

академия  
больших  
данных



Настоящий блок лекций подготовлен при  
поддержке «ЦРТ | Группа компаний» и  
Университета ИТМО



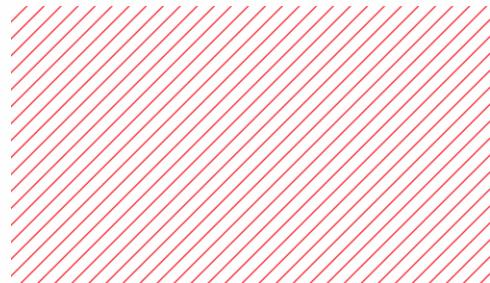
## Обработка речевых сигналов

### Блок 1. Цифровая обработка сигналов

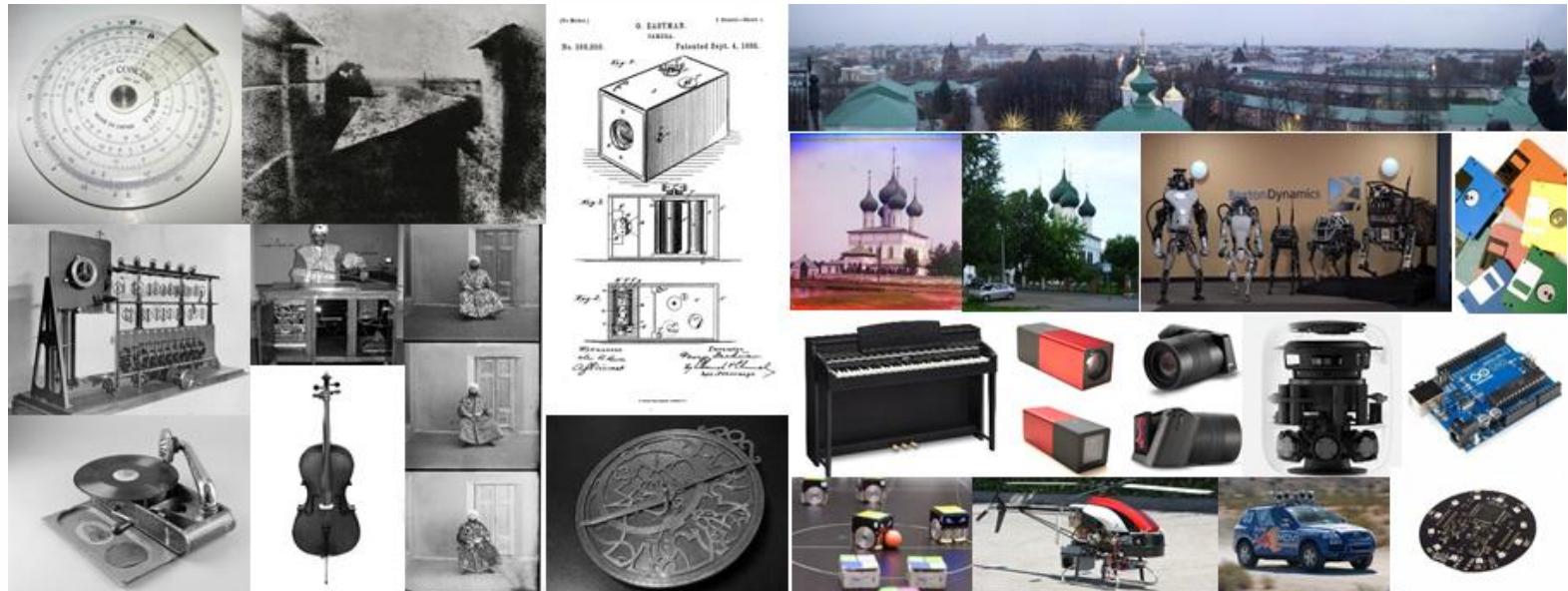
Владимир Волохов

Научный сотрудник ООО «ЦРТ-инновации»,  
доцент, к.т.н.

Цифровой сигнал      Нейронные сети      Преобразование Фурье  
Спектр и спектrogramма      Цифровой фильтр  
Теорема отсчетов      Частота дискретизации  
Акустические признаки



# Блок 1. Цифровая обработка сигналов



## Часть 2. Представление сигналов в спектральной области

262

263



# План лекции

---

- Что такое спектр?
- Почему синусоиды лежат в основе преобразования Фурье?
- Естественные детекторы синусоид у человека
- Формы преобразования Фурье для дискретного времени
- Преобразование Фурье для цифровых изображений
- Машина для восстановления сигнала
- Спектр сигнала после дискретизации
- Идея алгоритма быстрого преобразования Фурье
- Оконное преобразование Фурье
- Банк фильтров
- Вычисление мел-частотных кепстральных коэффициентов



# Идея понимания спектров является древней как сама наука!

---

- Древнегреческие математики и физики, а также арабские ученые пытались понять, почему радуга состоит из множества прекрасных цветов
- 14 век, доминиканский монах Теодор из Фрайберга провел эксперимент с бутылкой воды на солнце и обнаружил первичную и вторичную радугу
- 17 век, Р. Декарт и И. Ньютон дали полное изложение того, как белый свет может быть разложен на множество цветов радуги (явление дисперсии света)
- 19 век, Ж. Фурье описал математически то, что Р. Декарт и И. Ньютон описали будучи физиками. Ж. Фурье предложил подход, с использованием которого можно разложить любую функцию с регулярным поведением в сумму гармонических компонент
- Спектр – распределение энергии или другой физической величины по частотам



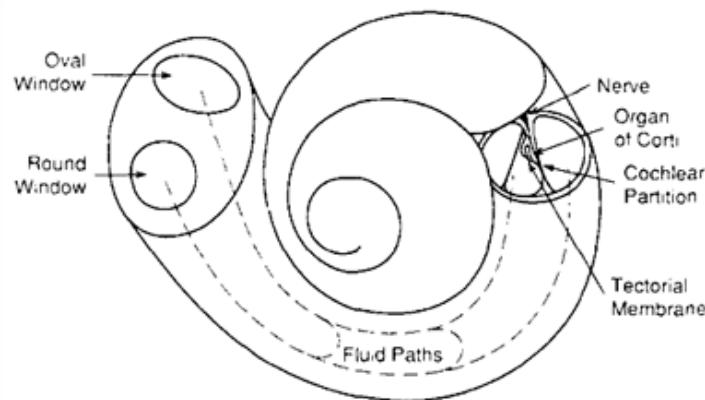
# Почему синусоиды лежат в основе преобразования Фурье!

---

- Осцилляции являются фундаментальным импульсом природы. Они есть везде!
- Устойчивые динамические системы демонстрируют осцилляционное поведение
  - Вращение колеса
  - Волны на берегу
  - Музыкальные инструменты
- Интуитивно, вещи, которые не движутся по окружности, не могут продолжаться
  - Бомбы
  - Ракеты
  - Жизнь человека
- Осцилляции легко описать и параметризовать (движение точки по окружности)!

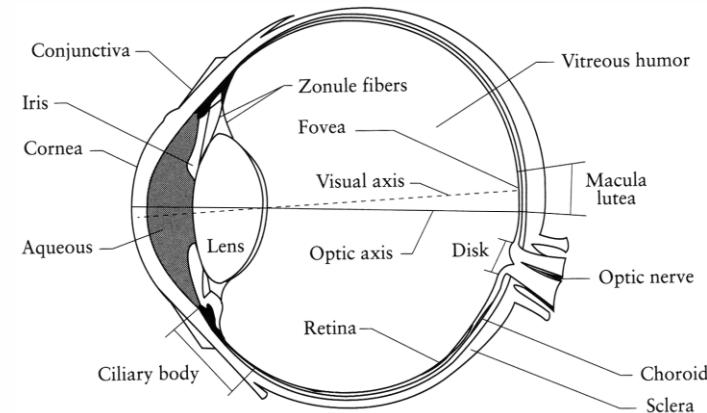
# Вы тоже можете детектировать синусоиды!

Улитка (внутреннее ухо)



Улитка детектирует колебания, связанные с воздушным давлением, на частотах от 20 Гц до 20 кГц

Палочки и колбочки (сетчатка глаза)



Палочки и колбочки детектируют электромагнитные колебания на частотах от 430 ТГц до 790 ТГц

# Можем ли мы разложить любой сигнал на синусоидальные компоненты?

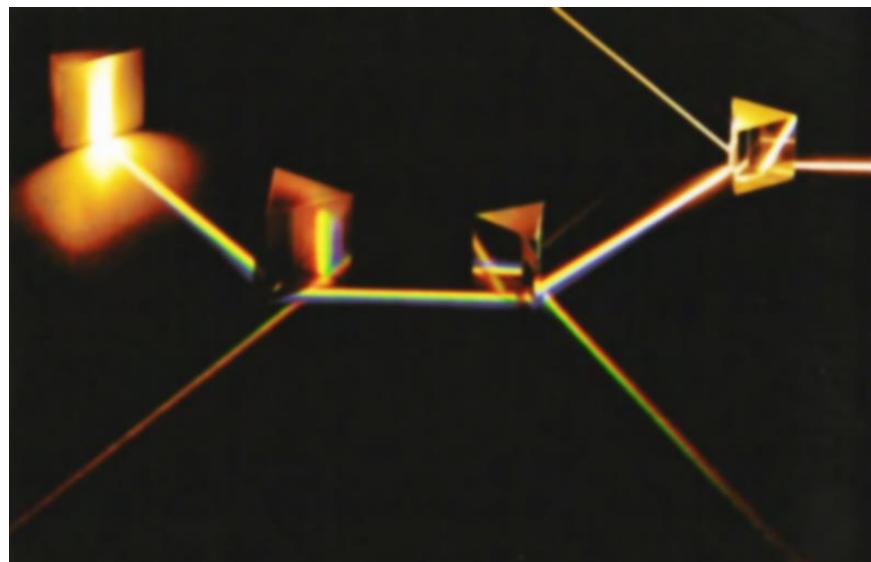
- Да, можем, и Ж. Фурье показал это!

