Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

ПРИКЛАДНЫЕ АСПЕКТЫ ИНФОРМАТИКИ
сборник учебно-методических материалов

для направления подготовки 10.03.01 Информационная безопасность

Благовещенск

2019

Печатается по решению
редакционно-издательского совета
факультета математики и информатики
Амурского государственного
университета

Составитель: Самохвалова С.Г.

Прикладные аспекты информатики: сборник учебно-методических материалов для направления подготовки 10.03.01 Информационная безопасность – Благовещенск: Амурский гос. ун-т, 2019

КРАТКОЕ ИЗЛОЖЕНИЕ ЛЕКЦИОННОГО МАТЕРИАЛА

Понятие сигнала. В XVIII веке в теорию математики вошло понятие функции, как определенной зависимости какой-либо величины **y** от другой величины – независимой переменной **x**, с математической записью такой зависимости в виде **y(x)**. Довольно скоро математика функций стала основой теории всех естественных и технических наук. Особое значение функциональная математика приобрела в технике связи, где временные функции вида s(t), v(f) и т.п., используемые для передачи информации, стали называть сигналами.

В технических отраслях знаний термин "сигнал" (signal, от латинского signum – знак) используется в широком смысловом диапазоне. Под ним понимают и техническое средство для передачи, обращения и использования информации - электрический, магнитный, оптический сигнал; и физический процесс, отображающий информационное сообщение - изменение какого-либо параметра носителя информации (электромагнитных колебаний, светового потока и т.п.) во времени, в пространстве или в зависимости от изменения значений каких-либо других аргументов (независимых переменных); и смысловое содержание определенного физического состояния или процесса, как, например, сигналы светофора, звуковые предупреждающие сигналы и т.п. Все эти понятия объединяет конечное назначение сигналов. Это определенные сведения, сообщения, информация о каких-либо процессах, состояниях или физических величинах объектов материального мира, выраженные в форме, удобной для передачи, обработки, хранения и использования этих сведений.

Термин “сигнал” часто отождествляют с понятиями “данные” (data) и “информация” (information). Действительно, эти понятия взаимосвязаны, но относятся к разным категориям.

Понятие информации имеет много определений, от наиболее широкого (информация есть формализованное отражение реального мира) до практического (сведения, являющиеся объектом хранения, передачи, преобразования, восприятия и управления). Мировая наука все больше склоняется к точке зрения, что информация, наряду с материей и энергией, принадлежит к фундаментальным философским категориям естествознания и относится к одному из свойств объективного мира. Что касается “данных” (от латинского datum – факт), то это совокупность фактов, результатов наблюдений, измерения каких-либо физических свойств объектов, явлений или процессов материального мира, представленных в формализованном виде. Это не информация, а сырье для получения информации путем соответствующей обработки и интерпретации (истолкования).

Термин "signal" в мировой практике является общепринятым для характеристики формы представления данных, при которой данные рассматриваются как результат некоторых измерений объекта исследований в виде последовательности значений скалярных величин (аналоговых, числовых, графических и пр.) в зависимости от изменения каких-либо переменных значений (времени, энергии, температуры, пространственных координат, и пр.). А так как данные содержат информацию, как об основных целевых параметрах объекта исследований, так и о различных сопутствующих и мешающих факторах измерений, то в широком смысле этого слова можно считать, что сигнал является носителем общей измерительной информации. При этом материальная форма носителей сигналов (механическая, электрическая, магнитная, акустическая, оптическая и любая другая), равно как и форма отображения данных в каких-либо физических параметрах или процессах носителей, значения не имеет. Информативным параметром сигнала может являться любой параметр носителя сигнала, функционально и однозначно связанный со значениями информационных данных.

Наиболее распространенное представление сигналов - в электрической форме в виде зависимости напряжения от времени U(t). Так, например, сигнал изменения напряженности магнитного поля по профилю аэросъемки – это и временная последовательность изменения электрического напряжения на выходе датчика аэромагнитометра, и запись этого напряжения на ленте регистратора, и последовательные значения цифровых отсчетов при обработке лент регистратора и вводе сигнала в ЭВМ.

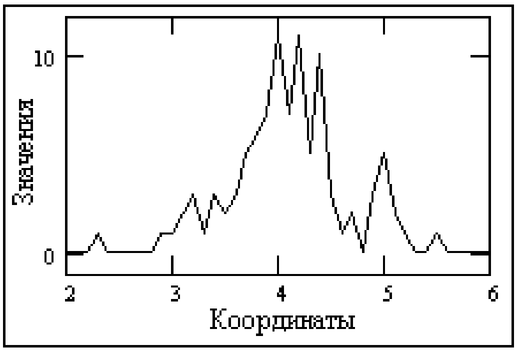


Рис. 1.1.1. Сигнал.

С математической точки зрения сигнал представляет собой функцию, т.е. зависимость одной величины от другой, независимой переменной. *По содержанию это информационная функция, несущая сообщение о физических свойствах, состоянии или поведении какой-либо физической системы, объекта или среды. А целью обработки сигналов можно считать извлечение определенных информационных сведений, которые отображены в этих сигналах (кратко - полезная или целевая информация) и преобразование этих сведений в форму, удобную для восприятия и дальнейшего использования.*

Под "анализом" сигналов имеется в виду не только их чисто математические преобразования, но и получение на основе этих преобразований выводов о специфических особенностях соответствующих процессов и объектов. Целями анализа сигналов обычно являются:

- Определение или оценка числовых параметров сигналов (энергия, средняя мощность, среднее квадратическое значение и пр.).
- Изучение изменения параметров сигналов во времени.
- Разложение сигналов на элементарные составляющие для сравнения свойств различных сигналов.
- Сравнение степени близости, "похожести", "родственности" различных сигналов, в том числе с определенными количественными оценками.

Математический аппарат анализа сигналов весьма обширен, и широко применяется на практике во всех без исключения областях науки и техники.

С понятием сигнала неразрывно связан термин *регистрации* сигналов, использование которого также широко и неоднозначно, как и самого термина сигнал. В наиболее общем смысле под этим термином можно понимать операцию выделения сигнала и его преобразования в форму, удобную для дальнейшего использования. Так, при получении информации о физических свойствах каких-либо объектов, под регистрацией сигнала понимают процесс измерения физических свойств объекта и перенос результатов измерения на материальный носитель сигнала или непосредственное энергетическое преобразование каких-либо свойств объекта в информационные параметры материального носителя сигнала (как правило - электрического). Но так же широко термин регистрации сигналов используют и для процессов выделения уже сформированных сигналов, несущих определенную информацию, из суммы других сигналов (радиосвязь, телеметрия и пр.), и для процессов фиксирования сигналов на носителях долговременной памяти, и для многих других процессов, связанных с обработкой сигналов.

Применительно к настоящему курсу под термином регистрации будем понимать *регистрацию данных* (data logging), которые проходят через конкретную систему или точку системы и определенным образом фиксируются на каком-либо материальном носителе или в памяти системы. Что касается процесса получения информации при помощи технических средств, обеспечивающих преобразование физических величин в сигналы, удобные для обработки и восприятия, то для этого процесса будем применять, в основном, термин *детектирования*.

<u>Шумы и помехи</u>. При детектировании сигналов, несущих целевую для данного вида измерений информацию, в сумме с основным сигналом одновременно регистрируются и мешающие сигналы - шумы и помехи самой различной природы (рис. 1.1.2). Шумы, как правило, имеют случайный (стохастический) характер. К помехам относят стационарные искажения полезных сигналов при влиянии на процессы измерений различных дестабилизирующих факторов (электромагнитные наводки, вибрация, и т.п.). Выделение полезных составляющих из общей суммы зарегистрированных сигналов или максимальное подавление шумов и помех в информационном сигнале при сохранении его полезных составляющих является одной из основных задач первичной обработки результатов наблюдений.

Рис. 1.1.2. Сигнал с помехами.

Виды шумов и помех разделяют по источникам их возникновения, по энергетическому спектру, по характеру воздействия на сигнал, по вероятностным характеристикам и другим признакам. Источники шумов и помех бывают внутренние и внешние.

Внутренние, как правило, присущи физической природе источников и детекторов сигналов, а также их материальных носителей. Например, флюктуации интенсивности излучения радионуклидов в силу статистической природы ядерных процессов, тепловые шумы электронных потоков в электрических цепях, и т.п.

Внешние источники шумов и помех бывают искусственного и естественного происхождения. К искусственным источникам относятся индустриальные помехи и помехи от работающей физико-технической аппаратуры. Естественными источниками являются молнии, флюктуации магнитных полей, всплески солнечной энергии, и т.д. Электрические и магнитные поля различных источников помех вследствие наличия индуктивных, емкостных и резистивных связей создают в цепях сигнальных систем паразитные разности потенциалов и токи, накладывающиеся на полезные сигналы.

Помехи подразделяются на флюктуационные, импульсные и периодические.

Флюктуационные помехи представляют собой хаотические и беспорядочные во времени процессы в виде нерегулярных случайных всплесков различной амплитуды. Как правило, флюктуационные помехи распределены по нормальному закону с нулевым средним.

Импульсные помехи проявляются как в виде отдельных импульсов, так и в виде последовательности импульсов, форма и параметры которых имеют случайный характер. Причинами импульсных помех являются резкие броски тока и напряжения в промышленных установках, транспортных средствах, а также природные электрические явления.

Периодические помехи вызываются электромагнитными полями линий электропередач, силовых электроустановок и др. Если основная мощность помех сосредоточена на отдельных участках диапазона частот, например, на частоте напряжения промышленной сети или кратна этой частоте, то такие помехи называют сосредоточенными.

В зависимости от характера воздействия на сигнал помехи разделяют на аддитивные и мультипликативные. Аддитивные (налагающиеся) помехи суммируются с сигналом, не зависят от его значений и формы и не изменяют информативной составляющей самого сигнала. Мультипликативные или деформирующие помехи могут изменять форму информационной части сигнала, иметь зависимость от его значений и от определенных особенностей в сигнале и т.п. При известном характере мультипликативных помех возможна коррекция сигнала на их влияние.

Следует заметить, что деление сигналов на полезные и мешающие (шумовые) является достаточно условным. Источниками мешающих сигналов также могут быть определенные физические процессы, явления или объекты. При выяснении природы мешающих сигналов они могут переводиться в разряд информационных. Так, например, вариации диаметра скважин и каверны является мешающим фактором практически для всех методов каротажа. Вместе с тем этот же фактор, при соответствующем методическом и аппаратурном обеспечении, может дать возможность бесконтактного определения диаметра скважин в качестве дополнительного информационного параметра.

Размерность сигналов. Простейшими сигналами геофизической практики являются одномерные сигналы, как, например, сейсмические импульсы s(t), измерения каких-либо параметров геофизических полей (электрических, магнитных, и пр.) по профилям на поверхности земли s(x) или по стволу скважины s(h), и т.п. Значения одномерных сигналов зависят только от одной независимой переменной, как, например, на рис. 1.1.1 и 1.1.2.

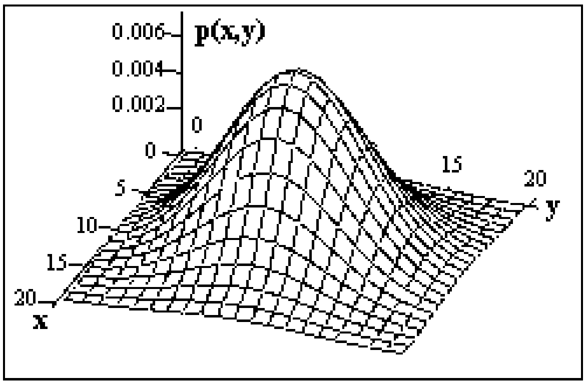


Рис. 1.1.3. Двумерный сигнал.

В общем случае сигналы являются многомерными функциями пространственных, временных и прочих независимых переменных - сейсмическая волна вдоль линии профиля s(x,t), аномалия гравитационного поля на поверхности наблюдений s(x,y), пространственно - энергетическое распределение потока ионизирующих частиц или квантов от источника излучения s(x,y,z,E) и т.п. Все большее применение находят также многомерные сигналы, образованные некоторым множеством одномерных сигналов, как, например, комплексные каротажные измерения нескольких физических параметров горных пород по стволу скважины одновременно.

Многомерные сигналы могут иметь различное представление по своим аргументам. Так, полный акустический сигнал сейсмического профиля дискретен по пространству (точкам расположения приемников) и непрерывен по времени.

Многомерный сигнал может рассматриваться, как упорядоченная совокупность одномерных сигналов. С учетом этого при анализе и обработке сигналов многие принципы и практические методы обработки одномерных сигналов, математический аппарат которых развит достаточно глубоко, распространяются и на многомерные сигналы. Физическая природа сигналов для математического аппарата их обработки значения не имеет.

Вместе с тем обработка многомерных сигналов имеет свои особенности, и может существенно отличаться от одномерных сигналов в силу большего числа степеней свободы. Так, при дискретизации многомерных сигналов имеет значение не только частотный спектр сигналов, но и форма растра дискретизации. Пример не очень полезной особенности - многомерные полиномы сигнальных функций, в отличие от одномерных, не разлагаются на простые множители. Что касается порядка размерности многомерных сигналов, то ее увеличение выше двух практически не изменяет принципы и методы анализа данных, и сказывается, в основном, только на степени громоздкости формул и чисто техническом усложнении вычислений.

Учитывая эти факторы, при рассмотрении общей теории анализа, преобразований и обработки сигналов ограничимся, в основном, одно- и двумерными сигнальными функциями. В качестве универсальных независимых переменных (аргументов функций) будем использовать, как правило, переменную "t" для одномерных сигналов и переменные "x,t" или "x,y" для двумерных сигналов, безотносительно к их физическому содержанию (пространство, время, энергия и пр.).

Математическое описание сигналов. Сигналы могут быть объектами теоретических исследований и практического анализа только в том случае, если указан способ их математического описания. Математическое описание позволяет абстрагироваться от физической природы сигнала и материальной формы его носителя, проводить классификацию сигналов, выполнять их сравнение, устанавливать степень тождества, моделировать системы обработки сигналов.

Большинство сигналов, встречающихся на практике, представлены во временной области функциями времени. При отображении сигналов на графике одной из координат (независимой) является ось времени, а другой координатой (зависимой) – ось амплитуд. Тем самым мы получаем амплитудно-временное представление сигнала. В общем случае описание сигнала задается функциональной зависимостью определенного информационного параметра сигнала от независимой переменной (аргумента) – s(x), y(t) и т.п. Такая форма описания и графического представления сигналов называется динамической (сигнал в реальной динамике его поведения по аргументам). Функции математического описания сигналов могут быть как вещественными, так и комплексными. Выбор математического аппарата описания определяется простотой и удобством его использования при анализе и обработке сигналов.

Отметим двойственность применения описания сигналов функциями типа s(t) и т.п. С одной стороны s(t) – это величина, равная значению функции в момент времени t. С другой стороны мы обозначаем через s(t) и саму функцию, т.е. то правило, по которому каждому значению t ставится в соответствие определенная величина s. В большинстве аналитических выражений это не вызывает недоразумений и при однозначном соответствии значений сигналов их аналитическим выражениям принимается по умолчанию.

Сделаем также одно замечание по терминологии описания сигналов. В теоретических работах по анализу сигналов конкретные значения величины сигнала (отсчеты значений по аргументу) часто именуют координатами сигнала. В отраслях знаний, связанных с геологией и горным делом, и в геофизической практике в том числе, этот термин используется по своему прямому смысловому назначению – пространственных координат результатов измерений, и является неизменным атрибутом всех геолого-геофизических данных. С учетом последнего фактора условимся применять термин “координата” по своему традиционному смысловому назначению в качестве обобщающего термина для независимых переменных сигнальных функций. При этом под понятием *координат* значений сигнала будем понимать не только какие-либо пространственные координаты, как это непосредственно имеет место для результатов измерений при геолого-геофизических съемках, но и любые другие аргументы, на числовой оси которых отложены значения или отсчеты сигнала и рассматривается динамика его изменения (пример на рис. 1.1.1).

Спектральное представление сигналов. Кроме динамического представления сигналов и функций в виде зависимости их значений от определенных аргументов при анализе и обработке данных широко используется математическое описание сигналов по аргументам, обратным аргументам динамического представления. Так, например, для времени обратным аргументом является частота. Возможность такого описания определяется тем, что любой сколь угодно сложный по своей форме сигнал, не имеющий разрывов второго рода (бесконечных значений на интервале своего задания), можно представить в виде суммы более простых сигналов, и, в частности, в виде суммы простейших гармонических колебаний, что выполняется при помощи преобразования Фурье. Соответственно, математически разложение сигнала на гармонические составляющие описывается функциями значений амплитуд и начальных фаз колебаний по непрерывному или дискретному аргументу – частоте изменения функций на определенных интервалах аргументов их динамического представления. Совокупность амплитуд гармонических колебаний разложения называют *амплитудным спектром* сигнала, а совокупность начальных фаз – *фазовым спектром*. Оба спектра вместе образуют полный частотный спектр сигнала, который по точности математического представления тождественен динамической форме описания сигнала.

Линейные системы преобразования сигналов описываются дифференциальными уравнениями, причем для них верен принцип суперпозиции, согласно которому реакция систем на сложный сигнал, состоящий из суммы простых сигналов, равна сумме реакций от каждого составляющего сигнала в отдельности. Это позволяет при известной реакции системы на гармоническое колебание с определенной частотой определить реакцию системы на любой сложный сигнал, разложив его в ряд гармоник частотного спектра сигнала. Широкое использование гармонических функций при анализе сигналов объясняется тем, что они являются достаточно простыми ортогональными функциями и определены при всех значениях непрерывных переменных. Кроме того, они являются собственными функциями времени, сохраняющими свою форму при прохождении колебаний через любые линейные системы и системы обработки данных с постоянными параметрами (изменяются только амплитуда и фаза колебаний). Немаловажное значение имеет и то обстоятельство, что для гармонических функций и их комплексного анализа разработан мощный математический аппарат.

Примеры частотного представления сигналов приводятся ниже (рис. 1.1.5 – 1.1.12).

Кроме гармонического ряда Фурье применяются и другие виды разложения сигналов: по функциям Уолша, Бесселя, Хаара, полиномам Чебышева, и др. Главное условие однозначности и математической идентичности отображения сигналов - ортогональность функций разложения. При качественном анализе сигналов могут применяться и неортогональные функции, выявляющие какие-либо характерные особенности сигналов, полезные для интерпретации физических данных.

Математические модели сигналов. Теория анализа и обработки физических данных базируется на математических моделях соответствующих физических полей и физических процессов, на основе которых создаются математические модели сигналов. Математические модели сигналов дают возможность обобщенно, абстрагируясь от физической природы, судить о свойствах сигналов, предсказывать изменения сигналов в изменяющихся условиях, заменять физическое моделирование процессов математическим. С помощью математических моделей имеется возможность описывать свойства сигналов, которые являются главными в изучаемых процессах, и игнорировать большое число второстепенных признаков. Знание математических моделей сигналов дает возможность классифицировать их по различным признакам, характерным для того или иного типа моделей. Так, сигналы разделяются на неслучайные и случайные в зависимости от возможности точного предсказания их значений в любые моменты времени. Сигнал является неслучайным и называется *детерминированным,* если математическая модель позволяет осуществлять такое предсказание. Детерминированный сигнал задается, как правило, математической функцией или вычислительным алгоритмом, а математическая модель сигнала может быть представлена в виде

$s = F(t, z, \square,\ldots; A, B, C,\ldots)$,

где s – информативный параметр сигнала; t, z, w, … – независимые аргументы (время, пространственная координата, частота и др.); A, B, C… – параметры сигналов.

Модель должна быть, по возможности, проще, минимизирована по количеству независимых аргументов и адекватна изучаемому процессу. Рассмотрим этот вопрос на примере геофизических данных.

Под геофизическим полем понимают собственное или индуцированное определенным внешним воздействием распределение какой-либо физической величины, создаваемое геологическим объектом или геологической структурой в пространстве, во времени или по любому другому аргументу (независимой переменной). В простейшем случае геофизический сигнал - это изменение какой-либо составляющей геофизического поля, т.е. сечение поля по одному из аргументов. В пределе геофизическое поле в целом может рассматриваться как первичный многомерный сигнал в прямом физическом отображении, с которого путем измерений могут сниматься формализованные копии определенных составляющих (сечений) сигнала на материальные носители информации.

Геофизическим полям в определенных условиях их регистрации соответствуют определенные математические модели сигналов, т.е. их описание на каком-либо формальном языке. Математическое описание не может быть всеобъемлющим и идеально точным и, по существу, всегда отображает не реальные объекты, а их упрощенные (гомоморфные) модели. Модели могут задаваться таблицами, графиками, функциональными зависимостями, уравнениями состояний и переходов из одного состояния в другое и т.п. Формализованное описание может считаться математической моделью оригинала, если оно позволяет с определенной точностью прогнозировать состояние и поведение изучаемых объектов путем формальных процедур над их описанием.

Неотъемлемой частью любой математической модели сигнала является область определения сигнала, которая устанавливается интервалом задания независимой переменной. Примеры задания интервала для переменных:

$a \le x \le b, \quad x \in [a, b].$

$a < y \le b, \quad y \in (a, b].$

$a < z < b, \quad z \in (a, b).$

Пространство значений независимой переменной обычно обозначается через индекс R. Так, например, $R:=(-\infty, +\infty)$, $x \in R$.

Кроме задания области определения сигнала могут быть также заданы виды численных значений переменных (целые, рациональные, вещественные, комплексные).

Математические модели полей и сигналов на первом этапе обработки и анализа результатов наблюдений должны позволять в какой-то мере игнорировать их физическую природу и возвращать ее в модель только на заключительном этапе интерпретации данных.

Виды моделей сигналов. При анализе физических данных используются два основных подхода к созданию математических моделей сигналов.

Первый подход оперирует с *детерминированными* сигналами, значения которых в любой момент времени или в произвольной точке пространства (а равно и в зависимости от любых других аргументов) являются априорно известными или могут быть определены (вычислены) с определенной степенью точности. Такой подход удобен в прямых задачах геофизики (расчеты полей для заданных моделей сред), в задачах активных воздействий на среду при заранее известных параметрах и форме сигнала воздействия (вибрационная сейсморазведка, электромагнитные методы каротажа и пр.), а также при использовании хорошо известных геолого-геофизических данных.

Второй подход предполагает случайный характер сигналов, закон изменения которых во времени (или в пространстве) носит случайный характер, и которые принимают конкретные значения с некоторой вероятностью. Модель такого сигнала представляет собой описание статистических характеристик случайного процесса путем задания закона распределения вероятностей, корреляционной функции, спектральной плотности энергии и др.

Случайность может быть обусловлена как собственной физической природой сигналов, что характерно, например, для методов ядерной геофизики, так и вероятностным характером регистрируемых сигналов как по времени или месту их появления, так и по содержанию. С этих позиций случайный сигнал может рассматриваться как отображение случайного по своей природе процесса или физических свойств объекта, которые определяются случайными параметрами или сложным строением геологической среды, результаты измерений в которой трудно предсказуемы.

