

М. Б. Столбов

**ОСНОВЫ АНАЛИЗА И ОБРАБОТКИ
РЕЧЕВЫХ СИГНАЛОВ**



**Санкт-Петербург
2021**

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

М. Б. Столбов

**ОСНОВЫ АНАЛИЗА И ОБРАБОТКИ
РЕЧЕВЫХ СИГНАЛОВ**

УЧЕБНОЕ ПОСОБИЕ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 09.04.02
«Информационные системы и технологии»
в качестве учебно-методического пособия для реализации основных
профессиональных образовательных программ высшего образования
магистратуры

 УНИВЕРСИТЕТ ИТМО

Санкт-Петербург
2021

Столбов М.Б., Основы анализа и обработки речевых сигналов – СПб.: НИУ ИТМО, 2021. – 101 с.

Рецензент:

Рыбин Сергей Витальевич, кандидат физико-математических наук, доцент (квалификационная категория «доцент практики») факультета информационных технологий и программирования, Университета ИТМО.

Пособие адресовано студентам магистратуры, обучающимся по направлению «Информационные системы и технологии» по профилю подготовки «Речевые информационные системы». В пособии изложены основы анализа и обработки речевых сигналов. Материал пособия представляет собой базу для последующего освоения углубленных курсов обработки речевых сигналов. Теоретический материал, подкрепляется вопросами и упражнениями.

Рекомендовано к использованию в Университете ИТМО по направлению подготовки (специальности) 09.04.02 Информационные системы и технологии в качестве учебно-методического пособия.

 **УНИВЕРСИТЕТ ИТМО**

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2021

© Столбов М. Б., 2021

© Коняхин В. В., оформление, 2021

ПРЕДИСЛОВИЕ

В связи с бурным развитием информационных, в частности, речевых технологий (РТ) возникла значительная дифференциация различных отраслей обработки речевых сигналов. К настоящему времени по различным направлениям речевых технологий издано большое число монографий. Для их прочтения необходимы базовые представления о речевых сигналах, методах их анализа и обработки. Такого издания на русском языке в настоящее время нет. Фундаментальные монографии по цифровой обработке речевых сигналов (ЦОРС) были изданы на русском языке около четырех десятилетий назад, когда обработка речевых сигналов еще представлялась единой областью. Предлагаемое пособие призвано восполнить существующий пробел и тем самым способствовать сокращению разрыва в развитии речевых технологий в России и за рубежом.

Цель данного пособия – прежде всего, познакомить читателя с основами анализа и базовыми методами первичной обработки речевых сигналов для дальнейшего использования в решении прикладных задач и углубленного изучения интересующих специальных вопросов. Название книги «Анализ и цифровая обработка речевых сигналов» должно подчеркнуть отличие данного курса от широко распространенных курсов по цифровой обработке сигналов и обозначить тот факт, что для анализа и обработки речевых сигналов в настоящее время должны применяться цифровые средства их обработки. Книга возникла на основе материалов лекций по курсу ЦОРС.

Для чтения данной книги необходимо предварительное знакомство с теорией и алгоритмами цифровой обработки сигналов (свертка, преобразование Фурье, основы линейной алгебры и др.). В пособии рассмотрены те понятия, характеристики дискретных речевых сигналов, методы их анализа и алгоритмы их обработки, которые мы считали важными, исходя из нашего опыта. Рассмотренные вопросы соприкасаются с различными дисциплинами, такими как обработка аудиосигналов, статистика и прикладная математика, физика и другие.

Материал пособия представляет собой базу для дальнейшего освоения углубленных курсов обработки и анализа речевых сигналов (автоматическое распознавание речи, синтез речи и др.). Большая часть теоретических материалов в книге приведена без доказательств, поскольку строгие доказательства привели бы к громоздким математическим выкладкам, что помешало бы восприятию физического смысла. При этом даются ссылки на фундаментальные работы. Углубленное изучение вопросов, затронутых в главе, можно осуществить, обратившись к литературе, список которой приводится в конце пособия. Для облегчения восприятия англоязычной литературы основные термины даются как на

русском, так и на английском языке. В конце пособия помещен англо-русский словарь терминов.

Материал, представленный в каждой главе, подкрепляется вопросами и упражнениями. Кроме того, навыки анализа и обработки РС закрепляются практически в ходе цикла лабораторных работ, выполняемых параллельно с прослушиванием курса лекций.

Пособие адресовано в первую очередь студентам магистратуры, обучающимся по направлению «Информационные системы и технологии» по профилю подготовки «Речевые информационные системы». Материалы пособия будут также полезны специалистам, работающим в области информационных технологий и студентам соответствующих специальностей при изучении курсов, посвященных методам обработки данных.

Подробное изложение многих из затронутых в пособии вопросов дано в переведенных на русский язык книгах Л. Рабинера [47, 48].

В список литературы на русском языке, представленный в конце пособия, включены как научные, так и научно-популярные книги. Надеемся, что знакомство с ними будет способствовать расширению кругозора в области речи и звука.

Структура пособия.

Глава 1 посвящена краткому обзору речевых технологий, месту ЦОРС в системе речевых технологий и обзору информационных источников.

В главе 2 приведены сведения по цифровым речевым сигналам.

В главе 3 излагаются необходимые основы теории линейных систем.

Главы 4 и 5 посвящены вопросам оценки базовых характеристик речевых сигналов.

В главе 6 рассматриваются простейшие и наиболее распространенные алгоритмы фильтрации речевых сигналов.

В заключение я приношу благодарность моим коллегам и друзьям, которые оказали мне неоценимую помощь при написании книги.

1 ВВЕДЕНИЕ В ПРЕДМЕТ

1.1 Краткая история речевых технологий

Речевые технологии (РТ) – это технологии, связанные с анализом и обработкой речевой информации. История развития речевых технологий тесно связана с эволюцией методов обработки речевых сигналов (РС).

На *первом этапе* развития РТ основной задачей являлось создание механических устройств воспроизведения звуков человеческой речи.

13 век. Немецкий философ Альбертус Магнус и английский учёный Роджер Бэкон пытались создать говорящие головы.

1769 г. Немецкий врач, механик и физик Христиан Кратценштейн конструирует резонансные полости, способные формировать некоторые звуки речи (гласные).

1791 г. Вольфганг фон Кёмпелен конструирует механический синтезатор речи (говорящая машина, способная формировать гласные и согласные).

1831 г. Чарльз Уитстон на основе идей Кёмпелена конструирует более совершенную говорящую машину.

На *втором этапе* (переход от механических к электрическим устройствам) развития РТ двигателем прогресса явилась разработка устройств передачи, записи и воспроизведения звука). Здесь можно выделить следующие достижения:

- телефон, Александр Белл, 1876 г.;
- фонограф с записью на восковых валиках, Томас Эдисон, 1877 г.;
- граммофон с записью на пластинках, Эмиль Берлингер, 1877 г.;
- радио, Александр Попов, 1895 г.;
- диктофон с записью на восковые цилиндры, Александр Белл, 1907 г.;
- патефон, Шарль Кро, 1907 г.;
- терменвокс, Лев Термен, 1920 г.;
- говорящая бумага, Борис Скворцов, 1930 г.;
- варифон-электрооптический синтезатор звуков, Евгений Шолло, 1931 г.;
- VOCODER – устройство электронного анализа-синтеза речи, Хомер Дадли, 1939 г.;
- беспроводная аналоговая телефония, 70–80 е годы.

На *третьем этапе* (70-годы – начало эры цифровой обработки речевых сигналов) развитие РТ значительно ускорилось. В начале 3-го этапа основными направлениями РТ стали следующие [48]:

- цифровая передача РС (цифровая телефония);
- синтез речи;
- верификация и идентификация дикторов;
- распознавание речи;
- улучшение качества РС (компенсация искажений);
- устранение дефектов речи.

В последующий период в связи с появлением компьютеров и микропроцессоров, обеспечивающих хранение больших объемов информации, быструю и гибкую обработку данных, РТ претерпели революционные изменения. В настоящее время сфера РТ быстро расширяется. В нее включаются когнитивные технологии, связанные с анализом и интерпретацией звуковых и акустических образов (анализ акустических сцен, обнаружение акустических событий, распознавание эмоций, интерпретация речевых сообщений и многие другие).

Современное состояние РТ можно охарактеризовать как *четвертый этап* развития РТ. К настоящему времени РТ нашли применение в повседневной жизни и стали неотъемлемой частью информационных технологий, включающих в себя распознавание образов, нейросети, технологии искусственного интеллекта, машинное обучение. Основные применения РТ на современном этапе:

- сжатие и кодирование речевых сигналов;
- верификация и идентификация дикторов, голосовая биометрия;
- речевая, текстовая и акустическая аналитика;
- автоматическое распознавание речи;
- синтез речи и клонирование голоса;
- контроль доступа;
- голосовая помощь и голосовое самообслуживание;
- коррекция нарушений слуха и речи;
- распознавания языка;
- распознавание эмоций и физического состояния;
- компенсация искажений РС;
- анализ акустических сцен, обнаружение и распознавание акустических событий;
- восстановление звукозаписей;
- понимание текстов и естественной речи;
- диалоговое и мультимодальное взаимодействие человека и машин;
- обучение иностранным языкам.

1.2 Предметная область цифровой обработки речевых сигналов

Речь реализуется в знаковой форме (письменность, тексты) и звуковой форме (сигналы). Поэтому РТ включают в себя как обработку звуков, так и обработку знаковой информации. Данное пособие посвящено вопросам обработки речевых сигналов.

Приведем определения базовых понятий сигнала и информации.

Информация – свойство материи, отличное от ее вещественных и энергетических свойств, являющееся содержательной характеристикой отражения. Являясь свойством материи, информация может рассматриваться как величина.

Сигнал (от лат. *Signum* – знак) – материальный носитель информации. Информация передается сигналом за счёт изменения характеристик или параметров сигнала во времени. Поэтому под обработкой сигналов мы будем подразумевать как непосредственно обработку сигналов, так и обработку изменяющихся во времени параметров.

Многие из РТ имеют математический, абстрактный характер и связаны с анализом и обработкой информации. Однако первичным источником аудиоинформации являются акустические сигналы, а первичным этапом преобразования данных является анализ и обработка акустических сигналов. Основное назначение цифровой обработки речевых сигналов (ЦОРС) – преобразование первичной аудиоинформации в форму, пригодную для других областей РТ. Анализ и обработка РС обычно выполняются в цифровой форме, поэтому ЦОРС может рассматриваться как специальный раздел цифровой обработки сигналов (ЦОС).

Целью обработки акустических сигналов является оценка их параметров для дальнейшего использования в решении прикладных задач. Цифровая обработка РС включает следующие задачи: фильтрация сигналов, компенсация шумов и искажений сигналов, сегментация сигналов на информативные участки, оценка параметров. Решение этих задач составляет основной предмет ЦОРС. Перечисленные методы преобразования и анализа акустических сигналов принято относить к первичной обработке (предобработке) сигналов (*Front-End Processing*).

В целом ЦОРС можно охарактеризовать с нескольких позиций:

- первичная обработка и анализ РС;
- элемент РТ;
- специальный раздел цифровой обработки сигналов.

В ЦОРС входит ряд задач, связанных с цепочкой преобразований акустических сигналов. Краткая запись основных этапов преобразования акустически сигналов дана на рисунке 1.1. Схема преобразований приведена на рисунке 1.2.

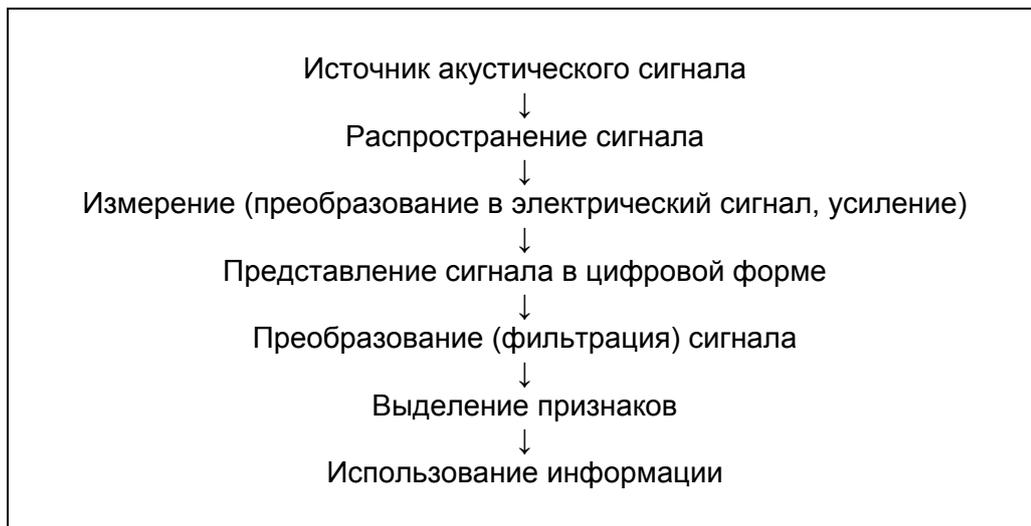


Рисунок 1.1 – Цепочка преобразований акустических сигналов

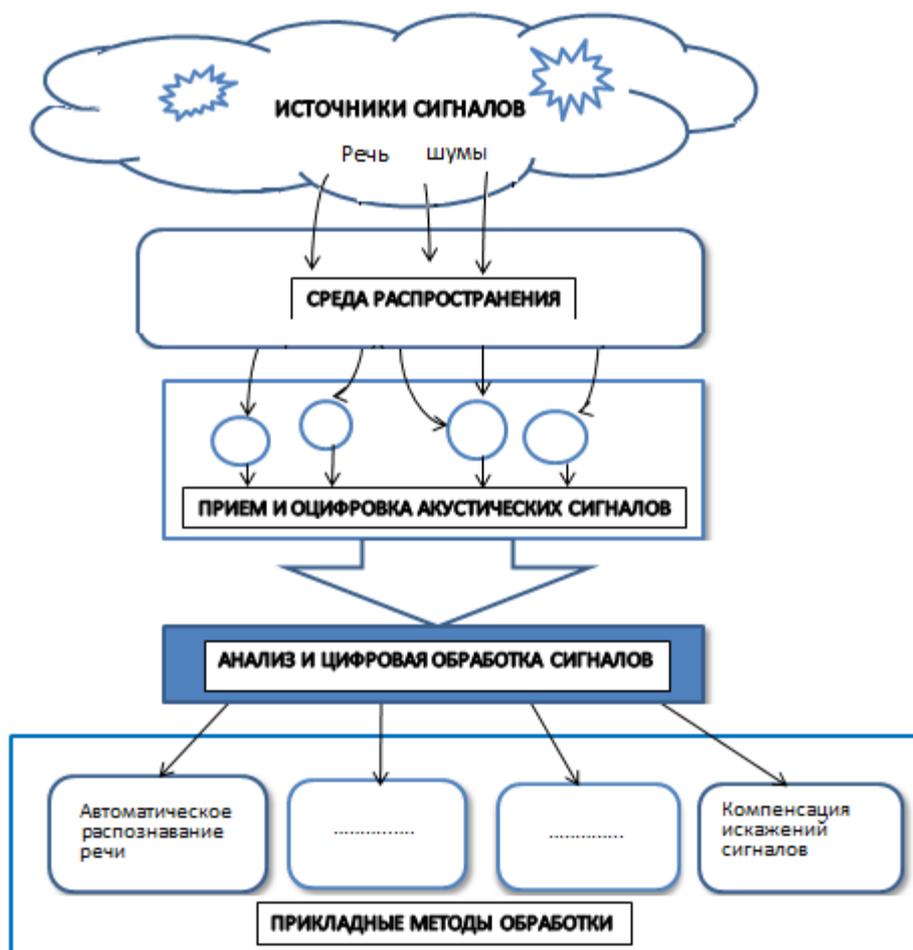


Рисунок 1.2 – Схема преобразований акустических сигналов

С каждым из элементов цепочки преобразований связан круг задач, относящихся к предметной области ЦОРС. Перечислим основные задачи.

Источники акустического сигнала

Акустический сигнал несет в себе информацию об источнике его формирования. В связи с этим ставятся задачи извлечения этой информации. Так, в

речевых сигналах содержится следующая информация:

- акустические фонетические символы, формирующие слова;
- просодика (ритм, интонация);
- гендерная принадлежность;
- возраст;
- акцент;
- индивидуальные признаки диктора;
- эмоции;
- физиологическое и психологическое состояние диктора.

Аналогично формулируются задачи выделения информации о других источниках акустических сигналов (например, в задачах технической диагностики).

Распространение сигнала

Задачи анализа акустической среды, условий формирования и распространения сигнала (например, определение акустического «отпечатка» или «портрета» помещения).

Измерение (преобразование в электрический сигнал, усиление)

Задачи оценки условий записи звука: время записи, характеристики аппаратуры записи, идентификация целостности записи.

Представление в цифровой форме

Задание частоты дискретизации и диапазона квантования аналоговых сигналов, соответствующих практическим задачам извлечения и хранения информации.

Преобразование цифрового сигнала

Задачи фильтрации сигнала, компенсации искажений, восстановления разборчивости и др.

Выделение признаков и использование информации

Сюда относятся задачи прикладных методов обработки: распознавания речи, идентификации диктора и др. Эти задачи относят к прикладной (вторичной) обработке РС (*Back-end processing*).

Обратим внимание еще на две особенности ЦОРС. Первая – это связь ЦОРС с алгоритмами и средствами обработки цифровых сигналов, вторая – это связь понятий и методов ЦОРС с другими отраслями науки.

Развитие технологий ЦОРС неразрывно связано с развитием методов, ал-

горитмов и элементной базы цифровой обработки сигналов. Многие методы ЦОРС получили развитие и практическое воплощение благодаря появлению новых средств обработки сигналов. Развитие методов и средств обработки сигналов проиллюстрировано в таблице 1.1.

Таблица 1.1 – Этапы развития методов и средств обработки сигналов

Годы	Методы обработки	Средства обработки
до 50-х	Аналоговые средства фильтрации	Механические и электронные аналоговые устройства
50-е	Фильтр Винера (1949) Адаптивный фильтр Уидроу (1959)	1-я ЭВМ (UNIAS) на лампах (1951) 1-ый компьютер IBM (1952)
60-70-е	Фильтр Калмана (1961) БПФ (1965) Спектральный анализ, методы синтеза цифровых фильтров	Цифровые схемы на дискретных элементах, первые однокристалльные процессоры
80-е	Методы нелинейного СА, методы спектрального вычитания Алгоритмы адаптивной фильтрации	Технология интегральных схем, процессоры ЦОС (ЦПОС) с фиксированной точкой
90-е	RASTA и другие методы обработки РС Алгоритмы для микрофонных решеток	ЦПОС с плавающей точкой Программируемые логические интегральные схемы
с 2000 по настоящее время	Широкое практическое использование РТ, цифровой звук, цифровая телефония, синтез и распознавание речи и пр.	Универсальные процессоры ЦОС Библиотеки программ ЦОС Международные стандарты ЦОС
Современный этап	Методы машинного обучения, нейронных сетей, искусственного интеллекта	Графические процессоры

Теория анализа и цифровой обработки речевых сигналов опирается на методы и алгоритмы ряда разделов математики, физики и других смежных областей (рисунок 1.3).

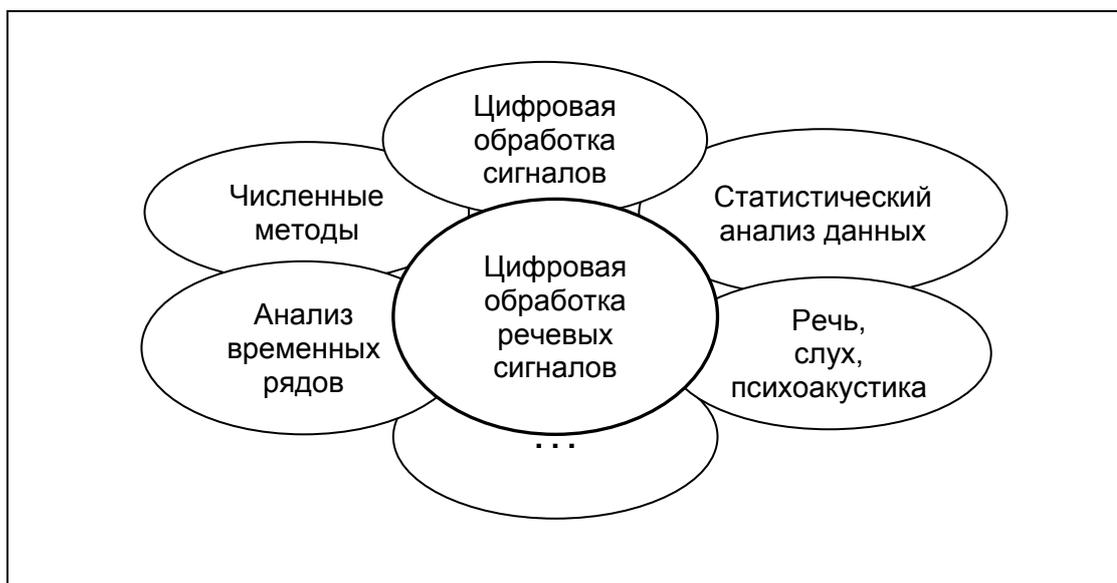


Рисунок 1.3 – Пограничные области ЦОРС

Трудно быть специалистом в РТ, не обладая определенными знаниями в соседних областях. Некоторые книги из этих областей включены в общий список литературы. «Чистые программисты», не имеющие достаточного кругозора, создают плохие программы. Важные мысли на эту тему можно найти в главе «Искусство вычислять для инженеров и ученых» книги Хэмминга [68].

1.3 Профессиональные сообщества и информационные ресурсы ЦОРС

Профессиональные сообщества объединяют ученых и инженеров компаний и университетов, работающих в различных отраслях РТ, и занимаются организацией обмена информацией: конференций, форумов, семинаров, конкурсов, изданием книг и журналов и формированием других информационных ресурсов. Список профессиональных сообществ приведен в приложении А.

Членство в таких сообществах помогает связям с коллегами, участию в мероприятиях и облегчает доступ к информационным ресурсам.

Перечислим некоторые из информационных ресурсов:

- книги (монографии, учебники, пособия, руководства, справочники), содержащие систематическое изложение отдельных тем;
- стандарты, описывающие общепринятые понятия, процедуры, алгоритмы (приложение Б);
- журналы, статьи (аналитические обзоры по конкретной теме, прикладные – опыт применения методов и алгоритмов, научные – новые теоретические и практические результаты) (приложение В);
- труды конференций, семинаров, конкурсов, содержащие текущие научные достижения в определенных областях (приложение Г);
- патенты, предназначенные для охраны авторских прав в области практической реализации методов и алгоритмов;
- базы и хранилища данных, содержащие общий материал для сравнения различных подходов для решения однотипных задач;
- форумы, где обсуждаются практические вопросы обработки сигналов;
- фреймворки и библиотеки программ, облегчающие разработку программных продуктов;
- конкурсы, имеющие целью сравнение передовых достижений по решению определенных задач.

Актуальная информация по РТ также распространяется через онлайн-курсы и практико-ориентированные тренинговые курсы компаний.

В систематизированном виде информация представлена в книгах. Список

научной и популярной литературы по анализу и обработке речевых и звуковых сигналов на русском языке приведен в конце пособия. Подавляющая часть изданий по речевой тематике опубликована на английском языке. В настоящее время по тематике РТ в мире ежегодно издается несколько монографий. Однако, наиболее полным и фундаментальным руководством по ЦОРС на данный момент остается книга Benesty J., Sondhi M., Huang Y. (Eds). Springer Handbook of Speech Processing, Springer Verlag, 2008.

Подавляющая часть периодической литературы по речевой тематике также издается на английском языке. В приложении В приведен список основных журналов. В Российской Федерации материалы по ЦОРС регулярно публикуются в журналах «Речевые технологии» и «Цифровая обработка сигналов», а также «Труды СПИИРАН». Список журналов приведен в приложении В.

1.4 Работа с пособием

Целью пособия является ознакомление с основами цифровой обработки речевых сигналов и освоение базовых практических навыков их анализа и обработки. Материал пособия может служить основой для последующего самостоятельного освоения различных областей речевых и информационных технологий. Пособие содержит элементарные понятия первичного анализа и первичной обработки РС.

Особенности пособия:

- предполагается предварительное знание основ цифровой обработки сигналов;
- широкий охват тем (каждая глава посвящена отдельной теме);
- справочный и обзорный характер материалов. Большинство формул приводятся без доказательств;
- каждый из разделов сопровождается задачами, необходимыми для практического усвоения теоретического материала.

Целью дальнейшего изучения методов анализа и обработки речевых сигналов обусловлен обширный список литературы на русском языке. Мы не рекомендуем по отдельной теме какую-то определенную книгу, полагая, что читатель воспользуется той, которая ему доступна и интересна. Для начала можно рекомендовать книги на русском языке: Бендатта и Пирсона [6, 7], Марпла [32], Отнеса и Эноксона [39], Рабинера и Гоулда [47], Рабинера и Шафера [48].

Для первого знакомств с англоязычной литературой может быть рекомендовано фундаментальное руководство по цифровой обработке речевых сигналов [73], написанное ведущими специалистами в области обработки речевых сигнала-

лов и охватывающее большое число вопросов, связанных с их обработкой. Для облегчения чтения англоязычных текстов в приложении Г дан список англоязычных терминов. Эти термины могут быть также полезны при формировании поисковых запросов по различным задачам ЦОРС.

Для написания текстов по речевой тематике на русском языке рекомендуем использовать приведенные в пособии *перечень сокращений* и *список обозначений*.

Для ознакомления с программами цифровой обработки и анализа РС могут быть рекомендованы пакеты VOICEBOX, PRAAT, НТК, а также книги [2, 31, 32, 39, 54].

Параллельно с изучением теоретического материала целесообразно выполнить курс лабораторных работ, имеющих цель практического освоения работы с аудиосигналами. Указания по лабораторным работам содержатся в методическом указании [59]. Материалы к лабораторным работам также содержат необходимые аудиофайлы. Лабораторные работы могут быть выполнены с использованием доступной версии программы SIS производства компании «Центр речевых технологий» [75] или частично с использованием свободно распространяемого редактора PRAAT [35, 74].

Необходимым элементом курса является самостоятельное решение упражнений (включающих вопросы и задачи), сопровождающих каждую из глав. Только так изложенная информация может превратиться в знания!

Поскольку в этом пособии неоднократно встречаются формулы с комплексными переменными, мы рекомендуем выполнить задания, приведенные ниже.

Вопросы и упражнения

1. X, Y – комплексные числа, $*$ – символ комплексного сопряжения. Правильно ли, что:

$$X Y = Y X$$

$$X^* Y = Y X^*$$

$$X^* Y = Y^* X$$

$$X^* Y = (Y^* X)^*$$

$$|X Y^*| = |X| |Y|$$

$$1/X = X^*/|X|^2$$

$$(XY)^* = X^* Y^*$$

Если $X = X^*$, то X – вещественное

$$X^* X = X^2$$

$$|X^*| = |X|$$

$$|X Y| = |X| |Y^*|$$

2. a, b, g, h, θ – реальные числа. Чему равно:

$$(a + jb) + (g + jh)$$

$$(a + jb) \times (g + jh)$$

$$(a + jb)^2$$

$$(a + jb) / (g + jh)$$

$$|a + jb|$$

$$\exp(j\theta)$$

$$\ln(a + jb)$$

3. Докажите, что:

$$|(a + jb)/(g + jh)| = |(a + jb)|/|(g + jh)^*|$$

$$|(a + jb)^* \times (g + jh)| = |(a + jb) \times (g + jh)^*|$$

$$(a + jb)^* \times (g + jh) = ((a + jb) \times (g + jh)^*)^*$$

$$1 + \exp(j\theta) = 2 \exp(j\theta/2) \cos(\theta)$$

$$1 - \exp(j\theta) = 2 \exp(j\theta/2) \sin(-\theta)$$

2 ЗВУК И ЕГО ЦИФРОВОЕ ПРЕДСТАВЛЕНИЕ

2.1 Звук

Звук – это колебательный процесс, возникающий в воздухе (или другой упругой среде) под действием каких-либо колеблющихся предметов. Источниками звука могут быть, например, голосовые связки человека, струны музыкальных инструментов или любой другой вибрирующий предмет, заставляющий колебаться окружающие его частицы воздуха. Плотность воздушной среды при этом то увеличивается, то уменьшается в соответствии с колебаниями источника звука. В простейшем случае это «чистый» тон (звук камертона), при котором источник излучает только одну частоту, и изменение мгновенных значений колебания строго подчиняются закону синуса.

Основные свойства звука перечислены ниже.

Звук – это колебательный процесс, возникающий в воздухе (или другой упругой среде) под действием каких-либо колеблющихся предметов.

Звук распространяется в пространстве в виде волн (сжатия и разряжения).

Распространение звука описывается волновым уравнением.

Звук описывается тремя основными параметрами: временем, амплитудой (звуковым давлением, интенсивностью), частотой волн.

Амплитуда сжатия и разряжения характеризуется звуковым давлением. Интенсивность звука пропорциональна квадрату амплитуды.

Амплитуды разных волн (сжатия и разряжения) складываются, например, $A + (-A) = 0$.

Волны распространяются в пространстве с определенной скоростью, в разных средах скорость звука разная.

Каждой частоте соответствует длительность по времени (период) и пространству (длина волны): $\lambda_0 = C \times T_0 = C / F_0$.

Синусоидальная звуковая волна за один период колебания T_0 проходит путь, равный длине волны $\lambda_0 = C \times T_0$. Длина волны – расстояние между «гребнями» (фронтами) звуковой волны, различающимися по фазе на один период.

Для одной и той же частоты длина волны в разных средах разная.