## Анализ

1. Позволяет выполнить переход от временной области к частотной области
2. Позволяет находить вклад различных частот
3. Позволяет выявить скрытые свойства сигнала

## Синтез

1. Позволяет выполнить переход от частотной области к временной области
2. Позволяет создать сигнал с известным частотным контентом
3. Позволяет подогнать сигнал к определенной частотной области, используя обработку сигналов





# Преобразование Фурье последовательности конечной длины

- Дискретное преобразование Фурье

## Утверждение

Множество из  $N$  сигналов в пространстве  $C^N$ ,  $w_k(n) = e^{j\frac{2\pi}{N}kn}$ ,  $n, k = 0, 1, \dots, N - 1$ , является ортогональным базисом в  $C^N$ . В обозначениях  $k$  – это номер вектора, а  $n$  – номер элемента вектора

### Формула анализа

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}, k = 0, 1, \dots, N - 1$$

$X(k)$  – это  $N$ -точечный сигнал в частотной области

### Формула синтеза

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}kn}, n = 0, 1, \dots, N - 1$$

$x(n)$  – это  $N$ -точечный сигнал во временной области

# Преобразование Фурье последовательности конечной длины

- Дискретное преобразование Фурье

## Утверждение

Множество из  $N$  сигналов в пространстве  $C^N$ ,  $w_k(n) = e^{j\frac{2\pi}{N}kn}$ ,  $n, k = 0, 1, \dots, N - 1$ , является ортогональным базисом в  $C^N$ . В обозначениях  $k$  – это номер вектора, а  $n$  – номер элемента вектора

### Формула анализа

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \dots, N - 1$$

$X(k)$  – это  $N$ -точечный сигнал в частотной области

Переход от временной области к частотной области не изменяет энергию сигнала!

### Формула синтеза

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn}, \quad n = 0, 1, \dots, N - 1$$

$x(n)$  – это  $N$ -точечный сигнал во временной области

### Равенство Парсеваля

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2$$

# Базисы бывают разные!

- Канонический базис, состоящий из  $N$  единичных импульсов  $w_k(n) = \delta(n - k)$ ,  $n, k = 0, 1, \dots, N - 1$ , можно использовать для представления сигнала длины  $N$

**Представление сигнала с помощью  
канонического базиса**

$$x(n) = \sum_{k=0}^{N-1} x(k)\delta(n - k), n = 0, 1, \dots, N - 1$$

$$\delta(n - k) = \begin{cases} 1, & n = k \\ 0, & n \neq k \end{cases}$$

**Формула синтеза для дискретного  
преобразования Фурье**

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}kn}, n = 0, 1, \dots, N - 1$$

- Базис вейвлетов, базис главных компонент и т.п.

# Дискретное преобразование Фурье некоторых последовательностей

- $x(n) = \delta(n), x(n) \in C^N$

$$X(k) = \sum_{n=0}^{N-1} \delta(n) e^{-j\frac{2\pi}{N}kn} = 1$$

- $x(n) = 1, x(n) \in C^N$

$$X(k) = \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}kn} = N\delta(k)$$

- $x(n) = 3 \cdot \cos\left(\frac{2\pi}{16}n\right), x(n) \in C^{64}$

$$X(k) = \begin{cases} 96, & k = 4, 60 \\ 0, & k \neq 4, 60 \end{cases}$$

$$x(n) = 3 \cdot \cos\left(\frac{2\pi}{64} \cdot 4n\right) = \frac{3}{2} \cdot (e^{j\frac{2\pi}{64} \cdot 4n} + e^{-j\frac{2\pi}{64} \cdot 4n}) = \frac{3}{2} \cdot (e^{j\frac{2\pi}{64} \cdot 4n} + e^{j\frac{2\pi}{64} \cdot 60n})$$

$$= \frac{1}{64} \cdot 64 \cdot \frac{3}{2} \cdot (e^{j\frac{2\pi}{64} \cdot 4n} + e^{j\frac{2\pi}{64} \cdot 60n}) = \frac{1}{64} \cdot X(4) \cdot e^{j\frac{2\pi}{64} \cdot 4n} + \frac{1}{64} \cdot X(60) \cdot e^{j\frac{2\pi}{64} \cdot 60n}$$

**Формула анализа**

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

**Формула синтеза**

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

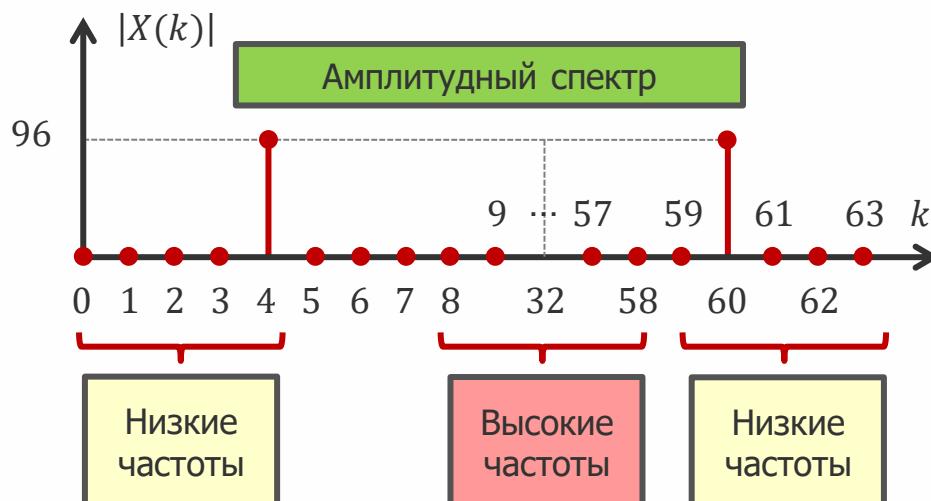
**Сумма  $N$  членов  
геом. прогрессии**

$$S_N = \frac{b_0(1-q^N)}{1-q}, q = \frac{b_n}{b_{n-1}}$$

# Разметка частотной оси

- $x(n) = 3 \cdot \cos\left(\frac{2\pi}{16}n\right)$ ,  $x(n) \in C^{64}$

$$X(k) = \begin{cases} 96, k = 4, 60 \\ 0, k \neq 4, 60 \end{cases}$$



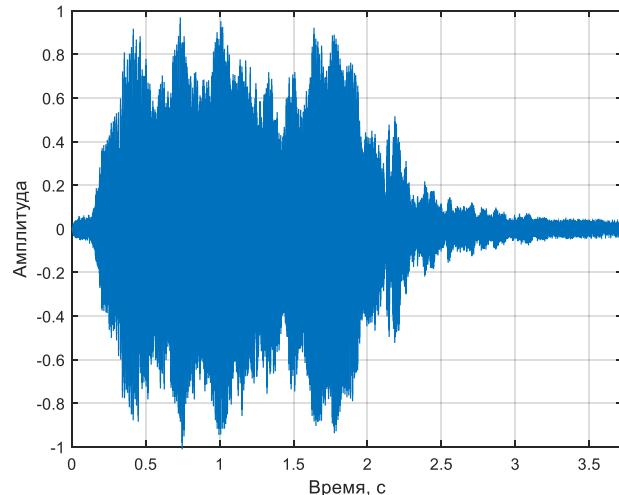
$$f_k = \frac{f_d}{N} \cdot k \text{ [Гц]}, f_d - \text{это частота дискретизации}$$

# Пример частотного анализа сигнала

- Сигнал паровозного гудка



$$f_{\text{д}} = 8000 \text{ [Гц]}$$
$$N = 29814$$

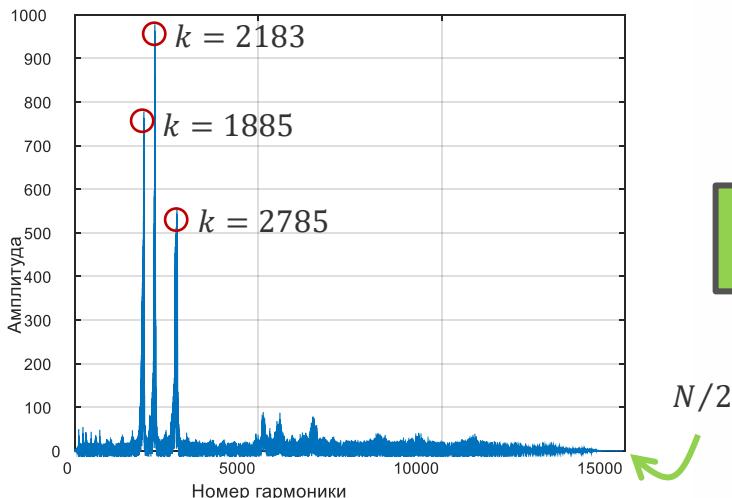


С.М. Прокудин-Горский «Паровоз Аб-132 с пароперегревателем Шмидта», 1910 год

Сигнал паровозного гудка  
(огибающая сигнала)

# Пример частотного анализа сигнала

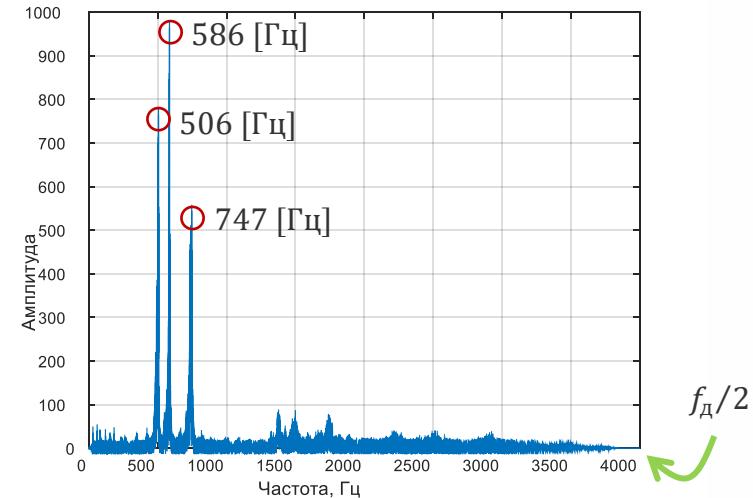
- Сигнал паровозного гудка



$$f_{\Delta} = 8000 \text{ [Гц]}$$
$$N = 29814$$

$N/2$

Огибающая амплитудного спектра  
(по оси абсцисс отложен номер гармоники)

