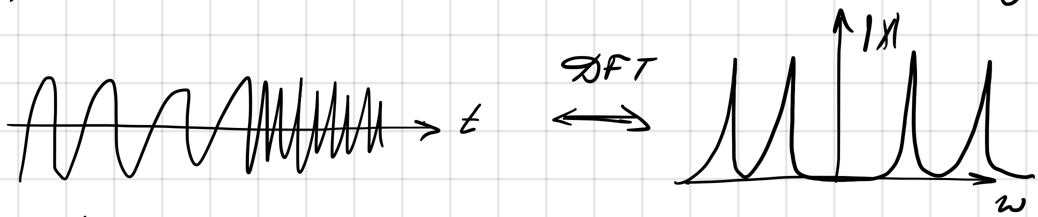
Современные методы распознавания и синтеза речи
Лекция 3

Лекция 3 Частотно-временной анализ

§1 Short-time Fourier transform

Преобразование Фурье показывает „что" находится в сигнале, но не „где". Например, DFS от музыки ⇒ узнаем все ноты, кот-е играли, но конкретный момент знать не будем.

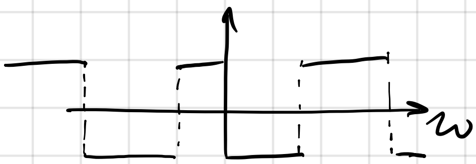
Пример



Временная область

(Time domain)

[сек.]



Частотная область

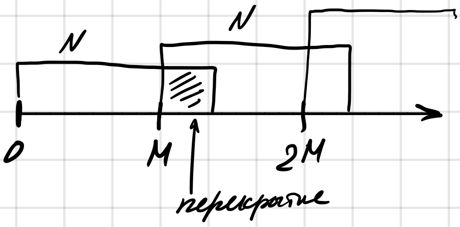
(Frequency domain)

[1/сек]

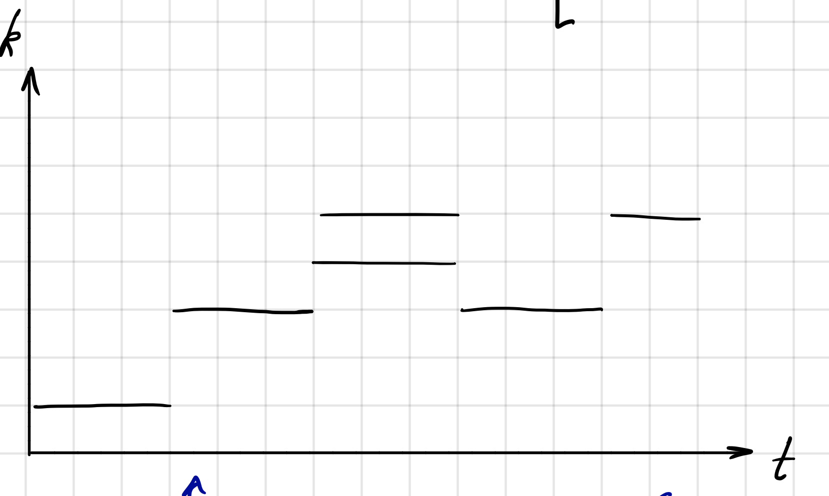
Спектрограмма:

$$S[k,m] = \sum_{i=0}^{N-1} x[mM+i]\, e^{j\frac{2\pi}{N}\cdot ik}$$

частота, время



$$S[k,m] = W_N \begin{bmatrix} x[0] & x[M] & \cdots \\ x[1] & x[M+1] & \cdots \\ x[2] & x[M+2] & \cdots \\ \vdots & \vdots & \ddots \\ x[N-1] & x[N+M-1] & \cdots \end{bmatrix}$$



можем восстановить ноты

Принцип неопределенности:

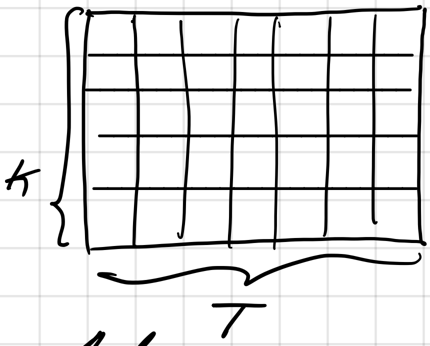
не можем одновременно точно узнать частоту и время.

