

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Факультет телекоммуникаций

Кафедра сетей и устройств телекоммуникаций

А. И. Митюхин

ЦИФРОВАЯ ОБРАБОТКА РЕЧИ И АНАЛИЗ ИЗОБРАЖЕНИЙ

*Рекомендовано УМО по образованию в области информатики
и радиоэлектроники в качестве учебно-методического пособия
для направлений специальности 1-45 01 01-02
«Инфокоммуникационные технологии (сети инфокоммуникаций)»,
1-45 01 01-06 «Инфокоммуникационные технологии
(лазерные информационно-измерительные системы)»,
1-45 01 02-01 «Инфокоммуникационные системы
(стандартизация, сертификация и контроль параметров)»,
специальности 1-98 01 02 «Защита информации в телекоммуникациях»*

Минск БГУИР 2016

УДК [621.391+004.932](076)
ББК 32.811я73+32.973.26-018.2я73
М67

Рецензенты:

кафедра высшей математики учреждения образования
«Военная академия Республики Беларусь»
(протокол №124 от 27.04.2015);

доцент кафедры телекоммуникационных систем учреждения
образования «Белорусская государственная академия связи»,
кандидат технических наук, доцент К. И. Пирогов

Митюхин, А. И.

М67 Цифровая обработка речи и анализ изображений : учеб.-метод.
пособие / А. И. Митюхин. – Минск : БГУИР, 2016. – 72 с. : ил.
ISBN 978-985-543-190-0.

Рассмотрены элементы цифровой обработки речи и анализа изображений, понятия теории анализа и синтеза речевых сигналов. Представлено одно из направлений цифровой обработки изображений – анализ с целью сжатия мультиспектральных изображений. Методы и алгоритмы цифровой обработки сигналов изложены на основе практического использования математического аппарата. Изложение материала сопровождается примерами решения задач, иллюстрациями, упражнениями.

УДК [621.391+004.932](076)
ББК 32.811я73+32.973.26-018.2я73

ISBN 978-985-543-190-0

© Митюхин А. И., 2016
© УО «Белорусский государственный
университет информатики
и радиоэлектроники», 2016

Содержание

Введение	4
1. Обработка речевых сигналов	6
1.1. Модель речевого сигнала	6
1.2. Последовательность операций цифровой обработки речи	8
1.3. Способы отсчета значений сигнала.....	10
1.4. Теорема дискретизации	14
2. Синтез речи	16
2.1. Линейные системы обработки речи	17
2.2. Представление систем и сигналов в частотной области	27
3. Анализ изображений	51
3.1. Статистические свойства изображений	51
3.2. Сжатие изображений на основе статистического анализа	61
3.3. Сжатие изображений с применением дискретных ортогональных преобразований.....	68
Литература	71

Введение

Обработка речевых сигналов

Методы обработки речевых сигналов имеют уже большую историю и в настоящее время применяются к широкому кругу задач – от телекоммуникаций до техники распознавания речи, когда можно управлять устройствами с помощью голоса и диктовать текст компьютеру, а не набирать текст, используя мышь и клавиатуру. Распознавание речи становится доступным практически для всех широко используемых языков. До 70-х годов XX века методы обработки речевых сигналов основывались на непрерывных принципах и алгоритмах, системы обработки строились на аналоговой аппаратуре. Величина сравнительно узкой полосы частот речевых сигналов хорошо согласовывалась с быстродействием существующей в то время цифровой процессорной техники. С появлением цифровой элементной базы со средней степенью интеграции и новых эффективных алгоритмов началось быстрое внедрение теоретических разработок в различные приложения.

Обработка речевых сигналов применяется для решения следующих основных задач:

- анализ речи для систем автоматического распознавания речевых сигналов, на основе которого работает классификатор образов для идентификации человека по голосу;
- синтез речи, когда входные текстовые, цифровые данные преобразуются в речевой сигнал для многих мультимедийных технологий.

Сначала выполняется анализ речи, а затем реализуется синтез речевого сигнала. Преобразование непрерывной формы речи в цифровую приводит к расширению спектра речевого сигнала. Выполняя спектральный анализ речи и эффективное кодирование спектральных коэффициентов, а также синтез (восстановление) исходной формы в приемнике, можно существенно уменьшить (сжать) полосу частот канала передачи информации.

Еще одной задачей обработки речевых сигналов является устранение искажений сигналов из-за воздействия шумов.

Для решения всех этих задач используются спектральные преобразования. В частности, комплексное преобразование Фурье на коротких интервалах времени. В целях эффективного кодирования речи (сжатия) применяется действительное дискретное косинусное преобразование. Иногда целесообразным является использование других спектральных преобразований, например, Хартли, Уолша – Адамара, Хаара.

Обработка изображений

Цифровая обработка изображений – это одна из областей цифровой обработки сигналов. При обработке изображений различают компьютеризированные процессы низкого, среднего и высокого уровня.

Начальный процесс включает в себя предварительную обработку изображений с целью уменьшения шума на изображении, повышения контраста, улучшения резкости и пр. На этапе предварительной обработки применяются надлежащие процедуры фильтрации, для того чтобы отличить объекты интереса от всех других и от фона. Для этого уровня характерны процессы с изображениями на входе и на выходе.

Для среднего уровня свойственны процессы с изображениями только на входе, на выход поступают признаки и атрибуты, извлекаемые из этих изображений, например, границы областей, линии контуров и пр. Обработка на среднем уровне охватывает следующие процессы анализа и извлечения признаков из изображений:

- сегментацию – разделение изображения на области или выделение на нем объектов;
- описание объектов и сжатие изображений;
- морфологическую обработку – анализ и модификация формы сегментированных объектов;
- распознавание отдельных объектов на изображении.

Для высокого уровня характерны процессы с изображением, связанные с машинным зрением, анализом распознанных объектов.

Пространственные данные могут иметь высокую размерность. Тогда на этапе получения признаков распознавания вычислительная и временная сложность обработки еще более возрастают. С целью упрощения обработки и процесса классификации образа объекта следует понижать размерность входа классификатора. Приемами снижения вычислительной сложности являются, например, устранение пространственной избыточности и уменьшение количества анализируемых признаков. Понижать размерность следует таким образом, чтобы сопутствующее этому увеличение ошибки распознавания было относительно невелико. Среди всех линейных преобразований данных оптимальным по отношению к критерию наименьшего значения среднеквадратичной ошибки при неполной размерности (неточном задании координат) является декоррелирующее преобразование Хотеллинга.

1. Обработка речевых сигналов

1.1. Модель речевого сигнала

Речь возникает из-за возбуждения акустической трубы – речевого тракта, который ограничен с одной стороны губами, а с другой – голосовой щелью.

Речь формируется одним из трех способов.

Первый – образование вокализованных звуков (звонких), когда вибрации голосовых связок в полостях речевого тракта производят квазипериодические колебания воздуха. Источник возбуждения звонких звуков – это квазипериодическая последовательность импульсов $e(t)$. Гласные буквы *и*, *э* являются примером вокализованных звуков.

Второй способ – образование фрикативных (глухих, невокализованных) звуков, когда воздух проталкивается сквозь сужение голосового тракта, т. к. воздушный поток блокируется языком и губами. Получается турбулентный, квазислучайный поток в области сужения, который является источником шума, возбуждающего речевой тракт. В образовании этих звуков голосовые связки не участвуют. Согласные буквы *ш*, *с*, *ф*, *з* являются примером невокализованных звуков.

Третий способ – образование фрикативных взрывных звуков, когда после закрытия речевого тракта следует резкий выдох (кратковременный всплеск). Примером взрывных звуков являются буквы *п*, *т*, *к*. Свойства взрывных звуков отчасти определяются предшествующими звуками.

Типичный речевой сигнал $x(t)$ напоминает либо квазипериодическое колебание с изменяющейся основной частотой в сравнительно широком диапазоне, либо шум. Речь – это нестационарный процесс. Напомним, свойства такого процесса меняются с течением времени. Различают нестационарные процессы с переменным во времени значением (рис. 1.1, а) и с переменным во времени средним квадратом (рис. 1.1, б).

Поскольку речевой тракт представляет собой полость, он характеризуется своими собственными резонансными частотами, называемыми формантами. Их величина определяется размерами и формой его полостей. Форманты соответствуют акустическим резонансам комплексной частотной характеристики (КЧХ) тракта. Только несколько основных формант определяют восприятие речи человеком. При изменении положения артикуляторов (губ, языка, челюсти, неба) во время произнесения речи форма полости речевого тракта изменяется. Из-за этого получается все разнообразие звуков. Спектры гласных содержат отчетливые резонансы. При нормальной речи форма акустической передающей системы под воздействием языка и губ изменяется медленно. На коротких интервалах времени (от 2 до 40 мс) характеристики речевого сигнала остаются постоянными.

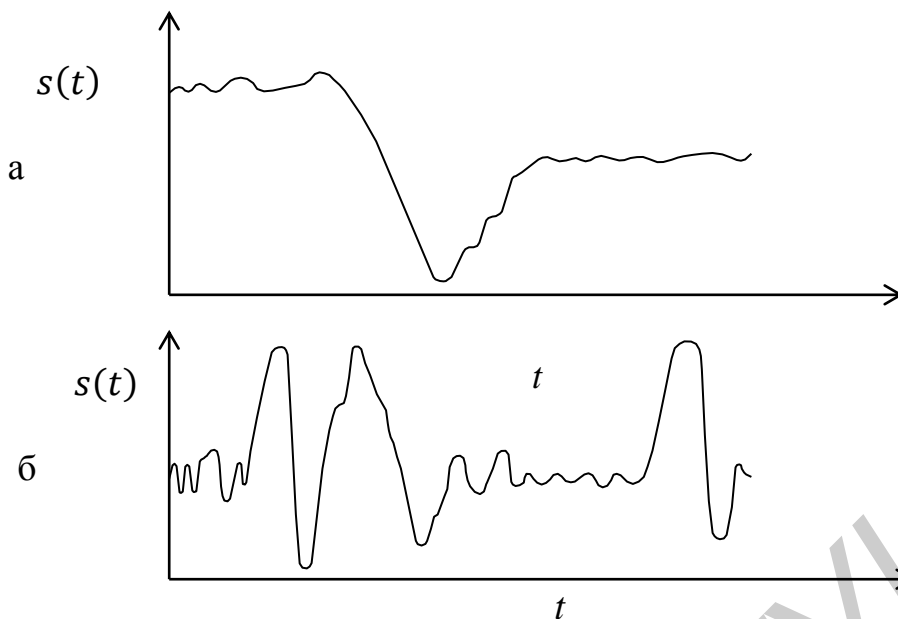


Рис. 1.1. Форма речевого сигнала

Можно принять, что речевой тракт соответствует квазистационарной линейной модели, в которой речевой сигнал является реакцией линейной системы с медленно изменяющимися (переменными) параметрами на шумовой возбуждающий сигнал. На коротких интервалах времени характеристика речевого сигнала $x(t)$ определяется сверткой возбуждающего сигнала $e(t)$ с импульсной характеристикой $h(t)$. В этом случае модели формирования речевого сигнала $x(t)$ соответствует фильтр с медленно изменяющимися параметрами (рис. 1.2).

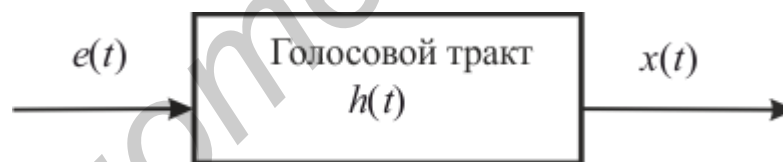


Рис. 1.2. Модель формирования речи

Представленная модель имеет и частотное описание. Преобразование Фурье речевого сигнала $x(t)$ равно произведению Фурье-образа возбуждающего сигнала $e(t)$ и Фурье-образа импульсной характеристики $h(t)$ речевого тракта:

$$X(e^{j\omega}) = E(e^{j\omega})H(e^{j\omega}).$$

В данной модели при входном периодическом сигнале на выходе формируется также периодический сигнал. На рис. 1.3 приведен амплитудный спектр $|X(e^{j\omega})|$ некоторого речевого сигнала $x(t)$.

Частоты всех гармоник спектра $X(e^{j\omega})$ кратны основной частоте (первой гармонике) возбуждающего сигнала $e(t)$:

$$\omega_0 = 2\pi f_0 = \frac{2\pi}{T},$$

где T – период возбуждающего сигнала.

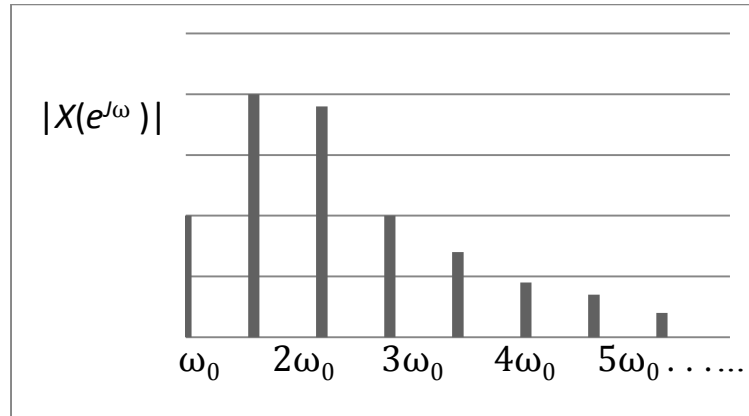


Рис. 1.3. Фурье-образ речевого сигнала

Основную гармонику ω_0 (с низшей частотой) называют основным тоном. Высота звука определяется значением гармоники ω_0 . Верхняя частота речевого сигнала может достигать 15 кГц и выше. Другие гармоники (обертоны) создают тембр звучания речи. По обертонам различаются голоса с одинаковым основным тоном. При изменении периода T возбуждающего сигнала $e(t)$ меняется частотное расстояние между гармониками спектра. Форма огибающей спектра $|X(e^{j\omega})|$ определяется формой воздушного импульса, выходящего из гортани. Спектр $X(e^{j\omega})$ также будет периодическим с основной частотой ω_0 . При изменении T частотная характеристика $|H(e^{j\omega})|$ речевого тракта является сравнительно гладкой функцией частоты. Вид огибающей спектра $|X(e^{j\omega})|$ определяется формой частотной характеристики $|H(e^{j\omega})|$ речевого тракта. Звукам, изменяющимся во времени, будут соответствовать изменяющиеся спектры $|X(e^{j\omega})|$.

Рассмотренная модель формирования речи используется для решения задач синтеза и анализа речи.

1.2. Последовательность операций цифровой обработки речи

Цифровая обработка речи связана с представлением непрерывного сигнала в виде последовательности чисел. Это означает, что речевой сигнал преобразуется в исходную последовательность чисел, которая по заданному алгоритму преобразуется процессором в другую последовательность чисел, однозначно соответствующую исходной. Далее из полученной новой последовательности формируется результирующий сигнал (рис. 1.4).

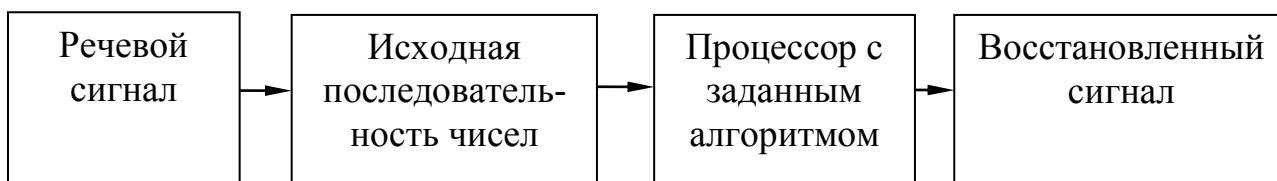


Рис. 1.4. Последовательность операций цифровой обработка речи

Эти преобразования должны производиться по определенным схемам (рис. 1.5).

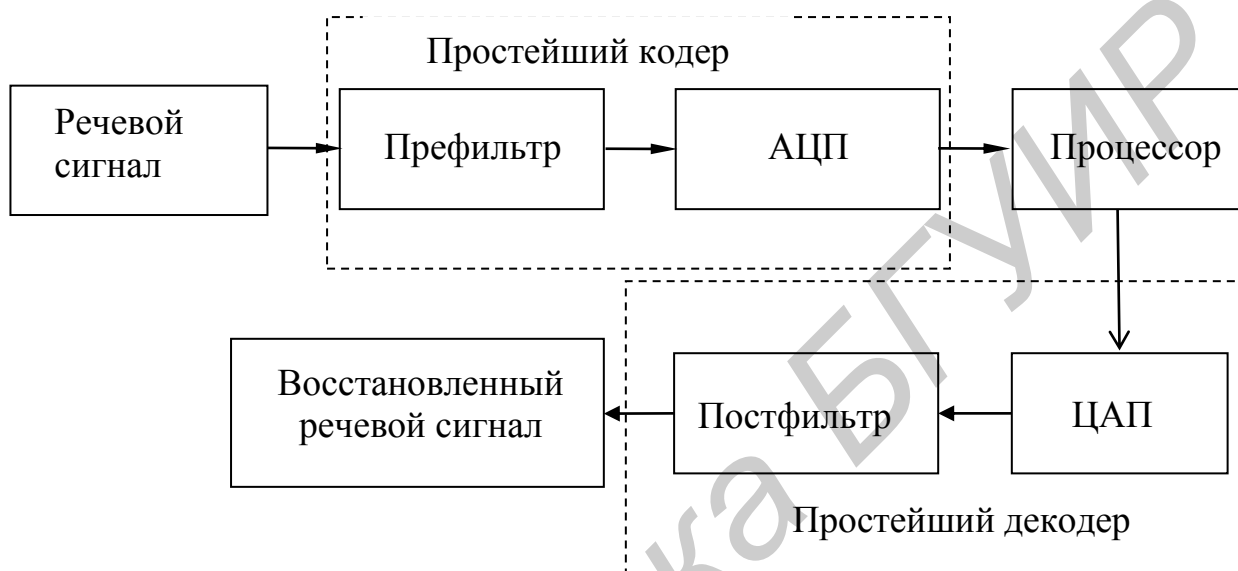


Рис. 1.5. Обобщенная схема цифровой обработка речи

Замечания:

1. Префильтрация обеспечивает удаление определенных верхних спектральных составляющих, лежащих выше некоторого значения. Например, для телефонного сигнала удаляются спектральные составляющие с $f > 3,4$ кГц. Префильтрация осуществляется аналоговым фильтром низких частот (ФНЧ), получившим название антиэлайсингового, поскольку он предотвращает искажения спектра типа наложения (aliasing), которые возникают в спектре цифрового сигнала при недостаточной частоте дискретизации. Это дает основание считать сигнал практически ограниченным по частоте и неподверженным эффекту наложения при частоте дискретизации f_d не менее $2W$, где W определяет ширину полосы частот речевого сигнала. Во временной области эффект наложения означает невозможность точного восстановления аналогового сигнала по его отсчетам. Желательно иметь в полосе сигнала амплитудную частотную характеристику ФНЧ с минимальными пульсациями. Подавление спектральных составляющих с частотой, превышающей половину частоты дискретизации речевого сигнала, должно составлять не менее 40 дБ.

2. Кодер может включать в себя устройства, устраняющие избыточность информационного сигнала для реализации сжатия. В простейшем случае в качестве кодера может быть использован аналого-цифровой преобразователь

(АЦП). Преобразование непрерывного речевого сигнала в последовательность отсчетов, каждый из которых является аппроксимацией соответствующего отсчета непрерывного сигнала, осуществляется АЦП.

3. Цифроаналоговый преобразователь (ЦАП) необходим для получения непрерывного сигнала из дискретного. Выходной сигнал ЦАП содержит нежелательные высокочастотные составляющие. Постфильтрация обеспечивает интерполяцию восстанавливаемого сигнала. В качестве постфильтра можно использовать аналоговый сглаживающий ФНЧ с частотой среза не выше половины частоты дискретизации речевого сигнала.

1.3. Способы отсчета значений сигнала

Поскольку реальные изменения физических процессов протекают во времени, то в качестве математической модели сигнала, представляющего эти процессы, используют функции времени. Независимая переменная может быть как непрерывной, так и дискретной. По способу отсчета значений сигнала во времени различают непрерывные, дискретные и цифровые сигналы.

Непрерывный сигнал описывается непрерывной функцией $x(t)$ и задается на непрерывных промежутках (рис. 1.6). Интервал изменения переменной времени может быть конечным $[0, t)$ или бесконечным $[0, \infty)$.

Дискретным называется сигнал, дискретный во времени и непрерывный по состоянию (рис. 1.7). Он задается решетчатой функцией $x_d(nT)$ дискретной переменной n . Величина $x_d(nT)$ – отсчет сигнала в точке n , указывающей порядковый номер отсчета дискретной последовательности. Для дискретного и цифрового сигналов это счетное множество, включающее N точек интервала $[0, N - 1]$. Интервал T называется периодом дискретизации, а обратная величина $f_d = \frac{1}{T}$ – частотой дискретизации.

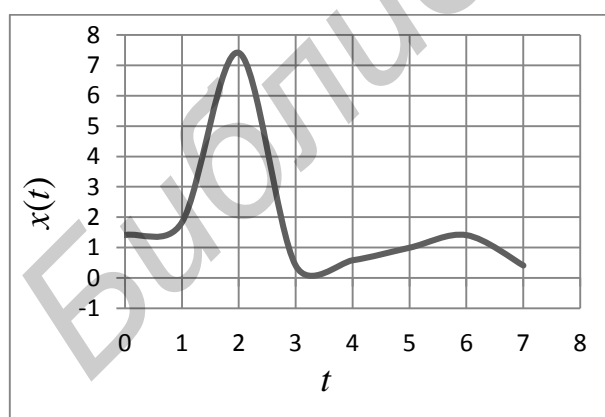


Рис. 1.6. Фрагмент речевого сигнала

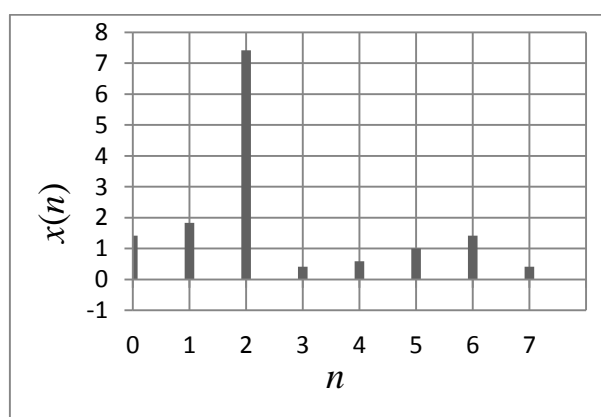


Рис. 1.7. Дискретный сигнал

Характеристиками дискретных сигналов являются энергия и мощность:

$$E = \sum_{n=0}^{N-1} |x(n)|^2, P = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2,$$

которые могут интерпретироваться как энергия и мощность ступенчатых огибающих дискретных сигналов (их континуальных аналогов). Дискретные сигналы ортогональны, если их взаимная энергия удовлетворяет условию

$$E = \sum_{n=0}^{N-1} x_1(n)x_2(n)^* = 0,$$

где знак * обозначает комплексное сопряжение.

