Некоторые дополнения и пояснения к теореме Шеннона

В.А. Зорич

2 декабря 2021 г.

Излагая ключевые понятия и идеи, относящиеся к теореме Шеннона, мы в спецкурсе не погружались в детальные обсуждения. Геометрические аспекты теоремы Шеннона служили лишь примером проявления эффектов многомерности, обсуждавшихся в спецкурсе. Изложенное ниже в какой-то мере восполнит этот пробел, и, заодно, покажет, как много полезного и важного, чего мы не касались, лежит рядом.

Дополнение 1.

Информация, энтропия и условная энтропия.

При передаче сообщений (значений случайной величины X) на приёмном конце могут появляться значения случайной величины Y. Какую информацию о системе X приносит знание системы Y ?

Конкретизируем постановку вопроса. Передатчик посылает сообщение $\bar{x} = (x_1, ..., x_n)$, приемник получает $\bar{y} = (y_1, ..., y_n)$. Как по полученному восстановить посланное? Если никаких искажений нет, т. е. всегда $y_i = x_i$, то и вопроса нет. Поэтому будем считать, что канал характеризуется некоторой матрицей (p_{ij}) вероятностей перехода посланного сигнала x_i в принятый сигнал y_j.

Какую информацию о сообщении $\bar{x}$ содержит сообщение $\bar{y}$? Или по-другому: как меняется (уменьшается) неопределенность $\bar{x}$, когда мы узнаем $\bar{y}$?

Обратимся к условным вероятностям и введем понятие условной энтропии $H(X|Y)$ случайной величины X на входе канала связи относительно случайной величины Y на выходе.

Тогда вслед за Шенноном можно будет рассмотреть величину

$$I(X;Y) = H(X) - H(X|Y),$$

и считать ее эффективным *количеством информации*, которое в среднем переносится одним посылаемым сигналом (значением случайной величины X) в таком канале связи.

Значит, *пропускная способность канала связи* определится как

$$C = \sup_{\{p_x\}} I(X;Y),$$

где супремум берется по всевозможным кодам, т. е. по всевозможным распределениям вероятностей $\{p_x\}$ входной случайной величины X, имеющей фиксированный конечный набор значений (алфавит).

Итак, определим условную энтропию $H(X|Y)$ одной случайной величины X относительно другой случайной величины Y.

Пусть $\{p_x\}$, $\{p_y\}$ и $\{p_{x,y}\}$ соответственно — распределения вероятностей случайных величин X, Y и совместной случайной величины $Z = (X,Y)$.

Если вероятность появления на входе значения x_i случайной величины X равна p_{x_i}, а вероятность $p(y_j|x_i)$ перехода x_i в y_j дана и равна p_{ij}, то вероятность p_{x_i,y_j} совместного события $z_{ij} = (x_i, y_j)$ равна $p(y_j|x_i)p_{x_i}$, а общая вероятность p_{y_j} появления на выходе значения y_j случайной величины Y равна $\sum_i p(y_j|x_i)p_{x_i}$.

Для облегчения записей формул и без потери их ясности мы не будем больше писать дополнительные нижние индексы. Например, стандартные формулы условной вероятности будем писать так: $p_{x,y} = p(y|x)p_x$ или $p_{x,y} = p(x|y)p_y$, поскольку $p_{x,y} = p_{y,x}$.

Найдем сначала условную энтропию $H(X|Y = y)$ случайной величины X при условии, что случайная величина Y имеет значение y. Иными словами, мы сейчас найдем, какой становится энтропия (неопределенность) X при условии, что случайная величина Y приняла значение y.

$$H(X|Y = y) = -\sum_x p(x|y)\log p(x|y) = -\sum_x \frac{p_{x,y}}{p_y}\log\frac{p_{x,y}}{p_y}.$$

Теперь можно найти и интересующую нас условную энтропию $H(X|Y)$ случайной величины X относительно случайной величины Y:

$$H(X|Y) = -\sum_y p_y H(X|Y=y) = -\sum_y p_y \sum_x \frac{p_{x,y}}{p_y} \log \frac{p_{x,y}}{p_y} =$$

$$= -\sum_{x,y} p_{x,y} \log p_{x,y} + \sum_y p_y \log p_y = H(X,Y) - H(Y).$$

Здесь $H(X,Y)$ — совместная энтропия пары $Z = (X,Y)$ случайных величин X и Y; распределение вероятностей пары есть $\{p_{x,y}\}$.

