представлении (формантный вокодер), показана на рис. 7.20. Параметры модели оценивались 100 раз в секунду и фильтровались для устранения шума. Частота дискретизации понижалась до удвоенной частоты среза фильтра и затем параметры квантовались. При синтезе каждый параметр интерполировался с целью получения частоты дискретизации 100 Гц и использовался в синтезаторе так, как это показано на рис. 7.14.

Для выявления подходящего множества параметров были проведены аудиторные испытания. Анализатор и синтезатор соединялись друг с другом для получения образцов сигнала. Затем

параметры подвергались фильтрации с помощью фильтров нижних частот и определялась такая полоса фильтра, при которой отсутствуют слышимые различия между синтезированными сигналами с отфильтрованными и неотфильтрованными параметрами.

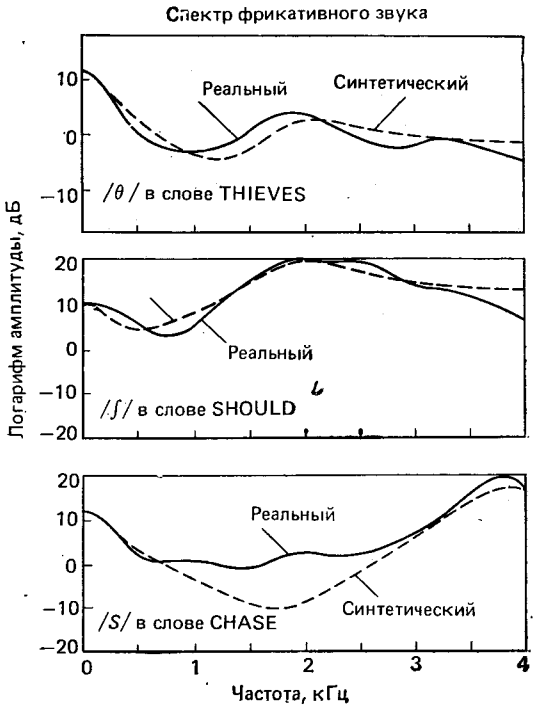


Рис. 7.18. Сравнение кепстрально сглаженного спектра и спектра модели для невокализованного сигнала

Обнаружено, что частота фильтра может быть выбрана равной 16 Гц без заметных различий в качестве. Полученные параметры затем дискретизировались с частотой около 33 Гц (прореживание 3 : 1). Затем был проведен эксперимент по определению требуемой скорости передачи. Формантные параметры и период основного тона квантовались с использованием линейного квантователя (подобранного для каждого параметра), а амплитудные параметры — с использованием логарифмического квантователя. Результаты данного эксперимента при анализе их с точки зрения качества восприятия представлены в табл. 7.1. При использовании частоты дискретизации 33 Гц и данных табл. 7.1 обнаружено, что для полностью вокализованной фразы качество сигнала по сравнению с синтезированным сигналом без кватования не снижается до скоростей порядка 600 бит/с. (Отметим, что для адекватного описания переходов тон/шум требуется передача соответствующего признака с частотой 100 Гц.)

На рис. 7.21*а* показан пример траектории параметров, оцененных по исходному речевому сигналу с частотой 100 раз в секунду. На рис. 7.21*б* представлены те же параметры после сглаживания с использованием КИХ-фильтра нижних частот с полосой

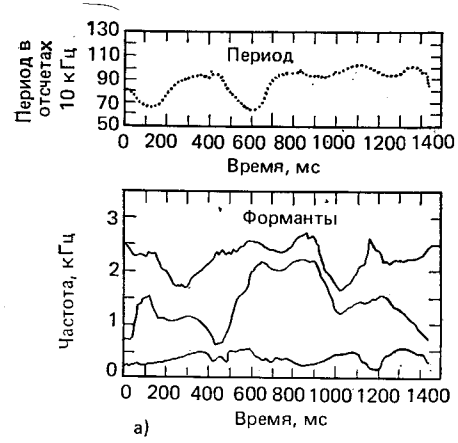
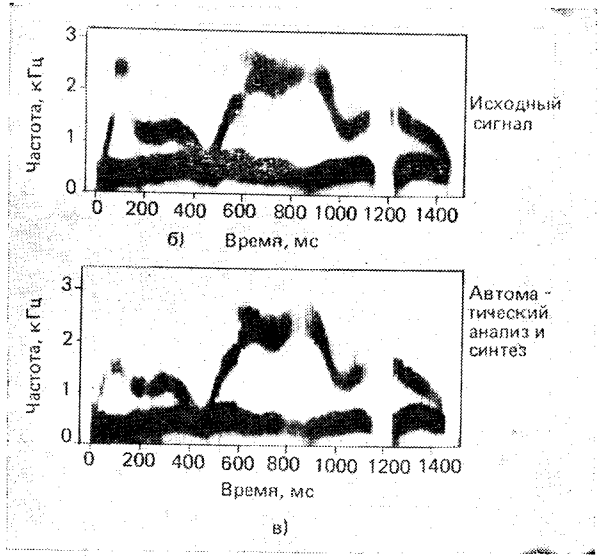


Рис. 7.19. Автоматический анализ и синтез фразы «We were away a year ago»:
а) период основного тона и формантные траектории, построенные на ЭВМ; *б*) широкополосные спектрограммы исходного сигнала; в) широкополосная спектрограмма синтетической речи [11]



16 Гц. На рис. 7.21*в* показаны траектории параметров после прореживания, квантования и интерполяции с коэффициентом 3. Хотя между трактрнями во всех трех случаях и имеется видимое различие, но различие между образцами по восприятию незначительно или вовсе отсутствует. Это представление сигнала использовано для экспериментов по синтезу речи в системах машинного ответа [16] (см. 9.1.3).

Таблица 7.1

Результаты анализа формантного вокодера по слуховому восприятию

Параметр	Необходимое количество двоичных единиц на отсчет
τ	6
F_1	3
F_2	4
F_3	3
$\log [A_V]$	2

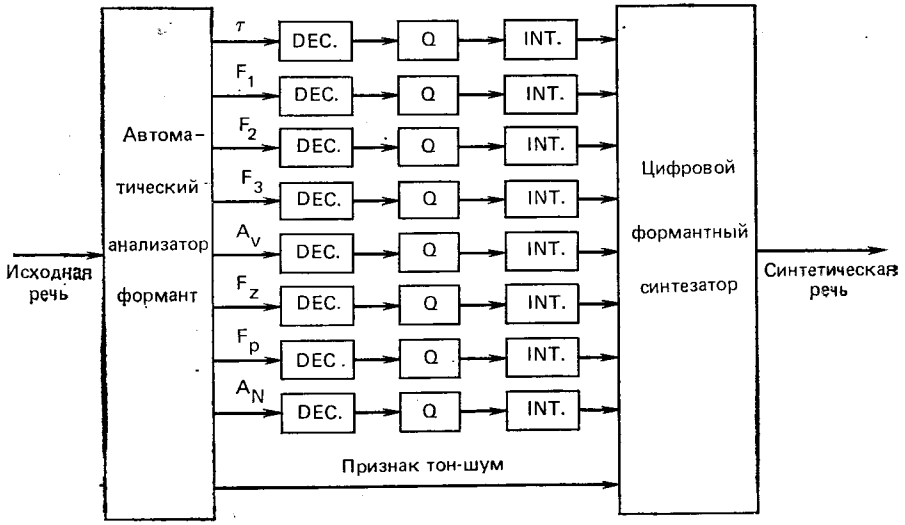


Рис. 7.20. Структурная схема формантного вокодера: DEC. — прореживание; INT. — интерполяция

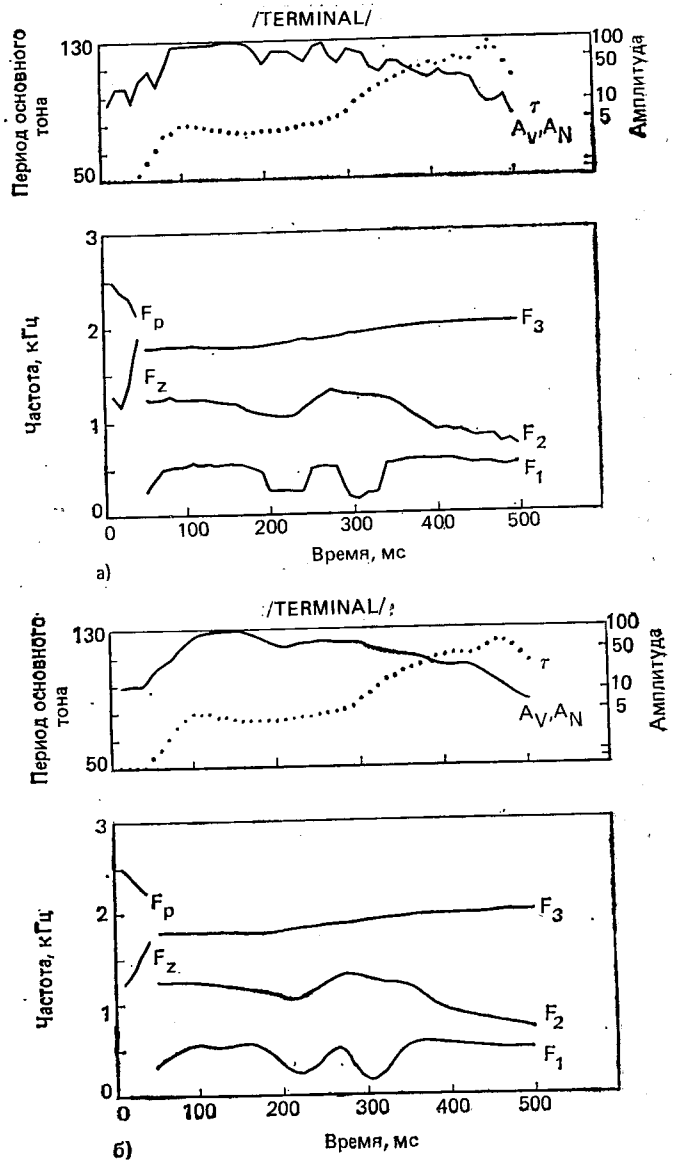
7.5. Гомоморфный вокодер

Как было показано выше, текущая гомоморфная обработка речевого сигнала приводит к весьма удобному описанию, где основные параметры сигнала отделены друг от друга, т. е. информация о сигнале возбуждения расположена в области больших времен, а информация о речевом тракте и форме импульса возбуждения — в области малых времен кепстра. Зависящий от времени комплексный спектр фактически содержит ту же информацию, что и текущий спектр сигнала, который, в свою очередь (см. гл. 6), является точным описанием речевого сигнала. Кепстральное представление, однако, не использует информации о фазе сигнала, содержащейся в преобразовании Фурье, и поэтому кратковременный кепстр не позволяет единственным образом описать речевое колебание. Тем не менее на основе кепстра можно оценить формантные частоты, период основного тона и классифицировать сигнал как вокализованный или невокализованный. Кепстр используется также для непосредственного описания речи в системах, называемых гомоморфными вокодерами [17].

В гомоморфном вокодере кепстр вычисляется 1 раз через каждые 10—20 мс. Период основного тона и признак тон/шум оцениваются по кепстру, а компоненты кепстра в области малых времен (примерно первые 30 отсчетов) квантуются и кодируются для передачи или хранения. По квантованным отсчетам кепстра в области малых времен в синтезаторе восстанавливается импульсная реакция $h_v(n)$ или $h_u(n)$ и вычисляется свертка с функцией возбуждения, восстановленной в синтезаторе по информации об основном тоне, признаке тон/шум и соответствующих амплитудах. Этот алгоритм представлен на рис. 7.22. На рис. 7.22*а* показан анализатор. Кепстр вычисляется в соответствии с 7.1.3. Затем с помощью кепстрального окна выделяется область малых времен. В [17] при моделировании использовались первые 26 отсчетов кепстра. Полный кепстр использовался также для выделения информации и об основном тоне и признаке тон/шум в соответствии с результатами § 7.3. Информация о сигнале возбуждения совместно с квантованными значениями кепстра использовалась для цифрового представления сигнала и передавалась по каналу 50—100 раз в секунду. Для синтеза входного сигнала по кепстральному описанию вычислялась импульсная реакция. Для того чтобы понять, как это делалось, вспомним, что кепстр — это четная функция времени и поэтому для построения кепстра достаточно знать лишь его часть, локализованную в области положительного времени.

Преобразование Фурье части кепстра в области малых времен приводит к логарифму передаточной функции, описывающей совместное влияние речевого тракта, формы импульса возбуждения и излучения. Однако фаза в данном случае равна нулю. В схеме рис. 7.22*б* преобразование Фурье изменяется для получения действительного четного преобразования, обратное преобразование которого представляет собой «импульсную характеристику», являющуюся четной функцией. Импульсная характеристика, полученная таким образом по кепстру (см. рис. 7.8*е*), показана на рис. 7.23*а*. Эту импульсную характеристику можно свернуть с последовательностью импульсов, отстоящих друг от друга на период основного тона для вокализованной речи, и с равноотстоящей последовательностью импульсов случайной полярности для невокализованных сегментов. (В [17] расстояние между импульсами для невокализованного сигнала превосходило единицу для уменьшения объема вычислений.)

По логарифмическому спектру можно получить и минимально-фазовую импульсную характеристику, для чего следует использовать кепстральное окно вида



$$l(n)=\begin{cases}1, & n=0,\\ 2, & 0<n\leqslant n_0,\\ 0, & \textit{в противном случае.}\end{cases} \qquad (7.48)$$

Результат преобразования, приводящий к минимально-фазовой характеристике, показан на рис. 7.23*б* [5]. Импульсная характе-

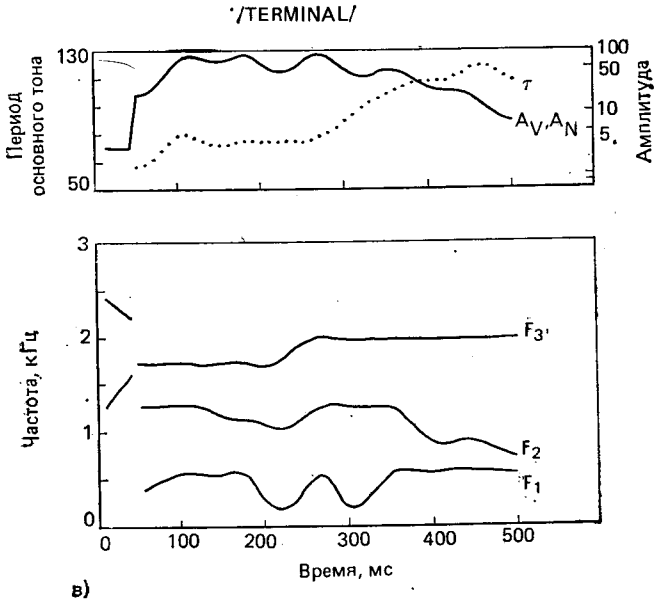


Рис. 7.21. Иллюстрация процесса квантования управляющих сигналов формантного вокодера:
а) исходные данные; *б*) сглаженные данные; *в*) квантованные и сглаженные данные

ристика, изображенная на рис. 7.23*б*, имеет такой же логарифм преобразования Фурье, как и исходная (рис. 7.23*а*). Оппенгейм [17] рассмотрел также случай максимально-фазового восстанов-

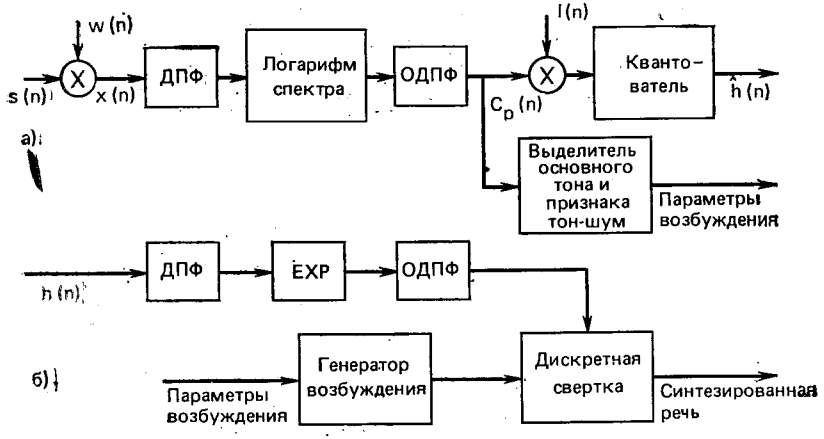


Рис. 7.22. Структурная схема гомоморфного вокодера:
а) анализатор; *б*) синтезатор

ления импульсной реакции, т. е.

$$l(n)=\begin{cases}1, & n=0,\\ 2, & -n_0\leqslant n<0,\\ 0, & \textit{в противном случае.}\end{cases} \tag{7.49}$$

Этот случай для нашего примера представлен на рис. 7.23*в*. Тесты на восприятие показали, что минимально-фазовое описание является наиболее предпочтительным. Это вполне естественно вследствие того, что минимально фазовый сигнал наиболее соответствует речевому сигналу.

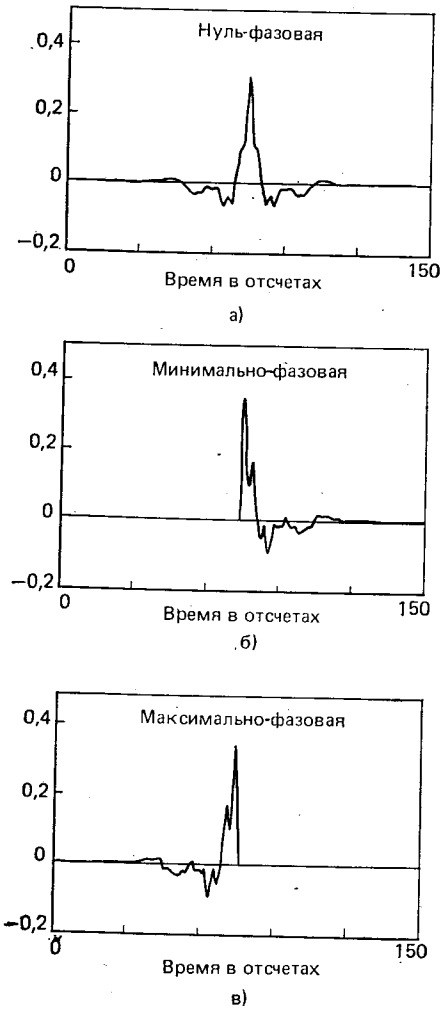


Рис. 7.23. Импульсная характеристика, вычисленная по кепстру:
а) нуль-фазовая; *б*) минимально-фазовая; *в*) максимально-фазовая

Гомоморфный вокодер с 26 значениями кепстра, квантованными с частотой 50 Гц, обеспечивает «очень высокое качество и натуральность речевого сигнала» [17]. Последующие исследования показали, что при преобразовании кепстральной информации перед квантованием скорость передачи может быть значительно понижена [18]. Другие исследования показали, что для повышения эффективности кепстральных методов целесообразно применять адаптацию протяженности временно́го окна, используемого при вычислении спектра сигнала [19].

Гомоморфный вокодер, как и любые вокодерные системы, в которых пытаются разделить параметры речи на сигнал возбуждения и параметры речевого тракта, позволяет достигнуть малой скорости передачи и дополнительной гибкости при обработке речи ценой усложнения в описании и потерь в качестве. Данная система обладает тем преимуществом, что кепстр, требующий для своего вычисления наибольших затрат, позволяет оценить как параметры речевого тракта, так и параметры возбуждения. Данный метод наиболее привлекателен, если имеется возможность использования БИС для вычисления ДПФ.

7.6. Заключение

В данной главе рассмотрены основные методы гомоморфной обработки сигналов применительно к речи. Основная идея гомоморфной обработки заключается в разделении или обратной свертке сегмента речевого сигнала на компоненты, представляющие импульсную характеристику и источник возбуждения. Это достигается путем линейной фильтрации обратного преобразования Фурье логарифма спектра сигнала, т. е. кепстра сигнала. Рассмотрены вычислительные аспекты применения гомоморфной обработки речи. В заключительной части главы изложены некоторые основные методы оценивания параметров сигнала на основе гомоморфной модели.

Задачи

7.1. Комплексный кепстр последовательности является обратным преобразованием Фурье комплексного логарифма спектра

$$\hat{X}(e^{i\omega})=\log|X(e^{i\omega})|+i\arg[X(e^{i\omega})].$$

Показать, что кепстр $c(n)$, определенный как обратное преобразование Фурье логарифма модуля, является четной частью $\hat{x}(n)$, т. е. показать, что $c(n)=\hat{x}(n)+\hat{x}(-n)/2$.

7.2. Рассмотрим полюсную модель речевого тракта в виде

$$V(z)=\frac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^{*} z^{-1})},$$

где $c_k=r_k e^{i\theta_k}$.

Показать, что соответствующий кепстр имеет вид

$$\hat{v}(n)=2\sum_{k=1}^{q}\frac{(r_k)^n}{n}\cos(\theta_k n).$$

7.3. Рассмотрим полюсную модель, описывающую речевой тракт, форму импульса возбуждения и излучение в виде

$$H(z)=\frac{G}{1-\sum_{k=1}^{p}\alpha_k z^{-k}}.$$

Предположим, что все полюса лежат внутри единичной окружности. Используя (7.22), получим рекурсивное соотношение между комплексным кепстром, $\hat{h}(n)$ и коэффициентами $\{\alpha_k\}$ (как комплексный кепстр $1/H(z)$ связан с $\hat{h}(n)$?).

7.4. Рассмотрим минимально-фазовую последовательность $x(n)$ *конечной длины* с кепстром $\hat{x}(n)$ и последовательность $y(n)=\alpha^n x(n)$ с комплексным кепстром $\hat{y}(n)$.

а) Если $0<\alpha<1$, то как $\hat{y}(n)$ связан с $\hat{x}(n)$?

б) Как следует выбрать α, чтобы $y(n)$ уже не был минимально-фазовым?

в) Как следует выбрать α, чтобы $y(n)$ был максимально-фазовым?

7.5. Показать, что если $x(n)$ — минимально-фазовый, то $x(-n)$ — максимально-фазовый.

7.6. Рассмотрим последовательность $x(n)$ с комплексным кепстром $\hat{x}(n)$. Преобразование $\hat{x}(n)$ имеет вид

$$\hat{X}(z)=\log[X(z)]=\sum_{m=-\infty}^{\infty}\hat{x}(m)z^{-m},$$

где $X(z)$ — z-преобразование $x(n)$.

Преобразование $\hat{X}(z)$ дискретизировано в N равностоящих точках на единичной окружности:

$$\hat{X}_p(k) = \hat{X}\left(e^{j\frac{2\pi}{N}k}\right), \quad 0 \leqslant k \leqslant N-1.$$

Используя ДПФ, вычисляем

$$\hat{x}_p(n) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{X}_p(k)\, e^{j\frac{2\pi}{N}kn}, \quad 0 \leqslant n \leqslant N-1,$$

что может служить аппроксимацией комплексного кепстра.

а) Выразить $\hat{X}_p(k)$ через действительный кепстр $\hat{x}(m)$.

б) Подставить выражение п.а) в обратное преобразование Фурье для $\hat{x}_p(n)$ и показать, что $\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN)$.

7.7. Рассмотрим последовательность

$$x(n) = \delta(n) + \alpha\delta(n - N_p).$$

а) Определить комплексный кепстр. Изобразить ваш результат.

б) Изобразить кепстр $c(n)$ для $x(n)$.

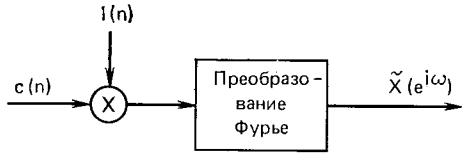


Рис. 3.7.1

в) Предположим, что по (7.30) вычислено приближение $x_p(n)$. Изобразить $x_p(n)$ для $0 \leqslant n \leqslant N-1$ в случае $N_p = N/6$. Что, если N не делится нацело на N_p?

г) Повторить п.в) для кепстральной аппроксимации $c_p(n)$ при $0 \leqslant n \leqslant N-1$ с использованием (7.33).

д) Если наибольшее значение кепстральной аппроксимации $c_p(n)$ используется для определения N_p, то как велико должно быть N для того, чтобы избежать ошибок?

7.8. Для сглаживания логарифма модуля спектра сигнала его кепстр часто взвешивают и преобразование Фурье имеет вид рис. 3.7.1.

а) Написать выражение, связывающее $X(e^{j\omega})$ с $\log|X(e^{j\omega})|$ и $L(e^{j\omega})$, где $L(e^{j\omega})$ — преобразование Фурье $l(n)$.

б) Какое кепстральное окно следует использовать для сглаживания функции $\log|X(e^{j\omega})|$?

в) Сравнить применение прямоугольного окна и окна Хемминга в качестве кепстральных окон.

г) Какова должна быть протяженность кепстрального окна и почему?

7.9. Рассмотрим сегмент вокализованного сигнала, который можно представить в виде $s(m) = p(m) * h_v(m)$, где $p(m) = \sum_{r=-\infty}^{\infty} \delta(m - rN_p)$. При вычислении комплексного кепстра (или кепстра) первый шаг заключается в умножении $s(m)$ на окно $w(m)$ для выделения сегмента $x_n(m) = s(m)w(n-m)$ входных данных для гомоморфной обработки.

а) Определить условия, при которых можно аппроксимировать $x_n(m)$ в виде $x(m) = p_n(m) * h_v(m)$, где $p_n(m) = p(m)w(n-m)$.

б) Для специального случая $n=0$ определить преобразование $p_0(m)$ через z-преобразование $w(m)$.

в) Выразить комплексный кепстр $\hat{p}_0(m)$ через $\hat{w}(m)$.

7.10. В задаче 7.9 показано, что периодичность взвешенного сегмента вокализованной речи может быть приближенно представлена выражением $p_n(m) = p(m)w(n-m)$, где $p(m) = \sum_{r=-\infty}^{\infty} \delta(m - rN_p)$. В этой задаче исследуется влияние

положения окна на комплексный кепстр $\hat{p}_n(m)$. Предположим, что имеется окно Хемминга вида

$$w(m) = \begin{cases} 0{,}54 - 0{,}46\cos(2\pi m/(2N_p)), & 0 \leqslant n \leqslant 2N_p; \\ 0, & \textit{в противном случае.} \end{cases}$$

а) Изобразить $p_n(m)$ как функцию m для $n = 3N_p/4;\ 9N_p/8;\ 5N_p/4;\ 3N_p/2$.

б) Для каждого из перечисленных выше случаев составить выражение для $p_n(m)$ и показать, что соответствующее z-преобразование имеет вид $P_n(z) = \alpha_1 z^{N_p} + \alpha_2 + \alpha_3 z^{-N_p}$.

в) Для каждого из перечисленных выше случаев определить и изобразить комплексный кепстр (указание: использовать разложение в ряд для $\log[P_n(z)]$). Опустить члены вида $\log[z^{\pm N_p}]$.

г) Для какого положения окна справедливы следующие утверждения:
последовательность $p_n(m)$ минимально-фазовая;
последовательность $p_n(m)$ максимально-фазовая;
первый кепстральный пик максимален;
первый кепстральный пик минимален.

д) Как изменятся ваши ответы на перечисленные выше вопросы, если окно удлинится? Укоротится?

7.11. Преобразование сигнала $x(n)$ определяется как

$$X(z) = \sum_{n=0}^{N-1} x(n)\, z^{-n}.$$

Вычислим $X(z)$ в последовательности точек $z_k = AW^{-k}k = 0, 1, \ldots, M-1$, где A и W — произвольные комплексные целые числа. Если сделать простую подстановку $nk = [n^2 + k^2 - (k-n)^2]/2$, то $X(z_k)$ можно записать в виде

$$X(z_k) = P(k) \sum_{n=0}^{N-1} y(n)\, g(k-n),$$

т. е. $X(z_k)$ — свертка $y(n)$ и $g(n)$.

а) Определить $P(k)$, $y(n)$ и $g(n)$ через $x(n)$, A и W.

б) Изобразить точки z_k на z-плоскости.

в) Можете ли Вы предложить способ применения БПФ для вычисления приведенного выше выражения?

8

Кодирование речевых сигналов на основе линейного предсказания

8.0. Введение

Линейное предсказание является одним из наиболее эффективных методов анализа речевого сигнала. Этот метод становится доминирующим при оценке основных параметров речевого сигнала, таких, как, например, период основного тона, форманты, спектр, функция площади речевого тракта, а также при сокращенном представлении речи с целью ее низкоскоростной передачи и экономного хранения. Важность метода обусловлена высокой точностью получаемых оценок и относительной простотой вычислений. В данной главе излагаются основные положения метода линейного предсказания и приводятся рекомендации по его практическому использованию.

Основной принцип метода линейного предсказания состоит в том, что текущий отсчет речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отсчетов. Коэффициенты предсказания при этом определяются однозначно минимизацией среднего квадрата разности между отсчетами речевого сигнала и их предсказанными значениями (на конечном интервале). Коэффициенты предсказания — это весовые коэффициенты, используемые в линейной комбинации.

Основные положения метода линейного предсказания хорошо согласуются с моделью речеобразования, рассмотренной в гл. 3, где показано, что речевой сигнал можно представить в виде сигнала на выходе линейной системы с переменными во времени параметрами, возбуждаемой квазипериодическими импульсами (в пределах вокализованного сегмента) или случайным шумом (на невокализованном сегменте). Метод линейного предсказания позволяет точно и надежно оценить параметры этой линейной системы с переменными коэффициентами.

Линейное предсказание уже обсуждалось в гл. 5 в связи с решением задачи квантования речи. Там предполагалось, что метод линейного предсказания можно применять для сокращения объема цифрового речевого сигнала.

Математические основы метода, используемого в адаптивном предсказателе высокого порядка при АРИКМ-кодировании, по существу совпадают с рассматриваемыми в этой главе. При АРИКМ-кодировании основная задача состоит в построении предсказателя, который минимизировал бы ошибку предсказания и, следовательно, шум квантования. В данной главе с более общих позиций будет показано, как основные идеи линейного предсказания приводят к ряду методов, которые можно использовать при оценке параметров речевых моделей.

Идеи и методы линейного предсказания довольно давно обсуждаются в технической литературе. Эти идеи используются в теориях автоматического управления и информации, где их называют методами оценивания систем, или методами идентификации систем. Под термином «идентификация» понимаются методы линейного предсказания (ЛП), основанные на оценивании параметров, однозначно описывающих систему при условии, что ее передаточная функция является полюсной. Применительно к обработке речевых сигналов методы линейного предсказания означают ряд сходных формулировок задачи моделирования речевого сигнала [1—18]. Эти формулировки часто отличаются в исходных предпосылках. Иногда они сводятся к различным методам вычисления, используемым для оценки коэффициентов предсказания. Так, применительно к речевым сигналам существуют следующие методы вычисления (часто равноценные): ковариационный [3], автокорреляционный [1, 2, 9], лестничного фильтра [11, 12], обратной фильтрации [1], оценки спектра [12], максимального правдоподобия [4, 6] и скалярного произведения [1]. В этой главе подробно рассматриваются сходства и различия трех первых из перечисленных выше методов, поскольку остальные подходы равноценны одному из этих трех.

Целесообразность использования линейного предсказания обусловлена высокой точностью описания речевого сигнала с помощью модели. Поэтому большая часть данной главы содержит методы оценивания различных параметров речевого сигнала с помощью линейного предсказания. Далее обсуждается ряд типичных примеров применения методов линейного предсказания, а в гл. 9 представлен ряд задач, для успешного решения которых также применяются методы линейного предсказания.

8.1. Методы анализа на основе линейного предсказания

В книге неоднократно использовалась основная модель речеобразования в дискретном времени, предложенная в гл. 3. На рис. 8.1 эта модель представлена в форме, наиболее удобной для решения задач линейного предсказания. В этом случае общий спектр, обусловленный излучением, речевым трактом и возбуждением, описывается с помощью линейной системы с переменными параметрами и передаточной функцией

$$H(z)=\frac{S(z)}{U(z)}=\frac{G}{1-\sum_{k=1}^{p}a_k z^{-k}}. \qquad (8.1)$$

Эта система возбуждается импульсной последовательностью для вокализованных звуков речи и шумом для невокализованных. Таким образом, модель имеет следующие параметры: классификатор вокализованных и невокализованных звуков, период основного тона для вокализованных сегментов, коэффициент усиления G и коэффициенты $\{a_k\}$ цифрового фильтра. Все эти параметры, разумеется, медленно изменяются во времени.

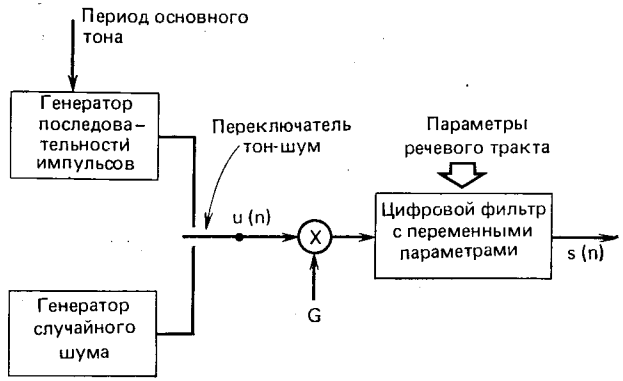


Рис. 8.1. Структурная схема упрощенной модели речеобразования

Определение периода основного тона и классификация тон/шум могут быть осуществлены на основе использования ряда методов, уже обсуждавшихся в данной книге, или с помощью рассматриваемых ниже методов линейного предсказания. Как отмечалось в гл. 3, для вокализованных звуков хорошо подходит модель, содержащая только полюсы в своей передаточной функции (чисто полюсная), но для носовых и фрикативных звуков требуется учитывать и нули. Однако из дальнейшего будет ясно, что если порядок p модели достаточно велик, то полюсная модель позволяет получить достаточно точное описание почти для всех звуков речи. Главное достоинство этой модели заключается в том, что как параметр G, так и коэффициенты можно оценить непосредственно с использованием очень эффективных с вычислительной точки зрения алгоритмов.

Для системы рис. 8.1 отсчет речевого сигнала $s(n)$ связан с сигналом возбуждения $u(n)$ простым разностным уравнением

$$s(n)=\sum_{k=1}^{p}a_k s(n-k)+Gu(n). \qquad (8.2)$$

Линейный предсказатель с коэффициентами α_k определяется как система, на выходе которой имеем

$$s(n)=\sum_{k=1}^{p}\alpha_k s(n-k). \tag{8.3}$$

Такие системы использовались в гл. 5 для уменьшения дисперсии погрешности предсказания. Системная функция предсказателя p-го порядка представляет собой полином вида

$$P(z)=\sum_{k=1}^{p}\alpha_k z^{-k}. \tag{8.4}$$

Погрешность предсказания определяется как

$$e(n)=s(n)-\tilde{s}(n)=s(n)-\sum_{k=1}^{p}\alpha_k s(n-k). \tag{8.5}$$

Из (8.5) видно, что погрешность предсказания представляет собой сигнал на выходе системы с передаточной функцией

$$A(z)=1-\sum_{k=1}^{p}\alpha_k z^{-k}. \tag{8.6}$$

Сравнение (8.2) и (8.5) показывает, что если сигнал точно удовлетворяет модели (8.2) и $\alpha_k=a_k$, то $e(n)=Gu(n)$. Таким образом, *фильтр погрешности предсказания* $A(z)$ является *обратным фильтром* для системы $H(z)$, соответствующей уравнению (8.1), т. е.

$$H(z)=G/A(z). \tag{8.7}$$

Основная задача анализа на основе линейного предсказания заключается в непосредственном определении параметров $\{\alpha_k\}$ по речевому сигналу с целью получения хороших оценок его спектральных свойств путем использования уравнения (8.7). Вследствие изменения свойств речевого сигнала во времени коэффициенты предсказания должны оцениваться на коротких сегментах речи. Основным подходом является определение параметров предсказания таким образом, чтобы минимизировать дисперсию погрешности на коротком сегменте сигнала. При этом *предполагается*, что полученные параметры являются параметрами системной функции $H(z)$ в модели речеобразования.

То, что подобный подход приводит к полезным результатам, возможно, не сразу очевидно, но его полезность будет неоднократно подтверждена различными способами. Во-первых, пусть $\alpha_k=a_k$, тогда $e(n)=Gu(n)$. Для вокализованной речи это означает, что $e(n)$ будет состоять из последовательности импульсов, т. е. $e(n)$ будет весьма мало́ почти все время. Поэтому в данном случае минимизация погрешности предсказания позволит получить требуемые коэффициенты. Другой повод, приводящий к тому же

подходу, вытекает из того, что даже если сигнал формируется системой (8.2) с постоянными во времени параметрами, которая возбуждается либо единичным импульсом, либо белым шумом, то можно показать, что коэффициенты предсказания, найденные по критерию минимизации среднего квадратического значения погрешности (в каждый момент времени), совпадают с коэффициентами в (8.2). Третьей, весьма важной для практики причиной является то, что подобная минимизация приводит к линейной системе уравнений, решение которых сравнительно легко приводит к получению параметров предсказания. Кроме того, полученные параметры, как это будет ясно из дальнейшего, составляют весьма плодотворную основу для точного описания сигнала.

Кратковременная энергия погрешности предсказания:

$$E_n=\sum_m e_n^2(m)= \tag{8.8}$$

$$=\sum_m (s_n(m)-\tilde{s}_n(m))^2= \tag{8.9}$$

$$=\sum_m\left[s_n(m)-\sum_{k=1}^{p}\alpha_k s_n(m-k)\right]^2, \tag{8.10}$$

где $s_n(m)$ — сегмент речевого сигнала, выбранный в окрестности отсчета n, т. е.

$$s_n(m)=s(m+n). \tag{8.11}$$

Пределы суммирования справа в (8.8)—(8.10) пока не определены, но поскольку предполагается использовать концепции кратковременного анализа, то эти пределы всегда предполагаются конечными. Кроме того, для получения среднего значения необходимо разделить полученный результат на длину речевого сегмента. Однако эти константы несущественны с точки зрения решения системы линейных уравнений и поэтому далее опускаются. Параметры α_k можно получить, минимизируя E_n в (8.10) путем вычисления $\frac{\partial E_n}{\partial \alpha_i}=0$, $i=1, 2, \ldots, p$, что приводит к системе уравнений

$$\sum_m s_n(m-i)\,s_n(m)=\sum_{k=1}^{p}\hat{\alpha}_k\sum_m s_n(m-i)\,s_n(m-k),\ 1\leqslant i\leqslant p, \tag{8.12}$$

где $\hat{\alpha}_k$ — значения α_k, минимизирующие E_n (поскольку значение α_k — единственное, далее знак $\wedge$ опускается и за величину, минимизирующую E_n, принимается α_k). Если ввести определение

$$\varphi_n(i,k)=\sum_m s_n(m-i)\,s_n(m-k), \tag{8.13}$$

тогда (8.12) можно переписать в более компактном виде:

$$\sum_{k=1}^{p} \alpha_k \varphi_n(i,k) = \varphi_n(i,0), i = 1, 2, \ldots p. \quad (8.14)$$

Эта система из p уравнений с p неизвестными может быть решена достаточно эффективным способом для получения неизвестных коэффициентов предсказания, минимизирующих средний квадрат погрешности предсказания на сегменте $s_n(m)$ [1]. Используя (8.10) и (8.12), можно показать, что средняя квадратическая погрешность предсказания имеет вид

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^{p} \alpha_k \sum_m s_n(m) s_n(m-k) \quad (8.15)$$

и, используя (8.14), можно выразить E_n в виде

$$E_n = \varphi_n(0,0) - \sum_{k=1}^{p} \alpha_k \varphi_n(0,k). \quad (8.16)$$

Таким образом, общая погрешность предсказания состоит из двух слагаемых, одно из которых является постоянным, а другое — зависит от коэффициентов предсказания.

Для решения системы уравнений относительно коэффициентов предсказания следует первоначально вычислить величины $\varphi_n(i,k)$, $1 \leqslant i \leqslant p$ и $0 \leqslant k \leqslant p$. Только после этого можно переходить к решению (8.14) и получению оценок α_k. Таким образом, принципиально анализ на основе линейного предсказания очень простой. Однако подробности, связанные с вычислением $\varphi_n(i,k)$ и последующим решением системы уравнений, являются достаточно запутанными и нуждаются в дальнейшем обсуждении.

Хотя пределы суммирования в (8.8)—(8.10) и (8.12) не определены, заметим, что в (8.12) они совпадают с соответствующими пределами в (8.8)—(8.10). Как было установлено, для кратковременного анализа соответствующие пределы должны охватывать конечный интервал. Имеется два подхода к этому вопросу, и, как это будет ясно из дальнейшего, в зависимости от пределов суммирования и выбора сегмента $s_n(m)$ различают два метода линейного предсказания.

8.1.1. Автокорреляционный метод [1, 2, 5]

Один из способов определения пределов в (8.8)—(8.10) и (8.12) основан на предположении, что сигнал равен нулю вне интервала $0 \leqslant m \leqslant N-1$. Это удобно записать в виде

$$s_n(m) = s(m+n) w(m), \quad (8.17)$$

[1] Очевидно, что α_k зависит от n (момент времени, для которого получена оценка), хотя эта зависимость не показана в явном виде. Далее индекс времени при E_n, $s_n(m)$ и $\varphi_n(i,k)$ опускается, если это не вызовет затруднений.

где $w(m)$ — окно конечной длительности (например, окно Хемминга), равное нулю вне интервала.

Значение этого предположения при решении вопроса о пределах суммирования в выражении для E_n можно рассмотреть на примере соотношения (8.5). Очевидно, что если $s_n(m)$ отличен от нуля только на интервале $0 \leqslant m \leqslant N-1$, то соответствующая погрешность предсказания $e_n(m)$ для предсказателя порядка p будет отлична от нуля на интервале $0 \leqslant m \leqslant N-1+p$. В этом случае E_n имеет вид

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m). \quad (8.18)$$

С другой стороны, легко показать, что пределы суммирования можно распространить на все ненулевые значения на интервале от $-\infty$ до $+\infty$ [2].

Возвращаясь к (8.5), можно отметить, что погрешность предсказания будет, вероятно, большой в начале интервала (т. е. $0 \leqslant m \leqslant p-1$), поскольку мы пытаемся предсказать сигнал по отсчетам, которые приравняли нулю. Очевидно, что погрешность будет большой и в конце интервала (т. е. $N \leqslant m \leqslant N+p-1$), поскольку здесь мы предсказываем нулевые значения по ненулевым. Поэтому в качестве окна $w(m)$ в уравнении (8.17) используется окно, которое стремится к нулю на концах интервала.

Пределы при вычислении $\varphi_n(i,k)$ в (8.13) совпадают с пределами (8.18). Но, поскольку $s_n(m)$ равно нулю вне интервала $0 \leqslant m \leqslant N-1$, легко показать, что

$$\varphi_n(i,k) = \sum_{m=0}^{N+p-1} s_n(m-i) s_n(m-k), 1 \leqslant i \leqslant p, 0 \leqslant k \leqslant p, \quad (8.19а)$$

можно выразить в виде

$$\varphi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k), 1 \leqslant i \leqslant p, 0 \leqslant k \leqslant p. \quad (8.19б)$$

Легко видеть, что в данном случае $\varphi_n(i,k)$ совпадает с кратковременной автокорреляционной функцией сигнала (4.30), вычисленной для $(i-k)$. Это означает, что

$$\varphi_n(i,k) = R_n(i-k), \quad (8.20)$$

где

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k). \quad (8.21)$$

Вычисление $R_n(k)$ детально обсуждалось в § 4.6 и здесь не рассматривается. Поскольку $R_n(k)$ — четная функция, то

$$\varphi_n(i,k) = R_n(|i-k|), i = 1, 2, \ldots, p, k = 0, 1, \ldots, p. \quad (8.22)$$

Таким образом, (8.14) можно представить в виде

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i)\ 1 \leqslant i \leqslant p. \tag{8.23}$$

Аналогично минимальный средний квадрат погрешности предсказания

$$E_n = R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k). \tag{8.24}$$

Систему уравнений (8.23) можно записать в матричной форме:

$$\begin{Vmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{Vmatrix} \begin{Vmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{Vmatrix} = \begin{Vmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ R_n(p) \end{Vmatrix}. \tag{8.25}$$

Матрица размером pxp является теплицевой, т. е. симметричной и такой, что элементы на любой диагонали равны между собой. Это особое свойство будет использовано в § 8.3 для получения эффективного алгоритма решения уравнений (8.23).

