Глубинное обучение в обработке звука

Карагодин Никита, *SberDevices*

План курса

Темы курса:

- *Digital Signal Processing*
- *Automatic Speech Recognition (ASR)*
- *Voice Activity Detection, Source Separation, Diarization*

- *Speech Synthesis (TTS)*
- *Music Synthesis*

4 *ДЗ*:

- *Digital Music Synthesis*
- *Automatic Speech Recognition (big)*
- *Speech Synthesis (big)*
- *Neural Music Synthesis (small)*

$$\text{Оценка} = 0{,}4 \cdot \text{МДЗ} + 0{,}6 \cdot \text{БДЗ}$$

Цифровая обработка сигналов

План лекции

- Классификация сигналов
- Звук и его характеристики
- Цифровые сигналы
- Преобразования Фурье
  - Разложение в ряд Фурье
  - Преобразование Фурье
- Дискретизация и теорема Котельникова
  - Преобразование Фурье с дискретным временем
  - Дискретное преобразование Фурье
- Оконное преобразование Фурье
- Спектрограммы и мел-спектрограммы

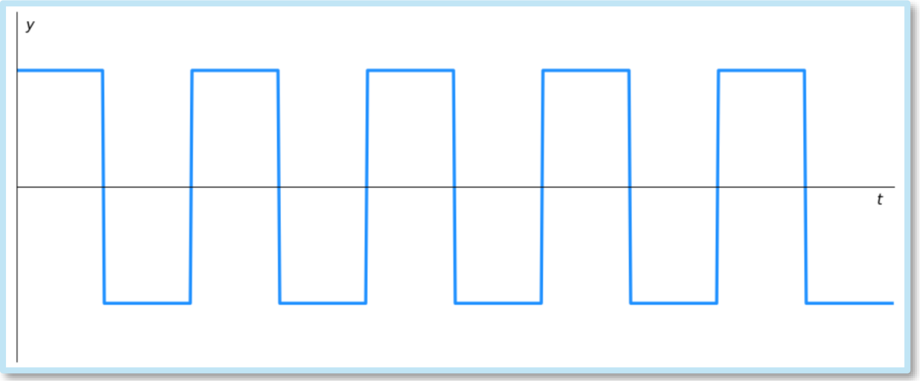
Классификация сигналов

Классификация сигналов

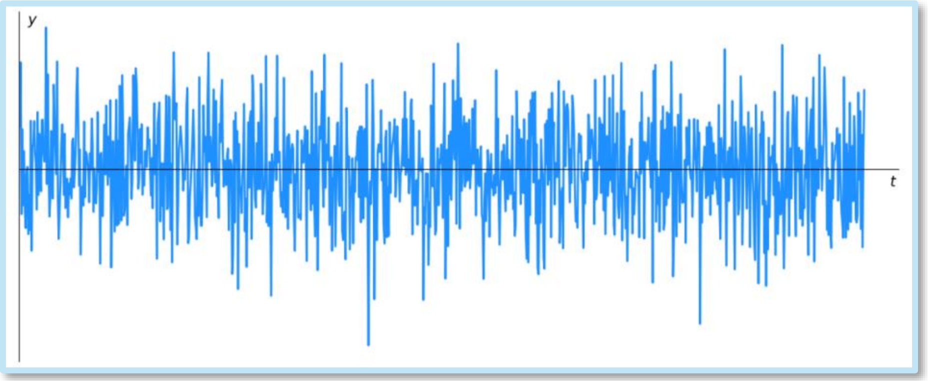
Сигналом называется физический процесс, передающий информацию о состоянии или поведении некоторого объекта или системы

Сигналы

Детерминированные



Случайные



Классификация сигналов

Детерминированные сигналы называется описываются аналитической функцией, и их поведение полностью известно в любой момент времени

Случайные сигналы – это сигналы, которые зависят от случайных факторов и не могут быть точно предсказаны

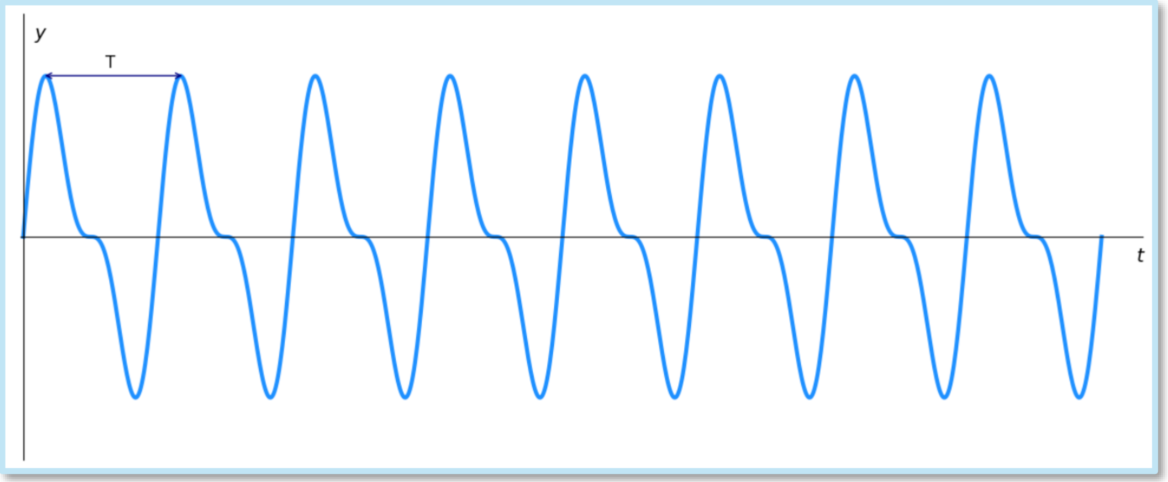
Мы можем лишь оценить вероятность того, какое значение будет принимать сигнал в некоторый момент времени

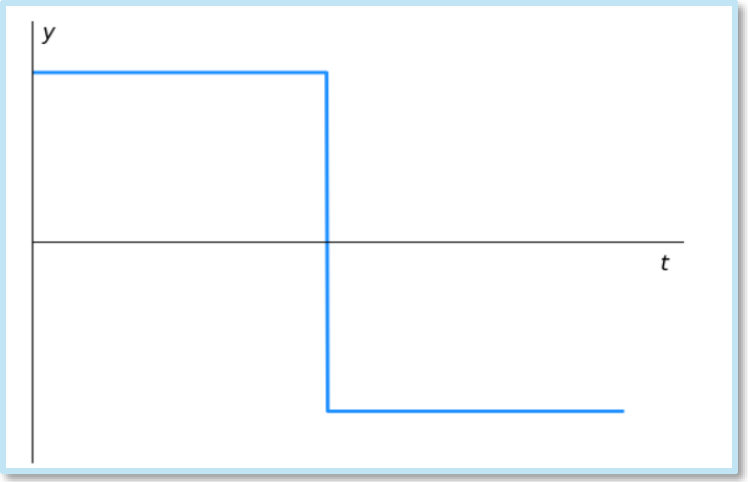
Классификация сигналов

Детерминированные сигналы

Периодические

Непериодические





Классификация сигналов

Периодический сигнал – это сигнал, который повторяется во времени с определённым периодом T, то есть для которого выполнено условие:

$$S(t + nT) = S(t), \qquad \forall n$$

где $n \in \mathbb{Z}$

Непериодические сигналы условию периодичности не удовлетворяют, и они, как правило, ограничены во времени

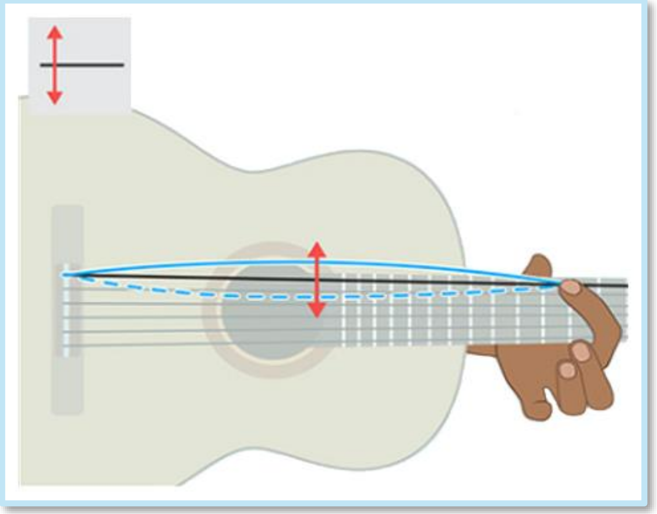
Звук и его характеристики

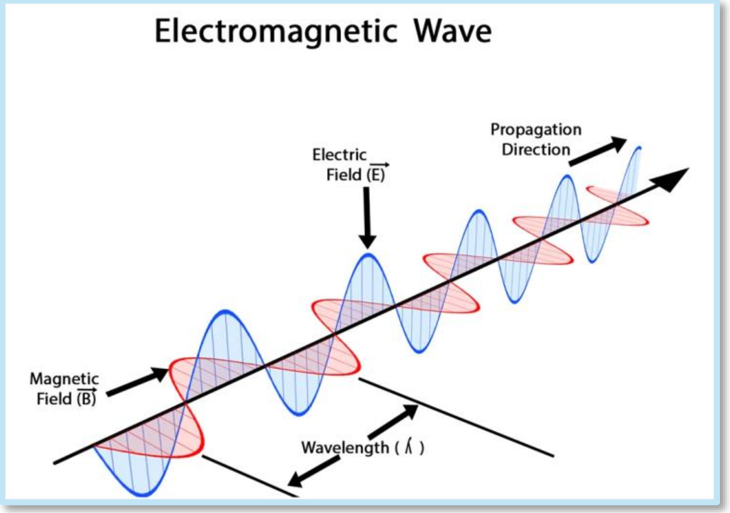
Звук и его характеристики

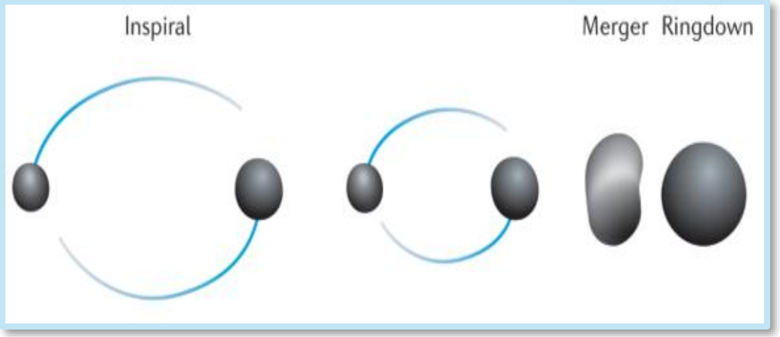
Волна – это изменение некоторой совокупности физических величин, которое способно перемещаться или колебаться внутри ограниченных областей пространства

Волны

Механические | Электромагнитные | Гравитационные





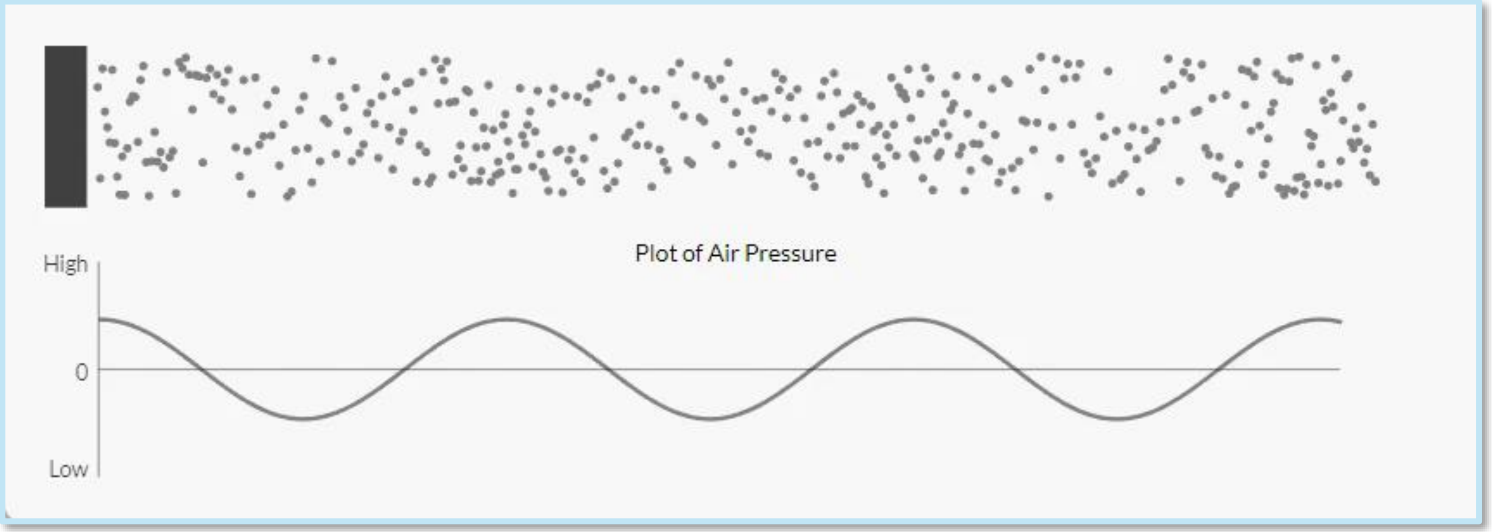


Звук и его характеристики

Механические волны – это периодические возмущения, которые распространяющиеся через физическую среду

Среда – это материал или вещество, которое служит для передачи волн

- Во время своего движения волна не переносит частицы
- Они лишь временно отклоняются от своего положения равновесия, а затем возвращаются обратно

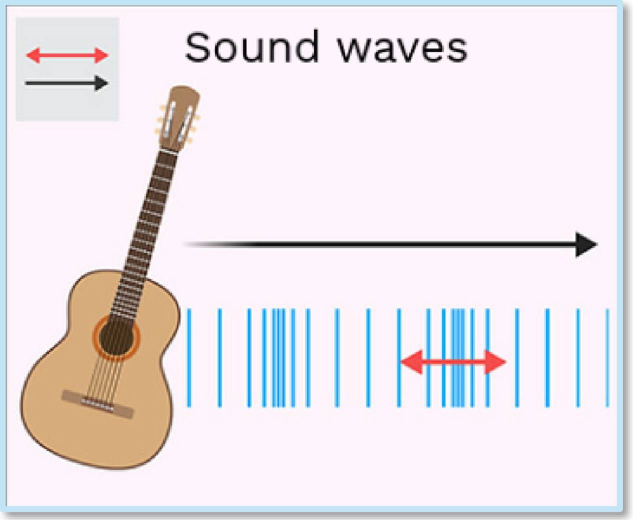


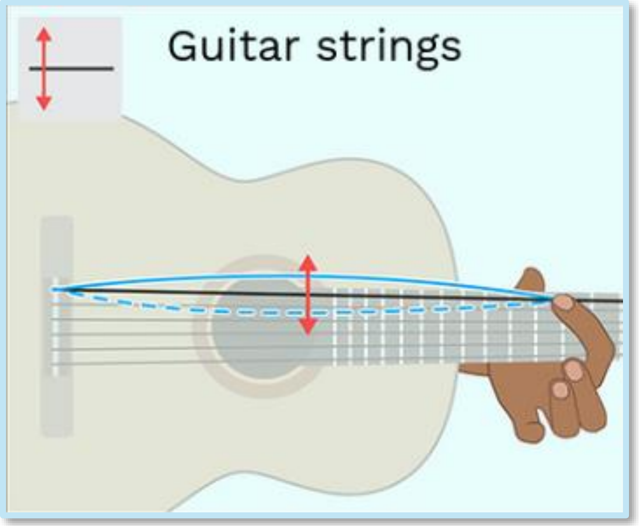
Звук и его характеристики

Механические волны

Продольные

Поперечные



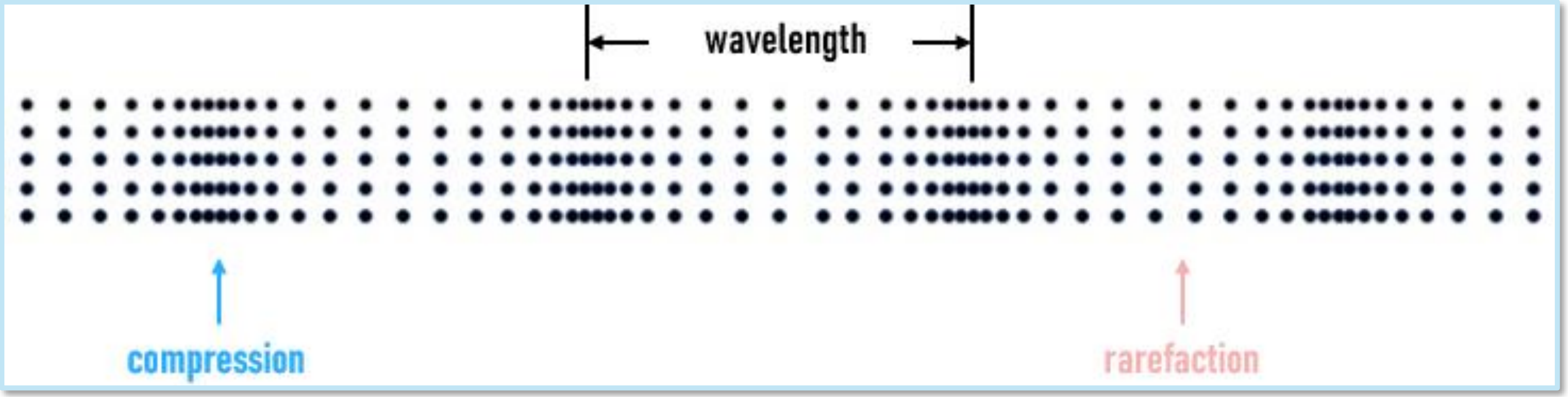


- ***В продольных волнах*** частицы среды колеблются вдоль направления распространения волны
- ***В поперечных волнах*** частицы среды колеблются перпендикулярно направлению распространения волны

Звук и его характеристики

- Когда звуковая волна проходит через воздух, его частицы начинают колебаться вдоль распространения волны

- При этом образуются области ***сжатия*** (***compression***) и ***разряжения*** (***rarefaction***)



Звуковая волна – это чередование участков с повышенным и пониженным давлениями

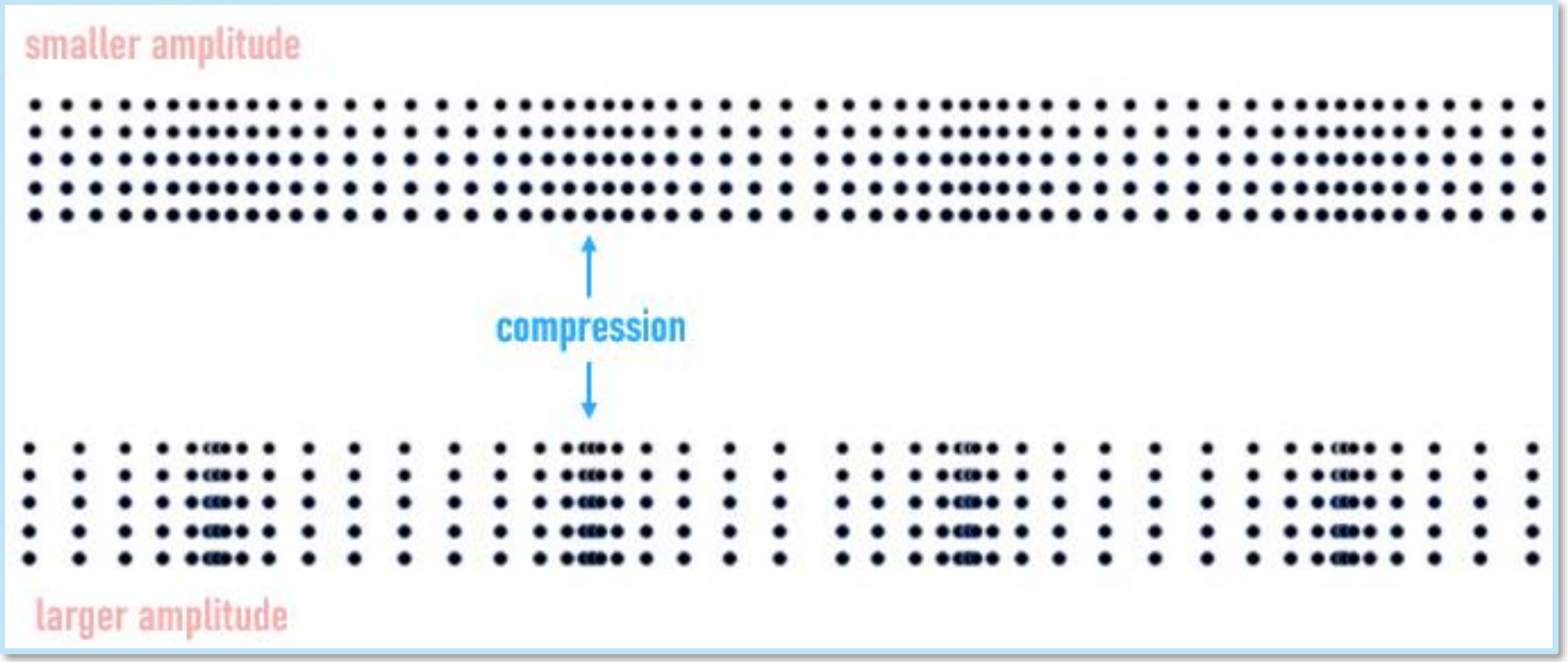
Звук – это волна давления, распространяющаяся через некоторую среду