Между этими двумя видами сигналов нет резкой границы. Строго говоря, детерминированных процессов и отвечающих им детерминированных сигналов в природе не существует. Даже сигналы, хорошо известные на входе в среду (при внешнем воздействии на нее), по месту их регистрации всегда осложнены случайными помехами, влиянием дестабилизирующих факторов и априорно неизвестными параметрами и строением самой среды. С другой стороны, модель случайного поля часто аппроксимируется методом суперпозиции сигналов известной формы. Детерминированные модели могут использоваться и для изучения случайных процессов, если уровень полезного сигнала в этом процессе значительно выше уровня статистических флюктуаций.

На выбор математической модели поля в немалой степени влияет также сложность математического аппарата обработки сигналов и сложившиеся традиции геологической интерпретации результатов наблюдений. Не исключается и изменение модели, как правило, с переводом из вероятностной в детерминированную, в процессе накопления информации об изучаемом объекте.

Классификация сигналов осуществляется на основании существенных признаков соответствующих математических моделей сигналов**.** Все сигналы разделяют на две крупных группы: детерминированные и случайные. Классификация сигналов внутри групп приведена на рис. 1.1.4.

Рис. 1.1.4. Классификация сигналов.

С математических позиций группы сигналов обычно называют множествами, в которые объединяют сигналы по какому-либо общему свойству. Принадлежность сигнала s к множеству L_P записывается в виде $L_P = \{s; P\}$, где P – определенное свойство данного множества сигналов.

Классификация детерминированных сигналов. Обычно выделяют два класса детерминированных сигналов: периодические и непериодические.

К множеству периодических относят гармонические и полигармонические сигналы. Для периодических сигналов выполняется общее условие $s(t) = s(t + kT)$, где $k = 1, 2, 3, \ldots$ - любое целое число (из множества целых чисел I от $-\infty$ до ∞), T - период, являющийся конечным отрезком независимой переменной. Множество периодических сигналов:

$L_P = \{s(t);\ s(t+kT) = s(t),\ -\infty < t < \infty,\ k \in I\}$.

Гармонические сигналы (синусоидальные), описываются следующими формулами:

$$s(t) = A\cdot\sin(2\pi f_o t+\phi) = A\cdot\sin(\omega_o t+\phi), \quad s(t) = A\cdot\cos(\omega_o t+\varphi), \qquad (1.1.1)$$

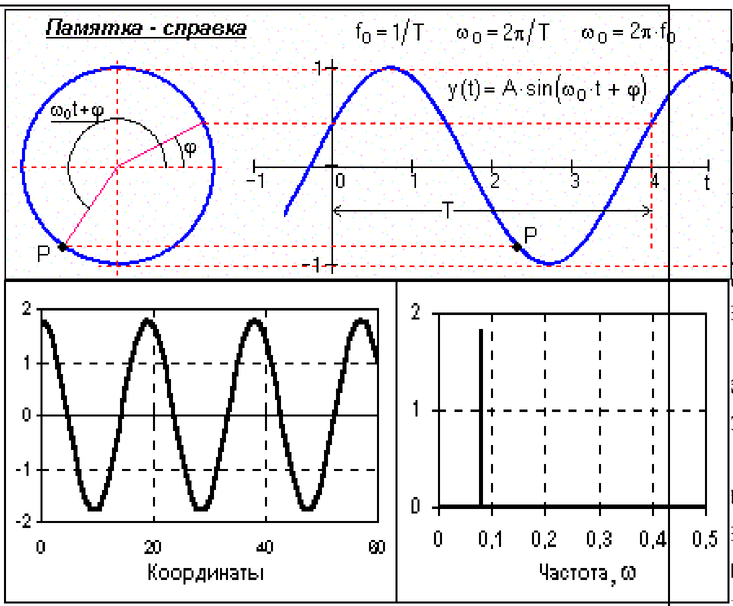


Рис. 1.1.5. Гармонический сигнал и спектр его амплитуд.

где A, f_o, □$_o$, □□□ϕ - постоянные ве-
чины, которые могут исполнять роль
формационных параметров сигнала: A -
плитуда сигнала, f_o - циклическая часто-
в герцах, □$_o$□= 2□f_o - угловая частота в
дианах, □ и ϕ- начальные фазовые углы
радианах. Период одного колебания T =
$_o$ = 2□/□$_o$. При φ = ϕ-π/2 синусные и ко-
нусные функции описывают один и тот
сигнал. Частотный спектр сигнала
едставлен амплитудным и начальным
зовым значением частоты f_o (при t = 0).

Полигармонические сигналы со-
вляют наиболее широко распростра-
нную группу периодических сигналов и
исываются суммой гармонических коле-
баний:

$$s(t) = \sum_{n=0}^{N} A_n \sin(2□f_nt+□_n) \equiv \sum_{n=0}^{N} A_n \sin(2□B_nf_pt+□_n), \quad B_n \in I, \qquad (1.1.2)$$

или непосредственно функцией $s(t) = y(t \pm kT_p)$, k = 1,2,3,..., где T_p - период одного полного колебания сигнала y(t), заданного на одном периоде. Значение $f_p = 1/T_p$ называют фундаментальной частотой колебаний.

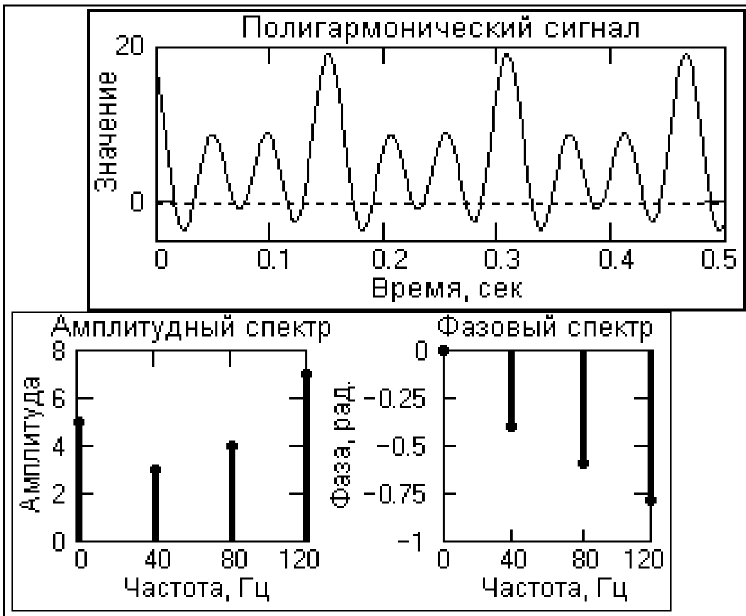


Рис. 1.1.6. Модель сигнала. Рис. 1.1.7. Спектр сигнала.

Полигармонические сигналы представляют собой сумму определенной постоянной составляющей (f_o=0) и произвольного (в пределе - бесконечного) числа гармонических составляющих с произвольными значениями амплитуд A_n и фаз φ_n, с частотами, кратными фундаментальной частоте f_p. Другими словами, на периоде фундаментальной частоты f_p, которая равна или кратно меньше минимальной частоты гармоник, укладывается кратное число периодов всех гармоник, что и создает периодичность повторения сигнала. Частотный спектр полигармонических сигналов дискретен, в связи с чем второе распространенное математическое представление сигналов - в виде спектров (рядов Фурье).

На рис. 1.1.6 приведен отрезок периодической сигнальной функции, которая получена суммированием постоянной составляющей и трех гармонических колебаний с разными значениями частоты и начальной фазы колебаний. Математическое описание сигнала задается формулой:

$$s(t) = \sum_{k=0}^{3} A_k \cdot \cos(2 \cdot \square \cdot f_k \cdot t + \square_k),$$

где: $A_k = \{5, 3, 4, 7\}$ - амплитуда гармоник; $f_k = \{0, 40, 80, 120\}$ - частота в герцах; $\square_k = \{0, -0.4, -0.6, -0.8\}$ - начальный фазовый угол колебаний в радианах; k = 0, 1, 2, 3. Фундаментальная частота сигнала 40 Гц.

Частотное представление данного сигнала (спектр сигнала) приведено на рис. 1.1.7. Обратим внимание, что частотное представление периодического сигнала s(t), ограниченного по числу гармоник спектра, составляет всего восемь отсчетов и весьма компактно по сравнению с временным представлением.

Периодический сигнал любой произвольной формы может быть представлен в виде суммы гармонических колебаний с частотами, кратными фундаментальной частоте колебаний $f_p\square = 1/T_p$. Для этого достаточно разложить один период сигнала в ряд Фурье по тригонометрическим функциям синуса и косинуса с шагом по частоте, равным фундаментальной частоте колебаний $\square f = f_p$:

$$s(t) = \sum_{k=0}^{K} (a_k \cos 2\square k\square ft + b_k \sin 2\square k\square ft), \qquad (1.1.3)$$

$$a_o = (1/T)\int_0^T s(t)\,dt, \quad a_k = (2/T)\int_0^T s(t) \cos 2\square k\square ft\,dt, \qquad (1.1.4)$$

$$b_k = (2/T)\int_0^T s(t) \sin 2\square k\square ft\,dt. \qquad (1.1.5)$$

Количество членов ряда Фурье $K = k_{max}$ обычно ограничивается максимальными частотами f_{max} гармонических составляющих в сигналах так, чтобы $f_{max} < K \cdot f_p$. Однако для сигналов с разрывами и скачками имеет место $f_{max} \to \infty$, при этом количество членов ряда ограничивается по допустимой погрешности аппроксимации функции s(t).

Одночастотные косинусные и синусные гармоники можно объединить и представить разложение в более компактной форме:

$$s(t) = \sum_{k=0}^{K} S_k \cos (2\square k\square ft - \square_k), \qquad (1.1.3')$$

$$S_k = \sqrt{a_k^2 + b_k^2}, \quad \square_k = \text{argtg}\,(b_k/a_k). \qquad (1.1.6)$$

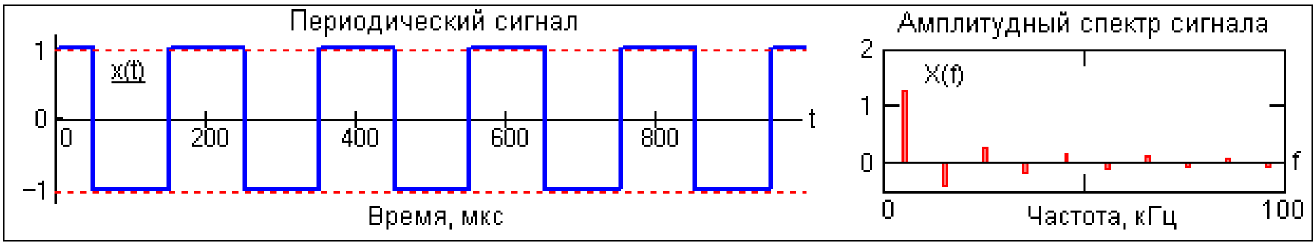


Рис. 1.1.8. Прямоугольный периодический сигнал (меандр).

Пример представления прямоугольного периодического сигнала (меандра) в виде амплитудного ряда Фурье в частотной области приведен на рис. 1.1.8. Сигнал четный относительно t=0, не имеет синусных гармоник, все значения $\square_k$ для данной модели сигнала равны нулю.

Информационными параметрами полигармонического сигнала могут быть как определенные особенности формы сигнала (размах от минимума до максимума, экстремальное отклонение от среднего значения, и т.п.), так и параметры определенных гармоник в этом сигнале. Так, например, для прямоугольных импульсов информационными параметрами могут быть период повторения импульсов, длительность импульсов, скважность импульсов (отношение периода к длительности). При анализе сложных периодических сигналов информационными параметрами могут также быть:

- Текущее среднее значение за определенное время, например, за время периода:

$(1/T)\int_{t}^{t+T} s(t)\,dt.$

- Постоянная составляющая одного периода:

$(1/T)\int_{0}^{T} s(t)\,dt.$

- Среднее выпрямленное значение:

$(1/T)\int_{0}^{T} |s(t)|\,dt.$

- Среднее квадратичное значение:

$\sqrt{\frac{1}{T}\int_{0}^{T} x(t)^2\,dt}\ .$

К непериодическим сигналам относят почти периодические и апериодические сигналы. Основным инструментом их анализа также является частотное представление.

Почти периодические сигналы близки по своей форме к полигармоническим. Они также представляют собой сумму двух и более гармонических сигналов (в пределе – до бесконечности), но не с кратными, а с произвольными частотами, отношения которых (хотя бы двух частот минимум) не относятся к рациональным числам, вследствие чего фундаментальный период суммарных колебаний бесконечно велик.

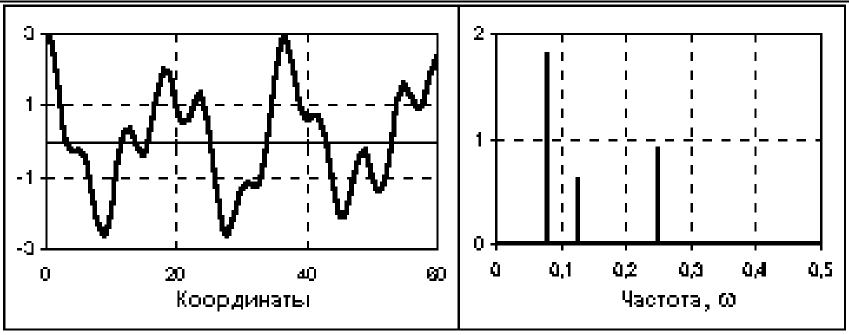


Рис. 1.1.9. Почти периодический сигнал и спектр его амплитуд.

Так, например, сумма двух гармоник с частотами 2f□ □и 3.5f□ дает периодический сигнал (2/3.5 – рациональное число) с фундаментальной частотой 0.5f□, на одном периоде которой будут укладываться 4 периода первой гармоники и 7 периодов второй. Но если значение частоты второй гармоники заменить значением $\sqrt{12}$ f□, то сигнал перейдет в разряд непериодических, поскольку отношение $2/\sqrt{12}$ не относится к числу рациональных чисел. Как правило, почти периодические сигналы порождаются физическими процессами, не связанными между собой. Математическое отображение сигналов тождественно полигармоническим сигналам (сумма гармоник), а частотный спектр также дискретен.

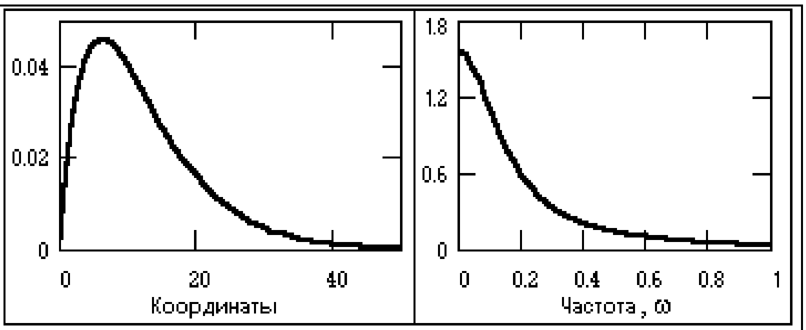


Рис. 1.1.10. Апериодический сигнал и модуль спектра.

Апериодические сигналы составляют основную группу непериодических сигналов и задаются произвольными функциями времени. На рис. 1.1.10 показан пример апериодического сигнала, заданного формулой на интервале (0, ∞):

$s(t) = \exp(-a\cdot t) - \exp(-b\cdot t),$

где a и b – константы, в данном случае a = 0.15, b = 0.17.

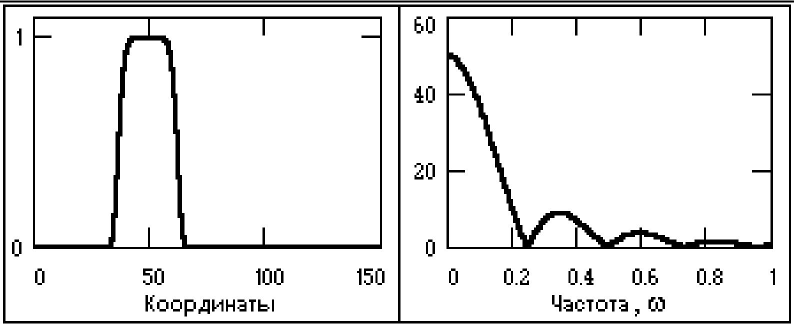


Рис. 1.1.11. Импульсный сигнал и модуль спектра.

К апериодическим сигналам относятся также импульсные сигналы, которые в радиотехнике и в отраслях, широко ее использующих, часто рассматривают в виде отдельного класса сигналов. Импульсы представляют собой сигналы определенной и достаточно простой формы, существующие в пределах конечных временных интервалов. Сигнал, приведенный на рис. 1.1.11, относится к числу импульсных.

Частотный спектр апериодических сигналов непрерывен и может содержать любые гармоники в частотном интервале [0, ∞]. Для его вычисления используется интегральное преобразование Фурье, которое можно получить переходом в формулах (1.1.3) от суммирования к интегрированию при $\Delta f \to 0$ и $k\Delta f \to f$.

$$s(t) = \int_0^\infty (a(f) \cos 2\pi ft + b(f) \sin 2\pi ft)\, df = \int_0^\infty S(f) \cos(2\pi ft-\varphi(f))\, df. \qquad (1.1.7)$$

$$a(f) = \int_0^T s(t) \cos 2\pi ft\, dt, \quad b(f) = \int_0^T s(t) \sin 2\pi ft\, dt, \qquad (1.1.8)$$

$$S(f) = \sqrt{a(f)^2 + b(f)^2}\,, \quad \varphi(f) = \text{argtg}\,(b(f)/a(f)). \qquad (1.1.9)$$

Частотные функции a(f), b(f) и S(f) представляют собой не амплитудные значения соответствующих гармоник на определенных частотах, а распределения спектральной плотности амплитуд этих гармоник по частотной шкале. Формулы (1.1.8-1.1.9) обычно называют формулами прямого преобразования Фурье, формулы (1.1.7) – обратного преобразования.

Если нас не интересует поведение сигнала за пределами области его задания [0, T], то эта область может восприниматься, как один период периодического сигнала, т.е. значение T принимается за фундаментальную частоту периодический колебаний, при этом для частотной модели сигнала может применяться разложение в ряды Фурье по области его задания (1.1.3-1.1.6).

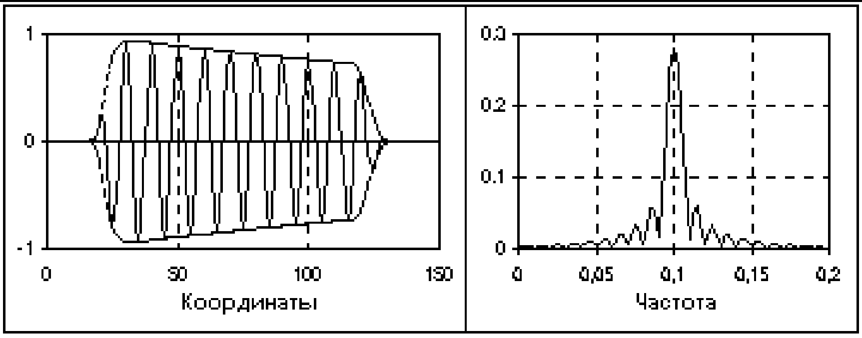


Рис. 1.1.12. Радиоимпульс и модуль его спектра.

В классе импульсных сигналов выделяют подкласс радиоимпульсов. Пример радиоимпульса приведен на рис. 1.1.12.

Уравнение радиоимпульса:

$s(t) = u(t) \cos(2\pi f_o t+\varphi_o)$.

где $\cos(2\pi f_o t+\varphi_o)$ – гармоническое колебание заполнения радиоимпульса, u(t) – огибающая радиоимпульса. Положение главного пика спектра радиоимпульса на частотной шкале соответствует частоте заполнения f_o, а его ширина определяется длительностью радиоимпульса. Чем больше длительность радиоимпульса, тем меньше ширина главного частотного пика.

С энергетических позиций сигналы разделяют на два типа: с ограниченной (конечной) энергией и с бесконечной энергией.

Для множества сигналов с ограниченной энергией должно выполняться условие:

$$L_2 = \{s;\ \int_{-\infty}^{\infty} |s(t)|^2\, dt < \infty\}.$$

О сигналах s(t) данного множества принято говорить, что они *интегрируемы с квадратом.* Очевидно, что этому множеству могут соответствовать только сигналы, стремящиеся к нулю на бесконечности: $\lim\limits_{|t| \Rightarrow \infty} s(t) \to 0$.

Как правило, к этому типу сигналов относятся апериодические и импульсные сигналы, не имеющие разрывов 2-го рода при ограниченном количестве разрывов 1-го рода. Любые периодические, полигармонические и почти периодические сигналы, а также сигналы с разрывами и особыми точками 2-го рода, уходящими в бесконечность, относятся к сигналам с бесконечной энергией. Для их анализа применяются специальные методы.

Для бесконечных по энергии сигналов, в том числе для периодических, ограничение по энергии может задаваться для определенного интервала (периода) T = t1-t2:

$$L_2(T) = \{s; \int_{t1}^{t2} |s(t)|^2\, dt < \infty\}.$$

Иногда в отдельный класс выделяют сигналы конечной длительности, отличные от нуля только на ограниченном интервале аргументов (независимых переменных). Такие сигналы называют *финитными.*

С позиций временной динамики сигналы подразделяются на стационарные и нестационарные. Стационарными называются сигналы, частотный спектр которых не изменяется во времени и не зависит от интервала задания сигналов. К ним относятся периодические и почти периодические сигналы. Большинство практических сигналов являются нестационарными на достаточно больших интервалах задания, но могут содержать в своем составе стационарные частотные составляющие. Так, модулированные сигналы радио и телевидения относятся к числу нестационарных, но имеют стационарные несущие частоты.

Выделяют следующие типы сигналов, которым соответствуют определенные формы их математического описания.

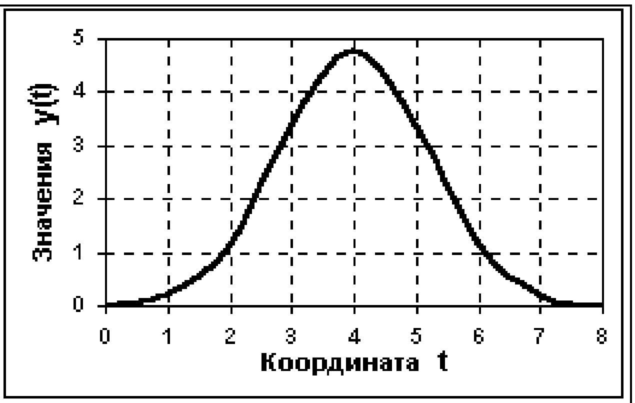


Рис. 1.2.1. Аналоговый сигнал.