Мы нашли, что $H(X|Y) = H(X,Y) - H(Y)$. Но поскольку $p_{x,y} = p_{y,x}$ и $p_{x,y} = p(y|x)p_x = p(x|y)p_y = p_{y,x}$, справедливы также соотношения $H(X,Y) = H(Y,X)$ и $H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.

Итак,

$$H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X) = H(Y,X).$$

Учитывая формулу $I(X;Y) = H(X) - H(X|Y)$ (определение Шеннона) количества информации, находим, что

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y).$$

Поскольку $H(X,Y) = H(Y,X)$, отсюда следует, что

$$I(X;Y) = I(Y;X).$$

Дополнение 2.

2.1. *Частота периодического процесса.*

Если совершается циклический (периодический) процесс, то время T одного полного цикла процесса называют его периодом. В наших масштабах, период обычно измеряют в секундах.

Например, если имеется колебательный процесс, то говорят о периоде T колебаний.

Количество $\nu = 1/T$ колебаний в единицу времени называется частотой колебаний или процесса.

Частоту ν обычно измеряют в Герцах, Герц (Гц, Hz) — одно колебание в секунду.

Гармонические колебания, типа $\cos\omega t$ обычно записывают, используя, так называемую, угловую (круговую) частоту ω, измеряемую в радианах в секунду. Связь указанных частот простая: $\nu = 2\pi/\omega$ или $\omega = 2\pi\nu$.

Человеческий слух воспринимает звуковые колебания в диапазоне 20 — 20000 Герц = 20 килогерц (20 Гц — 20 кГц , 20 Hz — 20 kHz).

Характерные диапазоны частот:

Зрение $3.85 \cdot 10^{14} - 7.89 \cdot 10^{14}$ Гц.

Частоты цветов в терагерцах (1 ТГц = 10^{12} Гц):

красный 405 — 480,

зелёный 530 — 600,

синий 620 — 680.

Радио 30 кГц — 3000 ГГц (1 кГц = 10^3 Гц , 1 ГГц = 10^9 Гц).

Телевидение (цифровое эфирное) 470 — 790 МГц (1 МГц = 10^6 Гц).

Рентген $3 \cdot 10^{16} - 6 \cdot 10^{19}$ Гц.

2.2. *Частота и комплексная экспонента.*

Комплексная запись тригонометрических функций

$\cos\omega t = \frac{1}{2}(e^{i\omega t} + e^{-i\omega t})$ и $\sin\omega t = \frac{1}{2i}(e^{i\omega t} - e^{-i\omega t})$

весьма удобна, например, в теории рядов и интегралов Фурье.

Следует, однако, иметь в виду, что при этом возникают парные отрицательные частоты. Наряду с частотой ω возникает частота $-\omega$. Так

$$\int_0^a 2\cos\omega t d\omega = \int_{-a}^a e^{i\omega t} d\omega = \frac{2\sin at}{t}.$$

Слева интеграл по отрезку $[0, a]$, справа интеграл по отрезку $[-a, a]$.

Дополнение 3.

Интеграл Фурье.

Напомним, что в терминах частот ω и ν преобразование Фурье $\hat{f}$ и $\check{f}$ соответственно и интеграл Фурье функции f (формула обращения преобразования Фурье в пространстве L_2) имеют вид

$$f(t) = \int_{\mathbb{R}} \hat{f}(\omega) e^{i\omega t} d\omega, \qquad \text{где} \quad \hat{f}(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} f(t) e^{-i\omega t} dt;$$

$$f(t) = \int_{\mathbb{R}} \check{f}(\nu) e^{i2\pi\nu t} d\nu, \qquad \text{где} \qquad \check{f}(\nu) = \int_{\mathbb{R}} f(t) e^{-i2\pi\nu t} dt.$$

Функцию f профессионалы по связи обычно называют сигналом, а её преобразование Фурье, функцию $\hat{f}$ или функцию $\check{f}$, называют спектром сигнала. Спектр показывает, какие частоты и в какой пропорции присутствуют в сигнале (или из каких простейших гармонических составляющих можно составить этот сигнал). Поскольку преобразование Фурье обратимо (в виде интеграла Фурье), то как сам сигнал, так и его преобразование Фурье вполне равносильны и являются различными ипостасями одной сущности — сигнала, в его представлении по времени и в его представлении по частоте.