Звук и его характеристики

Громкость – это наше субъективное ощущение силы звука (его звукового давления)

В основном, человеческое восприятие громкости основано на амплитуде звукового давления на барабанную перепонку

Если все остальные характеристики равны, то звуки более высокой амплитуды воспринимаются как более громкие



Звук и его характеристики

- Можно описать громкость через ***интенсивность***

- Если смотреть на звук, как на поток энергии от одного места к другому, то интенсивность описывает, насколько «концентрирован» этот поток энергии

Математически интенсивность определяется как количество энергии, проходящей через единицу площади в единицу времени:

$$Intensity = \frac{Energy}{Time \cdot Area} \qquad or \qquad Intensity = \frac{Power}{Area}$$

Измеряется интенсивность в ваттах на квадратный метр $\left(\frac{\text{Вт}}{\text{м}^2}\right)$

Звук и его характеристики

Уравнение связывающее интенсивность звука *I* и амплитуду давления *p* имеет вид:

$$I = \frac{p^2}{2\rho c}$$

ρ – плотность среды, c – скорость звука в среде

- Человеческое ухо может воспринимать звуку с очень низкой интенсивностью до $10^{-12}\,\frac{\text{Вт}}{\text{м}^2}$
- Самые громкие звуки могут превышать этот порог в триллионы раз
- Для удобства работы с таким широким диапазоном интенсивностей используется ***логарифмическая шкала децибел***

Звук и его характеристики

Существуют две почти идентичные логарифмические шкалы для описания амплитуды звука, которые часто заменяют друг друга

Уровень интенсивности звука $(\boldsymbol{Sound\ Intensity\ Level, SIL})$:

$$SIL = 10\log_{10}(\frac{I}{I_0})$$

- I – интенсивность звука, I_0 – интенсивность эталонного звука

Уровень звукового давления $(\boldsymbol{Sound\ Pressure\ Level, SPL})$:

$$SPL = 20\log_{10}(\frac{p}{p_0})$$

- p – звуковое давление, p_0 – эталонное звуковое давление

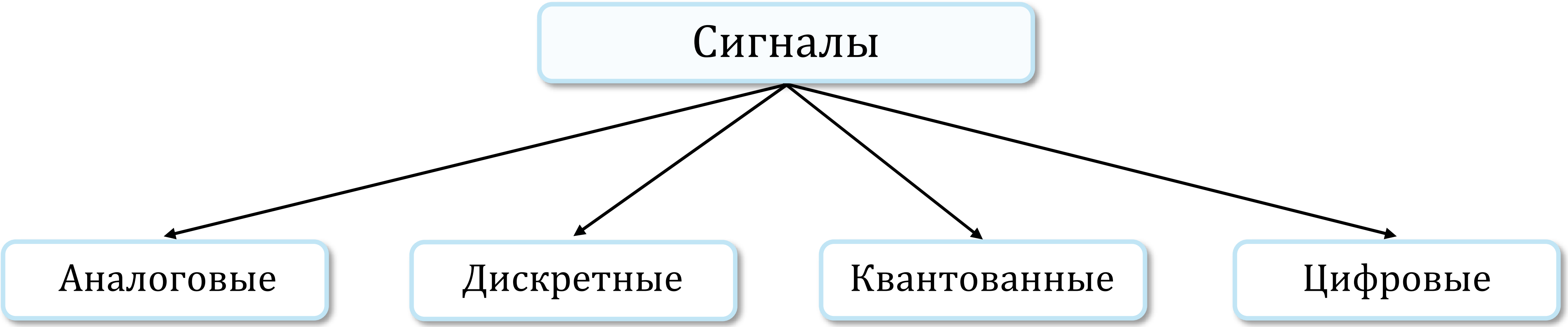
Звук и его характеристики

Источник	Уровень интенсивности (дБ)	Интенсивность (пВт/м²)	Амплитуда давления (пПа)
Порог слуха	0	1	20
Тихий лес	10	10	60
Тиканье часов	20	100	200
Рисовые хлопья	30	1,000	600
Библиотека	40	10,000	2,000
Вентилятор	50	100,000	6,000
Разговор	60	1,000,000	20,000
Шум в машине	70	10,000,000	60,000
Пылесос	80	100,000,000	200,000
Газонокосилка	90	1,000,000,000	600,000
Цепная пила	100	10,000,000,000	2,000,000
Механический цех	110	100,000,000,000	6,000,000
Закрытая арена	120	1,000,000,000,000	20,000,000
Реактивный взлет	130	10,000,000,000,000	60,000,000

Цифровые сигналы и их обработка

Аналоговые, цифровые и дискретные сигналы

Суть ***цифровой обработки сигналов*** состоит в том, что реальный физический сигнал (напряжение или ток) преобразуется в последовательность чисел, которая затем подвергается математическим преобразованиям

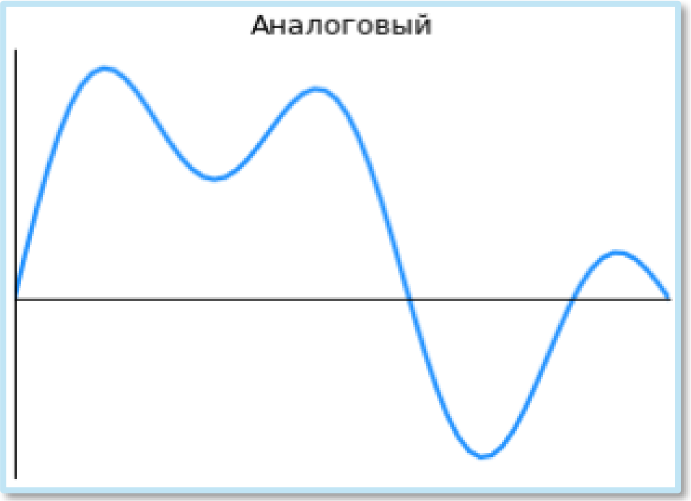


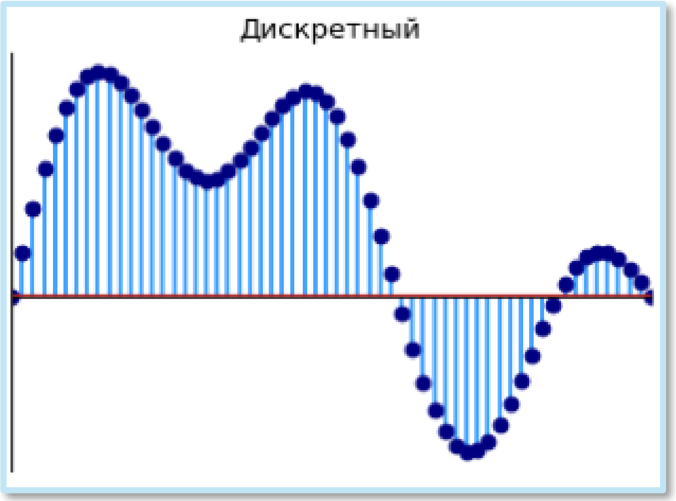
Аналоговые, цифровые и дискретные сигналы

Аналоговый сигнал **–** сигнал, которые описывается непрерывной функцией времени

Почти все физические процессы представляют собой аналоговые сигналы

- Для ***дискретного сигнала*** характерно прерывистое (***дискретное***) изменение во времени
- Эти изменения происходят скачкообразно через некоторые промежутки времени, которые называются интервалом или периодом дискретизации Δt





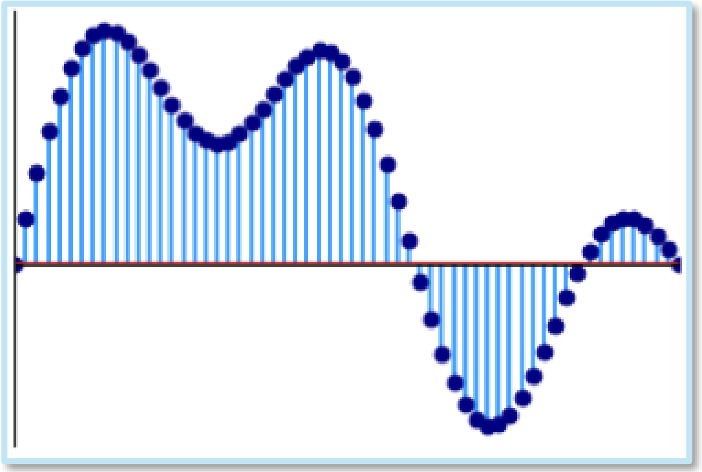
Аналоговые, цифровые и дискретные сигналы

- ***Дискретизация*** аналогового сигнала состоит в том, что сигнал представляется в виде последовательности значений, взятых в дискретные моменты времени, которые называются ***отсчётами*** (***samples***)

- Обычно отсчёты берутся через равные промежутки времени

Величина, обратная периоду дискретизации, называется ***частотой дискретизации*** (***sample rate***):

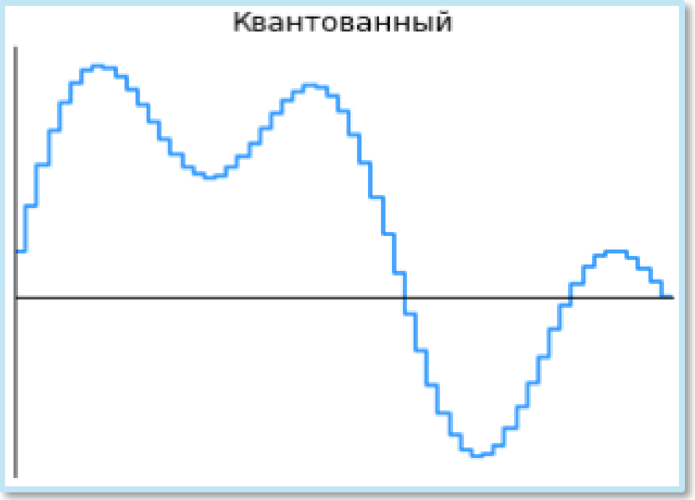
$$f_s = \frac{1}{\Delta t}$$



Аналоговые, цифровые и дискретные сигналы

- ***Квантованные сигналы*** принимают ряд конечных значений из диапазона непрерывных или дискретных величин
- Обычно сигналы квантуются по уровню, то есть по амплитуде

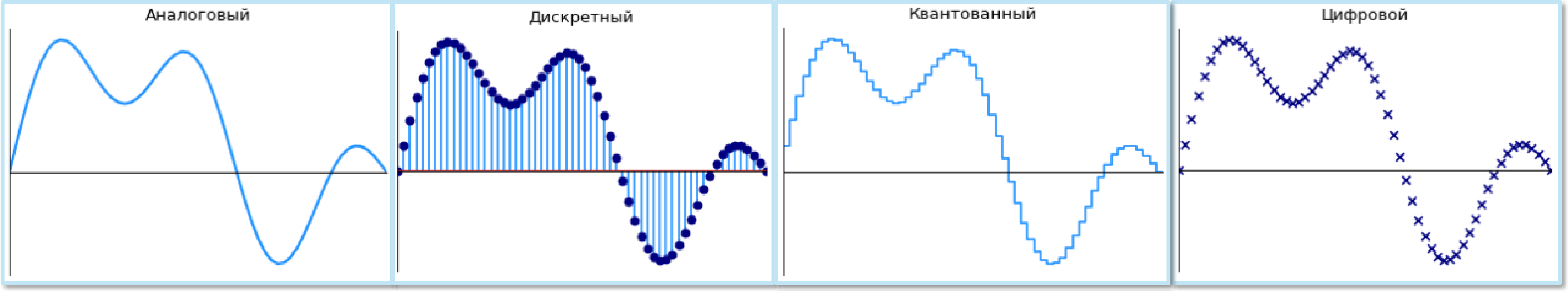
Процесс преобразования отсчётов сигнала в числа называется ***квантованием по уровню*** (***quantization***)



Аналоговые, цифровые и дискретные сигналы

Сигнал дискретный во времени, но не квантованный по уровню, называется ***дискретным*** (***discrete – time***)

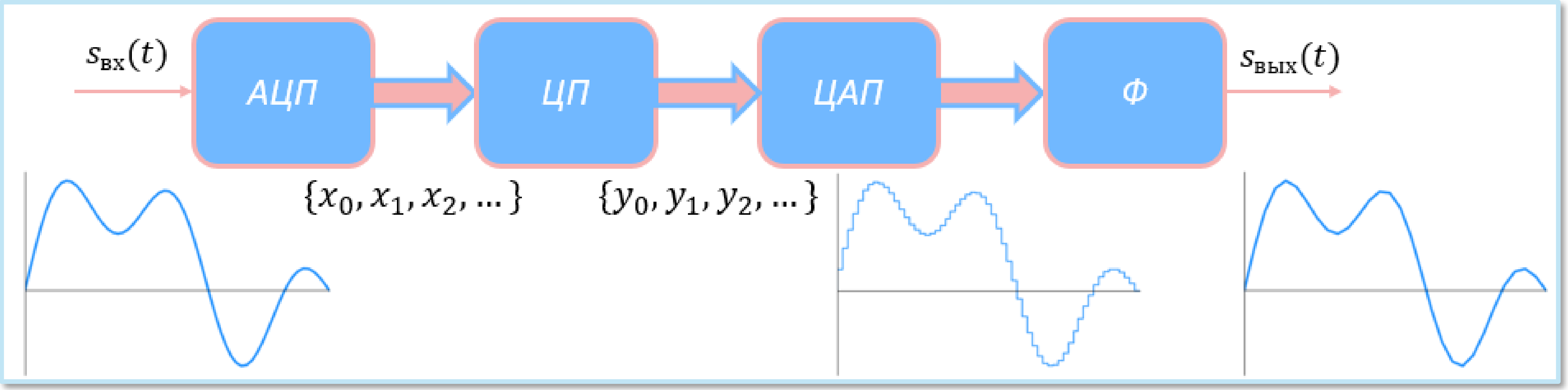
Сигнал дискретный во времени и квантованный по уровню называется цифровым (***digital***)



Преобразования сигналов

- Пусть на вход поступает аналоговый сигнал $s_{\text{вх}}(t)$
- Его дискретизация и квантование происходят в ***аналого-цифровом преобразователе*** (***АЦП, Analog-to-Digital Converter, ADC***)

Результатом работы АЦП является последовательность чисел, поступающая в цифровой процессор (ЦП), который выполняет обработку

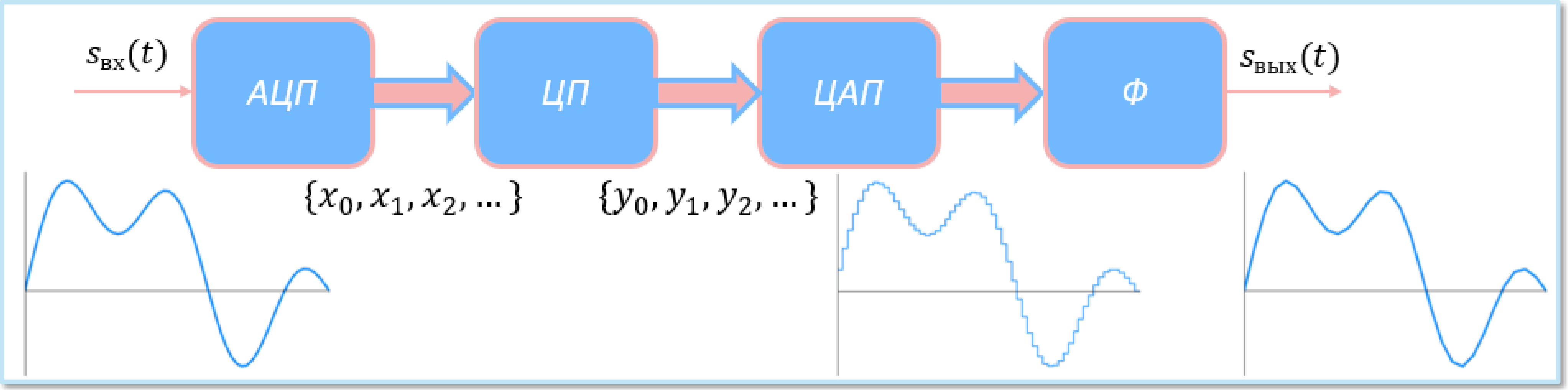


Преобразования сигналов

Результатом работы ЦП является новая последовательность чисел, представляющая собой отсчеты выходного сигнала

Аналоговый сигнал $s_{вых}(t)$ восстанавливается по этой последовательности с помощью ***цифро-аналогового преобразователя*** (***ЦАП***, ***Digital-to-Analog Converter***, ***DAC***)

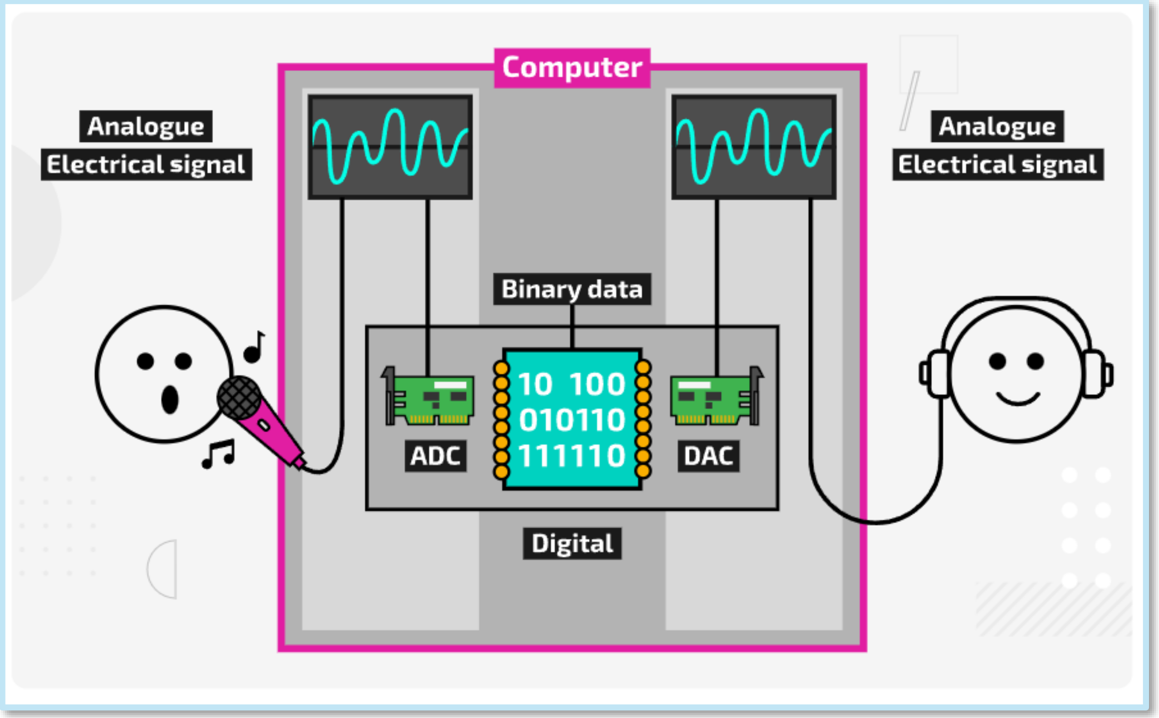
- Напряжение на выходе ЦАП имеет ступенчатую форму
- При необходимости оно может быть преобразовано в плавный сигнал с помощью низкочастотного ***сглаживающего фильтра*** Ф



Цифровое представление звука

Все современные методы обработки звука работают примерно одинаково:

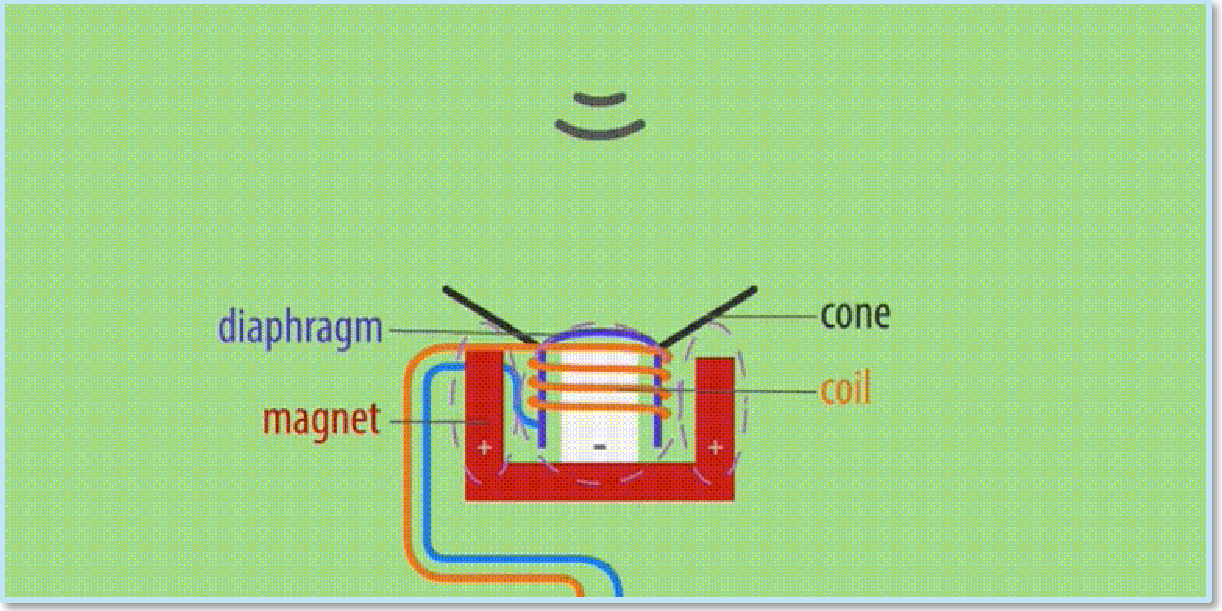
- Запись звука с помощью микрофона
- Сохранение звука на диске или компьютере
- Воспроизведение звука с помощью динамиков или наушников



Цифровое представление звука

Микрофон превращает звуковые волны в электрический сигнал

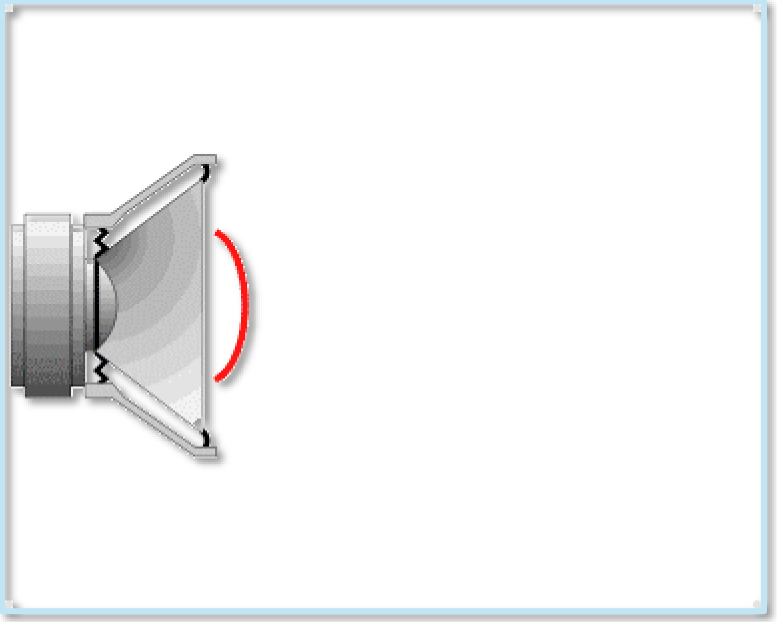
- ***Диафрагма*** внутри микрофона вибрирует по воздействием звуковые волн
- Эти вибрации преобразуются в электрический сигнал, которые изменяется в соответствии с частотой и амплитудой волн
- Затем этот сигнал обрабатывается АЦП и каждый отсчет сигнала записывается в виде двоичного числа



Цифровое представление звука

Для прослушивания аудио его необходимо преобразовать из цифровой формы в аналоговую

- ЦАП восстанавливает аналоговый сигнал из двоичного
- Затем аналоговый сигнал поступает на ***динамик,*** который вибрирует и создает звуковые волны в воздухе



Спектр сигнала и ряды Фурье

Спектр сигнала и ряды Фурье

Термин ***«преобразование Фурье»*** включает в себя 4 основных вида, каждый из которых соответствует определенному виду сигналов

Апериодические и непрерывные: распространяются на бесконечность и не имеют периода, используется стандартное преобразование Фурье

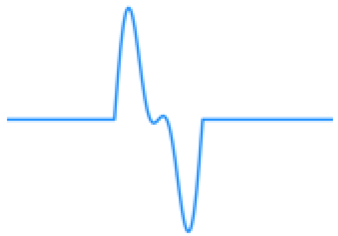
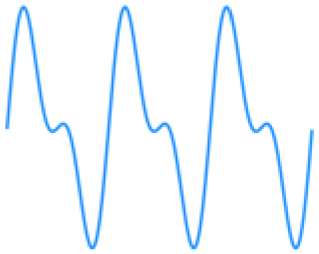
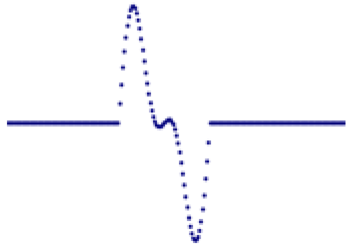
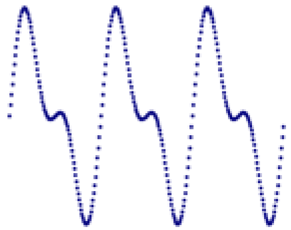
Периодические и непрерывные: повторяются с некоторым периодом на всей числовой оси, применяется разложение в ряд Фурье

Апериодические и дискретные: определены только в дискретных точках и не имеют периода, используется преобразование Фурье с дискретным временем

Периодические и дискретные: периодически повторяющиеся дискретные сигналы, применяется дискретное преобразование Фурье

Спектр сигнала и ряды Фурье

- В цифровой обработке используется только ***дискретное преобразование Фурье*** (***DFT***)
- В теории вы можете встретить первые три типа, но на практике будете использовать *DFT*

Тип преобразования	*Пример сигнала*
Преобразование Фурье (Fourier Transform)	
Ряды Фурье (Fourier Series)	
Преобразование Фурье с дискретным временем (Discrete Time Fourier Transform)	
Дискретное преобразование Фурье (Discrete Fourier Transform)	

Спектр сигнала и ряды Фурье

- Ряд Фурье был открыт французским математиком Жаном Фурье в начале *XIX* века, когда он исследовал уравнение теплопроводности
- С тех пор ряд Фурье стал одним из основных инструментов анализа периодических функций

Разложение в ряд Фурье – это способ представить периодический сигнал в виде суммы более простых функций, которые называются ***базисными***

В зависимости от конкретной формы базисных функций различают несколько форм записи ряда Фурье:

- ***Синусно-косинусная***
- ***Вещественная***
- ***Комплексная***

Спектр сигнала и ряды Фурье

Синусно-косинусная форма:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(k\omega_1 t) + b_k \sin(k\omega_1 t)\right), \qquad \omega_1 = \frac{2\pi}{T}$$

Кратные частоты $k\omega_1$ называются ***гармониками***, которые нумеруются в соответствии с индексом k

Коэффициенты ряда рассчитываются по формулам:

$$a_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)\cos(k\omega_1 t)\,dt, \qquad b_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)\sin(k\omega_1 t)\,dt, \qquad \frac{a_0}{2} = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)dt$$

$\frac{a_0}{2}$ имеет смысл среднего значения сигнала на периоде

Спектр сигнала и ряды Фурье

Спектр сигнала и ряды Фурье

Неудобство синусно-косинусной формы состоит в том, что в ней фигурируют 2 слагаемых – синус и косинус

Вещественная форма:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k \cos(k\omega_1 t + \varphi_k),$$

$$A_k = \sqrt{a_k^2 + b_k^2} \qquad \varphi_k = \begin{cases} -arctg\dfrac{b_k}{a_k}, & a_k \geq 0 \\ -arctg\dfrac{b_k}{a_k} \pm \pi, & a_k < 0 \end{cases}$$

Неудобство этой формы в том, что A_k и φ_k не рассчитываются напрямую, а выражаются через синусные или косинусные коэффициенты

Спектр сигнала и ряды Фурье

Комплексная форма получается из вещественной представлением косинуса в виде полусуммы комплексных экспонент:

$$\cos x = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$$

Комплексная форма:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \frac{A_k}{2} \left(e^{i(k\omega_1 t + \varphi_k)} + e^{-i(k\omega_1 t + \varphi_k)}\right)$$

Будем трактовать экспоненты со знаком «$-$» в показателе, как члены ряда с отрицательными номерами:

$$s(t) = \sum_{k=-\infty}^{\infty} \overline{C}_k e^{-ik\omega_1 t}$$

Спектр сигнала и ряды Фурье

Связь с коэффициентами вещественной формы:

$$\overline{C}_k = \frac{1}{2}A_k e^{i\varphi_k}, \qquad \mathrm{A}_k = 2\overline{C}_k, \qquad \varphi_k = \arg(\overline{C}_k)$$

Связь с коэффициентами синусно-косинусной формы:

$$\overline{C}_k = \frac{a_k}{2} - i\frac{b_k}{2}$$

$$a_k = 2Re(\overline{C}_k), \qquad b_k = -2Im(\overline{C}_k)$$

Отсюда сразу же следует формула расчёта коэффициентов $\overline{C}_k$:

$$\overline{C}_k = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)e^{-ik\omega_1 t}dt$$

Спектр сигнала и ряды Фурье

Совокупность амплитуд гармоник ряда Фурье называется ***амплитудным спектром***, а совокупность их фаз – ***фазовым спектром***

Если анализируемый сигнал $s(t)$ – вещественный, то его амплитудный и фазовый спектры обладают четной и нечётной симметрией соответственно:

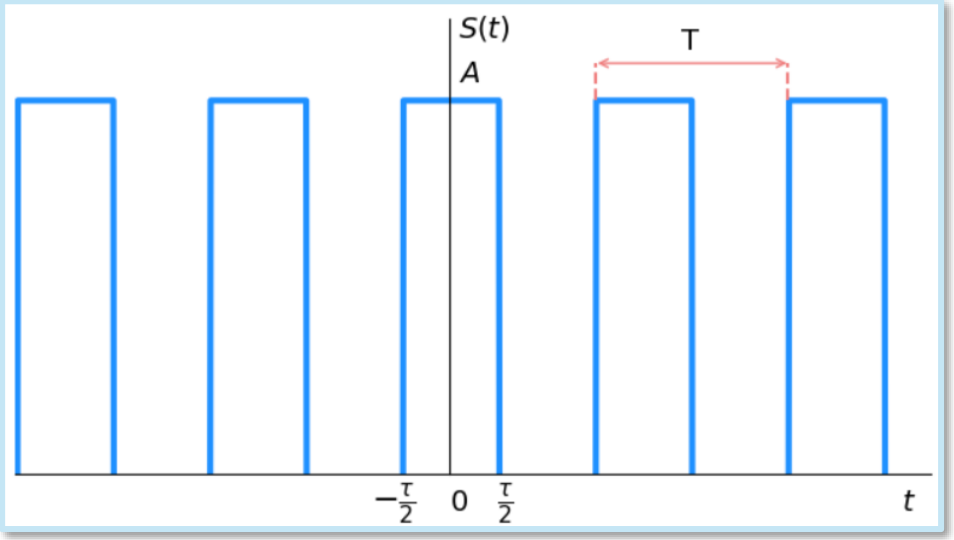
$$\mathrm{A}_{-k} = A_k, \qquad \varphi_{-k} = -\varphi_k$$

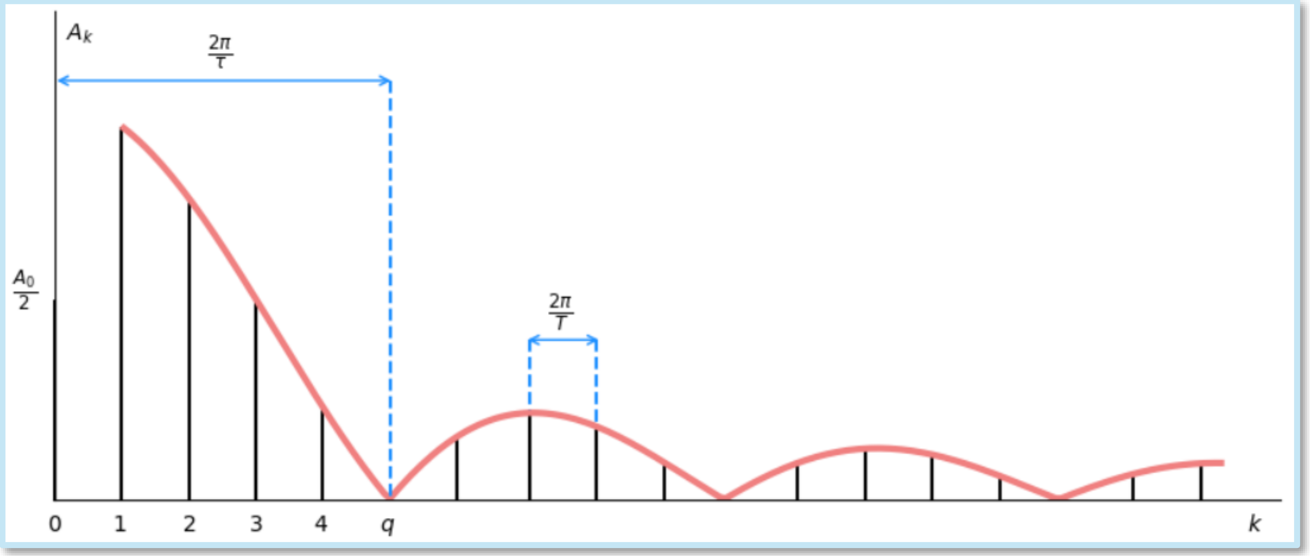
Коэффициенты комплексного ряда Фурье обладают комплексно-сопряженной симметрией:

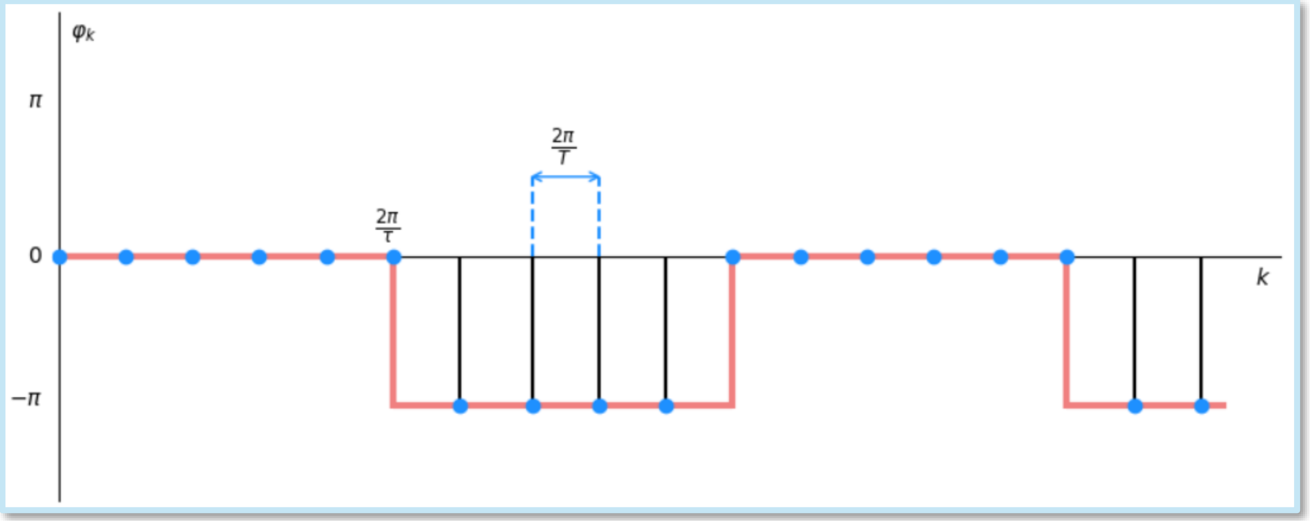
$$\overline{C}_{-k} = \overline{C}_k^*$$