Энергия и мощность ортогональных сигналов аддитивны, т. к.

$$E = \sum_{n=0}^{N-1} |x_1(n) + x_2(n)|^2 = \sum_{n=0}^{N-1} |x_1(n)|^2 + \sum_{n=0}^{N-1} |x_2(n)|^2.$$

Цифровая обработка речевых сигналов требует, чтобы дискретные сигналы были квантованы по величине, чтобы их амплитуда принимала только целочисленные значения. В этом случае дискретные сигналы называются цифровыми. Цифровой сигнал дискретен как во времени, так и по состоянию. Цифровой сигнал также описывается решетчатой функцией $x(nT)$, которая принимает конечное число значений на интервале $x'_d \leq x_d \leq x''_d$. Эти значения называются уровнями квантования, а соответствующая функция – квантованным сигналом. Квантование производится с целью получения конечной последовательности чисел. При попадании отсчета сигнала в пределы того или иного шага квантования производится его округление до уровня квантования, соответствующего этому шагу (рис. 1.8).

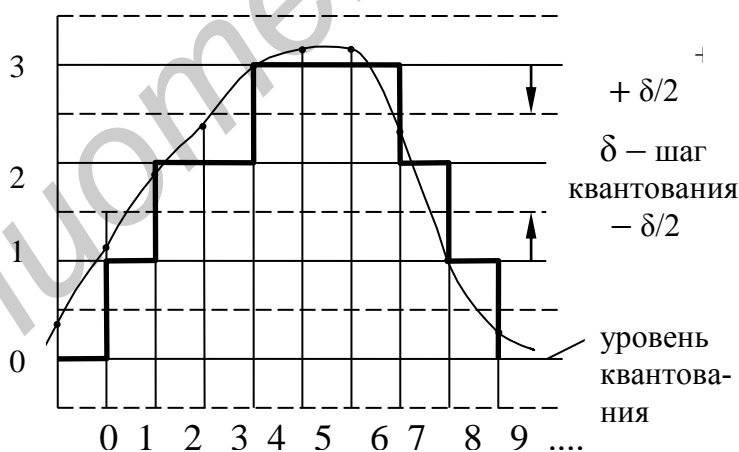


Рис. 1.8. Квантованный сигнал

Цифровой сигнал $x(nT)$ отличается от дискретного $x_d(nT)$ величиной, равной

$$\zeta_Q(nT) = x_d(nT) - x(nT),$$

где $\zeta_Q(nT)$ – ошибка квантования или искажение, возникающие за счет квантования. Максимальная ошибка квантования при использовании округления равна половине шага квантования:

$$\text{Max}|\sigma| = \frac{\delta}{2}.$$

На выходе устройства аналого-цифрового преобразования формируется последовательность двоичных чисел $\{x\}$, соответствующая значениям $x_d(nT)$ (рис. 1.8):

$$x(nT) = \{000\ 001\ 010\ 010\ 011\ 011\ 011\dots\}.$$

В результате получается цифровой образ сигнала, который называется волновой формой (Waveform). Другое название оцифрованного сигнала – сэмпл (Sample). Процесс цифрового преобразования называется сэмплированием. Сэмплы могут храниться в памяти, на жестком диске в файлах различных форматов.

Точность цифрового преобразования определяют динамический диапазон сигнала и соотношение сигнал – шум. Показатель динамического диапазона зависит от длины используемого двоичного кода, или битового разрешения (Bit Resolution) шума квантования δ . С увеличением числа уровней квантования повышается амплитудное разрешение сигнала и точность его воспроизведения (табл. 1.1).

Таблица 1.1

Зависимость динамического диапазона от разрядности квантования

Разрядность квантования (биты)	Уровень качества	Число уровней квантования (U_1)	Динамический диапазон $20 \log_{10} \frac{U_1}{\delta}$ [дБ]
8	Телефония	256	48
16	CD	65 536	96
24	DVD	16 777 216	144
32	HDA (High Definition)	4 294 967 296	192

При изучении цифровых систем удобно пользоваться нормированным временем:

$$\hat{t} = \frac{t}{T} = \frac{nT}{T} = n.$$

Нормированное время \hat{t} есть номер n отсчета дискретного или цифрового сигнала. Для описания дискретного (цифрового) сигнала могут быть использованы равнозначные обозначения:

$$x(nT), N_1 \leq n \leq N_2;$$

$$x(n), N_1 \leq n \leq N_2.$$

Первое обозначение применяется при равномерном расположении отсчетов, второе может применяться при неравномерном расположении отсчетов.

1.3.1. Представление дискретных последовательностей

Пусть задана произвольная временная дискретная последовательность $x(n) = \{x(0), x(1), \dots, x(N-1)\}$. Такая числовая последовательность может быть представлена как сумма взвешенных и задержанных цифровых единичных импульсов. Цифровой единичный импульс (отсчет) (рис. 1.9, а) определяется как

$$u_0(n) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0, \end{cases} \quad (1.1)$$

где $n = 0, \pm 1, \pm 2, \dots$.

На рис. 1.9, б изображен единичный отсчет, задержанный на k отсчетов, который определяется как

$$u_0(n-k) = \begin{cases} 1, & n = k, \\ 0, & n \neq k, \end{cases} \quad (1.2)$$

где $n = 0, \pm 1, \pm 2, \dots$.

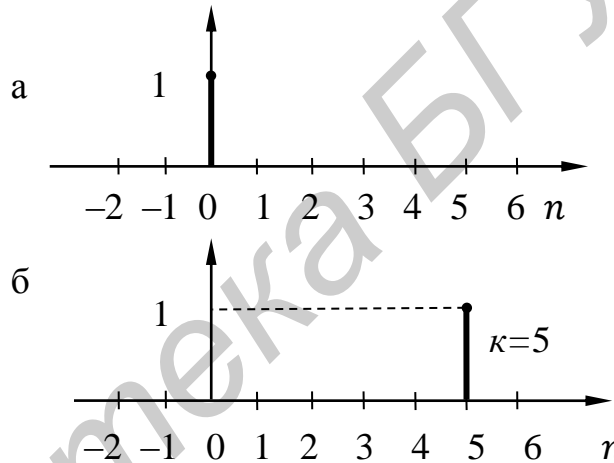


Рис. 1.9. Цифровой единичный импульс (отсчет)

В общем случае любой отсчет произвольной дискретной последовательности записывается в виде

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)u_0(n-k), \quad (1.3)$$

где $n = 0, \pm 1, \pm 2, \dots$.

Пример 1.1. На рис. 1.10 показана дискретная последовательность $x(n)$, которая равна

$$x(n) = \{1,41; 1,83; 7,41; 0,41; 0,58; 1; 1,41; 0,41\}.$$

Записать выражение для определения отсчета с номером $n = 6$.

Решение. Значение отсчета данной последовательности определяется как

$$\begin{aligned} x(6) &= \sum_{k=0}^7 x(k)u_0(6-k) = x(0)u_0(6-0) + x(1)u_0(6-1) + \\ &+ x(2)u_0(6-2) + x(3)u_0(6-3) + x(4)u_0(6-4) + x(5)u_0(6-5) + \\ &+ x(6)u_0(6-6) + x(7)u_0(6-7) = x(6)u_0(0) = x(6) = 1,41. \end{aligned}$$

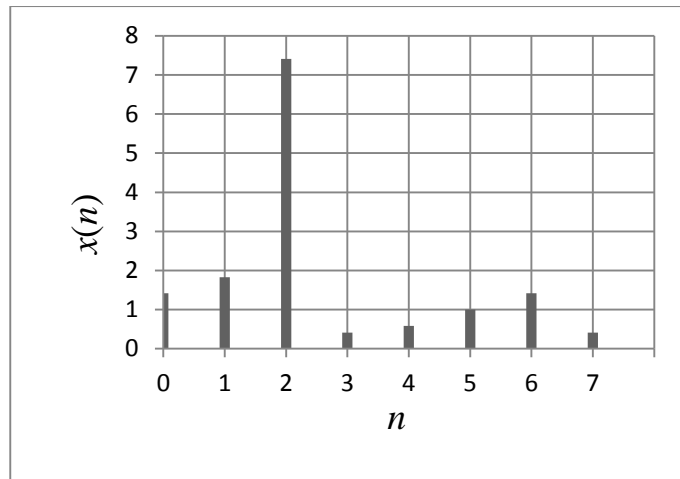


Рис. 1.10. Дискретная последовательность $x(n)$

Упражнение 1.1. Дана дискретная последовательность $x(n) = \{x(0), x(1), x(2), x(3), x(4)\} = \{1; 1,5; -1,2; 4; -4,8\}$. Записать выражение, определяющее значение отсчета последовательности с номером $n = 3$.

1.4. Теорема дискретизации

Дискретизация представляет собой процедуру взятия мгновенных значений непрерывного сигнала $x(t)$.

Положение 1. Задан сигнал, энергия которого полностью содержится внутри полосы частот от 0 до W Гц; сигнал дискретизируется с интервалами T секунд.

Теорема 1.1. Исходный сигнал может быть полностью восстановлен при частоте дискретизации $f_d = \frac{1}{T}$ не менее $2W$. Частоту $W = \frac{f_d}{2}$ называют частотой Найквиста.

В западной научной литературе принято считать, что принцип дискретизации был высказан Найквистом (Н. Nyquist) в 1928 году. По советским источникам теорему дискретизации сформулировал и доказал академик В. А. Котельников в 1931 году. Строгое определение и использование в качестве теоремы дано Шенноном (С. Е. Shannon) в 1948 году.

По теореме дискретизации максимальная частота непрерывного сигнала не должна превышать частоту Найквиста, следовательно, в частотной области все дискретные сигналы целесообразно рассматривать в диапазоне $[0, \frac{f_d}{2}]$ или $[0, \frac{\omega_d}{2}]$.

Положение 2. Задан сигнал длительностью t секунд с частотным спектром (приближенно), ограниченным частотой W Гц.

Теорема 1.2. Сигнал может быть полностью восстановлен (описан) $N = 2Wt$ отсчетами.

Замечание. Определение сигнала как приближенное введено для математической корректности. По самой природе преобразования Фурье никакая функция не может быть ограничена по частоте. Существует концептуальная

трудность строгого частотного ограничения реального сигнала частотой W . В силу конечной длительности сигнала (рис. 1.11), он имеет бесконечно широкий спектр. В моменты возникновения/окончания сигнала временные интервалы Δt его формирования очень малы и, следовательно, ΔW очень велико.

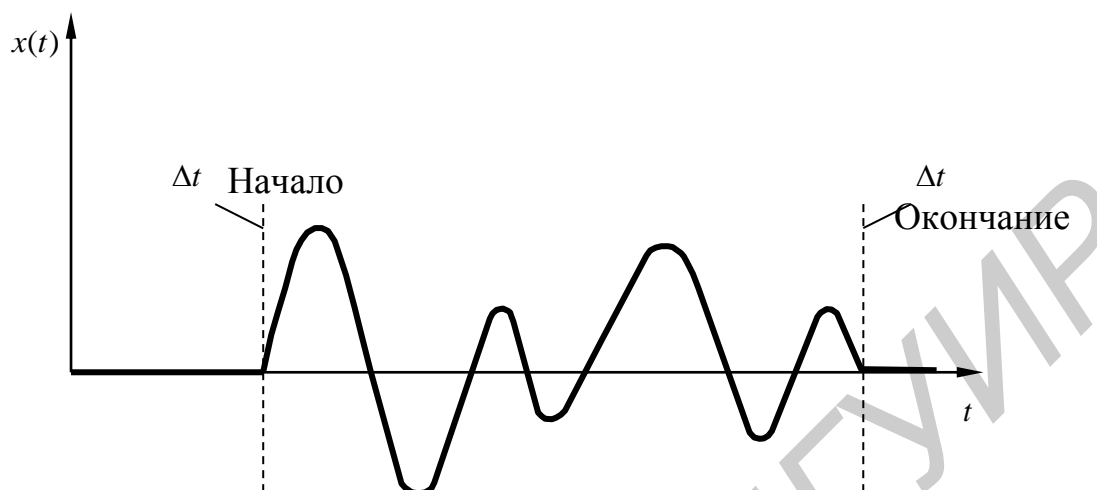


Рис. 1.11. Непрерывный сигнал

Фактически бесконечность спектра является препятствием для преобразования сигнала $x(t)$ в цифровую форму. Если $W \rightarrow \infty$, дискретизация невозможна. Тем не менее в спектре любого конечного сигнала есть такие высшие составляющие, значения которых имеют незначительные амплитуды, и потому ими можно пренебречь без заметного искажения самого сигнала.

Ограничение спектра до частоты $W = \frac{f_d}{2}$ осуществляется антиэлайсинговым фильтром. Для телефонного речевого сигнала полосой $W = 0,3-3,4$ кГц минимальная стандартная частота его дискретизации $f_d = 8$ кГц. Временной интервал дискретизации равен $T = \frac{1}{8 \cdot 10^3} = 125$ мкс. На временной интервал в 2–40 мс, когда статистические характеристики речи остаются практически постоянными, приходится

$$N = 2Wt = f_d t = 8 \cdot 10^3 (2-40) 10^{-3} = 16-320$$

дискретных отсчетов сигнала.

Теоретически точное восстановление сигнала после дискретизации может быть выполнено с помощью идеального ФНЧ с прямоугольной характеристикой и частотой среза $f_s = \frac{f_d}{2}$, однако практически это исключено, так как для создания фильтра потребуется затратить бесконечное время.

Упражнение 1.2. Определить ширину полосы частот телефонного цифрового сигнала с параметрами: $f_d = 8$ кГц, 256 уровней квантования.

Упражнение 1.3. Определить количество отсчетов N дискретного сигнала вокализованного звука «э», если сегмент звучания составляет 25 мс, $f_d = 8$ кГц.

2. Синтез речи

Для цифрового синтеза речи используются два метода: цифровая запись и эмуляция голосового тракта.

При цифровой записи речь оцифровывается, сжимается и записывается в память. При воспроизведении цифровые данные восстанавливаются и преобразовываются в непрерывный речевой сигнал. Этот подход имеет высокое качество речи, но он зависит от набора предварительно записанных слов.

Метод эмуляции имитирует физические процессы формирования речи с образованием посредством цифровой обработки сигналов вокализованных и фрикативных звуков. Характеристики резонансной полости эмулируются с помощью прохождения возбуждающих сигналов через цифровой фильтр с подобными резонансными (частотными) характеристиками.

Синтез речевых сигналов основывается на знании характеристик возбуждающего сигнала $e(t)$ и речевого тракта с импульсной характеристикой $h(t)$. В отношении сигнала $e(t)$ выясняется его форма: квазипериодическая или шумовая. Период сигнала $e(t)$ определяет частоту тона речи.

На коротком интервале длительностью 2...40 мс речевой сигнал можно синтезировать, если для возбуждения речевого тракта использовать два генератора. Источником возбуждения звонких звуков служит генератор последовательности импульсов $x(n)$. Этот генератор через каждые N отсчетов вырабатывает импульс, соответствующий очередной порции воздуха. Интервал между этими импульсами является периодом основного тона:

$$T = \frac{1}{f_0}.$$

Величина f_0 равна частоте колебания голосовых связок. Период T связан с системой синхронизации.

Выходная последовательность генератора шума (генератора случайных чисел) имитирует турбулентный поток, соответствующий фрикативным звукам. Каждый из источников возбуждения синтезатора соединяется с входом линейного фильтра с КЧХ $H(e^{j\omega})$ речевого тракта. Через каждые 10 мс коэффициенты $h(n)$ изменяются, отражая изменения формы тракта и $H(e^{j\omega})$. На выходе фильтра формируются дискретные сигналы $y(n)$ с частотой f_0 .

На рис. 2.1. представлена обобщенная структурная схема синтеза речи с использованием динамической квазилинейной дискретной системы (цифрового фильтра). Дискретная линейная система (фильтр) с изменяющейся комплексной частотной характеристикой синтезатора должна формировать кратковременный спектр, соответствующий разным звукам речевого сигнала. Связь между входом и выходом системы представляется операцией

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega}), \quad (2.1)$$

где $X(e^{j\omega}) = \sum_{n=0}^{\infty} x(n)e^{-j\omega n}$ – Фурье-образ входной дискретной последовательности $x(n)$; $Y(e^{j\omega})$ – Фурье-образ отклика $y(n)$ фильтра на $x(n)$.

По другой схеме синтезируемый сигнал получается при одновременном суммировании вокализированной и невокализированной составляющих. Существуют схемы синтезаторов речи, в которых используется принцип линейного прогнозирования речи.

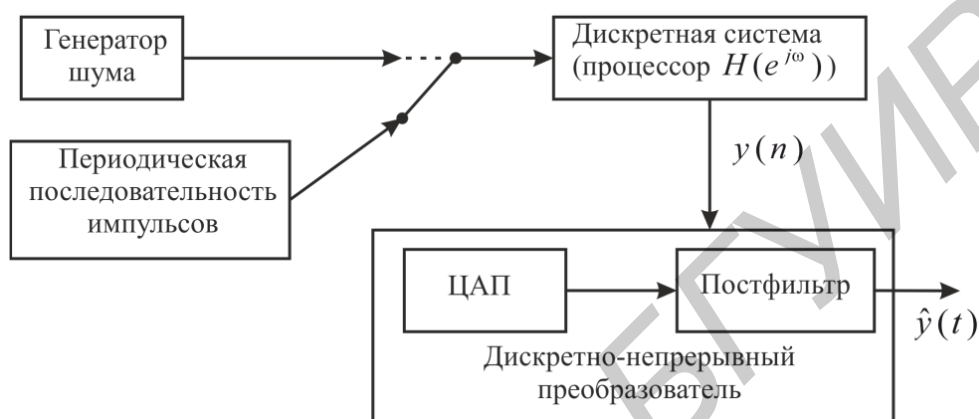


Рис. 2.1. Структурная схема синтезатора речи

2.1. Линейные системы обработки речи

2.1.1. Преобразование сигналов стационарными линейными системами с постоянными параметрами

Во многих приложениях применяется класс линейных систем с постоянными параметрами. Они сравнительно просты в математическом отношении, дают удобный вид обработки сигналов. Анализ физической системы, использующей технику обработки сигналов, часто приводит к схеме, которая представлена на рис. 2.2.

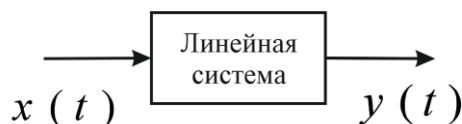


Рис. 2.2. Непрерывная линейная система

Линейная система определяется как однозначное преобразование или оператор, переводящий входной сигнал $x(t)$ (вход) в выходной $y(t)$ (выход, отклик, или реакция системы). Речевой сигнал $x(t)$ представляется как функция времени t . Математически это записывается в виде

$$y(t) = T[x(t)]. \quad (2.2)$$

Соотношение (2.2) – это правило, или формула, по которому вычисляется реакция системы через отсчеты сигнала, поданного на ее вход. Вид оператора $T[*]$ зависит от свойств конкретной системы. Оператор $T[*]$ определяет характер математических операций при отображении сигнала $x(t)$ в сигнал $y(t)$.

Класс линейных систем определяется линейными операциями или принципом суперпозиции. Если $x_1(t)$ и $x_2(t)$ – входные сигналы, а $y_1(t)$ и $y_2(t)$ – выходные сигналы, то при подаче на вход суммы сигналов вида

$$ax_1(t) + bx_2(t)$$

систему называют линейной тогда и только тогда, когда выполняется

$$T[ax_1(t) + bx_2(t)] = aT[x_1(t)] + bT[x_2(t)] = ay_1(t) + by_2(t), \quad (2.3)$$

где a и b произвольные постоянные параметры (константы) системы.

Выражение (2.3) характеризует свойство аддитивности линейной системы, в соответствии с которым реакция системы на сумму воздействий равна сумме реакций на каждое воздействие. Выражение (2.3) характеризует также свойство однородности линейной системы, в соответствии с которым умножение входного сигнала на постоянный параметр соответствует выходному сигналу, умноженному на тот же параметр:

$$T[ax(t)] = aT[x(t)] = ay(t). \quad (2.4)$$

Если система описывается линейным оператором $y(t) = T[x(t)]$, то это означает, что в данном преобразовании возможны только линейные операции сложения, вычитания и умножения на постоянный параметр и соотношение вход – выход линейной системы описывается линейным уравнением.

Очевидно, что условию линейности систем цифровой обработки сигналов (ЦОС) при замене непрерывного аргумента t на дискретный n удовлетворяют следующие выражения:

$$T[ax_1(n) + bx_2(n)] = aT[x_1(n)] + bT[x_2(n)] = ay_1(n) + by_2(n);$$

$$T[ax(n)] = aT[x(n)] = ay(n).$$

В ЦОС широко используется класс инвариантных к сдвигу во времени систем, называемых стационарными. В этих системах временной сдвиг (задержка) входного сигнала индуцирует сдвиг выходного. Если отклик $y(n)$ соответствует воздействию $x(n)$, $y(n - n_n)$ будет откликом на воздействие $x(n - n_n)$, где n_n – задержка входного сигнала. Выполняется соотношение

$$T[x(n - n_n)] = y(n - n_n).$$

Для того чтобы выяснить особенности линейной системы, зная сигналы на входе, необходимо измерить сигналы на выходе. Чтобы решить эту задачу, необходимо знать, как связаны входной и выходной сигналы с параметрами линейной системы.

2.1.2. Связь между входным и выходным сигналами в линейной системе

Наличие свойств линейности и стационарности позволяет представлять анализируемую систему в удобном виде. Предположим, на вход системы (рис. 2.3) подан единичный импульс $u_0(n)$.

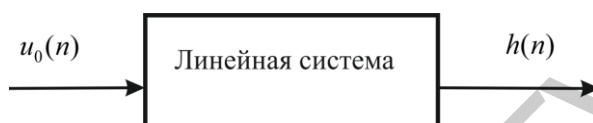


Рис. 2.3. К определению импульсной характеристики

Обычно выход линейной системы является запаздывающей и усиленной или подавленной версией входа. На рис. 2.4. показана форма возможного выходного сигнала линейной стационарной системы. Отклик системы на этот сигнал обозначим $h(n)$. Выход системы, инициированный единичным импульсом, не будет тем же импульсом, а будет меняться со временем, в некоторый момент, достигая максимального значения, например, как показано на рис. 2.4.