Если спектральная функция сигнала финитна, то в интегральном представлении сигнала интеграл по всей прямой можно заменить интегралом по тому конечному отрезку, где находится носитель спектральной функции.

Если функция f вещественнозначная, то $\bar{f} = f$, поэтому $\hat{f}(-\omega) = \hat{f}(\omega)$ и $\check{f}(-\nu) = \check{f}(\nu)$. В этом случае знание спектра сигнала на неотрицательной части оси частот вполне определяет весь спектр частот сигнала, что естественно (вспомните замечание о формуле Эйлера и комплексной записи гармонических колебаний).

Найдём, например, функцию (сигнал) с финитным, постоянным, единичным спектром угловых частот $0 \le \omega \le \Omega$.

$$f(t) = \int_{-\Omega}^{\Omega} e^{i\omega t} d\omega = \frac{2 \sin \Omega t}{t}.$$

Если бы речь шла о такой же функции по отношению к частотам в Герцах в диапазоне $0 \le \nu \le W$, то

$$f(t) = \int_{-W}^{W} e^{i2\pi\nu t} d\nu = \frac{2 \sin 2\pi W t}{2\pi t} = \frac{\sin 2\pi W t}{\pi t}.$$

Если $\Omega = 2\pi W$, то оба результата, вроде бы, должны совпадать? Они не совпадают! В чём дело?

Пояснение: $1 \neq 1$, если единицы разной размерности.

Заметим ещё, что $2\pi \hat{f}(\omega) = \check{f}(\nu)$, если $\omega = 2\pi\nu$.

То есть $\check{f}(\nu) = 2\pi \hat{f}(2\pi\nu)$ и, в частности, $\check{f}(\nu) \equiv 2\pi$ там, где $\hat{f}(\omega) \equiv 1$.

Дополнение 4.

Линейные приборы.

4.1. *Линейный прибор и его математическое описание (свертка).*

Хорошей математической моделью многих приборов и аппаратов является линейный оператор.

Слова «линейный прибор с инвариантными во времени свойствами» на математическом языке означают, что это линейный оператор A, действующий на функции от времени и коммутирующий с оператором сдвига T, т. е. $AT = TA$, где $T = T_\tau$, а $(T_\tau f)(t) = f(t - \tau)$.

Например, если A — проигрыватель, то он завтра должен выдать с того же диска ту же музыку, что и сегодня, лишь с естественным сдвигом во времени.

В соответствии с принятой у радиоинженеров терминологией, функцию f, на которую действует оператор A, будем называть *сигналом*, точнее, *входным сигналом* или *входом*, а результат Af обработки сигнала f прибором A будем называть *сигналом на выходе* или *выходом* и обозначать $\tilde{f}$.

Поскольку непрерывную функцию f можно хорошо аппроксимировать ступенчатой функцией, легко понять, что если знать отклик такого прибора A на элементарную ступеньку, то можно найти его отклик и на любой входной сигнал f.

В идеале ступенька превращается в единичный импульс — δ-функцию. Если написать тождество $f(t) = \int f(\tau)\delta(t - \tau)d\tau$, то сразу найдем, что $Af(t) = \int f(\tau)A\delta(t - \tau)d\tau = \int f(\tau)\tilde{\delta}(t - \tau)d\tau =: f * \tilde{\delta}$, где $*$ — символ операции свертки функций. Значит, $Af = f * \tilde{\delta}$.

Функция $A\delta = \tilde{\delta}$, т. е. отклик прибора на единичный импульс (δ-функцию), называется *аппаратной функцией прибора* и часто обозначается символом E

Итак, с математической точки зрения прибор A есть не что иное, как просто оператор свертки $Af = f * \tilde{\delta} = f * E$.

Решение уравнений в свертках имеет, таким образом, весьма конкретные прямые приложения (например, восстановление переданного сигнала f по принятому сигналу $Af = \tilde{f}$).

4.2. *Фурье-двойственное (спектральное) описание линейного прибора.*

Напомним, что частоту ν периодического процесса обычно измеряют количеством полных циклов в единицу времени (один Герц — это одно полное колебание в секунду; обозначается 1Гц или 1Hz). Угловая или круговая частота $\omega = 2\pi\nu$ отличается от частоты ν только множителем 2π, отвечающим за переход к измерению в радианах в единицу времени.