<u>Аналоговый сигнал</u> (analog signal) является непрерывной или кусочно-непрерывной функцией y=x(t) непрерывного аргумента, т.е. как сама функция, так и ее аргумент могут принимать любые значения в пределах некоторого интервала $y_1 \le y \le y_2$, $t_1 \le t \le t_2$. Если интервалы значений сигнала или его независимых переменных не ограничиваются, то по умолчанию они принимаются равными от $-\infty$ до $+\infty$. Множество возможных значений сигнала образует континуум - непрерывное пространство, в котором любая сигнальная точка может быть определена с точностью до бесконечности.

Источниками аналоговых сигналов, как правило, являются физические процессы и явления, непрерывные в динамике своего развития во времени, в пространстве или по любой другой независимой переменной, при этом регистрируемый сигнал подобен ("аналогичен") порождающему его процессу. Пример графического отображения сигнала приведен на рис. 1.2.1. Примеры сигналов, аналоговых по своей природе - изменение напряженности электрического, магнитного, электромагнитного поля во времени и в пространстве.

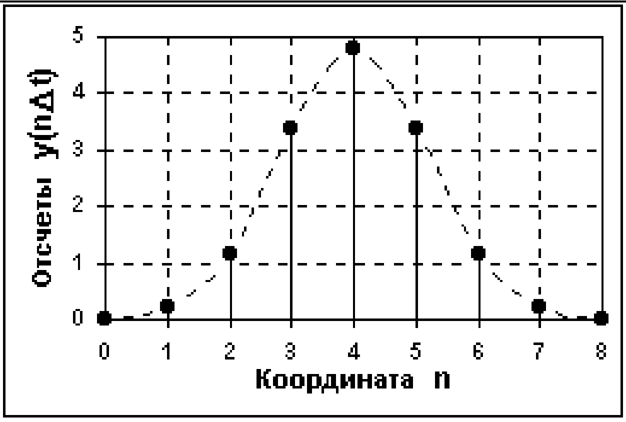


Рис. 1.2.2. Дискретный сигнал

<u>Дискретный сигнал</u> (discrete signal) по своим значениям также является непрерывной функцией, но определенной только по дискретным значениям аргумента. По множеству своих значений он является конечным (счетным) и описывается дискретной последовательностью отсчетов (samples) y(n□t), где $y_1 \leq$□$y \leq y_2$, □t - интервал между отсчетами (интервал или шаг дискретизации, sample time), n = 0, 1, 2,...,N. Величина, обратная шагу дискретизации: f = 1/□t, называется частотой дискретизации (sampling frequency). Если дискретный сигнал получен дискретизацией (sampling) аналогового сигнала, то он представляет собой последовательность отсчетов, значения которых в точности равны значениям исходного сигнала по координатам n□t.

Пример дискретизации аналогового сигнала (рис. 1.2.1) представлен на рис. 1.2.2. При □t = const (равномерная дискретизация данных) дискретный сигнал можно описывать сокращенным обозначением y(n). В технической литературе в обозначениях дискретизированных функций иногда оставляют прежние индексы аргументов аналоговых функций, заключая их в квадратные скобки - y[t]. При неравномерной дискретизации сигнала обозначения дискретных последовательностей обычно заключаются в фигурные скобки - $\{s(t_i)\}$, а значения отсчетов приводятся в виде таблиц с указанием значений координат t_i. Для числовых последовательностей (равномерных и неравномерных) применяется и следующее числовое описание: $s(t_i) = \{a_1, a_2, ..., a_N\}$, $t = t_1, t_2, ...,t_N$. Примеры дискретных геофизических сигналов - результаты вертикального электрического зондирования (дискретная величина разноса токовых электродов), профили геохимического опробования, и т.п.

<u>Цифровой сигнал</u> (digital signal) квантован по своим значениям и дискретен по аргументу. Он описывается квантованной решетчатой функцией $y_n = Q_k$[y(n□t)], где Q_k - функция квантования с числом уровней квантования k, при этом интервалы квантования могут быть как с равномерным распределением, так и с неравномерным, например - логарифмическим. Задается цифровой сигнал, как правило, в виде дискретного ряда (discrete series) числовых данных - числового массива по последовательным значениям аргумента при □t = const, но в общем случае сигнал может задаваться и в виде таблицы для произвольных значений аргумента.

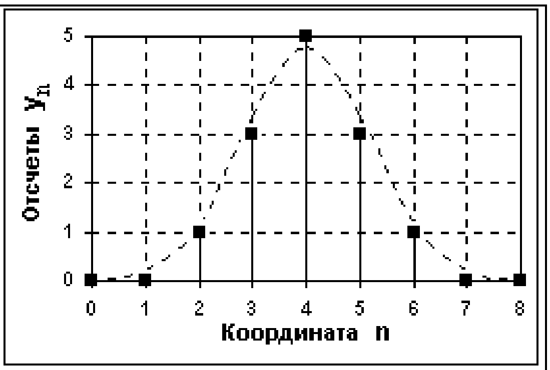


Рис. 1.2.3. Цифровой сигнал

По существу, цифровой сигнал по своим значениям (отсчетам) является формализованной разновидностью дискретного сигнала при округлении отсчетов последнего до определенного количества цифр, как это показано на рис 1.2.3. Цифровой сигнал конечен по множеству своих значений. Процесс преобразования бесконечных по значениям аналоговых отсчетов в конечное число цифровых значений называется квантованием по уровню, а возникающие при квантовании ошибки округления отсчетов (отбрасываемые значения) – шумами (noise) или ошибками (error) квантования (quantization).

В системах цифровой обработки данных и в ЭВМ сигнал всегда представлен с точностью до определенного количества разрядов, а, следовательно, всегда является цифровым. С учетом этих факторов при описании цифровых сигналов функция квантования обычно опускается (подразумевается равномерной по умолчанию), а для описания сигналов используются правила описания дискретных сигналов. Что касается формы обращения цифровых сигналов в системах хранения, передачи и обработки, то, как правило, они представляет собой комбинации коротких одно- или двуполярных импульсов одинаковой амплитуды, которыми в двоичном коде с определенным количеством числовых разрядов кодируются числовые последовательности сигналов (массивов данных).

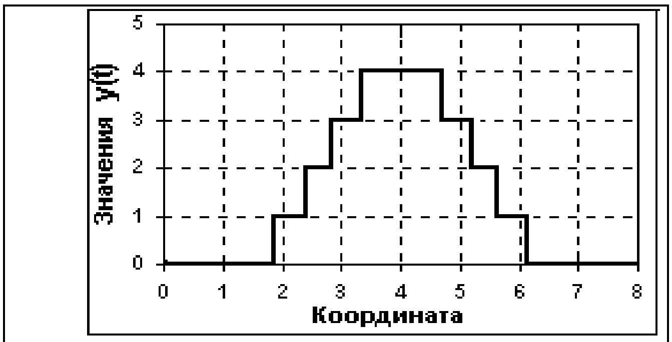


Рис. 1.2.4. Дискретно-аналоговый сигнал

В принципе, квантованными по своим значениям могут быть и аналоговые сигналы, зарегистрированные соответствующей аппаратурой (рис. 1.2.4), которые принято называть дискретно-аналоговыми. Но выделять эти сигналы в отдельный тип не имеет смысла - они остаются аналоговыми кусочно-непрерывными сигналами с шагом квантования, который определяется допустимой погрешностью измерений.

Большинство сигналов, с которыми приходится иметь дело при обработке геофизических данных, являются аналоговыми по своей природе, дискретизированными и квантованными в силу методических особенностей измерений или технических особенностей регистрации, т.е. преобразованными в цифровые сигналы. Но существуют и сигналы, которые изначально относятся к классу цифровых, как, например отсчеты количества гамма-квантов, зарегистрированных по последовательным интервалам времени.

Сигнал, значения которого отличны от нуля только на конечном интервале T, называют *финитным*. Если спектральная функция X(f) сигналов (преобразование Фурье) обращается в нуль вне некоторого конечного интервала частот, то они называются *сигналами с финитным спектром*. Если сигнал X(t) определен только для значений аргумента $t \geq 0$, то он считается *каузальным* (причинным).

Преобразования типа сигналов. Формы математического отображения сигналов, особенно на этапах их первичной регистрации (детектирования) и в прямых задачах описания геофизических полей и физических процессов, как правило, отражают их физическую природу. Однако последнее не является обязательным и зависит от методики измерений и технических средств детектирования, преобразования, передачи, хранения и обработки сигналов. На разных этапах процессов получения и обработки информации как материальное представление сигналов в устройствах регистрации и обработки, так и формы их математического описания при анализе данных, могут изменяться путем соответствующих операций преобразования типа сигналов.

Операция дискретизации (discretization) осуществляет преобразование аналоговых сигналов (функций), непрерывных по аргументу, в функции мгновенных значений сигналов по дискретному аргументу. Дискретизация обычно производится с постоянным шагом по аргументу (*равномерная дискретизация*), при этом $s(t) \Rightarrow s(n\Delta t)$, где значения $s(n\Delta t)$ представляют собой отсчеты функции s(t) в моменты времени $t = n\Delta t$, n = 0, 1, 2,..., N. Частота, с которой выполняются замеры аналогового сигнала, называется *частотой дискретизации*. В общем случае, сетка отсчетов по аргументу может быть произвольной, как, например, $s(t) \Rightarrow s(t_k)$, k=1, 2, …, K, или задаваться по определенному закону. В результате дискретизации непрерывный (*аналоговый*) сигнал переводится в последовательность чисел.

Операция восстановления аналогового сигнала из его дискретного представления обратна операции дискретизации и представляет, по существу, интерполяцию данных.

Дискретизация сигналов может приводить к определенной потере информации о поведении сигналов в промежутках между отсчетами. Однако существуют условия, определенные теоремой Котельникова-Шеннона, согласно которым аналоговый сигнал с ограниченным частотным спектром может быть без потерь информации преобразован в дискретный сигнал, и затем абсолютно точно восстановлен по значениям своих дискретных отсчетов.

Любая непрерывная функция на конечном отрезке может быть разложена в ряд Фурье, т.е. представлена в спектральной форме - в виде суммы ряда синусоид с кратными (нумерованными) частотами с определенными амплитудами и фазами. У относительно гладких функций спектр быстро убывает (коэффициенты модуля спектра быстро стремятся к нулю). Для представления "изрезанных" функций, с разрывами и "изломами", нужны синусоиды с большими частотами. Говорят, что сигнал имеет *ограниченный спектр*, если после определенной частоты F все коэффициенты спектра равны нулю, т.е. сигнал представляется в виде конечной суммы ряда Фурье.

Теоремой Котельникова-Шеннона устанавливается, что если спектр сигнала ограничен максимальной частотой f, то после дискретизации сигнала с частотой не менее 2f можно восстановить исходный непрерывный сигнал по полученному цифровому сигналу абсолютно точно. Для этого нужно выполнить интерполяцию цифрового сигнала "между отсчетами" специальной функцией (Котельникова-Шеннона).

Физический смысл теоремы Котельникова-Шеннона достаточно прост. Если максимальная частота в сигнале равна f, то достаточно на одном периоде этой гармоники иметь минимум 2 отсчета с известными значениями t_1 и t_2, как появляется возможность записать систему из двух уравнений ($y_1 = a \cos 2\pi f t_1$ и $y_2 = a \cos 2\pi f t_2$) и решить систему относительно 2-х неизвестных – амплитуды a и частоты f этой гармоники. Следовательно, частота дискретизации должна быть в 2 раза больше максимальной частоты f в сигнале. Для более низких частот это условие будет выполнено автоматически.

На практике эта теорема имеет огромное значение. Например, известно, что диапазон звуковых сигналов, воспринимаемых человеком, не превышает 20 кГц. Следовательно, при дискретизации записанных звуковых сигналов с частотой не менее 40 кГц мы можем точно восстановить исходный аналоговый сигнал по его цифровым отсчетам, что и выполняется в проигрывателях компакт-дисков для восстановления звука. Частота дискретизации звукового сигнала при записи на компакт-диск составляет 44100 Гц.

Операция квантования или аналого-цифрового преобразования (АЦП; английский термин Analog-to-Digital Converter, ADC) заключается в преобразовании дискретного сигнала $s(t_n)$ в цифровой сигнал $s(n) = s_n \approx s(t_n)$, $n = 0, 1, 2,.., N$, как правило, кодированный в двоичной системе счисления. Процесс преобразования отсчетов сигнала в числа называется квантованием по уровню (quantization), а возникающие при этом потери информации за счет округления – ошибками или шумами квантования (quantization error, quantization noise).

При преобразовании аналогового сигнала непосредственно в цифровой сигнал операции дискретизации и квантования совмещаются.

Операция цифро-аналогового преобразования (ЦАП; Digital-to-Analog Converter, DAC) обратна операции квантования, при этом на выходе регистрируется либо дискретно-аналоговый сигнал $s(t_n)$, который имеет ступенчатую форму (рис. 1.2.4), либо непосредственно аналоговый сигнал s(t), который восстанавливается из $s(t_n)$, например, путем сглаживания.

Так как квантование сигналов всегда выполняется с определенной и неустранимой погрешностью (максимум - до половины интервала квантования), то операции АЦП и ЦАП не являются взаимно обратными с абсолютной точностью.

Алиасинг. А что произойдет, если спектр аналогового сигнала был неограниченным или имел частоту, выше частоты дискретизации?

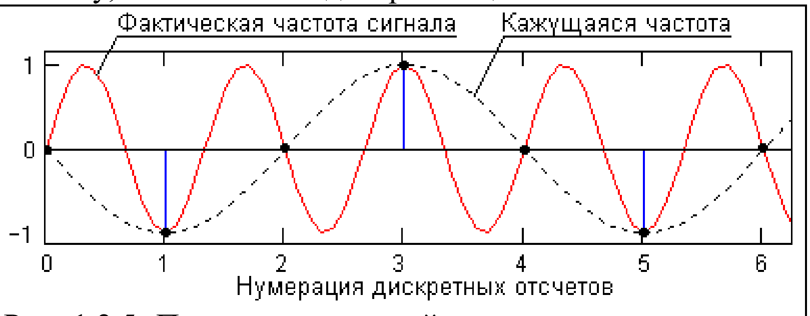


Рис. 1.2.5. Появление кажущейся частоты при дискретизации.

Предположим, что при записи акустического сигнала оркестра в помещении от какого-то устройства присутствует ультразвуковой сигнал с частотой 30 кГц. Запись выполняется с дискретизацией сигнала на выходе микрофона с типовой частотой 44.1 кГц. При прослушивании такой записи с использованием ЦАП мы услышим шумовой сигнал на частоте 30 – 44.1/2 ≈ 8 кГц. Восстановленный сигнал будет выглядеть так, как если бы частоты, лежащие выше половины частоты дискретизации, "зеркально" от нее отразились в нижнюю часть спектра и сложились с присутствующими там гармониками. Это так называемый эффект *появления ложных (кажущихся) частот* (aliasing). Эффект аналогичен известному эффекту обратного вращения колес автомобиля на экранах кино и телевизоров, когда скорость их вращения начинает превышать частоту смены кадров. Природу эффекта можно наглядно видеть на рис. 1.2.5. Аналогично в главный частотный диапазон дискретных сигналов "отражаются" от частоты дискретизации и все высокочастотные шумы, присутствующие в исходном аналоговом сигнале.

Для предотвращения алиасинга следует повышать частоту дискретизации или ограничить спектр сигнала перед оцифровкой *фильтрами низких частот* (*НЧ-фильтры*, *low-pass filters*), которые пропускают без изменения все частоты, ниже заданной, и подавляют в сигнале частоты, выше заданной. Эта граничная частота называется *частотой среза* (*cutoff frequency*) фильтра. Частота среза анти-алиасинговых фильтров устанавливается равной половине частоты дискретизации. В реальные АЦП почти всегда встраивается анти-алиасинговый фильтр.

Графическое отображение сигналов общеизвестно и особых пояснений не требует. Для одномерных сигналов график – это совокупность пар значений {t, s(t)} в прямоугольной системе координат (рис. 1.2.1 – 1.2.4). При графическом отображении дискретных и цифровых сигналов используется либо способ непосредственных дискретных отрезков соответствующей масштабной длины над осью аргумента, либо способ огибающей (плавной или ломанной) по значениям отсчетов. В силу непрерывности геофизических полей и, как правило, вторичности цифровых данных, получаемых дискретизацией и квантованием аналоговых сигналов, второй способ графического отображения будем считать основным.

Тестовые сигналы (test signal). В качестве тестовых сигналов, которые применяются при моделировании и исследовании систем обработки данных, обычно используются сигналы простейшего типа: гармонические синус-косинусные функции, дельта-функция и функция единичного скачка.

Дельта-функция или функция Дирака. По определению, дельта-функция описывается следующими математическими выражениями (в совокупности):

□(t-□) = 0 при t □ □,

$\int_{-\infty}^{\infty}$ □(t-□) dt = 1.

Функция □(t-□) не является дифференцируемой, и имеет размерность, обратную размерности ее аргумента, что следует из безразмерности результата интегрирования. Значение дельта-функции равно нулю везде за исключением точки □, где она представляет собой бесконечно узкий импульс с бесконечно большой амплитудой, при этом площадь импульса равна 1.

Дельта-функция является полезной математической абстракцией. На практике такие функции не могут быть реализованы с абсолютной точностью, так как невозможно реализовать значение, равное бесконечности, в точке t = □ на аналоговой временной шкале. Но во всех случаях, когда площадь импульса равна 1, длительность импульса достаточно мала, а за время его действия на входе системы сигнал на ее выходе практически не изменяется (реакция системы на импульс во много раз больше длительности самого импульса), входной сигнал можно считать *единичной импульсной функцией* со свойствами дельта - функции.

При своей абстрактности дельта - функция имеет вполне определенный физический смысл. Представим себе импульсный сигнал прямоугольной формы $\Pi(t-\tau)$ длительностью θ, амплитуда которого равна $1/\theta$, а площадь соответственно равна 1. При уменьшении значения длительности θ импульс, сокращаясь по длительности, сохраняет свою площадь, равную 1, и возрастает по амплитуде. Предел такой операции при $\theta \Rightarrow 0$ и носит название дельта - импульса. Этот сигнал □(t-□) сосредоточен в одной координатной точке t = □, конкретное амплитудное значение сигнала не определено, но площадь (интеграл) остается равной 1. Это не мгновенное значение функции в точке t = □, а именно импульс (импульс силы в механике, импульс тока в электротехнике и т.п.) – математическая модель короткого действия, значение которого равно 1.

Дельта-функция обладает *фильтрующим свойством*. Суть его заключается в том, что если дельта-функция □(t-□) входит под интеграл какой-либо функции в качестве множителя, то результат интегрирования равен значению подынтегральной функции в точке □ расположения дельта-импульса, т.е.:

$$\int_{-\infty}^{\infty} s(t)\ □(t-□)\ dt = s(□).$$

Интегрирование в выражении может ограничиваться ближними окрестностями точки □.

<u>Функция единичного скачка</u> или функция Хевисайда иногда называется также функцией включения. Полное математическое выражение функции:

$$\sigma(t) = \begin{cases} 0 & \text{если} \quad t < 0, \\ 1/2 & \text{если} \quad t = 0, \\ 1 & \text{если} \quad t > 0. \end{cases}$$

При моделировании сигналов и систем значение функции скачка в точке t=0 очень часто принимают равным 1, если это не имеет принципиального значения.

Функция единичного скачка используется при создании математических моделей сигналов конечной длительности. При умножении любой произвольной функции, в том числе периодической, на прямоугольный импульс, сформированный из двух последовательных функций единичного скачка

$s(t) = \sigma(t) - \sigma(t\text{-}T)$,

из нее вырезается участок на интервале 0-Т, и обнуляются значения функции за пределами этого интервала.

<u>Функция Кронекера.</u> Для дискретных и цифровых систем разрешающая способность по аргументу сигнала определяется интервалом его дискретизации □t. Это позволяет в качестве единичного импульса использовать дискретный интегральный аналог дельта-функции - функцию единичного отсчета □(k□t-n□t), которая равна 1 в координатной точке k = n, и нулю во всех остальных точках. Функция □(k□t-n□t) может быть определена для любых значений □t = const, но только для целых значений координат k и n, поскольку других номеров отсчетов в дискретных функциях не существует.

Математические выражения □(t-□) и □(k□t-n□t) называют также импульсами Дирака и Кронекера. Однако, применяя такую терминологию, не будем забывать, что это не просто единичные импульсы в координатных точках □ и n□t, а полномасштабные импульсные функции, определяющие как значения импульсов в определенных координатных точках, так и нулевые значения по всем остальным координатам, в пределе от $-\infty$ до ∞.

Количественная оценка информации

Для того чтобы оценить и измерять количество информации в соответствии с вышеизложенными аспектами, применяются различные подходы и методы. Среди них выделяются статистический, семантический, прагматический и структурный. Исторически наибольшее развитие получил статистический подход.

Статистический подход. Он изучается в обширном разделе кибернетики, называемой теорией информации. Основоположником этого подхода считается К.Шеннон, опубликовавший в 1948 г. Свою математическую теорию связи. Большой вклад в теорию информации до него внесли ученые Найквист и Хартли, которые соответственно в 1924 и 1928 гг. напечатали работы по теории телеграфии и передачи информации. Признаны во всем мире исследования по теории информации российских ученых А.Н.Колмогорова, А.Я.Хинчина, В.А. Котельникова, А.А.Харкевича и т.д.

К.Шенноном было введено понятие количества информации как меры неопределенности состояния системы, снимаемой при получении информации. Количественно выраженная неопределенность состояния получила название энтропии по аналогии с подобным понятием в статистической механике. При получении информации уменьшается неопределенность, т.е. энтропия системы. Очевидно, что, чем больше информации получает наблюдатель, тем больше снижается неопределенность, и энтропия системы уменьшается. При энтропии, равной нулю, о системе имеется полная информация, и наблюдателю она представляется целиком упорядоченной. Таким образом, получение информации связано с изменением степени неосведомленности получателя о состоянии этой системы.

До получения информации её получатель мог иметь некоторые предварительные (априорные) сведения о системе *X*. Оставшаяся неосведомленность и является для него мерой неопределенности состояния (энтропии) системы. Обозначим априорную энтропию системы *X* через *H(X*). после получения некоторого сообщения наблюдатель приобрел дополнительную информацию *I(X),*уменьшившую его начальную неопределенность так, что апостериорная (после получения информации) неопределенность состояния системы стала *H'(X).* Тогда количество информации I может быть определено как

I(X)=H(X)-H'(X)

Другими словами, количество информации измеряется уменьшением (изменением) неопределенности состояния системы.

Если апостериорная энтропия системы обратилась в нуль, то первоначально неполное значение заменится полным значением и количество информации, полученной в этом случае наблюдателем, будет

I(X)=H(X)

т.е. энтропия системы может рассматриваться как мера недостающей информации.

