Н.К. Шавенько

Основы теории информации и кодирования

Рекомендовано УМО по образованию в области геодезии и фотограмметрии
в качестве учебного пособия для студентов высших учебных заведений,
обучающихся по специальностям направления подготовки
120200 «Фотограмметрия и дистанционное зондирование»

Москва

Издательство МИИГАиК

2012

УДК 621.391.
ББК

Рецензенты:
с.н.с., к.т.н. И.В. Починок – зав. лабораторией мобильных и встраиваемых программных систем НИВЦ МГУ им. Ломоносова;
доц., к.т.н. А.Ф.Стеценко – зав кафедрой аэрокосмических съёмок МИИГАиК.

Шавенько Н.К.
Основы теории информации и кодирования. Учебное пособие. – М,: Изд-во МИИГАиК, 2012. – 125 с.

Содержит краткие теоретические положения курса «Основы теории информации и кодирования», а именно: основы теории информации, основы теории кодирования и передачи информации по каналам связи. Кроме этого пособие включает в себя лабораторный практикум, предназначенный для закрепления теоретического материала и представленный в виде подробного описания выполнения лабораторных работ. Для удобства пользования пособие содержит краткие справочные сведения по теории вероятностей, которые крайне необходимы при рассмотрении основных положений теории информации, теории кодирования и передачи информации по каналам связи.

Для студентов высших учебных заведений, обучающихся по направлению 120200 «Фотограмметрия и дистанционное зондирование».

УДК 621.391.
ББК

Введение.

В настоящем учебном пособии рассматриваются основы теории информации, основы теории кодирования и передачи информации по каналам связи, которые служат теоретическим базисом для специализированных курсов, связанных с автоматической обработкой информации и её передачи.

Содержание пособия соответствует программе курса «Основы теории информации и кодирования» для студентов высших учебных заведений, обучающихся по специальностям направления 120200 «Фотограмметрия и дистанционное зондирование».

Пособие состоит из четырёх теоретических глав, лабораторного практикума и приложения. В первой главе содержатся общие сведения по основам теории информации. Во второй главе рассматриваются основы теории кодирования, третья глава посвящена краткому изложению теории передачи информации по каналам связи и в четвёртой главе приведены примеры использования информационных моделей при анализе систем автоматической обработки изображений и их основных узлов.

. Изложение теоретических глав написано в предположении, что студенты знакомы с основными понятиями из курса высшей математики.

Лабораторный практикум включает в себя две лабораторные работы по основам теории информации, работу по основы теории кодирования и лабораторную работу по определению информационных характеристик оптических изображений и фотоизображений. Описание лабораторных работ выполнено из расчета на то, что к моменту их выполнения студентами изучен соответствующий теоретический материал, а также приобретены навыки работы на персональном компьютере в программном продукте «MathCad». В конце каждой работы приведены контрольные вопросы.

Приложение содержит краткий справочный материал по основам теории вероятностей, знание, которого необходимо при изучении основного теоретического материала представленного в пособии.

При подготовке пособия использованы теоретические положения, отраженные в научных статьях, в учебной литературе и монографиях.

Глава 1. Основы теории информации.

1.1. Информация. Общие понятия

При исследовании или наблюдении любого физического процесса, объекта или явления можно ввести понятие информации. Существует достаточно много определений понятия информация, существенно различающихся в зависимости от сферы деятельности в применении к которой даётся конкретное определение этого понятия. Такое положение, при котором существует множество различных определений и отсутствует единое общепринятое и точное определение этого понятия, вытекает из того, что понятие информации, наряду с такими фундаментальными понятиями как материя и энергия в физике или как множество или точка в математике, является первичным понятием и поэтому не может иметь четкого формализованного определения, т.е. не может быть определено через более простые известные объекты, имеющие четкие определения. Поэтому при определении основных фундаментальных первичных понятий используют интуитивный подход, определяя понятие через совокупность присущих ему свойств. Действительно, исходя из практического опыта, интуитивно, можно представить информацию как совокупность содержательных сведений, заключенных в том или ином объекте или явлении, и перечислить ее основные свойства, тем самым интуитивно, исходя из этих свойств, её определить.

Основные свойства информации:

1. Информация приносит знания об окружающем мире, которых в рассматриваемой точке пространства не было до ее получения.
2. Информация не материя, а свойство организованной материи. Информация такое же неотъемлемое свойство материи как масса и энергия, однако, в отличие от них она не подчиняется законам сохранения, подобным законам сохранения массы и энергии.
3. Информация не материальна, но она проявляется в виде материальных носителей — символов и сигналов. Причем символы — это реальные различимые получателем материальные объекты (буквы, цифры, изображения), а сигналы — это динамические процессы, т.е. изменяющиеся во времени или пространстве значения любой физической величины.
4. Информация может быть заключена как в самих символах, так ив их взаимном расположении (например, символы Т, Р, С, О могут принести информацию: торс, сорт, трос, рост и т.д.).

Символы и сигналы несут информацию только для получателя, способного их распознать, т.е. поставить в соответствие принятым символам и сигналам объекты реального мира и их отношения.

Исходя из вышеперечисленных свойств, информацию можно рассматривать как сведения (знания), полученные в результате моделирования (описания) реального мира или его исследуемой части, являющиеся объектом некоторых операций: передачи, распределения, преобразования, хранения или непосредственного использования.

Дальнейшее изложение требует определение некоторых устоявшихся понятий и терминов теории информации, которых необходимо строго придерживаться в дальнейшем.

Информация содержится в *сообщениях*, которые генерируются *источниками сообщений*.

Источник сообщений — любой процесс, объект или явление, который обладают способностью изменять свое состояние во времени или в пространстве.

Символ источника сообщений - это любое мгновенное состояние источника сообщений.

Сообщение — любая конечная последовательность символов.

Алфавит источника сообщений — все множество различных символов, генерируемых источником сообщений.

Объем алфавита источника сообщений — число различных символов, генерируемых источником сообщений.

Дискретный источник сообщений — источник сообщений, обладающий конечным алфавитом.

Непрерывный источник сообщений — источник сообщений, обладающий бесконечным алфавитом.

1.2.Измерение информации

Для практического использования понятия информации необходимо научиться ее измерять. Это можно сделать по аналогии с методикой численного измерения иных фундаментальных понятий, таких как материя (масса), энергия или пространство, для этого требуется установить, что принимается за меру количественной оценки информации и что принимать за единицу измерения этой меры. Под мерой количественной оценки понимают некоторое явление или объект, которые однозначно (пропорционально) связаны с определяемым первичным понятием и которые могут характеризовать количественное содержание этого понятия.

Таблица 1.1 иллюстрирует общепринятую методику измерения первичных понятий, таких как масса, пространство и т. п.

Таблица 1.1

Первичное понятие.	Мера количественной оценки.	Ед. измер. меры кол. оценки.
Материя.	Вес.	Грамм, тонна и т.д.
Пространство.	Расстояние.	Метр, километр и т.д.
Информация	?	?

Традиционно сложилось три основных подхода к выбору меры количественной оценки информации.

1. *Структурный подход,* при котором количественная оценка информации о событии оценивается путем определения объективной возможности этого события, входящего в некоторую полную группу событий.

2. *Статистический подход,* при котором количественная оценка информации о принятом сообщении производится на основе меры неопределенности, снимаемой с исследуемого информационного процесса (события) при получении данного сообщения.

3. *Семантический подход,* который в основном учитывает ценность полученной информации с точки зрения конкретного получателя этой информации.

Очевидно, что, для точного, технического, объективного использования, семантический подход не приемлем, так как он сугубо субъективен и не может дать общепринятой количественной меры информации, хотя этот подход и может быть использован в сфере гуманитарных и общественных наук.

Применительно к точным и техническим наукам для определения меры количественной оценки информации используют структурный и статистический подходы.

Выбор критерия для количественной оценки информации, независимо от выбранного полхода, должен удовлетворять условиям, вытекающим из практического опыта:

— сообщению большей длины (при одном и том же объеме алфавита) соответствует большее количество информации;

— большее количество информации содержится в тех сообщениях (одинаковой длины), которые составлены из символов большего алфавита;

— символы в сообщении могут появляться с различными вероятностями и могут быть статистически зависимыми.

Учитывая это, меру количественной оценки информации можно ввести исходя из следующих соображений. Предположим, что какое-то событие имеет m равновероятных исходов, например, появление какого-либо символа из алфавита, содержащего m таких символов. Измерить количество информации, содержащееся в сообщении из n таких символов можно, определив число N всех возможных сообщений, которые могут быть составлены из символов этого алфавита. Если сообщение формируется из одного символа, то $N=m$, если из двух, то $N = m \cdot m = m^2$, если из n символов, то $N = m^n$. Полученную меру количественной оценки информации можно понимать как меру неопределенности получения конкретного заданного сообщения, состоящего из n символов алфавита. Однако эта мера количественной оценки информации не совсем удобна. Действительно, при $m = 1$ (т.е. алфавит состоит из одного символа) неопределенности не существует и появление этого символа не несет никакой информации, однако значение N в этом случае ($N=1^n$) не обращается в нуль. Кроме этого, из практических соображений, целесообразно считать, что количество информации, полученное от двух независимых источников, равно сумме количеств информации, получаемых от каждого источника, а предлагаемая мера количества информации дает в этом случае произведение $N = m^{n_1+n_2} + m^{n_1} \cdot m^{n_2} = N_1 \cdot N_2$, где N_1, N_2 — число возможных сообщений от двух источников сообщений. Эти неудобства легко преодолимы, если в качестве меры количественной оценки информации (I) взять логарифм по какому-либо основанию от общего числа возможных сообщений (N)

$$I = \log N, \tag{1.1}$$

или логарифм вероятности появления конкретного сообщения (P)

$$I = \log N = \log \frac{1}{P} = -\log P, \tag{1.2}$$

при условии, что все сообщения равновероятны, т.е.

$$P = \frac{1}{N}.$$

Определенная по формулам (1.1) и (1.2) мера количественной оценки информации (I) называется *количеством информации.* Количество информации как мера количественной оценки информации может быть вписана в соответствующую графу таблицы 1.1.

В случае если сообщения не равновероятны, мера неопределенности будет зависеть не только от общего числа возможных сообщений, но и от распределения вероятности между возможными сообщениями.

В общем случае, при наличии шумов, понятие количества информации может быть определено из следующих соображений.

Если поступило сообщение о событии, априорная вероятность, которого равна P_1 (P_1 характеризует состояние системы до получения сообщения, т.е. до опыта), а после приема сообщения апостериорная вероятность этого события стала для получателя P_2 (P_2 характеризует состояние системы после получения сообщения), то прирост количества информации *(I),* связанный с приемом сообщения о событии, определяется выражением

$$I = \log \frac{P_2}{P_1}. \qquad (1.3)$$

Это выражение часто называют основным соотношением теории информации.

В частном случае, когда шумы при передаче и приеме сообщения отсутствуют, событие после приема сообщения о нем становится достоверным, т.е. $P_2 = 1$ и выражение (1.3) принимает вид:

$$I = -\log P_1. \qquad (1.4)$$

Таким образом, количество информации, содержащееся в сообщении, зависит от вероятности события до приема сообщения (P_1), и чем меньше эта вероятность, т.е. чем больше неопределенность исхода, тем больше количество информации о нем получается в результате приема сообщения. Поскольку $P \leq 1$, то определяемое формулой (1.4) количество информации всегда положительно. Следовательно, приём какого-либо сообщения никоим образом не может уменьшить количество уже имеющейся информации.

Единицы измерения количества информации определяются выбором основания логарифмов в выражениях (1.1) - (1.4). Если основание логарифмов берется равным 2, то получаем единицу количества информации *бит* (от английского *binary digit,* - двоичное число). Таким образом, один *бит* это количество информации, получаемое при приеме одного из двух равновероятных символов сообщения.

Если основание логарифмов равно 10, то единица количества информации носит название *дит,* а если используются натуральные логарифмы (по основанию *e*) – то *нат.*

Использование выше указанных формул часто бывает затруднено, поэтому на их основе получают иные выражения для определения количества информации, которые более просты и практичны, однако применимы лишь к конкретным видам сообщений.

1.3.Структурное (комбинаторное) определение количества информации (по Хартли).

Данное определение количества информации применимо лишь к дискретным сообщениям, причем таким, у которых символы равновероятны и взаимно независимы. Количество информации, содержащееся в такого рода сообщениях определяют из следующих соображений.

Пусть дан источник дискретных сообщений $A = (a_1, a_2, \ldots a_m)$, объем алфавита которого равен *m*. Предположим, что каждое сообщение включает в себя *n* символов, при этом сообщения различаются либо набором символов, либо их размещением. Число различных сообщений N_0, состоящих из *n* символов, будет $N_0 = m^n$. Предположим, что все сообщения равновероятны и одинакова ценность этих сообщений.

Тогда легко подсчитать количество информации, которое несет каждое сообщение.

Вероятность появления каждого такого сообщения (P_n) может быть легко найдена:

$$P_n = \frac{1}{N_0} = m^{-n}.$$

И, следовательно, количество информации в одном сообщении (I_n), в соответствии с (1.4), равно:

$$I_n = -\log_2 m^{-n} = n \cdot \log_2 m \text{ (бит).} \qquad (1.5)$$

Эту формулу предложил Р.Хартли в 1928 г., и она носит его имя. Разделив I_n на количество символов в сообщении (*n*), получим значение среднего количества информации (I_1), приходящееся на один символ сообщения:

$$I_1 = \log_2 m = -\log_2 P_m \text{ (бит / символ),} \qquad (1.6)$$

где P_m - вероятность появления одного символа сообщения.

Из соотношений (1.5) и (1.6) вытекают важные свойства дискретных сообщений, символы которых равновероятны и взаимно независимы.

1. Количество информации в сообщении пропорционально полному числу символов в нем – *n* и логарифму объема алфавита- *m*.
2. Среднее количество информации, приходящееся на один символ, зависит только от *m* – объема алфавита.

В реальных дискретных сообщениях символы часто появляются с различными вероятностями и, более того, часто существуют статистическая связь между символами, характеризующаяся условной вероятностью $P(a_i / a_j)$, которая равна вероятности появления символа a_i после символа a_j. Например, в тексте на русском языке вероятность появления различных символов (букв) различна. В среднем, в тексте из 1000 букв буква О появляется 110 раз, Е – 87, А – 75, Т – 65, Н – 65, С – 55, кроме того, существуют статистические связи между буквами, скажем, после гласных букв не может появиться Ь или Ъ.

Исходя из этого, применение формулы вычисления количества информации по Хартли (1.5) и (1.6) не всегда корректно.

1.4.Статистическое определение количества информации (по Шеннону).

Этот подход к определению количества информации в сообщениях, учитывающий не равновероятное появление символов сообщения и их статистическую связь, был предложен К.Шенноном в 1946 г.

Рассмотрение этого метода удобно начать с определения количества информации в дискретных сообщениях, символы которых появляются не равновероятно, однако статистическая связь между символами отсутствует.

Пусть, как и ранее, дан источник дискретных сообщений $A = (a_1, a_2, \ldots a_m)$ с объемом алфавита равным *m,* который генерирует сообщение, состоящее из *n* символов. Допустим, что в этом сообщении символ a_1 встречается n_1 раз, символ $a_2 - n_2$ раз и так далее вплоть до символа a_m, который встречается n_m раз, причем очевидно, что

$$n_1 + n_2 + \ldots + n_m = n$$

При приеме одного символа a_1, как следует из (1.4), получаем количество информации I^0_{a1}:

$$I^0_{a1} = -\log_2 P_{a1},$$

где P_{a1} - априорная вероятность появления символа a_1.

А количество информации I_{a1}, содержащееся в n_1 взаимно независимых символах a_1, будет равно:

$$I_{a1} = -n_1 \cdot \log_2 P_{a1}.$$

Аналогично, в n_2 символах a_2 содержится количество информации I_{a2}:

$$I_{a2} = -n_2 \cdot \log_2 P_{a2},$$

и так далее вплоть до

$$I_{am} = -n_m \cdot \log_2 P_{am}.$$

Очевидно, что полное количество информации (I_n), содержащееся в сообщении из *n* символов, равно сумме количеств информации содержащихся во всех *m* символах алфавита.

$$I_n = I_{a1} + I_{a2} + \ldots + I_{am} = -(n_1 \cdot \log_2 P_{a1} + n_2 \cdot \log_2 P_{a2} + \ldots + n_m \cdot \log_2 P_{am}) = -\sum_{i=1}^{m} n_i \cdot \log_2 P_{ai} \text{ (бит).}$$

Разделив и умножив это выражение на *n* ($n \neq 0$), приведем это выражение к виду:

$$I_n = -n \cdot \sum_{i=1}^{m} \frac{n_i}{n} \cdot \log_2 P_{ai} \text{ (бит)}$$

Ясно, что отношение $\frac{n_i}{n}$ – это априорная вероятность появления *i*-го символа. Таким образом, при достаточно большом *n*, имеем: $\frac{n_i}{n} = P_{ai}$, причем $\sum_{i=1}^{m} P_{ai} = 1$, как сумма вероятностей полной группы событий.

Окончательно получим:

$$I_n = -n \cdot \sum_{i=1}^{m} P_{a_i} \cdot \log_2 P_{a_i} \quad \text{(бит)} \tag{1.7}$$

При этом среднее количество информации, приходящееся на один символ (H), будет равно:

$$H = \frac{I_n}{n} = -\sum_{i=1}^{m} P_{a_i} \cdot \log_2 P_{a_i} \quad \left(\frac{бит}{символ}\right). \tag{1.8}$$

Определенная таким образом величина H называется *энтропией,* а формула (17) известна как формула Шеннона для энтропии источника дискретных сообщений. *Энтропия* определяет среднее количество информации, приходящееся на один символ дискретного сообщения.

В общем случае, символы, входящие в сообщения, могут появляться не только с различной вероятностью, но и быть статистически зависимыми. Статистическая зависимость может быть выражена условной вероятностью появления одного символа после другого.

Чтобы учесть статистические связи между символами, входящими в сообщение, вводят понятие *условной энтропии.*

Условная энтропия (H_k) определяется выражением

$$H_k = -\sum_{i=1}^{m} P_{a_i} \cdot \sum_{j=1}^{m} P\left(a_j / a_i\right) \cdot \log_2 P\left(a_j / a_i\right) \quad \left(\frac{бит}{символ}\right), \tag{1.9}$$

где $P\left(a_j / a_i\right)$ – условная вероятность появления символа a_j после символа a_i. Количество информации $\left(I_k\right)$, содержащееся в такого рода сообщении длиной *n* символов, равно:

$$I_k = n \cdot H_k = -n \sum_{i=1}^{m} P_{a_i} \cdot \sum_{j=1}^{m} P\left(a_j / a_i\right) \cdot \log_2 P\left(a_j / a_i\right) \quad \textit{(бит)} \tag{1.10}$$

1.5.Свойства функции энтропии источника дискретных сообщений.

Свойства функции энтропии можно наглядно продемонстрировать на примере источника дискретных сообщений A=($a1$, $a2$) с объемом алфавита m равного 2, т.е. m=2. В этом случае справедливо $P_{a1} + P_{a2} = 1$, и выражение (1.8) может быть записано в виде:

$$H = -P_{a1} \cdot \log_2 P_{a1} - P_{a2} \cdot \log_2 P_{a2} = -P_{a1} \cdot \log_2 P_{a1} - \left(1 - P_{a1}\right) \cdot \log_2\left(1 - P_{a1}\right) \quad \text{(бит/символ).}$$

График этой функции имеет вид, представленный на рис. 1.1.

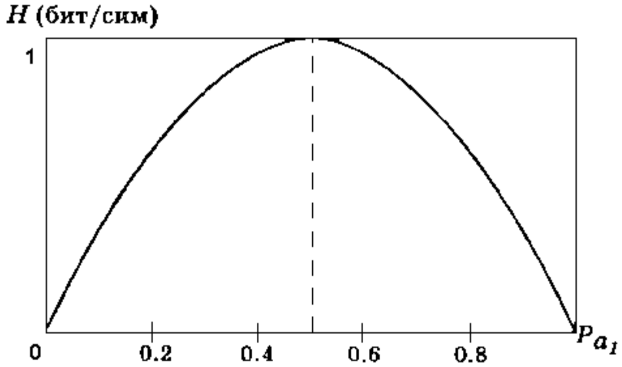


Рис.1.1. График функции энтропии.

Из графика видно, что обращение вероятности появления одного из возможных символов в 0 или 1 вносит полную определенность, энтропия обращается в 0 и сообщение о приёме такого символа не содержит в себе никакой информации.

При $P_{a1} = P_{a2} = 0{,}5$ получение конкретного символа наиболее неопределено и количество информации, содержащееся в поступившем символе, максимально.

Анализ формулы (1.8) и графика (Рис. 1.1) позволяет сформулировать основные свойства функции энтропии.

1 Энтропия источника дискретных сообщений есть величина вещественная, ограниченная и положительная.

2 Энтропия равна 0, если с вероятностью, равной единице, всегда выбирается один и тот же символ.

3 Энтропия максимальна, если все символы источника сообщений появляются независимо и равновероятно.

Интересно отметить, что сравнение выражений (1.5), (1.7) и (1.10) показывает, что формула Хартли является частным случаем формулы Шеннона при условии независимости и равновероятности появления символов в сообщении, а формула Шеннона, в свою очередь, является частным случаем условной энтропии при условии, что символы сообщения независимы. Действительно, из (1.10) следует, что количество информации (I_k),содержащееся в сообщении, состоящем из *n* неравновероятных и взаимно зависимых символов определяется выражением:

$$I_k = -n\sum_{i=1}^{m} P_a \cdot \sum_{j=1}^{m} P\left(a_j / a_i\right)\cdot 1 \quad {}_2 oP\left(a_j / a_i\right).$$

Если символы сообщения взаимно независимы, то $P\left(a_j/a_i\right) = P(a_j)$ и $\sum_{j=1}^{m} P(a_j) = 1$,

следовательно, это выражение преобразуется к виду:

$$I_k = -n\sum_{i=1}^{m} P_{a_i} \cdot \sum_{j=1}^{m} P\left(a_j/a_i\right)\cdot \log_2 P\left(a_j/a_i\right) = -n\sum_{j=1}^{m} P(a_j)\cdot \log_2 P(a_j) = I_n$$

Последнее выражение соответствует формуле Шеннона.

В случае равновероятного появления символов сообщения $P(a_j) = \frac{1}{m}$, при j= 1, 2,…, m.

(m – объём алфавита) и выше приведённое выражение для формулы Шеннона после соответствующего преобразования примет вид:

$$I_n = -n\sum_{j=1}^{m} P(a_j)\cdot \log_2 P(a_j) = -n\sum_{j=1}^{m} \frac{1}{m}\cdot \log_2 \frac{1}{m} = -n\cdot \log_2 \frac{1}{m} = n\cdot \log_2 m,$$

который соответствует формуле Хартли.

Таким образом, количество информации, определяемое по Хартли, т.е. при допущении полной независимости и равной вероятности появления отдельных символов сообщения, определяет максимально возможное количество информации в сообщении заданной длины (n).

При неравной вероятности появления символов (формула Шеннона) количество информации, содержащееся в сообщении заданной длины (n), снижается. Другим фактором, снижающим энтропию, а, следовательно, и количество информации в сообщении заданной длины (n), является наличие статистической зависимости между символами – корреляции.

Из-за корреляционных связей между символами и неравновероятного их появления количество информации в реальных сообщениях падает. Количественно эти потери информации характеризуются коэффициентом избыточности (R)

$$R = \frac{H_{max} - H}{H_{max}} = 1 - \frac{H}{H_{max}} = 1 - \frac{H}{\log_2 m}, \tag{1.11}$$

где H_{max} — максимальное количество информации, которое может содержать один символ сообщения, определяемое по формуле (1.6);

H — среднее количество информации, которое переносит один символ в реальных сообщениях;

m — число символов в алфавите источника сообщений (объём алфавита).

Избыточность $(R \neq 0)$ говорит о том, что число символов в сообщении больше, чем это требовалось бы при полном их использовании, т.е. при условии, что символы появляются равновероятно и взаимно независимо.

Интересно отметить, что, для европейских языков избыточность составляет не менее 0.5.

1.6 Информационная емкость дискретного сообщения.

При описании информационных процессов часто пользуются следующими информационными характеристиками.

Производительность источника сообщений (*P*) — среднее количество информации, генерируемое источником в единицу времени

$$P = I/t$$

Единицей измерения производительности источника сообщений является *бот*:

$$1 бот = \frac{бит}{сек}.$$

Информационная емкость сообщения (*R*) — это среднее количество информации, содержащееся в сообщении единичной длительности:

$$R = I/t,$$

где *I* — количество информации, содержащееся в сообщении;

t – длительность сообщения.

Скорость создания (генерации) информации (*Q*) — это среднее количество информации, генерируемое источником сообщений за единицу времени

$$Q = W \cdot H,$$

где *W* – скорость передачи символов (символ /сек);

H–среднее количество информации приходящееся на один символ (бит /символ).

Единицей измерения скорости генерации информации и информационной емкости сообщения служит бот.

В качестве примера, определим информационную емкость простейшего дискретного сообщения, состоящего из совокупности импульсов и пауз, причем длительность импульсов и пауз одинакова, а амплитуда импульсов постоянна, т.е. сообщение строится на использовании лишь двух символов (его объём алфавит равен 2).

Очевидно, если избыточность отсутствует, то на каждый символ такого сообщения приходится одна двоичная единица информации (1 бит), а на все сообщение, состоящее из *n* символов, приходится *n* бит информации.

Пусть длительность импульса и паузы одинакова и равна $t_и$, тогда длительность всего сообщения *(Т)* будят равна

$$T = n \cdot t_и.$$

Из радиотехники известно, что импульсу с длительностью $t_и$ соответствует определенная полоса частот Δw причем

$$\Delta w = k \cdot \frac{1}{t_{è}}, \qquad (1.12)$$

где k — коэффициент, зависящий от формы импульса $(k \approx 1)$.

Следовательно, количество информации в рассматриваемом двоичном сообщении длительностью *T* не может превышать предельного значения $(I_{пр})$

$$I_{пр} = n \cdot \log_2 2 = n = \frac{T}{t_и}.$$

И с учетом формулы (1.12) имеем:

$$I_{пр} = T \cdot \Delta w \quad \text{(бит)}.$$

Разделив $I_{пр}$ на *T,* получим предельную информационную ёмкость сообщения (*R*)

$$R \approx \Delta w \quad \text{(бот)}.$$

Таким образом, информационная емкость двоичного дискретного сообщения численно равна полосе частот сигнала, с помощью которого передаются символы сообщения, выраженной в герцах.

Рассмотрим дискретное сообщение состоящее из последовательности импульсов длительностью $t_и$, причем паузы между импульсами отсутствуют, а амплитуда каждого импульса соответствует одному из *L* заранее установленных дискретных уровней. Можно считать, что алфавит такого источника сообщений состоит из *L* равновероятных и независимых символов Количество информации в сообщении, состоящем из *n* таких символов (длительностью *T),* не может превышать предельного значения *(* $I_{пр}$*)*, определяемого по формуле Хартли

$$I_{пр} = n \cdot \log_2 L = \frac{T}{t_и} \cdot \log_2 L \approx T \cdot \Delta w \cdot \log_2 L \quad \text{(бит)}.$$

А информационная емкость такого сообщения (сигнала) (*R*) равна:

$$R = \frac{I_{пр}}{T} \approx \Delta w \cdot \log_2 L$$

Следует отметить, что полученные выражения имеют смысл только тогда, когда разность между двумя соседними символами значительно превышает уровень помехи и когда отсутствует избыточность. В противном случае, а также в случае, когда символы, формирующие сообщение имеют различную длительность, вычисление информационной емкости сообщения является более сложной задачей.

1.7. Информация в непрерывных сообщениях.

Под непрерывным сообщением подразумеваются сообщения, состоящие из символов бесконечного алфавита, т.е. символы такого сообщения представляют собой непрерывное множество из некоторого конечного интервала. Поэтому непрерывное сообщение можно рассматривать как некую непрерывную функцию.

Подходы, приведенные для оценки количества информации, содержащейся в дискретных сообщениях, непосредственно применить к непрерывным сообщениям не удаётся, так как это приводит к результату, противоречащему здравому смыслу.

Действительно, поскольку алфавит источника непрерывных сообщений (L) бесконечен и число символов в сообщении (n) также бесконечно, то и количество информации $(I_н)$, содержащееся в непрерывном сообщении даже конечной длины, равно бесконечности, т.е.

$$I_н = n \cdot \log_2 L \to \infty$$

Однако для реальных источников сообщений это противоречие устранимо, т.к. любые реальные источники сообщений обладают следующими особенностями.

Во-первых, реальные источники сообщений обладают инерционностью (т.е. для перехода их из одного состояния в другое требуется конечное время). Это выражается в ограниченности частотного спектра непрерывных сообщений, и, следовательно, к ним применима теорема Котельникова («теорема отсчетов»). Из этой теоремы следует, что любое непрерывное сообщение $x(t)$ длительностью T и верхней граничной частотой в спектре F_m может быть представлено последовательностью равноотстоящих мгновенных значений этого сообщения взятых с интервалом $\Delta t = 1/2F_m$. При этом общее число равноотстоящих мгновенных значений не будет превышать значения N:

$$N = \frac{T}{1/2F_m} + 1 = 2TF_m + 1.$$

Во-вторых, в любых реальных системах всегда присутствуют помехи (шумы), вызванные различными причинами, вплоть до молекулярныи и квантовых. Эти шумы ограничивают число различимых символов сообщения. Действительно, для надежного распознавания соседних символов необходимо, чтобы разность между ними была больше, чем уровень помехи (шума).

При заданной средней мощности помехи (P_n) и мощности сообщения $\left(P_c\right)$ число различимых символов (L) непрерывного сообщения $x(t)$ приближенно можно определить из соотношения:

$$L \approx \frac{\sqrt{P_c + P_п}}{\sqrt{P_п}} = \sqrt{\frac{P_c}{P_п} + 1}\,.$$

Таким образом, любое реальное непрерывное сообщение длительностью T и с частотным спектром, ограниченным частотой F_m, может быть представлено дискретным сообщением из N символов с алфавитом объемом L. В соответствии с формулой определения количества информации по Хартли (1.5), можно получить формулу для оценки максимального количества информации в таком сообщении ($I_н$):

$$I_н = N \cdot \log_2 L \approx (2TF_m + 1) \cdot \log_2 \sqrt{\frac{P_c + P_п}{P_п}} \quad \text{(бит)} \tag{1.13}$$

Можно так же подсчитать предельную информационную емкость реального непрерывного сообщения (R):

$$R = \frac{I_н}{T} \approx \frac{2TF_m + 1}{T} \cdot \frac{1}{2} \log_2 \frac{P_c + P_п}{P_п} \approx F_m \cdot \log_2 \frac{P_c + P_п}{P_п}\,,$$

при условии, что $2TF_m$>>1.

