E9 205 Machine Learning for Signal Processing

Introduction to Machine Learning of Sensory Signals

08-08-2018

Instructor - Sriram Ganapathy (sriramg@iisc.ac.in)

Teaching Assistant - Akshara Soman (aksharas@iisc.ac.in).

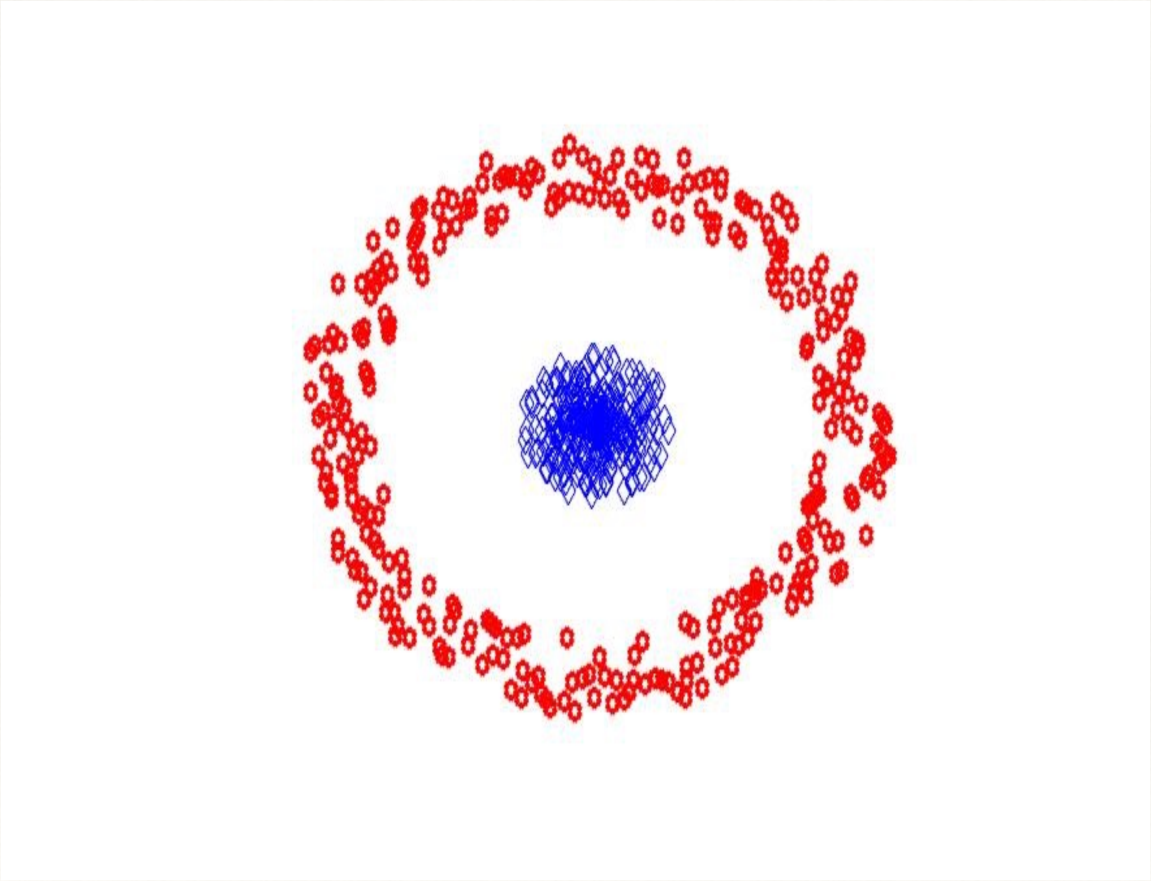
Feature Extraction

- Feature Extraction
  - Using measured data to build desirable values.
  - Attributes of the data that are informative and non-redundant.
  - Resilience to noise/artifacts.
  - Facilitating subsequent learning algorithm.