Амплитуда (интенсивность) звука в свободном пространстве уменьшается на 6 дБ при удвоении дистанции от источника (закон обратного квадрата).

Акустические приемники преобразуют амплитуду давления в амплитуду электрического сигнала.

Интенсивность (сила звука) – это величина, равная энергии звуковой волны, переносимой через единицу площади в единицу времени:

$$J = E/(t S) = P/S [\text{Дж}/\text{м}^2\text{с} = \text{Вт}/\text{м}^2],$$

где P – мощность звуковой волны.

2.2 Аналого-цифровое преобразование звуковых сигналов

Сигнал – зависимость одной величины от другой (функция). Например, зависимость давления воздуха в точке от времени можно рассматривать как звуковой сигнал. Зависимость напряжения в проводнике от времени тоже может представлять звуковой сигнал [29].

Акустическая энергия звука преобразуется в электрический сигнал с помощью микрофонов. Электрический сигнал преобразуется в цифровой с помощью аналого-цифрового преобразователя (АЦП). Обратное преобразование осуществляется с помощью цифро-аналогового преобразователя (ЦАП).

Аналого-цифровое преобразование – это дискретизация сигнала по времени и амплитуде. Дискретизация по амплитуде называется *квантованием сигнала*. Квантование ограничивает динамический диапазон цифрового сигнала, дискретизация по времени ограничивает частотный диапазон сигнала. На рисунке 2.1 представлена схема аналого-цифрового преобразования сигнала.

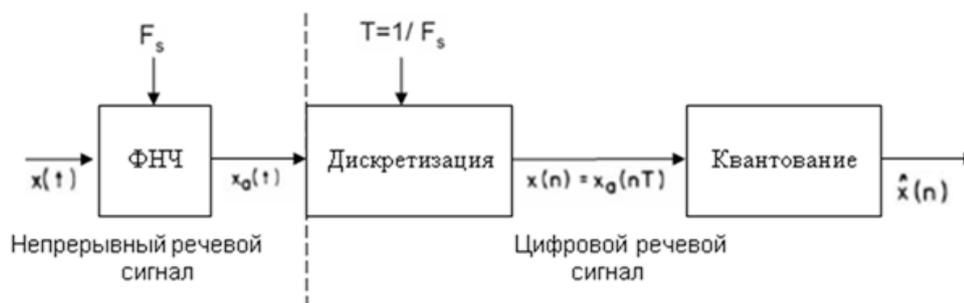


Рисунок 2.1 – Схема аналого-цифрового преобразования сигнала

2.3 Дискретизация во времени

Дискретизация непрерывного аналогового сигнала выполняется через равные промежутки T : $x[i] = x(iT)$. Выборочная функция представляет собой последовательность дискретных величин, соответствующих моментам времени (iT) .

Частота дискретизации (sampling rate) измеряется в герцах (Гц). Герцы – это количество отсчетов в секунду: $F_s = 1/T$. Временной шаг дискретизации и частота дискретизации однозначно связаны: $T=1/F_s$.

Пример: $F_s = 8000 \text{ Гц} \leftrightarrow T = 1/8000 = 125 \text{ мкс}$.

Перед дискретизацией частота сигнала ограничивается фильтром низких частот (ФНЧ) с граничной частотой Найквиста $F_n = F_s/2$. информация об исходном непрерывном с сигнале частично будет утрачена (не может быть восстановлена с помощью обратного цифро-аналогового преобразования, а будет восстановлена лишь в диапазоне частот, ограниченном частотой Найквиста).

2.4 Понятие частоты

Частота в герцах (Гц) – это число повторяющихся событий периодического процесса в секунду.

Например:

Частота вращения f – это число циклов в секунду.

Частота дискретизации сигнала F_s – это число отсчетов в секунду.

Помимо герц, используются и другие единицы измерения частоты.

Нормализованная частота (циклов на отсчет) связана с частотой f соотношением:

$$F = f \times T = f / F_s [-1/2, +1/2].$$

Радиальная (угловая) частота – это мера угловой скорости вращения, выражаемая числом радиан в секунду (рад/с). Радиан (безразмерная величина) – это угол, при котором окружность «провернулась» на длину своего радиуса. Угловая частота связана с частотой f соотношением:

$$\omega = 2\pi f.$$

Нормализованная радиальная частота (рад/отсчет) связана с частотой f соотношением:

$$\Omega = 2\pi f / F_s = \omega / F_s = \omega T [-\pi \dots +\pi], \omega = \Omega F_s.$$

В таблице 2.1 приведены соотношения между различными частотами.

Таблица 2.1 – Соотношение между разными частотами

Переменная	Единицы измерения	Соотношение	Интервал
Радиальная частота, ω	<i>рад/с</i>	$\omega = 2\pi f$	$-\infty < \omega < \infty$ $-2\pi F_n < \omega < 2\pi F_n$
Частота вращения, f	<i>циклов в секунду, Гц</i>	$f = F/T$	$-\infty < f < \infty$ $-F_n < f < F_n$
Нормализованная радиальная частота, Ω	<i>рад/отсчет</i>	$\Omega = 2\pi F = 2\pi f / F_s$	$-\pi \leq \Omega \leq \pi$
Нормализованная частота, F	<i>циклов на отсчет</i>	$F = f / F_s = f \times T$	$-1/2 \leq F \leq 1/2$

Выбор частоты дискретизации речевых сигналов и аудиосигналов определяется, как правило, задачами дальнейшего применения цифровой информации. Наиболее распространенные частоты дискретизации – кратные 8 кГц (16, 24, 48, 96 кГц) и кратные 11025 Гц (22,050 и 44,1 кГц). Наиболее часто в речевых приложениях используется частота 16 кГц.

2.5 Эффект наложения спектров

Если частота сигнала больше частоты Найквиста ($F_n = F_s/2$), то информация о сигнале частично будет потеряна (не может быть восстановлена обратным цифро-аналоговым преобразованием). Вместо сигнала с частотой $F_o > F_n$ будет восстановлен сигнал с частотой $F_i = F_s - F_o < F_n$. Этот эффект называется *наложением спектров (aliasing)*. Фильтр низких частот при дискретизации сигналов необходим, чтобы избежать эффекта наложения спектров.

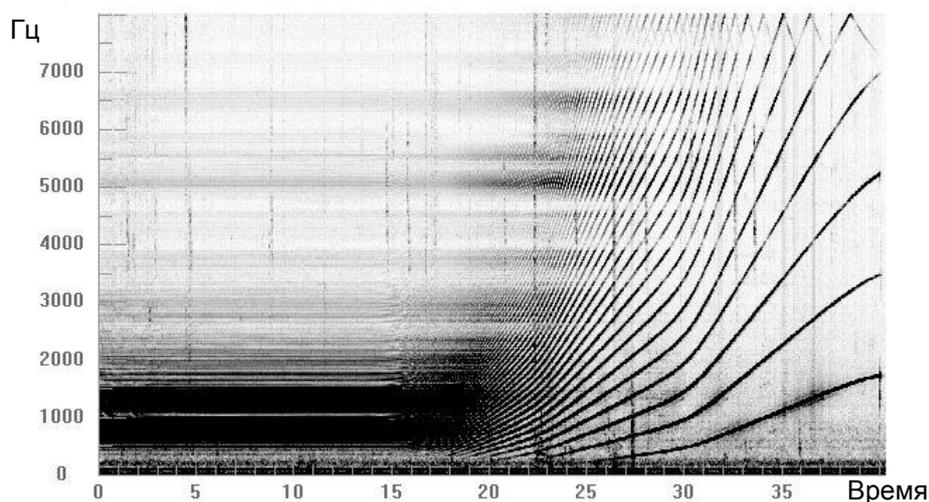


Рисунок 2.2 – Эффект отражения частот

На рисунке 2.2 приведена иллюстрация эффекта наложения спектров. Периодический импульсный сигнал формирует гармоники спектра. Выполняется его дискретизация с частотой $F_s = 16$ кГц. Однако ФНЧ, ограничивающий частоту сигнала частотой Найквиста $F_n = 8$ кГц, не подавляет сигнал в необходимой степени, и часть сигнала с частотой больше 8 кГц преобразуется в цифровой сигнал. При увеличении частоты следования импульсов частота гармоник выходит за границы частоты Найквиста, возникает эффект наложения спектров, который проявляется на спектрограмме в виде отраженных частот.

2.6 Квантование сигналов

Квантование сигналов – это дискретизация амплитуды и представление ее в двоичном коде (битах). При этом считается, что квантованный сигнал имеет целочисленные амплитуды, соответствующие двоичным числам. Обычно используют 8-, 16-, 24- и 32-разрядное квантование (1 байт, 2, 3, 4 байта). В за-

висимости от разрядности квантования (числа бит) диапазон кодируемых амплитуд будет быть разным. Так, амплитуда в интервале $[-2..+1]$ кодируется двумя битами: $(00 = 0, 10 = +1, 11 = -1, 01 = -2)$. В таблице 2.2 приведены интервалы квантования для разного числа разрядов квантования.

Таблица 2.2 – Интервалы квантования для разного числа разрядов квантования

Длина слова, бит	Диапазон чисел
2	-2..+1
8	-128... +127
16	-32768...+ 32767
24	-8388608...+8388607
M	$-2^{M-1} \dots 2^{M-1} - 1$

В некоторых (не во всех) случаях удобно считать, что старшие 16 разрядов кодируют целые амплитуды, а младшие – дробные части амплитуд. Тогда при переходе с 16-разрядного квантования к 24-разрядному диапазоны амплитуд сигналов будут совпадать.

Децибелы – это число, выражающее в логарифмической мере отношение R двух величин (измеряемой и эталонной), употребляется при большом отношении этих величин.

Децибелы для мощностей: $10 \lg(P/P_0)$.

Децибелы для амплитуд: $20 \lg(A/A_0) = 10 \lg(A/A_0)^2 = 10 \lg(P/P_0)$.

Рассмотрим, скольким дБ равно отношению 2. Отношению 10^3 соответствует, согласно определению, значение 60 дБ. Представим 10^3 как степень двойки 2^{10} . Тогда $60 \text{ дБ} \approx 20 \lg(2^{10}) = 10 \times 20 \lg(2)$.

Отсюда следует, что отношение 2 соответствует $20 \lg(2) = 6 \text{ дБ}$.

Тогда, представляя другие отношения в виде степени двойки, можно найти соответствующие им значения дБ.

Примеры.

$$3 \text{ дБ} = 20 \lg(2^{1/2}) \rightarrow R=1,41$$

$$12 \text{ дБ} \rightarrow R= 4$$

$$18 \text{ дБ} \rightarrow R= 8$$

$$2 \text{ дБ} = 20 \text{ дБ} - 18 \text{ дБ} \rightarrow R= 10/8 = 1,25$$

$$8 \text{ дБ} = 20 \text{ дБ} - 12 \text{ дБ} \rightarrow R= 10/4 = 2,5$$

В таблице 2.3 приведены наиболее часто используемые значения дБ для различных отношений амплитуд и соответствующих им мощностей.

Таблица 2.3 – Часто используемые отношения величин в дБ

Отношение амплитуд	1000	100	10	4	2	1	0,7	0,5	0,3	0,1	0,01	0,001
Отношение мощностей	10 ⁶	10 ⁴	100	16	4	0	0,5	0,25	0,1	0,01	10 ⁻⁴	10 ⁻⁶
Значение дБ	60	40	20	12	6	0	-3	-6	-10	-20	-40	-60

Различным разрядностям квантования соответствует различный диапазон изменения амплитуд, который можно выразить в децибелах. В качестве эталонной величины можно принять наименьшее кодируемое число (больше 0), например единицу. *Динамический диапазон* цифровых сигналов можно определить как отношение максимальной амплитуды к минимальной:

$$D = 20 \log_{10} (A_{max}/A_{min}) = \\ = 20 \log_{10} (A_{max}/I) \approx 20 \log_{10} (2^M) \approx 6 \times M \text{ дБ.}$$

Динамический диапазон связан с отношением сигнал-шум (ОСШ) записанного цифрового сигнала. Переход от непрерывных амплитуд к дискретным можно трактовать как аддитивный цифровой шум. Таким образом, оцифрованные сигналы могут быть охарактеризованы некоторой величиной ОСШ. Детали приведены, например, в книге [47].

В таблице 2.4 приведены основные сокращения, связанные с дискретизацией и квантованием сигналов.

Таблица 2.4 – Общепринятые сокращения для цифровых сигналов [17]

Наименование	Обозначение	Обозначение английское
Секунда	с	s
Миллисекунда	мс	ms
Микросекунда	мкс	μs
Герц	Гц	Hz
Килогерц	кГц	kHz
Децибел	дБ	dB

2.7 Цифровой звук и его запись

В литературе обычно проводится строгое разграничение между цифровыми и дискретными сигналами [32]. *Дискретные сигналы* – это сигналы, представляемые в виде счетных последовательностей, амплитуды которых являются непрерывными функциями.

Цифровые сигналы – это сигналы, у которых время и амплитуда дискретны.

Цифровой звук (Digital Sound) – результат дискретизации по времени и квантования по амплитуде сигналов микрофонов, то есть это сигналы, у которых время и амплитуда дискретны.

При записи и передаче цифрового звука важной характеристикой является

цифровой поток, зависящий от частоты дискретизации сигнала и разрядности квантования. Цифровой поток определяется как число бит или байт, передаваемых в секунду:

$$\text{Цифровой поток (Bit rate)} = M \times Fs \text{ [bps]}, \text{ где } bps = \text{bit/s}.$$

Цифровой поток позволяет оценить объем памяти (информации), необходимый для записи аудиоматериала, например при оцифровке аудиоархивов. Объем информации вычисляется по формуле:

$$V = \text{Bit rate} \times T.$$

Объем памяти принято характеризовать в байтах (Б).

Пример 1

Цифровой поток для стандарта MP3 составляет от 32 до 320 Кбит/с, что соответствует 4–40 КБ/с (KBps).

Пример 2

Оценим приблизительно (считая, что 1 КБ равен 1000 байт, а не 1024) объем памяти, необходимый для непрерывной записи аудиосигнала в течение суток.

При записи в формате 16 бит \times 16 кГц получим:

$$16 \text{ бит} \times 16 \text{ кГц} = 2 \text{ байт} \times 16 \text{ кГц} \approx 32 \text{ КБ/с}.$$

$$32 \times 60 = 1920 \text{ КБ/мин} = 1,92 \text{ МБ/мин}.$$

$$1,92 \times 60 = 114 \text{ МБ/час} = 0,114 \text{ ГБ/час}.$$

$$0,114 \times 24 = 2,726 \text{ ГБ/сутки}.$$

Ограничение объема пространства в файловой системе FAT16 составляет 2 ГБ. Поэтому накопителя с FAT16 недостаточно. Однако при записи на частоте 8 кГц карты памяти 2 ГБ хватит больше, чем на сутки.

2.8 Осциллограмма

Основной формой графического представления цифровых сигналов в звуковых редакторах является осциллограмма. Осциллограмма – это графическое представление цифрового сигнала в координатах «Амплитуда» – «Время». В разных звуковых редакторах используются различные осциллограммы. На рисунке 2.3 показаны осциллограммы цифровых сигналов в разных редакторах аудиосигналов.

Примечание

Отметим, что соединить дискретные отсчеты сигнала на графике прямыми линиями не совсем верно, так как при восстановлении непрерывного сигнала

по дискретному используются более сложные интерполирующие функции. Кусочно-постоянная интерполяция представляется более предпочтительной.

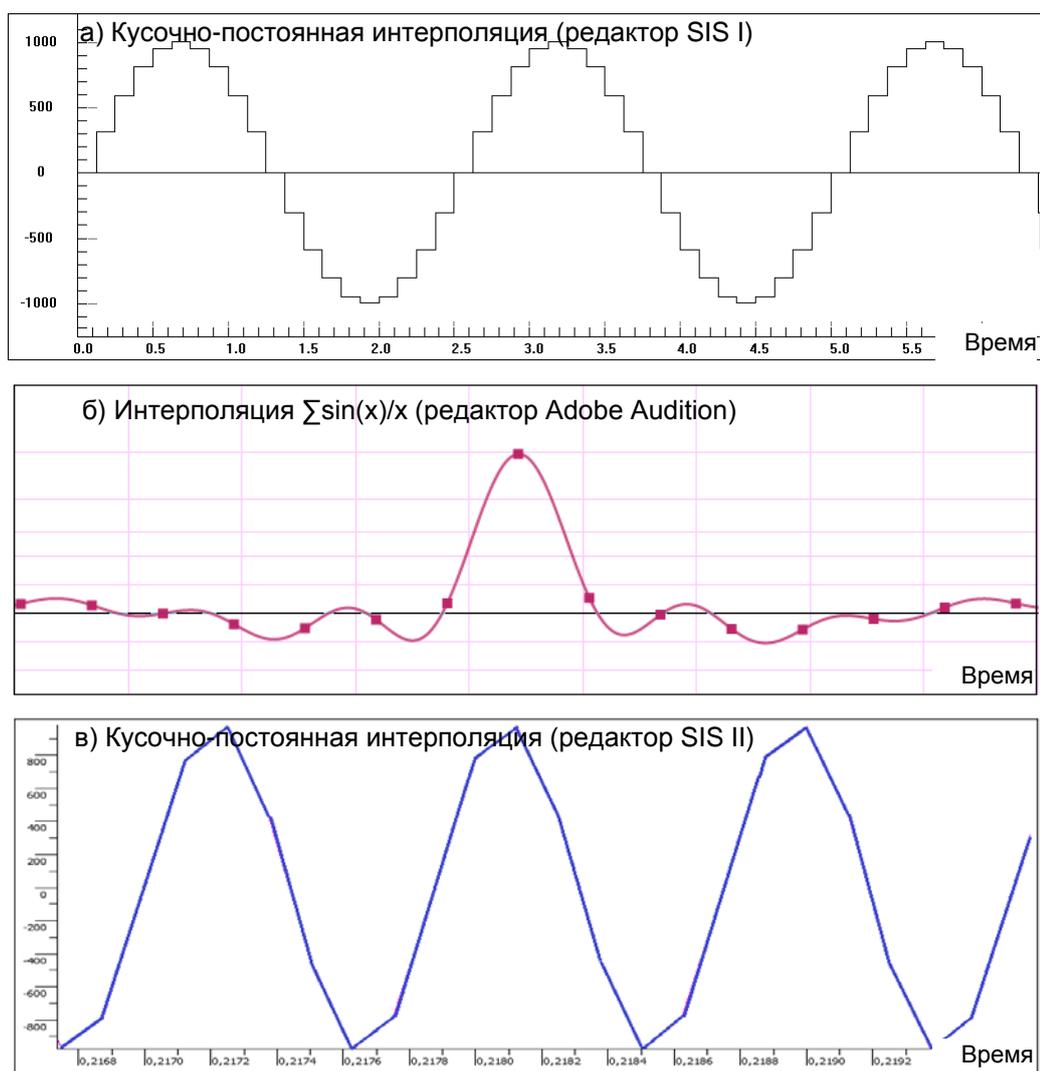


Рисунок 2.3 – Осциллограммы цифровых сигналов в разных редакторах аудиосигналов

Влияние частоты дискретизации на представление синусоидального сигнала в осциллограмме проиллюстрировано на рисунке 2.4.

Примечание

Отметим, что цифровые записи аудиосигнала, записанные на компьютере и цифровом диктофоне с одной той же частотой дискретизации, могут отличаться по длине и объему вследствие различия частот генераторов компьютера и диктофона (двух разных устройств). В заголовке файла будет указана одинаковая частота, например 16 кГц, однако реально она незначительно отличается. Поэтому фонограммы также будут отличаться.

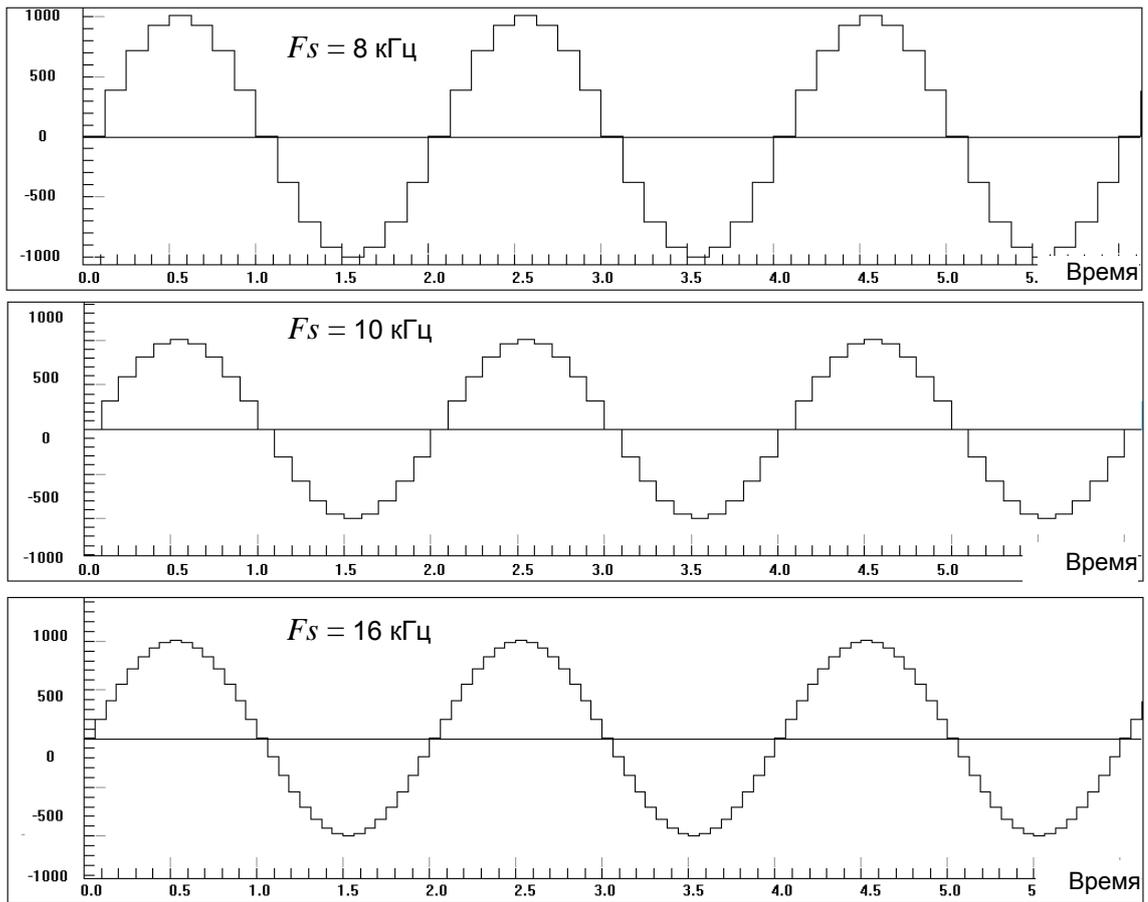


Рисунок 2.4 – Осциллограммы синусоидального сигнала для $F_s = 8$ кГц, 10 кГц, 16 кГц

Эффекты квантования заметны на осциллограмме сигнала небольшой амплитуды при небольшом масштабе времени (рисунок 2.5).

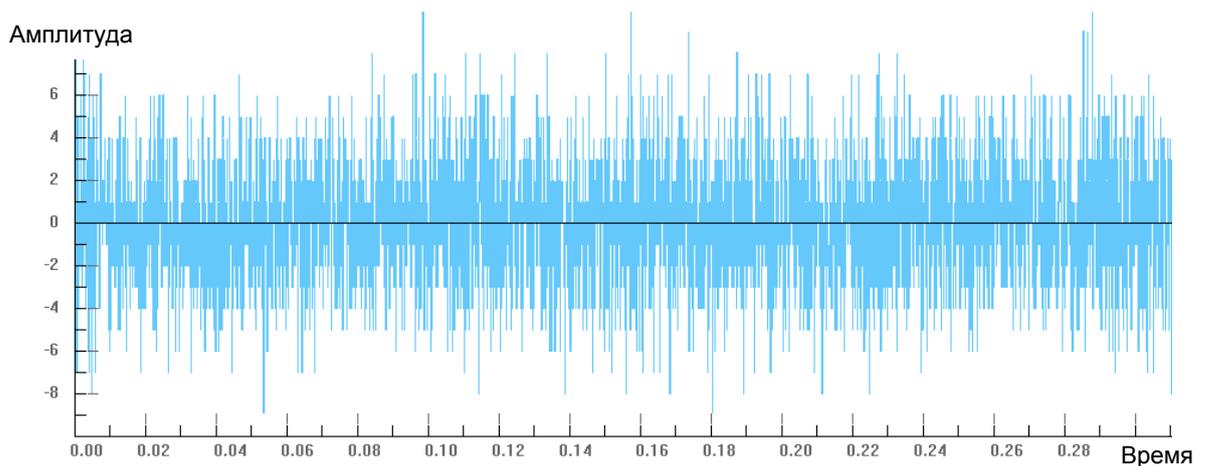


Рисунок 2.5 – Осциллограмма шума малого уровня

Более детально вопросы сбора и оцифровки данных рассмотрены в книге [39].

Вопросы и упражнения

1. В каких единицах измеряется фаза сигнала?
2. Скольким рад/с соответствует частота 500 Гц?
3. Сколько отсчетов сигнала приходится на длину волны 200 Гц при частоте дискретизации 16 кГц?
4. Какому расстоянию, пройденному звуком в помещении, соответствует кадр сигнала N отсчетов с частотой дискретизации 16 кГц?
5. Чему равен временной сдвиг между входным и выходным отсчетами линии задержки в N отсчетов сигнала с частотой дискретизации F_s ?
6. Какой величине соответствует нормализованная частота Найквиста?
7. В каких пределах 8-разрядное двоичное число кодирует целые числа?
8. Какое отношение амплитуд соответствует 8 децибелам?
9. Скольким децибелам соответствует отношение амплитуд, равное 7?
10. Изменится ли динамический диапазон цифровых сигналов при переходе от 16-разрядного квантования к 24-разрядному?
11. Чему равен динамический диапазон 16-разрядного сигнала?
12. Скольким байтам равен один килобайт?
13. Оцените объем записанной за сутки аудиоинформации для формата записи 16 бит \times 10 кГц.
14. Между двумя сигналами одной частоты f имеется фазовый сдвиг $\Delta\Phi(f)$. Каков пространственный сдвиг между этими сигналами в воздушной среде ($c = 334$ м/с)?
15. Расстояние от источника звука до ближайшего слушателя равно D_n , до дальнего слушателя – D_f . На сколько дБ необходимо усилить сигнал дальнего слушателя, чтобы его громкость была такой же, как у ближнего?

3 СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ РЕЧЕВЫХ СИГНАЛОВ

Целью оценивания параметров сигналов является установление соответствия между сигналами и их признаками (параметрами). Другой целью является представление сигнала компактным набором параметров вместо самого оцифрованного сигнала. Данная глава посвящена методам оценивания базовых параметров сигналов, в том числе речевых сигналов.

3.1 Случайная величина и ее характеристики

Статистическая *изменчивость наблюдений* означает то, что результаты повторных измерений случайной величины различны. Измеренные значения можно объединить в группы, которые образуют выборочное *пространство наблюдений* A . Элементы (точки) этого пространства можно группировать в подпространства A_1, \dots, A_k , называемые *событиями*. Эксперимент всегда приводит к одному из событий. Таким образом, выборочное пространство событий может быть представлено конечным множеством групп: $A = \{A_1 + A_2 + \dots + A_k\}$.

Количественной характеристикой некоторого события A_i может служить частота его появления $m(A_i)$. Частота появления события в случае n наблюдений определяется соотношением:

$$P(A_i) = m(A_i)/n.$$

Вероятность появления события определяется как предельное значение частоты при увеличении числа наблюдений:

$$P(A_i) = \lim P(A_i), n \rightarrow \infty.$$

Пусть измеряется некая числовая величина, то есть событием является значение измеренной случайной величины. *Функция распределения вероятностей* $F(\xi)$ случайной величины x определяется следующим соотношением:

$$F(\xi) = P(x \leq \xi) \leq 1.$$

Плотность вероятности случайной величины x – это вероятность распределения ее значений функции распределения по узким интервалам:

$$f(\xi) = \frac{dF(\xi)}{d\xi}.$$

С помощью плотности вероятности можно вычислить вероятность попадания значения случайной величины в заданный интервал значений:

$$P(x \leq a) = \int_{-\infty}^a f(\xi) d\xi = F(a) - F(-\infty) = F(a),$$

$$P(a < x \leq b) = \int_a^b f(\xi) d\xi = F(b) - F(a).$$

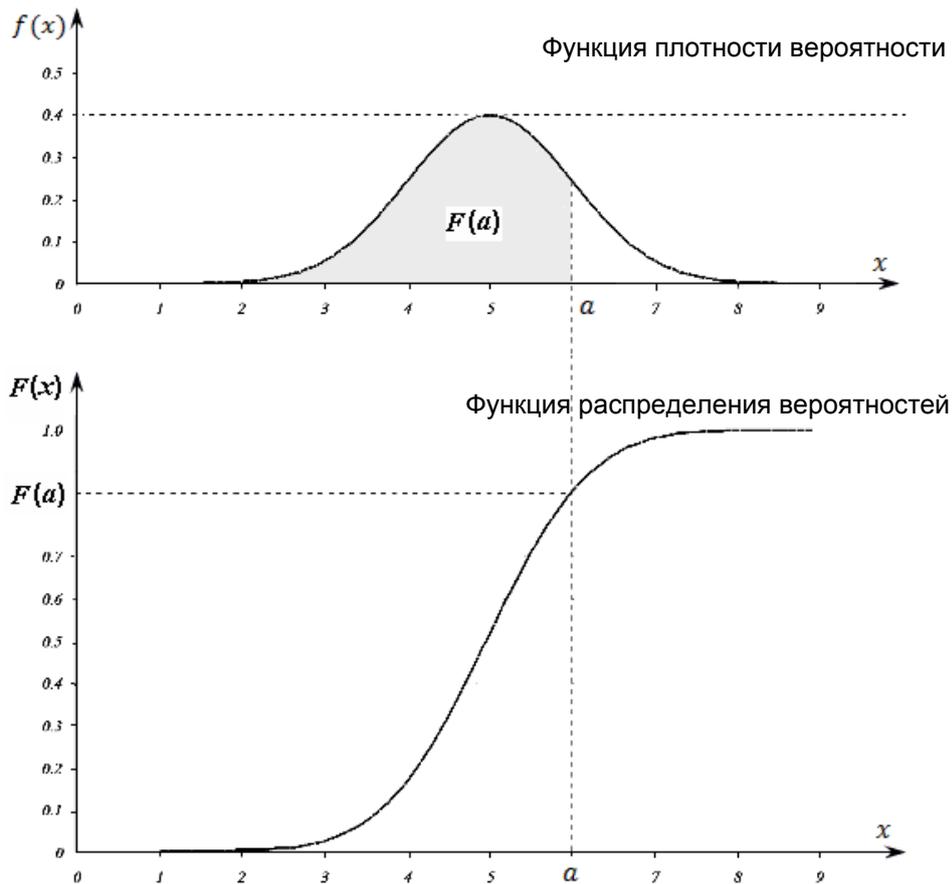


Рисунок 3.1 – Плотность вероятности и функция распределения случайной величины x [8]

Математическое ожидание однозначной непрерывной функции $g(x)$ определяется следующим выражением:

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Если мы знаем вероятность распределения величины по интервалам, то можно найти параметры распределения (характеристики случайных величин): центр тяжести, степень размытости и другие характеристики. Основные характеристики случайной величины приведены ниже.