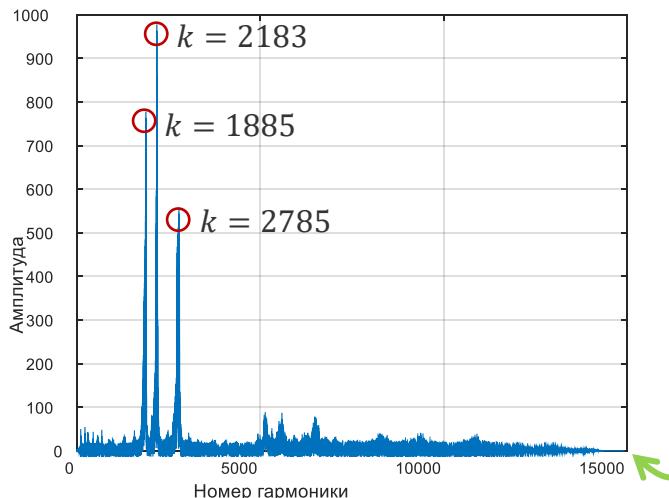


$f_{\Delta}/2$

Огибающая амплитудного спектра  
(по оси абсцисс отложена частота)

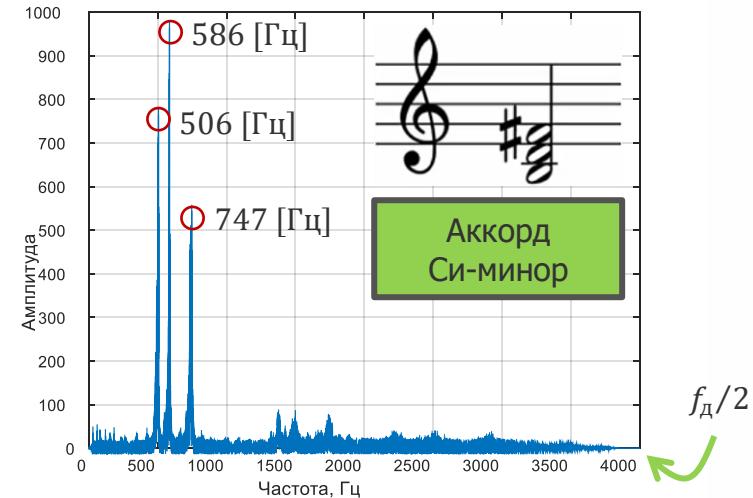
# Пример частотного анализа сигнала

- Сигнал паровозного гудка



Огибающая амплитудного спектра  
(по оси абсцисс отложен номер гармоники)

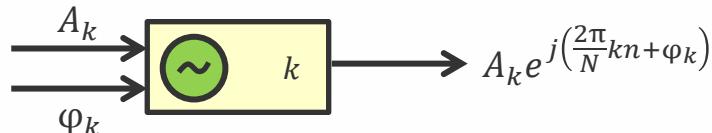
$$f_d = 8000 \text{ [Гц]} \\ N = 29814$$



Огибающая амплитудного спектра  
(по оси абсцисс отложена частота)

# Синтез сигнала и генератор синусоиды

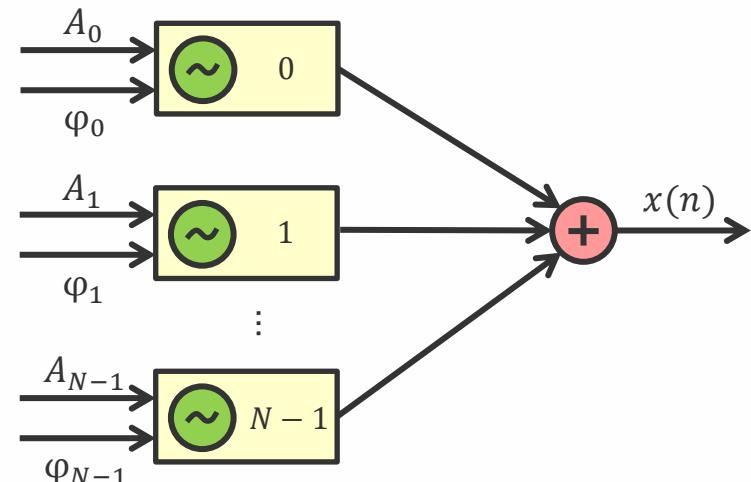
- Для генерации синусоиды необходимо задать ее амплитуду, частоту и начальную фазу ( $A_k$ ,  $\hat{\omega}_k$ ,  $\varphi_k$ )



$$A_k = \frac{|X(k)|}{N}, \quad \varphi_k = \tan^{-1} \left( \frac{\text{Im}(X(k))}{\text{Re}(X(k))} \right)$$

$\varphi_k$  – это компонента фазового спектра (начальная фаза гармоники)

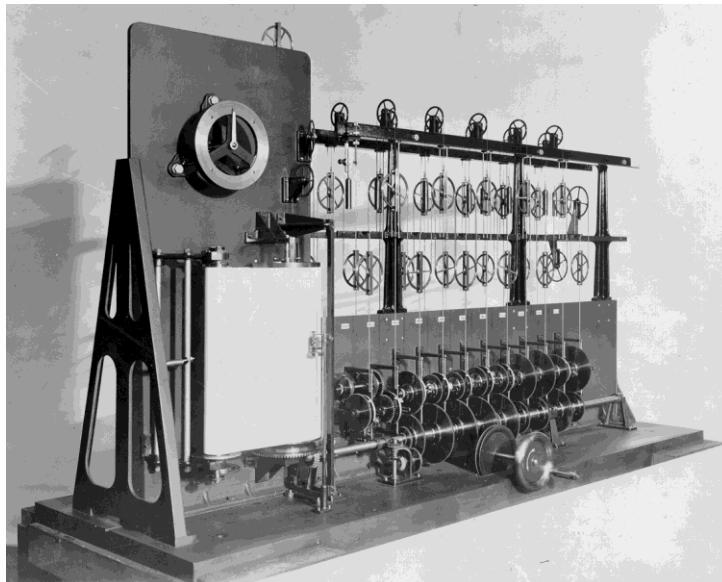
Генератор одной синусоиды



Машина для восстановления сигнала

# Машина для восстановления сигнала

- 19 век, У. Томсон (барон Кельвин) изобрел машину для предсказания приливов





# Преобразование Фурье периодической последовательности

- Дискретное преобразование Фурье обладает свойством скрытой периодичности, например,  $X(N) = X(0)$ ,  $X(N + 1) = X(1)$  и т.д.
- При явно заданной периодичности Фурье преобразования термин дискретное преобразование Фурье заменяется на **дискретный ряд Фурье**
  - **Дискретный ряд Фурье = дискретное преобразование Фурье с явно заданной периодичностью**
  - Дискретный ряд Фурье отображает  $N$ -периодический сигнал в  $N$ -периодическую последовательность коэффициентов
  - Обратный дискретный ряд Фурье отображает  $N$ -периодическую последовательность коэффициентов в  $N$ -периодический сигнал
  - Дискретный ряд Фурье  $N$ -периодического сигнала является математическим эквивалентом к дискретному преобразованию Фурье одного периода

# Преобразование Фурье для сигналов большей размерности

- Двумерное дискретное преобразование Фурье можно рассмотреть по отношению к цифровым изображениям

## Формула анализа

$$X(k, l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x(n, m) e^{-j2\pi\left(\frac{kn}{N} + \frac{lm}{M}\right)}$$

## Формула синтеза

$$x(n, m) = \frac{1}{N \cdot M} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} X(k, l) e^{j2\pi\left(\frac{kn}{N} + \frac{lm}{M}\right)}$$

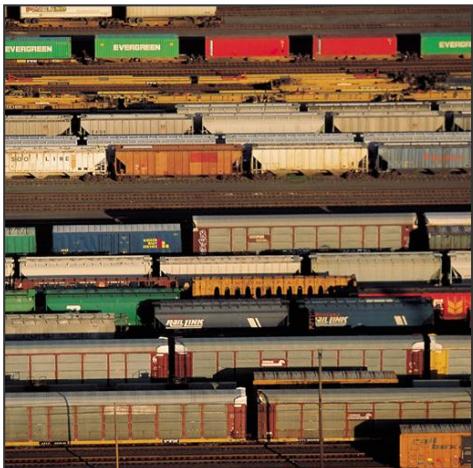
Вычисление двумерного дискретного преобразования Фурье предполагает, что изображение определено на торе



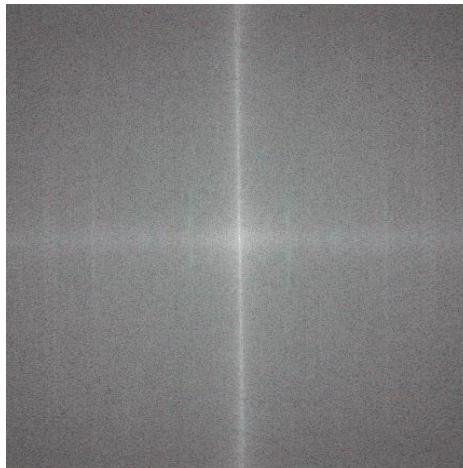
Photo: Bart Nagel [www.bartnagel.com](http://www.bartnagel.com)

# Преобразование Фурье для сигналов большей размерности

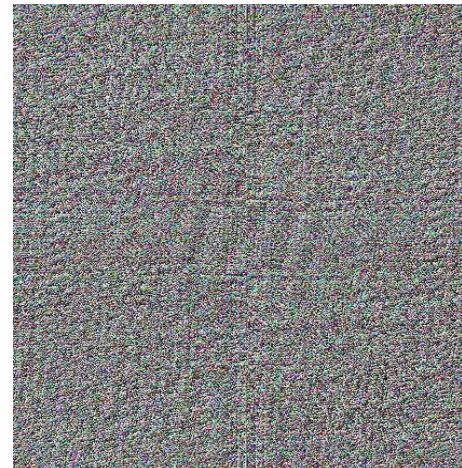
- Примеры амплитудных и фазовых спектров



Исходный сигнал



Амплитудный спектр



Фазовый спектр

# Преобразование Фурье для сигналов большей размерности

- Примеры амплитудных и фазовых спектров



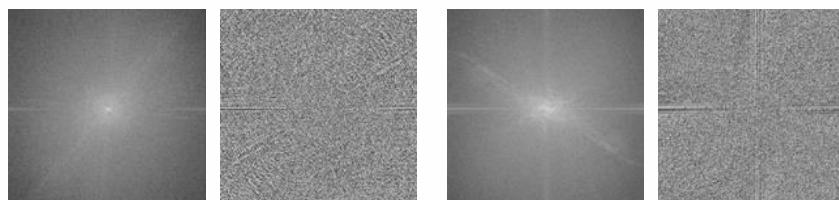
# Фазовый спектр содержит больше информации о сигнале, чем амплитудный!