Эта формула устанавливает предельную информационную емкость непрерывного сообщения. Из нее следует, что важнейшими характеристиками любого источника непрерывных сообщений (и канала передачи информации) являются ширина частотного спектра и отношение мощности сообщения (сигнала) к мощности помех.

Следует подчеркнуть, что приведенные выше соотношения справедливы только при условии взаимной независимости и равновероятности появления символов непрерывного сообщения т.к. их вывод основан на использовании формулы Хартли.

1.8. Энтропия непрерывных сообщений.

В приведенных выше рассуждениях предполагалось, что все возможные значения (символы) непрерывного сообщения равновероятны, однако это не всегда справедливо. Как правило, символы непрерывного сообщения $x(t)$ обладают некоей плотностью распределения вероятности *p(x)*, которая характеризует вероятность попадания символов непрерывного сообщения $x(t)$ в интервал Δx, примыкающий к точке x. И, если плотностью распределения вероятности значений непрерывного сообщения *p(x)*, не постоянна, то и вероятность появления отдельных символов непрерывного сообщения различна. Следовательно, в соответствии с формулой (1.4), количество информации, которое может нести отдельный символ непрерывного сообщения, так же не постоянно и будет зависеть от значений непрерывного сообщения $x(t)$ Поэтому важно выяснить, как зависят информационные характеристики непрерывного сообщения от присущего ему закона распределения его символов.

Положим, что непрерывное сообщение обладает известной функцией плотности распределения вероятности его символов *p(x)* (Рис. 1.2).

Рис. 1.2. Функция плотности распределения вероятности $p(x)$ сообщения(x).

Выберем интервал Δx_k. и обозначим середину этого интервала через x_k. Будем рассматривать все символы непрерывного сообщения, попадающие в интервал Δx_k , как k-тый символ дискретного сообщения. Вероятность появления этого символа дискретного сообщения (P_k) будет определяться выражением:

$$P_k = P\left(x_k - \frac{\Delta x_k}{2} \le x \le x_k + \frac{\Delta x_k}{2} \right) = \int\limits_{x_k - \frac{\Delta x_k}{2}}^{x_k + \frac{\Delta x_k}{2}} p(x)\,dx$$

Энтропию полученного таким образом дискретного сообщения ($H_Д$), на основании формулы Шеннона (1.8), можно записать в виде:

$$H_Д = -\sum_k P_k \cdot \log_2 P_k = -\sum_k \int\limits_{x_k - \frac{\Delta x_k}{2}}^{x_k + \frac{\Delta x_k}{2}} p(x)\,dx \cdot \log_2 \int\limits_{x_k - \frac{\Delta x_k}{2}}^{x_k + \frac{\Delta x_k}{2}} p(x)\,dx \qquad (1.14)$$

Предполагая, что функция *p(x)* вместе со своей производной непрерывная, имеем:

$$\int\limits_{x_k - \frac{\Delta x_k}{2}}^{x_k + \frac{\Delta x_k}{2}} p(x)\,dx \cong p(x_k) \cdot \Delta x_k \,.$$

Исходное непрерывное сообщение можно рассматривать как предел сформированного дискретного сообщения при $\Delta x_k \to 0$ (т.е. при стремлении объёма алфавита сформированного дискретного сообщения к бесконечности). Поэтому, подставив приведенное выше выражения в (1.14) и перейдя к пределу, получим выражение для энтропии непрерывного сообщения:

$$H_н(x) = \lim_{\Delta x_k \to 0} H_Д = \lim_{\Delta x_k \to 0}\left\{-\sum_k p(x_k)\cdot\Delta x_k\cdot\log_2\left[p(x_k)\cdot\Delta x_k\right]\right\}.$$

Заменяя в этом выражении логарифм произведения суммой логарифмов сомножителей, преобразуем это выражение к виду

$$H_н(x) = \lim_{\Delta x_k \to 0}\left\{-\sum_k\left[p(x_k)\cdot\log_2 p(x_k)\right]\Delta x_k\right\} + \lim_{\Delta x_k \to 0}\left\{-\sum_k\left[p(x_k)\cdot\Delta x_k\cdot\log_2\Delta x_k\right]\right\} =$$

$$= -\int_{-\infty}^{+\infty} p(x)\cdot\log_2 p(x)\,dx - \lim_{\Delta x_k \to 0}\sum_k p(x_k)\cdot\Delta x_k\cdot\log_2\Delta x_k \qquad (1.15)$$

Обозначим первое слагаемое в этом выражении как H_x, а второе - как H_Δ. Легко показать, что слагаемое H_Δ обращается в бесконечность при $\Delta x_k \to 0$ независимо от вида функции плотности распределения вероятности символов непрерывного сообщения ($p(x)$).

Действительно, если принять, что интервалы Δx_k одинаковы, т.е. $\Delta x_k = \Delta x$ для всех значений k, то для второго слагаемого в выражении (1.15) имеем:

$$H_\Delta = -\lim_{\Delta x \to 0}\left(\sum_k p(x_k)\cdot\Delta x_k\cdot\log_2\Delta x\right) = -\lim_{\Delta x \to 0}\left(\log_2\Delta x\cdot\sum_k p(x_k)\cdot\Delta x_k\right) =$$

$$= -\lim_{\Delta x \to 0}(\log_2\Delta x)\cdot\lim_{\Delta x \to 0}\left(\sum_k p(x_k)\cdot\Delta x_k\right) = -\lim_{\Delta x \to 0}(\log_2\Delta x)\cdot\int_{-\infty}^{+\infty} p(x)\,dx$$

С учетом того, что $\int_{-\infty}^{+\infty} p(x)\,dx = 1$, получаем

$$H_\Delta = -\lim_{\Delta x \to 0}(\log_2\Delta x) = \infty$$

На основании выше изложенного можно сделать следующие выводы:

1. Непрерывные сообщения не имеют абсолютной меры энтропии. Их полная энтропия равна бесконечности.
2. Второе слагаемое в выражении (1.15) стремится к бесконечности одинаковым образом для любых непрерывных сообщений независимо от присущего им закона распределения.

На практике при определении различных информационных характеристик непрерывных сообщений, которое в реальных условиях всегда действует на фоне шума, приходится вычислять разность энтропий сообщения и шума. При этом бесконечно большие слагаемые H_Δ в энтропии сообщения и энтропии шума взаимно уничтожаются, так как они не зависят от законов распределения символов сообщения и шума. И основное значение приобретает первое слагаемое H_x в выражении (1.15).

Таким образом, за меру энтропии непрерывного сообщения можно принять выражение

$$H_x = -\int_{-\infty}^{+\infty} p(x) \cdot \log_2 p(x) dx. \qquad (1.16)$$

Вычисленную таким образом энтропию называют *относительной* или *дифференциальной*, так как она определяет не полную энтропию непрерывного сообщения, равную, как было показано, бесконечности, а только ту ее часть (H_x), которая зависит от закона распределения символов непрерывного сообщения. И под энтропией непрерывного сообщения, как правило, понимают именно дифференциальную энтропию H_x.

1.9. Экстремальные свойства энтропии непрерывных сообщений.

Представляет интерес определение вида функции плотности распределения вероятности символов ($p(x)$) непрерывного сообщения $x(t)$, с заданными пределами изменения символов, которая обращает энтропию этого непрерывного сообщения (H_x) в максимум.

Пусть непрерывное сообщение $x(t)$ представляет собой ограниченную непрерывную функцию с областью значений из интервала $[a;b]$ (Рис. 1.3) и с неизвестной плотностью распределения вероятностей символов $p(x)$, которая удовлетворяет условию

$$\int_a^b p(x) dx = 1. \qquad (1.17)$$

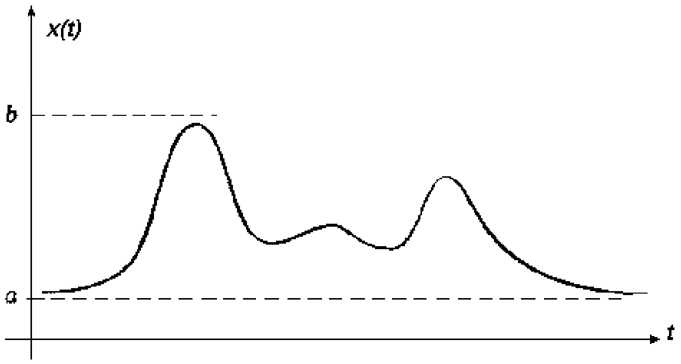


Рис. 1.3..График непрерывного сообщения $x(t)$.

Поставим задачу найти распределение $p_{max}(x)$, при котором дифференциальная энтропия этого сообщения

$$H_x = -\int_{-\infty}^{\infty} p(x) \cdot \log_2 p(x) dx$$

принимает максимальное значение.

Для решения этой задачи воспользуемся методом неопределенных множителей Лагранжа, который используется для нахождения локальных экстремумов функций нескольких переменных. Суть этого метода заключается в следующем. Если задана функция нескольких переменных, например $H = H(x, p)$, аргументы которой удовлетворяют некоему уравнению связи $\varphi(x, p) = 0$, то можно составить функцию Лагранжа $F(x, p, \lambda)$

$$F(x, p, \lambda) = H(x, p) + \lambda \cdot \varphi(x, p)$$

а локальные экстремумы определить из решения системы уравнений:

$$\begin{cases} \dfrac{\partial F(x, p, \lambda)}{\partial p} = 0 \\ \varphi(x, p) = 0 \end{cases}.$$

В данном случае функция H имеет вид:

$$H = H_x = -\int_{-\infty}^{\infty} p(x) \cdot \log_2 p(x)\, dx,$$

а уравнение связи получим из выражения (1.17).

Составив функцию Лагранжа

$$F(x, p, \lambda) = H + \lambda\left(\int_a^b p(x)dx - 1\right) = \int_a^b \left[- p(x) \cdot \log_2 p(x) + \lambda \cdot p(x)\right]dx - \lambda,$$

и продифференцировав ее по p (p не зависит от x, поэтому достаточно продифференцировать подынтегральное выражение), на основании необходимого условия существования экстремума получим:

$$\frac{\partial F(x, p, \lambda)}{\partial p} = -\log_2 p - \frac{1}{\ln 2} + \lambda = 0.$$

Откуда после несложных преобразований

$$\log_2 p = \lambda - \frac{1}{\ln 2} = \log_2 2^{\lambda} - \log_2 e = \log_2 \frac{2^{\lambda}}{e}$$

имеем:

$$p_{\max}(x) = \frac{2^{\lambda}}{e}, \text{ при } x \in [a; b]$$

Подставив это выражение в (1.17) получим:

$$\int_b^a \frac{2^{\lambda}}{e} dx = \frac{2^{\lambda}}{e}(b - a) = 1,$$

и используя предыдущее равенство находим:

$$p_{\max}(x) = \frac{2^{\lambda}}{e} = \frac{1}{b - a}, \text{ при } x \in [a; b].$$

Следовательно, искомая функция плотности распределения вероятности ($p_{max}(x)$) запишется в виде:

$$P_{max}(x)=\begin{cases}0, & при\ x<a,\\ \dfrac{1}{b-a}, & при\ a\le x\le b,\\ 0, & при\ x>b.\end{cases}$$

Таким образом, энтропия непрерывного сообщения принимает свое максимальное значение при равновероятном появлении всех символов, принадлежащих интервалу [a;b]. При этом максимально возможное значение энтропии будет определяться по формуле

$$H_{max}=-\int_{b}^{a}\frac{1}{b-a}\cdot\log_2\left(\frac{1}{b-a}\right)dx=\log_2(b-a).$$

1.10. Информация в непрерывных сообщениях при наличии шумов.

Рассмотрим модель источника непрерывных сообщений, представленную на рис. 1.4,

где: *1* — источник непрерывных зашумленных сообщений;

2 — идеальный источник непрерывных сообщений;

3 — сумматор.

Выходное сообщение такого устройства (y) представляет собой смесь идеального непрерывного сообщения (x) и шума (n). Будем рассматривать частный случай, когда шумы и помехи носят аддитивный характер, то есть когда выходное сообщение (y) есть сумма идеального сообщения (x) и шума (n) т.е.

$$y = x + n$$

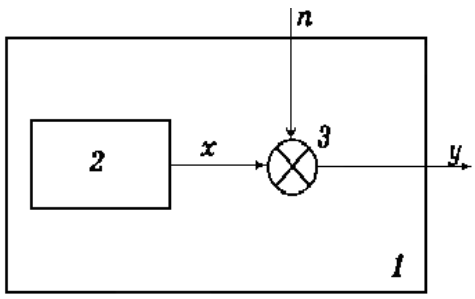


Рис. 1.4 Модель источника непрерывных сообщений.

Для определения качества информации, содержащегося в каком-либо конкретном символе непрерывного сообщения (I), можно воспользоваться основным соотношением теории информации (1.3):

$$I=\log\frac{P_2}{P_1},\tag{1.18}$$

где P_1 — априорная вероятность появления символа сообщения;

P_2 — апостериорная вероятность появления того же символа сообщения.

В случае непрерывного сообщения $x(t)$ под априорной вероятностью P_1 следует подразумевать вероятность того, что символ сообщения $x(t)$ заключен в интервале значений между x и $x+\Delta x$. Если известна функция плотности распределения вероятности символов исходного сообщения (*p(x)*), то априорная вероятность P_1 имеет вид:

$$P_1 = p(x)\cdot\Delta x\ ,$$

а апостериорную вероятность сообщения P_2 можно представить в виде:

$$P_2 = p_y(x)\cdot\Delta x,$$

где $p_y(x)$— условная плотность распределения вероятностей символов сообщения $x(t)$, когда известен конкретный символ y.

Следует отметить, что условная плотность распределения $p_y(x)$ представляет собой не что иное, как плотность распределения вероятностей шума $p(n)$. Действительно, если на выходе зашумлённого источника непрерывных сообщений зафиксирован какой-либо определённый символ y_0, то в сообщении $x(t)$ ему соответствует символ $x = y_0 - n$, который имеет плотность распределения тождественную плотности распределения шума n.

Подставляя в формулу (1.18) значения вероятностей P_1 и P_2 и переходя к пределу при $\Delta x \to 0$, получим следующие соотношения:

$$I = \lim_{\Delta x\to 0} \log_2\left[\frac{p_y(x)\cdot\Delta x}{p(x)\cdot\Delta x}\right] = \log_2\left[\frac{p_y(x)\cdot dx}{p(x)\cdot dx}\right] = \log_2\frac{p_y(x)}{p(x)} = \log_2\frac{p(n)}{p(x)} \qquad (1.19)$$

Предельный переход в этой формуле необходим, т.к. непрерывные сообщения обладают бесконечным алфавитом.

Определяемое формулой(1.19) количество информации (*I*) соответствует только одной паре возможных значений непрерывного сообщения x и шума n. Так как и сообщение и помеха могут принимать бесконечное число различных, не связанных между собой, значений, то для оценки среднего количества информации, получаемой при приеме сообщения на фоне аддитивного шума, необходимо усреднить выражение (1.19) по всем возможным значениям x и n.

Для этого используются известные соотношения теории вероятностей. Так, совместная плотность распределения вероятностей событий x и y ($p(x,y)$) представляется в виде:

$$p(x,y)=p_y(x)\cdot p(y)= p_x(y)\cdot p(x), \qquad (1.20)$$

где $p_x(y)$ - плотность распределения вероятности символов в сообщении $y=x+n$, когда переданное сообщение есть x ;

$p_y(x)$ - условная плотность распределения вероятности символов в сообщении x , когда принятое сообщение есть y.

Известны так же следующие соотношения:

$$\int_{+\infty}^{-\infty} p(x,y)\,dy = p(x);$$

$$\int_{+\infty}^{-\infty} p(x,y)\,dx = p(y); \qquad (1.21)$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\,dxdy = 1$$

Кроме того, из (1.20) следует

$$\frac{p_y(x)}{p(x)} = \frac{p_x(y)}{p(y)}.$$

Подставляя это отношение в формулу (1.19) получаем:

$$I = \log_2\left[\frac{p_x(y)}{p(y)}\right] = \log_2\left[\frac{p_x(y)\cdot dy}{p(y)\cdot dy}\right],$$

С учетом того, что $p_x(y)\cdot dy = P_x(y)$ есть условная вероятность появления сообщения y в интервале $[y; y+dy]$, когда передаваемое сообщение есть x, а $p(y)\cdot dy = P(y)$ - вероятность того, что сообщение y содержится в интервале $[y; y+dy]$ при неизвестном x, предыдущее выражение преобразуется к виду

$$I = \log_2\frac{P_x(y)}{P(y)} \qquad (1.22)$$

Это соотношение определяет количество информации, получаемой при передаче символа сообщения x, когда при приеме наблюдателю известен символ в виде суммы $y = x + n$.

Заметим, что $p_x(y)$·по существу есть не что иное, как плотность распределения вероятностей шума n. Действительно, если задан конкретный символ сообщения x_0, то сумма $y=x_0+n$ распределена по закону распределения шума n, который наложен на символ сообщения x_0, а следовательно,

$$p_x(y)\cdot = p(n). \qquad (1.23)$$

Учитывая (1.23), выражение(1.22) примет вид:

$$I = \log_2\frac{p(n)}{p(y)}. \qquad (1.24)$$

Для оценки среднего количества информации при приеме сообщения на фоне шума ($\widetilde{I}$) необходимо путем интегрирования по всем значениям x и y усреднить количество информации (I), приходящееся на каждое распределение вероятностей $p(x,y)$.

$$\begin{aligned}\widetilde{I} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I\cdot p(x,y)\,dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\log_2\left[\frac{p(n)}{p(y)}\right]\right\}\cdot p(x,y)\,dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{p(x,y)\cdot\log_2[p(n)]\}dxdy - \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{\log_2[p(y)]\}\cdot p(x,y)\,dxdy\end{aligned} \qquad (1.25)$$

Очевидно, что плотность распределения вероятностей шума (n) не зависит от x и y по отдельности, а определяется только их разностью, т.к. $n= y - x$. Кроме того, учитывая (1.23), соотношение (1.20) можно преобразовать к виду

$$P(x,y)= p_x(y)\cdot p(y)=p(n)\cdot p(x), \tag{56}$$

тогда первое слагаемое в (1.25) можно записать в следующем виде:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\cdot\log_2\left[p(n)\right]dxdy=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} p(n)\cdot\log_2\left[p(n)\right]dn\right]p(x)\,dx.$$

Во внутреннем интеграле dy заменено на dn , т.к. $y=x + n$, а при интегрировании по y величина x рассматривается как постоянная.

В такой записи внутренний интеграл, в соответствии с (1.16), представляет собой взятую со знаком минус энтропию шума ($H(n)$).

При стационарном шуме энтропия шума является величиной постоянной и может быть вынесена за знак интеграла, тогда первое слагаемое в выражении (1.25) преобразуется к виду:

$$\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} p(n)\cdot\log_2\left[p(n)\right]dn\right]p(x)\,dx=-H(n)\int_{-\infty}^{+\infty} p(x)\,dx=-H(n),$$

т.к.

$$\int_{-\infty}^{+\infty} p(x)\,dx=1\,.$$

Используя известное равенство (1.21), второе слагаемое выражения (1.25) легко привести к виду:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\log_2\left[p(y)\right]\cdot p(x,y)\,dx\,dy=\int_{-\infty}^{+\infty}\log_2\left[p(y)\right]\left[\int_{-\infty}^{+\infty} p(x,y)\,dx\right]dy=\int_{-\infty}^{+\infty}\log_2\left[p(y)\right]\cdot p(y)\cdot dy=-H(y)$$
(59)

Окончательно получаем:

$$\tilde{I}=H(y)-H(n)\,.$$

Таким образом, среднее количество информации в символе непрерывного сообщения x, которую можно извлечь из зашумлённого символа y ($y= x+ n$) равно разности энтропий принятого сообщения y и шума n.

Глава 2. Основы теории кодирования.

2.1. Кодирование. Основные понятия.

Различные методы кодирования широко используются в практической деятельности человека с незапамятных времён. Например, десятичная позиционная система счисления – это способ кодирования натуральных чисел. Другой способ кодирования натуральных чисел – римские цифры, причем этот метод более наглядный и естественный, действительно, палец – I, пятерня – V, две пятерни – X. Однако при этом способе кодирования труднее выполнять арифметические операции над большими числами, поэтому он был вытеснен способом кодирования основанном на позиционной десятичной системой счисления. Из этого примера можно заключить, что различные способы кодирования обладают присущими только им специфическими особенностями, которые в зависимости от целей кодирования могут быть как достоинством конкретного способа кодирования, так и его недостатком.

Широко известны способы числового кодирования геометрических объектов и их положения в пространстве: декартовы координаты и полярные координаты. И эти способы кодирования отличаются присущими им специфическими особенностями.

До XX века методы и средства кодирования играли вспомогательную роль, но с появлением компьютеров ситуация радикально изменилась. Кодирование находит широчайшее применение в информационных технологиях и часто является центральным вопросом при решении самых разных задач таких как:

– представление данных произвольной природы (чисел, текста, графики) в памяти компьютера;

– оптимальная передача данных по каналам связи;

– защита информации (сообщений) от несанкционированного доступа;

– обеспечение помехоустойчивости при передаче данных по каналам связи;

– сжатие информации.

С точки зрения теории информации кодирование — это процесс однозначного сопоставления алфавита источника сообщения и некоторой совокупности условных символов, осуществляемое по определенному правилу, а код (кодовый алфавит) — это полная совокупность (множество) различных условных символов (символов кода), которые могут использоваться для кодирования исходного сообщения и которые возможны при данном правиле кодирования. Число же различных кодовых символов составляющих кодовый алфавит называют объемом кода или объёмом кодового алфавита. Очевидно, что объём кодового алфавита не может быть меньше объёма алфавита кодируемого исходного сообщения. Таким образом, кодирование — это преобразование исходного сообщения в совокупность или последовательность кодовых символов, отображающих сообщение, передаваемое по каналу связи.

Кодирование может быть числовым (цифровым) и нечисловым, в зависимости от вида, в котором представлены кодовые символы: числа в какой-либо системе счисления или иные какие-то объекты или знаки соответственно.

В большинстве случаев кодовые символы представляют собой совокупность или последовательность неких простейших составляющих, например, последовательность цифр в кодовых символах числового кода, которые называются элементами кодового символа. Местоположение или порядковый номер элемента в кодовом слове определяется его позицией.

Число элементов символа кода, используемое для представления одного символа алфавита исходного источника сообщений, называют значностью кода. Если значность кода одинакова для всех символов алфавита исходного сообщения, то код называют равномерным, в противном случае — неравномерным. Число элементов входящих в кодовый символ иногда называют длиной кодового символа.

С точки зрения избыточности все коды можно разделить на неизбыточные коды и избыточные. В избыточных кодах число элементов кодовых символов может быть сокращено за счет более эффективного использования оставшихся элементов, в неизбыточных же кодах сокращение числа элементов в кодовых символах невозможно.

Задачи кодирования при отсутствии помех и при их наличии существенно различны. Поэтому различают эффективное (статистическое) кодирование и корректирующее (помехоустойчивое) кодирование. При эффективном кодировании ставится задача добиться представления символов алфавита источника сообщений минимальным числом элементов кодовых символов в среднем на один символ алфавита источника сообщений за счет уменьшения избыточности кода, что ведет к повышению скорости передачи сообщения. А при корректирующем (помехоустойчивом) кодировании ставится задача снижения вероятности ошибок в передаче символов исходного алфавита путем обнаружения и исправления ошибок за счет введения дополнительной избыточности кода.

Отдельно стоящей задачей кодирования является защита сообщений от несанкционированного доступа, искажения и уничтожения их. При этом виде кодирования кодирование сообщений осуществляется таким образом, чтобы даже получив их, злоумышленник не смог бы их раскодировать. Процесс такого вида кодирования сообщений называется шифрованием (или зашифровкой), а процесс декодирования – расшифрованием (или расшифровкой). Само кодированное сообщение называют шифрованным (или просто шифровкой), а применяемый метод кодирования – шифром.

Довольно часто в отдельный класс выделяют методы кодирования, которые позволяют построить (без потери информации) коды сообщений, имеющие меньшую длину по сравнению с исходным сообщением. Такие методы кодирования называют методами сжатия или упаковки данных. Качество сжатия определяется коэффициентом сжатия, который обычно измеряется в процентах и который показывает на сколько процентов кодированное сообщение короче исходного.

При автоматической обработке информации с использованием ЭВМ как правило используют числовое (цифровое) кодирование, при этом, естественно, возникает вопрос обоснования используемой системы счисления. Действительно, при уменьшении основания системы счисления упрощается алфавит элементов символов кода, но происходит удлинение символов кода. С другой стороны, чем больше основание системы счисления, тем меньшее число разрядов требуется для представления одного символа кода, а, следовательно, и меньшее время для его передачи, но с ростом основания системы счисления существенно повышаются требования к каналам связи и техническим средствам распознавания элементарных сигналов, соответствующих различным элементам символов кода. В частности, код числа, записанного в двоичной системе счисления в среднем приблизительно в 3,5 раза длиннее десятичного кода. Так как во всех системах обработки информации приходится хранить большие информационные массивы в виде числовой информации, то одним из существенных критериев выбора алфавита элементов символов числового кода (т.е. основания используемой системы счисления) является минимизация количества электронных элементов в устройствах хранения, а также их простота и надежность.

При определении количества электронных элементов требуемых для фиксации каждого из элементов символов кода необходимо исходить из практически оправданного предположения, что для этого требуется количество простейших электронных элементов (например, транзисторов), равное основанию системы счисления a. Тогда для хранения в некотором устройстве n элементов символов кода потребуется M электронных элементов:

$$M = a \cdot n. \tag{2.1}$$

Наибольшее количество различных чисел, которое может быть записано в этом устройстве N:

$$N = a^n .$$

Прологарифмировав это выражение и выразив из него n получим:

$$n = \ln N / \ln a.$$

Преобразовав выражение (2.1) к виду

$$M = a \cdot \ln N / \ln a \tag{2.2}$$

можно определить, при каком основании логарифмов a количество элементов M будет минимальным при заданном N. Продифференцировав по a функцию $M = f(a)$ и приравняв её производную к нулю, получим:

$$\frac{dM}{da} = \ln N \cdot \frac{\ln a - 1}{\ln^2 a} = 0.$$

Очевидно, что для любого конечного a

$$\ln N / \ln^2 a \neq 0$$

и, следовательно,

$$\ln a - 1 = 0,$$

откуда $a = e \approx 2{,}7$.

Так как основание системы счисления может быть только целым числом, то a выбирают равным 2 или 3. Для примера зададимся максимальной емкостью устройства хранения N=10^6 чисел. Тогда при различных основаниях систем счисления (a) количество элементов (M) в таком устройстве хранения будет, в соответствии с выражением (2.2), следующие (Таблица 2.1):

Таблица 2.1.

a	2	3	4	5	10	20
M	39,2	38,2	39,2	42,9	60	91,2

Следовательно, если исходить из минимизации количества оборудования, то наиболее выгодными окажутся двоичная, троичная и четверичная системы счисления, которые близки по этому параметру. Но так как техническая реализация устройств, работающих в двоичной системе счисления, значительно проще, то наибольшее распространение при числовом кодировании получили коды на основе системы счисления по основанию 2.

2.2. Избыточность кодов.

Понятие избыточности означает, что фактическая энтропия кода или сообщения *(H)* меньше, чем максимально возможная энтропия (H_{max}), т. е. число символов в сообщении или элементов в символе кода больше, чем это требовалось бы при полном их использовании.
Понятие избыточности легко пояснить следующим примером.

Выделение полезного сигнала на уровне помех — одна из основных проблем передачи информации. Одним из путей повышения надёжности передачи сообщений может быть передача дополнительных символов, т.е. повышение избыточности сообщений.

Действительно, по теореме Котельникова (§ 1.7), непрерывное сообщение (сигнал) можно передать последовательностью мгновенных отсчетов его значений с промежутками между ними Δt:

$$\Delta t = \frac{1}{2 \cdot f_{max}},$$

где f_{max} – верхняя граничная частота в спектре сигнала.

При наличии помех промежутки между отсчетами (Δt_n) необходимо уменьшать, т.е.

$$\Delta t_n < \frac{1}{2 \cdot f_{max}}.$$

В этом случае мы увеличиваем число отсчетов и, следовательно, увеличиваем избыточность сообщения и тем самым повышаем его помехозащищенность.

Пусть сообщение из n символов содержит количество информации I. Если сообщение обладает избыточностью, то его (при отсутствии шума) можно передать меньшим числом символов n_0 ($n_0 < n$). При этом, количества информации (I_1 и I_{1max}), приходящиеся на один символ сообщения, будут связаны соотношением:

$$I_1 = I/n < I_{1max} = I/n_0$$

и, следовательно,

$$n \cdot I_1 = n_0 \cdot I_{1max}.$$

За меру избыточности принимается величина R:

$$R = \frac{n - n_0}{n} = 1 - \frac{n_0}{n} = 1 - \frac{I_1}{I_{1max}} = 1 - \frac{H_1}{H_{max}}. \tag{2.3}$$

Таким образом, избыточность — это свойство, характеризующее возможность представления тех же сообщений в более экономной форме.

При кодировании избыточных сообщений возникает определенная исходная избыточность кодов. Наличие исходной избыточности уменьшает пропускную способность каналов и увеличивает формат сообщений. Вместе с тем в процессе передачи информации избыточность сообщений и кодов является средством, полезным для борьбы с внешними возмущающими воздействиями и помехами.

По наличию избыточности коды также делятся на избыточные и неизбыточные. Для неизбыточных кодов характерно то, что они позволяют просто определить различные символы сообщения. Переход от неизбыточного кода к избыточному осуществляется путем добавления позиций в кодовых символах, которые можно получить либо путем различных логических операций, выполняемых над основными информационными позициями, либо путем использования алгоритмов, связывающих неизбыточный и избыточный коды. Например, если есть символы сообщения A_1; A_2; A_3; A_4, то их можно закодировать в двоичном неизбыточном коде:

$$A_1 = 00; \quad A_2 = 01; \quad A_3 = 10; \quad A_4 = 11.$$

Для получения избыточного кода можно ввести еще одну позицию, значение которой будет определяться как сумма значений предшествующих символов по модулю два:

$$A_1 = 000; A_2 = 011; A_3 = 101; A_4 = 110.$$

Особенностью такого кода является то, что он позволяет обнаружить любую единичную ошибку (ошибку в одной из позиций кода), выявившуюся в процессе передачи кода.

2.3. Эффективное кодирование равновероятных символов сообщений.