Математическое ожидание случайной величины x : $\mu_x = E\{x\}$.

Дисперсия D_x (Variance) и среднее квадратическое (среднеквадратическое) отклонение σ_x (CKO, Standard deviation, Std): $D_x = \sigma_x^2 = E\{(x - \mu_x)^2\}$.

Стандартное отклонение (Std): $\sigma_x = D_x^{1/2}$.

Среднеквадратическое значение (Mean Square): $\Psi_x = E\{x^2\} = \mu_x^2 + \sigma_x^2$.

Корень среднеквадратического значения (Root Mean Square, RMS): $[E\{x^2\}]^{1/2}$.

Характеристики случайной случайных величин представляют собой неслучайные величины.

3.2 Случайный процесс и его характеристики

Процесс – это последовательность событий.

Случайный процесс – это функция $x(t)$, мгновенные значения которой есть случайные величины.

Процесс является стационарным, если функция распределения любых n его значений $F(x[i_{+1}], \dots, x[i_{+n}])$ не зависит от начального момента времени i .

Процесс является эргодическим, если усреднение по ансамблю можно заменить усреднением по времени.

Процесс является более общим понятием, чем сигнал, однако в дальнейшем мы будем использовать эти понятия как тождественные.

Рассмотрим частный вариант, когда случайная величина – это отсчеты амплитуды самого процесса. Тогда можно определить его характеристики. При описании случайного процесса обычно ограничиваются характеристиками, аналогичными числовым характеристикам случайных величин. Основные характеристики случайного процесса приведены ниже.

Среднее значение $\mu_x = \lim_{P \rightarrow \infty} 1/2P \int_{-P, P} x(t) dt$

Дисперсия процесса $Dx = \lim_{P \rightarrow \infty} 1/2P \int_{-P, P} [x(t) - \mu_x]^2 dt$

Стандартное (среднеквадратичное) отклонение $\sigma_x = Dx^{1/2}$

Среднеквадратическое значение: $\Psi_x = \mu_x^2 + \sigma_x^2$

3.3 Выборочные оценки параметров и их погрешности

В п. 3.2 были рассмотрены основные параметры случайного процесса. Рассмотрим практические вопросы оценки этих параметров. Для этого примем следующие определения.

Дискретный случайный процесс – это множество значений процесса для дискретных моментов времени nT : $x[1], x[2], \dots, x[n], \dots$

Реализация случайного процесса – это последовательность наблюдений $x[1], \dots$

Выборочные данные – измеренная реализация процесса.

Выборочная оценка параметра:

- оценка, сделанная по выборочным данным (случайная величина);
- оценка, сделанная по реализации процесса, а не по ансамблю;
- оценка, сделанная по конечной выборке.

Свойство эргодичности при выборочной оценке позволяет заменить операцию математического ожидания операцией усреднения по реализации про-

цесса:

$$E\{x\} \rightarrow \frac{1}{N} (\sum_{i=1}^N x[i]) = \langle x[i] \rangle,$$

где $E\{x\}$ – математическое ожидание случайной величины x ,

$\langle x[i] \rangle$ – среднее по времени значение выборки случайной величины x ,

$\langle \rangle$ – символ усреднения по выборке.

Выборочная оценка параметра является случайной величиной, которую можно охарактеризовать математическим ожиданием и дисперсией, что означает ее отличие от неслучайной оценки, вычисленной по бесконечному ансамблю значений.

Замена операции математического ожидания на операцию усреднения по реализации приводит к отличию выборочной оценки параметра от «истинного» значения, которое могло бы быть получено при бесконечном числе наблюдений. Это отличие описывается следующими характеристиками.

Смещением называют разность между истинным значением параметра и математического ожидания его оценки:

$$\Delta_{\hat{x}} = E\{\hat{x}\} - x = E\{\hat{x}\} - E\{x\} = E\{\hat{x} - x\},$$

где $E\{ \}$ – оператор математического ожидания,

\hat{x} – оценка случайной величины x .

Оценка \hat{x} параметра x смещена, если среднее значение \hat{x} не совпадает с x . Оценку называют несмещенной, если смещение $\Delta_{\hat{x}} = 0$, и асимптотически несмещенной, если при усреднении по ансамблю с увеличением числа реализаций смещение стремится к нулю.

Дисперсией оценки называется величина $D_{\hat{x}} = \sigma_{\hat{x}}^2 = E\{[\hat{x} - E\{\hat{x}\}]^2\}$, где $\sigma_{\hat{x}}$ – среднеквадратическое отклонение оценки (СКО).

Среднеквадратической ошибкой оценки называется величина:

$$\Psi_{\hat{x}}^2 = E\{[\hat{x} - x]^2\} = \sigma_{\hat{x}}^2 + \Delta_{\hat{x}}^2.$$

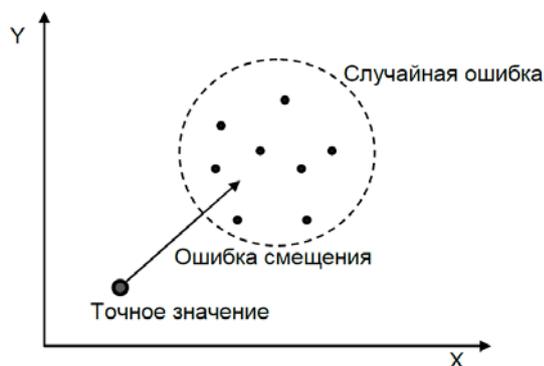


Рисунок 3.2 – Смещение и дисперсия оценки вектора $[x, y]$

На рисунке 3.2 показаны погрешности выборочной оценки вектора $[x, y]$.

Оценка *состоятельна*, если её дисперсия убывает с ростом объема выборки:

$$(\sigma_x^2) = 0, N \rightarrow \infty.$$

3.4 Выборочные оценки основных параметров случайного процесса

Речь является нестационарным процессом, значения ее параметров непрерывно меняются. Поэтому выборочные оценки параметров берутся на последовательных (обычно пересекающихся) интервалах (кадрах данных) в предположении, что на кадре процесс является стационарным и характеризуется фиксированным значением параметра. Такие характеристики называют *кратковременными* (*short-time*).

Обычно интервал оценки параметров РС выбирается равным 25–30 мс. При частоте дискретизации сигнала 16 кГц это соответствует 400–480 отсчетам.

Формулы выборочных оценок характеристик случайного процесса приведены ниже.

$$\text{Оценка среднего значения: } \mu_x = \frac{1}{N} \sum_{i=1}^N x[i] = \langle x[i] \rangle$$

$$\text{Центрирование процесса: } x_0[i] = x[i] - \mu_x,$$

$$\text{Оценка дисперсии: } D_x(N) = \langle x_0^2[i] \rangle = \frac{1}{(N-1)} \sum_{i=1}^N [x[i] - \mu_x]^2$$

$$\text{Смещенная оценка дисперсии: } D_x(N) = \frac{1}{N} \sum_{i=1}^N [x[i] - \mu_x]^2$$

$$\text{Быстрая оценка дисперсии: } D_x(N) = \frac{1}{N} \sum_{i=1}^N x^2[i] - \frac{1}{N} \sum_{i=1}^N x[i]^2$$

$$\text{Оценка СКО: } \sigma_x = \sqrt{D_x}$$

$$\text{Оценка энергии на интервале } N: E_x\{x(N)\} = \sum_{i=1}^N x^2[i]$$

$$\text{Оценка мощности: } P_x = \langle x^2[i] \rangle = \frac{1}{N} \sum_{i=1}^N x^2[i] = \frac{E\{x\}}{N} = \sigma_x^2 + \mu_x^2$$

$$\text{Корень из среднеквадратического значения: } RMS = \sqrt{P_x}$$

$$\text{Средний модуль: } MSF = \langle |x[i]| \rangle = \frac{1}{N} \sum_{i=1}^N |x[i]|$$

Обратим внимание на следующие свойства дисперсии.

Если сигнал является суммой или разностью сигналов *независимых источников*, тогда его дисперсия равна сумме дисперсий этих сигналов:

$$\begin{aligned} D_{x \pm y} &= \langle (x[i] \pm y[i])^2 \rangle = \langle x[i]^2 \pm 2x[i]y[i] + y[i]^2 \rangle = \\ &= \langle x[i]^2 + y[i]^2 \rangle = D_x + D_y, \end{aligned}$$

если $x[i] = y[i]$, то $D_{x+y} = 4D_x$.

3.5 Выборочная оценка функции распределения

Гистограмма – оценка вероятностной функции распределения, вычисленная по одной реализации процесса.

Параметры оценки гистограммы – интервал измерения величины $a < x[i] < b$, число K интервалов группировки равной длины $\Delta = (b-a)/K$, объем выборки N .

Виды оценок гистограммы [39]:

Гистограмма – последовательность числа попаданий в подинтервалы $\{N_j\}$, взятая без всяких изменений:

$$\sum_j N_j = N.$$

Выборочная вероятность попадания амплитуды в соответствующий интервал:

$$P_j = \frac{N_j}{N},$$

$$\sum_j P_j = 1.$$

Выборочная вероятностная функция распределения:

$$p_j = \frac{N_j}{N}; \frac{b-a}{K} = \frac{P_j}{\Delta},$$

$$\int p(x) dx \approx \sum p_j \Delta = 1.$$

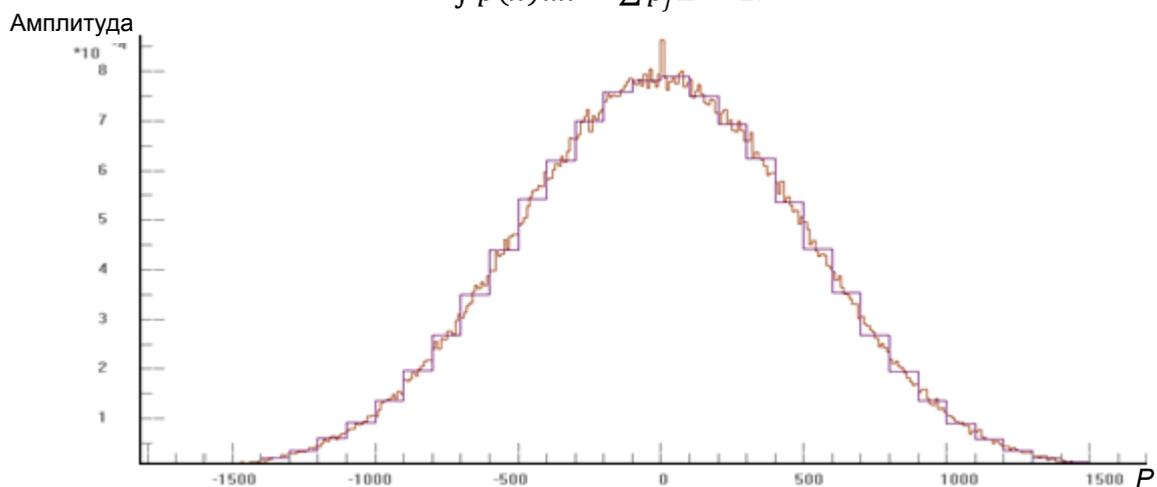


Рисунок 3.3 – Гистограммы (оценки функции распределения) белого шума ($\sigma = 500$, $\mu = 0$), вычисленные с интервалами 10 и 100

На рисунке 3.3 показаны графики выборочной функции распределения, вычисленные по одной реализации с разными интервалами группировки. При изменении интервала группировки форма функции распределения остается прежней.

Функции распределения получили широкое применение в анализе сигналов. Рассмотрим несколько примеров.

Функции совместного распределения нескольких параметров (многомерные функции распределения) широко используются в прикладных задачах об-

работки РС. Многомерные распределения сложной формы обыкновенно аппроксимируют набором гауссовых распределений, называемых *гауссовыми смесями*. На рисунке 3.4 приведен пример выборки двух параметров, аппроксимируемой моделью трех гауссовых функций распределения.

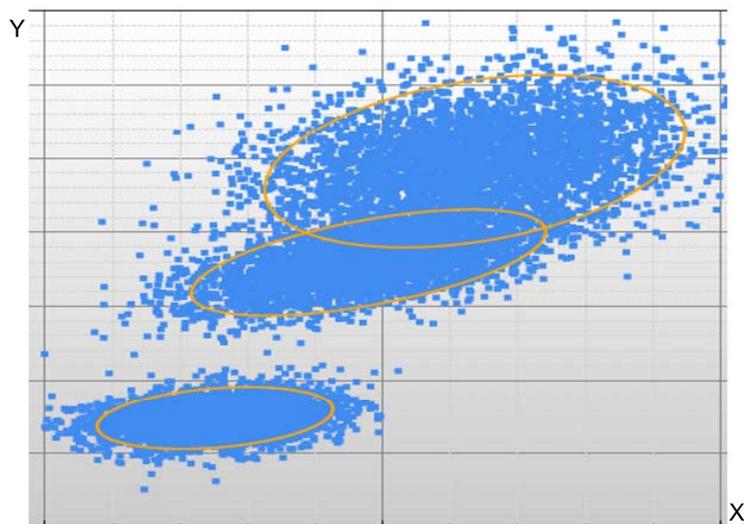


Рисунок 3.4. Выборка двух параметров, аппроксимируемая моделью трех гауссовых функций распределения

Пример необычной гистограммы сигнала показан на рисунке 3.5.

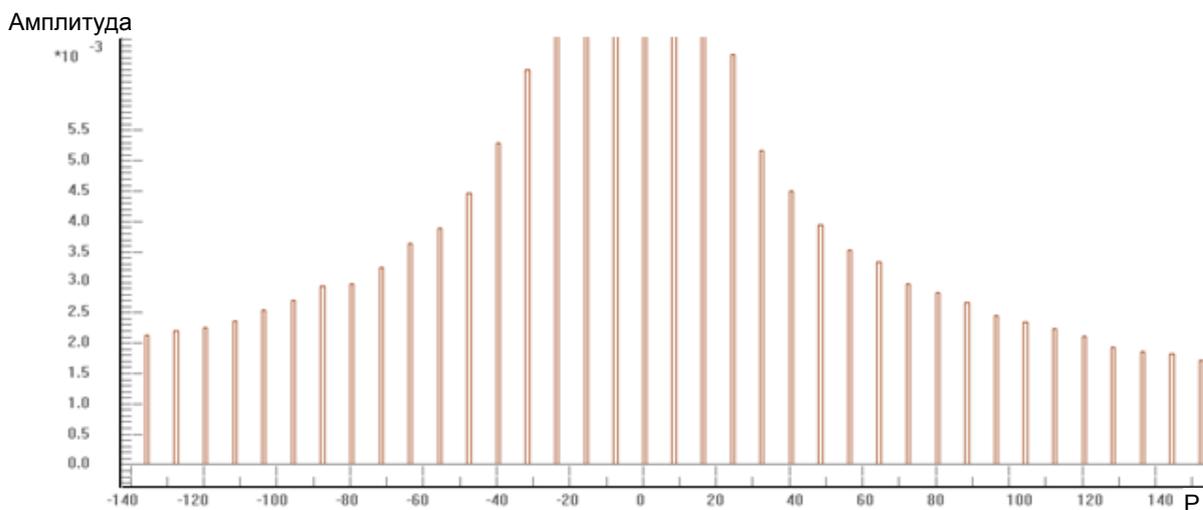


Рисунок 3.5 – Гистограмма сигнала

Анализ этой гистограммы показывает, что сигнал первоначально был записан в формате 12 бит, а затем был преобразован в формат 16 бит.

Гистограммы клипированных сигналов (речевого и шума) показаны на рисунке 3.6.

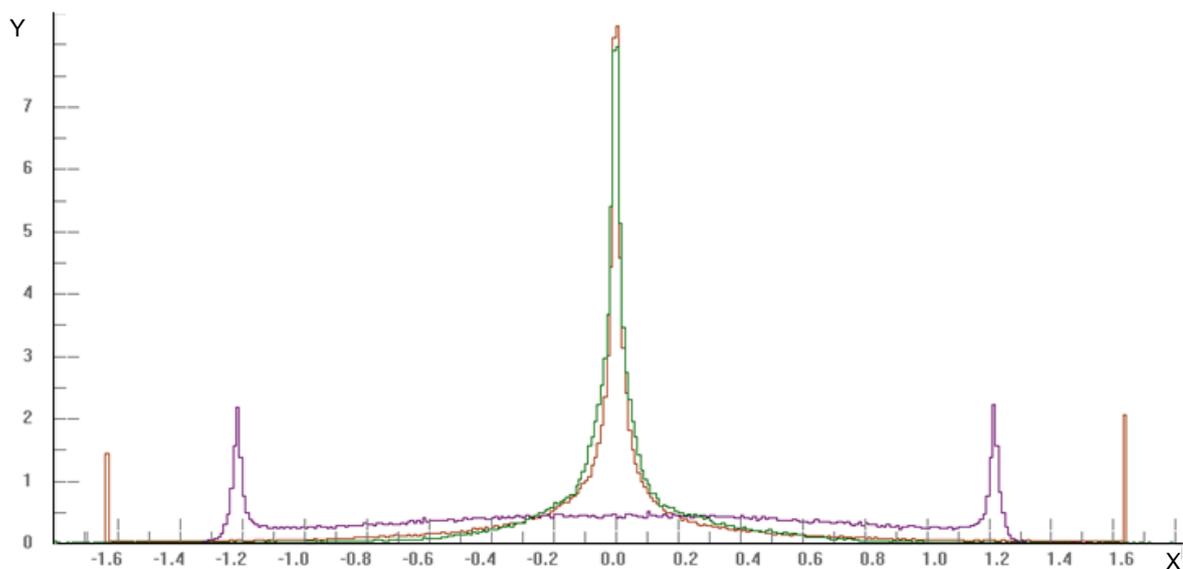


Рисунок 3.6 – Гистограммы чистого речевого сигнала и клипированных речевых сигналов

На гистограммах в области клиппирования сигналов появляется всплеск. Это свойство гистограммы позволяет обнаружить на фонограмме участки клиппирования сигнала.

3.6 Анализ статистической взаимосвязи случайных величин

Важной задачей первичного анализа сигналов является оценка их статистической взаимосвязи. Статистическая взаимосвязь оценивается с помощью нескольких мер. Рассмотрим некоторые.

Ковариация – это математическое ожидание произведения двух центрированных случайных величин x , y [39]:

$$\text{cov}(x, y) = \sigma_{xy} = E\{(x - \mu_x)(y - \mu_y)\}.$$

Для независимых величин $\text{cov}(x, y) = 0$.

Другой мерой подобия сигналов является дисперсия их разности. Соотношение между дисперсией и ковариацией определяется следующим соотношением:

$$\begin{aligned} D_{xy} &= E\{(x[i] - y[i])^2\} = E\{x[i]^2\} + E\{y[i]^2\} - E\{2x[i]y[i]\} = \\ &= D_x + D_y - 2\text{cov}(x, y). \end{aligned}$$

Из этого соотношения следует:

$$\min D_{xy} \Leftrightarrow \max \text{cov}(x, y).$$

Выборочная оценка ковариации (выборочная ковариация) двух дискретных сигналов вычисляется следующим образом:

$$C_{xy} = \frac{1}{N-1} \sum_{i=0}^{N-1} (x[i] - \mu_x)(y[i] - \mu_y).$$

Однако, если $x[i] = \gamma y[i]$, то ковариация будет зависеть от коэффициента γ . Более подходящей мерой является *коэффициент корреляции*. *Корреляция* – это нормированная ковариация [39]:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{E\{(x - \mu_x)(y - \mu_y)\}}{\sigma_x \sigma_y}.$$

Если между величинами x , y имеется линейное соотношение (простейшая линейная связь) $y[i] = \alpha x[i] + \beta$, то $|\rho_{xy}| = 1$.

Если величины x , y зашумлены, то $|\rho_{xy}| < 1$.

Если связь отсутствует, то $|\rho_{xy}| = 0$.

Выборочная оценка коэффициента корреляции центрированных процессов вычисляется по следующей формуле:

$$\rho_{xy} = \frac{\sum_i (x[i]y[i])}{\sqrt{\sum_i x[i]^2 \sum_i y[i]^2}}.$$

Быстрая выборочная оценка коэффициента корреляции процессов вычисляется по следующей формуле:

$$\rho_{xy} = (\sum_{xy} - 1/M \sum_x \sum_y) / (\sum_{xx} - 1/M \sum_x \sum_x) (\sum_{yy} - 1/M \sum_y \sum_y)^{1/2},$$

где

$$\sum_{xy} = \sum_i x[i] y[i], \sum_x = \sum_i x[i], \sum_y = \sum_i y[i].$$

Каким образом охарактеризовать статистическую взаимосвязь (степень «сходства») двух случайных процессов, например сигналов микрофонов? Чтобы учесть возможный временной сдвиг между сигналами, необходимо вычислить коэффициенты ковариации при различных значениях временного сдвига между сигналами по времени. Для этого используются ковариационные функции, рассмотренные в главе 4.

Вопросы и упражнения

1. $y[i] = c \times x[i] + a + n[i]$, где $x[i]$, $n[i]$ – независимые случайные процессы с нулевым средним. Как по измерениям $y[i]$, $x[i]$ вычислить оценки коэффициентов a , c ?
2. Вы измерили N реализаций (значений) некоторого параметра $x[i]$, $i=1, N$ и вычислили оценку среднего значения $A_x = \frac{1}{N} \sum x[i]$. Как оценить погрешность оценки среднего?
3. Докажите, что $\sigma x^2 = E\{x^2\} - E^2\{x\}$, где $E\{ \}$ – символ математического ожидания.
4. Чему равна дисперсия суммы двух независимых сигналов?
5. Изменится ли выборочная оценка мощности сигнала при увеличении частоты его дискретизации?
6. Необходимо контролировать текущую ошибку оценки, чтобы остановиться по достижению заданной точности. Как это сделать?
7. Докажите быструю формулу вычисления выборочной оценки ковариации двух дискретных процессов:

$$\begin{aligned} R_{xy} &= \frac{1}{N} \sum_0^{N-1} (x[i] - \mu_x) \times (y[i] - \mu_y) = \\ &= \frac{1}{N} \sum_0^{N-1} x[i]y[i] - \left(\frac{1}{N} \sum_0^{N-1} x[i] \right) \times \left(\frac{1}{N} \sum_0^{N-1} y[i] \right). \end{aligned}$$

8. Будут ли различаться выборочные оценки мощности двух синусоид одинаковой амплитуды, но разной частоты?

4 ВРЕМЕННОЙ И ЧАСТОТНЫЙ АНАЛИЗ СИГНАЛОВ

Для описания временных и частотных свойств речевых сигналов используются функции: корреляционная функция, спектр, коэффициенты линейного предсказания и ряд других. В данном разделе рассмотрены ковариационные функции, спектры, функция когерентности и соотношения между ними.

4.1 Автоковариационная функция процесса

Статистическая связь двух сдвинутых по времени значений стационарной последовательности $x[i]$ и $x[i+m]$ характеризуется ковариацией $R_m = R_{-m}$ и соответствует корреляции с задержкой m . Автоковариационная функция процесса (АКФ) описывает статистическую связь между временными отсчетами сигнала, разнесенными на величину задержки. Совокупность ковариаций с разным сдвигом называют автокорреляционной функцией процесса (*temporal autocorrelation*).

Несмещенная выборочная оценка автоковариационной функции процесса определяется формулой:

$$C_{xx}[m] = \frac{1}{(N-m)} \sum_{i=0}^{N-m-1} (x[i] - \mu_x) \times (x[i+m] - \mu_x),$$
$$m = -M+1, \dots, -1, 0, 1, \dots, M-1.$$

Свойства АКФ:

$$C_{xx}[-m] = C_{xx}[m],$$
$$C_{xx}[0] = D_x,$$
$$C_{xx}[m] \rightarrow 0 \text{ при } m \rightarrow \infty.$$

Если сигнал является суммой сигналов разных независимых источников, тогда его КФ будет равна сумме АКФ этих сигналов: $R_{ss}[m] = \sum R_{xxi}[m]$.

Чтобы сделать меру инвариантной к мощности (нормировке) сигналов, используют нормализованную АКФ (*energy normalized correlation*). Нормализованная АКФ определяется формулой:

$$R_{XX}[m] = \frac{C_{xx}[m]}{C_{xx}[0]},$$

или

$$R_{XX}[m] = \frac{\sum_{i=0}^{L-m-1} x[i]x[i+m]}{(\sum_{i=0}^{L-m-1} x^2[i] \sum_{i=0}^{L-m-1} x^2[i+m])^{\frac{1}{2}}}.$$

Параметрами оценки АКФ являются интервал суммирования L и интервал задержек M (максимальное запаздывание).

Основным применением автокорреляционной функции в анализе и обра-

ботке РС является оценка коэффициентов линейного предсказания при моделировании речевых сигналов [31]. Другим применением АКФ является обнаружение скрытых периодичностей процессов, например *основного тона* РС.

Пример

Обнаружение скрытых периодичностей процессов и оценка основного тона с помощью АКФ. Пусть

$$x[i] \approx x[i + k] + n[i],$$

тогда

$$\begin{aligned} C_{xx}[m] &= \frac{1}{N-m} \sum_{i=0}^{N-m-1} (x[i]x[m+k+i] + x[i]n[m+i]) = \\ &= C_{xx}[m+k] + C_{xn}[m]. \end{aligned}$$

Поскольку сигнал и шум не коррелированы, имеем:

$$C_{xx}[m] \approx C_{xx}[m+k].$$

Таким образом, АКФ является периодической функцией с периодом основного тона сигнала. На рисунке 4.1 представлен график АКФ на кадре речевого сигнала с периодом основного тона 37 мс.

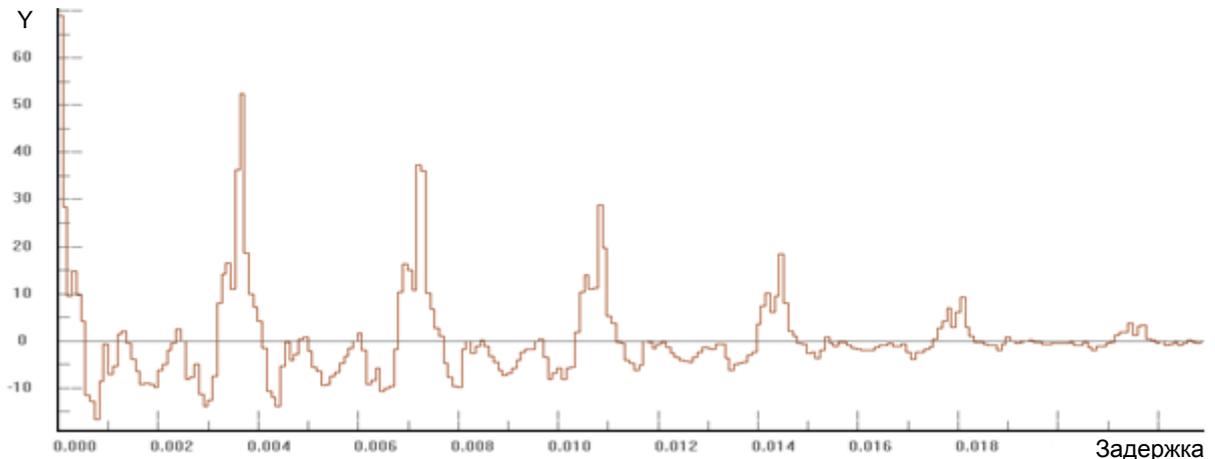


Рисунок 4.1 – АКФ на кадре тонального речевого сигнала

4.2 Ковариация и ее оценка

Кросс-ковариационный анализ необходим для идентификации систем, сравнения сигналов одного источника, принимаемых разными микрофонами и во многих других приложениях анализа и обработки РС. С его помощью можно определять степень сходства двух временных последовательностей. Иногда используется термин кросс-ковариация, чтобы не путать с автоковариацией.

