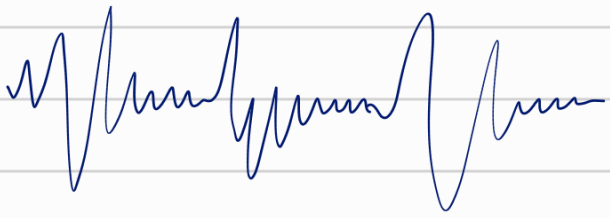
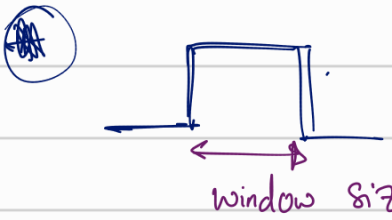
Short-time Fourier Transform.

①  ← audio signal

② 

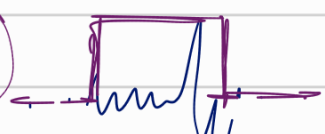
← window

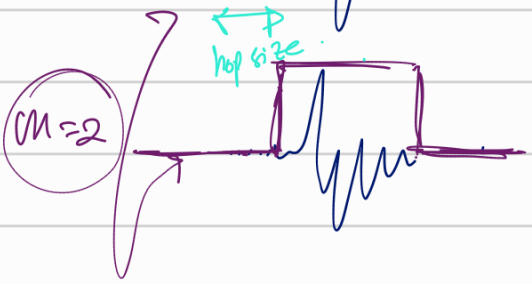
③ 

usually

frame size is the number of pts (samples.) inside the window.

$x_w(t) = x(t) \cdot w(t)$.

M=1  ← gets the dtft.



← gets the dtft.

overlapping

Hop size ← how many samples we slide to the right when we take the new frame.

$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i 2\pi n k / N}$$

$$S(u,k) = \sum_{n=0}^{N-1} x(n + m\,hop) \cdot w(n) \cdot e^{-i 2\pi n k / N}$$

- N is the frame size.
- $S(u,k)$: u = time / frame number; k = Fourier coeffic for kth slot / k frequency
- $m\,hop$: Starting sample of current frame.
- $w(n)$: windowing function

STFT ← Spectral Matrix.

(# frequency bins, # frames)

$$\#\text{frequency bins} = \frac{\text{frame size}}{2} + 1$$

↑ we only get the first half.

$$\#\text{frames} = \frac{\text{Samples} - \text{frame size}}{\text{hope size}} + 1$$

If frame size ↑

freq resolution ↑ time resolution ↓