Спектр сигнала и ряды Фурье

Синусно-косинусная форма

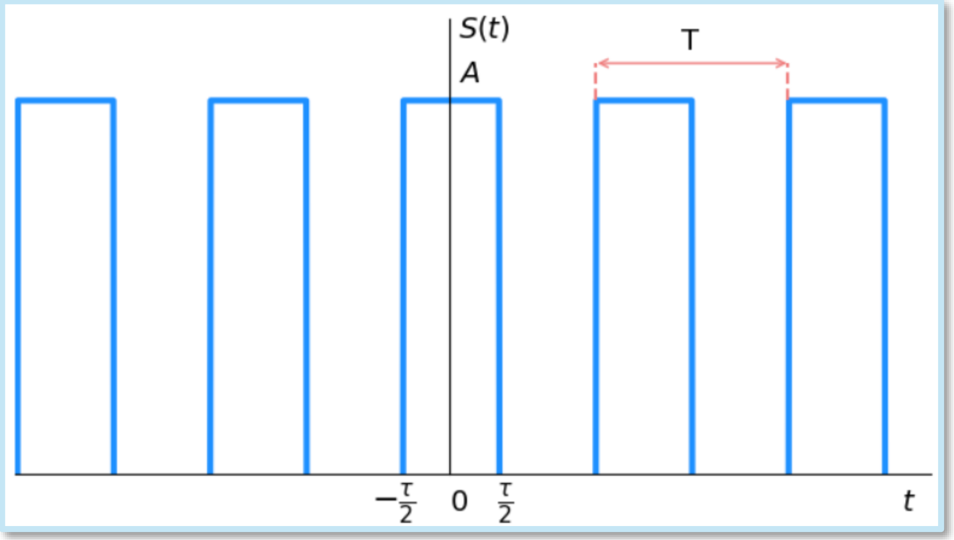


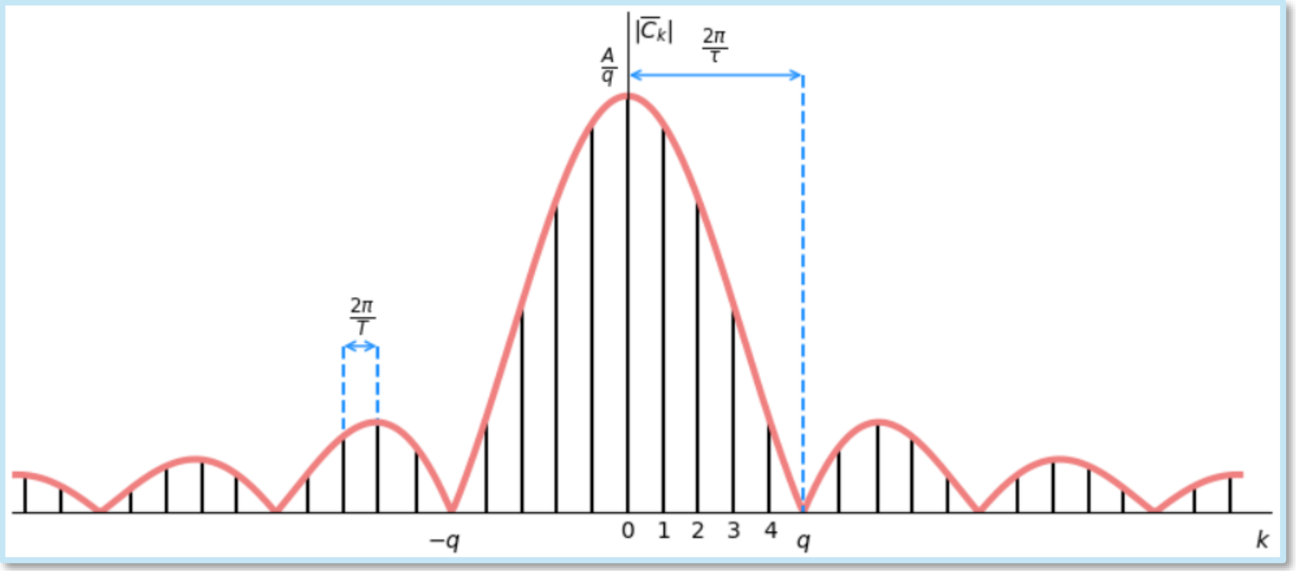


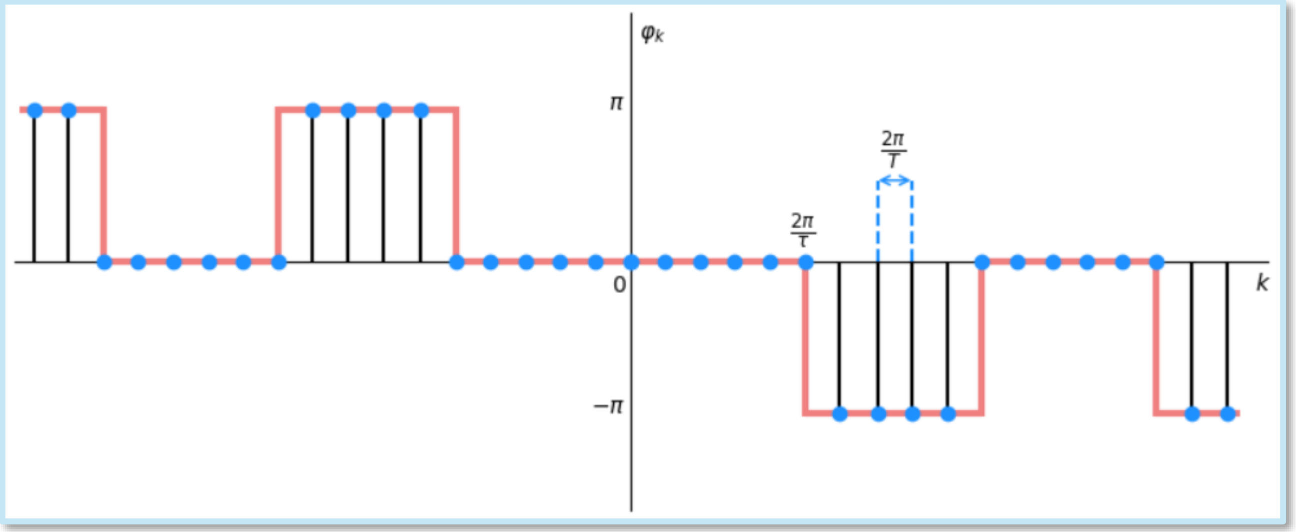


Спектр сигнала и ряды Фурье

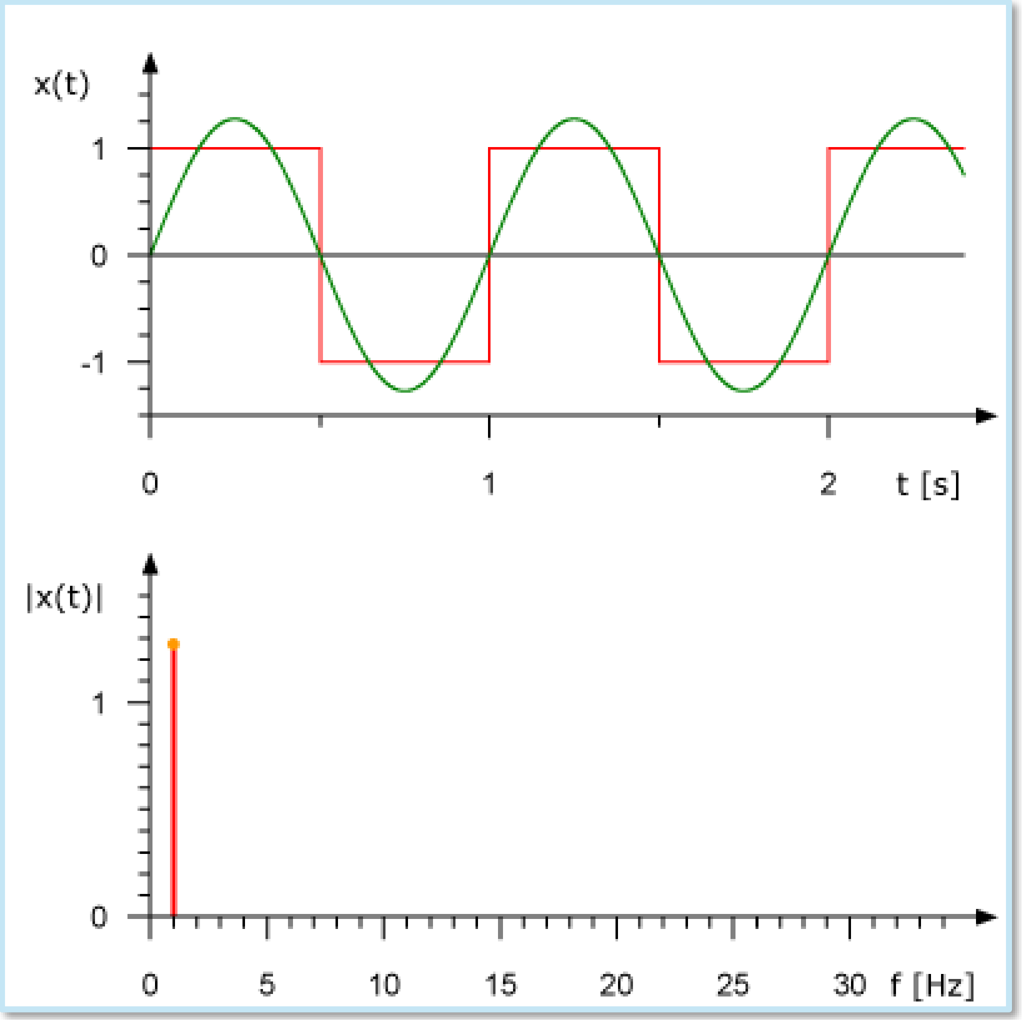
Комплексная форма







Спектр сигнала и ряды Фурье

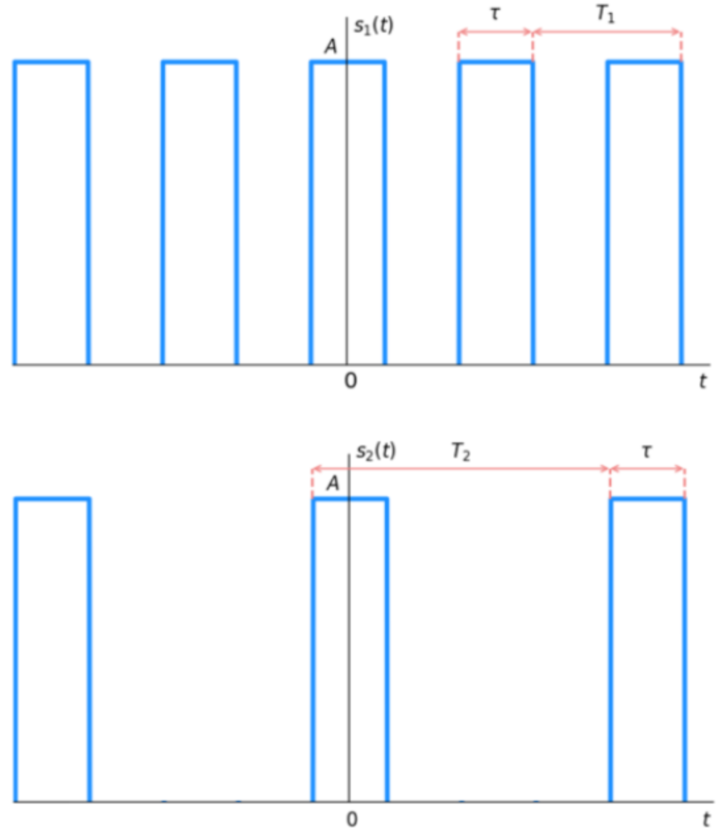


Преобразование Фурье

Преобразование Фурье

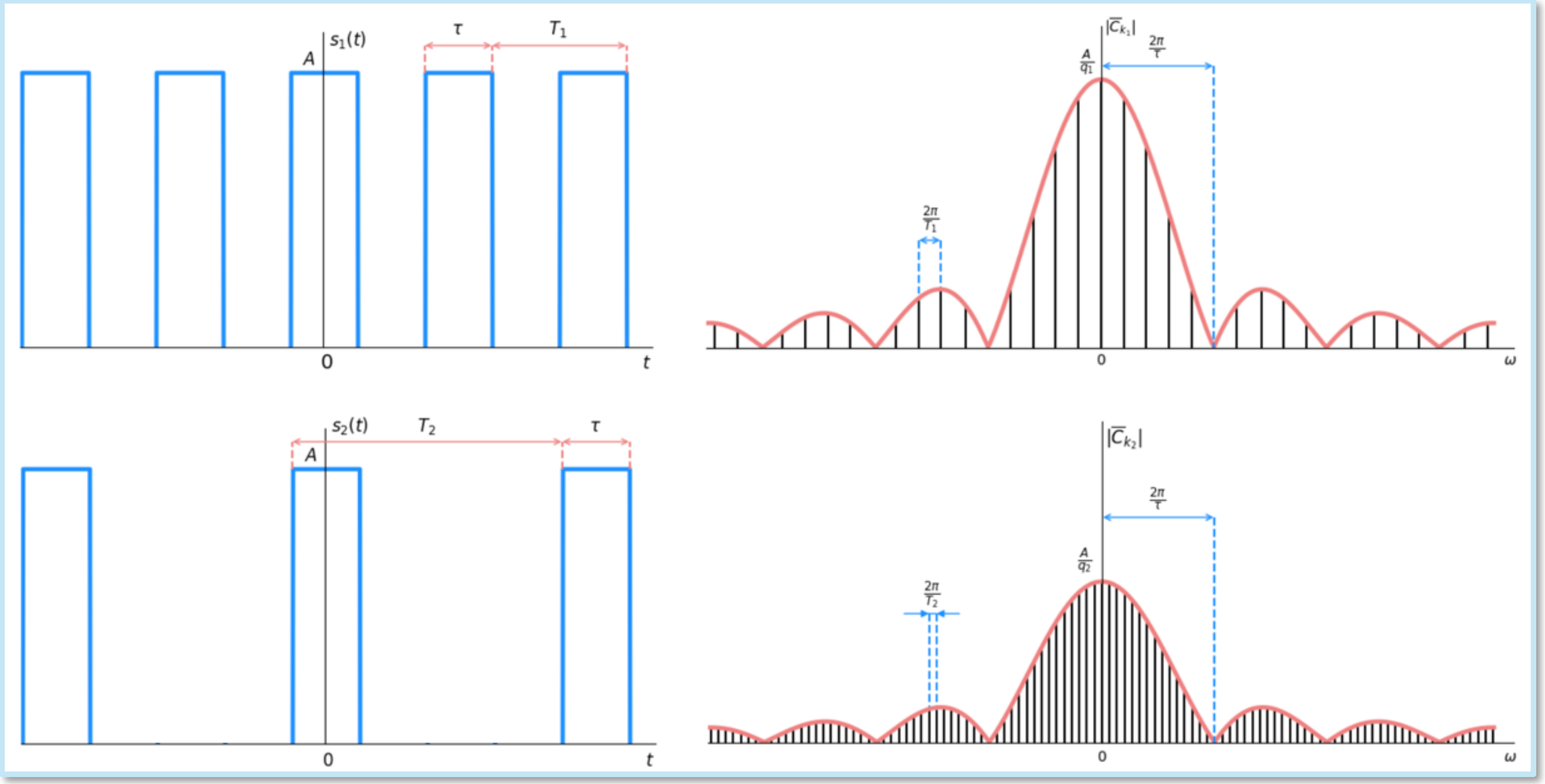
Идею рядов Фурье можно обобщить и на непериодические сигналы

- Будем увеличивать расстояние между импульсами, наша функция будет становится «менее периодической»
- В пределе $T \to \infty$ расстояние между частотами устремится к нулю, и мы получим непрерывный спектр



Преобразование Фурье

- Будем считать те же самые интегралы, но частоты станут ближе друг к другу
- Кроме того, уменьшатся амплитуды из-за множителя $\frac{1}{T}$
- Мы получим тот же ряд Фурье, но с другими параметрами



Преобразование Фурье

Формула для расчёта коэффициентов комплексного ряда Фурье меняется следующим образом:

- Частота перестаёт быть дискретной и становится непрерывной $k\omega_1 \rightarrow d\omega$
- Удаляется множитель $\frac{1}{\mathrm{T}}$

Результатом вычислений вместо коэффициентов ряда C_k становится функция частоты $S(\omega)$ – ***спектральная плотность***

В итоге мы получаем формулу прямого преобразования Фурье*:*

$$\overline{S}(\omega) = \int_{-\infty}^{\infty} s(t)e^{-i\omega t}dt$$

В формуле самого ряда Фурье суммирование заменяется интегрированием, появляется деление на 2π, и мы получаем обратное преобразование Фурье:

$$s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{S}(\omega)e^{i\omega t}d\omega$$

Свойства преобразования Фурье

Линейность: линейная комбинация сигналов имеет спектр в виде линейной комбинации их спектральных функций:

$$\text{если } s(t) = \alpha f(t) + \beta g(t), \qquad \text{то} \qquad \overline{S}(\omega) = \alpha \overline{F}(\omega) + \beta \overline{G}(\omega)$$

Задержка:

Пусть $\tau -$ время задержки:

$$s(t) = f(t - \tau),$$

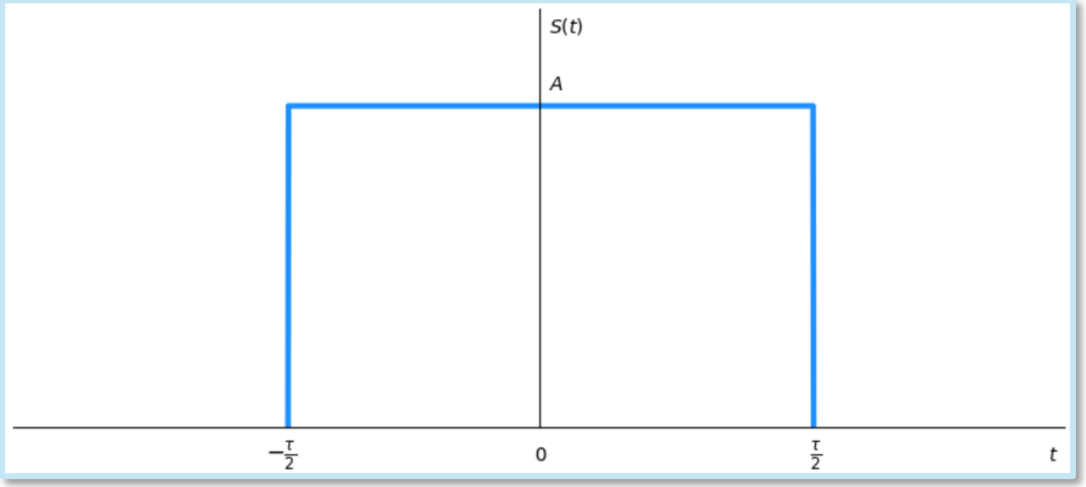
тогда спектральная плотность изменится следующим образом:

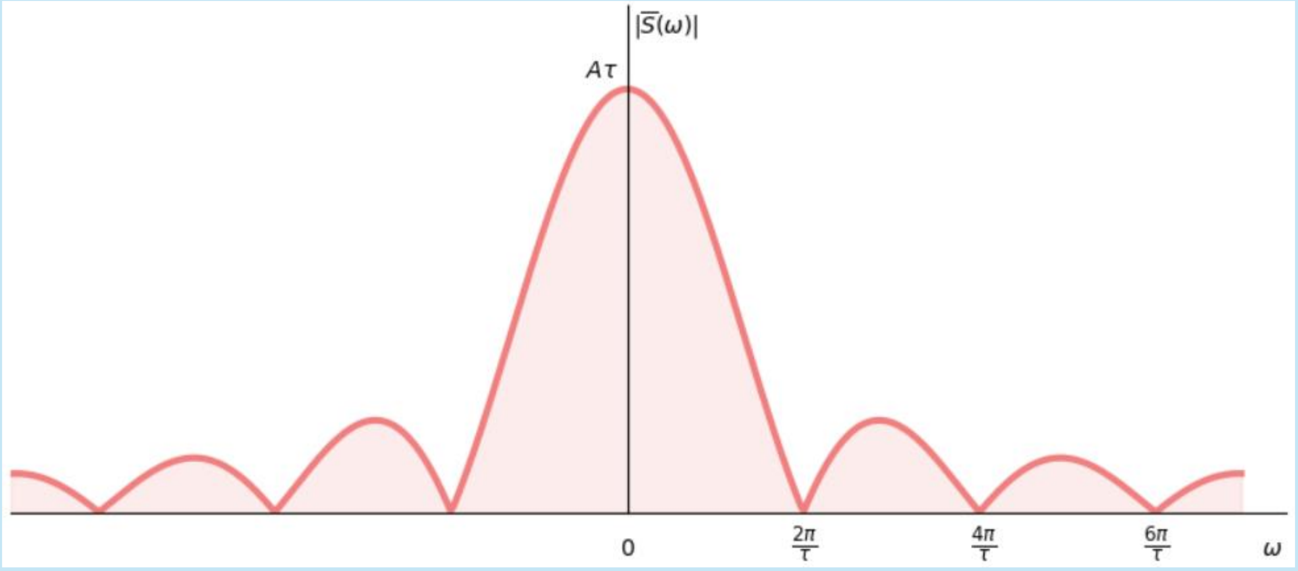
$$\overline{S}(\omega) = \int_{-\infty}^{\infty} f(t-\tau) e^{-i\omega t} dt = \int_{-\infty}^{\infty} f(t-\tau) e^{-i\omega(t-\tau)} d(t-\tau) e^{-i\omega\tau} = \overline{F}(\omega) e^{-i\omega\tau}$$

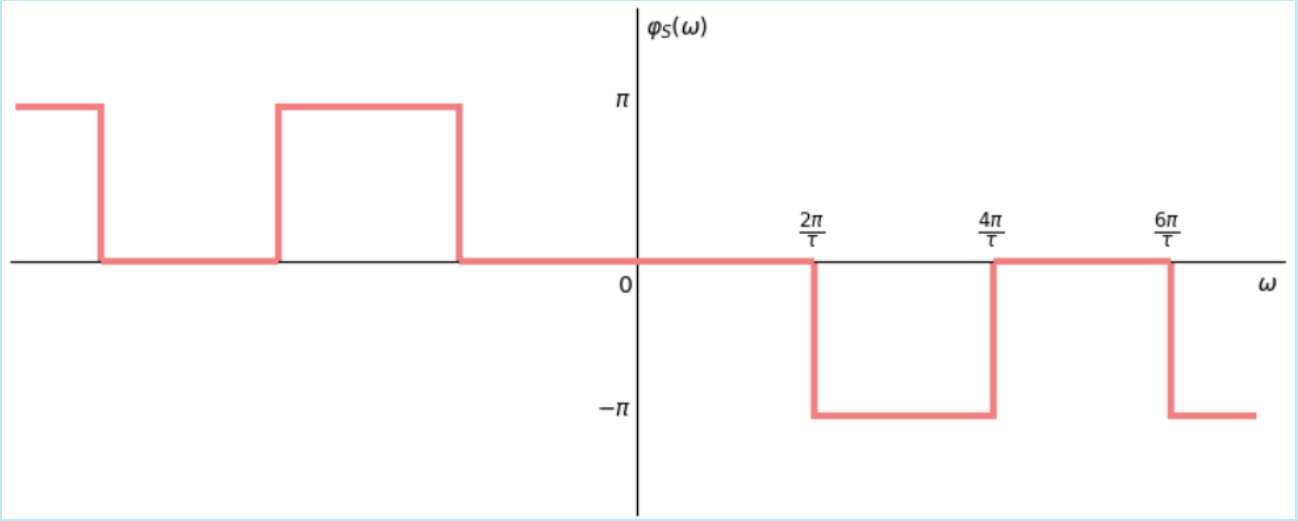
При задержке амплитудный спектр сигнала не меняется, а фазовый приобретает дополнительное слагаемое $-\omega\tau$

Преобразование Фурье

$$\overline{S}(\omega) = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} A e^{-i\omega t} dt = \frac{2A}{\omega} sin\left(\frac{\omega\tau}{2}\right) = A\tau \frac{sin\left(\frac{\omega\tau}{2}\right)}{\frac{\omega\tau}{2}}$$