$N\uparrow$ ⇒ лучше разрешение по частотам, но больше время

$N\downarrow$ ⇒ хуже частотное разрешение, но лучше локализация по времени

Применение: выделение признаков из звука.

- обучаться на исходном сигнале трудно: небольшая задержка или изменение громкости ⇒ сильное отличие сигнала
- спектрограмма более робастна к таким изменениям



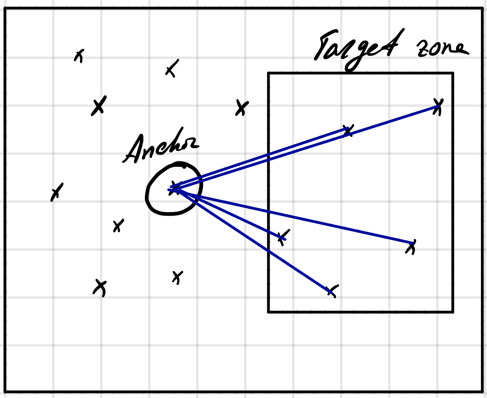
$\xrightarrow{ML}$

- детекция ключевой фразы
- идентификация голоса
- поиск конца фразы

Audio fingerprinting (Shazam)

1. Выделяем локальные пики в спектрограмме. Эти признаки робастны к сильному шуму.
   > Фильтр высоких частот
   > Глоб. max в окрестности

2. Выделение отпечатков (fingerprinting)

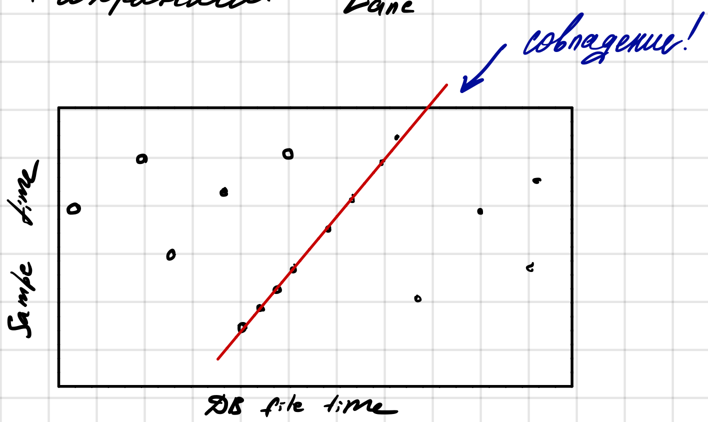


Для каждой пары: $[f_{anc}, f_{tg}, \Delta t]$

Снижение памяти: квантование f и Δt, запись одним числом, SHA для повышения энтропии (быстрее поиск)

\+ сохраняем t_{anc}

3. Поиск аудиофрагмента:

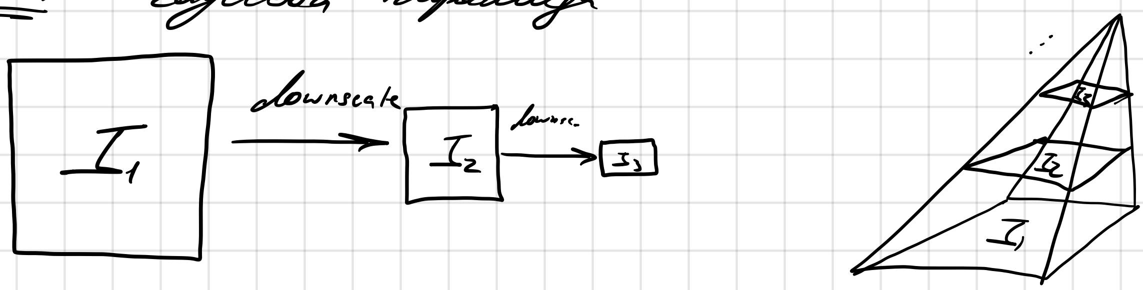


§2 Multiresolution & subband coding

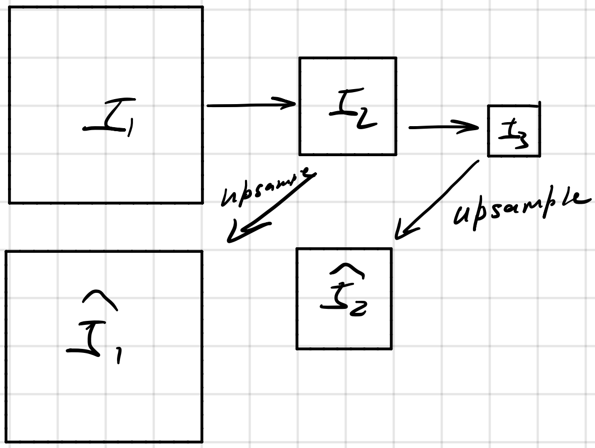
Идея: выделим низкочастотные компоненты, удалим их из сигнала. Выделим более высокие частоты, удалим их, etc.