- Поменяем местами фазовые спектры двух изображений и восстановим изображения

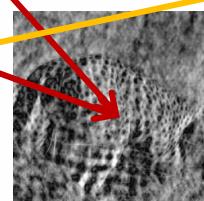
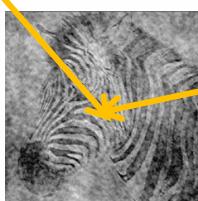
Исходные изображения



Амплитудный и фазовый спектры



Восстановленные изображения





# Преобразование Фурье непериодической последовательности

- Преобразование Фурье дискретного времени

- Если  $N \rightarrow \infty$ , то частоты  $\frac{2\pi}{N} k$  уплотняются на интервале  $[0, 2\pi]$  и  $\frac{2\pi}{N} k \rightarrow \hat{\omega}$ , где  $\hat{\omega} \in R$

### Формула анализа

$$X(e^{j\hat{\omega}}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\hat{\omega}n}, \hat{\omega} \in R$$

Для сходимости ряда  $X(e^{j\hat{\omega}})$  последовательность  $x(n)$  должна быть абсолютно или квадратично суммируемой

### Формула синтеза

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\hat{\omega}}) e^{j\hat{\omega}n} d\hat{\omega}, n \in Z$$

# Преобразование Фурье непериодической последовательности

## ■ Преобразование Фурье дискретного времени

- Если  $N \rightarrow \infty$ , то частоты  $\frac{2\pi}{N}k$  уплотняются на интервале  $[0, 2\pi]$  и  $\frac{2\pi}{N}k \rightarrow \hat{\omega}$ , где  $\hat{\omega} \in R$

### Формула анализа

$$X(e^{j\hat{\omega}}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\hat{\omega}n}, \hat{\omega} \in R$$

Для сходимости ряда  $X(e^{j\hat{\omega}})$  последовательность  $x(n)$  должна быть абсолютно или квадратично суммируемой

### Формула синтеза

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\hat{\omega}})e^{j\hat{\omega}n} d\hat{\omega}, n \in Z$$

### Пример

$$x(n) = \alpha^n u(n), |\alpha| < 1$$

$$\begin{aligned} X(e^{j\hat{\omega}}) &= \sum_{n=-\infty}^{\infty} \alpha^n e^{-j\hat{\omega}n} \\ &= \sum_{n=0}^{\infty} \alpha^n e^{-j\hat{\omega}n} \\ &= \frac{1}{1 - \alpha \cdot e^{-j\hat{\omega}}} \end{aligned}$$

# Преобразование Фурье непериодической последовательности

## ■ Преобразование Фурье дискретного времени

- Если  $N \rightarrow \infty$ , то частоты  $\frac{2\pi}{N} k$  уплотняются на интервале  $[0, 2\pi]$  и  $\frac{2\pi}{N} k \rightarrow \hat{\omega}$ , где  $\hat{\omega} \in R$

### Формула анализа

$$X(e^{j\hat{\omega}}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\hat{\omega}n}, \hat{\omega} \in R$$

Для сходимости ряда  $X(e^{j\hat{\omega}})$  последовательность  $x(n)$  должна быть абсолютно или квадратично суммируемой

### Формула синтеза

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\hat{\omega}}) e^{j\hat{\omega}n} d\hat{\omega}, n \in Z$$

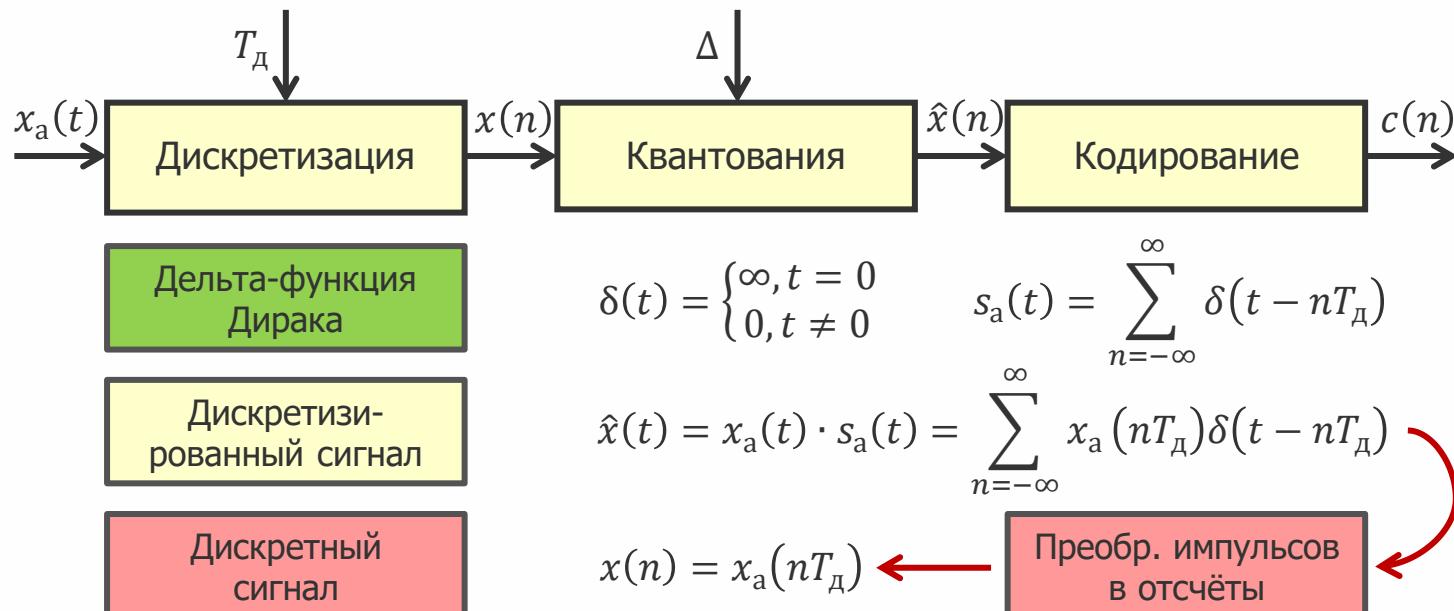
### Сумма членов беск. убыв. геом. прогрессии

$$S_N = \frac{b_0}{1-q}, q = \frac{b_n}{b_{n-1}}, |q| < 1$$

$$\begin{aligned} X(e^{j\hat{\omega}}) &= \sum_{n=-\infty}^{\infty} x(n)e^{-j\hat{\omega}n} \\ &= \sum_{n=0}^{\infty} \alpha^n e^{-j\hat{\omega}n} \\ &= \frac{1}{1 - \alpha \cdot e^{-j\hat{\omega}}} \end{aligned}$$

# Изменение спектра сигнала после дискретизации

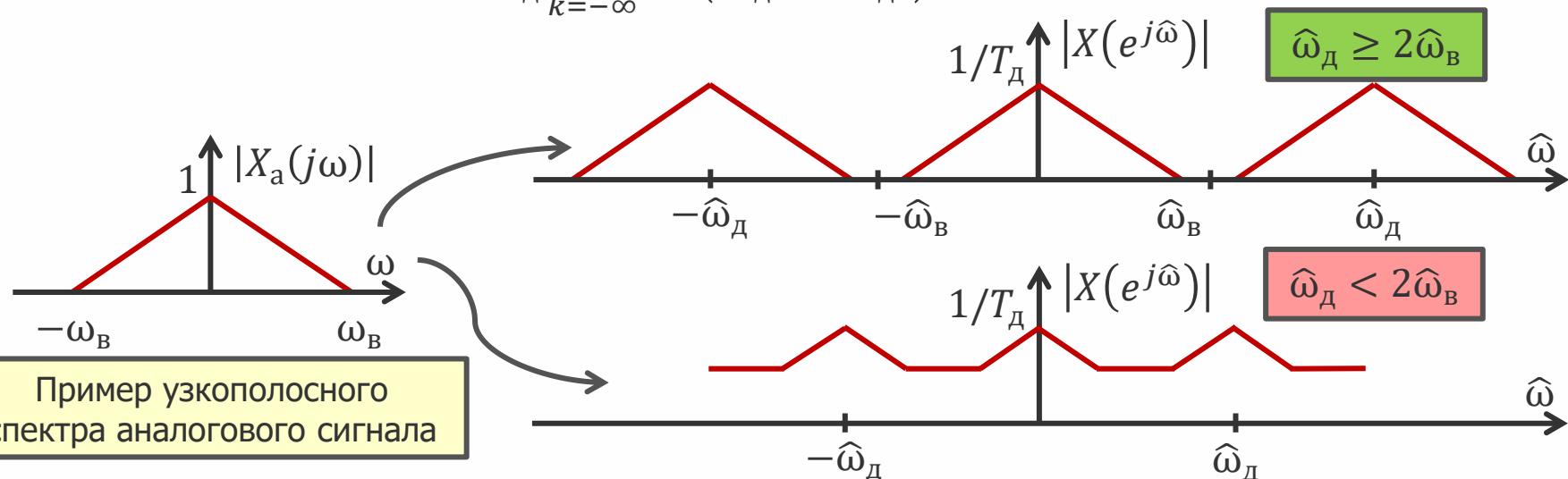
- Компоненты аналого-цифрового преобразователя (АЦП)