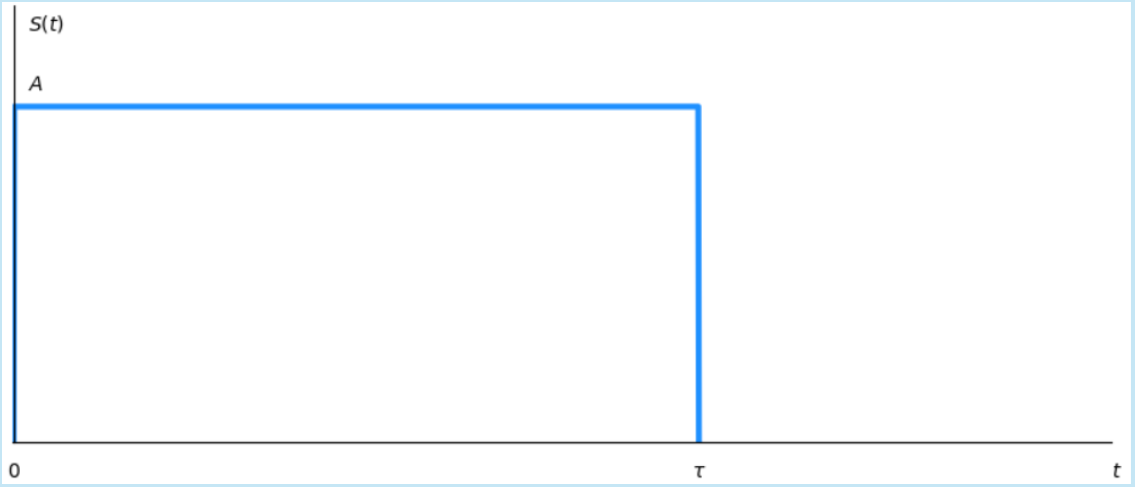


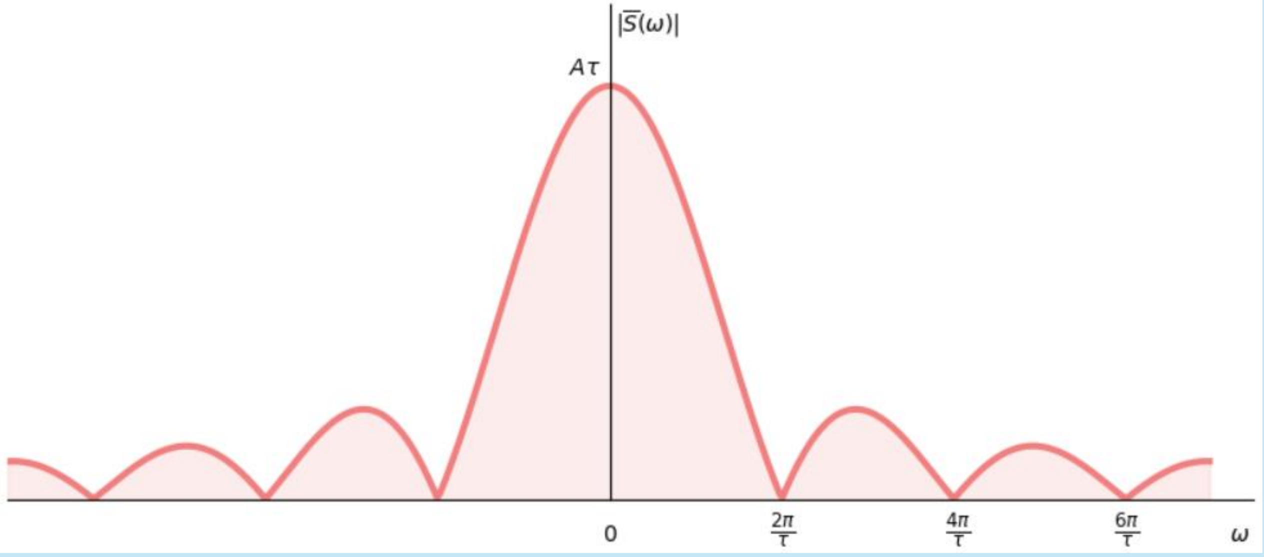


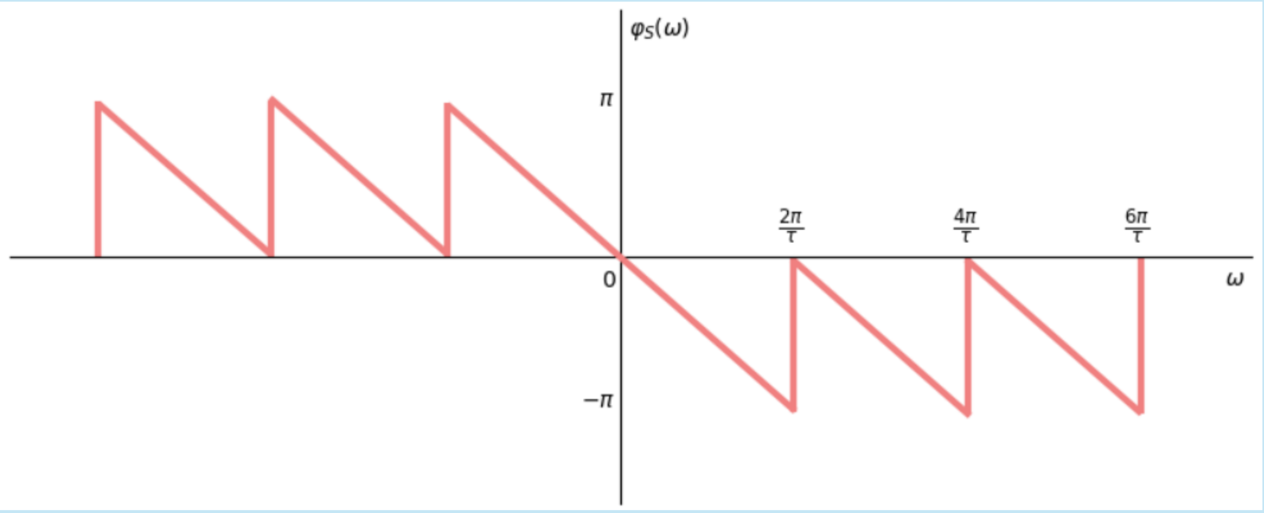


Преобразование Фурье

$$\overline{S}(\omega) = \int_0^{\tau} A e^{-i\omega t} dt = \frac{2}{i\omega}\left(1 - e^{-i\omega\tau}\right) = A\tau \frac{\sin\left(\frac{\omega\tau}{2}\right)}{\frac{\omega\tau}{2}} e^{-i\frac{\omega\tau}{2}}$$







Свойства преобразования Фурье

Изменение масштаба времени: если изменить длительность сигнала $f(t)$, сохраняя его форму, то новый сигнал $s(t)$ будет иметь вид:

$$s(t) = f(at)$$

При этом спектр будет вычисляться следующим образом:

$$\overline{S}(\omega) = \int_{-\infty}^{\infty} f(at)e^{-i\omega t}dt = \frac{1}{a}\int_{-\infty}^{\infty} f(at)e^{-i\frac{\omega}{a}at}d(at) = \frac{1}{a}\overline{F}\left(\frac{\omega}{a}\right)$$

Если учесть случай $a < 0$:

$$\overline{S}(\omega) = \frac{1}{|a|}\overline{F}\left(\frac{\omega}{a}\right), \qquad\qquad a \neq 0$$

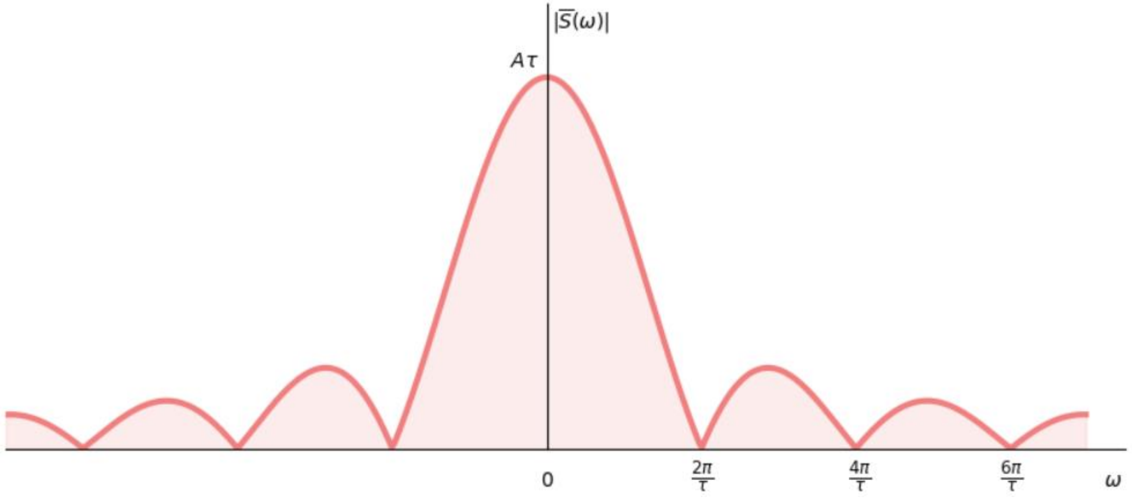
Изменение длительности сигнала приводит к изменению ширины спектра, причем эта зависимость обратно пропорциональная

Преобразование Фурье

- Спектр данного сигнала простирается до бесконечности, постепенно затухая

- Для определения протяженности спектра вводят понятие ***эффективной*** (***практической***) ***ширины спектра***

- При лепестковом характере спектра шириной спектра можно считать ширину главного лепестка



Преобразование Фурье

Произведение эффективной длительности сигнала на ширину его спектра называется базой сигнала, оно остаётся постоянным и равным некоторой константе:

$$B = \Delta t \cdot \Delta f$$

Для сигналов простой формы, не имеющих сложной внутриимпульсной структуры, величина базы составляет несколько единиц

Длительность сигнала и ширина его спектра связаны соотношением неопределенности, которое утверждает, что их произведение не может быть меньше единицы:

$$B > 1$$

Можно создать сигнал большой длительности, одновременно имеющий и широкий спектр, но короткий сигнал с узким спектром существовать не может

Дискретизация и теорема Котельникова

Спектр дискретного сигнала

Дискретный сигнал является последовательностью отсчётов, поэтому для анализа его спектра нужно сопоставить ей некоторую функцию

Обычно функция представляется в виде δ-функций с $T = 1$ и соответствующими множителями и задержками:

$$s(t) = \sum_{k=-\infty}^{\infty} x(k)\delta(t-k)$$

$\delta(t-k)$ – δ-функция, которая «активируется» в моменты времени $t - k$

Спектр дискретного сигнала

Преобразование Фурье линейно, спектр δ-функции равен единице, задержка сигнала приводит к умножению спектра на комплексную экспоненту:

$$\overline{S}(\omega) = \sum_{k=-\infty}^{\infty} x(k)e^{-i\omega k}$$

Мы считаем, что частота дискретизации $f_s = 1 \rightarrow \omega_s = 2\pi f_s = 2\pi$

Спектр $\overline{S}(\omega)$ нашей функции $s(t)$ – это и есть ***преобразование Фурье с дискретным временем*** $(\boldsymbol{Discrete - time\ Fourier\ Transform})$

Спектр дискретного сигнала

Изменим частоту на $\omega \pm 2\pi$:

$$\overline{S}(\omega \pm 2\pi) = \sum_{k=-\infty}^{\infty} x(k) e^{-i(\omega \pm 2\pi)k} = \sum_{k=-\infty}^{\infty} x(k) e^{-i\omega k} e^{-i(\pm 2\pi)k}$$

$e^{-i2\pi k} = 1$ для любого целого k:

$$\overline{S}(\omega \pm 2\pi) = \sum_{k=-\infty}^{\infty} x(k) e^{-i\omega k} = \overline{S}(\omega)$$

Вывод:

- Спектр $\overline{S}(\omega)$ является периодическим с периодом 2π

Связь спектров дискретного и аналогового сигналов

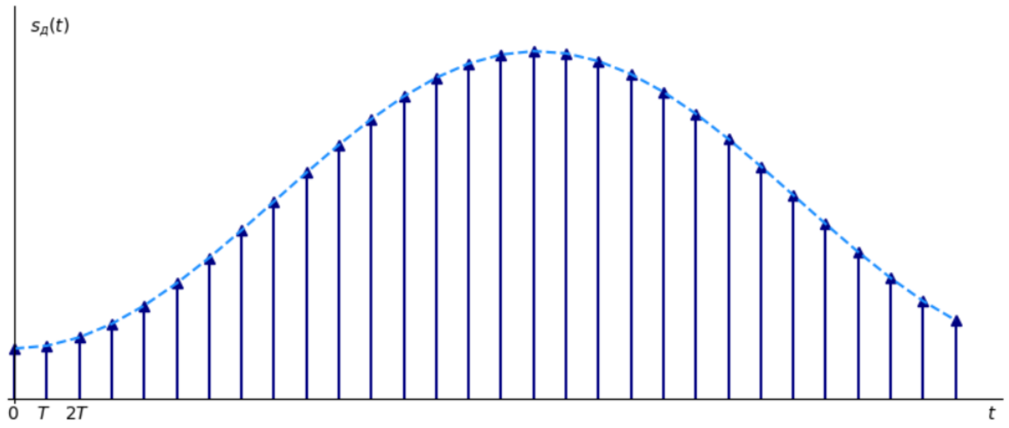
Пусть значения $x(k)$ являются отсчётами аналогового сигнала $s(t)$, взятыми с периодом T:

$$x(k) = s(kT)$$

Как спектр дискретного сигнала связан со спектром аналогового?

Рассмотрим дискретизированный сигнал в виде последовательности δ-функций, взвешенной значениями отсчетов $s(kT)$ аналогового сигнала $s(t)$:

$$s_{\text{д}}(t) = \sum_{k=-\infty}^{\infty} s(kT)\delta(t - kT)$$



Связь спектров дискретного и аналогового сигналов

Так как функция $\delta(t - kT)$ равна нулю везде, кроме моментов $t = kT$, можно заменить $s(kT)$ на исходный непрерывный сигнал $s(t)$:

$$s_{д}(t) = s(t) \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

Выражение с суммой является периодическим сигналом, поэтому можно разложить его в ряд Фурье и получить коэффициенты ряда:

$$\overline{C}_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t) e^{-i\omega_n t} dt = \frac{1}{T}$$

В итоге, периодическая последовательность δ-функций может быть представлена в виде комплексного ряда Фурье:

$$\sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{i\omega_n t},$$

- $\omega_n = \frac{2\pi n}{T}$

Связь спектров дискретного и аналогового сигналов

В итоге, мы получаем:

$$s_{д}(t) = \frac{s(t)}{T} \sum_{n=-\infty}^{\infty} e^{i\omega_n t} = \frac{1}{T} \sum_{n=-\infty}^{\infty} s(t) e^{i\omega_n t}$$

Умножение сигнала на $e^{i\omega_n t}$ соответствует сдвигу спектральной функции на ω_n, тогда спектр дискретизированного сигнала можно записать в следующем виде:

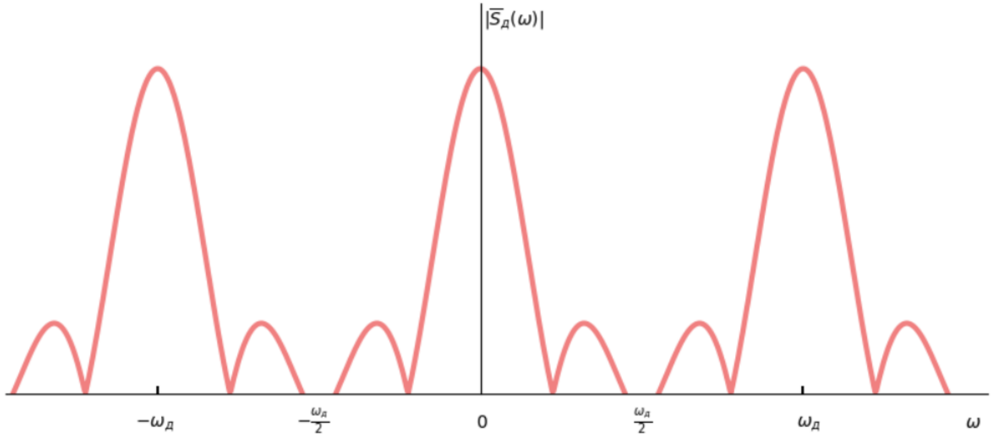
$$\overline{S}_{д}(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \overline{S}(\omega - \omega_n) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \overline{S}\left(\omega - \frac{2\pi n}{T}\right)$$

Связь спектров дискретного и аналогового сигналов

$$\overline{S}_{д}(\omega) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \overline{S}(\omega - \frac{2\pi n}{T})$$

Спектр дискретизированного сигнала представляет собой бесконечный ряд сдвинутых копий спектра исходного непрерывного сигнала $s(t)$

Расстояние между соседними копиями спектра равно частоте дискретизации $\omega_{д} = \frac{2\pi}{T}$



- Периодический сигнал → дискретный спектр
- Периодический спектр → дискретный сигнал

Теорема Котельникова

Любой сигнал $s(t)$, спектр которого не содержит частот больше некоторого верхнего значения $f_в$, может быть без потерь информации представлен своими дискретными отсчетами $\{s(kT)\}$, взятыми с интервалом T, удовлетворяющим следующему неравенству:

$$T < \frac{1}{2f_в}$$

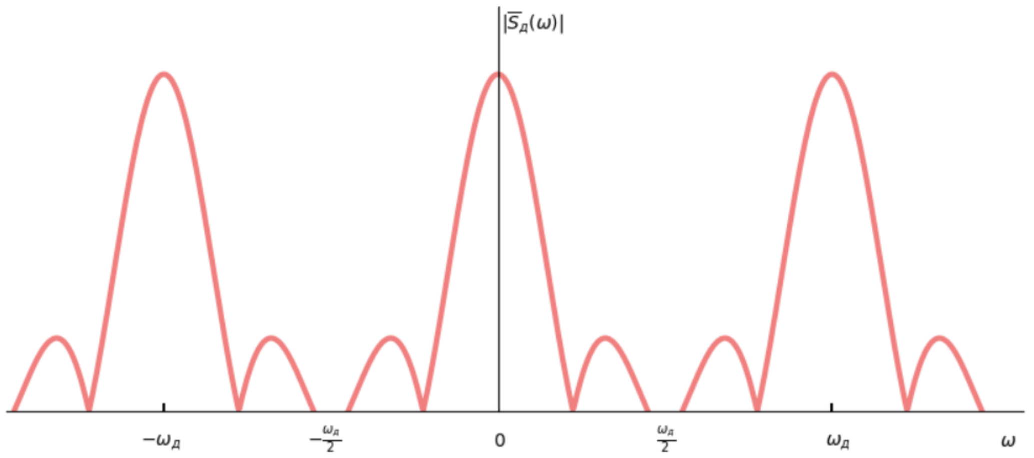
В зарубежных источниках данная теорема называется ***теоремой Найквиста-Шеннона*** (***Nyquist – Shannon theorem***) или теоремой отсчётов (***sampling theorem***)

Теорема Котельникова

Другие формулировки:

- Любой аналоговый сигнал может быть полностью восстановлен по своим дискретным отчётам, взятым с частотой $f > 2f_{\text{в}}$, где $f_{\text{в}}$ – максимальная частота спектра сигнала

- Для заданной частоты дискретизации f_s вы не сможете точно восстановить сигнал, содержащий частоты, большие или равные $\frac{f_s}{2}$



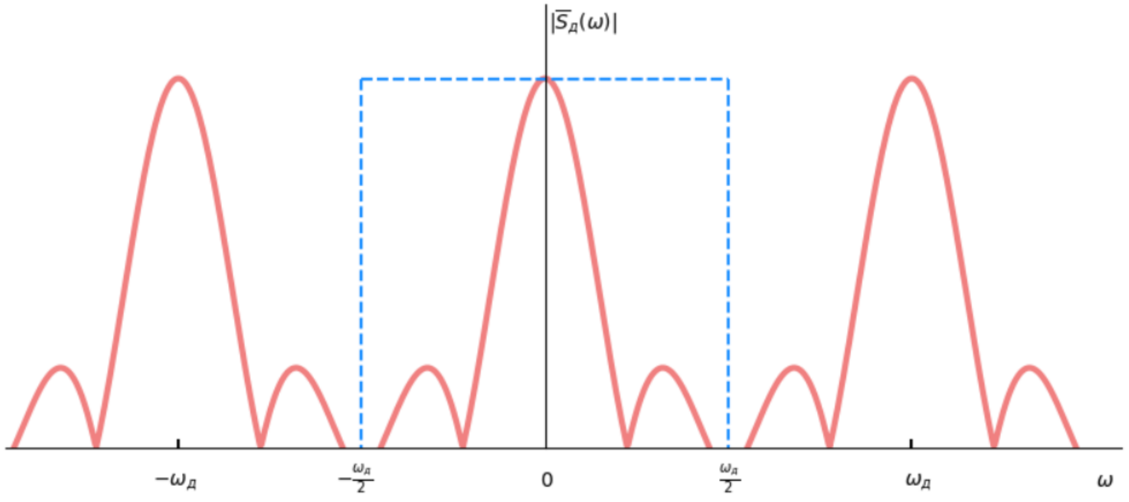
Значению $\frac{f_s}{2}$ дали особое название – ***частота Найквиста*** (***Nyquist frequency***)

Теорема Котельникова

Почему частота дискретизации должна быть именно в 2 раза больше? Почему не 5 или 100?

