

Інформація (повторення)

С терміном **інформація** пов'язані поняття:

- *Сигнал* представляє собою будь-який процес, несущий інформацію.
- *Сообщение* – це інформація, представлена в **определенной форме** і призначена для передачі.
- *Данные* – це інформація, представлена в **формалізованому вигляді** і призначена для обробки її технічними засобами.

Передача информации

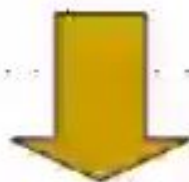
Источник информации – это субъект или объект, порождающий информацию и представляющий ее в виде сообщения.

Получатель информации – это субъект или объект, принимающий сообщение и способный правильно его интерпретировать.

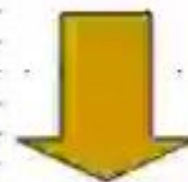
Сигнал – материальный носитель (субъект или объект), служащий для переноса информации от источника к потребителю посредством какой-нибудь среды (канала связи)



Канал связи — совокупность технических устройств, обеспечивающих передачу данных от **передатчика** к **приемнику**



Кодирующее
устройство



Декодирующее
устройство

Кодирующее устройство предназначено для кодирования информации.

Декодирующее устройство предназначено для преобразования полученного сообщения в исходное.



Схема передачи информации



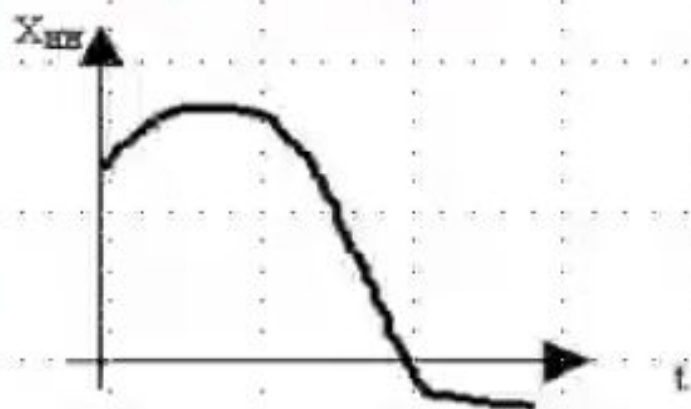
Виды сигналов

Сигнал называется *аналоговым* (или *непрерывным*), если его параметр может принимать *любое значение* в пределах некоторого промежутка (речь, музыка).

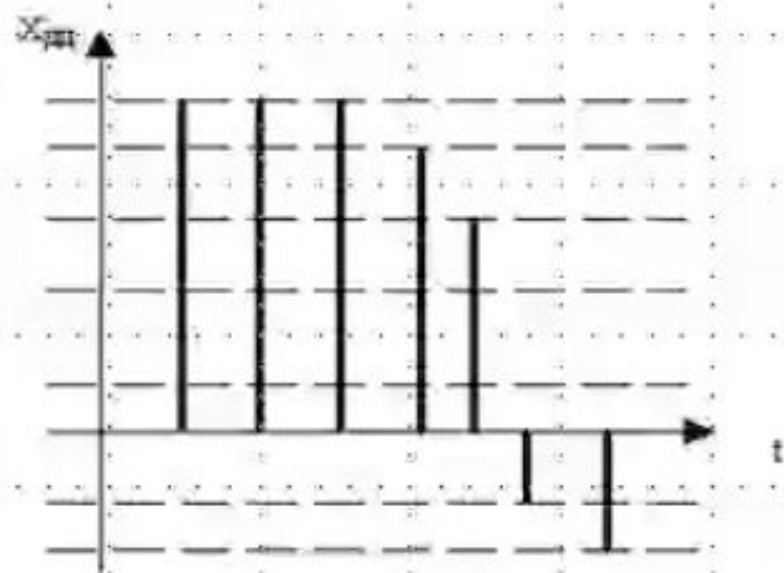
Сигнал называется *дискретным*, если его параметр может принимать *конечное число значений* в пределах некоторого промежутка (цифровое фото).