8.1.2. Ковариационный метод [3]

Другой основной подход к определению сегмента речевого сигнала и пределов суммирования заключается в том, что фиксируется интервал, на котором вычисляется средний квадрат погрешности, и рассматривается влияние этого обстоятельства на вычисление $\varphi_n(i, k)$. Другими словами, если определить

$$E_n = \sum_{m=0}^{N-1} e^2{}_n(m), \tag{8.26}$$

то $\varphi_n(i, k)$ выражается формулой

$$\varphi_n(i, k) = \sum_{m=0}^{N-1} s_n(m-i)\, s_n(m-k),\ 1 \leqslant i \leqslant p,\ 0 < k \leqslant p. \tag{8.27}$$

Изменив индекс суммирования, (8.27) можно выразить в виде

$$\varphi_n(i, k) = \sum_{m=-i}^{N-i-1} s_n(m)\, s_n(m+i-k),\ 1 \leqslant i \leqslant p,\ 0 < k < p, \tag{8.28а}$$

или

$$\varphi_n(i, k) = \sum_{m=-k}^{N-k-1} s_n(m)\, s_n(m+k-i),\ 1 \leqslant i \leqslant p,\ 0 < k \leqslant p. \tag{8.28б}$$

Полученные уравнения кажутся очень похожими на (8.19б), однако они имеют иные пределы суммирования. В (8.28) используется значение сигнала $s_n(m)$ вне интервала $0 \leqslant m \leqslant N-1$. Действительно, для вычисления $\varphi_n(i, k)$ для всех требуемых значений i и k необходимо использовать значения $s_n(m)$ на интервале $-p \leqslant m \leqslant N-1$. Для того чтобы это не противоречило пределам суммирования в (8.26), в данном случае используются необходимые значения сигнала без ограничения последовательности отсчетов окном конечной длительности, уменьшающимся к концам интервала, как это имело место в автокорреляционном методе. Таким образом, здесь используются отсчеты и вне интервала $0 \leqslant m \leqslant N-1$. Очевидно, что данный метод похож на метод вычисления модифицированной автокорреляционной функции (см. гл. 4). Как указывалось в § 4.6, это приводит не к автокорреляционной, а к взаимокорреляционной функции между двумя очень сходными, но не одинаковыми сегментами речевого сигнала конечной длительности. Хотя различие между (8.28) и (8.19) сводится к небольшим вычислительным подробностям, система уравнений

$$\sum_{k=1}^{p} \alpha_k \varphi_n(i, k) = \varphi_n(i, 0),\ i = 1, 2, \dots p, \tag{8.29а}$$

обладает свойствами, которые значительно влияют на метод решения и свойства получаемого оптимального предсказателя. В матричной форме система уравнений имеет вид

$$\begin{Vmatrix} \varphi_n(1,1) & \varphi_n(1,2) & \varphi_n(1,3) & \dots & \varphi_n(1,p) \\ \varphi_n(2,1) & \varphi_n(2,2) & \varphi_n(2,3) & \dots & \varphi_n(2,p) \\ \varphi_n(3,1) & \varphi_n(3,2) & \varphi_n(3,3) & \dots & \varphi_n(3,p) \\ \dots & \dots & \dots & \dots & \dots \\ \varphi_n(p,1) & \varphi_n(p,2) & \varphi_n(p,3) & \dots & \varphi_n(p,p) \end{Vmatrix} \begin{Vmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{Vmatrix} = \begin{Vmatrix} \varphi_n(1,0) \\ \varphi_n(2,0) \\ \varphi_n(3,0) \\ \dots \\ \varphi_n(p,0) \end{Vmatrix} \tag{8.29б}$$

В этом случае, поскольку $\varphi_n(i, k) = \varphi_n(k, i)$ [см. (8.28)], матрица размером pxp является квазикорреляционной симметричной, но не теплицевой. Действительно, можно сказать, что диагональные элементы связаны соотношением

$$\varphi_n(i+1, k+1) = \varphi_n(i, k) + s_n(-i-1)\, s_n(-k-1) - s_n(N-1-i)\, s_n(N-1-k). \tag{8.30}$$

Метод анализа, основанный на изложенном выше способе вычисления $\varphi_n(i, k)$, известен как *ковариационный метод*, поскольку матрица обладает свойствами ковариационной матрицы [5][1].

[1] Эта терминология отличается от общепринятой, поскольку термин «ковариация» обычно означает корреляцию сигнала, из которого предварительно вычитается среднее значение.

8.1.3. Заключение

Показано, что в зависимости от определения сегмента анализируемого сигнала можно получить две различные системы уравнений. Для автокорреляционного метода сигнал взвешивается с использованием N-точечного окна и величины $\varphi_n(i, k)$ получаются на основе кратковременной автокорреляционной функции. Полученная матрица корреляций является теплицевой и приводит к первой системе уравнений для параметров предсказания. При ковариационном методе сигнал предполагается известным на множестве значений $-p \leqslant n \leqslant N-1$. Никаких предположений о сигнале вне данного интервала не делается, поскольку только этот интервал необходим для вычислений. Полученная матрица корреляций в данном случае симметричная, но не теплицева. В результате два различных метода вычисления корреляции приводят к двум различным системам уравнений и к двум совокупностям коэффициентов предсказания с различными свойствами.

В последующих параграфах проводятся сравнение и противопоставление вычислительных процедур и результатов, даваемых обоими методами, наряду с рассматриваемыми ниже другими методами. Однако сначала рассмотрим определение коэффициента усиления G на рис. 8.1 на основе выражения для погрешности предсказания.

8.2. Вычисление коэффициента усиления модели [2]

Естественно ожидать, что коэффициент усиления G можно определить путем согласования энергии сигнала и линейно-предсказанных отсчетов. Это действительно верно, если сделать соответствующие предположения относительно сигнала возбуждения в модели с линейным предсказанием.

Постоянную G можно включить в сигнал возбуждения и ошибку предсказания (8.2) и (8.3)[1]. Сигнал возбуждения можно представить в виде

$$G\,u(n) = s(n) - \sum_{k=1}^{p} a_k\, s(n-k), \tag{8.31а}$$

при этом погрешность предсказания будет представлена в виде

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k\, s(n-k). \tag{8.31б}$$

В случае, когда $a_k = \alpha_k$, т. е. коэффициенты предсказания совпадают с коэффициентами модели,

$$e(n) = G\,u(n), \tag{8.32}$$

[1] Отметим, что усиление также является функцией времени.

т. е. входной сигнал модели пропорционален погрешности предсказания с коэффициентом пропорциональности G; подробнее свойства погрешности предсказания рассматриваются в § 8.5.

Поскольку (8.32) является приближенным (т. е. справедливо лишь при равенстве параметров модели и предсказателя), в общем случае определить G непосредственно по погрешности предсказания невозможно. Целесообразнее допустить, что энергия погрешности предсказания равна энергии сигнала возбуждения, т. е.

$$G^2 \sum_{m=0}^{N-1} u^2(m) = \sum_{m=0}^{N-1} e^2(m) = E_n. \tag{8.33}$$

В этом случае для определения G по каким-либо параметрам, например по a_k и коэффициентам корреляции, необходимы некоторые предположения относительно $u(n)$. Имеются два случая, для которых требуются соответствующие предположения. Для вокализованной речи естественно предположить, что $u(n) = \delta(n)$, т. е. возбуждение представляет собой единичный отсчет в нулевой момент времени[1]. Для справедливости этого предположения необходимо, чтобы линейный предсказатель с переменными во времени параметрами описывал как передаточную функцию речевого тракта, так и различие в форме между реальными импульсами возбуждения на вокализованном сегменте и единичными импульсами в модели. Для этого необходимо, чтобы порядок предсказателя был достаточным для описания как передаточной функции речевого тракта, так и эффекта возбуждения. Выбор порядка предсказателя будет рассмотрен в последующих разделах. Для невокализованных сегментов целесообразно предположить, что $u(n)$ представляет собой стационарный белый шум с нулевым средним и единичной дисперсией.

Используя эти предположения, определим G на основе соотношения (8.33). Для вокализованных сегментов на входе имеется сигнал $G\delta(n)$. Если обозначить сигнал на выходе для этого случая через $h(n)$ [поскольку в действительности это есть импульсная характеристика системы с передаточной функцией $H(z)$, как в (8.1)], получим соотношение

$$h(n) = \sum_{k=1}^{p} \alpha_k h(n-k) + G\delta(n). \tag{8.34}$$

Можно показать, что автокорреляционная функция для $h(n)$ (см. задачу 8.1), определяемая как

$$\widetilde{R}(m) = \sum_{n=0}^{\infty} h(n)\, h(m+n), \tag{8.35}$$

[1] Для справедливости этого предположения необходимо, чтобы интервал анализа совпадал с периодом основного тона.

удовлетворяет соотношениям

$$\widetilde{R}(m)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(|m-k|),\ m=1,2,\ \ldots,p, \tag{8.36а}$$

и

$$\widetilde{R}(0)=\sum_{k=0}^{p}\alpha_k\widetilde{R}(k)+G^2. \tag{8.36б}$$

Поскольку (8.36) совпадает с (8.32), следовательно,

$$\widetilde{R}(m)=R_n(m),\ 1\leqslant m\leqslant p. \tag{8.37}$$

Учитывая, что полные энергии сигнала $[R(0)]$ и импульсной реакции $[\tilde{R}(0)]$ должны быть равны, можно использовать (8.24), (8.33) и (8.34), чтобы получить

$$G^2=R_n(0)-\sum_{k=1}^{p}\alpha_k(R_n(k)=E_n. \tag{8.38}$$

Интересно, что требование равенства и (8.37) приводят к тому, что первые $p+1$ коэффициентов автокорреляции импульсной характеристики модели и сигнала должны совпадать.

В случае невокализованной речи корреляционная функция определяется статистическим усреднением. Предполагается, что сигнал возбуждения — белый шум с нулевым средним и единичной дисперсией, т. е.

$$E[u(n)\,u(n-m)]=\delta(m). \tag{8.39}$$

Если возбудить систему случайным процессом $Gu(n)$ и обозначить процесс на выходе через $g(n)$, то

$$g(n)=\sum_{k=1}^{p}\alpha_k g(n-k)+G\,u(n). \tag{8.40}$$

Если теперь $\tilde{R}(m)$ будет обозначать автокорреляционную функцию $g(n)$, то

$$\widetilde{R}(m)=E[g(n)\,g(n-m)]=\sum_{k=1}^{p}\alpha_k E[g(n-k)\,g(n-m)]+E[Gu(n)g\times$$

$$\times(n-m)]=\sum_{k=1}^{p}\alpha_k\widetilde{R}(m-k),\ m\neq 0, \tag{8.41}$$

поскольку $E[u(n)g(n-m)]=0$, $m>0$, вследствие некоррелированности предшествующим $u(n)$. Для $m=0$ получаем

$$\widetilde{R}(0)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(k)+GE[u(n)g(n)]=\sum_{k=1}^{p}\alpha_k\widetilde{R}(k)+G^2, \tag{8.42}$$

где $E[u(n)g(n)]=E[u(n)\,(Gu(n)+$ *члены, предшествующие* $n)]=$ $=G^2$. Поскольку энергия отклика на $Gu(n)$ должна совпадать с энергией сигнала, получаем

$$\widetilde{R}(m)=R_n(m),\ 0\leqslant m\leqslant p \tag{8.43}$$

или

$$G^2=R_n(0)-\sum_{k=1}^{p}\alpha_k R_n(k), \tag{8.44}$$

т. е. имеем то же самое, что и в случае импульсного возбуждения.

8.3. Решения уравнений линейного предсказания

Для эффективного использования метода линейного предсказания необходимо разработать эффективные алгоритмы решения системы линейных уравнений. Хотя можно использовать различные методы решения p уравнений с p неизвестными, все они оказываются различными по объему вычислений. Учитывая специальные свойства матрицы системы, решение можно получить значительно быстрее, чем в общем случае. В данном разделе подробно рассматриваются два метода получения параметров линейного предсказания и затем сравниваются и противопоставляются некоторые свойства полученных решений.

8.3.1. Решения на основе разложения Холецкого для ковариационного метода [3]

Система уравнений, решаемая при ковариационном методе, имеет вид

$$\sum_{k=1}^{p}\alpha_k\varphi_n(i,k)=\varphi_n(i,0),\ i=1,2,\ \ldots,p, \tag{8.45}$$

или в матричной форме

$$\boldsymbol{\Phi\alpha}=\boldsymbol{\psi}, \tag{8.46}$$

где $\boldsymbol{\Phi}$ — положительно определенная симметричная матрица из элементов $\varphi_n(i,j)$, а $\boldsymbol{\alpha}$ и $\boldsymbol{\psi}$ — вектор-столбцы с элементами α_i и $\varphi_n(i,0)$ соответственно. Решение системы уравнений (8.45) можно получить с использованием эффективного алгоритма с учетом симметрии и положительной определенности матрицы $\boldsymbol{\Phi}$. Получающийся при этом алгоритм называют методом Холецкого (или иногда методом квадратных корней) [3]. В этом методе матрица выражается в виде

$$\boldsymbol{\Phi}=\mathbf{VDV}^{\mathrm{t}}, \tag{8.47}$$

где $\mathbf{V}$ — нижняя треугольная матрица (элементы главной диагонали которой — единицы), а $\mathbf{D}$ — диагональная матрица. Верхний

индекс **t** означает транспонирование. Элементы матриц **V** и **D** легко определить из (8.47), приравнивая (i, j)-элементы для фиксированных (i, j) слева и справа, что позволяет получить

$$\varphi_n(i, j) = \sum_{k=1}^{j} V_{ik} d_k V_{jk}, \quad 1 \leqslant j \leqslant i-1, \tag{8.48}$$

или

$$V_{ij} d_j = \varphi_n(i, j) - \sum_{k=1}^{j-1} V_{ik} d_k V_{jk}, \quad 1 \leqslant j \leqslant i-1, \tag{8.49}$$

и для диагональных элементов

$$\varphi_n(i, i) = \sum_{k=1}^{i} V_{ik} d_k V_{ik} \tag{8.50}$$

или

$$d_i = \varphi_n(i,i) - \sum_{k=1}^{i-1} V^2_{ik} d_k, \quad i \geqslant 2, \tag{8.51}$$

если

$$d_1 = \varphi_n(1, 1). \tag{8.52}$$

Для иллюстрации использования соотношений (8.47)—(8.52) рассмотрим в качестве примера случай, когда $p=4$ и $\varphi_n(i, j) = \varphi_{ij}$. Уравнения (8.47) в этом случае имеют вид

$$\begin{Vmatrix} \varphi_{11} & \varphi_{21} & \varphi_{31} & \varphi_{41} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{41} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{43} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} \end{Vmatrix} = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{Vmatrix} \begin{Vmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{Vmatrix} \begin{Vmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Решение относительно d_1, d_4 и V_{ij} начнем с (8.52) при $i=1$, что дает $d_1=\varphi_{11}$. Используя (8.49) для $i=2, 3, 4$, получим V_{21}, V_{31} и V_{41} в виде $V_{21}d_1=\varphi_{21}$, $V_{31}d_1=\varphi_{31}$, $V_{41}d_1=\varphi_{41}$, $V_{21}=\varphi_{21}/d_1$, $V_{41}=\varphi_{41}/d_1$. Используя (8.51) для $i=2$, получаем $d_2=\varphi_{22}-V^2_{21}d_1$. Из (8.49) для $i=3$ и 4 имеем $V_{32}d_2=\varphi_{32}-V_{31}d_1V_{21}$, $V_{42}d_2=\varphi_{42}-V_{41}d_1V_{21}$ или $V_{32}=(\varphi_{32}-V_{31}d_1V_{21})/d_2$, $V_{42}=(\varphi_{42}-V_{41}d_1V_{21})/d_2$. Из (8.51) при $i=3$ получаем d_3, из (8.49) для $i=4$ находим V_{43} и, наконец, из (8.51) при $i=4$ определяем d_4. Получим матрицы **V** и **D**. Легко определить вектор-столбец $\boldsymbol{\alpha}$ с помощью двухшаговой процедуры. Из (8.46) и (8.47) находим

$$\mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{t}}\boldsymbol{\alpha} = \boldsymbol{\psi}, \tag{8.53}$$

что можно переписать в виде

$$\mathbf{V}\mathbf{Y} = \boldsymbol{\psi} \text{ и } \mathbf{D}\mathbf{V}^{\mathbf{t}}\boldsymbol{\alpha} = \mathbf{Y} \tag{8.54); (8.55}$$

или

$$\mathbf{V}^{\mathbf{t}}\boldsymbol{\alpha} = \mathbf{D}^{-1}\,\mathbf{Y}. \tag{8.56}$$

Таким образом, при известной матрице **V** уравнения (8.54) можно разрешить относительно вектор-столбца **Y**, используя простую рекурсивную процедуру

$$Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij} Y_j, \quad p \geqslant i \geqslant 2, \tag{8.57}$$

с начальным условием

$$Y_1 = \psi_1. \tag{8.58}$$

Аналогично, имея решение **Y**, можно из (8.56) получать α на основе рекурсивной процедуры вида

$$\alpha_i = Y_i/d_i - \sum_{j=i+1}^{p} V_{ji}\alpha_j, \quad 1 \leqslant i \leqslant p-1, \tag{8.59}$$

с начальными условиями

$$\alpha_p = Y_p/d_p. \tag{8.60}$$

Индекс i изменяется от $i=p-1$ до $i=1$ в порядке убывания.

Для иллюстрации использования алгоритма (8.57)—(8.60) продолжим рассмотрение примера. Сначала определим Y'_i, полагая, что **V** и **D** известны. В матричной форме имеем уравнения

$$\begin{Vmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{Vmatrix} \begin{Vmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{Vmatrix} = \begin{Vmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{Vmatrix}.$$

Из (8.57) и (8.58) получаем: $Y_1=\psi_1$, $Y_2=\psi_2-V_{21}Y_1$, $Y_3=\psi_3-V_{31}Y_{31}-V_{32}Y_2$, $Y_4=\psi_4-V_{41}Y_1-V_{42}Y_2-V_{43}Y_3$. Используя Y'_iS, решим уравнения (8.56) вида

$$\begin{Vmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{Vmatrix} \begin{Vmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{Vmatrix} = \begin{Vmatrix} 1/d_1 & 0 & 0 & 0 \\ 0 & 1/d_2 & 0 & 0 \\ 0 & 0 & 1/d_3 & 0 \\ 0 & 0 & 0 & 1/d_4 \end{Vmatrix} \begin{Vmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{Vmatrix} = \begin{Vmatrix} Y_1/d_1 \\ Y_2/d_2 \\ Y_3/d_3 \\ Y_4/d_4 \end{Vmatrix}.$$

Из (8.59) и (8.60) получаем: $\alpha_4=Y_4/d_4$, $\alpha_3=Y_3/d_3-V_{43}\alpha_4$, $\alpha_2=Y_2/d_2-V_{32}\alpha_3-V_{42}\alpha_4$, $\alpha_1=Y_1/d_1-V_{21}\alpha_2-V_{31}\alpha_3-V_{41}\alpha_4$, что и завершает решение ковариационных уравнений.

Используя разложение Холецкого, можно получить простое выражение для минимальной погрешности ковариационного метода через вектор **Y** и матрицу **D**. Напомним, что погрешность предсказания в ковариационном методе имеет вид

$$E_n = \varphi_n(0, 0) - \sum_{k=1}^{p} \alpha_k \varphi_n(0, k), \tag{8.61}$$

или в матричных обозначениях

$$E_n = \varphi_n(0, 0) - \boldsymbol{\alpha}^{\mathbf{t}}\boldsymbol{\psi}. \tag{8.62}$$

Подставляя вместо α^t его выражение из (8.56) в виде $\mathbf{Y}^t\mathbf{D}^{-1}\mathbf{V}^{-1}$, получим

$$E_n = \varphi_n(0, 0) - \mathbf{Y}^t \mathbf{D}^{-1} \mathbf{V}^{-1} \boldsymbol{\psi}. \tag{8.63}$$

Используя (8.54), имеем

$$E_n = \varphi_n(0, 0) - \mathbf{Y}^t \mathbf{D}^{-1} \mathbf{Y}, \tag{8.64}$$

или

$$E_n = \varphi_n(0, 0) - \sum_{k=1}^{p} Y_k^2/d_k. \tag{8.65}$$

Таким образом, минимальное значение среднего квадрата ошибки можно непосредственно определить через вектор $\mathbf{Y}$ и матрицу $\mathbf{D}$. Более того, (8.65) можно использовать для вычисления E_n при различных значениях p вплоть до значения, используемого при решении уравнений. Следовательно, можно проследить, как изменяется мощность погрешности предсказания при увеличении числа коэффициентов предсказания.

8.3.2. Алгоритм Дарбина для рекурсивного решения автокорреляционных уравнений [2]

Для автокорреляционного метода матричное уравнение относительно параметров предсказания имеет вид

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i), \quad 1 \leqslant i \leqslant p, \tag{8.66}$$

так как матрица системы имеет теплицеву форму, существует ряд специальных алгоритмов решения этой системы уравнений. Хотя наиболее известными и популярными являются методы Левинсона и Робинсона [1], эффективнее с точки зрения изложения является алгоритм Дарбина [2] (для простоты индекс у автокорреляционной функции опущен):

$$E^{(0)} = R(0); \tag{8.67}$$

$$k_i = [R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)]/E^{(i-1)}, \quad 1 \leqslant i \leqslant p; \tag{8.68}$$

$$\alpha_i^{(i)} = k_i; \tag{8.69}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}, \quad 1 \leqslant j \leqslant i-1; \tag{8.70}$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \tag{8.71}$$

Уравнения (8.67)—(8.71) решаются рекурсивно для $i = 1, 2, \ldots, p$, и окончательное решение принимает вид

$$\alpha_j = \alpha_j^{(p)}, \quad 1 \leqslant j \leqslant p. \tag{8.72}$$

Отметим, что при вычислении параметров предсказания для модели порядка p получаются и все параметры для предсказания меньших порядков, т. е. $\alpha_j^{(i)}$ — это j-й параметр предсказателя порядка i.

Для иллюстрации применения рассмотренного выше алгоритма вычислим параметры предсказателя второго порядка. Исходные матричные уравнения имеют вид

$$\begin{Vmatrix} R(0) & R(1) \\ R(1) & R(0) \end{Vmatrix} \begin{Vmatrix} \alpha_1 \\ \alpha_2 \end{Vmatrix} = \begin{Vmatrix} R(1) \\ R(2) \end{Vmatrix}.$$

Используя приведенный алгоритм, имеем:

$$E^{(0)} = R(0); k_1 = R(1)/R(0); \alpha_1^{(1)} = R(1)/R(0);$$

$$E^{(1)} = \frac{R^2(0) - R^2(1)}{R(0)}; k_2 = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)};$$

$$\alpha_2^{(2)} = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)}; \alpha_1^{(2)} = \frac{R(1)R(0) - R(1)R(2)}{R^2(0) - R^2(1)};$$

$$\alpha_1 = \alpha_1^{(2)}; \alpha_2 = \alpha_2^{(2)}.$$

Величины $E^{(i)}$ в (8.71) представляют собой мощности погрешностей предсказания для предсказателя порядка i. Таким образом, на каждом шаге вычислений можно контролировать мощность погрешности. Коэффициенты корреляции можно заменить нормированными коэффициентами $r(i)$, что не изменит решения уравнения, но при этом величины $E^{(i)}$ следует интерпретировать как нормализованную погрешность. Если обозначить ее через $V^{(i)}$, то

$$V^{(i)} = \frac{E^{(i)}}{R(0)} = 1 - \sum_{k=1}^{i} \alpha_k r(k), \tag{8.73}$$

где

$$0 < V^{(i)} \leqslant 1, i \geqslant 0. \tag{8.74}$$

Можно показать, что нормированная погрешность при $i = p$ может быть представлена в виде

$$V^{(p)} = \prod_{i=1}^{p} (1 - k_i^2), \tag{8.75}$$

где величины k_i удовлетворяют условию

$$-1 \leqslant k_i \leqslant 1. \tag{8.76}$$

Это ограничение параметров k_i весьма важно, ибо [1, 18] оно является необходимым и достаточным условием того, чтобы все корни полинома $A(z)$ лежали внутри единичной окружности. Это гарантирует устойчивость системы $H(z)$. К сожалению, доказательство этого факта увело бы нас далеко в сторону, но его отсутствие в данной книге не снижает важности приведенного результата. Более того, можно показать, что в ковариационном методе нет таких условий устойчивости.

8.3.3. Постановка задачи и ее решение на основе лестничного фильтра [11]

Как было показано выше, оба метода вычисления параметров предсказания включают в себя два этапа: оценивание матрицы корреляций и решение системы линейных уравнений. Эти методы широко и успешно используются применительно к обработке речевых сигналов. Вместе с тем к настоящему времени развит другой класс методов, называемых методами на основе лестничного фильтра, в которых оба этапа в известном смысле объединены в один рекурсивный алгоритм оценивания параметров линейного предсказания. Для того чтобы проследить связь между этими методами, полезно начать с алгоритма Дарбина. Прежде всего отметим, что коэффициенты предсказания на i-й интерации являются параметрами предсказателя i-го порядка. Используя эти коэффициенты, можно определить

$$A^{(i)}(z) = 1 - \sum_{k=1}^{i} \alpha_k^{(i)} z^{-k} \tag{8.77}$$

как передаточную функцию обратного фильтра порядка i (или фильтр погрешности предсказания). Если на входе этого фильтра имеется сигнал $s_n(m) = s(n+m)w(m)$, то на его выходе погрешность предсказания $e_n^{(i)}(m) = e^{(i)}(n+m)$, где

$$e^{(i)}(m) = s(m) - \sum_{k=1}^{i} \alpha_k^{(i)} s(m-k). \tag{8.78}$$

Отметим, что далее для простоты опускается индекс n, который означает, что рассматривается сегмент сигнала, расположенный на n-м отсчете. Используя z-преобразование, запишем (8.78) в виде

$$E^{(i)}(z) = A^{(i)}(z)\, s(z). \tag{8.79}$$

Подставляя (8.70) в (8.77), получаем рекурсивную формулу для $A^{(i)}(z)$ через $A^{(i-1)}(z)$, т. е.

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \tag{8.80}$$

(см. задачу 8.5). Подставляя (8.80) в (8.79), получаем

$$E^{(i)}(z) = A^{(i-1)}(z)\, s(z) - k_i z^{-i} A^{(i-1)}(z^{-1})\, s(z). \tag{8.81}$$

Первый член в (8.81), очевидно, является z-преобразованием погрешности предсказания для предсказателя $(i-1)$-го порядка. Второй член можно интерпретировать аналогичным образом, если ввести обозначения

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1})\, s(z). \tag{8.82}$$

Обратное преобразование от $B^{(i)}(z)$ есть

$$b^{(i)}(m) = s(m-i) - \sum_{k=1}^{i} \alpha_k^{(i)} s(m+k-i). \tag{8.83}$$

Это уравнение предполагает, что мы предсказываем $s(m-i)$ по i отсчетам входного сигнала $\{s(m-i+k),\ k=1, 2, \ldots, i\}$, которые следуют за $s(m-i)$. Таким образом, можно сказать, что $b^{(i)}(m)$ — погрешность возвратного предсказания. На рис. 8.2 показано, что

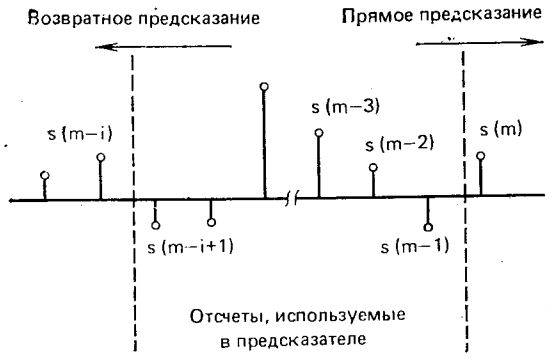


Рис. 8.2. Иллюстрация прямой и возвратной погрешностей для предсказателя i-го порядка

i отсчетов, используемых для предсказания, в данном случае совпадают с отсчетами, которые применяются для предсказания текущего значения (8.78). Возвращаясь к (8.81), видим, что погрешность предсказания может быть представлена в виде

$$e^{(i)}(m) = e^{(i-1)}(m) - k_i\, b^{(i-1)}(m-1). \tag{8.84}$$

Подставляя (8.80) в (8.82), получим

$$B^{(i)}(z) = z^{-i} A^{(i-1)}(z^{-1})\, s(z) - k_i A^{(i-1)}(z)\, s(z), \tag{8.85}$$

или

$$B^{(i)}(z) = z^{-1} B^{(i-1)}(z) - k_i E^{(i-1)}(z). \tag{8.86}$$

Таким образом, i-я возвратная погрешность

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i\, e^{(i-1)}(m). \tag{8.87}$$

Уравнения (8.84) и (8.85) определяют погрешность прямого и возвратного предсказаний для предсказателя порядка i через соответствующие погрешности для предсказателя порядка $(i-1)$. Использование предсказателя нулевого порядка эквивалентно отсутствию предсказания вообще, т. е.

$$e^{(0)}(m) = b^{(0)}(m) = s(m), \tag{8.88}$$

Таким образом, уравнения (8.84) и (8.85) можно представить в виде структурной схемы (рис. 8.3). Такая схема называется лестничной. Очевидно, используя p каскадов лестничного фильтра, на выходе последнего из них можно получить погрешность предсказания, как это изображено на рис. 8.3. Таким образом, на рис. 8.3 представлена цифровая реализация фильтра погрешности предсказания с передаточной функцией $A(z)$.

Полученная схема является непосредственным следствием алгоритма Дарбина, и параметры k_i можно вычислять с использованием уравнений (8.67)—(8.72). Параметры предсказания как таковые на рис. 8.3 отсутствуют. Итакура [4, 6] показал, что k_i

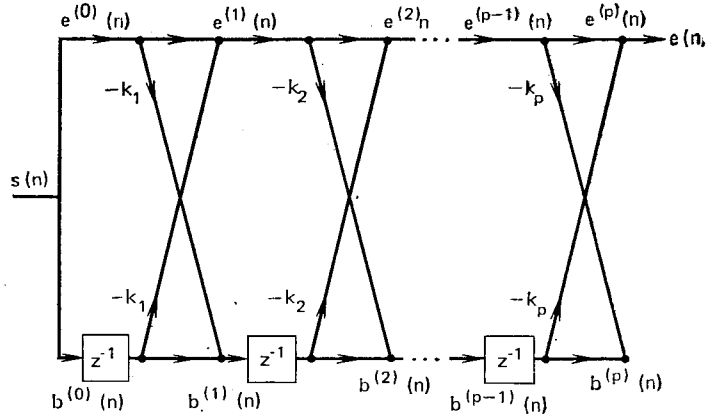


Рис. 8.3. Структурная схема реализации лестничного фильтра

можно получить непосредственно по погрешностям предсказания и в силу особенностей лестничной схемы совокупность коэффициентов k_i, $i=1, 2, ..., p$, можно получить без вычисления коэффициентов предсказания в соответствии с соотношением

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m)\, b^{(i-1)}(m-1)}{\left\{\sum_{m=0}^{N-1} (e^{(i-1)}(m))^2 \sum_{m=0}^{N-1} (b^{(i-1)}(m-1))^2\right\}^{1/2}}. \qquad (8.89)$$

Это выражение — разновидность нормированной взаимно-корреляционной функции, т. е. показатель корреляции между погрешностью предсказания и возвратной погрешностью. Поэтому параметры k_i называют частными корреляциями [4, 6]. Уравнение (8.89) эквивалентно (8.68) после подстановки (8.78) и (8.83) в (8.89).

Заменив (8.68) на (8.89) в алгоритме Дарбина, можно, как и ранее, получить параметры предсказания. Таким образом, анализ на основе частных корреляций приводит к несколько иному подходу, чем при обращении матриц, и дает результаты, совпадающие с автокорреляционным методом, т. е. совокупность параметров частных корреляций эквивалентна совокупности параметров предсказания, минимизирующих средний квадрат погрешности предсказания. Этот метод открывает новый класс методов, основанных на лестничном фильтре (рис. 8.3) [11].

Аналогично Бург [12] разработал алгоритм, основанный на минимизации суммы среднего квадрата прямой и возвратной погрешностей на рис. 8.3, т. е.

$$\widetilde{E}^{(i)} = \sum_{m=0}^{N-1} [(e^{(i)}(m))^2 + (b^{(i)}(m))^2]. \qquad (8.90)$$

Подставляя (8.84) и (8.87) в (8.90) и дифференцируя $\widetilde{E}^{(i)}$ по k_i, получаем

$$\frac{\partial \widetilde{E}^{(i)}}{\partial k_i} = -2 \sum_{m=0}^{N-1} [e^{(i-1)}(m) - k_i\, b^{(i-1)}(m-1)]\, b^{(i-1)}(m-1) -$$
$$-2 \sum_{m=0}^{N-1} [b^{(i-1)}(m-1) - k_i\, e^{(i-1)}(m)]\, e^{(i-1)}(m). \qquad (8.91)$$

Приравнивая производную нулю, находим

$$k_i = \frac{2 \sum_{m=0}^{N-1} [e^{(i-1)}(m)\, b^{(i-1)}(m-1)]}{\sum_{m=0}^{N-1} [e^{(i-1)}(m)]^2 + \sum_{m=0}^{N-1} [b^{(i-1)}(m-1)]^2}. \qquad (8.92)$$

Можно показать, что оценки k_i на основе (8.92) удовлетворяют соотношению

$$-1 \leqslant k_i \leqslant 1. \qquad (8.93)$$

Следует, однако, иметь в виду, что оценки k_i в соответствии с (8.92) отличаются от оценок (8.89) или, что то же самое, от автокорреляционного метода.

Отметим в заключение, что вычисление коэффициентов предсказания и параметров k включает в себя следующие шаги:

1. Начальные условия $e^{(0)}(m) = s(m) = b^{(0)}(m)$.
2. Вычислить $k_1 = \alpha_1^{(1)}$ из (8.92).
3. Определить прямую и возвратную погрешности предсказания $e^{(1)}(m)$ и $b^{(1)}(m)$ по (8.84) и (8.87).
4. Установить $i=2$.
5. Определить $k_i = \alpha_i^{(i)}$ из (8.92).
6. Определить $\alpha_j^{(i)}$ для $j=1, 2, ..., i-1$ из (8.70).
7. Определить $e^{(i)}(m)$ и $b^{(i)}(m)$ из (8.84) и (8.87).
8. Установить $i=i+1$.
9. Если i меньше или равно p, идти к 5.
10. Алгоритм закончен.

Между изложенным методом, а также автокорреляционным и ковариационным методами существует ряд различий. Основное из них состоит в том, что в лестничном методе коэффициенты предсказания оцениваются непосредственно по речевому сигналу без промежуточного вычисления автокорреляционной функции.

Кроме того, метод гарантирует получение устойчивого фильтра без использования окон. По этим причинам подход на основе лестничного фильтра является важным и предпочтительным способом реализации линейного предсказания.

8.4. Сравнение методов решения уравнений линейного предсказания

Выше уже обсуждались различия между ковариационным, корреляционным подходами и методом на основе лестничного фильтра. Здесь рассматриваются вопросы практического использования полученных уравнений. Эти вопросы включают в себя вычислительную сторону задачи, цифровую и физическую устойчивость получающихся решений и вопросы выбора количества полюсов или каскадов модели. Начнем с вычислительных аспектов задачи, связанных с получением параметров предсказания по речевому сигналу.

Два основных вопроса при вычислении коэффициентов предсказания состоят в объеме памяти и количестве умножения. В табл. 8.1 (по результатам Портнова и др. [13] и Макхоула

Таблица 8.1

Объем вычислений при решении уравнений линейного предсказания

	Ковариационный метод	Автокорреляционный метод	Лестничный метод
	Разложение Холецкого	Метод Дарбина	Метод Бурга
Память:			
данные	N_1	N_2	$3N_3$
матрица	Пропорционально $p^2/2$	Пропорционально p	—
окно	0	N_2	—
Вычисления (умножения):			
взвешивание	0	N_2	—
корреляция	Пропорционально N_1p	Пропорционально N_2p	—
матричное решение	Пропорционально p^3	Пропорционально p^2	$5N_3p$

[11]) содержится требуемый объем вычислений для трех рассмотренных методов. С точки зрения памяти, ковариационный метод требует наличия N_1 ячеек для данных и $p^2/2$ ячеек для ковариационной матрицы. Для автокорреляционного метода требуется N_2 ячеек памяти как для данных, так и для окна из p ячеек для хранения автокорреляционной матрицы. Для лестничного метода необходимо наличие $3N_3$ ячеек памяти для исходных данных и двух погрешностей предсказания. Для того чтобы подчеркнуть

различие в объеме исходных данных, в случае ковариационного метода этот объем обозначен через N_1, автокорреляционного — N_2, лестничного — N_3. Вопрос выбора этого объема будет обсуждаться далее. А сейчас, полагая, что N_1, N_2 и N_3 сравнимы, получаем, что автокорреляционный и ковариационный методы требуют меньше памяти, чем лестничный метод.

Количество умножений, необходимое для реализации каждого метода, показано в нижней части табл. 8.1. Для ковариационного метода вычисление матрицы требует около N_1p умножений, а решение матричного уравнения (с использованием разложения Холецкого) осуществляется с использованием p^3 умножений [Портнов и другие дали точное значение: $(p^3+9p^2+2p)/6$ операций умножения, p — деления и p — извлечения квадратного корня]. Для автокорреляционного метода вычисление корреляционной матрицы требует проведения N_2p умножений, а решение автокорреляционных уравнений — около p^2 умножений. Таким образом, если $N_1 \approx N_2$ и $N_1 \gg p$, $N_2 \gg p$, то автокорреляционный метод требует меньше вычислений, чем ковариационный. Однако в большинстве случаев при обработке речевых сигналов количество операций умножения для вычисления автокорреляционной функции значительно превосходит объем вычислений при решении уравнения, поэтому время вычислений в обоих случаях примерно одинаково. Для вычисления частных корреляций в лестничном методе необходимо осуществить $5N_3p$ умножений[1]. Таким образом с точки зрения эффективности вычислительной процедуры лестничный метод наименее эффективен. Однако при решении вопроса о его использовании следует иметь в виду ряд преимуществ этого метода по сравнению с другими.

Другой аспект сопоставления различных подходов связан с устойчивостью полученной системы:

$$H(z) = G/A(z). \tag{8.94}$$

Этот фильтр устойчив, если все его полюсы лежат строго внутри единичной окружности. Полюсы фильтра $H(z)$ совпадают с нулями полинома знаменателя $A(z)$, т. е.

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}. \tag{8.95}$$

Как утверждалось выше, для автокорреляционного метода все корни лежат внутри единичной окружности, т. е. устойчивость $H(z)$ гарантирована. Следует отметить, что эти теоретические гарантии могут нарушаться на практике, если автокорреляционная функция вычисляется с недостаточной точностью. Встречающиеся в этом случае округления при вычислении автокорреляционной функции могут привести к плохой обусловленности автокорреляционной матрицы. Маркел и Грей показали, что этот эф-

[1] Макхоул [11] предложил метод вычисления частных корреляций с той же эффективностью, что и при ковариационном методе.

фект можно устранить применением предыскажений, приводящих к выравниванию спектра [1]. Использование предыскажений позволяет уменьшить разрядность вычислителя при сохранении устойчивости предсказателя. Алгоритм Дарбина позволяет легко проверить устойчивость, поскольку параметры k_i (частные корреляции) должны удовлетворять условию

$$-1 \leqslant k_i \leqslant 1. \qquad (8.96)$$

Таким образом, если при вычислении коэффициентов предсказания любая из величин k_i не удовлетворяет (8.96), то, как известно, корни лежат вне единичной окружности.

Для ковариационного метода невозможно гарантировать устойчивость фильтра-предсказателя. Однако на практике при достаточно большом числе отсчетов на интервале оценивания получаемый предсказатель почти всегда устойчив. Это объясняется тем, что при большом числе отсчетов на интервале оценивания ковариационный и автокорреляционный методы дают почти одинаковый результат.

Для лестничного метода устойчивость гарантирована, ибо коэффициенты предсказания получаются по коэффициентам частных корреляций, которые по определению удовлетворяют (8.96). Кроме того, устойчивость сохраняется и при вычислениях с использованием вычислителя с конечной разрядной сеткой [1]. При возникновении сомнений в устойчивости предсказателя необходимо определить корни полинома предсказателя. Если обнаружено, что корни расположены вне единичной окружности, то простая процедура, состоящая в отражении корней внутрь окружности, позволяет получить устойчивый полином с той же частотной характеристикой, что и неустойчивый.

Два других аспекта проблемы решения уравнений линейного предсказания сводятся к выбору порядка предсказателя p и длины интервала анализа N. Выбор p определяется частотой дискретизации и не зависит от используемого метода. Поскольку подлежащий анализу речевой спектр характеризуется в общем случае вкладом голосового тракта со средней плотностью примерно два полюса (или один комплексный полюс) на каждый килогерц, то для представления полного вклада в речевой спектр необходимо наличие F_s полюсов, где F_s — частота дискретизации в килогерцах. Таким образом, при дискретизации с частотой 10 кГц необходимо десять полюсов для представления голосового тракта. Кроме того, для описания источника возбуждения и излучения губами необходимо еще три-четыре полюса. Следовательно, общее число полюсов при частоте дискретизации 10 кГц составляет 13—14. Для подтверждения этого факта на рис. 8.4 представлена зависимость нормированной погрешности предсказания от порядка предсказателя p для вокализованной и невокализованной речей при частоте дискретизации 10 кГц. Хотя при возрастании p погрешность уменьшается, можно заметить, что уже при $p \approx 13 \div 14$ погрешность изменяется незначительно. Интересно, что

для невокализованного сигнала погрешность значительно больше, чем для вокализованного. Это, конечно, не является чем-то неожиданным, так как модель для невокализованного сигнала менее точна, чем для вокализованного. Дополнительное экспериментальное исследование поведения погрешности предсказания при различных p дано в последующих параграфах.

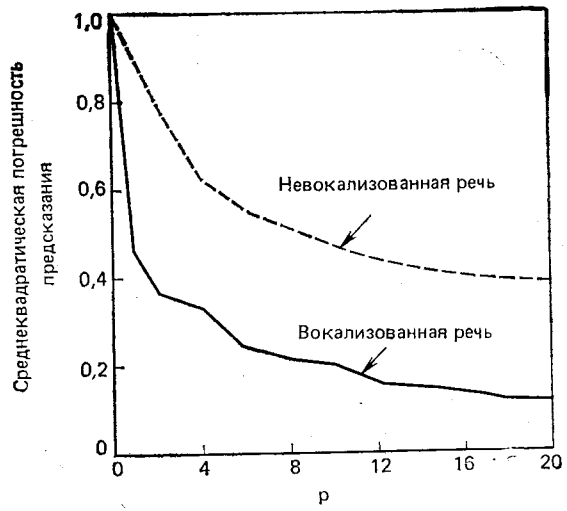


Рис. 8.4. Среднеквадратическая погрешность предсказания в зависимости от числа коэффициентов p [3]

Выбор длины интервала анализа N является чрезвычайно важным при использовании линейного предсказания. Выбирать следует как можно меньше, поскольку полный объем вычислений непосредственно связан с N. Для автокорреляционного метода показано, что для достижения хороших результатов значения N должны составлять несколько периодов основного тона [1, 2]. Поскольку при использовании автокорреляционного метода применяется временно́е окно, то для того чтобы эффект спадания окна на границах интервала не влиял на результаты, этот интервал должен быть достаточно большим. Так, в большинстве систем анализа на основе линейного предсказания используются окна от 100 до 400 отсчетов (при частоте дискретизации 10 кГц), причем предпочтение отдается более длинным окнам. Как для ковариационного, так и для лестничного методов выбор протяженности окна зависит от ряда условий. Поскольку взвешивание данных в этих случаях не требуется, нет и ограничений снизу на протяженность окна. Если анализ можно ограничить интервалом между импульсами основного тона (синхронно с ним), то достаточно, чтобы значение N составляло около $2p$. Но при использовании столь короткого интервала можно получить неудовлетворитель-

ные результаты, как только импульс основного тона появится внутри интервала анализа. Поэтому в большинстве практических систем, в которых нельзя организовать анализ синхронно с основным тоном, используются интервалы того же порядка, что и в корреляционном методе. В последующих параграфах будут представлены результаты экспериментальной оценки влияния длины и положения интервала анализа на погрешность предсказания для ковариационного и автокорреляционного методов[1]. Однако сначала коротко рассмотрим свойства погрешности предсказания и нормированной погрешности, получаемой на ее основе.