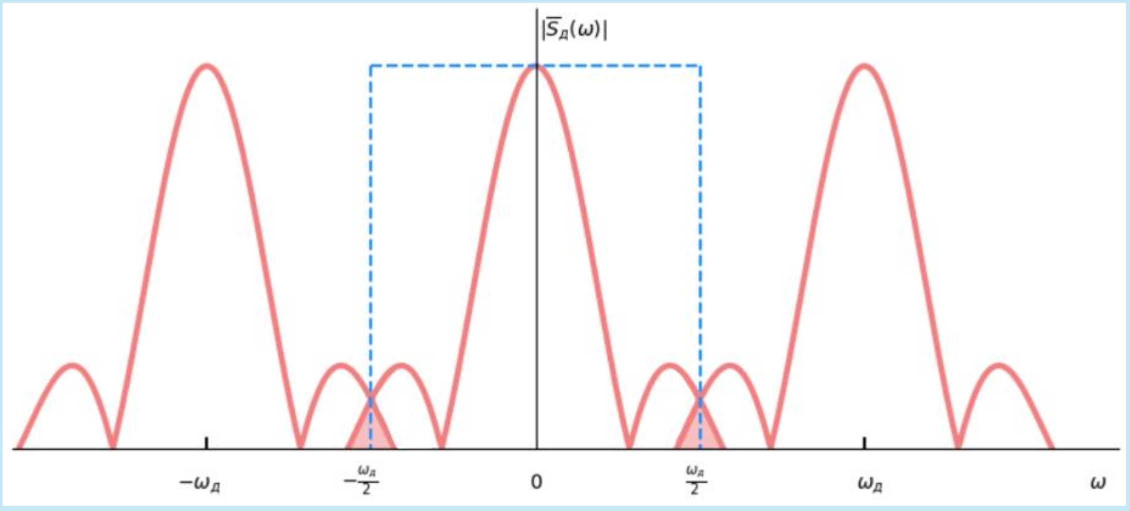
- Чтобы восстановить непрерывный сигнал по его отсчётам, необходимо пропустить дискретный сигнал через фильтр нижних частот (ФНЧ) с частотой среза, равной половине частоты дискретизации

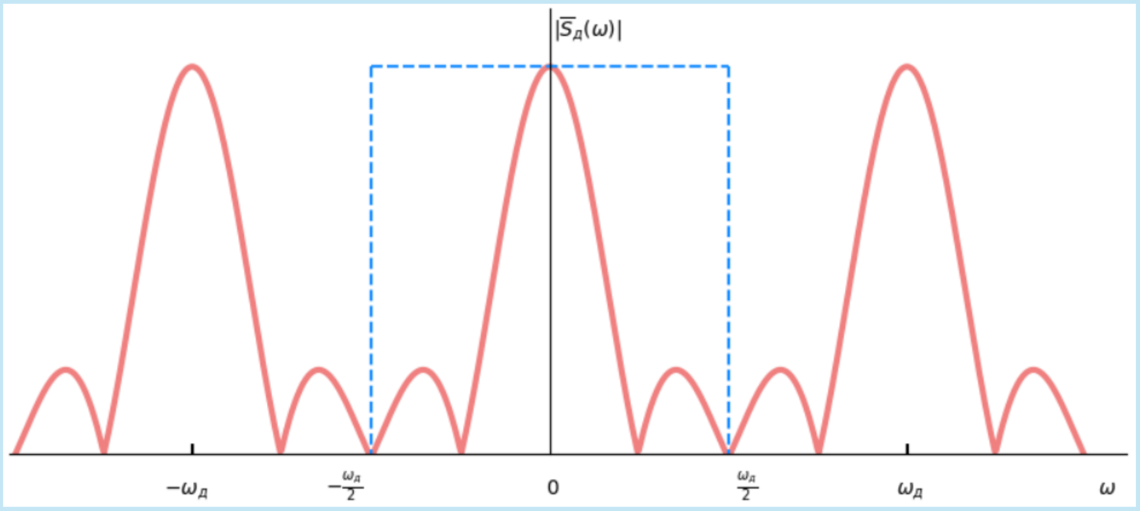
- Точное восстановление сигнала возможно, если сдвинутые копии спектра не перекрываются, то есть когда $\omega_д > 2\omega_в$

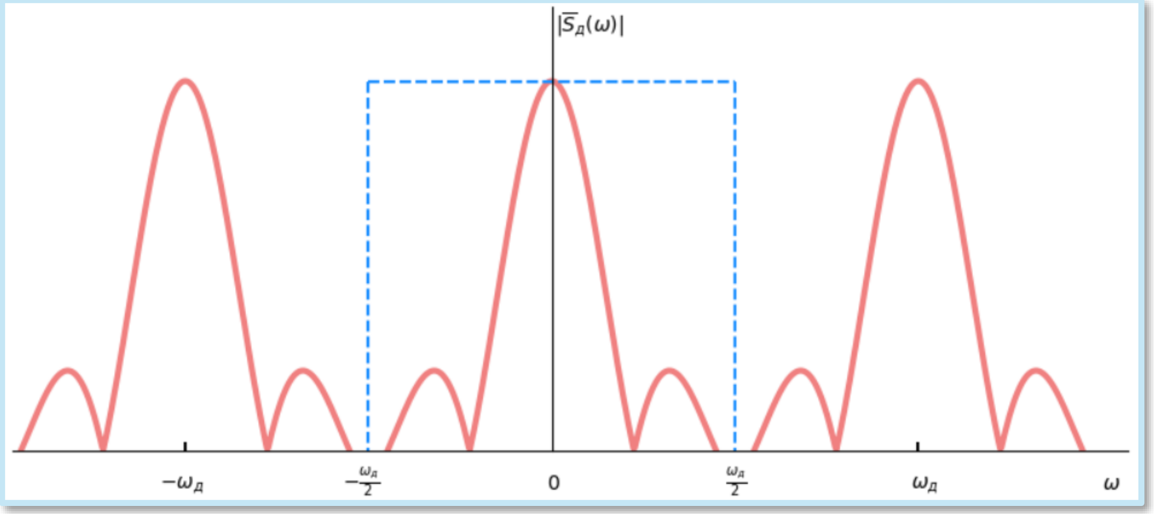


Теорема Котельникова

Если это условие не выполняется, сдвинутые копии спектра будут накладываться друг на друга, что приведет к искажениям при восстановлении непрерывного сигнала







Дискретное преобразование Фурье

Дискретное преобразование Фурье

Мы рассмотрели преобразование Фурье с дискретным временем, которое даёт нам непрерывный спектр $\overline{S}(\omega)$

Мы не можем вычислить или сохранить эту непрерывную функцию

Идея:

Сделаем исходный сигнал еще и периодическим

Дискретное преобразование Фурье

Пусть у нас есть периодическая последовательность отсчётов $\{x(k)\}$ с периодом N и соответствующий ей сигнал из смещенных δ-функций:

$$x(k+N) = x(k), \qquad s(t) = \sum_{k=-\infty}^{\infty} x(k)\delta(t-kT)$$

Сигнал $s(t)$ тоже будет периодическим с периодом NT

- Поскольку сигнал $s(t)$ дискретный, то его спектр будет периодическим с периодом $\frac{2\pi}{T}$
- Поскольку этот же сигнал периодический, его спектр будет дискретным с расстоянием между гармониками $\frac{2\pi}{NT}$

Дискретное преобразование Фурье

Найдем спектр этого сигнала, разложив его в ряд Фурье:

$$\overline{X}(n) = \frac{1}{NT}\int_0^{NT} s(t)e^{-i\omega_n t}dt = \frac{1}{NT}\int_0^{NT}\sum_{k=0}^{N-1} x(k)\delta(t-kT)\,e^{-i\omega_n t}dt =$$

$$= \frac{1}{NT}\sum_{k=0}^{N-1} x(k)\int_0^{NT}\delta(t-kT)e^{-i\omega_n t}dt = \frac{1}{NT}\sum_{k=0}^{N-1} x(k)e^{-i\omega_n kT} = \frac{1}{NT}\sum_{k=0}^{N-1} x(k)e^{-i\frac{2\pi nk}{N}}$$

При рассмотрении дискретных последовательностей обычно не привязываются ко действительному масштабу времени, поэтому множитель $\frac{1}{T}$ удаляют ($f_s = 1$) и $\frac{1}{N}$ тоже обычно удаляют и получают дискретное преобразование Фурье (***Discrete Fourier Transform, DFT***):

$$\overline{X}(n) = \sum_{k=0}^{N-1} x(k)e^{-i\frac{2\pi nk}{N}}$$

Дискретное преобразование Фурье

Существует и ***обратное дискретное преобразование Фурье***:

$$x(k) = \frac{1}{N}\sum_{n=0}^{N-1} \bar{X}(n)e^{i\frac{2\pi nk}{N}}$$

Если представить произвольный отсчёт ДПФ в виде суммы действительной и мнимой частей*:*

$$X(n) = X_{real}(n) + iX_{imag}(n),$$

То амплитуда и фаза вычисляются следующим образом*:*

$$X_{mag}(n) = |X(n)| = \sqrt{X_{real}^2(n) + X_{imag}^2(n)}$$

$$X_{\varphi}(n) = arctan\frac{X_{imag}(n)}{X_{real}(n)}$$

Смысл формулы ДПФ

Перепишем комплексную экспоненту через формулу Эйлера:

$$\overline{X}(n) = \sum_{k=0}^{N-1} x(k)\left[\cos\left(\frac{2\pi kn}{N}\right) - i\sin\left(\frac{2\pi kn}{N}\right)\right],$$

- $\overline{X}(n)$ – n-ая компонента ДПФ
- n – индекс ДПФ в частотной области, $n = 0, 1, 2, \dots, N-1$
- $x(k)$ – последовательность отсчётов $x(1)$, $x(2)$, $x(3)$,…
- k – временной индекс входных отсчётов, $k = 0, 1, 2, \dots, N-1$
- N – количество отсчётов входной последовательности, оно же количество отсчётов ДПФ

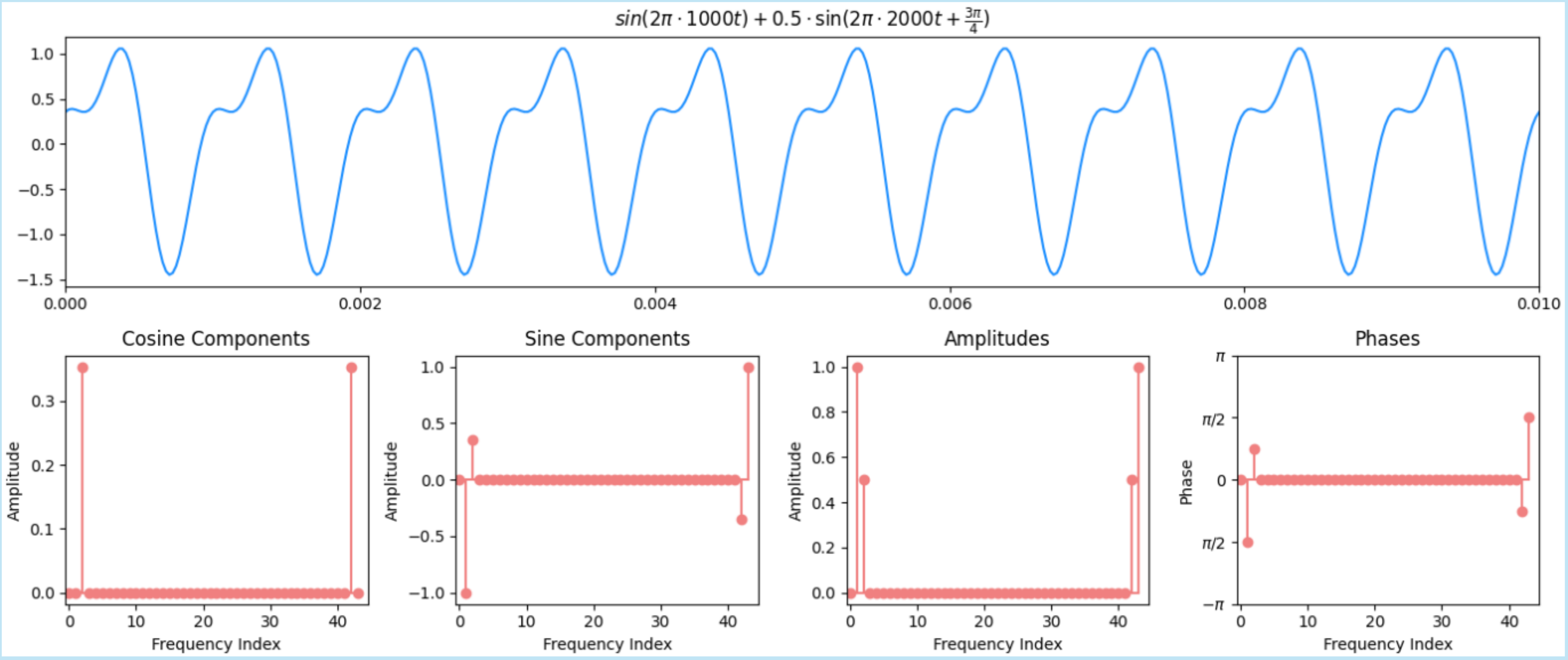
Тык

Свойства ДПФ

Возьмём сигнал:

$$x(t) = \sin(2\pi \cdot 1000t) + 0.5 \cdot \sin(2\pi \cdot 2000t + \frac{3\pi}{4})$$

с частотой дискретизации $f_s = 44000$ Гц и вычислим его ДПФ по 44 точкам



Свойства ДПФ

Симметрия:

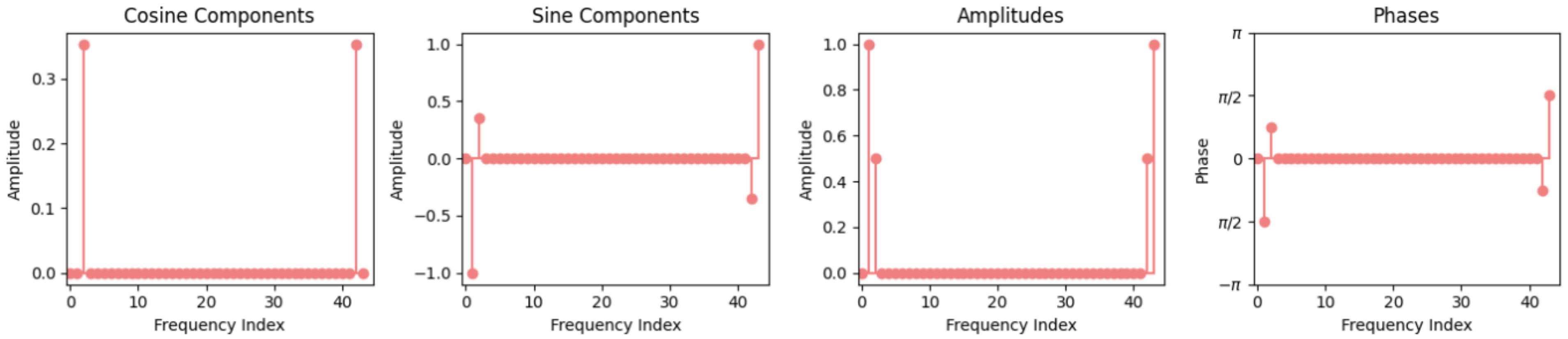
Если входной сигнал $x(k)$ действительный, то

- Амплитуда ДПФ с номером n равна амплитуде ДПФ с номером $N-n$:

$$|X(n)| = |X(N-n)|$$

- Фаза n-го отсчёта равна фазе $(N-n)$-го отсчёта, взятой с противоположным знаком:

$$X_\varphi(n) = -X_\varphi(N-n)$$



Свойства ДПФ

Если входная последовательность ДПФ действительна, то $X(n)$ и $X(N-n)$ образуют комплексно-сопряженную пару:

$$X(n) = X^*(N-n)$$

Действительная часть $X(n)$ будет обладать четной симметрией, а мнимая - нечетной

Если мы делаем ДПФ по N точкам, мы получим N комплексных отсчётов, но только $\frac{N}{2}+1$ из них будут независимыми

- Чтобы получить ДПФ действительной последовательности $x(k)$, нам нужно вычислить только первые $\frac{N}{2}+1$ отсчётов
- Оставшиеся коэффициенты не дают дополнительной информации о спектре

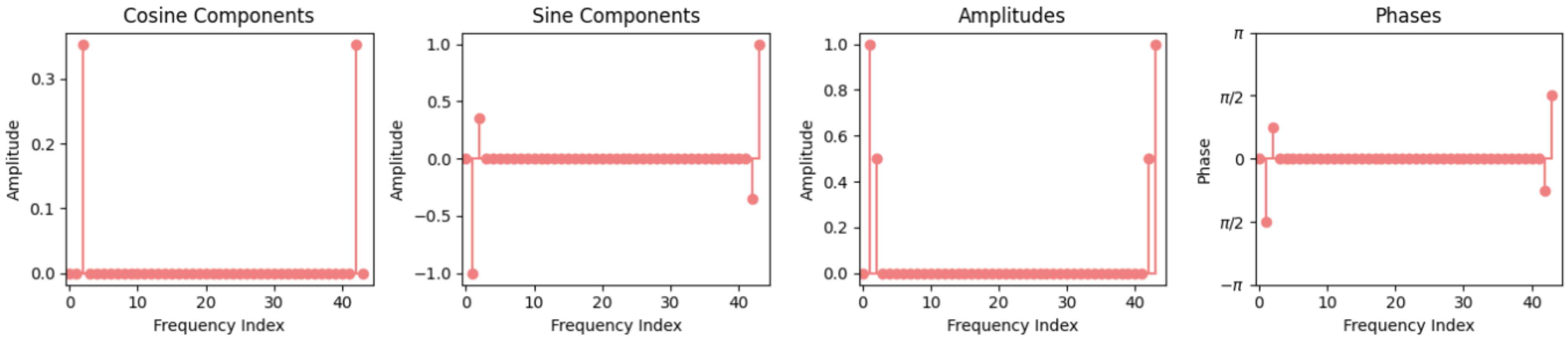
Свойства ДПФ

Частотное разрешение:

Если мы дискретизируем непрерывный сигнал с частотой f_s Гц, а затем считаем его ДПФ по N точкам, то его n –ая частота вычисляется следующим образом:

$$f_{analysis}(n) = \frac{nf_s}{N}$$

Расстояние между соседними отсчетам равно $\frac{f_s}{N}$

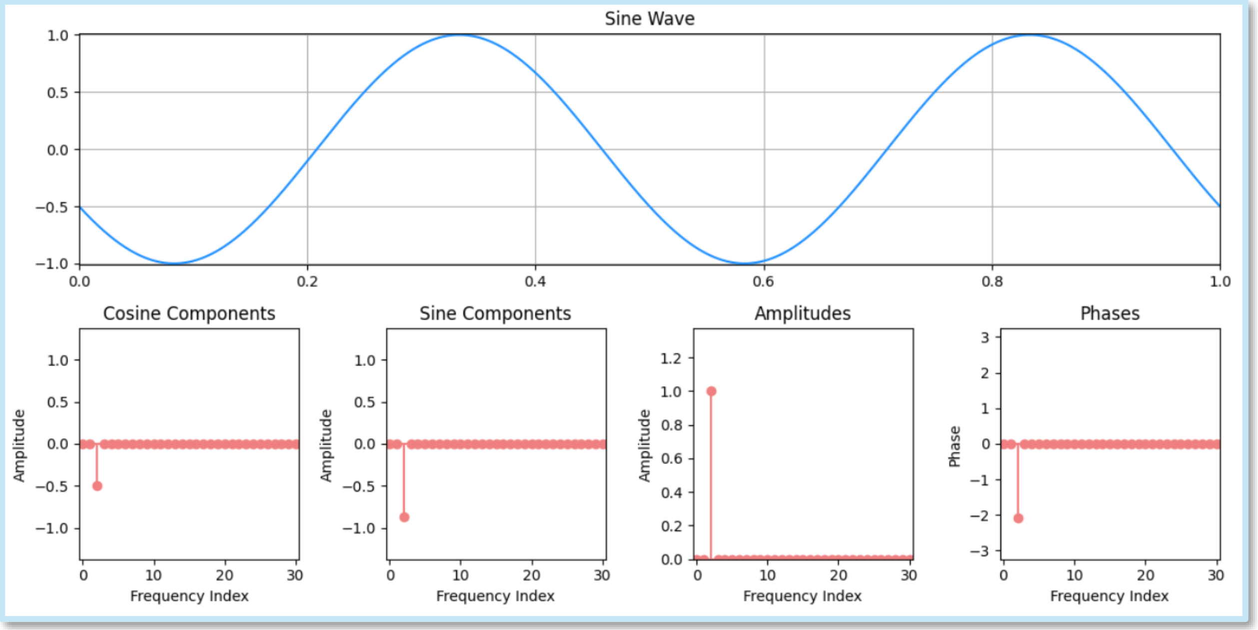


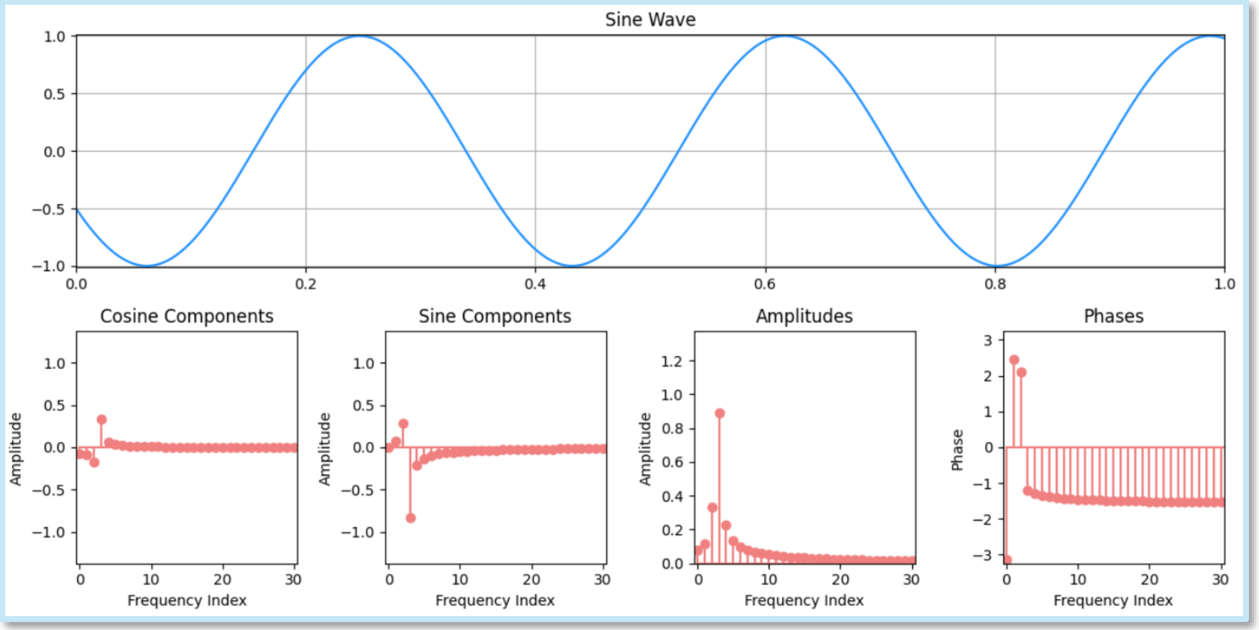
Тык

Свойства ДПФ

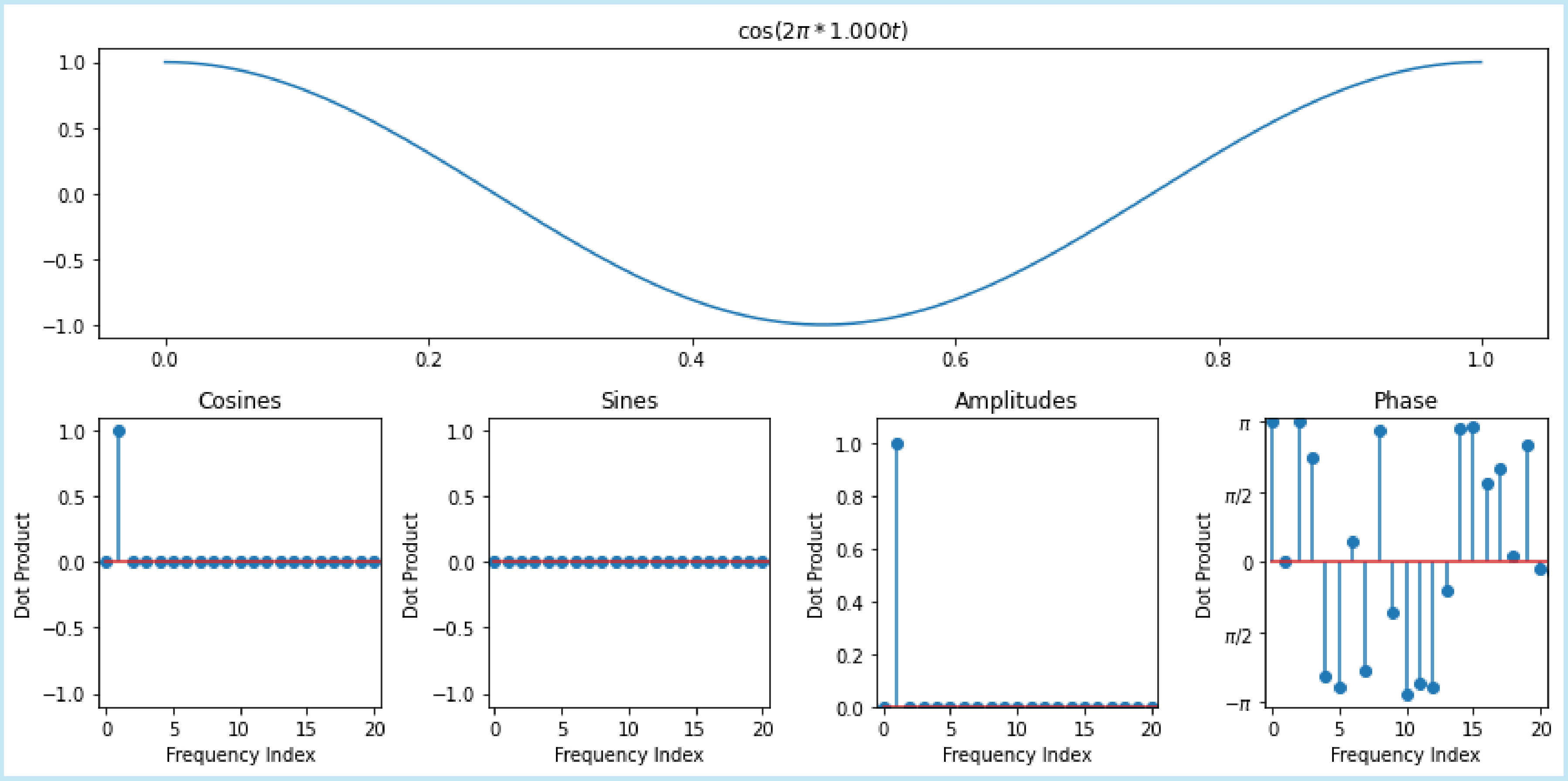
Растекание спектра:

- При ДПФ предполагается, что последовательность отсчетов сигнала является периодической
- Если значения начальных и конечных отсчётов сигнала сильно различаются, то на стыках возникают скачки, из-за которых спектр сигнала расширяется (***spectrum leakage***)





Свойства ДПФ



Оконное преобразование Фурье

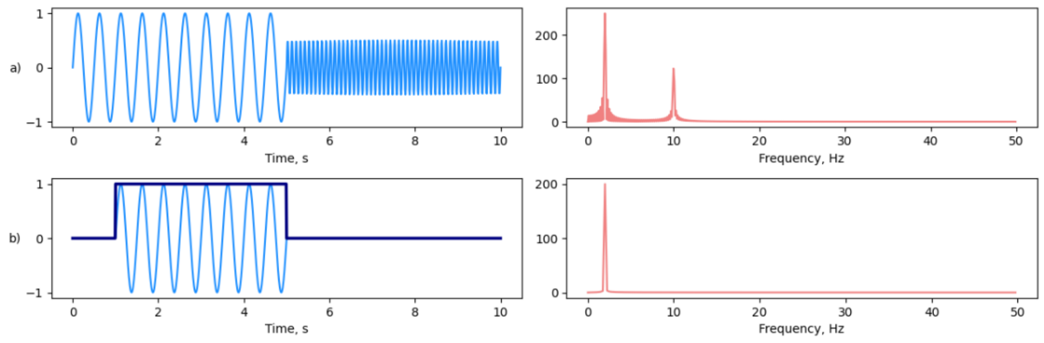
Оконное преобразование Фурье

Что если наш сигнал меняется со временем? Как представить спектр таких сигналов?

Денис Габор в 1946 году предложил метод ***оконного преобразования Фурье*** (***Short – time Fourier transform, STFT***), которое является компромиссом между временным и частотным представлением сигнала

Идея ***STFT*** заключается в рассмотрении небольшого участка сигнала:

- Фиксируется оконная функция, которая отлична от нуля только в течение короткое периода времени
- Исходный сигнал умножается на оконную функцию и получают «оконный сигнал»
- Перемещая это окно во времени, вычисляют ***FFT*** для каждого оконного сигнала

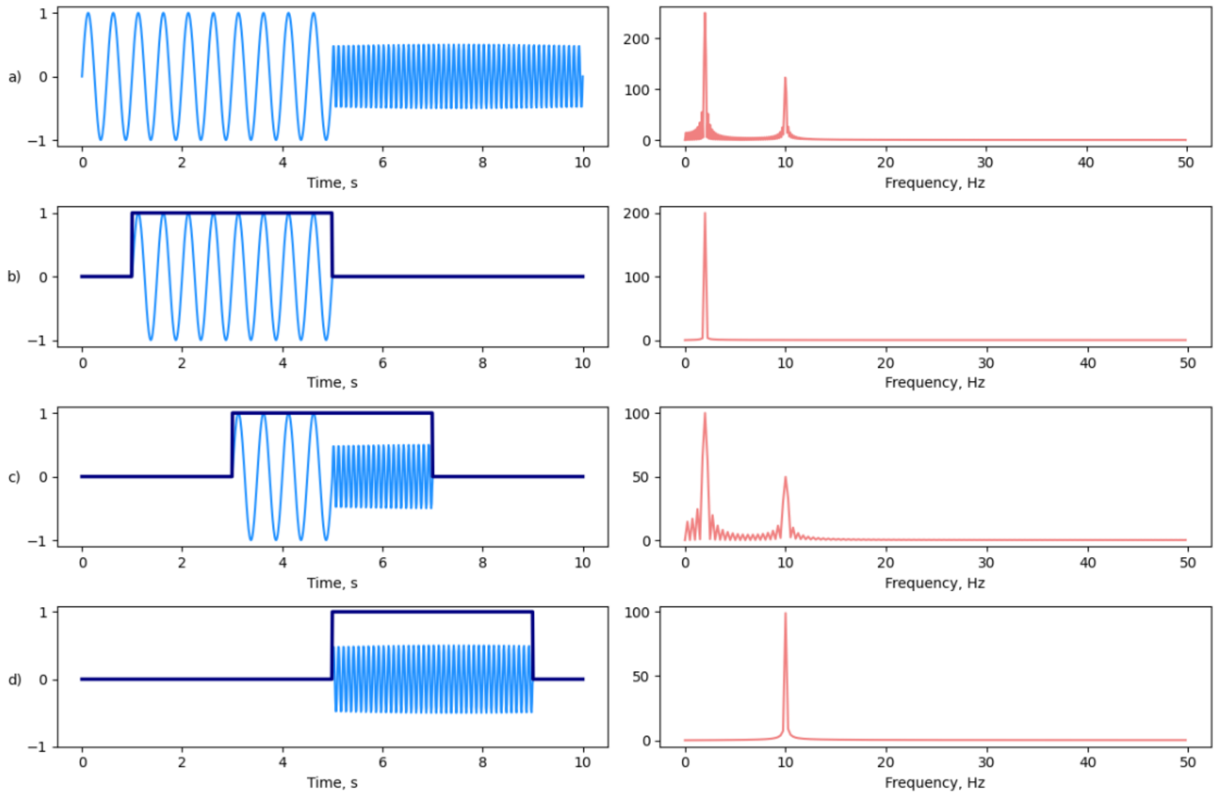


Оконное преобразование Фурье

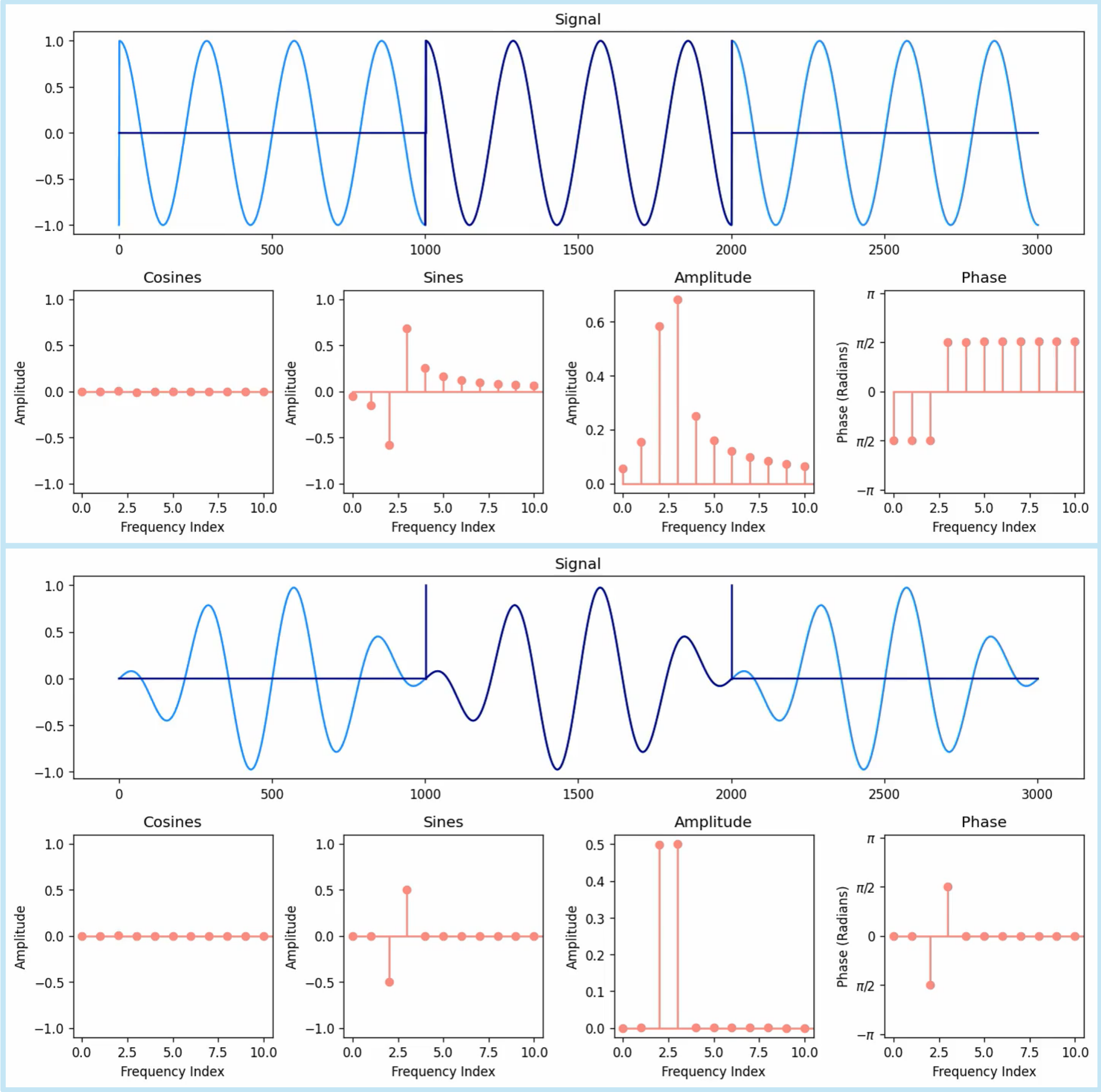
Чтобы получить отдельные сегменты сигнала, исходный сигнал перемножается с оконной функций:

$$x_\omega(t) = x(t) \cdot \omega(t),$$

- $x(t)$ – исходный сигнал, $\omega(t)$ – оконная функция



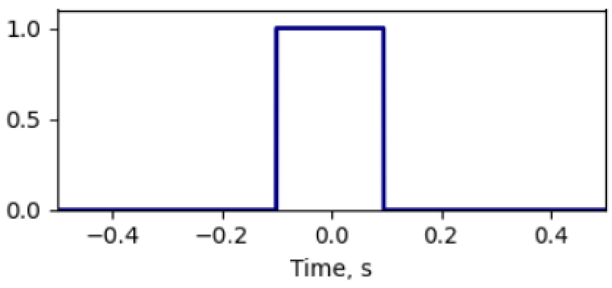
Оконное преобразование Фурье

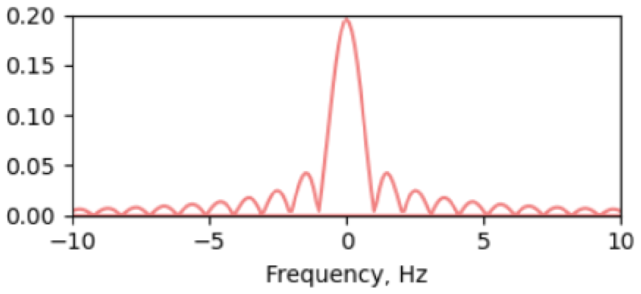


Роль оконной функции

Прямоугольное окно:

- Оставляет сигнал неизменным внутри интересующего участка и обнуляет за его пределами



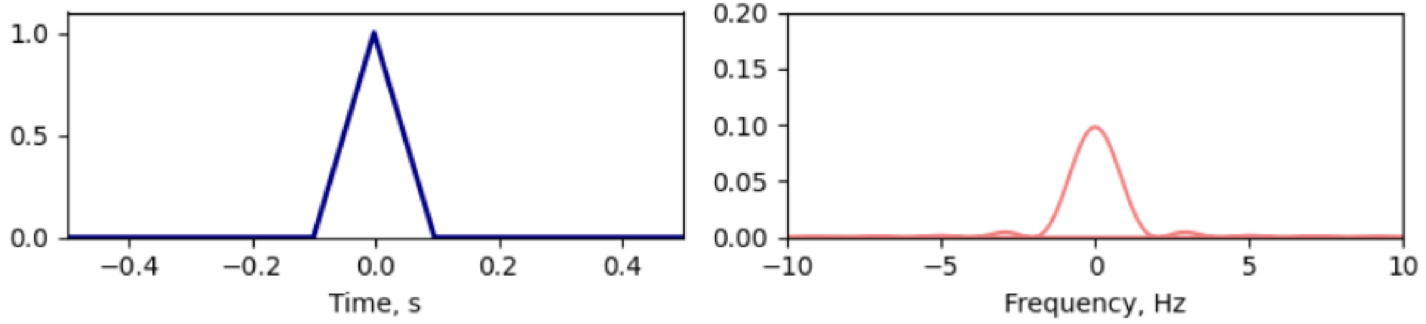


Преобразование Фурье прямоугольного окна – это функция *sinc*, которая имеет вид медленно затухающей кривой

Использование прямоугольного окна приводит к разрывам на границах участка, что вызывает артефакты, распространяющиеся по всем спектру

Роль оконной функции

Треугольное окно:



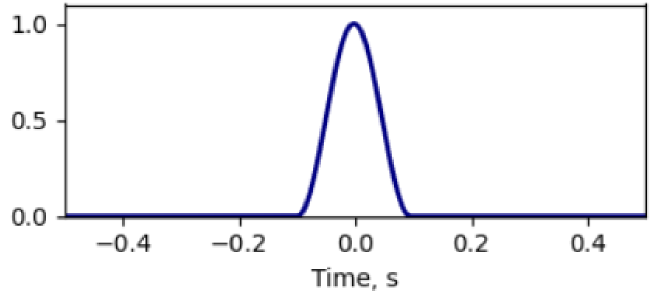
Приводит к гораздо меньшим артефактам, так как более плавно спадает по краям

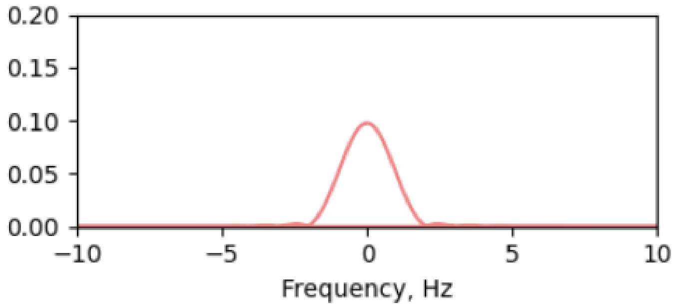
Роль оконной функции

Окно Ханна:

- Представляет собой часть косинуса:

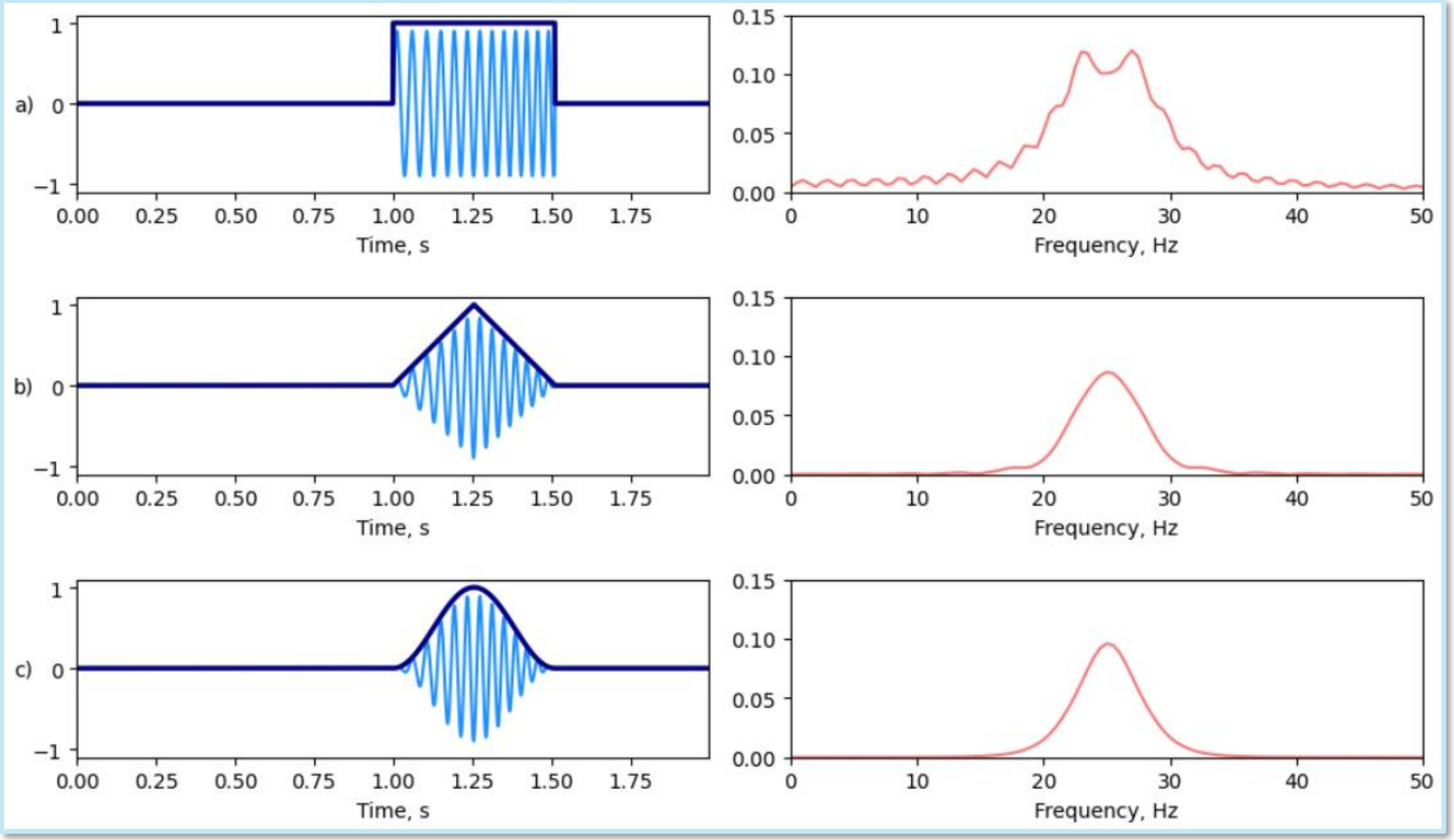
$$\omega(t) = \begin{cases} \frac{1}{2}(1 + \cos(\pi t)), & \text{если } -0.5 \le t \le 0.5 \\ 0, & \text{иначе} \end{cases}$$