Следует различать непрерывность или дискретность сигнала *по уровню* и *по времени*

Примеры сигналов







Аналоговый по уровню и по времени сигнал X



Дискретный по времени сигнал X



Преобразования сигналов

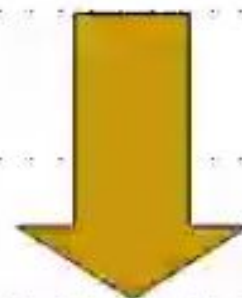
1. Непрерывный  Непрерывный
При таком преобразовании из-за помех, образуемых самим техническим устройством, всегда происходит потеря информации
2. Дискретный  Дискретный
Переход при представлении сигналов к другому алфавиту
3. Дискретный  Непрерывный
Теорема отсчетов дает ответ на вопрос о возможности преобразования без потерь
4. Непрерывный  Дискретный
Два способа квантования



Дискретизация непрерывного сигнала



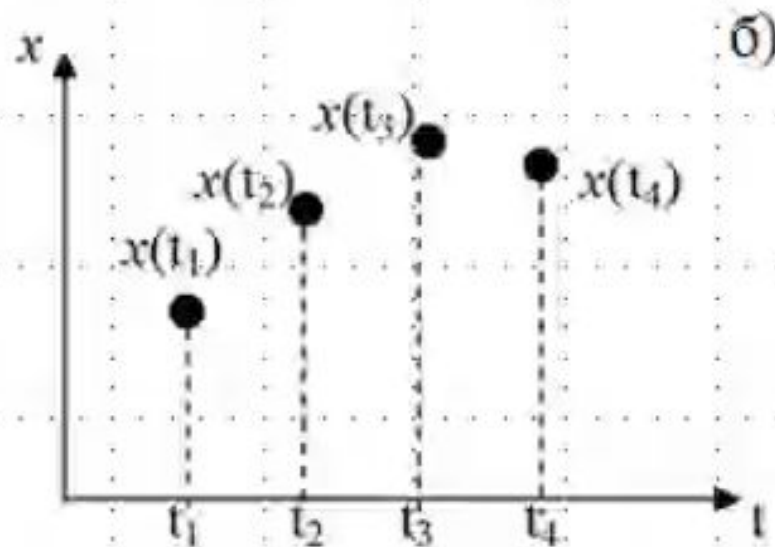
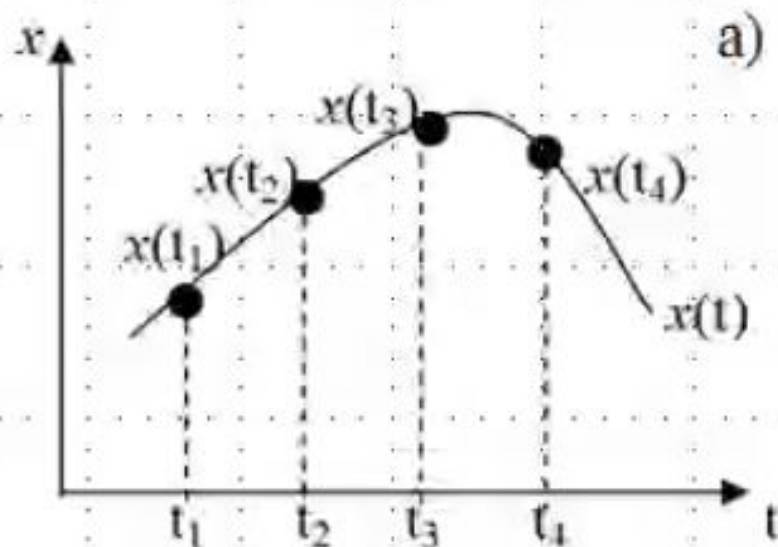
Квантование по времени



Квантование по уровню

Квантование по времени

Замена непрерывной (по времени и по уровню) функции $x(t)$ некоторым множеством непрерывных (по уровню) функций $x(t_i)$



а) аналоговый сигнал $x(t)$ до квантования;

б) дискретный (по времени) сигнал $x(t)$ – результат квантования.