Если система X обладает дискретным состояниями (т.е. переходит из состояния в состояния скачком), их количество равно N, а вероятность нахождения системы в каждом из состояний – $p_1, p_2, p_3, ..., p_N$ (причем $\sum_{i=1}^{N} p_i = 1$ и $p_i \le 1$), то согласно теореме Шеннона энтропия системы равна

$$H(X) = -K_0 \sum_{i=1}^{N} p_i \log_a p_i$$

Здесь коэффициент K_0 и основание логарифма а определяют систему единиц измерения количества информации. Логарифмическая мера информации была предложена Хартли для представления технических параметров систем связи как более удобная и более близкая к восприятию человеком, привыкшим к линейным сравнениям с принятыми эталонами. Например, каждый чувствует, что две однотипные дискеты должны обладать вдвое большей емкостью, чем одна, а два идентичных канала связи должны иметь удвоенную пропускную способность.

Знак минус поставлен для того, чтобы значение энтропии было положительным, так как $p_i \le 1$ и логарифм в этом случае отрицательный.

Если все состояния системы равновероятны, т.е. $p_i = \frac{1}{N}$, её энтропия

Энтропия H обладает рядом интересных свойств. Вот некоторые из них.

Энтропия H равна нулю только тогда, когда все вероятности p_i , кроме одной, равны нулю, а эта единственная вероятность равна единице. Таким образом, *H=0* только в случае полной определенности состояния системы.

При заданном числе состояний системы N величина H максимальна и равна $K_0 \log_a N$, когда все p_i равны.

Определим единицы измерения количества информации с помощью выражения для энтропии системы с равновероятными состояниями.

Пусть система имеет два равновероятных состояния, т.е. N=2. Будем считать, что снятие неопределенности о состоянии такой системы дает одну единицу информации, так как при полном снятии неопределенности энтропия количественно равна информации *H=I.* Тогда

$$1 = K_0 \log_a 2$$

Очевидно, что правая часть равенства будет тождественно равна единице информации, если принять $K_0 = 1$ и основание логарифма a=2. В общем случае при N равновероятных состояний количество информации будет

$$I = \log_2 N$$

Эта формула получила название формулы Хартли и показывает, что количество информации, необходимое для снятия неопределенности о системе с равновероятными состояниями, зависит лишь от количества этих состояний.

Информация о состояниях системы передается получателю в виде сообщений, которые могут быть представлены в различной синтаксической форме, например в виде кодовых комбинаций, использующих m различных символов и n разрядов, в каждой из которых может находиться любой из символов. Если код не избыточен, то каждая кодовая комбинация отображает одно из состояний системы. Количество кодовых комбинаций будет

$$N = m^n$$

Подставив это выражение в формулу для I.

$$I = n \log_2 m$$

Если код двоичный, т.е. используются лишь два символа (0 или 1), то m=2 и I=n. В этом случае количество информации в сообщении составит n двоичных единиц, называемых битами (binary digit (bit) – двоичная цифра).

При использовании в качестве основания логарифма числа десять единиц измерения информации могут быть десятичными, или дитами. Так как $\log_2 N = \log_{10} \dfrac{N}{\log_{10} 2} = 3{,}321 \log_{10} N$, то десятичная единица составляет примерно 3,33 бита.

Иногда удобно применять натуральное основание логарифма ℓ . В этом случае получающие единицы информации называются натуральными или *натами*. Переход от основания a к основанию b требует лишь умножения на $\log_b a$.

Введенная количественная статистическая мера информации широко используется в теории информации для оценки собственной, взаимной, условной и других видов информации. Рассмотрим в качестве примера собственную информацию. Под *собственной информацией* будем понимать информацию, содержащуюся в данном конкретном сообщении. А конкретное сообщение, как указывалось, дает получателю информацию о возможном существовании конкретного состояния системы. Тогда количество собственной информации, содержащееся в сообщении X_i , определяется как

$$I(X_i) = -\log_2 p(X)$$

Собственная информация имеет следующие свойства:

- собственная информация неотрицательна
- чем меньше вероятность возникновения сообщения, тем больше информации оно содержит. Именно поэтому неожиданные сообщения так воздействуют на психику человека, что содержащаяся в них большое количество информации создает информационный психологический удар, иногда приводящий к трагическим последствиям.
- Если сообщение имеет вероятность возникновения, равную единице, то информация, содержащаяся в нем, равна нулю, так как заранее известно, что может прийти только это сообщение, а значит, ничего нового потребитель информации не получает.
- Собственная информация обладает свойством аддитивности, т.е. количество собственной информации нескольких независимых сообщений равно их сумме. Например, для собственной информации двух сообщений X_i и Y_i может быть записано:

$$I(X_i, Y_i) = -\log_2 P(X_i) - \log_2 P(Y_i) = I(X_i) + I(Y_i).$$

Следует еще раз отметить, что статистический подход к количественной оценке информации был рассмотрен для дискретных систем, случайным образом переходящих из состояния в состояние, и, следовательно, сообщение об этих состояниях также возникает случайным образом.

Кроме того, статистический метод определения количества информации практически не учитывает семантического и прагматического аспектов информации.

Семантический подход. Этот подход является наиболее трудно формализуемым и до сих пор окончательно неопределившимся.

Наибольшее признание для измерения смыслового содержания информации получила тезаурусная мера, предложенная Ю.И. Шнейдером. Идеи тезаурусного метода были сформулированы ещё основоположником кибернетики Н. Винером. Для понимания и использования информации её получатель должен обладать определенным запасом знаний.

Если индивидуальный тезаурус потребителя $S_П$ отражает его знания о данном предмете, то количество смысловой информации I_C, содержащаяся в некотором сообщении, можно оценить степенью изменения этого тезауруса, произошедшего под воздействием данного сообщения. Очевидно, что количество информации I_C нелинейно зависит от состояния индивидуального тезауруса пользователь, и хотя смысловое содержание сообщения S постоянно, пользователи, имеющие отличающиеся тезаурусы, будут получать *неодинаковое* количество информации.

В самом деле, если индивидуальный тезаурус получателя информации близок к нулю, $S_П \approx 0$, то в этом случае и количество воспринятой информации равно нулю: $I_C = 0$.

Иными словами, получатель не понимает принятого сообщения, и, как следствие, для него количество воспринятой информации равно нулю. Такая ситуация эквивалентна прослушиванию сообщения на неизвестном иностранном языке. Несомненно, сообщение не лишено смысла, однако оно непонятно, а значит, не имеет информативности.

Количество семантической информации I_C в сообщении также будет равно нулю, если пользователь информации абсолютно все знает о предмете, т.е. его тезаурус $S_П$, и сообщение не дает ему ничего нового.

Интуитивно мы чувствуем, что между этими полярными значениями тезауруса пользователя существует некоторое оптимальное значение , $S_{П.ОПТ}$, при котором количество информации I_C, извлекаемое из сообщения, становиться для получателя максимальным.

Тезаурусный метод подтверждает тезис о том, что информация обладает свойством относительности и имеет, таким образом, относительную, субъективную ценность. Для того чтобы объективно оценить научную информации, появилось понятие общечеловеческого тезауруса, степень изменения которого и определяет значительность получаемых человечеством новых знаний.

Прагматический подход. Он определяет количество информации как меру, способствующую достижению поставленной цели. Одной из первых работ, реализующих этот подход, явилась статья А.А. Харкевича. В ней он предлагал принять за меру ценности информации количество информации, необходимое для достижения поставленной цели. Этот подход базируется на статической теории Шеннона и рассматривает количество информации как приращение вероятности достижения цели. Так, если принять вероятность достижение цели до получения информации равной p_0, а после её получения - p_1, то прагматическое количество информации $I_П$ определяется как

$$I_П = \log \frac{p_1}{p_0}.$$

Если основание логарифма сделать равным двум, то $I_П$ будет измеряться в битах, как и при статистическом подходе.

При оценке количества информации в семантическом и прагматическом аспектах необходимо учитывать и временную зависимость информации. Дело в том, что информация, особенно в системах управления экономическими объектами, имеет свойство стареть, т.е. её ценность со временем падает, и важно использовать её в момент наибольшей ценности.

Структурный подход. Он связан с проблемами хранения, реорганизации и извлечения информации и по мере увеличения объемов накапливаемой в компьютерах информации приобретает все большее значение.

При структурном подходе абстрагируются от субъективности, относительно ценности информации и рассматривают логические и физические структуры организации информации. С изобретением компьютеров появилось возможность хранить на машинных носителях громадные объемы информации. Но для её эффективного использования необходимо определить такие структуры организации информации, чтобы существовала возможность быстрого поиска, извлечения, записи, модификации информационной базы.

При машинном хранении структурной единицей информации является один байт, содержащий восемь бит (двоичных единиц информации). Менее определенной, но также переводимой в байты является неделимая единица экономической информации – реквизит.

Реквизиты объединяют в показатели, показатели – в записи, записи – в массивы, из массивов создают комплексы массивов, а из комплексов – информационные базы. Структурная теория позволяет на логическом уровне построить оптимальную структуру информационной базы, которая затем с помощью определенных средств реализуется на физическом уровне – уровне технических устройств хранения информации. От выбранной структуры хранения зависит такой важный параметр, как время доступа к данным, т.е. структура влияет на время записи и считывания информации, а значит, и на время создания и реорганизации информационной базы.

Информационная база совместно с системой управления базой данных (СУБД) формирует автоматизированный банк данных.

Значение структурной теории информации растет при переходе от банков данных к банкам знаний, в которых информация подвергается ещё более высокой степени структуризации.

После преобразования информации в машинную форму (рис.2.) её аналитический и прагматический аспекты как бы уходят в тень, и дальнейшая обработка информации происходит по «машинным законам», одинаковым для информации любого смыслового содержания. Информация в машинном виде, т.е. в форме электрических, магнитных и тому подобных сигналов и состояний, носит название данных. Для того, чтобы понять их смысловое содержание, необходимо данные снова преобразовать в информацию.

Преобразование “информация – данные ” производятся в устройствах ввода-вывода ЭВМ.

Базисным понятием всей теории информации является понятие энтропии. Энтропия – мера неопределенности некоторой ситуации.

Энтропия как мера неопределенности

Подойдем к описанию случайных событий несколько с иной стороны. То, что событие случайно, означает отсутствие полной уверенности в его наступлении, что, в свою очередь, создает ***неопределенность*** в исходах опытов, связанных с данным событием. Безусловно, <u>степень неопределенности</u> различна для разных ситуаций. Например, если опыт состоит в определении возраста случайно выбранного студента 1-го курса дневного отделения вуза, то с большой долей уверенности можно утверждать, что он окажется менее 30 лет; хотя по положению на дневном отделении могут обучаться лица в возрасте до 35 лет, чаще всего очно учатся выпускники школ ближайших нескольких выпусков. Гораздо меньшую определенность имеет аналогичный опыт, если проверяется, будет ли возраст произвольно выбранного студента меньше 20 лет. Для практики важно иметь возможность произвести <u>численную оценку</u> неопределенности разных опытов. Попробуем ввести такую количественную меру неопределенности.

Начнем с простой ситуации, когда опыт имеет n равновероятных исходов. Очевидно, что неопределенность каждого из них зависит от n, т.е.

неопределенность = f(n)

Можно указать некоторые свойства этой функции:

(1) f(1)=0, поскольку при n=1 исход опыта не является случайным и, следовательно, неопределенность отсутствует;

(2) f(n) возрастает с ростом n, т.к. ввиду большого числа возможных исходов предсказание результата опыта становится весьма затруднительным.

Для определения явного вида функции f(n) рассмотрим два независимых опыта A и B, с количествами равновероятных исходов, соответственно n_A и n_B. Рассмотрим сложный опыт C, который состоит в одновременном выполнении опытов A и B. Число возможных исходов опыта C равно $n_A \cdot n_B$, причем, все они равновероятны. Очевидно, неопределенность исхода такого опыта будет больше неопределенности опыта A, поскольку к ней добавляется неопределенность B. Естественно допустить, что мера неопределенности C равна сумме неопределенностей опытов A и B, т.е. неопределенность аддитивна:

$$f(n_A n_B) = f(n_A) + f(n_B) \qquad (1.1)$$

Теперь можно задуматься о том, каким может быть явный вид функции f(n), чтобы он удовлетворял свойствам (1) и (2). Легко увидеть, что такому набору свойств удовлетворяет функция log(n), причем, можно показать, что она единственная из всех возможных классов функций. Таким образом:

за меру неопределенности опыта с n равновероятными исходами можно принять число log(n).

Следует заметить, что выбор основания логарифма в данном случае значения не имеет, поскольку в силу известной формулы перехода от одного основания логарифма к другому

$$\log_b n = \log_b a \cdot \log_a n \quad ,$$

переход к другому основанию состоит во введении одинакового для обеих частей выражения постоянного множителя $\log_b a$, что равносильно изменению масштаба (т.е. размера единицы) измерения неопределенности. Поскольку это так, мы имеет возможность выбрать удобное для нас (из каких-то дополнительных соображений) основание логарифма. Таким удобным основанием оказывается 2, поскольку в этом случае за единицу измерения принимается неопределенность, содержащаяся в опыте, имеющем лишь два равновероятных исхода, которые можно обозначить, например, ИСТИНА (True) и ЛОЖЬ (False) и использовать для анализа таких событий аппарат математической логики.

*(а) Единица измерения неопределенности при двух возможных исходах опыта называется **бит**.*

(Название **бит** происходит от английского **binary digit**, что в дословном переводе означает «двоичный разряд» или «двоичная единица».)

Таким образом, нами установлен явный вид функции, описывающей неопределенность опыта, имеющего n равновероятных исхода:

$$f(n) = \log_2 n \qquad (1.2)$$

На основании формул ($p = \frac{1}{n}$) несложно найти неопределенность, вносимую каждым отдельным исходом в общую. Поскольку исходов n и все они равновероятны (и, следовательно, равнозначны), а общая неопределенность равна $\log_2 n$, из свойства аддитивности неопределенности следует, что неопределенность, вносимая одним исходом составляет

$$\frac{1}{n}\log_2 n = -\frac{1}{n}\log_2\frac{1}{n} = -p \cdot \log_2 p \quad ,$$

где $p = \frac{1}{n}$ – вероятность любого из отдельных исходов.

Таким образом, неопределенность (обозначим, наконец, ее H), вносимая каждым из равновероятных исходов, равна:

$$H = -p\log_2 p = \log n \qquad (1.3)$$

Данную формулу в 1926 г предложил Хартли.

Теперь попробуем обобщить формулу на ситуацию, когда исходы опытов не равновероятны, например, $p(A_1)$ и $p(A_2)$. Тогда:

$$H_1 = -p(A_1) \cdot \log_2 p(A_1) \quad \text{и} \quad H_2 = -p(A_2) \cdot \log_2 p(A_2)$$
$$H_0 = H_1 + H_2 = -p(A_1) \cdot \log_2 p(A_1) - p(A_2) \cdot \log_2 p(A_2)$$

Обобщая это выражение на *n* неравновероятных исходов, получим:

$$H(A) = -\sum_{i=1}^{n} p(A_i) \log_2 p(A_i) \qquad (1.4)$$

Введенная таким образом величина получила название ***энтропия***.

Впервые мера (1.4) была предложена Клодом Шенноном в его фундаментальной работе "Математические основы теории связи" опубликованной в 1948г в которой были заложены основы современной ТИ. Предполагающая мера была названа энтропией не случайно. Дело в том, что вид формулы (1.4) совпадает с полученным ранее результатом Больцманом выражением для энтропии термодинамической системы.

Рассмотрим взаимосвязь меры Шеннона с мерой Хартли если в источнике может быть реализовано h равновероятных состояний, то вероятность каждого из них, с учетом этого меру неопределенности источника Хартли можно трактовать, как количество информации приходящей на одно дискретное сообщение (поскольку все сообщения источника равновероятные количества информации в каждом из них равны) в тоже время энтропия по Шеннону это среднее количество информации содержащееся в одном из не равновероятных состояний. Она позволяет учесть статистические свойства источника информации.

Наряду с рассмотренными мерами Хартли и Шеннона существуют и другие подходы к определению количества информации. Наиболее интересной, наиболее новой явилась информационная концепция Колмогорова, ее основным тезисом является то, что на основании определения энтропии (1.4) количество информации связывается с вероятностью наступления Pi, т.к. понятие вероятности имеет смысл лишь в связи с массовыми явлениями количества единиц информации в единичном акте и представляющих интерес в связи с данным исходом, оказывается выраженным через вероятности массовых явлений. Шенноновская мера интересна не сама по себе, а как основание встроенной теории позволяющей изменить и расширить существующие предположения о возможностях в технике связи, которая и подлежит в рассмотрении ТИ.

*(b) энтропия является **мерой неопределенности** опыта, в котором проявляются случайные события, и равна **средней неопределенности** всех возможных его исходов.*

Впервые понятие энтропии было введено в 1865 г. немецким физиком Рудольфом Клаузиусом как функции состояния термодинамической системы, определяющей направленность самопроизвольных процессов в системе. Клаузиус сформулировал II начало термодинамики. В частности, он показал, что максимума энтропия достигает при полной раз упорядоченности в системе, чему соответствует состояние равновесия. Другими словами, в физике энтропия оказывается мерой беспорядка в системе. Позднее (в 1872 г.) Людвиг Больцман, развивая статистическую теорию, связал энтропию системы с вероятностью ее состояния, дал статистическое (вероятностное) толкование II-му началу термодинамики и, в частности, показал, что вероятность максимальна у полностью разупорядоченной (равновесной) системы, причем, энтропия и термодинамическая вероятность оказались связанными логарифмической зависимостью! Сходство понятий и соотношений между ними в теории информации и статистической термодинамике, как оказалось позднее, имеет глубокий смысл.

Что дает понятие энтропии для решения практических задач? Рассмотрим одну из них. Пусть имеются два ящика, в каждом из которых по 12 шаров. В первом – 3 белых, 3 черных и 6 красных; во втором – каждого цвета по 4. Опыты состоят в вытаскивании по одному шару из каждого ящика. Что можно сказать относительно неопределенностей этих опытов? Согласно находим энтропии обоих опытов:

$$H_1 = -\frac{3}{12}\log_2\frac{3}{12} - \frac{3}{12}\log_2\frac{3}{12} - \frac{6}{12}\log_2\frac{6}{12} = 1{,}5$$

$$H_2 = -\frac{4}{12}\log_2\frac{4}{12} - \frac{4}{12}\log_2\frac{4}{12} - \frac{4}{12}\log_2\frac{4}{12} = \log_2 3 = 1{,}58$$

Ясно, что $H_2>H_1$, т.е. во втором опыте неопределенность исхода выше, что, кстати, иллюстрирует справедливость формулы.

Чем больше энтропия источника, тем больше степень неожиданности выдаваемых им сообщений в среднем, т.е. тем более неопределенным является ожидание сообщений.

Вернемся к понятию энтропии как меры неопределенности некоторого опыта, исход которого зависит от выбора одного элемента из множества исходных. Множество исходных элементов называется выборочным пространством. Вероятности нахождения элементов исходного множества в том или ином состоянии есть числа положительные, сумма их равна 1.

Выборочное пространство и его вероятностные характеристики представляют собой ансамбль сообщений. Для дискретного ансамбля вероятность события равна сумме вероятностей элементов выборочного пространства, содержащихся в этом событии.

Ансамбль сообщений на выходе источника будем называть ансамблем источника сообщений и обозначать буквой А. Абстрактный алфавит, при помощи которого мы представляем исходное множество элементов источника сообщений, обозначается $\{a_1, a_2, \ldots, a_i, \ldots, a_m\}$. Вероятности появления буквы на выходе источника сообщений обозначают $p(a_1)$, $p(a_2)$, …, $p(a_i)$,… $p(a_m)$. $\sum_{i=1}^{m} p(a_i) = 1$. В этом случае энтропия источника сообщений

$$H(A) = -\sum_{i=1}^{M} p(a_i)\log p(a_i)$$

и представляет собой неопределенность появления на выходе источника сообщений буквы первичного алфавита.

Ансамбль сообщений на выходе приемника будем называть ансамблем приемника сообщений и обозначать буквой В. Для того чтобы отличить переданные и принятые сигналы, абстрактный алфавит в котором представлен ансамбль приемника сообщений, обозначается $\{b_1, b_2, \ldots, b_j, \ldots, b_n\}$, а соответствующие вероятности - $p(b_1)$, $p(b_2)$, …, $p(b_i)$,… $p(b_n)$.

Энтропия приемника сообщений

$$H(B) = -\sum_{j=1}^{N} p(b_j)\log p(b_j)$$

и представляет собой неопределенность появления на входе приемника буквы после ее появления на выходе источника сообщений. Если в канале связи не происходит потерь информации, то всегда буква a_1 соответствует букве b_1, $a_2 - b_2$ и т.д. При этом $H(A)=H(B)$.

Свойства энтропии.

1. Энтропия является вещественной и неотрицательной величиной, так как для любого i ($1<i<N$) p_i изменяется в интервале от 0 до 1, $\log p_i$ отрицателен и, следовательно, - $p_i \log p_i$ положительна.

2. Энтропия – величина ограниченная.

3. Энтропия обращается в нуль лишь в том случае, если вероятность одного из состояний равна единице; тогда вероятности всех остальных состояний, равны нулю.

4. Энтропия максимальна, когда все состояния источника равновероятны.

$H_{\max} = \log N$

5. Энтропия объединения нескольких статистически независимых источников информации равна сумме энтропий исходных источников.

Условная энтропия

Если состояния элементов системы не зависят друг от друга, если состояние одной системы не зависит от состояния другой системы, то неопределенность того, что некоторый элемент системы будет находиться в одном из k возможных состояний полностью определялась бы вероятностными характеристиками отдельных элементов системы, либо вероятностными характеристиками состояний самих систем. При этом подразумевается, что символы сообщения взаимонезависимы, т.е. с приходом одного символа распределение вероятностей последующих символов не изменяется. На практике же чаще всего встречаются взаимозависимые символы и сообщения. Если передавать не просто отдельные буквы алфавита, а смысловые сообщения, то можно убедиться, что существует взаимозависимость передаваемых символов. Одни буквы встречаются чаще. другие реже, одни буквы и слова часто следуют за другими, другие редко.

Понятие условной энтропии широко используется для определения информационных потерь при передаче информации.

Если элементы источника сообщений принимают состояния a_1, a_2, …..a_n с вероятностями соответственно $p(a_1)$, $p(a_2)$,….$p(a_n)$, а элементы адресата – состояния $в_1$, $в_2$,…..$в_m$, с вероятностями соответственно $p(в_1)$, $p(в_2)$,…..$p(в_m)$, то понятие условной энтропии $H(в_j/a_i)$ выражает неопределенность того, что отправив a_i, мы получим $в_j$. Если в канале связи присутствуют помехи, то с различной степенью вероятности может быть принят любой из сигналов b_j, и наоборот, принятый сигнал $в_j$ может появится в результате отправления любого из сигналов a_i. Если в канале связи помехи отсутствуют, то всегда посланному сигналу а соответствует принятый сигнал в и т.д. При этом энтропия источника $H(A)$ равна энтропии приемника $H(B)$. Если в канале связи присутствуют помехи, то они уничтожают часть передаваемой информации.