# Изменение спектра сигнала после дискретизации

- Результат дискретизации аналогового сигнала для спектральной области

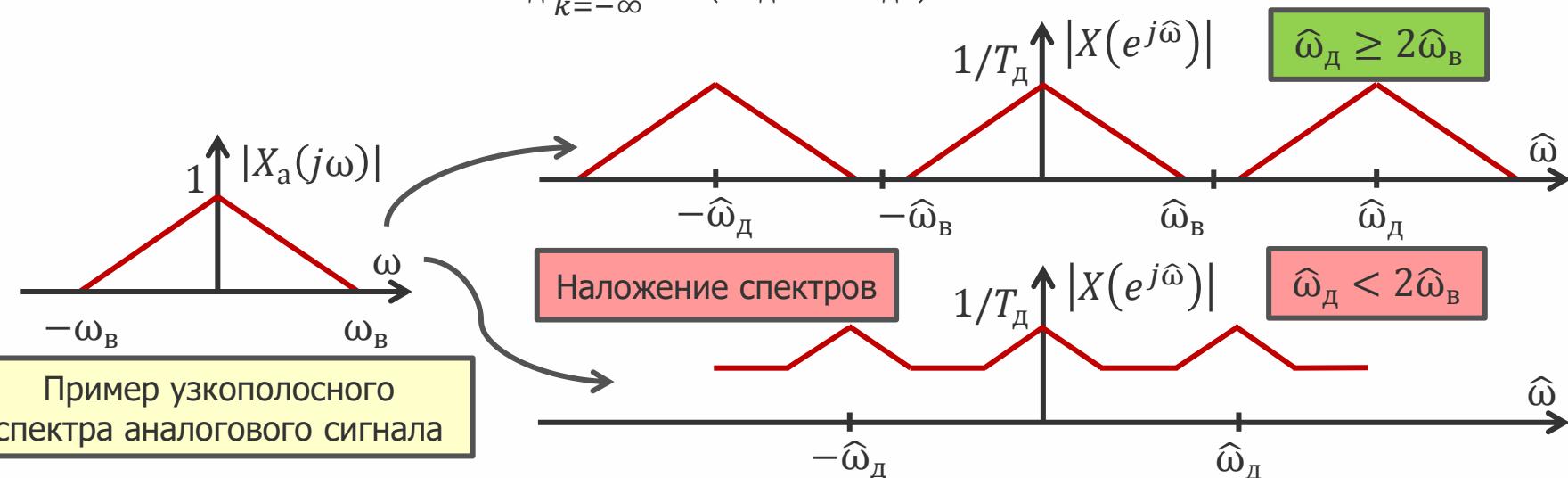
$$X(e^{j\hat{\omega}}) = \frac{1}{T_D} \sum_{k=-\infty}^{\infty} X_a \left( j \frac{\hat{\omega}}{T_D} - j \frac{2\pi k}{T_D} \right), \quad \hat{\omega} = \omega T_D \text{ [рад]}$$



# Изменение спектра сигнала после дискретизации

- Результат дискретизации аналогового сигнала для спектральной области

$$X(e^{j\hat{\omega}}) = \frac{1}{T_D} \sum_{k=-\infty}^{\infty} X_a\left(j\frac{\hat{\omega}}{T_D} - j\frac{2\pi k}{T_D}\right), \quad \hat{\omega} = \omega T_D \text{ [рад]}$$



# Преобразование отсчётов в импульсы

- Цифро-аналоговое преобразование (ЦАП)

Дискретный  
сигнал

Дискретизи-  
рованный сигнал

Аналоговый  
сигнал

$$x(n) = x_a(nT_d)$$

Преобр. отсчётов  
в импульсы

$$\hat{x}(t) = \sum_{n=-\infty}^{\infty} x(n) \delta(t - nT_d)$$

$$x_a(t) = \sum_{n=-\infty}^{\infty} x(n) \frac{\sin\{\pi(t - nT_d)/T_d\}}{\pi(t - nT_d)/T_d}$$

Ряд Котельникова – фильтрация с помощью идеального низкочастотного фильтра. На практике ряд Котельникова заменяется на интерполятор нулевого порядка с компенсационным фильтром



# Идея алгоритма быстрого преобразования Фурье

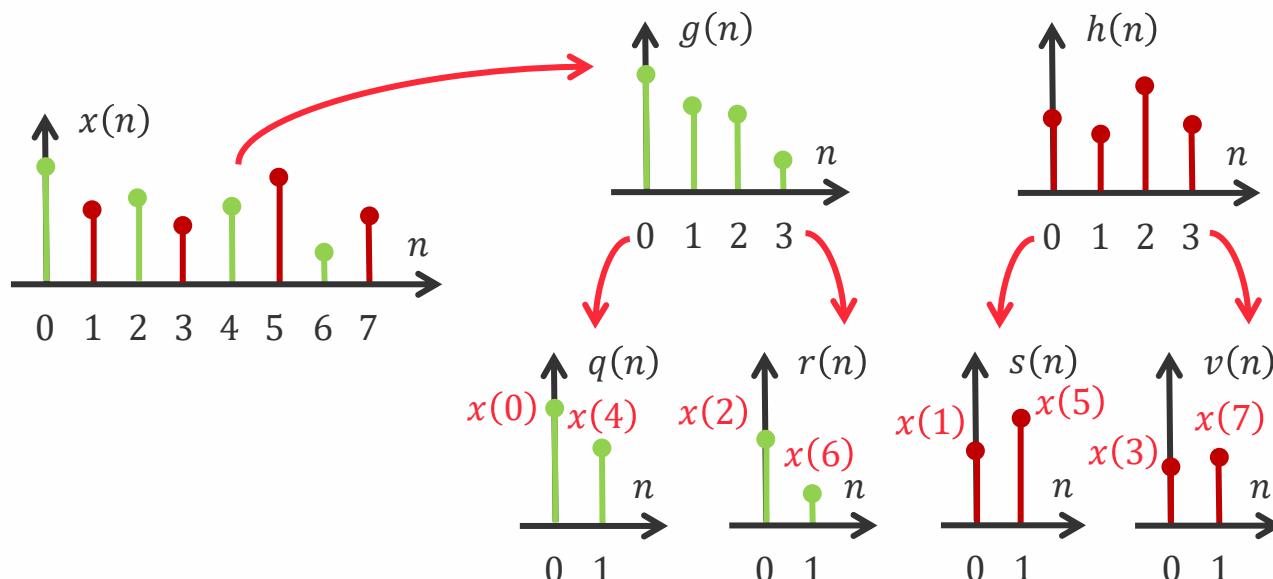
- 1965 год, Д.У. Кули и Дж. Тьюки предложили первый алгоритм (**быстрое преобразование Фурье с прореживанием по времени**), реализующий дискретное преобразование Фурье (ДПФ) «быстро»
  - Основной принцип – «разделяй и властвуй», разбить сложную задачу на несколько простых, решить простые задачи, объединить решение простых задач в решение сложной задачи
  - Сложность дискретного преобразования Фурье последовательности длины  $N$  составляет  $O(N^2)$
  - Если разделить последовательность длины  $N$  на две последовательности длины  $N/2$ , то сложность ДПФ для каждой из последовательностей по отдельности составит  $O(N^2/4)$ , а суммарная сложность составит  $O(N^2/2)$
  - Если продолжать деление вплоть до двухточечных ДПФ, то сложность составит  $O(N \log_2 N)$

**ДПФ для 2-точечной последовательности**

$$X(0) = x(0) + x(1), X(1) = x(0) - x(1)$$

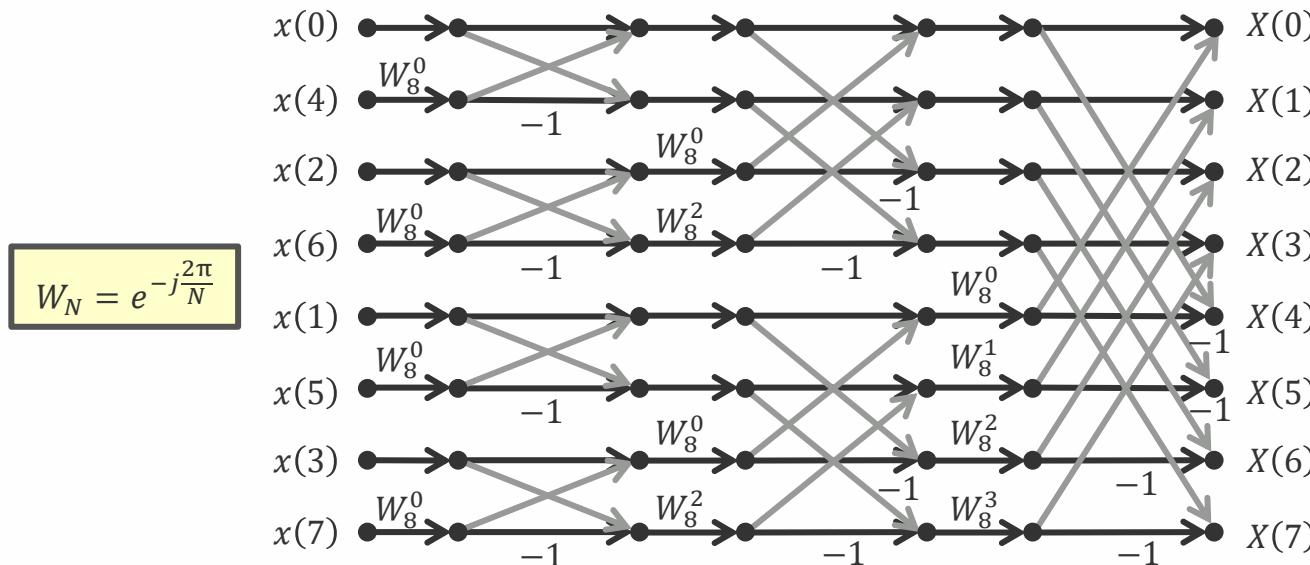
# Идея алгоритма быстрого преобразования Фурье

- Деление на отсчёты с чётными и нечётными номерами для 8-точечной последовательности