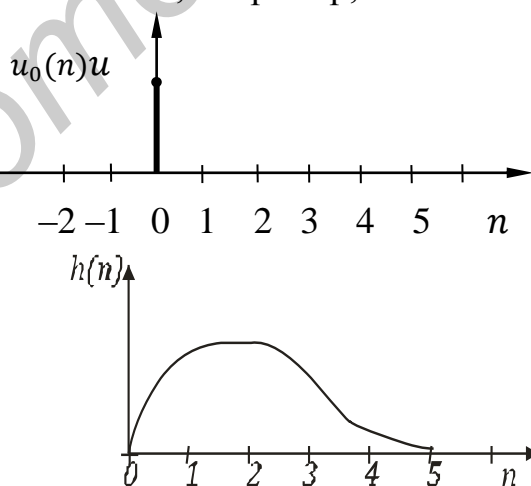


Рис. 2.4. Входной единичный импульс и импульсная характеристика системы

Из рисунка видно, что в момент времени $n = 0$ выход равен $y(0) = h(0)$. В момент времени n выход равен $y(n) = h(n)$. Вход можно представить импульсом разной амплитуды (мощности). В момент времени $n = 0$ подадим на вход системы отсчет высотой $x(0) = x(0)u_0(0)$. Поскольку входной сигнал

увеличился в $x(0)$ раз, то согласно свойствам линейной системы, отклик на этот сигнал также увеличился в $x(0)$ раз. Выходной сигнал будет равен

$$y(0) = x(0) h(0).$$

В моменты n дискретного времени выходной сигнал $y(n)$ определяется значениями отсчетов:

$$y(0) = x(0) h(0);$$

$$y(1) = x(0) h(1);$$

$$y(n) = \dots x(0) h(n).$$

Пусть теперь на вход линейной системы подается последовательность импульсов $x(n)$ (рис. 2.5).

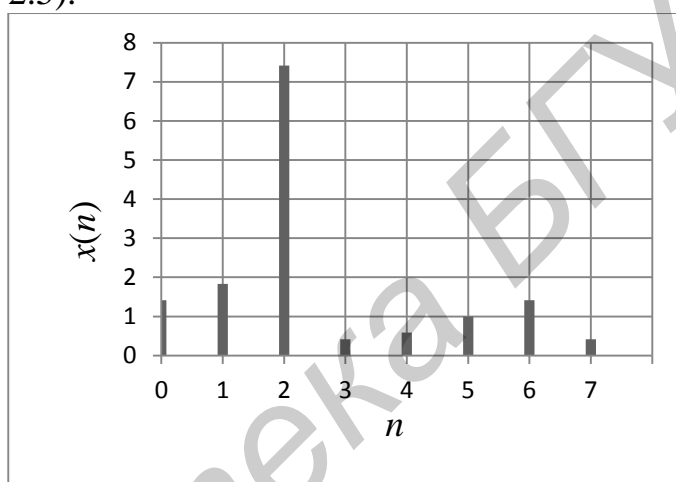


Рис. 2.5. Входная последовательность линейной системы

В этом случае на каждое импульсное воздействие формируется отклик, соответствующий импульсной характеристике системы. Выход системы получается суммированием всех отдельных импульсных откликов (рис. 2.6).

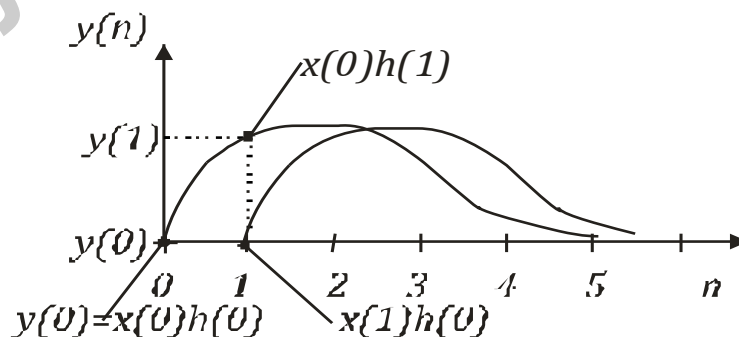


Рис. 2.6. Два отклика на воздействия $x(0), x(1)$

Из рисунка видно, что в момент времени $n = 0$ выход равен $y(0) = x(0)h(0)$. В следующий момент времени $n = 1$ выход состоит из суммы двух компонент:

– первое слагаемое $x(1)h(0)$ – из-за текущего импульса $x(1)$;

– второе слагаемое $x(0)h(1)$ – из-за запаздывающего влияния входа в момент времени $n = 0$. Выходной сигнал системы в момент времени $n = 1$ равен

$$y(1) = x(0)h(1) + x(1)h(0).$$

Последующие выходы запишутся как

$$y(2) = x(0)h(2) + x(1)h(1) + x(2)h(0);$$

$$y(n) = x(0)h(n) + x(1)h(n-1) + \dots + x(n)h(0).$$

Выводы

1. Если система линейна, выход можно записать как линейную сумму влияния предыдущих входов.

2. Выход получается умножением входной последовательности на соответствующие отсчеты обращенной во времени функции импульсной характеристики.

3. Учитывая свойство линейности и возможность представления любой последовательности в форме линейной комбинации сдвинутых единичных импульсов (1.7), можно утверждать, что линейная стационарная система определяется своей реакцией на единичный импульс или импульсной характеристикой при нулевых начальных условиях.

Замечание. Признаком нулевых начальных условий является отсутствие отклика при отсутствии входной последовательности $x(n)$. Если начальный момент времени обозначим $n = 0$, то ему соответствует

$$\begin{cases} x(n) \equiv 0, \\ y(n) \equiv 0, \end{cases} \text{ при } n < 0. \quad (2.5)$$

Нулевые начальные условия отвечают принципу причинности, в соответствии с которым отклик не может возникнуть раньше воздействия.

Для описания во временной области линейных стационарных дискретных систем широко применяются следующие линейные уравнения:

– свертка, где используется импульсная характеристика;

– разностные уравнения, где используются параметры линейных дискретных систем (ЛДС).

2.1.3. Уравнение свертки

Из свойств стационарности линейной системы (инвариантности к сдвигу) следует, что откликом на последовательность $u_0(n - k)$ будет $h_0(n - k)$ (рис. 2.7).

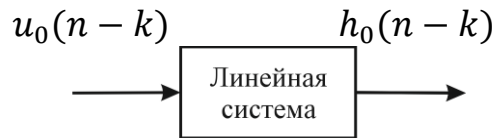


Рис. 2.7. Стационарная линейная система

Рассмотрим отклик ЛДС на входную последовательность, представленную в виде (1.3):

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)u_0(n-k).$$

Из свойств линейности и стационарности ЛДС следует, что откликом на эту последовательность должна быть последовательность

$$\begin{aligned} y(n) &= T\left[\sum_{k=-\infty}^{\infty} x(k)u_0(n-k)\right] = \sum_{k=-\infty}^{\infty} x(k)T[u_0(n-k)] = \\ &= \sum_{k=-\infty}^{\infty} x(k)h(n-k), \quad -\infty < n < \infty. \end{aligned} \quad (2.6)$$

Если $y(n)$ – последовательность, значения которой зависят от отсчетов двух последовательностей $h(n)$ и $x(n)$ и определяется выражением (2.6), то говорят, что $y(n)$ есть свертка. Обозначение операции свертки

$$y(n) = x(n) \cdot h(n). \quad (2.7)$$

С учетом нулевых начальных условий (2.5) линейное уравнение (2.6) записывается в виде

$$y(n) = \sum_{k=0}^{\infty} x(k)h(n-k), \quad -\infty < n < \infty. \quad (2.8)$$

Согласно этим формулам отклик $y(n)$ вычисляется как дискретная свертка воздействия и импульсной характеристики. Определим значение коэффициента свертки в точке $n = 3$. По формуле (2.8) имеем

$$y(3) = x(0)h(3) + x(1)h(2) + x(2)h(1) + x(3)h(0).$$

Операцию вычисления дискретной свертки по формуле (2.8) называют линейной (апериодической) сверткой. Из формулы свертки следует, что $y(n)$, т. е. n -й член выходной последовательности получается в результате умножения входной последовательности (записанной как функция от k) на последовательность $h(n-k)$, $-\infty < k < \infty$. Затем при каждом фиксированном n суммируются все произведения $x(k)h(n-k)$ с параметром k в качестве параметра суммирования. Следовательно, при свертке двух последовательностей для вычисления n -го члена последовательности $y(n)$ используются все отсчеты входной последовательности $x(n)$ и $h(n)$.

Выходной сигнал (реакция) линейной стационарной дискретной системы на воздействие $x(n)$ представляется в виде комбинации сдвинутых во времени импульсных характеристик:

$$y(n) = \sum_{k=-\infty}^{\infty} x(n)h(n - k).$$

2.1.3.1. Графическое представление свертки

Свертка – это одна из наиболее используемых операций в ЦОС. Свертка – основная операция цифровой фильтрации. Для более полного понимания формулы свертки рассмотрим процесс образования последовательности $h(n - k)$. На рис. 2.8 показана последовательность $h(k) = (h(0), h(1), h(2), h(3)) = (5, 4, 3, 2)$.

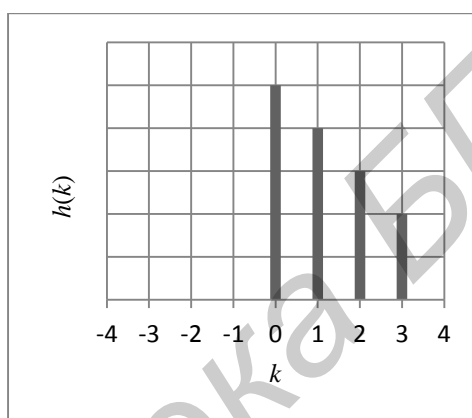


Рис. 2.8. Дискретная последовательность $h(k)$

Определим последовательность $h_1(k) = h(-k) = h(0 - k)$. Форма этой последовательности показана на рис. 2.9. Например, если $k = -1$, то $h_1(k) = h((0 - (-1)) = h(1) = 4$; $h_1(-2) = h(2) = 3$ и т. д.

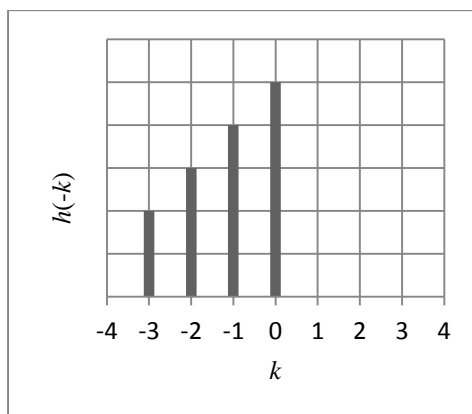


Рис. 2.9. Дискретная последовательность $h(-k)$

Далее определим последовательность

$$h_2(k) = h_1(k - n) = h(-(k - n)) = h(n - k).$$

Форма этой последовательности для $n = 3$ показана на рис. 2.10. Например, если $n = 3$, имеем $h_2(k) = h_1(k - 3)$. Для $k = 3$ получаем $h_2(3) = h_1(0) = 5$. Для $k = 2$ получаем $h_1(-1) = 4$ и т. д.

Вывод. Для вычисления последовательности $h(n - k), -\infty < k < \infty$ необходимо:

- обратить по времени последовательность $h(k)$ относительно точки $k = 0$ для получения последовательности $h(-k)$;
- сдвинуть последовательность $h(-k)$ на n интервалов дискретизации вправо.

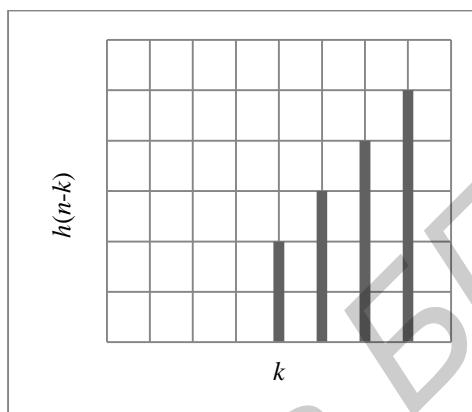


Рис. 2.10. Дискретная последовательность $h(n - k)$

Пример 2.1. Задана входная последовательность $x(n) = (x(0), x(1), x(2)) = (1, 1, 1)$ (рис. 2.11) и импульсная характеристика ЛДС $h(n) = (h(0), h(1), h(2), h(3)) = (5, 4, 3, 2)$ (рис. 2.8.) Вычислим свертку.

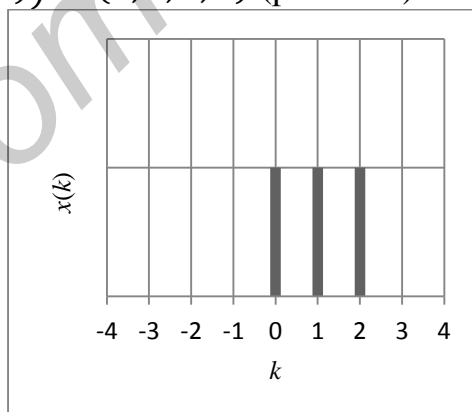


Рис. 2.11. Дискретная последовательность $h(n)$

Для $n = 0$ по формуле свертки вычисляем

$$\begin{aligned}
 y(0) &= \sum_{k=-3}^5 x(k)h(0 - k) = \\
 &= x(-3)h(3) + x(-2)h(2) + x(-1)h(1) + x(0)h(0) + x(1)h(-1) + \\
 &\quad + x(2)h(-2) + x(3)h(-3) + x(4)h(-4) + x(5)h(-5) = 5.
 \end{aligned}$$

Условно процесс вычисления свертки отображается на рис. 2.12. Временное соотношение сворачиваемых отсчетов показано ниже

$$\begin{array}{r} 1\ 1\ 1 \\ \underline{2\ 3\ 4\ 5} \\ 5. \end{array}$$

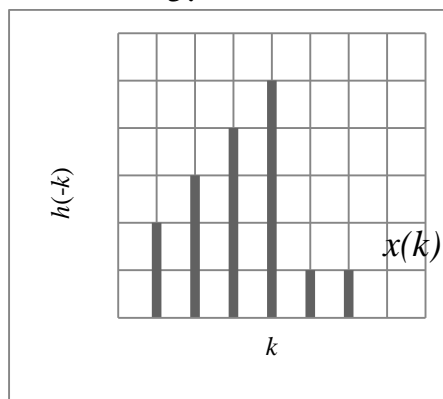


Рис. 2.12. Процесс вычисления свертки

Для $n = 1$ по формуле свертки вычисляем

$$y(1) = \sum_{k=-3}^5 x(k)h(1-k) = x(0)h(1) + x(1)h(0) = 1 \cdot 4 + 1 \cdot 5 = 9.$$

Условно процесс вычисления свертки отображается на рис. 2.13. Временное соотношение сворачиваемых отсчетов показано ниже:

$$\begin{array}{r} 1\ 1\ 1 \\ \underline{2\ 3\ 4\ 5} \\ 4+5 = 9. \end{array}$$

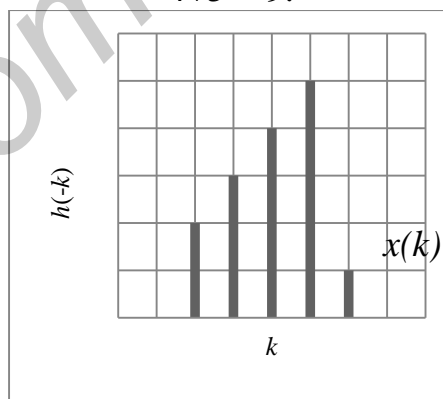


Рис. 2.13. Процесс вычисления свертки

Аналогично вычисляются остальные коэффициенты свертки.

На рис. 2.14 показана реакция линейной дискретной системы с импульсной характеристикой $h(n) = (5, 4, 3, 2)$ на входную последовательность $x(n) = (1, 1, 1)$.

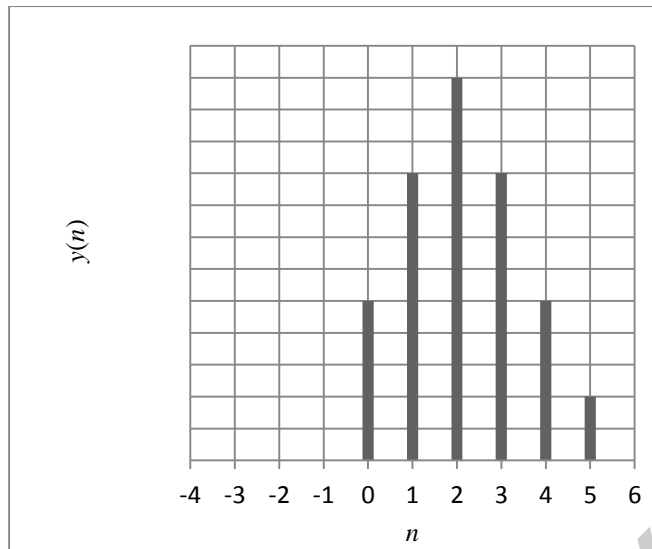


Рис. 2.14. Реакция линейной дискретной системы

Выводы

1. Если длительность входной последовательности и (или) импульсной характеристики бесконечна, длительность отклика также бесконечна.

2. Если длительность входной последовательности и импульсной характеристики конечны и равны N_1 и N_2 , то длительность отклика $y(n)$ также конечна и равна $N = N_1 + N_2 - 1$.

Для конечных последовательностей $x(n)$ и $h(n)$ формулы свертки в двух вариантах записи приобретают вид

$$y(n) = \sum_{k=0}^{N_1+N_2-2} x(k)h(n-k), n = 0, 1, \dots, N_1 + N_2 - 2. \quad (2.9)$$

$$y(n) = \sum_{k=0}^{N_1+N_2-2} h(k)x(n-k), n = 0, 1, \dots, N_1 + N_2 - 2. \quad (2.10)$$

Упражнение 2.1. Вычислить линейную свертку двух дискретных конечных последовательностей $x(n) = (1, 1, 1, 1, 1)$ и $h(n) = (0; -0,02; 0; 0,17; 0,5; 0,85; 1; 0,85; 0,5; 0,17; 0; -0,02; 0)$, где $h(n)$ – импульсная характеристика ЛДС, $x(n)$ – входная последовательность. Построить график свертки.

Упражнение 2.2. Вычислить отклик ЛДС по формуле свертки. Задана импульсная характеристика $h(n) = (1, 2, 2, 3, -1)$ и входное воздействие $x(n) = (1, 2)$. Изобразить график свертки.

Упражнение 2.3. Найти реакцию ЛДС на единичный импульс:

а) идеальной системы задержки $y(n) = x(n - n_d)$, $-\infty < n < \infty$, где n_d – целое число;

б) сумматора $y(n) = \sum_{k=-\infty}^n x(k)$.

2.2. Представление систем и сигналов в частотной области

Помимо временной области сигналы и линейные системы могут описываться в областях других независимых переменных. В этом случае соответствующие функции времени преобразуются в функции другого аргумента, например функции частоты.

Замечание. Следует различать подобные преобразования и преобразования функций одной и той же независимой переменной.

2.2.1. Спектральный (гармонический) анализ сигналов

Гармонический анализ – это раздел математики, который изучает представления функций в виде тригонометрических рядов и интегралов. В 1807 году Жан Батист Жозеф Фурье высказал идею о том, что любая сложная периодическая функция $x(t) = x(t + kT)$, $-\infty < t < \infty$, где T – период может быть представлена в виде суммы синусоидальных и/или косинусоидальных функций с частотами (кратными основной частоте), умноженных на некоторые коэффициенты.

Основным понятием в спектральном анализе является гармоническое колебание. Его временное представление имеет следующий вид:

$$x(t) = x_m \cos(\omega_0 t - \varphi_0); \quad (2.11)$$

или
$$x(t) = x_m \cos(2\pi f_0 t - \varphi_0) = x_m \cos\left(\frac{2\pi}{T} t - \varphi_0\right),$$

где x_m – амплитуда колебания (модуль наибольшего его отклонения от нуля);

T_0 [с] – период колебания;

$f_0 = \frac{1}{T_0}$ [с⁻¹], [Гц] – линейная частота;

$\omega_0 = 2\pi f_0$ $\left[\frac{\text{рад}}{\text{с}}\right]$ – круговая частота;

$\varphi_0 = \frac{\Delta T}{T} 2\pi$ [рад] – начальная фаза колебания.

Для представления такого колебания в частотной области достаточно задать две функции частоты, показывающие, что на частоте $\omega = \omega_0 = 2\pi f_0 = \frac{2\pi}{T}$ амплитуда сигнала равна x_m , а начальная фаза равна φ_0 , т. е.

$$x(\omega) = \begin{cases} x_m, & \omega = \omega_0, \\ 0, & \omega \neq \omega_0; \end{cases}$$

$$\varphi(\omega) = \begin{cases} \varphi_0, & \omega = \omega_0, \\ 0, & \omega \neq \omega_0. \end{cases}$$

Значение $x_m = x(\omega_0)$ называется амплитудным спектром гармонического колебания, а $\varphi_0 = \varphi(\omega_0)$ – фазовым спектром гармонического колебания. Их совокупность – просто спектр.

Определение 2.1. Гармоника синусоидальных функций периода T – это функции синуса или косинуса со значениями частоты, представляющими целые кратные основной частоты ω_0 . В гармоническом анализе вводится понятие n -й гармоники периодического колебания частоты, под которой понимают гармоническое колебание с частотой, в n раз превышающей основную частоту ω_0 . Например,

$$x'(t) = x_m \cos(2\omega_0 t - \varphi_0) = x_m \cos(2 \cdot 2\pi f_0 t - \varphi_0) = x_m \cos\left(\frac{2 \cdot 2\pi}{T} t - \varphi_0\right)$$

выполняет два колебания за каждые T секунд.

2.2.2. Инвариантность синусоиды для линейной системы

Если входной сигнал – гармоническое колебание $x(t) = x_m \sin(\omega_0 t - \varphi_0)$, то выходной сигнал линейной системы будет также синусоидальным (косинусоидальным) той же частоты ω_0 , хотя амплитуда x_m и начальная фаза φ_0 могут отличаться от первоначальных значений. Форма сигнала сохраняется, так как в линейной системе с сигналом возможны только такие операции, как умножение на постоянную величину, дифференцирование, интегрирование, задержка и суммирование.

2.2.3. Комплексная экспоненциальная функция

На практике наряду с синусоидальной функцией в отображении сигналов широко применяется комплексная экспоненциальная функция вида

$$x(t) = e^{j\omega t} = \cos \omega t + j \sin \omega t. \quad (2.12)$$

На рис. 2.15 показано представление этой функции.