8.5. Погрешность предсказания

В результате анализа сигнала с помощью линейного предсказания возникает погрешность предсказания, определяемая как (8.97):

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) = Gu(n). \qquad (8.97)$$

Если речевой сигнал действительно порождается моделью линейного предсказания порядка p с переменными во времени параметрами, то $e(n)$ должна представлять собой хорошее приближение для сигнала на выходе источника возбуждения. Основываясь на этом обстоятельстве, можно ожидать, что для вокализо-

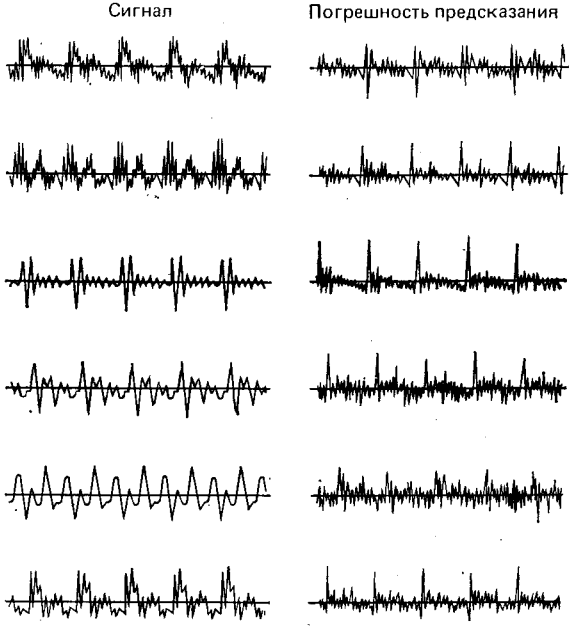


Рис. 8.5. Примеры сигналов и погрешностей предсказания для гласных (*i*, *e*, *a*, *o*, *u*, *y*) [14]

[1] Исследование Рабинера и др. [16] показали, что выбор параметров для лестничного и ковариационного методов совпадают. Поэтому в данном параграфе различия между ними не делается.

ванных сегментов речевого сигнала погрешность предсказания должна быть большой в начале каждого периода основного тона. Таким образом, период основного тона можно определить, оценивая координаты достаточно больших отсчетов и определяя период как разность координат во времени двух соседних отсчетов погрешности, превысивших соответствующий порог. С другой стороны, период основного тона можно определить на основе корреляционного анализа погрешности предсказания путем обнаружения максимального пика в подходящем диапазоне задержек. Другим объяснением того, что погрешность предсказания удобна для оценивания периода основного тона, является тот факт, что спектральная плотность погрешности практически равномерна во всей полосе частот, что обусловлено устранением из нее формант.

Чтобы показать особенности сигнала погрешности предсказания, на рис. 8.5 представлен ряд фрагментов гласных звуков соответствующих фрагментов погрешности (Страубе [14]). Для всех этих фрагментов гласных звуков в сигнале погрешности явно видны импульсы на интервалах, соответствующих периоду основного тона. На рис. 8.6—8.9 представлен ряд других экспериментов с погрешностью предсказания. На всех рисунках в части *а*) показан сегмент обрабатываемого речевого сигнала, в части *б*) — погрешность предсказания, в части

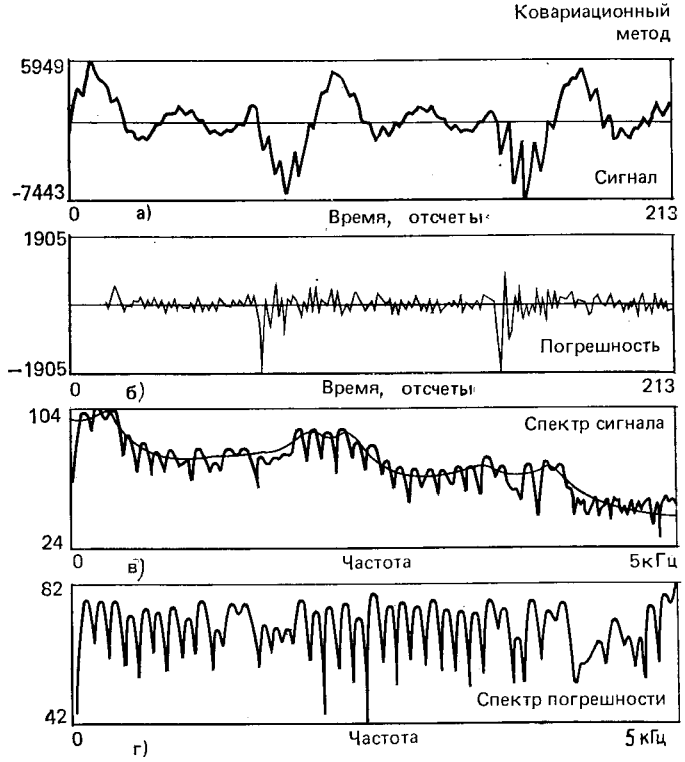


Рис. 8.6. Типичные сигналы и спектры ковариационного метода линейного предсказания для мужского голоса [16]

в) — логарифм дискретного преобразования Фурье сигнала из *а*) (вычисленный с использованием БПФ) совместно с логарифмом $H(e^{i\omega T})$ в качестве огибающей, в части *д*) представлен логарифм спектральной плотности погрешности предсказания (рассчитанный на основе БПФ). Рисунки 8.6 и 8.7 содержат результаты обработки гласного звука *i* (как в слове we), произнесенного муж-

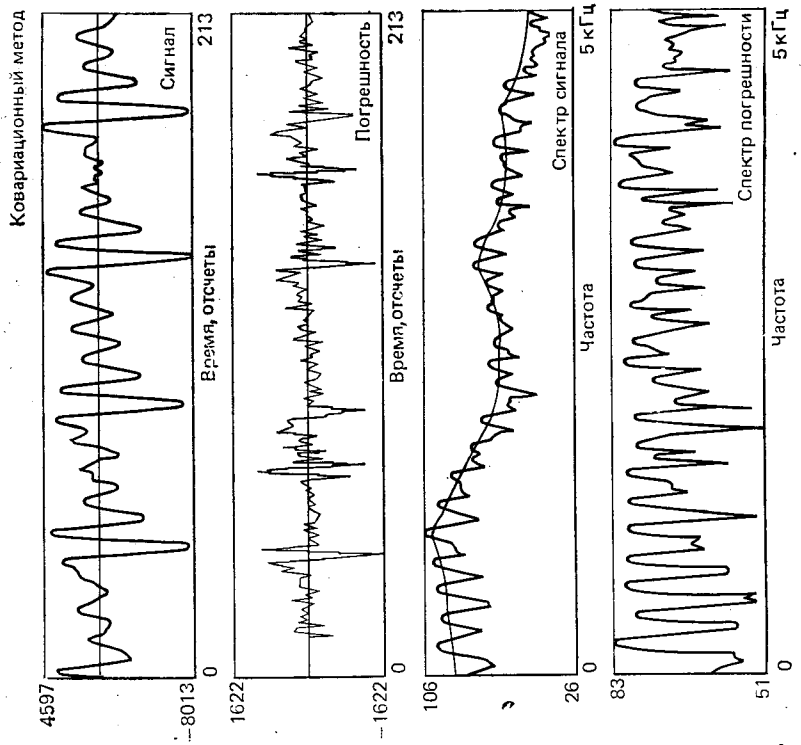


Рис. 8.8. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для женского голоса [16]

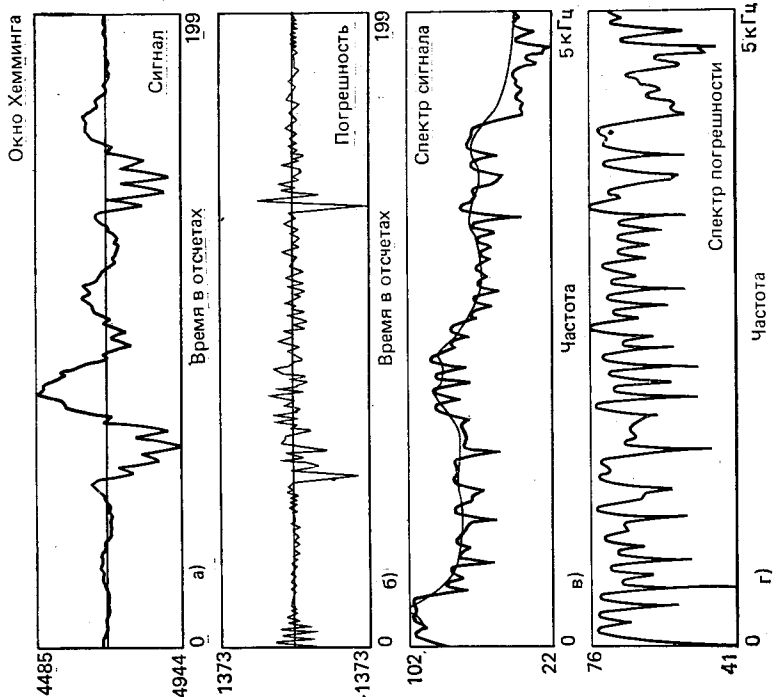


Рис. 8.7. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для мужского голоса [16]

ским голосом, с использованием ковариационного и автокорреляционного методов соответственно (с окном Хемминга). Продолжительность интервала анализа составляла 20 мс. Легко видеть, что в погрешности предсказания имеются пики точно в начале каждого периода основного тона, а спектральная плотность достаточно равномерна, хотя в ней просматривается линейчатая структура, возникающая за счет основного тона. Отметим слишком большую погрешность предсказания в начале сегмента на рис. 8.7 при использовании автокорреляционного

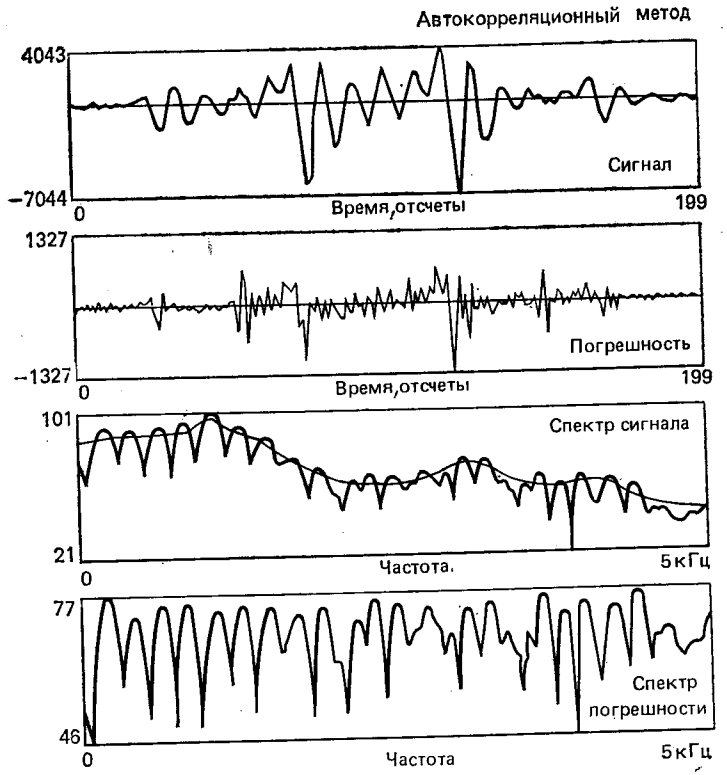


Рис. 8.9. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для женского голоса [16]

метода. Это, конечно, обусловлено попыткой предсказать отсчеты сигнала по нулевым значениям вне интервала $0 \leqslant m \leqslant 199$. Скорость убывания окна Хемминга в данном случае оказалась недостаточной для эффективного устранения этой ошибки.

На рис. 8.8 и 8.9 показаны аналогичные результаты, полученные на гласном *a* (как в слове father), произнесенном женским голосом. У этого диктора в интервал анализа попадает около пяти периодов основного тона. Таким образом, на рис. 8.8 в сигнале погрешности наблюдается большое количество острых пиков в начале каждого периода основного тона при ковариационном методе анализа. Однако использование окна Хемминга в автокорреляционном методе привело к тому, что пики в погрешности предсказания уменьшились вследствие убывания функции окна к концу интервала.

Поведение сигналов погрешности предсказания, представленного на предыдущих рисунках, позволяет рассчитывать, что они являются как раз теми сигналами, по которым наиболее просто оценить период основного тона. В [5] пока-

зано, что для звуков, не имеющих явно выраженной гармонической структуры, например плавных, как r, l, или носовых, как m, n, пики в погрешности предсказания выражены не столь отчетливо. Кроме того, на переходах между вокализованными и невокализованными звуками выбросы за счет основного тона в погрешности часто просто не проявляются.

Короче говоря, хотя сигнал погрешности предсказания $e(n)$ кажется очень подходящим для построения на его основе детектора основного тона, однако имеется ряд специфических трудностей в определении положения импульса для широкого класса гласных звуков и, таким образом, сигнал погрешности предсказания не следует считать чем-то исключительным при решении указанной задачи. В 8.10.1 рассмотрен один метод определения основного тона по погрешности предсказания.

8.5.1. Другие выражения для нормированного среднего квадрата погрешности предсказания

Нормированная погрешность предсказания для автокорреляционного метода определяется как

$$V_n = \frac{\sum_{m=0}^{N+p-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)}, \qquad (8.98а)$$

где $e_n(m)$ — погрешность, соответствующая речевому сегменту $s_n(m)$ для момента n. Для ковариационного метода соответствующее определение имеет вид

$$V_n = \frac{\sum_{m=0}^{N-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)}. \qquad (8.98б)$$

Полагая, что $\alpha_0 = -1$, погрешность предсказания можно записать в виде

$$e_n(m) = -\sum_{k=0}^{p} \alpha_k s_n(m-k). \qquad (8.99)$$

Подставляя (8.99) в (8.98) и используя (8.13), получим, что

$$V_n = \sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i \frac{\varphi_n(i,j)}{\varphi_n(0,0)} \alpha_j, \qquad (8.100а)$$

и подставляя уравнение (8.14) в (8.100), получим

$$V_n = -\sum_{i=0}^{p} \alpha_i \frac{\varphi_n(0,i)}{\varphi_n(0,0)}. \qquad (8.100б)$$

Еще одно выражение для V_n получено в алгоритме Дарбина, т. е.

$$V_n = \prod_{i=1}^{p} (1-k_i^2). \qquad (8.101)$$

Не все полученные выше выражения эквивалентны между собой. Они требуют дополнительного анализа с учетом используемого метода. Например, (8.101), основанное на алгоритме Дарбина, справедливо лишь для автокорреляционного и лестничного методов. Аналогично, поскольку лестничный метод не использует вычисления автокорреляционной функции в явном виде, выражения (8.100) в данном случае непосредственно неприменимы. В табл. 8.2 объединены все перечисленные выше выражения для нормированного среднего квадрата погрешности и показана область применения каждого из выражения. (Для упрощения таблицы индексы n и p опущены.)

8.5.2. Экспериментальное определение погрешности предсказания

Для получения рекомендаций по выбору параметров p и N при практическом использовании алгоритма линейного предсказания Чандра и Лин [15] провели серию исследований. Они измеряли нормированную среднюю квадратическую погрешность предсказания для предсказателя порядка p при различных параметрах алгоритма для следующих случаев: ковариационный и автокор-

Таблица 8.2

Выражения для нормированной погрешности

Выражение	Ковариационный метод	Автокорреляционный метод	Лестничный метод
$V = \frac{\sum_m e^2(m)}{\sum_m s^2(m)}$	Справедливо	Справедливо *	Справедливо
$V = \sum_i \sum_j \alpha_i \frac{\varphi(i,j)}{\varphi(0,0)} \alpha_j$	Справедливо	Справедливо **	Несправедливо
$V = \sum_i \alpha_i \frac{\varphi(i,j)}{\varphi(0,0)}$	Справедливо	Справедливо **	Несправедливо
$V = \prod_i (1-k_i^2)$	Несправедливо	Справедливо	Справедливо

* Это выражение вычисляется по взвешенному сигналу при верхнем пределе $N-1+p$.
** В этом случае $\varphi(i,j) = R(i-j)$.

реляционный методы; синтетические гласные и натуральная речь; синхронный с основным тоном и асинхронный анализ. Нормированная погрешность определялась в соответствии с табл. 8.2. На рис. 8.10—8.15 показаны результаты, полученные Чандрой и Лином [15].

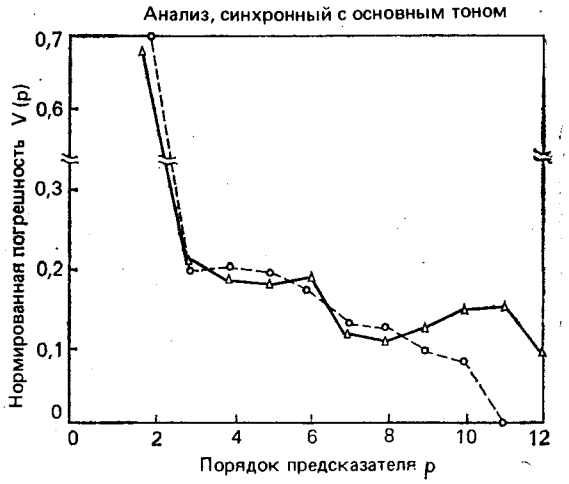


Рис. 8.10. Зависимость погрешности предсказания от порядка предсказателя для вокализованного сегмента синтетического гласного звука при анализе, синхронизированном сигналом основного тона [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

На рис. 8.10 показана нормированная дисперсия V при различном порядке модели для сегмента синтетического звука $|i|$ (в слове «heed») при периоде основного тона, равном 83 отсчетам. Интервал анализа составлял 60 отсчетов и начинался в начале периода основного тона, т. е. результаты соответствуют синхронному анализу. Для ковариационного метода погрешность

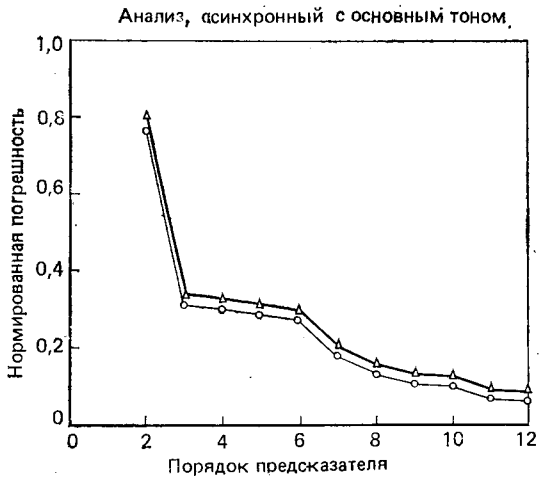


Рис. 8.11. Зависимость погрешности предсказания $V(p)$ от порядка предсказателя p при анализе, синхронизированном сигналом основного тона [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

предсказания монотонно убывает при увеличении порядка модели от 0 до 11, т. е. до порядка модели, используемой при синтезе данного звука. Для автокорреляционного метода погрешность остается на уровне 0,1 при больших p [7]. Это объясняется тем об-

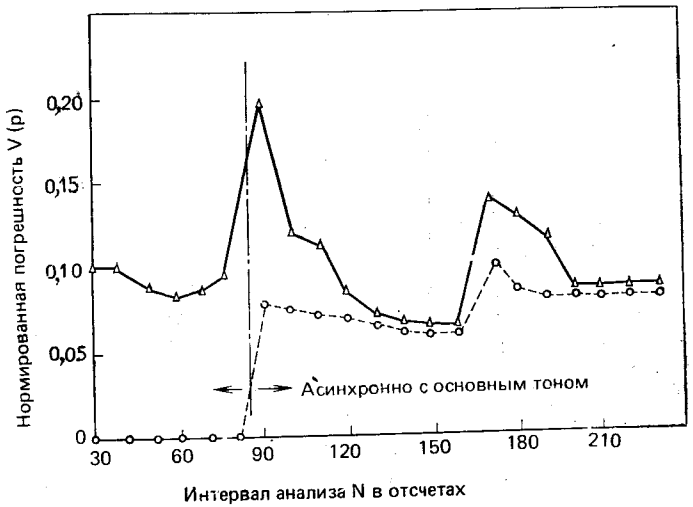


Рис. 8.12. Погрешность предсказания как функция длины интервала анализа N для вокализованного сегмента синтетического речевого сигнала [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

стоятельством, что для случая автокорреляционного анализа при малой протяженности интервала анализа погрешность предсказания в начале сегмента составляет значительную часть общего

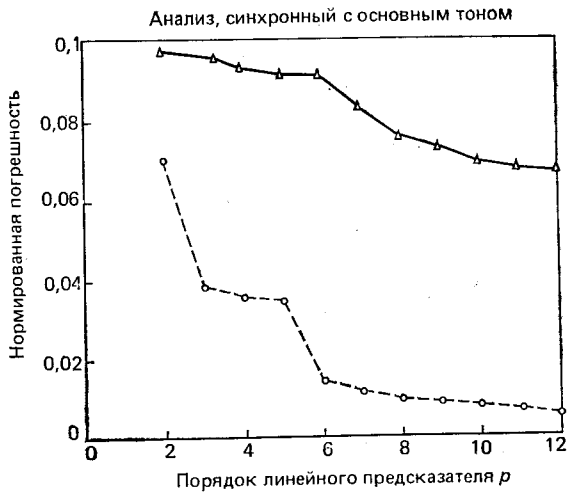


Рис. 8.13. Погрешность предсказания $V(p)$ как функция порядка предсказателя p для вокализованного сегмента натурального речевого сигнала при анализе, синхронизированном основным тоном [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

среднего квадрата погрешности. Этого, конечно, не происходит при ковариационном методе, где для предсказания используются отсчеты вне интервала, на котором ведется предсказание.

На рис. 8.11 показаны результаты анализа с асинхронной обработкой того же сегмента, что и на рис. 8.10. В данном случае

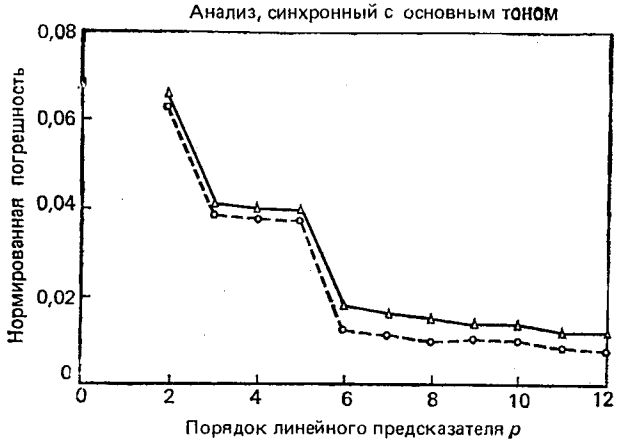


Рис. 8.14. Погрешность предсказания $V(p)$ как функция порядка предсказателя для вокализованного сегмента гласного звука при асинхронном анализе [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

длительность окна составляла 120 отсчетов. При этом ковариационный и автокорреляционный методы дают примерно одинаковые результаты при различных значениях p. Более того, значе-

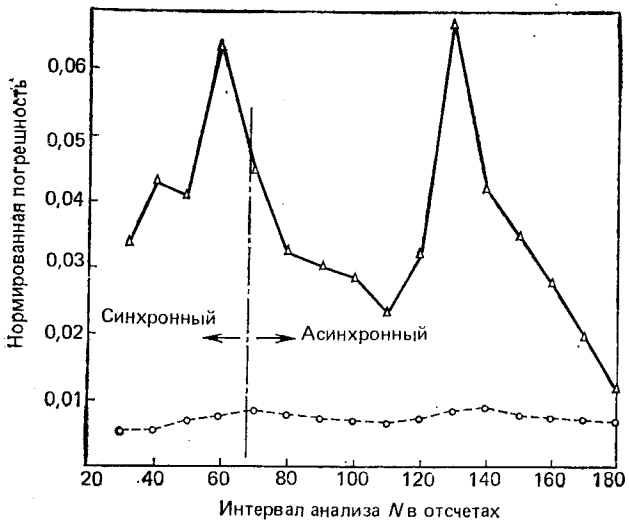


Рис. 8.15. Погрешность предсказания $V(p)$ как функция интервала анализа для вокализованного сегмента реального речевого сигнала [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

ния V монотонно убывают приблизительно до 0,1 при $p=11$. Таким образом, в данном случае при асинхронной обработке, по крайней мере, синтетического звука оба метода приводят к сходным результатам.

На рис. 8.12 показана зависимость V от N для предсказателя при $p=12$ на сегменте синтетической речи. Как и предполагалось, для значений N, значительно меньших периода основного тона (83 отсчета), ковариационный метод приводит к значительно меньшим V, чем автокорреляционный. Величина V резко возрастает в области гармоник основного тона и содержит большие скачки из-за высокой погрешности предсказания в случае использования импульсов для возбуждения системы. Но при достаточно больших значениях N (два и более периодов основного тона) оба подхода приводят к сравнимым значениям V. На рис. 8.13—8.15 приведены аналогичные результаты для вокализованного сегмента реального речевого сигнала. Из рис. 8.13 видно, что при синхронном анализе нормированная погрешность для ковариационного метода значительно меньше, чем для автокорреляционного, а при асинхронном анализе (см. рис. 8.14) результаты сравнимы. Наконец, рис. 8.15 показывает изменение V в зависимости от N при анализе сигнала с $p=12$. В области периода основного тона значение V для автокорреляционного метода заметно изменяется, в то время как при ковариационном методе анализа изменения незначительны. При больших N кривые V для обоих методов приближаются друг к другу.

8.5.3. Зависимость нормированной погрешности предсказания от положения интервала анализа

В 8.5.2 рассмотрены некоторые свойства нормированной погрешности предсказания, а именно зависимость дисперсии погрешности от протяженности временно́го окна N и порядка модели. Остался еще один источник изменения V — изменение дисперсии при изменении положения интервала анализа. Для иллюстрации этого эффекта на рис. 8.16 представлены результаты анализа сегмента длительностью 40 мс гласного звука $|i|$, произнесенного мужским голосом в случае, когда интервал анализа последовательно перемещался каждый раз на один отсчет. На рис. 8.16*а* представлена энергия сигнала (частота дискретизации 10 кГц), на рис. 8.16*б* — нормированная средняя квадратическая погрешность V (частота дискретизации по-прежнему 10 кГц) при использовании ковариационного метода для модели с 14 полюсами ($p=14$) и интервалом анализа 20 мс ($N=200$). На рис. 8.16*в* показана нормированная средняя квадратическая погрешность при использовании окна Хемминга, на рис. 8.16*г* — средняя квадратическая погрешность при использовании автокорреляционного метода в случае прямоугольного окна. Средний период основного тона для этого диктора равен 84 отсчетам (т. е. 8,4 мс), т. е. интервал анализа составляет приблизительно 2,5 периода основного тона, или 20 мс.

Для ковариационного метода имеются значительные изменения погрешности предсказания в зависимости от положения временно́го окна (т. е. погрешность не является гладкой функцией

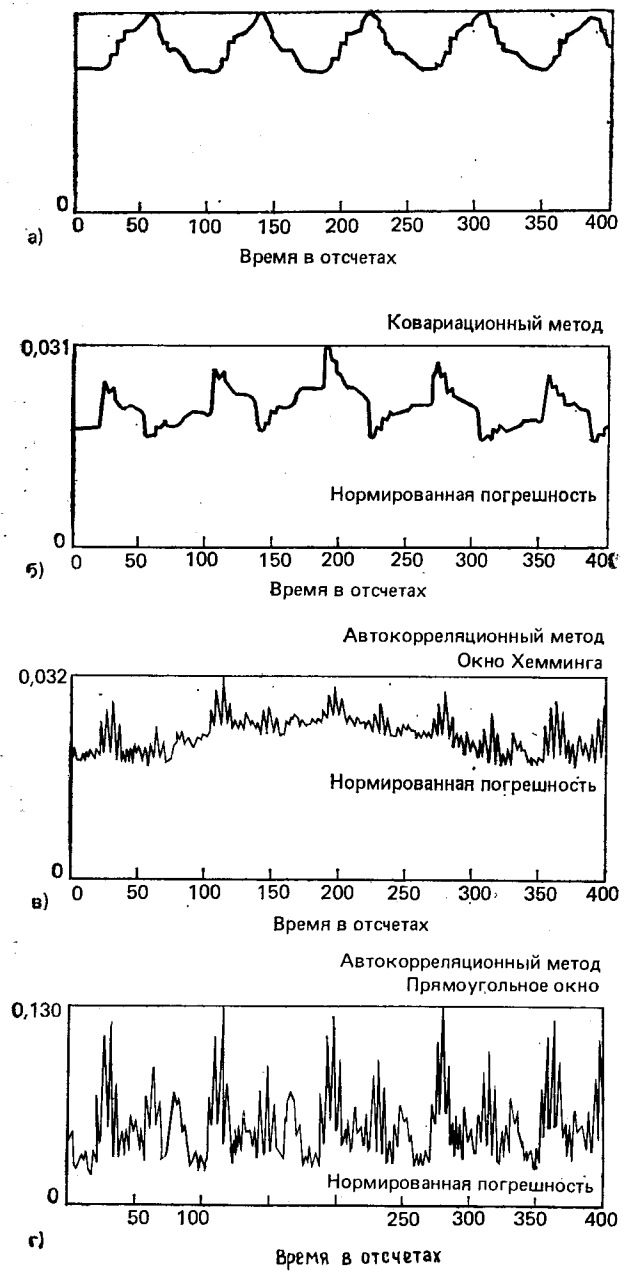


Рис. 8.16. Последовательность погрешности предсказания для 200 отсчетов речевого сигнала и трех систем линейного предсказания [16]

времени). Этот эффект обусловлен наличием значительных пиков погрешности предсказания в начале каждого периода основного тона. Когда в интервал анализа попадают три пика погрешности, нормированная ошибка оказывается больше, чем в случае двух пиков погрешности предсказания. Этим и объясняется наличие резких разрывов в погрешности при попадании в интервал анализа очередного пика в ошибке предсказания. Каждый скачок нормированной погрешности следует непосредственно за участком сглаженной нормированной погрешности предсказания. Точное поведение нормированной погрешности между скачками зависит от особенностей сигнала и метода анализа.

На рис. 8.16*в* и *г* показан различный до некоторой степени характер поведения погрешности при использовании автокорреляционного метода анализа для окна Хемминга и прямоугольного окна. Как видно из рисунков, в данном случае нормированный средний квадрат погрешности предсказания содержит в основном высокочастотные компоненты и слабо зависит от наличия импульсов основного тона в интервале анализа. Высокочастотные компоненты объясняются наличием в каждом интервале анализа первых p отсчетов, которые линейно непредсказуемы. Эти флуктуации при использовании окна Хемминга значительно меньше, чем в случае прямоугольного окна, поскольку окно Хемминга уменьшается к концу интервала анализа. Другая компонента высокочастотных флуктуаций определяется взаимным расположением импульсов основного тона и начала интервала анализа, как это выше излагалось для ковариационного метода. Однако в автокорреляционном методе этот эффект сказывается слабее, чем в ковариационном, особенно при использовании окна Хемминга, поскольку новые импульсы основного тона, попадающие в интервал анализа, ослабляются вследствие применения окна.

Изменения, подобные показанным на рис. 8.16, типичны для большинства гласных звуков [16]. Флуктуации в погрешности из-за различного положения окна могут быть ослаблены применением фильтрации и спектрального предыскажения перед обработкой сигнала с помощью линейного предсказания [16].

8.6. Анализ линейного предсказания в частотной области

До сих пор методы линейного предсказания рассматривались на основе разностных уравнений и корреляционных функций, т. е. с позиций представления сигнала во временной области. Однако предполагалось, что коэффициенты линейного предсказания являются коэффициентами знаменателя передаточной функции, описывающей действие речевого тракта, формы сигнала возбуждения и излучения. Таким образом, располагая совокупностью параметров предсказания, можно определить частотную характеристику модели речеобразования путем простой подстановки в $H(z)$ значения $z=e^{i\omega}$, т. е.

$$H(e^{i\omega}) = \frac{G}{1-\sum_{k=1}^{p}\alpha_k e^{-i\omega_k}} = \frac{G}{A(e^{i\omega})}. \qquad (8.102)$$

Если изобразить $H(e^{i\omega})$ как функцию частоты[1], то можно ожидать, что на формантных частотах будут видны максимумы, как и при рассмотрении спектральных представлений в предыдущей главе. Таким образом, линейное предсказание можно рассматривать как метод кратковременной оценки спектра. Действительно, подобные методы широко применяются не только при обработке сигналов речи [12]. В данном параграфе метод линейного предсказания по минимуму среднего квадрата ошибки трактуется на основе частотного представления и проводится сравнение данного метода с другими методами представления речевого сигнала в частотной области.

8.6.1. Спектральная трактовка среднего квадрата погрешности предсказания

Рассмотрим параметры линейного предсказания, полученные с помощью автокорреляционного метода. В этом случае погрешность представления во временно́й области будет иметь вид

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m), \tag{8.103а}$$

или в частотной области на основе теоремы Парсеваля

$$E_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S_n(e^{i\omega})|^2 |A(e^{i\omega})|^2 d\omega, \tag{8.103б}$$

где $S_n(e^{i\omega})$ — преобразование Фурье для сегмента сигнала $s_n(m)$, а

$$A(e^{i\omega}) = 1 - \sum_{k=1}^{p} \alpha_k e^{-i\omega k}. \tag{8.104}$$

Вспоминая, что

$$H(e^{i\omega}) = G/A(e^{i\omega}), \tag{8.105}$$

и используя (8.103), можно получить

$$E_n = \frac{G^2}{2\pi}\int_{-\pi}^{\pi} \frac{|S_n(e^{i\omega})|^2}{|H(e^{i\omega})|^2} d\omega. \tag{8.106}$$

Поскольку подынтегральное выражение в (8.106) положительно, то минимизация E_n эквивалентна минимизации отношения энергетического спектра сигнала к квадрату модуля частотной характеристики линейной системы в модели речеобразования.

В § 8.2 показано, что автокорреляционная функция $R_n(m)$ сегмента речевого сигнала $s_n(m)$ и автокорреляционная функция $\tilde{R}(m)$ импульсной характеристики $h(m)$, соответствующей системной функции $H(z)$, совпадают для первых $(p+1)$ значений. Таким образом, при $p \to \infty$ соответствующие автокорреляционные функции совпадают при всех значениях и, следовательно,

$$\lim_{p\to\infty} |H(e^{i\omega})|^2 = |S_n(e^{i\omega})|^2. \tag{8.107}$$

[1] См. задачу 8.2, где рассмотрен метод вычисления $H(e^{i\omega})$ с использованием БПФ.

Это означает, что при достаточно большом p можно аппроксимировать спектр сигнала с любой точностью.

Для иллюстрации возможностей спектрального анализа на основе линейного предсказания на рис. 8.17 [7] показаны кривые $20\log_{10}|H(e^{i\omega})|$ и $20\log_{10}|S_n(e^{i\omega})|$. Спектр сигнала получен по

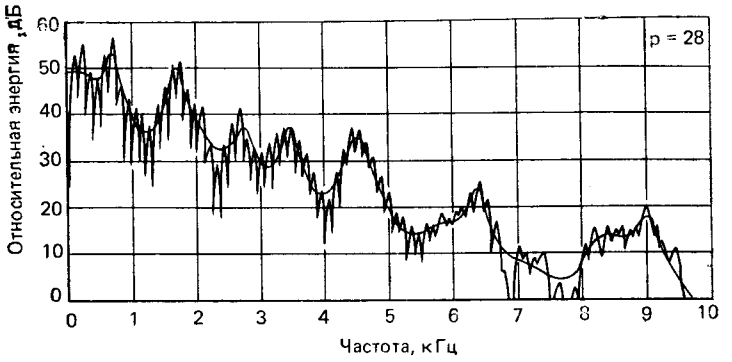


Рис. 8.17. Спектр речевого сигнала и 28-полюсной модели [17]

алгоритму БПФ на сегменте речевого сигнала длительностью 20 мс (при частоте дискретизации 20 кГц) с использованием окна Хемминга (см. гл. 6). Расчет сделан для звука $|ae|$. Спектр системы с линейным предсказанием получен для случая, когда используется предсказатель 28-го порядка ($p=28$), а его коэффициенты рассчитаны по автокорреляционному методу [2]. На рисунке отчетливо видна гармоническая структура спектра сигнала. На том же рисунке проявляется важная особенность анализа спектра с помощью линейного предсказания. Спектральное описание линейного предсказания лучше согласуется со спектром сигнала: более высокая точность обеспечивается в области больших значений спектральной плотности (т. е. вблизи максимумов спектра), более низкая точность — в области малых значений (т. е. в области провалов спектра). Это не является неожиданным, если учесть, что в соответствии с (8.106) в общую погрешность предсказания большой вклад вносят области, где $|S_n(e^{i\omega})| > |H(e^{i\omega})|$, по сравнению с областями, в которых $|S_n(e^{i\omega})| < |H(e^{i\omega})|$. Таким образом, спектральное описание на основе линейного предсказания, отвечая критерию оптимальности, приводит к хорошим результатам в области спектральных максимумов и к значительно худшим — в области минимумов спектра.

Проведенное выше обсуждение позволяет считать, что выбором порядка предсказателя p можно добиться нужной степени сглаживания спектра. Это утверждение иллюстрирует рис. 8.18, на котором показан сегмент речевого сигнала, его преобразование Фурье и спектры, полученные на основе линейного предсказания при различном порядке p. Очевидно, что увеличение p приводит к более детальному описанию спектральной плотности. Поскольку

наша задача сводится к получению лишь спектральных изменений, обусловленных совместным действием источника возбуждения, речевого тракта и излучения, требуется выбирать p таким образом, чтобы сохранить положение формантных максимумов и особенности формы спектральной плотности.

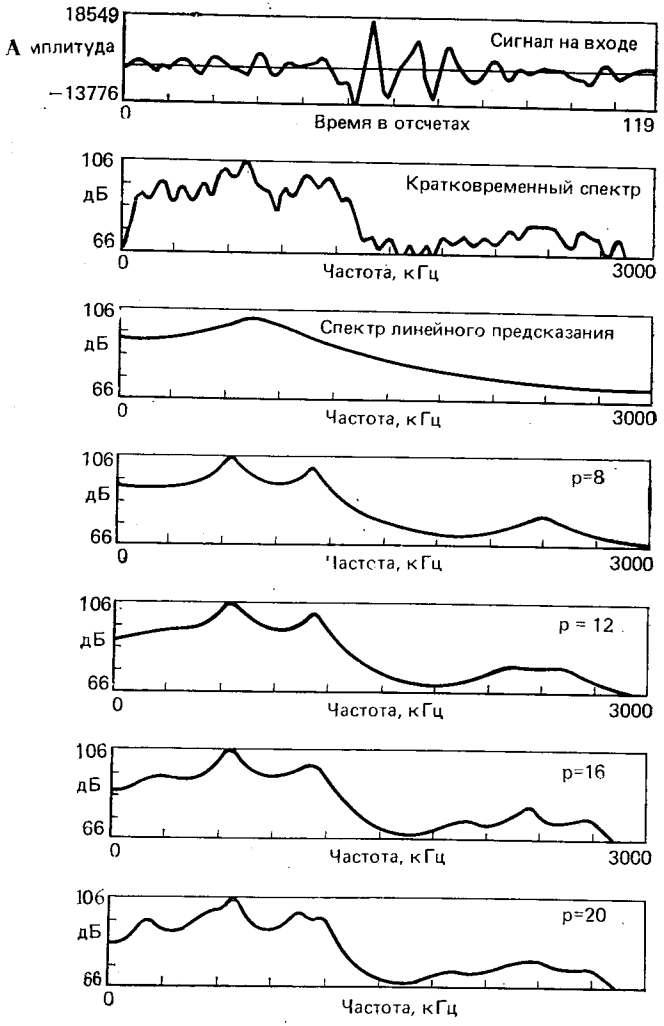


Рис. 8.18. Спектр гласного /a/ при частоте дискретизации 6 кГц для различных значений порядка предсказателя p

Предполагается, что коэффициенты предсказания оцениваются по автокорреляционному методу. Только в этом случае преобразование Фурье кратковременной автокорреляционной функции совпадает с квадратом модуля кратковременного преобразования

Фурье сигнала. Однако это не исключает использования $H(e^{i\omega})$ в качестве оценки спектра, даже если коэффициенты предсказания оцениваются по ковариационному методу.

8.6.2. Сравнение кратковременного спектрального анализа с оценкой спектра на основе линейного предсказания

В качестве примера на рис. 8.19 показаны четыре логарифмических спектра сегмента синтетического гласного |a| [10]. Первые две кривые получены с использованием методов кратковременного анализа спектра, рассмотренных в гл. 6. В первом случае использовался взвешенный сегмент сигнала длительностью 512 отсчетов (51,2 мс), который преобразовывался (с использованием 512-точечного БПФ) для осуществления относительно узкополосного спектрального анализа, результат которого показан в верхней части рис. 8.19. В данном спектре хорошо просматриваются отдельные гармоники сигнала возбуждения, что объясняется большой протяженностью интервала анализа.

На втором рисунке интервал анализа уменьшен до 128 отсчетов (12,8 мс), что привело к широкополосному спектральному анализу сигнала. Здесь уже отдельные гармоники неразличимы, просматривается огибающая спектра в целом. Хотя в данном случае формантные частоты в спектре видны хорошо, но их однозначная идентификация или оценка затруднительна.

Третий спектр получен на основе гомоморфного сглаживания, рассмотренного в гл. 7. Несглаженный спектр был рассчитан по 300 отсчетам (30 мс) с использованием БПФ. Сглаженный спектр, показанный на рисунке, был получен линейным сглаживанием логарифма спектра. В данном случае отдельные форманты легко разделяются и могут быть измерены с использованием простых методов

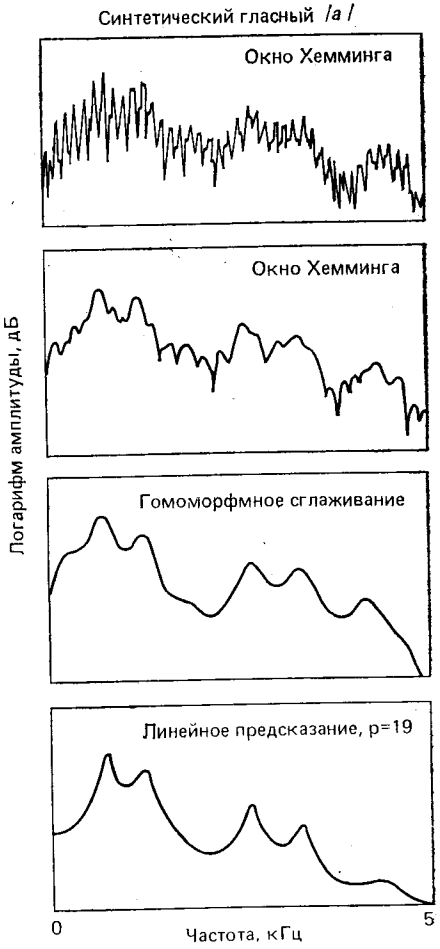


Рис. 8.19. Спектр синтетического звука /a/

поиска экстремумов. Однако оценка полосы форманты в данном случае очень сложна из-за использования сглаживания, применяемого для получения окончательного спектра.