Квантование по времени

Дискретизация связана с потерей информации.

Погрешность дискретизации $\varepsilon(t) = x(t) - v(t)$, где $v(t)$ – функция восстановления, которая по дискретным значениям восстанавливает $x(t)$.

Виды дискретизации различаются по **регулярности** отсчетов:

1. **равномерная** дискретизация – Δt постоянно;

2. **неравномерная** дискретизация – Δt переменна:

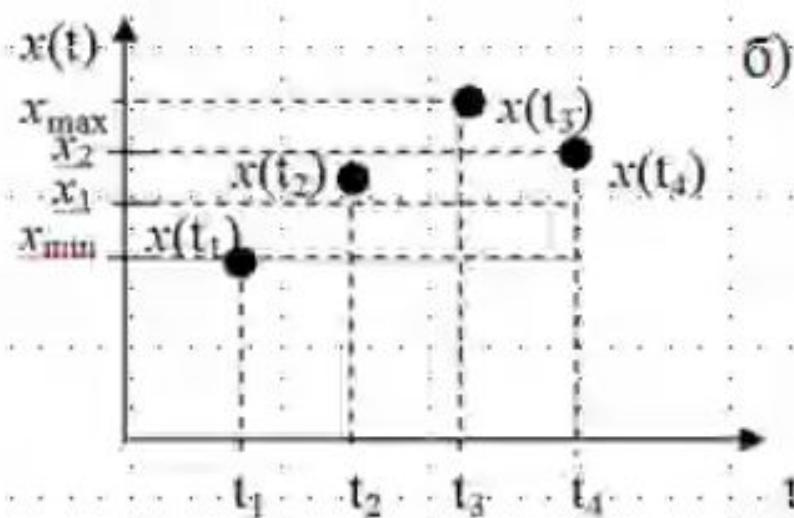
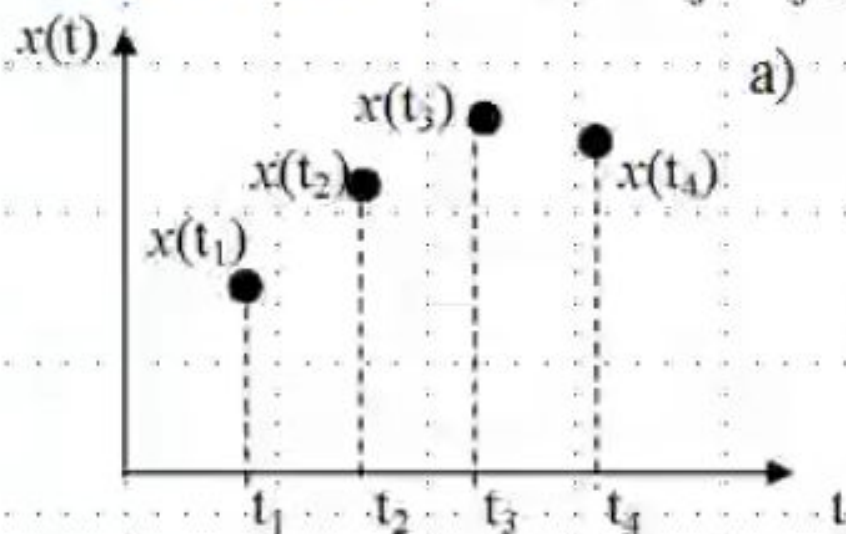
- **адаптивная** – Δt меняется автоматически в зависимости от текущего изменения сигнала. Это позволяет увеличивать шаг дискретизации, когда изменения сигнала $x(t)$ незначительны, и уменьшать – в противном случае;

- **программируемая** – Δt изменяется оператором или в соответствии с заранее выставленными условиями, например, в фиксированные моменты времени.