Эффективное кодирование используется в каналах без шума, т.е. в таких каналах, где помехи отсутствуют, либо ими можно пренебречь. Основной задачей кодирования в таком канале является обеспечение максимальной скорости передачи информации, близкой к пропускной способности канала передачи (§3.4).

В случае, если все символы алфавита кодируемого сообщения независимы и их появление равновероятно, построение оптимального эффективного кода не представляет трудностей. Действительно, пусть $H(x)$ — энтропия исходного сообщения. Будем считать, что символы сообщения (x_i) равновероятны и объем алфавита исходного источника сообщений равен m. Следовательно, вероятность появления любого i-го символа данного сообщения $(P(x_i))$ будет одинакова, т.е.

$$P(x_i) = \frac{1}{m}, \qquad i = 1,.., m,$$

а энтропия сообщения равна $(H(x))$:

$$H(x) = -\sum_{i=1}^{m} P(x_i) \cdot \log_2 P(x_i) = \log_2 m.$$

Если для кодирования используется числовой код по основанию k (объем алфавита элементов кодовых символов равен k), то энтропия элементов кодовых символов (H_1), при условии, что элементы символов кода появляются равновероятно и взаимнонезависимо, определится из соотношения:

$$H_1 = \log_2 k.$$

Тогда длина эффективного равномерного кода, т.е. число элементов в кодовом символе ($l_{эфф.}$), может быть найдена из соотношения:

$$l_{эф} = \frac{H_x}{H_1} = \frac{\log_2 m}{\log_2 k} = \frac{\log_2 k^n}{\log_2 k} = n, \qquad (2.4)$$

где $m = k^n$.

2.4. Эффективное кодирование неравновероятных символов сообщений

В случае, если символы кодируемого сообщения неравновероятны, в общем виде правило получения оптимального эффективного кода неизвестно. Однако из общих соображений можно представить принципы его построения.

Очевидно, что эффективное кодирование будет оптимальным, если неравномерное распределение вероятностей появления символов алфавита источника сообщений с помощью определенным образом выбранного кода переводят в равновероятное распределение вероятностей появления независимых элементов кодовых символов. В этом случае среднее количество информации, приходящееся на один элемент символа кода, будет максимальным. Для определения вида кода, удовлетворяющего этому требованию, можно рассмотреть «функцию стоимости» (цены передачи символов сообщения) в виде:

$$Q = \sum_{i=1}^{m} P(x_i) \cdot W_i\ , \qquad (2.5)$$

где $P(x_i)$ — вероятность появления i-го символа алфавита исходного кодируемого сообщения;

m — объем алфавита;

W_i — стоимость передачи i-го символа алфавита, которая пропорциональна длине кодового слова.

Эффективный код должен минимизировать функцию Q. Если передача всех элементов символов кода имеет одинаковую стоимость, то стоимость кодового символа будет пропорциональна длине соответствующего кодового символа. Следовательно, в общем случае (при неравновероятных символах исходного сообщения) код должен быть неравномерным, поэтому построение эффективного кода должно основываться на следующих принципах:

1. Длина кодового символа (n_i) должна быть обратно пропорциональна вероятности появления соответствующего символа исходного кодируемого сообщения (x_i);
2. Начало более длинного кодового символа не должно совпадать с началом более короткого (для возможности разделения кодовых символов без применения разделительных знаков);
3. В длинной последовательности элементы символов кода должны быть независимы и равновероятны.

Теоретическое обоснование возможности эффективного кодирования передаваемых по каналу сообщений обеспечивает теорема, доказанная К. Шенноном и которая носит название первая теорема Шеннона или основная теорема Шеннона о кодировании для каналов без помех. Эта теорема гласит, что если источник сообщений имеет энтропию H [бит/символ], а канал обладает пропускной способностью C [бит/сек] (пропускная способность характеризует максимально возможное значение скорости передачи информации), то всегда можно найти способ кодирования, который обеспечит передачу символов сообщения по каналу со средней скоростью

$$V_{cp} = \left(\frac{C}{H} - \varepsilon\right) \text{ [символ/сек]},$$

где ε — сколь угодно малая величина.

Обратное утверждение говорит, что передача символов сообщения по каналу со средней скоростью $V_{cp.} > H$ невозможна и, следовательно,

$$V_{\max} = \frac{C}{H} \text{ [символ/сек]}.$$

Эта теорема часто приводится в иной формулировке: сообщения источника сообщений с энтропией H всегда можно закодировать последовательностями элементов кодовых символов с объемом алфавита k так, что среднее число элементов символов кода на один символ кодируемого сообщения ($l_{\text{ср.}}$) будет сколь угодно близко к величине $H / log_2 k$, но не менее ее.

Не вдаваясь в доказательство этой теоремы, отметим, что эта теорема показывает возможность наилучшего эффективного кодирования, при котором обеспечивается равновероятное и независимое поступление элементов символов кода, а следовательно, и максимальное количество переносимой каждым из них информации, равное $log_2 k$ (бит/элемент) . К сожалению, теорема не указывает конкретного способа эффективного кодирования, она лишь говорит о том, что при выборе каждого элемента кодового символа необходимо чтобы он нес максимальное количество информации, а, следовательно, все элементы символов кода должны появляться с равными вероятностями и взаимнонезависимо.

Исходя из изложенных принципов, разработаны ряд алгоритмов эффективного кодирования как для взаимнонезависимых символов сообщения, так и для взаимнозависимых. Суть их состоит в том, что они сокращают среднюю длину кодовых символов путем присвоения кодовых символов минимальной длины символам исходного сообщения, которые встречаются наиболее часто.

2.5. Алгоритмы эффективного кодирования неравновероятных взаимнонезависимых символов источников сообщений

Алгоритм Шеннона-Фено. Суть этого алгоритма, при использовании двоичного кода (объем алфавита элементов символов кода равен 2), заключается в следующем.

Все символы алфавита источника сообщений ранжируют, т. е. располагают в порядке убывания вероятностей их появления. Затем символы алфавита делятся на две группы приблизительно равной суммарной вероятности их появления. Все символы первой группы получают «0» в качестве первого элемента кодового символа, а все символы второй группы — «1». Далее группы делятся на подгруппы, по тому же правилу примерно равных суммарных вероятностей, и в каждой подгруппе присваивается вторая позиция кодовых символов. Процесс повторяется до закодирования всех символов алфавита кодируемого источника сообщений. В кодовый символ, соответствующий последней группе, добавляется в качестве последнего элемента «0» для того, чтобы начальный элемент символов кода не совпадал с конечным, что позволяет исключить разделительные элементы между символами кода.

Таблица 2.2 иллюстрирует процесс построения кода Шеннона–Фено на примере источника сообщений, алфавит которого состоит из восьми символов.

Таблица 2.2

Номер символа. (i)	Символы алфавита. (m_i)	Вероятности **(P_i).**	Номера Разбиений.	Кодовые Символы.
1	m_1	1/2	I	0
2	m_2	1/4		10
3	m_3	1/8	II	110
4	m_4	1/16	III	1110
5	m_5	1/32		11110
6	m_6	1/64	IV	111110
7	m_7	1/128	V	1111110
8	m_8	1/256		11111110

На рис. 2.1 представлен граф кодирования (кодовое дерево), который показывает, как «расщепляется» ранжированная последовательность символов кодируемого источника сообщений на группы и отдельные символы и какие кодовые символы присваиваются группам и отдельным символам алфавита источника сообщений на каждом шаге разбиения.

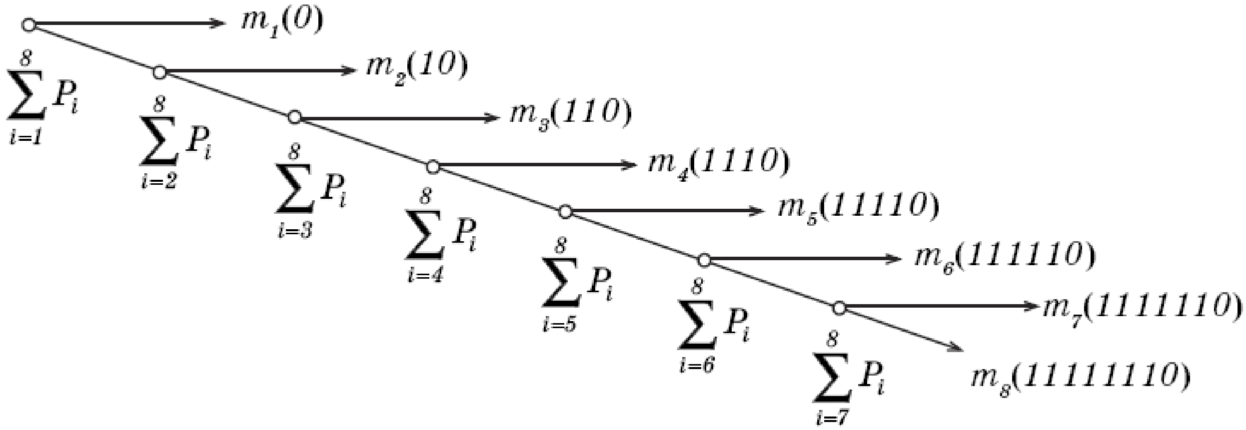


Рис 2.1. Граф кодирования по алгоритму Шеннона–Фено.

Алгоритм Шеннона–Фено применим и при иных числовых основаниях кода ($k > 2$). В этом случае алгоритм получения кода аналогичен рассмотренному примеру, только алфавит кодируемого источника сообщений разбивается на k групп и подгрупп примерно одинаковой суммарной вероятности.

Представляет интерес сравнение эффективного кодирования равномерным кодом и неравномерным кодом по алгоритму Шеннона–Фено.

В качестве примера рассмотрим предложенный выше (Табл. 2.2) источник сообщений с объемом алфавита равным 8 и соответствующими вероятностями появления отдельных символов (P_i). Для кодирования используем двоичный код ($k = 2$).

Энтропия рассмотренного источника сообщений (H_u) определяется по формуле Шеннона:

$$H_u = -\sum_{i=1}^{8} P_i \cdot \log_2 P_i = 1{,}9 \quad \text{(бит/символ).}$$

Максимально же возможное значение энтропии источника сообщений ($H_{u.\max}$), при условии равновероятного и взаимно независимого появления символов, находится по формуле Хартли:

$$H_{u.\max} = -\log_2 \frac{1}{m} = \log_2 8 = 3.$$

Следовательно, избыточность рассматриваемого источника сообщений (R_u) может быть найдена из соотношения:

$$R_u = 1 - \frac{H_u}{H_{u\,max}} = 0.3 \quad .$$

Используя формулу для эффективного равномерного кода (2.4) при $k = 2$, получим значность равномерного двоичного кода *(n_p):*

$$n_p = \log_k m = \log_2 8 = 3,$$

и избыточность равномерного кода *($R_{рк}$)*:

$$R_{рк} = 1 - \frac{H_u}{n_p \cdot \log_2 k} \cong 0{,}3 \quad .$$

Энтропия элементов символов равномерного кода $\left(H_{1.P}\right)$, т.е. количество информации, приходящееся на один элемент символа кода, будет равна:

$$H_{1\,P.} = \frac{H_u}{n_P} \cong 0{,}6 \quad \text{(бит / элемент символа кода).} \qquad (2.6)$$

При использовании эффективного кодирования по алгоритму Шеннона-Фено соответствующие информационные параметры кода будут следующие.

Средняя длина неравномерного кода *(n_H)* определяется выражением:

$$n_H = \sum_{i=1}^{m} n_i \cdot p_i = 2\,, \qquad (2.7)$$

где n_i — значность i - го кодового символа, соответствующего символу алфавита m_i.

Избыточность неравномерного кода *($R_{НК}$)* определим из соотношения:

$$R_{НК} = 1 - \frac{H_u}{n_H \cdot \log_2 K} \cong 0{,}0$$

Энтропия элементов символов эффективного неравномерного кода $\left(H_{1H}\right)$ может быть легко найдена:

$$H_{1\,H} = \frac{H_u}{n_H} = 0.9\ .8 \quad \text{(бит/элемент символа кода).} \qquad (2.8)$$

Из сравнения выражений (2.6) и (2.8) видно, что, при использовании эффективного кодирования по алгоритму Шеннона-Фено, энтропия элементов символов такого неравномерного кода на 50% выше чем энтропия элементов символов в случае использования равномерного кода. Если предположить, что скорость передачи по каналу элементов символов кода (W) одинакова для равномерного и неравномерного кода, то скорость передачи информации (V), определяемая выражением

$$V = H \cdot W\,,$$

где H — энтропия элементов символа кода, также будет на 50% выше при использовании эффективного кодирования по алгоритму Шеннона–Фено по сравнению с равномерным кодированием.

Алгоритм Шеннона–Фено часто применяют и для блочного кодирования. При этом также существенно повышается эффективность кодирования. Для иллюстрации такого кодирования рассмотрим процедуру эффективного кодирования двоичным числовым кодом сообщений, генерируемых источником сообщений с объемом алфавита равным 2 (m = 2), т.е. с алфавитом, состоящим только из двух символов m_1 и m_2 с вероятностями появления $P(m_1)$ = 0,9 и $P(m_2)$ = 0,1 и, следовательно, с энтропией H = 0,47.

При посимвольном кодировании по алгоритму Шеннона–Фено эффект отсутствует, так как на каждый символ сообщения будет приходится один символ кода, состоящий из одного элемента.

Произведем теперь кодирование по алгоритму Шеннона–Фено блоков, состоящих из комбинаций двух символов источника сообщений, считая символы взаимнонезависимыми. Результат приведен в таблице 2.3.

Таблица 2.3.

Блоки	Вероятности	Номера разбиений	Кодовые комбинации
m_1m_1	0,81		1
		I	
m_1m_2	0,09		01
		II	
m_2m_1	0,09		001
		III	
m_2m_2	0,01		0001

Среднее число элементов символов кода на один символ исходного сообщения, вычисленное по формуле (2.7), равно 0,645, что значительно ниже, чем при посимвольном кодировании.

Кодирование блоков, соответствующих комбинациям из трех символов источника сообщений, дает еще больший эффект. Результат приведен в таблице 2.4.

Таблица 2.4.

Блоки.	Вероятности.	Номера разбиений.	Кодовые комбинации.

$m_1 m_1 m_1$	0,729	I	1
$m_2 m_1 m_1$	0,081	III	011
$m_1 m_2 m_1$	0,081	II	010
$m_1 m_1 m_2$	0,081	IV	001
$m_2 m_2 m_1$	0,009	VI	00011
$m_2 m_1 m_2$	0,009	V	00010
$m_1 m_2 m_2$	0,009	VII	00001
$m_2 m_2 m_2$	0,001		00000

В этом случае среднее число элементов символов кода на один символ исходного источника сообщений равно 0,53.

Теоретический минимум H = 0,47 может быть достигнут при кодировании блоков неограниченной длины.

Алгоритм Шеннона-Фено не всегда приводит к однозначному построению кода, так как при разбиении на подгруппы можно сделать большей по суммарной вероятности как верхнюю, так и нижнюю подгруппу. Этого недостатка лишен алгоритм Хаффмена, который гарантирует однозначное построение эффективного кода.

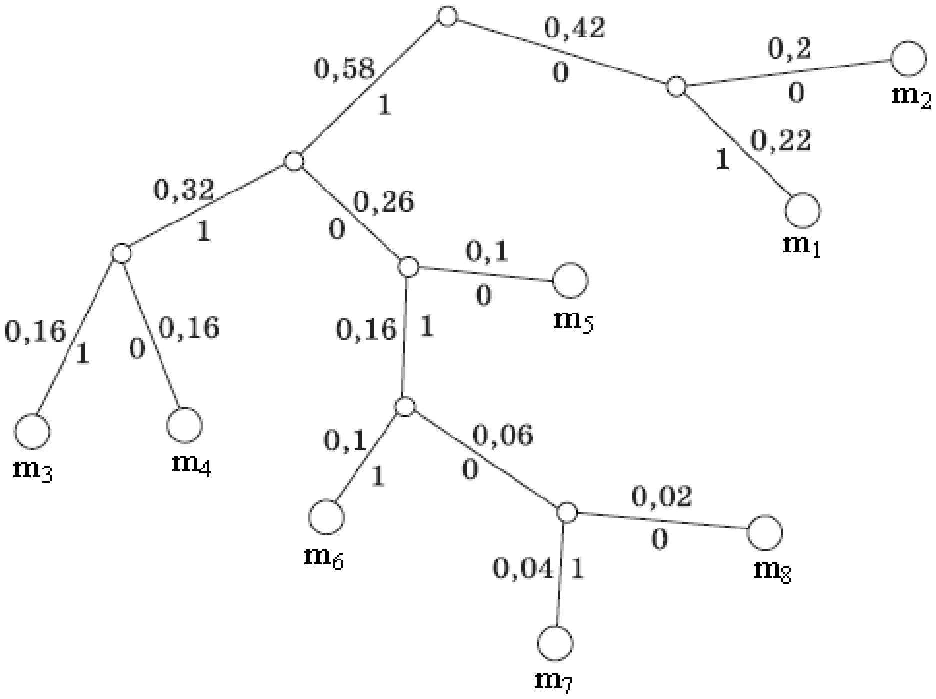
Алгоритм Хаффмена. Суть этого алгоритма, при использовании двоичного кода, состоит в следующем. Все символы алфавита источника сообщений ранжируют, т.е. выписывают в столбец в порядке убывания вероятностей их появления. Два последних символа объединяют в один вспомогательный символ, которому приписывают суммарную вероятность.

Вероятности символов, не участвовавших в объединении, и вероятность вспомогательного символа вновь ранжируют, т.е. располагают в порядке убывания вероятностей в дополнительном столбце и два последних символа объединяются. Процесс продолжается до тех пор, пока не получим единственный вспомогательный символ с вероятностью равной единице. Пример кодирования по алгоритму Хаффмена приведен в таблице 2.5.

Таблица 2.5

Символы	Вероятности	Вспомогательные столбцы						
		1	2	3	4	5	6	7
m_1	0,22	0,22	0,22	0,26	0,32	0,42	0,52	1
m_2	0,20	0,20	0,20	0,22	0,26	0,32	0,42	
m_3	0,16	0,16	0,16	0,20	0,22	0,26		
m_4	0,16	0,16	0,16	0,16	0,20			
m_5	0,10	0,10	0,16	0,16				
m_6	0,10	0,10	0,10					
m_7	0,04	0,06						
m_8	0,02							

На рис. 2.2 показан граф кодирования (кодовое дерево), который иллюстрирует ранжирование символов на группы и отдельные символы, причем из точки, соответствующей вероятности 1, направляем две ветви: одной из них (с большей вероятностью) присваиваем символ 1, а второй – символ 0.



Такое последовательное ветвление продолжим до тех пор, пока не дойдем до вероятности каждого символа. Двигаясь по кодовому дереву сверху вниз, можно записать для каждого символа источника сообщений соответствующую ему комбинацию (кодовый символ).

Рис 2.2. Граф кодирования по алгоритму Хаффмена .

Различные символы, генерируемые источником сообщения, и соответствующие им кодовые символы представлены в таблице 2.6.

Таблица 2.6.

m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
01	00	111	110	100	1011	10101	10100

Этот алгоритм можно использовать и при ином числовом основании кода, а также использовать блоки, как это рассмотрено в алгоритме Шеннона-Фено.

Эффективность рассмотренных алгоритмов достигается благодаря присвоению более коротких кодовых комбинаций (кодовых символов) символам источника сообщений, вероятность которых более высока, и более длинных кодовых комбинаций - символам источника сообщений с малой вероятностью. Это ведет к различиям в длине кодовых символов и, как следствие, к трудностям при их декодировании. Для разделения отдельных кодовых символов можно использовать специальный разделительный элемент, но при этом существенно снижается эффективность кода, т.к. средняя длина кодового символа фактически увеличивается на один элемент символа кода.

Целесообразнее обеспечить декодирование без введения дополнительных элементов символов. Этого можно добиться, если в эффективном коде ни одна кодовая комбинация не будет совпадать с началом более длинной кодовой комбинации. Коды, удовлетворяющие этому условию, называют префиксными кодами (префиксом или началом называют первый элемент в кодовом символе, а последний элемент – окончанием или постфиксом).

Легко заметить, что коды, построенные по алгоритмам Шеннона–Фено или Хаффмена, являются префиксными.

2.6. Алгоритмы эффективного кодирования неравновероятных взаимозависимых символов сообщений

Устранение взаимной зависимости символов источника сообщения может быть осуществлено путем укрупнения алфавита исходного источника сообщения. Для этого подлежащие кодированию сообщения последовательно разбиваются на двух-, трех- или *n*-знаковые сочетания (блоки), вероятности которых известны, а затем эти сочетания кодируются в соответствии с алгоритмами Шеннона-Фено или Хаффмена.

Недостаток этого алгоритма состоит в том, что при его использовании не учитываются связи между символами, входящими в состав соседних сочетаний (блоков).

Этот недостаток может быть устранен кодированием по методу диаграмм, триграмм или в общем случае *k*-грамм. *k*-граммой называют последовательность из *k* смежных символов сообщения. При $k=2$ сочетание смежных знаков называют диаграммой, при $k=3$ — триграммой и т.д.

В процессе кодирования по методу k-грамм производят непрерывное последовательное перемещение k-граммы по сообщению с шагом равным одному символу. Этот процесс (получение 3-х k-грамм) иллюстрируется рис. 2.3, где x_i - символы сообщения.

Рис 2.3. Процесс непрерывного последовательного перемещения k-граммы по сообщению.

Если вероятности появления различных k-грамм известны, то их эффективное кодирование, в частности, может быть выполнено по алгоритмам Шеннона-Фено или Хаффмена. Конкретное значение k выбирается исходя из степени взаимозависимости между символами сообщения и сложности технической реализации кодирующих и декодирующих устройств.

2.7. Недостатки алгоритмов эффективного кодирования.

Основным недостатком этих алгоритмов является специфическое влияние помех на достоверность декодирования, которое проявляется в том, что одиночная ошибка в кодовой комбинации может перевести ее в другую кодовую комбинацию, не равную ей по длительности. Это может привести к неверному декодированию ряда последующих комбинаций, что называют треком ошибки, хотя существуют методы, позволяющие свести трек ошибки к минимуму.

Существенным недостатком является также сложность технической реализации систем эффективного кодирования, которые должны включать в себя буферные устройства и устройства накопления. Использование этих устройств вызвано тем, что длина кодовых комбинаций различна, а каналы связи эффективно работают только в том случае, если символы поступают на них с постоянной скоростью. Кроме этого, при кодировании блоками необходимо накапливать символы, прежде чем присвоить их совокупности какую-либо кодовую комбинацию.

2.8. Помехоустойчивое (корректирующее) кодирование. Общие понятия

Высокие требования к достоверности и надежности передачи, обработки и хранения информации в системах передачи данных, в вычислительных системах и сетях, в региональных системах управления и различного рода информационных системах требуют такого кодирования информации, которое обеспечивало бы безошибочную ее передачу, а в случае появления ошибок - их обнаружение и исправление.

Коды, обладающие такой способностью, называют помехоустойчивыми или корректирующими. Подавляющее большинство существующих в настоящее время помехоустойчивых кодов обладают требуемыми свойствами благодаря их алгебраической структуре. Поэтому их называют алгебраическими кодами. Хотя существуют и иные коды, корректирующее действие которых основано на оценке вероятности искажения каждого символа кода.

Кодовые комбинации (кодовые символы) алгебраических кодов включают в себя две группы элементов кодовых символов: информационные элементы и проверочные элементы. Совокупность информационных элементов кодового символа соответствуют символу кодируемого сообщения, а проверочные (избыточные) элементы добавляются к информационным элементам и служат для обнаружения и исправления ошибок.

Все алгебраические коды можно разделить на два больших класса: блочные (блоковые) и непрерывные.

Блочные коды представляют собой совокупность кодовых символов, состоящих из отдельных комбинаций (блоков) элементов символов кода, которые кодируются и декодируются независимо. При этом каждому символу кодируемого исходного сообщения ставится в соответствие блок (комбинация) из n элементов символов кода, куда включаются информационные и проверочные элементы. Блочный код называют равномерным, если n для всех блоков одинаково.

Непрерывные (древовидные) коды представляют собой непрерывную последовательность кодовых символов, причем введение проверочных элементов производится непрерывно, без разделения ее на независимые блоки.

Как блочные коды, так и непрерывные могут быть разделимыми и неразделимыми.

В разделимых кодах информационные и проверочные элементы символов кода отчетливо разграничены и всегда занимают одни и те же определенные позиции (разряды). Такие коды часто называют (n,k) коды, где n — длина кодового символа, k — число информационных элементов в нем.

При кодировании неразделимыми кодами разделение кодового символа на информационные элементы и проверочные невозможно.

Среди разделимых кодов выделяют систематические (линейные) и несистематические. Систематическими кодами называют коды, в которых проверочные элементы являются линейными комбинациями информационных. Эти коды наиболее распространены, т.к. их использование существенно упрощает техническую реализацию кодирующих и декодирующих устройств.

2.9. Теоретические основы помехоустойчивого кодирования

Теоретические основы помехозащищенного кодирования сообщений базируются на основной теореме Шеннона о кодировании для канала с помехами. Эта теорема гласит:

1. При любой производительности источника сообщений, меньшей чем пропускная, способность канала, существует такой способ кодирования, который позволяет обеспечить передачу всей информации, создаваемой источником сообщений, со сколь угодно малой вероятностью ошибки.
2. Не существует способа кодирования, позволяющего вести передачу информации со сколь угодно малой вероятностью ошибки, если производительность источника сообщений больше пропускной способности канала.

Анализируя теорему, важно отметить, что она устанавливает теоретический предел процесса передачи информации, создаваемой источником сообщений, со сколь угодно малой вероятностью ошибки и показывает, что помехи в канале не накладывают ограничений на точность и достоверность передачи сообщений, а ограничения накладываются лишь на скорость передачи, при которой может быть достигнута сколь угодно высокая достоверность передачи.

Теорема, к сожалению, не указывает конкретных путей построения кодов, обеспечивающих надежность передачи, однако она доказывает, что такие пути существуют.

Для определения конкретных направлений, обеспечивающих повышение надёжности и достоверности передачи сообщений, обратимся к общей информационной модели системы передачи сообщений, представленной на рис. 2.4.

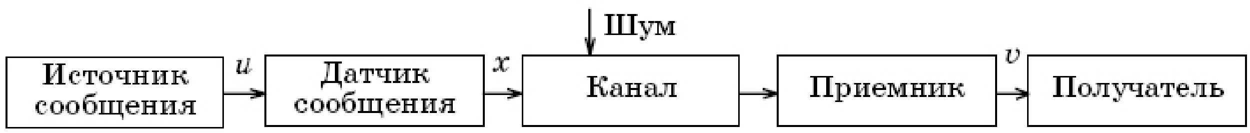


Рис. 2.4. Общая информационная модель системы передачи сообщений.

Пусть u – исходное сообщение, генерируемое источником сообщения, датчик сообщения преобразует его в сигнал, $x = \Psi(u)$ затем этот сигнал проходит по каналу и поступает на приёмник, в котором преобразовывается в сообщение v: $v = \Psi^{-1}(x)$.

Таким образом, процесс передачи сообщений описывается в следующем образом (Рис.2.5.):

— на передающем конце канала пространство сообщений U преобразуется в пространство сигналов X ;

— на приёмном конце канала из пространства сигналов X восстанавливается пространство сообщений V .

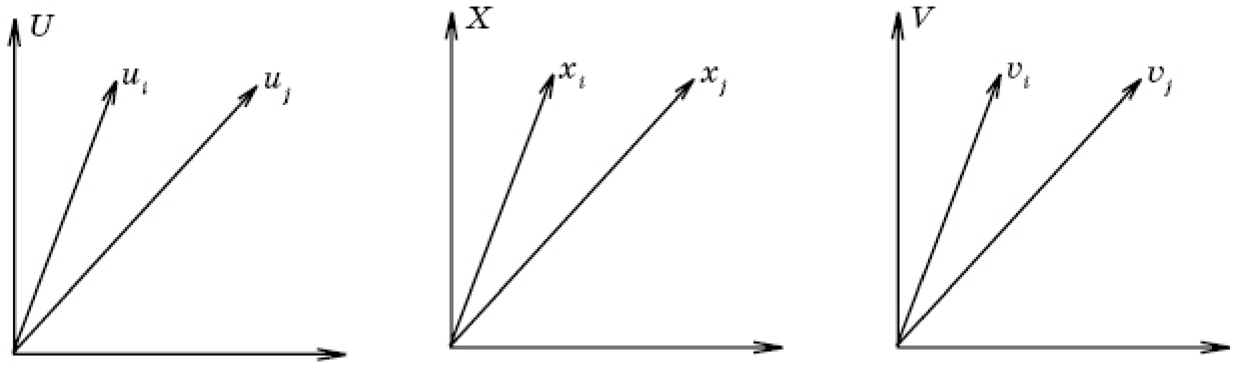


Рис. 2.5. Иллюстрация процесса передачи сообщений по каналу.

При отсутствии помех выполняется однозначное соответствие между исходным символом сообщения (u_i) и принятым символом (v_i). Действительно,

$$v_i = \Psi^{-1}(x_i) = \Psi^{-1}[\Psi(u_i)] = u_i,$$

При наличии помех сигнал X при прохождении через канал будет искажаться и, в общем случае, $v_i \neq u_i$, что приводит к искажению исходного сообщения и вызывает трудности в разделении и опознании исходных символов сообщений. Для безошибочного приёма необходимо идентифицировать принятое сообщение v, то есть определить, какому исходному сообщению u_i оно соответствует.

Выполнить эту идентификацию можно следующим образом. Если собственные шумы в приёмнике отсутствуют, то можно однозначно определить принятый сигнал x, соответствующий полученному сообщению v:

$$v = \Psi^{-1}(x)$$

и определить положение сигнала x в пространстве сигналов X (Рис. 2.6.).

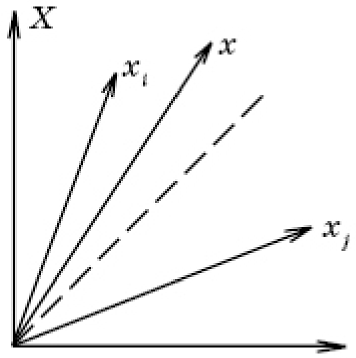


Рис. 2.6. Положение сигнала *x* в пространстве сигналов *X*.

Вектору *x* можно присвоить значение ближайшего к нему известного вектора из пространства сигналов *X* , то есть $x = x_i$, если

$$|x - x_i| < |x - x_j|$$

для любого $i \neq j$.

Приемник, работающий по выше рассмотренному принципу идентификации, называют идеальным приемником.