Пусть имеем два дискретных сигнала $x[i]$, $y[i]$, $i = 0, 1, \dots, N-1$. Функция кросс-ковариации (ФКК) определяется выражением:

$$C_{xy}[m] = \frac{1}{L-m} \sum_{i=0}^{L-m-1} (x[i] - \mu_x) \times (y[i+m] - \mu_y),$$

$$m = -M, \dots, -1, 0, 1, \dots, M,$$

где $\mu_x = \langle x[i] \rangle$, $\mu_y = \langle y[i] \rangle$, m – «запаздывание». M – «максимальное запаздывание», $i = 0, 1, \dots, L$.

Свойства функции кросс-ковариации представлены ниже.

$$C_{xy}[m] \rightarrow 0 \text{ при } m \rightarrow \infty,$$

$$C_{xy}[-m] = C_{yx}[m],$$

$$C_{xx}[-m] = C_{xx}[m],$$

$$C_{xy}[m] + C_{xy}[-m] \text{ – четная функция,}$$

$$C_{xy}[m] - C_{xy}[-m] \text{ – нечетная функция.}$$

Во многих случаях удобно пользоваться краткой векторной записью ковариационной и автокорреляционной функций.

Определим вектор отсчетов $\mathbf{X}[i] = [x[i], x[i-1], \dots, x[i-M+1]]^T$, и вектор ковариаций $\mathbf{C}_{xy} = [C_{xy}[0], C_{xy}[1], \dots, C_{xy}[M-1]]^T$. Оценку вектора ковариаций можно записать следующим образом:

$$\mathbf{C}_{xy} = \langle \mathbf{X}[i] \mathbf{Y}[i] \rangle.$$

В случае $y[i] = x[i]$ получим оценку вектора автокорреляций:

$$\mathbf{C}_{xx} = \langle \mathbf{X}[i] \mathbf{X}[i] \rangle.$$

Обобщением ФКК для случая многомерных сигналов (микрофонных решеток) является *кросс-корреляционная матрица* сигналов:

$$A_{XY}[i, j, m] = \frac{1}{L-m} \sum_{n=0}^{L-m-1} (x_i[n] - \mu_{x_i}) \times (x_j[n+m] - \mu_{x_j}),$$

$$m = -M, \dots, -1, 0, 1, \dots, M.$$

В векторной форме:

$$\mathbf{A}_{xy} = \langle \mathbf{Y}[i] \mathbf{X}^T[i] \rangle.$$

Инвариантная к нормировке сигналов нормализованная *кросс-корреляционная функция* определяется формулой:

$$R_{xy}[m] = C_{xy}[m] / \sigma_x \sigma_y \leq 1.$$

Пример

Измерение временных задержек (сдвигов) между сигналами.

Сигнал $s(t)$ распространяется из точки A в точку B с задержкой T_{xy} :

$$A: x(t) = s(t) + n(t),$$

$$B: y(t) = \alpha s(t - T_{xy}) + v(t).$$

Можно показать, что функция кросс-корреляции $C_{xy}(\tau)$ между сигналами $x(t)$, $y(t)$ будет иметь пик, соответствующий задержке сигнала между этими точками:

$$C_{xy}(\tau) = \alpha C_{xx}(\tau - T_{xy}),$$

$$T_{xy} = \operatorname{argmax}\{C_{xy}(\tau)\}.$$

По положению максимума ФКК можно измерить временной сдвиг между сигналами.

4.3 Спектральный анализ сигналов

Спектральный анализ – это один из методов, который позволяет охарактеризовать частотный состав измеряемого сигнала и описать сигнал в спектральном пространстве [32].

Преобразование Фурье (Fourier transform) – это разложение функций на синусоиды. Существует несколько видов преобразования Фурье [29, 32]:

- а) непериодический непрерывный сигнал можно разложить в интеграл Фурье: $x(t) \rightarrow X(f)$;
- б) периодический непрерывный сигнал можно разложить в бесконечный ряд Фурье $x(t) \rightarrow X[n]$;
- в) непериодический дискретный сигнал можно разложить в интеграл Фурье: $x[i] \rightarrow X(f)$;
- г) периодический дискретный сигнал можно разложить в конечный ряд Фурье: $x[i] \rightarrow X[n]$.

Речевой сигнал является нестационарным процессом, значения параметров которого непрерывно меняются. Поэтому выборочные оценки спектра берутся на последовательных (обычно пересекающихся) кадрах в предположении, что на кадре процесс является стационарным и характеризуется фиксированным спектром. Дискретный вектор сигнала на кадре преобразуется с помощью преобразования Фурье в дискретный спектр той же размерности (конечный ряд Фурье). Таким образом, при обработке РС используется только последний вид преобразования Фурье (дискретное преобразование Фурье, ДПФ). В этой модели сигнала неявно подразумевается его периодичность, то есть периодическое повторение за пределами интервала наблюдения.

Рассмотрим дискретное преобразование Фурье подробнее.

Спектр – представление функции $x[i]$ на интервале $[0, N-1]$ в виде суммы периодических компонент:

$$x[i] = \sum_{n=0}^{N-1} Ax[n] \cos\left(\frac{2\pi in}{N} + \varphi_x[n]\right) =$$

$$= \sum_{n=0}^{N-1} (Ac[n] \cos(2\pi in/N) + As[n] \sin(2\pi in/N)),$$

где N – размерность спектра.

Для компактности записи можно представить пары коэффициентов спектра в виде комплексных чисел (коэффициентов комплексного спектра):

$$X[n] = Ac[n] + jAs[n],$$

тогда:

$$x[i] = \sum X[n] \exp[j2\pi in/N].$$

В этом случае вектор коэффициентов дискретного комплексного спектра вычисляются из вектора отсчетов сигнала на кадре с помощью *дискретного преобразования Фурье (Discrete Fourier Transform, DFT)*:

$$\begin{aligned} X[n] &= DFT\{x[i]\} = \sum_{i=0}^{N-1} x[i] \exp\left[-\frac{j2\pi ni}{N}\right] = \\ &= \sum_{i=0}^{N-1} x[i] \left[\cos\left(\frac{2\pi ni}{N}\right) - j \sin\left(\frac{2\pi ni}{N}\right) \right], n = 0, \dots, N-1. \end{aligned}$$

Коэффициенты дискретного спектра соответствуют частотам $n \times Fs/N$.

Свойства комплексного спектра вещественной последовательности $x[i]$:

$$\operatorname{Re}\{X[n]\} = \operatorname{Re}\{X[N-n]\}, n=1, \dots, N/2,$$

$$\operatorname{Im}\{X[n]\} = -\operatorname{Im}\{X[N-n]\}, n=1, \dots, N/2,$$

$$\operatorname{Im}\{X[n]\} = 0, n=1, N/2.$$

Реальные коэффициенты спектра *симметричны*, мнимые – *антисимметричны* относительно номера $n = N/2$. Таким образом, вектор N отсчетов сигнала преобразуется в вектор N независимых отсчетов комплексного спектра: $N/2+1$ значений реальной части и $N/2-1$ значений мнимой части. В разных программах отсчеты комплексного спектра могут размещаться («упаковываться») в векторе спектра различными способами.

На рисунке 4.2 приведен пример дискретного спектра для $N = 8$.



Рисунок 4.2 – Реальная и мнимая части спектра вещественного сигнала ($N=8$)

Комплексные коэффициенты спектра могут быть представлены в полярных координатах модуля и фазы:

$$X[n] = |X[n]| \exp(j2\pi \Phi x[n]) = Ax[n] \exp(j2\pi \Phi x[n])$$

$$Ax[n] = (\operatorname{Re}\{X[n]\} + j\operatorname{Im}\{X[n]\})^{1/2},$$

$$\Phi x[n] = \operatorname{arctg}\left(\frac{\operatorname{Im}\{X[n]\}}{\operatorname{Re}\{X[n]\}}\right),$$

где обозначение arctg введено для главного значения.

Амплитудные коэффициенты $Ax[n]$ называют *амплитудным спектром*, фазовые $\Phi x[n]$ – *фазовым спектром*.

Вектор коэффициентов спектра может быть преобразован в вектор отсчетов сигнала на кадре с помощью обратного дискретного преобразования Фурье (ОДПФ). Оператор ОДПФ будем обозначать IDFT (*Inverse Discrete Fourier Transform*):

$$x[i] = \operatorname{IDFT}\{X[n]\} = \frac{1}{N} \sum_{n=0}^{N-1} X[n] \exp\left(j2\pi \frac{ni}{N}\right) =$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} [\operatorname{Re}\{X[n]\} + j\operatorname{Im}\{X[n]\}] \left[\cos\left(\frac{2\pi ni}{N}\right) + j \sin\left(\frac{2\pi ni}{N}\right) \right], i=0, N-1.$$

Обратное дискретное преобразование Фурье выполняется с помощью той же самой процедуры, что и ДПФ. Необходимо только вместо действительных отсчетов сигнала подставить комплексные значения спектра и поменять знак при мнимой части, а результат поделить на N .

Поскольку реальная часть спектра является симметричной функцией, а мнимая – антисимметричной, то последовательность $x[i]$ является вещественной.

Замечание.

В принципе, имея N отсчетов сигнала, можно вычислить спектр для любой частоты (в интервале $0-Fs/2$):

$$X(f) = \sum_{n=0}^{N-1} x[n] \exp(-j2\pi fTn).$$

Поэтому в дальнейшем мы также будем использовать такую запись:

$$X(f) = \operatorname{FT}\{x[j]\}.$$

Переход от непрерывного спектра к дискретному осуществляется выбором частот $fn = n/NT$, $n = 0, \dots, N-1$.

4.4 Спектр мощности

Спектральная плотность мощности (СПМ) показывает распределение дисперсий случайного процесса в непрерывном диапазоне частот $0 - \pi$. Величину $P_{xx}(\omega)d\omega$ можно интерпретировать как приближенное значение дисперсии

процесса в диапазоне частот $\omega - (\omega + d\omega)$.

Рассмотрим дискретную оценку СПМ. *Периодограмма* – оценка дискретного спектра мощности сигнала на кадре определяется соотношением:

$$P_{xx}[n] = |X[n]|^2 = \text{Re}^2\{X[n]\} + \text{Im}^2\{X[n]\}.$$

Периодограмма имеет смысл распределения мощности сигнала по частотным полосам. Например, в белом шуме средний спектр мощности равномерный.

Оценка спектра мощности на кадре (на конечной выборке) имеет большую дисперсию, которая не уменьшается с увеличением длины выборки, поскольку при этом увеличивается размерность спектра. Таким образом, оценка СПМ является несостоятельной. Для решения этой проблемы разработан ряд методов оценки дискретного спектра мощности, основанных на усреднении (сглаживании) периодограмм [39]:

- а) *метод Бартлетта*: усреднение периодограмм по времени на неперекрывающихся кадрах $P_{xx}[n] = \langle |X[n, k]|^2 \rangle = \frac{1}{K} \sum_k |X[n, k]|^2$, где k – индекс кадров;
- б) *метод Уэлча*: сглаживание периодограмм на перекрывающихся кадрах;
- в) *метод Даньелла*: сглаживания периодограмм по частотам;
- г) совместное сглаживание периодограмм по частоте и времени;
- д) *метод Блэкмана-Тьюки*: основан на соотношениях Винера-Хинчина, связывающих дискретные оценки спектра мощности и корреляционной функции $P_{xx}[n] = \text{DFT}\{C_{xx}[m]\}$, $C_{xx}[m] = \text{IDFT}\{P_{xx}[n]\}$.

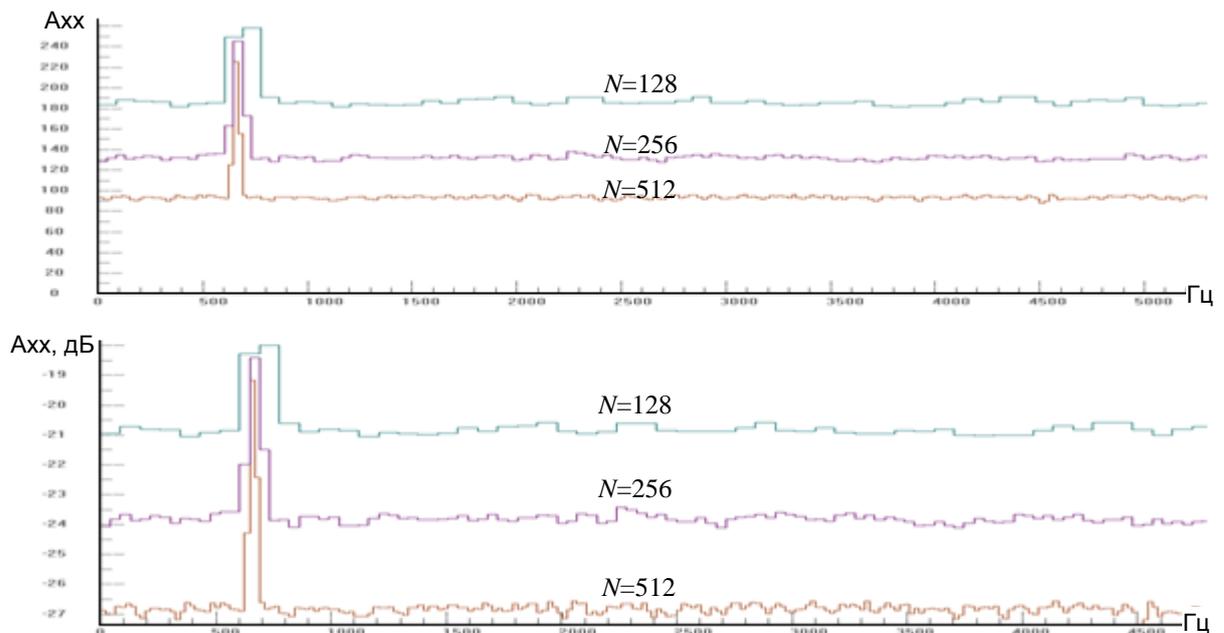


Рисунок 4.3 – Средние амплитудные спектры смеси белого шума и гармоники: наверху – в линейном масштабе, внизу – в логарифмическом масштабе

Обычно на графиках показывают средний спектр мощности, вычисленный по методу Бартлетта или соответствующий ему амплитудный спектр $Ax[n] = \sqrt{P_{xx}[n]}$. В шкале дБ эти представления не различаются.

Отметим различие дискретных оценок спектра случайного и детерминированного тонального сигнала. На рисунке 4.3 показаны средние спектры разной размерности белого шума и гармоники.

Из графиков видно, что с изменением числа полос (размерности спектра) амплитуда спектра шума меняется и почти не меняется для гармоники.

4.5 Кросс-спектр

В анализе дискретных сигналов кросс-спектр используется как элемент частотного анализа кросс-корреляции двух дискретных сигналов. *Плотность кросс-спектра* (КСП) – это разложение ковариации по частотам:

$$X[n] = \sum_{i=0}^{N-1} x[i] \exp(-j2\pi in/N) = DFT\{X\},$$

$$Y[n] = \sum_{i=0}^{N-1} y[i] \exp(-j2\pi in/N) = DFT\{Y\}.$$

Кросс-спектр является комплексной функцией и определяется следующим выражением:

$$P_{xy}[n] = \langle X^*[n] Y[n] \rangle = R_{xy}[n] - jQ_{xy}[n].$$

Модуль и фаза кросс-спектра определяются следующими выражениями:

$$|P_{xy}[n]|^2 = R_{xy}^2[n] + Q_{xy}^2[n],$$

$$|P_{xy}[n]| = \left[|P_{xy}[n]|^2 \right]^{1/2},$$

$$\Phi_{xy}[n] = \arctg \left[\frac{Q_{xy}[n]}{R_{xy}[n]} \right],$$

$$P_{xy}[n] = |P_{xy}[n]| \exp(j\Phi_{xy}[n]).$$

КСП характеризует степень и характер связи между частотными составляющими двух временных рядов. Модуль КСП характеризует соотношение амплитуд спектральных составляющих рядов. Характер опережения или запаздывания между рядами характеризуется фазовым спектром (фазовым сдвигом).

Свойства КСП [21]:

$$|P_{xy}[n]|^2 \leq P_{xx}[n] P_{yy}[n],$$

$$|P_{yx}[n]|^2 = |P_{xy}[n]|^2,$$

$$P_{yx}[n] = P_{xy}^*[n].$$

Обобщением кросс-спектров для случая многомерных сигналов (например, микрофонных решеток) является матрица кросс-спектров [39]:

$$P_{xx}(f) = \langle X^*(f) X^T(f) \rangle.$$

4.6 Соотношение между ковариационной функцией и кросс-спектром

Ковариация и кросс-спектр связаны преобразованием Фурье:

$$P_{xy}[n] = \sum_{m=0}^{N-1} C_{xy}[m] \exp\left(-j2\pi \frac{mn}{N}\right) = DFT\{C_{xy}[m]\}, n = 0, 1, \dots, N-1,$$

$$C_{xy}[m] = \frac{1}{N} \sum_{n=0}^{N-1} P_{xy}[n] \exp\left(j2\pi \frac{mn}{N}\right) = IDFT\{P_{xy}[n]\}, m = 0, 1, \dots, N-1$$

Для отдельного кадра:

$$C_{xy}[m] = \frac{1}{N} \left[\sum_{n=0}^{N-1} [X^*[n]Y[n]] \exp\left(j2\pi \frac{mn}{N}\right) \right] =$$

$$= IDFT\{X^*[n]Y[n]\} = IDFT\{DFT\{x[i]\} * DFT\{y[i]\}\}.$$

Для вещественных сигналов $x[i]$, $y[i]$ функция $C_{xy}[m]$ вещественная, однако она не является симметричной относительно $m=0$.

4.7 Функция когерентности и ее оценка

Когерентность (Coherence) – от латинского *Coherens* («находящийся в связи» или «взаимосвязь»). Когерентность объединяет спектральные и корреляционные характеристики процессов и характеризует корреляцию между процессами отдельно в каждой спектральной полосе. Функция когерентности является мерой (нормированной 0..1) линейной связи двух сигналов на заданной частоте f .

Комплексная функция когерентности

Комплексная функция когерентности определяется выражением:

$$\Gamma_{xy}(f) = (P_{xy}(f) / [P_{xx}(f) P_{yy}(f)]^{1/2}) =$$

$$= \langle X(f, k) * Y(f, k) \rangle / [\langle |X(f, k)|^2 \rangle \langle |Y(f, k)|^2 \rangle]^{1/2} = |\Gamma_{xy}(f)| \exp(\varphi_{xy}(f)),$$

Отсюда следует $\Gamma_{xy}(f) = \Gamma_{yx}^*(f)$.

Функция когерентности может принимать как положительные, так и отрицательные значения.

Вещественная функция когерентности определяется выражением:

$$|\Gamma_{xy}(f)| = |P_{xy}(f) / [P_{xx}(f) P_{yy}(f)]^{1/2}|.$$

Производным от функции когерентности является *квадрат модуля когерентности (КМК)*, определяемый выражением:

$$G_{xy}(f) = |\Gamma_{xy}(f)|^2 = |P_{xy}(f)|^2 / [P_{xx}(f) P_{yy}(f)].$$

На каждой частоте КМК соответствует коэффициенту корреляции:

$$\rho_{xy}^2 = \text{cov}_{xy}^2 / \sigma_x^2 \sigma_y^2 \leftrightarrow |\Gamma_{xy}(f)|^2 = |P_{xy}(f)|^2 / [P_{xx}(f) P_{yy}(f)],$$

где $\text{cov}_{xy} = \langle xy \rangle$, $\sigma_x^2 = \langle x^2 \rangle$, $\sigma_y^2 = \langle y^2 \rangle$.

Рассмотрим роль усреднения в оценке когерентности. Физический смысл усреднения значения заключается в суммировании мгновенных единичных векторов на комплексной плоскости, в результате чего получается комплексный вектор произвольной длины. Если мгновенные компоненты были синфазны, то отдельные вектора имеют одинаковое направление, и результирующий вектор будет большим. Если мгновенные вектора имели разные направления, компенсирующие друг друга, то результирующий вектор будет иметь малую длину. Таким образом, коэффициент когерентности характеризует среднюю взаимосвязь фаз двух процессов.

Обыкновенно статистическая связь существует между детерминированными процессами или процессами, связанными линейным преобразованием.

В случае, когда исследуется связность сигнала во времени, используются сдвинутые на интервал ΔT кадры сигнала. Тогда функцию когерентности можно интерпретировать как функцию автокорреляции на частоте f со сдвигом ΔT . Таким образом, функция когерентности является обобщением АКФ для сигнала на частоте f .

В общем случае функция когерентности зависит как от временной, так и от пространственной координаты. Различают *пространственную и временную когерентность*, которые характеризуют взаимосвязь процессов, разнесенных в пространстве или во времени. Если рассматривать эти процессы как сигналы на входе и выходе системы, то изменение процесса во времени или в пространстве можно рассматривать как результат прохождения через систему.

Временная функция когерентности позволяет судить о корреляции спектральных компонент сигнала во времени, пространственная – о корреляции спектральных компонент сигналов в пространстве.

Пространственная функция когерентности используется при анализе акустических характеристик помещений и в методах обработки сигналов микрофонных решеток.

Оценка функции когерентности

Функцию временной когерентности целесообразно вычислять на непересекающихся кадрах сигнала. При пересечении кадров сигналы на кадрах становятся коррелированными, и значения оценки функции когерентности начинают расти даже для некоррелированного процесса, например белого шума.

Особенность оценки функции когерентности заключается в том, что операция усреднения выполняется раньше вычисления модуля. Если усреднение не производится, то независимо от природы наблюдаемых процессов функция когерентности равна единице:

$$|\Gamma_{xy}(f, k)| = |X^*(f, k) Y(f, k)| / [|X(f, k)|^2 |Y(f, k)|^2]^{1/2} = \\ = \{|X^*(f, k)/|X(f, k)| \times Y(f, k)/|Y(f, k)|\} = |\exp[j(\Phi_Y(f, k) - \Phi_X(f, k))]| = 1.$$

В определении функции когерентности используются средние спектры. При этом усредняется не отношение мгновенных спектров, а отдельно спектры в числителе и знаменателе.

Операция усреднения означает сложение комплексных векторов (единичной длины). Если фазовый угол между векторами постоянен, то результирующий комплексный вектор будет иметь длину единица, если фазовый угол меняется случайным образом, то длина результирующего вектора стремится к нулю. Таким образом, в результате усреднения шум, который можно считать случайным вектором, подавляется, так как фазы двух сигналов не связаны между собой. Следовательно, функция когерентности измеряет степень общности между сигналами в полосе частот.

Более детально вопросы оценки функции когерентности дискретных последовательностей конечной длины и их случайная и систематическая погрешности рассмотрены, например, в книге [39].

Функцию когерентности можно оценивать на реальных данных либо вычислить для модели сигналов. Рассмотрим примеры оценки временной и пространственной когерентности.

Пример 1

На рисунке 4.4 приведена экспериментальная оценка среднего спектра периодической импульсной помехи, на рисунке 4.5 – оценка модуля временной функции когерентности ($\times 10^4$) периодической импульсной помехи между сигналами соседних кадров.

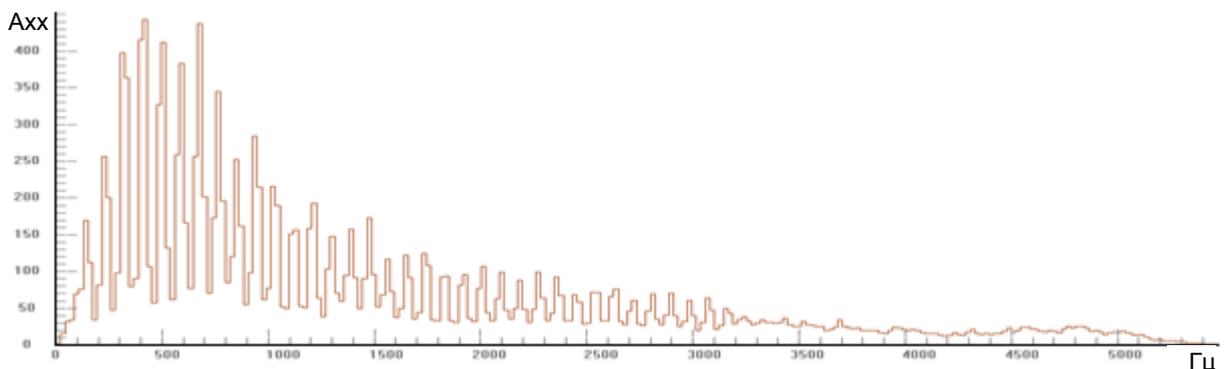


Рисунок 4.4 – Средний спектр периодической импульсной помехи

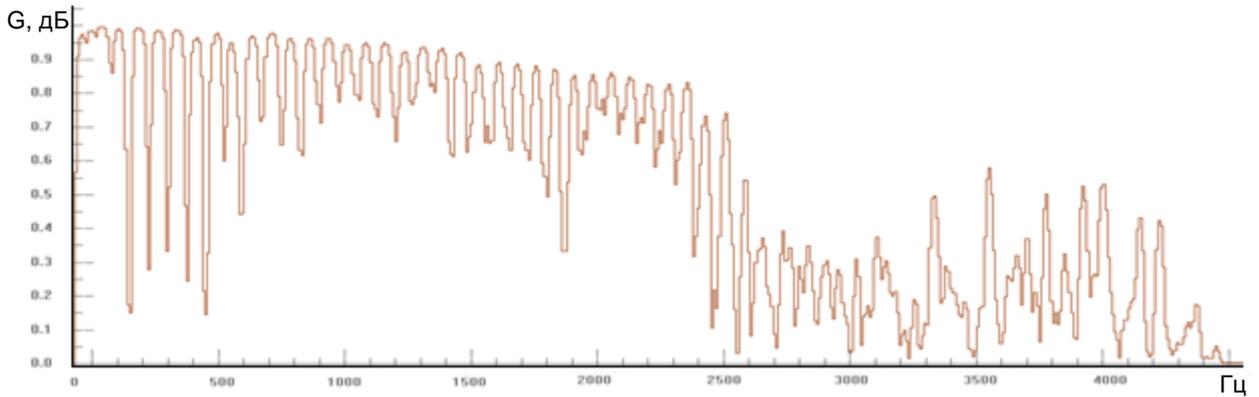


Рисунок 4.5 – Оценка модуля временной функции когерентности периодической импульсной помехи на соседних кадрах

Пример 2

Теоретическая оценка пространственной функции когерентности. На микрофоны 1, 2, расположенные на расстоянии d один от другого под углом θ относительно оси, проходящей через микрофоны, поступает широкополосный сигнал. На микрофон 2 сигнал поступает с задержкой $\tau = d \times \cos\theta / c$ относительно сигнала микрофона 1:

$$x_2(t) = x_1(t + \tau).$$

Комплексный спектр сигнала 2-го микрофона отличается от спектра 1-го микрофона комплексным множителем: $X_2(f, k) = X_1(f, k) e^{-j2\pi f \tau}$.

Тогда

$$\begin{aligned} \Gamma_{12}(f, d) &= \langle X_1(f, k) * X_2(f, k) \rangle / [\langle |X_1(f, k)|^2 \rangle \langle |X_2(f, k)|^2 \rangle]^{1/2} = \\ &= \exp(-j2\pi d \cos\theta f / c), \end{aligned}$$

где $c = 340$ м/с – скорость звука в воздухе.

$$G_{12}(f) = |\Gamma_{12}(f)|^2 = 1.$$

Вопросы и упражнения

1. Имеем дискретный сигнал (белый шум). Оцениваем его средний спектр с разным размером окна (ДПФ 256, 512, 1024). Изменится ли амплитуда выборочной оценки амплитудного спектра?
2. Может ли функция когерентности принимать отрицательные значения?

5 СИСТЕМЫ И ПРЕОБРАЗОВАНИЯ СИГНАЛОВ

В данной главе описаны некоторые типы систем и рассматриваются две основные задачи: первая – представление (моделирование) аналоговых систем в цифровой форме, вторая – оценка параметров дискретных систем.

Общие вопросы моделирования сигналов и систем изложены в монографии [42].

5.1 Общие сведения о системах

При рассмотрении сигналов наряду с понятием «сигнала» используется понятие (или концепция) «системы». Системы описывают преобразования сигналов (аналоговых и цифровых). Понятие системы удобно, когда мы хотим связать два (или более) сигнала, возникающих при их преобразовании или реализовать такое преобразование. Часто удобно рассматривать сигнал как выход системы, которая может быть «реальной» или просто удобной моделью. Основное назначение систем – моделирование и выполнение преобразований.

Акустические сигналы претерпевают различные преобразования: распространение сигнала в пространстве, преобразование микрофоном акустического сигнала в электрический, преобразования электрического сигнала. После преобразования электрических сигналов в цифровую форму они обрабатываются уже в дискретных системах.

Система – это некоторое преобразование сигнала. Система преобразует входной сигнал $x(t)$ в выходной сигнал $y(t)$. Например, зависимость давления воздуха в точке от времени можно рассматривать как звуковой сигнал. Зависимость напряжения в проводнике от времени тоже может представлять звуковой сигнал. Микрофон преобразует акустическую энергию звукового сигнала в напряжение.

Модель системы – математический способ описания преобразования сигналов и их характеристик. Модель системы описывает связи между сигналами на входе и выходе (в дальнейшем – входом и выходом) и может быть описана в символьной форме следующим соотношением:

$$Y = F\{X\},$$

где X – вход системы, Y – выход, $F\{\}$ – преобразование, выполняемое системой.

Примером системы является аппарат речеобразования, преобразующий звук колебаний голосовых связок в звуки речи. Различные модели речеобразования рассмотрены, например, в книгах [31, 46, 47].

Перечислим основные задачи, связанные с применением систем.