Спектр, показанный в нижней части рисунка, получен в результате анализа на основе линейного предсказания с использованием модели при $p=12$ для сегмента длительностью 128 отсчетов (12,8 мс). Сравнивая этот спектр с другими, можно отметить, что параметрическое описание позволяет четко выявить формантную структуру без дополнительных побочных экстремумов и флуктуаций. Это объясняется тем, что модель линейного предсказания хорошо описывает речевой тракт в случае гласных звуков при правильном выборе порядка p. Поскольку порядок модели можно определить по полосе сигнала, метод линейного предсказания приводит к хорошей оценке спектральных свойств источника возбуждения, речевого тракта и излучателя.

На рис. 8.20 показано сравнение спектров сегмента натурального гласного звука, полученных гомоморфным сглаживанием и

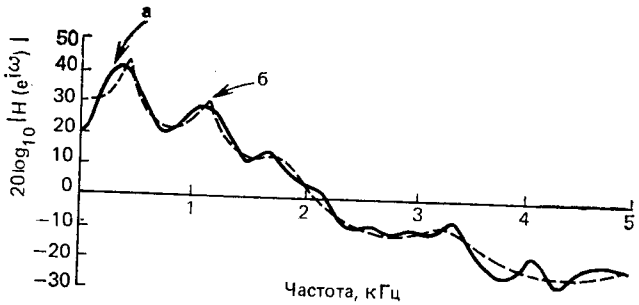


Рис. 8.20. Сравнение спектра речи, полученного с помощью кепстрального сглаживания (*а*) и линейного предсказания (*б*)

с использованием линейного предсказания. Хотя формантные частоты в обоих случаях практически совпадают, спектр линейного предсказания имеет меньше побочных пиков. Это объясняется тем, что при анализе на основе линейного предсказания и порядке модели $p=12$ в спектре может возникнуть не более шести пиков. При гомоморфном анализе такое ограничение отсутствует. Как отмечалось выше, спектральные пики при анализе на основе линейного предсказания более острые, чем при гомоморфном анализе, что обусловлено применением сглаживания спектра в последнем случае.

8.6.3. Селективное линейное предсказание

Изложенные выше идеи можно применять не ко всей полосе спектра сигнала, а только к ее части. Такой метод назван Макхоулом методом селективного линейного предсказания [8]. Появление этого метода объясняется тем, что в ряде случаев необхо-

димо использовать лишь часть спектральной плотности сигнала. Например, для адекватного описания фрикативов в системах распознавания речи используется частота дискретизации, равная 20 кГц. Для вокализованных сегментов при этом требуется диапазон от 0 до 4 кГц, а для невокализованных сегментов более важен диапазон от 4 до 8 кГц. Используя селективное представление, спектр в диапазоне от 0 до 4 кГц можно сформировать на основе предсказателя порядка p_1, а спектр в диапазоне от 4 до 8 кГц — на основе другого предсказателя — порядка p_2.

Селективное линейное предсказание осуществляется следующим способом. Чтобы сформировать сигнал лишь в диапазоне от $f=f_A$ до $f=f_B$, требуется лишь линейно отобразить эту область так, что $f=f_A$ отображается в $f'=0$, а $f=f_B$ в $f'=\omega'/2\pi=0{,}5$ (т. е. в половину частоты дискретизации). Параметры предсказания являются решением системы уравнений предсказания, коэффициенты корреляции в которой получены по формуле

$$R'(i)=\frac{1}{2\pi}\int_{-\pi}^{\pi}|S_n(e^{i\omega'})|^2 e^{i\omega' i}\,d\omega'. \qquad (8.108)$$

На рис. 8.21 представлены результаты селективного линейного предсказания [8]. Исходный сигнал был тем же, что и на рис. 8.17. В диапазоне от 0 до 5 кГц использовалась модель с $p_1=14$,

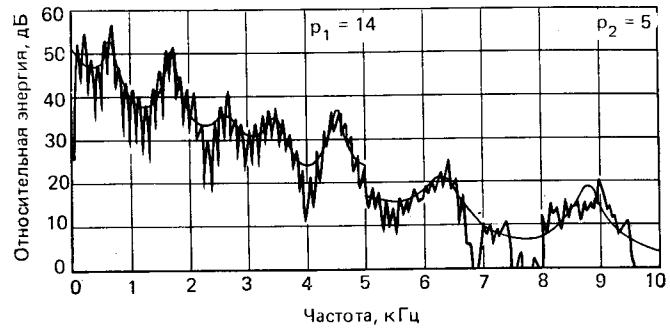


Рис. 8.21. Применение селективного линейного предсказания к спектру сигнала, изображенному на рис. 8.17, с использованием 14-полюсной модели в диапазоне 0—5 кГц и 5-полюсной модели в диапазоне 5—10 кГц [2]

а диапазон от 5 до 10 кГц описывался независимо моделью с $p_2=5$. На частоте 5 кГц имеется разрыв в спектральной плотности, что объясняется отсутствием условий согласования обоих спектров.

8.6.4. Сравнение методов линейного предсказания с методами анализа через синтез

Как излагалось в гл. 6, мера погрешности, обычно используемая в методе анализа через синтез, представляет собой логарифм

отношения спектральной плотности мощности сигнала к квадрату модуля спектральной плотности мощности модели:

$$E' = \int_{-\pi}^{\pi} \left\{ \log \left[\frac{|S_n(e^{i\omega})|^2}{|H(e^{i\omega})|^2} \right] \right\}^2 d\omega. \qquad (8.109)$$

Таким образом, минимизация E' в методе анализа через синтез эквивалентна минимизации среднего квадратического отклонения между логарифмами спектров.

Сравнение мер погрешности моделирования на основе методов линейного предсказания и анализа через синтез приводит к следующим выводам:

1. Оба метода связаны с соотношением спектров сигнала и модели.

2. Оба метода одинаково учитывают различные частотные диапазоны.

3. Оба метода пригодны для минимизации погрешности в некотором выбранном диапазоне.

4. Критерий качества в линейном предсказании более чувствителен к тем участкам спектра, на которых $|S_n(e^{i\omega})|^2 > |H(e^{i\omega})|^2$, в то время как в методе анализа через синтез критерий качества одинаково чувствителен во всем диапазоне.

Из сказанного следует, что при анализе несглаженного спектра (см. рис. 8.17) критерий метода линейного предсказания приводит к лучшим результатам, чем критерий метода анализа через синтез [7]. Более того, объем вычислений, необходимых при использовании линейного предсказания, значительно меньше. Если же анализируется сигнал с гладким спектром (например, с помощью набора фильтров), то как анализ на основе линейного предсказания, так и метод анализа через синтез приводят к хорошему совпадению спектров. На практике для спектральных плотностей такого типа почти всегда применяется метод анализа через синтез.

8.7. Применение анализа на основе линейного предсказания к моделям речевого тракта в виде труб без потерь

В гл. 3 рассматривалась модель речеобразования, включавшая в себя последовательное соединение N акустических труб без потерь (рис. 8.22). Коэффициенты отражения r_k на рис. 8.22 связаны с площадями поперечного сечения соотношением

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}. \qquad (8.110)$$

В 3.3.4 получена передаточная функция такой системы в предположении, что коэффициент отражения от источника возбуждения $r_G = 1$, т. е. сопротивление источника предполагается бесконечно

большим. Передаточная функция системы, представленной на рис. 8.22, как показано в 3.3.4, имеет вид

$$V(z) = \frac{\prod_{k=1}^{N} (1 + r_k)\, z^{-N/2}}{D(z)}, \qquad (8.111)$$

где $D(z)$ удовлетворяет соотношениям:

$$D_0(z) = 1; \qquad (8.112а)$$

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}); \qquad (8.112б)$$

$$D(z) = D_N(z). \qquad (8.112в)$$

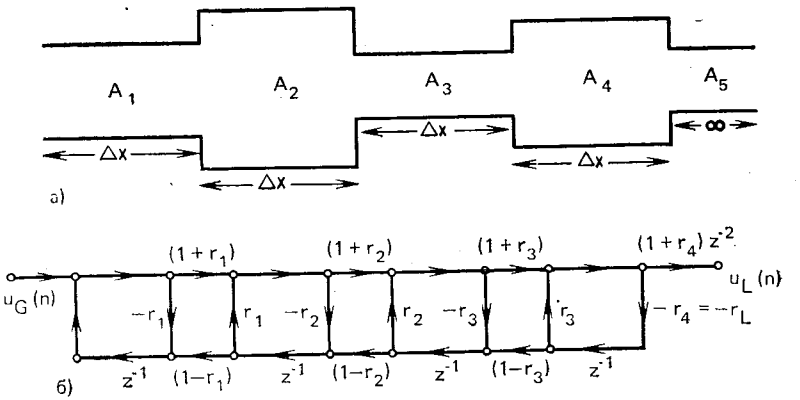


Рис. 8.22. Модель трубы без потерь, нагруженной на трубу бесконечной длины (*а*) и граф прохождения сигнала при бесконечном сопротивлении источника (*б*)

Все это весьма напоминает обсуждение лестничного метода в 8.8.3. Действительно, там было показано, что полином

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}, \qquad (8.113)$$

полученный при анализе на основе линейного предсказания, можно получить с использованием рекурсивной процедуры:

$$A^{(0)}(z) = 1; \qquad (8.114а)$$

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}); \qquad (8.114б)$$

$$A(z) = A^{(p)}(z), \qquad (8.114в)$$

где параметры $\{k_i\}$ названы коэффициентами частной корреляции. Сравнивая уравнения (8.112) и (8.114), замечаем, что передаточная функция

$$H(z) = G/A(z), \qquad (8.115)$$

полученная на основе линейного предсказания, имеет тот же вид, что и передаточная функция акустической трубы без потерь, имеющей p секций. Если

$$r_i = -k_i, \tag{8.116}$$

то очевидно, что

$$D(z) = A(z). \tag{8.117}$$

Используя (8.110) и (8.116), легко показать, что эквивалентные площади поперечных сечений модели в виде неоднородной трубы связаны с коэффициентами частных корреляций соотношением

$$A_{i+1} = \left[\frac{1-k_i}{1+k_i}\right] A_i. \tag{8.118}$$

Заметим, что частные корреляции определяют соотношение площадей соседних секций. Таким образом, площади поперечного сечения модели в виде неоднородной трубы не определяются абсолютно точно, а все модели с подходящими условиями нормализации дают одинаковую передаточную функцию.

«Функция площади», полученная с использованием (8.118), не является соответствующей функцией для речевого тракта человека. Однако Вакита [17] показал, что при использовании предыскажений, устраняющих влияние источника возбуждения и излучения, функции площади, описывающие речевой тракт, часто бывают весьма сходными с конфигурацией голосового тракта, используемого человеком при речеобразовании.

8.8. Соотношения между различными параметрами речи

Хотя коэффициенты предсказания α_k, $1 \leqslant k \leqslant p$, часто считаются основными параметрами при анализе речи на основе линейного предсказания, обычно сразу же возникает задача преобразования этих параметров в некоторые другие для получения иных представлений речевого сигнала. Эти представления часто оказываются более удобными при применении линейного предсказания. В этом разделе рассматриваются методы получения других полезных описаний сигнала на основе непосредственного использования параметров линейного предсказания [1, 2].

8.8.1. Корни полинома передаточной функции предсказателя

Вместо коэффициентов линейного предсказания можно использовать корни полинома

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \sum_{k=1}^{p} (1 - z_k z^{-1}). \tag{8.119}$$

Множество корней $\{z_i,\ i=1,\ 2,\ \ldots,\ p\}$ представляет собой эквивалентное представление $A(z)$. При необходимости пересчитать корни на z-плоскости в корни на s-плоскости воспользуемся подстановкой

$$z_i = e^{s_i T}, \tag{8.120}$$

где $s_i = \sigma_i + i\,\Omega_i$ — корень на s-плоскости, соответствующий корню z_i на z-плоскости. Если $z_i = z_{ir} + i\,z_{ii}$, то

$$\Omega_i = (1/T)\,\tan^{-1}(z_{ii}/z_{ir}) \tag{8.121}$$

и

$$\sigma_i = (1/2T)\,\log(z_{ir}^2 + z_{ii}^2). \tag{8.122}$$

Соотношения (8.121) и (8.122) полезны в случаях применения линейного предсказания в формантном анализе.

8.8.2. Кепстр

Другим представлением сигнала является кепстр импульсной характеристики всей системы линейного предсказания. Если система с линейным предсказанием имеет передаточную функцию $H(z)$ с импульсной реакцией $h(n)$ и комплексным кепстром $\hat{h}(n)$, то можно показать, что кепстр $\hat{h}(n)$ получается с помощью рекурсивных соотношений

$$\hat{h}(n) = \alpha_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \hat{h}(k)\,\alpha_{n-k},\ n \geqslant 1, \tag{8.123}$$

где

$$H(z) = \sum_{n=0}^{\infty} h(n)\,z^{-n} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}. \tag{8.124}$$

8.8.3. Импульсная характеристика полюсной системы

Импульсная характеристика $h(n)$ полюсной системы с передаточной функцией (8.124) может быть определена на основе рекурсивного уравнения

$$h(n) = \sum_{k=1}^{p} \alpha_k\,h(n-k) + G\,\delta(n), \tag{8.125}$$

где $h(n)$ предполагается (по определению) равной 0 для $n<0$ и G — амплитуда возбуждения.

8.8.4. Автокорреляционная функция импульсной характеристики

Как отмечалось в § 8.2, автокорреляционная функция импульсной характеристики определяется выражением (см. задачу 8.1)

$$\widetilde{R}(i)=\sum_{n=0}^{\infty} h(n)\,h(n-i)=\widetilde{R}(-i). \tag{8.126}$$

и удовлетворяет соотношениям

$$\widetilde{R}(i)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(|i-k|). \tag{8.127}$$

и

$$\widetilde{R}(0)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(k)+G^2. \tag{8.128}$$

Уравнения (8.127) и (8.128) можно использовать для определения $\widetilde{R}(i)$ по коэффициентам предсказания и наоборот.

8.8.5. Коэффициенты автокорреляции полиномиальной передаточной функции предсказателя

Полином передаточной функции предсказателя (обратного фильтра) имеет вид

$$A(z)=1-\sum_{k=1}^{p}\alpha_k z^{-k}, \tag{8.129}$$

а импульсная характеристика равна

$$a(n)=\delta(n)-\sum_{k=1}^{p}\alpha_k\delta(n-k).$$

Автокорреляционная функция импульсной характеристики обратного фильтра определяется соотношением

$$R_a(i)=\sum_{k=0}^{p-i}a(k)\,a(k+i),\ 0\leqslant i\leqslant p. \tag{8.130}$$

8.8.6. Коэффициенты частной корреляции

Для автокорреляционного метода коэффициенты предсказания можно получить по коэффициентам частной корреляции, используя рекурсивные соотношения:

$$a_i^{(i)}=k_i; \tag{8.131а}$$

$$a_j^{(i)}=a_j^{(i-1)}-k_i\,a_{i-j}^{(i-1)},\ 1\leqslant j\leqslant i-1, \tag{8.131б}$$

решая (8.131а) и (8.131б) для $i=1, 2, \ldots, p$ и устанавливая последний набор коэффициентов равным

$$\alpha_j=a_j^{(p)},\ 1\leqslant j\leqslant p. \tag{8.131в}$$

Аналогично частные корреляции можно рассчитать по коэффициенту линейного предсказания, используя возвратную рекурсию в виде:

$$k_i=a_i^{(i)}; \tag{8.132а}$$

$$a_j^{(i-1)}=\frac{a_j^{(i)}+a_i^{(i)}\,a_{i-j}^{(i)}}{1-k_i^2},\ 1\leqslant j\leqslant i-1, \tag{8.132б}$$

где i изменяется от p к $p-1$ и т. д. до 1 и финальное решение есть

$$a_j^{(p)}=\alpha_j,\ 1\leqslant j\leqslant p. \tag{8.132в}$$

8.8.7. Логарифм отношения площадей

Важной совокупностью эквивалентных параметров, которые можно получить по коэффициентам частных корреляций, является совокупность логарифмов отношений площадей поперечного сечения, определяемых как

$$g_i=\log\left[\frac{A_{i+1}}{A_i}\right]=\log\left[\frac{1-k_i}{1+k_i}\right],\ 1\leqslant i\leqslant p. \tag{8.133}$$

Параметры g_i эквивалентны логарифму отношения площадей поперечных сечений соседних секций в модели неоднородной акустической трубы без потерь, которая имеет ту же передаточную функцию, что и модель линейного предсказания (см. § 8.7). Параметры g_i, как установлено в [2] и другими [1], наиболее удобны для квантования вследствие незначительной спектральной чувствительности величин g'_i.

Параметры k_i можно непосредственно получить по параметрам g_i с помощью обратного преобразования:

$$k_i=\frac{1-e^{g_i}}{1+e^{g_i}},\ 1\leqslant i\leqslant p. \tag{8.134}$$

8.9. Синтез речевого сигнала по параметрам линейного предсказания

Речевой сигнал может быть синтезирован по параметрам линейного предсказания различными способами. Простейший способ состоит в использовании для синтеза системы, описываемой теми же параметрами, которые применялись при анализе. На рис. 8.23 изображена структурная схема такого синтезатора. Для синтеза речевого сигнала в данном случае используются такие меняющиеся во времени параметры, как период основного тона, признак тон—шум, коэффициент усиления или минимальное среднее квадратическое значение и p коэффициентов линейного предсказания. Импульсный генератор в данном случае работает как источник возбуждения на вокализованных звуках, формирующий импульсы

в начале каждого периода основного тона. Генератор шума представляет собой источник возбуждения на невокализованных сегментах, формирующий некоррелированный равномерно распределенный случайный процесс с единичной дисперсией и нулевым

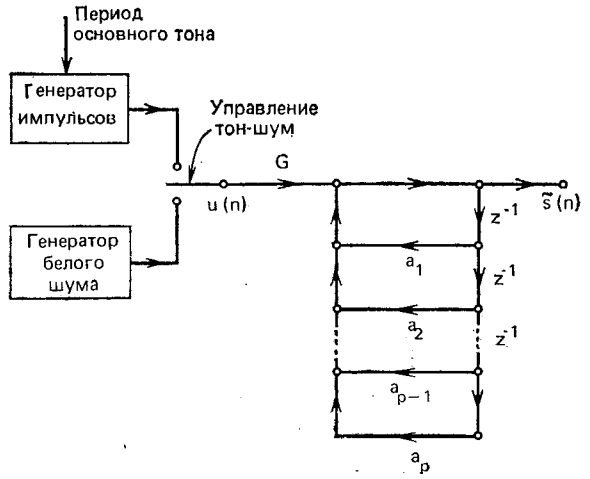


Рис. 8.23. Структурная схема синтезатора на основе линейного предсказания

средним. Выбор между двумя источниками обеспечивается с помощью признака тон—шум. Коэффициент усиления G определяет полную амплитуду возбуждения. Отсчет синтезированной речи определяется соотношением

$$\tilde{s}(n)=\sum_{k=1}^{p}\alpha_k\,\tilde{s}(n-k)+G\,u(n). \qquad (8.135)$$

Схема устройства, реализующего (8.135), представлена на рис. 8.23. Эта схема представляет собой простой и непосредственный способ реализации синтезатора речи по параметрам предсказания. Для воспроизведения каждого отсчета требуется p умножений и p сложений.

В модели синтеза, представленной на рис. 8.23, параметры синтезатора должны изменяться во времени. Хотя обычно оценка параметров производится периодически на интервалах вокализованной речи, управляющие параметры синтезатора изменяются в начале каждого периода основного тона. Для невокализованной речи они изменяются 1 раз на интервале (т. е. через каждые 10 мс для скорости 100 отсчетов на интервал анализа). Установлено, что подстройка управляющих параметров в начале каждого периода основного тона (называемая синтезом, синхронным с основным тоном) является более эффективной по сравнению с подстройкой 1 раз на интервале анализа (которая называется асинхронной). Это, в свою очередь, требует интерполяции параметров

для того, чтобы получить их значения в начале любого периода основного тона.

Установлено, что параметры усиления и основного тона следует интерполировать геометрически (т. е. линейно в логарифмическом масштабе) [3], однако вследствие требования устойчивости параметры линейного предсказания непосредственно интерполировать нельзя. Это обусловлено тем, что интерполяция между двумя устойчивыми множествами параметров может привести к неустойчивым параметрам. Одним из путей преодоления этой трудности, в соответствии с результатами Атала, является интерполяция первых p отсчетов автокорреляционной функции импульсной реакции фильтра (см. рис. 8.21). Используя соотношения, полученные в § 8.4, коэффициенты предсказания можно получить по первым p отсчетам автокорреляции импульсной характеристики и наоборот. Более того, интерполяция автокорреляционных коэффициентов всегда приводит к получению устойчивого фильтра[1].

Синтезатор, изображенный на рис. 8.23, используется в ряде приложений при моделировании систем с линейным предсказанием. Его основным достоинством является простота технической реализации. Существенный недостаток заключается в том, что синтезатор представляет собой прямую форму рекурсивного цифрового фильтра, что требует высокой точности при вычислении коэффициентов, ибо прямая форма программирования весьма чувствительна к изменениям коэффициентов. Другой, более удобный способ синтеза речевого сигнала может быть основан на использовании коэффициентов отражения или коэффициентов частных корреляций в рамках модели неоднородной трубы без потерь. Другими словами, схема фильтра, изображенного на рис. 8.23, может быть заменена схемой рис. 8.22. Преимущество этого подхода заключается в том, что в данном случае синтез проводится на основе ограниченных коэффициентов отражения $r_i=-k_i$ ($|k_i|<1$), которые можно интерполировать непосредственно, без нарушения устойчивости фильтра. Такая структура также менее чувствительна к погрешностям квантования, возникающим при цифровой реализации, чем фильтр в прямой форме.

Из рис. 8.22*б* очевидно также, что для реализации предсказателя порядка p в данном случае требуется $4p+2$ умножений и $2(p-1)$ сложений на отсчет по сравнению с p сложениями и умножениями в фильтре с прямой формой. В 3.3.3 показано, что секции с четырьмя умножениями можно заменить секциями с двумя умножениями за счет увеличения числа сложений. Осуществляя преобразования, показанные на рис. 3.41, граф рис. 8.22 можно изобразить в виде рис. 8.24. Рисунок 8.24*а* приводит к фильтру с $2p-1$ умножениями и $4p-1$ сложениями, а рис. 8.24*б*

[1] Аналогично можно интерполировать частные корреляции и логарифмы площади; устойчивость сохраняется при условии устойчивости исходных совокупностей параметров.

изображает фильтр с p умножениями и $3p-2$ сложениями. При использовании модели в виде неоднородной трубы без потерь для целей синтеза выбор той или иной формы зависит от ряда факторов, так что невозможно утверждать однозначно, какая форма является более эффективной.

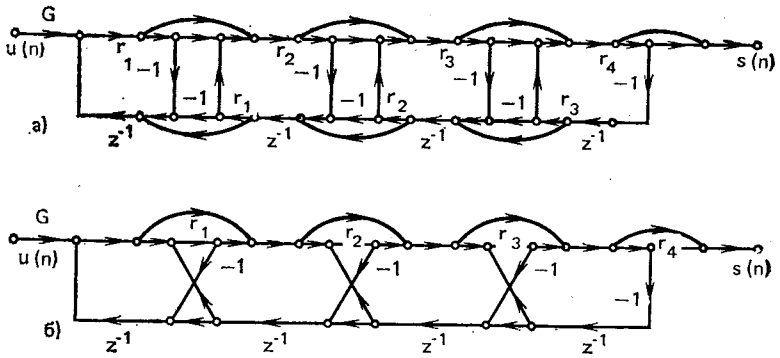


Рис. 8.24. Эквивалентные модели акустической трубы без потерь:
а) с двумя операциями умножения; *б*) с одной операцией умножения

8.10. Применение параметров линейного предсказания

Как следует из результатов предыдущего параграфа, теория линейного предсказания достаточно хорошо разработана. Разработаны методы оценки всех основных параметров речевого сигнала. На основе такого анализа проведены обширные исследования вокодеров, что привело к пониманию свойств различных представлений линейного предсказания применительно к квантованию. Эти методы, наконец, получили распространение при решении задач верификации и идентификации дикторов, распознавания и классификации речи, устранения реверберации и т. д. В § 8.10 будут представлены методы оценивания параметров речевого сигнала на основе линейного предсказания.

8.10.1. Оценивание основного тона на основе коэффициентов линейного предсказания

Выше уже обсуждался вопрос о том, как на основе использования сигнала погрешности можно, по крайней мере теоретически, построить оценку основного тона. Хотя этот метод, вообще говоря, позволяет оценить период основного тона достаточно точно, в [19] предложен несколько иной алгоритм, называемый SIFT-методом (метод обратной фильтрации). Сходный подход предложен в [20].

На рис. 8.25 представлена структурная схема SIFT-алгоритма. Входной сигнал $s(n)$ поступает на вход фильтра нижних частот с частотой среза около 900 Гц и затем обычная частота дискретизации 10 кГц снижается до 2 кГц путем прореживания (т. е. каждые четыре из пяти отсчетов выбрасываются). Прореженный

выход $x(n)$ затем анализируется с использованием автокорреляционного метода. Обратного фильтра четвертого порядка оказывается вполне достаточно для этих целей, поскольку в диапазоне до 1 кГц имеется не более двух формант. В результате анализа на выходе обратного фильтра получается сигнал с почти равномерным спектром[1]. Цель линейного предсказания, таким образом, заключается в выравнивании спектра подобно тому, как это делалось при клиппировании (см. гл. 4). Затем вычисляются кратковременная автокорреляционная функция погрешности предсказания и положение максимума в ней в подходящем интервале задержек выбирается в качестве оценки периода основного тона. Для получения дополнительной точности при оценке основного тона применяется интерполяция автокорреляционной функции в области максимального значения. Сегмент речи классифицируется как невокализованный, если максимальное значение автокорреляционной функции (нормированной соответствующим образом) оказывается ниже некоторого выбранного порога.

На рис. 8.26 [19] показаны колебания, полученные в различных точках анализатора. На рис. 8.26*а* изображен отрезок анализируемого входного сигнала, на рис. 8.26*б* — спектр входного сигнала и спектр сигнала на выходе обратного фильтра. В данном примере имеется лишь одна форманта на частоте 250 Гц. На рис. 8.26*в* показан спектр, а на рис. 8.26*г* — временна́я диаграмма сигнала на выходе обратного фильтра. Наконец, на рис. 8.26*д* представлена нормированная автокорреляционная функция входного сигнала, на которой хорошо виден период основного тона длительностью 8 мс.

Линейное предсказание в алгоритме SIFT используется для выравнивания спектра с целью облегчения оценивания основного тона. Метод дает весьма точные оценки периода основного тона до тех пор, пока спектр сигнала выравнивается достаточно хорошо. Однако для голосов с малым периодом основного тона (например, детских) этот метод выравнивания спектра приводит к плохим результатам

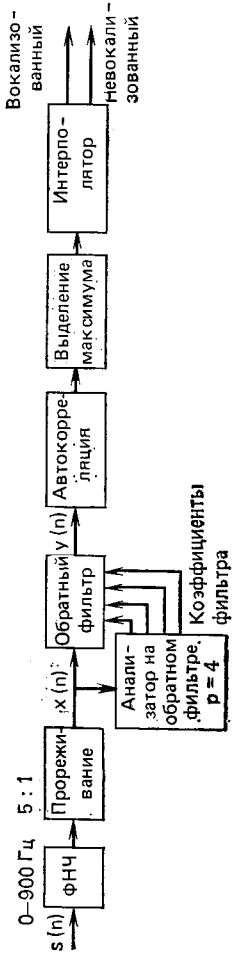


Рис. 8.25. Структурная схема алгоритма SIFT для выделения основного тона

[1] Сигнал на выходе фильтра — это погрешность предсказания для предсказателя четвертого порядка.

из-за отсутствия высших гармоник основного тона в полосе от 0 до 900 Гц (особенно при использовании сигналов с телефонных ли-

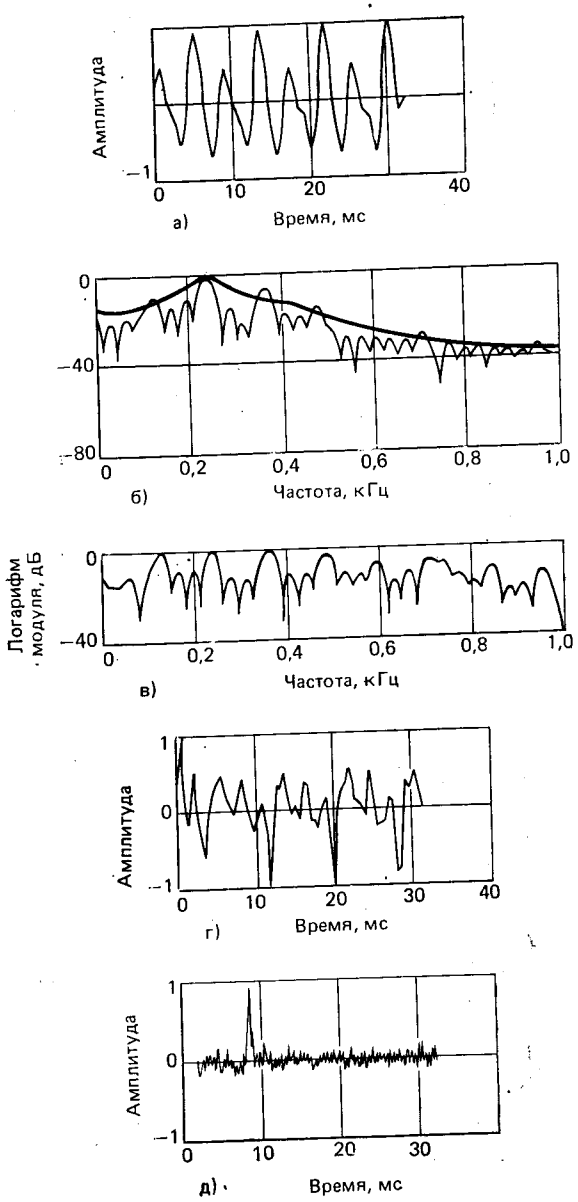


Рис. 8.26. Типичные сигналы алгоритма SIFT [19]

ний). Для таких дикторов и условий передачи лучшие результаты могут дать другие методы выделения основного тона.

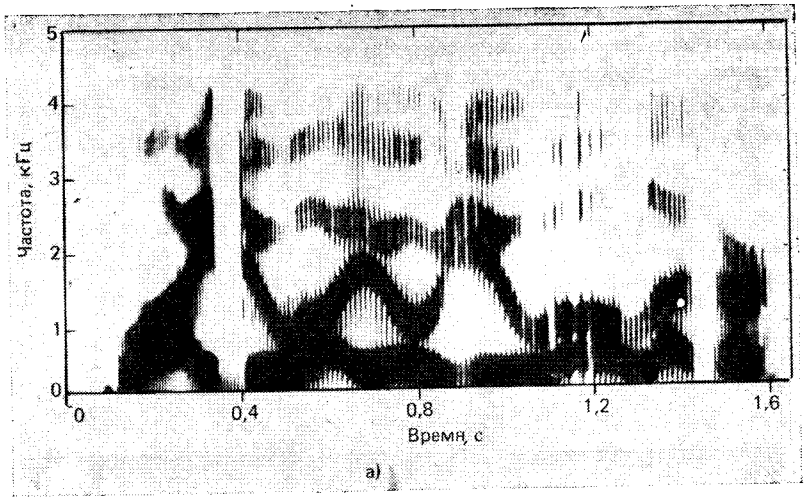
8.10.2. Формантный анализ с использованием коэффициентов линейного предсказания

Анализ речи на основе линейного предсказания при использовании его доли оценивания формантных частот вокализованного сигнала имеет как преимущества, так и недостатки. Форманты можно оценить по коэффициентам предсказания двумя способами. Первый состоит в факторизации полинома предсказания на основе полученных корней и вынесении решения о том, какие из корней описывают форманты, а какие — форму спектра [21, 22]. Другой способ заключается в оценивании спектра и использовании метода выделения максимумов, рассмотренного в гл. 7 [23].

Особое преимущество, присущее методу линейного предсказания в формантном анализе, состоит в том, что как центральные частоты формант, так и их полосы можно оценивать достаточно точно с помощью факторизации полинома предсказателя. Поскольку порядок полинома p выбирается заранее, количество комплексно-сопряженных полюсов составляет $p/2$. Таким образом, упомянутая выше проблема классификации корней полинома с целью определения того, какие из корней описывают форманты, в данном случае оказывается значительно менее сложной, чем при использовании сходных методов, например кепстрального сглаживания. Кроме того, побочные полюсы легко устраняются вследствие того, что полоса соответствующих им формант оказывается во много раз больше, чем можно ожидать для обычного речевого сигнала. На рис. 8.27 показан пример, показывающий, что положение полюсов в самом деле дает хорошее представление о формантных частотах [3].

Недостатком метода линейного предсказания является использование для описания спектра сигнала полюсной модели. Так, хотя для носовых звуков и получается неплохое описание спектра полюсной моделью, совпадение корней полинома и действительных формантных частот неочевидно. Совершенно неясно, чему соответствуют получающиеся корни: нулям и полюсам носовой полости или искомым резонансным частотам. Другая трудность заключается в том, что хотя оценки ширины формант можно определить с использованием полученных корней, однако непонятно, как они соотносятся с истинными формантами. Это объясняется тем, что полученная оценка зависит от расположения и длительности интервала анализа и метода анализа.

С учетом этих достоинств и недостатков предложен ряд методов оценивания формантных частот с использованием линейного предсказания как на основе метода выделения максимумов в спектре, так и на основе факторизации полинома предсказателя. После выбора совокупности формантных параметров устанавливается соответствие между формантными параметрами и номерами формант, как и во всех других методах анализа. Сюда входят и требования непрерывности формант, необходимости предыскажений для исключения взаимного поглощения одной фор-



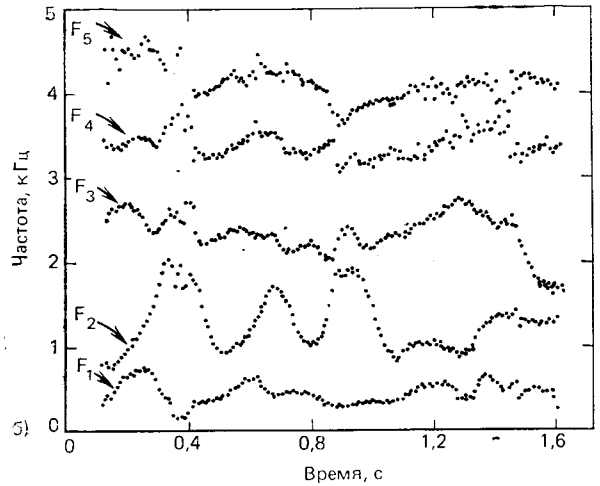


Рис. 8.27. Спектрограмма исходного сигнала (*а*) и центральные частоты (*б*) расположения комплексных полюсов 12-полюсной модели линейного предсказания [3]

манты другой и использования методов обострения пиков в спектре с помощью перемещения старшего параметра линейного предсказания к границе единичной окружности. Обсуждение различных методов содержится в работах Маркела [21, 22], Атала [3], Маххоула и Вольфа [5] и Мак-Кандлесса [23].

8.10.3. Вокодер на основе линейного предсказания

Наиболее важными областями применения линейного предсказания являются низкоскоростная передача речи (вокодеры) и ее хранение (для систем с машинным речевым ответом). На рис. 8.28

представлена структурная схема вокодера, построенного на основе линейного предсказания. Вокодер состоит из передатчика, канала связи и приемника. В передатчике вычисляются коэффициенты линейного предсказания и основной тон, которые затем

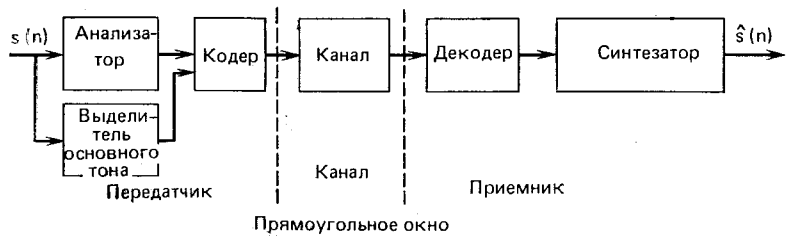


Рис. 8.28. Структурная схема вокодера с линейным предсказанием

кодируются для передачи по каналу связи. В приемнике происходят декодирование параметров и синтезирование выходного речевого сигнала. Выше уже рассматривались вопросы как анализа, так и синтеза сигнала. Для простоты предположим, что канал не вносит ошибок в передаваемое сообщение. Таким образом, в данном разделе рассматриваются различные множества параметров с точки зрения их пригодности для кодирования речи при заданной скорости передачи.

Основными параметрами являются: p коэффициентов линейного предсказания, период основного тона, признак тон—шум и коэффициент усиления. Методы подходящего кодирования периода основного тона, признака тон—шум и коэффициента усиления достаточно хорошо известны. Так, для представления периода основного тона необходимо 6 бит, для признака тон—шум — 1 бит, а для коэффициента усиления — 5 бит при логарифмическом квантовании [3].

Хотя принципиально можно и непосредственно квантовать параметры предсказания, такой подход из-за условий устойчивости требует относительно высокой точности представления (8—10 бит на параметр). Это связано с тем, что малые изменения параметров предсказания приводят к большим изменениям в расположении полюсов. Поэтому непосредственное квантование параметров предсказания не находит широкого применения.

Естественно, возникает вопрос выбора подходящей совокупности параметров, удобных для кодирования и передачи. Среди известных параметров наиболее подходящими являются корни полинома и коэффициенты отражения. Корни полинома можно квантовать таким образом, чтобы обеспечить устойчивость системы: расположение корней внутри единичного круга гарантирует устойчивость. Используя этот подход, Атал показал, что необходимо 5 бит на корень (т. е. 5 бит на полосу и 5 бит на центральную частоту). При этом синтетическая речь практически не отличима от синтетической речи, полученной без квантования параметров.

Используя такой метод кодирования, получаем, что скорость передачи составляет $72F_s$ бит/с, где F_s — количество интервалов анализа в секунду. Обычно значение F_s составляет 100, 67 и 33, что дает скорости 7200, 4800 и 2400 бит/с соответственно.

Другими параметрами, которые легко квантовать и для которых просто проверяется условие устойчивости, являются частные корреляции k_i. В данном случае условие устойчивости легко обеспечить при квантовании параметров. Макхоул и Висвансан [25] показали, что распределения частных корреляций весьма асимметричны, поэтому для правильного распределения фиксированного количества двоичных единиц необходимо предварительное преобразование параметров перед квантованием. Используя меру спектральной чувствительности, Макхоул и Висвансан [25] определили оптимальное преобразование вида

$$g_i = f(k_i) = \log\left[\frac{1-k_i}{1+k_i}\right] = \log\left[\frac{A_{i+1}}{A_i}\right], \quad 1 \leqslant i \leqslant p, \qquad (8.136)$$

где A_i — функции площади поперечного сечения неоднородной акустической трубы без потерь. Таким образом, оптимальными параметрами при кодировании речевого сигнала являются логарифмы отношений площадей поперечного сечения в рамках модели неоднородной трубы без потерь. Легко видеть, что соотношение (8.136) отображает интервал $-1 \leqslant k_i \leqslant 1$ в интервал $-\infty \leqslant g_i \leqslant \infty$. Используя это преобразование, Атал [27] показал, что коэффициенты g_i имеют почти равномерное распределение и малую корреляцию между параметрами и, таким образом, являются весьма удобными для цифрового представления. При использовании этих параметров для кодирования речи требуется 5—6 бит на параметр для получения такого же качества синтезированного сигнала, как и без квантования.

Во всех перечисленных выше случаях предполагалось, что для кодирования параметров используется один из методов ИКМ. В [26] показано, что использование методов кодирования различных параметров при линейном предсказании (см. гл. 5) позволяет дополнительно уменьшить скорость передачи сигнала. Используя АРИКМ при кодировании параметров предсказания, можно получить хорошее восприятие речи при скоростях 1000—2000 бит [26].

8.10.4. Полувокодер с линейным предсказанием [1]

Ранее было показано, что наиболее слабым звеном всех известных вокодеров является необходимость точной оценки источника возбуждения. В гл. 6 рассматривались некоторые вокодерные системы, которые не требовали непосредственной оценки основного тона и признака тон—шум, но описывали возбуждение через фазу или производную фазы сигнала. Другой подход, позволяющий избежать непосредственной оценки параметров возбуждения, состоит в использовании вокодеров, возбуждаемых речевым сигналом. Системы такого типа исследовались Аталом и др. [26] и Винстеном [27].

[1] Имеется в виду вокодер, в котором в качестве сигнала возбуждения применяется преобразованный речевой сигнал. (Прим. ред.)

На рис. 8.29 представлена структурная схема вокодера с речевым возбуждением. В этой системе имеются два отдельных подканала передачи, один из которых используется для передачи

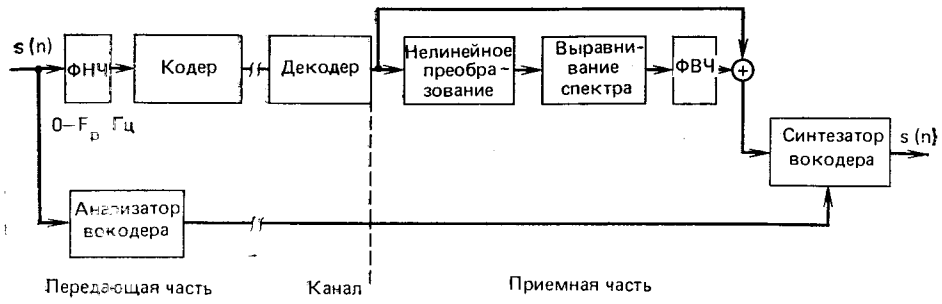


Рис. 8.29. Структурная схема вокодера, возбуждаемого речью (полувокодер)

узкополосного речевого сигнала, а другой — для передачи параметров обычного вокодера (например, коэффициентов предсказания, огибающей спектра и т. д.). Узкополосный сигнал, который при передаче можно закодировать любым из методов гл. 5, используется в синтезаторе для образования сигнала возбуждения путем соответствующих нелинейных преобразований и выравнивания спектра. Причина высокой эффективности данного метода заключается в том, что низкочастотная часть сигнала содержит всю необходимую информацию о возбуждении, т. е. она синхронна с точным периодом основного тона при периодическом возбуждении и шумоподобна в других случаях.

При использовании такого метода можно избежать оценивания основного тона и признака тон—шум. Однако в данном случае по каналу передается дополнительная информация, необходимая для описания низкочастотной части спектра, поэтому вокодеры с возбуждением речевым сигналом требуют более высоких скоростей передачи, чем обычные вокодеры. Так, например, при использовании речевого возбуждения скорость передачи составляет около 3000—4000 бит/с, т. е. на 1000—2000 бит/с больше, чем в обычных вокодерах. Выигрыш, получаемый за счет увеличения скорости передачи, сводится к меньшей зависимости качества передачи от замены диктора или изменений условий передачи. Это объясняется отсутствием устройств выделения основного тона и классификаторов тон—шум. Более детальное обсуждение вокодера, возбуждаемого речевым сигналом, можно найти в [27], [28].

8.11. Заключение

В данной главе рассматривались методы линейного предсказания речи. Основное внимание уделялось подходам, позволяющим в наибольшей мере понять процессы речеобразования. Были рассмотрены некоторые аспекты применения этих методов, а также предпринята попытка выявить там, где это возможно, сходства и различия между основными методами обработки сигналов.