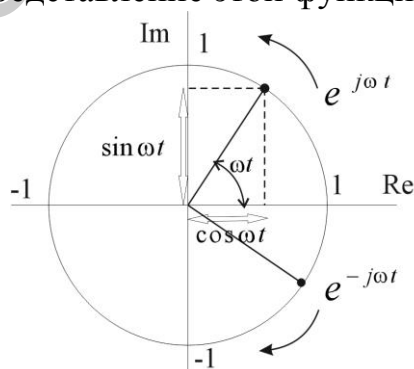


Рис. 2.15. Графическое представление функции $e^{j\omega t}$

Функция $x(t) = e^{j\omega t}$ отражает положение комплексного числа $e^{j\omega t}$ на единичной окружности в комплексной плоскости. Выражению $x(t) = e^{j\omega t}$ соответствует точка, расположенная на единичной окружности в комплексной плоскости. Прямая, соединяющая эту точку с началом координат комплексной

плоскости, образует с действительной осью угол $\varphi = \omega t$. Точка движется по окружности против часовой стрелки со скоростью $\omega \left[\frac{\text{рад}}{\text{с}} \right]$. Выражение $x(t) = e^{j\omega t}$ представляет собой единичный вектор, угол которого линейно нарастает со временем со скоростью $\omega \left[\frac{\text{рад}}{\text{с}} \right]$. Выражение $x(t) = e^{-j\omega t}$ соответствует вектору, угол которого линейно нарастает со временем в противоположном направлении с той же скоростью $\omega \left[\frac{\text{рад}}{\text{с}} \right]$. Поскольку

$$e^{-j\omega t} = \cos(-\omega t) + j \sin(-\omega t) = \cos(\omega t) - j \sin(\omega t) = (e^{j\omega t})^*,$$

где $(e^{j\omega t})^*$ обозначает комплексно-сопряженную функцию, $e^{-j\omega t}$ и $e^{j\omega t}$ – сопряженные функции.

Следовательно, сигналы вида $x_1(t) = e^{-j\omega t}$ и $x_2(t) = e^{j\omega t}$ изменяются с одинаковой частотой. Понятие отрицательной частоты следует из данных представлений. Сумма этих сигналов выражает действительную функцию времени:

$$x_1(t) + x_2(t) = e^{-j\omega t} + e^{j\omega t} = 2 \cos \omega t.$$

2.2.3.1. Нормированная частота

При анализе дискретных сигналов удобно пользоваться нормированной частотой:

$$\hat{\omega} = \frac{\omega}{f_d} = \frac{2\pi f}{f_d} = 2\pi fT \text{ или } \hat{f} = \frac{f}{f_d} = f.$$

В соответствии с теоремой дискретизации обрабатываемый сигнал ограничивается частотной областью $\omega = (0, \omega_{\max})$,

$$\hat{\omega}_{\max} = \frac{\omega_{\max}}{f_d} = \frac{2\pi f_{\max}}{f_d}, f_{\max} = \frac{f_d}{2}.$$

$$\hat{\omega}_{\max} = \frac{2\pi f_d}{2f_d} = \pi, \hat{f}_{\max} = \frac{f_{\max}}{f_d} = \frac{f_d}{2f_d} = 0,5.$$

Тогда дискретный сигнал можно рассматривать в области нормированных частот $\hat{\omega} \in [0, \pi]$ или $\hat{f} \in [0, 0,5]$. Применение нормированных частот позволяет исследовать частотные характеристики дискретных систем и спектры дискретных сигналов в единой полосе частот. Для ЦОС важны не абсолютные значения частоты сигнала и частоты дискретизации, а их отношение, т. е. значение нормированной частоты. Покажем это на примере.

Пример 2.2. На рис. 2.16 изображена дискретная синусоида $x(n) = \sin \frac{2\pi n}{N}$, где $N = 8, n = 0, 1, \dots, N - 1$. Дискретная синусоида связана с непрерывным гармоническим сигналом выражением

$$x(t) = \sin(\omega_0 t) \Leftrightarrow \sin(2\pi f_0 nT),$$

где T – период дискретизации.

Сравним две дискретные синусоиды, каждая из которых имеет следующие значения параметров:

$$x_1(t) = \sin(2\pi f_1 n T_1) = \sin(2\pi f_1 n \frac{1}{f_{d_1}}), \quad f_1 = 5 \text{ Гц}, f_{d_1} = 25 \text{ Гц}; \quad (2.13)$$

$$x_2(t) = \sin(2\pi f_2 n T_2) = \sin(2\pi f_2 n \frac{1}{f_{d_2}}), \quad f_2 = 5 \text{ кГц}, f_{d_2} = 25 \text{ кГц}. \quad (2.14)$$

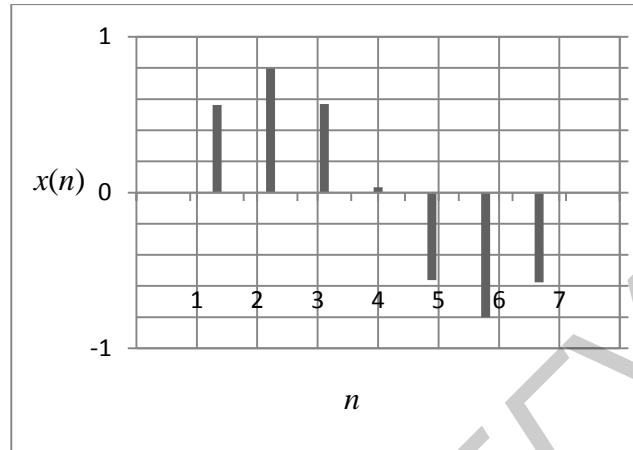


Рис. 2.16. Дискретная синусоида

Подставляя значения $f_1, f_{d_1}, f_2, f_{d_2}$ соответственно в (2.13) и (2.14), получим:

$$x_1(t) = \sin\left(2\pi 5n \frac{1}{25}\right) = \sin\left(2\pi n \frac{1}{5}\right);$$

$$x_2(t) = \sin\left(2\pi 5 \cdot 10^3 n \frac{1}{25 \cdot 10^3}\right) = \sin\left(2\pi n \frac{1}{5}\right).$$

Формы этих дискретных сигналов одинаковы, так как равны их нормированные частоты

$$\hat{f}_1 = \frac{f_1}{f_{d_1}} = \frac{5}{25} = 0,2; \quad \hat{f}_2 = \frac{f_2}{f_{d_2}} = \frac{5 \cdot 10^3}{25 \cdot 10^3} = 0,2.$$

2.2.4. Дискретная комплексная экспоненциальная последовательность

В цифровых системах сигналы определяются лишь для дискретных значений времени $t = nT$. В этом случае сигнал $x(t)$, записанный как комплексная экспоненциальная функция, преобразуем следующим образом:

$$x(t) = e^{j\omega t} = \cos \omega t + j \sin \omega t \Leftrightarrow x(nT) = e^{j\omega nT} = \cos \omega nT + j \sin \omega nT, \quad -\infty < n < \infty. \quad (2.15)$$

Для нормированной частоты $\hat{\omega}$ выражение (2.15) можно представить в виде

$$x(n) = e^{j\hat{\omega} f_{d_1} n T} = e^{j\hat{\omega} \frac{1}{T} n T} = e^{j\hat{\omega} n} = \cos \hat{\omega} n + j \sin \hat{\omega} n. \quad (2.16)$$

Определение 2.2. Функция $x(n) = e^{j\hat{\omega}n}$ называется дискретной комплексной экспоненциальной последовательностью.

Данная система функций составляет счетное бесконечное множество и определена на бесконечном интервале нормированных частот. Вещественная и мнимая часть последовательности (2.16) меняется синусоидально в зависимости от n . По аналогии с непрерывным временем параметр $\hat{\omega}$ называется круговой частотой дискретной комплексной экспоненты. В формуле (2.16) частота $\hat{\omega}$ измеряется в радианах.

2.2.5. Конечная дискретная комплексная экспоненциальная последовательность

Экспоненциальная последовательность $x(n) = e^{j\hat{\omega}n}$ может быть задана на конечном интервале времени $[0, N)$, где N – целое положительное число. Тогда величина NT определяет основной период дискретной комплексной экспоненциальной последовательности. В этом случае значение $\Delta f = \frac{1}{NT}$ – основная линейная частота последовательности. Круговая частота

$$\Delta\omega = 2\pi\Delta f = \frac{2\pi}{NT}$$

определяет период дискретизации по частоте. Абсолютное значение непрерывной частоты с учетом дискретизации переменной частоты равно

$$\omega = k\Delta\omega, k = 0, 1, \dots, N - 1.$$

Далее выполним следующее преобразование последовательности:

$$\begin{aligned} x(n) &= e^{j\hat{\omega}n} = e^{j\frac{\omega n}{f}} = e^{j\omega n T} = e^{jk\Delta\omega n T} = e^{j\frac{2\pi}{NT}knT} = e^{j\frac{2\pi}{N}kn} = \\ &= \cos\frac{2\pi}{N}kn + j \sin\frac{2\pi}{N}kn. \end{aligned} \quad (2.17)$$

Определение 2.3. Функция вида

$$\text{def}(kn) = e^{-j\frac{2\pi}{N}kn} = \exp\left(-j\frac{2\pi}{N}kn\right) = \cos\frac{2\pi}{N}kn - j \sin\frac{2\pi}{N}kn$$

называется конечной дискретной комплексной экспоненциальной функцией (ДЭФ) (последовательностью).

Введем обозначение $W = e^{-j\frac{2\pi}{N}}$. Тогда функция

$$W^{kn} = \text{def}(kn) = e^{-j\frac{2\pi}{N}kn}, k = 0, 1, \dots, N - 1, n = 0, 1, \dots, N - 1$$

называется поворачивающим множителем. Так как показатель степени комплексного числа W^{kn} со знаком «плюс», то функция $\text{def}(kn) = e^{-j\frac{2\pi}{N}kn}$ описы-

вает точку, которая движется по окружности в направлении часовой стрелки. В выражении (2.17) переменные времени n и k частоты изменяются дискретно, в отличие от (2.16), где время изменяется дискретно, а частота ω непрерывно. Огибающая дискретных значений функции $e^{j\frac{2\pi}{N}kn}$ соответствует функции $e^{j\hat{\omega}n}$. На рис. 2.17 показано графическое представление функции W^{kn} . Когда переменная n последовательно принимает значения $0, 1, \dots, N - 1$, то через N шагов комплексный вектор ДЭФ проходит 2π радиан или совершает один оборот на комплексной плоскости. Вращаясь, вектор ДЭФ занимает на плоскости только N положений.

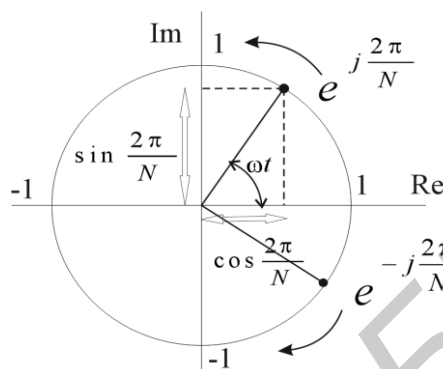


Рис. 2.17. Графическая иллюстрация функции $W^{kn} = e^{-j\frac{2\pi}{N}kn}$

Функция W^{kn} соответствует единичному вектору на комплексной плоскости, угол которого линейно нарастает со временем. Модуль комплексного числа W^{kn} равен

$$|W^{kn}| = \left| e^{-j\frac{2\pi}{N}kn} \right| = 1, \text{ а его аргумент } \varphi = \tan^{-1} \frac{\text{Im}W^{kn}}{\text{Re}W^{kn}}.$$

Пример 2.3. Вычислим значения аргумента φ (фазы) вектора ДЭФ для $k = 1, n = 1, 2, \dots, 8, N = 8$.

Решение. Значения фазы определяются как

$$\varphi_n = n \varphi, \text{ где } \varphi = -\frac{2\pi}{N} \text{ и}$$

соответственно равны $\varphi = (-45^\circ, -90^\circ, \dots, -315^\circ, 360^\circ)$. Угол φ линейно нарастает со временем $t = nT$, где T – интервал дискретизации.

Пример 2.4. Пусть $k = 2, n = 1, 2, \dots, 8, N = 8$. Значения фазы φ вектора ДЭФ соответственно равны

$$\varphi = (-90^\circ, -180^\circ, -270^\circ, -360^\circ, -450^\circ, -540^\circ, -630^\circ, -720^\circ).$$

Как видно, через 8 шагов комплексный вектор совершает два оборота на комплексной плоскости за то же время t , что и в примере 2.3. Вектор ДЭФ движется со скоростью $\frac{2\pi}{N}k \left[\frac{\text{рад}}{\text{временной отсчет}} \right]$. Параметр k значения $\frac{2\pi}{N}k$ определяет скорость нарастания (изменения) фазы или частоту функции W^{kn} . Величина фазы вектора ДЭФ за дискретное время n определяется как $\varphi = \left(\frac{2\pi}{N} \right) n$.

Таким образом, частота функции W^{kn} – это число оборотов, совершаемых вектором ДЭФ на интервале ее определения N . На рис. 2.18 показаны положения вектора ДЭФ на комплексной плоскости примера 2.3.

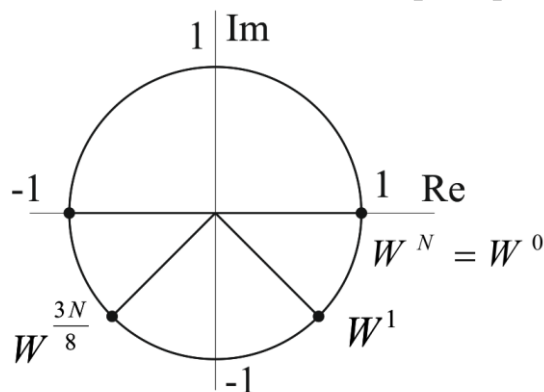


Рис. 2.18. Положения вектора ДЭФ

2.2.6. Представление периодической функции рядом Фурье

Следующим важным понятием гармонического анализа сигнал является спектр сигнала. Под спектром сигнала понимают совокупность его гармонических составляющих.

Любой сигнал $x(t)$, описываемый периодической функцией времени, удовлетворяющей условиям Дирихле (модели реальных сигналов им удовлетворяют), можно представить в виде суммы гармонических колебаний, называемой рядом Фурье:

$$x(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos k\omega_0 t + \sum_{k=1}^{\infty} b_k \sin k\omega_0 t, \quad (2.18)$$

где $a_0 = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt$ – среднее значение сигнала за период, или постоянная составляющая сигнала; $\{a_k\}, \{b_k\}$ – множества коэффициентов.

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos k\omega_0 t dt, \quad k = 0, 1, \dots; \quad (2.19)$$

$$b_k = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin k\omega_0 t dt, \quad k = 1, 2, \dots. \quad (2.20)$$

Из формул (2.18)–(2.20) следует, что функцию $x(t)$ можно представить множеством действительных чисел $\{a_0, a_k, b_k\}$.

2.2.6.1. Комплексная форма ряда Фурье

С целью упрощения расчетов часто используют вместо тригонометрической формы записи ряда Фурье его комплексную форму. Расчет спектров сигналов в комплексной области значительно проще, поскольку нет необходимо-

сти рассматривать отдельно коэффициенты a_k и b_k тригонометрической формы записи ряд Фурье. С учетом формул Эйлера

$$\begin{aligned}\cos k\omega_0 t &= \frac{1}{2}(e^{jk\omega_0 t} + e^{-jk\omega_0 t}); \\ \sin k\omega_0 t &= \frac{1}{2j}(e^{jk\omega_0 t} - e^{-jk\omega_0 t}); \\ x(t) &= \sum_{k=-\infty}^{k=\infty} C_k e^{jk\omega_0 t}.\end{aligned}$$

В этом случае $x(t)$ определяется множеством комплексных чисел $\{C_k\}$:

$$C_0 = \int_{-T/2}^{T/2} x(t) dt = a_0;$$

$$C_k = \frac{1}{2}(a_k - jb_k) = |C_k|e^{j\varphi_k};$$

$$\text{где } |C_k| = \sqrt{a_k^2 + b_k^2}; \quad \varphi_k = \begin{cases} \operatorname{arctg}\left(\frac{\operatorname{Im}C_k}{\operatorname{Re}C_k}\right) = \tan^{-1}\frac{-b_k}{a_k}, & k = 1, 2, \dots; \\ 0, & k = 0. \end{cases}$$

Совокупность комплексных амплитуд C_k называют комплексным спектром периодического сигнала. Совокупность значений $|C_k|$ и φ_k называется спектром периодического сигнала. Амплитуды гармоник $|C_k|$ характеризуют амплитудный спектр, а начальные фазы φ_k – фазовый спектр. Угол φ_k отражает ориентацию комплексного вектора относительно направления действительной оси. На рис. 2.19 показана геометрическая интерпретация C_k .

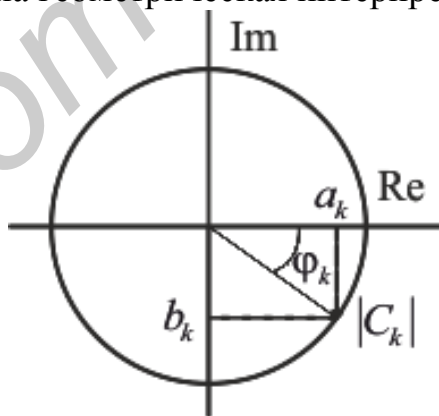


Рис. 2.19. Геометрическая интерпретация спектра

Таким образом, спектр периодического сигнала представляется в виде постоянной составляющей и бесконечного числа гармонических колебаний (синусоидальных или косинусоидальных) с соответствующими амплитудами и начальными фазами. На рис. 2.20 и 2.21 показаны амплитудный и фазовый спектры некоторого периодического фрагмента речевого сигнала.

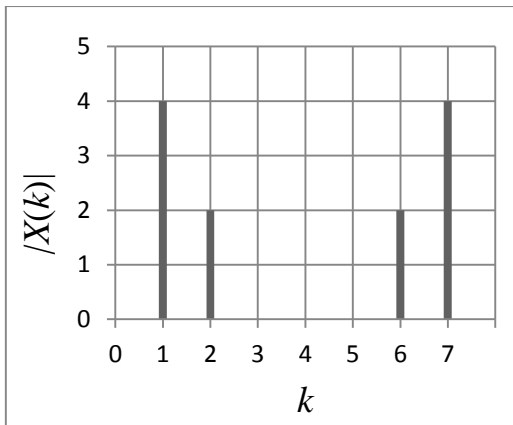


Рис. 2.20. Амплитудный спектр

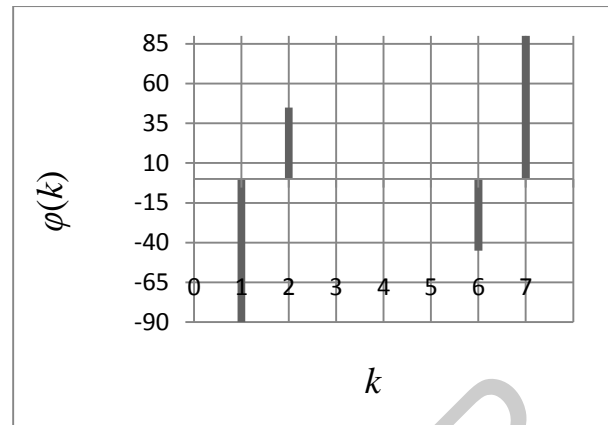


Рис. 2.21. Фазовый спектр

Каждая гармоническая составляющая изображена вертикальными отрезками, длины которых равны ее амплитуде и фазе. Как видно, спектр периодического сигнала является дискретным. Частоты всех гармоник кратны основной частоте. Это означает, что если периодический сигнал следует с частотой, например, 1 кГц, то в его спектре могут быть только частоты 0, 1, 2 кГц и т. д. В спектре такого периодического сигнала не могут присутствовать частоты 1,5 или 1,2 кГц.

2.2.7. Преобразование Фурье

Когда непрерывная функция $x(t)$ не является периодической (но площадь под графиком ее модуля конечна), она может быть описана в частотной области (аргументом является непрерывная частота ω) при помощи прямого преобразования Фурье:

$$X(j\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt. \quad (2.21)$$

Поскольку $e^{j\omega t} = \cos \omega t + j \sin \omega t$, преобразование использует множество синусоидальных функций. Значение функции $X(j\omega)$ в области ее определения определяется интегралом по всем значениям функции $x(t)$.

Замечание. Значения переменной t также влияют на частоты, но так как по этой переменной производится интегрирование, это влияние одинаково для всех значений переменной ω .

Комплексная функция $X(j\omega)$ соответствует комплексным коэффициентам C_n ряда Фурье. Различие состоит в том, что при разложении в ряд Фурье получается набор гармоник в значениях $k = \pm 1, \pm 2, \dots$. Преобразование Фурье формирует непрерывный спектр $X(j\omega)$, соответствующий огибающей этих гармоник. Преобразование Фурье можно представить призмой, которая разлагает функцию $x(t)$ на различные составляющие в зависимости от ее частотного содержания.

Функция $x(t)$ может быть полностью восстановлена при помощи обратного преобразования Фурье:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega) e^{j\omega t} d\omega, \quad (2.22)$$

где $X(j\omega)$ указывает вес, который придается выражению $e^{j\omega t}$.

Таким образом, преобразование Фурье позволяет работать в Фурье-области, а затем вернуться во временную область определения функции без потери информации.

2.2.8. Комплексная частотная характеристика линейной дискретной системы

Вычислим отклик линейной стационарной системы на входную последовательность $x(n) = e^{j\hat{\omega}n}$. Эта последовательность функционально связана с дискретизированной синусоидой с частотой ω . Отличительной особенностью сигнала (со свойством инвариантности синусоиды) является то, что откликом ЛДС остается синусоидальная последовательность с той же частотой, но с другой амплитудой и начальной фазой, которые определяются параметрами ЛДС. Далее для упрощения записи формул будем использовать обозначение нормированной частоты $\hat{\omega}$ как ω .

Если система характеризуется импульсной характеристикой $h(n)$, то

$$\begin{aligned} y(n) &= \sum_{k=-\infty}^{\infty} x(n)h(n-k) = \sum_{k=-\infty}^{\infty} h(k)e^{j\omega(n-k)} = \\ &= \sum_{k=-\infty}^{\infty} h(k)e^{j\omega n}e^{-j\omega k} = e^{j\omega n} \sum_{k=-\infty}^{\infty} h(k)e^{-j\omega k}. \end{aligned}$$

Вводя обозначение

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k)e^{-j\omega k}, \quad (2.23)$$

отклик системы можно записать как

$$y(n) = e^{j\omega n} H(e^{j\omega}). \quad (2.24)$$

Таким образом, для последовательности $x(n) = e^{j\omega n}$ отклик совпадает с входной последовательностью с точностью до комплексного множителя $H(e^{j\omega})$. Множитель $H(e^{j\omega})$ называется комплексной частотной характеристикой (КЧХ) системы, импульсная характеристика которой $h(n)$.

Из формулы (2.24) следует, что $H(e^{j\omega})$ описывает изменение комплексной амплитуды комплексной экспоненты (входного сигнала) в зависимости от значений нормированной частоты ω .

Замечание. Выражение (2.23) определяет прямое преобразование Фурье функции $h(n)$.