Информационные потери полностью описываются через частную и общую условную энтропию. Вычисление частных и общей условной энтропии удобно производить при помощи канальных матриц. **Если канал связи описывается со стороны источника сообщений (т.е. известен посланный сигнал)**, то вероятность того, что при передаче сигнала а по каналу связи с помехами мы получим сигнал в, обозначается как условная вероятность $p(в_j/a_i)$, а канальная матрица имеет вид:

B A	b_1 b_2 … b_j …….. b_m
a_1 a_2	$p(b_1/a_2),\ \ p(b_2/a_1),...,\ \ p(b_j/a_1),\ \ ...,\ \ p(b_m/a_1)$ $p(b_1/a_2),\ \ p(b_2/a_2),...,\ \ p(b_j/a_2),\ \ ...,\ \ p(b_m/a_2)$ …………………………………………………………..
a_i	$p(b_1/a_i),\ \ p(b_2/a_i),...,\ \ p(b_j/a_i),\ \ ...,\ \ p(b_m/a_2)$ ……………………………………………………………
a_m	$p(b_1/a_m),\ \ p(b_2/a_m),\ \ ...,\ \ p(b_j/a_m),...,\ \ p(b_m/a_m)$

Вероятности, которые расположены по диагонали, определяют вероятности правильного приема, остальные – ложного. Значения цифр, заполняющих колонки канальной матрицы, обычно уменьшаются по мере удаления от главной диагонали и при полном отсутствии помех все, кроме цифр, расположенных на главной диагонали, равны нулю.

Прохождение данного вида сигнала со стороны источника сообщений в данном канале связи описывается распределением условных вероятностей вида $p(в_j/a_i)$. Например, для сигнала a_1 распределением вида

$$p(b_1/a_1)+p(b_2/a_1)+\ldots+p(b_j/a_i)+\ldots+p(b_m/a_1)=1$$

Потери информации, приходящиеся на долю сигнала a_i описываются при помощи частной условной энтропии. Например, для сигнала a_1

$$H(b_j/a_1)=-\sum p(b_j/a_1)\log p(b_j/a_1)$$

Суммирование производится по j, так как i-е состояние остается постоянным.

Потери при передаче всех сигналов по данному каналу связи описываются при помощи общей условной энтропии. Для ее вычисления следует просуммировать все частные условные энтропии, т.е. произвести двойное суммирование по *i* и по *j*. При этом, в случае равновероятных появлений сигналов на выходе источника сообщений

$$H(B/A) = -\frac{1}{N}\sum_{j}\ \sum_{i} p(b_j/a_i)\log p(b_j/a_i)$$

В случае неравновероятного появления символов источника сообщений следует учесть вероятность появления каждого символа, умножив на нее соответствующую частную условную энтропию. При этом общая условная энтропия

$$H(B/A) = -\sum_{i}\sum_{j} p(a_i)p(b_j/a_i)\log p(b_j/a_i)$$

Если исследовать канал связи ***со стороны приемника сообщений*** (т.е. известен принятый сигнал), то с получением сигнала $в_j$ предполагаем, что был послан какой-то из сигналов a_i. При этом канальная матрица имет вид:

B A	b_1 b_2 … b_j …….. b_m
a_1	$p(a_1/b_2),\quad p(a_2/b_1),\ ...,\quad p(a_1/b_j),$
a_2	$p(a_1/b_2),\quad p(a_2/b_2),\ ...,\quad p(a_2/b_j),$
	$...,\quad p(a_1/b_m)$
a_i	$...,\quad p(a_1/b_m)$
	………………………………………………………..
a_m	
	$p(a_i/b_1),\quad p(a_i/b_2),\ ...,\quad p(a_i/b_j),\quad ...,\quad p(a_i/b_m)$
	………………………………………………………
	$p(a_m/b_1),\quad p(a_m/b_2),\quad ...,\quad p(a_m/b_j),\ ...,\quad p(a_m/b_m)$

В этом случае единице должны равняться суммы условных вероятностей не по строкам, а по столбцам канальной матрицы

$$p(a_1/b_j) + p(a_2/b_j) + \ldots + p(a_i/b_j) + \ldots + p(a_m/b_j) = 1$$

Частная условная энтропия

$$H(a_i/b_j) = -\sum_{i=1}^{m} p(a_i/b_j)\log p(a_i/b_j)$$

Общая условная энтропия

$$H(A/B) = -\sum_{j}\sum_{i} p(b_j)p(a_i/b_j)\log p(a_i/b_j).$$

Понятие условной энтропии в теории информации используется при определении взаимозависимости между символами кодируемого алфавита, для определения потерь при передаче информации по каналам связи, при вычислении энтропии объединения.

Во всех случаях при вычислении условной энтропии в том или ином виде используются условные вероятности.

Если в канале связи помехи отсутствуют, то все элементы канальной матрицы, кроме элементов, расположенных на главной диагонали, равны нулю. Вероятность получения правильного сигнала станет безусловной, а условная энтропия будет равна нулю. Канальная матрица будет иметь вид

$$p(a/b) = \begin{vmatrix} 1\ 0\ 0 \ldots . 0 \\ 0\ 1\ 0 \ldots . 0 \\ \ldots\ldots\ldots\ldots \\ 0\ 0\ 0 \ldots . 1 \end{vmatrix}$$

В этом случае условная энтропия будет равна нулю.

Энтропия объединения

Взаимная энтропия, или как ее часто называют, *энтропия объединения* используется для вычисления энтропии совместного появления статистических зависимых сообщений.

Пусть $(a_1, a_2, \ldots, a_i, \ldots, a_n)$ есть выборочное пространство *A*, характеризующее источник сообщений, а $(b_1, b_2, \ldots, b_j, \ldots, b_m)$ есть выборочное пространство *B*, характеризующее приемник сообщений. При этом *а* есть сигнал на входе шумящего канала, а *в* – сигнал на его выходе. В этом случае взаимная энтропия представляет собой информацию о переданном сигнале a_i, содержащегося в принятом сигнале b_j. Взаимосвязь переданных и принятых сигналов описывается вероятностями совместных событий вида $p(a_i, b_j)$, а взаимосвязь выборочных пространств *A* и *B* описывается матрицей объединения вида:

$$p(a_i, b_j) = \begin{vmatrix} p(a_1,b_1) & p(a_1,b_2) & \ldots & p(a_1,b_m) \\ p(a_2,b_1) & p(a_2,b_2) & \ldots & p(a_2,b_m) \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots\ldots \\ p(a_m,b_1) & p(a_m,b_2) & \ldots & p(a_m,b_m) \end{vmatrix}$$

Если матрица описывает канал связи, то число строк матрицы равно числу столбцов, *m*=*n*, и пределы суммирования по *i* и по *j* одинаковы.

Независимо от равенства или неравенства числа строк числу столбцов матрица объединения обладает следующими свойствами

1) $\sum_i p(a_i, b_j) = p(b_j)$

Сумма вероятностей по столбцам равна вероятности приёмника.

2) $\sum_j p(a_i, b_j) = p(a_i)$

Сумма вероятностей по строкам рана вероятности источника.

3) $\sum_i p(a_i) = \sum_j p(b_j) = 1$, т. е. $\sum_i \sum_j p(a_i, b_j) = 1$

Сумма всех элементов равна 1.

Условные вероятности при помощи матрицы объединения находятся следующим образом

$$p(a_i / b_j) = \frac{p(a_i, b_j)}{\sum_i p(a_i, b_j)} = \frac{p(a_i, b_j)}{p(b_j)}.$$

$$p(b_j / a_i) = \frac{p(a_i, b_j)}{\sum_j p(a_i, b_j)} = \frac{p(a_i, b_j)}{p(a_i)}.$$

Взаимная энтропия ансамблей A и B про помощи матрицы объединения вычисляется путем последовательного суммирования по строкам или по столбцам всех вероятностей вида *p(a,b)*, умноженных на логарифм этих же вероятностей

$$H(A,B) = -\sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j) \quad \text{бит/два символа}$$

Размерность «бит/два символа» объясняется тем, что взаимная энтропия представляет собой неопределенность возникновения пары символов, то есть неопределенность на два символа.

Взаимная энтропия передаваемого ансамбля A и принимаемого ансамбля B равна сумме безусловной энтропии H(A) и условной энтропии H(B/A)

$$H(A,B) = H(A) + H(B/A)$$

H(B/A) в данном случае представляет ту добавочную информацию, которую дает сообщение B после того как стала известна информация, содержащаяся в сообщении A.

Таким образом, условная энтропия представляет собой неопределенность того, что при приеме *в* было послано *а*, а взаимная энтропия отражает неопределенность возникновения пары вида *ав*.

Так как взаимная энтропия есть неопределенность относительно пары символов, сигналов, состояний, в общем случае, относительно пары элементов взаимосвязанных выборочных пространств *A* и *B*, то не имеет значения имеет ли эта пара вид *ав* или *ва,* так как неопределенность возникновения такого сочетания – одинакова. *Взаимная энтропия обладает свойством симметрии.*

$$H(A,B) = H(B,A)$$

Если построена матрица вероятностей p(a,в), описывающая взаимосвязь двух произвольных выборочных пространств, в частности взаимосвязь входа и выхода шумящего канала связи, то остальные информационные характеристики могут не задаваться, так как матрица объединения обладает информационной полнотой.

Определение. Набор информационных характеристик произвольного канала связи считается информационно полным, если с помощью этого набора, путем алгебраических преобразований, можно получить любую другую информационную характеристику того же канала связи.

Согласование физических характеристик сигнала и канала.

Конкретный канал связи обладает определенными физическими параметрами, от которых зависит возможность передачи по нему тех или иных сигналов. Независимо от назначения непрерывного канала его можно характеризовать тремя основными параметрами: время, в течении которого он представляется для передачи сигнала Tk, шириной полосы пропускания сигнала Fk и допустимым превышением сигнала над помехами Hk . Превышение допустимого превышения сигнала Hk характеризуется разностью максимально допустимого сигнала в канале Pu max и уровня помех Ps. Для проводных каналов превышение в основном определяется пробивным напряжением и уровнем перекрестных помех, для радиоканалов - возможностями выявления сигнала на соответствующих расстояниях.

Произведение указанных основных параметров канала связи принято называть объемом канала и обозначать Vk

Vk = Tk Fk Hk

При оценке возможностей передачи сигнала по каналу с заданными физическими характеристиками также ограничиваются рассмотрением трех основных параметров сигнала: его длительности Tc ширины спектра Fc и превышением над помехой Hc причем

Hc =log(Pu /P s)

где Ps - средняя мощность помехи в канале.

Превышение Hc связано с возможностями передатчика и дальностью передачи. Чем больше превышение Hc, тем меньше вероятность ошибочного приема. Аналогично объему канала вводится понятие объема V 4c 0 передаваемого сигнала:

Vc=TcFcHc

Как объем сигнала, так и объем канала могут быть представлены в трехмерном пространстве с соответствующими координатами T, F, H

Необходимым условием принципиальной возможности неискаженной передачи сигнала выполнения соотношения

$$V_c \le V_k$$

При этом, однако могут потребоваться преобразования для обеспечения достаточных условий передачи, а именно:

$$T_c \le T_k;\ F_c \le F_k;\ H_c \le H_k$$

Когда канал имеет меньшую полосу пропускания, чем практическая ширина спектра, подлежащего передаче сигнала, последнюю можно уменьшить за счет увеличения длительности сигнала. Объем сигнала при этом сохраняется неизменным. Практически также преобразование можно осуществить, например, посредством записи сигнала на магнитную ленту с высокой скоростью и последующего воспроизведения со скоростью, при которой ширина его спектра равна полосе пропускания канала.

Если, наоборот, широкополосный канал представляется на время меньшее длительности сигнала, то согласование осуществляется за счет расширения спектра сигнала. Для реализации также может использоваться накопитель на магнитной ленте, однако в данном случае скорость воспроизведения должна быть выше скорости записи.

При низком допустимом уровне превышения сигнала в канале преобразование заключается в уменьшении уровня превышения передаваемого сигнала с одновременным увеличением его длительности путем многократного повторения передачи. Возможны и другие виды преобразования.

Информационные характеристики источника дискретных сообщений и канала связи

1. Производительность ИС – величина, определяемая КИ, выдаваемой ИС в единицу времени:

$$\Pi(X) \equiv \frac{I_0(X)}{T} \qquad \left[\frac{бит}{c}\right].$$

Если ИС с Э $H(X)$ за время Т выдает *n* сообщений каждая длительностью τ_c, то общее КИ: $I_0(X) = n * H(X)$.

$$T = n * \tau_c, \text{ следовательно } \Pi(X) = \frac{n * H(X)}{n * \tau_c} = \frac{H(X)}{\tau_c} \qquad \left[\frac{бит}{c}\right].$$

При $\tau_c = \text{const}$ $\Pi(X)$ зависит только от Э ИС, т. е. от статистической структуры источника:

$$\max \Pi(X) = \frac{\max H(X)}{\tau_c} = \frac{\log_2 m}{\tau_c}.$$

Таким образом, максимальной производительностью обладают источники, сообщения которых независимы и равновероятны. Это оптимальные источники. Чем больше отличается распределение в источнике от независимого и равновероятного, тем меньше производительность ИС, т. е. источник содержит сообщения с меньшей информацией.

2. Избыточность ИС – величина, показывающая, насколько сообщение от источника с Э $H(X)$ отличается по количеству содержащейся в них информации от оптимального источника:

$$R(X) \equiv 1 - \frac{H(X)}{H\max(X)} = 1 - \frac{H(X)}{\log_2 m}.$$

Избыточность заключена в $0 \le R(X) \le 1$.

При оптимальном источнике R(X) = 0, при $H(X) = 0$ R(X) = 1.

Наличие в источнике избыточности показывает, что сообщение от этого источника не является оптимальным, и производительность источника и скорость передачи информации меньше максимально возможных.

Для русского языка избыточность $\max H(X)=\log_2 32 = 5 \quad \left[\frac{бит}{символ}\right]$. Численное значение Э с учетом неравномерного появления букв:

$H(X)= 4.42 \quad \left[\frac{бит}{символ}\right]$,

$R(X) = 1 - \frac{4.42}{5} = 0.116.$

Вероятности появления отдельных букв с учетом предшествующих букв уменьшает Э.

С учетом двухбуквенных сочетаний $H(X)$=3.52, R(X)=0.3.

Общим способом уменьшения избыточности ИС является переход от первичного ИС (алфавита) к искусственному вторичному источнику (алфавиту) с R(X) = 0, статистическая структура сообщений в котором оптимальна.

3. Техническая скорость передачи C_m определяется числом посылок сигнала, проходящих по каналу в единицу времени (с):

$$C_m = \frac{n}{T} = \frac{n}{n*\tau_c} = \frac{1}{\tau_c} \qquad \text{[baud] или [бод].}$$

1 бод – 1 посылка в секунду.

Техническая скорость выбирается из условия согласования ширины спектра сигнала и полосы пропускания канала: $\frac{1}{\tau_c} = \Delta F_c = \Delta F_k$.

Техническая скорость не зависит от информационных характеристик ИС.

4. Информационная скорость передачи $C_и$ определяется количеством информации, проходящей по каналу передачи в единицу времени:

$$C_и \equiv \frac{I_0(X)}{T} = \frac{n*H(X)}{T} = C_m * H(X) \qquad \left[\frac{бит}{c}\right].$$

Информационная скорость является более полной характеристикой канала сообщений, чем техническая скорость.

5. Пропускная способность канала C_0. Пропускная способность определяется как предельное значение информационной скорости передачи, достижимое в определенных условиях передачи.

$$C_0 = \max C_и = \max[C_m * H(X)] \qquad \left[\frac{бит}{c}\right].$$

Пропускная способность наиболее полно характеризует возможности передачи по каналу и зависит как от производительности ИС, так и от мощности помех, действующих в канале связи.

а) При отсутствии помех, действующих в канале связи (телефонная линия), КИ, проходящей по каналу, определяется понятием собственной информации, и информационная скорость определяется как отношение общего КИ ко времени передачи:

$$C_r = \frac{I_0(X)}{T} = \frac{nH(X)}{n\tau_c} = \frac{H(X)}{\tau_c} = \Pi(X).$$

Информационная скорость численно совпадает с производительностью ИС:

$$C_0 = \max C_r = \max \Pi(X) = C_m \log_2 m.$$

Пропускная способность в этом случае реализуется тогда, когда производительность источника максимальна, т. е. в источниках независимых и равновероятных сообщений.

Если ИС является двоичным m = 2, то $C_0 = C_m \log_2 2 = C_m$.

Формально в этом случае пропускная способность равна технической скорости передачи.

В большинстве реальных случаев ИС обладают избыточностью, т. е. R(X) > 0, следовательно, $H(X) < \log_2 m$. Э меньше максимально возможной:

$$C_и = \Pi(X) < \frac{\log_2 m}{\tau_c} = C_0 .$$

Максимальное значение информационной скорости меньше пропускной способности. Пропускная способность не реализуется, канал используется не полностью.

Для повышения информационной скорости передачи следует стремиться уменьшить избыточность ИС, что достигается путем преобразования сообщений источника в оптимальные (независимые, равновероятные).

К. Шеннон доказал несколько теорем, первая из которых применима к каналу без помех.

1 Теорема Шеннона. Если производительность ИС меньше пропускной способности канала, т. е.

$$\Pi(X) < \frac{\log_2 m}{\tau_c},$$

то сообщение от источника можно преобразовать так, чтобы передавать их сколь угодно точно по дискретному идеальному каналу без шумов со скоростью достаточно близкой к пропускной способности.

Комментарий. Шеннон дает идеологическую основу для преобразования ИДС с избыточностью в оптимальный источник. В теории информации для этого используется так называемое эффективное (статистическое) кодирование, в результате применения которого $C_и = C_0$, избыточность равна нулю, производительность максимальна.

Технический комментарий. Согласно этой теореме, с точки зрения технической реализации существует способ кодирования и декодирования сообщений, при котором вероятность ошибочного декодирования сколь угодно мала. Кроме того, Шенноном доказано, что, если производительность источника превышает пропускную способность $\Pi(X) > C_0$, таких способов кодирования не существует.

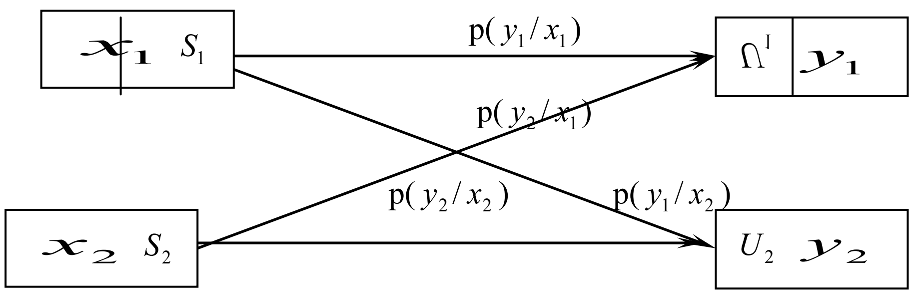
b) В реальных каналах с помехами КИ определяется понятием взаимной информации, поэтому пропускная способность при условии C_m = const:

$$C_0 = \frac{I(Y,X)}{T} = \frac{I(X,Y)}{T} = C_m * \max[H(X) - H(X/Y)] = C_m * \max[H(Y) - H(Y/X)].$$

Величины $H(X)$ и $H(X/Y)$, определяющие пропускную способность, зависят от многих факторов, в частности от статистической структуры ИС, от вида применяемой модуляции, от мощности помех, действующих в канале и т. д., т. е. зависит от группы факторов, влияющих на вероятность ошибки при приеме.

Введем выражение для пропускной способности бинарного симметричного канала информации без памяти (любая компьютерная сеть). Канал не обладает памятью, если нет зависимости от предыдущих символов.

Модель канала



где x - сообщение, S - сигнал, U- сигнал с шумом, y - принятое сообщение.

Вероятности правильного приема: p(y_1/x_1) - вероятность принятия 0 при посылке 0, p(y_2/x_2) - вероятность принятия 1 при посылке 1. Пусть $p_{ош}$ = p(y_2/x_1) = p(y_1/x_2) ≡ p, $p_{прав}$ = p(y_1/x_1) = p(y_2/x_2) = 1 – p. Такой канал называется симметричным.

Как связать вероятность ошибочного приема

$C_0 = C_m * \max[H(Y) - H(Y/X)]$,

$$H(Y/X) = M\{H(Y/x_i)\} = \sum_{i=1}^{2} p(x_i)H(Y/x_i) = -\sum_{i=1}^{2} p(x_i)\sum_{j=1}^{2} p(y_j/x_i)\log_2 p(y_j/x_i) =$$

$$= -\,p(0)[(1-p)\log_2(1-p) + p\log_2 p] - p(1)[p\log_2 p + (1-p)\log_2(1-p)] =$$

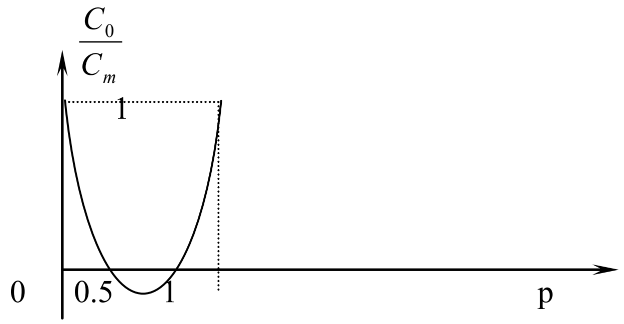
$$= -\,[p(0)+p(1)][p\log_2 p + (1-p)\log_2(1-p)] = -\,p\log_2 p - (1-p)\log_2(1-p).$$

Считается, что p задано системой, следовательно, требуется максимизировать $H(Y)$. Если канал бинарный, сообщения равновероятны и независимы, то $\max H(Y) = 1$,

$C_0 = C_m\,[p\log_2 p + (1-p)\log_2(1-p)]$.

Для источника с m состояниями

$$C_0 = C_m\,[\log m + p\log\frac{p}{m-1} + (1-p)\log(1-p)].$$



p очень мало, например, 0.01. С точки зрения технической реализации правая ветвь не используется. С увеличением p пропускная способность падает. Если вероятность ошибки совпадает с вероятностью правильного приема p = $p_{прав}$= 0.5, пропускная способность C_0= 0. Говорят, что в этом случае наступает обрыв канала, по каналу передавать информацию бессмысленно.

Вторая теорема Шеннона относится к каналу с помехами.

2 Теорема Шеннона. Если производительность ИС меньше пропускной способности канала $\Pi(X) < C_0$, то сообщение от этого источника можно преобразовать так, чтобы передавать их по каналу с помехами с любой степенью точности, т. е. за счет существования избыточности в сообщениях, вводимой специальным образом, можно уменьшить вероятность ошибки до сколь угодно малой величины.