Задачи

8.1. Рассмотрим разностное уравнение

$$h(n) = \sum_{k=1}^{p} \alpha_k h(n-k) + G\delta(n).$$

Автокорреляционная функция $h(n)$ определяется как

$$\tilde{R}(m) = \sum_{n=0}^{\infty} h(n)\, h(n+m).$$

а) Показать, что $\tilde{R}(m) = \tilde{R}(-m)$.
б) Подстановкой разностного уравнения в $\tilde{R}(-m)$ показать, что

$$\tilde{R}(m) = \sum_{k=1}^{p} \alpha_k \tilde{R}(|m-k|), \quad m = 1, 2, \ldots, p.$$

8.2. Системная функция $H(z)$, вычисленная в N равноотстоящих точках единичной окружности, есть

$$H\left(e^{i\frac{2\pi}{N}k}\right) = \frac{G}{1 - \sum_{n=1}^{p} \alpha_n e^{-i\frac{2\pi}{N}kn}}, \quad 0 \leqslant k \leqslant N-1.$$

Описать процедуру использования алгоритма БПФ для вычисления $H\left(e^{i\frac{2\pi}{N}k}\right)$

8.3. Уравнение (8.30) можно использовать для сокращения объема вычислений, необходимых для получения ковариационной матрицы в ковариационном методе.

а) Используя определение $\varphi_n(i, k)$ в ковариационном методе, показать, что $\varphi_n(i+1, k+1) = \varphi_n(i, k) + s_n(-i-1)s_n(-k-1) - s_n(N-1-i)s_n(N-1-k)$, считая, что $\varphi_n(i, 0)$ вычислено для $i = 0, 1, 2, \ldots, p$.

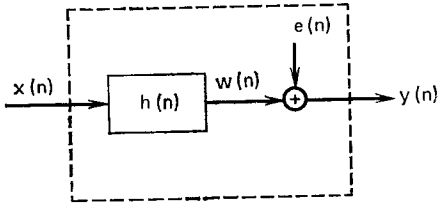


Рис. 3.8.1

б) Показать, что элементы на главной диагонали можно вычислить, основываясь на $\varphi_n(0, 0)$, т. е. получить рекурсивную формулу для $\varphi_n(i, i)$.

в) Показать, что элементы на нижних диагоналях также вычисляются рекурсивно, начиная с $\varphi_n(i, 0)$.

г) Как получить элементы на верхних диагоналях?

8.4. Линейное предсказание можно рассматривать как оптимальный метод оценивания линейной системы, основанный на ряде предположений. На рис. 3.8.1 представлен другой способ оценки параметров линейной системы. Предположим, что наблюдению доступен как $x(n)$, так и $y(n)$ и что $e(n)$ — белый гауссовский шум с нулевым средним и дисперсией σ^2_e, статистически не связанный с $x(n)$. Оценка импульсной характеристики линейной системы должна быть такой, чтобы минимизировать квадрат ошибки $\varepsilon = E[(y(n) - \hat{h}(n) * x(n))^2]$, где $\hat{h}(n)$, $0 \leqslant n \leqslant M-1$ — оценка $h(n)$.

а) Определить систему линейных уравнений относительно $\hat{h}(n)$ через автокорреляционную функцию $x(n)$ и взаимно-корреляционную функцию между $y(n)$ и $x(n)$.

б) Как решить систему уравнений, полученную в п. а)? Как соотнести полученную систему с методом линейного предсказания, рассмотренным в данной главе?

в) Получить выражение для ε-минимального среднего квадрата ошибки.

8.5. При выводе лестничного алгоритма фильтр погрешности i-го порядка определялся как

$$A^{(i)}(z) = 1 - \sum_{k=1}^{i} \alpha_k^{(i)} z^{-k}.$$

Коэффициенты предсказания удовлетворяют соотношениям (8.131). Подставляя выражения $\alpha_j^{(i)}$, $1 \leqslant j \leqslant i$, в выражение для $A^{(i)}(z)$, получить

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}).$$

8.6. Дан отрезок речевого сигнала $s(n)$, который имеет период N_p отсчетов, при этом $s(n)$ можно представить в виде дискретного преобразования Фурье

$$s(n) = \sum_{k=1}^{M} \left(\beta_k e^{i\frac{2\pi}{N_p}kn} + \beta_k^* e^{i\frac{2\pi}{N_p}kn} \right),$$

где M — число имеющихся гармоник основной частоты $(2\pi/N_p)$. С целью спектрального выравнивания сигнала (для выделения основного тона) запишем сигнал $y(n)$ в виде

$$y(n) = \sum_{k=1}^{M} \left(e^{i\frac{2\pi}{N_p}kn} + e^{-i\frac{2\pi}{N_p}kn} \right).$$

Эта задача связана с процедурой выравнивания спектра сигнала с использованием комбинированного метода, основанного на линейном предсказании и гомоморфной обработке.

а) Показать, что выравненный по спектру сигнал можно выразить в виде

$$y(n) = \frac{\sin\left[\frac{\pi}{N_p}(2M+1)n\right]}{\sin\left[\frac{\pi}{N_p}n\right]} - 1.$$

(Отметим, что эта последовательность изображена на рис. 6.20 для $N_p = 15$ и $M = 2$.)

Теперь предположим, что линейное предсказание сделано по сигналу $s(n)$ с использованием окна длиной в несколько периодов основного тона, а значение p в анализе таково, что $p = 2M$. При этом получена системная функция вида

$$H(z) = 1/(1 - \sum_{k=1}^{p} \alpha_k z^{-k}) = 1/A(z).$$

Знаменатель можно представить в виде $A(z) = \prod_{k=1}^{(p)} (1 - z_k z^{-1})$.

б) Как связаны $p = 2M$ полюсов в $A(z)$ с частотами, представленными в сигнале?

Кепстр $\hat{h}(n)$ импульсной характеристики $h(n)$ определяется как последовательность, z-преобразование которой имеет вид $\hat{H}(z)=\log H(z)=-\log A(z)$. Отметим, что $\hat{h}(n)$ можно вычислить по α_k, используя (8.123). Показать, что $\hat{h}(n)$ связано с нулями $A(z)$ соотношением

$$\hat{h}(n)=\sum_{k=1}^{p}\frac{z_k^n}{n},\quad n>0.$$

в) Используя результаты пп. а) и б), доказать, что $y(n)=n\hat{h}(n)$ является сигналом с выравненным спектром, таким, как это необходимо для выделения основного тона.

8.7. «Стандартный» метод вычисления кратковременного спектра для сегмента речевого сигнала показан на рис. 3.8.2*а*. Более сложный метод, требующий больше вычислений для получения $\log|X(e^{i\omega})|$, показан на рис. 3.8.2*б*.

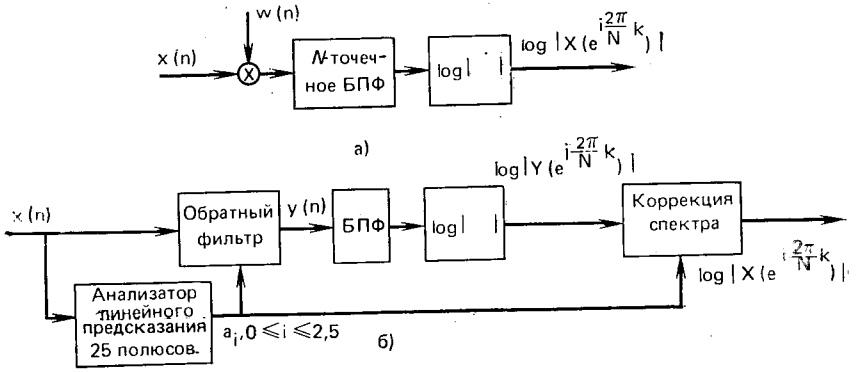


Рис. 3.8.2

а) Рассмотрите новый метод вычисления спектра и объясните функции устройства коррекции спектра.

б) В чем возможные преимущества нового метода? Рассмотрите использование окон, присутствие нулей в спектре и т. д.

8.8. Предлагается метод определения основного тона на основе линейного предсказания с использованием автокорреляционной функции погрешности предсказания $e(n)$. Вспомним, что $e(n)$ можно представить в виде

$$e(n)=\hat{s}(n)-\sum_{i=1}^{p}\alpha_i\hat{s}(n-i),$$

и если обозначить $\alpha_0=-1$, тогда

$$e(n)=-\sum_{i=0}^{p}\alpha_i\hat{s}(n-i),$$

где сигнал взвешивается с окном $\hat{s}(n)=s(n)w(n)$ и отличен от нуля на интервале $0\leqslant n\leqslant N-1$.

а) Показать, что автокорреляционная функция $e(n)$, $R_e(m)$, может быть выражена в виде

$$R_e(m)=\sum_{l=-\infty}^{\infty}R_a(l)\,R_{\hat{s}}(m-l),$$

где $R_a(l)$ — автокорреляционная функция параметров предсказания: $R_{\hat{s}}(l)$ — автокорреляционная функция сигнала $\hat{s}(n)$.

б) Определить число сложений и умножений для вычисления, если частота дискретизации равна 10 кГц, а $R_e(m)$ заключено в интервале от 3 до 15 мс?

8.9. В этой книге рассматривался ряд вокодеров: канальный последовательный формантный, параллельный формантный, гомоморфный, фазовый и вокодер на основе линейного предсказания. Чисто теоретически упорядочите эти вокодеры по качеству сигнала. Объясните подробно полученную очередность. При обсуждении следует рассмотреть вопросы используемой модели, теряемую при анализе информацию, необходимость слежения за основным тоном и т. д.

8.10. Предположим, что два диктора пытаются установить связь, используя вокодеры различного типа, как это показано на рис. 3.8.3. У диктора *1* имеются анализатор вокодера с линейным предсказанием типа рассмотренного в 8.3.10 и синтезатор в прямой форме, который описан в § 8.9. Диктор *2* располагает гомоморфным вокодером, обсуждавшимся в § 7.5.

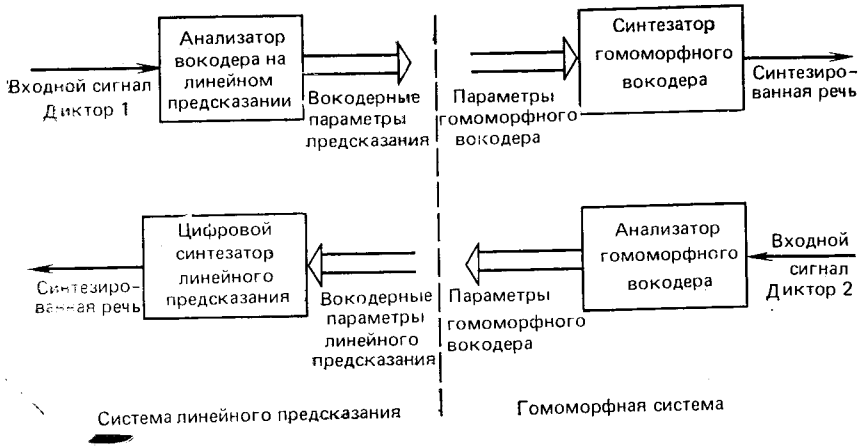


Рис. 3.8.3

а) Чтобы диктор *1* мог связаться с диктором *2*, необходимо преобразовать описание сигнала на основе линейного предсказания в гомоморфное описание для синтеза речи с помощью гомоморфного синтезатора. Придумать метод такого преобразования.

б) Придумать метод преобразования гомоморфного описания в описание на основе линейного предсказания с тем, чтобы диктор *2* мог установить связь с диктором *1*.

8.11. Рассмотрим два взвешенных сегмента речи $x(n)$ и $\hat{x}(n)$, определенные на интервале $0\leqslant n\leqslant N-1$ (вне этого интервала оба сегмента равны нулю). Осуществим анализ на основе линейного предсказания на каждом из сегментов. Таким образом, получим автокорреляционные функции, определяемые как

$$R(k)=\sum_{n=0}^{N-1-k}x(n)\,x(n+k),\quad 0\leqslant k\leqslant p;$$

$$\hat{R}(k)=\sum_{n=0}^{N-1-k}\hat{x}(n)\,\hat{x}(n+k),\quad 0\leqslant k\leqslant p.$$

На основе автокорреляционных функций найдем параметры предсказания $\boldsymbol{\alpha}=(\alpha_0,\alpha_1,\ldots,\alpha_p)$ и $\hat{\boldsymbol{\alpha}}=(\hat{\alpha}_0,\ldots,\hat{\alpha}_1,\hat{\alpha}_p)$, $(\alpha_0=\hat{\alpha}_0=1)$.

а) Показать, что погрешность предсказания

$$E^{(p)} = \sum_{n=0}^{N-1+p} e^2(n) = \sum_{n=0}^{N-1+p} \left[-\sum_{i=0}^{p} \alpha_i x(n-i) \right]^2$$

может быть записана в виде $E^{(p)} = \boldsymbol{\alpha} \mathbf{R}_{\alpha} \boldsymbol{\alpha}^t$, где $\mathbf{R}_{\alpha}$ — матрица $(p+1) \times (p+1)$. Определить $\mathbf{R}_{\alpha}$.

б) Предположим, что сигнал $\hat{x}(n)$ пропущен через обратный фильтр с коэффициентами α, что дает погрешность предсказания $\hat{e}(n)$, определяемую выражением

$$\tilde{e}(n) = -\sum_{i=0}^{p} \alpha_i \hat{x}(n-i).$$

Показать, что средняя квадратическая ошибка $\tilde{E}^{(p)}$, определяемая как $\tilde{E}^{(p)} = \sum_{n=0}^{N-1+p} [\tilde{e}(n)]^2$, может быть записана в виде $\tilde{E}^{(p)} = \mathbf{d} \underset{\alpha}{\mathbf{R}} \boldsymbol{\alpha}^t$, где $\underset{\alpha}{\mathbf{R}}$ — матрица $(p+1) \times (p+1)$. Определить $\underset{\alpha}{\mathbf{R}}$.

в) Если определить отношение $D = \tilde{E}^{(p)}/E^{(p)}$, то что можно сказать о диапазоне значения D?

8.12. Предложена следующая мера различимости между двумя сегментами речевого сигнала с параметрами предсказания $\boldsymbol{\alpha}$ и $\hat{\boldsymbol{\alpha}}$ и корреляционными матрицами $\mathbf{R}_{\alpha}$ и $\underset{\alpha}{\mathbf{R}}$ (см. задачу 8.11):

$$D(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = \frac{\boldsymbol{\alpha} \underset{\alpha}{\mathbf{R}} \boldsymbol{\alpha}^t}{\hat{\boldsymbol{\alpha}} \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t}.$$

а) Показать, что мера различимости $D(\alpha, \hat{\alpha})$ может быть записана в следующей удобной для вычислений форме:

$$D(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = \left[\frac{\left(b(0) \hat{R}(0) + 2 \sum_{i=1}^{p} b(i) \hat{R}(i) \right)}{\hat{\boldsymbol{\alpha}} \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t} \right],$$

где $b(i)$ — автокорреляционная функция вектора $\boldsymbol{\alpha}$ — равна:

$$b(i) = \sum_{j=0}^{p-i} \alpha_j \alpha_{j+i}, \quad 0 \leqslant i \leqslant p.$$

б) Предположим, что величины (вектора, матрицы, скаляры) $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, $\mathbf{R}$, $\hat{\mathbf{R}}$, $(\alpha \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t)$, $\underset{\alpha}{\mathbf{R}}$ и $\mathbf{b}$ вычислены заранее, т. е. известны к моменту расчета меры расстояния. Сравнить объем вычислений, необходимый для определения $D(\alpha, \hat{\alpha})$, используя оба выражения для D, рассмотренные в данной задаче.

9

Цифровая обработка речи в системах речевого общения человека с машиной[1]

9.0. Введение

В предыдущих главах внимание было сконцентрировано на основных теоретических вопросах, необходимых для понимания современных методов цифровой обработки речевых сигналов. Еще не рассматривалась обширная область применения разработанных методов, т. е. способов использования моделей и связанных с ними параметров в системах передачи или автоматического выделения информации из сигнала речи. В данной главе приведены характерные примеры цифровой обработки речи применительно к системам общения между человеком и машиной (ЭВМ) посредством голоса. Существует ряд причин, по которым имеет смысл ограничиться рассмотрением примеров связи между человеком и машиной. Прежде всего, эта область наиболее плодотворна с точки зрения возможностей использования методов цифровой обработки речи и позволяет, таким образом, проиллюстрировать почти все рассмотренные выше методы обработки. Кроме того, эта область является чрезвычайно важной, дающей все новые и новые приложения, область, которая только еще развивается и демонстрирует огромные возможности для широкого применения.

Системы речевого обмена между человеком и машиной можно подразделить на три класса: с речевым ответом, распознавания диктора и распознавания речи.

Системы с речевым ответом предназначаются для выдачи информации пользователю в форме речевого сообщения. Таким образом, системы с речевым ответом — это системы односторонней связи, т. е. от машины к человеку. С другой стороны, системы второго и третьего классов — это системы связи от человека к машине. В системах распознавания диктора задача состоит в верификации диктора (т. е. в решении задачи о принадлежности данного диктора к некоторой группе лиц) или идентификации диктора из некоторого известного множества. Таким образом, класс задач распознавания диктора распадается на два подкласса: верификации и идентификации говорящего. Различия и сходство между этими задачами будут рассмотрены в последующем.

Последний класс задач распознавания речи также можно разделить на подклассы в зависимости от таких факторов, как размер словаря, количество дикторов, условия произнесения слов и т. д. Основная задача распознающей системы сводится либо к точному распознаванию произнесенной на входе фразы (т. е. система фонетической или орфографической печати произнесенного текста), либо к «пониманию» произнесенной фразы (т. е. к правильной реакции на сказанное диктором). Именно задача понимания, а не распознавания наиболее важна для систем с достаточно большим словарем непрерывных речевых сигналов, в то время как задача точного распознавания более важна для систем с ограниченным словарем, малым количеством дикторов, систем распознавания изолированных слов. Различные аспекты построения систем распознавания речи также рассматриваются в данной главе.

В заключительной части главы рассмотрены некоторые системы, типичные для каждой области речевого общения человека с машиной. Более подробно рассматриваются особенности обработки речевого сигнала с целью закрепления результатов предшествующих глав. Однако как для полноты обсуждения, так и для более глубокого понимания здесь излагаются и общие аспекты обработки информации в системе, поскольку часто они оказываются достаточно важными для успешной работы системы в целом.

[1] Имеются в виду цифровые ЭВМ. (Прим. ред.)

9.1. Системы с речевым ответом

На рис. 9.1 представлена общая структурная схема системы с речевым ответом на базе ЭВМ. Элементами этой системы являются блоки: памяти для хранения словаря системы с речевым ответом; хранения правил синтеза сообщений по элементам словаря; программ формирования речевого ответа.

На вход системы с речевым ответом поступает сообщение о содержании вопроса, порождаемого либо другой системой обработки информации, либо непосредственно от человека, обратившегося с интересующим его вопросом к информационной системе.

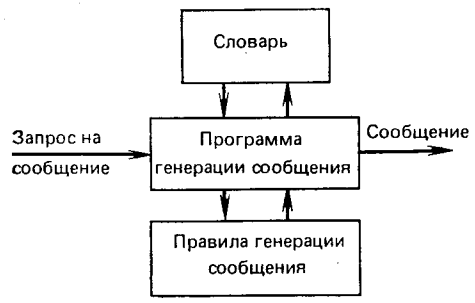


Рис. 9.1. Структурная схема системы с речевым ответом

Откликом системы на поставленный вопрос служит выходное сообщение в виде речевой фразы. Простым примером такой системы является автоматическая справочная телефонная служба, которая обнаруживает неправильно набранный номер, определяет причину ошибки (например, телефон отключен или ему присвоен новый номер и т. д.) и посылает на выход системы с речевым ответом сообщение, содержащее необходимые абоненту указания. В таких системах словарь обычно состоит из ограниченного набора изолированных слов (например, цифр с различными окончаниями).

В качестве другого примера рассмотрим информационную систему о состоянии курса акций. Здесь абонент должен с помощью кнопочного набора ввести код интересующего его курса. Система декодирует набор, определяет текущий курс акций и затем выдает соответствующую информацию в систему с речевым ответом для составления требуемой фразы. В данном случае словарь должен содержать достаточно широкий набор различных слов и фраз.

Существуют два основных подхода к построению систем с речевым ответом. Один из них заключается в попытке построения системы, речевые возможности которой сравнимы с возможностями человека. Такие системы (называемые часто системами синтеза речи по правилам) основаны на модели речеобразования, рассмотренной в гл. 3. В этом случае для синтеза достаточно хранить словарь произношений элементов. Сигналы, необходимые для управления речевым синтезатором, в соответствии с моделью речеобразования формируются на основе правил синтеза. Такие системы представляют интерес в том случае, если требуется словарь весьма большого объема. Реализация подобных систем — это проблема, требующая чрезвычайно трудоемких исследований, и на этапе синтеза сигнала имеются обширные возможности применения рассмотренных выше методов цифровой обработки сигналов. Однако

основная трудность при построении подобных систем состоит в разработке правил управления синтезатором. В данной книге примеры таких систем не рассматриваются, поскольку это увело бы нас в область лингвистики. Интересующихся читателей отсылаем к работам [2—6].

В системах с речевым ответом второго типа используется ограниченный словарь и сигнал на выходе таких систем формируется посредством сочленения отдельных элементов реального речевого сигнала, взятых из словаря. На рис. 9.2 представлена структурная схема системы, в которой словарь, состоящий из отдельных

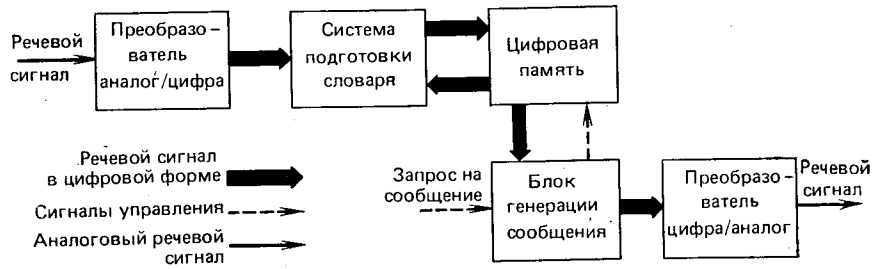


Рис. 9.2. Структурная схема системы с речевым ответом [8]

слов, представленных в цифровой форме, хранится в памяти. Сообщения конструируются в этом случае путем отыскания требуемых слов и фраз в памяти и воспроизведения их в требуемой последовательности. При разработке систем подобного типа следует учитывать три основных соображения. Во-первых, способ представления и хранения словаря должен быть выбран таким образом, чтобы в разработанной системе имелась возможность свободного доступа к любому элементу словаря. Во-вторых, должен быть выбран способ редактирования речевого материала словаря совместно со способом записи его элементов в память. В-третьих, необходимо обеспечить заданную последовательность выбора и воспроизведения элементов словаря (т. е. способ формирования сообщения).

Поскольку назначение систем с речевым ответом состоит в формировании речевых сообщений, предназначенных для человека, требование к разборчивости становится определяющим. Не менее важное значение, однако, имеют и такие параметры речи, как качество восприятия и натуральность. Таким образом, в разрабатываемой системе необходимо с предельной полнотой реализовать все три основных условия с тем, чтобы добиться максимально возможной разборчивости и натуральности речевого сигнала.

9.1.1. Основные аспекты построения систем с речевым ответом

Развитие методов цифрового представления и цифровой обработки сигналов, а также методов построения цифровых устройств

позволяет создавать системы речевого ответа, выполненные полностью на базе цифровой техники. В показанной на рис. 9.2 цифровой системе необходимо, прежде всего, осуществить аналого-цифровое преобразование, т. е. представить речевой сигнал в цифровой форме. Аналогично для преобразования цифрового представления в аналоговую форму требуется цифроаналоговый преобразователь. Поскольку словарь представлен в цифровой форме, его можно хранить в цифровой памяти. Для доступа к элементам словаря в нужной последовательности и составления из них требуемой фразы необходима система формирования сообщений. Полученное цифровое представление синтезированной фразы, в свою очередь, поступает на цифроаналоговый преобразователь.

Центральным фактором, определяющим сложность систем с речевым ответом, является выбор способа цифрового представления речи при составлении словаря. Как будет ясно из дальнейшего, здесь имеются широкие возможности использования различных способов цифрового представления, начиная со способов кодирования речевых колебаний (см. гл. 5) и кончая системами «анализ—синтез» (см. гл. 6—8). Выбор способа цифрового представления оказывает большое влияние на объем и тип цифровой памяти, а также на способ синтеза речевого сообщения.

При рассмотрении способа цифрового представления речевого сигнала применительно к системам с речевым ответом полезно остановиться на трех основных моментах:

скорость передачи информации (в битах в секунду), необходимая для получения приемлемого качества;

сложность способа кодирования и декодирования;

гибкость представления, т. е. возможность модификации элементов словаря.

На рис. 9.3 показаны результаты сравнительного анализа методов цифрового представления, рассмотренных в гл. 5—8, по трем

Рис. 9.3. Обзор методов кодирования речи

перечисленным выше показателям. Представление на основе кодирования речевого колебания требует наибольших скоростей передачи и, следовательно, максимального объема памяти для хранения элементов словаря. Эти способы являются простейшими с точки зрения алгоритмов кодирования—декодирования. С другой стороны, способы анализа—синтеза, которые буквально «разбирают речевой сигнал на части», обладают широкими возможностями полезной модификации элементов словаря. Два первых фактора, т. е. скорость передачи и сложность реализации, оказывают существенное влияние на технико-экономические показатели при разработке полностью цифровых систем речевого ответа. Рассмотрим типовой словарь, объем которого составляет 100 слов со средней протяженностью 1 с. В табл. 9.1 показаны оценки (весьма осторожные) объема памяти, необходимого для хранения произноси-

Таблица 9.1

Объем памяти, необходимой для хранения цифрового представления речи

Метод кодирования	Скорость, кбит/с	Объем памяти, бит	Примерная стоимость, долл.
ИКМ	40	4 000 000	4000
АРИКМ	24	2 400 000	2400
Линейное предсказание	2,4	240 000	240
Форманты	0,6	60 000	60

мой в течение 100 с речи. Даже при использовании для кодирования логарифмической ИКМ объем памяти остается вполне приемлемым. Основной вопрос заключается в соотношении между стоимостями цифровой памяти и аппаратуры кодирования—декодирования для наиболее сложных систем кодирования. В последнем столбце табл. 9.1 приведена стоимость памяти системы речевого ответа из расчета 0,1 цента за бит[1]. С учетом того, что один блок памяти может быть использован для ряда каналов, стоимость устройства кодирования—декодирования должна быть достаточно малой с тем, чтобы не являться определяющей частью общей стоимости системы. Совершенно очевидно, например, что стоимость формантного синтезатора, необходимого для высококачественного воспроизведения речевого сигнала, окажется значительно большей, чем стоимость соответствующего блока памяти для словаря малого объема.

Другой важной задачей, решаемой при построении систем с речевым ответом, являются создание и редактирование словаря. При решении этой задачи, т. е. подготовке элементов словаря и обеспечении высококачественного сигнала на выходе, цифровые методы оказываются чрезвычайно эффективными и гибкими. Обычно слова и фразы, включаемые в словарь, произносятся специально обученным диктором и записываются с высоким качеством. Затем слова или фразы подвергаются аналого-цифровому преобразова-

[1] Это довольно грубая верхняя граница действительной стоимости одного бита памяти.

нию и кодированию. Цифровое представление (которое может быть как описанием формы сигнала, так и основанным на представлении типа «анализ—синтез») оперативно хранится в цифровой форме в ЭВМ. Для исключения пауз между фразами используется специальный метод поиска начала и конца фразы. Как показано в гл. 4, при высококачественной записи начало и конец каждой фразы можно определить с высокой точностью. При этом можно точно сказать, удовлетворяет ли протяженность данной фразы заданной. Фраза, кроме того, может быть воспроизведена для проверки окончаний слов или фразы на слух. Записи можно легко повторять, пока не будут достигнуты требуемые длительность и окончание вводимой фразы.

Заключительным шагом в создании словаря являются сравнение энергетических уровней всех слов в словаре и соответствующее изменение уровней для получения некоторого единого уровня или такого распределения уровней, которое предопределяется предполагаемым использованием словаря. Это может быть сделано или на основе вычисления максимального значения сигнала, или на основе использования других мер, таких, как кратковременная энергия.

Если слово или фраза записаны с требуемым качеством, то они хранятся в определенном месте памяти словаря. Это достигается простой установкой файлов в речевой системе и указанием адресов, которые используются системой синтеза фраз для определения начала и окончания каждого элемента словаря.

Помимо рассмотренных методов создания словаря система с речевым ответом включает в себя методы синтеза фраз по элементам словаря. В этом случае методы цифрового представления также обладают значительными преимуществами. Если используется метод кодирования формы речевого колебания, то все, что здесь необходимо, — это сочленить речевые сигналы элементов словаря. Если элементом словаря является отдельное слово, то такой метод может привести к некоторой потере натуральности звучания, но подобный подход обладает важным преимуществом, состоящим в том, что система синтеза фраз оказывается очень простой. В самом деле, такая система легко может быть выполнена на основе микропроцессора. Пример подобной системы рассматривается в 9.1.2.

С другой стороны, представление, основанное на преобразовании типа «анализ—синтез», обладает большой гибкостью по отношению к изменяющимся свойствам элементов словаря, например временным соотношениям, окончаниям и т. д. Это свойство является даже более важным, чем малая скорость передачи (объем описания), которую можно достигнуть при использовании описания на основе преобразования «анализ-синтез». Поскольку элементы словаря представлены в виде набора основных параметров речевого сигнала, можно, например, изменять период основного тона и длительность слов таким образом, чтобы привести их в соответствие с контекстом. Более интересной представляется воз-

можность такого изменения параметров на границах слов, чтобы добиться как можно большего сходства между синтезированными и реальными речевыми сигналами. Достигнуть такого эффекта даже в простейших случаях можно лишь на основе использования правил для определения требуемого периода основного тона и протяженности во времени, а также алгоритмов изменения параметров в соответствии с изменяющейся протяженностью слов и поглощением их границ в слитной речи. Вследствие малого объема параметрического описания словаря для получения удовлетворительного качества синтезированного сигнала системы с речевым ответом на основе преобразования «анализ—синтез» следует реализовывать аппаратурно с использованием микропроцессоров. Пример системы такого рода рассмотрен в 9.1.3.

9.1.2. Многоканальная цифровая система с речевым ответом

На рис. 9.4 показана структурная схема многоканальной цифровой системы, созданная в лабораториях Белла с применением малого спецвычислителя [7, 8]. В этой системе элементы словаря представлены с помощью АРИКМ со скоростью 24 кбит/с. Кодер и декодеры АРИКМ реализованы в виде отдельных устройств.

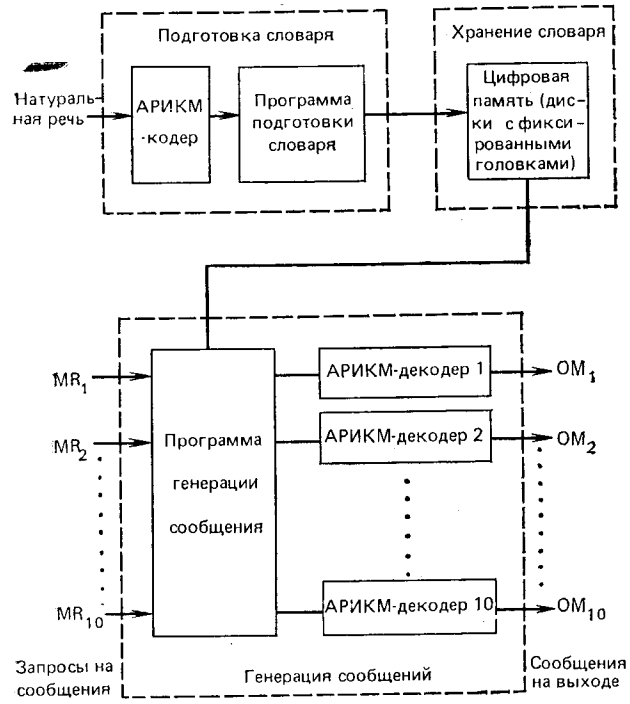


Рис. 9.4. Структурная схема многоканальной системы с речевым ответом [7]

Начало и конец каждого слова определяются автоматически с использованием алгоритма, основанного на вычислении кратковременной «энергии» АРИКМ кодовых слов [7]. Словарь хранится на диске с фиксированным положением головок для быстрого поиска элементов словаря по командам программы синтеза фраз. Эта часть системы выполняет, главным образом, логические операции и операции пересылки данных. Важной особенностью, свойственной многим вычислительным машинам и системам памяти, является возможность одновременного обслуживания ряда информационных каналов единой словарной памятью. Синтезированные в ЭВМ фразы запрашиваются другими компьютерами либо по телефонным линиям, либо непосредственно через цифровые устройства ввода—вывода. Программа синтеза сообщений определяет необходимые элементы словаря и пересылает их в память. Цифровое представление требуемых сообщений хранится в буферной памяти машины со свободным прямым доступом. Буферные устройства связаны с соответствующими АРИКМ декодерами с помощью каналов прямого доступа в память. При такой организации система обеспечивает одновременное обслуживание ряда каналов. В лабораториях Белла создана десятиканальная система с речевым ответом. Она применяется в ряде приложений, как это рассмотрено в 9.1.4.

9.1.3. Система синтеза речи на основе последовательного объединения слов, закодированных формантами

В качестве примера использования преобразования типа «анализ—синтез» рассмотрим структурную схему, представленную на рис. 9.5. В этом случае элементы словаря сформированы так, как это описано в § 7.4, где речевой сигнал представлен набором параметров, например период основного тона, признак вокализованный — невокализованный, интенсивность и формантные частоты. Таким образом, для хранения слов и фраз словаря достаточно объема памяти 600 бит, приходящейся на 1 с длительности сигнала.

Помимо элементов словаря система синтеза сообщений должна включать методы обработки, обеспечивающие получение требуемой длительности слов и периода основного тона в синтезируемой фразе. На основе этих данных формантные частоты соседних слов сглаживаются так, чтобы они непрерывно переходили одна в другую аналогично слитной речи. На рис. 9.6 показано, как использование формантного описания позволяет преобразовывать элементы словаря для достижения натуральности звучания. Длительность можно изменять посредством интерполяции. Кроме того, на стыке второго и третьего слов формантные частоты подстраиваются таким образом, чтобы их траектории не разрывались при переходе через границу, как это и должно быть в слитной речи. Заметим, наконец, что контуры основного тона каждого из элементов словаря могут быть изменены или опущены для получения единого контура основного тона, отвечающего фразе в целом.

На рис. 9.7 представлен пример [10], иллюстрирующий различие между сочленением осциллограмм речевого сигнала и сочленением слов, представленных формантами. На рис. 9.7*а* показана

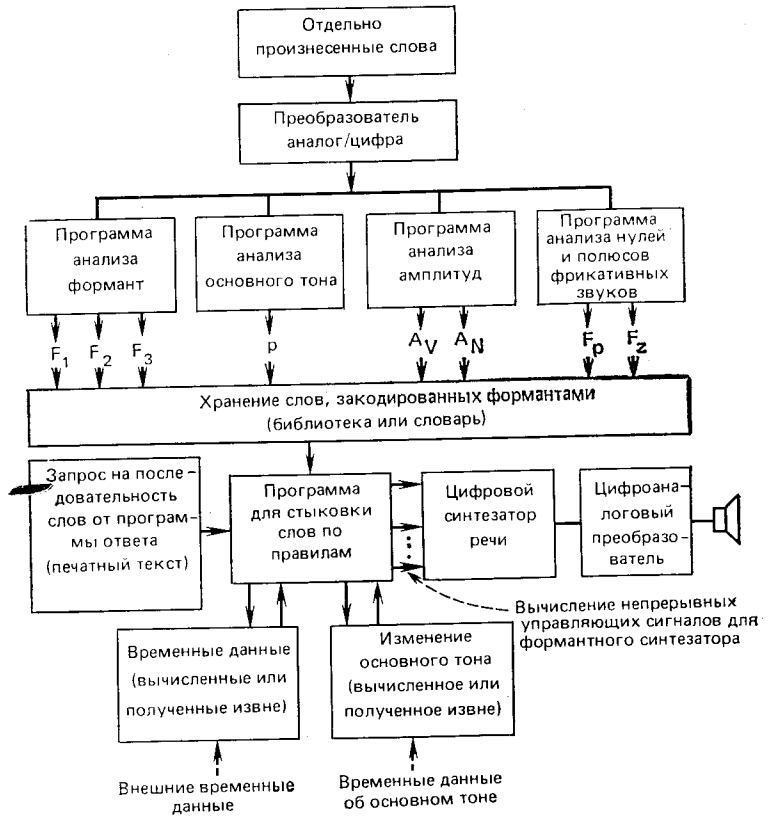


Рис. 9.5. Структурная схема системы с речевым ответом, основанная на формантном представлении [9]

спектрограмма исходной фразы «I am an aspiring orator». На рис. 9.7*в* показана спектрограмма той же фразы, полученная путем сочленения отдельно произнесенных слов, закодированных формантами без изменения периода основного тона или их длительности. В данном случае видны разрывы в формантных траекториях. На рис. 9.7*б* показана спектрограмма фразы, полученной из последовательности слов, закодированных формантами путем соответствующего согласования частот на границах слов. Контур основного тона и длительность слов, показанных на рис. 9.7*б*, рассчитаны по данным, полученным в результате исследования системы синтеза речи по правилам, разработанным Кокером и

Умедой [3, 5, 6]. Длительности слов на рис. 9.7*а* и *б* вполне соизмеримы, а соответствующие формантные траектории весьма похожи друг на друга.

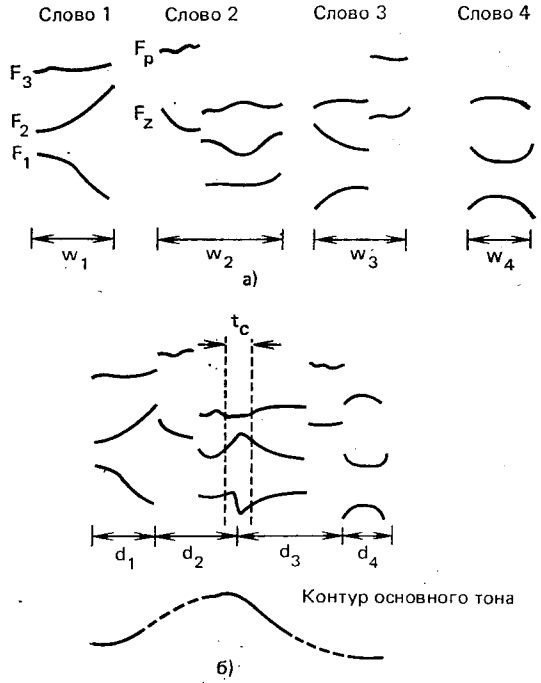


Рис. 9.6. Типичный пример, иллюстрирующий процесс генерации сигналов управления по словам из словаря. Представлен случай сообщения из четырех слов; все параметры — функции времени [9]

Система, изображенная на рис. 9.5, использовалась для синтеза телефонных сообщений вида «Номер 135-3201» [9]. Правила управления основным тоном и длительностью произнесения семизначной последовательности цифр подбирались эмпирически по измерениям параметров реального речевого сигнала. При этом оказалось, что синтезированная речь, формируемая системой, изображенной на рис. 9.5, является более предпочтительной по сравнению с речью, полученной путем простого сочленения последовательности слов. Это связано с наличием «машинного» акцента у синтезированной речи. Хотя эти результаты и являются обнадеживающими, однако для построения методики синтеза, приводящей к натурально звучащей высококачественной синтезированной речи, сформированной по словарю фраз или слов, представленных в цифровой форме, требуются обширные исследования [11].

В заключение отметим, что имеется целый ряд способов цифрового представления элементов словаря, позволяющих столь же гибко манипулировать с параметрами речевых фраз.

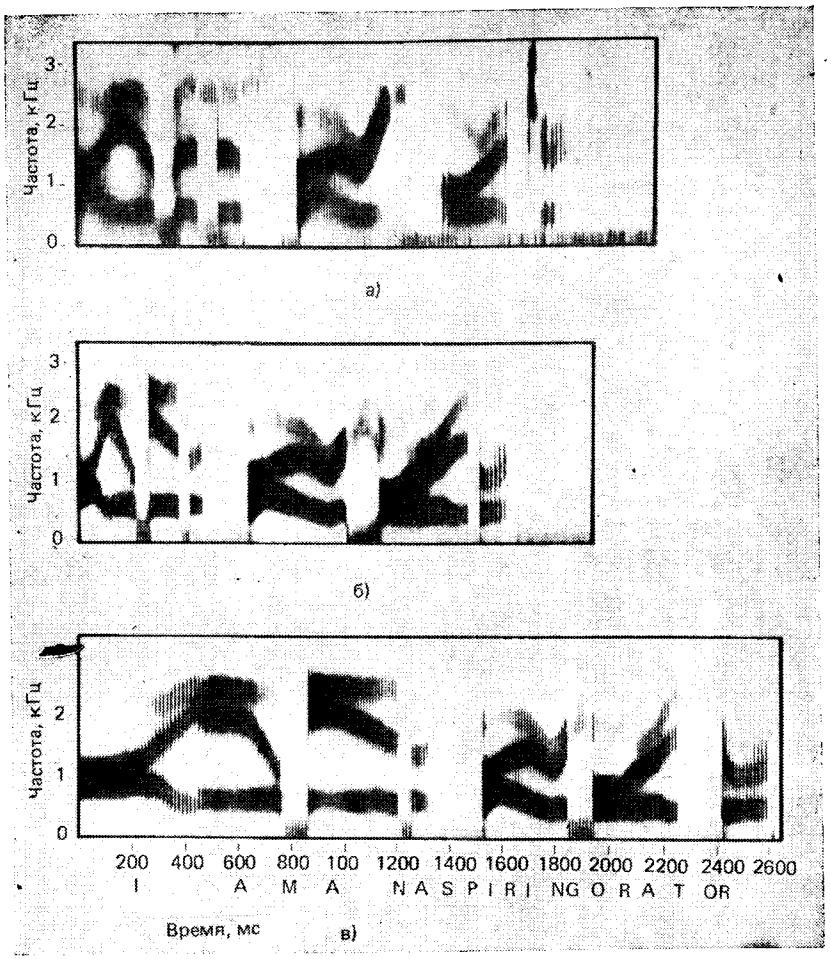


Рис. 9.7. Сравнение спектрограмм:
а) исходного сигнала; *б*) слов, модифицированных по просодике, периоду основного тона и временным соотношениям; *в*) последовательности изолированных слов [10]

9.1.4. Применение систем с речевым ответом

Гибкость систем с речевым ответом облегчает их использование в ряде экспериментальных работ, выполняемых в лабораториях Белла. В настоящее время созданы и исследованы: система речевых команд для выполнения межблочных соединений аппаратуры связи; вспомогательная справочная система; система информации о текущем курсе акций; система информации и контроля банков данных; справочная авиаслужба; система верификации дикторов. Ниже описываются две из перечисленных систем.

Применение систем речевого ответа при производстве работ по монтажу оборудования электросвязи. Обычно монтажник работает по отпечатанному тексту, содержащему указания по каждому межблочному соединению. Однако в ряде случаев при монтаже оборудования монтажнику бывает неудобно отрывать глаза от работы для чтения инструкции. В подобных случаях удобнее записать инструкцию на кассетный магнитофон и дать возможность монтажнику работать с указаниями в форме магнитофонной записи. Обычно для запуска магнитофона используют ножной выключатель, а останавливается он по тональному сигналу, записанному в конце каждого указания.

Таблица соединений может быть продиктована человеком. Однако для этого один человек должен прочитать указания (затратив на это, может быть, несколько часов), а другой — проверить запись во избежание ошибок. Обнаруженные неточности затем исправляются. Даже после успешной записи инструкции через некоторое время может потребоваться ее уточнение, что приведет к необходимости повторения всего процесса. Потребность в уточнении указаний может возникать несколько раз на протяжении нескольких дней или недель. Повышение утомляемости приводит к возрастанию числа ошибок, допускаемых людьми.

Таким образом, система с речевым ответом обладает рядом преимуществ:

1. Указания по монтажу обычно состоят из набора простых команд, содержащих лишь необходимую информацию, такую, как цвет и длина провода, точки его подключения. В этих случаях не требуется сглаженная или слитная речь.

2. Инструкции могут быть сформированы на основании относительно малого словаря — около 50 слов достаточно для части аппаратуры и около 100 слов — для комплекса оборудования связи, монтируемого фирмой «Вестерн Электрик».