Подсчитаем отклик $\tilde{f}$ прибора на входной сигнал $f = e^{i\omega t}$ (тем самым, в силу формулы Эйлера $e^{i\omega t} = \cos\omega t + i\sin\omega t$ узнаем отклик прибора и на простейшие гармонические колебания $\sin\omega t$ круговой частоты ω; комплексный язык удобен):

$$\begin{aligned} &Af = f * \tilde{\delta} = f * E = \int f(t-\tau)\tilde{\delta}(\tau)d\tau = \\ &\int \mathrm{e}^{i\omega(t-\tau)}E(\tau)d\tau = \left(\int E(\tau)\mathrm{e}^{-i\omega\tau)}d\tau\right)\mathrm{e}^{i\omega t} = P(\omega)\mathrm{e}^{i\omega t}. \end{aligned}$$

Мы получили колебание той же частоты, что и на входе, но, возможно, измененное по амплитуде в $|P(\omega)|$ раз и по фазе на аргумент $\arg P(\omega)$ комплексного числа $P(\omega)$.

Величина P как функция от ω называется *спектральной характеристикой прибора.* Спектральная характеристика прибора, как видно, является (с точностью до нормирующего множителя) преобразованием Фурье $\hat{E}$ аппаратной функции E этого прибора: $P = 2\pi\hat{E}$.

Вычислим тогда Af, зная P. Представляя f интегралом Фурье, находим представление Af в виде интеграла Фурье:

$$f(t) = \int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega t}d\omega \quad \text{и} \quad Af(t) = \int_{\mathbb{R}} \hat{f}P(\omega)e^{i\omega t}d\omega.$$

В частности, если $f = \delta$, то

$$E(t) = \tilde{\delta}(t) = A\delta(t) = \frac{1}{2\pi}\int_{\mathbb{R}} P(\omega)e^{i\omega t}d\omega \quad \left(= \int_{\mathbb{R}} \hat{E}(\omega)e^{i\omega t}d\omega\right).$$

4.3. *Функции и приборы с финитным спектром.*

Приборы, с которыми реально приходится иметь дело, подобно слуху и зрению, способны воспринимать сигнал только в определенном диапазоне частот. Поэтому представляется естественным обратить особое внимание на функции и приборы с финитным спектром.

Если спектр (преобразование Фурье $\hat{f}$) функции f финитен (тождественно равен нулю вне некоторого компакта), то в ее представлении в виде интеграла Фурье участвует интеграл, взятый только по конечному промежутку:

$$f(x) = \int_{-a}^{a} \hat{f}(\omega)e^{i\omega x}d\omega.$$

Функции с финитным спектром ввиду их важности были предметом самостоятельного математического исследования.

Если функция g принадлежит пространству $L_2(\mathbb{R})$, то, как известно из теории преобразования Фурье, функция

$$f(x) = \int_{-a}^{a} g(\omega)e^{i\omega x}d\omega$$

тоже принадлежит пространству $L_2(\mathbb{R})$. Более того, применив неравенство Коши—Буняковского—Шварца, легко увидеть, что она ограничена на вещественной оси и распространяется на всю комплексную плоскость как целая функция и при этом $|f(x+iy)| \leq ce^{a|y|}$. Класс целых функций такого вида называют классом Винера и обозначают W_a.

Теорема Пале—Винера (Paley—Wiener) утверждает, что $f \in W_a$ тогда и только тогда, когда эта функция допускает представление

$$f(z) = \int_{-a}^{a} g(\omega)e^{i\omega z}d\omega,$$

где $g \in L_2(\mathbb{R})$.

4.4. *Идеальный фильтр и его аппаратная функция.*

Спектральная характеристика прибора, как видно, является (с точностью до нормирующего множителя) преобразованием Фурье $\hat{E}$ аппаратной функции E этого прибора: $P = 2\pi\hat{E}$.

Теперь обратимся к прибору с финитным спектром. Простейшим и основным примером тут может служить прибор, спектральная функция которого $P = 2\pi\hat{E}$ постоянна на промежутке $[-a, a] = [-\Omega, \Omega]$ и равна нулю вне этого промежутка. Пусть $\hat{E} \equiv \frac{1}{2\mathrm{a}}$ на промежутке $[-a, a] = [-\Omega, \Omega]$.