С точки зрения технической реализации эта теорема означает, что существует способ кодирования и декодирования, при котором вероятность ошибочного декодирования может быть сколь угодно малой. Если $\Pi(X) > C_0$, то таких способов не существует.

Вторая теорема Шеннона является идеологической основой для существования помехоустойчивого (корректирующего) кодирования в каналах связи.

Модели дискретных каналов

Дискретным каналом называют совокупность средств, предназначенных для передачи дискретных сигналов. Такие каналы широко используются при передачи данных, в телеграфии, радиолокации.

Дискретные сообщения, состоящие из последовательности знаков алфавита источника сообщений первичного алфавита Z1 Z2 ... ZL преобразуются в кодирующем устройстве в последовательности символов.

Объём *m* алфавита символов (вторичным алфавитом) U1,U 2 ... Um, как правило меньше объема *L* алфавита знаков, но они могут и совпадать.

Материальным воплощением символа является элементарный сигнал, получаемый в процессе манипуляции - дискретного изменения определенного параметра переносчика информации. Элементарные сигналы формируются с учетом физических ограничений, накладываемых конкретной линией связи. В результате манипуляции последовательности символов ставится в соответствие сложный сигнал. Множество сложных сигналов конечно. Они различаются числом, составом и взаимным расположением элементарных сигналов.

Термины "элементарный сигнал" и "символ", так же как "сложный сигнал" и "последовательность символов", в дальнейшем будут использоваться как синонимы.

Информационная модель канала с помехами задается множеством символов на его входе и выходе и описанием вероятностных свойств передачи отдельных символов. В общем случае канал может иметь множество состояний и переходить из одного состояния в другое как с течением времени, так и в зависимости от последовательности передаваемых символов.

В состоянии канал характеризуется матрицей условных вероятностей $p(v_j / u_i)$ того, что переданный символ *Ui* будет воспринят на выходе как символ *Vj*. Значение вероятностей в реальных каналах зависит от многих различных факторов: свойств сигналов, являющихся физическими носителями символов (энергия, вид модуляции и т.д.), характера и интенсивности воздействующих на канал помех, способа определения сигнала на приемной стороне.

При наличии зависимости переходных вероятностей канала от времени, что характерно практически для всех реальных каналов, он называется нестационарным каналом связи. Если эта зависимость несущественна, используется модель в виде стационарного канала, переходные вероятности которого не зависят от времени. Нестационарный канал может быть представлен рядом стационарных каналов, соответствующих различным интервалам времени.

Канал называется с "памятью" если переходные вероятности в данном состоянии канала зависят от его предыдущих состояний. Если переходные вероятности постоянны, т.е. канал имеет только одно состояние, он называется стационарным каналом без памяти.

Стационарный дискретный двоичный канал без памяти однозначно определяется четырьмя условными вероятностями: р*(0/0) ,p(1/0),p(0/1), p(1/1).*

Если вероятности искажения можно принять равными, т.е. *p(0/1) ~ p(1/0)=q,* то такой канал называется двоичным симметричным каналом при *p(0/1) -p(1/0)* канал называется несимметричным). Символы на его выходе правильно принимают с вероятностью *1-p=q*. Математическая модель упрощается.

Именно этот канал исследовался наиболее интенсивно не столько в силу своей практической значимости (многие реальные каналы описываются им весьма приближенно), сколько в силу простоты математического описания.

Важнейшие результаты, полученные для двоичного симметрического канала, распространены на более широкие классы каналов.

Следует отметить еще одну модель канала, которая в последнее время приобретает все большее значение. Это дискретный канал со стиранием. Для него характерно, что алфавит выходных символов отличается от алфавита входных символов. На входе, как и ранее, символы 0 и 1, а на выходе канала фиксируются состояния, при которых сигнал с равным основанием может быть отнесен как к единице так и к нулю. На месте такого символа не ставится ни нуль, ни единица: состояние отмечается дополнительным символом стирания S. При декодировании значительно легче исправить такие символы, чем ошибочно определенные.

Информационные характеристики непрерывного источника сообщений и канала связи

К информационным характеристикам источников непрерывных сообщений (ИНС) относятся дифференциальная энтропия (ДЭ) и эпсилон-энтропия (ЭЭ).

Обобщим понятие энтропии и КИ на ансамбль непрерывных сообщений.

ИНС может вырабатывать любую реализацию сообщения x(t) из неограниченного множества. В этом случае на множестве реализаций задана плотность распределения вероятности или интегральный закон:

$x(t) \in X \to w(x)\ [F(x)]$.

Заменим непрерывное состояние ИС x(t) дискретными $x_1, x_2, \ldots, x_m$ отстоящими друг от друга на интервал Δx. В этом случае можно определить вероятность нахождения (появления) отсчета $x_1, x_2, \ldots, x_m$ как $p_i = w(x_i)\Delta x$.

Поэтому на основании определения среднего КИ можно определить эту величину в данном случае следующим образом:

$$M\{-\log p_i\} = H(X) = -\sum_{i=1}^{m} w(x_i)\Delta x \log(w(x_i)\Delta x).$$

Устремляя к нулю и учитывая условия нормировки для плотности распределения вероятности

$$\sum_{i=1}^{m} w(x_i)\Delta x = 1,$$

мы получим, что

$$H(X) = -\int_{-\infty}^{+\infty} w(x)\log w(x)dx - \lim_{\Delta x\to 0}\log \Delta x.$$

$-\lim\limits_{\Delta x\to 0}\log \Delta x \to +\infty$ и не зависит от w(x), следовательно, среднее КИ и Э непрерывного сообщения являются бесконечно большими величинами. Но на практике интересуются величиной приращения Э. Поэтому величину $\log \Delta x$, которая зависит от Δx, не учитывают и вводят понятие ДЭ

$$h(x) \equiv -\int_{-\infty}^{+\infty} w(x)\log w(x)dx. \qquad (1)$$

Замечание. ДЭ в отличие от обычной Э нельзя рассматривать как меру собственной информации. Она не обладает многими свойствами обычной Э, в частности она может принимать и отрицательные значения. Информационный смысл имеет разность двух ДЭ, чем и объясняется название. Что касается свойства аддитивности, то оно сохраняется, т. е. ДЭ нескольких сечений случайного процесса равна сумме их ДЭ, вычисленной с учетом вероятностной зависимости между сечениями. Если решить вариационную задачу поиска распределения вероятности, доставляющую максимум функции ДЭ случайного процесса при ограничениях вида:

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

$$\int_{-\infty}^{+\infty} x^2 w(x)dx = \sigma_x{}^2,$$

окажется, что решением является

$$w(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\exp(-\frac{x^2}{\sigma_x{}^2})$$ - Нормальный закон или закон Гаусса.

Максимальное значение ДЭ: $\max k(x) = \log\sqrt{2\pi e\sigma_x^2}$.

Только дисперсия влияет на величину $\max k(x)$.

Понятие УЭ для ИНС для фиксированного состояния вводится по аналогии с соотношением (1).

$$h(X/y) = -\int_{-\infty}^{+\infty} w(x/y)\log w(x/y)dx, \qquad (2)$$

y - параметр.

КИ, содержащееся во всех возможных сообщениях источника для всех возможных состояний источника X имеет вид:

$$h(X/Y) = M\{h(X/y)\} = \int_{-\infty}^{+\infty} w(y)h(X/y)dy = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} w(y)w(x/y)\log w(x/y)dxdy. \quad (3)$$

Энтропия объединения ИНС определяется согласно приведенным ранее формула следующим образом:

$h(X*Y) = h(X) + h(Y/X)$,

$h(X)$ - ДЭ, $h(Y/X)$ - УЭ, определенная по (3).

Взаимное КИ I(X,Y) между двумя ИНС X и Y вводят как разность безусловной и условной ДЭ:

$$I(X,Y) \equiv h(X) - h(X/Y). \qquad (4)$$

Две реализации сообщений: принятое *y(t)* и переданное *x(t)* называются эквивалентными, если различие между ними несущественно в смысле выбранного критерия (обычно это критерий СКО). Вводится понятие отклонения $\varepsilon(t) \equiv x(t) - y(t)$, задается ε_0^2. Если $\overline{\varepsilon^2}(t) < \varepsilon_0^2$, то реализации считаются эквивалентными.

ЭЭ $H_\varepsilon(x)$ называется минимальное среднее КИ, содержащееся в одном отсчете сообщения y(t) относительно сообщения x(t), при котором они еще эквивалентны. В соответствии с соотношением (4)

$H_\varepsilon(x) \equiv \min I(x,y)$,

где I(x,y) – взаимное КИ между реализациями x и y.

$H_\varepsilon(x) = h(x) - \max\limits_{\{w(x/y)\}} h(x/y)$.

ЭЭ определяет количество существенной информации, содержащейся в одном отсчете непрерывного сообщения.

Рассмотрим ИНС x(t), представляющий собой стационарный гауссовский процесс с ограниченной дисперсией σ_x^2. Процесс x(t) может быть представлен как

x(t) = y(t) – $\varepsilon(t)$,

где $\varepsilon(t)$ – шум воспроизведения. Поэтому УДЭ $h(x/y)$, присутствующая в определении ЭЭ, при заданном сообщении x(t) полностью определяется шумом воспроизведения $\varepsilon(t)$. Поэтому $\max h(x/y) = \max h(\varepsilon)$. Ранее было показано, что максимум $\max h(\varepsilon)$ достигается на гауссовском случайном процессе $\varepsilon(t)$:

$\max h(x/y) = \max h(\varepsilon) = \log_2 \sqrt{2\pi e \sigma_\varepsilon^2}$.

В этом случае выражение для ЭЭ достигает максимума, если x(t) и $\varepsilon(t)$ – гауссовские процессы. ЭЭ гауссовского источника с ограниченной дисперсией:

$$H_\varepsilon(x) = \log_2 \sqrt{2\pi e \sigma_x^2} - \log_2 \sqrt{2\pi e \sigma_\varepsilon^2} = 0.5 \log_2 \frac{\sigma_x^2}{\sigma_\varepsilon^2}. \qquad (5)$$

Соотношение характеризует КИ, приходящееся на один отсчет. Отношение двух дисперсий характеризует отношение сигнал/шум, при котором сообщения y(t) и x(t) можно считать эквивалентными. Это величина зависит от характера передаваемых сообщений, и при независимых отсчетах сообщений содержащаяся в них информация суммируется.

Производительность ИНС определяется как КИ, выдаваемое этим источником в единицу времени (в системе СИ 1 с) при заданном критерии эквивалентности. Это означает, что здесь мы будем оперировать понятием ЭЭ.

Если средняя скорость выдачи независимых отсчетов сообщений (техническая скорость) равняется C_m, то ε – производительность ИС для гауссовского процесса:

$\Pi_\varepsilon(x) = C_m\, H_\varepsilon(x) = C_m\,(\, h(x) - \log_2 \sqrt{2\pi e \sigma_\varepsilon^2}\,)$.

Отсчеты берутся в соответствии с теоремой Котельникова:

При равном спектре сообщения в полосе $0 \ldots F_в$ отсчеты являются некоррелированными, а для гауссовского ИС – независимыми. Для гауссовского процесса:

$$\Pi_\varepsilon(x) = 2\, F_в\, H_\varepsilon(x),$$

с учетом (5)

$$\Pi_{\varepsilon}(x) = F_{в} \log(\frac{\sigma_x^2}{\sigma_{\varepsilon}^2}) = F_{в} \log \rho .$$

КИ, выдаваемое гауссовским ИС за время Т, определяется:

$$I_0(x) = \mathrm{T}^* \, \Pi_{\varepsilon}(x) = \mathrm{T}^* F_{в} \log \rho . \qquad (6)$$

Величина $I_0(x)$ совпадает по определению с объемом сигнала, если считать, что динамический диапазон сигнала равен $\log \rho$. Выражение определяет максимальное КИ, выдаваемое ИС за время Т, т. к. производительность гауссовского источника больше производительности любого другого источника с той же мощностью.

Согласование статистических свойств источника сообщений и канала связи.

Согласование статистических свойств и отражающих их информационных характеристик источника сообщений и канала связи проводится с целью улучшения качества осуществляется по трем основным показателям: достоверности, средней скорости передачи и сложности технической реализации системы, определяющей ее стоимость и надежность. Хотя с точки зрения практики сложность технической реализации может иметь решающее значение, при определении предельных возможностей системы целесообразно ограничиться только первыми двумя показателями.

Достоверность дискретного канала обычно оценивается значением вероятности ошибочного приема одного символа (элементарного сигнала). В случае передачи непрерывных сообщений о достоверности судят по значению среднеквадратической ошибки при воспроизведении сообщения

M[E]=M[(W(t)-Z(t))]

где W(t) - сообщение, поступающее с выхода канала;

Z(t) - сообщение на выходе канала;

Достоверность характеризует помехоустойчивость информационной системы.

Под скоростью передачи подразумевают среднее количество информации, передаваемое по каналу в единицу времени. Именно эта (а не техническая) скорость формирования символов подлежит согласованию с пропускной способностью канала.

Скорость передачи информации характеризует эффективность системы.

Если высоких требований в отношении скорости передачи и достоверности к системе передачи не предъявляется, то согласование статистических (информационных) характеристик источника сообщений и канала связи не является принципиально необходимым.

При преобразовании сообщений в сигналы в этом случае могут преследоваться две основные цели. Одна из них заключается в том, чтобы преобразовать сообщения в такую систему символов (код), чтобы она обеспечивала простоту и надежность аппаратурной реализации информационных устройств и приемлемую их эффективность:

простоту аппаратуры различения элементарных сигналов, соответствующих отдельным символам, приемлемое время при их передаче, простоту выполнения в этой системе арифметических и логических действий. Техническая реализация процесса кодирования в таком простейшем виде при непрерывном входном сигнале осуществляется аналого-цифровыми преобразователями.

Другой целью преобразования сообщения является защита их от санкционированного доступа. Такое преобразование называют шифрованием. Оно может проводится как на уровне знаков, так и на уровне символов.

В случае отсутствия необходимости в статистическом согласовании источника сообщений с каналом связи вопросы повышения качества функционирования системы решаются для дискретного канала о входа модулятора до выхода демодулятора.

Считается, что символы на вход модулятора поступают равновероятно и статистические связи между ними отсутствуют.

Из множества сигналов, удовлетворяющих заданным ограничениям по мощности и полосе частот, для отображения символов, отбираются такие, которые в предположении воздействия на них аддитивного гауссова шума обеспечивают наибольшую достоверность приема отдельного символа. Одновременно определяется и структура оптимального приемника. Наиболее полно эти вопросы рассмотрены для случая двоичного канала ($m=2$)

Увеличение эффективности и помехоустойчивости системы передачи информации, как показал Шеннон, возможно за счет введения в канал связи кодирующего, а следовательно и декодирующего устройств, цел которых в статистическом согласовании свойств источника сообщений и канала связи.

Доказанными им теоремами обоснованно существование оптимального способа кодирования, при котором достигается скорость передачи информации, сколь угодно близкая к пропускной способности данного канала связи. Под способом кодирования при этом подразумевается совокупность операций по преобразованию сообщений в сигналы и обратного преобразования смеси сигнала с помехами в сообщения, включая операции в части канала <модулятор - демодулятор>.

К сожалению, указанные теоремы не дают конструктивных рекомендаций относительно путей реализации оптимального способа кодирования. Определить соответствующую совокупность операций, а следовательно, и структуру оптимальной системы связи пока не удалось даже при ряде допущений, существенно упрощающих модели каналов. Для упрощения задачи переходят к оптимизации системы по частям путем нахождения наилучшего кода при условии оптимально спроектированной части канала <модулятор - демодулятор>.

Выяснить также целесообразность разделения процедур кодирования, обусловленных статистическими свойствами источника сообщений, и процедур кодирования, зависящих от статистических свойств канала связи. Такое разделение способствует лучшему пониманию существа процессов преобразования. С практической точки зрения оно ценно тем, что позволяет реализовать как кодирующие, так и декодирующие устройства из двух из двух фактически независимых блоков: кодера источника (КИ) и декодера источника (ДКИ), кодера канала (КК) и декодера канала (ДКК).

Рассмотрим теперь особенности статистического согласования различных источников сообщений и каналов связи.

Предположим, что дискретные сообщения, поступающие с источника, обладают избыточностью, а вредным действием помех в канале можно пренебречь, что будет близко к реальности при отношении сигнал/помеха, значительно превышающим единицу. В этом случае учитывать проблему обеспечения помехоустойчивости нет необходимости и остается задача повышения эффективности.

В основной теореме Шеннона о кодировании для дискретного канала без помех утверждается, что посредством преобразования сообщений в статистической независимые и равновероятные символы можно повысить скорость передачи вплоть до пропускной способности этого канала (подробно об этом будем рассматривать позже).

Техническая реализация указанной возможности осуществляется кодером источника, обеспечивающим такое кодирование, при котором за счет устранения избыточности снижается среднее число символов, требующихся для выражения знака сообщения. При отсутствии помех это непосредственно дает выигрыш во времени передачи (или в объеме запоминающего устройства), что повышает эффективность системы. Поэтому такое кодирование получило название эффективного или оптимального.

При наличии помех в канале оно позволяет преобразовать входную информацию в последовательность символов, наилучшим образом (в смысле максимального сжатия) подготовленную для дальнейших преобразований.

Кодирование информации по каналу без помех

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 .

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 $(H(x))$:

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

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:

$$l_{эфф}=\frac{H(x)}{H_1}=\frac{\log_2 m}{\log_2 k}=\frac{\log_2 k^n}{\log_2 k}=n,$$

где $m = k^n$.

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:

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

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 $V_{ср.} > H$ 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.

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.

Алгоритм Шеннона-Фано. 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 1 иллюстрирует процесс построения кода Шеннона–Фано на примере источника сообщений, алфавит которого состоит из восьми символов.

Таблица 1

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	II	10
3	m_3	1/8	III	110
4	m_4	1/16	IV	1110
5	m_5	1/32	V	11110
6	m_6	1/64	VI	111110
7	m_7	1/128	VII	1111110
8	m_8	1/256		11111110

На рис. 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.

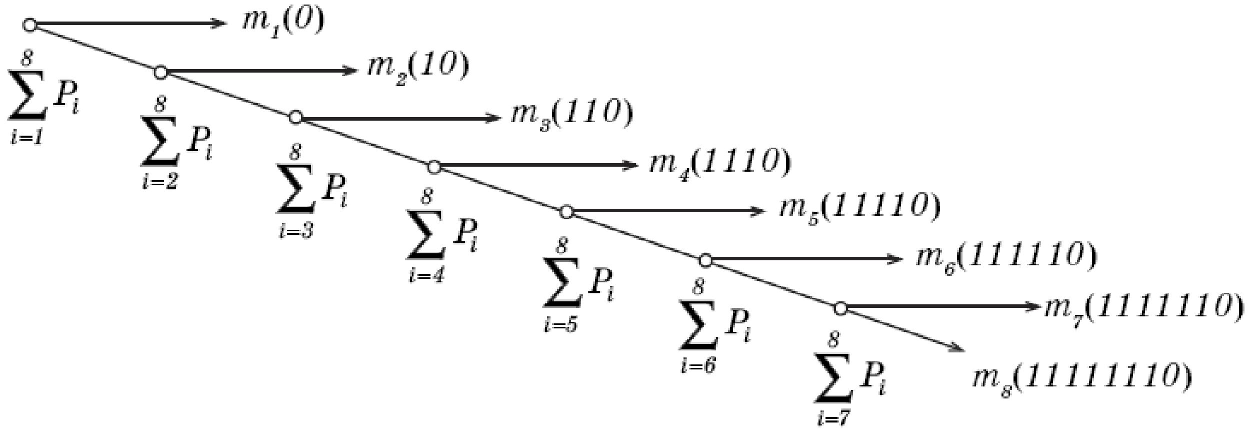


Рис 1. Граф кодирования по алгоритму Шеннона–Фано.

Алгоритм Шеннона–Фано 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–Фано.

В качестве примера рассмотрим предложенный выше (Табл. 1) 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:

$$H_u = -\sum_{i=1}^{8} P_i \log_2 P_i$$ (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:

$$R_u = 1 - \frac{H_u}{H_{u\max}} = 0{,}3.$$

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

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

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

$$R_{рк} = 1 - \frac{H_u}{n_p \log_2 2} \approx 0{,}3.$$

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

$$H_{1P} = \frac{H_u}{n_p} \approx 0{,}6 \text{ (бит / элемент символа кода).}$$

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

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

$$n_H = \sum_{i=1}^{m} n_i \cdot P_i = 2,$$

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

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

$$R_{HK} = 1 - \frac{H_u}{n_H \log_2 2} \approx 0{,}0$$

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

$$H_{1H} = \frac{H_u}{n_H} = 0{,}9 \text{ (бит/элемент символа кода).}$$

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.

Таблица 2.

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

Среднее число элементов символов кода на один символ исходного сообщения, 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 3.

Таблица 3.

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.

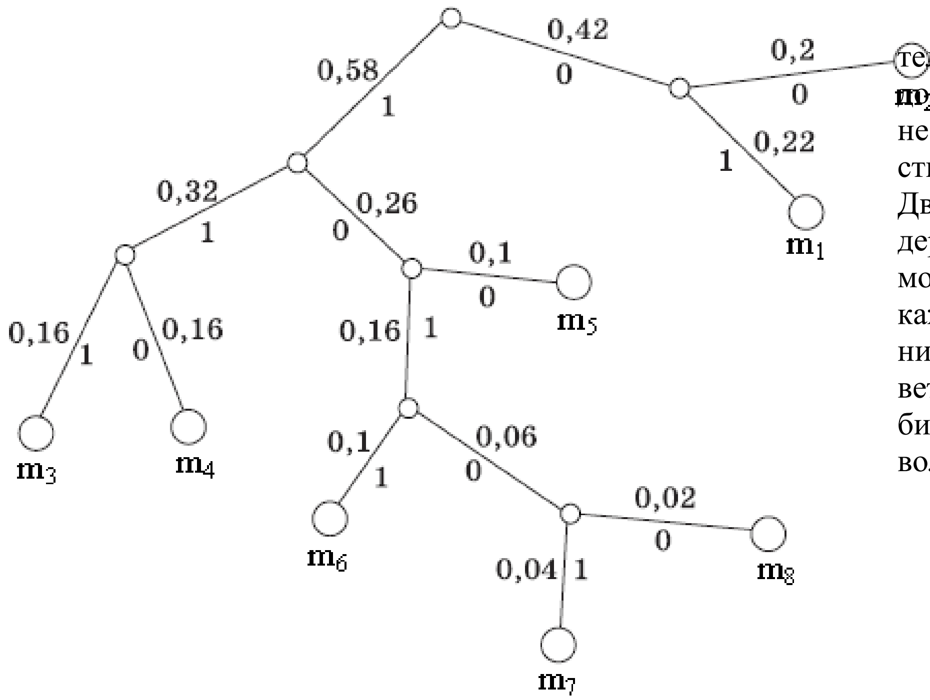
Алгоритм Хаффмена. Суть этого алгоритма, при использовании двоичного кода, состоит в следующем. Все символы алфавита источника сообщений ранжируют, т.е. выписывают в столбец в порядке убывания вероятностей их появления. Два последних символа объединяют в один вспомогательный символ, которому приписывают 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 4.