С учетом нулевых начальных условий ($x(n) \equiv 0, y(n) \equiv 0$ при $n < 0$), прямое преобразование Фурье

$$H(e^{j\omega}) = \sum_{n=0}^{\infty} h(n)e^{-j\omega n}. \quad (2.25)$$

Из формулы (2.25) следует, что комплексная частотная характеристика $H(e^{j\omega})$ является непрерывной периодической функцией с периодом 2π . Свойство непрерывности объясняется тем, что частота изменяется непрерывно. Периодичность КЧХ следует из подстановки $(\omega \pm 2\pi k)$ в уравнение (2.25):

$$H(e^{j(\omega \pm k)}) = \sum_{n=-\infty}^{\infty} h(n)e^{-j(\omega \pm 2\pi k)n}, k \in \mathbb{Z}.$$

Так как

$$e^{-j(\omega \pm 2\pi k)n} = e^{-j\omega n} e^{\pm j2\pi kn} = e^{-j\omega n} (\cos 2\pi kn \pm \sin 2\pi kn) = e^{-j\omega n},$$

при целых k получаем

$$H(e^{j(\omega \pm k)}) = H(e^{j\omega}).$$

Из этого же следует, что ЛДС будет иметь тот же отклик на комплексные экспоненты частот ω и $(\omega \pm 2\pi k)$.

Функция $H(e^{j\omega})$ 2π периодична, поэтому достаточно определения значений $H(e^{j\omega})$ на полуинтервале длиной 2π , т. е.

$$0 \leq \omega < 2\pi \text{ или } -\pi < \omega \leq \pi.$$

Для согласования с непрерывным случаем функцию $H(e^{j\omega})$ удобно задавать на полуинтервале $-\pi < \omega \leq \pi$. При таком выборе периода нижними частотами называют частоты, близкие к нулю, а верхними – частоты, лежащие около $\pm\pi$.

В общем случае $H(e^{j\omega})$ – комплексная функция. В алгебраической форме она выражается через действительную и мнимую части

$$H(e^{j\omega}) = \operatorname{Re}H(e^{j\omega}) + j\operatorname{Im}H(e^{j\omega})$$

или в показательной форме через модуль и аргумент, т. е.

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\arg H(e^{j\omega})} = |H(e^{j\omega})|e^{j\varphi},$$

где $\arg H(e^{j\omega}) = \varphi$ – значение фазочастотной характеристики на частоте ω .

2.2.9. Дискретизированное по времени преобразование Фурье

Произвольную входную дискретную последовательность вещественных или комплексных чисел $x(n)$ можно представить в виде интеграла Фурье:

$$x(n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) e^{j\omega n} d\omega, \quad (2.26)$$

$$\text{где } X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}, \quad (2.27)$$

что и есть дискретизированное по времени прямое преобразование Фурье последовательности $x(n)$. Преобразование $X(e^{j\omega})$ называют также спектральной функцией. Поскольку $X(e^{j\omega})$ непрерывная периодическая функция частоты, она может быть выражена рядом Фурье. Тогда формула (2.27) представляет собой разложение периодической функции $X(e^{j\omega})$ в виде ряда Фурье, в котором коэффициентами Фурье являются значения $x(n)$.

Обратное преобразования Фурье (2.26) спектральной функции $X(e^{j\omega})$ можно трактовать как представление последовательности $x(n)$ через непрерывную периодическую функцию частоты $X(e^{j\omega})$. Последовательность $x(n)$ можно рассматривать в виде суперпозиции (интеграла) экспоненциальных сигналов с комплексными амплитудами $X(e^{j\omega})$.

Замечание. Пара преобразований Фурье существует только тогда, когда ряд (2.27) сходится.

В алгебраической форме спектральная функция $X(e^{j\omega})$ выражается через действительную и мнимую части

$$X(e^{j\omega}) = \operatorname{Re}X(e^{j\omega}) + j\operatorname{Im}X(e^{j\omega}) \quad (2.28)$$

или в показательной форме через модуль и аргумент (фазу), т. е.

$$X(e^{j\omega}) = |X(e^{j\omega})| e^{j\arg H(e^{j\omega})} = |X(e^{j\omega})| e^{j\varphi},$$

где $\arg X(e^{j\omega}) = \varphi$ – значение фазочастотной характеристики на частоте ω .

Совокупность значений $|X(e^{j\omega})|$ $\varphi = \arg H(e^{j\omega})$ характеризуют амплитудный и фазовый спектры последовательности $x(n)$.

2.2.9.1. Теорема о свертке

Одним из свойств дискретизированного преобразования Фурье является свойство, определяемое как теорема о свертке.

Ранее было показано, что отклик линейной стационарной системы на последовательность $x(n) = e^{j\omega n}$ равен

$$y(n) = e^{j\omega n} H(e^{j\omega}),$$

где $H(e^{j\omega})$ – КЧХ системы, которая представляет собой комплексный коэффициент передачи для каждого значения частоты ω . Вычислим отклик линейной системы с импульсной характеристикой $h(n)$ на входную последовательность $x(n) = e^{j\omega n}$. В этом случае свертка

$$y(n) = \sum_{k=-\infty}^{\infty} x(n)h(n-k) = e^{j\omega n} H(e^{j\omega}) = x(n)H(e^{j\omega}), \quad (2.29)$$

где $x(n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) e^{j\omega n} d\omega$.

Подставляя в (2.29) выражение для $x(n)$ (2.26), получаем

$$y(n) = \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) e^{j\omega n} d\omega \right] H(e^{j\omega}). \quad (2.30)$$

Отклик линейной стационарной системы на $x(n)$ является суперпозицией откликов на каждую комплексную экспоненту, входящую в представление сигнала $x(n)$. В то же время отклик на каждую комплексную экспоненту получается умножением на комплексный коэффициент пропорциональности $H(e^{j\omega})$, зависящий от значения частоты ω . Промежуток интегрирования в (2.30) совпадает с периодом функции $H(e^{j\omega})$, поэтому можно записать

$$y(n) = \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{j\omega}) H(e^{j\omega}) e^{j\omega n} d\omega \right]. \quad (2.31)$$

Вводя обозначение

$$Y(e^{j\omega}) = X(e^{j\omega}) H(e^{j\omega}), \quad (2.32)$$

получаем КЧХ свертки. Формула (2.32) отражает связь между входом и выходом в частотной области. Выражение (2.32) – одно из важнейших свойств преобразования Фурье. Соотношение (2.31), записанное в виде

$$y(n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(e^{j\omega}) e^{j\omega n} d\omega,$$

представляет собой обратное преобразование Фурье от Фурье-образа свертки.

Вывод. Связь между входом и выходом линейной системы в виде операции свертки во временной области представляется операцией произведения в частотной области.

2.2.9.2. Преобразование Фурье на коротких интервалах времени

Моделирование речевого сигнала с помощью свертки импульсной характеристики изменяющегося во времени речевого тракта и нестационарного входного сигнала приводит к необходимости рассматривать процесс фильтрации на том временном интервале, в течение которого характеристики тракта практически остаются постоянными. В этом случае вид КЧХ $H(e^{j\omega})$ будет определять спектр речи.

Нестационарный временной процесс можно разбить на временные сегменты длиной $M \ll N$. В течение этого времени процесс можно определять как квазистационарный. В частотной области речевому сигналу будет соответствовать преобразование Фурье на коротком интервале (текущий спектр).

Пусть имеется последовательность $x(n) = \{x(0), x(1), \dots, x(N - 1)\}$. Разобьем $x(n)$ на K сегментов длиной M отсчетов:

$$x_k(n) = x(n + (k - 1)M), 0 \leq n \leq M - 1, 1 \leq k \leq K.$$

Пример 2.5. Пусть $x(n) = \{x(0), x(1), \dots, x(11)\}$, $K = 4$, $M = 3$. Получаем сегменты:

$$\begin{aligned} x_1(n) &= \{x_1(0), x_1(1), x_1(2)\} = \{x(0), x(1), x(2)\}; \\ x_2(n) &= \{x(3), x(4), x(5)\}; \\ x_3(n) &= \{x(6), x(7), x(8)\}; \\ x_4(n) &= \{x(9), x(10), x(11)\}. \end{aligned}$$

Рассмотренный алгоритм сегментации сигнала может интерпретироваться как применение к нему скользящего окна.

Преобразование Фурье на коротком интервале вычисляется по выражению

$$X_k(e^{j\omega}) = \sum_{n=0}^{M-1} x_k(n)e^{-j\omega n}, k = 1, 2, \dots, K.$$

После вычисления всех преобразований $X_k(e^{j\omega})$ получается двумерный массив спектров, который описывает изменяющиеся во времени спектральные характеристики речевого сигнала. Двумерная спектральная функция зависит от дискретной временной переменной n и непрерывной частотной переменной ω .

Сегментация последовательности $x(n)$ равносильна умножению $x(n)$ на прямоугольную весовую функцию $w(n)$ конечной длины, называемую прямоугольным окном:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq M - 1, \\ 0 & \text{для других } n. \end{cases}$$

Тогда требуемый сегмент вычисляется по формуле

$$x_k(n) = x(n)w(n - (k - 1)M), 0 \leq n \leq M - 1, 1 \leq k \leq K.$$

При использовании данных примера 2.5

$$w(n) = \begin{cases} 1, & 0 \leq n \leq 2, \\ 0 & \text{для других } n; \end{cases}$$

$$w(n) = \{w(0), w(1), w(2)\} = \{1, 1, 1\}.$$

Один из сегментов, например при $k = 3$, равен

$$x_3(n) = x(n)w(n - (3 - 1)3) = x(n)w(n - 6).$$

При $n = 0, x_3(0) = x(0)w(0 - 6) = 0;$

$n = 1, x_3(1) = x(1)w(1 - 6) = 0;$

...

$n = 6, x_3(6) = x(6)w(6 - 6) = x(6);$

$n = 7, x_3(7) = x(7)w(7 - 6) = x(7);$

$n = 8, x_3(8) = x(8)w(8 - 6) = x(8);$

$n = 9, x_3(9) = x(9)w(9 - 6) = 0;$

$n = 10, x_3(10) = x(10)w(10 - 6) = 0;$

$n = 11, x_3(11) = x(11)w(11 - 6) = 0.$

Результат сегментации $x_3(n) = \{x(6), x(7), x(8)\}$. Введем обозначение $m = (k - 1)M$. Из рассмотренного примера следует, что значение m определяет положение окна в момент времени t . При изменении n окно $w(n)$ сдвигается вправо и формируется следующий элемент последовательности $x(n)$:

$$x(n)w(n - m). \quad (2.33)$$

Пример 2.6. Пусть $w(n) = \{1, 1, 1\}$. Образует $w_1(n) = w(n - m)$.

Если $m = 0, n = 0, 1, 2$, формируется нулевой сдвиг последовательности:

$$w_1 = (111);$$

$m = 1, n = 0, 1, 2, 3$, формируется сдвиг на один временной интервал:

$$w_1 = (0111);$$

$m = 2, n = 0, 1, 2, 3, 4$, $w_1 = (00111)$ – сдвиг на два интервала;

...

$m = 6, n = 0, 1, 2, 3, 4, 5, 6, w_1 = (000000111).$

При использовании выражения (2.33) преобразование Фурье на коротком интервале можно записать как

$$X(n, e^{j\omega}) = \sum_{m=0}^n x(m)w(n - m)e^{-j\omega m}, \quad (2.34)$$

где $0 \leq \omega < 2\pi$. Напомним, что ω определяется так же, как нормированная частота $\frac{\omega}{f_d} = \hat{\omega} \Rightarrow \omega$ и измеряется в радианах.

Для осуществления спектрального описания речи в диапазоне $0 \leq \omega < 2\pi$ выбирается L частот анализа $\omega_l, l = 0, 1, \dots, L - 1$.

Определение 2.3. Представление текущего спектра речи в координатах время – частота называется спектрограммой речевого сигнала.

2.2.10. Спектральный анализ речевого сигнала

Одной из задач анализа речевых сигналов является определение спектральных характеристик возбуждающего сигнала $e(t)$. При этом выясняется форма возбуждающего сигнала $e(t)$ и частота основного тона речи (для квазипериодического сигнала). Рассмотрим пример спектрального описания речи на одной из частот ω_l в момент времени n .

Пример 2.7. Задано значение частоты анализа $\omega_l = \frac{\pi}{2}$ и текущее значение времени $n = 2$. Сегментированная последовательность (рис. 2.5) $x(n) = \{1, 2, 7\}$; $w(n) = \{1, 1, 1\}$. Подставляя в формулу $X(n, e^{j\omega_l}) = \sum_{m=0}^n x(m)w(n-m)e^{-j\omega_l m}$ исходные данные, получаем

$$\begin{aligned} X\left(2, e^{j(\omega_l = \frac{\pi}{2})}\right) &= \sum_{m=0}^2 x(m)w(2-m)e^{-j\frac{\pi}{2}m} = \\ &= x(0)w(2)e^{-j\frac{\pi}{2}0} + x(1)w(1)e^{-j\frac{\pi}{2}1} + x(2)w(0)e^{-j\frac{\pi}{2}2} = \\ &= x(0)e^{-j\frac{\pi}{2}0} + x(1)e^{-j\frac{\pi}{2}1} + x(2)e^{-j\frac{\pi}{2}2} = x(0) - jx(1) - x(2) = \\ &= (x(0) - x(2)) - jx(1) = -6 - j2. \end{aligned}$$

Амплитуда спектральной составляющей на частоте $\omega_l = \frac{\pi}{2}$ для дискретного времени $n = 2$ равна

$$\left|X\left(2, e^{j\omega_l = \frac{\pi}{2}}\right)\right| = \sqrt{(-6)^2 + (-2)^2} = \sqrt{40}.$$

Нетрудно заметить, что формула (2.34) описывает свертку последовательностей $x'(n) = x(n)e^{-j\omega n}$ и $w(n)$:

$$X(n, e^{j\omega}) = x'(n) \cdot w(n) = (x(n)e^{-j\omega n}) \cdot w(n). \quad (2.35)$$

В этом случае $X(n, e^{j\omega})$ является откликом линейной системы с импульсной характеристикой $h(n) = w(n)$ речевого тракта. Последовательность $x'(n)$ получается процессом модуляции $x(n)$ последовательностью $e^{-j\omega n}$ в диапазоне изменения частоты ω . Подставляя в (2.35) выражение $e^{-j\omega n} = (\cos \omega n - j \sin \omega n)$, получаем

$$\begin{aligned} X(n, e^{j\omega}) &= [x(n)(\cos \omega n - j \sin \omega n)] * h(n) = \\ &= x(n) \cos \omega n \cdot h(n) - jx(n) \sin \omega n \cdot h(n). \end{aligned} \quad (2.36)$$

Выражение (2.36) дает структурное описание метода спектрального анализа речевого сигнала, если комплексный спектр представить выражением

$$X(n, e^{j\omega}) = \text{Re}X(n, e^{j\omega}) + j\text{Im}X(n, e^{j\omega}),$$

где
$$\begin{cases} \text{Re}X(n, e^{j\omega}) = \sum_{m=0}^n x(m)h(n-m) \cos \omega n, \\ \text{Im}X(n, e^{j\omega}) = \sum_{m=0}^n x(m)h(n-m) \sin \omega n. \end{cases} \quad (2.37)$$

На рис. 2.22 показан метод спектрального анализа речевого сигнала $x(n)$.

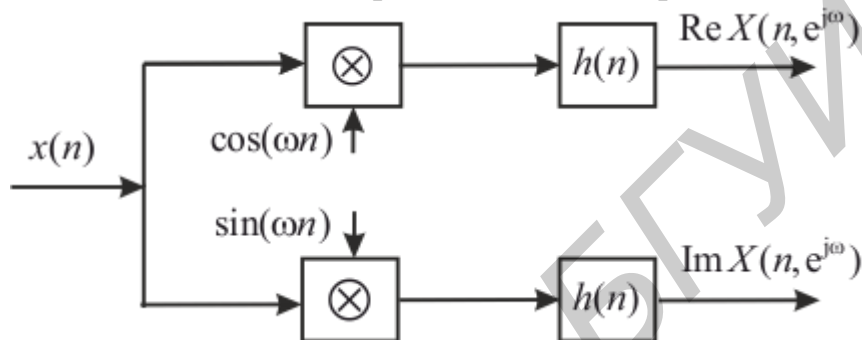


Рис. 2.22. Метод спектрального анализа речевого сигнала

Частотная характеристика $H(e^{j\omega})$ дискретного фильтра анализатора соответствует комплексному коэффициенту передачи идеального ФНЧ с частотой среза ω_s . Периодическая частотная характеристика идеального ФНЧ с периодом 2π показана на рис. 2.23, а.

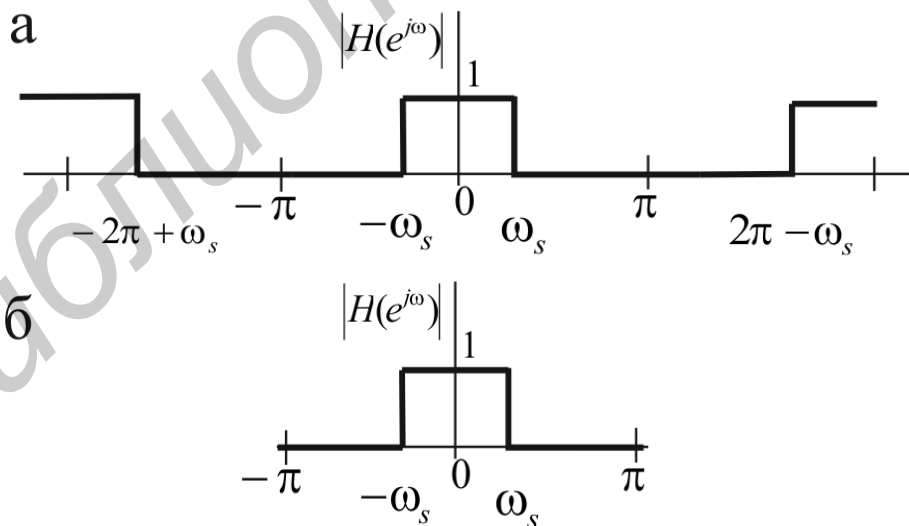


Рис. 2.23. Характеристика фильтра идеального ФНЧ

Поскольку КЧХ полностью определяется своими значениями на полуинтервале $-\pi < \omega \leq \pi$, частотную характеристику ФНЧ можно рассматривать только на этом полуинтервале, рис. 2.23, б.

2.2.10.1. Вычисление импульсной характеристики идеального ФНЧ

Частотная характеристика фильтра с частотой среза $\omega_s = \frac{\pi}{2}$ на интервале $-\pi \leq \omega \leq \pi$ определяется выражением

$$H(e^{j\omega}) = \begin{cases} 1, & -\omega_s \leq \omega \leq \omega_s, \\ 0, & \omega_s \leq |\omega| \leq \pi. \end{cases}$$

Импульсная характеристика вычисляется с помощью обратного преобразования Фурье:

$$\begin{aligned} h(n) &= \frac{1}{2\pi} \int_{-\omega_s}^{\omega_s} H(e^{j\omega}) e^{j\omega n} d\omega = \frac{1}{2\pi} \int_{-\omega_s}^{\omega_s} e^{j\omega n} d\omega = \\ &= \frac{1}{2j\pi n} e^{j\omega n} \Big|_{-\omega_s}^{\omega_s} = \frac{1}{2j\pi n} (e^{j\omega_s n} - e^{-j\omega_s n}) = \\ &= \frac{1}{2j\pi n} (\cos \omega_s n + j \sin \omega_s n - \cos \omega_s n + j \sin \omega_s n) = \\ &= \frac{2j \sin \omega_s n}{2j\pi n} = \frac{\sin \omega_s n}{\pi n}, \quad -\infty \leq n \leq \infty. \end{aligned}$$

Вычислим несколько значений коэффициентов $h(n)$. Для $n = 0$ имеем

$$h(n) = \frac{\sin \omega_s 0}{\pi 0} = \frac{0}{0}$$

неопределенность вида $\frac{0}{0}$. Применяя правило Лопиталья, дифференцируем числитель и знаменатель по $\frac{\pi}{2}n$:

$$h(0) = \frac{\cos \frac{\pi}{2} 0}{2} = \frac{1}{2}.$$

Для других значений n получаем

$$h(1) = \frac{\sin \frac{\pi}{2}}{\pi} = \frac{1}{\pi}, \quad h(2) = \frac{\sin \frac{\pi}{2} 2}{2\pi} = 0, \quad h(3) = \frac{\sin \frac{\pi}{2} 3}{3\pi} = -\frac{1}{3\pi}, \quad h(4) = 0, \quad h(5) = \frac{1}{5\pi}$$

и т. д. На рис. 2.24 показана импульсная характеристика идеального ФНЧ.

Как видно, коэффициенты импульсной характеристики стремятся к нулю при неограниченном возрастании n .

Более эффективная реализация метода анализа речи, показанного на рис. 2.22, достигается применением алгоритма дискретного преобразования Фурье (ДПФ) или с помощью алгоритма быстрого преобразования Фурье (БПФ).

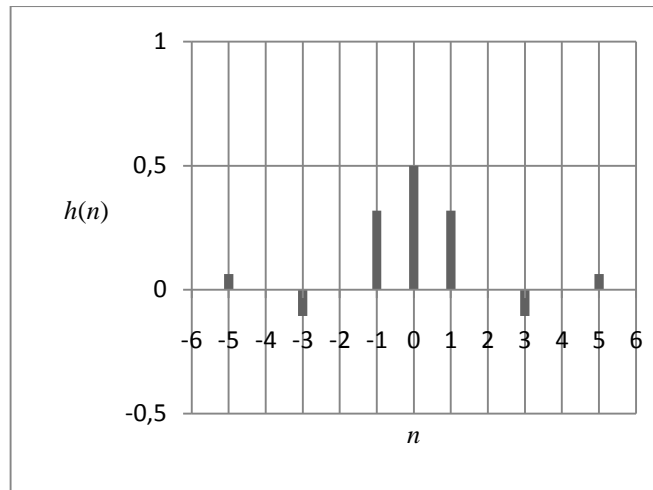


Рис. 2.24. Импульсная характеристика идеального ФНЧ (частота среза $\omega_s = \frac{\pi}{2}$)

2.2.11. Дискретное преобразование Фурье

Дискретное преобразование Фурье устанавливает связь между временным и частотным представлением сигнала при разложении его по конечным дискретным экспоненциальным функциям. Прямое дискретное преобразование Фурье (ДПФ) последовательности $x(n)$ определяет дискретную последовательность $X(k)$ в частотной области:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}} = \sum_{n=0}^{N-1} x(n) W^{kn}, k = 0, 1, \dots, N-1, \quad (2.38)$$

где k – индекс ДПФ в частотной области.