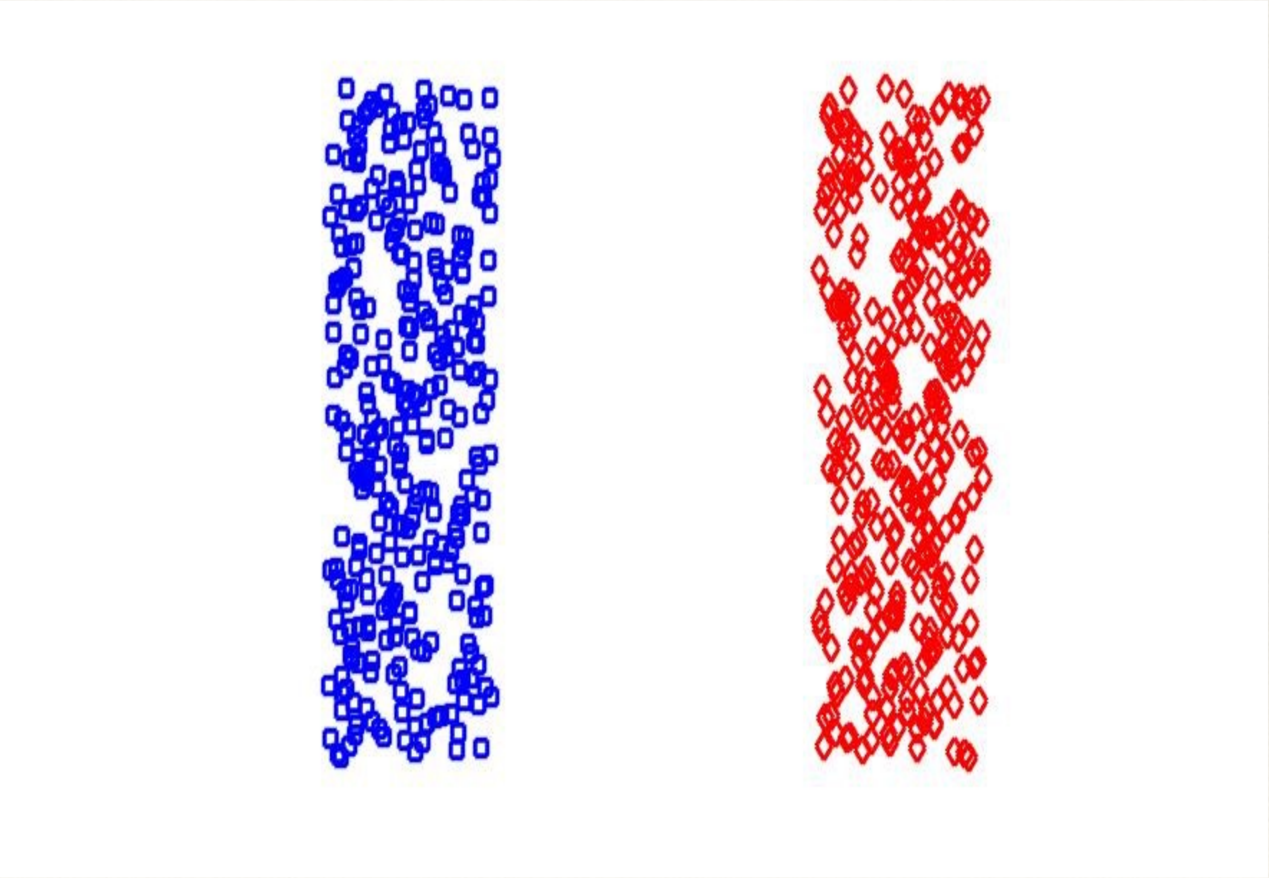
Feature Extraction

- Representation Problem

Cartesian Coordinates



Polar Coordinates



Feature Extraction

Scope for this course

I. Feature Extraction in Text.

II. Feature Extraction in Speech and Audio signals.

III. Feature Extraction for Images.

Text Modeling - Introduction to NLP

- Definitions
    - Documents, Corpora, Tokens (Terms)
- Term Frequency (TF)
- Collection Frequency (CF)
- Document Frequency (DF)
- TF-IDF
- Bag of words model