Таблица 4

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 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. Граф кодирования по алгоритму Хаффмена .

Различные символы, генерируемые источником сообщения, и соответствующие им кодовые символы представлены в таблице 5.

Таблица 5.

m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	00	111	110	100	1011	1010	10

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.

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 рисунке, где x_i - символы сообщения.

Рис. 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.

Кодирование информации по каналу с помехами

В реальных условиях прием двоичных символов всегда происходит с ошибками из-за действия помех. В этом случае вместо символа 1 принимается 0 и наоборот. Ошибки могут возникать из-за помех, действующих в канале связи, из-за изменений характеристик канала связи (например, из-за замераний), из-за снижения уровня мощности ПРД, из-за нестабильности АЧХ и ФЧХ канала. В дискретных каналах общепринятым критерием оценки качества передачи является допустимая вероятность ошибки приема, нормированная на символ. Например, при телеграфной передаче эта величина составляет 10^{-3} на знак, при передаче данных – порядка 10^{-6}, при записи на CD-диск - 10^{-16}, в спутниковых каналах - 10^{-8}. Для обеспечения таких значений вероятностей используются различные меры, основной из которых служит повышение качества приема информации. Эти методы можно разбить на две группы:

1. Методы увеличения помехоустойчивости приема дискретной информации, связанные с выбором уровня сигнала, увеличения отношения $\frac{сигнал}{помеха + шум}$, увеличения ширины полосы пропускания канала и т. д. Все это – энергетические характеристики.

2. Методы обнаружения и исправления ошибок, основанные на искуственном введении избыточности в передаваемом сообщении. Избыточность передаваемого сообщения (сигнала) можно увеличить различными способами.

Последний метод базируется на том факте, что объем сигнала:

$$V_c = P_c * \Delta F * T .$$

Возможность увеличения мощности и полосы сигнала ограничена, поэтому идут на увеличение третьего сомножителя. Увеличение длительности сигнала осуществляется следующими способами:

a) многократными повторениями кодовых комбинаций;

b) одновременной передачей кодовых комбинации по нескольким каналам на разных частотах;

c) помехоустойчивым кодированием.

Здесь в (c) избыточность используется наиболее эффективно. Введение избыточности при использовании помехоустойчивых кодов обязательно связано с увеличением числа разрядов кодовой комбинации. Самыми распространенными помехоустойчивыми кодами являются так называемые разделимые коды, в которых кодовые комбинации состоят из двух частей: информационной и проверочной. Как правило, такие коды имеют кодовые комбинации, содержащие n символов, из которых первые k символов являются информационными, а за ними располагаются n - k проверочных символов. Эти коды обозначаются (n,k).

Основные характеристики корректирующих кодов (КК):

1. число разрешенных и запрещенных кодовых комбинаций;
2. избыточность;
3. минимальное кодовое расстояние (хэмминговое);
4. число обнаруживаемых и число исправляемых ошибок.

Число разрешенных и запрещенных кодовых комбинаций. Для блочных двоичных кодов, в которых информация передается словами (блоками) с числом символов равном n, общее число возможных кодовых комбинаций определяется:

$$N_0 = 2^n$$

Кодовые комбинации, используемые при кодировании, называются разрешенными, все другие комбинации называются запрещенными. Очевидно, что число разрешенных кодовых комбинаций при наличии k кодовых разрядов равно:

$$N_k = 2^k .$$

Поэтому число запрещенных кодовых комбинаций определяется:

$$N_з = N_0 - N_k = 2^n - 2^k .$$

Если кодовая комбинация на выходе канала связи оказывается запрещенной, то это указывает на помеху при передаче.

Избыточность корректирующего кода:

$$R_k \equiv \frac{r}{n} = \frac{n-k}{n} = 1 - \frac{k}{n} .$$

Минимальное кодовое расстояние d_0. Для того, чтобы можно было обнаруживать и исправлять ошибки, разрешенная кодовая комбинация должна как можно больше отличаться от запрещенной. Если ошибки в канале связи возникают независимо, то вероятность преобразования одной кодовой комбинации в другую будет тем меньше, чем большим числом символов они различаются.

Кодовым расстоянием (КР) называется количество единиц в сумме двух кодовых комбинаций по модулю 2.

Пример.

Сумма по модулю 2 кодовых пятиразрядных слов:

$01011 \oplus 10010 = 11001$,

Получено 3 единицы, следовательно, кодовое расстояние d = 3.

Минимальное КР d_0 - наименьшее из КР, найденных по всем разрешенным кодовым комбинациям при попарном их сравнении. В безызбыточном коде все комбинации являются разрешенными, поэтому d_0= 1. Например, если n = 3, $N_0 = 2^3 = 8$, кодовые комбинации:

000, 001, 010, 011, 100, 101, 110, 111.

У такого кода любая разрешенная кодовая комбинация при искажении одного символа переходит в другую разрешенную комбинацию. Такой код называется примитивным (или первичным), он не обладает корректирующей способностью. Для того чтобы код обладал корректирующими свойствами, необходимо ввести избыточность, которая увеличила бы d от двух и выше. Таким образом, минимальное КР – важнейшая характеристика КК.

Число обнаруживаемых и число исправляемых ошибок. При применении двоичных кодов учитывают только дискретные искажения, при которых 1 переходит в 0 или наоборот. Такой переход в одном элементе кодовой комбинации называется единичной ошибкой. В общем число позиций (элементов) кодовой комбинации, на которых под действием помехи одни символы оказались замененными на другие, называют кратностью ошибки g, 0 ≤ g ≤ n. Таким образом, фактически кратность ошибки -хэмминговое расстояние между переданной и принятой кодовой комбинацией:

$000 \xrightarrow[g=2]{} 110$.

Если d_0=2, то ни одна из разрешенных кодовых комбинаций при одиночной ошибке (g = 1) не переходит в другую разрешенную кодовую комбинацию.

Например, для трехразрядного кода, приведенного выше, можно организовать подмножество разрешенных кодовых комбинаций по принципу четности в них единиц.

000, 011, 101, 110 - разрешенные комбинации,

001, 010, 100, 111 - запрещенные комбинации.

В этом случае d_0= 2, такой код обнаруживает только одиночные ошибки и другие ошибки с нечетной кратностью.

В общем случае, при необходимости обнаруживать ошибки кратностью до g включительно хэмминговое расстояние d_0 должно быть по крайней мере на единицу больше, чем g, т. е.

$d_0 \geq g_0 + 1$,

где g_0 - кратность обнаруживаемой ошибки.

Теорема 1 (характеризующая обнаружительную способность кода). Если код имеет хэмминговое расстояние $d_0 > 1$, и используется декодирование по методу обнаружения ошибки, то все ошибки кратностью $g_0 < d_0$ обнаруживаются гарантированно, а что касается ошибок кратностью $g \geq d_0$, то одни из них обнаруживаются, другие – нет. Таким образом, условие обнаружения ошибок кратностью g_0 формулируется:

$g_0 \leq d_0 - 1$.

Рассмотрим процедуру исправления ошибок на примере симметричного канала передачи информации без памяти (модель рассматривалась ранее в разделе ПС дискретного канала). В этом случае оптимальным является правило декодирования по наименьшему кодовому расстоянию. Это эквивалентно максимизации функции правдоподобия. Согласно этому критерию, запрещенную кодовую комбинацию декодируют как ту разрешенную, которая находится на наименьшем кодовом расстоянии от нее.

Декодер Виттерби. Исправляющая способность при этом правильном декодировании определяется соотношением:

$d_0 \geq 2 g_u + 1$,

где g_u - кратность исправляемых ошибок. Корректирующий код подбирается под канал, в котором он используется.

2 Теорема (о исправляющей способности кода). Если код имеет хэмминговое расстояние d_0 > 2, и используется декодирование с исправлением ошибок по наименьшему кодовому расстоянию (алгоритм максимального правдоподобия им. Виттерби), то все ошибки кратностью $g_u < \frac{d_0}{2}$ исправляются гарантированно. Что касается ошибок большей кратности, то после исправления кодовая комбинация может приобрести вид, который не соответствует исходной кодовой комбинации. Таким образом, сама процедура исправления выполнена. Окончательно:

$$g_u \le \frac{d_0 - 1}{2}.$$

Из 1 и 2 Теорем следует, что кратность обнаруживаемых ошибок $g_o < d_0$, кратность исправляемых ошибок $g_u < \frac{d_0}{2}$. Это означает, что, если код исправляет ошибки кратности g_u, то он может обнаружить количество ошибок в два раза больше, чем g_u, т. е. $g_o = 2g_u$.

Корректирующие возможности кодов. Вопрос о максимально необходимой избыточности, при которой код обладает необходимыми корректирующими свойствами, является одним из важнейших в теории кодирования, и не получил до сих пор полного решения. Получен лишь ряд верхних и нижних границ (оценок), которые устанавливают связь между максимально возможным хэмминговым расстоянием и избыточностью кода.

Верхняя граница Плоткина

$$d_0 \le \frac{n * 2^{k-1}}{2^k - 1}$$

дает выражения для максимального значения d_0.

Верхняя граница Хэмминга

$$2^k \le \frac{2^n}{\sum_{i=0}^{(d_o-1)/2} C_k^i}$$

устанавливает максимальное число разрешенных кодовых комбинаций при заданных n и d_0.

В частности, для кодов Хэмминга, обладающими кодовыми расстояниями 3 и 4, справедливы соотношения: $d_0 = 3 \rightarrow$ r $\ge \log_2$(n+1), $d_0 = 4 \rightarrow$ r $\ge \log_2 2$*n,

r – число проверочных символов, n – число информационных символов.

Граница Варшамова-Гильберта определяет нижнюю границу для числа проверочных разрядов r = n – k в случае кодов большой разрядности, необходимого для обеспечения заданного кодового расстояния d: $r \ge \log_2 \sum_{k=0}^{d-2} C_{n-k}^i$.

Определены оценки, которые дают представление о верхней границе d_0 при фиксированных n и k и оценка снизу для числа проверочных символов r при заданных k и d.

Корректирующие коды Хэмминга

Построение этих кодов базируется на принципе проверки на четность числа единиц в информационной группе кодового блока.

Поясним эту идею на примере простейшего корректирующего кода, называемого кодом с проверкой на четность (иначе с контролем по паритету). В таком коде к кодовым комбинациям безызбыточного первичного k-разрядного кода добавляется один дополнительный разряд (проверочный символ для проверки на четность). Если число единиц исходного кодового сообщения четное, то в дополнительном разряде формируют проверочный символ 0, а если нечетное, то в дополнительный разряд пишут 1. В результате общее количество символов 1 в любой передаваемой кодовой комбинации всегда будет четным. Таким образом, правило формирования проверочного символа строится следующим образом: $r_1 = i_1 \oplus i_2 \oplus ... \oplus i_k$,

где i – информационный символ, k – число информационных символов.

Добавление дополнительного разряда увеличивает общее число возможных комбинаций вдвое по сравнению с числом комбинаций исходного первичного кода, а условие четности разделяет все комбинации на разрешенные и запрещенные. Следовательно, код с проверкой на четность позволяет обнаруживать одиночную ошибку при приеме, т. к. она нарушает условие четности, переводя разрешенную комбинацию в запрещенную. Критерием правильности принятой кодовой комбинации является равенство нулю результата суммирования S по модулю 2 всех n символов кода, включая проверочный символ r_1. При наличии одиночной ошибки величина S принимает следующее значение:

S = $r_1 \oplus i_1 \oplus i_2 \oplus ... \oplus i_k$,

S – синдром , локатор ошибки. При S = 0 ошибки нет.

Этот код описывается согласно нашей терминологии: (n,k) = (n,n - 1). Так как d_0 = 2, то может быть обнаружена однократная ошибка, исправление производиться не может. Код с проверкой на четность может быть использован только для обнаружения однократных ошибок. Увеличивая число дополнительных проверочных символов и формируя их по определенным правилам, равными 0 или 1, можно усилить корректирующие свойства кода и позволить ему исправлять ошибки. На этом основано построение кодов Хэмминга.

Рассмотрим коды Хэмминга для случая исправления одиночной ошибки. В этом случае для каждого целого существует n = $2^r - 1$, k = $2^r - 1 - r$ разрядов.

Например, при r = 3 код Хэмминга имеет вид: (7,4). В этом случае четыре информационных символа дополняются тремя проверочными:

$(i_1, i_2, i_3, i_4) - (r_1, r_2, r_3)$.

Алгоритм кодирования Хэмминга, т. е. правило получения проверочных символов из информационных, имеет следующий вид:

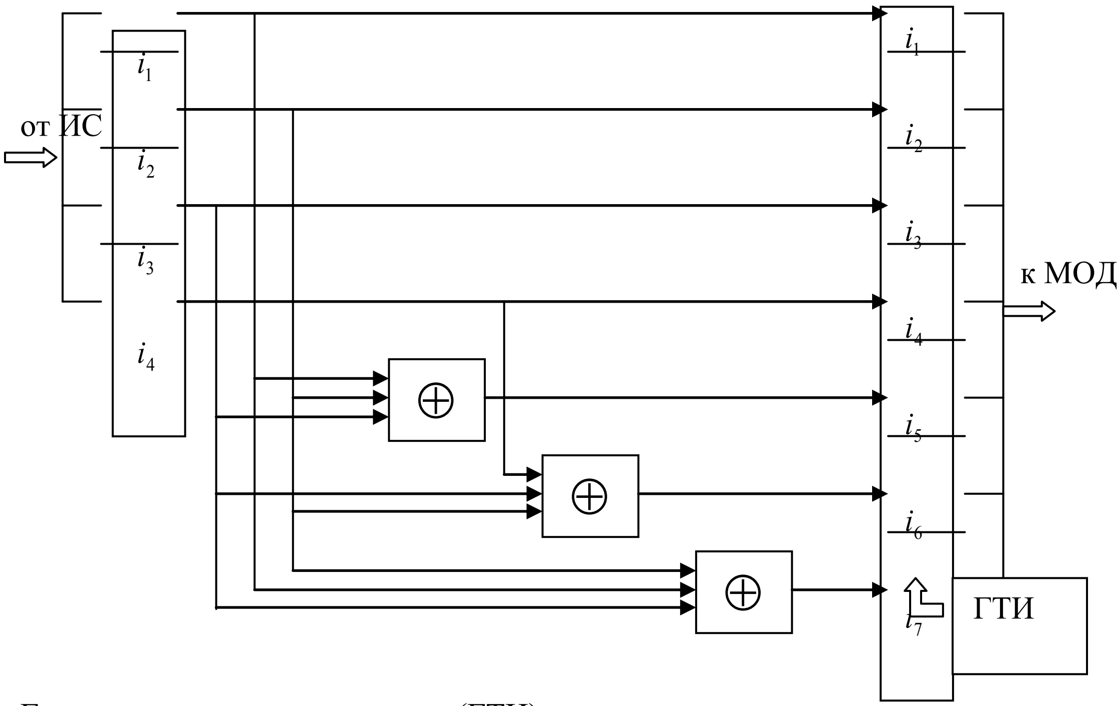
$$\begin{cases} r_1 = i_1 \oplus i_2 \oplus i_3, \\ r_2 = i_2 \oplus i_3 \oplus i_4, \\ r_3 = i_1 \oplus i_2 \oplus i_4. \end{cases}$$

Таким образом, существует $2^4 = 16$ всевозможных кодовых слов в коде x(7,4).

Номер перестановки	i_1	i_2	i_3	i_4	r_1	r_2	r_3
1	0	0	0	0	0	0	0
2	0	0	0	1	0	1	1
3	0	0	1	0	1	1	0
4	0	0	1	1	1	0	1
5	0	1	0	0	1	1	1
6	0	1	0	1	1	0	0
7	0	1	1	0	0	0	1
8	0	1	1	1	0	1	0
9	1	0	0	0	1	0	1
10	1	0	0	1	1	1	0
11	1	0	1	0	0	1	1
12	1	0	1	1	0	0	0
13	1	1	0	0	0	1	0
14	1	1	0	1	0	0	1
15	1	1	1	0	1	0	0
16	1	1	1	1	1	1	1

Принцип построения этой таблицы может быть технически реализован с помощью нижеприведенной схемы кодера Хэмминга (7,4).

4-символьное информационное слово 7-символьное кодовое слово



Генератор тактовых импульсов (ГТИ) проталкивает кодовые слова к модулятору.

Декодер Хэмминга. На вход декодера поступает искаженное кодовое слово

V = $(i_1', i_2', i_3', i_4', r_1', r_2', r_3')$.

Далее в декодере в режиме исправления ошибок строится трехсимвольная последовательность $\vec{S} = (S_1, S_2, S_3)$ - синдром ошибок:

$$\begin{cases} S_1 = r_1' \oplus i_1' \oplus i_2' \oplus i_3', \\ S_2 = r_2' \oplus i_2' \oplus i_3' \oplus i_4', \\ S_3 = r_3' \oplus i_1' \oplus i_2' \oplus i_4'. \end{cases}$$

Этот синдром представляет собой сочетание результатов проверки на четность соответствующих символов кодовой группы и характеризует определенную конфигурацию ошибок. При r = 3 имеется $2^3 = 8$ всевозможных синдромов, при этом $\vec{S} = (0,0,0)$ указывает на отсутсвие ошибок при приеме.

Всякому ненулевому синдрому $\vec{S}$ соответствует определенная конфигурация ошибок, которая и исправляется далее в декодере. Например, для (7,4) ненулевые синдромы и соответствующие конфигурации ошибок имеют следующий вид:

Синдром	001	010	011	100	101	110	111
Конфигурация ошибок	0000001	0000010	0001000	0000100	1000000	0010000	0100000
Ошибка в символе	r_3	r_2	i_4	r_1	i_1	i_3	i_2

Каждая из ошибок имеет свой единственный синдром. При технической реализации декодера возможно создание такого цифрового корректора ошибок (дешифратора синдрома), который по соответствующему синдрому исправляет соответствующий символ в принятой кодовой комбинации. После внесения исправлений проверочные символы можно не выводить на выход декодера.

Циклические коды

Циклический код – такой групповой код, все базовые комбинации которого могут быть получены из одной путем циклического сдвига ее элементов.

Циклический сдвиг кодовой комбинации – перемещение ее элементов справа налево без нарушения порядка их следования, так что крайний левый элемент занимает место крайнего правого.

В теории циклических кодов принято записывать кодовые комбинации в виде полинома некоторой фиктивной переменной x:

$$C_1a_1 \oplus C_2a_2 \oplus C_3a_3 \oplus \ldots \oplus C_ia_i \oplus \ldots \oplus C_qa_q \neq 0$$

где ai – значение символа кодовой комбинации на позиции i при нумерации справа налево;

xi – 1 – фиктивная переменная в степени номера позиции i без единицы.

Пример

Представить в виде полинома кодовую комбинацию 1011101.

$$a(x) = a^7x^{7-1} + a^6x^{6-1} + a^5x^{5-1} + a^4x^{4-1} + a^3x^{3-1} + a^2x^{2-1} + a^1x^{1-1} =$$

$$= a^7x^6 + a^6x^5 + a^5x^4 + a^4x^3 + a^3x^2 + a^2x + a^1 =$$

$$= 1\cdot x^6 + 0\cdot x^5 + 1\cdot x^4 + 1\cdot x^3 + 1\cdot x^2 + 0\cdot x + 1 =$$

$$= x^6 + x^4 + x^3 + x^2 + 1.$$

Максимальная степень x в полиноме на единицу меньше числа элементов в кодовой комбинации.

Запись комбинации в виде полинома понадобилась для того, чтобы отобразить формализованным способом операцию циклического сдвига исходной кодовой комбинации. Допустим, задана исходная кодовая комбинация и соответствующий ей полином

$$a(x) = a^nx^{n-1} + a^{n-1}x^{n-2} + \ldots + a^ix^{i-1} + \ldots + a^2x + a^1.$$

Умножим a(x) на x:

$$a(x)\cdot x = a^nx^n + a^{n-1}x^{n-1} + \ldots + a^ix^i + \ldots + a^2x^2 + a^1x$$

Так как максимальная степень x в кодовой комбинации длиной n не превышает (n – 1), то из правой части полученного выражения для получения исходного полинома необходимо вычесть an(xn – 1). Вычитание an(xn – 1) называется взятием остатка по модулю (xn – 1). Сдвиг исходной комбинации на i тактов можно представить следующим образом: a(x)xi – an(xn – 1), то есть умножением a(x) на xi и взятием остатка по модулю (xn – 1). Взятие остатка необходимо при получении многочлена степени, большей или равной n.

Возвращаясь к определению циклического кода и учитывая запись операций циклического сдвига кодовых комбинаций, можно записать порождающую матрицу циклического кода в следующем виде:

$$V = \begin{bmatrix} p(x) \\ p(x)\cdot x - C_2\left(x^n - 1\right) \\ p(x)\cdot x^2 - C_3\left(x^n - 1\right) \\ \ldots \\ p(x)\cdot x^i - C_{i+1}\left(x^n - 1\right) \\ \ldots \\ p(x)\cdot x^{m-1} - C_m\left(x^n - 1\right) \end{bmatrix},$$

где p(x)– исходная кодовая комбинация, на базе которой получены все остальные (m – 1) базовые комбинации;

Ci = 0 или Ci = 1 (“0”, если результирующая степень полинома p(x)xi не превосходит (n – 1), “1”, если превосходит).

Комбинация *p(x)* называется порождающей (образующей, генераторной) комбинацией. Для построения циклического кода достаточно верно выбрать *p(x)*. Затем все остальные кодовые комбинации получаются такими же, как и в групповом коде.

Порождающий полином должен удовлетворять следующим требованиям:

- p(x) должен быть ненулевым;
- вес p(x) не должен быть меньше минимального кодового расстояния: $\lambda[p(x)] \geq d_{\min}$;
- p(x) должен иметь максимальную степень k (k – число избыточных элементов в коде);
- p(x) должен быть делителем полинома (xn – 1).

Выполнение условия 4 приводит к тому, что все рабочие кодовые комбинации циклического кода приобретают свойство делимости на *p(x)* без остатка. Учитывая это, можно дать другое определение циклического кода.

Циклический код – код, все рабочие комбинации которого делятся на порождающую без остатка.

