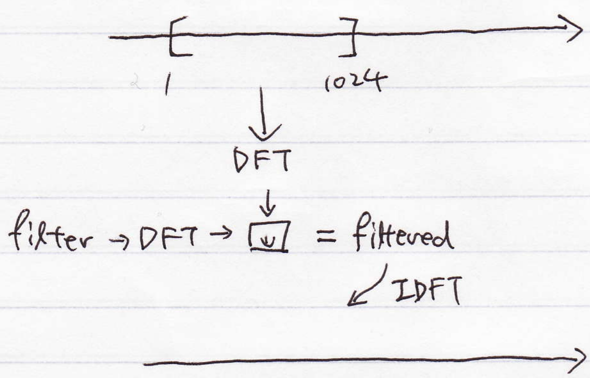
2/8 STFT

$$F(s) = \int_{-\infty}^{\infty} f(t)\, e^{-j2\pi st}\, dt$$

in case of image : entire image (replicated DFT)

in case of audio ?

voice



Same in images : JPEG and MPEG
work over windows of 16 x16 pixels

How to capture "locality"?
→ Dennis Gabor (inventor of holography)
Result of a Fourier - Transform should be a function of frequency as well as location b

* Short Term Fourier Transform

$$STFT\{f(t)\} = F(\underset{\text{freq}}{s}, \underset{\text{location}}{b}) = \int_{-\infty}^{\infty} f(t)\, rect(t-b)\, e^{-j2\pi s(t-b)}\, dt$$

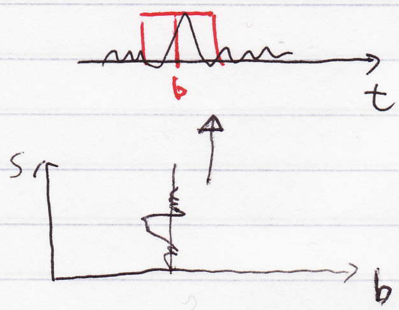
Remarks : 1. This is not a convolution (correlation)
2. Shift by b in $e^{-j\pi s(t-b)}$ does not affect the integration

$$e^{j\pi sb} \cdot \int_{-\infty}^{\infty} f(t)\, rect(t-b)\, e^{-j2\pi st}\, dt$$

3. Inverse STFT

$$f(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(s,b)\, e^{j\pi st}\, ds\, dt$$

$* \quad STFT\{f(t)\}$



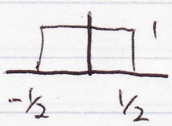
(spectrogram)

TWO Issues with STFT or spectrogram

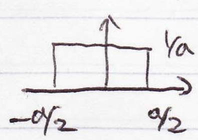
1. "Instantaneous frequency"* what it matters here is <u>the length of the window</u> $\left(^* \; w = \dfrac{d\varphi_{Fourier}}{dt}\right)$

Ideally, we want a varying window rect(t) in order to be able to detect all frequencies.

Introduce also a scale 'a' in the STFT



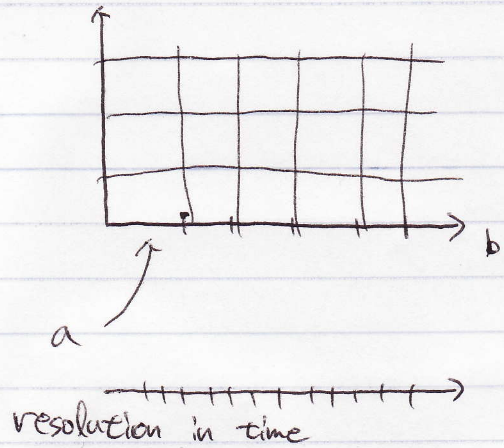
$\rightarrow \quad \frac{1}{a}\,\mathrm{rect}\left(\frac{t}{a}\right)$



$$STFT\{f(t)\} = F(s, a, b) = \frac{1}{a}\int_{-\infty}^{\infty} f(t)\,\mathrm{rect}\left(\frac{t-b}{a}\right) e^{-j2\pi(t-b)}\,dt$$

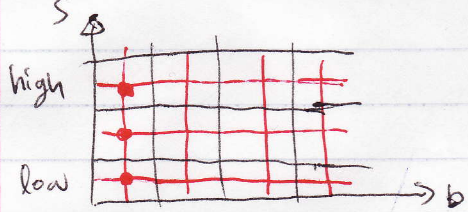
frequency ↗ scale ↑ location ↖

2. How we can ensure that scale 'a' is enough to capture slower frequencies



resolution in frequency fixed.