Квантование по уровню

Преобразование непрерывных (по уровню) сигналов $x(t)$ в диапазоне от x_{\min} до x_{\max} в моменты отсчета t_i в дискретное множество значений x_k — *уровней квантования*.

Шаг квантования $\Delta x = x_j - x_{j-1}$



- а) аналоговые по уровню (но дискретные по времени) сигналы $x(t_i)$ до квантования;
- б) квантованные по уровню сигналы $x(t_i)$.



Квантование по уровню

При квантовании по уровню не всегда сигнал $x(t_i)$ совпадает с уровнем квантования. В таком случае $x(t_i)$ отождествляют с ближайшим значением – с ближайшим меньшим (или большим) значением.

Погрешность квантования по уровню $\varepsilon(x_k) \equiv x(t_i) - x_k$.

Погрешность квантования по уровню тем меньше, чем меньше шаг квантования.

Виды квантования по уровню:

1. **равномерное**, когда диапазон изменения сигнала разбивается на m одинаковых частей. Тогда, зная размер шага квантования, для представления x_k достаточно знать число k .
2. **неравномерное**, когда диапазон изменения сигнала разбивается на m различных частей.

Достоинства дискретной формы

- Возможность дискретизации непрерывного сигнала с любой желаемой точностью (для возрастания точности достаточно уменьшить шаг) принципиально важна с точки зрения информатики.
- **Компьютер** (англ. *computer* – вычислитель) – это программируемое электронное устройство, способное обрабатывать данные и производить вычисления, а также выполнять другие задачи манипулирования **символами**. Это цифровая машина, внутреннее представление информации в которой **дискретно**.
- Дискретизация входной информации (если она непрерывна) позволяет сделать ее пригодной для компьютерной обработки.



Мера информации

Семиотика (греч. знак, признак) – наука, занимающаяся исследованием свойств знаков и знаковых систем.

Семиотика выделяет три отношения между знаками:

- **Синтаксис** рассматривает внутренние свойства сообщений, т. е. определяет способ представления информации на носителе (в сигнале).
- **Семантика** анализирует смысловое содержание сообщения, его отношение к источнику информации. Семантика может рассматриваться как некоторое соглашение, известное получателю информации, о том, что означает каждый сигнал (так называемое правило интерпретации).
- **Прагматика** рассматривает потребительское содержание сообщения, его отношение к получателю, т. е. определяет влияние информации на поведение потребителя.

Мера информации

- Определить понятие «*количество информации*» довольно сложно.
- В соответствии с семиотикой при оценке количества *переданной* информации выделяют три уровня:
 - синтаксический,
 - семантический,
 - прагматический.

Синтаксический уровень

Идея заключается в рассмотрении технических проблем совершенствования методов передачи сообщений и их материальных носителей – сигналов, проблем доставки получателю сообщений. Полностью абстрагируются от смыслового содержания сообщений и их целевого предназначения.

информация = данные

Учитывают:

- Тип носителя.
- Способ представления информации.
- Скорость передачи и обработки.
- Размеры кодов представления информации и т.д.

Семантический уровень

Идея: Проблемы связаны с формализацией и учетом смысла передаваемой информации. Проблемы этого уровня чрезвычайно сложны, так как смысловое содержание информации больше зависит от получателя, чем от семантики сообщения, представленного на каком-либо языке.

На данном уровне:

- Анализируются сведения, которые отражает информация.
- Выявляется смысл информации.
- Выявляется содержание информации.
- Осуществляется обобщение.

Прагматический уровень

Идея: Проблемы этого уровня связаны с определением ценности и полезности информации для потребителя. Интересуют последствия от получения и использования данной информации потребителем.

На данном уровне:

- Ценность информации может быть различной для разных потребителей.
- Фактор актуальности доставки и использования.

Классификация мер информации

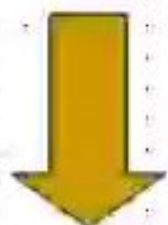
Мера информации – критерий оценки количества информации. Обычно она задана некоторой неотрицательной функцией, определенной на **множестве событий** и являющейся **аддитивной**, то есть мера конечного объединения событий (множеств) равна сумме мер каждого события.