Вносит некоторое размытие частот, что может привести к более гладкому спектру

Роль оконной функции



Спектрограммы и мел-спектрограммы

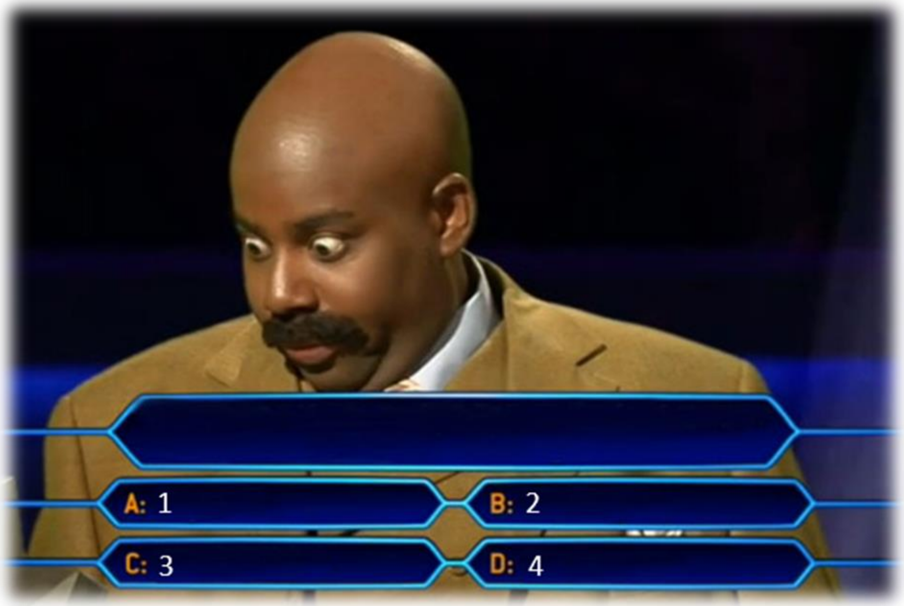
Спектрограмма

1. Спектрограмма – изображение, показывающее зависимость спектральной плотности сигнала от времени

2. Спектрограмма – двумерное изображение мощности сигнала

3. Спектрограмма – мгновенный спектр сигнала, зависящий от времени

4. Спектрограмма – изображение спектра сигнала по мере его изменения во времени



Спектрограмма

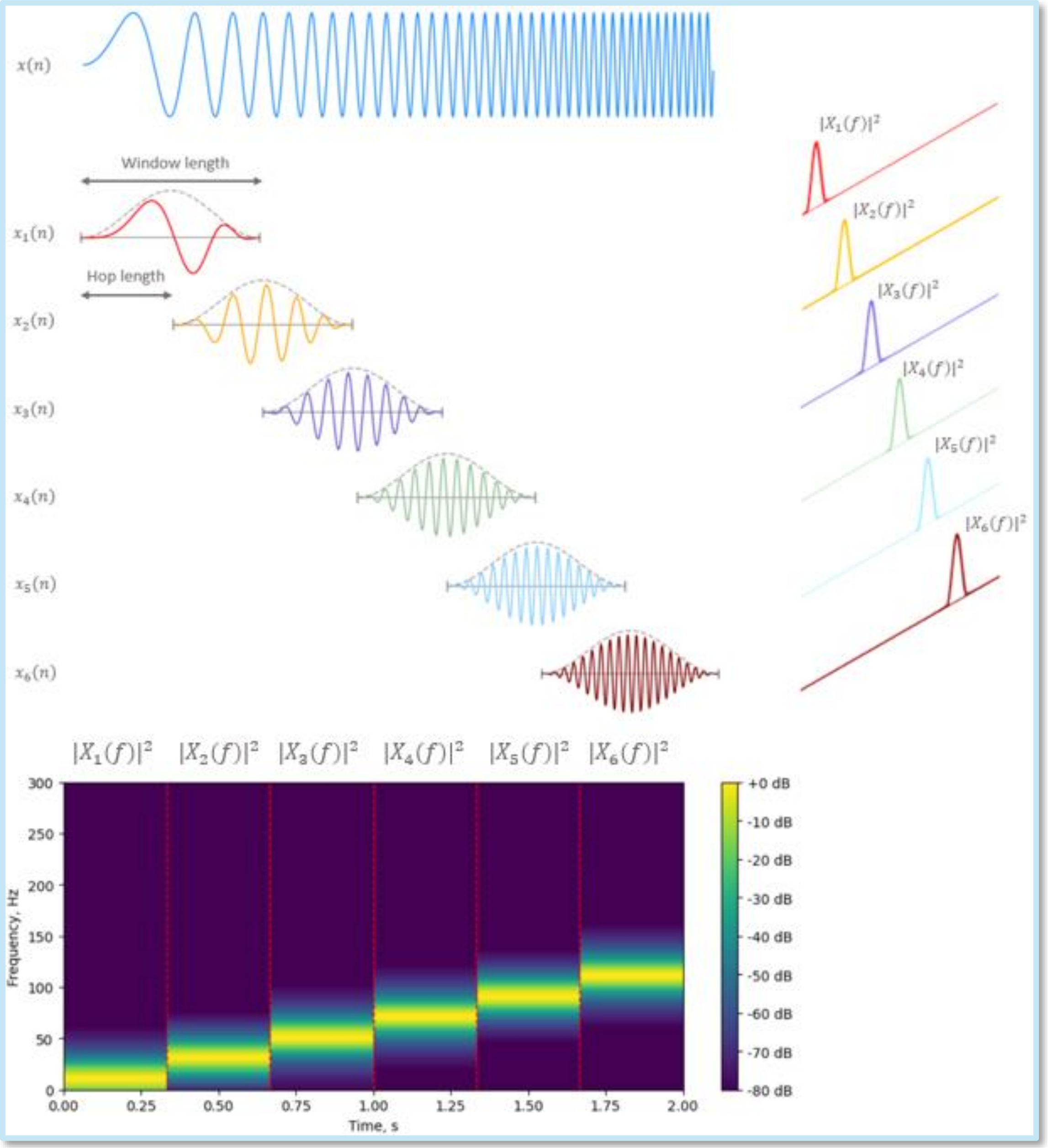
Когда мы применяем $STFT$ к сигналу, мы получаем на выходе комплексные значения для каждого момента времени t и частоты ω:

$$STFT = \begin{pmatrix} 1.2 - 2.7i & 4.5 - 0.3i & \cdots\cdots & 1.4 + 2.3i \\ 0.3 + 0.8i & 0.2 - 1.6i & & 0.5 + 1.1i \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ 0.4 + 0.3i & 0.3 + 0.8i & \cdots\cdots & 0.4 - 1.7i \end{pmatrix}$$

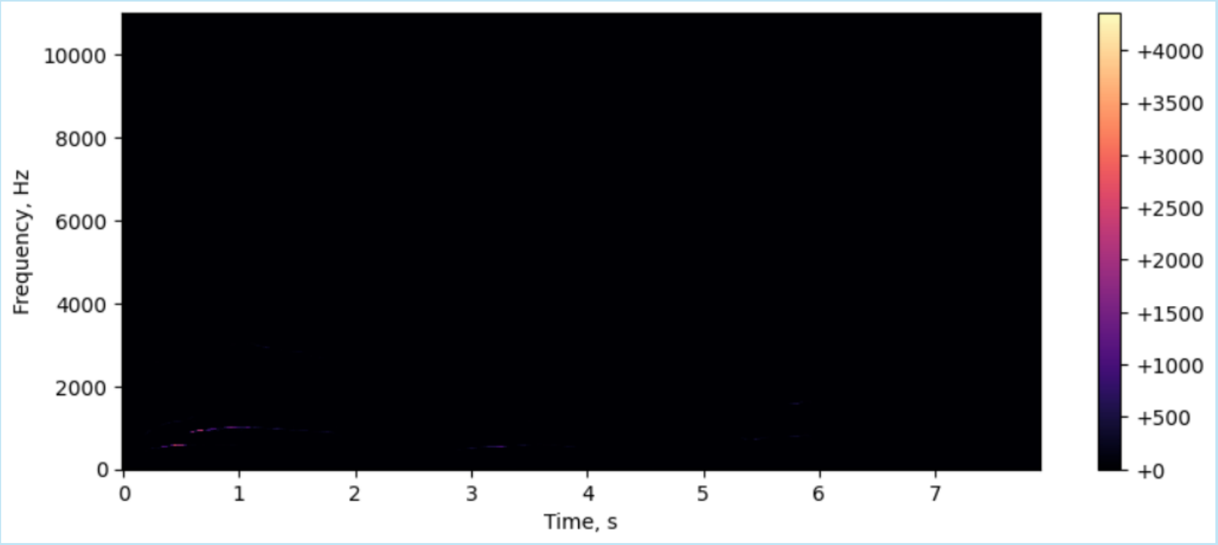
Для визуализации этих данных используется ***спектрограмма***, которая обычно определяется следующим образом:

$$spectrogram(t, \omega) = |STFT(t, \omega)|^2$$

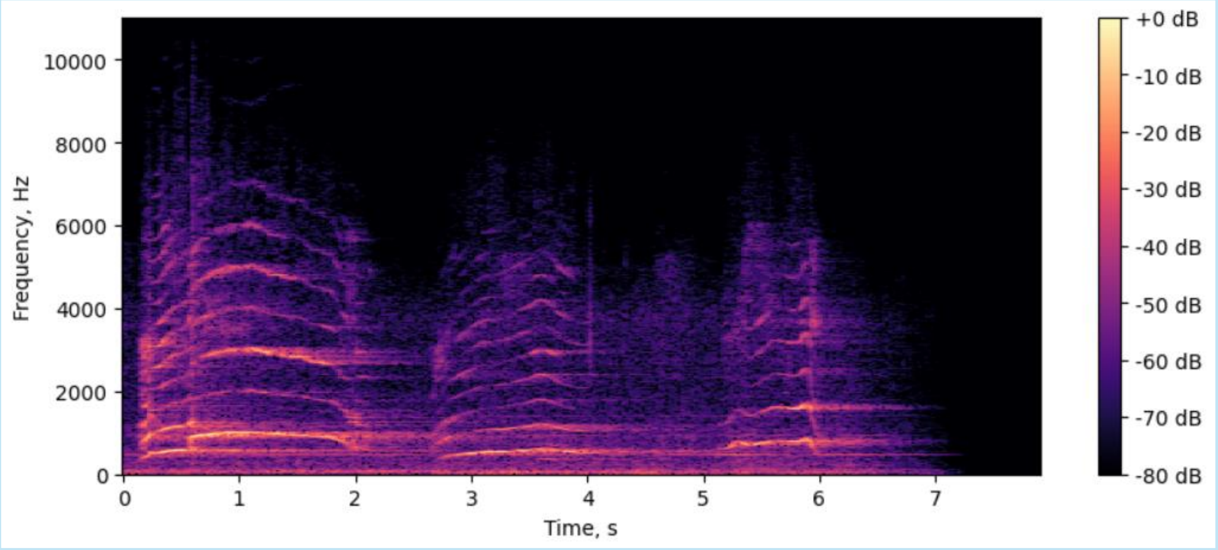
Спектрограмма



Спектрограмма



Переход к децибелам



Восприятие частот

Наш слух более чувствителен к изменениям низких частот

400 Гц 450 Гц

5050 Гц 5000 Гц

Мы легко отличаем разницу между 400 Гц и 450 Гц, но отличить звуки 5000 Гц и 5050 Гц сложнее

- Поэтому был придуман способ измерять высоту звука так, чтобы одинаковые интервалы по частоте воспринимались нами так же одинаково, вне зависимости от области частот

- Единицу измерения такой величины назвали ***мел***

Мел-спектрограмма

Мел - это психофизическая единица высоты звука

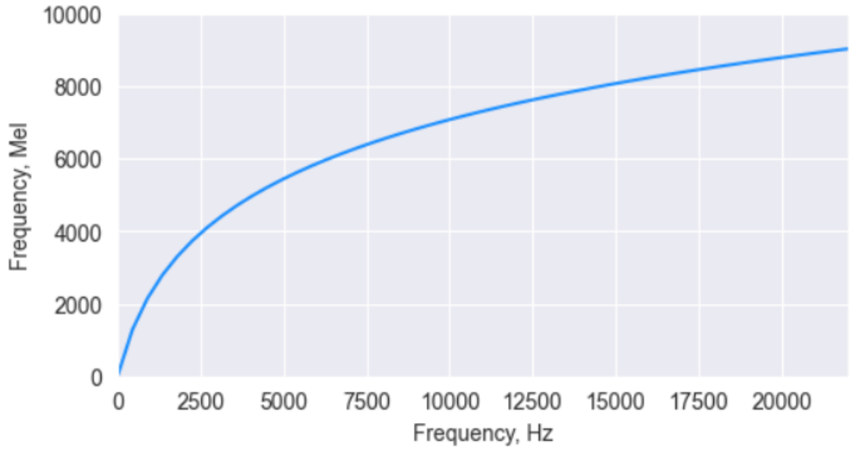
Формула перехода от герц к мел-шкале была выведена эмпирически с помощью проведения большого количества психоакустических тестов

- Для перевода из герц в мел-шкалу используют формулу:

$$m = 1127 \cdot \ln\left(1 + \frac{f}{700}\right), \qquad m = 2595 \cdot log_{10}\left(1 + \frac{f}{700}\right)$$

- Обратный переход:

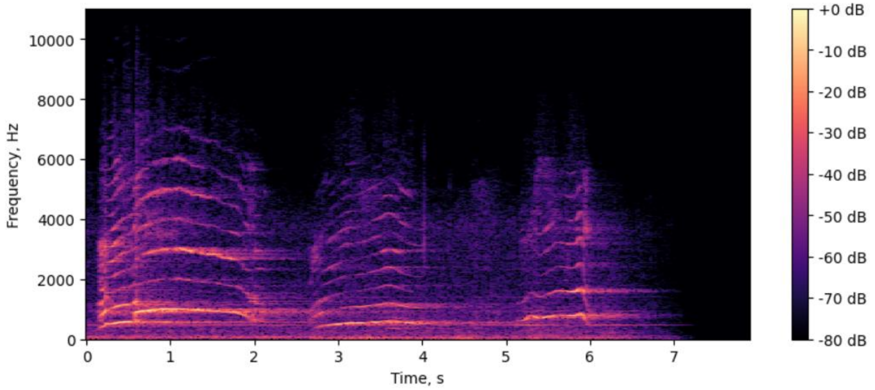
$$f = 700 \cdot \left(e^{\frac{m}{1127}} - 1\right) \qquad f = 700 \cdot \left(10^{\frac{m}{2295}} - 1\right)$$

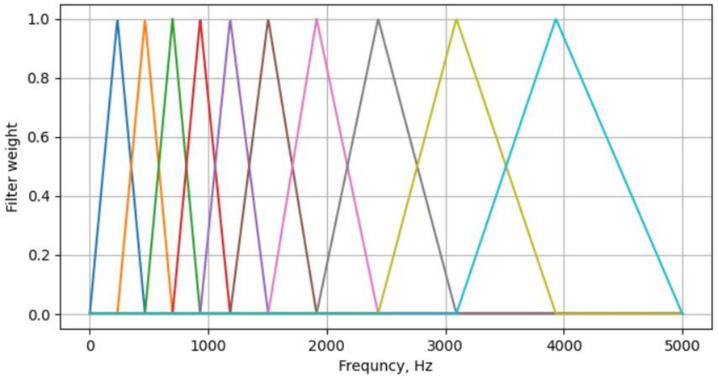


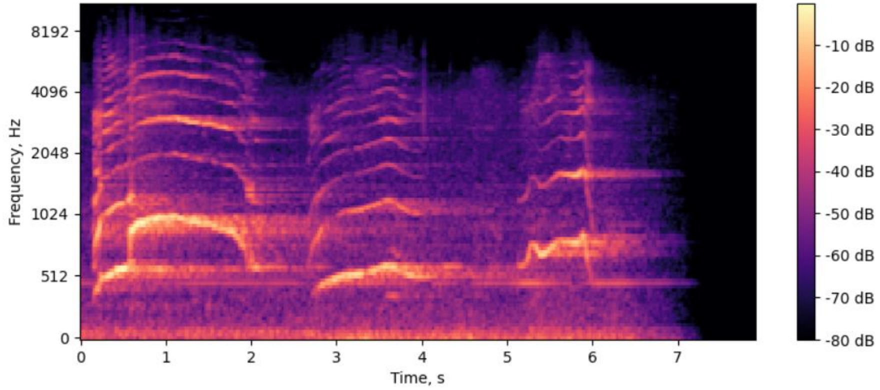
Мел-спектрограмма

Алгоритм вычисления:

- Вычислить спектрограмму
- Построить мел-фильтры
- Преобразовать частоту в мел-шкалу



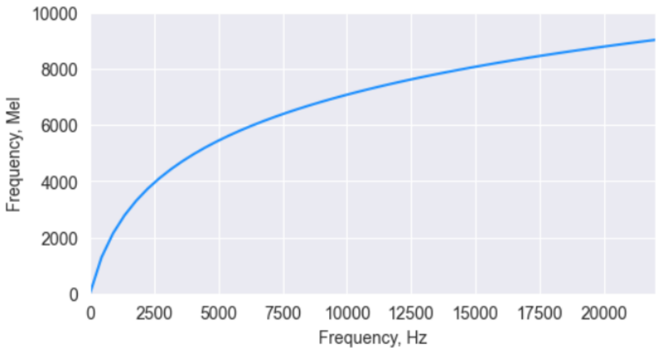


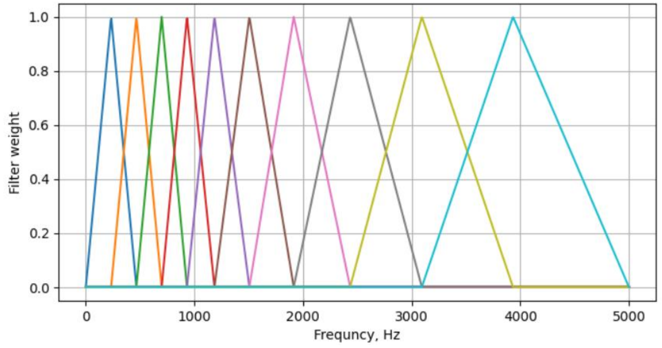


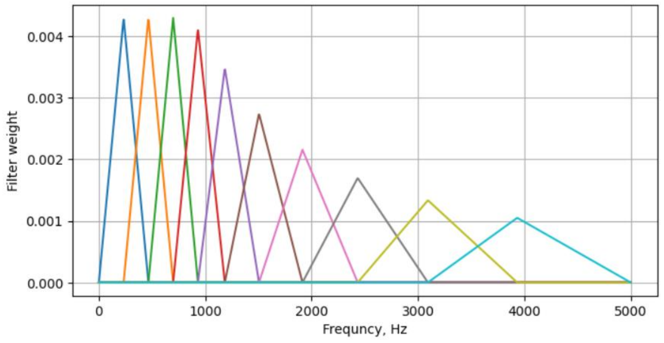
Мел-спектрограмма

Банк мел-фильтров вычисляется следующим образом:

- Переводим максимальную и минимальную частоту в мел-шкалу
- Создаём n равномерно расположенных точек
- Преобразуем соответствующие частоты каждой точки обратно в герцы и округляем до ближайшего отсчёта
- Находим координаты вершин треугольника и строим фильтр
- Нормализация







Спасибо за внимание!