3. Инструкции обычно формируются с помощью ЭВМ, поэтому их удобно использовать в системах с речевым ответом.

4. Инструкции часто изменяются. Использование систем с речевым ответом позволяет упростить утомительную работу по пересмотру содержания инструкций.

На рис. 9.8 представлена структурная схема системы с речевым ответом, предназначенная для формирования указаний по монтажу оборудования электросвязи. В качестве исходной информации для создания инструкции используется колода перфокарт, полученная с ЭВМ фирмы «Вестерн Электрик». Отперфорированные символы описывают слова, предназначенные для конкретной инструкции. Например, фраза

КРАСНЫЙ-ПАУЗА-20-7-ПАУЗА-4-Р-ПАУЗА-7-Z-КОНЕЦ

сообщает монтажнику, что красный провод длиной 27 дюймов следует пропустить от точки 4Р к точке 7Z. Используя соответствующий словарь, система с речевым ответом формирует нужное сообщение и посылает его к АРИКМ-декодеру. Полученный фрагмент речевого сообщения (указание монтажнику) записывается на кассету магнитофона.

В рассмотренном примере не используются возможности системы, связанные с передачей сообщений по каналу связи, однако можно в полной мере использовать достоинства многоканального

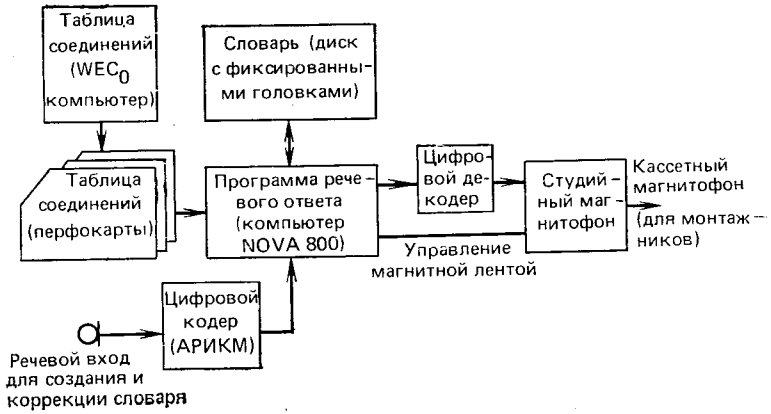


Рис. 9.8. Система речевого ответа для автоматического синтеза инструкции по монтажу [8]

режима работы как для одновременного формирования нескольких инструкций, так и для параллельного формирования одной слишком длинной инструкции с целью сокращения времени, затрачиваемого на ее запись [7, 8].

Системы с изменяющимся содержанием информационного банка. При производстве монтажных работ с применением систем речевого ответа между возможным пользователем и системой формирования сообщений нет взаимодействия. Это связано с тем, что кнопочный ввод в систему заменен предварительно отперфорированным набором карт, которые определяют сообщение, требуемое на выходе системы. При использовании систем с речевым ответом в качестве вспомогательной справочной службы, выдающей справки о кредитах, состоянии текущего счета, наличии товаров, необходим доступ к содержанию информационного банка. Система должна находить нужную информацию и формировать соответствующее сообщение, поступающее к абоненту. На рис. 9.9 представлена структурная схема системы с изменяющимся содержанием информационного банка. Здесь предпо-

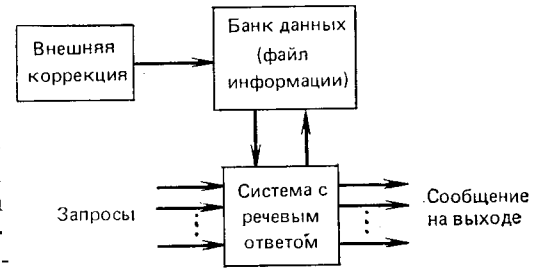


Рис. 9.9. Структурная схема системы речевого ответа с изменяющимся словарем [8]

лагается, что содержание банка данных может быть изменено как с помощью дополнительного внешнего источника, так и самой системой.

Предположим, что содержанием банка информации являются опись количества товаров, производимых компанией, и объем реализованной продукции с распределением по наименованиям. После каждой проведенной сделки информационная система с речевым ответом должна вносить изменения в содержание банка данных. По мере производства товаров банк данных также должен уточняться. В этом примере информационная система с речевым ответом не только позволяет вести учет товаров, но и предотвращает возможность продажи несколькими агентами одного и того же изделия в случае, когда их количество ограничено. Учет товаров также дает возможность компании всегда иметь текущую статистику спроса и, таким образом, корректировать ассортимент производимой продукции в соответствии со спросом.

Интересно также применение системы с коррекцией информационного банка в качестве системы информации о текущем курсе акций. Содержание банка данных составляют сведения о рыночной стоимости любой акции. Внешняя коррекция данных производится непосредственно с телеграфной или телетайпной ленты, содержащей последние биржевые новости.

Система информации о текущем курсе акций используется примерно таким образом. Абонент вызывает систему, которая отвечает: «Это система информации о текущем курсе акций. Стоимость приводится также и по отношению к ближайшему прошлому рабочему дню. Пожалуйста, введите рыночное обозначение интересующей вас акции». Абонент вводит: А-Т-Т-* — и система отвечает: «Американские Телефонные и Телеграфные, 62 и 3/8, вверх на 1/4».

Несомненно, что в будущем системы с речевым ответом на основе ЭВМ найдут широкое применение. Очевидно также, что ключевую роль при построении таких систем будут играть методы цифровой обработки речевых сигналов.

9.2. Системы распознавания дикторов

При распознавании дикторов цифровая обработка речи является тем первым шагом, с которого начинается решение задачи распознавания образов. Как видно из рис. 9.10, речевой сигнал (представление образа вектором) представлен с использованием таких методов цифровой обработки, которые сохраняют индивидуальные особенности диктора. Полученный образ сравнивается с предварительно подготовленными эталонными образами, а затем применяется соответствующая логика принятия решений для определения голоса заданного диктора среди возможного множества. Системы распознавания дикторов подразделяются на два вида: идентификация и верификация. При верификации диктора требуется установить его идентичность данному эталону. Устройство верификации принимает одно из двух возможных решений: диктор является тем, за кого он себя выдает, или не является. Для вынесения такого решения используется совокупность параметров, содержащих необходимую информацию об индивидуальности диктора и измеряемых по одной или нескольким фразам. Измеренные значения сравниваются (часто с использованием некоторых существенно нелинейных

метрик близости) с аналогичными параметрами эталонных образов подлежащего опознанию диктора.

Таким образом, при верификации диктора требуется однократное сравнение совокупности (совокупностей) измеренных значений со значениями параметров эталонов, на основе которого выносится решение о принятии или отклонении.

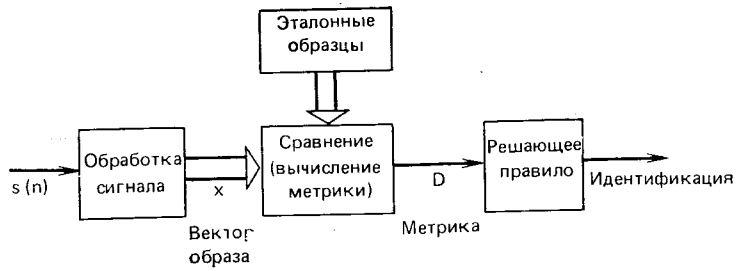


Рис. 9.10. Общее представление задачи распознавания диктора

предполагаемой идентичности. В общем случае вычисляется расстояние между измеренными значениями и распределением эталонов. На основе распределения потерь между возможными типами ошибок (т. е. верификации «самозванца» и отклонении «подлинного» диктора) устанавливается соответствующий порог различимости (расстояния). Вероятность перечисленных выше ошибок практически не зависит от N (числа эталонов, хранимых в системе), поскольку все эталоны голосов других дикторов используются для формирования устойчивого распределения, характеризующего всех дикторов. Записывая сказанное выше в математической форме, обозначим распределение вероятности измеренных значений вектора $\mathbf{x}$ для диктора как $p_i(\mathbf{x})$, что приводит к простому решающему правилу вида

$$\begin{aligned} &\text{Верифицировать диктора } i, \text{ если } p_i(\mathbf{x}) > c_i\, p_{av}(\mathbf{x}); \\ &\text{Отклонить диктора } i, \text{ если } p_i(\mathbf{x}) < c_i\, p_{av}(\mathbf{x}), \end{aligned} \tag{9.1}$$

где c_i — константа для i-го диктора, определяющая вероятности ошибок i-го диктора, а $p_{av}(\mathbf{x})$ — среднее (по всему ансамблю дикторов) распределение вероятности измеренных значений вектора $\mathbf{x}$. Изменяя порог c_i, можно изменять вероятность ошибки, определяемую вероятностями ошибок обоих типов.

Задача идентификации диктора существенно отличается от задачи верификации. В этом случае система должна точно указать одного из дикторов среди N дикторов данного множества. Таким образом, вместо однократного сравнения измеряемых параметров с хранимым в системе эталоном необходимо провести N сравнений. Решающее правило в этом случае сводится к *выбору такого диктора i, для которого*

$$\begin{aligned} &p_i(\mathbf{x}) > p_j(\mathbf{x}), \\ &j = 1,\ 2,\ \ldots.\ N,\ j \neq i, \end{aligned} \tag{9.2}$$

т. е. выбирается диктор с минимальной абсолютной вероятностью ошибки. С увеличением количества дикторов в ансамбле возрастает и вероятность ошибки, поскольку большое число вероятностных распределений в ограниченном пространстве параметров не может не пересекаться. Все более вероятным становится то, что два или более дикторов в общем ансамбле будут иметь распределения вероятностей, которые близки друг к другу. При таких условиях приемлемая идентификация дикторов становится практически невозможной.

Приведенный выше анализ позволяет сделать вывод, что между задачами идентификации и верификации имеется много общего и много различий. В каждом случае диктор должен произнести одну или несколько тестовых фраз. По этим фразам проводятся некоторые измерения, и затем вычисляются одна или

несколько мер различимости («расстояния») между предъявленным и эталонным векторами. Таким образом, с позиции методов цифровой обработки обе эти задачи сходны. Основное различие возникает на этапе вынесения решений.

9.2.1. Система верификации диктора

На рис. 9.11 показана структурная схема системы верификации диктора в реальном масштабе времени [13—16]. Лицо, желающее быть верифицированным, сначала вводит в систему данные, подтверждающие его право на идентификацию, а затем по

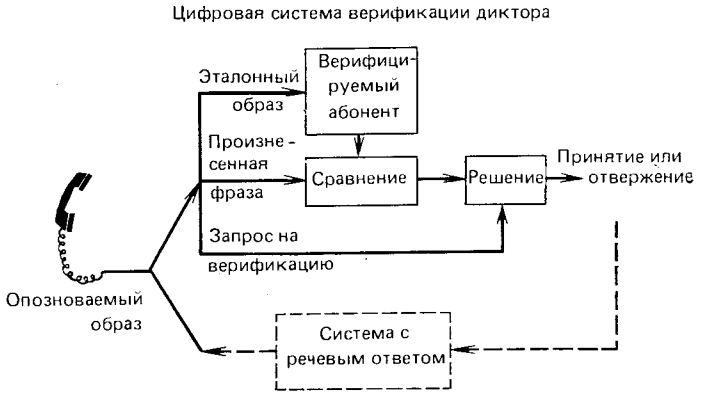


Рис. 9.11. Структурная схема системы верификации диктора [13]

запросу системы (сформированному, например, системой с речевым ответом) произносит эталонные фразы и в случае его верификации поручает системе выполнить необходимые операции. Система обрабатывает произнесенную диктором тестовую фразу с целью получения образа, сравниваемого затем с эталонным образом, соответствующим указанному при передаче права на идентификацию. Затем вычисляется матрица потерь [c_i в (9.1)], определяющая полную ошибку, и выносится решение о принятии или отклонении утверждения абонента об идентичности.

На рис. 9.12 представлена та часть системы верификации, в которой непосредственно осуществляется обработка сигнала. Отсчеты речевого сигнала, возникающие где-либо внутри выбранного интервала, обрабатываются с целью определения начала и конца фразы. Это осуществляется в устройстве анализа моментов начала и конца фразы. Такое устройство описано в гл. 4. После определения начала и конца фразы проводится ряд измерений и оценок параметров для формирования образа, описывающего данную фразу. Обычно используются измерения следующих параметров: периода основного тона для получения траектории периода основного тона данной фразы; кратковременной энергии для получения траектории кратковременной энергии; коэффициентов линейного предсказания для получения траектории передаточной

функции речеобразующего тракта и, наконец, оценивание формантных частот. (Все показанные на рис. 9.12 параметры одновременно используются в системе верификации, однако вследствие

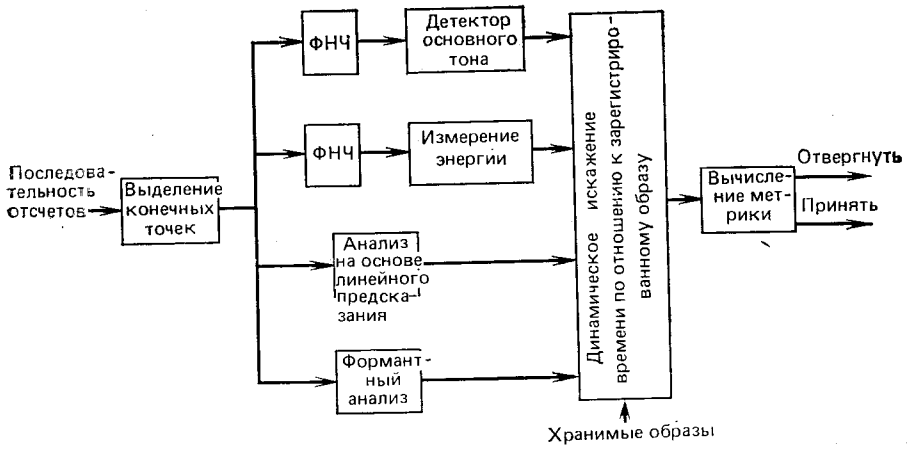


Рис. 9.12. Основные этапы обработки речевого сигнала в системе верификации диктора

большого объема вычислений, которые необходимо осуществить для анализа по методу линейного предсказания или для оценивания формантных частот при верификации в реальном масштабе

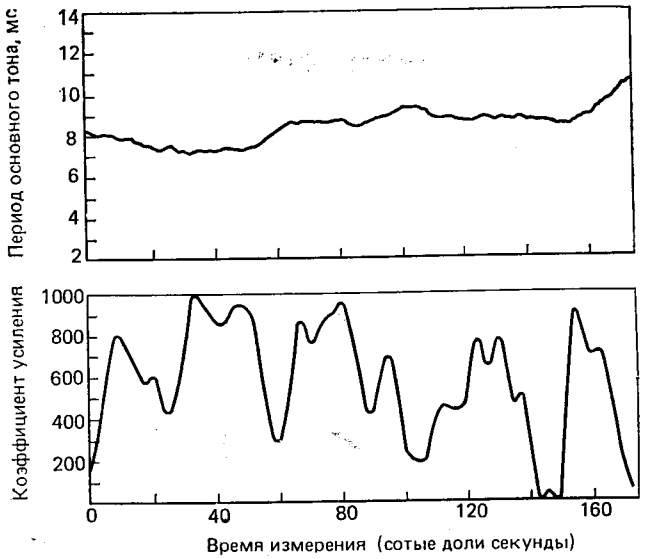


Рис. 9.13. Период основного тона и энергетический контур, используемые при верификации диктора [13]

времени, ограничиваются только измерением основного тона и энергии.)

Для иллюстрации широких возможностей использования методов цифровой обработки в таких системах полезно упомянуть конкретные алгоритмы, использованные Розенбергом (и другими разработчиками этой системы) при проведении соответствующих измерений. Выделение основного тона осуществлялось на основе методов параллельной обработки во временно́й области (см. гл. 4). Интенсивность измерялась с использованием кратковременного усреднения абсолютного значения в соответствии с результатами гл. 4. Для анализа по методу линейного предсказания использовался автокорреляционный алгоритм, рассмотренный в гл. 8. Наконец, формантный анализ проведен на основе гомоморфной фильтрации, описанной в гл. 7.

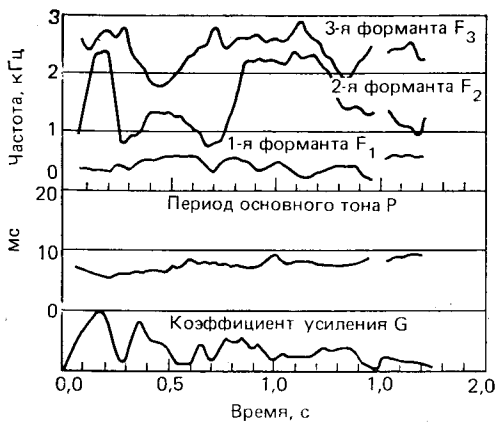


Рис. 9.14. Траектории первых трех формант, период основного тона, интенсивность для верификации диктора [15]

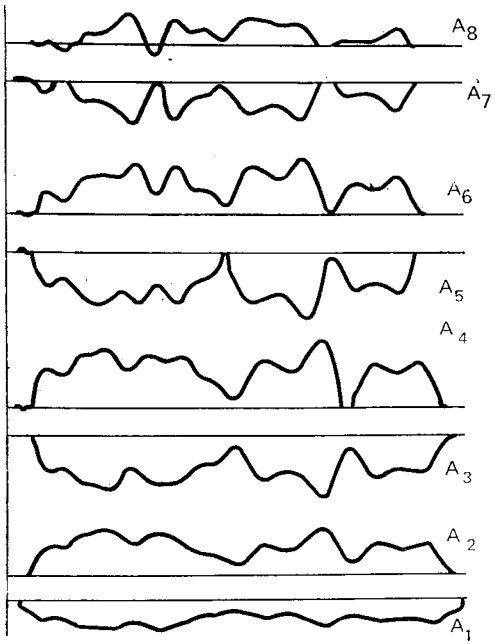


Рис. 9.15. Траектории первых восьми коэффициентов линейного предсказания для верификации диктора по фразе [16]

На рис. 9.13—9.15 показаны типичные траектории измерений для тестовой фразы: «We were away a year ago», произнесенной мужским голосом. На рис. 9.13 представлены траектории периода основного тона и интенсивности по всей фразе [13]. Эти данные оценивались периодически 100 раз/с и сглаживались фильтром нижних частот КИХ-типа с полосой 16 Гц. Для данного диктора флуктуации траектории интенсивности значительно превосходят флуктуации в траектории основного тона. На рис. 9.14 для той же фразы, произнесенной другим диктором, представлены траектории трех первых формант совместно с контуром основного тона и интенсивности [15]. Формантные траектории сглажены таким же КИХ-фильтром с частотой среза 16 Гц. Наконец, на рис. 9.15 представлены первые восемь коэффициентов предсказания 12-полюсной модели [16]. Из этого рисунка видно, что для данной фразы описание с помощью параметров линейного предсказания обладает значительной избыточностью. Таким образом, при использовании этих данных с целью верификации можно утверждать, что коэффициенты линейного предсказания внесут меньший вклад в уменьшение ошибок верификации. Можно считать поэтому, что при правильном подборе оцениваемых параметров при верификации можно получить почти такую же вероятность ошибки, как и при совместном использовании всех оценок, перечисленных выше.

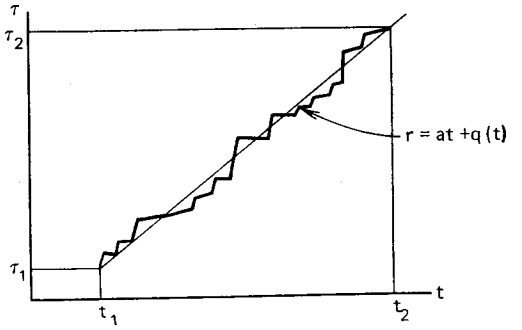


Рис. 9.16. Функция преобразования временно́го масштаба

После вычисления необходимых оценок параметров их необходимо сравнить с соответствующими эталонами голоса идентифицируемого диктора. Поскольку диктор не в состоянии повторить абсолютно точно в том же темпе одну и ту же фразу, нецелесообразно сравнивать такие временны́е параметры, как период основного тона, интенсивность и изменение во времени формантных частот. Эту трудность можно преодолеть путем нелинейного преобразования временно́го масштаба входного множества параметров для получения более точного соответствия между эталоном и последующими оценками параметров для одного и того же диктора. Процесс преобразования временно́го масштаба чрезвычайно важен и часто используется при обработке речевого сигнала.

Процесс преобразования временно́го масштаба схематически представлен на рис. 9.16. Временно́й масштаб следует трансформировать таким образом, чтобы характерные точки измеренной траектории $a(t)$ совпали с характерными точками эталонной траектории $r(t)$. Предполагается, что преобразующая функция имеет вид

$$\tau = \alpha t + q(t), \tag{9.3}$$

где $q(t)$ — нелинейная функция трансформации масштаба, а α представляет собой средний наклон характеристики преобразования. Отсутствие $q(t)$ соответствует простой линейной модификации. Граничные условия накладываются таким образом, чтобы

начальная и конечная точки исходной и эталонной фраз точно совпали. Эти условия имеют вид

$$\tau_1 = \alpha t_1 + q \mid t_1); \tau_2 = \alpha t_2 + q \mid t_2). \qquad (9.4а, б)$$

Теперь осталось только выбрать константу и параметр $q(t)$ таким образом, чтобы достичь наилучшего совпадения сравниваемых траекторий. Один из наиболее простых подходов заключается в определении $q(t)$ как кусочно-линейной функции с конечным числом точек излома по оси t, в которых изменяется наклон $q(t)$. Точки излома и наклон $q(t)$ (как и средний наклон α) определяются затем методом наискорейшего спуска, где в качестве критерия выступает мера различимости либо в виде расстояния между обрабатываемой и эталонной траекториями, либо в виде корреляции между ними.

Значительно более простым и эффективным с точки зрения вычислений является метод динамического программирования для оптимального выбора функции, преобразующей временной масштаб. Наложив вместо условия кусочной линейности условие непрерывности, относительно несложно определить оптимальный алгоритм преобразования для множества траекторий [17].

Рассмотрим работу алгоритма преобразования масштаба для двух траекторий, представленных в виде дискретной последовательности отсчетов. Обозначим точки на измеренной траектории через $n=1, 2, ..., N$, а точки на эталонной траектории через $m=1, 2, ..., M$. Требуется выбрать такую функцию преобразования масштаба w, чтобы выполнялись условия:

$$\left.\begin{array}{l} m = w(n);\ w(1) = 1 \text{ в начальной точке,} \\ w(N) = M \text{ в конечной точке,} \end{array}\right\} \qquad (9.5),\ (9.6)$$

Если используемая преобразующая функция линейна, то она имеет вид

$$w(n) = \left[\left(\frac{M-1}{N-1}\right)(n-1) + 1\right]. \qquad (9.7)$$

Если для преобразования используется нелинейная функция, то в соответствии с граничными условиями следует рассмотреть стратегию движения из начальной точки $n=1$, $m=1$ в конечную точку $n=N$, $m=M$ по дискретной сетке точек на плоскости. Для ограничения степени нелинейности преобразующей функции целесообразно предположить, что w не может изменяться более чем на два шага дискретной сетки при любом n. Иными словами,

$$w(n+1) - w(n) = \begin{cases} 0, 1, 2, & w(n) \neq w(n-1); \\ 1, 2, & w(n) = w(n-1). \end{cases} \qquad (9.8)$$

Таким образом, при изменении значения функции в предшествующей точке ее приращения в данной точке могут составлять 0, 1, 2, а в противном случае — только 1 и 2. Чтобы определить, какие из условий (9.8) выполняются, необходимо иметь меру сходства между эталонной траекторией в точке n и входной траекторией в точке m. Мера сходства (или расстояние) между траекториями используется для определения вида преобразующей функции, которая доставляет локальный минимум максимальному значению расстояния по всей траектории в соответствии с ограничениями (9.8).

Для примера на рис. 9.17 [17] показаны область возможных значений дискретной сетки (n, m) и типичная функция преобразования масштаба $w(n)$ (сплошная линия внутри сетки) для трансформации 20 точек ($N=20$) эталонной траектории в 15 точек ($M=15$) обрабатываемой траектории. Вследствие ограничений непрерывности получающаяся функция должна лежать внутри параллелограмма, изображенного на рисунке.

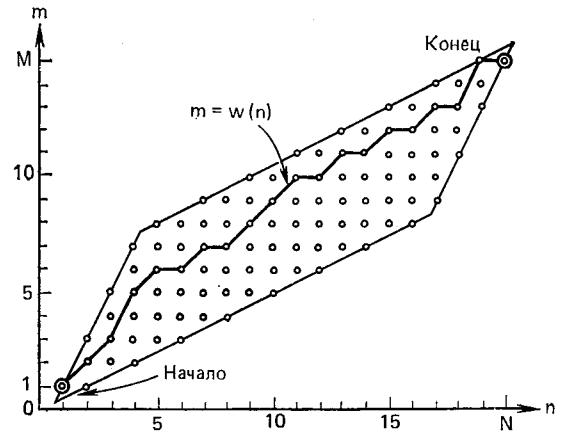


Рис. 9.17. Пример построения функции преобразования временно́го масштаба [17]

Метод трансформации временно́го масштаба был использован как для верификации [13], так и для распознавания речи [17]. В качестве примера на рис. 9.18 представлены траектории интенсивности как до, так и после трансформации масштаба [13]. В данном случае сближение траекторий весьма заметно.

Заключительным шагом в процессе верификации (см. рис. 9.12) являются вычисление некоторой полной меры различимости (на основе частных мер различимости отдельных траекторий) и сравнение ее с выбранным соответствующим образом порогом. Простейшей мерой различимости двух траекторий может служить нормированная сумма квадратов; например, для j-й траектории мера различимости d_j будет иметь вид

$$d_j = \sum_i [[a_{js}(i) - a_{jr}(i)]/\sigma_{aj}(i)]^2, \qquad (9.9)$$

где $a_{js}(i)$ — значение j-й траектории входного сигнала в момент i; $a_{jr}(i)$ — значение j-й траектории эталона в момент i; $\sigma_{aj}(i)$ —

стандартное отклонение j-й траектории в момент i. Полная мера различимости обычно представляет собой взвешенную сумму корней, т. е.

$$D = \sum_i w_j d_j, \tag{9.10}$$

где w_j — вес, выбираемый на основе значимости j-го измеренного значения траектории верифицируемого диктора.

Рассмотренная выше система верификации диктора была всесторонне исследована, и полученные результаты позволили сделать вывод о том, что они являются потенциально достижимыми для такого рода систем. Был проведен целый ряд экспериментов по проверке системы как при использовании высококачественных фраз при малом числе дикторов, так и при использовании фраз «телефонного» качества при очень большом числе дикторов. Использовался даже хорошо тренированный имитатор (подражатель), пытавшийся «обмануть» систему. Результаты этих экспериментов показали, что в случае высококачественного сигнала равные вероятности ошибок в системе (т. е. когда вероятность ошибочного отклонения диктора равна вероятности ошибочного отождествления) могут быть сделаны сколь угодно близкими к нулю, если выбрано достаточное количество наблюдений и подобраны весовые коэффициенты для каждого диктора. Равновероятные ошибки в таких системах при использовании профессионального имитатора составляют 4,1%. При использовании сигнала телефонного качества и применении в качестве признаков интенсивности и периода основного тона вероятность ошибок составила примерно 7%. Добавление более сложных признаков, таких, как форманты или параметры линейного предсказания, приводит к значительному уменьшению вероятности ошибки.

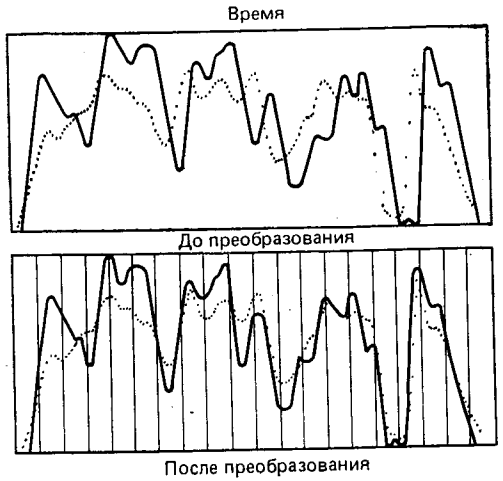


Рис. 9.18. Пример преобразования временно́го масштаба траектории кратковременной энергии речи [13]

9.2.2. Система идентификации диктора

Задачи идентификации и верификации во многом сходны. С точки зрения обработки сигналов обе рассматриваемые задачи почти совпадают и практически все, что изображено на рис. 9.12,

в равной мере подходит как для верификации, так и для идентификации. Основное отличие заключается в тех параметрах, которые используются для построения меры различимости, а также в необходимости вычисления N мер различимости вместо одной. Решение, формируемое системой при идентификации диктора, сводится к выбору того диктора, чье эталонное описание наиболее близко к описанию, полученному по входному сигналу. При верификации требуется решить задачу бинарного выбора, т. е. принять или отклонить утверждение о том, что голос опознаваемого диктора идентичен данному эталону, и это достигается на основе сравнения значения меры различимости с выбранным порогом.

Хотя для целей верификации систем вполне пригодна классическая мера различимости (9.10), для идентификации диктора обычно используют более сложные и устойчивые к различным аномалиям меры различимости [18, 20]. Напомним, что значение меры различимости вычисляется с целью сравнения входного и эталонного образов. Мера различимости, используемая Аталом [18, 10], может быть получена следующим образом. Пусть $\mathbf{x}$ представляет собой вектор-столбец измеренных значений входного сигнала размерностью L, причем элементом $\mathbf{x}$ является k-е измеренное значение. Предполагается, что совместная функция плотности вероятности измеренных значений для i-го диктора представляет собой многомерное распределение Гаусса со средним значением m_i и ковариационной матрицей $\mathbf{W}_i$. Таким образом, L-мерная плотность распределения Гаусса для $\mathbf{x}$ имеет вид

$$g_i(\mathbf{x}) = (2\pi)^{-1/2} |\mathbf{W}_i|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-m_i)^t W_i^{-1}(\mathbf{x}-m_i)\right], \tag{9.11}$$

где $\mathbf{W}_i^{-1}$ — матрица, обратная $\mathbf{W}_i$ ($\mathbf{W}_i$ предполагается неособенной); $|\mathbf{W}_i|$ — детерминант $\mathbf{W}_i$; t — транспонирование вектора. Решающее правило, минимизирующее вероятность ошибки, состоит в том, что вектор измеренных значений $\mathbf{x}$ следует отнести к классу i, если

$$p_i g_i(\mathbf{x}) \geqslant p_j g_j(\mathbf{x}), \; i \neq j, \tag{9.12}$$

где p_i — априорная вероятность принадлежности вектора $\mathbf{x}$ к классу i. Поскольку $\ln y$ — монотонно возрастающая функция своего аргумента, решающее правило (9.12) можно значительно упростить, переписав в виде

$$d_i(\mathbf{x}) = \begin{cases} \frac{1}{2}(\mathbf{x}-m_i)^t \mathbf{W}_i^{-1}(\mathbf{x}-m_i) + \frac{1}{2}\ln|W_i| = \ln p_i \leqslant \\ \leqslant d_j(\mathbf{x}), \quad i \neq j. \end{cases} \tag{9.13}$$

Последние два члена в правой части (9.13) не зависят от вектора $\mathbf{x}$, и поэтому можно считать, что они представляют собой смещение i-го класса. Для большинства практически важных случаев установлено, что решающее правило со смещением в правой части не имеет преимуществ перед решающим правилом, основанным

только на первом члене (9.13). Таким образом, функцию различимости можно определить как

$$\hat{d}_i = (\mathbf{x} - m_i)^t \mathbf{W}_i^{-1} (\mathbf{x} - m_i), \qquad (9.14)$$

а индекс i выбран таким образом, чтобы минимизировать по этому индексу.

Решающее правило предполагает вычисление вектора средних и ковариационной матрицы для каждого класса i на множестве решения. Вектор средних и ковариационная матрица определяются по обучающей последовательности $\mathbf{x}_i(n)$ векторов, принадлежащих i-му классу:

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{x}_i(n) \qquad (9.15)$$

и

$$\mathbf{W}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{x}_i(n)\, \mathbf{x}_i^t(n) - \mathbf{m}_i\, \mathbf{m}_i^t. \qquad (9.16)$$

На рис. 9.19 представлены типичные примеры распределений параметров, измеренных по речевому сигналу, и их приближений одномерными гауссовскими распределениями. В одних случаях степень согласия больше, чем в других.

Следует сказать несколько слов относительно предположений и необходимых вычислений, приводящих к решающему правилу (9.14). Предположение относительно нормальности распределения измеренных значений можно подтвердить рядом соображений. Во-первых, чтобы решающее правило оставалось в силе, распределение не обязательно должно быть строго нормальным. Эта особенность часто проявляется в физических измерениях. Например, в случае унимодальных распределений достаточно, чтобы распределение было нормально в центральной области возможных значений. Более того, как отмечалось выше, решающее правило оказывается оптимальным для целого класса распределений, которые могут быть получены из нормального с помощью монотонных преобразований. Наконец, решающее правило требует знания только первых двух моментов распределения. Точное оценивание высших моментов оказывается задачей трудноразрешимой на практике.

Важное преимущество функции различимости (9.14) состоит в ее инвариантности к несингулярным линейным преобразованиям [20]. Это свойство инвариантности оказывается чрезвычайно важным, поскольку использование некоторой совокупности параметров и их линейного преобразования приводит к одним и тем же результатам, например одинаковые результаты можно получить по некоторым траекториям и их преобразованию Фурье. Второе важное свойство меры различимости (9.14) состоит в том, что она построена на основе взвешивания различных компонент опознаваемого вектора в соответствии с их значимостью [20].

Используя решающее правило (9.14), Атал исследовал эффективность различных параметрических представлений речевого сигнала применительно к задаче идентификации диктора [20]. Каждый из десяти дикторов произносил один и тот же текст по 6 раз.

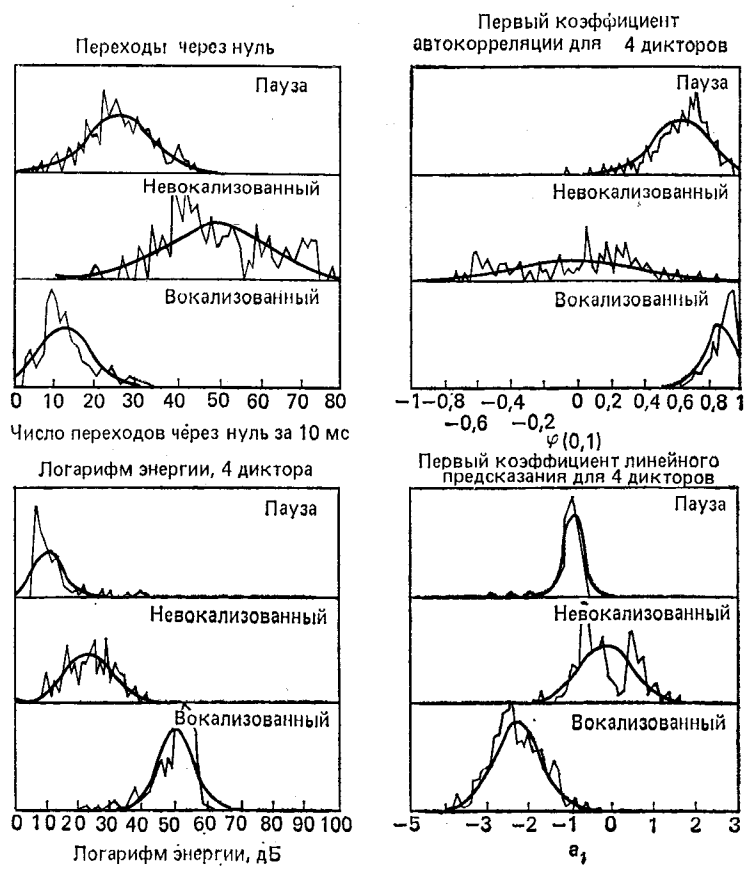


Рис. 9.19. Измеренные распределения некоторых параметров речевого сигнала с подогнанными к ним гауссовыми кривыми [27]

Каждая фраза разделялась на 40 сегментов равной длины, чем обеспечивалось грубое выравнивание масштаба времени. Средняя длина каждого сегмента составляла около 50 мс. Затем проводился анализ на основе линейного предсказания на каждом из 40 сегментов для каждого из 60 предложений. Таким образом был получен вектор образов для каждого интервала анализа. По коэффициентам линейного предсказания рассчитывались импульсная характеристика, автокорреляционная функция, функция площадей поперечного сечения в неоднородной акустической трубе без потерь и кепстр. Затем проверялась точность идентификации, для чего одна фраза служила опознаваемой, а остальные использова-

лись как эталонные для каждого диктора. Решающее правило (9.14) использовалось для идентификации диктора на каждом из 40 интервалов анализа с целью определения вероятности правильной идентификации. Результаты для каждого из параметров представлены на рис. 9.20.

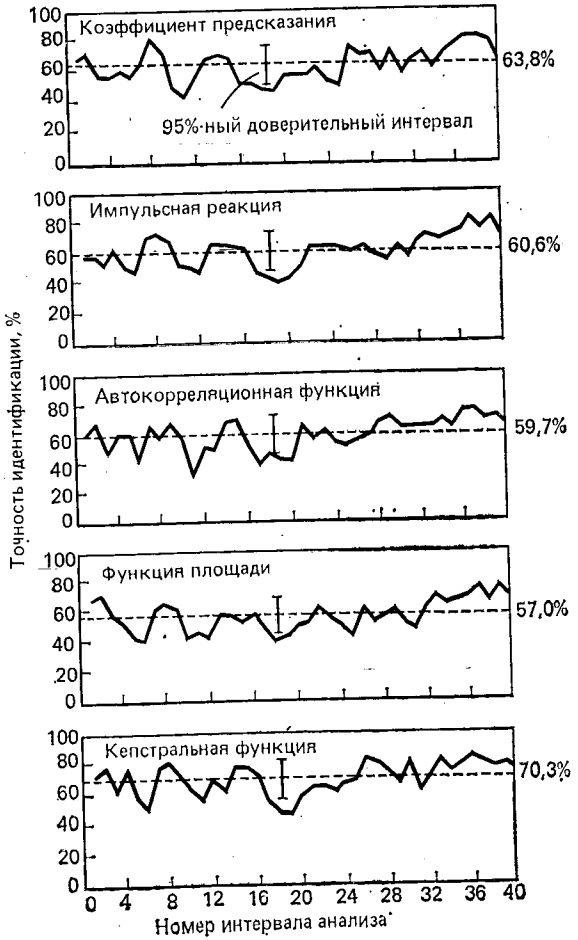


Рис. 9.20. Точность идентификации диктора в зависимости от множества используемых параметров [20]

Все параметры дают примерно одинаковую вероятность ошибки, однако точность кепстрального метода несколько выше, чем всех других. Объединяя несколько интервалов анализа для получения описания опознаваемого вектора большей размерности, можно добиться меньшей вероятности ошибки. На рис. 9.21 представлены кривые точности идентификации, достигнутой с использованием кепстрального метода в зависимости от длительности

сегмента речевого сигнала, используемого для вычисления различимости. Полученные результаты показывают, что для данного ансамбля дикторов 95%-ная точность идентификации может быть достигнута на сегментах сигнала длительностью около 0,5 с.

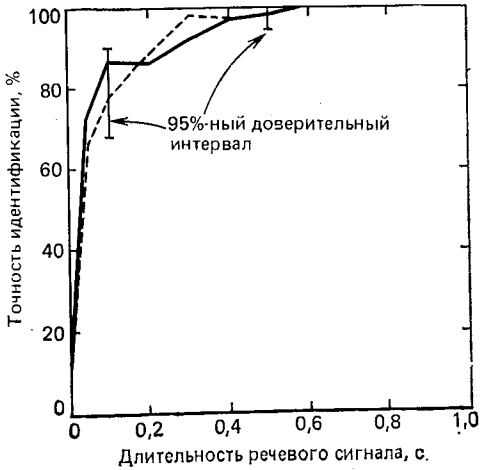


Рис. 9.21. Точность идентификации диктора (с использованием кепстральных параметров) в зависимости от длительности речевого сигнала [20]

9.3. Системы распознавания речи [17, 21—26]

Как и при распознавании диктора, методы цифровой обработки применяются при распознавании речевого сигнала для получения описания распознаваемого образа, которое затем сравнивается с хранимыми в памяти эталонами. Задача распознавания речевого сигнала состоит в определении того, какое слово, фраза или предложение были произнесены.

В отличие от областей машинного речевого ответа и распознавания диктора, где задача в общем случае достаточно определена, область распознавания слов является одной из тех, где, прежде чем поставить задачу, требуется ввести большое число предположений, например:

тип речевого сигнала (изолированные слова, непрерывная речь и т. д.);

число дикторов (система для одного диктора, нескольких дикторов, неограниченного числа дикторов);

тип диктора (определенный, случайный, мужчина, женщина, ребенок);

условия произнесения фраз (звукоизолированное помещение, машинный зал, общественное место);

система передачи (высококачественный микрофон, узконаправленный микрофон, телефон);

тип и число циклов обучения (без обучения, с ограниченным числом циклов обучения, с неограниченным числом циклов обучения);

размер словаря (малый объем 80—20 слов, средний объем 20—100 слов и большой объем — более 100 слов).

формат произносимых фраз (ограниченный по длительности текст, свободный речевой формат).

Из приведенного перечня условий следует, что при создании систем распознавания речи реализация некоторых из условий может оказаться более предпочтительной. В данном параграфе будут рассмотрены три наиболее распространенных типа систем распознавания речи, в которых широко используются методы цифровой обработки сигналов. Все они являются системами распознавания с ограниченным словарем, не содержащим контекста. Хотя в системах распознавания слитной речи также широко используются цифровые методы обработки [21, 22], однако большая часть усилий при разработке таких систем затрачивается на синтаксический и семантический анализ фраз. Эти области близко примыкают к лингвистической теории речи, поэтому изложение подобных вопросов увело бы нас в сторону от рассматриваемых здесь задач. Интересующегося читателя мы отсылаем к соответствующей литературе, содержащей обсуждение систем, «понимающих» речь.

9.3.1. Система распознавания изолированных цифр [25]

Система распознавания изолированных цифр обладает следующими свойствами:

1. Словарь малого объема состоит из изолированных слов, обозначающих десять цифр (0—9).
2. Отсутствуют ограничения на количество дикторов, а также на их пол и возраст.
3. Условия произнесения: машинный зал, микрофон — узконаправленный или высококачественный.
4. Обучение не предусмотрено.
5. Формат на входе — однословный с паузами между словами.

На рис. 9.22 представлена структурная схема системы распознавания изолированных цифр. Как видно из этого рисунка, основными элементами системы являются устройство анализа моментов

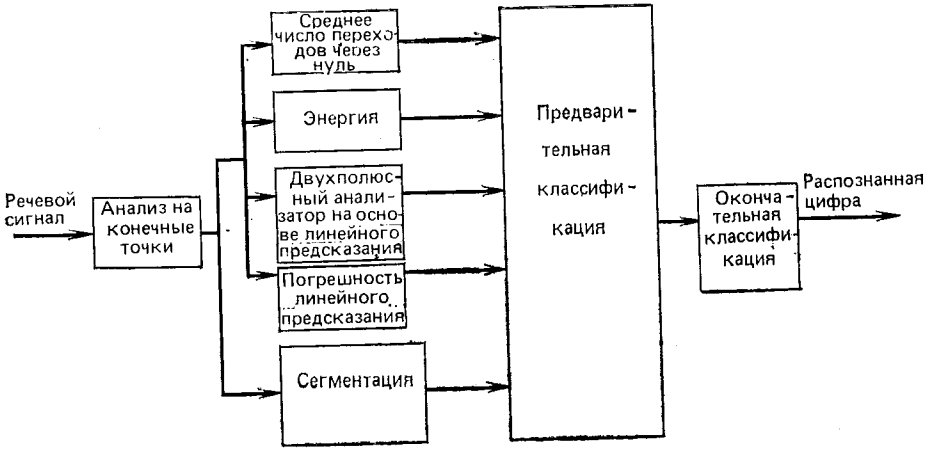


Рис. 9.22. Структурная схема системы распознавания отдельно произнесенных цифр [25]

начала и окончания слова (как и в системе распознавания), устройство обработки, формирующее образ или вектор измеренных значений, устройство сегментации фразы на интервалы и блок предварительных и окончательных решений относительно произнесенной цифры.