Постепенно раскручиваем сигнал, разделяя его на компоненты.

Пример Гауссова пирамида

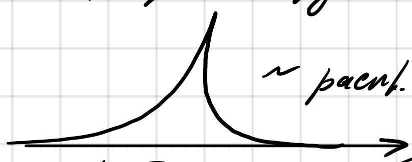


Получим представление на разных уровнях. Посмотрим на ошибку реконструкции



$D_k = I_k - \text{upsample}(I_{k+1})$

ошибка реконструкции.



~ распр. Лапласа.

Удобно использовать для сжатия изображений!

Пирамида Лапласа:

$I_K,\ D_{K-1},\ D_{K-2}, \ldots, D_1.$

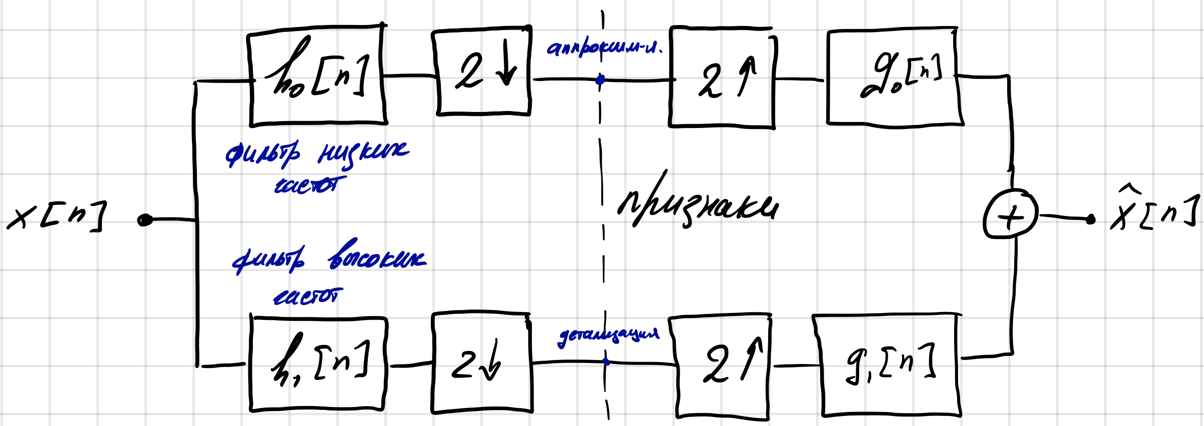
передаем так как есть

сильно квантуем.

Техника „multiresolution analysis“

Subband coding

Разделяем на низкочастотную (аппроксимация) и высокочастотную (детализация) компоненты.



Как должны быть связаны h_0, h_1, g_0, g_1? Хотим восстанавливать без ошибок.

$$X(z) = \sum_{n=-\infty}^{+\infty} x[n] z^{-n}$$

$$x_{down}[n] = x[2n] \Longleftrightarrow X_{down}(z) = \frac{1}{2}\left(X(z^{1/2}) + X(-z^{1/2})\right)$$

$$X_{down}(z) = \sum_{n=-\infty}^{+\infty} x_{down}[n] z^{-n} = \sum_{n=-\infty}^{+\infty} x[2n] z^{-n} = \sum_{n=-\infty}^{+\infty} x[2n] (z^{1/2})^{-2n} = \frac{1}{2}\left(X(z^{1/2}) - X(-z^{1/2})\right)$$

$$x^{up}[n] = \begin{cases} x[n/2], & n = 0, 2, 4, \dots \\ 0, & \text{otherwise} \end{cases} \Longleftrightarrow X^{up}(z) = X(z^2)$$

$$\hat{X}[n] = \frac{1}{2}\left(\underbrace{X(z)}_{\text{исходный сигнал}} + \underbrace{X(-z)}_{\text{искаженная версия}}\right)$$