В соответствии с уровнями оценки количества переданной информации выделяют три меры информации:

- синтаксическая мера,
- семантическая мера,
- прагматическая мера.

Синтаксическая мера информации

На синтаксическом уровне для измерения информации вводятся два параметра



V_D

Объем информации
(объемный подход)



I

Количество информации
(вероятностный подход)

Объемный подход (V_D)

Идея: Если количество информации, содержащейся в сообщении из одного символа, принять за единицу, то объем информации (данных) V_D в любом другом сообщении будет равен количеству символов (разрядов) в этом сообщении. В памяти компьютера объем информации записывается двоичными знаками и равен количеству требуемых для этой записи двоичных кодов.

Объем данных (V_D) в техническом смысле этого слова понимается как информационный объем сообщения или как объем памяти, необходимый для хранения сообщения без каких-либо изменений.

Единицы измерения информации

кол-во *бит* = кол-во двоичных цифр (0 и 1)

bit = **binary digit** – двоичное число

Пример: код 11001011 имеет объем данных V= 8 бит

1 байт = 8 бит

1 Кбайт = 1024 байт = 2^{10} байт

1 Мбайт = 1024 Кбайт = 2^{20} байт = 1 048 576 байт;

1 Гбайт = 1024 Мбайт = 2^{30} байт = 1 073 741 824 байт;

1 Тбайт = 1024 Гбайт = 2^{40} байт;

1 Пбайт = 1024 Тбайт = 2^{50} байт;

1 Эбайт = 1024 Пбайт = 2^{60} байт;

Вероятностный подход

События, о которых нельзя сказать произойдут они или нет, пока не будет осуществлен эксперимент, называются **случайными**.

Отдельный повтор случайного события называется **опытом**, а интересующий нас исход этого опыта – **благоприятным**.

Если N – общее число опытов, а N_A – количество благоприятных исходов случайного события A , то отношение N_A/N , называется **относительной частотой появления события A** .

В разных сериях опытов частота может быть различна, но при увеличении количества опытов относительная частота все меньше отклоняется от некоторой константы, ее наличие называется **статической устойчивостью частот**.

Если все исходы опыта конечны и равновозможные, то их вероятность равна

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

где N - число исходов.

Энтропия

Энтропия – численная величина, измеряющая неопределенность.

$$H = f(N)$$

Некоторые свойства:

1. $f(1)=0$, так как при $N=1$ исход не является случайным и неопределенность отсутствует.
2. $f(N)$ возрастает с ростом N , чем больше возможных исходов, тем труднее предсказать результат.
3. Если a и b два независимых опыта с количеством равновероятных исходов N_a и N_b , то мера их суммарной неопределенности равна сумме мер неопределенности каждого из опытов:

$$f(N_a) + f(N_b) = f(N_a \cdot N_b)$$

Количество информации – разность неопределенностей "ДО" (априори) и "ПОСЛЕ" (апостериори) опыта:

$$I = H_1 - H_2$$

Какой вид имеет функция f ?

$$X = N^M$$

общее число исходов

N – число исходов, M – число опытов (пример: $X = 2^3 = 8$)

Ситуацию с M попытками можно рассматривать как некоторую сложную систему, состоящую из независимых друг от друга подсистем - «однократных» попыток. Энтропия такой системы будет в M раз больше, чем энтропия одной системы (*принцип аддитивности энтропии*):

$$f(N^M) = M \cdot f(N),$$

т.к. $\ln X = M \cdot \ln N \Rightarrow M = \frac{\ln X}{\ln N},$

получаем

$$f(X) = \frac{f(N)}{\ln N} \cdot \ln X.$$

Формула Хартли

Обозначим через K следующее отношение:

$$K = \frac{f(N)}{\ln N}$$

обычно принимают

$$K := \frac{1}{\ln 2}$$

Получим $H = f(N) = K \cdot \ln N$, тогда

формула Хартли для **равновозможных** исходов

$$H = \log_2 N$$

Важным при введении какой-либо величины является вопрос о том, что принимать за единицу ее измерения. Очевидно, H будет равно единице при $N = 2$. Иначе говоря, в качестве единицы принимается количество информации, связанное с проведением опыта, состоящего в получении одного из двух равновероятных исходов (примером такого опыта может служить бросание монеты при котором возможны два исхода: «орел», «решка»). Такая единица количества информации называется «**бит**».

Количество информации (в битах), заключенное в двоичном слове, равно числу двоичных знаков в нем.

Формула Шеннона

Мера Хартли подходит лишь для идеальных, абстрактных систем, так как в реальных системах состояния системы не одинаково осуществимы (**не равновероятны**).

Для них используют более подходящую **меру Шеннона** для **неравновозможных** исходов (вероятности исходов p_i могут быть различны):

$$H = \sum_{i=1}^N p_i \cdot \log_2 \left(\frac{1}{p_i} \right) = - \sum_{i=1}^N p_i \cdot \log_2 (p_i).$$

N – исходов (состояний), p_i – вероятность (относительная частота) перехода системы в i -е состояние, сумма всех p_i должна равняться 1.

Формула Шеннона

Положительная сторона *формулы Шеннона* – ее отвлеченность от смысла *информации*.

В отличие от *формулы Хартли*, она учитывает различность состояний, что делает ее пригодной для практических вычислений.

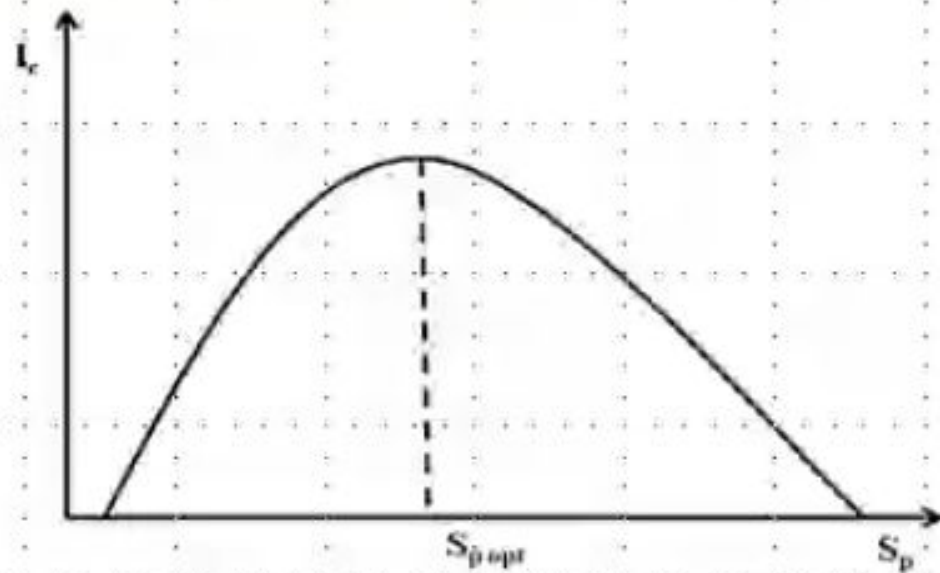
Основная отрицательная сторона *формулы Шеннона* – она не распознает различные состояния системы с одинаковой вероятностью.

Семантическая мера информации

Тезаурус — это совокупность сведений, которыми располагает пользователь или система.

при $S_p \rightarrow 0$ пользователь не воспринимает, не понимает поступающую информацию;

при $S_p \rightarrow \infty$ пользователь все знает, и поступающая информация ему не нужна.



Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса.

$$C = \frac{I_c}{V_D}$$

Относительной мерой количества семантической информации может служить коэффициент содержательности C , который определяется как отношение количества семантической информации к ее объему.