Такой прибор, как мы теперь понимаем, будет пропускать все гармоники угловой частоты, не большей Ω $(0 \leq \omega \leq \Omega = a)$, и не будет реагировать на более высокие частоты (частоты большей $\frac{1}{2\pi}\Omega = W$ Герц).

Прибор с этими свойствами называют *фильтром низких частот с верхней граничной частотой* Ω (или с верхней граничной частотой $\frac{1}{2\pi}\Omega = W$ Герц).

Найдем аппаратную функцию такого фильтра низких частот с верхней граничной угловой частотой a :

$$E(t) = \frac{1}{2a}\int_{-a}^{a} e^{i\omega t}d\omega = \frac{\sin at}{at} = \frac{\sin \Omega t}{\Omega t} = \frac{\sin 2\pi W t}{2\pi W t}.$$

Функция $\frac{\sin t}{t}$, которая, как мы видим, имеет постоянный спектр $\hat{f}(\omega) \equiv \frac{1}{2}$, на единичном промежутке угловых частот $0 \leq \omega \leq \Omega = 1$, является аппаратной функцией идеального фильтра низких частот с единичной полосой пропускания. Она реализуется как ответ такого фильтра на единичный импульс, осуществленный в момент $t = 0$.

Важность функции $\frac{\sin x}{x}$ в электротехнике и теории передачи информации по каналу связи привела к тому, что появилось следующее специальное обозначение:

$$\operatorname{sinc} x := \frac{\sin x}{x}.$$

Некоторое объяснение того, почему на конце тут появилась буква c (от count), как раз и даёт теорема (формула) отсчётов.

Напомним представление интегралом Фурье функции с финитным спектром

$$f(t) = \int_{-\Omega}^{\Omega} \hat{f}(\omega) e^{i\omega t} d\omega = \int_{-W}^{W} \check{f}(\nu) e^{i2\pi\nu t} d\nu \ .$$

Здесь $\omega = 2\pi\nu, \quad \Omega = 2\pi W, \quad 2\pi\hat{f}(\omega) = \check{f}(\nu)$.

В частности сигнал $f(t) = \frac{\sin 2\pi W t}{2\pi W t}$ имеет финитный спектр $\check{f}(\nu) \equiv \frac{1}{2W}$ в полосе частот $0 \leq \nu \leq W \quad (|\nu| \leq W)$ Герц.

Дополнение 5.

Теорема отсчётов и около.

5.1. *Энергия и средняя мощность сигнала.*

Напомним, что теорема отсчетов

$$f(t) = \sum_{-\infty}^{\infty} f\left(\frac{k}{2W}\right) \frac{\sin 2\pi W(t - \frac{k}{2W})}{2\pi W(t - \frac{k}{2W})}$$

восстанавливает сигнал — функцию $f \in L_2(\mathbb{R})$ с финитным спектром частот ν, не превышающих W Герц, — по совокупности отсчетных значений $f(t_k)$ в точках $t_k = k\Delta$, где $\Delta = \frac{1}{2W}$ — отсчетный интервал времени (интервал Найквиста), зависящий от W.

Заметим, что сейчас мы записали формулу отсчётов по отношению к частотам, измеряемым в Герцах. В математике и в теоретических исследованиях, где встречается аппарат рядов и преобразований Фурье, чтобы не осложнять формулы многократным повторением множителя 2π, обычно пользуются угловыми частотами.

Чем шире полоса частот, тем сложнее может быть функция f, тем чаще надо брать отсчеты для ее адекватного дискретного кодирования и восстановления, но зато тем больше информации она (т. е. такой сигнал) может нести.

Базовой в формуле отсчётов является функция $\operatorname{sinc} x = \frac{\sin x}{x}$, точнее, функция $\operatorname{sinc}(2\pi W t) = \frac{\sin(2\pi W t)}{2\pi W t}$. Как мы уже знаем, она имеет финитный спектр $\check{f}(\nu) = \frac{1}{2W}$ на промежутке $-W \leq \nu \leq W$ и является аппаратной функцией идеального фильтра низких частот $0 \leq \nu \leq W$. Она реализуется как ответ такого фильтра на единичный импульс, осуществленный в момент $t = 0$.

Сдвиг по времени, как видно из формул преобразования Фурье, приводит к тому, что функция $e_k(t) = \operatorname{sinc} 2\pi W(t - \frac{k}{2W}) = \frac{\sin 2\pi W(t-\frac{k}{2W})}{2\pi W(t-\frac{k}{2W})}$ имеет спектр $\check{e}_k(\nu) = \frac{1}{2W}\exp(-i\frac{\pi}{W}k\nu)$ в полосе частот $0 \leq \nu \leq W$ $(|\nu| \leq W)$.