Хотим избежать искажений

$$\hat{X}(z) = \frac{1}{2} G_0(z)\left[H_0(z)X(z) + H_0(-z)X(-z)\right] + \frac{1}{2} G_1(z)\left[H_1(z)X(z) + H_1(-z)X(-z)\right] =$$

$$= \frac{1}{2}\underbrace{\left[G_0(z)H_0(z) + G_1(z)H_1(z)\right]}_{= 2} X(z) + \frac{1}{2}\underbrace{\left[G_0(z)H_0(-z) + G_1(z)H_1(-z)\right]}_{= 0} X(-z)$$

$$\begin{bmatrix} G_0(z) & G_1(z) \end{bmatrix} \underbrace{\begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix}}_{H_m} = \begin{bmatrix} 2 & 0 \end{bmatrix}$$

$$H_m^{-1} = \frac{1}{\det H_m}\begin{bmatrix} H_1(-z) & -H_1(z) \\ -H_0(-z) & H_0(z) \end{bmatrix}$$

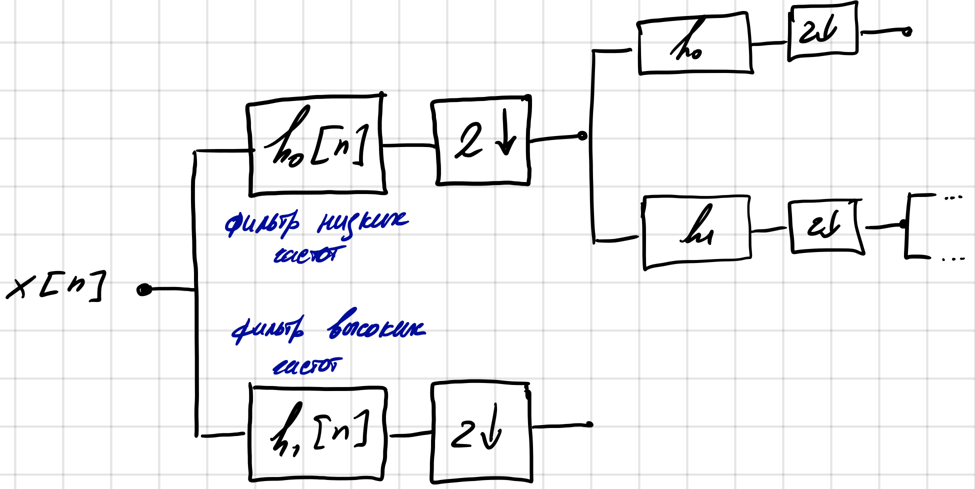
$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \frac{2}{\det(H_m(z))}\begin{bmatrix} H_1(-z) \\ -H_0(-z) \end{bmatrix}$$

Интересно что H_1 опр-т G_0, H_0 — G_1

для FIR $\alpha \cdot z^{-(2k-1)}$

Высокочастотную компоненту можно дополнить раскладывать: вновь разделить на несколько полос частот.

Применение: разное кодирование для разных полос



§3 Вейвлет преобразование

Пусть $\varphi(x)$ — квадратично суммируемая функция. Применим к ней целочисленные сдвиги и растяжения:

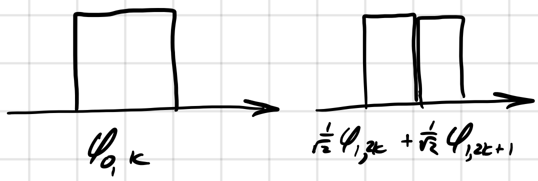
$$\varphi_{J,k}(x) = 2^{J/2}\varphi(2^J x - k)$$

$\varphi(x)$ – scaling function / father wavelet.

Порождаемые пространства: $V_J = \operatorname{Span}_k \{\varphi_{J,k}(x)\} \subset L_2(R)$

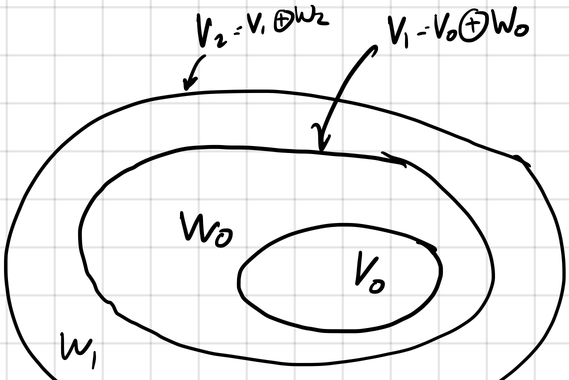
Пример: $\varphi(x) = \begin{cases} 1, & x \in [0, 1) \\ 0, & \text{иначе} \end{cases}$