Отдельное значение $X(k)$ есть сумма почленных произведений входной последовательности отсчетов на косинусоиду и синусоиду, частота которых такова, что на интервале наблюдения из N отсчетов укладывается k их полных периодов. В алгебраической форме

$$X(k) = \text{Re}X(k) + j\text{Im}X(k).$$

В показательной форме

$$X(k) = |X(k)| e^{j\arg X(k)}, \arg X(k) = \varphi.$$

Обратное ДПФ (ОДПФ) имеет следующий вид:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi kn}{N}} = \sum_{k=0}^{N-1} X(k) W^{-kn}, n = 0, 1, \dots, N-1. \quad (2.39)$$

Пример 2.8. Вычисление ДПФ фрагмента квазипериодического речевого сигнала $x(t)$. Форма сигнала показана на рис. 2.25. Исходные параметры дискретизации: $f_d = \frac{1}{T_d} = 8000$ Гц ($T_d = 125$ мкс), $N = 8$.

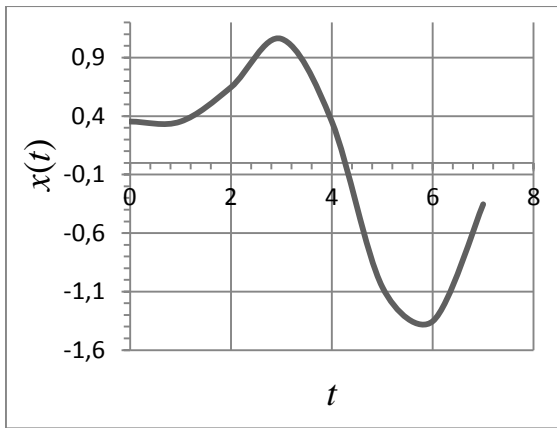


Рис. 2.25. Фрагмент речевого сигнала

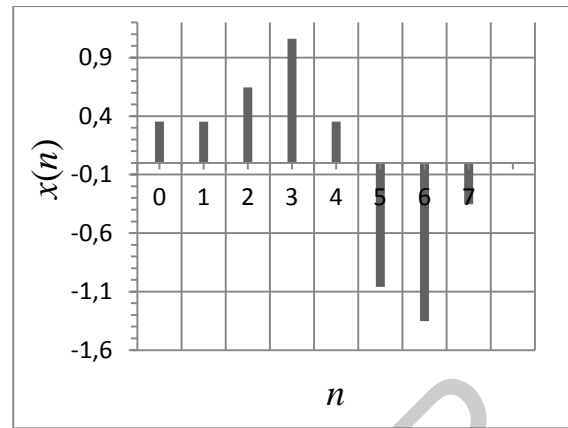


Рис. 2.26. Дискретный сигнал

Решение. После дискретизации $x(t)$ получена дискретная вещественная последовательность

$$x(n) = \{x(0), x(1), \dots, x(7)\} = \\ = (0,35; 0,35; 0,64; 1,06; 0,35; -1,06; -1,35; -0,35).$$

ДПФ определяет спектральный состав фрагмента речи в 8 точках частотной оси по формуле (2.38). N разных составляющих спектрального анализа определяется выражением

$$\frac{kf_d}{N}, k = 0, 1, \dots, N - 1.$$

Коэффициент ДПФ $X(0)$ характеризует амплитуду постоянной составляющей сигнала:

$$X(0) = \sum_{n=0}^{N-1} x(n)W^{0n} = \sum_{n=0}^7 x(n) = 0; \\ X(0) = 0,0 + j0,0, \quad \varphi(0) = 0.$$

Коэффициент $X(1)$ задает амплитуду основной гармоники частотой $\frac{f_d}{N} = \frac{8000}{8} = 1000$ Гц. Для $k = 1$ получаем

$$X(1) = \sum_{n=0}^7 x(n) \cos \frac{\pi}{8} n - jx(n) \sin \frac{2\pi}{8} n = 0,0 - j4,0.$$

Величина амплитуды основной гармоники $|X(1)| = 4$. Фазовый угол этой гармоники равен

$$\varphi(1) = \operatorname{arctg} \frac{\operatorname{Im}X(1)}{\operatorname{Re}X(1)} = \operatorname{arctg} \left(-\frac{4}{0} \right) = -\frac{\pi}{2};$$

$$X(2) = \sum_{n=0}^7 x(n) \cos \frac{2\pi}{8} 2n - jx(n) \sin \frac{2\pi}{8} 2n = 1,414 + j1,414;$$

$$|X(2)| = \sqrt{(\sqrt{2})^2 + (\sqrt{2})^2} = 2.$$

Фазовый угол второй гармоники

$$\varphi(2) = \operatorname{arctg} \frac{1,414}{1,414} = \operatorname{arctg} 1 = \frac{\pi}{4}.$$

Аналогично вычисляются остальные коэффициенты ДПФ:

$$X(3) = X(4) = X(5) = 0,0 - j0,0;$$

$$X(6) = 1,414 - j1,414, |X(6)| = 2, \varphi(6) = \arctg \frac{-1,414}{1,414} = \arctg(-1) = -\frac{\pi}{4};$$

$$X(7) = 0,0 + j4,0, \varphi(7) = \frac{\pi}{2}.$$

На рис. 2.27 показан график амплитудного спектра $|X(k)|$ фрагмента речевого сигнала. Как видно, исследуемый сигнал имеет компоненты с частотами $f_1 = 1$ кГц и $f_2 = 2$ кГц. Амплитуда первой гармоники в 2 раза выше амплитуды второй гармоники с частотой $f_2 = 2$ кГц. На рис. 2.28 отображены фазочастотные свойства сигнала. Между первой и второй гармониками имеется фазовый сдвиг величиной $\varphi = 135^\circ = \frac{3\pi}{4}$.

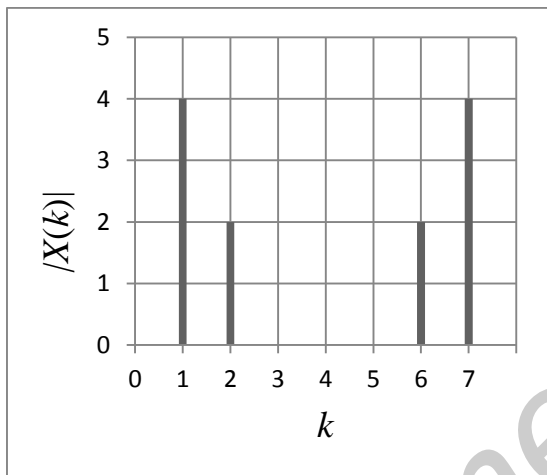


Рис. 2.27. Амплитудный спектр

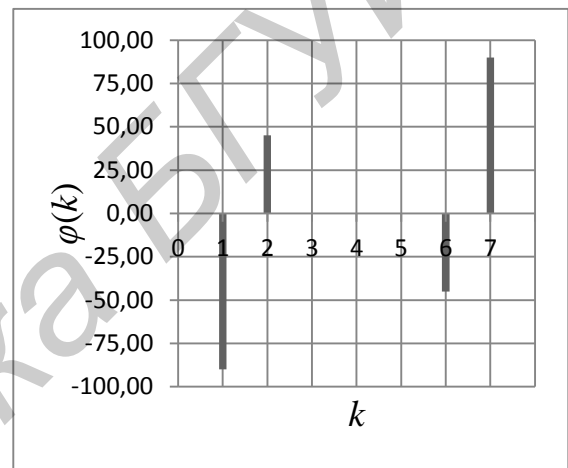


Рис. 2.28. Фазовый спектр

Замечания:

1. Если сигнал содержит синусоидальную компоненту с амплитудой A_k и целым количеством периодов на N отчетах, то реальное значение амплитуды A_k равно

$$A_k = \frac{2|X(k)|}{N}.$$

В примере были получены значения $|X(1)| = 4$, $|X(2)| = 2$. Отсюда реальное значение

$$A_1 = \frac{2 \cdot 4}{8} = 1, A_2 = \frac{2 \cdot 2}{8} = 0,5.$$

2. Если сигнал имеет постоянную составляющую A_0 , то

$$A_0 = \frac{X(0)}{N}.$$

2.2.11.1. Синтез непрерывного речевого сигнала

Данные спектрального анализа позволяют произвести синтез сигнала. Пусть были получены следующие значения анализа:

- спектральные компоненты с частотами $f_1 = 1$ кГц и $f_2 = 2$ кГц;
- амплитуды компонент $A_0 = 0, A_1 = 1$ и $A_2 = 0,5$;
- фазовый сдвиг $\varphi = \frac{3\pi}{4}$.

Тогда непрерывный сигнал формируется в виде суммы постоянной составляющей и двух гармоник:

$$x(t) = x_0 + x_1(t) + x_2(t),$$

где $x_0 = 0$; $x_1(t) = \sin(\omega_1 t) = \sin(2\pi f_1 t) = \sin(2\pi 1000t)$;

$$x_2(t) = 0,5 \sin(\omega_2 t + \frac{3\pi}{4}) = 0,5 \sin(2\pi f_2 t + \frac{3\pi}{4}) = 0,5 \sin(2\pi 2000t + \frac{3\pi}{4}).$$

Синтезированный сигнал описывается выражением

$$x(t) = \sin(2\pi 1000t) + 0,5 \sin(2\pi 2000t + \frac{3\pi}{4}).$$

Форма непрерывного речевого сигнала $x(t)$ показана на рис. 2.25.

2.2.12. Система анализа – синтеза речи

С целью более эффективной передачи речи по каналу обработка речевого сигнала разбивается на три этапа: спектральный анализ, компактное описание (сжатие) и синтез. Сжатие речи уменьшает скорость потока цифровых данных (сужает полосу частот канала).

Синтез речи во многом обусловлен особенностями реализации метода спектрального анализа. В п. 2.2.10 был рассмотрен метод анализа текущего спектра речи на коротком временном интервале, когда форма речевого тракта не изменяется. Такой форме соответствует импульсная характеристика $h(n)$ с постоянными параметрами. В этом случае цифровой фильтр нижних частот имитирует речевой тракт на одной (центральной) частоте анализа ω_k . Так как речевой тракт характеризуется изменяющимися параметрами, спектральные измерения для системы анализа должны выполняться для L частот ω_k , $k = 0, 1, \dots, L - 1$. Спектральный анализ становится параллельным L -канальным, как показано на рис. 2.29.

Рассмотренные методы анализа – синтеза можно использовать для сжатия полосы речевого сигнала и тем самым более эффективно расходовать частотный ресурс. Если передаваемые параметры $\text{Re } X(n, e^{-j\omega_k})$ и $\text{Im } X(n, e^{-j\omega_k})$ квантовать на меньшее число уровней и дискретизировать с меньшей частотой $f_d = \frac{1}{T_d}$, ширина спектра речи уменьшится. Напомним, полоса частот W_d цифрового сигнала равна величине

$$W_d = (\log_2 Q) f_d,$$

где Q – число уровней квантования.

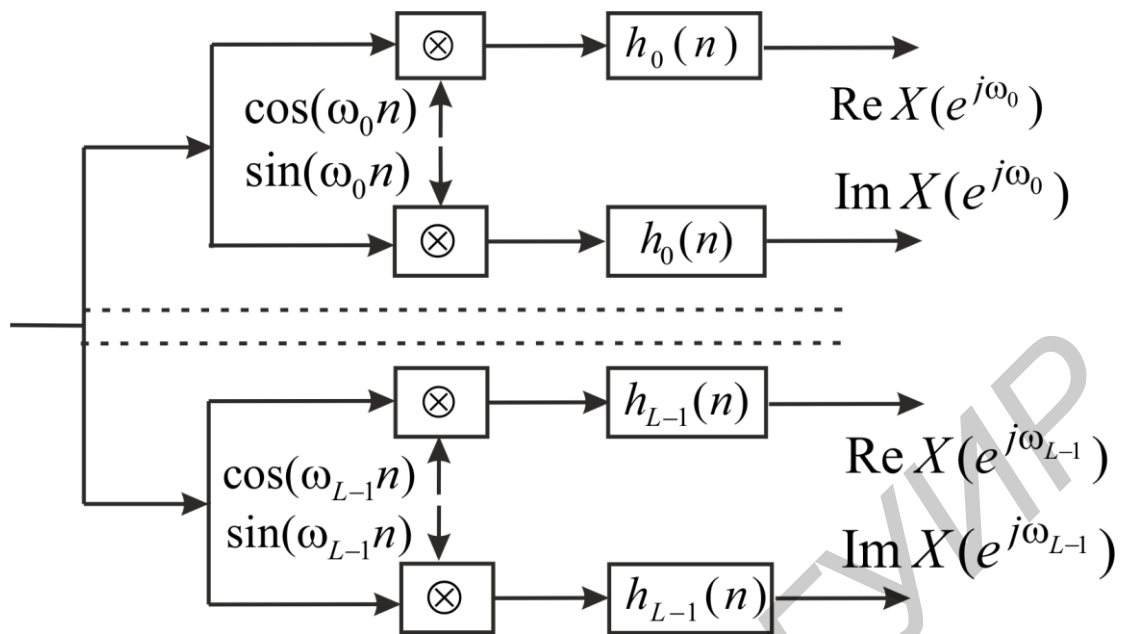


Рис. 2.29. Схема системы анализа речи

На рис. 2.30 показана схема системы анализа с возможностью сужения полосы частот речевого сигнала за счет увеличения периода дискретизации $T_d \rightarrow T_{d_1}$ и уменьшения значения $Q \downarrow$.

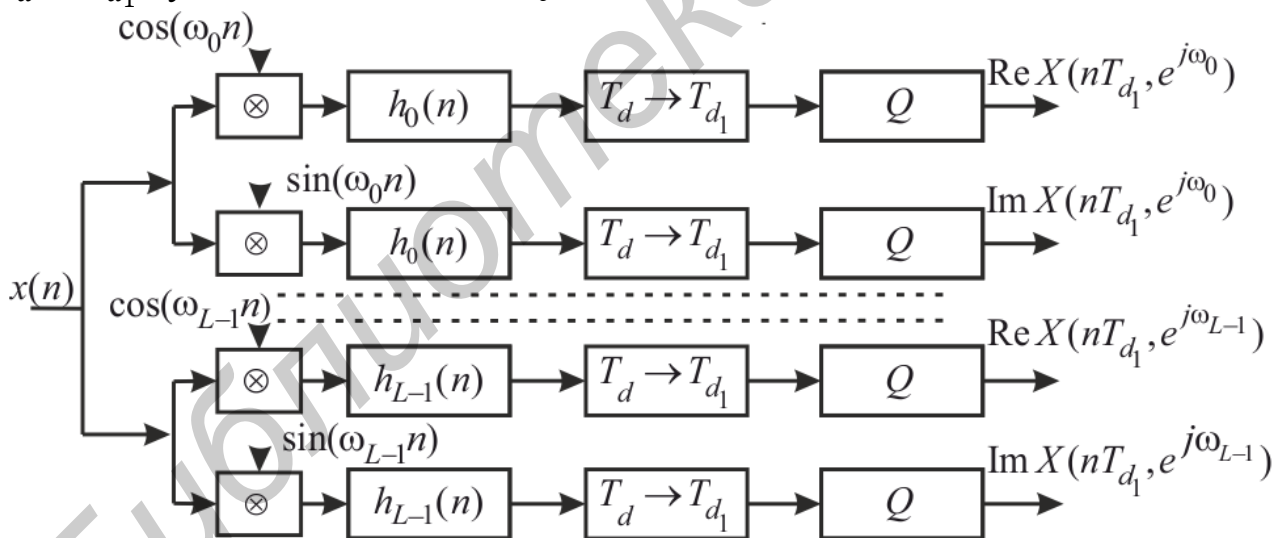


Рис. 2.30. Схема системы анализа со сжатием спектра речи

Экспериментальные исследования системы показывают, что хорошая разборчивость речи достигается при скорости цифрового потока 15 кбит/с. В отличие от обычной телефонной системы с импульсно-кодовой модуляцией со скоростью передачи 64 кбит/с получается снижение скорости в четыре раза (коэффициент сжатия $K = 4$).

Синтезируемая речь получается после суммирования результатов обратного преобразования Фурье сигналов L -канальных полосовых фильтров. Спектр речевого сигнала на выходе синтезатора определяется суммарной импульсной характеристикой:

$$h_{\Sigma}(n) = \sum_{k=0}^{L-1} h_k(n) = \sum_{k=0}^{L-1} h(n) \cos(\omega_k n) = h(n) \sum_{k=0}^{L-1} \cos(\omega_k n),$$

где $h(n)$ – импульсная характеристика фильтра нижних частот.

Синтезируемое колебание определяется выражением

$$y_k(n) = \operatorname{Re}X(n, e^{j\omega_k}) \cos(\omega_k n) + \operatorname{Im}X(n, e^{j\omega_k}) \sin(\omega_k n). \quad (2.40)$$

Синтез одного из каналов синтезатора выполняется по схеме, показанной на рис. 2.31. Импульсная характеристика каждого полосового фильтра получается в результате модуляции импульсной характеристики фильтра спектрального анализатора гармоническим колебанием $\cos(\omega_k n)$:

$$h_k(n) = h(n) \cos(\omega_k n).$$

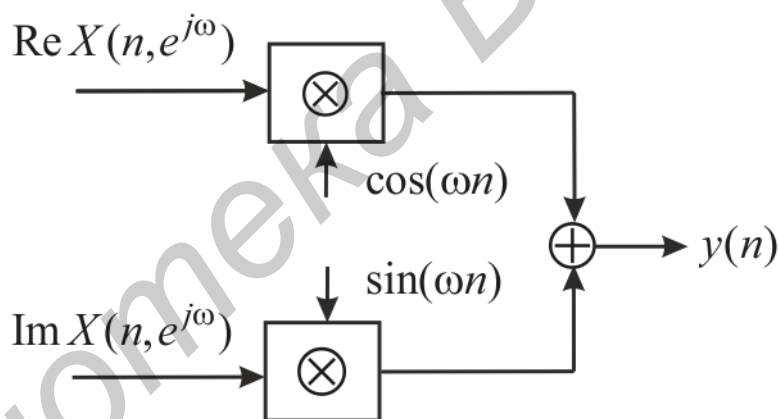


Рис. 2.31. Схема системы синтеза речи

3. Анализ изображений

Изображения описываются пространственным распределением освещенности на плоскости функцией $E(x_1, x_2)$ двух пространственных переменных и представляются двумерными массивами точек. Точка на 2-D плоскости называется пикселем – сокращение от picture element (элемент изображения). Измеряемой величиной в пикселе является освещенность. Из-за статистической природы наблюдаемого изображения g измерения покажут различные значения g освещенности. Многие алгоритмы обработки изображений оперируют над большим количеством пикселей. В этом случае данные, полученные из изображений, представляются многомерными случайными величинами. Значение каждого пикселя в m -й строке и n -м столбце можно представить в виде случайной величины $g_{m,n}$. Если обработка изображений производится блоками размером $M \times N$ (или размером $N \times N$), получается матрица $[g_{m,n}]$, элементы которой случайные величины. Случайные величины можно записывать в виде вектора $\mathbf{g} = (g_1, g_2, \dots, g_m)^T$, где T обозначает операцию транспонирования. Вектор \mathbf{g} можно рассматривать как выборку из случайного процесса.

3.1. Статистические свойства изображений

Параметрами случайных данных $\mathbf{g} = \{g_1, g_2, \dots, g_m\}$ являются среднее значение и дисперсия.

Среднее значение $m_g = E\{g\}$ определяется формулой

$$m_g = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_m^M g_m. \quad (3.1)$$

Величина отклонения значений данных g от среднего значения оценивается дисперсией

$$\sigma^2 = E\{(g - m_g)^2\}.$$

Для анализа изображений и последующей реализации определенных алгоритмов обработки, например, эффективного кодирования (сжатия), требуется знать и другие параметры случайного процесса. Изображение как случайный процесс характеризуют также ковариация и корреляция.

Ковариация двух случайных величин g_i и g_j определяется как

$$\sigma_{ij}^2 = E\{(g_i - m_i)(g_j - m_j)\}.$$

Ковариация показывает степень зависимости случайных величин g_i и g_j . Случайные величины g_i и g_j являются зависимыми, если вероятность появления величины g_i зависит от значения g_j .

Корреляции двух случайных величин g_i и g_j определяется как

$$c_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}, |c_{ij}| \leq 1. \quad (3.2)$$

Корреляция показывает степень «похожести» случайных величин g_i и g_j . Чем больше значение c_{ij} , тем более подобны случайные величины. Величины g_i и g_j являются некоррелированными, если $c_{ij} = 0$.

Для N случайных величин ковариация определяется как

$$\text{cov}(\mathbf{g}) = E\{(\mathbf{g} - \mathbf{m}_g)(\mathbf{g} - \mathbf{m}_g)^T\}, \quad (3.3)$$

где вектор $\mathbf{g} = (g_1, g_2, \dots, g_m)^T$ – случайный вектор.

Ковариация показывает, в какой степени случайные величины g_i и g_j совокупности векторов связаны друг с другом. Случайные величины g_i и g_j называются некоррелированными, если ковариация σ_{ij}^2 равна нулю.

Замечание. Для сжатия изображений с использованием дисперсионного принципа отбора спектральных коэффициентов, когда строки и столбцы обрабатываются независимо, ковариация вычисляется для строк и столбцов.

3.1.1. Статистический анализ фрагмента данных

Рассмотрим пример вычисления параметров случайных данных, представленных матрицей \mathbf{G} размером $M \times N = 2 \times 5$:

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 \\ 2 & 0 & 1 & -1 & 3 \end{bmatrix}.$$

Каждый столбец можно представить в виде вектора $\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$. Массив данных состоит из векторов множества $\{\mathbf{g}\}$:

$$\mathbf{g}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}; \mathbf{g}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \mathbf{g}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \mathbf{g}_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}; \mathbf{g}_5 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

1. Находим средний вектор данных по формуле (3.1):

$$\mathbf{m}_g = \frac{1}{N} \sum_{n=1}^N \mathbf{g}_i = \frac{1}{5} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Ковариационная матрица $\text{cov}(\mathbf{g})$ столбцов матрицы \mathbf{G} определяется из (3.3):

$$\text{cov}(\mathbf{g}) = \frac{1}{N} \sum_{n=1}^N (\mathbf{g}_n - \mathbf{m}_g)(\mathbf{g}_n - \mathbf{m}_g)^T = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1M}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2M}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{M1}^2 & \sigma_{M2}^2 & \dots & \sigma_{M2}^2 \end{bmatrix}. \quad (3.4)$$

Диагональные элементы матрицы (3.4) являются дисперсиями случайных величин векторов $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N$. Элемент σ_{ii}^2 – это дисперсия i -й координаты случайных векторов \mathbf{g} . Каждый внедиагональный элемент σ_{ij}^2 матрицы соответствует ковариации случайных величин (g_i и g_j) с i -й и j -й координатой векторов \mathbf{g} . Ковариационная матрица симметрична, так как $\sigma_{ij}^2 = \sigma_{ji}^2$.

