

С. И. Киркоров

**ЦИФРОВАЯ ОБРАБОТКА
РЕЧИ И ИЗОБРАЖЕНИЯ**

Минск
«Колорград»
2019

УДК 004.93
ББК 32.973.26-018.2
К43

Киркоров, С. И.
К43 Цифровая обработка речи и изображения / С. И. Киркоров. – Минск: Колорград, 2019. – 236 с.
ISBN 978-985-596-419-4.

Рассматриваются методы и алгоритмы цифровой обработки речевых и аудиосигналов, способы повышения качественных показателей современных систем связи.

Издание подготовлено на основе лекций одноименного курса, преподаваемого автором более 5 лет для специальностей: 1-45 01 01 – «Инфокоммуникационные технологии»; 1-45 01 02 – «Инфокоммуникационные системы». Требуется знания по дисциплине «Теория электрической связи» и начальные знания программирования в MATLAB. Возрастные ограничения – от 18 лет, связаны с ориентацией на профессиональных программистов, преподавателей вузов и студентов старших курсов.

УДК 004.93
ББК 32.973.26-018.2

ISBN 978-985-596-419-4

© Киркоров С. И., 2019
© Оформление. ООО «Колорград», 2019

Оглавление

Введение	4
Лекция № 1. Предмет и задачи курса, его взаимосвязь с другими курсами специальности.....	5
Лекция № 2. Цифровое представление речевого сигнала и его параметры	14
Лекция № 3. Анализ речевых сигналов во временной и частотной областях	27
Лекция № 4. Гомоморфная обработка речи	37
Лекция № 5. Сжатие речевых и аудиосигналов на основе модели речеобразования.....	44
Лекция № 6. Сжатие речевых сигналов на основе психоакустической модели слухового анализатора человека.....	67
Лекция № 7. Методы и средства восстановления разборчивости зашумленной речи ...	84
Лекция № 8. Критерии оценивания качества воспроизведения речи.....	99
Лекция № 9. Цифровое представление изображений.....	108
Лекция № 10. Предварительная обработка изображений	123
Лекция № 11. Сжатие изображений на основе психовизуальной модели зрительного анализатора.	144
Лекция № 12. Вейвлет-преобразования речи и изображений	169
Лекция № 13. Компьютерная обработка звуков и изображений (2 часа).	195
Лекция № 14. Критерии оценивания качества воспроизведения изображений (2 часа).	212
Вопросы и варианты ответов по курсу для тестирования. Методика оценки тестирования	218
Приложение А. Основная и дополнительная литература:.....	224
Приложение Б. Тематика реферативных работ.....	225
Приложение В. Содержание учебного материала	227
Сведения об авторе	229

Введение

Студенты при изучении дисциплины «Цифровая обработка речи и изображений знакомятся с методами и алгоритмами цифровой обработки речевых, аудиосигналов, способами повышения качественных показателей современных систем связи.

В соответствии с основными требованиями квалификационной характеристики инженеров электросвязи студент должен уметь разрабатывать и эксплуатировать современные средства цифровой обработки аудио- и видеосигналов в телекоммуникационных системах.

Важнейшим видом услуг систем телекоммуникаций является обеспечение эффективной передачи аудио- и видеосигналов. В связи с этим повышенные требования предъявляются к быстродействию и качеству обработки речевых сигналов и изображений с целью улучшения характеристик современных систем связи.

Программой дисциплины предусматривается изучение как теоретических, так и практических вопросов разработки, эксплуатации и обслуживания средств обработки речевых, аудиосигналов и сигналов изображения.

Теоретические занятия ведутся путем рассмотрения вопросов от общего к частному. Полученные знания углубляются и закрепляются на лабораторно-практических занятиях по курсу, а также во время прохождения специальной эксплуатационной практики. После изучения дисциплины студент должен знать принципы разработки, эксплуатации и обслуживания средств обработки аудио- и видеосигналов.

Дисциплина «Цифровая обработка речи и изображений» базируется на знании студентами разделов обязательных курсов: «Теория электрической связи», «Цифровая обработка сигналов», «Теория электрических цепей».

Лекция № 1. Предмет и задачи курса, его взаимосвязь с другими курсами специальности

Краткая история развития теории и техники ЦОС.

Вся история развития теории и техники ЦОС, как и информатики в целом, напрямую связана с достижениями в области дискретной схемотехники и компьютерных технологий. Дискретизация непрерывной информации во времени и квантование по уровню являются основой ее эффективного кодирования, преобразования, передачи и архивации. С появлением в 40-х годах первых ЭВМ стало реальным возникновение нового фундаментального научного направления вычислительной математики, одним из разделов которой можно считать машинные алгоритмы цифровой обработки данных. Однако ограниченные вычислительные ресурсы используемых в те годы ЭВМ не позволяли проводить обработку данных в реальном времени. Речь могла идти лишь о моделировании реальных процессов.