Ошибка приема идеального приемника (прием сообщения u_i вместо переданного u_j) происходит лишь тогда, когда сигнал *x* под действием помех переходит границу раздела областей этих сообщений, то есть тогда, когда

$$|x_i - x| > \frac{1}{2}|x_i - x_j|$$

Вероятность появления такой ошибки может служить мерой помехоустойчивости или надежности системы связи (П):

$$\Pi = \log_2\left(\frac{1}{p_0}\right) = -\log_2 p_0 \qquad (2.9.)$$

где p_0 – вероятность появления ошибки на выходе приемника.

Очевидно, что с увеличением расстояния между векторами x_i и x_j помехоустойчивость повышается.

Если величина помехи ξ определяется как $\xi = x_i - x$, то для безошибочного приема сообщения *x* необходимо выполнение условия:

$$|x - x_i| < |x - x_j|$$

или

$$|\xi| = |x_i - x - x_j| = |\Delta x_{i\ j} - \xi|,$$

где $\Delta x_{ij} = |x_i - x_j|$

И тогда условие безошибочного приема запишется в виде:

$$|\Delta x_{ij}| = 2|\xi| .$$

В случае, если помеха (ξ) имеет плотность распределения вероятностей $f(\xi)$, то вероятность появления ошибки P_0 можно определить из выражения:

$$P_0 = P\left\{|\xi| \geq \frac{\Delta x_{i\ j}}{2}\right\} = \int_{-\infty}^{-\frac{\Delta x_{i\ j}}{2}} f(x)dx + \int_{+\frac{\Delta x_{i\ j}}{2}}^{+\infty} f(x)dx$$

В общем случае пространство сообщений *U* имеет размерность *n* (например, каждый символ этого сообщения состоит из *n* элементов). Тогда условие безошибочного приема получим из следующих соображений.

Так как

$$X = \Psi(U) ,$$

то

$$\Delta X_k = \frac{\partial \Psi}{\partial U_k} \cdot \Delta U_k ,$$

а

$$|\Delta X|^2 = \sum_{k=1}^{n} \Delta X_k^2 ,$$

где X_k – *k*-я координата пространства *X*.

Следовательно,

$$|\Delta X|^2 = \sum_{k=1}^{n} \Delta X_k^2 \left(\frac{\partial \Psi}{\partial u_k} \cdot \Delta U_k\right)^2 .$$

Если $\frac{\partial \Psi}{\partial U_k}$ и ΔU_k – независимы, то минимальное расстояние между сигналами $(d_{\min})$ должно удовлетворять условию:

$$(d_{\min})^2 = |\Delta X|^2_{\min} = \sum_{k=1}^{n}\left(\frac{\partial \Psi}{\partial U_k} \cdot \Delta U_k\right)^2 \leq \left(\frac{\partial \Psi}{\partial U}\right)^2_{\max} \cdot \sum_{k=1}^{n}(\Delta U_k) = \dot{\Psi}^2_{\max} \cdot r^2 ,$$

где $\dot{\Psi}_{\max} = \left(\frac{\partial \Psi}{\partial U}\right)_{\max}$;

r – расстояние между двумя ближайшими сообщениями.

Корректирующее кодирование применяют в тех случаях, когда возможности других более простых способов повышения помехозащищенности (таких как многократное повторение или фильтрация) исчерпаны. Это обусловлено усложнением систем передачи информации при введении кодирующих устройств.

Как следует из выше приведённых выводов теории помехоустойчивости, основной практический принцип построения корректирующих (помехозащищенных) кодов основывается на допущении, что принятый кодовый символ после воздействия на него помехи всегда остается в некоторой близости от кодового символа, однозначно соответствующего передаваемому символу исходного сообщения. Это позволяет в дальнейшем отождествить принятый символ с ближайшим к нему символом кодового алфавита путём нахождения минимального расстояния из всех расстояний между принятым символом и всеми символами кодового алфавита.

Создание специальных помехоустойчивых цифровых кодов, позволяющих не только обнаружить, но и при определенных условиях исправить возникающие ошибки, основано на увеличении минимального расстояния между символами кодового алфавита (d_{min}) путём. увеличения избыточности кодов.

Это достигается введением при кодировании дополнительных (избыточных) элементов символов кода, которые удовлетворяют некоторым дополнительным условиям и проверка которых дает возможность обнаружить и исправить ошибки.

Принципы практического построения корректирующих кодов удобно продемонстрировать на примере блочных систематических кодов (в частности алгебраических кодов), как наиболее наглядных и широко применяемых.

Для алгебраических (n, k) корректирующих кодов избыточность кода (R_n, R_k) определяется выражениями:

$$R_n = \frac{n-k}{n} = 1 - \frac{k}{n} = 1 - \frac{H}{H_m}, \qquad R_n = (0-1);$$

$$R_k = \frac{n-k}{k} = \frac{n}{k} - 1 = \frac{H_m}{H} - 1, \qquad R_k = (0-\infty), \tag{2.9}$$

где n — число элементов в кодовом символе;

k — число информационных элементов в кодовом символе (или минимально возможное число элементов в кодовом символе.

Величина R_k, с областью изменения от 0 до ∞ предпочтительнее, так как лучше отвечает смыслу понятия избыточности.

Коды, обеспечивающие заданную помехоустойчивость при минимально возможной избыточности, называют оптимальными.

Очевидно, что при оптимальном кодировании, когда лишние элементы в кодах отсутствуют, т.е. когда $n = k$, избыточность равна 0.

Для оценки кодов с точки зрения возможности обнаружения и исправления ошибок используют понятие кодового расстояния (число кодовых переходов для двоичных кодов) d_{ij}.

Кодовым расстоянием (d_{ij}) между кодовыми символами k_i и k_j называют суммарный результат сложения по модулю m_k одноименных разрядов кодовых символов k_i и k_j:

$$d_{ij} = \sum_{k=1}^{n} k_{ik} \oplus k_{jk},$$

где $k_{ik} \oplus k_{jk} = \begin{cases} k_{ik} + k_{jk}, & \text{если } k_{ik} + k_{jk} < m_k \\ k_{ik} + k_{jk} - m_k, & \text{если } k_{ik} + k_{jk} \geq m_k \end{cases}$;

k_{ik} и k_{jk} — k разряд кодовых символов k_i и k_j;

$\oplus$ — символ сложения по модулю m_k;

m_k — основание кода (основание системы счисления цифрового кода).

Например, если даны символы двоичного числового кода k_i = 10111 и k_j = 01010, то кодовое расстояние между этими символами равно 4 (d_{ij}= 4).

Кодовое расстояние может быть посчитано и для числовых кодов с иными основаниями, отличными от 2.

Кодовое расстояние иногда называют расстоянием Хемминга.

2.10. Некоторые методы построения блочных корректирующих кодов

1. Коды, построенные на основе увеличения кодового расстояния

Идея обнаружения и исправления ошибок в таких кодах заключается в следующем. Для передачи символов сообщения используют не все N возможных комбинаций элементов символов кода длиной n, а только часть из них N_0, которые называются разрешенными символами кода. Оставшиеся ΔN комбинаций ($\Delta N = N - N_0$) называют запрещенными. Ошибка обнаруживается тогда, когда приемник получает запрещенную комбинацию элементов символов кода. Всякий код, у которого $\Delta N > 0$, способен обнаруживать ошибки в ΔN случаях из N. Доля обнаруживаемых ошибок (s) определяется выражением:

$$\sigma = \frac{\Delta N}{N} = 1 - \frac{N_0}{N}.$$

По формуле (2.9) легко определить избыточность такого корректирующего кода (R_k):

$$R_k = \frac{H(N)}{H(N_0)} - 1 = \frac{n \cdot \log_2 m_k}{\log_2 N_0} - 1,$$

где $H(N_0)$—энтропия кода с объемом алфавита N_0 ;

$H(N)$—энтропия кода с объемом алфавита N;

n— число элементов в кодовом символе;

m_k— основание кода.

Исправление ошибок в корректирующих кодах заключается в том, что, обнаружив ошибку, определяют кодовое состояние от полученной запрещенной комбинации k_i до всех разрешенных символов кода k_j (j=1, 2, ..., N_0) и в качестве переданного символа кода считается тот из разрешенных символов кода, до которого расстояние минимально. Таким образом, исправление ошибок корректирующими кодами производится с помощью трех последовательных операций:

- определения кодового расстояния между принятой кодовой комбинацией и разрешенными кодовыми символами;
- отыскания минимального кодового расстояния;
- присвоение принятой кодовой комбинации значения ближайшего к ней (по кодовому расстоянию) разрешенного кодового символа.

Например, в трехзначном равномерном двоичном коде число возможных кодовых комбинаций 8:

000, 001, 010, 011, 100, 101, 010, 111,

кодовое расстояние между которыми (d_{ij}) равно 1. Ошибка в любом одном элементе символа кода превращает передаваемый разрешенный кодовый символ в другой, но так же разрешенный.

Если увеличить кодовое расстояние (d_{ij}=2), т.е. сделать разрешенными кодовые символы отличающимися в двух элементах (001, 111, 010, 100), то помехозащищенность кода увеличивается. Действительно, если в процессе передачи возникает ошибка только в одном элементе любого разрешенного символа, то эта комбинация превращается в запрещенную, что свидетельствует о появлении одиночной ошибки в символе кода, хотя и неизвестно какой именно. Такой код не позволяет выяснить какой конкретно из передаваемых элементов искажен, а значит и не позволяет исправить ошибку.

Если еще больше увеличить кодовое расстояние (d_{ij}=3), т. е. сделать разрешенными кодовые символы отличными друг от друга в трех элементах (например, 111, 000), то помехозащищенность кода еще более возрастает.

Действительно, если в процессе передачи произойдет ошибка в одном или двух элементах любого из двух разрешенных символов кода, то возникнет запрещенная комбинация элементов. Таким образом, применение этого кода дает возможность обнаружить две или, если произойдет только одна ошибка, что более вероятно, то можно восстановить переданный символ кода.

Из приведенного примера видно, что для обнаружения единичной ошибки в символе кода требуется один избыточный элемент в символах кода, чтобы обеспечить минимальное кодовое расстояние (расстояние между разрешенными символами (d_{ij})) равное 2, а для исправления в символах кода одной ошибки необходимо увеличить минимальное кодовое расстояние между символами кода (d_{ij}) до 3, т. е. ввести в символы кода два избыточных элемента. В общем случае для избыточных кодов справедливо соотношение:

$$d_{ij} = 1 + p + q\ ,$$

где $p \geq q$;

p— количество обнаруживаемых ошибок в символах;

q— количество исправляемых ошибок в символах кода.

Свойства корректирующих кодов по обнаружению и исправлению ошибок, в зависимости от величины минимального кодового расстояния между разрешенными символами, приведены в таблице 2.7.

Таблица 2.7

Характеристика кодов.			**Свойства кодов.**
d	*p*	*q*	
1	0	0	Отличает одну кодовую комбинацию от другой.
2	1	0	Обнаруживает одну ошибку.
3	1	1	Обнаруживает и исправляет одну ошибку.
	2	0	Обнаруживает две ошибки.
4	2	1	Обнаруживает две ошибки и исправляет одну.
	3	0	Обнаруживает три ошибки.
5	2	2	Обнаруживает и исправляет две ошибки.
	3	1	Обнаруживает три ошибки и исправляет одну.
	4	0	Обнаруживает четыре ошибки.
			и т. д.

2. Коды, построенные на основе проверки на четность.

Этот метод построения помехозащищенных кодов заключается в том, что в каждый кодовый символ двоичного кода добавляется для проверки один избыточный элемент (0 или 1) так, чтобы общее количество единиц в каждом символе кода было четным. Построение такого кода на примере двухэлементного двоичного кода представлено в Таблице 2.8.

Таблица 2.8.

Неизбьггочные кодовые символы.	Контрольный избыточный элемент.	Полные кодовые символы, обнаруживающие единичную ошибку.
00	0	000
01	1	011
10	1	101
11	0	110

При получении символов, таким образом построенных избыточных кодов, перед их раскодированием производится проверка их на четность. При одиночной ошибке количество единичных элементов в кодовом символе станет нечетным, что дает возможность обнаружить ошибку и осуществить повторную передачу.

Близким к рассмотренному способу построения корректирующих кодов является способ добавления контрольных избыточных элементов, на основе суммирования по основанию «2» элементов символов двоичного кода и на основе простейших логических операций над элементами символов кода.

3. На основе защиты сдвоенными элементами.

В этом коде количество элементов в символах двоичного кода удваивается, причем к каждому элементу «0» приписывается «1», а к каждому элементу «1» приписывается элемент «0». Например, исходный кодовый символ 11010 по этому методу кодируется следующим образом:

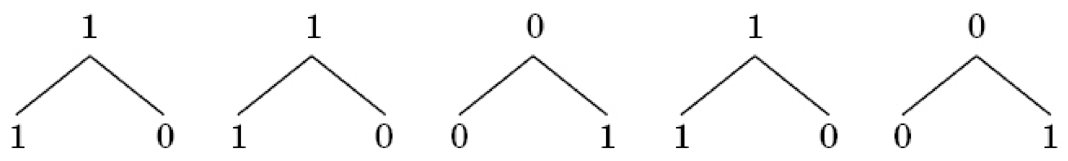


Рис. 2.7. Метод построения корректирующего кода на основе защиты сдвоенными элементами.

Полученный корректирующий код позволяет обнаружить одиночную ошибку в каждом разряде путем сложения по модулю 2 каждой пары элементов символа корректирующего кода. Если вследствие единичного сбоя в какой-либо паре сумма будет равна 0, то это является сигналом ошибки в данном разряде.

2.11. Кодирование как средство защиты информации от несанкционированного доступа.

Часто хранимая и передаваемая информация может представлять интерес для лиц, желающих использовать ее в своих интересах, поэтому важную роль играет информационная безопасность, которая должна обеспечить защиту конфиденциальной информации от ознакомления с ней конкурирующих групп. Это вынуждает использовать различные виды защиты информации от несанкционированного доступа, т. е. защиты информации от лиц, которым она не предназначена.

Существуют три основных способа защиты информации. Один из них предполагает защиту ее чисто силовыми методами – охрана документа (носителя информации) физическими лицами, передача его специальным курьером и т.п. Второй способ заключается в сокрытии самого факта наличия информации. Реализация этого метода чаще всего связана с использованием, так называемых, симпатических чернил, которые проявляются при соответствующем воздействии на документ, но могут использоваться и более экзотические методы. Например, один из них приведён в трудах древнегреческого историка Геродота. На голове раба, которая брилась наголо, записывалось нужное сообщение и, когда волосы достаточно отрастали, раба отправляли к адресату, который снова брил его голову и считывал сообщение. Третий способ защиты информации основывался на кодировании, т. е. на преобразовании сообщений (данных) в форму (код), при которой исходные сообщения становятся доступными только при наличии у получателя некоторой специфической информации (ключа), позволяющий выполнить обратное преобразование и получить исходное сообщение. Такой вид защиты информации называют криптографической защитой информации и осуществляют её специфическими операциями кодирования и декодирования, которые носят названия шифрования и дешифрования соответственно. Зашифрованное сообщение называют криптограммой, а область знаний о шифрах, методах их создания и раскрытия называется криптографией (или тайнописью). Свойство шифра противостоять раскрытию называется криптостойкостью (или надёжностью) и обычно измеряется сложностью алгоритма дешифрирования.

В практической криптографии криптостойкость шифра оценивается из экономических соображений. Если раскрытие шифра стоит (в денежном отношении, включая необходимые компьютерные ресурсы, специальные устройства и т.п.) больше, чем сама зашифрованная информация, то шифр считается достаточно надёжным.

Основное требование к шифру состоит в том, чтобы расшифровка была возможна только при наличии санкции, то есть некоторой дополнительной информации (или устройства), которая называется ключом шифра. Процесс декодирования шифровки без ключа называется дешифрированием (или просто раскрытием шифра).

Шифры, в которых для зашифровки и расшифровки используется один и тот же ключ, называются симметричными.

В настоящее время широкое распространение получили шифры с открытым ключом. Эти шифры не являются симметричными - для зашифровки и расшифровки используется различные ключи. При этом ключ, используемый для зашифровки, является открытым (не секретным) и может быть сообщен всем желающим отправить шифрованное сообщение, а ключ, используемый для расшифровки, является закрытым и хранится в секрете получателем шифрованных сообщений. Даже знание всего зашифрованного сообщения и открытого ключа, с помощью которого оно было зашифровано, не позволяет дешифрировать сообщение без знания закрытого ключа.

Важнейшей задачей криптографической защиты является обеспечение невозможности доступа к информации при условии, что потенциальный противник обладает любым техническим оборудованием, способным перехватить, записать криптограммы и ему известны некоторые фрагменты криптограмм и соответствующие им части исходного сообщения.

Методы криптографического закрытия могут иметь как программную, так и аппаратную реализацию.

Программная реализация осуществляется на основе вычислительных процессов, как на этапе шифрования, так и на этапе дешифрирования. Аппаратная реализация основана на использовании специальной аппаратуры.

Применение криптографии известно с глубокой древности и с тех пор известно большое число различных методов криптографического закрытия (шифров), как чисто информационных, так и механических, которые имеют различные степени сложности и надёжности защиты.

Представляет интерес рассмотрение некоторые из них на примере шифрования текстов.

<u>Шифр простой подстановки.</u> При этом методе шифровки все символы алфавита однозначно заменяют другими символами этого же или иного алфавита. Если объём алфавита исходного сообщения равен n и замена производится из этого же алфавита, то существует n! способов замены символов исходного сообщения, т. е. существует n! различных ключей.

Историческим примером шифра подстановки является шифр Цезаря (I век до н. э.). Применительно к тексту на русском языке он состоит в следующем. Выписывается алфавит, а затем под ним выписывается тот же алфавит, но со сдвигом, например, на 3 буквы:

а б в г д .... *ъ э ю я*

г д е ё ж .... *я а б в*

При шифровании буква *а* заменяется буквой *г, б – д* и так далее. Получатель сообщения проделывал обратную последовательность операций и восстанавливал исходное сообщение. Ключом в шифре Цезаря является величина сдвига алфавита в нижней строке, которая в принципе может быть любой.

В художественной литературе классическим примером шифра простой подстановки является шифр «Пляшущие человечки» (К. Дойль).

Примером такого типа шифров может являться, так называемый, книжный шифр, который заключается в том, что каким-либо малозаметным способом помечаются буквы секретного сообщения в тексте книги или другого печатного текста. Интересно отметить, что во время первой мировой войны германские шпионы использовали этот шифр, нанося симпатическими чернилами точки на букве газетного текста. Книжный шифр в современном его виде имеет несколько иной вид. Суть его состоит в замене на номер строки и номер этой буквы на заранее оговоренной странице некоторой книги. Ключом такого шифра является книга и используемая страница в ней. Этот шифр применялся даже во времена второй мировой войны.

Ещё одним примером шифра подстановки может служить, так называемый, квадрат Полибия. Шифрование в этом случае заключается в следующем. В квадратную матрицу с числом элементов равным или большему объёму алфавита на место каждого элемента в произвольном порядке вписываются все буквы алфавита. Шифруемая буква заменяется её координатами в матрице. При расшифровке каждая пара чисел определяла соответствующую букву сообщения. Ключом является расположение букв в исходной матрице. Отметим, что при произвольном расположении букв в матрице возникает некоторое затруднение: либо надо помнить отправителю и получателю сообщения заданное расположение букв в матрице (ключ шифра), что затруднительно, либо иметь при себе запись этого ключа, что представляет опасность ознакомления с ключом посторонних лиц. Поэтому в ряде случаев ключ представляют следующим образом. Берётся какое-либо «ключевое слово», которое легко запомнить, удаляют из него повторы букв и записывают его в начальных элементах матрицы. На место остальных элементов записывают остальные буквы алфавита в естественном порядке. Разновидностью такого шифра является, так называемый, «тюремный шифр» при котором матрица заполняется буквами в порядке их следования в алфавите.

Методы шифрования простой подстановкой достаточно просты, но не обеспечивают высокой степени защиты, так как буквы любого языка обладают определенной и различной вероятностью появления. В зашифрованном таким образом тексте статистические свойства исходного сообщения сохраняются, поэтому, анализируя криптограммы достаточной длительности, можно их дешифрировать исходя из их статистических свойств.

<u>Шифры перестановки.</u> Суть этого шифра заключается в следующем: берётся некоторое число n и записывается в строку ряд чисел 1, 2, …, n затем под ними записываются те же цифры в произвольном порядке. Например, для $n = 5$:

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \end{matrix}$$

После этого записывается шифруемое сообщение без пропусков и разбивается на группы по n букв. Если число букв до n не кратно n, то последняя группа дополняется до n любыми буквами. После этого буквы каждой группы переставляются в соответствии с выбранной двухстрочной таблицей: первая буква становится четвёртой, вторая – третьей и так далее. После выполнения перестановки в каждой группе, полученный текст записывается без пропусков. Ключом этого шифра является таблица перестановок. При дешифрировании криптограмма разбивается на группы по n букв и буквы перестанавливаются в обратном порядке.

Шифр Вижинера. Шифрование по этому методу заключается в следующем. Каждая буква алфавита нумеруется, скажем, для русского языка ставятся в соответствие цифры от 1 (А = 1) до 33 (Я = 33).

А	Б	В	…	Я
1	2	3	…	33

В качестве ключа используется некоторое слово или просто определённая последовательность букв. Этот ключ подписывается с повторением под шифруемым сообщением, так чтобы под каждой буквой исходного сообщения находилась одна буква ключа. Криптограмма формируется как последовательность цифр, получаемых в результате суммирования числовых эквивалентов, соответствующих букве исходного сообщения и букве стоящего под ней ключа и приведённой по модулю 33 (объём алфавита). Степень надёжности закрытия сообщений достаточно высока, так как этот шифр нарушает статистическое распределение вероятностей появления отдельных букв в сообщении.

Для обеспечения достаточно высокой надёжности закрытия необходимо использовать весьма длинные ключи, что сопряжено с определёнными трудностями. Шифр Вижинера с неограниченным неповторяющимся ключом известен как шифр Вернама.

Шифрование гаммированием. Суть этого метода шифрования заключается в том, что цифровые эквиваленты символов сообщения (букв) складываются с псевдослучайной последовательностью чисел, именуемой гаммой, и приводятся по модулю k, где k – объём алфавита источника сообщений. Таким образом, ключом в этом способе шифрования является псевдослучайная последовательность чисел.

Псевдослучайную последовательность генерируют на основе регистров сдвига с обратными связями. Соответствующим выбором обратных связей можно добиться генерирования последовательностей с периодом повторения $T = 2^n - 1$ символов, где n – число разрядов регистра. Получаемые таким образом последовательности чисел являются псевдослучайными, так как они удовлетворяют ряду основных тестов на случайность, что существенно затрудняет раскрытие такого ключа, и в то же время являются детерминированными, что позволяет обеспечить однозначность дешифрирования сообщений.

Надёжность криптографического закрытия методом гаммирования, в основном, зависит от длины периода неповторяющейся части гаммы и, если длина периода превышает длину шифруемого сообщения, то раскрыть криптограмму, опираясь только на статистические результаты обработки, теоретически невозможно.

Однако, если известно некоторое число цифровых эквивалентов символов сообщения и соответствующие им символы криптограммы, дешифрирование довольно легко выполнить, так как преобразование, осуществляемое при гаммировании, является линейным. Для полного раскрытия криптограммы достаточно всего $2n$ соответствующих пар символов исходного сообщения и символов криптограммы.

Глава 3. Передача информации по каналам связи.

3.1. Канал связи. Общие понятия.

Каналом связи или просто *каналом* называют совокупность технических средств и физических сред, обеспечивающих передачу сообщений из одной точки пространства в другую. В зависимости от того, для передачи какого вида сообщений (дискретных или непрерывных) используется канал, он называется дискретным или непрерывным соответственно.

Анализ процессов передачи сообщений по каналам во многом определяется влиянием шумов или помех. Если вредным воздействием помех в канале можно пренебречь, то при его анализе используют модель канала в виде идеализированного канала, называемого каналом без помех. Такому каналу присуща тождественность сообщений на входе и выходе канала.

Если же влияние помех значительно и пренебрежение несоответствия входного и выходного сообщения недопустимо, то используется более сложная модель канала – канал с помехами.

Теоретически, при передаче сообщений по каналом без помех никаких проблема связи, в этом идеальном случае, не возникает. Однако на практике, техническая реализация таких каналов трудноосуществима, да и сам процесс передачи сообщений по таким каналам проблематичен. Действительно, в этом случае любой непрерывный сигнал (импульс) конечной длительности, значение информационного параметра которого на выходе канала может быть измерено с неограниченной точностью, обладает бесконечным алфавитом, а, следовательно, с его помощью может быть передано бесконечно большое количество информации, что противоречит здравому смыслу.

Поэтому рассмотрение процесса передачи сообщений по каналу без помех нецелесообразно, и в практической деятельности имеет смысл рассматривать только процессы передачи сообщений по каналам с шумами.

Оценка качества канала, как правило, осуществляется по трём основным показателям: достоверности, средней скорости передачи и сложности технической реализации. Хотя с практической точки зрения сложность технической реализации может иметь решающее значение, при определении предельных (потенциальных) информационных возможностей системы целесообразно оценивать качество канала достоверностью и средней скоростью передачи. Достоверность дискретного канала обычно оценивается вероятностью ошибочного приема одного символа, а о достоверности непрерывного канала судят по значению среднеквадратической ошибки $(M(\varepsilon^2))$ при передаче сообщения.

$$M(\varepsilon^2) = M\left((x(t)-u(t))^2\right) = \frac{1}{T}\int_0^T (x(t)-u(t))^2 dt,$$

где $x(t)$ – сообщение на входе канала;

$u(t)$ – сообщение на выходе канала;

T – длительность сигнала.

Средняя скорость передачи (скорость передачи) характеризует среднее количество информации, передаваемое по каналу в единицу времени. Этот показатель тоже зависит от наличия шумов и помех. В любом реальном канале связи всегда имеют место помехи и искажения как детерминированного, так и случайного характера, что ведет к потере информации в канале. Действительно, если на передающем конце канала количество информации, приходящейся на один символ сообщения равно I_1, то вследствие воздействия шума после прохождения канала на один символ принятого сообщения будет приходиться количество информации равное I_2, причем

$$I_2 = I_1 - \Delta I,$$

где ΔI представляет собой количественную меру потери информации при передаче одного символа. Так как среднее количество информации, переносимое одним символом по каналу с шумом, уменьшается, то, очевидно, что снижается и пропускная способность канала с шумом по сравнению с каналом без шумов.

Достоверность или помехозащищенность канала – одна из важнейших характеристик канала. Повысить достоверность или помехозащищенность канала и тем самым застраховаться от ошибок при передаче сообщений по каналу связи с шумами можно путем многократных повторных передач, очевидно, чем большее число раз будет повторена передача одного и того же сообщения, тем больше достоверность принятого сообщения. Действительно, помехоустойчивость канала при многократных повторных передачах (Π_n) в соответствии с формулой (2.9) может быть найдена из выражения:

$$\Pi_n = \log_2 \left(\frac{1}{p_0} \right)^n = n \cdot \Pi ,$$

где n – число повторений;

p_o – вероятность появления ошибки;

Π – помехоустойчивость канала без повторения передач.

Очевидно, что многократное повторение сообщений ведет к росту помехозащищенности, но в тоже время значительно возрастает избыточность. Для полной достоверности передачи сообщений необходимо бесконечно увеличивать число повторных передач, при этом бесконечно возрастает избыточность и, следовательно, пропускная способность канала будет стремиться к нулю, что неприемлемо с практической точки зрения.

Другим путем обеспечения достоверной передачи сообщений по каналу с шумами является рациональное помехоустойчивое кодирование, теоретическим обоснованием которого служит теорема Шеннона (§ 1.9). Эта теорема, применительно к каналам передачи информации, может быть сформулирована следующим образом.

Пусть дискретный канал обладает пропускной способностью C, а дискретный источник сообщения имеет производительность H (бит/сек), т. е. источник сообщения создает H единиц информации в секунду. Если $H \leq C$, то существует такая система кодирования, что сообщения могут быть переданы по каналу с произвольно малой вероятностью ошибок (со сколь угодно малой ненадежностью). Однако, если $H > C$, то наименьшая ненадежность, которая может быть достигнута применением того или иного способа кодирования ограничена пределом разности $(H - C)$.

Из теоремы следует, что для того, чтобы осуществить передачу с наперед заданной малой вероятностью ошибок, нужно уменьшить скорость передачи сообщений так, чтобы она стала равной или меньшей пропускной способности канала.

3.2. Передача дискретных сообщений по каналам связи.

Канал связи представляет собой совокупность технических средств и физических сред, предназначенную для передачи сообщений из одной точки пространства в другую. Эта передача чаще всего осуществляется в условиях неизбежных помех. В результате воздействия помех каждый отправленный символ x_i может быть опознан получателем как символ y_k, причем $y_k \neq x_i$. Такое событие называют ошибкой.

Передачу символов сообщения можно рассматривать как составной эксперимент, состоящий в отправлении символов сообщения x_i и получения символов y_k. С точки зрения теории информации физическое устройство канала несущественно, а свойства канала при этом полностью описываются матрицей переходных вероятностей $p\left(x_i/y_k\right)$ или $p\left(y_k/x_i\right)$,

где $p\left(x_i/y_k\right)$ есть вероятность передачи символа x_i, если зафиксирован полученный символ y_k,

$p\left(y_k/x_i\right)$ – вероятность получения символа y_k, если зафиксирован (передается) символ x_i.

При этом предполагается, что новые символы (сверх заданного объема алфавита m) не могут быть созданы под влиянием помех.

Следовательно,

$$\sum_{k=1}^{m} p\left(y_k/x_i\right) = 1;$$

$$\sum_{i=1}^{m} p\left(x_i/y_k\right) = 1.$$

Если помехи отсутствуют, то все диагональные элементы матрицы $p\left(y_k/x_i\right)$ или матрицы $p\left(x_i/y_k\right)$ равны единице, а остальные – нулю. При очень больших помехах все элементы матриц могут быть приблизительно одинаковыми.

При наличии помех, передача символа x_i не снимает полностью неопределенность относительно полученного символа y_k. Таким образом, передача символов по каналу описывается ниже перечисленными мерами неопределенности (энтропиями).

Неопределенность передаваемых символов при условии их независимости $(H(x))$:

$$H(x) = -\sum_{i=1}^{m} P(x_i) \cdot \log_2 P(x_i).$$

Неопределенность полученных символов $(H(y))$:

$$H(y) = -\sum_{k=1}^{m} P(y_k) \cdot \log_2 P(y_k).$$

Неопределенность получения символов при зафиксированном символе $x_i \left(H\left(y/x_i\right)\right)$:

$$H\left(\frac{y}{x_i}\right) = -\sum_{k=1}^{m} P\left(\frac{y_k}{x_i}\right) \cdot \log_2 P\left(\frac{y_k}{x_i}\right).$$

Эта величина называется частной энтропией принятых символов.