Задачи анализа. Зная характеристики системы, необходимо определить выход по входу или вход по выходу:

$$X, F \rightarrow Y,$$

$$Y, F \rightarrow X.$$

Задачи идентификации системы. Имея вход и выход, необходимо определить характеристики системы:

$$X, Y \rightarrow F.$$

Задачи «слепой» идентификации системы (blind system identification). Зная выход, необходимо определить вход и характеристики системы:

$$Y \rightarrow F, X.$$

Задачи синтеза. Необходимо определить характеристики системы, реализующей заданное преобразование между входом и выходом:

$$X, Y \rightarrow F.$$

Модели (наряду с аналитическими выражениями) являются формой представления знаний о системах. Для одного и того же физического объекта можно использовать разные модели. Всякая модель имеет характер проекции [42]. Модель позволяет «представить» или «заменить» систему (рисунок 5.1).

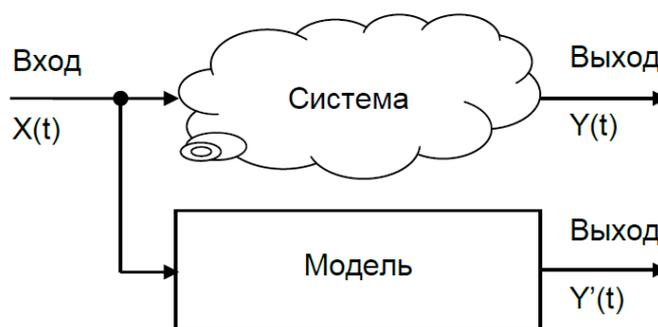


Рисунок 5.1 – Система и модель

В дальнейшем, говоря о системе, мы будем подразумевать наши представления о ней, то есть ее модель.

5.2 Линейные системы

Линейная система – система, осуществляющая линейное преобразование сигнала, например, электрическая цепь с элементами, не зависящими от напряжений и токов. Линейные системы – первое приближение (модель) преобразований сигналов. Математическое определение линейности:

$$y(t) = F\{a x_1(t) + b x_2(t)\} = a F\{x_1(t)\} + b F\{x_2(t)\}.$$

Большое количество реальных систем по преобразованию сигналов мож-

но считать линейными. Например, микрофон является линейной системой (с достаточной степенью точности), так как если в него будут говорить одновременно два человека с разной громкостью, то электрический сигнал на выходе будет взвешенной суммой сигналов (от каждого человека в отдельности) на входе, а коэффициенты будут означать громкость разговора первого и второго человека [29]. Однако в обработке РС широко применяются и нелинейные системы. Например, *компрессор*, *экспандер* и ряд других мгновенных преобразований являются нелинейными преобразованиями.

Инвариантная к временному сдвигу система – это система, сдвиг сигнала на входе которой порождает тот же отклик, только сдвинутый по времени, т.е. если $x(t) \rightarrow y(t)$, то $x(t+\tau) \rightarrow y(t+\tau)$. Это означает, что свойства системы, например микрофонов, не меняются во времени и не зависят от входного сигнала.

Далее мы будем рассматривать линейные инвариантные к сдвигу системы, называя их просто линейными.

Свойства линейных систем:

- а) постоянный (константа) сигнал переводится любой линейной системой в постоянный сигнал;
- б) при прохождении через линейную систему синусоида остается синусоидой. Могут измениться лишь ее амплитуда и фаза (сдвиг во времени).

5.3 Линейные дискретные системы

Дискретные системы – системы, преобразующие сигналы, дискретные или по времени, или по уровню, или уровню и времени (то есть цифровые сигналы). В дальнейшем мы будем называть дискретными системами системы, преобразующие цифровые сигналы. Дискретные системы являются дискретным аналогом непрерывных систем. Соотношение между непрерывной и дискретной системами схематически представлено на рисунке 5.2.

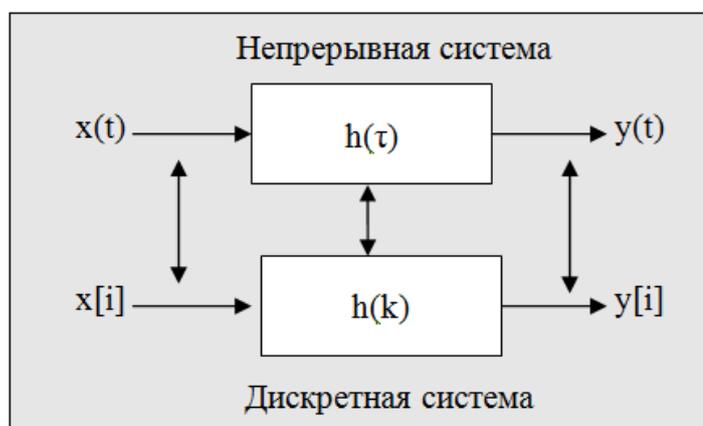


Рисунок 5.2 – Соотношение между непрерывной и дискретной системой

Под *дискретизацией системы* подразумевается преобразование непре-

рывной динамической модели к дискретной форме описания в разностных уравнениях. При этом предполагается, что в моменты $t = iT$ дискретные сигналы $y[i] = y(iT)$ полученной дискретной модели с определенной степенью точности повторяют значения сигналов $y(t)$ исходной непрерывной системы.

Дискретная модель описывает преобразование дискретного входного сигнала $x[i]$ в выходной сигнал $y[i]$:

$$x[i] \rightarrow y[i] = F\{x[i]\},$$

где $F\{ \}$ представляет собой дискретную систему (вычислительный процесс) преобразования входного сигнала $x[i]$ в выходной сигнал $y[i]$.

Простейшими дискретными системами являются *линейные дискретные системы* (ЛДС). Основными операциями в ЛДС являются суммирование, умножение и задержка на один отсчет (рисунок 5.3).

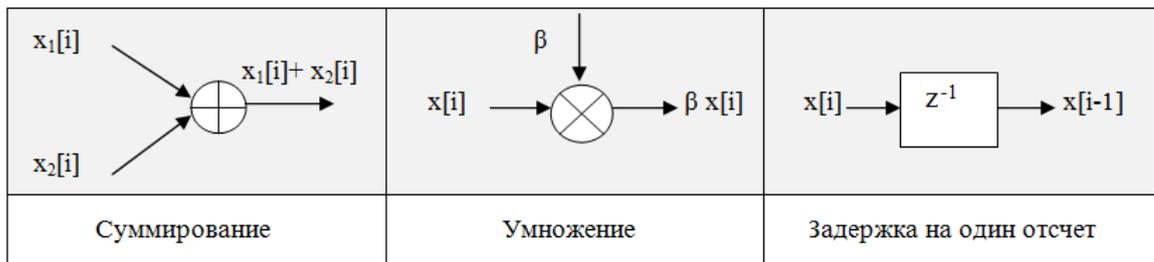


Рисунок 5.3 – Основные операции в ЛДС.
 z^{-1} – оператор задержки сигнала на один отсчет

Свойства линейности ЛДС формулируются так же, как и для непрерывных линейных систем.

Линейность: $y[i] = F\{ax_1[i] + bx_2[i]\} = a F\{x_1[i]\} + b F\{x_2[i]\}$.

Инвариантность во времени: $y[i - K] = F\{x[i - K]\}$.

5.4 Дискретная свёртка

Рассмотрим формирование выходного сигнала в ЛДС. Для этого рассмотрим реакцию системы на цифровую *дельта-функцию* – сигнал вида $\delta[0] = 1$, $\delta[i \neq 0] = 0$.

Любой дискретный сигнал можно разложить в сумму таких функций, сдвинутых во времени. Например, бесконечный сигнал $x[i]$ можно представить в виде:

$$x[i] = \sum_{n=-\infty}^{+\infty} x[n] \delta[i - n] = \sum_{n=-\infty}^{+\infty} x[i - n] \delta[n].$$

Рассмотрим отклик (выходной сигнал) линейной системы на цифровую дельта-функцию. Для этого подадим дельта-функцию в систему и измерим выходной сигнал. Пусть выходной сигнал равен $h[n]$: $\delta[n] \rightarrow h[n]$.

Сигнал $h[n]$ называется *импульсной характеристикой* или реакцией системы (*impulse response*), поскольку он является откликом системы на единичный

импульс. Зная $h[n]$ (отклик системы на дельта-функцию), можно вычислить отклик системы на любой входной сигнал. Действительно, так как любой входной сигнал является линейной комбинацией сдвинутых во времени дельта-функций, то выходной сигнал будет той же самой линейной комбинацией сдвинутых во времени функций $h[n]$. Это следует из линейности системы и инвариантности к сдвигу по времени:

$$y[i] = \sum_{n=-\infty}^{+\infty} x[i-n] h[n] .$$

Таким образом, каждая точка сигнала превращается в функцию h (сдвинутую в нужную позицию и умноженную на величину данной точки сигнала), а потом все эти функции складываются.

Дополнительно будем считать систему *каузальной*, если ее отклик следует за входом (что соответствует физически реализуемым системам).

Рассмотренная операция получения результирующего сигнала по исходному называется *сверткой (convolution)*. Итак, любая линейная система осуществляет свертку входного сигнала со своей импульсной характеристикой. Это записывается так: $y[i] = x[i] * h[i]$. Функция $h[n]$ называется импульсной характеристикой линейной системы.

Свойства свертки представлены ниже.

Для каузальной системы $h[i < 0] = 0$.

*Свертка с единичным импульсом: $x[i] = x[i] * \delta[i]$.*

*Коммутативность: $x[i] * h[i] = h[i] * x[i]$.*

*Ассоциативность: $x[i] * h[i] * g[i] = (x[i] * h[i]) * g[i] = x[i] * (h[i] * g[i])$.*

Дистрибутивность:

$x[i] * (g[i] + h[i]) = x[i] * g[i] + x[i] * h[i]$,

$(x[i] + y[i]) * h[i] = x[i] * h[i] + y[i] * h[i]$.

Периодичность:

если $h[i]$ является периодической, то $h[i] * x[i]$ также периодическая.

5.5 Система с конечной импульсной характеристикой

Систему с конечным числом коэффициентов импульсной характеристики называют системой с конечной импульсной характеристикой (КИХ). КИХ-систему (в отличие от свертки с бесконечным числом коэффициентов) можно представить в следующем виде:

$$y[i] = x[i] * h[i] = \sum_{k=M1}^{M2} x[i-k] h[i].$$

Для каузальной системы $y[i] = x[i] * h[i] = \sum_{k=0, +M} x[i-k] h[i]$.

Для некаузальной системы $y[i] = x[i] * h[i] = \sum_{k=-M1, +M2} x[i-k] h[i]$.

Пример некаузальной системы: $y[i] = 1/4 (x[i-1] + 2x[i] + x[i+1])$.

Частотное описание преобразования КИХ-системы дано ниже.

Представим уравнение свертки в частотной области. ДПФ свертки преобразуется в произведение ДПФ импульсной реакции и ДПФ сигнала.

$$y[i] = h[i] * x[i] \rightarrow \text{DFT}\{h[i] * x[i]\} = \text{DFT}\{h[i]\} \text{DFT}\{x[i]\}.$$

Разобьем входной и выходной сигнала на кадры конечного размера (*framing*). Применим ДПФ к кадрам входного сигнала конечного размера:

$$Y(f, k) = H(f) X(f, k), H(f) = Y(f, k) / X(f, k),$$

где k – индекс кадров, f – индекс частоты.

$$H(f) = \text{DFT}\{h[i]\}, X(f, k) = \text{DFT}\{x[i]\}, Y(f, k) = \text{DFT}\{y[i]\},$$

где $H(f)$ – передаточная функция ЛДС,

$$H(f) = \text{DFT}\{h[i]\} = \sum_{n=0, L-1} h[n] \exp(-j2\pi f T n).$$

Рассмотрим выходной сигнал системы:

$$\begin{aligned} Y(f, k) &= H(k) X(k) = \text{Re}\{Y\} + j \text{Im}\{Y\} = \\ &= (\text{Re}\{H\} + j \text{Im}\{H\}) (\text{Re}\{X\} + j \text{Im}\{X\}) = \\ &= (\text{Re}\{H\} \text{Re}\{X\} - \text{Im}\{H\} \text{Im}\{X\}) + j (\text{Re}\{H\} \text{Im}\{X\} + \text{Im}\{H\} \text{Re}\{X\}). \end{aligned}$$

Поскольку $h[n], x[n]$ – вещественные функции, то $\text{Re}\{H\} \text{Re}\{X\}$ – симметричные, $\text{Im}\{H\} \text{Im}\{X\}$ – антисимметричные функции, поэтому $\text{Re}\{Y\}$ – симметричная, $\text{Im}\{Y\}$ – антисимметричная.

Следовательно, вектор кадра выходного сигнала $Y[k] = \text{IDFT}\{Y(f, k)\}$.

Последовательность кадров выходного сигнала преобразуется в вещественную функцию $y[i]$ – выходного сигнала.

5.6 Система с бесконечной импульсной характеристикой

В общем случае преобразование сигнала в ЛДС может быть представлено линейным разностным уравнением:

$$y[i] = \sum_{n=0, P} b_n x[i-n] - \sum_{m=1, Q} a_m y[i-m].$$

Схема этого преобразования приведена на рисунке 5.4.

Система имеет обратную связь, поскольку задержанный сигнал с выхода системы поступает обратно в систему. Отклик такой системы на единичный им-

пульс может длиться бесконечно. Поэтому такие системы называют системами с бесконечной импульсной характеристикой (БИХ).

Рассмотрим вкратце описание систем с БИХ.

Временное представление: $y[i]*a[i] = b[i]*x[i]$

Частотное представление: $Y(f) A(f) = X(f) B(f)$

АЧХ: $Y(f) = X(f) B(f)/A(f) = H(f) X(f)$, $H(f) = B(f)/A(f)$

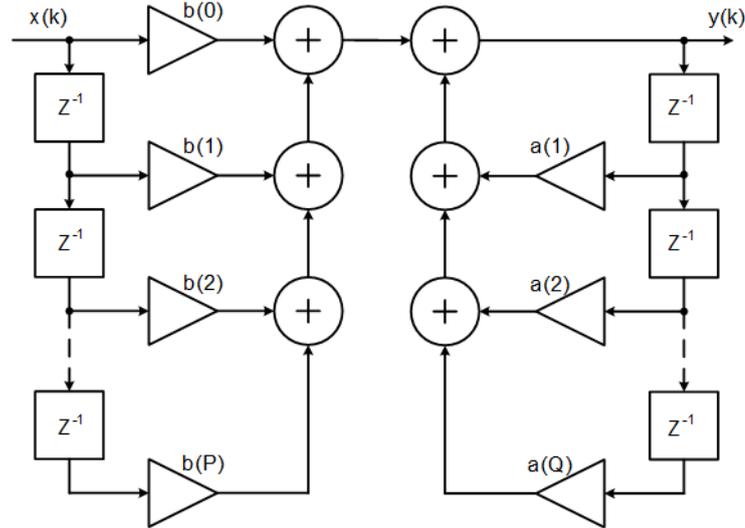


Рисунок 5.4 – Схема линейной дискретной системы

Пример 1

Простая БИХ-система: $y[i] = -a_1 y[i-1] + b_0 x[i] + b_1 x[i-1]$.

Пример 2

Частный случай БИХ-системы (уравнение авторегрессии): $y[i]*a[i] = x[i]$.

5.7 Спектральное описание линейных систем с одним входом и выходом

Предположим, что на вход линейной системы подается некоторый стационарный случайный процесс $x(t)$ с нулевым средним. Тогда выход $y(t)$ будет иметь такие свойства:

$$y(t) = h(t) * x(t),$$

$$Y(f) = H(f) X(f) = |H(f)| e^{j\Phi(f)} X(f),$$

где $|H(f)|$ – модуль передаточной функции системы (амплитудно-частотная характеристика, АЧХ), $\Phi(f)$ – фазовая характеристика передаточной функции (фазово-частотная характеристика, ФЧХ).

Функции плотности спектра мощности $P_{xx}(f)$, $P_{yy}(f)$ и функцию кросс-спектра $P_{xy}(f)$ связывают соотношения:

$$P_{yy}(f) = |H(f)|^2 P_{xx}(f),$$

$$P_{xy}(f) = H(f) P_{xx}(f).$$

Если спектр мощности входа и кросс-спектр мощности известны, то частотная функция отклика определена:

$$H_{xy}(f) = P_{xy}(f)/P_{xx}(f).$$

В случае идеальной линейной системы без шума:

$$\begin{aligned} \Gamma_{xy}^2(f) &= |P_{xy}(f)|^2 / P_{xx}(f) P_{yy}(f) = \\ &= |H(f) P_{xx}(f)|^2 / P_{xx}(f) |H(f)|^2 P_{xx}(f) = 1. \end{aligned}$$

Следовательно, в случае линейных систем функция когерентности достигает своего теоретического максимума, равного единице на всех частотах. Если же функция когерентности меньше единицы, то одной из возможных причин этого может служить отсутствие линейной зависимости выхода от входа рассматриваемой системы, то есть нелинейность системы.

5.8 Оценка характеристик линейных систем

Оценку характеристик систем выполняют по наблюдениям сигнала на входе и выходе системы. Другие методы выполняют по наблюдениям сигнала на выходе системы в предположении, что на вход системы поступает сигнал с известными свойствами, например, белый шум. Рассмотрим некоторые из методов.

Оценка передаточной функции с использованием тестового сигнала

Рассмотрим линейную систему:

$$Y(f) = |H(f)| e^{j\Phi(f)} X(f).$$

Задача заключается в оценке амплитудной и фазовой характеристики передаточной функции $H(f)$.

Оценка при помощи синусоид

Метод реализуется с использованием инструментальных средств анализа аналоговых сигналов. На вход системы подается косинус частоты f_0 :

$$x(t) = A \cos(2\pi f_0 t).$$

Спектр входного сигнала имеет вид:

$$X(f) = A/2[\delta(f-f_0) + \delta(f+f_0)].$$

Тогда спектр выходного сигнала будет иметь вид:

$$Y(f) = |H(f)| e^{j\Phi(f)} A/2[\delta(f-f_0) + \delta(f+f_0)] = |H(f_0)| A/2 (e^{j\Phi(f_0)} + e^{j\Phi(-f_0)}),$$

а сигнал будет содержать информацию об АЧХ и ФЧХ системы:

$$y(t) = A|H(f_0)| \cos(2\pi f_0 t + \Phi(f_0)).$$

Располагая инструментом оценки фаз и амплитуд, можно оценить АЧХ и ФЧХ системы [39]. Изменяя частоту тестового, сигнала можно измерить передаточную функцию в нужном диапазоне.

Оценка при помощи случайного сигнала

Другой метод основывается на цифровой обработке сигналов. На вход системы подается широкополосный сигнал. Дискретные спектры входного и выходного сигналов на кадрах (k) связаны соотношением:

$$Y(f, k) = H(f) X(f, k).$$

Кросс-спектр входного и выходного сигналов определяется соотношением:

$$P_{xy}(f) = \langle X^*(f, k) Y(f, k) \rangle = \langle X^*(f, k) H(f) X(f, k) \rangle = H(f) P_{xx}(f),$$

где $()^*$ – символ комплексного сопряжения.

Отсюда может быть вычислена оценка передаточной функции:

$$\hat{H}(f) = P_{xy}(f) / P_{xx}(f).$$

По передаточной функции можно вычислить импульсную характеристику системы:

$$h[l] = \text{FT}^{-1}\{\hat{H}(f)\}.$$

Необходимо, чтобы входной сигнал имел спектральные компоненты во всем интересующем диапазоне частот, например белый шум.

Описанный метод особенно удобен для качественной экспресс-оценки АЧХ системы, если на ее вход подавать белый шум.

Оценка коэффициентов КИХ-системы по входу и выходу

Рассмотрим процедуру оценки импульсной характеристики КИХ-системы. Запишем отклик системы в векторной форме:

$$y[n] = \mathbf{h}^T \mathbf{X}[k] = \mathbf{X}^T[k] \mathbf{h},$$

где k – индекс кадров, $\mathbf{X}[k] = [x[k], x[k-1], \dots, x[k-L+1]]^T$ – вектор входного сигнала, $\mathbf{h} = [h[0], h[1], \dots, h[L-1]]^T$ – импульсная характеристика.

Запишем соотношения между вектором кросс-корреляции и импульсным откликом:

$$\langle \mathbf{X}[k] y[n] \rangle = \mathbf{C}_{xy} = \langle \mathbf{X}[k] \mathbf{X}^T[k] \mathbf{h} \rangle = \mathbf{A}_{xx} \mathbf{h},$$

где \mathbf{C}_{xy} – вектор кросс-корреляции, \mathbf{A}_{xx} – матрица автокорреляции.

Решая систему линейных уравнений, получим оценку коэффициентов импульсного отклика:

$$\mathbf{h} = \mathbf{A}_{xx}^{-1} \mathbf{C}_{xy}.$$

Оценка коэффициентов авторегрессионной модели по входу и выходу

Представим авторегрессионную модель в векторной форме:

$$y[l] = \mathbf{A}^T \mathbf{Y}[l] + e[l] = \mathbf{Y}^T[l] \mathbf{A} + e[l],$$

где $\mathbf{Y}[l] = [y[l-1], y[l-2], \dots, y[l-p]]^T$ – вектор отсчетов сигнала,

$\mathbf{A} = [a[1], a[2], \dots, a[p]]^T$ – вектор коэффициентов.

Запишем ошибки между сигналом $y[l]$ и выходом фильтра;

$$E\{e[l]^2\} = \langle (y[l] - \mathbf{A}^T \mathbf{Y}[l])^2 \rangle.$$

Минимизация СКО ошибки приводит к оптимальному решению:

$$\langle \mathbf{Y}[l] y[l] \rangle = \mathbf{C}_y = \langle \mathbf{Y}[l] \mathbf{Y}^T[l] \mathbf{A} \rangle = \mathbf{C}_{yy} \mathbf{A}.$$

Решая систему линейных уравнений, получим оценку БИХ-коэффициентов:

$$\mathbf{A} = \mathbf{C}_{yy}^{-1} \mathbf{C}_y.$$

Всеобъемлющее описание методов идентификации систем читатель найдет в книге Гропа [18].

Вопросы и упражнения

1. Система задана следующим уравнением: $y[n] = x[n] + 2x[n+1] + 3$. Является ли система линейной, инвариантной к сдвигу, каузальной?
2. Приведите пример некоммутативной операции.
3. Как вычислить фазовую характеристику линейной системы по входному и выходному сигналам?
4. Почему АЧХ системы вычисляются по усредненным спектрам:

$$\hat{H}(f) = \langle X^*(f, k) Y(f, k) \rangle / \langle X^*(f, k) X(f, k) \rangle,$$

а не мгновенным: $\hat{H}(f) = Y(f, k) / X(f, k)$?

5. Как оценить фазовый сдвиг сигнала на заданной частоте между двумя последовательными кадрами сигнала?
6. При фиксированном временном сдвиге между кадрами фазовая задержка на разных частотах линейно увеличивается пропорционально частоте. Как по функции фазовой задержки построить функцию временной задержки между кадрами или на различных частотах?
7. Запишите соотношения между АЧХ и импульсной характеристикой КИХ-фильтра.

6 ФИЛЬТРАЦИЯ ДИСКРЕТНЫХ СИГНАЛОВ

Одним из применений цифровых систем является преобразование сигналов (и временных рядов) заданным образом. Эти преобразования могут быть выполнены с помощью фильтров. Слово «фильтр» возникло из физических явлений, изучаемых теорией электрических цепей, и получило дальнейшее развитие в теории цифровой обработки сигналов.

Причины популярности фильтров заключаются в следующем:

- а) фильтры – простейшая модель преобразования процессов, сигналов. Когда мы захотим преобразовывать сигналы, то естественно обратиться именно к фильтрам;
- б) линейные фильтры являются простейшими устройствами для преобразования цифровых сигналов;
- в) цифровые фильтры являются преемниками аналоговых фильтров.

Задачи, решаемые с помощью фильтров:

- а) сглаживание;
- б) выделение и подавление отдельных спектральных компонент;
- в) обнаружение событий (импульсов, тональных компонент и пр.);
- г) выделение сигнала в шуме;
- д) моделирование систем.

Теория и методы конструирования цифровых фильтров детально рассмотрены в многочисленных курсах и учебниках по цифровой обработке сигналов, например в книгах [2, 47, 54] и многих других. К настоящему времени процедуры конструирования многих фильтров автоматизированы. Однако в конкретных областях фильтры конструируются с учетом специфики сигналов и применения. В данной главе рассматриваются лишь некоторые фильтры, широко применяемые в ЦОРС.

6.1 Рекурсивные фильтры первого порядка

Простейшие рекурсивные фильтры

Общий вид рекурсивного фильтра первого порядка определяется соотношением:

$$y[i] = \sum_{k=0}^K b[k]x[i - k] - ay[i - 1].$$

Простейшие фильтры первого порядка описываются тремя коэффициентами, значения которых определяются всего одним параметром ($\alpha = 1 - \beta$):

$$y[i] = b_0x[i] + b_1x[i - 1] - ay[i - 1].$$

Формулы для некоторых простейших фильтров первого порядка приведены в таблице 6.1.

Таблица 6.1 – Простейшие фильтры первого порядка [39]

Наименование фильтра	Значения коэффициентов		
	b_0	b_1	a
ФНЧ	$(1 - \alpha) = \beta$	0	$-\alpha = (\beta - 1)$
ФНЧ усиленный	$(1 + \alpha)/2 = (1 - \beta)/2$	$(1 + \alpha)/2 = (1 - \beta)/2$	$-\alpha = -(1 - \beta)$
ФВЧ	$1 - \alpha = \beta$	0	$\alpha = (1 - \beta)$
ФВЧ Конелла (выкалывающий)	$(1 - \alpha/2) = (1 - \beta)/2$	$-(1 + \alpha/2) = -(1 - \beta)/2$	$-(1 - \alpha) = -\beta$
Дифференциатор	1	$-(1 - \alpha) = -\beta$	0
Симметричный дифференциатор	$(1 - \alpha) = \beta$	$-(1 - \alpha) = -\beta$	0

Фильтр экспоненциального сглаживания

Фильтр экспоненциального сглаживания (ФЭС) является самым распространенным фильтром в цифровой обработке речевых сигналов благодаря своей простоте. ФЭС определяется следующими соотношениями:

$$\begin{aligned} b_0 &= \beta, \\ b_1 &= 0, \\ a &= b_0 - 1 = \beta - 1 = -\alpha, \end{aligned}$$

где α – коэффициент забывания /усреднения (*forgetting/averaging factor*), β – постоянная сглаживания (*smoothing factor*).

Используются различные формы записи ФЭС:

$$\begin{aligned} y[i] &= \beta x[i] + \alpha y[i-1], \\ y[i] &= (1 - \beta) y[i-1] + \beta x[i], \\ y[i] &= y[i-1] + \beta (x[i] - y[i-1]), \\ y[i] &= \alpha y[i-1] + (1 - \alpha) x[i] = \beta x[i] + \alpha y[i-1], \\ y[i] &= x[i] - \alpha (x[i] - y[i-1]). \end{aligned}$$

Импульсная реакция ФЭС представляет собой: суммирующее экспоненциальное окно, характеризующее экспоненциальное забывание информации. Отсюда название – фильтр экспоненциального сглаживания.

Пример

Реакция ФЭС на импульсную последовательность с периодом T_0 $x[i] = \sum_k$

$20000 \times \delta(i - kT_0)$, $F_s = 8$ кГц показана на рисунке 6.1.

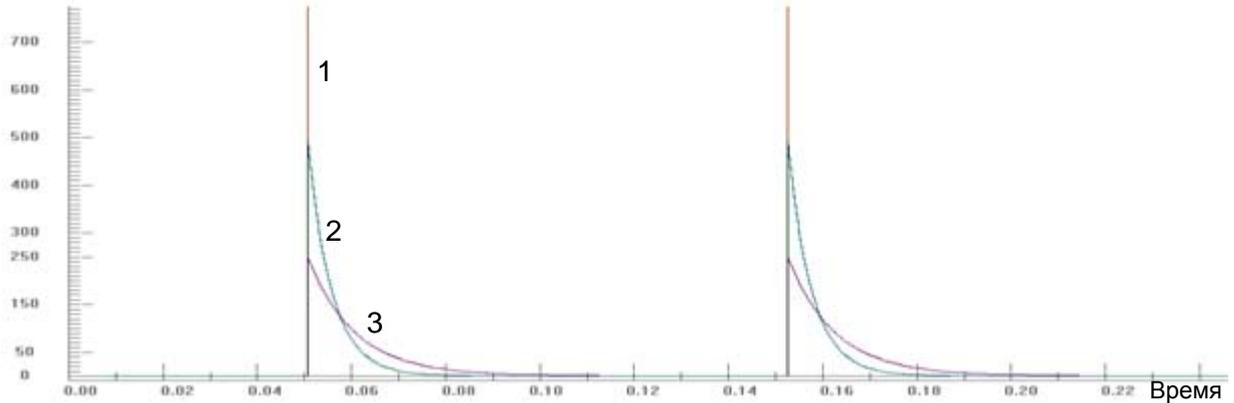


Рисунок 6.1 – Исходный импульсный сигнал (1) и отклики ФЭС: (2) $\beta = 0,0125$ и (3) $\beta = 0,0062$

Из графиков видно, что уменьшение величины постоянной сглаживания β в два раза приводит к уменьшению амплитуды отклика в момент импульса в два раза и увеличению времени отклика. Таким образом, постоянная сглаживания β характеризует способность ФЭС сглаживать случайные вариации сигнала.

В случае ступенчатого изменения амплитуды сигнала (процесса) параметр β характеризует время реакции фильтра на это изменение.