Положение начало радикально изменяться с появлением в 60-х годах класса малых ЭВМ, ориентированных преимущественно на решение задач управления и обработки данных в реальном времени. Потенциальная возможность обработки, преобразования и передачи аналоговых по природе сигналов цифровыми методами с помощью малых ЭВМ привлекла внимание специалистов, работающих во многих областях, и прежде всего в области связи, гидроакустики и обработки речевых сигналов. С этого времени формулируется круг проблем и задач теории ЦОС как самостоятельного научного направления, которому предстоит в дальнейшем свой многоэтапный путь становления и развития.

Цифровая фильтрация и спектральный анализ.

Основной предметной областью теории ЦОС (1965–1975 гг.) были цифровая фильтрация и спектральный анализ (рисунок 1.1), причем оба направления рассматривались с общей позиции частотных представлений. Общей основой развивающихся направлений был синтез цифровых фильтров частотной селекции. Базовые положения теории ЦОС закладывались и апробировались фактически на теории дискретных систем и теории цепей с использованием известного к тому времени набора машинных алгоритмов и прежде всего алгоритма быстрого преобразования Фурье (БПФ).

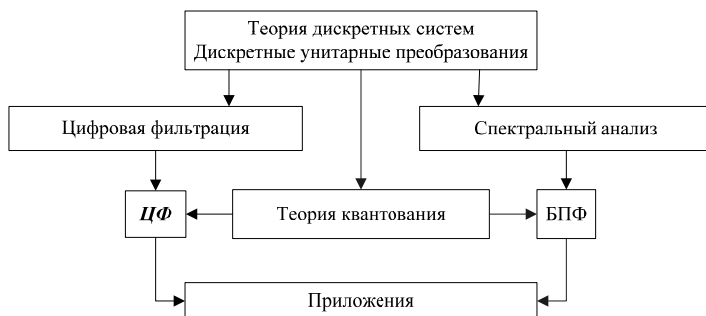


Рисунок 1.1. Взаимосвязь цифровой фильтрации и спектрального анализа

Первые цифровые устройства с позиции сегодняшних представлений обладали низкой эффективностью и имели крайне ограниченное применение, связанное, как правило, с военными технологиями.

Многоскоростная фильтрация и адаптивная обработка сигналов.

В начале 70-х годов начинается новый этап становления техники ЦОС и компьютерных технологий. Открываются новые возможности, и возникают новые проблемы. Теория ЦОС входит в

очередной этап своего развития. Именно в этот период формируются четыре основных взаимосвязанных направления современной теории ЦОС (рисунок 1.2).

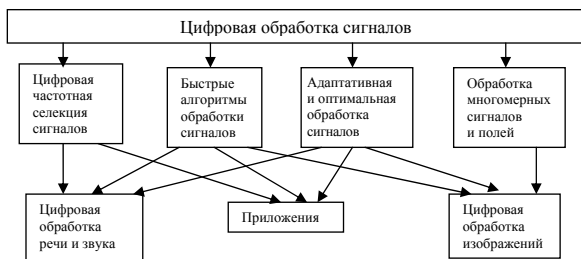


Рисунок 1.2. Основные взаимосвязанные направления современной теории ЦОС

Первое направление – цифровая частотная селекция сигналов, – закрепляет и систематизирует достижения в области проектирования цифровых полосовых фильтров и их наборов. Наиболее оригинальные работы в этом направлении были связаны с развитием теории многоскоростной обработки сигналов на основе эффектов прореживания по времени и по частоте.

Второе направление – быстрые алгоритмы обработки сигналов, – ориентировано на построение высокоскоростных алгоритмов ЦОС путем исключения «избыточности» операций преобразования и замены трудоемких операций умножения операциями сложения и сдвига (многочисленные модификации алгоритма БПФ и методы теоретико-числовых преобразований).

Третье направление – адаптивная и оптимальная обработка сигналов, – охватывает широкий спектр методов решения задач оптимальной фильтрации (фильтры Винера, Калмана и др.) и обработки сигналов в условиях априорной неопределенности о характере исследуемого динамического процесса.

Четвертое направление – обработка многомерных сигналов и полей, – является естественным развитием обработки одномерных сигналов на случай многомерных цифровых систем.

Указанные направления взаимосвязаны друг с другом, и эта взаимосвязь базируется как на общей математической основе, «питающей» все четыре направления, так и на прямом использовании основных положений и методов одних направлений в других. В эти годы успешно начали развиваться системы цифровой обработки изображений и звуковой информации – самостоятельные области науки и техники.