Из ортогональности функций $\check{e}_k$ на промежутке $[-W, W]$ (или любом промежутке длины $2W$) и равенства Парсеваля для преобразования Фурье можно заключить, что сами функции e_k ортогональны в пространстве $L_2(\mathbb{R})$, причем $\|e_k\|^2 = \frac{1}{2W}$.

Значит из равенства $f = \sum_{-\infty}^{\infty} x_k e_k$ можно заключить, что $\|f\|^2 = \frac{1}{2W}\sum_{-\infty}^{\infty} x_k^2$.

На практике сигнал f имеет некоторую конечную продолжительность T, т. е. $f(t) \equiv 0$ вне отрезка $0 \leq t \leq T$. Это условие несовместимо с условием финитности спектра функции f. Однако можно считать, что значения $f(t)$ функции малы вне отрезка $[0, T]$ и отсчетные значения функции вне этого отрезка полагаются равными нулю.

Тогда равенство $f = \sum_{-\infty}^{\infty} x_k e_k$ заменится на $f(t) = \sum_{k=1}^{2WT} x_k e_k(t)$, где $t \in [0, T], \quad x_k = f(k\Delta), \quad \Delta = \frac{1}{2W}$.

Такой сигнал f записывается вектором $x = (x_1, ..., x_n) \in \mathbb{R}^n$ его отсчетных значений, и при этом $n = 2WT$

При тех же условиях равенство Парсеваля

$$\int_{-\infty}^{\infty} f^2(t)dt = \sum_{-\infty}^{\infty} x_k^2 \|e_k\|^2 = \frac{1}{2W}\sum_{-\infty}^{\infty} x_k^2$$

заменится на равенство

$$\int_0^T f^2(t)dt = \sum_1^n x_k^2 \|e_k\|^2 = \frac{1}{2W} \sum_1^n x_k^2 = \frac{1}{2W} \|x\|^2.$$

Интеграл здесь, с точностью до конкретного размерного множителя, дает энергию (работу) сигнала f (например, когда f реализуется перепадом напряжения на единичном сопротивлении). Значит, средняя мощность P сигнала f на временно́м промежутке $[0,T]$ есть

$$P = \frac{1}{T} \int_0^T f^2(t)dt = \frac{1}{2WT} \|x\|^2 = \frac{1}{n} \|x\|^2.$$

Таким образом, $\|x\|^2 = nP = 2WTP$, и можно интерпретировать P как среднюю мощность, которая приходится на одну координату вектора x, т. е. на одно отсчетное значение сигнала f.

Сигналы f продолжительности T с финитным спектром в полосе частот W, средняя мощность которых не больше P, в векторном представлении $x = (x_1, ..., x_n)$ оказываются, таким образом, расположенными внутри шара $B(\mathrm{o}, r) = B(r) \subset \mathbb{R}^n$ радиуса $r = \sqrt{2WTP} = \sqrt{nP}$ с центром в начале координат евклидова пространства $\mathbb{R}^n$ размерности $n = 2WT$.

Для сравнения с основным результатом теоремы Шеннона, дающим асимптотику скорости (в битах в секунду) передачи информации по каналу связи при наличии помех (белого шума), приведём два небольших расчёта.

5.2. *Квантование по уровням.*

Измерение отсчетного значения сигнала f производится с некоторой пороговой (предельной) точностью ε. Если амплитуда любого передаваемого сигнала не больше A (т. е. $|f|(t) \le A$ при $t \in [0,T]$), то, разделив отрезок $[-A, A]$ равномерной сеткой точек (уровней) с шагом ε, мы в качестве $f(t)$ будем брать ближайшую к значению $f(t)$ точку этой сетки. Значения $f(t)$ оказываются проквантованными по уровням, количество которых $\alpha = \frac{2A}{\varepsilon}$ (считаем α целым и большим 1). Отвечающее сигналу f слово $x = (x_1, ..., x_n)$, состоящее из n букв x_k, будет записано в алфавите, имеющем α различных букв. Всего имеется α^n таких различных слов x. Если $n = 2WT$, а W и T — большие числа, то число α^n огромно.