Хотя существует много способов представления сигнала, которые можно использовать в системах распознавания речи, представления, применяемые в системах, инвариантных к диктору, должны быть достаточно устойчивыми [25]. Измерения параметров должны быть простыми и однозначными, а их измеренные значения должны наиболее полно отражать различия в звуках речи.

Кроме того, измерения должны допускать достаточно простую интерпретацию с позиций систем, инвариантных к диктору. В одной из таких устойчивых систем (см. рис. 9.22) использованы следующие параметры: среднее число переходов через нуль, энергия, коэффициенты линейного предсказания с использованием двухполюсной модели и погрешность предсказания.

Измерения первых двух параметров рассматривались в гл. 4. Хотя в гл. 8 рассматривался общий метод линейного предсказания, использование двухполюсной модели является несколько необычным. Использование двухполюсной модели было предложено Макоулом и Вульфом [26] как удачный метод описания основных свойств кратковременного спектра. Частота полюса характеризует основную концентрацию энергии в спектре, а погрешность предсказания показывает общий наклон спектра.

Для иллюстрации сказанного на рис. 9.23 [26] представлены спектры отдельных звуков речи, полученные с использованием

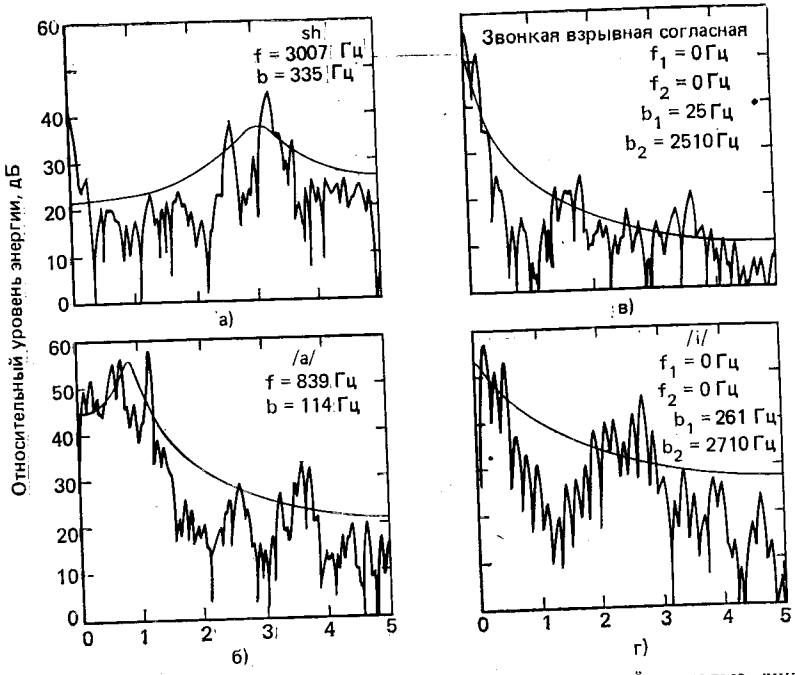


Рис. 9.23. Согласие между спектром сигнала и двухполюсной моделью линейного предсказания для некоторых звуков речи [26]

БПФ, и спектры соответствующей этим звукам двухполюсной модели. Для случая двухполюсного анализа полином имеет либо один комплексно-сопряженный корень, либо два действительных корня. На рис. 9.23*а* изображен спектр звука */sh/* в слове «short». В данном примере двухполюсная модель дает комплексно-сопряженный корень на частоте около 3 кГц, т. е. в области максимальной концентрации энергии в спектре. На 9.23*б* представлены аналогичные результаты для гласного звука */a/*, где концентрация

основной энергии в спектре имеет место на частоте около 800 Гц. В примерах рис. 9.23*в* основная часть энергии спектра сосредоточена в области нулевых частот и, таким образом, модель имеет два действительных полюса в правой полуплоскости z-плоскости.

Из рис. 9.23 видно, что вычисленные частоты двухполюсной модели хорошо описывают распределение энергии в спектре звука и могут, таким образом, быть использованы для описания звуков с относительно высоко- или низкочастотным распределением энергии. Так, например, шумоподобные звуки характеризуются относительно высокочастотной концентрацией энергии, тогда как носовые и гласные звуки имеют относительную концентрацию в области низких частот.

Для иллюстрации типичных результатов анализа на рис. 9.24 и 9.25 представлены траектории параметров, построенных по словам «девять» и «шесть». Предварительное распознавание основано

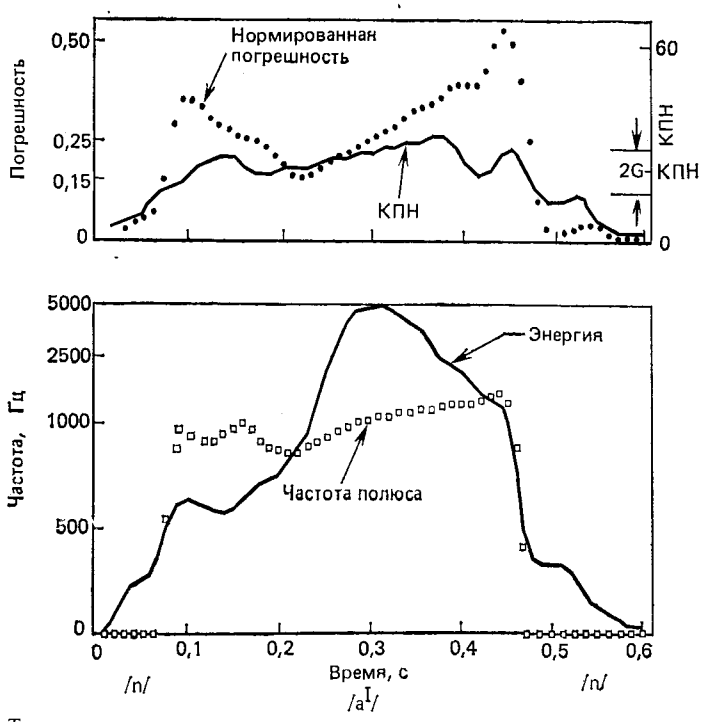


Рис. 9.24. Траектории сигналов для распознавания отдельно произнесенной цифры /nine/ [25]

на грубой классификации цифр на основе анализа каждого отдельного измерения (в различных точках на протяжении всего слова), а окончательное решение выносится путем объединения отдельных решений по каждому из измеренных значений. Так, например, начальный носовой сегмент слова «девять» на рис. 9.24

характеризуется малой нормированной погрешностью предсказания и положением полюсов на нулевой частоте, в то время как фрикативным началу и концу слова «шесть» на рис. 9.25 соответствуют значительная нормированная погрешность предсказания, отличная от нуля частота полюсов спектра и большое количество переходов через нуль (КПН).

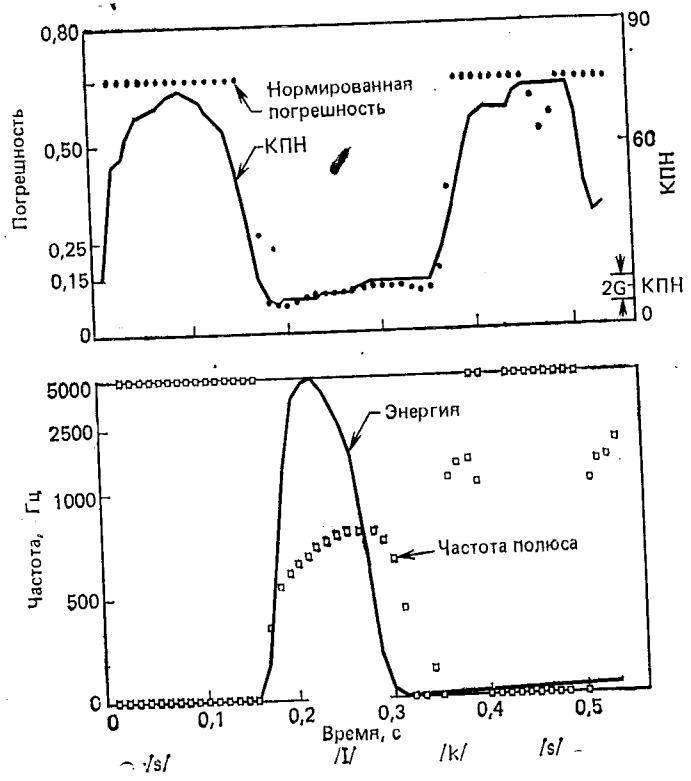


Рис. 9.25. Траектории сигналов для распознавания отдельно произнесенной цифры /six/ [25]

Самбуром и Рабинером [25] предложен древовидный алгоритм принятия окончательного решения на основе совместной обработки частных решений, полученных по каждому измеренному значению на каждом интервале анализа. При использовании этого алгоритма точность распознавания для 65 дикторов составила от 94,4 до 97,3%.

9.3.2. Система распознавания слитной последовательности цифр

Приведем решение более сложной задачи распознавания слитной последовательности цифр при произнесении их произвольным диктором. Свойства, которыми должна обладать эта система, в

основном совпадают со свойствами системы распознавания, рассмотренной в 9.3.1, с одним важным исключением. Свойство 5 в данном случае состоит в необходимости распознавания слитной последовательности из трех слов (цифр) без пауз между ними.

Хотя между системами распознавания изолированных цифр и слитной последовательности цифр много общего, реализация этих систем распознавания существенно различается, особенно в блоке анализа или обработки сигнала. Это связано с необходимостью предварительной сегментации слитной последовательности на отдельные цифры перед их распознаванием. Задача сегментации является чрезвычайно сложной, и в настоящее время не найдено простого решения для общего случая.

На рис. 9.26 представлена структурная схема блока обработки сигнала в системе распознавания слитной последовательности

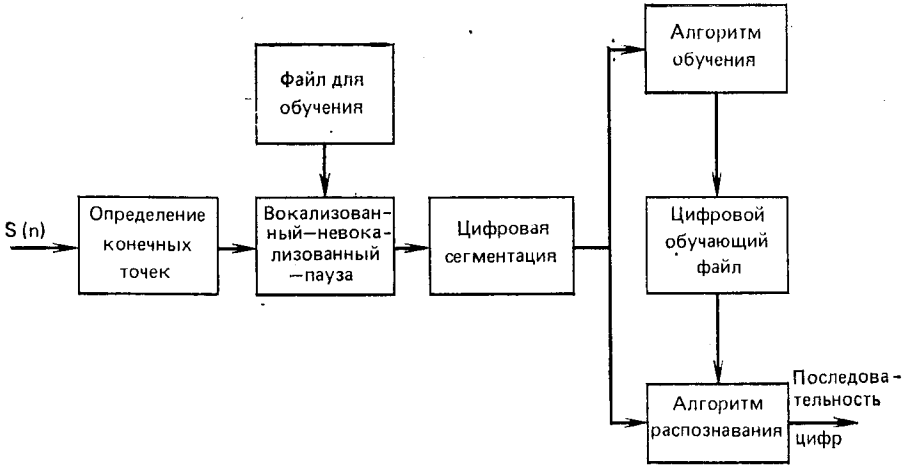


Рис. 9.26. Структурная схема системы распознавания последовательности цифр [27]

цифр. Записанная последовательность цифр первоначально подвергается анализу с целью определения моментов начала и окончания фразы. Для этого используется метод анализа во временной области, изложенный в гл. 4. Вслед за определением моментов начала и окончания фразы речевой сигнал подвергается обработке с целью оценивания следующих параметров (100 раз/с): среднего числа переходов через нуль, логарифма энергии, коэффициентов линейного предсказания, логарифма погрешности линейного предсказания и первого коэффициента автокорреляции.

Измеренные параметры используются затем в качестве входных сигналов решающего устройства, которое классифицирует каждый 10-миллисекундный интервал как вокализованный, невокализованный или паузу на основе неевклидовой метрики [такого типа, как (9.14)]. Для сегментации слитного потока на отдельные

цифры используется нелинейно сглаженная траектория признака «вокализованный—невокализованный—пауза» совместно с некоторой статистической информацией о надежности классификации на каждом интервале и измеренными значениями энергии сигнала. Для правильной сегментации в систему необходимо ввести информацию о количестве цифр в фразе. Для всех примеров, рассматриваемых в данном разделе, предполагается, что последовательность содержит ровно три цифры.

Сегментация осуществляется на основе использования известных результатов по различным измерениям на каждом 10-миллисекундном интервале. Например, известно, что невокализованный интервал соответствует интервалу, в пределах которого находится искомая граница, поскольку ни одна из цифр не содержит невокализованных звуков внутри слова. Известно также, что глубокие провалы в траектории энергии на вокализованном сегменте почти всегда соответствуют границе между цифрами. Основываясь на этих наблюдениях, можно синтезировать простые и более сложные правила сегментации фразы. Хотя существуют отдельные случаи, для которых точная сегментация затруднительна, существует возможность сегментации слитной последовательности цифр с полной ошибкой менее 1%, т. е. имеется менее 1% случаев, в которых определение границ цифр на слух показывает, что при автоматической сегментации часть сигнала данной цифры включена в сигнал другой цифры или, наоборот, к сигналу данной цифры добавлена часть сигнала следующей.

На рис. 9.27 и 9.28 показаны два примера последовательностей цифр, сегментированных системой распознавания. На этих рисунках через *а*, *б*, *в*, *г* обозначены траектории числа переходов через нуль, логарифма энергии, статистического параметра, на основе которого проводятся классификация «вокализованный—невокализованный—пауза» и принятие решения. Статистический параметр, используемый для классификации типа сегмента и представленный на рисунках кривыми *в*, означает вероятность того, что классификатор выносит правильное решение, поэтому он изменяется от нуля до единицы. Траектория *г* признака «вокализованный—невокализованный—пауза» является трехуровневой, где уровень *1* соответствует паузе, уровень *2* — невокализованному сигналу, а уровень *3* — вокализованному сигналу.

На рис. 9.27 показаны результаты сегментации последовательности цифр /721/. Первая граница расположена на начальном участке невокализованного сегмента, т. е. /s/ в слове «seven», следующая граница — в начале следующего невокализованного сегмента, соответствующего звуку /t/ в слове «two», а третья граница — в области локального минимума логарифма энергии внутри второй вокализованной зоны. Точная граница не совпадает с локальным минимумом логарифма энергии, но расположена вблизи этого минимума. Она определена с помощью ряда совместных решений алгоритма сегментации. Третья граница точно не определена, но для распознавания цифр ее точное положение внутри вокализованно-

го сегмента и не требуется. Граница последней цифры расположена в начале последней паузы. Следует отметить, что другое возможное расположение границы на рис. 9.27 соответствует точному локальному минимуму траектории логарифма энергии в звуке /v/

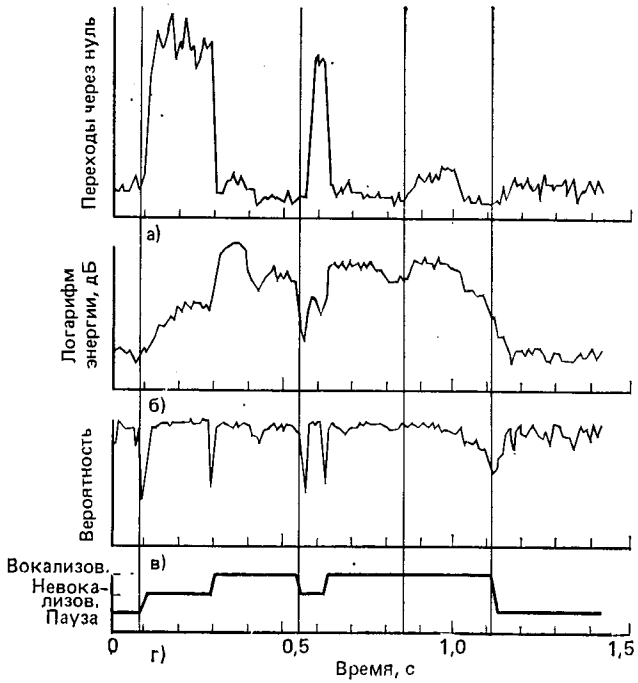


Рис. 9.27. Сегментация последовательности цифр /721/ [27]

слова «seven». Однако правило сегментации исключает этот минимум, он заменяется минимумом на следующем вокализованном сегменте.

На рис. 9.28 представлен более сложный для сегментации отрезок, соответствующий последовательности цифр /191/. Входной сигнал полностью вокализован, поэтому нет невокализованных граничных сегментов. Кроме того, отсутствуют четкие минимумы траектории логарифма энергии, т. е. эти минимумы недостаточно глубокие и имеют большую протяженность. Таким образом, расположение границ выбрано с использованием алгоритма сегментации на основе логических правил. Прослушивания показали, что расположение границ на полностью вокализованных сегментах некритично, что объясняется наличием значительной коартикуляции при произнесении таких полностью вокализованных последовательностей цифр.

Следующим шагом после сегментации является реализация алгоритма распознавания. Для каждого сегмента цифры вокализованный участок (который определяется по признаку «вокализованный—невокализованный—пауза») подвергается анализу на

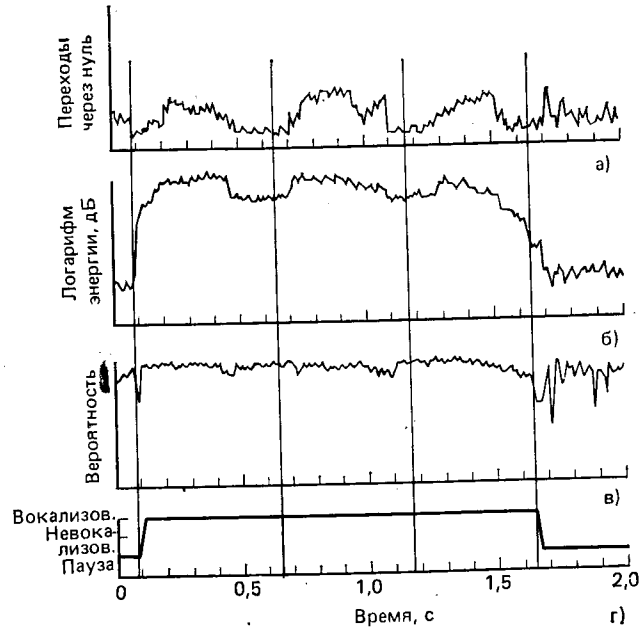


Рис. 9.28. Сегментация последовательности цифр /191/ [27]

основе линейного предсказания с использованием десятиполюсной модели. Метод распознавания основан на статистическом решающем правиле, по которому для каждого интервала обрабатываемой фразы (совокупности параметров линейного предсказания) проводится сравнение с соответствующим эталоном и выносится решение о соответствии обрабатываемой цифры той из эталонных, с которой имеется наибольшее сходство (наименьшее расстояние).

Эталонные файлы содержат статистическую информацию о коэффициентах линейного предсказания на каждом интервале анализа и для каждой цифры. В этих файлах содержится информация о среднем и дисперсии по множеству произнесений и множеству дикторов. При распознавании обрабатываемая фраза подвергается преобразованию масштаба времени выравниванием длительностей опознаваемых и эталонных цифр. Все методы преобразования временно́го масштаба, рассмотренные в § 9.1, пригодны и в данном случае. Поскольку все цифры (за исключением 7) моносиллабические, то в большинстве систем распознавания используется линейное преобразование масштаба. Для каждой эталонной фразы вычисляется среднее расстояние между ее коэффициента-

ми линейного предсказания и соответствующими коэффициентами обрабатываемой фразы, а за произнесенную принимается та цифра, для которой это расстояние оказывается минимальным. Выбор меры различимости, определяющей расстояние между совокупностями параметров линейного предсказания, является одним из наиболее важных факторов, определяющих качество работы подобных систем. Известен ряд мер различимости, используемых при обработке параметров линейного предсказания. В следующем разделе рассматриваются некоторые из них и обсуждается связь между их свойствами и статистическими свойствами параметров линейного предсказания.

Испытания систем, аналогичных представленной на рис. 9.26, показали, что, если система настроена на определенного диктора, точность распознавания достигает 98—100%, а для произвольного диктора точность составляет около 95% [27, 28].

9.3.3. Меры различимости в пространстве параметров линейного предсказания

Для систем распознавания диктора и речи требуется количественно и достаточно эффективно с вычислительной точки зрения сравнивать два сегмента речевого сигнала, имеющих различные коэффициенты линейного предсказания. Таким образом, необходима мера различимости $D(\mathbf{a}, \hat{\mathbf{a}})$, где D — расстояние между сегментами речи с параметрами линейного предсказания $\mathbf{a}=(1, a(1), a(2), \ldots, a(p))$ и $\hat{\mathbf{a}}=(1, \hat{a}(1), \hat{a}(2), \ldots, \hat{a}(p))$. Поскольку D — расстояние, следует потребовать, чтобы

$$D(\hat{\mathbf{a}}, \mathbf{a}) \geqslant 0 \text{ и } D(\hat{\mathbf{a}}, \mathbf{a}) = 0 \text{ при } \mathbf{a} = \hat{\mathbf{a}}. \qquad (9.17); (9.18)$$

Одну из таких мер различимости $D(\hat{\mathbf{a}}, \mathbf{a})$ предложил Итакура [17]. Эта мера может быть получена на основе следующих рассуждений. Предположим, что вследствие шума и неполной адекватности модели линейного предсказания речи невозможно точно оценить коэффициенты линейного предсказания на соответствующем сегменте сигнала. Оценки (измеренные значения) можно получить лишь приближенно. Предположим, что имеется сегмент сигнала с оценками параметров предсказания $\hat{\mathbf{a}}$. Задача состоит в определении вероятности того, что коэффициенты $\hat{\mathbf{a}}$ являются оценками, полученными на сегменте с истинными параметрами $\mathbf{a}$. Если такая вероятность оценена, то можно получить эффективную меру различимости сегментов.

Манн и Вальд [29] показали, что оценки $\hat{\mathbf{a}}$ распределены по многомерному закону Гаусса со средним $\mathbf{a}$ и ковариационной матрицей $\mathbf{\Lambda}$, определяемой выражением

$$\mathbf{\Lambda} = (\mathbf{R}^{-1}/N)(\hat{\mathbf{a}}\,\mathbf{R}\,\hat{\mathbf{a}}^t), \qquad (9.19)$$

где $\mathbf{R}$ — корреляционная матрица речевого сигнала размером $(p+1)\times(p+1)$; N — протяженность интервала оценивания (числе отсчетов); индекс t означает транспонирование. Таким образом, вероятность получения оценки $\hat{\mathbf{a}}$ при условии, что коэффициенты линейного предсказания соответствуют истинному сигналу $\mathbf{a}$ имеет вид

$$P(\hat{\mathbf{a}}/\mathbf{a}) = [(2\pi)^{p/2}\,|\mathbf{\Lambda}|^{1/2}]^{-1} \exp[-0{,}5\,(\hat{\mathbf{a}}-\mathbf{a})\,\mathbf{\Lambda}^{-1}\,(\hat{\mathbf{a}}-\mathbf{a})^t], \qquad (9.20)$$

где $|\mathbf{\Lambda}|$ — определитель матрицы $\mathbf{\Lambda}$. Соответствующая мера различимости получается, если вычислить логарифмы выражения (9.20) и пренебречь смещением за счет $|\mathbf{\Lambda}|$. Окончательное выражение для меры различимости имеет вид

$$D(\hat{\mathbf{a}}, \mathbf{a}) = (\hat{\mathbf{a}}-\mathbf{a})\left(N\frac{\mathbf{R}}{\hat{\mathbf{a}}\,\mathbf{R}\,\hat{\mathbf{a}}^t}\right)(\hat{\mathbf{a}}-\mathbf{a})^t. \qquad (9.21)$$

Чем больше вероятность того, что $\hat{\mathbf{a}}$ получено из распределения с истинными параметрами $\mathbf{a}$, тем меньше расстояние, вычисленное с использованием меры различимости (9.21). С целью преодоления вычислительных трудностей Итакура предложил весьма похожую меру различимости

$$D'(\hat{\mathbf{a}}, \mathbf{a}) = \log\left(\frac{\mathbf{a}\,\mathbf{R}\,\mathbf{a}^t}{\hat{\mathbf{a}}\,\mathbf{R}\,\hat{\mathbf{a}}^t}\right). \qquad (9.22)$$

Предложение, лежащее в основе проведенного анализа, заключается в том, что одним и тем же звукам речи соответствуют одни и те же параметры линейного предсказания на любом сегменте. Различия в оценках параметров линейного предсказания для таких сегментов целиком относят за счет статистической природы обрабатываемого сигнала. Для большого числа систем такое предположение вполне справедливо. Однако в том случае, когда коэффициенты линейного предсказания изменяются вследствие некоторых эффектов, таких, как замена диктора, коартикуляция и т. д., истинные параметры также меняются. Эти изменения лучше описывать через статистическое распределение с некоторым средним значением.

Таким образом, для полного описания некоторого сегмента речевого сигнала необходимо определить распределение $\mathbf{a}$. Целесообразно предположить, что $\mathbf{a}$ является гауссовым со средним значением $\mathbf{m}$ и ковариационной матрицей $\mathbf{S}$. На основе такого описания $\mathbf{a}$ расстояние между $\hat{\mathbf{a}}$ и $\mathbf{a}$ определяется выражением

$$\hat{D}(\hat{\mathbf{a}}, \mathbf{a}) = (\hat{\mathbf{a}}-\mathbf{m})\,\mathbf{C}^{-1}\,(\hat{\mathbf{a}}-\mathbf{m})^t, \qquad (9.23)$$

где $\mathbf{C}$ — полная ковариационная матрица вида

$$\mathbf{C} = \mathbf{S} + (R^{-1}/N)(\hat{\mathbf{a}}\,\mathbf{R}\,\hat{\mathbf{a}}^t). \qquad (9.24)$$

Для использования меры различимости (9.23) необходимо оценить величины $\mathbf{m}$ и $\mathbf{S}$ для каждого интервала анализа и каждого

эталона. Величина $m=(1, m(1), m(2), ..., m(p))$ является средним значением **a** и определяется выражением

$$m(n)=\frac{1}{y}\sum_{j=1}\hat{a}_j(n),\ n=1, 2, \ldots, p, \qquad (9.25)$$

где $\hat{a}_j(n)$, $j=1, 2, ..., j$ — оценки параметров из выборки с одним и тем же распределением **a**. Аналогично ковариационная матрица **S** с элементами $s(n, p)$ имеет вид

$$s(n, p)=\frac{1}{y}\sum_{j=1}\hat{a}_j(n)\hat{a}_j(p)-m(n)m(p). \qquad (9.26)$$

9.3.4. Система распознавания с большим объемом словаря

Третья из описываемых здесь систем распознавания обладает словарем, объем которого значительно превосходит объемы словарей двух первых систем. Однако платой за увеличение объема словаря является то, что система перестает быть не зависимой от диктора, т. е. система должна быть предварительно обучена применительно к каждому предполагаемому пользователю. С учетом обсуждения, проведенного во введении, разработанная Итакурой [17] система с большим словарем обладает следующими свойствами:

1. Словарь состоит из изолированных слов, количество которых составляет 100—500.
2. Система предназначена для одного диктора, но после соответствующего обучения может быть настроена на любого диктора.
3. Отсутствуют ограничения на пол и возраст диктора.
4. Отсутствуют жесткие ограничения на условия произнесения.
5. Система работает с сигналом телефонного качества.
6. Предусмотрено обучение системы в виде одно- или многократного произнесения каждого слова словаря.
7. Форматом произнесения являются слова, разделенные паузами.

На рис. 9.29 представлена структурная схема обработки сигнала в системе распознавания слов. Для повышения эффективности и снижения объема вычислений Итакура использовал частоту дискретизации 6,67 кГц. Поскольку полоса частот входного сигнала составляет 3 кГц, такая частота дискретизации вполне подходит для данного случая.

После определения моментов начала и окончания слов на основе использования методов обработки во временной области (см. гл. 4) оцениваются первые восемь коэффициентов корреляции со скоростью 67 раз/с. Для компенсации искажений спектра, вносимых телефонной линией, Итакура вычислял спектр, усредненный на большом интервале времени, что достигалось усреднением коэффициентов корреляции по всей фразе и подгонкой к усредненному по фразе спектру двухполюсной модели. Параметр-

ры двухполюсной модели использовались для построения обратного фильтра. Средний по фразе спектр затем нормировался по входу путем свертки исходных автокорреляционных коэффициентов и коэффициентов корреляции импульсной характеристики обратного фильтра. Первые шесть нормированных автокорреляцион-

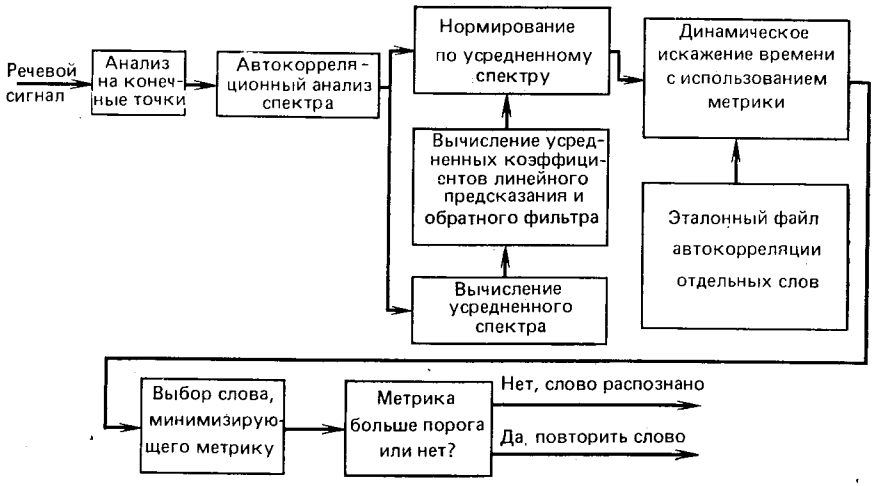


Рис. 9.29. Структурная схема системы распознавания слов, снабженная большим словарем и ориентированная на отдельного диктора

ных коэффициентов использовались затем как для создания эталонных образов, так и для распознавания.

После нормализации спектра начинается процедура распознавания. Неизвестная фраза сравнивается с каждым эталоном из имеющихся в файле. Сравнение происходит на основе меры различимости [см. 9.3.3, ф-ла (9.22)]. Эта мера использовалась также и для динамического согласования временно́го масштаба входной фразы при минимизации расстояния с каждой из эталонных фраз. На основе вычисления расстояний до каждого слова из каталога эталонов выбирается то слово, для которого полученное расстояние минимально. Если абсолютное значение расстояния превышает некоторый порог, то решение не принимается. В этом случае выбирается другое слово с минимальным расстоянием, оно принимается в качестве решения и поступает на выход системы распознавания.

Эта система исследовалась с использованием двух различных словарей. Применяя словарь объемом примерно 120 слов (названия различных городов Японии), Итакура получил частость правильного распознавания, равную 97,3%, а частость отклонения предъявленных слов 1,65%. Для словаря, соответствующего 26 буквам и цифрам от 0 до 9, полученная частость правильного распознавания равна 88,6%. Такое увеличение частости ошибок

(11,4% при частости отклонения 0%) обусловлено большим сходством между некоторыми элементами словаря, например *b* и *d*, *m* и *n* или *i* и *y*.

9.4. Комбинированная система речевого общения с машиной

Чтобы кратко охарактеризовать метод построения системы речевого общения человека и машины в будущем, завершим эту главу элементарным описанием системы, сочетающей верификацию и идентификацию с речевым ответом. Рассмотрим, в частности, экспериментально проверенную систему, используемую при управлении полетами самолетов и продаже авиабилетов[1].

Для описания системы приведем простой «сценарий» ее работы, позволяющий понять, как она может функционировать и быть полезной в различных сферах обслуживания. Клиент авиакомпании вызывает компьютер. Отвечает система речевого ответа.

Компьютер. *Это система информации и обслуживания авиакомпании ABC. Введите, пожалуйста, номер вашего текущего счета.* (Абонент вводит, компьютер проверяет.)

Компьютер. *Доброе утро, мистер XYZ. Пожалуйста, произнесите Вашу контрольную фразу для кредитного обслуживания.* (Клиент произносит, компьютер проверяет ее.)

Компьютер. *Спасибо, мистер XYZ. Я подтверждаю ваш текущий счет. Куда вы желаете полететь?*

Клиент. *Вашингтон.*

Компьютер. *Из какого аэропорта вы желаете лететь на Вашингтон*[2]*?*

Клиент. *Ньюарк.*

Компьютер. *Какого числа вы хотели бы вылететь из Ньюарка?* (Клиент сообщает дату или говорит, что без даты.)

Компьютер. *В какое время (такого-то числа или в такой-то день) вы бы хотели уехать?* (Клиент указывает время.)

Компьютер. *Минуточку, я сейчас дам вам номера рейсов самолетов, которые вылетают в пределах двух часов от указанного времени.* (Компьютер анализирует файл с расписанием, записанным на дисках, выбирает требуемые рейсы и зачитывает их.)

Компьютер. *Какой рейс вы желаете заказать?*

Клиент. *Рейс 123.*

Компьютер. *Вы желаете первый класс или дешевле рейсом 123?*

Клиент. *Первым классом.*

Компьютер. *Сколько мест первого класса вам нужно?*

Клиент. *Одно.*

Компьютер. *Минуточку, пожалуйста. Я проверю наличие одного места.* (Проверяет.)

Компьютер. *Я с удовольствием сообщаю* (повторяет все сначала). *Ваш билет появится на вашем телетайпе в течение часа. Счет будет прислан вам к концу месяца. Желаете ли вы заказать еще что-либо?*

Клиент. *Нет.*

Компьютер. *Благодарим за пользование услугами авиакомпании ABC.* (Отключается.)

Вариант этой системы проходит проверку в лабораториях Белла в коммутируемой сети с обусловленным заранее множеством телефонов и локальным обменом.

[1] Пример взят из статьи Дж. Л. Фланагана.

[2] Отметим два обстоятельства: 1. На каждом шаге ответное сообщение является одновременным подтверждением распознавания на последнем шаге. Если последняя фраза распознана неправильно, то клиент нажимает кнопку и система повторяет последний вопрос. 2. Словарь для распознавания на каждом шаге соответствует множеству возможных ответов на поставленный вопрос. Такая организация словаря существенно повышает устойчивость системы распознавания.

9.5. Заключение

В данной главе предпринята попытка иллюстрировать, каким образом методы цифровой обработки сигналов, рассмотренные в книге, могут быть использованы для построения интересных и полезных систем обработки речи. Чем больше известно о методах цифровой обработки речевых сигналов, тем более сложные системы обработки речевых сигналов можно создать и тем более широкое применение найдут полученные знания как в обычных системах передачи информации, так и в системах общения между человеком и машиной.

Курсовые проекты

Ниже предлагаются темы курсовых проектов по цифровой обработке речевых сигналов для трех основных направлений:

I. Литературные обзоры и доклады.

II. Проекты по реализации.

III. Проекты по моделированию.

I. Литературные обзоры и доклады

Студент должен выбрать тему и рассмотреть следующие вопросы:

1. Суть проблемы.
2. Что является наиболее важным в этой проблеме, например, области применения и т. д.
3. В чем состоит основной подход?
4. Что уже достигнуто в данной области?
5. Необходимы ли новые подходы?
6. Какие проблемы не решены? Что требует дальнейшей проработки?
7. Что необходимо для дальнейшего прогресса, например, технология, необходимость новых фундаментальных исследований и т. д.

Предлагается несколько тем для литературного обзора:

1. Методы выделения основного тона.
2. Методы классификации речи на вокализованную и невокализованную.
3. Влияние телефонного канала на анализ речи.
4. Описание фонетических особенностей английского языка.
5. Фонетические характеристики и моделирование источников звуков для речеобразования.
6. Методы формантного анализа.
7. Синтез речи по правилам.
8. Методы адаптивного квантования.

9. Методы анализа площади поперечного сечения речевого тракта.
10. Методы идентификации дикторов.
11. Машинные системы речевого ответа.
12. Распознавание цифр с помощью ЭВМ.
13. Передача речи в гелиевой среде.
14. Системы обучения глухих речи.
15. Проблема подавления реверберации.
16. Методы подавления отражений.
17. Системы синтеза на основе коэффициентов линейного предсказания.
18. Линейное предсказание и методы идентификации систем.
19. Применение гомоморфной обработки речи.
20. Ускорение и замедление речи.
21. Нуль-полюсный анализ речи.
22. Методы анализа через синтез при обработке речи.
23. Артикуляторная модель речи.
24. Реализация методов кодирования речевой волны.
25. Системы сокращения полосы частот сигнала речи.

II. Проекты по технической реализации

Такие проекты, если это возможно, должны быть доведены до стадии технической реализации или хотя бы до стадии логических схем. Основные вопросы, решаемые в проектах такого типа, состоят в следующем:

1. В чем заключается выбранная Вами задача? Заметим, что проекты такого типа позволят Вам проявить свою изобретательность и придумать новое и более удачное решение какой-либо задачи, которая может быть уже решена.

2. Чем Вы располагаете для решения данной проблемы, например теорией и технологией?

3. В чем заключаются тонкости предлагаемого решения? Это должно быть сделано настолько подробно, насколько это возможно в пределах имеющегося времени.

4. Было ли возможным предлагаемое решение ранее? Если нет, то почему?

5. Какие требования по технической реализации необходимы для внедрения системы?

6. Желательно оценить сложность реализации (в виде количества умножителей, сумматоров, микропроцессоров, памяти и других средств хранения) и примерную стоимость устройства. Темы, предлагаемые для проектов по реализации:

1. Разработать преобразователь ИКМ в АРИКМ, ИКМ в АДМ и т. д.
2. Разработать устройство выделения основного тона.
3. Предложить систему обработки речевых сигналов, которую можно было бы реализовать на широко распространенных микропроцессорах.
4. Разработать устройство обнаружения речевого сигнала в зашумленном телефонном канале.
5. Разработать четырехполосный анализатор спектра речевого сигнала.
6. Разработать систему отображения спектрограмм речи.
7. Разработать параллельный формантный речевой анализатор.
8. Разработать устройство шифрования речи.
9. Разработать цифровое устройство выделения основного тона.
10. Разработать устройство для обнаружения вокализованной и невокализованной речи.
11. Разработать устройство различения речевого сигнала от шума.

III. Проекты по моделированию

Студентам следует браться за эти проекты, если они достаточно хорошо владеют моделированием на ЭВМ и могут получить достаточное количество машинного времени для их реализации. В рамках проектов этого типа требуется кратко описать проблему, включая математическую теорию и цели исследования, распечатки программ (с соответствующей документацией и комментариями), а также результаты контрольного счета. Для этого типа проектов предлагаются следующие темы:

1. Устройства выделения основного тона во временной области (автокорреляционные, кепстральные, на основе линейного предсказания и т. д.).
2. Устройства анализа речи на вокализованную и невокализованную.
3. Устройства определения начала и конца элементов речи.
4. Формантные анализаторы.
5. Системы анализа на основе линейного предсказания — преобразователь сигнала в спектр модели линейного предсказания.
6. N-канальный спектральный анализатор — фазовый и спектральный вокодеры.
7. Кодеры речевого колебания, т. е. АРИКМ (адаптивная разностная ИКМ), АДМ (адаптивная дельта-модуляция) и т. д.
8. Расчет функции площади поперечного сечения голосового тракта.
9. Исследование влияния формы и протяженности временного окна на энергию, автокорреляцию и спектрограмму речевого сигнала.
10. Речевые синтезаторы: спектральные, параллельные, прямые, лестничные.
11. Программа преобразования функции площади поперечного сечения в формантные частоты.
12. Кодопреобразователь между двумя любыми кодовыми форматами.
13. Кепстрально сглаженный спектр для сигнала речи.
14. Преобразование параметров линейного предсказания в другие параметры и исследование их спектральных свойств.
15. Сравнение спектров линейного предсказания, кепстрального и БПФ.

СПИСОК ЛИТЕРАТУРЫ

К главе 1

1. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Tech. J.*, Vol. 27, pp. 623-656, October 1968.
2. J. L. Flanagan, *Speech Analysis, Synthesis, and Perception,* 2nd Edition, Springer Verlag, New York, 1972.
3. J. L. Flanagan, "Computers That Talk and Listen: Man-Machine Communication by Voice," *Proc. IEEE,* Vol. 64, No. 4, pp. 416-432, April 1976.
4. H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. AIEE,* Vol. 47, pp. 617-644, February 1928.
5. H. Dudley, "Remaking Speech," *J. Acoust. Soc. Am.*, Vol. 11, pp. 169-177, 1939.
6. L. R. Rabiner and R. W. Schafer, "Digital Techniques for Computer Voice Response: Implementations and Applications," *Proc. IEEE,* Vol. 64, pp. 416-433, April 1976.
7. C. H. Coker, "A Model of Articulatory Dynamics and Control," *Proc. IEEE,* Vol. 64, No. 4, pp. 452-460, April 1976.
8. B. S. Atal, "Automatic Recognition of Speakers from Their Voices," *Proc. IEEE,* Vol. 64, No. 4, pp. 460-475, April 1976.
9. D. R. Reddy, "Speech Recognition by Machine: A Review," *Proc. IEEE,* Vol. 64, No. 4, pp. 501-531, April 1976.
10. H. Levitt, "Speech Processing Aids for the Deaf: An Overview," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, pp. 269-273, June 1973.

К главе 2

1. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.
2. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.
3. A. Peled and B. Liu, *Digital Signal Processing, Theory, Design and Implementation,* John Wiley and Sons, New York, 1976.
4. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Computation of Complex Fourier Series," *Math Computation,* Vol. 19, pp. 297-381, April 1965.
5. H. D. Helms, "Fast Fourier Transform Method of Computing Difference Equations and Simulating Filters," *IEEE Trans. Audio and Electroacoustics,* Vol. 15, No. 2, pp. 85-90, 1967.
6. T. G. Stockham, "High-Speed Convolution and Correlation," *1966 Spring Joint Computer Conference,* AFIPS Proc., Vol. 28, pp. 229-233, 1966.
7. J. F. Kaiser, "Nonrecursive Digital Filter Design Using the I_0-Sinh Window Function," *Proc. 1974 IEEE Int. Symp. on Circuits and Systems,* San Francisco, pp. 20-23, April 1974.
8. L. R. Rabiner, B. Gold, and C. A. McGonegal, "An Approach to the Approximation Problem for Nonrecursive Digital Filters," *IEEE Trans. Audio and Electroacoustics,* Vol. 19, No. 3, pp. 200-207, September 1971.
9. T. W. Parks and J. H. McClellan, "Chebyshev Approximation for Nonrecursive Digital Filter with Linear Phase," *IEEE Trans. Circuit Theory,* Vol. CT-19, pp. 189-194, March 1972.
10. J. H. McClellan, T. W. Parks, and L. R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-21, pp. 506-526, December 1973.
11. L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR Digital Filter Design Techniques Using Weighted Chebyshev Approximation," *Proc. IEEE,* Vol. 63, No. 4, pp. 595-609, April 1975.
12. A. G. Deczky, "Synthesis of Recursive Digital Filters Using the Minimum p-Error Criterion," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-20, No. 5, pp. 257-263, October 1972.
13. L. R. Rabiner, J. F. Kaiser, O. Herrmann, and M. T. Dolan, "Some Comparisons Between FIR and IIR Digital Filters," *Bell Syst. Tech. J.*, Vol. 53, No. 2, pp. 305-331, February 1974.
14. R. W. Schafer and L. R. Rabiner, "A Digital Signal Processing Approach to Interpolation," *Proc. IEEE,* Vol. 61, No. 6, pp. 692-702, June 1973.
15. L. R. Rabiner and R. E. Crochiere, "A Novel Implementation for FIR Digital Filters," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-23, pp. 457-464, October 1975.
16. R. E. Crochiere and L. R. Rabiner, "Optimum FIR Digital Filter Implementation for Decimation, Interpolation and Narrowband Filters," *IEEE Trans. Acoust. Speech, and Signal Proc.*, Vol. ASSP-23, pp. 444-456, October 1975.
17. R. E. Crochiere and L. R. Rabiner, "Further Considerations in the Design of Decimators and Interpolators," *IEEE Trans. Acoustics, Speech, and Signal Processing,* Vol. ASSP-24, No. 4, pp. 269-311, August 1976.
18. D. J. Goodman, "Digital Filters for Code Format Conversion," *Electronics Letters,* Vol. 11, February 1975.