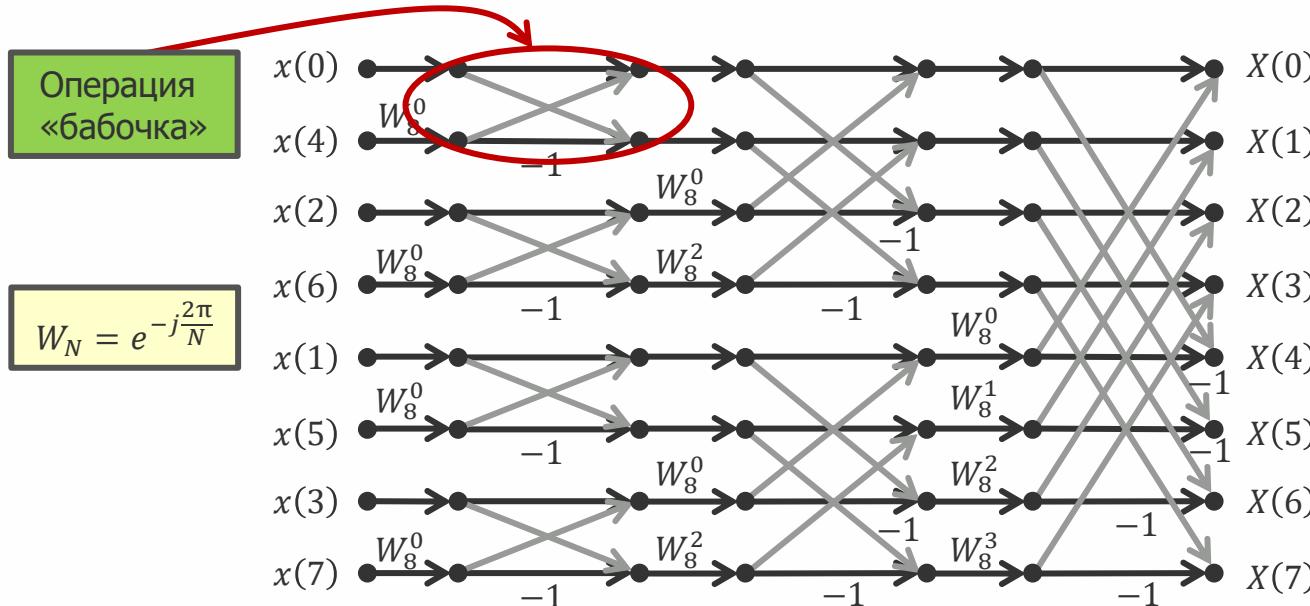
# Идея алгоритма быстрого преобразования Фурье

- Граф 8-точечного быстрого преобразования Фурье с прореживанием по времени



# Идея алгоритма быстрого преобразования Фурье

- Граф 8-точечного быстрого преобразования Фурье с прореживанием по времени





# Оконное преобразование Фурье

- Основная идея оконного преобразования Фурье состоит в следующем:
  - взять маленькие кусочки длины  $L$  из некоторого сигнала длины  $N$
  - вычислить дискретное преобразование Фурье каждого кусочка

## Формула анализа

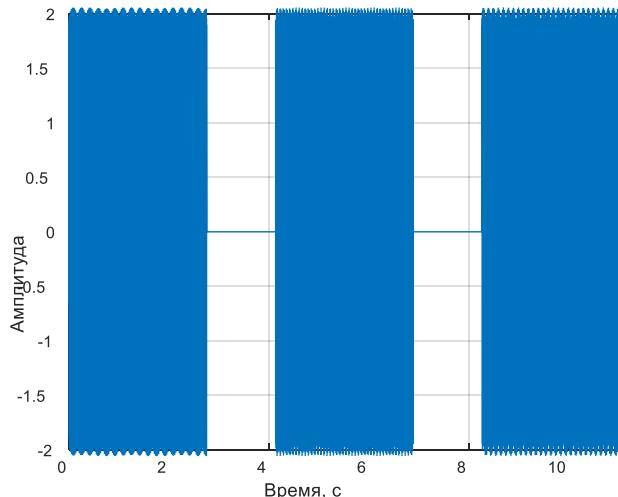
$$X(k, m) = \sum_{n=0}^{L-1} x(n + m)w(n)e^{-j\frac{2\pi}{L}kn},$$

$w(n)$  – это оконная функция,  $m$  – шаг окна

- Основные сложности в использовании оконного преобразования Фурье
  - Как выбрать длину окна?
  - Как выбрать сдвиг окна?
  - Как выбрать форму окна?

# Двухтональный многочастотный сигнал

- Используется в телефонии для коммутации абонентов



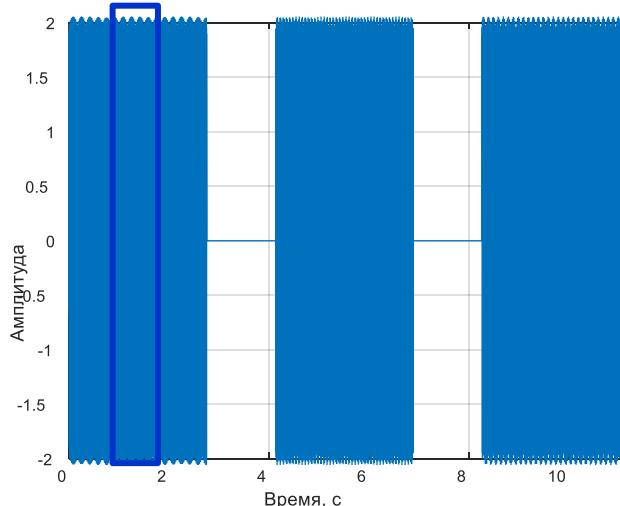
Набор цифр «1-5-9»

697	<b>1</b>	<b>2</b>	<b>3</b>
770	<b>4</b>	<b>5</b>	<b>6</b>
852	<b>7</b>	<b>8</b>	<b>9</b>
941	*	<b>0</b>	#
$f, \text{ Гц}$	1209	1336	1477

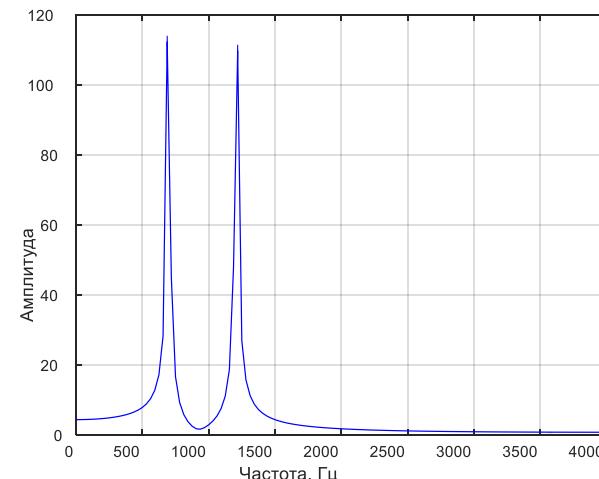
Таблица генерации

# Двухтональный многочастотный сигнал

- Используется в телефонии для коммутации абонентов



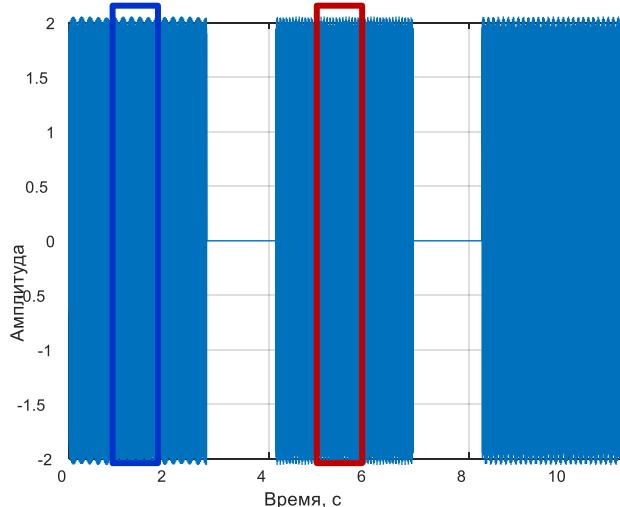
Набор цифр «1-5-9»



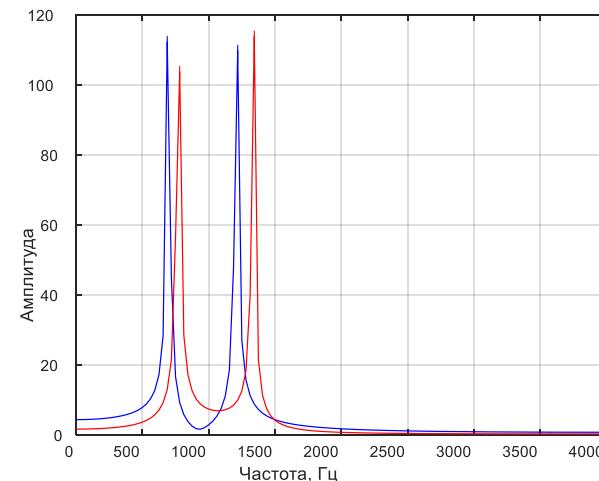
Амплитудный спектр в окне

# Двухтональный многочастотный сигнал

- Используется в телефонии для коммутации абонентов



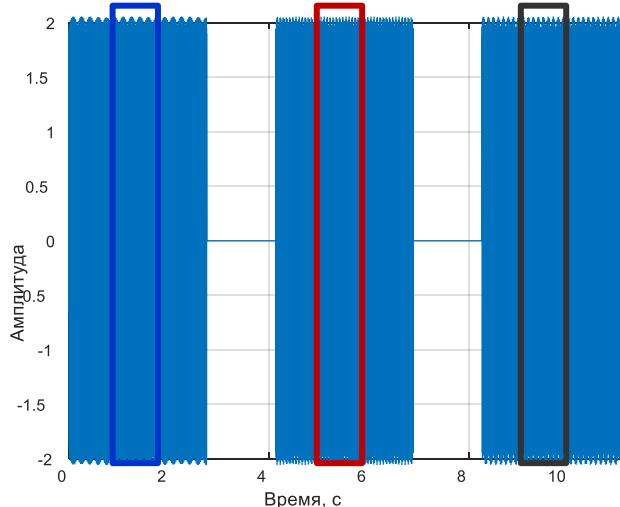
Набор цифр «1-5-9»