Цифровые методы обработки изображений намного превосходят аналоговые по гибкости и эффективности. При линейной обработке изображений широко применяются различные двумерные унитарные преобразования: Фурье, косинусное, синусное, наклонное, Адамара, Хаара, сингулярное и другие, позволяющие в той или иной степени сократить объем информации. Цифровая обработка изображений применяется для их реставрации и улучшения, выделения признаков деталей изображений, обнаружения и распознавания образов, цифрового внутри- и межкадрового кодирования.

В настоящее время интенсивно развиваются *мультимедийные интерактивные системы*, обеспечивающие работу с неподвижными изображениями и движущимся видео, анимированной компьютерной графикой и текстом, речью и высококачественным звуком. Основная проблема – совместная обработка разнородных данных: цифровых и аналоговых, видео и неподвижных изображений и т. п. В компьютере все данные хранятся в цифровой форме, в то время как теле-, видео- и большинство аудиоаппаратуры имеют дело с аналоговым сигналом. Однако выходные устройства компьютера – мониторы и динамики – имеют аналоговый выход. Поэтому простейший и наиболее дешевый путь построения первых систем мультимедиа состоял в стыковке разнородной аппаратуры с компьютером, предоставлении компьютеру возможностей управления этими устройствами, совмещении выходных сигналов компьютера с видео- и аудиоустройствами и обеспечении их нормальной совместной работы. Дальнейшее развитие мультимедиа происходит в

направлении объединения разнородных типов данных в цифровой форме на одной среде-носителе, в рамках одной системы.

Мультимедиа – это сумма технологий, позволяющих компьютеру вводить, обрабатывать, хранить, передавать и отображать (выводить) такие типы данных, как текст, графика, анимация, оцифрованные неподвижные изображения, видео, звук, речь. Появление систем мультимедиа подготовлено как требованиями практики, так и развитием теории цифровой обработки. Резкий рывок, произошедший в этом направлении за последние несколько лет, обеспечен, прежде всего, развитием технических и системных средств. Это и прогресс в развитии ПЭВМ – резко возросшие объем памяти, быстродействие, графические возможности, характеристики внешней памяти, – и достижения в области видеотехники, лазерных дисков – аналоговых и CD-, DVD-дисков, – а также их массовое внедрение. В настоящий момент внешние носители хранения данных и средства связи позволяют распространять стереоизображение высокого качества по приемлемой для массового потребителя цене. Важную роль в этом сыграла также разработка методов быстрого и эффективного сжатия/развертки данных.

Три основных принципа мультимедиа:

1. Представление информации с помощью комбинации множества воспринимаемых человеком сред (собственно термин происходит от англ. multi – много, и media – среда).
2. Наличие нескольких сюжетных линий в содержании продукта (в том числе и выстраиваемых самим пользователем на основе «свободного поиска» в рамках предложенной в содержании продукта информации).
3. Художественный дизайн интерфейса и средств навигации.

Несомненным достоинством и особенностью технологии являются следующие возможности мультимедиа, которые активно используются в представлении информации:

1. возможность хранения большого объема самой разной информации на одном носителе (до 20 томов авторского текста, около 2000 и более высококачественных изображений, 30–45 минут видеозаписи, до 7 часов звука);
2. возможность увеличения (детализации) на экране изображения или его наиболее интересных фрагментов, иногда в двадцатикратном увеличении (режим «лупа») при сохранении качества изображения. Это особенно важно для презентации произведений искусства и уникальных исторических документов;
3. возможность сравнения изображения и обработки его разнообразными программными средствами с научно-исследовательскими или познавательными целями;
4. возможность выделения в сопровождающем изображении текстовом или другом визуальном материале «горячих слов (областей)», по которым осуществляется немедленное получение справочной или любой другой пояснительной (в том числе визуальной) информации (технологии гипертекста и гипермедиа);
5. возможность осуществления непрерывного музыкального или любого другого аудиосопровождения, соответствующего статичному или динамичному визуальному ряду;
6. возможность использования видеофрагментов из фильмов, видеозаписей и т. д., функции «стоп-кадра», показдрова «пролистывания» видеозаписи;
7. возможность включения в содержание диска баз данных, методик обработки образов, анимации (к примеру, сопровождение рассказа о композиции картины графической анимационной демонстрацией геометрических построений ее композиции) и т. д.;
8. возможность подключения к глобальной сети Internet;
9. возможность работы с различными приложениями (текстовыми, графическими и звуковыми редакторами, картографической информацией);
10. возможность создания собственных «галерей» (выборки) из представляемой в продукте информации (режим «карман» или «мои пометки»);
11. возможность «запоминания пройденного пути» и создания «закладок» на заинтересовавшей экранной «странице»;
12. возможность автоматического просмотра всего содержания продукта («слайд-шоу») или создания анимированного и озвученного «путеводителя-гида» по продукту («говорящей и

показывающей инструкции пользователя»); включение в состав продукта игровых компонентов с информационными составляющими;

- возможность «свободной» навигации по информации и выхода в основное меню (укрупненное содержание), на полное оглавление или вовсе из программы в любой точке продукта.