2. Находим отклонения $(\mathbf{g} - \mathbf{m}_g)$ каждого вектора от среднего вектора:

$$(\mathbf{g}_1 - \mathbf{m}_g) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix};$$

$$(\mathbf{g}_2 - \mathbf{m}_g) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix};$$

$$(\mathbf{g}_3 - \mathbf{m}_g) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$(\mathbf{g}_4 - \mathbf{m}_g) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix};$$

$$(\mathbf{g}_5 - \mathbf{m}_g) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

3. Вычислим матричные произведения векторов отклонений:

$$(\mathbf{g}_1 - \mathbf{m}_g)(\mathbf{g}_1 - \mathbf{m}_g)^T = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix};$$

$$(\mathbf{g}_2 - \mathbf{m}_g)(\mathbf{g}_2 - \mathbf{m}_g)^T = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix};$$

$$(\mathbf{g}_3 - \mathbf{m}_g)(\mathbf{g}_3 - \mathbf{m}_g)^T = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix};$$

$$(\mathbf{g}_4 - \mathbf{m}_g)(\mathbf{g}_4 - \mathbf{m}_g)^T = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \begin{bmatrix} 0 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix};$$

$$(\mathbf{g}_5 - \mathbf{m}_g)(\mathbf{g}_5 - \mathbf{m}_g)^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

4. По формуле (3.3) получаем

$$\text{cov}(\mathbf{g}) = \frac{1}{5} \left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \right) = \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix}.$$

Используя выражение (3.2), вычисляем корреляционную матрицу случайных векторов:

$$C = \begin{bmatrix} 1 & 0,224 \\ 0,224 & 1 \end{bmatrix},$$

где $c_{12} = c_{21} = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} = \frac{0,2}{\sqrt{0,4 \cdot 2}} = 0,224$.

Как видно, случайные величины g_1 и g_2 всей совокупности векторов $\{g\} = \{g_1, \dots, g_5\}$ сравнительно мало коррелированы.

3.1.2. Собственные векторы и собственные значения матрицы данных

Сформулируем следующую задачу. Требуется найти скаляры λ , для которых существует такой не равный нулю вектор X , что выполняется равенство

$$AX = \lambda X, \quad (3.5)$$

где A обозначает некоторую матрицу размером $n \times n$.

Выражение (3.5) означает, что преобразование вида

$$T[X] = \lambda X$$

отражает вектор X на некоторое кратное λX . Когда это происходит, X называют собственным вектором, который соответствует собственному значению λ , а вместе они образуют собственную пару (λ, X) для матрицы A . Уравнение (3.5) является нелинейным, поскольку содержит произведение обоих неизвестных λ и X . Если число λ известно, то уравнение относительно X становится линейным.

Используя единичную матрицу I , уравнение (3.5) перепишем как

$$\begin{aligned} AX &= \lambda IX; \\ AX - \lambda IX &= (A - \lambda I)X = 0. \end{aligned} \quad (3.6)$$

Из выражения (3.6) следует, что произведение матрицы $(A - \lambda I)$ и вектора X равно нулевому вектору. Еще раз повторим, нас интересуют только те значения λ , для которых имеется ненулевой собственный вектор X . Другими словами, матрица $(A - \lambda I)$ должна быть вырожденной. Матрица L называется невырожденной или несингулярной (лат. singularis – особый), если существует такая обратная матрица M , что

$$LM = ML = I.$$

Критерий вырожденности дается определителем матрицы.

Теорема 3.1. Число λ является собственным значением матрицы A с соответствующим ненулевым собственным вектором тогда и только тогда, когда

$$\det(A - \lambda I) = 0. \quad (3.7)$$

Так как I – единичная матрица, определитель (3.7) можно представить выражением

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n1} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0.$$

Если определитель записан в виде полинома степени n , его называют характеристическим полиномом

$$p(\lambda) = \det(A - \lambda I) = (-1)^n (\lambda^n + c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \cdots + c_{n-1} \lambda + c_n). \quad (3.8)$$

Полином (3.8) имеет n корней. Каждый их корней λ можно подставить в уравнение (3.6), чтобы получить конкретную систему уравнений, которая имеет собственный вектор X . Для $\lambda \in \mathbb{R}$ можно получить действительный собственный вектор X .

Определение 3.1. Если A – действительная матрица размером $n \times n$, то ее n собственных значений $\lambda_1, \lambda_2, \dots, \lambda_n$ это действительные и комплексные корни характеристического полинома (3.8).

Теорема 3.2. Для каждого собственного значения λ существует по крайней мере один собственный вектор, соответствующий λ .

Теорема 3.3. Если собственные значения матрицы A различны, то существуют n собственных векторов $X_j, j = 1, 2, \dots, n$.

Пример 3.1. Задана матрица $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$. Вычислить собственные пары (λ_j, X_j) матрицы A .

Решение. Характеристическое уравнение определяется выражениями

$$p(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{bmatrix} = 0;$$

$$(4 - \lambda)(-3 - \lambda) + 10 = \lambda^2 - \lambda - 2 = 0.$$

Напомним, что для квадратного уравнения $x^2 + px + q = 0$ произведение корней $(x_1 \cdot x_2) = q$, а сумма корней $(x_1 + x_2) = -p$.

Характеристический многочлен раскладывается в произведение двучленов:

$$(\lambda + 1)(\lambda - 2) = \lambda^2 - \lambda - 2.$$

Тогда $\lambda_1 = -1, \lambda_2 = 2$. Каждому из этих значений соответствует пространство собственных векторов, удовлетворяющих уравнению

$$p(\lambda) = \det(A - \lambda I) = 0.$$

Для первого корня (собственного значения) вычисляем собственный вектор:

$$(A - \lambda_1 I)X_1 = 0;$$

$$\begin{bmatrix} 4 + 1 & -5 \\ 2 & -3 + 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{cases} 5x_1 - 5x_2 = 0, \\ 2x_1 - 2x_2 = 0, \end{cases} x_1 = 1, x_2 = 1.$$

Собственным вектором является любой, кратный вектору $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Аналогично для второго собственного значения $\lambda_2 = 2$ вычисляем собственный вектор:

$$(A - \lambda_2 I)X_2 = 0;$$

$$\begin{bmatrix} 4 - 2 & -5 \\ 2 & -3 - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{cases} 2x_1 - 5x_2 = 0, \\ 2x_1 - 2x_2 = 0, \end{cases} x_1 = 5, x_2 = 2.$$

Собственным вектором является любой, кратный вектору $X_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$.

3.1.2.1. Диагональная матрица собственных значений

Теорема 3.4. Для действительной и симметричной матрицы всегда существует ортонормированный базис, состоящий из n собственных векторов.

Пример 3.2. Задана ковариационная матрица $K_g = \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix}$. Вычислить собственные пары (λ_j, X_j) матрицы K_g .

Характеристическое уравнение определяется выражениями:

$$p(\lambda) = \det(K_g - \lambda I) = 0;$$

$$\det \begin{bmatrix} 0,4 - \lambda & 0,2 \\ 0,2 & 2 - \lambda \end{bmatrix} = 0;$$

$$(0,4 - \lambda)(2 - \lambda) - 0,04 = \lambda^2 - 2,4\lambda + 0,76 = 0;$$

$$\lambda_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q} = 1,2 \pm \sqrt{1,44 - 0,76} = 1,2 \pm \sqrt{0,68} = \\ = 1,2 \pm 0,824; \lambda_1 = 2,024, \lambda_2 = 0,375.$$

Проверим правильность вычисленных значений корней характеристического уравнения:

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = (\lambda - 2,024)(\lambda - 0,375) = \lambda^2 - 2,4\lambda + 0,76.$$

Каждому из этих значений корней соответствует пространство собственных векторов, удовлетворяющих уравнению

$$p(\lambda) = \det(\mathbf{K}_g - \lambda \mathbf{I}) = 0.$$

Для первого собственного значения $\lambda_1 = 2,024$ вычисляем собственный вектор:

$$(\mathbf{K}_g - \lambda_1 \mathbf{I})\mathbf{X}_1 = 0;$$

$$\begin{bmatrix} 0,4 - 2,024 & 0,2 \\ 0,2 & 2 - 2,024 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{bmatrix} -1,624 & 0,2 \\ 0,2 & -0,024 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{cases} -1,624x_1 + 0,2x_2 = 0, \\ 0,2x_1 - 0,024x_2 = 0. \end{cases}$$

Выберем $x_2 = a$, где a – произвольная постоянная, и вычислим x_1 :

$$0,2a = 1,624x_1, \quad x_1 = \frac{0,2a}{1,624} = 0,123a.$$

Собственным вектором является любой, кратный вектору

$$\mathbf{X}_1 = \begin{bmatrix} 0,123a \\ a \end{bmatrix} = a \begin{bmatrix} 0,123 \\ 1 \end{bmatrix}.$$

Проверим правильность вычисленного значения координат вектора по второму уравнению системы:

$$0,2 \cdot 0,123a - 0,024a \cong 0.$$

Проверим также тождество $\mathbf{K}_g \mathbf{X}_1 = \lambda_1 \mathbf{X}_1$:

$$\mathbf{K}_g \mathbf{X}_1 = \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix} \begin{bmatrix} 0,123 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,249 \\ 2,024 \end{bmatrix};$$

$$\lambda_1 \mathbf{X}_1 = 2,024 \begin{bmatrix} 0,123 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,249 \\ 2,024 \end{bmatrix}.$$

Далее вычисляем собственный вектор для второго собственного значения $\lambda_2 = 0,375$:

$$(K_g - \lambda_2 I)X_2 = 0;$$

$$\begin{bmatrix} 0,4 - 0,375 & 0,2 \\ 0,2 & 2 - 0,375 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{bmatrix} 0,02462 & 0,2 \\ 0,2 & 1,624 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{cases} 0,02462x_1 + 0,2x_2 = 0, \\ 0,2x_1 + 1,624x_2 = 0. \end{cases}$$

Выберем $x_2 = b$, где b – произвольная постоянная, и вычислим x_1 :

$$0,2b = -0,02462x_1, \quad x_1 = -\frac{0,2b}{0,02462} = -8,134b.$$

Собственным вектором является любой, кратный вектору

$$X_2 = \begin{bmatrix} -8,134b \\ b \end{bmatrix} = b \begin{bmatrix} -8,134 \\ 1 \end{bmatrix}.$$

Проверим правильность вычисленного значения координат вектора по второму уравнению системы:

$$0,2 \cdot (-8,134b) + 1,624b \cong 0.$$

Образуем матрицу $Z = [X_1 \ X_2]$, состоящую из собственных векторов ковариационной матрицы K_g :

$$Z = [X_1 \ X_2] = \begin{bmatrix} 0,123 & -8,134 \\ 1 & 1 \end{bmatrix}. \quad (3.9)$$

Векторы X_1 и X_2 ортогональны и линейно независимы, поскольку

$$X_1^T X_2 = 0.$$

Далее запишем матрицу (3.9) в ортонормированной форме. Это означает, что каждый вектор делится на свою длину $\frac{X}{\|X\|}$,

где $\|X\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

Координатам вектора X_1 соответствуют значения:

$$x_1 = \frac{0,123}{\sqrt{0,123^2 + 1}} \cong 0,122; \quad x_2 = \frac{1}{\sqrt{0,123^2 + 1}} \cong 1.$$

Координатам вектора X_2 соответствуют значения:

$$x_1 = \frac{-8,134}{\sqrt{-8,134^2 + 1}} \cong -1; \quad x_2 = \frac{1}{\sqrt{-8,134^2 + 1}} \cong 0,122.$$

Матрица (3.9) в ортонормированной форме имеет вид

$$\Phi = \begin{bmatrix} 0,122 & -1 \\ 1 & 0,122 \end{bmatrix}.$$

Заметим, что если столбцы матрицы являются ортонормированными, то

$$\Phi^T \Phi = I.$$

Обратная матрица равна $\Phi^{-1} = \Phi^T$.

Теорема 3.5. Если K_g – действительная симметричная квадратная матрица, а $\Phi^T \Phi = I$, то существует такая ортогональная матрица, что

$$\Phi^T K_g \Phi = \Phi^{-1} K_g \Phi = D_\lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \quad (3.10)$$

где D_λ – диагональная матрица, состоящая из собственных значений матрицы K_g .

Выводы

1. Из теоремы следует, что если выполнять линейное преобразование действительной симметричной квадратной матрицы с коррелированными значениями посредством использования ядра преобразования, составленного из собственных векторов исходной матрицы, то достигается ее полная декорреляция.

2. В результате получается матрица специального вида, имеющая в качестве элементов главной диагонали положительные числа $\lambda_1, \lambda_2, \dots, \lambda_n$ и остальные элементы которой равны нулю.

Пример 3.3. Покажем, что матрицу $K_g = \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix}$ можно привести к диагональному виду и, следовательно, можно вычислить матрицу собственных значений D_λ матрицы K_g .

Решение. Подставляя данные в формулу (3.10), получаем

$$\begin{aligned} \Phi^T K_g \Phi &= \begin{bmatrix} 0,122 & 1 \\ -1 & 0,122 \end{bmatrix} \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix} \begin{bmatrix} 0,122 & -1 \\ 1 & 0,122 \end{bmatrix} \cong \begin{bmatrix} 2,024 & 0 \\ 0 & 0,375 \end{bmatrix} = \\ &= D_\lambda = \text{diag}(2,024; 0,375). \end{aligned}$$

Теорема 3.6. Сумма n собственных значений равняется сумме n диагональных элементов матрицы K_g :

$$\lambda_1 + \lambda_2 + \dots + \lambda_n = \sigma_{11}^2 + \sigma_{22}^2 + \dots + \sigma_{nn}^2.$$

Эта сумма называется следом матрицы. Для приведенного выше решения получаем

$$\lambda_1 + \lambda_2 = \sigma_{11}^2 + \sigma_{22}^2, \quad 2,024 + 0,375 \cong 0,4 + 0,2.$$

3.1.3. Преобразование подобия

В общем подходе линейное обратимое преобразование переводит вектор \mathbf{g} в некоторый вектор $\hat{\mathbf{g}}$. Результат преобразования определяется ядром \mathbf{W} :

$$\hat{\mathbf{g}} = \mathbf{W}\mathbf{g}. \quad (3.11)$$

Через $\mathbf{m}_{\hat{\mathbf{g}}}$ обозначим средний вектор в области преобразований:

$$\mathbf{m}_{\hat{\mathbf{g}}} = E\{\hat{\mathbf{g}}\},$$

где E – оператор математического ожидания.

Оценка вектора $\mathbf{m}_{\hat{\mathbf{g}}}$ вычисляется по формуле усреднения:

$$\mathbf{m}_{\hat{\mathbf{g}}} = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_m^M \hat{\mathbf{g}}_m.$$

С учетом (3.11) математическое ожидание в области преобразований

$$\mathbf{m}_{\hat{\mathbf{g}}} = E\{\hat{\mathbf{g}}\} = E\{\mathbf{W}\mathbf{g}\} = \mathbf{W}E\{\mathbf{g}\} = \mathbf{W}\mathbf{m}_g. \quad (3.12)$$

Выборочная оценка ковариационной матрицы $\text{cov}(\hat{\mathbf{g}})$ для множества векторов $\{\hat{\mathbf{g}}\}$, т. е. в области линейных обратимых преобразований, определяется как

$$\text{cov}(\hat{\mathbf{g}}) = E\{(\hat{\mathbf{g}} - \mathbf{m}_{\hat{\mathbf{g}}})(\hat{\mathbf{g}} - \mathbf{m}_{\hat{\mathbf{g}}})^T\}. \quad (3.13)$$

Используя формулы (3.11) и (3.12), выражение (3.13) запишем следующим образом:

$$\begin{aligned} \text{cov}(\hat{\mathbf{g}}) &= E\{(\mathbf{W}\mathbf{g} - \mathbf{W}\mathbf{m}_g)(\mathbf{W}\mathbf{g} - \mathbf{W}\mathbf{m}_g)^T\} = \\ &= E\{\mathbf{W}(\mathbf{g} - \mathbf{m}_g)((\mathbf{W}\mathbf{g})^T - (\mathbf{W}\mathbf{m}_g)^T)\} = \\ &= \mathbf{W}E\{(\mathbf{g} - \mathbf{m}_g)(\mathbf{g}^T\mathbf{W}^T - \mathbf{m}_g^T\mathbf{W}^T)\} = \\ &= \mathbf{W}E\{(\mathbf{g} - \mathbf{m}_g)(\mathbf{g} - \mathbf{m}_g)^T\}\mathbf{W}^T = \mathbf{W}\mathbf{K}_g\mathbf{W}^T. \end{aligned}$$

Таким образом,

$$\text{cov}(\hat{\mathbf{g}}) = \mathbf{K}_{\hat{\mathbf{g}}} = \mathbf{W}\mathbf{K}_g\mathbf{W}^T. \quad (3.14)$$

Выражение (3.14) называется преобразованием подобия.

Теорема 3.7. Ковариационная матрица $\mathbf{K}_{\hat{\mathbf{g}}}$ векторов в области преобразований $\{\hat{\mathbf{g}}\}$ будет иметь диагональный вид, если матрицу \mathbf{W} составить из транспонированных собственных векторов ковариационной матрицы \mathbf{K}_g :

$$\mathbf{K}_{\hat{g}} = \Phi^T \mathbf{K}_g \Phi = \Phi^{-1} \mathbf{K}_g \Phi = \mathbf{D}_{\hat{\lambda}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}. \quad (3.15)$$

Формула (3.15) совпадает с выражением (3.10). Собственные значения ковариационных матриц в области исходных данных \mathbf{K}_g и в области преобразований $\mathbf{K}_{\hat{g}}$ совпадают. Это же относится к собственным векторам. Так как $\mathbf{D}_{\hat{\lambda}} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, это означает, что после преобразования по (3.15) ковариационная матрица в области преобразований имеет диагональный вид и состоит из некоррелированных элементов.

Вывод. Если в выражении (3.11) вместо ядра преобразования \mathbf{W} использовать ядро Φ^T , построенное из транспонированных собственных векторов ковариационной матрицы \mathbf{K}_g исходных данных, то преобразованные данные будут некоррелированными. Для всех элементов вектора $\hat{\mathbf{g}} = (\hat{g}_1, \hat{g}_2, \dots, \hat{g}_m)$ коэффициенты корреляции $c_{ij} = 0$.

3.2. Сжатие изображений на основе статистического анализа

3.2.1. Преобразование Хотеллинга

Преобразование Хотеллинга обеспечивает процесс декорреляции коррелированных сигналов в набор некоррелированных. Пусть ковариационной матрице \mathbf{K}_g исходных данных соответствует набор собственных векторов и соответствующие им собственные значения $\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_M$. Упорядочим собственные значения по убыванию, так что

$$\lambda_j \leq \lambda_{j+1}, \text{ для } j = 1, 2, \dots, M - 1. \quad (3.16)$$

Для примера 3.2 это будет означать

$$\lambda_2 = 0,375 < \lambda_1 = 2,024.$$

Пусть матрица \mathbf{W} в выражении $\hat{\mathbf{g}} = \mathbf{W}\mathbf{g}$ составлена из транспонированных собственных векторов ковариационной матрицы \mathbf{K}_g таким способом, чтобы векторы соответствовали своим упорядоченным собственным значениям (3.16). Тогда преобразование Хотеллинга с ядром $\mathbf{W} = \Phi^T$ определяется выражением

$$\hat{\mathbf{g}} = \mathbf{W}(\mathbf{g} - \mathbf{m}_g). \quad (3.17)$$

Формулу (3.17) можно записать как

$$\begin{aligned} \hat{\mathbf{g}} &= \sum_{i=1}^M (g_i - m_i) \mathbf{W}_i = (g_1 - m_1) \mathbf{W}_1 + \\ &+ (g_2 - m_1) \mathbf{W}_2 + \dots + (g_M - m_M) \mathbf{W}_M. \end{aligned} \quad (3.18)$$

Обратное преобразование Хотеллинга определяется по формуле

$$\mathbf{g} = \mathbf{W}^T (\hat{\mathbf{g}} + \mathbf{m}_g). \quad (3.19)$$

Выражение (3.19) можно записать как

$$\mathbf{g} = (\sum_{i=1}^M \hat{g}_i \mathbf{W}_i^T) + \mathbf{m}_g = \hat{g}_1 \mathbf{W}_1^T + \hat{g}_2 \mathbf{W}_2^T + \dots + \hat{g}_M \mathbf{W}_M^T + \mathbf{m}_g. \quad (3.20)$$

Пример 3.4. Вычислить преобразование Хотеллинга вектора $\mathbf{g}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, если $\mathbf{W} = \Phi^T$, $\mathbf{m}_g = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Использовать матрицу Φ^T из примера 3.3.

$$\begin{aligned} \text{Решение. } \hat{\mathbf{g}}_1 &= \mathbf{W}(\mathbf{g}_1 - \mathbf{m}_g) = \begin{bmatrix} 0,122 & 1 \\ -1 & 0,122 \end{bmatrix} \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \\ &= \begin{bmatrix} 0,122 & 1 \\ -1 & 0,122 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,878 \\ 1,122 \end{bmatrix}. \end{aligned}$$

Используя форму (3.18), получаем

$$\begin{aligned} \hat{\mathbf{g}}_1 &= \sum_{i=1}^M (g_i - m_i) \mathbf{W}_i = (g_1 - m_1) \mathbf{W}_1 + \\ &+ (g_2 - m_2) \mathbf{W}_2 = -1 \begin{bmatrix} 0,122 \\ -1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0,122 \end{bmatrix} = \begin{bmatrix} 0,878 \\ 1,122 \end{bmatrix}. \end{aligned}$$

Пример 3.5. Восстановить вектор \mathbf{g}_1 с помощью обратного преобразования Хотеллинга. Исходные данные – из примера 3.4.

$$\text{Решение. } \mathbf{g}_1 = \mathbf{W}^T (\hat{\mathbf{g}}_1 + \mathbf{m}_g) = \begin{bmatrix} 0,122 & -1 \\ 1 & 0,122 \end{bmatrix} \begin{bmatrix} 0,878 \\ 1,122 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Используя форму (3.20), получаем

$$\begin{aligned} \mathbf{g}_1 &= \left(\sum_{i=1}^M \hat{g}_i \mathbf{W}_i^T \right) + \mathbf{m}_g = \hat{g}_1 \mathbf{W}_1^T + \hat{g}_2 \mathbf{W}_2^T + \mathbf{m}_g = \\ &= 0,878 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + 1,122 \begin{bmatrix} -1 \\ 0,122 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 0 \\ 2 \end{bmatrix}. \end{aligned}$$

Упражнение 3.1. Массив данных состоит из векторов множества $\{\mathbf{g}\}$:

$$\mathbf{g}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad \mathbf{g}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad \mathbf{g}_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}; \quad \mathbf{g}_5 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Вычислить преобразования Хотеллинга векторов: \mathbf{g}_2 , \mathbf{g}_3 , \mathbf{g}_4 , \mathbf{g}_5 .

Упражнение 3.2. Показать, что векторы множества $\{\hat{\mathbf{g}}\}$, получаемые посредством преобразования Хотеллинга, имеют нулевое математическое ожидание.