Частотная характеристика ФЭС

Рассмотрим модель ФЭС и запишем ее z -преобразование:

$$y[i] = \alpha y[i-1] + (1-\alpha) x[i],$$

$$Y(z) = \alpha z^{-1} Y(z) + (1-\alpha) X(z).$$

Отсюда получаем передаточную функцию ФЭС:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1-\alpha}{(1-\alpha z^{-1})}.$$

Тогда квадрат АЧХ будет иметь вид:

$$H^2(z) = H(z)H^*(z) = \frac{(1-\alpha)^2}{(1-\alpha z^{-1})(1-\alpha z)}.$$

Учитывая соотношение $\{z = e^{j\Omega}\}$ между z -преобразованием и Фурье-преобразованием [47], получим формулу АЧХ в терминах частоты:

$$H^2(z) = \frac{(1-\alpha)^2}{(1-\alpha e^{-j\Omega})(1-\alpha e^{j\Omega})} = \frac{(1-\alpha)^2}{1-2\alpha \cos(\Omega) + \alpha^2}.$$

Значения АЧХ на границах рабочего диапазона частот (нулевой частоте и частоте Найквиста) такие:

$$\Omega = 0; \cos(0) = 1, H(0) = (1-\alpha)/(1-\alpha) = 1,$$

$$\Omega = \pi; \cos(\pi) = -1, H(\pi) = (1-\alpha)/(1+\alpha).$$

Таким образом, ФЭС сохраняет амплитуду сигнала на нулевой частоте и подавляет высокие частоты, то есть может применяться как ФНЧ, в частности,

для оценки среднего значения процесса или параметра.

Пример

Рассмотрим процесс, представляющий собой сумму константы и белого шума: $x[i] = C + v[i]$. На рисунке 6.2 представлены средние спектры исходного сигнала и сигнала на выходе ФЭС при двух значениях постоянной сглаживания.

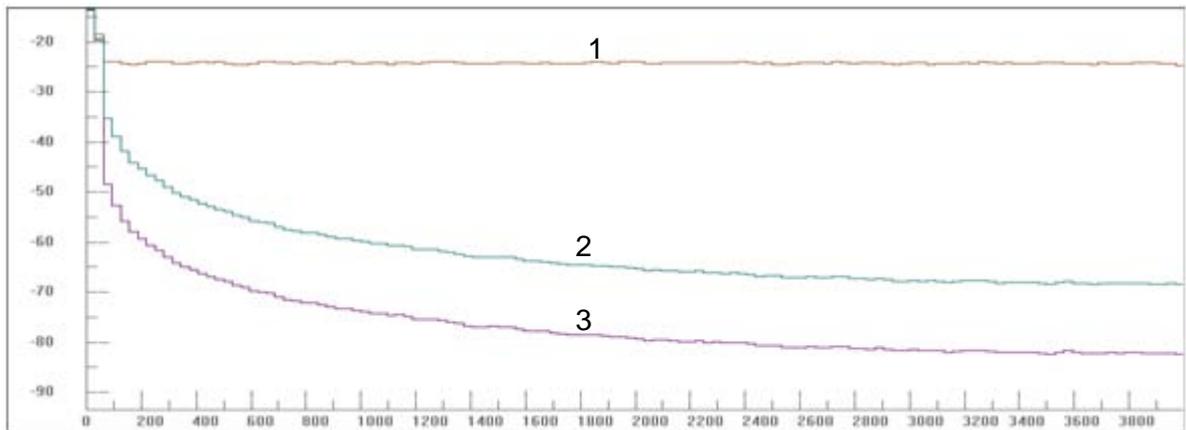


Рисунок 6.2 – Спектры сигналов ($F_s = 8$ кГц): (1) – исходный сигнал, (2) – ФЭС $\beta = 0,0125$ ($\Delta F = 100$ Гц), (3) – ФЭС $\beta = 0,0062$ ($\Delta F = 50$ Гц)

Следует иметь в виду, что ФЭС имеет нелинейную фазовую характеристику, то есть сигналы различной частоты будут иметь на выходе ФЭС не только разную амплитуду, но и разную фазовую задержку. Для компенсации различий фазовой задержки следует профильтровать сигнал (фонограмму) в прямом и обратном направлении.

Задание значений параметров ФЭС

ФЭС работает как ФНЧ с некоторой эффективной полосой частот и временем реакции фильтра на изменение характеристики процесса. Необходимо обеспечить значения этих параметров для произвольной частоты дискретизации F_s .

Связь между коэффициентом сглаживания β и полосой пропускания ΔF ФНЧ выражается следующими соотношениями: $\beta \approx 2\Delta F / F_s \ll 1$. Если $\Delta F = Fn$, то $\beta = 1$ и $y[i] = x[i]$.

Связь между коэффициентом сглаживания β и постоянной времени (адаптации) фильтра T_a выражается соотношением $\beta \approx 1 / TaFs$. Если, $T_a = T$, то $y[i] = x[i]$.

ФЭС является сглаживающим фильтром со скользящим экспоненциальным окном. Этому окну можно сопоставить эквивалентное прямоугольное окно в N_e отсчетов:

$$x_{av}(k) = \frac{1}{N_e} \sum_{k-N_e}^k x[i] = x_{av}(k-1) + \beta \times (x[k] - x_{av}(k-1)).$$

Эквивалентная длина прямоугольного окна равна $Ne = 2/\beta - 1$. Постоянную сглаживания можно вычислить через эквивалентную длину окна:

$$\beta = 2/(Ne + 1) \approx 2/Ne.$$

Примеры

$$Ta = 100 \text{ мс}, Fs = 10 \text{ кГц} \Rightarrow \beta = 0,001, Ne = 1999.$$

$$Ta = 10 \text{ мс}, Fs = 10 \text{ кГц} \Rightarrow \beta = 0.01, Ne = 199.$$

$$\Delta F = 50 \text{ Гц}, Fs = 10 \text{ кГц} \Rightarrow \beta = 0.01, Ne = 1999.$$

Применения ФЭС

Оценка среднего значения и удаление постоянной составляющей

Поскольку $H(0) = 1$, то ФЭС сохраняет амплитуду сигнала на нулевой частоте и поэтому может быть использован для оценки текущего среднего значения зашумленного сигнала.

Пример

Рассмотрим процесс, представляющий собой сумму константы и белого шума: $x[i] = -500 + v[i]$, $\sigma v = 2000$, $Fs = 8 \text{ кГц}$. На рисунке 6.3 показана осциллограмма исходного сигнала $x[i]$ и сигнала на выходе ФЭС.

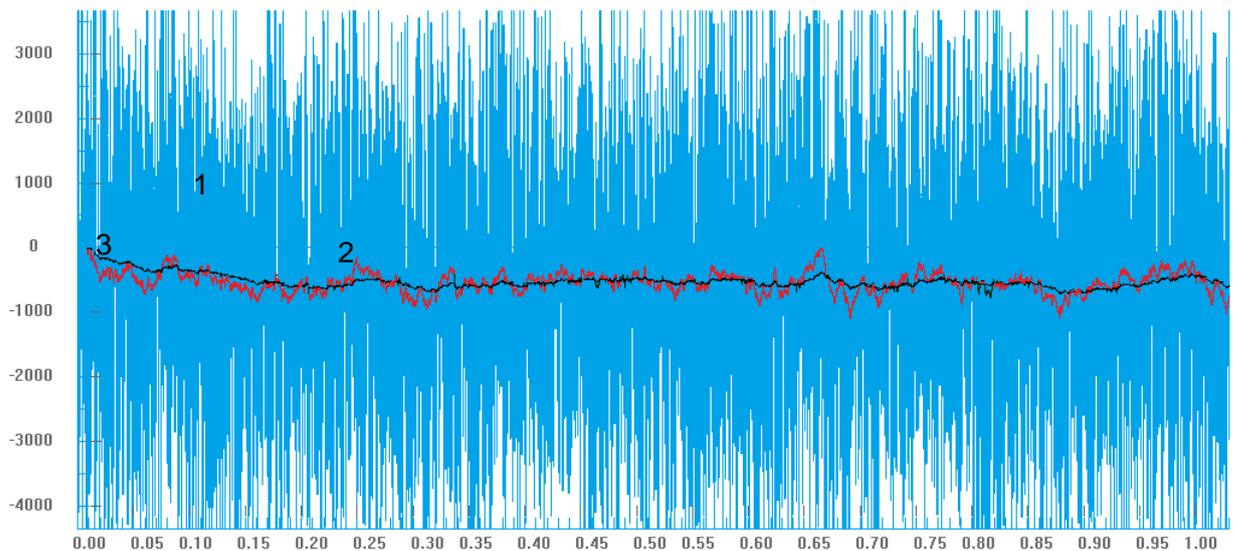


Рисунок 6.3 – Осциллограммы сигнала:

(1) – исходный сигнал, (2) – ФЭС $\Delta F = 50 \text{ Гц}$, (3) – ФЭС $\Delta F = 10 \text{ Гц}$

Графики показывают, что на выходе ФЭС формируется оценка среднего значения (-500). Вариативность оценки зависит от коэффициента сглаживания.

Удаление медленно меняющейся постоянной составляющей сигнала (DC cut)

Модель сигнала: $y[i] = a[i] + x[i]$, где $a[i]$ – тренд.

Оценка тренда: $\hat{a}[i] = \hat{a}[i-1] + \beta (y[i] - \hat{a}[i-1])$, $\beta \ll 1$.

Удаление тренда: $x[i] = y[i] - \hat{a}[i]$.

Оценка средневыпрямленного значения фона

Задача заключается в следующем. Имеем РС в шуме. Необходимо оценить средний уровень амплитуд фонового шума, например средневыпрямленную амплитуду. *Средневыпрямленное значение* сигнала (огibaющая амплитуд) вычисляется с помощью ФЭС:

$$Mx[i+1] = Mx[i] + \beta (|x[i]| - Mx[i]).$$

На интервалах шума алгоритм дает оценку огibaющей амплитуд фона, на интервалах речи – огibaющую РС. Необходимо замедлить обновление оценки огibaющей на интервалах, где амплитуды РС существенно превосходит амплитуду фона. Это может быть осуществлено с помощью управления коэффициентом сглаживания. Если $|x[i]| \gg Mx[i]$, то $\beta[i] = \beta_0 \times (Mx[i]/|x[i]|)$. В этом случае амплитудная огibaющая на интервалах РС будет обновляться замедленно: $Mx[i+1] = Mx[i] (1 + \beta_0)$, $\beta_0 \ll 1$.

Результат работы обычного ФЭС и ФЭС с управляемым коэффициентом сглаживания представлен на рисунке 6.4.

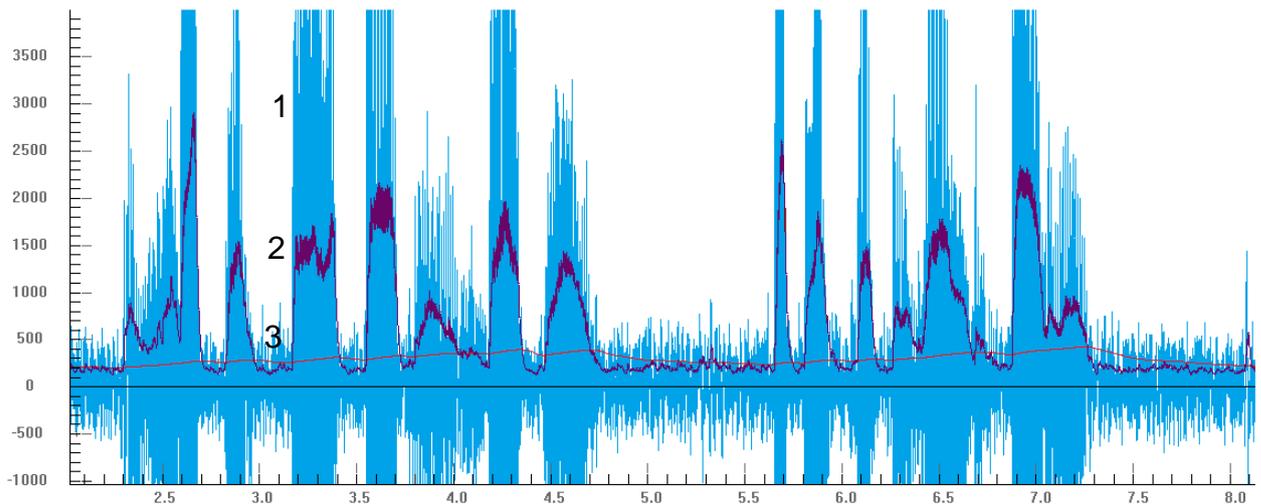


Рисунок 6.4 – Осциллограмма и средневыпрямленные значения сигнала:

- (1) – исходный сигнал, (2) – ФЭС $Ta = 10$ мс,
- (3) – ФЭС с управляемым коэффициентом $Ta = 1$ с

Интегратор

Рекурсивный алгоритм интегратора:

$$y[i] = x[i] + (1 - \beta)y[i - 1] = y[i - 1] + \beta \left(\frac{x[i]}{\beta} - y[i - 1] \right),$$

где $\beta \ll 1$ – коэффициент интегрирования входного сигнала.

Из приведенной формулы следует, что интегратор – это ФЭС с усилением. Усиление обратно пропорционально постоянной сглаживания, то есть памяти фильтра. Таким образом, величина накопленного сигнала будет возрастать с ростом памяти фильтра.

Рассмотрим произвольный рекурсивный фильтр 1-го порядка ($0 < a < 1$):

$$y[i] = bx[i] + ay[i - 1] = y[i - 1] + (1 - a) \left(\frac{b}{1 - a} x[i] - y[i - 1] \right) = \\ = y[i-1] + (1-a)(gx[i] - y[i-1]).$$

Из приведенной формулы следует, что любой рекурсивный фильтр 1-го порядка можно свести к ФЭС с усилением.

Представим рекурсивный фильтр 1-го порядка как усилитель с обратной связью [68]:

$$y[i] = b \times \left(x[i] + \frac{a}{b} y[i - 1] \right) = b \times (x[i] + g \times y[i - 1]).$$

Коэффициент g описывает обратную связь, то есть использование части выхода в качестве входа в более позднее время. Выход фильтра есть сумма двух входов, умноженная на коэффициент усиления b . Мы имеем:

$$(x + gy) \times b = y.$$

Тогда $y = bx/(1 - gb)$.

Пусть $b = -10^9$, $g = 1/10$. Тогда $y = -10 x/(1 + 10^{-8}) \approx -10 x$.

Таким образом, выходной сигнал равен входному, умноженному на минус 10, и не зависит от небольших изменений в характеристиках усилителя (т.е. нечувствителен к неточному значению коэффициента усиления).

6.2 Дифференцирующие фильтры

Основное назначение дифференцирующих фильтров (ДФ) – контрастирование и детектирование коротких временных событий. Соотношения для простейших ДФ приведены в таблице 6.2.

Таблица 6.2 – Простейшие фильтры первой производной

Временные соотношения	z -преобразование передаточной функции
$d[i] = x[i] - \beta x[i-1]$	$H(z) = (1 - \beta z^{-1})$
$d[i] = 0,5(x[i] - x[i-1])$	$H(z) = 0,5(1 - z^{-1})$
$d[i] = x[i] - (x[i+1] + x[i-1])/2$	$H(z) = z(z^{-1} - 0,5(1 - z^{-2}))$
$d[i] = (2x[i+2] + x[i+1] - x[i-1] - 2x[i-2])/10$	$H(z) = 0,1(2z^2 + z - z^{-1} - 2z^{-2})$
$d[i] = (x[i] - x[i-1]) + (1 - \beta) d[i-1]$	$H(z) = (1 - z^{-1}) / (1 - (1 - \beta) z^{-1})$
$d[i] = x[i+1] - x[i-1] + 2(x[i+2] - x[i-2]) + a d[i-1]$	$H(z) = 0,1z^4 (2 + z^{-1} - z^{-3} - 2z^{-4}) / (1 - a z^{-1})$

Поскольку дифференцирование приводит к усилению шумов, спектр которых располагается преимущественно в диапазоне высоких частот, то для ограничения частотной полосы ДФ используется ослабление ВЧ компонент выхода дифференцирующего фильтра, например с помощью ФЭС. Комбинация ДФ и ФЭС представляет собой усреднение разностей, что приводит к ослаблению влияния медленных изменений параметра. Такой фильтр аналогичен фильтру высоких частот.

На рисунке 6.5 представлены спектры сигналов на выходе ДФ и комбинации ДФ и ФЭС.

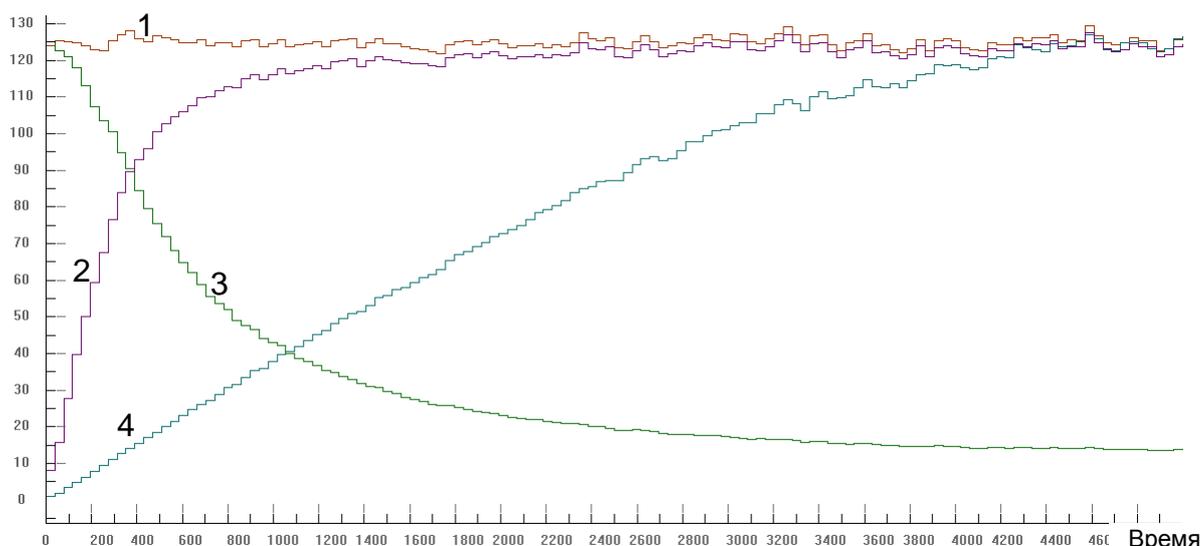


Рисунок 6.5 – Спектр белого шума и сигналов на выходе фильтров:
 (1) – БШ, (2) – совместного ФЭС и ДФ, (3) – ФЭС, (4) – ДФ

Дифференциальные фильтры детально рассмотрены в книге Хэмминга [69].

Операция дифференцирования в некотором смысле является обратной операцией экспоненциального сглаживания. Рассмотрим ФЭС:

$$y[i] = (1 - \beta) y[i-1] + \beta x[i] = \alpha y[i-1] + \beta x[i].$$

Отсюда получаем формулу восстановления исходного сигнала с помощью дифференцирующего фильтра:

$$x[i] = \frac{1}{\beta} (y[i] - \alpha y[i - 1]).$$

Аналогично, если мы применим операцию интегрирования к дифференцированному сигналу $y[i] = x[i] - a x[i-1]$, то сможем восстановить исходный сигнал $x[i] = y[i] + a x[i-1]$.

6.3 Сглаживающие КИХ-фильтры

Фильтр экспоненциального сглаживания является простейшим сглаживающим фильтром и фильтром низких частот. Недостатком этого фильтра является неравномерная по частоте фазовая задержка. От этого недостатка свободны

КИХ-фильтры с симметричной импульсной характеристикой. Сглаживающие КИХ-фильтры также одновременно являются фильтрами низких частот. Рассмотрим физический смысл одного из сглаживающих КИХ-фильтров.

Предположим, что на интервале $[i-M \dots i+M]$ мы аппроксимируем последовательность отсчетов $x[i]$ с помощью многочлена второй степени:

$$x[m] = A + Bm + Cm^2.$$

Найдем коэффициенты из условия минимума суммы квадратов остаточных ошибок:

$$Q(A, B, C) = \sum_{m=-M}^M [x[m] - (A + Bm + Cm^2)]^2 = \\ = \langle [x[m] - (A + Bm + Cm^2)]^2 \rangle_M,$$

где $\langle \rangle_M$ – обозначение операции суммирования: $(\sum_{m=-M, M})$.

$$\min Q(A, B, C) \Leftrightarrow dQ/dA=0, dQ/dB=0, dQ/dC=0 \rightarrow$$

$$A\langle 1 \rangle + B\langle m \rangle + C\langle m^2 \rangle = \langle x[m] \rangle,$$

$$A\langle m \rangle + B\langle m^2 \rangle + C\langle m^3 \rangle = \langle m x[m] \rangle,$$

$$A\langle m^2 \rangle + B\langle m^3 \rangle + C\langle m^4 \rangle = \langle m^2 x[m] \rangle.$$

Взяв симметричный интервал точек $\{-M, \dots, -1, 0, 1, \dots, M\}$, обращаем в нули суммы нечетных степеней. Решив систему уравнений, получим:

$$A = (\langle m^4 \rangle \langle x \rangle - \langle m^2 \rangle \langle m^2 x[m] \rangle) / (\langle 1 \rangle \langle m^4 \rangle - \langle m^2 \rangle^2).$$

Тогда сглаженное значение последовательности $x[m]$ в центральной точке $m = 0$ равно:

$$\hat{x}[0] = A = \sum_{m=-M}^M h[m] x[m].$$

То есть сглаженное значение вычисляется как КИХ-фильтрация входного сигнала. Для произвольного момента времени имеем:

$$\hat{x}[i] = \sum_{m=-M}^M h[m] x[i - m] = x[i] * h[m].$$

Ниже приведены коэффициенты некоторых фильтров, построенные на основе этих соотношений [69]:

а) для полинома 1-й степени: $M = 1, h[m] = 1/4[1, 2, 1]$;

б) для полиномов 2-й степени:

$$M = 2, h[m] = 1/35[-3, 12, 17, 12, -3];$$

$$M = 3, h[m] = 1/21[-2, 3, 6, 7, 6, 3, -2];$$

$$M = 4, h[m] = 1/231[-21, 14, 39, 54, 59, 54, 39, 14, -21];$$

$$M = 5, h[m] = 1/429[-36, 9, 44, 69, 84, 89, 84, 69, 44, 9, -36];$$

в) для полинома 4-й степени: $M = 3, h[m] = 1/231[5, -30, 75, 131, 75, -30, 5]$.

Порядок сглаживающего полинома и ширина окна определяют крутизну переходной области (границы полосы пропускания) и глубину подавления вне полосы пропускания.

Между сглаживающими КИХ-фильтрами (ФНЧ) и КИХ-фильтрами высоких частот существует простая мнемоническая связь:

$$\Phi ВЧ = 1 - \Phi НЧ,$$

$$\Phi НЧ = 1 - \Phi ВЧ.$$

При этом для коэффициентов фильтров выполняются следующие соотношения:

$$\Phi НЧ: \sum h(n) = 1,$$

$$\Phi ВЧ: \sum h(n) = 0.$$

Примеры

$$\Phi НЧ = [\frac{1}{2}, \frac{1}{2}] \leftrightarrow \Phi ВЧ = [1, 0] - [\frac{1}{2}, \frac{1}{2}] = [\frac{1}{2}, -\frac{1}{2}],$$

$$\Phi НЧ = \frac{1}{3} [1, 1, 1] \leftrightarrow \Phi ВЧ = [0, 1, 0] - \frac{1}{3} [1, 1, 1] = \frac{1}{3} [1, -2, 1].$$

6.4 Полосовые фильтры и фильтры-гребенки

Поскольку акустические сигналы часто имеют тональную или гармоническую структуру, то в ряде приложений стоит задача выделения или, напротив, подавления сигнала в узкой полосе частот или для ряда гармоник. Простейшими фильтрами выделения или подавления тонального сигнала являются полосовые фильтры второго порядка: *резонансный фильтр* и *фильтр-пробка*.

Полосовые фильтры второго порядка

Резонансный щелевой фильтр (*Slot, One-formant filter*):

$$y[i] = A \times x[i] + B \times y[i - 1] + C \times y[i - 2].$$

Фильтр-пробка (*Notch, One-zero filter*):

$$y[i] = A_1 \times x[i] + B_1 \times y[i - 1] + C_1 \times y[i - 2],$$

где $C = -\exp(-2\pi \Delta F / Fs)$,

$$B = 2 \exp(-2\pi \Delta F / A) \times \cos(2\pi F_0 / Fs),$$

$$A = 1 - B - C,$$

$$A1 = 1/A,$$

$$B1 = -B/A,$$

$$C1 = -C/A, \Delta F - \text{ширина частотной полосы полосового фильтра,}$$

F_0 – центральная частота полосового фильтра,

Fs – частота дискретизации.

Пример

Графики АЧХ щелевого фильтра и фильтра-пробки приведены на рисунке 6.6.

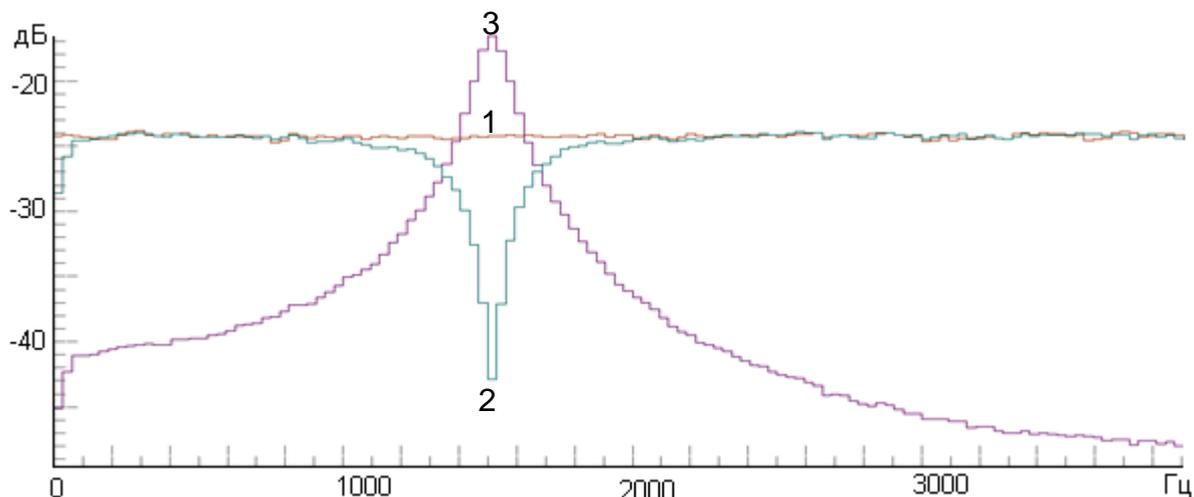


Рисунок 6.6 – АЧХ фильтров: (1) – средний спектр сигнала
(2) – спектр сигнала на выходе фильтра пробки,
(3) – спектр сигнала на выходе щелевого фильтра

Недостаток фильтра-пробки – большая ширина полосы заграждения, щелевого фильтра – большая ширина полоса пропускания.

Фильтры-гребенки

Многие акустические (в том числе речевые) и электрические сигналы представляют собой периодически повторяющиеся импульсы. Спектры таких сигналов является линейчатым и представляют собой последовательность максимумов (гармоник), повторяющихся с шагом (частотой основного тона) F_0 обратно пропорциональным периоду импульсов (основного тона) T_0 :

$$F_n = n \times F_0 = n / T_0, n = 0, 1, 2, \dots$$

Для выделения или подавления таких сигналов используют КИХ- или БИХ-фильтры-гребенки (*comb-filters*). Соотношения для простейших фильтров-гребенок приведены в таблице 6.4, где использованы следующие обозначения:

Z_n – частоты минимумов АЧХ,

M_n – частоты максимумов АЧХ,

$$T = 1/F_s,$$

$F_0 = 1/KT = F_s/K$ – частота ОТ гармоник.

Физическая интерпретация фильтров-гребенок заключается в том, что на выход через фильтр проходят сигналы, оказавшиеся «в фазе», и подавляются сигналы, оказавшиеся «в противофазе». Остальные компоненты сигнала, с разными фазами, в результате микширования частично ослабляются.

Таблица 6.4 – Фильтры-гребенки и их параметры

Формулы фильтров	Нули (Z_n) и максимумы (M_n) АЧХ $n = 0, 1, \dots$
$y[i] = x[i] + x[i-K]$ <i>нечетный (Odd comb)</i>	$Z_n = Fs \times (1/2 + n)/K = (Fs/2K, 3Fs/2K, \dots)$ $M_n = n \times Fs/K = (0, Fs/K, 2Fs/K, \dots)$
$y[i] = x[i] - x[i-K]$ <i>четный (Even comb)</i>	$Z_n = n \times Fs/K = (0, Fs/K, 2Fs/K, \dots)$ $M_n = Fs \times (1/2 + n)/K = (Fs/2K, 3Fs/2K, \dots)$
$y[i] = x[i] + g y[i-K],$ $0 < g < 1$	$M_n = n \times Fs/K = (0, Fs/K, 2Fs/K, \dots)$ <i>Многополосный щелевой (Multiple slot)</i>
$y[i] = x[i] - g y[i-K],$ $0 < g < 1$	$Z_n = n \times Fs/K = (0, Fs/K, 2Fs/K, \dots)$ <i>Многополосный запирающий (Multiple notch)</i>

Временной отклик КИХ-фильтра-гребенки равен времени задержки одного из сигналов. Временной и частотный отклик БИХ-фильтра зависит от времени задержки и величины коэффициента g . Чем ближе значение коэффициента к единице, тем больше время реакции фильтра.

При некоторых обстоятельствах, если существует достаточная задержка сигнала при возвращении с выхода на вход, система начинает осциллировать. Такие ситуации могут возникнуть в больших аудиториях, которые имеют микрофон, сигнал с которого через выход усилителя поступает на акустические колонки системы. Звук акустической колонки задерживается в пространстве на некоторое время, поступает на микрофон и далее после усиления снова поступает на колонку. В результате в аудитории возникает характерный резкий шум. Частота колебаний зависит от характеристик помещения, микрофона, колонки и усилителя.