Области применения речевых технологий

Задача обработки сигналов.

Задача обработки сигналов схематически представлена на рисунке 1.3. В случае речевых сигналов источником информации является человек. Измерению или наблюдению обычно подвергается акустическое колебание. Обработка сигнала предполагает в первую очередь формирование описания на основе некоторой модели с последующим преобразованием полученного представления 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;
- Преобразовать полученное представление в более удобную для решаемой задачи форму.

Цифровая обработка включает как получение дискретных представлений сигнала, так и теорию, расчет и применение цифровых алгоритмов для преобразования полученных дискретных представлений. Первые методы цифровой обработки речевых сигналов имитировали сложные аналоговые системы. 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.

Цифровые методы в настоящее время широко применяются при решении задач обработки речевых сигналов.

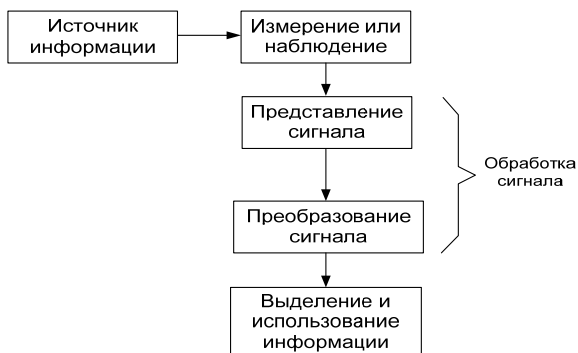


Рисунок 1.3. Схема обработки речевой информации

Способы представления речевых сигналов и их применение.

При рассмотрении вопросов применения цифровой обработки речевых сигналов полезно сконцентрировать внимание на трех основных направлениях:

1. представление речевых сигналов в цифровой форме;
2. цифровая реализация аналоговых методов обработки;
3. методы, основанные исключительно на цифровой обработке.

Представление речевых сигналов в цифровой форме является, конечно, одним из центральных вопросов. Одной из самых основных теорем является теорема дискретизации, или теорема Котельникова, утверждающая, что всякий ограниченный по полосе частот сигнал может быть представлен в виде последовательности равноотстоящих отсчетов, взятых с достаточно высокой частотой. Таким образом, процедура дискретизации лежит в основе теории и приложений цифровой обработки. Существует ряд способов дискретного представления речевых сигналов. Как показано на рисунке 1.4, эти способы могут быть разбиты на две большие группы – цифровое и параметрическое представление речевого колебания.

Цифровое представление речевого колебания 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 речевого колебания подвергается дискретизации и квантованию, а затем обрабатывается для получения параметров модели. Параметры модели: разделяются на параметры возбуждения (относящиеся к источнику звуков речи) и параметры голосового тракта (относящиеся непосредственно к отдельным звукам речи).

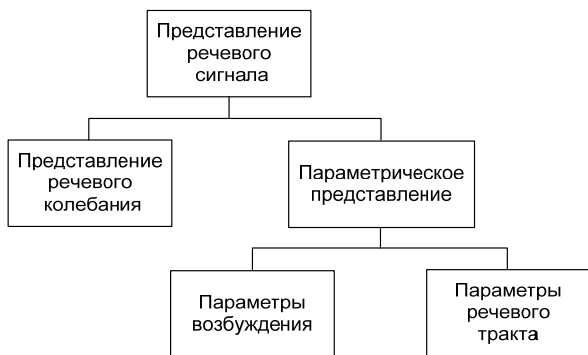


Рисунок 1.4. – Способы представления речевого сигнала

На рисунке 1.5 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. Как следует из рисунка, требуемая скорость передачи информации (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. Скорость передачи далеко не единственный фактор, определяющий выбор типа цифрового представления. Другими факторами являются стоимость, гибкость цифрового представления, качество восприятия речи и т. д.

Наиболее важным фактором, определяющим выбор цифрового представления сигнала и методов цифровой обработки, является специфика решаемой прикладной задачи. На рисунке 1.6 приведено несколько примеров из обширной области передачи и обработки речевых сигналов.



Рисунок 1.5. Диапазон скоростей передачи при различном представлении речевого сигнала



Рисунок 1.6. Области применения речевой связи

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 в 1930-х годах. Целью разработки вокодера являлось уменьшение полосы частот, необходимой для передачи речи. Эта задача актуальна и в настоящее время, несмотря на наличие широкополосных спутниковых, оптических систем связи и т. д. 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.

Системы синтеза речи.

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

Системы речевого обмена между человеком и компьютером можно подразделить на три класса:

1. с речевым ответом;
2. распознавания диктора:
 - верификация диктора;
 - идентификация диктора;
3. 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