скейлинг функция Хаара



$$\varphi_{0,k}(x) = \frac{1}{\sqrt{2}}\varphi_{1,2k}(x) + \frac{1}{\sqrt{2}}\varphi_{1,2k+1}(x)$$

$$V_1 \subset V_2 \subset V_3 \subset \ldots \subset L_2(R) = V_\infty$$



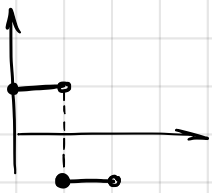
Будем раскладывать функцию по подпр-м:

$$L^2(\mathbb{R}) = V_0 \oplus W_1 \oplus W_2 \oplus \ldots$$

Как построить базис W_k? Мы знаем базис V_k, V_{k+1}, $V_{k+1} = V_k \oplus W_k$. Ортогонализация Грама-Шмидта!

$$\langle \varphi_{J,k}(x), \psi_{J,l}(x) \rangle = 0 \quad \forall J, k, l \in \mathbb{Z}$$

Для скейлинг-функции Хаара: $\psi(x) = \begin{cases} 1, & 0 \le x < 0.5 \\ -1, & 0.5 \le x < 1 \\ 0, & \text{иначе} \end{cases}$



$$\psi_{J,l}(x) = 2^{J/2}\psi(2^J x - l)$$

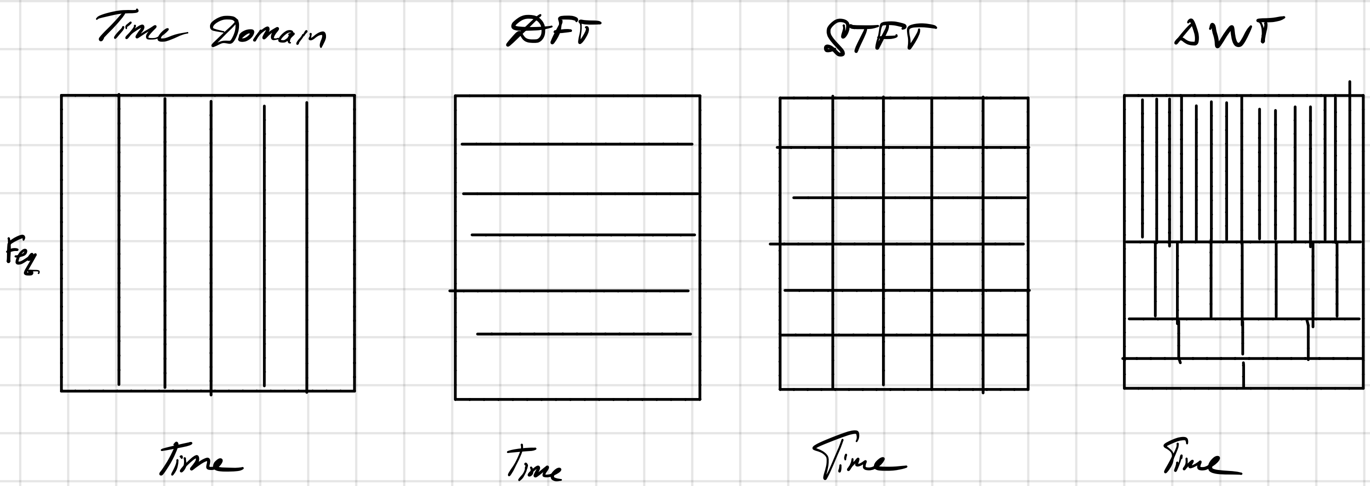
материнский вейвлет

Разложение: $f(x) = \underbrace{\sum_k c_{J_0}(k)\,\varphi_{J_0,k}(x)}_{\text{аппроксимация}} + \underbrace{\sum_{J=J_0}^{+\infty}\sum_k d_J(k)\,\psi_{J,k}(x)}_{\text{детализация}}$

$$c_{J_0} = \langle \varphi_{J_0,k}(x), f(x) \rangle$$

$$d_J(k) = \langle \psi_{J,k}(x), f(x) \rangle$$

По аналогии с DTFT: непр. вейвлет преобразование: $\psi_{s,\tau}(x) = \frac{1}{\sqrt{s}}\psi\left(\frac{x-\tau}{s}\right)$



Применение: сжатие (JPEG-2000), шумоподавление.

Анализ: CWT показывает схожесть участка сигнала сдвинутому и промасштабированному вейвлету.