Таким образом, при использовании систем с обратной связью есть риск возникновения самовозбуждения системы.

Задание параметров фильтра-гребенки

Коэффициент g определяет степень ослабления основного тона. Если $g=1$, то фильтр пропускает все гармоники основного тона (F_0), и подавляет нечетные гармоники $n \times F_0/2$, если $g=-1$, фильтр пропускает нечетные гармоники основного тона $n \times F_0/2$, подавляет четные гармоники.

Допустим необходимо подавить гармоники основного тона с частотой F_0 , то есть нули должны быть на частотах $n \times F_0$, $n=0, 1, 2, \dots$. Тогда модель КИХ-фильтра-гребенки должна быть такой:

$$y[i] = x[i] - g x[i-K], g > 0.$$

Задержка K определяющая частоту следования максимумов и минимумов фильтра, задается на основе следующего соотношения:

$$K = \text{round}(Fs/F_0).$$

Рекурсивный фильтр-гребенка позволяет по сигналу на выходе КИХ-фильтра-гребенки восстановить сигнал на входе:

$$x[i] = y[i] + g x[i-K].$$

Аналогично можно восстановить сигнал на входе БИХ-фильтра-гребенки:

$$x[i] = y[i] + g y[i-K].$$

Ширина максимумов КИХ-фильтра равна ширине минимумов. Минимумы (нули) являются глубокими, поэтому этот фильтр можно эффективно использовать для подавления гармоник. В БИХ-фильтре более эффективно достигается выделение гармоник.

Пример 1

Рассмотрим КИХ-фильтры, аппроксимирующие гребенку гармоник ОТ речевого сигнала в заданном интервале $Fo=100 \dots 330$ Гц. Пусть $Fs = 10$ кГц, тогда $K = 10000/(100 \dots 330) = 100 \dots 33$. Таким образом, длина фильтра соответствует длине обычно применяемых фильтров.

Пример 2

$Fs = 10$ кГц, $Fo = 500$ Гц. На рисунке 6.7 показаны характеристики КИХ-фильтров выделения и подавления гармоник ОТ.

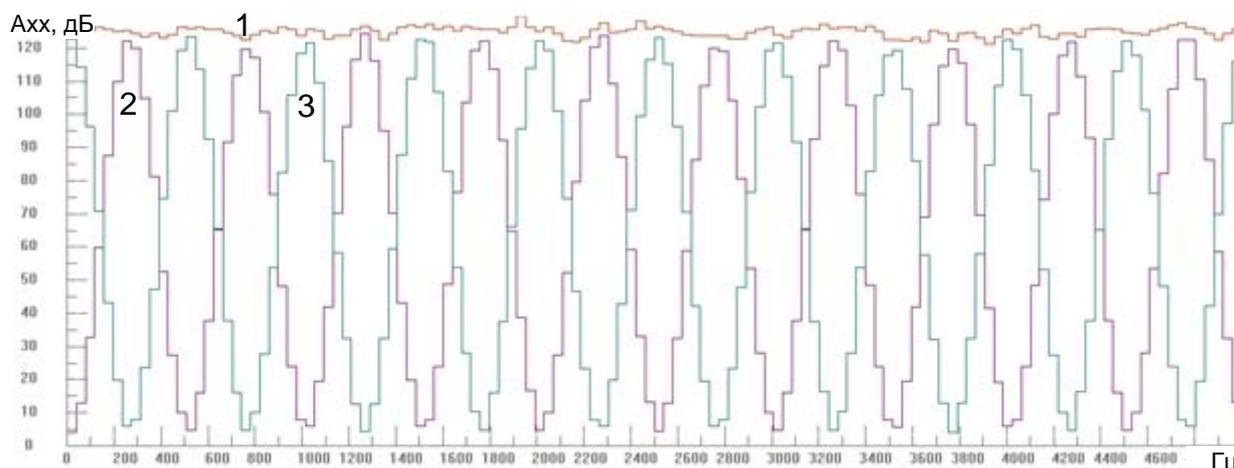


Рисунок 6.7 – Средние спектры белого шума: (1) – до фильтрации, (2) – после фильтра подавления четных гармоник ОТ, (3) – после фильтра подавления нечетных гармоник ОТ

Обобщение

Пусть сигналы поступают на два микрофона с временным сдвигом ΔT . Тогда у сигнала

$$Y_+(t) = X(t) + X(t-\Delta T)$$

нули будут на частотах

$$Zn = (1/2 + n)/\Delta T, n = 0, 1, 2, \dots,$$

а для сигнала

$$Y(t) = X(t) - X(t-\Delta T)$$

нули будут на частотах $Zn = n/\Delta T$. Положение нулей не зависит от частоты дискретизации.

Пример

Данный пример демонстрирует неожиданный эффект микширования сигналов диктофона «Гном» при записи базы данных для задачи распознавания речи с использованием внутренних микрофонов диктофона. Для улучшения качества фонограммы записанные сигналы правого и левого каналов сложили. Однако в результате качество распознавания неожиданно ухудшилось. Причиной оказался комб-эффект, возникший из-за того, что диктор располагался на разном расстоянии от микрофонов, и между сигналами возникла небольшая временная задержка.

Рисунок 6.8 иллюстрирует этот эффект. На рисунке представлены средние спектры сигналов микрофонов и средний спектр микшированного сигнала.

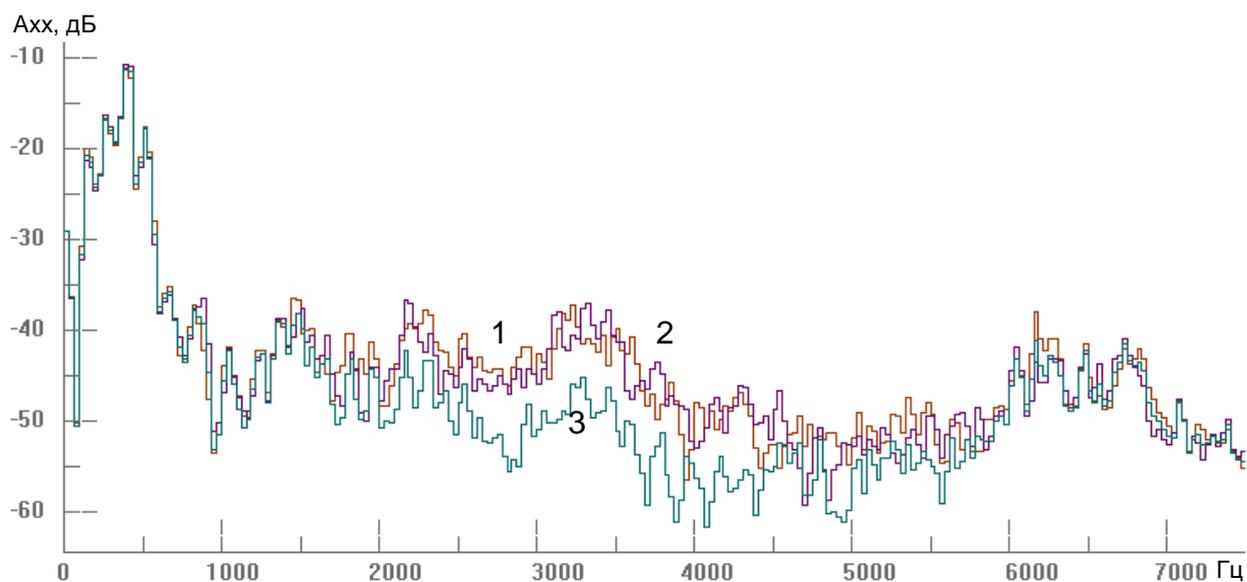


Рисунок 6.8 – Средние спектры речи:
(1) и (2) – сигналы $x_1[i]$ и $x_2[i]$ микрофонов диктофона,
(3) – сигнал $y[i] = \frac{1}{2}(x_1[i] + x_2[i])$

Из рисунка видно, что средние спектры сигналов микрофонов приблизительно одинаковые, а средний спектр микшированного сигнала имеет значительный провал в интервале 3000–4000 Гц. Таким образом, микширование сигналов привело к искажению результирующего сигнала.

Применение фильтров-гребенок

Основными применениями фильтров-гребенок является детектирование гармонических сигналов и подавление гармонических помех. Например, филь-

ры-гребенки могут применяться для выделения РС в присутствии периодических импульсных наводок с известной фиксированной частотой. Основным достоинством фильтров является простота их реализации. Основным недостатком КИХ-фильтров – большая ширина заграждения, что приводит к подавлению значительной части РС. Основным недостатком БИХ-фильтров – большое время реакции (затянутый во времени отклик).

6.5 Фильтры в частотной области

Речевые сигналы занимают широкий спектральный диапазон (несколько октав), поэтому во многих случаях необходимо в разных на разных частотах применять разные правила обработки. Это удобно делать с помощью фильтрации сигналов в частотной области. Другим достоинством обработки в частотной области является ее вычислительная эффективность.

Частотная фильтрация — это процесс обработки звукового сигнала с целью изменения спектрального состава сигнала. Задачами такой обработки могут быть: амплитудно-частотная коррекция сигнала (усиление или ослабление отдельных частотных составляющих); полное подавление спектра сигнала или шумов в определенной полосе частот. Рассмотрим базовые алгоритмы обработки в частотной области.

Принципы фильтрации в частотной области

В некоторых случаях АЧХ-фильтра $H(f)$ задана лишь в частотной области, а фильтрацию необходимо проводить во временной области. Тогда импульсная характеристика фильтра может быть вычислена на основе соотношения:

$$h[i] = \text{FT}[H(f)].$$

Свойства $H(f)$ и $h[i]$ связаны следующими соотношениями:

$H(f)$ – вещественная и симметричная:

$$\text{Im}\{H(f)\} = 0, \text{Re}\{H(f)\} = \text{Re}\{H(-f)\} \Leftrightarrow h[i] \text{ симметричная, вещественная.}$$

$H(f)$ – комплексно сопряженная:

$$\text{Im}\{H(f)\} = -\text{Im}\{H(-f)\}, \text{Re}\{H(f)\} = \text{Re}\{H(-f)\} \Leftrightarrow h[i] \text{ вещественная (но не симметричная).}$$

Фильтрация в частотной области реализуется следующим образом. Сигнал сегментируется на кадры, каждый из которых последовательно преобразуется в комплексный спектр. Комплексные спектры поэлементно умножаются на коэффициенты фильтра, после чего спектры преобразуются в последовательность кадров результирующего сигнала:

$$X(f, k) = \text{FT}\{X[k]\},$$

$$Y(f, k) = H(f) X(f, k),$$

$$Y[k] = \text{IFT}\{Y(f, k)\},$$

где k – индекс кадров, f – индекс частот.

Комплексные спектры сигнала преобразуются в кадры выходного сигнала, из которых «собирается» сигнал на выходе фильтра. Сигнал на выходе фильтра будет вещественным, поскольку выполняются следующие соотношения:

$X[k]$ – вещественная $\Leftrightarrow \text{Re}\{X(f, k)\}$ – симметричная,

$\text{Im}\{X(f, k)\}$ – антисимметричная,

$H(f)$ – вещественная, $\text{Re}\{H(f)\}$ – симметричная,

$\text{Im}\{H(f)\}$ – антисимметричная.

Тогда:

$\text{Re}\{Y(f, k)\}$ – симметричная, $\text{Im}\{Y(f, k)\}$ – антисимметричная

$\Leftrightarrow y[i]$ вещественный (но не симметричный).

Отметим еще одно важное свойство частотного фильтра. В случае, когда АЧХ фильтра является вещественной функцией, фаза выходного сигнала совпадает с фазой входного:

$$\text{tg}[\Phi_y(f)] = \left[\frac{\text{Im}\{Y(f)\}}{\text{Re}\{Y(f)\}} \right] = \left[\frac{H(f)\text{Im}\{X(f)\}}{H(f)\text{Re}\{X(f)\}} \right] = \left[\frac{\text{Im}\{X(f)\}}{\text{Re}\{X(f)\}} \right] = \text{tg}[\Phi_x(f)].$$

Явление Гиббса

В случае покадровой обработки сигнала в частотной области кадр входного сигнала преобразуется фильтром в кадр выходного сигнала. Однако временной отклик фильтра может не укладываться в один кадр. В этом случае на границах кадров обработанного сигнала возникают скачки амплитуды. Это называется явлением Гиббса. Детали описаны, например, в книгах Хемминга [39, 69]. На слух этот эффект проявляется как периодический стук.

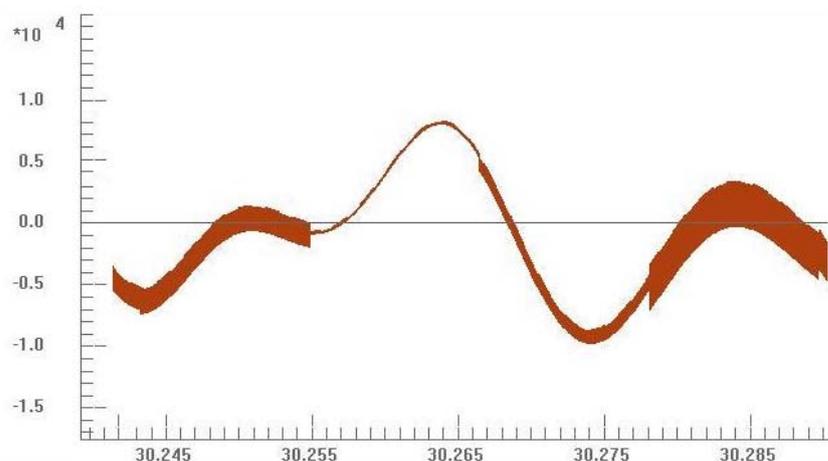


Рисунок 6.9 – Осциллограмма с явлением Гиббса

На рисунке 6.9 приведен фрагмент осциллограммы, на которой проявляется явление Гиббса.

Ослабление эффекта Гиббса достигается с помощью сглаживания частотной характеристики фильтра (когда это возможно) либо с помощью метода пересечения и суммирования (*overlap-add method, OLA*) или пересечения и накопления (*overlap-save method*). Описание алгоритма пересечения и суммирования приведено, например, в [37, 38].

Вопросы и упражнения

1. Докажите что, если импульсная реакция $h[k]$ симметричная, то передаточная функция $H(f)$ вещественная.
2. Выведите формулу передаточной функции дифференцирующего фильтра.
3. Выведите формулу передаточной функции интегратора с утечкой (интегрирующего фильтра).
4. Выведите формулу АЧХ ФЭС.
5. Выведите формулу ФЧХ ФЭС. Как фазовая задержка ФЭС зависит от частоты?
6. Докажите, что при фильтрации вещественного сигнала в частотной области выходной сигнал также будет вещественным, если импульсная характеристика фильтра $h[k]$, по которой с помощью ДПФ построена передаточная функция $H(f)$, является вещественной.
7. Определите, с помощью какого фильтра-гребенки можно подавить тональную помеху с частотами 250, 350, 450, ... Гц дискретного сигнала с $F_s = 16$ кГц.
8. Докажите, что ФЭС является линейным фильтром.
9. Докажите, что для КИХ-фильтров ФВЧ: $\sum_k h[k] = 0$, ФНЧ: $\sum_k h[k] = 1$.
10. Частота дискретизации сигнала равна F_s . Усредненная характеристика вычисляется с использованием алгоритма экспоненциального сглаживания на окнах размера N , следующих с шагом L . Чему должен быть равен коэффициент сглаживания, если мы хотим оценивать среднюю характеристику на временном интервале T_a ?
11. Частота дискретизации сигнала равна 16 кГц. Через каждые L отсчетов измеряется мгновенное значение параметра $P[k]$, после чего мгновенные значения усредняются с помощью алгоритма экспоненциального сглаживания:

$$Pa[k] = Pa[k-1] + \beta [P[k] - Pa[k-1]].$$

Какова величина параметра сглаживания β , если постоянная времени сглаживания T_a составляет 1 с?

ЗАКЛЮЧЕНИЕ

Речевые технологии продолжают оставаться активно развивающейся областью, в которой трудятся большие коллективы ученых и инженеров. Каждый год в этой области разрабатываются новые методы, алгоритмы и устройства и публикуется несколько книг по различным вопросам РТ.

Цифровая обработка речевых сигналов как элемент РТ не является исключением. ЦОРС – это одна из активно развивающихся областей речевых технологий. Остается большое число нерешенных проблем, в частности, связанных с задачами дистанционного распознавания речи, человеко-машинного взаимодействия, слухом роботов, многоканальной обработки сигналов микрофонных решеток и многие другие.

Интенсивно развивающимся направлением речевых технологий на данном этапе являются методы на основе нейронных сетей. Наряду с этим направлением интенсивно развиваются технологии пространственной фильтрации РС с использованием микрофонных решеток, методы нелинейной фильтрации и ряд других. Все эти направления базируются на цифровой обработке сигналов.

Описанные в пособии характеристики и алгоритмы цифровой обработки сигналов могут быть полезны для работы не только с речевыми сигналами, но и в других областях, использующих методы анализа и обработки цифровых сигналов.

Ограниченный объем пособия не позволил рассмотреть углубленно многие важные вопросы ЦОРС. Некоторые из них (по материалам других модулей курса) мы планируем осветить в следующих изданиях. Однако большую часть тематики ЦОРС придется в дальнейшем осваивать самостоятельно.

В заключение хочу дать совет. Столкнувшись с необходимостью решить какую-то проблему, попытайтесь сначала предложить ваше собственное решение, а после обращайтесь к другим источникам. Это полезно по трем причинам.

Во-первых, это развивает ваши творческие способности.

Во-вторых, это нередко помогает понять и оценить идеи других авторов.

В-третьих, наконец, ваша собственная идея может оказаться более плодотворной, чем те, что вы найдете у других.

Желаю творческих успехов!

ПРИЛОЖЕНИЕ А
ПРОФЕССИОНАЛЬНЫЕ СООБЩЕСТВА В ОБЛАСТИ
РЕЧЕВЫХ ТЕХНОЛОГИЙ

Наименование ресурса	Адрес ресурса в сети Интернет
DSP groups: + SP processing	http://www.dsprelated.com http://docs.yahoo.com/info/term
Defence Advanced Research Programs Agency (DAPRA)	www.darpa.mil
International Speech Communication Association (ISCA)	www.isca-speech.org
Acoustical Society of America (ASA)	www.acousticalsociety.org
Applied Voice Input/Output Society (AVIOS)	www.avios.org
Institute of Electrical and Electronics Engineers (IEEE)	www.ieee.org
Human Language Technologies Agency (HLT)	www.hltcentral.org
HMM Tool Kit Software (HTK)	www.htk.eng.com.ac.uk
IEEE Signal Processing Society (IEEE SPS)	www.signalprocessingsociety.org
Audio Engineering Society (AES)	www.aes.org
European Association for Speech, Signal, and Image Processing (EURASIP)	www.eurasip.org
European Language Resources Association (ERLA)	www.icp.inpq.fr/ERLA
European Acoustics Association (EAA)	www.eaa-fenestra.org
Forensic analysis of audio evidence	www.ic.orni.gov
International Computer Science Institute (ICSI)	www.icsi.berkeley.edu
IEEE Signal Processing Society (IEEE SPS)	www.signalprocessingsociety.org

ПРИЛОЖЕНИЕ Б
СИСТЕМЫ СТАНДАРТИЗАЦИИ

Наименование ресурса	Адрес ресурса в сети Интернет
РОССТАНДАРТ Федеральное агентство по техническому регулированию и метрологии	https://www.rostandart.ru https://www.gost.ru
ITU International Telecommunication Unite	www.itu.int
ISO International Organization for Standardization	www.iso.org
ANSI American National Standards Institute	www.ansi.org
NIST National Institute of Standards and Technology	www.nist.gov
ETSI European Telecommunications Standards Institute	http://www.etsi.org http://portal.etsi.org
DIN Deutsches Institut für Normung	https://www.din.de
COCOSDA The International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques	www.cocosda.org
MPEG Moving Picture Experts Group	https://mpeg.chiariglione.org
AES Standards Committee	http://www.aes.org/standards/
IEC International Electrotechnical Commission Международная электротехническая комиссия (МЭК)	https://www.iec.ch/homepage

ПРИЛОЖЕНИЕ В
ОСНОВНЫЕ ЖУРНАЛЫ ПО РЕЧЕВЫМ ТЕХНОЛОГИЯМ

IEEE Proceedings
IEEE Transactions on Signal Processing (IEEE Signal Processing Society)
IEEE Transactions on Acoustics Speech and Signal Processing
IEEE Transactions on Speech and Audio Processing
IEEE Journal of Selected Topics in Signal Processing
IEEE Acoustics, Speech, and Signal Processing Magazine
IEEE Transactions on Audio, Speech, and Language Processing
IEEE Transactions on Information Forensics and Security
IEEE Transactions on Image Processing
IEEE Transactions on Multimedia
IEEE Journal on Machine Learning Applications in Speech
Journal of the Acoustical Society of America (JOSA)
EURASIP Journal on Advances in Signal Processing
Speech Communication (EURASIP and the ISCA)
EURASIP Journal on Audio, Speech, and Music Processing
Journal of the Audio Engineering Society (AES)
Journal of Signal Processing (Japan Advanced Institute of Science and Technology)
Acta Acustica (The European Acoustics Association)
Applied Acoustics
International Journal of Computer Science
International Journal of Speech Technology
Information and communication technologies and services
Circuits, Systems, and Signal Processing (Springer)
Acoustic Science & Technology (The Acoustical Society of Japan)

Журналы на русском языке:

Речевые технологии (уклон в лингвистику).

Труды СПИИРАН (Санкт-Петербургский институт информатики и автоматизации Российской академии наук).

Цифровая обработка сигналов (статьи по речевым сигналам редко).

Научно-технический вестник информационных технологий, механики и оптики (статьи по речевым сигналам редко).

Известия вузов. Приборостроение (статьи по речевым сигналам редко).

Радиотехника (статьи по речевым сигналам редко).

ПРИЛОЖЕНИЕ Г
ОСНОВНЫЕ КОНФЕРЕНЦИИ ПО РЕЧЕВЫМ ТЕХНОЛОГИЯМ

INTERSPEECH

International Conference Acoustic Signal and Speech Processing (ICASSP)

International Convention of the Audio Engineering Society (AES)

EURASIP conferences

European Signal Processing Conference (EUSIPCO)

SIGNAL PROCESSING

Military Speech Technology Conferences

International Workshop on Acoustic Signal Enhancement (IWAENC)

International Workshop on Acoustic Echo and Noise Control

International Conference Speech and Language Processing (ICSLP)

International Congress on Sound and Vibration (ICSV)

SpeechTEK

International Conference on Signals and Systems

Neural Information Processing Systems (NIPS)

International Conference on Auditory Displays

International Conference on Digital Audio Effects

International Conference on Spatial Audio

Digital Audio Processing for Loudspeakers and Headphones

Конференции в Российской Федерации

Цифровая обработка сигналов, организует Институт проблем управления (Москва)

Speech and Computer (SPECOM), организует СПИИРАН (Санкт-Петербург)

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аграновский, А. В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов / А. В. Аграновский, Д. А. Леднов. – М.: Радио и связь, 2004. – 164 с.
2. Айфичер Э., Джервис Б. Цифровая обработка сигналов. Практический подход // М.: Изд. дом «Вильямс», 2004. – 992 с.
3. Андерсон Т. Статистический анализ временных рядов. – М., Мир, 1976.
4. Ахмад, Х. М., Жирков В. Ф. Введение в цифровую обработку речевых сигналов: учеб. пособие // Владим. гос. ун-т. – Владимир: Изд-во Владим. гос. ун-та, 2007. – 192 с.
5. Бендатт Дж., Пирсон А. Измерение и анализ случайных процессов // М.: Мир, 1974.
6. Бендатт Дж., Пирсон А. Применения корреляционного и спектрального анализа // М.: Мир, 1983.
7. Бендатт Дж., Пирсон А. Прикладной анализ случайных данных // М.: Мир, 1989. – 540 с.
8. Богоявленский С.Б. Теоретические и практические аспекты принятия решений в условиях неопределенности и риска. – СПб.: Изд-во СПбГЭУ, 2014.
9. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление. М., Мир, 1974, вып. 1.
10. Болл Р. М. и др. Руководство по биометрии // М.: Техносфера, 2007. – 368 с.
11. Брэгг У. Мир света. Мир звука // М.: Наука, 1967. — 336 с.
12. Винцюк Т.К., Анализ, распознавание и интерпретация речевых сигналов — Киев: Наукова думка, 1987. – 264 стр.
13. Васильев В., Гуров И. Компьютерная обработка сигналов в приложении к интерферометрическим системам – СПб.: БВХ-Санкт-Петербург, 1998. – 240 с.
14. Борискевич А.А. Цифровая обработка речи и изображений. Минск, 2010. [Электронный ресурс]: URL: <http://gendocs.ru/v35045/?cc=8> (дата обращения 15.05.2020)
15. Гоулд Б., Рейдер Ч. Цифровая обработка сигналов // М.: Сов. Радио, 1973. — 367 с.

16. Горшков Ю. Г. Обработка речевых и акустических биомедицинских сигналов на основе вейвлетов. 2017
17. ГОСТ 8.417-81. Государственная система обеспечения единства измерений. Единицы величин
18. Гроп Д. Методы идентификации систем // М.: Мир, 1979. – 302 с.
19. Дворянкин С.В. Цифровая шумоочистка аудиоинформации. Под ред. А.В. Петракова. // М.: ИП РадиоСофт, 2011. – 208 с.
20. Дегтяров Н.П. Параметрические и информационное описание речевых сигналов.– Минск, 2003 – 216 с.
21. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения // М.: Мир, 1971. Вып. 1. – 316 с.
22. Дьяконов В. П., Абраменкова И. MATLAB. Обработка сигналов и изображений. Специальный справочник // СПб: Питер, 2002.
23. Кирн П. Цифровой звук. Реальный мир.// М.: «Изд. дом Вильямс», 2008. – 720 с.
24. Ковалгин Ю.А. Вологдин Э.И. Цифровое кодирование звуковых сигналов. – СПб.: Корона-принт, 2004.– 240с
25. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров, – М.: Мир, 1970.
26. Клюкин И. И. Удивительный мир звука // Л.: Судостроение, 1978.
27. Ли У. А. Методы автоматического распознавания речи. М., Мир, 1983.
28. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. – Минск: «Белорусская наука», 2008.
29. Лукин А. Введение в цифровую обработку сигналов (математические основы) // Лаборатория компьютерной графики и мультимедиа, МГУ, 2007.
30. Макс Ж. Методы и техника обработки сигналов при физических измерениях. В 2-х томах – М.: Мир, 1983.
31. Маркел Дж. Д., Грей А. Х. Линейное предсказание речи. – М.: Связь, 1980. – 308 с.
32. Марпл С. Л. Цифровой спектральный анализ и его применение. – М.: Мир, 1990.

33. Матвеев Ю.Н., Симончик К.К., Тропченко А.Ю., Хитров М.В. Цифровая обработка сигналов. Учебное пособие // СПб.: СПбНИУ ИТМО, 2013. – 166 с.
34. Михайлов, В. Г. Измерение параметров речи / В. Г. Михайлов, Л. В. Златоустова; под ред. М. А. Сапожкова. – М.: Радио и связь, 1979. – 416 с.
35. Мороз Г. Компьютерная фонетика. Как пользоваться PRAAT (2017). [Электронный ресурс]: URL: http://web-corpora.net/~agricolamz/talks/17.03.22.Praat/17.03.22_moroz_Praat.pdf. (дата обращения 15.05.2020)
36. Назаров М.В. Методы цифровой обработки и передачи речевых сигналов / М.В. Назаров, Ю.Н. Прохоров. – М.: Радио и связь, 1985 – 176 с.
37. Оппенгейм А., Шафер Р. Цифровая обработка сигналов. – М.: Связь, 1979. – 416 с.
38. Оппенгейм А., Шафер Р. Цифровая обработка сигналов [Электронный ресурс]: издание 3-е, исправленное / А. Оппенгейм. – М.: Техносфера, 2012. – 1048 с. – Режим доступа: ЭБС «Айбукс». – Неогранич. доступ
39. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. Основные методы // М.: Мир, 1982. – 432 с.
40. Применение цифровой обработки сигналов // Под ред. А. Оппенгейма. – М.: Мир, 1980. – 552 с.
41. Папулис А. Теория систем и преобразований в оптике // М.: Мир, 1971.
42. Пешель М. Моделирование сигналов и систем // М.: Мир, 1981.
43. Плескунин В.И., Воронина Е.Д. Теоретические основы организации и анализа выборочных данных в эксперименте. Л., Изд-во Ленингр. Ун-та, 1979.
44. Попов О. Б. Компьютерный практикум по цифровой обработке аудиосигналов. Учебное пособие для ВУЗов. – М.: Горячая линия – Телеком, 2010. – 176 с.
45. Прокудин А, Oetzmann A., Mazzoni D. Audacity 1.2.1. Вводный курс, 2004. [Электронный ресурс]: URL: <http://ikonnikovo.narod.ru/Soft/Audacity/audacity.pdf>. (дата обращения 15.05.2020)
46. Прохоров Ю. Н. Статистические модели и рекуррентное предсказание речевых сигналов. – М.: Сов. Радио, 1984. – 240 с.
47. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. – М.: Мир, 1978. – 848 с.