Полная энтропия принятых символов вычисляется усреднением $H\left(\frac{y}{x_i}\right)$ по вероятностям передаваемых символов x_i:

$$H\left(\frac{y}{x}\right) = \sum_{i=1}^{m} H\left(\frac{y}{x_i}\right) \cdot P(x_i)$$

Величину $H\left(\frac{y}{x}\right)$ называют средней условной энтропией принимаемых символов.

Неопределенность передаваемых символов при зафиксированном принятом символе $y_k \left(H\left(\frac{x}{y_k}\right)\right)$:

$$H\left(\frac{x}{y_k}\right) = -\sum_{i=1}^{M} P\left(\frac{x_i}{y_k}\right) \cdot \log_2 P\left(\frac{x_i}{y_k}\right)$$

Эта величина является частной энтропией передаваемых символов.

Полную энтропию передаваемых символов находят усреднением энтропии $H\left(\frac{x}{y_k}\right)$ по вероятностям принимаемых символов y_k:

$$H\left(\frac{x}{y}\right) = \sum_{k=1}^{m} H\left(\frac{x}{y_k}\right) \cdot P(y_k).$$

В соответствии с основным соотношением теории информации (1.3), прирост количества информации (I), связанный с приемом одного символа сообщения, определяется выражением:

$$I = \log_2 \frac{P_2}{P_1} = \log_2 P_2 - \log_2 P_1 = H(x) - H\left(\frac{x}{y}\right) \qquad (3.1)$$

где P_1 – априорная вероятность появления символа;

P_2 – апостериорная вероятность появления этого же символа.

Справедливо также соотношение:

$$I = H(y) - H\left(\frac{y}{x}\right). \qquad (3.2)$$

Из выражений (3.1) и (3.2) видно, что, по мере уменьшения помех, величина I будет стремиться к $H(x)$, а при увеличении помех будет стремиться к нулю.

3.3. Передача непрерывных сообщений по каналам связи.

Так же как в случае передачи дискретных сообщений, при наличии шумов непрерывное сообщение $x(t)$ после прохождения канала может быть принято как сообщение $y(t)$, причем $y(t) \neq a \cdot x(t+\tau)$ для всех или некоторых моментов времени, где a и τ - константы, характеризующие ослабление и запаздывание сигнала, которые обычно несущественны с точки зрения определения количества информации, содержащегося в сообщении. Такое событие называют ошибкой.

Источник непрерывных сообщений удобно представить в виде непрерывной случайной величины X , которая после прохождения канала преобразуется в непрерывную случайную величину Y . Количество информации в непрерывных случайных величинах X и Y , при наличии шумов, можно определить предварительно проведя их дискретизацию по уровню (т.е. преобразовав их в дискретные случайные величины), а после определения содержащегося в них количества информации выполнив предельный переход, когда число уровней дискретизации стремится к бесконечности. Эта операция подробно рассмотрена в основах теории информации (§1.8.).

В результате выполнения операций дискретизации по уровню и предельного перехода можно сделать вывод: дифференциальные энтропии случайных непрерывных величин X и Y конечны, хотя их полная энтропия и стремится к бесконечности.

Аналогичным образом можно определить количество информации (I) случайной величины X , содержащейся в непрерывной случайной величине Y:

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{XY}(x,y)\cdot \log_2 \frac{P_{XY}(x,y)}{P_X(x)\cdot P_Y(y)}\cdot d \tag{3.3}$$

где $P_{XY}(x,y) = P_X \cdot P_Y\left(y/x\right)$ – совместная плотность распределения вероятности величин X и Y;

$P_X(x), P_Y(y)$ – плотность распределения случайных величин X и Y соответственно.

Из анализа формулы (3.3) можно сделать следующие выводы.

1. Если помехи столь велики, что случайная величина Y практически не зависит от случайной величины X , т.е. $P_Y\left(y/x\right) = P_Y(y)$ то

$$P_{XY}(x,y) = P_X(x)\cdot P_Y(y)$$

и, следовательно, количество получаемой информации (I), вычисленное по формуле (3.3), равно нулю.

2. Если помеха (n), представляемая в виде непрерывной случайной величины. носит аддитивный характер, то есть

$$Y = X + n,$$

причем, случайная величина X и помеха n независимы (что обычно выполняется), то

$$P_Y\left(y/x\right) = P_n(y - x),$$

где $P_n(y - x) = P_n(n)$ – плотность распределения вероятности помехи.

В этом случае,

$$I = H(y) - H(n),$$

где,

$H(y) = -\int_{-\infty}^{\infty} P_Y(y) \cdot \log_2 P_Y(y) dy$ – дифференциальная энтропия сигнала Y;

$H(n) = -\int_{-\infty}^{\infty} P_n(n) \cdot \log_2 P_n(n) dn$ – дифференциальная энтропия шума n.

3.4. Согласование каналов с сигналами.

Наряду с информацией, передаваемой в виде дискретных сигналов, в ряде случаев приходится передавать по каналам связи информацию в непрерывной или аналоговой форме. Такие сигналы характеризуются мощностью (или интенсивностью), частотным диапазоном и длительностью. Важнейшее значение имеет мощность сигнала, однако практически, из-за наличия шумов, свойства сигнала как переносчика информации определяются не абсолютной его величиной, а превышением уровня сигнала над уровнем помех. За количественную меру интенсивности сигнала принимают отношение (H_c) мощности сигнала (P_c) к мощности действующей в канале помехи ($P_П$):

$$H_c = \frac{P_c}{P_П},$$

или же логарифм этого отношения в виде:

$$H_c = 10 \lg \frac{P_c}{P_П},$$

единица измерения которого носит название децибел (дБ).

Следующим важнейшим параметром любого сигнала является ширина его полосы частот или частотный диапазон F_c, представляющий собой разность между максимальной $f_{\max}$ и минимальной $f_{\min}$ частотами, которые содержатся в спектре сигнала.

$$F_c = f_{\max} - f_{\min}$$

Сигнал характеризуется также длительностью T_c равной разности между временем окончания сигнала t_k и временем его начала $t_н$:

$$T_c = t_k - t_н$$

Произведение $V_c = H_c F_c T_c$ называется объемом сигнала.

Аналогичными параметрами можно характеризовать и канал связи (канал передачи сообщений), скажем H_k – допустимый диапазон изменения мощности в канале, F_k– ширина спектра частот, пропускаемых каналом, T_k – время занятия канала.

Тогда свойства канала можно характеризовать произведением

$$V_k = H_k F_k T_k,$$

которое называют емкостью канала.

Очевидно, что передача сигнала по каналу связи возможна лишь при соблюдении условий:

$$H_k \geq H_c,$$
$$F_k \geq F_c, \qquad (3.4)$$
$$T_k \geq T_c.$$

Из этих ограничений следует, что $H_k \cdot F_k \cdot T_k \geq H_c \cdot F_c \cdot T_c$ и, следовательно, $V_k \geq V_k$. Таким образом, сигнал можно передать по каналу связи без искажений только в том случае, если емкость канала больше или равна объему сигнала. Ограничения (3.4), касающиеся отдельных параметров сигнала: мощности, частоты и времени, являются относительными и при определенных условиях, за счет некоторых преобразований исходного сигнала, могут быть преодолены.

Основным назначением любой информационной системы, а следовательно и канала связи, является передача сообщений, содержащих какую-либо информацию, с заданным качеством.

По аналогии с информационными характеристиками источников сообщений (§ 1.6), для каналов передачи сообщений так же можно использовать информационные параметрами, характеризующие их качество. Одной из важнейших характеристик, определяющих качество канала, является *скорость передачи информации* (Q), под которой понимают среднее количество информации передаваемое по каналу в единицу времени. Для вычисления значения Q можно использовать соотношение:

$$Q = W \cdot H,$$

где W - техническая скорость передачи символов сообщений, под которой подразумевают число элементарных сигналов (символов), передаваемых по каналу в единицу времени (символ / сек.).

H – среднее количество информации, приходящиеся на один элементарный сигнал или символ (бит / символ).

Техническая скорость передачи символов сообщений (*W*) зависит от свойств и технических средств канала. С учетом возможных различий в длительностях символов, можно считать, что

$$W = \frac{1}{\tau_{cp}},$$

где - среднее значение длительности символа.

За единицу измерения скорости передачи информации (*Q)* принимается *бот* = [бит/сек].

Для теории и практики важно знать, до какого предела и каким путем можно повысить скорость передачи информации по конкретному каналу, а поэтому другой важнейшей характеристикой канала является предельное количество информации, которое может быть передано по нему за единицу времени. Данная характеристика получила название *пропускной способности канала (С)*, т.е.

$$C = \max Q = \max(W \cdot H) = W \cdot \max H = W \cdot \log_2 m,$$

где m – объём алфавита передаваемого сообщения.

Пропуская способность канала измеряется также как и скорость передачи информации в ботах.

Пропускная способность канала имеет ограниченное значение, как при передаче дискретных сообщений, так и непрерывных сигналов. Это объясняется тем, что при увеличении числа возможных значений сигнала (x_k) увеличение количества передаваемой информации возможно лишь до тех пор, пока возможно уверенное распознавание этих значений. Если же отличия между соседними значениями сигнала (Δx) становится меньше среднего уровня шума, то увеличение числа возможных значений сигнала не увеличивает количества передаваемой информации. Поэтому пропускная способность канала растет с увеличением отношения полезного сигнала к шуму (отношение сигнал / шум). Количественное соотношение между пропускной способностью канала связи (C) при передаче непрерывных сообщений, его частотными характеристиками и средними мощностями сигнала и шума (P_c и $P_ш$) определяется соотношением:

$$C \le f_B \cdot \log_2\left(1 + \frac{P_c}{P_ш}\right) \text{ (бит/сек)} \qquad (3.26)$$

де f_B – верхняя граничная частота в спектре частот канала.

Между пропускной способностью канала и скоростью передачи информации существуют предельные соотношения, вытекающие из второй теоремы Шеннона (см. § 2.9): если скорость создания информации (H) меньше пропускной способности (C), то имеется правило кодирования (возможно неизвестное), при котором вероятность ошибок может быть сколь угодно малой, а скорость передачи информации может сколь угодно приближаться к пропускной способности канала.

Глава 4. Использование информационных моделей при анализе систем автоматической обработки изображений и их основных узлов.

4.1. Информационные модели систем автоматической обработки изображений

В современной науке и технике широко используются системы, служащие для сбора, обработки или передачи информации. Основное отличие таких систем состоит в том, что результатом их функционирования являются не материальные объекты реального мира, а лишь сведения или знания о них (информация) и поэтому основными характеристиками этих систем являются информационные характеристики. Такого рода системы носят название информационных систем.

К информационным системам следует отнести и системы автоматической обработки результатов дистанционного зондирования. Исходным источником информации в таких системах чаще всего служат оптические или фотографические изображения, полученные в результате регистрации распределения отраженных или излученных электромагнитных волн оптического диапазона, а цель обработки заключается в получении каких-либо параметров изображения, его передачи или преобразовании. Поэтому такие системы принято называть системами автоматической обработки изображений.

Сложность проектирования и исследования информационных систем требует их моделирования с целью оценки их информационных свойств, причем, учитывая специфику этих систем, основным видом их моделирования должно быть математическое информационное моделирование.

Информационные модели - это вид математических моделей, которые позволяют моделировать информационные процессы в различных системах.

Общая информационная модель любой системы может быть представлена в виде, изображенном на рис. 4.1.

Источни
сообщени

Рис. 4.1. Общая информационная модель системы.

Отдельные блоки, входящие в общую информационную модель, определяются следующим образом.

Источник сообщений - физический объект, обладающий возможностью генерировать сообщение.

Датчик сообщений - техническое устройство, преобразующее исходное сообщение в изоморфную (аналогичную) форму, удобную для дальнейшей передачи или обработки.

Канал - техническое устройство, представляющее собой последовате6льность преобразователей, которые передают или обрабатывают исходное сообщение по заданному алгоритму.

Приемник сообщений - техническое устройство, с помощью которого переданное или обработанное сообщение преобразуется в форму, удобную для получателя.

Получатель - некий объект, способный принимать решения на основе полученной или обработанной соответствующим образом информации. Так, в качестве получателя сообщения может быть или человек, или логическое устройство, или регистрирующий прибор.

Источник помех - совокупность помех и шумов, действующих на отдельные блоки информационной модели, приведенных к каналу.

Следует отметить, что в конкретных информационных моделях могут отсутствовать один или даже несколько блоков, приведенных в общей информационной модели.

Информационная модель в общем виде может быть применена при моделировании самых различных устройств и процессов, связанных с передачей и обработкой информации. Однако при моделировании конкретного устройства или процесса гораздо удобнее пользоваться такими информационными моделями, в которых указаны конкретные функциональные блоки, входящие непосредственно в эти устройства. На рис. 4.2 в качестве примера представлена информационная модель системы автоматической обработки фотоизображений на основе ЭВМ. При этом наряду с функциональными блоками системы показано и их обозначение в терминах информационной модели общего вида.

Р

ис.

4.2.

Инфо

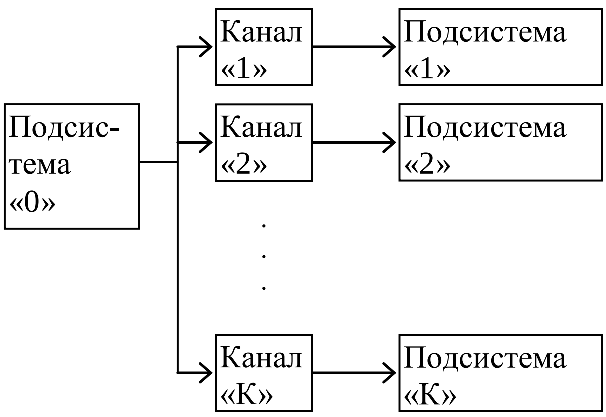
рмаци

онная модель системы автоматической обработки фотоизображений на основе ЭВМ.

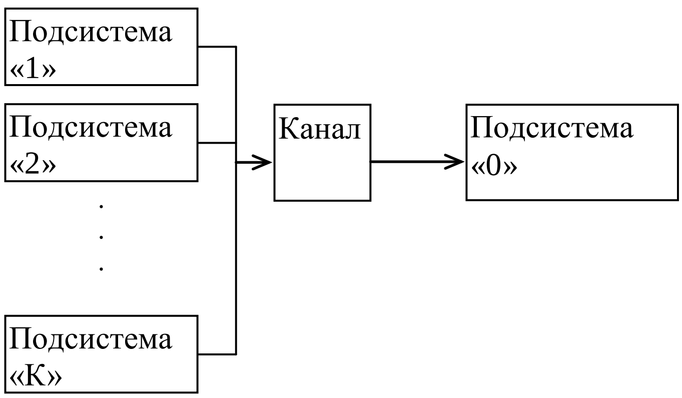
Современные информационные системы, в частности, системы, используемые для исследования природных ресурсов, представляют собой довольно сложные структуры. Поэтому построение их информационных моделей, как правило, требует их предварительного разбиения на несколько подсистем, после чего можно переходить к информационному моделированию каждой отдельной подсистемы с учетом их информационной согласованности.

В большинстве случаев построение информационной модели сложной системы можно представить в виде одной из следующих трех основных блок-схем, изображенных на рис. 4.3, или их комбинаций:

а). Последовательное соединение подсистем



б). Соединение подсистем с параллельной передачей информации



в). Соединение подсистем с параллельным сбором информации

Рис. 4.3. Основные блок-схемы построения информационных моделей сложных систем.

Следует отметить, что каналы связи между отдельным и подсистемами могут и отсутствовать (обычно в тех случаях, когда получатель сообщения одной подсистемы служит источником сообщения другой подсистемы). Например, если требуется построить информационную модель автоматической системы получения изображений земной поверхности из космоса, которая состоит из фоторегистрирующей системы, установленной на борту космического аппарата, и системы передачи изображений с космического аппарата на Землю, то удобно рассмотреть информационную модель этой системы как последовательную комбинацию двух информационных моделей: информационной модели получения фотоизображения поверхности Земли и информационной модели передачи изображений на Землю. В этом случае фотоизображение земной поверхности, принятое на борт КА, будет для первой информационной системы получателем сообщения, а для второй - источником сообщения.

Широкое применение систем автоматической обработки изображений и необходимость при их разработке и эксплуатации учета информационных характеристик требует более подробного рассмотрения присущих этим системам специфических составляющих и процессов в них происходящих с точки зрения информационного моделирования.

Источники сообщений, входящие в информационные модели систем автоматической обработки изображений, с целью подчеркивания их специфики, обыкновенно называют источниками визуальных сообщений. При этом источник визуальных сообщений определяют как любой физический процесс, объект или явление, происходящий в некоей физической среде и сопровождающийся излучением или отражением лучистой энергии, при условии, что мгновенные состояния этого процесса с достаточной полнотой и точностью характеризуются мгновенными пространственными распределениями мощностей и спектральных составов элементарных потоков лучистой энергии, излучаемых (или отражаемых) различными областями физической среды, в которой протекает данный процесс или явление.

Приведенное определение требует дополнительных пояснений. Строго говоря, термин «источник визуальных сообщений» следует применять только к тем источникам, спектр лучистой энергии которых не выходит за пределы видимой части оптического спектра, то есть от 0,450 до 0,700мкм. Однако, в связи с тем, что основные физические свойства электромагнитных излучений сохраняются в гораздо более широком диапазоне длин волн и что аналогичные по принципу действия преобразователи лучистой энергии, выступающие в качестве датчиков сообщений, также могут быть использованы в более широком частотном диапазоне, к источникам визуальных сообщений относят источники сообщений, спектр которых принадлежит всей оптической области электромагнитных излучений с длинами волн от 0,01мкм до 300мкм.

За элементарный поток лучистой энергии будем принимать поток, излучаемый (или отражаемый) элементарной площадкой таких размеров, что его мощность и спектральный состав могут восприниматься только как интегральные величины. Примером элементарного потока лучистой энергии может служить поток лучистой энергии, излучаемый звездой или точечным источником света.

По аналогии с терминологией, принятой в теории информации, символом визуального сообщения называют любое мгновенное состояние источника визуальных сообщений, то есть отдельное (мгновенное) пространственное распределение мощностей и спектральных составов потоков лучистой энергии, излучаемых элементарными площадками источника визуальных сообщений. В этом случае все множество различных (неповторяющихся) символов, генерируемых источником визуальных сообщений, называют алфавитом источника визуальных сообщений, число различных символов – объёмом алфавита источника визуальных сообщений, а любую последовательность символов визуального сообщения, имеющую для получателя сообщения законченное смысловое (семантическое) значение - визуальным сообщением.

Если взять в качестве примера аэрофотосъемку земной поверхности, то в терминах информационных моделей можно считать земную поверхность источником визуальных сообщений, последовательность аэрофотоснимков, относящихся к одному маршруту, - визуальным сообщением, а каждый отдельный аэроснимок - символом визуального сообщения. Однако отметим, что при информационном моделировании других процессов аэрофотоснимок уже можно рассматривать как источник визуальных сообщений.

Поверхность любого источника визуальных сообщений *Q* (Рис. 4.4) можно разбить на элементарные площадки таким образом, что вся поверхность источника визуальных сообщений будет равна сумме всех элементарных площадок, то есть .

Рис. 4.4. Разбиение поверхности источника визуальных сообщений на совокупность элементарных площадок, При этом каждый символ визуального сообщения ( ), генерируемый таким источником в момент времени , в общем случае будет описываться мгновенным пространственным распределением мощностей и спектральных составов элементарных потоков лучистой энергии, излучаемой элементарными площадками:

$$E_k = E(t_k) = \{W(x, y, z); \Lambda(x, y, z)\}_{t=t_k}, \qquad (4.1)$$

где - декартовы координаты центров тяжести элементарных площадок.

Следовательно, математической моделью визуального сообщения будет последовательность во времени дискретных распределений при $k = 0, 1, 2, ..n$. Такая модель позволяет охватить все множество источников визуальных сообщений (ИВС), различных по физической природе и информационным свойствам, однако на практике представляет интерес разбиение всего множества ИВС на классы, в которые входят ИВС однородные по своей структуре, физической природе и информационным свойствам.

Представляется рациональным положить в основу классификации ИВС выражение (5.4) и классифицировать их по размерности дискретных функций и и по размерности дискретных распределений , описывающих символы визуальных сообщений.

Очевидно, что ИВС могут быть одномерными - , и двумерными - , а размерность функций и может принимать значения от 0 до 3.

В особый класс можно выделить источники визуальных сообщений, для которых

и , (4.2)

то есть распределения W и Λ не зависят от координат пространства. По своим физическим свойствам они аналогичны точечным источникам и полностью характеризуются значением величин W и Λ. Поэтому их называют интегральными источниками визуальных сообщений.

Все остальные источники, для которых хотя бы одно условие (4.2) не выполняется - относятся к классу дифференциальных источников визуальных сообщений.

Если для дифференциальных ИВС справедливо

(4.3)

то такие ИВС называют дифференциальными объемными.

Соответственно, ИВС, для которых

и (4.4)

называют дифференциальными плоскими (площадными) источниками.

Дифференциальные ИВС, для которых

; и ; (4.5)

называют дифференциальными линейными источниками.

Исходя из этого на рис.4.5 представлена классификация источников визуальных сообщений.

Рис. 4.5. Классификация источников визуальных сообщений.

Примерами различных классов источников визуальных сообщений (ИВС) могут служить:

– интегральные одномерные: оптический телеграф ( ), светофор ( );
– интегральный двумерный: излучение звезд;
– дифференциальный одномерный линейный: оптический плоскостной клин ( ), цветовой клин ( );
– дифференциальный одномерный плоский: черно-белое фотоизображение ( ), цветовое поле ( );
– дифференциальный одномерный объемный: черно-белое стереоизображение ( ), цветовое тело ( );
– дифференциальный двумерный линейный: спектрограф;

– дифференциальный двумерный плоский: цветное фотоизображение;

– дифференциальный двумерный объемный: цветное стереоизображение.

При дистанционных методах зондирования в качестве источников сообщения используют дифференциальные плоские и объемные (как одномерные, так и двумерные) ИВС. При технической же реализации систем автоматической обработки результатов дистанционного зондирования объемные ИВС, как правило, приводят путем соответствующих проективных преобразований к совокупности нескольких плоских ИВС, например, объемный ИВС заменяют стереопарой плоских ИВС, а различные участки оптической области электромагнитных излучений ( ), которым принадлежат спектры исследуемых ИВС, однозначно преобразуются в видимую часть оптического диапазона с помощью соответствующих датчиков и регистрирующих устройств, например, инфракрасные ИВС визуализируют в видимой области. Указанные преобразования выполняют с целью получения ИВС приемлемых для субъективного визуального контроля и использования, а также для сокращения номенклатуры видов ИВС и, следовательно, числа различных датчиков, используемых в системах автоматической обработки результатов дистанционного зондирования.

Таким образом, в системах автоматической обработки изображений в качестве источников сообщений в подавляющем числе случаев используются (или приводятся к ним) дифференциальные плоские одномерные и двумерные ИВС со спектром электромагнитных излучений, лежащем в видимой части диапазона оптического излучения, а символы таких ИВС традиционно называют оптическими изображениями или просто изображениями (черно-белыми или цветными соответственно). В дальнейшем при рассмотрении информационных аспектов систем автоматической обработки изображений именно такие ИВС и генерируемые ими символы (изображения) используются в них в качестве источников сообщения, при этом полученные результаты могут быть распространены и на другие виды ИВС.

4.2. Информационная оценка качества оптических изображений.

При проектировании автоматизированных систем обработки и передачи изображений различного назначения одним из основных требований, предъявляемых к ним, является обеспечение заданного качества получения и воспроизведения изображений, что требует выбора количественного критерия для его объективной оценки. По общепринятой методологии количественная оценка (измерение) характеристик любого физического объекта или явления подразумевает определение некоей меры их оценки и выбор единиц измерения этой меры. В связи с чем и количественная оценка качества изображения, как характеристики оптического изображения, требует выбора меры для оценки качества изображения и единиц ее измерения.

В большинстве применяемых в настоящее время критериев оценки качества черно-белого изображения за меру этой оценки принимается разрешающая способность (QR), которая зависит как от пространственного разрешения изображения, характеризующегося максимально возможным количеством различимых черных и белых линий на единицу длины, так и от фотометрического разрешения (контраста), оценивающегося числом различимых уровней (градаций) используемой оптической характеристики изображения, например, оптической плотности. Числовой интегрированной оценкой качества изображения при указанном подходе выступает частотно-контрастная характеристика (ЧКХ), определяющая максимальное число различимых черных и белых линий на единицу длины при заданном контрасте.

Однако из анализа критериев оценки качества изображений, основанных на разрешающей способности, следует ряд замечаний.

Во-первых, практическое применение критериев QR часто ведет к завышению разрешающей способности изображений и, как следствие, их качества, вызванное субъективными особенностями этих методов, которые проявляются в следующем:

– линейные объекты разрешаются глазом значительно лучше, чем точечные ;
– часто интересующие наблюдателя объекты имеют меньший контраст, чем контраст черного и белого при использовании стандартных мир, например при обработке изображений звездного неба;
– анализ изображения, как правило, проводится на некотором его фрагменте и в силу специфики зрения, заключающейся в обобщении и домысливании, получаемое разрешение оказывается больше, чем разрешение изображения того же качества, состоящее из двух отдельных точек (семантическая оценка).

Во-вторых, критерии на основе QR не позволяют однозначно учесть влияние условий получения изображений на их качество. Например, установить, какое из двух изображений с точки зрения качества лучше: полученное на фотоматериале с высоким пространственным разрешением и низким фотометрическим разрешением или изображение, зарегистрированное на фотоматериале с низким пространственным разрешением и высоким фотометрическим разрешением.

В-третьих, следует отметить, что хотя критерии QR и позволяют оценить качество конкретного изображения, которое может варьироваться в весьма широких пределах в зависимости от объекта и условий съемки, тем не менее они не позволяют установить граничные значения параметров, характеризующих качество изображения с учетом заданных условий и целей их получения.

В-четвертых, оценка качества изображения по критерию QR, как правило, не является корректной, так как оба вида разрешения: пространственное и фотометрическое взаимозависимы.

И, наконец, следует отметить, что практическая оценка качества изображения по критерию QR (например, путем впечатывания в него стандартных мир) подразумевает, что определяемое таким образом качество изображения, как и само изображение, считается детерминированным и статическим, хотя реально оцениваемое качество изображения, как было указано выше, может варьироваться в зависимости от условий получения изображения и целей использования.

Перечисленные замечания не позволяют осуществить всесторонне объективную оценку качества изображения по критерию QR.

Другой подход к выбору критерия оценки качества изображения может быть основан на том факте, что любое изображение представляет собой заданного вида сообщение, содержащее информацию о каком-либо объекте или процессе и зависящее от условий формирования этого изображения. При этом за меру оценки качества изображения предполагается принять наибольшее количество информации, которое может в нем содержаться, а за единицу измерения этой меры количество информации, выраженное в битах, отнесенное к единице площади изображения, то есть энтропию изображения. В дальнейшем введенный таким образом критерий оценки качества изображения будем определять как энтропийный и обозначать его QI.

Критерий QI позволяет, во-первых, для каждого конкретного изображения при оценке его качества учитывать условия его формирования, и, во-вторых, если данное изображение подвержено дальнейшей обработке или преобразованию, то повторное применение этой оценки к изображению, полученному после указанных операций, дает возможность установить потерю информации или степень ее использования.

С физической точки зрения изображение можно охарактеризовать как результат отображения на фиксированный момент времени и заданной ограниченной поверхности (плоскости) совокупности интегральных источников излучения или отражения, расположенных вне заданной поверхности. Данное толкование позволяет перейти к информационному подходу представления изображения, по которому изображение - это результат преобразования, передачи и воспроизведения информации, содержащейся в источнике визуального сообщения (излучающей или отражающей поверхности). При этом предполагается, что источник визуального сообщения (ИВС) в каждый момент времени соответствует одному из символов алфавита, рассматриваемого источника (изображению), характеризующему пространственное распределение интегральных потоков лучистой энергии, излучаемых (или отражаемых) отдельными элементарными участками излучающей поверхности.

Информационный подход представления изображения и введенный критерий *QI* позволяют отождествить понятие качества оптического изображения с максимальным количеством информации, содержащимся в символе ИВС, где под символом ИВС (изображением) традиционно понимается любое мгновенное состояние ИВС, характеризующееся мгновенным распределением энергии и спектрального состава излучения, генерируемого элементарными участками ИВС.

Для получения расчетных формул оценки качества оптического изображения по критерию *QI* ограничим область задания изображения прямоугольным участком *Q*, разбитым регулярной квадратной сеткой на элементарные участки (пиксели, элементы разложения) такого размера, что в каждом из них значение рассматриваемого информационного параметра оптического изображения, например, яркости (*U*) является постоянным (Рис. 4.6).

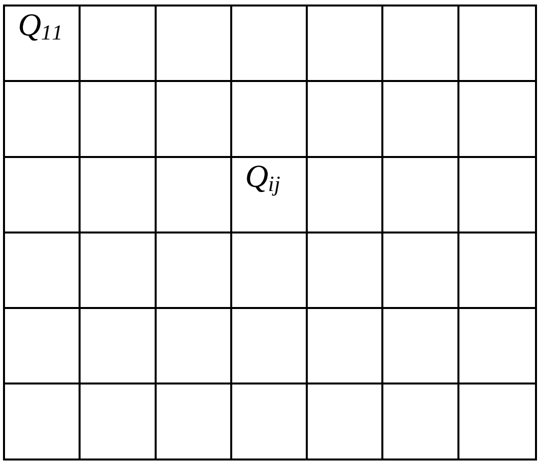


Рис. 4.6. Разбиение прямоугольного участка изображения *Q* регулярной квадратной сеткой на элементарные участки .

Отметим, что такое разбиение всегда можно выполнить, так как любому реальному процессу или явлению свойственна инерционность (то есть конечное время или пространство переходного процесса) и поэтому всегда будет существовать достаточно малая, но конечная часть изображения, которую можно считать интегральным источником оптического сообщения. Предполагая, что значения равновероятны и некоррелированы (в этом случае энтропия максимальна), количество информации (*I*), содержащееся в любой подобласти , обладающей площадью , можно найти исходя из определения количества информации по Хартли согласно формуле:

$$, \qquad (4.6)$$

где - общее число пикселов, содержащихся в подобласти ;

, - значения площадей соответственно подобласти и пиксела ;

m - объем алфавита пикселов, то есть количество различимых дискретных значений информационного параметра *U*.