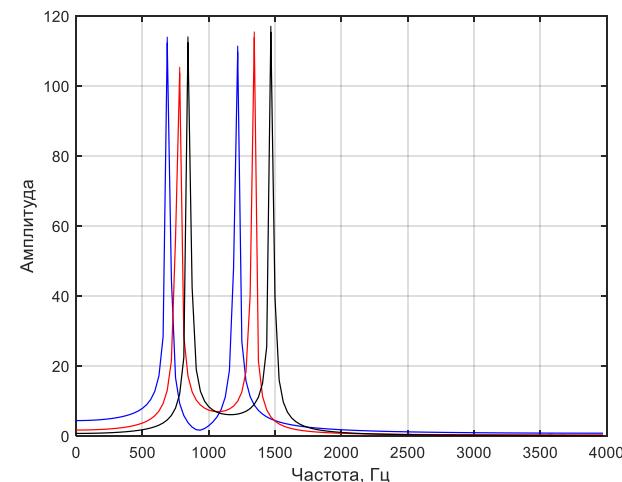
Амплитудный спектр в окне

# Двухтональный многочастотный сигнал

- Используется в телефонии для коммутации абонентов



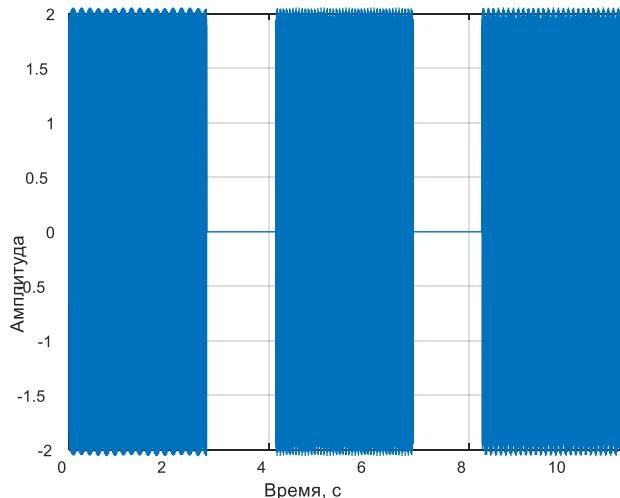
Набор цифр «1-5-9»



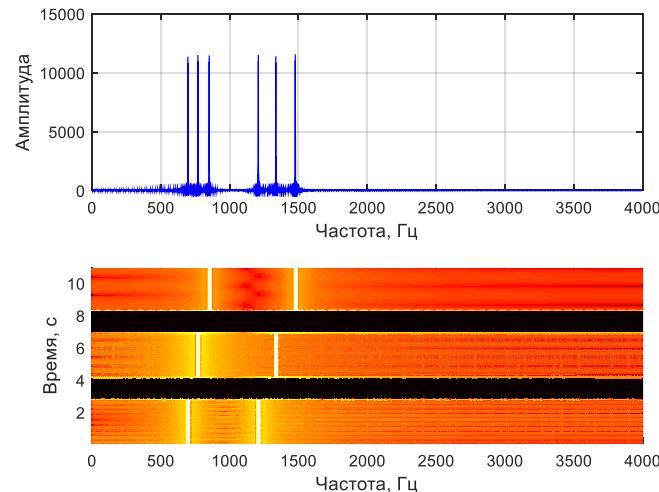
Амплитудный спектр в окне

# Двухтональный многочастотный сигнал

- Используется в телефонии для коммутации абонентов



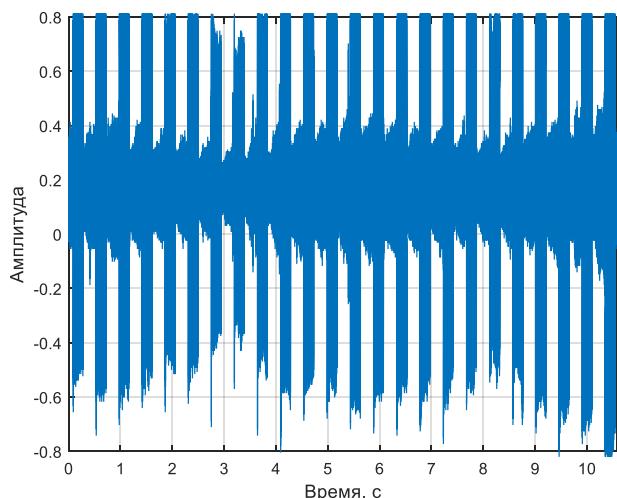
Набор цифр «1-5-9»



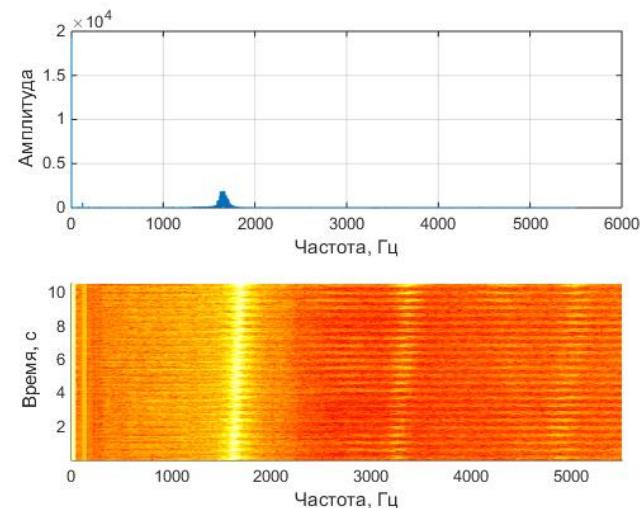
Амплитудный спектр и спектrogramма

# Сигнал первого спутника

- 4 октября 1957 года в СССР был запущен первый искусственный спутник Земли «Спутник-1»



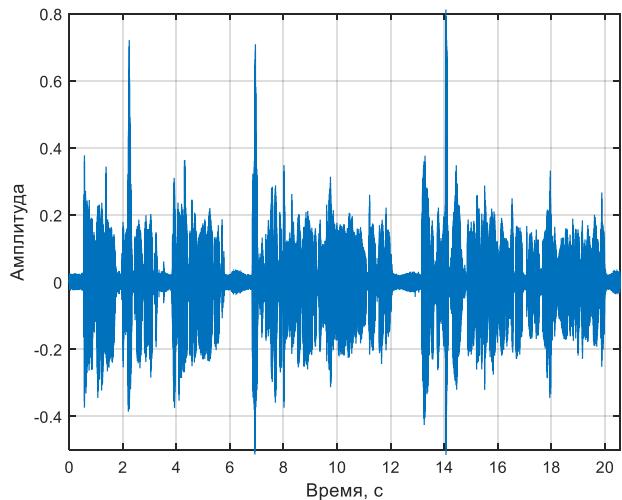
«Бип-бип» сигнал «Спутника-1»



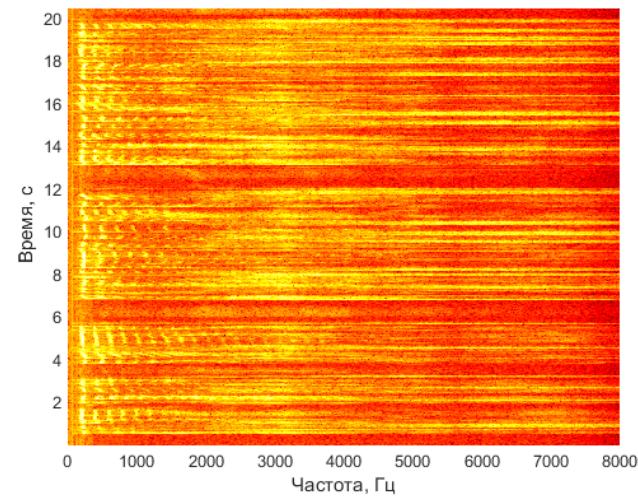
Амплитудный спектр и спектrogramма

# Речевой сигнал

- Обработка речевого сигнала с большим размером окна



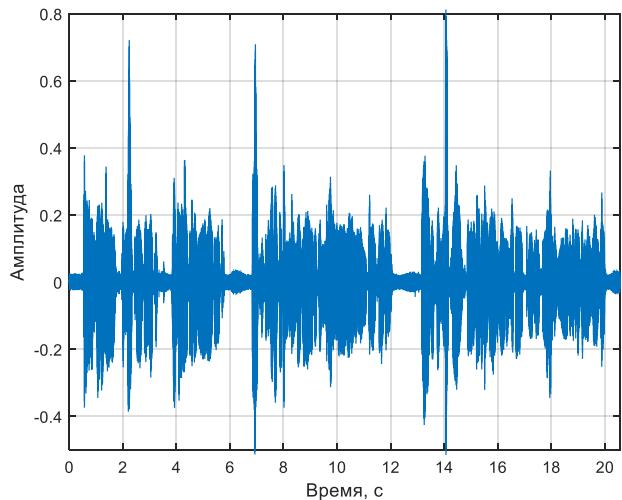
Речевой сигнал



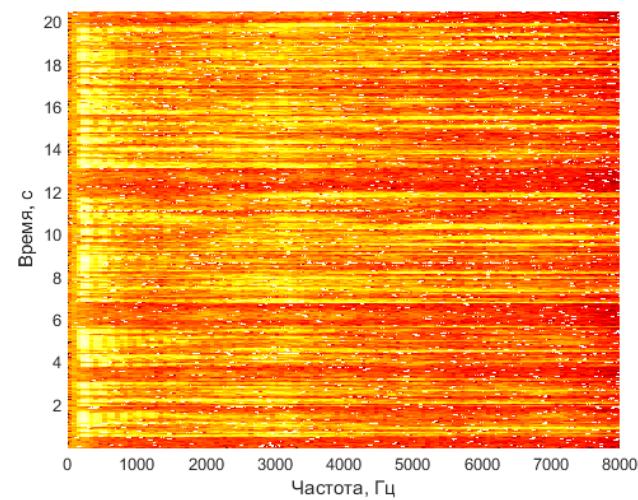
Спектрограмма

# Речевой сигнал

- Обработка речевого сигнала с малым размером окна



Речевой сигнал

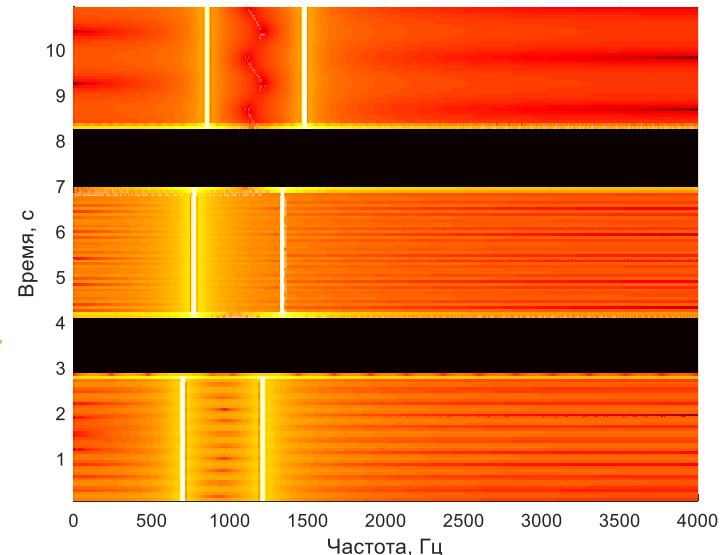


Спектрограмма

# Спектрограмма и ее разметка

- Если  $T_d$  известен, то возможно разметить частотную и временную оси
  - Наибольшая положит. частота:  $f_d/2$  [Гц]
  - Частотное разрешение:  $f_d/L$  [Гц]
  - Ширина временного куска:  $L T_d$  [с]