5.3. *Идеальный многоуровневый канал связи.*

При этих условиях за время T можно различить $M = \alpha^{2WT}$ (и не более) различных сигналов $f \sim x = (x_1, ..., x_n)$, т. е. указать один определенный сигнал-слово-сообщение из M возможных.

Двоичная запись $x^0 = (x_1^0, ..., x_m^0)$, различающая M объектов, требует $m = \log_2 M$ символов $0, 1$ (считаем m целым). Информацию о значении очередной координаты двоичного вектора (если координаты равноправны, а их возможные значения $0, 1$ равновероятны) принимают за элементарную единицу информации и называют *бит*. Если мы могли бы без каких-либо ошибок получать и передавать векторы (слова), кодирующие наши M сообщений, то за время T мы могли бы различить M объектов (сигналов, сообщений). Скорость передачи информации (о выборе одного из M возможных объектов) по такому идеальному каналу связи (и при такой кодировке), измеренная в битах в секунду, была бы равна $\frac{1}{T} \log_2 M = 2W \log_2 \alpha$.

При $\alpha = 2$ и описанной кодировке мы имели бы скорость $2W$.

Теорема Шеннона, при указанной в ней кодировке, гарантирует предельную скорость $C = W \log_2(1 + \frac{P}{N})$, где, как обычно, P и N — удельные (на одно отсчётное значение) мощности сигнала и помехи, характерные для рассматриваемого канала связи, при наличии в нём помех.

5.4. *Иная запись формулы отсчётов.*

Мы получили следующее представление функции с финитным спектром, известное под названием *формулы Котельникова—Шеннона*, или *теоремы отсчетов*:

$$f(t) = \sum_{-\infty}^{\infty} f\left(\frac{\pi}{a} n\right) \frac{\sin a(t - \frac{\pi}{a} n)}{a(t - \frac{\pi}{a} n)} \ .$$

Отметим, что формулу можно преобразовать и к виду

$$f(t) = \frac{\sin at}{a} \sum_{-\infty}^{\infty} f\left(\frac{\pi}{a} n\right) \frac{(-1)^n}{t - \frac{\pi}{a} n} \ .$$

Подобно тому, как интерполяционный полином Лагранжа даёт полиномиальную интерполяцию значений произвольной функции по её значениям в выделенных точках, так и формула отсчётов может рассматриваться как интерполяционная. Если взять произвольную функцию, то правая часть формулы отсчётов всё равно имеет смысл, и в системе точек $\{\frac{\pi}{a} n\}$ даёт значения, совпадающие со значениюми исходной функции.

Равенства левой и правой частей во всех точках уже может не быть, а интерполяцию исходной функции между её значениями в точках указанной системы правая часть формулы отсчётов будет осуществлять.

5.5. *Торема Шеннона и комментарий Шеннона.*

Ниже мы приведём цитату из работы Шеннона,[1] где он доказывает и обсуждает свою широко известную теперь теорему (ниже это Теорема 2). Несколько слов он пишет и по поводу теоремы отсчётов (Теорема 1).

Theorem 1. *If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $1/2$ seconds apart.*

This is a fact which is common knowledge in the communication art. ...

Theorem 1 has been given previously in other forms by mathematicians but in spite of its evident importance seems not to have appeared explicitly in the literature of communication theory.

Theorem 2. *Let P be the average transmitter power, and suppose the noise is white thermal noise of power N in the band W. By sufficiently complicated encoding systems it is possible to transmit binary digits at a rate*

$$C = W \log_2 \frac{P+N}{N}$$

with as small a frequency of errors as desired. It is not possible by any encoding method to send at a higher rate and have an arbitrarily low frequency of errors.

This shows that the rate $C = W \log_2 \frac{P+N}{N}$ measures in a sharply defined way the capacity of the channel for transmitting information. It is a rather surprising result, since one would expect that reducing the frequency of errors would require reducing the rate of transmission, and that the rate must approach zero as the error frequency does. Actually, we can send at the rate C but reduce errors by using more involved encoding and longer delays at the transmitter and receiver. The transmitter will take long sequences of binary digits and represent this entire sequence by a particular signal function of long duration. The delay is required because the transmitter must wait for the full sequence before the signal is determined. Similarly, the receiver must wait for the full signal function before decoding into binary digits.

[1] C.E.Shannon, Communication in the Presence of Noise. Proceedings of the IRE vol. 37, no. 1, pp. 10–21, Jan. 1949.