Очевидно, что для любых реальных ИВС число дискретных значений информационного параметра - конечно, так как при их формировании всегда присутствуют шумы, обусловленные молекулярной структурой физических объектов и квантовой природой лучистой энергии. Поэтому, если известен интервал допустимых значений информационного параметра (*U*) и величина шума ( ), то значение *m* может быть определено из соотношения:

$$.$$

В результате соотношение (4.6) преобразуется к виду:

$$, \qquad (4.7)$$

а максимально возможное количество информации, содержащееся в ИВС площадью , будет определяться выражением:

$$. \qquad (4.8)$$

Если выбрать подобласть *Q* единичной площади ( ), то максимально возможное количество информации (*H*), содержащейся в ней

$$. \qquad (4.9)$$

может служить численной оценкой качества изображения.

Таким образом, величина *H*, характеризующая качество изображения по критерию *QI*, определяется двумя параметрами: площадью минимально различимого элемента разложения (пиксела) и максимальным значением отношения сигнал/шум .

Нахождение этих параметров на практике затруднено, так как они, во-первых, коррелированы, а, во-вторых, зависят от условий формирования изображения. Действительно, при уменьшении площади элемента разложения ИВС отношение сигнал/шум возрастает, а уменьшение освещенности ведет и к падению отношения сигнал/шум, и к увеличению размеров минимального различимого элемента разложения площадью .

Определить функциональную связь между площадью элемента разложения ( ) и отношением сигнал/шум , а также их зависимость от условий освещенности возможно, исходя из квантовой теории света, согласно которой оптический луч представляет собой поток фотонов. Таким образом, любой ИВС можно рассматривать как некую поверхность, каждый элемент разложения которой излучает (или отражает) фотоны с различной интенсивностью ( ), т.е.

,

где - декартовы координаты элемента разложения .

Несамосветящиеся ИВС могут генерировать символы визуальных сообщений только в том случае, если они освещены и отражают падающий на них некий равномерный поток фотонов.

Интенсивность ( ) определяется как количество фотонов ( ), излученных элементом разложения площадью за время :

.

Таким образом, любой источник оптических сообщений (F) при заданном можно представить в виде совокупности элементов разложения , каждый из которых излучает определенное число ( ) квантов света, то есть

$$F = \sum_{i,j} \Phi_{ij} = \sum_{i,j} n_{ij} = n\sum_{i,j} (x_i, y_i),$$ при

Представляет интерес определение минимального числа фотонов в этом потоке, которое обеспечит получение ИВС и соответствующего изображения с заданным качеством.

Очевидно, что минимальное количество равномерно распределенных падающих фотонов ( ), необходимое для построения любого контурного (содержащего только два уровня яркости) ИВС, состоящего из N элементов разложения, будет

.

Но если требуется получить полутоновый ИВС и соответствующее ему полутоновое изображение с некоторой заданной минимальной различимой разностью яркостей (минимальным контрастом) элементов разложения (c), или с заданным числом градаций и отношением сигнал/шум , то минимальное количество фотонов ( ), необходимое для построения такого изображения, должно быть равным:

,

где

, при i=1,2,….m; j=1,2,….k.

Излучение фотонов, а следовательно, и их падение на некоторую площадку подчиняется квантовым законам и носит случайный характер, подчиняющийся, как известно, закону распределения Пуассона, из которого следует, что вероятность количества фотонов , испускаемых каким-либо светящимся телом за время T, описывается выражением:

,

где - среднее значение количества испускаемых фотонов.

Поэтому очевидно, что в действительности число фотонов, падающих на каждый элемент разложения за время T, распределено вокруг среднего значения так, что среднеквадратичное отклонение (в соответствии с распределением Пуассона) равно , то есть среднеквадратическое значение шума будет равно . Если при этом отношение сигнал/шум ( ) будет больше c, то получить изображение с заданным минимальным контрастом (c) невозможно. Поэтому необходимо повышать минимальное число падающих фотонов, чтобы уровень шума не превышал заданный минимальный контраст, то есть минимальное число фотонов, падающих на один элемент разложения ИВС, должно быть увеличено до величины n_c, такой, чтобы выполнить условие:

и, следовательно,

.

В этом случае общее число фотонов, приходящееся на весь ИВС, возрастет до :

.

n_3 – это минимальное количество фотонов, падающих на один элемент разложения ИВС, которое обеспечит получение изображения, состоящего из N элементов разложения, с заданным отношением сигнал / шум (контраст) в соответствии с распределением Пуассона.

Для подсчета минимального количества фотонов, падающих на ИВС и обеспечивающих получение изображения с заданным минимальным контрастом необходимо учитывать еще один фактор.

При обработке изображений, соответствующих символам визуальных сообщений, как правило, считают, что их минимально различимый контраст соответствует среднеквадратическому значению шумовых флуктуаций. Однако это не значит, что мгновенное значение шума в этом случае не будет превышать принятого уровня минимально различимого контраста, так как мгновенное значение шума может значительно превышать его среднеквадратичное значение.

Поэтому при обработке изображений, состоящих из большого числа элементов разложения, необходимо, чтобы минимально различимый контраст соответствующих символов визуального сообщения существенно превышал среднеквадратичный уровень шума, вызванного флуктуацией потока падающих фотонов.

Величину превышения легко оценить, аппроксимируя статистическое распределение числа падающих фотонов нормированным нормальным законом распределения, кривая распределения которого ( ) представлена на рис.4.7, а ее аналитическое выражение может быть приведено к виду

, (4.10)

где σ - среднеквадратическое отклонение от среднего значения ( );

k =1,2,...

Рис. 4.7. Кривая распределения числа падающих фотонов аппрокси мированная нормированным нормальным законом распределения.

Вероятность того, что значение числа падающих световых фотонов на отдельный пиксел ( ) будет отличаться от среднего значения ( ) на величину более чем $k\sigma$ определяется выражением:

.

В Табл. 4.1 приведены численные значения этой вероятности при различных *k*.

Табл. 4.1

k	1	2	3	4	5	6	7
	0,32	0,046	$0{,}27\cdot 10^{-3}$	$6{,}3\cdot 10^{-5}$	$5{,}7\cdot 10^{-7}$	$2{,}0\cdot 10^{-9}$	$2{,}5\cdot 10^{-12}$

Таблица 4.1 позволяет определить необходимую кратность превышения минимально различимым контрастом среднеквадратичного значения шума флуктуации числа падающих фотонов при заданной вероятности ложного его измерения. Например, для получения изображения с заданным минимальным контрастом и вероятностью ложного его измерения не более 1% необходимо, чтобы значение заданного минимального контраста не менее чем в 3 раза превышало среднеквадратическое значение флуктуации числа падающих фотонов.

Следовательно, для уменьшения до допустимого значения вероятности ложного измерения информационного параметра элементов разложения ИВС необходимо еще больше увеличивать среднее число падающих на них фотонов, тем самым увеличивая отношение сигнал/шум.

В случае распределения Пуассона при среднем значении числа падающих фотонов среднеквадратичное значение шума равно , а отношение сигнал/шум -

.

Поэтому для увеличения отношения сигнал/шум ( ) в *k* раз среднее число падающих фотонов необходимо увеличить в раз, таким образом, минимальное число фотонов ( ), которое необходимо для получения изображения, состоящего из *N* элементов разложения, с заданным контрастом *c* и при заданной вероятности ложного измерения, можно найти из выражения:

, (4.11)

а соответственно число фотонов, приходящихся на один элемент разложения ( ) будет определяться равенством:

. (4.12)

Подтверждением правильности полученного теоретического соотношения (4.11) могут служить экспериментальные исследования, которые заключались в фоторегистрации изображения специального тест-объекта при различных условиях съемки.

Тест-объект представляет собой белое поле, на котором размещались черные круги, причем вдоль строк круги имели по сравнению с фоном одинаковый контраст, но различные диаметры ( ): диаметр каждого последующего в два раза меньше предыдущего, а по столбцам были размещены круги одного диаметра, но контраст ( ) каждого последующего был в два раза меньше предыдущего. В этом случае произведение остается постоянным вдоль линий, параллельных диагонали тест-объекта.

На рис.4.8 представлены зарегистрированные изображения тест-объекта, полученные при постоянной освещенности, но различных временах экспозиции, причем каждое последующее изображение получено при экспозиции в 4 раза большей, чем предыдущее.

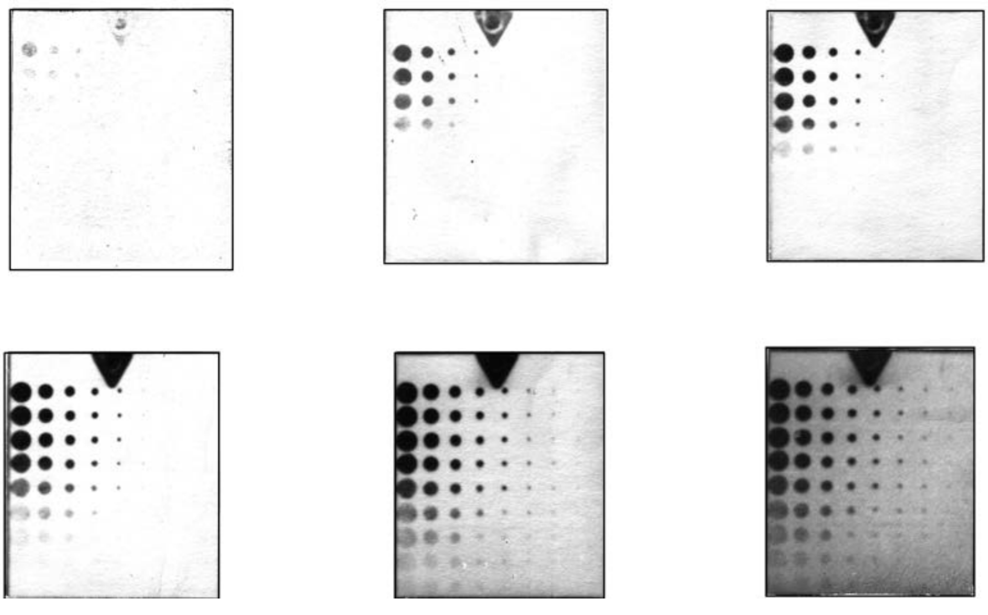


Рис. 4.8. Изображения тест-объекта, полученные при постоянной освещенности, но различных временах экспозиции.

Из соотношения (4.11) следует, что при постоянных условиях съемки

.

А так как число падающих на тест-объект фотонов ( ) соответствует величине экспозиции (*B*), то можно считать, что произведение обратно пропорционально величине экспозиции:

.

Следовательно, для того, чтобы граница различимых дисков на зарегистрированном изображении тест-объекта сдвинулась на один диск вправо (;) или чтобы граница различимых дисков сдвинулась на один диск вниз (;) необходимо увеличить экспозицию в 4 раза.

Анализируя полученные фотоизображения тест-объекта, можно сделать вывод, что разрешение элементов изображения при заданной экспозиции зависит как от их размера ( ), так и от контраста ( ), и хорошо согласуется с полученным теоретическим соотношением (4.11).

Действительно, при увеличении экспозиции в 4 раза граница различимых дисков смещается на один шаг вправо, в сторону дисков с диаметром в 2 раза меньше, и на один шаг вниз, в сторону дисков с контрастом в 2 раза меньшим, как и следует из соотношения (4.11).

Для практического использования полученных соотношений (4.11, 4.12) целесообразно привести их к общепринятому виду, то есть к виду с использованием энергетических и светотехнических характеристик. Известно, что падение за время на некоторую площадку площадью числа фотонов n с частотой ν соответствует притоку энергии Q:

, (4.13)

где - энергия одного фотона;

- постоянная Планка.

В случае немонохроматического излучения в диапазоне частот от до

, (4.14)

где - распределение числа фотонов по частоте ν.

В этом случае энергетический и светотехнический поток излучения ($\Phi_э$ и $\Phi_с$ соответственно) определяют по формулам

$$\Phi_э = \frac{Q}{\Delta t}[\text{Вт}]; \qquad \Phi_с \cong 683 \cdot \Phi_э [\text{лм}], \tag{4.15}$$

а освещенность энергетическая и светотехническая (и   ) по формулам

$$E_э = \frac{\Phi_э}{S_0} = \frac{Q}{\Delta t \cdot S_0}\left[\text{Вт} \cdot \text{м}^{-2}\right]; \qquad E_с = \frac{\Phi_с}{S_0}[\text{лк}]. \tag{4.16}$$

Освещенность ИВС легко может быть измерена и тем самым на основании (4.13,4.15 и 4.16) определено соответствующее число фотонов, падающих на ИВС. Действительно, для монохроматического излучения

$$n = \frac{E_э \cdot \Delta t \cdot S_0}{h \cdot \nu} \tag{4.17}$$

и тогда соотношения (4.11 и 4.12) могут быть представлены в следующем виде:

$$\frac{E_э \cdot \Delta t}{h \cdot \nu} = \frac{k^2}{\overline{S} \cdot c^2}. \tag{4.18}$$

Полученное соотношение (4.18) позволяют получить жесткую функциональную зависимость между легко измеряемыми или конструктивно задаваемыми (как в устройствах считывания) размером элемента разложения (   ), параметрами получения символа визуального сообщения ($E_э, \Delta t, \nu$) и предельно возможным при заданной допустимой погрешности (k) минимальным контрастом (c), практическое измерение которого трудно выполнимо.

Выразив из (4.18) c - предельно возможный минимальный контраст и подставив его в (4.9), с учетом того, что   и   , получим выражение для максимального предельного количества информации (H), которое может содержаться в символе ИВС единичной площади

$$H = \frac{1}{\overline{S}} \cdot \log_2\left[\sqrt{\frac{E_э \cdot \Delta t \cdot \overline{S}}{k^2 \cdot h \cdot \nu}} + 1\right]\left(\text{бит}/\text{м}^2\right) \tag{4.19}$$

или с учетом того, что

$$H \cong \frac{1}{2\overline{S}} \cdot \log_2\left(\frac{E_э \cdot \Delta t \cdot \overline{S}}{k^2 \cdot h \cdot \nu}\right)\left(\text{бит}/\text{м}^2\right). \tag{4.20}$$

А максимальное предельное среднее количество информации, приходящееся на один элемент разложения, определяется соотношением

$$h = \frac{H}{N} = \frac{1}{2}\log_2\left(\frac{E_э \cdot \Delta t \cdot \overline{S}}{k^2 \cdot h \cdot \nu}\right)\left(\text{бит}/\text{элемент}\right). \tag{4.21}$$

Именно выражения 4.20 и 4.21 могут служить численной мерой качества изображения, потому что предельная энтропия (H) и предельное среднее количество информации, приходящейся на один элемент разложения (h) изображения характеризуют изображения с основополагающей информационной точки зрения, которая во многом определяет их потребительские свойства. Практическое использование соотношений (4.20 и 4.21) не вызывает трудностей, так как все входящие в них величины могут быть либо легко измерены, либо заданы конструктивно. В качестве примера определим предлагаемым способом качество изображения, которое может быть получено в процессе считывания фотоизображения размером при освещении его равномерным световым потоком , генерируемым гелий-неоновым лазером с длиной волны ( ). Считывание производится квадратной аппертурой с линейным размером d (мкм) ( ) с быстродействием 10^5 элементов/сек ( ) и вероятность ложного измерения различимых уровней контраста не превышающей 0,01% (k=4). В этом случае исходя из выше приведенных соотношений легко найти:

- число различимых уровней контраста (m):

$$m(d)=\sqrt{\frac{\Phi\cdot\Delta t\cdot d^2}{S_0\cdot k^2\cdot h\cdot\nu}}+1=7{,}0\ \cdot 1\ {}^{6}4\ d\theta 1;$$

- среднее количество информации, приходящейся на один элемент разложения $h(d)$:

;

- предельное количество информации, содержащейся на всем изображении $I(d)$:

;

- предельная энтропия изображения $H(d)$ при :

.

В Табл. 4.2 приведены значения , , и при различных значениях d.

Табл. 4.2

$d(10^{-6}$м)	1	5	10	15	25
$\Delta S(10^{-12}$м)	1	25	100	215	625
m	8	36	71	106	177
h(бит/эл.разл.)	3,0	5,2	6,2	6,7	7,5
I(бит)	$1{,}2\cdot10^{11}$	$8{,}3\cdot10^{9}$	$2{,}5\cdot10^{9}$	$1{,}2\cdot10^{9}$	$4{,}8\cdot10^{8}$
H(бит/м^2)	$3{,}0\cdot10^{12}$	$2{,}1\cdot10^{11}$	$6{,}2\cdot10^{10}$	$3{,}0\cdot10^{10}$	$1{,}2\cdot10^{10}$

H(бит/мм^2)	$3{,}0 \cdot 10^6$	$2{,}1 \cdot 10^5$	$6{,}2 \cdot 10^4$	$3{,}0 \cdot 10^4$	$1{,}2 \cdot 10^4$

Важно так же отметить, что в случае регистрации изображений с помощью реальных регистрирующих устройств в регистрирующее устройство попадает лишь часть (   ) фотонов, отраженных символом визуального сообщения, но и эти попадающие в регистрирующее устройство фотоны регистрируются не все, так как квантовый выход регистрирующих устройств (   ), как правило, меньше 1.

Поэтому в этом случае для регистрации изображения символа визуального сообщения с заданным качеством необходимо корректировать полученные соотношения, еще больше увеличивая общее число фотонов, падающих на символ визуального сообщения (   ):

. (4.22)

Все выкладки, приведенные выше, касались прежде всего излучения в видимой области спектра (0,4 - 0,7 мкм), однако, так как исходными предпосылками при выводе соотношений является исключительно квантовая теория, то очевидно, что все вышесказанное справедливо и к системам, работающим во всем диапазоне электромагнитного излучения.

Следует также учитывать, что полученные соотношения справедливы только при условии, что линейные размеры элемента разложения значительно превышают длину волны светового излучения, падающего на ИВС, так как они не учитывают волновые эффекты света.

Такой способ определения качества изображения является более корректным по сравнению с использованием для этих целей полос заданного контраста, потому что именно с помощью элементов разложения (а не полос) можно построить любое изображение, и современные автоматические методы обработки изображений предполагают разбиение изображения на совокупность элементов, форма которых близка к кругу или квадрату.

Предложенный подход к определению качества изображения хорошо согласуется с методологией современных физических представлений, которые запрещают обсуждать явления сами по себе, независимо от способа их наблюдения. При этом различные явления относятся к объективной реальности (не зависящей от сознания субъекта), а их наблюдение позволяет получить физическую реальность, то есть ту часть объективной реальности, которая может быть познана опытным путем и сознанием субъекта, причем величина этой части зависит от способа наблюдения.

Поэтому очевидно, что качество изображения (его информативность) как физическая реальность, отражающая некую часть объективной реальности, будет определяться условиями наблюдения - степенью освещенности изучаемого объекта.

4.3. Информационная оценка качества фотоизображений.

Как отмечалось ранее, изображение представляет собой символ дифференциального плоского или объемного (одномерного или двумерного) ИВС, получаемый в результате отображения на фиксированный момент времени и на заданной ограниченной плоскости и зарегистрированный на некотором носителе. В дальнейшем при его обработке зарегистрированное изображение уже само может играть роль источника сообщения при информационном моделировании процессов его обработки.

Очевидно, что процессы отображения и регистрации изображения с точки зрения их информационных характеристик могут характеризоваться потерями информации, которые будут зависеть как от метода регистрации, так и от свойств носителя изображения. В настоящее время наиболее универсальным, быстродействующим и общепринятым методом регистрации оптических изображений является фоторегистрация, при котором изображение регистрируется (фиксируется) на светочувствительном фотоносителе в виде фотоизображения.

Фотоизображения очень часто используются в качестве источников визуальных сообщений в геодезии, картографии, исследовании природных ресурсов, экологии и других областях человеческой деятельности. Во многих из них информационную емкость фотоизображений определяют через их изобразительные и измерительные возможности применительно к решению какой-либо конкретной задачи. В этом случае, как правило, применяют семантический (субъективный) подход, так как оценивают количество информации с точки зрения возможности решения конкретной задачи (например, измерение или дешифрирование) при конкретной технологии использования фотоизображения. Кроме того, изобразительные и измерительные свойства фотоизображения определяются методами его получения, поэтому судить об информационных характеристиках самого фотоизображения не корректно.

С точки зрения теории информации более объективным является рассмотрение фотоизображения как источника визуальных сообщений, входящего в некоторую информационную модель, и оценка его предельных (потенциальных) возможностей с целью его информационного согласования с остальными блоками информационной модели.

Для примера рассмотрим черно-белое фотоизображение, которое, как известно, представляет собой дифференциальный одномерный плоский источник визуальных сообщений. В дальнейшем полученные выводы можно будет применить и к цветным фотоизображениям, которые являются дифференциальными двумерными плоскими источниками визуальных сообщений (см. рис. 4.5).

Согласно избранной модели (§4.2) источник визуальных сообщений можно разбить на элементарные площадки, причем, если размеры всех площадок одинаковы, то единица площади фотоизображения содержит *N* элементарных площадок:

,

где - площадь элементарной площадки.

Положив, что возможные плотности почернения каждой площадки независимы, равновероятны и могут принимать одно из *m* значений, можно найти максимальное количество информации, содержащейся в единице площади фотоизображения (*I*),

. (4.23)

Применять это выражение для подсчета информационной емкости конкретного фотоизображения не корректно, так как конкретное фотоизображение имеет определенный закон распределения возможных плотностей элементарных площадок. Однако, при оценке максимально возможной информационной емкости, которая достигается при равновероятном распределении плотностей площадок, вполне допустимо. Затруднения при использовании этого выражения возникают при выборе *m* - числа возможных значений оптической плотности элементарных площадок, так как известно, что из-за своей природы фотоизображение немыслимо без наличия шумов, вызванных зернистостью фотослоя.

Наличие шумов ограничивает число различимых градаций плотности и тогда максимальная информационная емкость фотоизображения ($I\phi$), при условии равновероятности градаций плотности и их взаимной независимости, может быть определена по формуле Хартли:

$$I\phi = N \cdot \log_2\left(\frac{\Delta D}{\sigma_D}+1\right), \qquad (4.24)$$

где - диапазон возможного изменения оптической плотности;

- максимально возможная плотность фотоносителя;

- плотность вуали;

- среднеквадратическое значение шумов.

Формула (4.24) справедлива, если допустить, что сигнал и шум независимы и стационарны. Однако для фотоизображений эти допущения некорректны, так как известны эмпирическая зависимость, устанавливающая связь между средней оптической плотностью

(   ) участка фотоизображения площадью (   ) и среднеквадратическим значением флуктуации оптической плотности (   ):

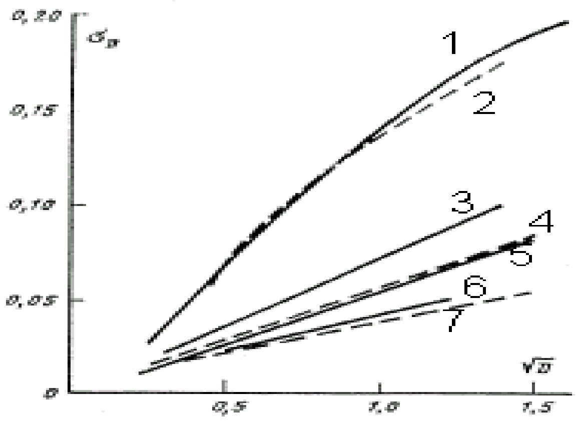
$$\sigma_D = \sqrt{\frac{0,4\quad 3\bar{D}4\cdot S_з}{\Delta S}}, \tag{4.25}$$

где $S_з$ - средняя площадь проекции одного зерна используемого фотоматериала,
и выражение для определения соответствующего отношения сигнал/шум (ψ):

(4.26)

где D_0 – плотность вуали.

Графики зависимости и от $\sqrt{D}$ при $\Delta S = 1\ 0м0\ к^2м$ для различных фотоматериалов (1 – астропластинка; 2- панхром-10; 3 – панхром средней чувствительности; 4 – диапозитив; 5 – микрат-200; 6 – панхром мелкозернистая; 7 – кинопозитив МЗ) представлены на рис.4.9 и рис. 4.10 соответственно.



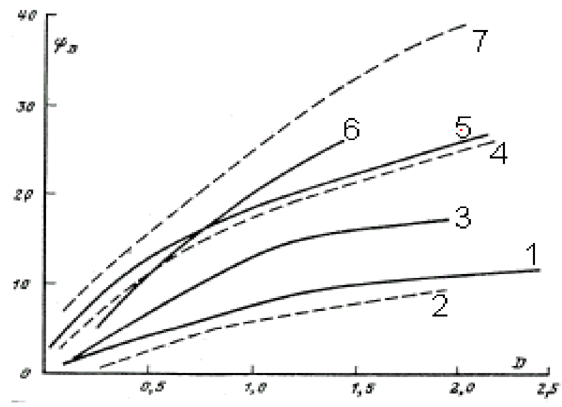


Рис.4.9 Графики зависимости от $\sqrt{D}$. Рис. 4.10 Графики зависимости от $\sqrt{D}$.

Анализ формул (4.24) и (4.25) показывает, что число различимых градаций будет функцией как средней оптической плотности (   ), так и площади элементарной площадки (   ), при изменения оптической плотности от до .

Найти в этом случае предельную информационную емкость фотоизображения можно, разбив весь диапазон изменения плотности на *L* равноотстоящих уровней (и в дальнейшем устремив *L* к ∞), при этом разность между соседними уровнями равна

.

Средняя плотность элементарных площадок (   ), принадлежащих *k*-ому уровню, можно определить из выражения:

$$\overline{D}_k = D_0 + k\cdot\delta D, \qquad (4.27)$$

где $k = 1,2,\ldots,L$.

Число элементарных площадок (   ) со средней плотностью   ,приходящихся на участок фотоизображения единичной площади, учитывая равновероятное распределение оптической плотности (для обеспечения максимального количества информации), может быть легко найдено:

$$. \qquad (4.28)$$

Из выражений (4.24), (4.25), (4.27) и (4.28) легко найти соотношение для нахождения максимального количества информации (I_k), которое может содержать совокупность точек числом   , принадлежащая *k*-ому уровню оптической плотности. Действительно,

$$I_k = \frac{1}{\Delta S\cdot L}\cdot\log_2\left(\frac{\overline{D}_k^2}{\sqrt{0,4\cdot\overline{D}_k^3\frac{S_з}{\Delta S}}}+1\right) = \frac{1}{\Delta S\cdot L}\cdot\log_2\left(\frac{(D_0+k\cdot\delta D)^2}{\sqrt{0,4\cdot\frac{S_з}{\Delta S}\cdot(D_0+k\cdot\delta D)^3}}+1\right). \qquad (4.29)$$

Общее количество информации, содержащееся в фотоизображении единичной площади ($I\Phi$), находится путем суммирования всех количеств информации (I_k), содержащихся в элементах разложения с плотностью   по *k*:

$$I\Phi = \sum_{k=1}^{L} I_k = \frac{1}{\Delta S\cdot L}\cdot\sum_{k=1}^{L}\log_2\left(\frac{(D_0+k\cdot\delta D)^2}{\sqrt{0,4\cdot\frac{S_з}{\Delta S}\cdot(D_0+k\cdot\delta D)^3}}+1\right)(б/еи д). \qquad (4.30)$$

В полученном соотношении в качестве параметра используется *L* – число градаций по уровню оптической плотности. Представляет интерес зависимость предельного количества информации ($I\Phi$) от избранного числа градаций оптической плотности (*L*). На рис.4.11 представлен график зависимости предельного количества информации, содержащегося на участке фотоизображения площадью 1 мм^2, от числа градаций *L* при $D_{max} = 2,7$, $D_0 = 0,1$, $S_з = 5\cdot10^{-6}$ мм^2 и $d = 10\cdot10^{-3}$ мм.

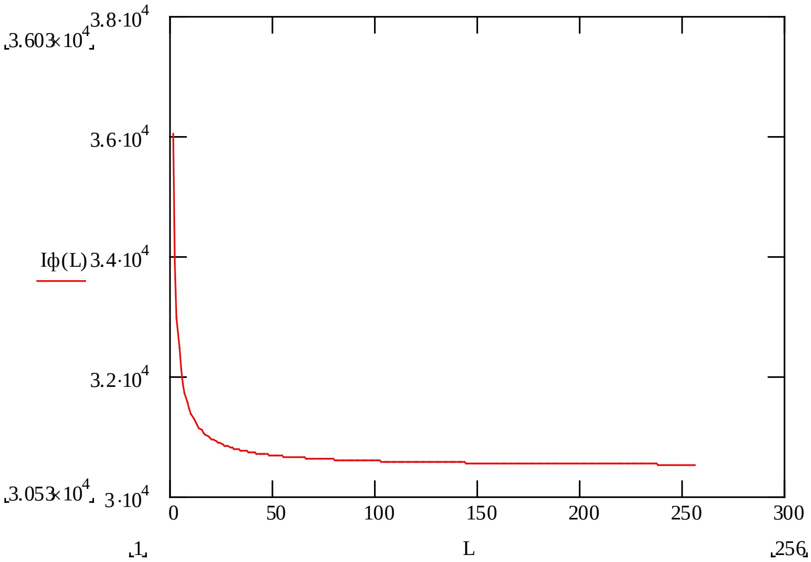


Рис. 4.11. График зависимости предельного количества информации, содержащегося на участке фотоизображения, от числа градаций *L*.

Таблица 4.3 содержит численные значения этой зависимости при некоторых заданных значениях *L*.

Табл.4.3

L	2	4	8	16	32	64	128	256
$I\phi(L)$(бит)	$3{,}39\cdot10^4$	$3{,}25\cdot10^4$	$3{,}16\cdot10^4$	$3{,}11\cdot10^4$	$3{,}08\cdot10^4$	$3{,}06\cdot10^4$	$3{,}06\cdot10^4$	$3{,}05\cdot10^4$

Из графика (Рис.4.11) видно, что в предельном случае, при $L \to \infty$, можно найти точное предельное количество информации, содержащееся на участке фотоизображения заданной площади ($I_{ф.п}$). Для оценки погрешности вычисления предельного количества информации на участке фотоизображения построен график (Рис.4.12) зависимости отношения $M = I\phi(L)/I_{ф.п}$ от числа градаций оптической плотности (*L*), при этом в качестве $I_{ф.п}$ бралось $I\phi(256)$, а в Таблице 4.4 представлены численные значения отношения *M* при некоторых значениях *L*.