3.2.2. Описание изображений на основе анализа главных компонент

Предположим, что вместо использования всех собственных векторов ковариационной матрицы \mathbf{K}_g строится ядро преобразования \mathbf{W}_l Хотеллинга, состоящее лишь из l собственных векторов, которым отвечают l наибольших значений. В этом случае матрица преобразования имеет размерность $l \times N$. Вектор $\hat{\mathbf{g}}$ будет содержать l ненулевых координат. Восстановление исходного вектора через обратное преобразование Хотеллинга определяется соотношением

$$\mathbf{g}' = \mathbf{W}_l^T \hat{\mathbf{g}}_l + \mathbf{m}_g. \quad (3.21)$$

Ошибка восстановления, возникающая при отбрасывании $M - l$ координат, определяется как

$$\Delta \mathbf{g} = \mathbf{g} - \mathbf{g}'.$$

Среднеквадратичная ошибка, которая возникает при восстановлении \mathbf{g} , задается выражением

$$\varepsilon = E\{\|\Delta \mathbf{g}^T \Delta \mathbf{g}\|\} = \sum_{j=1}^M \lambda_j - \sum_{j=1}^l \lambda_j = \sum_{j=l+1}^M \lambda_j \quad (3.22)$$

Если в преобразовании используются все собственные векторы ковариационной матрицы, то ошибка $\varepsilon = 0$. Так как собственные значения $\lambda_j \leq \lambda_{j+1}$ упорядочены по убыванию, ошибку можно минимизировать, выбирая l собственных векторов, которым соответствуют наибольшие собственные значения. Если координата \hat{g}_k вектора $\hat{\mathbf{g}}$ не используется, то среднеквадратичная ошибка увеличивается на соответствующее значение λ_k .

Пример 3.6. Требуется реализовать компактное описание мультиспектрального изображения с использованием анализа главных компонент.

Мультиспектральное изображение состоит из наборов изображений, которые пространственно совмещены, но имеют отличающиеся значения яркостей пикселей. Такие изображения формируются мультиспектральным сканером, работающим в разных диапазонах длин волн (табл. 3.1).

На рис. 3.1 показан мультиспектральный набор, состоящий из двух изображений размером 1×5 . Вектор, описывающий один пиксель такого изображения, содержит две координаты $\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$.

При регистрации n изображений вектор, описывающий один пиксель мультиспектрального изображения, становится n -мерным. Если размер мультиспектрального изображения составляет $M \times N$, то все мультиспектральное изображение описывается $K = (M \times N)$ n -мерными векторами.

Таблица 3.1

Спецификация мультиспектрального сканера

Номер канала	Диапазон длин волн (мкм)	Назначение
1	0,45–0,52 Синий цвет	Максимальная фильтрация влаги
2	0,52–0,6 Зеленый цвет	Измерение плотности растительного покрова
3	0,63–0,69 Красный цвет	Различение формы растительности
4	0,76–0,9 Ближнее ИК-излучение	Распределение биомассы
5	1,55–1,75 Средний ИК-диапазон	Содержание влаги в почве и растительности
6	2,08–2,35 Средний ИК-диапазон	Поиск полезных ископаемых
7	10,4–12,5 Тепловое ИК-излучение	Влажность почвы и температурная карта

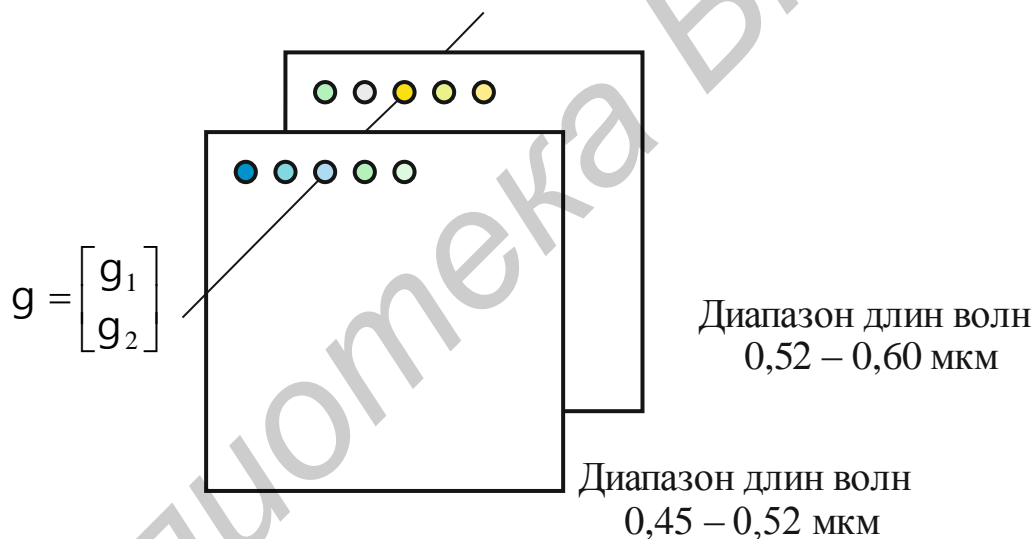


Рис. 3.1. Составляющие мультиспектрального изображения

При регистрации n изображений вектор, описывающий один пиксель мультиспектрального изображения, становится n -мерным. Если размер мультиспектрального изображения составляет $M \times N$, то все мультиспектральное изображение описывается $K = (M \times N) n$ -мерными векторами.

Пусть мультиспектральное изображение описывается пятью двумерными векторами:

$$\{g\} = \{g_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, g_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, g_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, g_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, g_5 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}\}.$$

По K векторам g рассчитывается ковариационная матрица $\text{cov}(g)$ (3.4):

$$\text{cov}(\mathbf{g}) = \mathbf{K}_g = \frac{1}{5} \sum_{k=1}^5 (\mathbf{g}_k - \mathbf{m}_g)((\mathbf{g}_k - \mathbf{m}_g)^T).$$

После подстановки исходных данных получаем

$$\mathbf{K}_g = \begin{bmatrix} 0,4 & 0,2 \\ 0,2 & 2 \end{bmatrix}.$$

По полученной ковариационной матрице \mathbf{K}_g вычисляются собственные пары $(\lambda_j, \mathbf{X}_j)$:

$$(\lambda_1, \mathbf{X}_1) = (2,024, \begin{bmatrix} 0,123 \\ 1 \end{bmatrix}); \quad (\lambda_2, \mathbf{X}_2) = (0,375, \begin{bmatrix} -8,134 \\ 1 \end{bmatrix}).$$

Для ортонормированных собственных векторов собственные пары записываются в виде

$$(\lambda_1, \mathbf{X}_1) = (2,024, \begin{bmatrix} 0,122 \\ 1 \end{bmatrix}), \quad (\lambda_2, \mathbf{X}_2) = (0,375, \begin{bmatrix} -1 \\ 0,122 \end{bmatrix}).$$

Ортонормированные собственные векторы образуют матрицу

$$\Phi = \begin{bmatrix} 0,122 & -1 \\ 1 & 0,122 \end{bmatrix}.$$

Далее строится матрица $\mathbf{W} = \Phi^T$ размером 2×2 , составленная из транспонированных собственных векторов ковариационной матрицы \mathbf{K}_g :

$$\mathbf{W} = \begin{bmatrix} 0,122 & 1 \\ -1 & 0,122 \end{bmatrix}. \quad (3.23)$$

Матрица \mathbf{W} используется в качестве ядра линейного преобразования Хотеллинга для отражения исходных векторов \mathbf{g} в векторы $\hat{\mathbf{g}}$ по формуле (3.17):

$$\hat{\mathbf{g}} = \mathbf{W}(\mathbf{g} - \mathbf{m}_g).$$

Выполняя преобразования Хотеллинга векторов $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \mathbf{g}_4$ и \mathbf{g}_5 , получим следующие значения:

$$\hat{\mathbf{g}}_1 = \begin{bmatrix} 0,878 \\ 1,122 \end{bmatrix}, \quad \hat{\mathbf{g}}_2 = \begin{bmatrix} -1 \\ -0,122 \end{bmatrix}, \quad \hat{\mathbf{g}}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\hat{\mathbf{g}}_4 = \begin{bmatrix} -2 \\ -0,244 \end{bmatrix}, \quad \hat{\mathbf{g}}_5 = \begin{bmatrix} 2 \\ -0,756 \end{bmatrix}.$$

Как видно, первой строке матрицы \mathbf{W} (3.23) соответствует собственное значение, абсолютная величина которого значительно превышает величину второго собственного значения ($2,024 \gg 0,375$). Упорядочение собственных

значений по убыванию соответствует последовательности собственных значений:

$$\begin{pmatrix} \lambda^1 & \lambda^2 \\ 2,024 & -0,375 \end{pmatrix}.$$

В нашем случае упорядочение строк матрицы \mathbf{W} (3.23) в соответствии с собственными значениями приводит к исходной же матрице \mathbf{W} .

При реализации сжатия мультиспектрального изображения (см. рис. 3.1) в два раза размерность ядра преобразования \mathbf{W} Хотеллинга уменьшается в два раза. Используем собственный вектор, соответствующий собственному значению $\lambda^1 = 2,024$.

Напомним, собственные значения ковариационной матрицы в области преобразований $\mathbf{K}_{\hat{g}}$ соответствуют дисперсиям коэффициентов преобразований $g_i, i = 1, 2, \dots, M$. Собственное значение $\lambda^1 = 2,024$ соответствует дисперсии яркости в области преобразования Хотеллинга первого составного изображения мультиспектрального изображения. Так как $\lambda^1 \gg \lambda^2$, то на первое изображение приходится 84 % – большая часть всей дисперсии мультиспектрального изображения. Следовательно, первое изображение характеризуется много большей яркостью в сравнении со вторым.

Таким образом, вместо сохранения всего набора изображений мультиспектральных данных достаточно хранить только изображения в области преобразований, характеризующиеся максимальными дисперсиями.

Усеченное ядро преобразования Хотеллинга будет иметь вид

$$\mathbf{W}_i = \mathbf{W}_1 = [0,122 \quad 1].$$

Восстановление мультиспектрального изображения \mathbf{g} через выполнение обратного преобразования Хотеллинга определяется соотношением (3.20):

$$\mathbf{g}' = \left(\sum_{i=1}^1 \hat{g}_i \mathbf{W}_i^T \right) + \mathbf{m}_g,$$

где вектор \mathbf{g}' обозначает оценку \mathbf{g} .

Оценка первого вектора, описывающего первый пиксель мультиспектрального изображения, равна

$$\mathbf{g}'_1 = \left(\sum_{i=1}^1 \hat{g}_1 \mathbf{W}_1^T \right) + \mathbf{m}_g = \hat{g}_1 \mathbf{W}_1^T + \mathbf{m}_g = 0,878 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 1,107 \\ 1,878 \end{bmatrix}.$$

Аналогично вычисляются остальные оценки составляющих мультиспектрального изображения:

$$\mathbf{g}'_2 = \left(\sum_{i=1}^1 \hat{g}_1 \mathbf{W}_1^T \right) + \mathbf{m}_g = -1 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 0,878 \\ 0 \end{bmatrix};$$

$$\mathbf{g}'_3 = (\sum_{i=1}^1 \hat{g}_1 \mathbf{W}_1^T) + \mathbf{m}_g = 0 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

$$\mathbf{g}'_4 = (\sum_{i=1}^1 \hat{g}_1 \mathbf{W}_1^T) + \mathbf{m}_g = -2 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 0,756 \\ -1 \end{bmatrix};$$

$$\mathbf{g}'_5 = (\sum_{i=1}^1 \hat{g}_5 \mathbf{W}_1^T) + \mathbf{m}_g = 2,122 \begin{bmatrix} 0,122 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cong \begin{bmatrix} 1,258 \\ 3,122 \end{bmatrix}.$$

Ошибки (векторы ошибок $\Delta \mathbf{g}$), возникающие при отбрасывании $M - L$ координат, равны

$$\Delta \mathbf{g}_1 = \mathbf{g}_1 - \mathbf{g}'_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1,107 \\ 1,878 \end{bmatrix} = \begin{bmatrix} -1,107 \\ 0,122 \end{bmatrix};$$

$$\Delta \mathbf{g}_2 = \mathbf{g}_2 - \mathbf{g}'_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0,878 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,122 \\ 0 \end{bmatrix};$$

$$\Delta \mathbf{g}_3 = \mathbf{g}_3 - \mathbf{g}'_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\Delta \mathbf{g}_4 = \mathbf{g}_4 - \mathbf{g}'_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0,756 \\ -1 \end{bmatrix} = \begin{bmatrix} 0,244 \\ 0 \end{bmatrix};$$

$$\Delta \mathbf{g}_5 = \mathbf{g}_5 - \mathbf{g}'_5 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1,258 \\ 3,122 \end{bmatrix} = \begin{bmatrix} 0,742 \\ -0,122 \end{bmatrix}.$$

Далее вычисляются соответствующие среднеквадратичные ошибки (3.22), возникающие при отбрасывании $M - L$ координат:

$$\Delta \mathbf{g}_1^T \Delta \mathbf{g}_1 = [-1,107 \quad 0,122] \begin{bmatrix} -1,107 \\ 0,122 \end{bmatrix} = 1,241;$$

$$\Delta \mathbf{g}_2^T \Delta \mathbf{g}_2 = [0,122 \quad 0] \begin{bmatrix} 0,122 \\ 0 \end{bmatrix} = 0,0148;$$

$$\Delta \mathbf{g}_3^T \Delta \mathbf{g}_3 = 0;$$

$$\Delta \mathbf{g}_4^T \Delta \mathbf{g}_4 = [0,244 \quad 0] \begin{bmatrix} 0,244 \\ 0 \end{bmatrix} = 0,059;$$

$$\Delta \mathbf{g}_5^T \Delta \mathbf{g}_5 = [0,742 \quad -0,122] \begin{bmatrix} 0,742 \\ -0,122 \end{bmatrix} = 0,565.$$

Суммарная среднеквадратичная ошибка восстановления мультиспектрального изображения при $l = 1$ равна

$$\begin{aligned}\varepsilon &= \sum_{j=l+1}^M \lambda_j = \lambda_2 = E\{\|\Delta \mathbf{g}^T \Delta \mathbf{g}\|\} = \frac{1}{K} \sum_{i=1}^K \Delta g_i^T \Delta g_i = \\ &= \frac{1}{5} (1,241 + 0,014 + 0,059 + 0,565) = 0,375.\end{aligned}$$

Как видно, ошибка точно соответствует ранее вычисленному собственному значению λ_2 .

Вывод. Для компактного описания (передачи) мультиспектрального изображения достаточно хранить только изображения в области преобразований с максимальными значениями λ_j , вектор математического ожидания \mathbf{m}_g и усеченную матрицу \mathbf{W}_l .

Упражнение 3.3. Реализовать компактное описание мультиспектрального изображения с использованием анализа главных компонент. Изображение описывается пятью двумерными векторами:

$$\{\mathbf{g}\} = \left\{ \mathbf{g}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mathbf{g}_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \mathbf{g}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{g}_4 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mathbf{g}_5 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Упражнение 3.4. Сжать мультиспектральное изображение с использованием анализа главных компонент. Изображение описывается шестью двумерными векторами:

$$\{\mathbf{g}\} = \left\{ \mathbf{g}_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \mathbf{g}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{g}_3 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mathbf{g}_4 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{g}_5 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{g}_6 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Упражнение 3.5. Сжать мультиспектральное изображение с использованием анализа главных компонент. Изображение описывается шестью трехмерными векторами:

$$\{\mathbf{g}\} = \left\{ \mathbf{g}_1 = \begin{bmatrix} 5 \\ 2 \\ 2 \end{bmatrix}, \mathbf{g}_2 = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}, \mathbf{g}_3 = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}, \mathbf{g}_4 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \mathbf{g}_5 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \mathbf{g}_6 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

3.3. Сжатие изображений с применением дискретных ортогональных преобразований

Среднеквадратичная ошибка восстановленного изображения посредством преобразования Хотеллинга будет минимальной по сравнению с ошибками при всех других возможных дискретных ортогональных преобразованиях. В этом смысле кодирование изображения с помощью преобразования Хотеллинга достигает предельной эффективности, являясь оптимальным. При этом на статистический анализ (вычисления ковариационной матрицы \mathbf{K}_g) обрабатываемых данных размером $N \times N$ потребуется выполнить N^2 умножений. Это число увеличится в K раз в случае необходимости обработки массива величиной $K(N^2)$ пикселей, где K – это число обрабатываемых изображений (кадров). Кроме того, на выполнение прямого $\hat{\mathbf{g}} = \mathbf{W}(\mathbf{g} - \mathbf{m}_g)$ и обратного $\mathbf{g} = \mathbf{W}^T(\hat{\mathbf{g}} + \mathbf{m}_g)$ преобразований Хотеллинга потребуется $2N^4$ умножений. Особенно чрезмерно

большим, требующим значительных временных затрат, является вычисление собственных значений и собственных векторов ковариационной матрицы K_g по мере увеличения значения N . Так как разным изображениям соответствуют различные ковариационные матрицы, процесс преобразования существенно усложняется. Чтобы уменьшить вычислительную сложность обработки изображений, используют идею разбиения массива данных на меньшие размеры (фрагменты), а затем для каждого из фрагментов производится процедура преобразования. Экспериментальные исследования показали, что корреляция между двумя пикселями быстро стремится к нулю с ростом расстояния между ними. Поэтому можно кодировать изображение, разбив его на отдельные равные блоки $L_0, L_1, \dots, L_i, \dots$. Вычисление ковариационных матриц и преобразование производятся в пределах каждого L_i блока.

Эффективный анализ и описание изображений можно реализовать в пространстве быстрых линейных ортогональных преобразований на основе спектральных методов. Для этого используются следующие двумерные дискретные преобразования:

- преобразование Фурье;
- преобразование Уолша – Адамара;
- косинусное преобразование;
- преобразование Хартли, Хаара и др.

Сущность спектральных методов заключается в том, что кодируется и передается (хранится) не само изображение, а значения трансформант, получающихся при ортогональном преобразовании этого изображения. В процессе преобразования изображения, имеющего сильные корреляционные связи между смежными элементами, происходит процесс декорреляции, так что значения коэффициентов преобразования оказываются практически некоррелированными. В отличие от исходного изображения, для которого характерно в среднем равномерное распределение энергии между элементами, распределение энергии между спектральными коэффициентами резко неравномерно. Основная доля энергии приходится на коэффициенты с малыми номерами (низкие пространственные частоты), и лишь небольшая ее часть – на прочие. Коэффициенты, имеющие малую амплитуду, опускаются либо квантуются на малое число уровней, что позволяет использовать меньшее число разрядов кода для передачи или хранения. При приеме (считывании) выполняются обратные преобразования. Предварительно по принятому коду восстанавливается матрица трансформант, а затем путем обратного ортогонального преобразования изображение восстанавливается. Нелинейное квантование коэффициентов преобразования позволяет существенно сократить объем передаваемой (сохраняемой) информации без заметного ухудшения качества передаваемого изображения при его восстановлении.

В общем случае кодирование посредством линейных дискретных ортогональных преобразований реализуется путем двух последовательных операций. Первая операция состоит в линейном преобразовании статистически зависимых

элементов изображения. С помощью второй операции осуществляется кодирование определенным способом выбранных коэффициентов преобразования. Кодер выполняет четыре операции.

1. Разбиение изображения размером $N \times N$ на $\left(\frac{N}{n}\right)^2$ блоков размером $n \times n$.

2. Линейное ортогональное преобразование с целью декорреляции значений элементов в каждом блоке и уплотнения как можно большего количества информации в наименьшее число трансформант.

3. На этапе квантования те коэффициенты, которые несут минимальную информацию, удаляются или же квантуются грубо (они дают наименьший вклад в качество восстанавливаемого изображения). Затем идет операция помехоустойчивого кодирования. Декодер выполняет обратную последовательность операций (за исключением квантования).

4. На заключительном этапе осуществляется кодирование квантованных коэффициентов, с помощью неравномерных (эффективных) и помехоустойчивых кодов.

Все или некоторые из указанных этапов могут быть адаптированы к содержанию блока, т. е. к локальным статистическим характеристикам изображения; такой вариант называют адаптивным трансформационным кодированием. В противном случае говорят о неадаптивном (пороговом) трансформационном кодировании. Декодер выполняет обратную последовательность операций (за исключением квантования).

Большинство из названных преобразований позволяет снизить вычислительную сложность до величины порядка $N \log_2 N$. Преобразование изображений на основе линейных ортогональных преобразований вместо преобразования Хотеллинга приводит к определенным качественным потерям. Эти потери можно оценить, сравнивая их с ошибками, которые получаются при анализе на основе преобразования Хотеллинга.

Литература

1. Оппенгейм, А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер. – М. : Техносфера, 2006.
2. Теория прикладного кодирования: учеб. пособие. В 2 т. / В. К. Конопелько [и др.]. – Минск : БГУИР, 2004.
3. Айфичер, Э. С. Цифровая обработка сигналов : практический подход / Э. С. Айфичер, Б. У. Джервис; пер. с англ. – М. : Издательский дом «Вильямс», 2008.
4. Лайонс, Р. Цифровая обработка сигналов / Р. Лайонс. – М. : Бином-Пресс, 2006.
5. Смит, С. Цифровая обработка сигналов. Практическое руководство для инженеров и научных работников / С. Смит; пер. с англ. – М. : Додека-XXI, 2008.
6. Сергиенко, А. Б. Цифровая обработка сигналов / А. Б. Сергиенко. – СПб. : Питер, 2003.
7. Лосев, В. В. Микропроцессорные устройства обработки информации. Алгоритмы цифровой обработки : учеб. пособие для вузов / В. В. Лосев. – Минск : Выш. школа, 1990.
8. Яне, Б. Цифровая обработка изображений / Б. Яне. – М. : Техносфера 2007.
9. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005.
10. Гонсалес, Р. Цифровая обработка изображений в среде Matlab / Р. Гонсалес, Р. Вудс, С. Эддинс. – М. : Техносфера, 2006.
11. Прэтт, У. Цифровая обработка изображений. В 2 т. / У. Прэтт. – М. : Мир, 1982.
12. Шапиро, Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман. – М. : Бином, 2006.
13. Сэломон, Д. Сжатие данных, изображений и звука : учеб. пособие для вузов / Д. Сэломон. – М. : Техносфера, 2004.

Учебное издание

Митюхин Анатолий Иванович

**ЦИФРОВАЯ ОБРАБОТКА РЕЧИ
И АНАЛИЗ ИЗОБРАЖЕНИЙ**

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

Редактор *М. А. Зайцева*

Корректор *Е. Н. Батурчик*

Компьютерная правка, оригинал-макет *В. М. Задоя*

Подписано в печать 24.03.2016. Формат 60x84 1/16. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 4,3. Уч.-изд. л. 4,0. Тираж 75 экз. Заказ 150.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».

Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,

№2/113 от 07.04.2014, №3/615 от 07.04.2014.

ЛП №02330/264 от 14.04.2014.

220013, Минск, П. Бровки, 6