48. Рабинер Л. Р., Шафер Р.В. Цифровая обработка речевых сигналов // М.: Радио и связь, 1981. – 395 с.
49. Радзишевский А. Основы аналогового и цифрового звука // М., СПб., Киев, Изд-во дом «Вильямс», 2006/ – 281.
50. Рамишвили Г. С. Автоматическое распознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.
51. Рылов А.С. Анализ речи в распознающих системах // Минск: Бестпринт, 2003. – 262 с.
52. Сапожков М. А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. – 367 с.
53. Сиберт У. Цепи, сигналы, системы // М.: Мир, 1988.
54. Солонина А.И., Клионский Д.М., Меркучева Т.В., Перов С.Н. Цифровая обработка сигналов и MATLAB. СПб: «БХВ-Петербург», 2013.
55. Сорокин В.Н. Теория речеобразования. М.: Радио и Связь. 1985.
56. Сорокин В.Н. Синтез речи. М.: Наука, 1992.
57. Сорокин В.Н. Теоретические основы речевых технологий. Тезисы краткого курса лекций // Институт проблем передачи информации РАН. [Электронный ресурс]: URL: <http://iitp.ru/upload/content/1374/Sorokin.pdf>. (дата обращения 15.05.2020).
58. Сорокин В.Н. Речевые процессы. М.: Народное образование, 2012.
59. Столбов М. Б. Кассу А.-Р. М. Цифровая обработка речевых сигналов: Учебно-методическое пособие по лабораторному практикуму – СПб.: НИУ ИТМО, 2016. – 71 с.
60. Тампель И.Б., Карпов А.А. Автоматическое распознавание речи. Учебное пособие. – СПб.: Университет ИТМО, 2016. – 138 с.
61. Топников, А. И. Цифровая обработка речевых сигналов: практикум / А. И. Топников ; Яросл. гос. ун-т им. П. Г. Демидова.– Ярославль : ЯрГУ, 2018. – 40 с.
62. Тропченко А Ю., Тропченко А.А. Цифровая обработка сигналов. Методы предварительной обработки. Учебное пособие по дисциплине «Теоретическая информатика». – СПб.: СПбГУ ИТМО, 2009 – 100 с.
63. Тэйлор Р. Шум. – М.: Мир, 1978. – 308 с.

64. Уидроу Б., Стирнз С. Адаптивная обработка сигналов. – М.: Радио и связь, 1989.
65. Фант. Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 283 с.
66. Фланаган Дж. Анализ, синтез и восприятие речи. – М.: Связь, 1968.
67. Фролов А. В. Синтез и распознавание речи. Современные решения / А. В. Фролов, Г. В. Фролов. – М. : Связь, 2003. – 216 с.
68. Хемминг Р. В. Численные методы для научных работников и инженеров. – М.: Наука, 1972. – 400 с.
69. Хемминг Р. В. Цифровые фильтры. – М.: Недра, 1987.
70. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации. – М.: Сов. Радио, 1971. – 256 с.
71. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи.– М.: Радио и связь, 2000.– 454 с.
72. Audacity. [Электронный ресурс]: URL: <https://audacity-pro.site>. (дата обращения 15.05.2020).
73. Benesty J., Sondhi M., Huang Y. (Eds). Springer Handbook of Speech Processing // Springer, 2008, 1161 P.
74. PRAAT Short Tutorial (2003). [Электронный ресурс]: URL: https://web.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf. (дата обращения 15.05.2020)
75. SIS II Специализированный звуковой редактор STC-S521. Руководство пользователя. ЦВАУ 00432-01 94
76. VOICEBOX: Speech Processing Toolbox for MATLAB. [Электронный ресурс]: URL: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.htm> (дата обращения 15.05.2020).

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

- АКФ – автокорреляционная функция
- АЧХ – амплитудно-частотная характеристика
- АР – авторегрессия (модель временного ряда)
- АРСС – авторегрессия-скользящее среднее (модель временного ряда)
- АФ – адаптивный фильтр
- АЦП – аналого-цифровой преобразователь
- БИХ – бесконечная импульсная характеристика
- БПФ – быстрое преобразование Фурье
- БШ – белый шум
- ВЧ – высокие частоты
- Гц – герцы (частота)
- дБ – децибел
- ДПФ – дискретное преобразование Фурье
- ДФ – дифференцирующий фильтр
- ДШ – дисперсия шума
- КАС – кратковременный анализ сигналов
- Кбит – килобит
- КБ – килобайт
- кГц – килогерц
- КИХ – конечная импульсная характеристика
- КСА – кратковременный спектральный анализ сигналов
- КСП – кросс-спектр (плотность кросс-спектра)
- КЛП – коэффициенты линейного предсказания
- КФ – корреляционная функция
- КМК – квадрат модуля когерентности
- ЛДС – линейная дискретная система
- МБ - мегабайт
- МНК – метод наименьших квадратов
- МР – микрофонная решетка
- НЧ – низкие частоты
- НЧ-фильтрация – выделение низкочастотной составляющей сигнала
- ОБПФ – обратное быстрое преобразование Фурье

ОДПФ – обратное дискретное преобразование Фурье
ОСШ – отношение сигнал-шум
ОТ – основной тон гармоник
ПКС – плотность кросс-спектра
ПЛИС – программируемые логические интегральные схемы
ПСМ – плотность спектра мощности
ПФ – преобразование Фурье
РС – речевой сигнал
РТ – речевые технологии
СА – спектральный анализ
СИФ – спектральный инверсный фильтр
СКО – средняя квадратическая ошибка оценки
СПМ – спектральная плотность мощности
СС – скользящее среднее (модель временного ряда)
ФКК – функция кросс-ковариации
ФЛП – фильтр линейного предсказания
ФНЧ – фильтр низких частот
ФВЧ – фильтр высоких частот
ФШШ – фильтр широкополосного шума
ФЧХ – фазово-частотная характеристика
ФЭС – фильтр экспоненциального сглаживания
ЦАП – цифро-аналоговый преобразователь
ЦОС – цифровая обработка сигналов
ЦОРС – цифровая обработка речевых сигналов
ЦПОС – цифровой процессор обработки сигналов

СПИСОК СИМВОЛОВ И ОБОЗНАЧЕНИЙ

A

A_p – среднее значение параметра p

$A_p[k]$ – среднее значение параметра на кадре k

A_k – коэффициенты линейного предсказания (КЛП)

A – вектор КЛП $A = [A_1, A_2, \dots, A_p]^T$

$A[k]$ – вектор КЛП на кадре k

$Ax[f]$ – средний амплитудный спектр сигнала $x[i]$

$Ax[n]$ – средний дискретный амплитудный спектр сигнала $x[i]$

$Ax[n, k] = |Xn(k)|$ – дискретный амплитудный спектр сигнала $x[i]$ на кадре k

a_i – рекурсивные коэффициенты фильтра

A_{xx} – матрица автокорреляции

B

b – шаг (бин) дискретных отсчетов в области частот ($= 1/NT$) в герцах

b_i – нерекурсивные коэффициенты фильтра

C

c – скорость звука в воздухе (340 м/с)

$C_{xy}[m]$ – дискретная кросс-ковариационная функция двух процессов, соответствующая временному сдвигу m

C_{xy} – вектор кросс-ковариации двух процессов

$CCx[m]$ – коэффициенты кепстра

$CCx[m, k]$ – коэффициенты кепстра на k -м кадре

$cov(x, y)$, C_{xy} – ковариация двух случайных величин

D

D – дистанция

D_x – дисперсия случайной переменной x

DFT{ } – дискретное прямое преобразование Фурье

E

e – 2,7182818...

$\exp(x) = e^x$

E_x – энергия сигнала

$E_x[k]$ – энергия сигнала на k -м кадре

$e_p[i]$ – огибающие мощности

$e_x[i]$ – огибающие сигнала

$E\{x\}$ – математическое ожидание случайной величины x

$E\{x[i]\}$ – математическое ожидание дискретного сигнала $x[i]$

F

f – частота (циклов, выборок, операций в секунду)

F_s – частота дискретизации сигнала (Гц) ($= 1/T$)

F_n – частота Найквиста (Гц)

F_k – частота следования кадров (Гц)

F_m – модуляционная частота (Гц)

F_o – частота основного тона (Гц)

G

$G_{xy}[n]$ – дискретная квадратичная функция когерентности

$G[n, k]$ – частотный коэффициент передачи, целевая функция фильтра широкополосного шума для бина n на кадре k

$G(f, k)$ – частотный коэффициент передачи, целевая функция фильтра широкополосного шума на частоте f на кадре k

g – коэффициент передачи

$g[i]$ – коэффициент передачи для дискретного момента времени i

$gSNR$ – изменение отношения сигнал-шум

H

$h[i]$ – функция импульсного отклика

$H(f)$ – амплитудно-частотная характеристика

$H[n]$ – дискретная амплитудно-частотная характеристика

$H\{ \}$ – преобразование, выполняемое системой

I

i – дискретное время (временной индекс)

$\text{Im}\{ \}$ мнимая часть числа, заключенного в скобки

$\text{IDFT}\{ \}$ – дискретное обратное преобразование Фурье

J

$j = \sqrt{-1}$ – мнимая единица

K

k – индекс кадров сигнала

L

L – индекс шага кадров

$L\{ \}$ – оператор линейного преобразования

\lg – десятичный логарифм

\ln – натуральный логарифм

$LPCCx[m]$ – коэффициенты кепстра линейного предсказания

M

Mx – средневывпрямленное значение сигнала (магнитуда)

$Mx[k]$ – средневывпрямленное значение сигнала на k -м кадре

$$(Mx[k] = \frac{1}{N} \sum_{i=1}^{i=N} x[i])$$

$Mx[i]$ – средневывпрямленное значение сигнала для дискретного момента времени i

$MFCScx[m]$ – мел-спектральные коэффициенты кепстра

$MTF(Fm)$ – модуляционная передаточная функция

m – индекс

N

N – размерность (количество бинов) дискретного спектра

n – индекс для дискретных частот (бинов) спектра

P

p – параметр

$P_{xx}(f)$ – спектральная плотность мощности (спектр мощности)

$P_{nn}(f)$ – обычно спектр мощности шума

$P_{ss}(f)$ – обычно спектр мощности речевого сигнала

$P_{xx}[m]$ – дискретная спектральная плотность мощности

$P_{xy}(f)$ – кросс-спектр (взаимная спектральная плотность мощности)

$P_x[i]$ – мощность сигнала для дискретного момента времени i

$P_x[k]$ – мощность сигнала на k -м кадре

P_s – мощность речевого сигнала

P_n – мощность шума

R_{CCxy} – коэффициент корреляции Пирсона

Q

q – индекс

R

$r_{xy}[m]$ – дискретная кросс-корреляционная функция двух процессов (нормированная кросс-ковариация), соответствующая временному сдвигу m

$R_{xx}[m]$ – дискретная автоковариационная (автокорреляционная) функция, соответствующая временному сдвигу m

$r_{xx}[m]$ – дискретная нормализованная (нормированная) КФ $R_{xx}[m]/R_{xx}[0]$

R_{xy} – коэффициент корреляции между векторами

$\text{Re}\{ \}$ действительная часть числа, заключенного в скобки

S

$s[i]$ – временной ряд (обычно речевой сигнал)

SNR – отношение сигнал-шум

SNR_{seg} – сегментное отношение сигнал-шум

$S_{xx}(f)$ – спектральная плотность энергии

$S_{xx}[n]$ – дискретная спектральная плотность энергии

T

t – непрерывное время

T – шаг (интервал, период) дискретизации (выборки) по времени ($= 1/F_s$)

THR – порог

T_0 – период основного тона, период процесса

T_a – постоянная времени сглаживания

T_r, T_{60} – время реверберации

V

$v[i]$ – возбуждающий случайный процесс (в т.ч. белый шум)

W

$W(f)$ – АЧХ фильтра (обычно фильтра Винера)

$w[i]$ – коэффициенты фильтра Винера

X

$x[i]$ – временной ряд, дискретный сигнал (обычно вход системы)

$x(t)$ – непрерывный сигнал (функция времени)

$x(iT)$ – дискретные временные отсчеты сигнала $x(t)$

$\mathbf{X}[k] = [x[kL], x[kL-1], \dots, x[kL-N+1]]^T$ – вектор отсчетов сигнала на кадре

$X[n, k], X_n[k]$ – коэффициенты дискретного комплексного спектра сигнала на кадре k

$X_r[n, k], X_i[n, k]$ – реальная и мнимая части дискретного комплексного спектра на кадре k

$|X_n[k]|, |X[n, k]|$ – модуль дискретного комплексного спектра на кадре k

$\langle x \rangle$ – оценка средней величины сигнала $x[i]$

Y

$y[i]$ – временной ряд, дискретный сигнал (обычно выход системы)

Z

z – оператор сдвига вперед: $zx[i] = x[i+1]$

z^{-1} – оператор сдвига назад: $z^{-1}x[i] = x[i-1]$

СИМВОЛЫ ГРЕЧЕСКОГО АЛФАВИТА

- α – коэффициент забывания /усреднения (*forgetting/averaging factor*) ФЭС
- β – постоянная сглаживания (*smoothing factor*) ФЭС
- λ – длина звуковой волны
- $\delta[x]$ – дельта-функция Дирака
- $\delta[i]$ – единичная цифровая дельта-функция, функция Кронекера
- Δ_x – смещение оценки величины x
- Δ – оператор разности назад первого порядка: $\Delta x[i] = x[i] - x[i-1] = (1 - z^{-1}) x[i]$
- ΔF (BW) – ширина полосы пропускания фильтра
- ΔT – временной интервал
- ε_t – возбуждающий случайный процесс, в т.ч. белый шум
- $\Gamma_{xy}(f)$, $\Gamma_{xy}(n)$ – комплексная функция когерентности
- μ_x – среднее значение случайной величины x
- ρ_{xy} – корреляция, нормированная ковариация: $\sigma_{xy} / \sigma_x \sigma_y$
- ρ_t – нормированная автокорреляционная функция
- σ – стандартное отклонение
- τ – временная задержка, запаздывание
- π – 3,14159265...
- φ – фазовый угол
- $\Phi_x(f)$, $\Phi_x[n]$ – фазовый спектр непрерывный и дискретный
- $\Phi_{xy}(f)$, $\Phi_{xy}[n]$ – фазовый кросс-спектр непрерывный и дискретный
- Ψ_x – среднее квадратичное значение случайной величины x
- ω – угловая частота, рад/с
- Ω – нормализованная угловая частота, рад/отсчет
- $(\)^*$ – символ комплексного сопряжения
- $(*)$ – символ свертки
- $\langle \rangle$ – обозначение операции усреднения по времени
- $[]^T$ – обозначение операции транспонирования вектора
- \prod – знак произведения
- \sum – знак суммы

АНГЛО-РУССКИЙ СЛОВАРЬ ТЕРМИНОВ

Термин	Определение
A	
Access Control and Authentication	Контроль доступа
Adaptive Filter (AF)	Адаптивный фильтр
Adaptive line enhancer (ALE)	Адаптивный линейный компенсатор
Adaptive null-forming (ANF)	Адаптивный формирователь нуля
Adaptive Noise Canceller (ANC)	Адаптивный компенсатор шума
Amplitude-Frequency Response (AFR)	Амплитудно-частотная характеристика фильтра (АЧХ)
Analogue-to-Digital Converter (ADC)	Аналого-цифровой преобразователь (АЦП)
Audio-to-Audio Alignment	Приведение среднего спектра сигнала к среднему спектру образцового сигнала
Autocorrelation (ACF)	Автокорреляционная функция (АКФ)
Automatic Speech Recognition (ASR)	Автоматическое распознавание речи
B	
Back-end	Прикладная (вторичная) обработка РС
Band	Диапазон частот, полоса
Bandwidth (BW)	Ширина полосы пропускания
Band-pass filter (BPF)	Пропускающий полосовой фильтр
Band-stop filter (BSF)	Заграждающий полосовой фильтр
Bit rate	Цифровой поток
bps = bit/s	Мера цифровой потока – число бит, передаваемых в секунду
Blind equalization	Слепое выравнивание канала
C	
Clipping	Клиппирование
Channel deconvolution, equalization, compensation	Компенсация, нормализация канала
Cepstral Distance (CD)	Кепстральная дистанция
Cepstral Mean Normalization (CMN)	Нормализация канала на основе вычитания среднего кепстра
Cepstral Mean Subtraction (CMS)	Вычитание среднего кепстра
Comb-filter	Фильтр-гребенка
Computer Acoustic Scene Analysis (CASA)	Компьютерный анализ акустических сцен

Cross correlation function (CCF)	Кросс-ковариационная функция двух процессов (ККФ)
Cross spectrum	Кросс-спектр
Cross power spectrum density (CSPD).	Плотность кросс-спектра (ПКС)
D	
Data compression	Сжатие данных
Dropouts	Пропадание сигнала
Digital-to-Analogue Converter (DAC)	Цифро-аналоговый преобразователь (ЦАП)
Direct current (DC), DC offset	Постоянный ток, постоянная составляющая (среднее значение) сигнала
DC cut	Удаление медленно меняющейся постоянной составляющей сигнала
Discrete Cosine Transformation (DCT)	Дискретное косинусное преобразование
Digital Signal Processing (DSP)	Цифровая обработка сигналов (ЦОС)
Direction of Arrival (DOA)	Направление прихода (сигнала)
Decibel (dB)	Децибел (дБ)
Direct to Reverberation Ratio (DRR)	Отношение интенсивностей прямого звука и реверберации
Discrete Fourier Transform (DFT)	Дискретное преобразование Фурье (ДПФ)
Distortion	Искажение
Deep Neural Network (DNN)	Глубокая нейронная сеть
Dual-Tone Multi-Frequency (DTMF)	Двухтональный аналоговый сигнал для кодирования знаков
E	
Equaliser/Equalizer (EQ)	Эквалайзер
Early Decay Time (EDT)	Раннее время затухания реверберации
Echo Return Loss Enhancement (ERLE)	Мера эффективности подавления эха компенсатором эха, дБ
F	
Frequency-Domain Adaptive Filtering (FDAF)	Частотный адаптивный фильтр
Finite Impulse Response (FIR)	Конечная импульсная характеристика (КИХ)
Fast Fourier Transform (FFT)	Быстрое преобразование Фурье (БПФ)
Fourier Transform (FT)	Преобразование Фурье (ПФ)
Framing	Разбиение сигнала на кадры
Front-End	Первичная обработка РС
Fundamental Frequency	Частота основного тона

G

Graphic EQ	Графический эквалайзер
Gain	Коэффициент передачи (коэффициент усиления уровня звука или электрического сигнала)
Gaussian Mixture Model (GMM)	Модель гауссовых смесей
Generalized Spectral Subtraction (GSS)	Обобщенное спектральное вычитание
Graphics Processing Unit (GPU),	Графические процессоры

H

Harmonic cancellation	Подавление гармоник
Harmonic enhancement	Выделение гармоник
Hearing Aids	Коррекция нарушений слуха и речи
Hertz (Hz)	Герц
High-pass filter (HPF)	Фильтр высоких частот (ФВЧ)
Harmonic Tone Suppression (HTS)	Подавление гармоник тона

I

Independent Component Analysis(ICA)	Анализ независимых компонент (метод слепого разделения сигналов независимых источников)
Impulse response (IR)	Импульсный отклик, характеристика
Instant Tone Suppression (ITS)	Подавление тональных импульсов
Integrated Circuit (IC)	Технология интегральных схем
Interactive Voice Response (IVR) systems	Системы голосового самообслуживания
Infinite Impulse Response (IIR)	Бесконечная импульсная характеристика (БИХ)
Inverse Discrete Fourier Transform (IDFT)	Дискретное обратное преобразование Фурье
Inverse Fast Fourier Transform (IFFT)	Обратное быстрое преобразование Фурье (ОБПФ)
Inverse Discrete Cosine Transformation (IDCT)	Дискретное обратное косинусное преобразование

K

kHz	Килогерц (кГц)
Kurtosis	Экцесс

L

Language Recognition	Распознавания языка
Least Mean Square (LMS)	Наименьшее среднеквадратическое значение

Least-squares error (LSE)	Ошибка наименьших квадратов
Linear Prediction (LP)	Линейное предсказание
Linear prediction coefficients (LPC)	Коэффициенты линейного предсказания (КЛП)
Linear Prediction Cepstral Coefficients (LPCC)	Коэффициенты кепстра спектра линейного предсказания
Linear Time-Invariant (LTI) Systems	Инвариантные по времени линейные системы
Linear Spectrum Frequency (LSF) domain	Пространство линейного спектра частот
Low- pass filter (LPF)	Фильтр низких частот (ФНЧ)
M	
Magnitude-squared Coherence function (MSC)	Квадрат модуля когерентности (КМК)
Magnitude Spectral Subtraction (MSS)	Амплитудное спектральное вычитание
Mean	Среднее
Mean Square	Среднеквадратическое значение
Mel-Frequency Cepstral Coefficients (MFCC)	Коэффициенты кепстра мел-частотного спектра
Mean Opinion Score MOS	Средняя экспертная оценка
Mel-Frequency Cepstral Coefficients (MFCC)	Коэффициенты кепстра мел-частотного спектра
Multiple-Input, Multiple-Output (MIMO) system	Система с одним входом и одним выходом
Multiple-Input, Single-Output MISO system	Система со многими входами и многими выходами
Multi-notch filter (MNF)	Многополосный целевой фильтр
Magnitude Sum Function (MSF)	Средний модуль, среднев्यпрямленное значение сигнала (магнитуда)
Mean Square Error (MSE)	Среднеквадратическая (средняя квадратическая ошибка (СКО)
Minimum Mean Square Error (MMSE)	Критерий минимума СКО
MMSE Short-term Spectral Analysis (MMSE-STSA)	Оценка кратковременной амплитуды спектра по критерию минимума СКО
MMSE Logarithmic Spectral Analysis (MMSE-LSA)	Оценка логарифмической амплитуды спектра по критерию минимума СКО
Modulation Transfer Function (MTF)	модуляционная передаточная функция
Moving average estimator (filter)	Фильтр скользящего среднего, в том числе фильтр экспоненциального сглаживания
N	
Natural Language Processing (NLP)	Понимание текстов и естественного языка
Noise Cancellation	Компенсация помехи
Noise Reduction/Suppression (NR/NS)	Подавление шума

Normalized Least Mean Square (NLMS)	Нормализованный минимум СКО (алгоритм минимума СКО с нормализацией)
Non-negative Matrix Factorization (NMF)	Неотрицательное матричное разложение (НМР)
Nonlinear Spectral Subtraction (NSS)	Нелинейное спектральное вычитание
Notch filter	Выкалывающий фильтр (подавляющий) фильтр-пробка
P	
Pulse Code Modulation (PCM)	Импульсно-кодовая модуляция (цифровое представление звука)
Pitch	Основной тон гармонического сигнала
Pearson's Correlation Coefficient (PCC)	Коэффициент корреляции Пирсона
Power Spectral Density (PSD)	Спектральная плотность мощности (СПМ)
Power Spectral Subtraction (PSS)	Метод спектрального вычитания
R	
Root Mean Square (RMS)	Квадратный корень из среднеквадратического значения
Room Impulse Response (RIR)	Импульсный отклик помещения
Room Transfer Function (RTF)	Передаточная функция помещения
Real Time (RT)	Масштаб реального времени
Reverberation Time (TR)	Время реверберации
S	
Speech enhancement	Выделение речи
Speech coding, data compression	Сжатие и кодирование речевых сигналов
Standard Deviation (Std)	Стандартное отклонение, σ
Single-Input, Multiple-Output (SIMO) system	Система с одним входом и многими выходами
Single-Input, Single-Output (SISO) system	Система с одним входом и одним выходом
Short-Time	Кратковременная оценка
Short-Time Energy (STE)	Кратковременная энергия
Signal to Interference Ratio (SIR)	Отношение сигнал-помеха, интерференция
Signal-to-Noise Ratio (SNR)	Отношение сигнал-шум (ОСШ)
Signal-to-Noise Ratio Enhancement (SNRE)	Улучшение отношения сигнал-шум
Slot-, one-formant filter	Узкополосный полосовой фильтр
Spectral-domain	Спектральное пространство
Spectral Subtraction Method (SSM)	Метод спектрального вычитания
Spectral Power Density (SPD)	Спектральная плотность мощности (СПМ)

Spectral Flux (SF)	Спектральный поток
Speech/Text Analytics	Речевая, текстовая и акустическая аналитика
Speech coding	Сжатие и кодирование речевых сигналов
Speech enhancement	Компенсация искажений РС
Short-Time Fourier Transform (STFT)	Кратковременное преобразование Фурье
Short-Time Spectral Amplitude (STSA)	Кратковременная амплитуда спектра
Snapshots	Измерения (замеры)
Standard deviation (Std)	Стандартное отклонение
Speech Transmission Index (STI)	Индекс передачи речи
T	
Teager Energy Operator (TEO)	Энергия Тигера (Тигера-Кайзера)
Text-to-Speech Synthesis (TSS)	Автоматический синтез речи по тексту
Time-Domain Methods (TD)	Временные методы
Time-Domain Adaptive Filtering (TDAF)	Временной адаптивный фильтр
Time waveform	Осциллограмма
Tonal Noise Suppression (TNS)	Подавление тонального шума
Tone Spectral Suppression (TSS)	Спектральное подавление тонального шума
W	
Wiener Filter (WF)	Фильтр Винера
Windowing	Оконное взвешивание
Weighted Prediction Error (WPE)	Взвешенная ошибка предсказания
White Noise (WN)	Белый шум (БШ)
V	
Voice Activity Detector (VAD)	Детектор (обнаружитель) речевой активности
Variance (Var)	Дисперсия
Voice Recognition/Biometric Authentication	Верификация и идентификация дикторов, голосовая биометрия
Voice Assistants	Голосовая помощь
Voice Cloning	Клонирование голоса
Voice over Internet Protocol (VoIP)	Передача голоса по Интернет-протоколу
Z	
Zero-Crossing Rate (ZCR)	Частота пересечения нуля

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ.....	3
1 ВВЕДЕНИЕ В ПРЕДМЕТ.....	5
1.1 Краткая история речевых технологий.....	5
1.2 Предметная область цифровой обработки речевых сигналов.....	7
1.3 Профессиональные сообщества и информационные ресурсы ЦОРС.....	11
1.4 Работа с пособием.....	12
Вопросы и упражнения.....	14
2 ЗВУК И ЕГО ЦИФРОВОЕ ПРЕДСТАВЛЕНИЕ.....	15
2.1 Звук.....	15
2.2 Аналого-цифровое преобразование звуковых сигналов.....	16
2.3 Дискретизация во времени.....	16
2.4 Понятие частоты.....	17
2.5 Эффект наложения спектров.....	18
2.6 Квантование сигналов.....	18
2.7 Цифровой звук и его запись.....	20
2.8 Осциллограмма.....	21
Вопросы и упражнения.....	24
3 СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ РЕЧЕВЫХ СИГНАЛОВ.....	25
3.1 Случайная величина и ее характеристики.....	25
3.2 Случайный процесс и его характеристики.....	27
3.3 Выборочные оценки параметров и их погрешности.....	27
3.4 Выборочные оценки основных параметров случайного процесса.....	29
3.5 Выборочная оценка функции распределения.....	30
3.6 Анализ статистической взаимосвязи случайных величин.....	32
Вопросы и упражнения.....	34
4 ВРЕМЕННОЙ И ЧАСТОТНЫЙ АНАЛИЗ СИГНАЛОВ.....	35
4.1 Автоковариационная функция процесса.....	35
4.2 Ковариация и ее оценка.....	36
4.3 Спектральный анализ сигналов.....	38
4.4 Спектр мощности.....	40
4.5 Кросс-спектр.....	42
4.6 Соотношение между ковариационной функцией и кросс-спектром.....	43
4.7 Функция когерентности и ее оценка.....	43
Вопросы и упражнения.....	46
5 СИСТЕМЫ И ПРЕОБРАЗОВАНИЯ СИГНАЛОВ.....	47
5.1 Общие сведения о системах.....	47

5.2	Линейные системы	48
5.3	Линейные дискретные системы	49
5.4	Дискретная свёртка	50
5.5	Система с конечной импульсной характеристикой	51
5.6	Система с бесконечной импульсной характеристикой.....	52
5.7	Спектральное описание линейных систем с одним входом и выходом	53
5.8	Оценка характеристик линейных систем	54
	Вопросы и упражнения	56
6	ФИЛЬТРАЦИЯ ДИСКРЕТНЫХ СИГНАЛОВ.....	57
6.1	Рекурсивные фильтры первого порядка.....	57
6.2	Дифференцирующие фильтры	63
6.3	Сглаживающие КИХ-фильтры.....	64
6.4	Полосовые фильтры и фильтры-гребенки	66
6.5	Фильтры в частотной области	71
	Вопросы и упражнения	74
	ЗАКЛЮЧЕНИЕ.....	75
	ПРИЛОЖЕНИЕ А ПРОФЕССИОНАЛЬНЫЕ СООБЩЕСТВА В ОБЛАСТИ РЕЧЕВЫХ ТЕХНОЛОГИЙ	76
	ПРИЛОЖЕНИЕ Б СИСТЕМЫ СТАНДАРТИЗАЦИИ	77
	ПРИЛОЖЕНИЕ В ОСНОВНЫЕ ЖУРНАЛЫ ПО РЕЧЕВЫМ ТЕХНОЛОГИЯМ	78
	ПРИЛОЖЕНИЕ Г ОСНОВНЫЕ КОНФЕРЕНЦИИ ПО РЕЧЕВЫМ ТЕХНОЛОГИЯМ	80
	СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	81
	ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	86
	СПИСОК СИМВОЛОВ И ОБОЗНАЧЕНИЙ	88
	СИМВОЛЫ ГРЕЧЕСКОГО АЛФАВИТА	93
	АНГЛО-РУССКИЙ СЛОВАРЬ ТЕРМИНОВ	94

Столбов Михаил Борисович

Основы анализа и обработки речевых сигналов

Учебное пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49, литер А