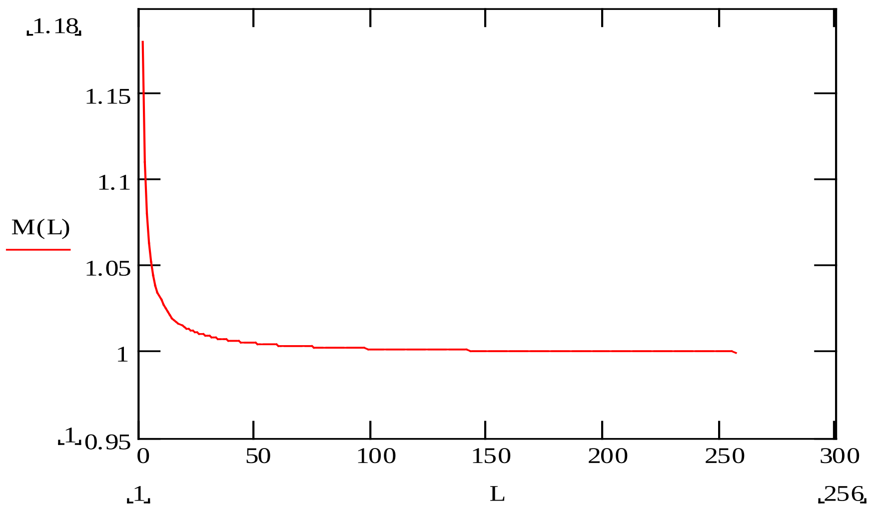


Рис. 4.12. График зависимости отношения $M = I\phi(L)/I_{\phi}$ от числа градаций оптической плотности (L),

Табл. 4.4

L	1	2	4	8	16	32	64	128	256
M	1,18	1,11	1,06	1,03	1,017	1,008	1,004	1,001	1,000

Из рис.4.12 и таблицы 4.4 видно, что для подсчета предельного количества информации, содержащейся на участке фотоизображения с погрешностью не превышающей 1%, число градаций оптической плотности должно быть не менее 32.

Пользуясь полученным выражением (4.30), можно рассчитать предельную информационную емкость некоего абстрактного фотоносителя как функцию площади элементарной площадки. Пусть фотоноситель имеет следующие параметры: ; ; .

Выберем единицу площади фотоизображения , а число уровней градаций оптической плотности $L = 32$.

На рис.4.13 представлен график зависимости предельной информационной емкости фотоизображения площадью 1мм^2 ($I\Phi$) в зависимости от линейного размера элементарной площадки (d).

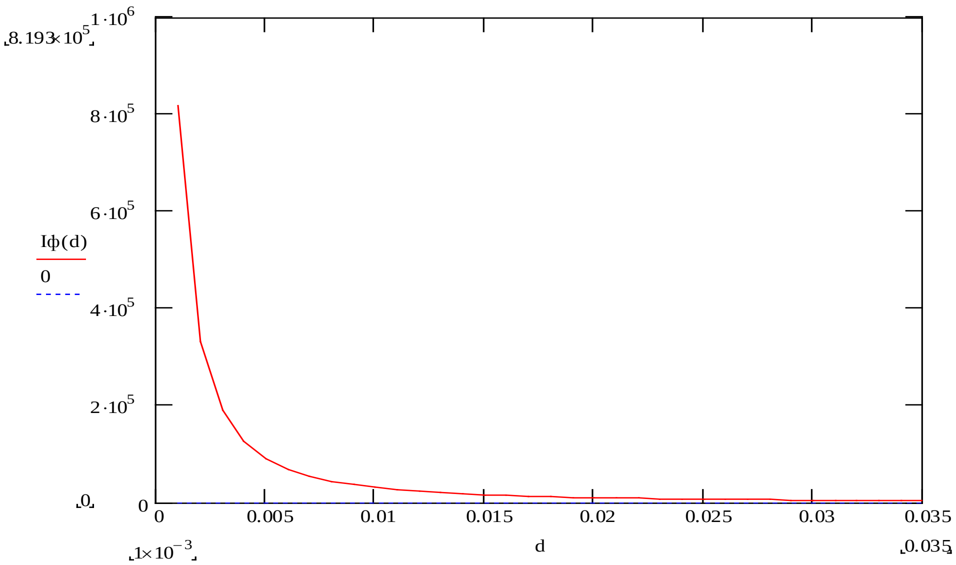


Рис. 4.13. График зависимости предельной информационной емкости фотоизображения площадью 1мм^2 ($I\Phi$) в зависимости от линейного размера элементарной площадки (d).

А на рис.4.14 представлен график зависимости предельного среднего количества информации, приходящегося на одну элементарную площадку от линейных размеров элементарной площадки (d).

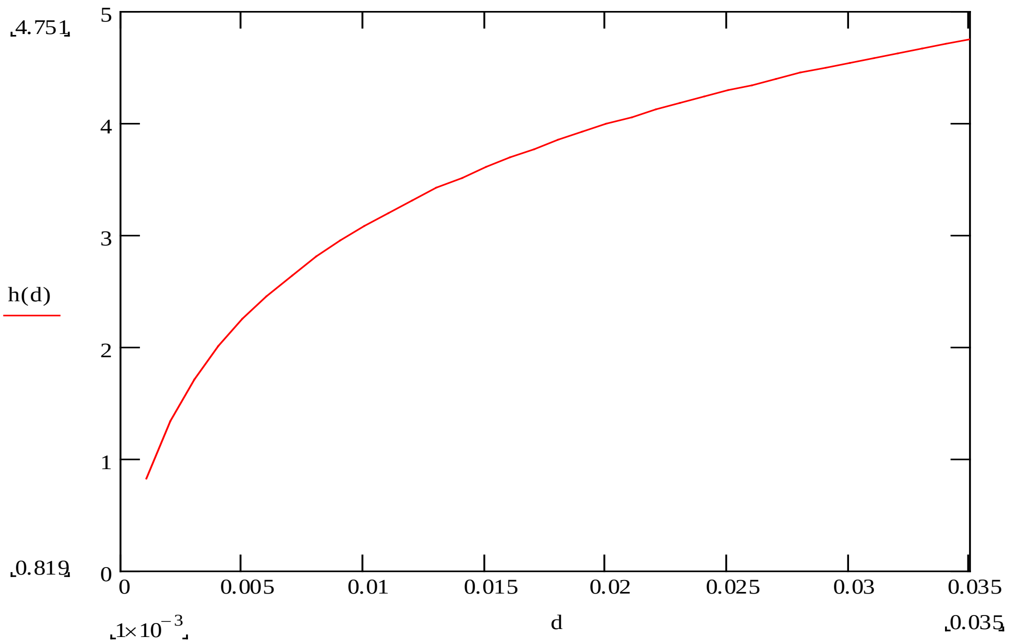


Рис.4.14. График зависимости предельного среднего количества информации, приходящегося на одну элементарную площадку от линейных размеров элементарной площадки (d).

В таблицу 4.5 сведены численные значения предельной информационной емкости участка изображения площадью 1мм^2 ($I\phi$), среднее количество информации, приходящееся на одну элементарную площадку (h) при различных линейных размерах (d) и площадях ( ) элементарной площадки.

Табл. 4.5

d (10^{-3}мм)	1	2	3	5	10	15	25
ΔS (10^{-6}мм^2)	1	4	9	25	100	225	625
$I\phi$ (бит/мм^2)	$8{,}19\cdot10^5$	$3{,}22\cdot10^5$	$1{,}87\cdot10^5$	$8{,}98\cdot10^4$	$3{,}08\cdot10^4$	$1{,}60\cdot10^4$	$6{,}81\cdot10^3$
h (бит/эл.пл.)	0,82	1,33	1,70	2,24	3,08	3,60	4,26

Из таблицы 4.5 видно, что при линейных размерах элементарной площадки (d) менее 2 мкм среднее количество информации, приходящейся на одну элементарную площадку становится менее 1 бит, то есть число различимых уровней плотности элементарной площадки становится менее двух, что не позволяет различать отдельные элементарные площадки.

Поэтому для рассмотренного фотоносителя использование элементарных площадок с линейными размерами (d) менее 2мкм нецелесообразно.

Сравнение данных Табл. 4.2 и Табл. 4.5 показывает, что информационная емкость оптического изображения единичной площади, приведенного в § 4.2, существенно выше информационной емкости фотоносителя такой же площади.

Представляет интерес определение зависимости относительных потерь информации, содержащейся на единичной площади оптического изображения, приведенного в Табл.4.2 при регистрации его на фотоносителе такой же площади (Табл.4.5), от линейных размеров элементарной площадки (d). В этом случае относительные потери информации $(\eta(d))$ вычислялись по формуле:

$$\eta(d) = \frac{H(d)}{I\phi(d)}. \tag{4.31}$$

Результаты вычислений представлены в Табл. 4.6.

Табл.4.6

d(10^{-3}мм)	1	5	10	15	25
$\eta(d)$	3,66	2,34	2,01	1.87	1,76

Из этой таблицы видно, что зернистость фотоносителя особенно сильно влияет на относительные потери информации при малых линейных размерах элементарных площадок.

Следует заметить, что рассчитанная по формуле 4.30 информационная емкость фотоносителя является предельно возможной. Реальные фотоизображения обладают значительно меньшей информационной емкостью, так как их разрешение ограничено фотографическими системами, условиями съемки и обработки фотоматериала, а также достаточно произвольным распределением значений оптической плотности.

В случае, если известен конкретный закон распределения числа элементарных площадок (   ), попадающих в каждый из L уровней плотности (гистограмма распределения) из выражения (4.30) может быть получено соотношение максимального количества информации (   ), содержащейся на единичной площади такого изображения

$$I' = \frac{1}{\Delta S}\sum_{k=1}^{L} P_k \cdot \log_2\left(\frac{(D_0 + k\cdot\delta D)}{\sqrt{0,4 \cdot \frac{S_з}{\Delta S}\cdot(D_0 + k\cdot\delta D)}} + 1\right) (б\ /еи\ д), \tag{4.32}$$

где P_k - вероятность попадания элементарных площадок в k-тый уровень оптической плотности.

Следует подчеркнуть, что полученные соотношения (4.31) и (4.32) характеризуют фотоноситель только как носитель информации, в случае формирования оптического изображения, соответствующего заданному фотоизображению, необходимо учитывать и квантовые свойства источника света (§4.3).

4.4. Информационная оценка датчиков сообщений

Согласно принятой общей информационной модели (Рис.4.1) первичное восприятие сообщений осуществляется датчиками сообщений. Датчик сообщений преобразует первичную естественную форму визуального сообщения в форму, удобную для последующей обработки, чаще всего в электрическую. Датчики принято характеризовать следующими основными параметрами:

1. Коэффициент преобразования $K_п$, который определяется соотношением

$$K_п = \frac{\Delta y}{\Delta x}, \tag{4.33}$$

где - приращение первичного сообщения;

- приращение выходной величины датчика.

Этот параметр характеризует лишь безынерционные и линейные датчики сообщений.

Датчик называют безынерционным, если $K_п$ не зависит от длительности приращения , и линейным, если $K_п$ не зависит от величины . Линейность датчика характеризует степень приближения функции к линейной зависимости.

Степень нелинейности датчика сообщений характеризуется коэффициентом нелинейных искажений.

2. Передаточная функция - отношение преобразования Лапласа от выходной величины к преобразованию Лапласа от первичного сообщения

, (4.34)

где

, .

3. Чувствительность - это минимальное значение сообщения, которое может быть преобразовано датчиком

$$X_{min} = \frac{\sigma_ш}{K_п}, \tag{4.35}$$

где $\sigma_{ш}$ - эффективная величина шумов датчика.

Зная основные технические параметры датчика можно найти основные информационные характеристики, присущие процессу преобразования исходного сообщения *x* в выходной сигнал датчика *y*, входящего в информационную модель (Рис.4.1).

Действительно, при неоднократном преобразовании символов исходного сообщения можно найти и - минимальное и максимальное значение входного сигнала датчика, соответствующего исходному сообщению. По этим значениям можно определить диапазон преобразуемого входного сигнала :

. (4.36)

Из (4.35) следует, что для каждого датчика существует конкретное значение (чувствительность датчика), определяемое уровнем шумов датчика. Чувствительность датчика определяет разрешающую способность датчика , то есть ту минимальную разность между уровнями входного сигнала, при которой можно достоверно распознать отдельные символы исходного сообщения.

Так как диапазон входного сообщения (D_x) и разрешающая способность ( ) конечны, то возможно различить лишь конечное число уровней (*N)*

$$N = \frac{D_x}{\Delta x} + 1 .$$

Таким образом, объем алфавита входного сообщения не может быть больше *N*.

В случае отсутствия данных о законе распределения символов источника сообщения можно считать, что их появление равновероятно, и энтропия входного сообщения датчика *H* может быть найдена по формуле Хартли

$$H = \log_2\left(\frac{D_x}{\Delta x} + 1\right)(\text{бит/сим}). \tag{4.37}$$

Если же появление символов неравновероятно, но известен закон распределения символов входного сообщения, то энтропия (*H*) определяется по формуле Шеннона

, (4.38)

где - вероятность попадания в *k* уровень;

k - номер уровня.

При известной частоте измерений датчика - *F* (сек$^{-1}$) легко подсчитать входную производительность датчика

$$P_д = H \cdot F(\text{б о}). \tag{4.39}$$

Для информационного согласования источника сообщений с датчиком сообщений необходимо, чтобы производительность источника сообщений $(P_и)$ не превышала входную производительность датчика сообщений $(P_д)$, в противном случае неминуемы потери информации.

Воздействие шумов на датчик выражается не только в ограничении энтропии воспринимаемого сообщения, но и в потерях информации в процессе преобразования. Известно, что если на датчик воздействует некий аддитивный шум (*n*), то энтропия полезного сигнала , которую можно извлечь из смеси , равна разности энтропий принятого сообщения и шума , то есть

. (4.40)

Заметим, что влияние шума (*n*) нарушает однозначную зависимость между входным сигналом *x* и выходным сигналом *y* и делает ее случайной (статистической).

В случае аддитивного шума закон распределения выходного сигнала *y* при фиксированном входном сигнале *x* равен закону распределения шума, то есть

,

где - плотность распределения шума;

- условная плотность распределения выходного сигнала *y* при известном входном сигнале *x*.

Тогда выражение (4.40) может быть записано в следующем виде:

. (4.41)

Подставив в это выражение соответствующие значения и , рассчитанные по формуле Шеннона, получим:

, (4.42)

где - совместная плотность вероятности сигналов *x* и *y*;

- плотность вероятности выходного сигнала *y*;

- плотность вероятности входного сигнала *x*;

и - соответствующие условные плотности вероятности при фиксированном входном сигнале и фиксированном выходном сигнале соответственно.

Для случая непрерывных сигналов *x* и *y* выражение (4.42) после предельного перехода приобретет вид

, (4.43)

где - совместная плотность вероятности сигналов *x* и *y*;

- условная плотность вероятности сигнала *y* при фиксированном *x*;

, - плотность вероятности сигналов *x* и *y* соответственно.

Таким образом, датчик сообщения как элемент информационной системы требует согласования по информационным параметрам с источником сообщений с целью уменьшения (а в идеале устранения) потерь информации в процессе преобразования сообщений источника в иную форму, удобную для дальнейшего использования.

Для практического использования рассмотренные выше соотношения (4.40-4.43) не всегда удобны, так законы распределения символов входного сообщения (*x*) и шума (*n*) чаще всего неизвестны. Поэтому на практике целесообразно информационные свойства датчиков сообщений выражать через максимально возможную для данного датчика энтропию входного сообщения - $H_{\max}$, считая, что символы входного сообщения равновероятны и некоррелированы

$$H_{\max} = \log_2\left(\frac{D_x}{\Delta x}+1\right)_{\max} = \log_2\left(\left(\frac{D_x}{\Delta x}\right)_{\max}+1\right), \quad (4.44)$$

а входную производительность датчика характеризовать ее максимальным предельным значением $P_{д.\max}$:

$$P_{д.\max} = H_{\max} \cdot F_{\max} \quad (бит/с), \quad (4.45)$$

где $F_{\max}$ - максимальная частота измерения датчика.

В сканирующих устройствах в качестве фотоэлектрического преобразователя (фотодатчика) используют полупроводниковые фотоприемники или фотоэлектронные умножители (ФЭУ), для которых основной причиной ограничения энтропии входного сообщения ($H_{\max}$) и входной производительности ($P_{д.\max}$) являются шумы фотометрирования, вызванные квантовой природой света и физической природой фотоприемников, причем энтропия входного сообщения ($H_{\max}$) и частота измерения фотометрических параметров отдельных элементов изображения ($F_{\max}$) взаимозависимы.

Применительно к фотоприемникам диапазон преобразуемого входного сообщения $D_x = \Phi_{\max} - \Phi_{\min}$, где $\Phi_{\max}$ и $\Phi_{\min}$ - максимальное и минимальное возможные значения светового потока, падающего на фотоприемник, а в качестве разрешающей способности Δx целесообразно использовать среднеквадратическое эффективное значение шума фотоприемника, приведенное к его входу - $\delta_{ф.п}$. Тогда выражение (4.44) преобразуется к виду:

$$H_{ф.max} = \log_2\left(\frac{\Phi_{max} - \Phi_{min}}{\delta_{ф.п}} + 1\right) = \log_2(\psi_ф + 1), \qquad (4.46)$$

где $H_{ф.max}$ - максимальная входная энтропия фотоприемника,

$\psi = \dfrac{\Phi_{max} - \Phi_{min}}{\delta_{ф.п}}$ - отношение сигнал/шум фотоприемника.

В фотоприемных устройствах основным видом шума является дробовой шум, представляющий собой флуктуации во времени электрического сигнала, вызванные тем, что электрический ток является потоком дискретных частиц, число которых подчиняется квантовым законам. Среднеквадратическое (эффективное) значение дробового шума (δ_i) на выходе фотоприемника при заданном значении светового потока, падающего на фотоприемник (Φ) пропорционально полосе частот преобразуемых сигналов (Δf) или обратно пропорционально времени измерения светового потока (τ). Известно выражение для нахождения среднеквадратического значения дробового шума на выходе фотоприемника (ФЭУ):

$$\delta_i = \sqrt{2e \cdot i \cdot \Delta f \cdot M(1+B)} = \sqrt{\frac{e \cdot i \cdot M(1+B)}{\tau}} = \sqrt{\frac{e \cdot \Phi \cdot \Sigma_a \cdot M(1+B)}{\tau}}, \qquad (4.47)$$

где $e = 1{,}6 \cdot 10^{-19} A \cdot c$ - заряд электрона;

i - ток ФЭУ [А];

$M = \dfrac{\Sigma_a}{\Sigma_к}$ - коэффициент усиления ФЭУ;

Σ_a - анодная чувствительность ФЭУ [А/Вт];

$\Sigma_к$ - катодная чувствительность ФЭУ [А/Вт];

$(1+B)$ - коэффициент, зависящий от конструкции ФЭУ, в среднем равный 2,5.

Аналогичные соотношения, связывающие среднеквадратическое значение дробового шума с техническими характеристиками фотоприемника и временем измерения характеристик элемента изображения известны и для полупроводниковых фотоприемников.

Приведя среднеквадратическое (эффективное) значение шума ко входу фотоприемника

$$\delta_{ф.п} = \frac{\delta_i}{\Sigma_a},$$

и задаваясь отношением сигнал/шум на входе фотоприемника

$$\psi_{ф} = \frac{\Phi_{max} - \Phi_{min}}{\delta_{ф.п}},$$

с учетом (4.47) можно получить выражение для нахождения отношения сигнал/шум фотоприемника при заданной длительности измерения светового потока (τ):

$$\psi_{ф} = \frac{\Phi_{max} - \Phi_{min}}{\delta_{ф.ш.п}} = \frac{\left(\Phi_{max} - \Phi_{min}\right)\cdot \sum_{a}}{\sqrt{\dfrac{e \cdot i \cdot M \cdot (1+B)}{\tau}}}. \tag{4.48}$$

Следовательно, из (4.46) вытекает соотношение для определения максимальной входной энтропии фотоприемника:

$$H_{ф.max} = \log_2(\psi_{ф}+1) = \log_2\left(\frac{\left(\Phi_{max} - \Phi_{min}\right)\cdot \sum_{a}}{\sqrt{\dfrac{e \cdot i \cdot M \cdot (1+B)}{\tau}}} + 1\right) (бит). \tag{4.49}$$

Исходя из общего вида информационной модели (Рис. 4.1), условием информационной согласованности источника сообщения и датчика сообщения, при выполнении которого обеспечивается преобразование исходного сообщения без потерь информации, будет:

$$H_{и.max} \le H_{ф.max}, \tag{4.50}$$

где $H_{и.max}$ - максимальное количество информации, которое может содержаться в символе источника сообщений (элементарной площадке).

Следовательно, должно выполняться условие:

$$\log_2(\psi_{ф}+1) \ge H_{и.max}. \tag{4.51}$$

В случае использования в качестве источника сообщения оптических или фотоизображений их $H_{и.max}$ может быть найдено на основе полученных соотношений (§4.2, §4.3), и из (4.51) легко определить требуемое отношение сигнал/шум фотоприемника ($\psi_{ф}$):

$$\psi_{ф} \ge 2^{H_{и.max}} - 1 = m - 1, \tag{4.51}$$

где m – объем алфавита источника сообщения,

а зная требуемое отношение сигнал/шум и значения технических параметров используемого фотоприемника из 4.48 легко получить соотношение для определения времени измерения характеристик элемента изображения (τ) при заданном отношении сигнал/шум:

$$\tau \geq \frac{\psi_{\text{ф}}^2 \cdot e \cdot i \cdot M(1+B)}{(\Phi_{max} - \Phi_{min})^2 \sum_{i\,min}^2} = \frac{(m-1)^2 \cdot e \cdot i \cdot M(1+B)}{(\Phi_{max} - \Phi_{min})^2 \sum_{i\,min}^2}. \quad (4.52)$$

Таким образом, максимально возможное значение входной производительности фотоприемника ($P_{\text{ф.}max}$) может быть найдено из следующих соотношений:

– при заданном τ

$$P_{\text{ф.}max} = H_{\text{ф.}max} \cdot F = \frac{1}{\tau} \cdot \log_2\left(\frac{(\Phi_{max} - \Phi_{min}) \cdot \sum_{i\,min}}{\sqrt{\frac{e \cdot i \cdot M(1+B)}{\tau}}} + 1\right); \quad (4.53)$$

– при заданном $H_{\text{и}\,max} = \log_2 m$

$$P_{\text{ф}\,max} = H_{\text{и}.max} \cdot F_{max} = \log_2 m \cdot \frac{(\Phi_{max} - \Phi_{min})^2 \cdot \sum_{i\,min}^2}{(m+1)^2 \cdot e \cdot i \cdot M(1+B)}; \quad (4.54)$$

где $F = \frac{1}{\tau}$, $F_{max} = \frac{1}{\tau_{min}}$ - частота измерения датчика;

m - объем алфавита источника сообщения.

Лабораторный практикум.

Лабораторная работа №1.
Информация в дискретных сообщениях.

Цель работы. Научиться практически определять количество информации в различного вида дискретных сообщениях.

Теоретическое обоснование. Количество информации, содержащееся в дискретном сообщении (I) можно найти из простого соотношения

$$I = n \cdot H,$$

где n — число символов в сообщении,

H- энтропия источника сообщений, то есть среднее количество информации, приходящееся на один символ сообщения.

Энтропия источника сообщения определяется из основного соотношения теории информации (1.4), которое для удобства практического использования преобразуется к виду наиболее простому и удобному в зависимости от свойств дискретного источника сообщений.

В случае, если символы источника сообщения появляются равновероятно и взаимно независимо, то для подсчета энтропии такого рода сообщений используют формулу Хартли:

$$I = n \cdot \log_2 m \ (бит); \qquad H_1 = \log_2 m \ (бит/симв),$$

где m- объем алфавита источника дискретных сообщений.

Если же символы источника сообщения генерируются с различными вероятностями, но взаимно независимы, то используют формулу Шеннона:

$$I = -n \cdot \sum_{i=1}^{m} P_{a_i} \cdot \log_2 P_{a_i} \ (бит),$$

$$H_2 = -\sum_{i=1}^{m} P_{a_i} \cdot \log_2 P_{a_i} \ (бит/симв),$$

где P_{ai} — вероятность появления символа a_i.

В случае же неравновероятного появления символов источника сообщения и наличия статистических зависимостей между соседними символами энтропию такого рода источника можно определить с помощью формулы Шеннона с условными вероятностями:

$$H_2 = -\sum_{i=1}^{m} P_{a_i} \cdot \sum_{i=1}^{m} P\left(a_j / a_i\right) \cdot \log_2 P\left(a_j / a_i\right) \ (бит/симв)$$

где $P\left(\frac{a_j}{a_i}\right)$ — условная вероятность появления символа a_j после символа a_i.

Содержание работы.

1.Посчитать среднее количество информации, приходящееся на один символ источника дискретных сообщений (энтропию) в случаях:

а —равновероятного и взаимно независимого появления символов;

б —неравновероятного и взаимно независимого появления символов;

в —при неравновероятном появлении символов и наличии статистических связей между соседними символами.

В качестве дискретного источника сообщений взять источник с объемом алфавита m = 34 (аналогичный по объему алфавита тексту на русском языке: 33 буквы и пробел), а его статистические характеристики смоделировать с помощью генератора случайных чисел.

2.Подсчитать количество информации в сообщении, представляющим собой Вашу фамилию, имя и отчество, считая, что символы сообщения появляются неравновероятно и независимо. Закон распределения символов найти путем анализа участка любого текста на русском языке длиной не менее 300 символов.

Выполнение работы. Работа выполняется на персональном компьютере в программном средстве «Mathcad». Так как в этом программном продукте в качестве встроенных функций используются только функции натуральных и десятичных логарифмов, то в процессе выполнения работы необходимо выполнить переход к логарифмам по основанию 2 по формуле перехода к иному основанию:

$$\log_b N = \frac{\log_a N}{\log_a b},$$

где a — основание известных логарифмов;

б — основание требуемых логарифмов;

N — логарифмируемая величина.

П.1.а. Используя формулу Хартли, найти энтропию указанного источника дискретных сообщений (H_1).

П.1.б. Смоделировать закон распределения символов дискретного источника сообщений, используя оператор *rnd (A)*, который генерирует случайные числа из диапазона [0,A] по следующей программе:

m : =34 — задание объема алфавита (m);

i : =1, 2,…,m — i- порядковый номер символа алфавита;

r(i) :=rnd (1) — генерирование 34 случайных чисел в интервале от 0 до 1;

$$I := \sum_i r(i)$$ — нахождение суммы всех *r(i);*

$$P(i) := \frac{r(i)}{I}$$ — *P(i)* – вероятность появления *i*-го символа (a_i).

Проверить правильность вычислений, найдя сумму всех *P(i)* при *i = 1,2,…,m*.

Построить график закона распределения *P(i)* Используя формулу Шеннона, определить энтропию смоделированного источника дискретных сообщений (H_2).

П.1.в. Смоделировать матрицу условных вероятностей появления символа a_j после символа a_i по следующей программе:

m : =34 -— задание объема алфавита (*m*);

$$\left.\begin{array}{l} i := 1,2,..m \\ j := 1,2,..m \end{array}\right\}$$ — порядковый номер символа алфавита;

r(i,j) := rnd(1) — генерирование матрицы (34×34) случайных чисел в интервале от 0 до1;

$$W_i := \sum_i r(i,\ j)$$ — нахождение суммы элементов в каждой строке матрицы *r(i,j)*;

$$S(i,\ j) := \frac{r(i,\ j)}{W_i}$$ —нормировка по строкам матрицы *r(i,j)* с целью получения суммы элементов в каждой строке, равной 1;

$$U_j := \sum_i S(i,\ j)$$ — нахождение сумм элементов в каждом столбце матрицы *S(i,j)*;

$$PP(i,\ j) := \frac{S(i,\ j)}{U_j}$$ — нормировка по столбцам матрицы *S(i,j)* с целью получения суммы элементов в каждом столбце равной 1.

Полученные значения элементов матрицы *PP(i,j)* приближенно можно считать условными вероятностями появления символа под номером *j* после *i*-го символа.

Используя формулу Шеннона с условными вероятностями определить энтропию смоделированного источника дискретных сообщений (H_3).

П.2.Определить вероятность появления каждого символа (буквы) P_i путем деления числа появлений этого символа (a_i) на общее число символов (не менее 300), входящих в сообщение. В случае, если какой-либо символ (из *m*= 34) в сообщении не встретился, считать, что он встретился 1 раз, иначе может возникнуть неопределенность в формуле Шеннона. Отсутствие в исследуемом сообщении какого-либо символа из состава алфавита источника сообщений свидетельствует лишь о том, что анализируемое сообщение не содержит достаточного числа символов (не достаточно длинное), чтобы появились все символы входящие в алфавит.

Построить график закона распределения символов (букв)в сообщении.

Проверить правильность полученного закона распределения, для чего найти сумму вероятностей появления каждого символа. Эта сумма должна быть равна 1.

С помощью формулы Шеннона найти энтропию (H_4) дискретного источника (текста на русском языке). Подсчитав число символов в Вашей фамилии, имени и отчестве (включая пробелы), найти количество информации, содержащейся в этом сообщении.

Контрольные вопросы.

1. Какие источники сообщений называют дискретными?
2. Для каких источников дискретных сообщений применимы формулы Хартли, Шеннона?
3. Каким образом описывается статистическая зависимость между соседними символами в дискретных сообщениях?
4. Дайте определение энтропии источника дискретных сообщений.
5. Как проверить правильность нахождения закона распределения символов источника дискретных сообщений?
6. Какой вид дискретных сообщений обладает наибольшей энтропией?

Лабораторная работа №2.
Информация в непрерывных сообщениях.

Цель работы. Изучение методов определения количества информации в непрерывных сообщениях.

Теоретическое обоснование. Непрерывные сообщения - это сообщения, построенные на основе бесконечного алфавита, поэтому соотношения, используемые для определения количества информации в дискретных сообщениях для них в общем случае не применимы. Исключение составляют лишь непрерывные сообщения, у которых символы появляются равновероятно, спектр ограничен, а сами они проявляются на некотором уровне шумов.

Действительно, такие сообщения могут быть представлены дискретными сообщениями с объемом алфавита, равным числу различимых уровней, и числом символов в сообщении, определяемым теоремой Котельникова. Количество информации содержащейся в них может быть посчитано с помощью формулы Хартли.

Число различимых уровней (L) определяют на основе соотношения:

$$L = \sqrt{\frac{P_c + P_ш}{P_ш}},$$

где P_c — мощность сообщения (сигнала);

$P_ш$ —мощность шума.

В соответствии с теоремой Котельникова, для передачи непрерывного сообщения длительностью T(сек) и граничной частотой в спектре F_m(сек$^{-1}$) достаточно передать его равноотстоящие мгновенные значения с интервалом Δt (сек) и общим числом N , причем $\Delta t \leq 1/2F_m$. и $N \geq 2TF_m+1$.