К главе 3

1. G. Fant, *Acoustic Theory of Speech Production,* Mouton, The Hague, 1970.

2. J. L. Flanagan, *Speech Analysis, Synthesis and Perception,* 2nd Ed., Springer-Verlag, New York, 1972.

3. H. Fletcher, *Speech and Hearing in Communcation,* original edition, D. Van Nostrand Co., New York, 1953. Reprinted by Robert E. Krieger Pub. Co. Inc., New York, 1972.

4. T. Chiba and M. Kajiyama, *The Vowel, Its Nature and Structure,* Phonetic Society of Japan, 1958.

5. I. Lehiste, Ed., *Readings in Acoustic Phonetics,* MIT Press, Cambridge, Mass., 1967.

6. J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer, and N. Umeda, "Synthetic Voices for Computers," *IEEE Spectrum,* Vol. 7, No. 10, pp. 22-45, October 1970.

7. W. Koenig, H. K. Dunn, and L. Y. Lacy, "The Sound Spectrograph," *J. Acoust. Soc. Am.,* Vol. 17, pp. 19-49, July 1946.

8. R. K. Potter, G. A. Kopp, and H. C. Green, *Visible Speech,* D. Van Nostrand Co., New York, 1947. Republished by Dover Publications, Inc., 1966.

9. R. Jakobson, C. G. M. Fant, and M. Halle, *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates,* M.I.T. Press, Cambridge, Mass., 1963.

10. N. Chomsky and M. Halle, *The Sound Pattern of English,* Harper & Row, Publishers, New York, 1968.

11. G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," *J. Acoust. Soc. Am.,* Vol. 24, No. 2, pp. 175-184, March 1952.

12. A. Holbrook and G. Fairbanks, "Diphthong Formants and Their Movements," *J. of Speech and Hearing Research,* Vol. 5, No. 1, pp. 38-58, March 1962.

13. O. Fujimura, "Analysis of Nasal Consonants," *J. Acoust. Soc. Am.,* Vol. 34, No. 12, pp. 1865-1875, December 1962.

14. J. M. Heinz and K. N. Stevens, "On the Properties of Voiceless Fricative Consonants," *J. Acoust. Soc. Am.,* Vol. 33, No. 5, pp. 589-596, May 1961.

15. P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," *J. Acoust. Soc. Am.,* Vol. 27, No. 4, pp. 769-773, July 1955.

16. L. L. Beranek, *Acoustics,* McGraw-Hill Book Co., New York, 1954.

17. P. M. Morse and K. U. Ingard, *Theoretical Acoustics,* McGraw-Hill Book Co., New York, 1968.

18. M. R. Portnoff, "A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal Tract," M. S. Thesis, Dept. of Elect. Engr., MIT, Cambridge, Mass., June 1973.

19. M. R. Portnoff and R. W. Schafer, "Mathematical Considerations in Digital Simulations of the Vocal Tract," *J. Acoust. Soc. Am.,* Vol. 53, No. 1 (Abstract), p. 294, January 1973.

20. M. M. Sondhi, "Model for Wave Propagation in a Lossy Vocal Tract," *J. Acoust. Soc. Am.,* Vol. 55, No. 5, pp. 1070-1075, May 1974.

21. J. S. Perkell, *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study,* MIT Press, Cambridge, Mass., 1969.

22. M. M. Sondhi and B. Gopinath, "Determination of Vocal-Tract Shape from Impulse Response at the Lips," *J. Acoust. Soc. Am.,* Vol. 49, No. 6 (Part 2), pp. 1847-1873, June 1971.

23. B. S. Atal, "Towards Determining Articulator Positions from the Speech Signal," *Proc. Speech Comm. Seminar,* Stockholm, Sweden, pp. 1-9, 1974.

24. R. B. Adler, L. J. Chu, and R. M. Fano, *Electromagnetic Energy Transmission and Radiation,* John Wiley and Sons, Inc., New York, 1963.

25. D. T. Paris and F. K. Hurd, *Basic Electromagnetic Theory,* McGraw-Hill Book Co., New York, 1969.

26. A. M. Bose and K. N. Stevens, *Introductory Network Theory,* Harper and Row, New York, 1965.

27. H. K. Dunn, "Methods of Measuring Vowel Formant Bandwidths," *J. Acoust. Soc. Am.,* Vol. 33, pp. 1737-1746, 1961.

28. J. L. Flanagan and L. L. Landgraf, "Self Oscillating Source for Vocal-Tract Synthesizers," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-16, pp. 57-64, March 1968.

29. J. L. Flanagan and L. Cherry, "Excitation of Vocal-Tract Synthesizer," *J. Acoust. Soc. Am.,* Vol. 45, No. 3, pp. 764-769, March 1969.

30. K. Ishizaka and J. L. Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords," *Bell Syst. Tech. J.,* Vol. 50, No. 6, pp. 1233-1268, July-August 1972.

31. J. L. Flanagan, K. Ishizaka, and K. L. Shipley, "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract," *Bell Sys. Tech J.,* Vol. 54, No. 3, pp. 485-506, March 1975.

32. J. L. Kelly, Jr. and C. Lochbaum, "Speech Synthesis," *Proc. Stockholm Speech Communications Seminar,* R.I.T., Stockholm, Sweden, September 1962.

33. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

34. F. Itakura and S. Saito, "Digital Filtering Techniques for Speech Analysis and Synthesis," *7th Int. Cong. on Acoustics,* Budapest, Paper 25 C1, 1971.

35. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Am.,* Vol. 50, No. 2 (Part 2), pp. 637-655, August 1971.

36. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech,* Springer-Verlag, New York, 1976.

37. H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-21, No. 5, pp. 417-427, October 1973.

38. B. Gold and L. R. Rabiner, "Analysis of Digital and Analog Formant Synthesizers," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-16, pp. 81-94, March 1968.

39. A. E. Rosenberg, "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," *J. Acoust. Soc. Am.,* Vol. 49, No. 2, pp. 583-590, February 1971.

40. L. R. Rabiner, "Digital Formant Synthesizer for Speech Synthesis Studies," *J. Acoust. Soc. Am.,* Vol. 43, No. 4, pp. 822-828, April 1968.

41. G. Winham and K. Steiglitz, "Input Generators for Digital Sound Synthesis," *J. Acoust. Soc. Am.,* Vol. 47, No. 2, pp. 665-666, February 1970.

К главе 4

1. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

2. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

3. J. M. Baker, "A New Time-Domain Analysis of Human Speech and Other Complex Waveforms," Ph.D Dissertation, Carnegie-Mellon Univ., Pittsburgh, PA., 1975.

4. P. J. Vicens, "Aspects of Speech Recognition by Computer," Ph.D. Thesis, Stanford Univ., AI Memo No. 85, Comp. Sci. Dept., Stanford Univ., 1969.

5. L. D. Erman, "An Environment and System for Machine Understanding of Connected Speech," Ph.D. Dissertation, Carnegie-Mellon Univ., Pittsburgh, PA, 1975.

6. L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," *Bell Syst. Tech. J.,* Vol. 54, No. 2, pp. 297-315, February 1975.

7. M. R. Sambur and L. R. Rabiner, "A Speaker Independent Digit-Recognition System," *Bell Syst. Tech. J.,* Vol. 54, No. 1, pp. 81-102, January 1975.

8. J. L. Flanagan, *Speech Analysis, Synthesis and Perception,* 2nd Ed., Springer Verlag, N.Y., 1972.

9. B. S. Atal, "Automatic Speaker Recognition Based on Pitch Contours," *J. Acoust. Soc. Am.,* Vol. 52, pp. 1687-1697, December 1972.

10. A. E. Rosenberg and M. R. Sambur, "New Techniques for Automatic Speaker Verification," *IEEE Trans. Acoust., Speech, and Signal Proc.,* Vol. ASSP-23, pp. 169-176, April 1975.

11. H. Levitt, "Speech Processing Aids for the Deaf: An Overview," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-21, pp. 269-273, June 1973.

12. L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," *IEEE Trans. Acoust., Speech, and Signal Proc.,* Vol. ASSP-24, No. 5, pp. 399-418, October 1976.

13. B. Gold, "Computer Program for Pitch Extraction," *J. Acoust. Soc. Am.,* Vol. 34, No. 7, pp. 916-921, 1962.

14. B. Gold and L. R. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain," *J. Acoust. Soc. Am.,* Vol. 46, No. 2, Pt. 2, pp. 442-448, August 1969.

15. T. P. Barnwell, J. E. Brown, A. M. Bush, and C. R. Patisaul, "Pitch and Voicing in Speech Digitization," Res. Rept. No. E-21-620-74-B4-1, Georgia Inst. of Tech., August 1974.

16. M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average Magnitude Difference Function Pitch Extractor," *IEEE Trans. Acoust., Speech and Signal Proc.,* Vol. ASSP-22, pp. 353-362, October 1974.

17. M. M. Sondhi, "New Methods of Pitch Extraction," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-16, No. 2, pp. 262-266, June 1968.

18. L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," *IEEE Trans. Acoust., Speech and Signal Proc.,* Vol. ASSP-25, No. 1, pp. 24-33, February 1977.

19 J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-Time Digital Hardware Pitch Detector," *IEEE Trans. Acoust., Speech, and Signal Proc.,* Vol. ASSP-24, No. 1, pp. 2-8, February 1976.

20. J. W. Tukey, "Nonlinear (Nonsuperposable) Methods for Smoothing Data," *Congress Record, 1974 EASCON,* p. 673, 1974.

21. L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Applications of a Nonlinear Smoothing Algorithm to Speech Processing," *IEEE Trans. Acoust., Speech, and Signal Proc.,* Vol. ASSP-23, No. 6, pp. 552-557, December 1975.

22. W. A. Blankenship, "Note on Computing Autocorrelation," *IEEE Trans. Acoust., Speech, and Signal Proc.,* Vol. ASSP-22, No. 1, pp. 76-77, February 1974.

23. W. B. Kendall, "A New Algorithm for Computing Autucorrelations," *IEEE Trans. Computers,* Vol. C-23, No. 1, pp. 90-93, January 1974.

24. T. G. Stockham, Jr., "High-Speed Convolution and Correlation," 1966 Spring Joint Computer Conf., AFIPS Conf. Proc., Vol. 28, pp. 229-233, 1966.

К главе 5

1. Robert V. Bruce, *Bell,* Little Brown and Co., Boston, p. 144, 1973.

2. W. B. Davenport, "An Experimental Study of Speech-wave Probability Distributions," *J. Acoust. Soc. Am.,* Vol. 24, pp. 390-399, July 1952.

3. M. D. Paez and T. H. Glisson, "Minimum Mean Squared-Error Quantization in Speech," *IEEE Trans. Comm.*, Vol. Com-20, pp. 225-230, April 1972.

4. P. Noll, "Non-adaptive and Adaptive DPCM of Speech Signals," *Polytech. Tijdschr. Ed. Elektrotech/Elektron* (The Netherlands), No. 19, 1972.

5. H. K. Dunn and S. D. White, "Statistical Measurements on Conversational Speech," *J. Acoust. Soc. Am.*, Vol. 11, pp. 278-288, January 1940.

6. A. V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

7. P. Noll, "A Comparative Study of Various Schemes for Speech Encoding," *Bell System Tech. J.*, Vol. 54, No. 9, pp. 1597-1614. November 1975.

8. N. S. Jayant, "Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantizers," *Proc. IEEE*, Vol. 62, pp. 611-632, May 1974.

9. W. R. Bennett, "Spectra of Quantized Signals," *Bell System Tech. J.*, Vol. 27, No. 3, pp. 446-472, July 1948.

10. B. Smith, "Instantaneous Companding of Quantized Signals, *Bell System Tech. J.*, Vol. 36, No. 3, pp. 653-709, May 1957.

11. J. Max, "Quantizing for Minimum Distortion," *IRE Trans. Inform. Theory*, Vol. IT-6, pp. 7-12, March 1960.

12. P. Noll, "Adaptive Quantizing in Speech Coding Systems," *Proc. 1974 Zurich Seminar on Digital Communications*, Zurich, March 1974.

13. T. P. Barnwell, A. M. Bush, J. B. O'Neal, and R. W. Stroh, "Adaptive Differential PCM Speech Transmission," *RADC-TR-74-177*, Rome Air Development Center, July 1974.

14. A. Croisier, "Progress in PCM and Delta Modulation: Block-Companded Coding of Speech Signals," *Proc. 1974 Zurich Seminar on Digital Communication*, March 1974.

15. N. S. Jayant, "Adaptive Quantization With a One Word Memory," *Bell System Tech. J.*, pp. 1119-1144, September 1973.

16. C. C. Cutler, "Differential Quantization of Communications," U. S. Patent 2, 605, 361, July 29, 1952.

17. R. A. McDonald, "Signal to Noise and Idle Channel Performance of DPCM Systems — Particular Applications to Voice Signals," *Bell System Tech. J.*, Vol. 45, No.7, pp. 1123-1151, September 1966.

18. J. S. Schouten, F. E. DeJager, and J. A. Greefkes, "Delta Modulation, a New Modulation System for Telecommunications," *Philips Tech. Rept.* pp. 237-245, March 1952.

19. F. E. DeJager, "Delta Modulation, a Method of PCM Transmission Using a 1-Unit Code," *Phillips Res. Rep.*, pp. 442-466, December 1952.

20. H. R. Schindler, "Delta Modulation," *IEEE Spectrum*, Vol. 7, pp. 69-78, October 1970.

21. J. E. Abate, "Linear and Adaptive Delta Modulation," *Proc. IEEE*, Vol. 55, pp. 298-308, March 1967.

22. N. S. Jayant, "Adaptive Delta Modulation With a One-Bit Memory," *Bell System Tech. J.*, pp. 321-342, March 1970.

23. J. A. Greefkes, "A Digitally Companded Delta Modulation Modem for Speech Transmission," *Proc. IEEE Int. Conf. Comm.*, pp. 7-33 to 7-48, June 1970.

24. R. Steele, *Delta Modulaton Systems*, Halsted Press, London, 1975.

25. P. Cummiskey, Unpublished work, Bell Laboratories.

26. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," *Bell System Tech. J.*, Vol. 52, No. 7, pp. 1105-1118, September 1973.

27. P. Noll, "Effect of Channel Errors on the Signal-to-Noise Performance of Speech Encoding Systems," *Bell System Tech. J.*, Vol. 54, No. 9, pp. 1615-1636, November 1975.

28. N. S. Jayant, "Step-Size Transmitting Differential Coders for Mobile Telephony," *Bell System Tech. J.*, Vol. 54, No. 9, pp 1557-1582, November 1975.

29. B. S. Atal and M. R. Schroeder, "Adaptive Predictive Coding of Speech Signals,"
Bell System Tech. J., Vol. 49, No. 8, pp. 1973-1986, October 1970.

30. R. W. Stroh, "Optimum and Adaptive Differential PCM," Ph.D. Dissertation, Polytechnic Inst. of Brooklyn, Farmingdale, N.Y., 1970.

31. G. L. Baldwin and S. K. Tewksbury, "Linear Delta Modulator Integrated Circuit With 17-Mbit/s Sampling Rate," *IEEE Trans. on Comm.*, Vol. COM-22, No. 7, pp. 977-985, July 1974.

32. D. J. Goodman, "The Application of Delta Modulation to Analog-to-PCM Encoding," *Bell System Tech. J.*, Vol. 48, No. 2, pp. 321-343, February 1969.

33. S. L. Bates, "A Hardware Realization of a PCM-ADPCM Code Converter," M. S. Thesis, MIT, Cambridge, Mass., January 1976.

34. *Waveform Quantization and Coding*, N. S. Jayant, Editor, IEEE Press, 1976.

К главе 6

1. R. W. Schafer and L. R. Rabiner, "Design of Digital Filter Banks for Speech Analysis," *Bell Syst. Tech. J.*, Vol. 50, No. 10, pp. 3097-3115, December 1971.

2. R. W. Schafer, L. R. Rabiner, and O. Herrmann, "FIR Digital Filter Banks for Speech Analysis," *Bell Syst. Tech. J.*, Vol. 54, No. 3, pp. 531-544, March 1975.

3. J. B. Allen, "Short-Term Spectral Analysis and Synthesis and Modification by Discrete Fourier Transform," *IEEE Trans. Acoustics, Speech, and Signal* Proc., Vol. ASSP-25, No. 3, pp. 235-238, June 1977.

4. J. B. Allen and L. R. Rabiner, "A Unified Theory of Short-Time Spectrum

Analysis and Synthesis," *Proc. IEEE,* Vol. 65, No. 11, pp. 1558-1564, November 1977.

5. J. F. Kaiser, "Nonrecursive Digital Filter Design Using the I_0-SINH Window Function," *Proc. 1974 IEEE Int. Symp. on Circuits and Syst.*, pp. 20-23, April 1974. (Also in *Digital Signal Processing, II,* IEEE Press, 1976.)

6. R. W Schafer and L. R. Rabiner, "Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-21, No. 3, pp. 165-174, June 1973.

7. M. R. Portnoff, "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-24, No.3, pp. 243-248, June 1976.

8. R. K. Potter, G. A. Kopp, and H. G. Kopp, *Visible Speech,* Dover Publications, New York, 1966.

9. R. H. Bolt et al., "Speaker Identification by Speech Spectrograms," *Science,* 166, pp. 338-343, 1969.

10. F. Poza, "Voiceprint Identification: Its Forensic Application," *Proc. 1974 Carnahan Crime Countermeasures Conference,* April 1974.

11. R. W. Schafer and L. R. Rabiner, "System for Automatic Formant Analysis of Voiced Speech," *J. Acoust. Soc. Am.*, Vol. 47, No. 2, pp. 634-648, February 1970.

12. D. W. Tufts, S. E. Levinson, and R. Rao, "Measuring Pitch and Formant Frequencies for a Speech Understanding System," *Proc. 1976 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pp. 314-317, April 1976.

13. R. Koenig, H. K. Dunn, and L. Y. Lacey, "The Sound Spectrograph," *J. Acoust. Soc. Am.*, Vol. 18, pp. 19-49, 1946.

14. A. V. Oppenheim, "Speech Spectrograms Using the Fast Fourier Transform," *IEEE Spectrum,* Vol. 7, pp. 57-62, August 1970.

15. H. F. Silverman and N. R. Dixon, "A Parametrically Controlled Spectral Analysis System for Speech," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-22, No. 5, pp. 362-381, October 1974.

16. M. R. Schroeder, "Period Histogram and Product Spectrum: New Methods for Fundamental Frequency Measurement," *J. Acoust. Soc. Am.*, Vol. 43, No. 4, pp. 829-834, April 1968.

17. A. M. Noll, "Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum, and a Maximum Likelihood Estimate," *Proc. Symp. Computer Proc. in Comm.*, pp. 779-798, April 1969.

18. C. G. Bell, H. Fujisaki, J. M. Heinz, K. N. Stevens, and A. S. House, "Reduction of Speech Spectra by Analysis-by-Synthesis Techniques," *J. Acoust. Soc. Am.*, Vol. 33, pp. 1725-1736, December 1961.

19. E. N. Pinson, "Pitch Synchronous Time Domain Estimation of Formant Frequencies and Bandwidths," *J. Acoust. Soc. Am.*, Vol. 35, No. 8, pp. 1264-1273, August 1963.

20. A. V. Oppenheim, "A Speech Analysis-Synthesis System Based on Homomorphic Filtering," *J. Acoust. Soc. Am.*, Vol. 45, pp. 458-465, February 1969.

21. J. Olive, "Automatic Formant Tracking in a Newton-Raphson Technique," *J. Acoust. Soc. Am.*, Vol. 50, pp. 661-670, August 1971.

22. M. Halle and K. N. Stevens, "Analysis by Synthesis," *Proc. Sem. Speech Compression,* Vol. II, Paper D7, December 1959.

23. M. V Mathews, J. E. Miller and E. E. David, Jr., "Pitch Synchronous Analysis of Voiced Sounds," *J. Acoust. Soc. Am.*, Vol. 33, pp. 179-186, 1961.

24. A. E. Rosenberg, "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," *J. Acoust. Soc. Am.*, Vol. 49, pp. 583-590, 1971.

25. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing,* Chapter 3, pp. 105-123, Prentice-Hall, Englewood Cliffs, N.J., 1975.

26. N. S. Jayant, "Adaptive Delta Modulation With a One-Bit Memory," *Bell Syst. Tech. J.*, Vol. 49, pp. 321-342, 1970.

27. R. E. Crochiere, "On the Design of Sub-Band Coders for Low Bit Rate Speech Communication," *Bell Syst. Tech. J.*, Vol. 65, No. 5, pp. 747-770, May-June 1977.

28. J. L. Flanagan and R. M. Golden, "Phase Vocoder," *Bell Syst. Tech. J.*, Vol. 45, pp. 1493-1509, 1966.

29. J. P. Carlson, "Digitalized Phase Vocoder," *Proc. Conf. on Speech Comm. and Proc.*, Boston, Mass., November 1967.

30. H. Dudley, "The Vocoder," *Bell Labs Record,* Vol. 17, pp. 122-126, 1939.

31. J. L. Flanagan, *Speech Analysis, Synthesis and Perception,* Second Edition, Chapter 8, pp. 321-385, Springer-Verlag, New York, 1972.

32. M. R. Schroeder, "Vocoders: Analysis and Synthesis of Speech," *Proc. IEEE,* Vol. 54, pp. 720-734, May 1966.

33. B. Gold and C. M. Rader, "Systems for Compressing the Bandwidth of Speech," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-15, No. 3, pp. 131-135, September 1967.

34. B. Gold and C. M. Rader, "The Channel Vocoder," *IEEE Trans. Audio and Electroacoustics,* Vol. AU-15, No. 4, pp. 148-160, December 1967.

К главе 7

1. A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear Filtering of Multiplied and Convolved Signals," *Proc. IEEE,* Vol. 56, No. 8, pp. 1264-1291, August 1968.

2. R. W. Schafer, "Echo Removal by Discrete Generalized Linear Filtering," Technical Report 466, Research Lab of Electronics, MIT, February 1969.

3. A. V. Oppenheim, "Superposition in a Class of Nonlinear Systems," Tech. Report No. 432, Research Lab. of Electronics, MIT, Cambridge, Massachusetts, March 1965.

4. B. Bogert, M. Healy, and J. Tukey, "The Quefrency Alanysis of Time Series for Echoes," *Proc. Symp. on Time Series Analysis*, M. Rosenblatt, Ed., Ch. 15, pp. 209-243, J. Wiley, New York, 1963.

5. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Chapter 10, pp. 480-531, Prentice-Hall, Englewood Cliffs, N.J., 1975.

6. J. M. Tribolet, "A New Phase Unwrapping Algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, Vol. ASSP-25, No. 2, pp. 170-177, April 1977.

7. K. Steiglitz and B. Dickinson, "Computation of the Complex Cepstrum by Factorization of the z-Transform," *Proc. 1977 ICASSP*, pp. 723-726, May 1977.

8. A. V. Oppenheim and R. W. Schafer, "Homomorphic Analysis of Speech," *IEEE Trans. on Audio and Electroacoustics*, Vol. AU-16, No. 2; pp. 221-226, June 1968.

9. A. M. Noll, "Cepstrum Pitch Determination," *J. Acoust. Soc. Am.*, Vol. 41, pp. 293-309, February 1967.

10. L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," *IEEE Trans on Acoustics, Speech, and Signal Proc.*, Vol. ASSP-26, No. 1, pp. 24-33, February 1977.

11. R. W. Schafer and L. R. Rabiner, "System for Automatic Formant Analysis of Voiced Speech," *J. Acoust. Soc. Am.*, Vol. 47, No. 2, pp. 634-648, February 1970.

12. J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer, and N. Umeda, "Synthetic Voices for Computers," *IEEE Spectrum*, Vol. 7, No. 10, pp. 22-45, October 1970.

13. L. R. Rabiner, R. W. Schafer, and C. M. Rader, "The Chirp z-Transform Algorithm and Its Application," *Bell System Tech. J.*, Vol. 48, pp. 1249-1292, 1969.

14. J. Olive, "Automatic Formant Tracking in a Newton- Raphson Technique," *J. Acoust. Soc. Am.*, Vol. 50, pp. 661-670. August 1971.

15. A. E. Rosenberg, R. W. Schafer, and L. R. Rabiner, "Effects of Smoothing and Quantizing the Parameters of Formant-Coded Voiced Speech," *J. Acoust. Soc. Am.*, Vol. 50, No. 6, pp. 1532-1538, December 1971.

16. L. R. Rabiner, R. W. Schafer, and J. L. Flanagan, "Computer Synthesis of Speech by Concatentation of Formant-Coded Words," *Bell System Tech. J.*, Vol. 50, No. 5, pp. 1541-1558, May-June 1971.

17. A. V. Oppenheim, "A Speech Analysis-Synthesis System Based on Homomorphic Filtering," *J. Acoust. Soc. Am.*, Vol. 45, pp. 458-465, February 1969.

18. C. J. Weinstein and A. V. Oppenheim, "Predictive Coding in a Homomorphic Vocoder," *IEEE Trans. on Audio and Electroacoustics*, Vol. AU-19, No. 3, pp. 243-248, September 1971.

19. C. R. Patisaul and J. C. Hammett, "Time-Frequency Resolution Experiment in Speech Analysis and Synthesis," *J. Acoust. Soc. Am.*, Vol. 58, No. 6, pp. 1296-1307, December 1975.

К главе 8

1. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.

2. J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, Vol. 63, pp. 561-580, 1975.

3. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Am.*, Vol. 50, pp. 637-655, 1971.

4. F. I. Itakura and S. Saito, "Analysis-Synthesis Telephony Based Upon the Maximum Likelihood Method," *Proc. 6th Int. Congress on Acoustics*, pp. C17-20, Tokyo, 1968.

5. J. Makhoul, and J. Wolf, "Linear Prediction and the Spectral Analysis of Speech," *BBN Report No. 2304*, August 1972.

6. F. I. Itakura and S. Saito, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," *Elec. and Comm. in Japan*, Vol. 53-A, No. 1, pp. 36-43, 1970.

7. J. Makhoul, "Spectral Linear Prediction: Properties and Applications," *IEEE. Trans. on Acoustics, Speech, and Signal Proc.*, Vol. ASSP-23, No. 3, pp. 283-296, June 1975.

8. J. Makhoul, "Spectral Analysis of Speech by Linear Prediction," *IEEE Trans. on Audio and Electroacoustics*, Vol. AU-21, No. 3, pp. 140-148, June 1973.

9. J. D. Markel and A. H. Gray Jr., "On Autocorrelation Equations as Applied to Speech Analysis," *IEEE Trans. on Audio and Electroacoustics*, Vol. AU-21, pp. 69-79, April 1973.

10. V. Zue, "Speech Analysis by Linear Prediction," *MIT QPR No. 105, Research Lab of Electronics*, April 1972.

11. J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-25, No. 5, pp. 423-428, October 1977.

12. J. Burg, "A New Analysis Technique for Time Series Data," *Proc. NATO Advanced Study Institute on Signal Proc.*, Enschede Netherlands, 1968.

13. M. R. Portnoff, V. W. Zue, and A. V. Oppenheim, "Some Considerations in the Use of Linear Prediction for Speech Analysis," *MIT QPR No. 106., Research Lab of Electronics*, July 1972.

14. H. Strube, "Determination of the Instant of Glottal Closure from the Speech Wave," *J. Acoust. Soc. Am.*, Vol. 56, No. 5, pp. 1625-1629, November 1974.

15. S. Chandra and W. C. Lin, "Experimental Comparison Between Stationary and Non-stationary Formulations of Linear Prediction Applied to Speech," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-22, pp. 403-415, 1974.

16. L. R. Rabiner, B. S. Atal, and M. R. Sambur, "LPC Prediction Error-Analysis of Its Variation with the Position of the Analysis Frame," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, Vol. ASSP-25, No. 5, pp. 434-442, October 1977.

17. H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, No. 5, pp. 417-427, October 1973.
18. E. M. Hofstetter, "An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis," *Tech. Note 1973-36, MIT Lincoln Labs,* July 1973.
19. J. D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-20, No. 5, pp. 367-377, December 1972.
20. J. N. Maksym, "Real-Time Pitch Extraction by Adaptive Prediction of the Speech Waveform," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, No. 3, pp. 149-153, June 1973.
21. J. D. Markel, "Application of a Digital Inverse Filter for Automatic Formant and F_0 Analysis," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, No. 3, pp. 149-153, June 1973.
22. J. D. Markel, "Digital Inverse Filtering — A New Tool for Formant Trajectory Estimation," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-20, No. 2, pp. 129-137, June 1972.
23. S. S. McCandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra," *IEEE Trans. on Acoustics, Speech, and Signal Proc.,* Vol. ASSP-22, No. 2, pp. 135-141, April 1974.
24. J. D. Markel and A. H. Gray Jr., "A Linear Prediction Vocoder Simulation Based Upon the Autocorrelation Method," *IEEE Trans. on Acoustics, Speech, and Signal Proc.,* Vol. ASSP-22, No. 2, pp. 124-134, April 1974.
25. R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," *IEEE Trans. on Acoustics, Speech, and Signal Proc.,* Vol. ASSP-23, No. 3, pp. 309-321, June 1975.
26. M. R. Sambur, "An Efficient Linear Prediction Vocoder," *Bell Syst. Tech. J.,* Vol. 54, No. 10, pp. 1693-1723, December 1975.
27. B. S. Atal, M. R. Schroeder, and V. Stover, "Voice-Excited Predictive Coding System for Low Bit-Rate Transmission of Speech," *Proc. ICC,* pp. 30-37 to 30-40, 1975.
28. C. J. Weinstein, "A Linear Predictive Vocoder with Voice Excitation," *Proc. Eascon,* September 1975.

К главе 9

1. J. L. Flanagan, "Computers that Talk and Listen: Man-Machine Communication by Voice," *Proc. IEEE,* Vol. 64, No. 4, pp. 405-415, April 1976.
2. N. R. Dixon and H. D. Maxey, "Terminal Analog Synthesis of Continuous Speech Using the Diphone Method of Segment Assembly," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-16, No. 1, pp. 40-50, January 1968.
3. J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schafer and N. Umeda, "Synthetic Voices for Computers," *IEEE Spectrum,* Vol. 7, pp. 22-45, January 1970.
4. J. Allen, "Synthesis of Speech from Unrestricted Text," *Proc. IEEE,* Vol. 64, No. 4, pp. 433-442, April 1976.
5. N. Umeda, "Linguistic Rules for Text-to-Speech Synthesis," *Proc. IEEE,* Vol. 64, No. 4, pp. 443-451, April 1976.

6. C. H. Coker, "A Model of Articulatory Dynamics and Control," *Proc. IEEE,* Vol. 54, No. 4, pp. 452-459 April 1976.
7. L. H. Rosenthal, L. R. Rabiner, R. W. Schafer, P. Cummiskey, and J. L. Flanagan, "A Multiline Computer Voice Response System Utilizing ADPCM Coded Speech," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* Vol. ASSP-22, No. 5, pp. 339-352, October 1974.
8. L. R. Rabiner and R. W. Schafer, "Digital Techniques for Computer Voice Response: Implementations and Applications," *Proc. IEEE,* Vol. 64, No. 4, pp. 416-433, April 1976.
9. L. R. Rabiner, R. W. Schafer, and J. L. Flanagan, "Computer Synthesis of Speech by Concatenation of Formant-Coded Words," *Bell System Tech. J.,* Vol. 50, No. 5, pp. 1541-1548, May-June 1971.
10. D. S. Levinstone, "Speech Synthesis System Integrating Formant-Coded Words and Computer-Generated Stress Parameters," M.Sc. Thesis, Dept. of Electrical Engr., MIT, Cambridge, 1972.
11. J. P. Olive and L. H. Nakatani, "Rule-Synthesis of Speech by Word Concatenation; A First Step," *J. Acoust. Soc. Am.,* Vol. 55, No. 3, pp. 660-666, March 1974.
12. J. L. Flanagan, L. R. Rabiner, R. W. Schafer, and J. D. Denman, "Wiring Telephone Apparatus from Computer Generated Speech," *Bell System Tech. J.,* Vol. 51, pp. 391-397, February 1972.
13. A. E. Rosenberg, "Automatic Speaker Verification: A Review," *Proc. IEEE,* Vol. 64, No. 4, pp. 475-487, April 1976.
14. G. R. Doddington, "A Method of Speaker Verification," Ph.D. dissertation, Univ. Wisconson, Madison, 1970.
15. R. C. Lummis, "Speaker Verification by Computer using Speech Intensity for Temporal Registration," *IEEE Trans. on Audio and Electroacoustics,* Vol. AU-21, pp. 80-89, 1973.
16. A. E. Rosenberg and M. R. Sambur, "New Techniques for Automatic Speaker Verification," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* Vol. ASSP-23, pp. 169-176, 1975.
17. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* Vol. ASSP-23, No. 1, pp. 67-72, February 1975.
18. B. S. Atal, "Automatic Recognition of Speakers from Their Voices," *Proc. IEEE,* Vol. 64, No. 4, pp. 460-475, April 1976.
19. P. D. Bricker et al., "Statistical Techniques for Talker Identification," *Bell System Tech. J.,* Vol. 50, pp. 1427-1454, April 1971.
20. B. S. Atal, "Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification," *J. Acoust. Soc. Am.,* Vol. 55, pp. 1304-1312, June 1974.
21. A. Newell et al., *Speech Understanding Systems,* Academic Press, New York, 1975.

22. D. R. Reddy, Editor, *Speech Recognition: Invited Papers of the IEEE Symposium,* Academic Press, New York, 1975.

23. T. B. Martin, "Practical Applications of Voice Input to Machines," *Proc. IEEE,* Vol. 64, No. 4, pp. 487-501, April 1976.

24. D. R. Reddy, "Speech Recognition by Machine: A Review," *Proc. IEEE,* Vol. 64, No. 4, pp. 501-531, April 1976.

25. M. R. Sambur and L. R. Rabiner, "A Speaker Independent Digit Recognition System," *Bell System Tech. J.,* Vol. 54, No. 1, pp. 81-102, January 1975.

26. J. Makhoul and J. Wolf, "The Use of a Two-Pole Linear Prediction Model in Speech Recognition," Report 2537, Bolt, Beranek, and Newman, September 1973.

27. L. R. Rabiner and M. R. Sambur, "Some Preliminary Results on the Recognition of Connected Digits," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* Vol. ASSP-24, No. 2, pp. 170-182, April 1976.

28. M. R. Sambur and L. R. Rabiner, "A Statistical Approach to the Recognition of Connected Digits," *IEEE Trans. Acoustics, Speech, and Signal Proc.,* Vol. ASSP-24, No. 6, December 1976.

29. H. B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica;* Vol. 11, Nos. 3 and 4, pp. 173-220, July-October 1943.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

К главе 1

1. **Сапожков М. А.** Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963. 452 с.
2. **Фланаган Дж.** Анализ, синтез и восприятие речи: Пер. с англ./Под ред. А. А. Пирогова. М.: Связь, 1968. 396 с.
3. **Фланаган Дж.** Вычислительные машины говорят и слушают. Речевое общение человека с машиной. — ТИИЭР, 1976, т. 64, № 4.
4. **Харкевич А. А.** Очерки общей теории связи. М.: Гостехиздат, 1955. 268 с.
5. **Шеннон К.** Работы по теории информации и кибернетике: Пер. с англ./Под ред. Р. Л. Добрушина, О. Б. Лупанова с предисловием А. Н. Колмогорова. М.: ИЛ, 1963. 243 с.

К главе 2

1. **Голд Б., Рэйдер Ч.** Цифровая обработка сигналов: Пер. с англ./Под ред. А. М. Трахтмана. М.: Сов. радио, 1973. 367 с.
2. **Введение** в цифровую фильтрацию/Под ред. Р. Богнера, А. Константинидиса: Пер. с англ./Под ред. Л. И. Филиппова. М.: Мир, 1976. 216 с.
3. **Оппенгейм А. В., Шафер Р. В.** Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. М.: Связь, 1979. 416 с.
4. **Рабинер Р., Гоулд Б.** Теория и приложение цифровой обработки сигналов: Пер. с англ./Под ред. Ю. Н. Александрова. М.: Мир, 1978. 848 с.

К главе 3

1. **Бондарко Л. В.** Звуковой строй современного русского языка. М.: Просвещение, 1977. 175 с.
2. **Оппенгейм А. В., Шафер Р. В.** Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. М.: Связь, 1979. 416 с.
3. **Сорокин В. Н.** Потери в речевом тракте. — Акустический журнал, 1977, т. XXIII, № 6, с. 939—946.
4. **Сорокин В. Н.** О роли подглоточной области в процессе речеобразования. — В кн.: Проблемы построения систем понимания речи. М.: Наука, 1980, с. 125.
5. **Прохоров Ю. Н.** Новые модели речевых сигналов и рекуррентное оценивание параметров. — В кн.: Проблемы построения систем понимания речи. М.: Наука, 1980.
6. **Фант Г.** Акустическая теория речеобразования. М.: Наука, 1964. 283 с.
7. **Фланаган Дж.** Анализ, синтез и восприятие речи: Пер. с англ./Под ред. А. А. Пирогова. М.: Связь, 1968. 396 с.
8. **Харкевич А. А.** Основы радиотехники. М.: Связьиздат, 1963. 560 с.

К главе 4

1. **Оппенгейм А. В., Шафер Р. В.** Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. М.: Связь, 1979. 416 с.
2. **Рабинер Р., Гоулд Б.** Теория и применение цифровой обработки сигналов: Пер. с англ./Под ред. Ю. Н. Александрова. М.: Мир, 1978. 848 с.
3. **Фланаган Дж.** Анализ, синтез и восприятие речи: Пер. с англ./Под ред. А. А. Пирогова. М.: Связь, 1968. 396 с.

К главе 5

1. **Джаянт С.** Цифровое кодирование речевых сигналов. Квантизаторы для ИКМ, ДИКМ и ДМ. — ТИИЭР, 1975, т. 62, № 5, с. 83—107.
2. **Дельта-модуляция:** Теория и применение/Венедиктов М. Д., Женевский Ю. П., Марков В. В., Эйдус Г. С. М.: Связь, 1976. 271 с.
3. **Назаров М. В., Пономарев Е. П.** Адаптивная разностная ИКМ с линейным предсказанием на основе лестничного фильтра. — Электросвязь, 1979, № 11, с. 47—51.
4. **Назаров М. В., Прохоров Ю. Н.** и др. Цифровая реализация устройств первичной обработки речевых сигналов с линейным предсказанием. — В кн.: Тезисы докладов 11-го Всесоюзного семинара АРСО-11. Ереван, 1980.
5. **Оппенгейм А. В., Шафер Р. В.** Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. М.: Связь, 1979. 416 с.
7. **Эбейт Дж.** Линейная и адаптивная дельта-модуляция. — ТИИЭР, 1967, т. 55, № 3, с. 59—71.

К главам 6, 7

1. **Вокодерная** телефония/Под ред. А. А. Пирогова. М.: Связь, 1974. 535 с.
2. **Куля В. И.** Ортогональные фильтры. Киев: Техніка, 1967. 240 с.
3. **Оппенгейм А. В., Шафер Р. В.** Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. М.: Связь, 1979. 416 с.
4. **Рабинер Р., Гоулд Б.** Теория и применение цифровой обработки сигналов. Пер. с англ./Под ред. Ю. Н. Александрова. М.: Мир, 1978. 848 с.
5. **Сапожков М. А.** Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963. 452 с.
6. **Фланаган Дж.** Анализ, синтез и восприятие речи: Пер. с англ./Под ред. А. А. Пирогова. М.: Связь, 1968. 396 с.
7. **Шрёдер М. Р.** Вокодеры: Анализ и синтез речи. — ТИИЭР, 1966, т. 54, № 5, с. 5—21.

К главе 8

1. **Акинфиев Н. Н.** К вопросу построения теории речевых сообщений. — Труды Гос. НИИ МРТП СССР, 1957, вып. 4, с. 3—25.
2. **Баронин С. П.** Спектральный анализ и проблема сокращенного описания речевых сигналов. — В кн.: Спектральный анализ звуков речи и интонации. М. 1969, с. 13—30.
3. **Коротаев Г. А.** Системы анализа и синтеза речевого сигнала с линейным предсказанием. — Зарубежная радиоэлектроника, 1976, № 10, с. 3—14.
4. **Маркел Дж. Д., Грэй А. Х.** Линейное предсказание речи: Пер. с англ./Под ред. Ю. Н. Прохорова, В. С. Звездина. М.: Связь, 1980. 308 с.

5. **Прохоров Ю. Н.** Рекуррентное оценивание параметров речевых сигналов. — В кн.: Распознавание образов: Теория и приложение. М.: Наука, 1977, с. 67.

К главе 9

1. **Вокодерная** телефония/Под ред. А. А. Пирогова. М.: Связь, 1974. 535 с.
2. **Анализ** и распознавание речевых сигналов на ЭВМ. М.: Изд. ВЦ АН СССР, 1975. 167 с.
3. **Гудонавичюс Р. В., Кемешис П. П., Читавичюс А. Б.** Распознавание речевых сигналов по их структурным свойствам. Л.: Энергия, 1977. 61 с.
4. **Винцюк Т. К.** и др. Система реального времени для распознавания слов и слитной речи. — В кн.: Автоматическое распознавание слуховых образов. Материалы Всесоюзной школы-семинара АРСО-10. Тбилиси: Мецниереба, 1978, с. 176—178.
5. **Деркач М.** и др. Восприятие речи в распознающих моделях. Львов: ЛГУ, 1971. 187 с.
6. **Загоруйко Н. Г.** Методы распознавания и их применение. М.: Сов. радио, 1972. 206 с.
7. **Звездин В. С.** Речевое общение человека и ЭВМ. М.: Знание, 1980. 64 с.
8. **Куля В. И., Смирнов Ю. М.** и др. Методы и средства построения систем речевого обмена между ЦВМ и человеком. — В кн.: Автоматическое распознавание слуховых образов. Материалы Всесоюзной школы-семинара АРСО-10. Тбилиси: Мецниереба, 1978, с. 152—154.
9. **Лабутин В. К., Молчанов А. П.** Модели механизмов слуха. М.: Энергия, 1973. 200 с.
10. **Лейтес Р. Д., Соболев В. Н.** Цифровое моделирование систем синтетической телефонии. М.: Связь, 1969. 118 с.
11. **Лобанов Б. М.** Принципы автоматического синтеза интонационных структур. — В кн.: Автоматическое распознавание слуховых образов. Материалы Всесоюзной школы-семинара АРСО-10. Тбилиси: Мецниереба, 1978, с. 158—161.
12. **Махонин В. А.** О психоморфизме в автоматике. М.: Наука, 1971. 127 с.
13. **Описание** и распознавание объектов в системах искусственного интеллекта. М.: Наука, 1980. 87 с.
14. **Проблемы** построения систем понимания речи. М.: Наука, 1980. 144 с.
15. **Распознавание** образов. М.: Наука, 1977. 127 с.
16. **Рамишвили Г. С.** Автоматическое опознавание говорящего по голосу. М.: М.: Радио и связь, 1981. 224 с.
17. **Речевое** общение в автоматизированных системах. М.: Наука, 1975. 130 с.
18. **Турбович И. Т., Гитис В. Г., Маслов О. К.** Опознавание образов. М.: Наука, 1971. 246 с.
19. **Цемель Г. И.** Опознавание речевых сигналов. М.: Наука, 1971. 147 с.
20. **Чистович Л. А., Венцов А. В.** и др. Физиология речи. Восприятие речи человеком. Л.: Наука, 1976. 386 с.

Предметный указатель

ОГЛАВЛЕНИЕ