Свойство делимости всех комбинаций на *p(x)* позволяет просто записать порождающую матрицу кода в каноническом виде:

$$V_{KAH} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & c_1^1 & c_1^2 & \dots & c_1^k \\ 0 & 1 & 0 & \dots & 0 & c_2^1 & c_2^2 & \dots & c_2^k \\ 0 & 0 & 1 & \dots & 0 & c_3^1 & c_3^2 & \dots & c_3^k \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & \dots & 1 & c_m^1 & c_m^2 & \dots & c_m^k \end{bmatrix} = \begin{bmatrix} E_m & C_{m\times k} \end{bmatrix}.$$

Для определения строк контрольной подматрицы $C_{m\times k}$поступают следующим образом:

- любая строка единичной подматрицы записывается в виде полинома: $E_i(x) = x_m$-i;
- справа к ней приписывается *k* нулей, что равносильно умножению на x_k: $E_i(x)x_k$;
- результат делится на порождающий полином *p(x):* $E_i(x)\cdot x^k = \xi(x)\cdot p(x) + r_i(x)$; при этом остаток $r_i(x)$ имеет степень не выше (k – 1) и содержит *k* элементов;
- рабочая комбинация циклического кода состоит из *m* элементов единичной подматрицы и из *k* элементов остатка $r_i(x)$ (он приписывается в “чистом” виде): $E_i(x)\cdot x^k \oplus r_i(x) = \xi(x)\cdot p(x)$

В настоящее время для облегчения процедуры построения циклических кодов их авторами найдены различные порождающие полиномы *p(x)* в зависимости от требований к коду. В частности, существуют таблицы с полиномами *p(x)* для циклических кодов с $d_{\min} = 3$ (коды с $d_{\min} = 3$ наиболее часто применяются на практике). Их можно найти в литературе по теории информации. Для построения кодов с $d_{\min} = 4$ достаточно умножить выбранный полином *p(x),* найденный в таблице порождающих полиномов кодов с $d_{min} = 3$, на (x + 1), что равносильно проверке на общую четность.

Суммируя изложенное, приведем процедуру выбора *p(x):*

определить число информационных элементов *m* (M = 2m) и число избыточных элементов *k* (если $d_{\min} = 3$, 2k = m + k + 1);

найти *p(x)* степени *k* по таблице (если полиномов этой степени несколько, то выбирается любой).

Разработаны циклические коды, обеспечивающие произвольное минимальное кодовое расстояние $d_{\min} = 5$. Они получили название БЧХ (Боуза-Чоудхури-Хоквингема, по имени разработчиков). Порождающие полиномы для таких кодов в зависимости от предъявляемых к ним требований, можно найти в таблице.

k	n	m	s	d_{min}	Образующий полином	
					Символическая запись	Запись в виде полинома
3	7	4	1	3	13	1011
4	15	11	1	3	23	10011
8	15	7	1	3	721	111010001
10	15	5	3	7	2467	10100110111
5	31	26	1	3	45	100101
10	31	21	2	5	3551	11101101001
15	31	16	3	7	107657	1000111110101111
25	31	11	5	11	5423325	101100010011011010101

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ К ЛАБОРАТОРНЫМ РАБОТАМ

Лабораторные работы имеют различный уровень сложности и на их выполнение требуется различное количество часов. Каждая предполагает самостоятельную работу студентов по освоению лекций и теоретического материала, вынесенного на самостоятельное изучение. Текущий контроль знаний осуществляется путем опроса студентов перед началом лабораторного занятия по вопросам, перечень которых приведен в каждой лабораторной работе.

Лабораторное занятие 1. Энтропия

ЗАДАНИЕ

1. Ознакомиться с теоретической частью, используя дополнительную литературу.
2. Исходя из полученных у преподавателя исходных данных (количества передаваемых сообщений N), рассчитать энтропию передаваемого сообщени.
3. Провести программный контроль выполнения пункта 2 на примере исходных данных полученных у преподавателя.
4. Отчет.

Контрольные вопросы

1.Дать определение энтропии.

2.Запишите формулу Шеннона и формулу Хартли.

3.Перечислите основные свойства энтропии.

4. Что является единицей измерения энтропии?

5. В каких случаях энтропия равна нулю?

6. При каких условиях энтропия принимает максимальное значение?

7. В чем состоит правило сложения энтропий для независимых источников?

Лабораторное занятие 2. Скорость передачи информации и пропускная способность каналов связи

ЗАДАНИЕ

1. Ознакомиться с теоретической частью, используя дополнительную литературу.

2. Исходя из полученных у преподавателя исходных данных (количества передаваемых сообщений N), рассчитать скорость передачи и пропускную способность дискретного канала связи с шумами и непрерывного канала связи без шумов.

3. Провести программный контроль выполнения пункта 2 на примере исходных данных полученных у преподавателя.

4. Отчет.

Контрольные вопросы.

1. Сформулируйте теорему Шеннона для канала без помех.

2. Как отличается трактовка величины Hz для случаев "посимвольного" и "цепочечного" эффективного кодирования?

3. Почему при вероятности ошибки $p_0 = 1$ пропускная способность канала имеет ту же величину, что и при $p_0 = 0$? Как практически можно использовать такой канал?

4. В чем суть теоремы Шеннона для канала с помехами?

5. Как практически можно избежать потери информации в канале с помехами?

Лабораторное занятие 3. Построение оптимального кода по методу Шеннона-Фано и методу Хаффмена.

ЗАДАНИЕ

1. Получить у преподавателя задание и ознакомиться с ним.

2. Построить код по методу Шеннона-Фано и проверить его оптимальность.

3. Построить код по методу Хаффмена и кодовое дерево.

4. Провести программный контроль выполнения 2, 3 пунктов на примере случайных сообщений.

5. Подготовить отчет и сдать работу.

Контрольные вопросы.

1. Какой код называется оптимальным?

2. Способы построения оптимальных кодов.

3. В чем заключается сущность оптимального кодирования и практический результат его применения?

4. Как оценивается эффективность ОНК?

5. Какие коды называются оптимальными неравномерными кодами?

Лабораторное занятие 4. Построение двоичного группового кода (кода Хемминга).

ЗАДАНИЕ

1. Ознакомиться с теоретической частью, используя дополнительную литературу.

2. Построить код Хемминга по заданным исходным данным (число информационных разрядов k).

3. Составить систему уравнений кодирования для определения проверочных разрядов для кода Хемминга по пункту 2.

4. Провести программный контроль выполнения 2 и 3 пунктов на примере случайных кодовых комбинаций.

5. Отчет.

Контрольные вопросы.

1. На каких позициях проверочные символы в коде Хэмминга?

2. Что такое информационные и проверочные символы?

3. Какими графическими и геометрическими способами можно представить коды? Приведите пример.

4. Что такое кодовое расстояние, как оно определяется между двумя комбинациями двоичного кода?

5. Каким соотношением связаны информационные, проверочные символы и минимальное кодовое дерево?

Лабораторное занятие 5. Образование циклического кода

ЗАДАНИЕ

1. Получить у преподавателя задание и ознакомиться с ним.

2. Вычислить параметры кода *d, m, k, p, l, S*. Найти образующий многочлен, воспользовавшись таблицей неприводимых многочленов.

3. Проверить, имеются ли ошибки в исследуемой комбинации, при наличии ошибок - исправить их.

4. Провести программный контроль выполнения 4, 5 пунктов на примере случайных кодовых комбинаций.

5. Подготовить отчет и сдать работу.

Контрольные вопросы.

1. В чем заключаются основные идеи обнаружения и исправления ошибок циклическим кодом?

2. Что такое кодовое расстояние?

3. Чем отличается представление циклическим кодом для d = 3 и d = 5? где d - кодовое расстояние?

4. Какие существуют способы формирования комбинаций циклического кода?

5. В чем достоинство циклических кодов?

6. Что такое транспонированная матрица для циклического кода и ее размерность?

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ

Практическое занятие 1. Энтропия. Условная энтропия. Взаимная энтропия.

Задача 1. Чему равна условная энтропия и энтропия объединения, если канальная матрица имеет вид

$$p(b/a) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Задача 2. Канал связи, в котором передаются сигналы A1, A2, A3, A4, описан следующей канальной матрицей:

$$p(a,в) = \begin{vmatrix} 0.01 & 0.1 & 0.11 & 0.02 \\ 0.02 & 0.02 & 0.05 & 0.07 \\ 0.2 & 0.08 & 0.07 & 0.03 \\ 0.02 & 0.03 & 0.06 & 0.01 \end{vmatrix}$$

Найти долю информационных потерь, которые припадают на сигнал A2 при передаче сигналов A1 - A4 по данному каналу связи.

Задача 3. Определить пропускную способность дискретного канала связи, описанному матрицей

$$p(A,B) = \begin{vmatrix} 0.1 & 0 & 0 \\ 0.1 & 0.3 & 0 \\ 0 & 0.1 & 0.4 \end{vmatrix}$$

Задача 4. Определить все возможные информационные характеристики канала связи, в котором взаимосвязь источника с приемником может быть описана матрицей вида:

$$p(A,B) = \begin{vmatrix} 0.2 & 0.1 & 0 \\ 0.2 & 0 & 0.2 \\ 0.1 & 0.1 & 0.1 \end{vmatrix}$$

Задача 5. Вероятность появления сигналов на входе приемника сообщений равна соответственно рв1=0.2, рв2=0.3, рв3=0.5. Канал связи описан следующей канальной матрицей:

$$p(A/B) = \begin{vmatrix} 0.97 & 0 & 0.01 \\ 0.02 & 0.98 & 0.01 \\ 0.01 & 0.02 & 0.98 \end{vmatrix}$$

Определить информационную скорость.

Задача 6. В результате статистических испытаний канала связи получены следующие условные вероятности перехода одного сигнала в другой: $p(b_1/a_1)=0.85$, $p(b_2/a_1)=0.1$, $p(b_3/a_1)=0.05$,

$p(b_1/a_2)=0.09$, $p(b_2/a_2)=0.91$, $p(b_3/a_2)=0$, $p(b_1/a_3)=0.08$, $p(b_3/a_3)=0.92$. Построить канальную матрицу и определить общую условную и взаимную энтропию сообщений, передаваемых по данному каналу связи, если на выходе источника сигналы появились с равной вероятностью.

Практическое занятие 2. Количественная оценка информации.

Задача 1. Влияние помех в канале связи описывается следующим распределением условных вероятностей:

$$p(b/a)=\begin{vmatrix} 0.98 & 0.01 & 0.01 \\ 0.15 & 0.75 & 0.1 \\ 0.3 & 0.2 & 0.5 \end{vmatrix}$$

Вычислить количество информации, которое переносится одним символом сообщения при равновероятном появлении символов в сообщении. Вычислить количество информации в сообщении, составленном из 400 букв первичного алфавита.

Задача 2. Влияние помех в канале связи описывается следующим распределением условных вероятностей:

$$p(b/a)=\begin{vmatrix} 0.98 & 0.01 & 0.01 \\ 0.15 & 0.75 & 0.1 \\ 0.3 & 0.2 & 0.5 \end{vmatrix}$$

Вычислить количество информации, которое переносится одним символом сообщения при вероятностях $p(a1)=0.7$; $p(a2)=0.2$; $p(a3)=0.1$. Вычислить количество информации в сообщении, составленном из 400 букв первичного алфавита.

Задача 3. Определить количество информации при передаче k сообщений по каналу связи, описанному следующей канальной матрицей:

$$p(b/a)=\begin{vmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{vmatrix}$$

если на выходе источника сообщений символы встречаются с вероятностями $p(A1)=0.8$, $p(A2)=0.1$, $p(A3)=p(A4)=0.5$.

Задача 4. Взаимодействие двух систем A и B описывается следующей матрицей:

$$p(A,B)=\begin{vmatrix} 0.4 & 0.1 & 0 \\ 0 & 0.2 & 0.1 \\ 0 & 0 & 0.2 \end{vmatrix}$$

Определить безусловную энтропию системы A и системы B, и H(a /b).

Задача 5. В сообщении, составленном из пяти качественных признаков, последние используются с разной частотой, т.е. вероятности их различны и равны соответственно $p1=0.8$, $p2=0.15$, $p3=0.03$, $p4=0.015$ и $p5=0.005$. Всего в сообщении принято 20 знаков. Определить количества информации на букву сообщения и во всем сообщении. Каково было бы количество информации в данном сообщении, если бы все признаки имели равную вероятность?

Задача 6. Определить частную условную энтропию относительно каждого символа источника сообщений при передаче по каналу связи, описанному следующей канальной матрицей:

$$p(A,B)=\begin{vmatrix} 0.2 & 0 & 0 \\ 0.1 & 0.2 & 0 \\ 0 & 0.1 & 0.4 \end{vmatrix}$$

Практическое занятие 3. Построение оптимального кода по методу Шеннона-Фано и методу Хаффмена.

Задача 1. Построить код для передачи сообщений, составленных из алфавита с распределением вероятностей: $A_1=0{,}18$; $A_2=0{,}18$; $A_3=0{,}18$; $A_4=0{,}18$; $A_5=0{,}1$; $A_6=0{,}09$; $A_7=0{,}09$. Построение провести по методу Шеннона-Фано.

Задача 2. Построить ОНК для передачи сообщений, составленных из некоторого условного словаря со следующими вероятностями появления слов в тексте: запятая – 0,37; товарищ – 0,13; свою – 0,125; в – 0,11; и – 0,08; труд – 0,06; бой - 0,05; хранить – 0,023; беззаветно – 0,002.

Задача 3. Построить код для передачи сообщений, составленных из алфавита с распределением вероятностей: A_1=0,18; A_2=0,18; A_3=0,18; A_4=0,18; A_5=0,1; A_6=0,09; A_7=0,09. Построение провести по методу Хаффмена с числом качественных признаков равное 3.

Задача 4. Построить ОНК по методу Хаффмена для вторичного алфавита с числом качественных признаков равное 3, если символы кодируемого алфавита имеют следующее распределение вероятностей: A_1=0,37; A_2=0,25; A_3=0,18; A_4=0,1; A_5=0,06; A_6=0,02; A_7=0,02.

Задача 5. Построить ОНК по методам Шеннона-Фано и Хаффмена, если символы источника сообщений появляются с вероятностями: $A_1=A_2=A_3=A_4$=0,19; $A_5=A_6=A_7$=0,08. Сравнить, насколько полученные коды близки к оптимальному.

Практическое занятие 4. Построение двоичного группового кода.

Задача 1. Пусть требуется передать 16 сообщений. Построить систематический код, исправляющий одну ошибку.

Задача 2. Требуется исправить любую одиночную ошибку при передаче комбинации 0101, т.е. $n_u=4$.

Задача 3. Построить матрицу для группового кода, способного исправлять одиночную ошибку при передаче 32 символов первичного алфавита. И показать процесс исправления ошибки.

Задача 4. Определить кодовое расстояние между двоичными векторами: 110011011; 100110010.

Задача 5. Какой вид имеет производящая матрица группового кода, оптимального с точки зрения минимума корректирующих разрядов при максимуме информационных разрядов, для использования его в системе телемеханики, проектируемой для передачи не менее 2000 различных сообщений.

Практическое занятие 5. Построение циклического кода по заданным характеристикам.

Задача 1. Используя метод образующих матриц, построить циклический код, исправляющий одинарные ошибки при передаче комбинаций четырехзначного двоичного кода на все сочетания.

Задача 2. Методом умножения образующего многочлена на многочлены четырехзначного двоичного кода на все сочетания построить циклический код.

Задача 3. Методом циклического сдвига строки образующей матрицы и зеркальной ее комбинации построить циклический код с $d_0=3$ и $n_u=5$.

Задача 4. Показать процесс исправления одиночной ошибки в принятой кодовой комбинации 1100101.

Задача 5. При помощи образующей матрицы, полученной в результате умножения единичной матрицы на образующий многочлен x^3+x+1, построить циклический код, исправляющий одиночную ошибку в любом из четырех информационных разрядов.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ

Самостоятельная работа студентов по дисциплине проводится с целью:

систематизации и закрепления полученных теоретических знаний и практических умений по дисциплине;

углубления и расширения теоретических знаний;

формирования умений использовать полученные знания в новых условиях;

развития познавательных и творческих способностей;

формирования самостоятельности мышления, способности к саморазвитию, самореализации.

В учебном процессе выделяют два вида самостоятельной работы – **аудиторную,** которая выполняется под руководством преподавателя, и **внеаудиторную**, которая выполняется по заданию преподавателя, но без его непосредственного участия в определенные сроки и с последующей проверкой результатов на занятиях.

Перед выполнением обучающимися внеаудиторной самостоятельной работы преподаватель проводит инструктаж по выполнению заданий, которые включают цель задания, его содержание, сроки выполнения, объем работы, основные требования к результатам работы, критерии оценки результатов внеаудиторной самостоятельной работы.

В качестве форм контроля внеаудиторной самостоятельной работы обучающихся используется тестирование, самоотчеты, контрольные работы.

Основные формы самостоятельной учебной работы:

работа над конспектом лекции: лекции – основной источник информации по многим предметам, позволяющий не только изучить материал, но и получить представление о наличии других источников, сопоставить разные взгляды на основные проблемы данного курса. Лекции предоставляют возможность «интерактивного» обучения, когда есть возможность задавать преподавателю вопросы и получать на них ответы. Поэтому имеет смысл находить время для хотя бы беглого просмотра информации по материалу лекций (учебники, справочники и пр.) и непонятные, а также дискуссионные моменты обсуждать с преподавателем, другими студентами;

подготовка к практическому занятию: производится, как правило, с использованием методических пособий, состоит в теоретической подготовке (особенно для семинаров) и выполнении практических заданий (решение задач, ответы на вопросы и т.д.).

доработка конспекта лекции с применением учебника, методической литературы, дополнительной литературы: этот вид самостоятельной работы студентов особенно важен в том случае, когда изучаемый предмет содержит много неоднозначно трактуемых вопросов, проблем. Тогда преподаватель заведомо не может успеть изложить различные точки зрения, и студент должен ознакомиться с ними по имеющейся литературе. Кроме того, рабочая программа дисциплины предполагает рассмотрение некоторых относительно несложных тем только во время самостоятельных занятий, без чтения лектором;

подбор, изучение, анализ и конспектирование рекомендованной литературы;

самостоятельное изучение отдельных тем, параграфов;

консультации по сложным, непонятным вопросам лекций, семинаров, зачетов;

подготовка к зачету: данная форма СРС может быть весьма разнообразной по своей сути, так как сам зачет бывает различным. Он проводится обычно по итогам семестра перед сессией в письменной или устной форме, причем преподаватель может включать в него вопросы как практических занятий, так и лекционных (что особенно уместно, когда по данному предмету не сдается экзамен). Главное отличие зачета от экзамена – почти всегда не пяти-, а двухбалльная система оценки (сдал – не сдал), что делает его получение несколько более простым делом. С другой стороны, порой процедура его сдачи достаточно сложна, а иногда применяется и пятибалльная оценка (так называемый зачет с оценкой). Таким образом, для сдачи зачета необходимо, прежде всего, выполнить все требования преподавателя, что предполагает знание этих требований. Нужно как можно раньше выяснить, какие вопросы предстоит готовить и каковы правила самой процедуры (учитывается ли посещаемость, надо ли пропущенные занятия отрабатывать, а если надо, то каким образом и т.д.). Практика показывает, что хорошее посещение занятий является почти полной гарантией получения зачета, так как тогда можно быть в курсе всех требований преподавателя. И, напротив, большое количество пропусков может осложнить жизнь даже сильному студенту. Кроме того, необходимо учитывать, что проблемы могут появиться при распространенном подходе студента к практическим занятиям, когда многие работают первые месяцы вполсилы, накапливая задолженности по выполнению рефератов, практических заданий, конспектов и пр., а перед сессией пытаются все это сделать за одну неделю. Старайтесь распределять силы равномерно по всей дистанции семестра, и тогда зачетная неделя перед сессией будет не самой напряженной, а самой разгрузочной;

подготовка к экзамену: один из самых ответственных видов самостоятельной работы, и в то же время возможность сэкономить большое количество времени в период сессии, если эту подготовку начинать заблаговременно. Одно из главных правил – представлять себе общую логику предмета, что достигается проработкой планов лекций, составлением опорных конспектов, схем, таблиц. Фактически основной вид полготовки к экзамену – «свертывание» большого объема информации в компактный вид, а также тренировка в ее «развертывании» (примеры к теории, выведение одних закономерностей из других и т.д.). Владение этими технологиями обеспечивает, пожалуй, более половины успеха. Тем более что преподаватель обычно замечает в течение семестра целенаправленную подготовку такого студента и может поощрить его тем или иным способом. Надо также правильно распределить силы, не только готовясь к самому экзамену, но и позаботившись о допуске к нему (часто это хорошее посещение занятий, выполнение в назначенный срок практических заданий, активность на семинарах). Наконец, необходимо выяснить условия проведении, самого экзаменационного испытания, использовав для этой цели прежде всего консультацию (хотя преподаватель обычно касается этой темы заранее): количество и характер вопросов, форма проведения (устно или письменно), возможность использовать при подготовке различные материалы и пособия (таблицы, схемы, тетради для практических занятий и т.д.).

ЛИТЕРАТУРА

1. Гуменюк А.С. Прикладная теория информации [Электронный ресурс]: учебное пособие/ А.С. Гуменюк, Н.Н. Поздниченко— Электрон. текстовые данные.— Омск: Омский государственный технический университет, 2015.— 189 с.— Режим доступа: http://www.iprbookshop.ru/58097.html.— ЭБС «IPRbooks»

2. Санников В.Г. Теория информации и кодирования [Электронный ресурс]: учебное пособие/ В.Г. Санников— Электрон. текстовые данные.— М.: Московский технический университет связи и информатики, 2015.— 95 с.— Режим доступа: http://www.iprbookshop.ru/61558.html.— ЭБС «IPRbooks»

3. Акулиничев Ю.П. Теория и техника передачи информации [Электронный ресурс]: учебное пособие/ Ю.П. Акулиничев, А.С. Бернагрдт— Электрон. текстовые данные.— Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2012.— 210 с.— Режим доступа: http://www.iprbookshop.ru/13984.html.— ЭБС «IPRbooks»

4. Чернышев А.Б. Теория информационных процессов и систем [Электронный ресурс]: учебное пособие/ А.Б. Чернышев, В.Ф. Антонов, Г.Б. Суюнова— Электрон. текстовые данные.— Ставрополь: Северо-Кавказский федеральный университет, 2015.— 169 с.— Режим доступа: http://www.iprbookshop.ru/63140.html.— ЭБС «IPRbooks»

5. Математические методы теории передачи сигналов. Часть 3 [Электронный ресурс]: учебное пособие/ — Электрон. текстовые данные.— М.: Московский технический университет связи и информатики, 2012.— 48 с.— Режим доступа: http://www.iprbookshop.ru/61494.html.— ЭБС «IPRbooks»

6. Балюкевич Э.Л. Теория информации [Электронный ресурс]: учебное пособие/ Э.Л. Балюкевич— Электрон. текстовые данные.— М.: Евразийский открытый институт, 2009.— 215 с.— Режим доступа: http://www.iprbookshop.ru/10863.html.— ЭБС «IPRbooks»

7. Гультяева Т.А. Основы теории информации и криптографии [Электронный ресурс]: конспект лекций/ Т.А. Гультяева— Электрон. текстовые данные.— Новосибирск: Новосибирский государственный технический университет, 2010.— 88 с.— Режим доступа: http://www.iprbookshop.ru/44987.html.— ЭБС «IPRbooks»

СОДЕРЖАНИЕ