Remedy : Scale appropriately the frequencies:
Double the resolution in time (smaller windows)
Resolution in frequency is halved.

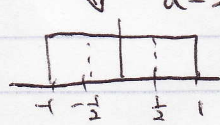


We cannot have infinitely small resolution both in time and frequency

Spectrogram has to be in OCTAVES

$$F(s,a,b) = \int_{-\infty}^{\infty} f(t)\,\mathrm{rect}\left(\frac{t-b}{a}\right) e^{-j2\pi s\left(\frac{t-b}{a}\right)}\,dt$$

$a=2 \Rightarrow \frac{s}{2}$



(wave expanded)

Motivation for <u>wavelets</u>

Def. Wavelet is any function $\Psi(t)$ where fourier tranform satisfies

$$\int_{-\infty}^{\infty} \frac{|\Psi(s)|^2}{|s|}\,ds < \infty$$

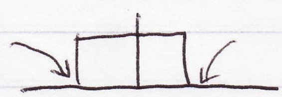
*Wavelet transform of f(t)

$$WT\{f(t)\} = F(s,a,b) = \int_{-\infty}^{\infty} f(t)\,\Psi\left(\frac{t-b}{a}\right)dt$$

* Best way to represent both location and frequency (ex: JPEG 2000)

THIRD ISSUE

Rect has a discontinuity

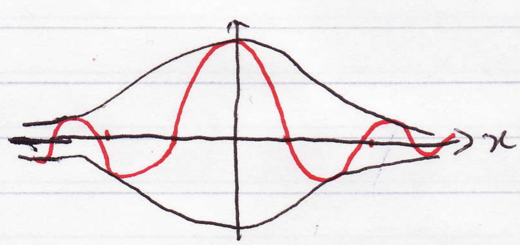


We prefer smooth boundaries : The Gaussian

$$STFT\{f(t)\} = \int_{-\infty}^{\infty} f(t)\, \underbrace{e^{-\frac{(t-b)^2}{2a^2}} \quad e^{-j2\pi s \frac{t-b}{a}}}_{g(t)}\, dt$$

$g(t) = e^{-\frac{t^2}{2}} e^{-j2\pi st}$: Gabor function

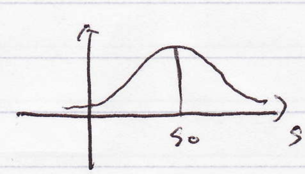
$g(t, a, b)$: Gabor Wavelet



$$e^{\frac{-x^2}{2\sigma^2}} \cos 2\pi s x$$

Quiz What is Fourier of Gabor?

$$e^{-\frac{t^2}{2}} e^{-j2\pi s_0 t} \quad \circ\!\!-\!\!\circ \quad e^{-\frac{(s-s_0)^2}{2}}$$



(positive gaussian only)

$$\boxed{\text{Image Matching Slides}}$$

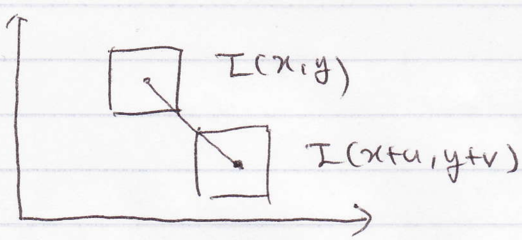
* Image Matching

Features → detection, descriptor

invariance issue → detection, descriptor

invariance issue ↔ trade off ↔ discriminability

Mathematical expression of which point in ☐ is more unique?
(image patch)



$$E(u,v) = \sum_{x,y \in W} \left[I(x+u, y+v) - I(x,y) \right]^2$$

: SSD

Sum of Square Difference

Taylor expansion (first order) $\left(\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v \right)^2$

$$= \left(\begin{pmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \right)^2$$

$$E(u,v) = \sum_{x,y \in W} (u \; v) \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \frac{(\partial I)^2}{\partial x \partial y} \\ \frac{(\partial I)^2}{\partial x \partial y} & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$