Используя формулу Хартли, количество информации, содержащееся в таком непрерывном сообщении, (I) и его энтропия (H) могут быть найдены по формулам:

$$I = N \cdot \log_2 L \cong (2TF_m + 1) \cdot \log_2 \sqrt{\frac{P_c + P_ш}{P_ш}} \quad (бит);$$

$$H = \frac{I}{N} \cong \log_2 \sqrt{\frac{P_c + P_ш}{P_ш}} \quad \left(\frac{бит}{символ}\right).$$

В общем случае, когда символы непрерывного сообщения появляются неравновероятно и их закон распределения описывается некоторой функцией плотности распределения вероятности *p(x)*, (*p(x)* характеризует вероятность попадания символа непрерывного сообщения x в некий интервал Δx), для подсчета информационных характеристик пользуются формулой для подсчета энтропии непрерывных сообщений.

Общая энтропия непрерывных сообщений, как показано в §1.8, равна бесконечности, однако выражение, ее описывающее, представляет собой сумму двух слагаемых, одно из которых стремится к бесконечности одинаковым образом для любых непрерывных сообщений, а второе является конечным и зависит от закона распределения символов непрерывного сообщения. Это слагаемое и называется дифференциальной или относительной энтропией.

Дифференциальная энтропия (H_x) определяется выражением:

$$H_x = -\int_{-\infty}^{\infty} p(x) \cdot \log_2 p(x)\, dx$$

Практическое использование дифференциальной энтропии основано на предположении, что непрерывные сообщения проявляются на некотором уровне аддитивных шумов (что всегда справедливо для реальных сообщений). В этом случае энтропия реального непрерывного сообщения H равна разности энтропий принятого (зашумленного) сообщения (H_c) и шума ($H_ш$), то есть

$$H = H_c + H_ш.$$

Содержание работы.

1.Подсчитать количество информации в реальном непрерывном сообщении при наличии аддитивного шума при условии, что символы сообщения появляются равновероятно.

В качестве такого непрерывного сообщения использовать сигнал $U(t)$ вида:

$$U(t) = A\sin(k \cdot t) + B\cos(n \cdot t) + C\sin(m \cdot t)$$

где k— Ваш номер по списку;

$n=k+2$;

$m=k+4$;

А, В, С берется из таблицы и определяется Вашим номером по списку.

Длительность сигнала $U(t)$ равна 2 сек. (T = 2сек.). В качестве аддитивного шума использовать случайный сигнал $X(t)$, получаемый с помощью генератора случайных чисел. Значения шума лежат в интервале от $-D/2$ до $D/2$ (D=1,2).

2.Подсчитать дифференциальную энтропию зашумленного аддитивным шумом сигнала $Q(t)=U(t) + X(t)$ и дифференциальную энтропию шума $X(t)$ и определить среднее количество информации, приходящееся на один символ, которое может быть извлечено из зашумленного сигнала $Q(t)$,считая, что все сигналы принимают свои значения с равной вероятностью.

Выполнение работы Работа выполняется на персональном компьютере в среде программного продукта «Mathcad»:

П.1.а). Построить графики сигналов $U(t)$, $X(t)$ и $Q(t) = U(t) + X(t)$.

б). Представить непрерывный сигнал $Q(t)$ в виде последовательности отсчетов, в соответствии с теоремой Котельникова. Следует учесть, что максимальная частота в спектре сигнала $U(t)$ определяется круговой частотой m (размерность: рад/сек), а максимальная частота спектра, входящая в выражение теоремы Котельникова, является циклической частотой (размерность $сек^{-1}=гц$), поэтому справедливо соотношение

$$F_m = \frac{m}{2\pi}.$$

в). Построить на одном экране графики $Q(t)$ и $Q_k(t)$, где $Q_k(t)$- график взятия выборок сигнала $Q(t)$.

г). Вычислить мощность полезного сигнала P_c и мощность шума $P_ш$ по формулам

$$P_c = \frac{1}{T}\int_0^T [Q(t)]^2 dt; \quad P_ш = \frac{1}{T}\int_0^T [X(t)]^2 dt$$

д). Найти количество информации (I), содержащееся в непрерывном сообщении при наличии аддитивного шума, и его энтропию (H).

П.2. а). Выразить аналитически функцию плотности распределения вероятности значений сигнала $Q(t)$ и построить ее график.

Так как сигнал *U(t)* и шум *X(t)*, по условию, принимают свои значения с одинаковой вероятностью, то и сигнал *Q(t)=U(t)+ X(t)* будет принимать все свои значения равновероятно во всем диапазоне значений от *-(A+B+C+D/2)* до *(A+B+C+D/2).*

Поэтому его функция плотности распределения вероятности *p(Q)* (с учетом того, что $\int_{-\infty}^{+\infty} p(Q)\,dQ = 1$) имеет следующий вид:

$$p(Q)=\begin{cases} 0. & \text{при } Q<-\left(A+B+C+\dfrac{D}{2}\right); \\ \dfrac{1}{2(A+B+C)+D} & \text{при } |Q| \le -\left(A+B+C+\dfrac{D}{2}\right); \\ 0. & \text{при } Q>\left(A+B+C+\dfrac{D}{2}\right). \end{cases}$$

Программа вычисления этой функции может быть реализована на основе оператора условного перехода *if* следующим образом:

Q := -10, –9,9….10

$$p(Q):=if\left(Q<\left(A+B+C+\frac{D}{2}\right),0,if\left(Q<\left(A+B+C+\frac{D}{2}\right),\frac{1}{2(A+B+C)+D},0\right)\right).$$

б). Выразить аналитически функцию плотности распределения вероятности шума *p(x)* и построить её график аналогично тому, как это сделано для функции плотности распределения вероятности сигнала *Q(t):*

$$x:=-2,-1.9..2.$$

$$p(x):=if\left(x<-\frac{D}{2},0,if\left(x<\frac{D}{2},\frac{1}{D},0\right)\right)$$

в). Определить дифференциальную энтропию сигнала *Q(t)* и дифференциальную энтропию шума *X(t)* в соответствии с определением дифференциальной энтропии:

$$H_c = -\int_{-\infty}^{+\infty} p(Q)\cdot \log_2 p(Q)\,dQ \quad \left(\frac{бит}{сим}\right)$$

$$H_ш = -\int_{-\infty}^{+\infty} p(x)\cdot \log_2 p(x)\,dx \quad \left(\frac{бит}{сим}\right).$$

г). Найти среднее количество информации, приходящееся на один символ зашумленного сигнала, как разность между H_c и $H_ш$ и сравнить ее с энтропией зашумленного сигнала *H*, вычисленного в п.1 этой лабораторной работы.

Контрольные вопросы

1. Какие источники сообщений называют непрерывными?
2. Сформулируйте теорему Котельникова.
3. Какое соотношение определяет число различимых уровней непрерывного сообщения при наличии аддитивного шума?
4. Дайте определение дифференциальной энтропии.
5. Чему равна полная энтропия непрерывного сообщения и из чего она слагается?

Лабораторная работа № 3.
Эффективное кодирование неравновероятных символов источника дискретных сообщений.

Цель работы. Ознакомление с алгоритмами эффективного кодирования неравновероятных символов источника дискретных сообщений и сравнение их эффективности.

Содержание работы. По номеру в списке группы (k) из Таблицы 1 выбрать закон распределения вероятности появления символов источника дискретных сообщений с объёмом алфавита $M=8$.

Произвести эффективное кодирование заданного источника дискретных сообщений по алгоритму Шеннона-Фено и по алгоритму Хафмена.

Рассчитать для построенных на основе этих алгоритмов кодов:

а). среднюю длину неравномерного кода ($n_н$);

б). избыточность неравномерного кода($R_{нк}$);

в). энтропию элементов символов полученных кодов($H_{1н}$);

Сравнить полученные результаты с соответствующими параметрами равномерного двоичного цифрового кода.

Таблица 1.

m_i	1	2	3	4	5	6	7	8	9	10
m_1	0.10	0.18	0.07	0.65	0.55	0.60	0.01	0.15	0.30	0/01
m_2	0.51	0.10	0.03	0.05	0.05	0.06	0.02	0.10	0.20	0.05
m_3	0.02	0.47	0.11	0.06	0.16	0.02	0.02	0.30	0.10	0.03
m_4	0.10	0.07	0.33	0.03	0.03	0.10	0.15	0.35	0.05	0.02
m_5	0.02	0.03	0.25	0.02	0.02	0.02	0.02	0.02	0.15	0.10
m_6	0.20	0.02	0.01	0.15	0.02	0.15	0.45	0.02	0.10	0.14
m_7	0.01	0.04	0.17	0.02	0.02	0.03	0.30	0.01	0.07	0.25
m_8	0.04	0.09	0.03	0.02	0.15	0.05	0.03	0.05	0.03	0.40

Контрольные вопросы.

1. Какой вид кодирования называют эффективным и в чем его специфика?
2. Что такое избыточность кодов?
3. Какие коды называются равномерными?
4. На каких принципах основано построение эффективных кодов при неравновероятном появлении символов сообщения?
5. Принцип построения эффективного кода по алгоритму Шеннона-Фено.
6. Принцип построения эффективного кода по алгоритму Хафмена.

Лабораторная работа №4.
Информационное моделирование источников визуальных сообщений и фотоизображений

Цель работы. Изучение методов информационной оценки качества источников визуальных сообщений и фотоизображений, используемых в системах автоматической обработки изображений.

Теоретическое обоснование. Качество считанных цифровых изображений в огромной степени зависит от качества исходных изображений, которое может характеризоваться их информационными емкостями, т.е. максимальным количеством информации приходящиеся на единицу их площади. Поэтому, исходя из информационных моделей, необходимо рассчитать информационные характеристики

- исходного оптического изображения ;
- фотоизображения.

1. Информационные характеристики оптического изображения, как функция линейного размера (d) элемента разложения, определяются из следующих соотношений:

   - число различимых уровней контраста (m):

$$m(d) = \sqrt{\frac{\Phi \cdot \Delta t \cdot d^2}{S_0 \cdot k^2 \cdot h \cdot \nu}} + 1 = \sqrt{\frac{E \cdot \Delta t \cdot \bar{s}}{k^2 \cdot h \cdot \nu}} + 1;$$

   - среднее максимальное количество информации приходящееся на один элемент разложения $h(d)$:

$$h(d) = \log_2 m(d) \quad \text{(бит/символ);}$$

   - предельное количество информации $I(d)$, содержащееся во всем изображении площадью S:

$$I(d) = \frac{S}{d^2} \cdot h(d) \quad \text{(бит);}$$

   - информационная емкость изображения (предельное количество информации, содержащееся в изображении единичной площади) $H(d)$

$$H(d) = \frac{1}{d^2} \cdot h(d) \quad \text{(бит/ед. площади),}$$

где Φ - световой поток падающий на изображение;

S_0 - площадь изображения;

$E = \dfrac{\Phi}{S_0}$ - освещенность изображения;

Δt - время экспозиции;

$h = 6{,}625 \cdot 10^{-34}$ Дж·сек. - постоянная Планка;

ν - частота падающего монохроматического света;

k - кратность превышения минимально различимым контрастом среднеквадратического уровня шума, вызванного флуктуацией потока падающих фотонов, которая находится из Таблицы 1

Таблица 1

k	1	2	3	4	5
$P(\lvert n_m - n_0 \rvert \geq k \cdot \sigma)$	0,32	0,046	0,00027	0,000063	0,00000057

2. Характерной особенностью фотоизображения является наличие на них шумов, вызванного их зернистостью фотослоя, которые ведут к флуктуациям оптической

плотности, среднеквадратическое значение которой (σ_D) определяется выражением:

$$\sigma_D = \sqrt{\frac{0{,}434\,\overline{D}\,S_з}{\Delta S}},$$

где $S_з$ - средняя площадь проекции одного зерна фотослоя, зависящая от типа фотоносителя;

$\overline{D}$ - средняя оптическая плотность фотоизображения;

ΔS - площадь элементарного участка фотоизображения.

Отношение сигнал \ шум фотоизображения (Ψ_D) определяется соотношением:

$$\Psi_D = \frac{\overline{D} - D_0}{\sigma_D},$$

где D_0 - плотность вуали.

Предельное количество информации, содержащееся в фотоизображении единичной площади (т.е. информационная емкость фотоизображения) $I_ф$ можно найти из выражения:

$$I_ф = \frac{1}{\Delta S}\cdot\sum_{k=1}^{L}\log_2\left(\frac{D_0 + k\cdot\delta D}{\sqrt{\frac{0{,}434\cdot S_з}{\Delta S}\cdot(D_0 + k\cdot\delta D)}}+1\right) \quad \text{(бит/ед.площади)},$$

где L - число равноотстоящих уровней плотности;

$\delta D = \frac{D_{max} - D_0}{L}$ - величина разности между соседними уровнями.

Количество информации, содержащееся в одной элементарной площадке фотоизображения, определяется выражением:

$$h = \frac{I_ф}{N} \quad \text{(бит)},$$

где $N = \frac{1}{\Delta S}$.

Выполнение работы:

1. Определить информационные характеристики оптического изображения размером S=0,2*0,2 м^2 при его равномерной освещенности гелий-неоновым лазером мощностью $Ф$ и длиной волны света λ=0,63 мкм.($\nu = 4{,}76\cdot 10^{14}\ с^{-1}$). Считывание производится квадратной апертурой с линейным размером d

($d^2 = \bar{S}$) с быстродействием N элементов/сек. ($\Delta t = \frac{1}{N}$) и допустимой погрешностью измерения различных уровней контраста не превышающей ε. Значения Φ, $d, \Delta t, \varepsilon$ взять из Таблицы 2 в соответствии с заданным вариантом.

Построить графики зависимостей: $m(d),\ h(d),\ I(d),\ H(d)$ при $\Delta t = 1\cdot 10^{5}$ сек.

Найти и распечатать значения этих характеристик при $d =$ 2, 5, 10, 15, 25 мкм.

Построить графики зависимостей: $m(\Delta t),\ h(\Delta t),\ I(\Delta t),\ H(\Delta t)$ при $d =$ 10 мкм.

Найти и распечатать значения этих характеристик при $\Delta t = 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}$ сек.

Проанализировать полученные результаты.

2. Произвести анализ выражения для определения информационной емкости фотоизображения на примере фотоизображения площадью 1 мм^2 при $D_{\max}$=2,7; D_0=0,1; $S_з$=$5\cdot 10^{-6}$ мм^2; d=$10\cdot 10^{-3}$ мм. При этих значениях параметров найти значение I_Φ при L=2^n (n=1, 2,….8).

Построить графики информационных характеристик I_Φ и h от линейного размера элементарной площадки $d(\Delta S = d^2)$ при значении L=64.

Найти и распечатать значения этих характеристик при d=2, 5, 10, 15, 25 мкм.

3. Сравнить полученные результаты с п.1, т. е. определить относительные потери информации, содержащейся в оптическом изображении, при регистрации его на фотоносителе при заданном d. Для этого вычислить коэффициент $\eta(d)$ при $d =$ 2, 5, 10, 15, 25 мкм.

$$\eta(d) = \frac{H(d)}{I_\Phi(d)}.$$

Таблица 2.

№	1	2	3	4	5	6	7	8	9	10	11	12	13
Ф(вт.)	1	1	1	1	2	2	2	2	3	3	3	3	1
d(мкм)	10	20	25	30	10	15	20	25	30	10	15	20	25
Δt (сек.)	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-6}	10^{-5}	10^{-4}
ε	0,01	0,01	0,1	0,1	0,01	0,01	0,1	0,1	0,01	0,01	0,1	0,1	0,01

Контрольные вопросы.

1. Что такое информационное моделирование?
2. Общий вид информационной модели и её составляющие.
3. Дайте определение пространственного разрешения и разрешения по плотности.
4. Информационные характеристики оптического изображения.
5. От каких основных параметров зависит информационная ёмкость оптического изображения.
6. Что является причиной появления шумов фотоизображения, как они проявляются и от каких параметров зависят.
7. От каких основных параметров зависит информационная ёмкость фотоизображения.

Приложение.

Некоторые полезные сведения из теории вероятностей.

1. Случайные события.

1.1. Событием (U) называют всякий факт, который может произойти или не произойти.

1.2. Вероятностью события P(U) называется численная мера степени объективной возможности этого события.

1.3. *Достоверным событием* называют событие *(U)*, которое в результате опыта непременно должно произойти. Вероятность достоверного события принимается равной единице, т.е. *P(U)=1*.

1.4 *Невозможным событием* называют событие *Q*, которое в результате опыта никоим образом не может произойти. Вероятность невозможного события принимается равной нулю, т.е. *P(Q)=0*.

1.5 Из п.1.3. и 1.4. следует, что вероятность любого реального события (A) заключена в интервале от 0 до 1, т.е. $0 \le P(A) \le 1$.

1.6 Несколько событий в данном опыте называют *несовместными*, если никакие два из них не могут произойти одновременно.

*1.7.*Несколько событий в данном опыте называют *равновозможными*, если по условиям симметрии опыта нет оснований считать какое-либо из них более предпочтительным или возможным.

1.8. *Условной вероятность события А* при наличии события *В* называют вероятность события *А*, вычисленную при условии, что событие *В* произошло. Условная вероятность в этом случае обозначается как *P(A/B).*

1.9. События называются *независимыми*, если появление одного из них никоим образом не меняет вероятности появления другого. Для независимых событий *А* и *В* справедливо:

$$P\left(A/B\right) = P(A), \qquad P\left(B/A\right) = P(B).$$

1.10. Полной группой событий называют несколько событий таких, что в результате опыта непременно произойдет хотя бы одно из них. Если события A_k *(k=1, 2, …,n)* составляют полную группу событий и несовместны, то

$$\sum_k P(A_k) = 1,$$

так как $\sum_k A_k$ - достоверное событие.

1.11. Если несколько событий

а) образуют полную группу событий;

б) несовместны;

в) равновозможны,

то вероятность события *А* можно вычислить по формуле

$$P(A) = \frac{m}{n},$$

где n – общее число возможных событий (исходов опыта),

m – число событий, благоприятствующих событию *А*.

2. Алгебра событий

2.1. Случайные события обозначают большими буквами алфавита.

2.2. *Равенство двух событий А и В (А=В)* означает, что появление одного события непременно влечет за собой появление другого.

2.3. *Суммой двух событий А и В* называют событие *С (С=А=В)*, заключающееся в появлении хотя бы одного из событий *А* и *В*.

2.4. *Произведением двух событий А и В* называют событие *С(С=А·В),* заключающееся в одновременном наступлении обоих событий *А* и *В*.

2.5. Если результат какого- либо опыта может иметь два взаимно исключающих события *А* и *В* (полная группа событий), одно из этих событий называют *противоположным событием* другому, например, событие *В* - противоположное событию *А* и обозначается *Ă* (читается как «не *A*»). Для прямого события *A* и противоположного события *Ă* (как составляющих полную группу событий) справедливо тождество:

$$P(A) + P(\breve{A})=1.$$

2.6. *Теорема сложения вероятностей.* Вероятность суммы двух событий *A* и *B* равна сумме вероятностей этих событий без вероятности произведения этих событий, т.е.

$$P(A+B)=P(A) + P(B)- P(A\cdot B).$$

В случае, если *A* и *B* несовместны, то $P(A\cdot B)=0$ и, следовательно,

$$P(A+B)=P(A)+P(B).$$

2.7.Теорема умножения вероятностей. Вероятность произведения двух событий *A* и *B* равна вероятности одного из них, умноженной на условную вероятность другого при наличии первого, т.е.

$$P(A\cdot B) = P(A\cdot B) = P(A)\cdot P(B/A) = P(B)\cdot P(A/B).$$

В случае, если события *A* и *B* независимы, то с учетом п.1.9 справедливо:

$$P(A\cdot B)=P(A)\cdot P(B)= P(B)\cdot P(A).$$

2.8. Полезное правило решения задач на нахождение вероятности случайных событий.

Для того, чтобы найти вероятность интересующего события, необходимо:

а) с помощью алгебры событий описать интересующее событие через известные события, т.е. через такие события, вероятности которых известны;

б) с помощью теорем вероятности суммы и вероятности произведения найти вероятность интересующего события.

3. Случайные величины.

3.1. *Случайной величиной* называют величину, которая в результате опыта может принимать то или иное численное значение, не известное заранее.

3.2. *Дискретной (прерывной) случайной величиной* называют случайную величину, которая может принимать только конечное число различных значений.

3.3. *Непрерывной случайной величиной* называют случайную величину, которая может принимать бесконечное число различных значений, заполняющих какой- либо промежуток.

3.4. *Законом распределения случайной величины* называют всякое соотношение, устанавливающее связь между всеми возможными значениями случайной величины и соответствующими им вероятностями.

3.5. Для дискретных случайных величин, в качестве закона распределения, чаще всего используют ряд распределения. Ряд распределения представляет собой таблицу, в верхней части которой размещены все возможные значения дискретной случайной величины, а в нижней части - соответствующие им вероятности.

X	x_1	x_2	….	x_k	….	x_n
$P(x_i)$	$P(x_1)$	$P(x_2)$	….	$P(x_k)$	….	$P(x_n)$

Формальным признаком правильности составления ряда распределения является выполнение условия:

$$\sum_{i=1}^{n} P(x_i) = 1.$$

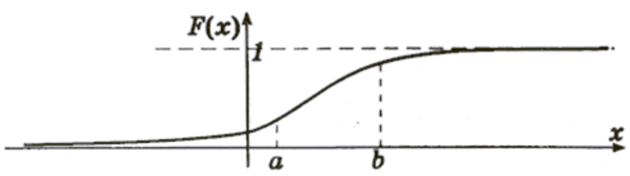
3.6. Для непрерывных случайных величин в качестве закона распределения часто используют функцию распределения (функция распределения может быть использована в качестве закона распределения и для дискретных величин). *Функцией распределения случайной величины* X называют функцию $F(x)$, выражающую вероятность того, что случайная величина X примет значение меньше, чем x, т.е.

$$F(x) = P(X<x).$$

Из определения функции распределения вытекают ее основные свойства:

$$1.\ 0 \le F(x) \le 1;$$
$$2.\ F(-\infty) = 0;$$
$$3.\ F(+\infty) = 1;$$
$4.\ F(x) -$ н е у б ы я в ф а ю н

В общем случае график функции распределения имеет вид:



Из определения $F(x)$ вытекает важное соотношение, выражающее вероятность попадания случайной величины X в интервал от a до b:

$$P(a \le X \le b) = F(b) - F(a).$$

3.7. Наряду с функцией распределения, в качестве закона распределения непрерывных случайных величин используют и функцию плотности распределения вероятности ($f(x)$), которая определяется как производная функции распределения, т.е.

$$f(x)=F'(x).$$

Из определения функции плотности распределения вероятности вытекают ее основные свойства:

$$1.\ f(-\infty)=0;$$
$$2.\ f(+\infty)=0;$$
$$3.\ f(x)\geq 0;$$
$$4.\ \int_{-\infty}^{+\infty} f(x)dx=1.$$

В общем случае график функции плотности распределения вероятности *f(x)* имеет вид:

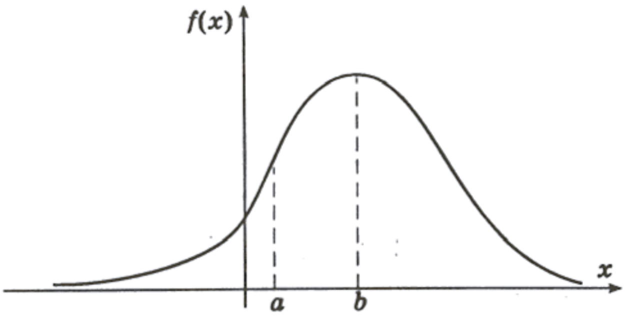


График функции *f(x)* называют *кривой распределения.*

3.8. *Элементом вероятности* для случайной величины *X* называют величину *f(x)dx*, выражающую вероятность попадания случайной величины *X* в элементарный отрезок *dx*, примыкающий к точке *x*.

3.9. Вероятность попадания случайной величины *X* в конечный промежуток от *a* до *b* определяется соотношением:

$$P(a\leq X\leq b)=\int_{a}^{b} f(x)d \quad ,$$

из которого следует свойство 4 п.3.7.

3.10. Из соотношения п. 3.9. совместно с п.3.6, можно получить выражение, связывающее функцию распределения (*F(x)*) и функцию плотности распределения вероятности (*f(x)*) случайной величины:

$$F(x)=\int_{-\infty}^{x} f(x)d \ .$$

4.Статистические характеристики случайных величин.

4.1. *Математическое ожидание* случайной величины *X* (обозначают *M[X]* или m_x) – это среднее значение случайной величины, вычисленное по формулам:

$M[X]=\sum_{i=1}^{n} x_i\cdot p(x_i)$ - для дискретных случайных величин;

$M[X]=\int_{-\infty}^{+\infty} x \cdot f(x)d$ - для непрерывных случайных величин.

4.2. *Центрированной* случайной величиной ^{o}X называют разность между самой случайной величиной X и ее математическим ожиданием $M[X]$, т.е.

$$^{o}X = X - M[X].$$

4.3. *Дисперсия* случайной величины X (обозначается $D[X]$ или d_x) — это есть математическое ожидание квадрата соответствующей ей центрированной случайной величины, т.е.

$$D[X]=M[^{o}X^2],$$

которая вычисляется по формулам:

$D[X]=\sum_{i=1}^{n}(x_i - M[X])^2 \cdot p(x_i)$ - для дискретных случайных величин;

$D[X]=\int_{-\infty}^{+\infty}(x - M[X])^2 \cdot f(x)d$ - для непрерывных случайных величин.

Дисперсия характеризует среднее отклонение значений случайной величины от её математического ожидания. Размерность дисперсии не совпадает с размерностью характеризуемой случайной величины. Размерность дисперсии есть квадрат размерности соответствующей случайной величины.

4.4. *Среднее квадратическое отклонение ($\sigma_ч$)* случайной величины X—это есть квадратный корень из ее дисперсии, т.е.

$$\sigma_ч = \sqrt{D[X]}.$$

Среднее квадратическое отклонение иногда называют стандартом. Среднее квадратическое отклонение имеет ту же размерность, что и сама случайная величина.

4.5. *Начальным моментом порядка k (v_k)* случайной величины X называют математическое ожидание k-той степени этой случайной величины, т.е.

$$v_k = M[X^k].$$

4.6. *Центральным моментом порядка k* (μ_k) случайной величины X называют математическое ожидание k-той степени отклонения этой случайной величины от ее математического ожидания, т.е.

$$\mu_k = M[(X - M[X])^k]$$

4.7. *Эксцесс* случайной величины *(E)* – это есть величина, вычисленная по формуле:

$$E=\frac{\mu_4}{(\mu_2)^2}-3.$$

Для нормального закона распределения $E=0$, отличие эксцесса от нуля указывает на отклонение эмпирического закона распределения от нормального закона распределения.

4.8. *Ассиметрия* характеризует симметричность кривой распределения случайной величины X. Показатель ассиметрии (S) вычисляется по формуле:

$$S = \left(\frac{\mu_3}{(\sigma_x)^3} \right).$$

Для симметричных распределений S=0.

5.Случайные функции.

5.1. Случайной функцией X(t) называют функцию, которая в результате опыта может принимать тот или иной конкретный вид, неизвестный заранее.

5.2. Конкретный вид, принимаемый случайной функцией, называют *реализацией* случайной функции.

5.3. Сечением случайной функции называют случайную величину $X(t_k)$, в которую обращается случайная функция $X(t)$ при фиксированном аргументе ($t = t_k$).

5.4. Одномерным законом распределения случайной функции $X(t)$ называют закон распределения $f(x,t_k)$ сечения $X(t)$ случайной функции.

5.5. Математическим ожиданием случайной функции $X(t)$ называют неслучайную функцию $m_x(t)$, которая при каждом значении аргумента t представляет собой математическое ожидание соответствующего сечения этой случайной функции.

5.6. Дисперсией случайной функции $X(t)$ называют неслучайную функцию $d_x(t)$, которая при каждом значении аргумента t представляет собой дисперсию соответствующего сечения этой случайной функции.

5.7. Корреляционной функцией случайной функции $X(t)$ называют неслучайную функцию двух аргументов $R_x(t_k,t_1)$, которая при каждой паре значений аргументов t_k и t_1 равна корреляционному моменту соответствующих сечений случайной функции, т. е.

$$R_x(t_k, t_1) = M\left[\dot{X}(t_k) \cdot \dot{X}(t_l)\right],$$

где $\dot{X}(t) = X(t) - m_x(t)$ - центрированная случайная функция.

Корреляционная функция характеризует статистическую связь между сечениями случайной функции, т.е. внутреннюю структуру случайной функции. При $t_k = t_1$ корреляционная функция обращается в дисперсию, действительно,

$$R_x(t_k, t_k) = M\left[(\dot{X}(t_k)\ ^2\right].$$

5.8. Нормированной корреляционной функцией случайной функции *X(t)* называют неслучайную функцию двух аргументов $r_x(t_k,t_{k1})$, определяемую по формуле:

$$r_x(t_k,t_l)=\frac{R_x(t_k,t_l)}{\sigma_x(t_k)\cdot\sigma_x(t_l)}=\frac{R_x(t_k,t_1)}{\sqrt{D_x(t_k)\cdot D_x(t_l)}}\ ,$$

при $t_k = t_1$ $\qquad r_x(t_k\ ,t_{k1}) = 1.$

5.9. Стационарной случайной функцией называют случайную функцию, математическое ожидание которой постоянно ($m_x(t)=c\ o\ n$), а её корреляционая функция зависит только от разности между аргументами:

$$R_x(t_1,t_2)=R_x(\tau),$$

где $\tau = t_1 - t_2$.

Литература.

1.Журкин И.Г. Шавенько Н.К Сигналы Учебное пособие по курсу «Автоматизированная обработки аэрокосмической информации. –М.: Изд. МИИГАиК, 2007.

2.Дмитриев В.И. Прикладная теория информации.- М.:Изд.«Высшая школа»,1999г.

3.Хэмминг Р.В. Теория кодирования и теория инфорции.—М.:Изд. «Радио и связь»,1998г.

4.Колесник В.Д., Полтырев Г.Ш. Курс теории информации. -М.: Изд.»Наука», 1994г.

5.Мощиль В.И., Шавенько Н.К. Основы теории информации. Учебное пособие. –М.: Изд. МИИГАиК, 2006.

6. Мощиль В. И., Шавенько Н. К. Основы теории кодирования. Учебное пособие. –М.: Изд. МИИГАиК, 1999.

7. Новиков Ф.А. Дискретная математика для программистов. –Санкт–Петербург; Изд.дом «Питер», 2008.

Содержание-