Пример спектрограммы двухтонального многочастотного сигнала, имеющего частоту дискретизации 8000 [Гц] и длительность около 11 [с]





# Основные сложности в использовании оконного преобразования Фурье

---

- Как выбрать длину окна?
  - **Широкое окно:** узкополосная спектrogramма; больше точек ДПФ и выше частотное разрешение; большей событий может произойти и малая точность по времени
  - **Узкое окно:** широкополосная спектrogramма; много временных кусков и точная позиция смены событий во временной области; несколько точек ДПФ и бедное частотное разрешение
- Как выбрать сдвиг окна?
  - **Малый сдвиг:** большая вычислительная сложность и малая вероятность пропуска событий
  - **Большой сдвиг:** малая вычислительная сложность и большая вероятность пропуска событий
- Как выбрать форму окна?
  - Форма окна влияет на искажение получившегося спектра, которые особенно заметны в областях резкой смены спектрального уровня, а также областях с константными уровнями. **Наиболее обычным на практике для речевой обработки является окно Хэмминга!**
  - Читаем про **вейвлет-преобразование**, как дальнейшее развитие идей оконного преобразования Фурье (**Вейвлеты в обработке сигналов [Малла С., 2005]**)!

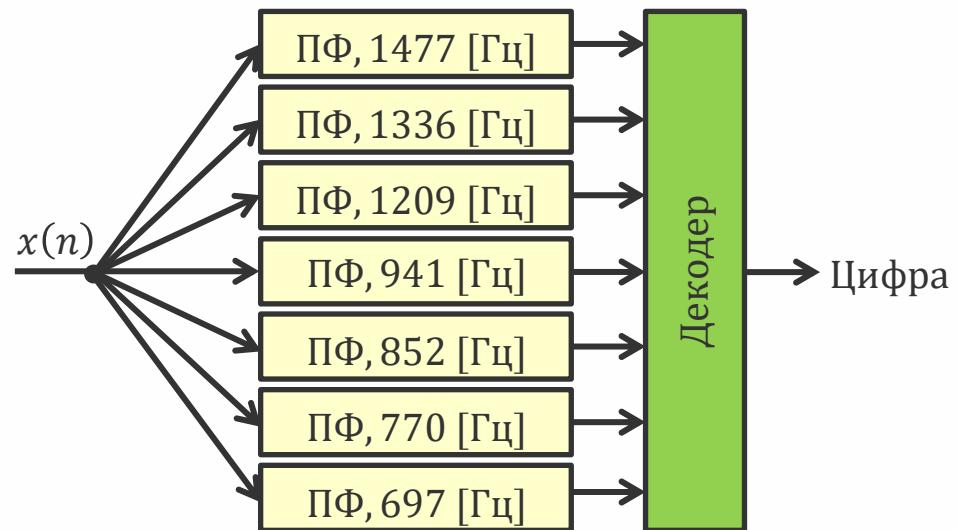
# Банк фильтров

- Банк фильтров – набор полосовых фильтров (ПФ) с общим входом (банком фильтров анализа) или суммирующим выходом (банком фильтров синтеза)

Банк фильтров можно использовать для декодирования двухтонального многочастотного сигнала

Полосовой фильтр используется для выделения требуемой полосы частот

Декодер для принятия решения подсчитывает энергию каждого из семи сигналов на выходе фильтров



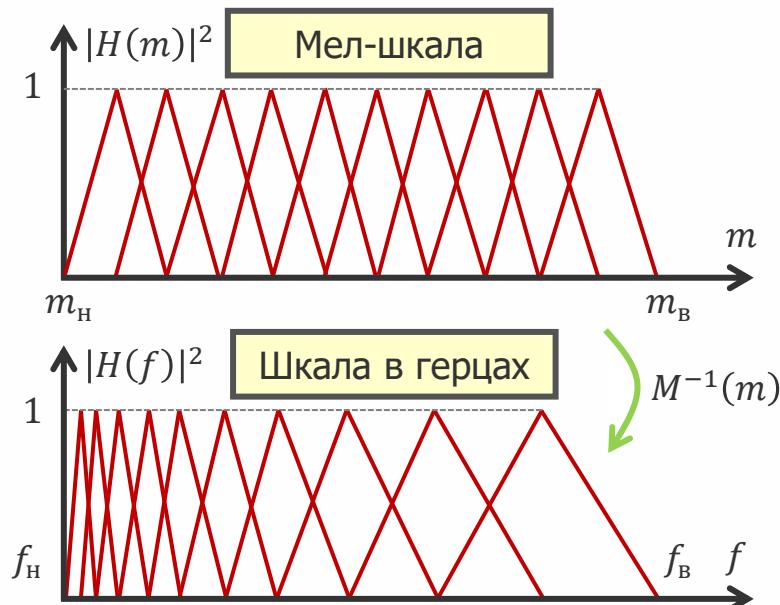


# Вычисление мел-частотных кепстральных коэффициентов

- Обычно мел-частотные кепстральные коэффициенты для речевого сигнала вычисляются внутри скользящего окна, например Хэмминга (25 [мс] – ширина, 10 [мс] – шаг)
- **Мел-шкала:**  $M(f) = 1125 \cdot \ln(1 + f/700)$  [мел],  $M^{-1}(m) = 700 \cdot e^{m/1125-1}$  [Гц]
- Вычислить спектр для куска сигнала, с частотой дискретизации  $f_d$ , умноженного на окно длины  $L$
- Рассчитать мел-банк фильтров, задав число фильтров, верхнюю и нижнюю частоты на  $[0, f_d/2]$
- Перемножить квадрат амплитудно-частотной характеристики каждого фильтра со спектром мощности сигнала  $|X(k)|^2/L$  и просуммировать коэффициенты получившихся спектров, рассчитав энергии внутри соответствующих полос банка фильтров
- Вычислить логарифм от значений энергий на предыдущем шаге
- Вычислить дискретное косинусное преобразование от логарифмов коэффициентов
- **За подробностями можно обратиться сюда:** <http://practicalcryptography.com/miscellaneous/>

# Вычисление мел-частотных кепстральных коэффициентов

- А вот небольшие пояснения к сказанному!



Почему длина окна 25 [мс]? Это интервал стационарности речевого сигнала!

Зачем мел-шкала? Человек более восприимчив к изменению высоты тона на низких частотах, чем на верхних

Почему логарифм от значений энергий? Для удвоения восприятия громкости звука человеком требуется увеличить энергию звукового колебания в 8 раз

Зачем дискретное косинусное преобразование? Появляется возможность уменьшить число используемых коэффициентов; коэффициенты становятся декоррелизованными



# Благодарности

---

- В лекции использовались материалы следующих книг:
  - Солонина А.И. и др. Основы цифровой обработки сигналов, 2005
  - Prandoni P., Vetterli M. Signal processing for communications, 2008
  - Hayes M.H. Schaum's outlines of digital signal processing, 2011
  - Николенко С., Кадурин А., Архангельская Е. Глубокое обучение, 2018
- В лекции использовались материалы следующих онлайн-курсов:
  - Oppenheim A.V. Digital signal processing (RES.6-008), 1975. Massachusetts Institute of Technology
  - Peters R.A. Image processing (EECE/CS 253), 2007. Vanderbilt University School of Engineering
  - Prandoni P., Vetterli M. Digital signal processing, 2014. Coursera's online course of École polytechnique fédérale de Lausanne
  - Ng A. Convolutional neural networks, 2019. Coursera's online course of Stanford University
  - Ng A. Sequence models, 2019. Coursera's online course of Stanford University
- Отдельное спасибо коллегам по работе, без которых бы эта лекция не состоялась!

# Группа компаний ЦРТ

## О НАС

В группу компаний ЦРТ входят компании ЦРТ, ЦРТ-инновации и SpeechPro.

**ЦРТ** – российская компания, разработчик инновационных систем в сфере технологий синтеза и распознавания речи, анализа аудио- и видеинформации, распознавания лиц, голосовой и мультимодальной биометрии.

**ЦРТ-инновации** – научно-исследовательская компания, передовой разработчик голосовых и бимодальных биометрических систем. Резидент Фонда «Сколково». Области научно-исследовательской деятельности компании: биометрия по голосу и лицу, распознавание речи, анализ больших данных.

**SpeechPro** – представительство Группы ЦРТ в Северной Америке с главным офисом в Нью-Йорке. SpeechPro взаимодействует с клиентами и партнерами ЦРТ из США и Канады.

## КОНТАКТНАЯ ИНФОРМАЦИЯ

### Санкт-Петербург

Адрес: 194044, г. Санкт-Петербург,  
ул. Гельсингфорская, 3-11, лит. Д  
Телефон: (+7 812) 325-88-48  
Факс: (+7 812) 327-92-97  
Эл. почта: [stc-spb@speechpro.com](mailto:stc-spb@speechpro.com)

### Москва

Адрес: Москва, ул. Марксистская, д.3, стр.2,  
Бизнес-центр «Таганский»  
Телефон: +7 (495) 669-74-40  
Факс: +7 (495) 669-74-44  
Эл. почта: [stc-msk@speechpro.com](mailto:stc-msk@speechpro.com)