Text Processing

Example [Manning and Schutze, 2006]

Word	cf	df
try	10422	8760
insurance	10440	3997

► **Figure 6.7** Collection frequency (cf) and document frequency (df) behave differently, as in this example from the Reuters collection.

term	df_t	idf_t
car	18,165	1.65
auto	6723	2.08
insurance	19,241	1.62
best	25,235	1.5

► **Figure 6.8** Example of idf values. Here we give the idf's of terms with various frequencies in the Reuters collection of 806,791 documents.

https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

Perplexity

- Measuring the goodness of language modeling
-

$$\mathrm{PP}(W) = P(w_1 w_2 \ldots w_N)^{-\frac{1}{N}} = \sqrt[N]{\frac{1}{P(w_1 w_2 \ldots w_N)}}$$

On a Wall-street Journal Corpus

	Unigram	**Bigram**	**Trigram**
Perplexity	962	170	109

https://web.stanford.edu/~jurafsky/slp3/4.pdf

Speech and Audio Processing

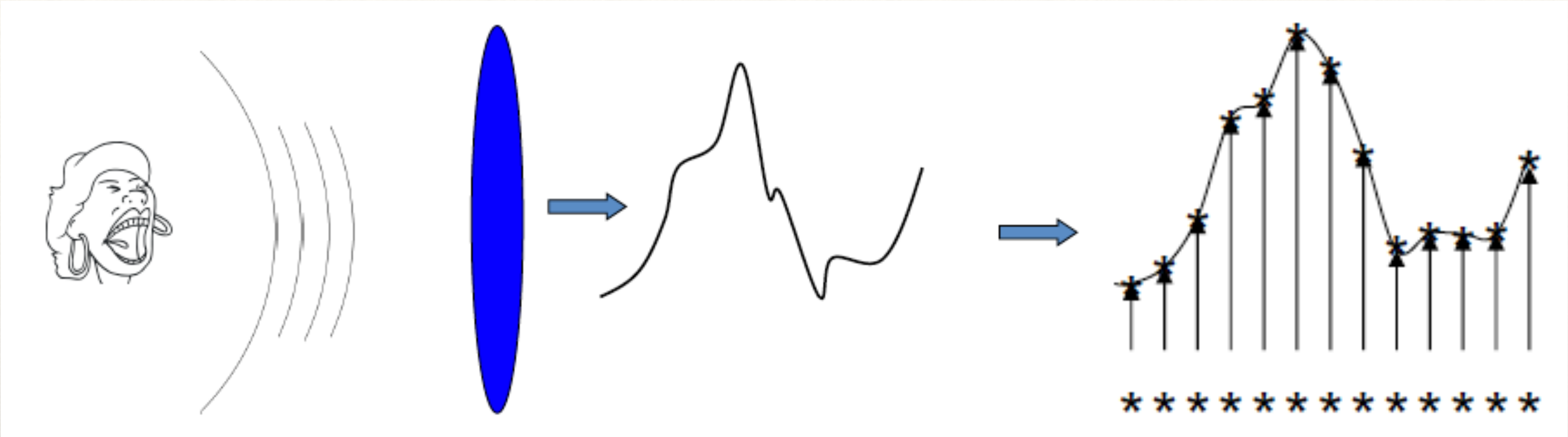
Speech and Audio

- Speech/Audio - 1D signals
  - Generated by pressure variations producing regions of high pressure and low pressure.
  - Travels through a medium of propagation (like air, water etc).
  - Human sensory organ - eardrum.
    - Converting pressure variations to electrical signals.
    - Action mimicked by a microphone.

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

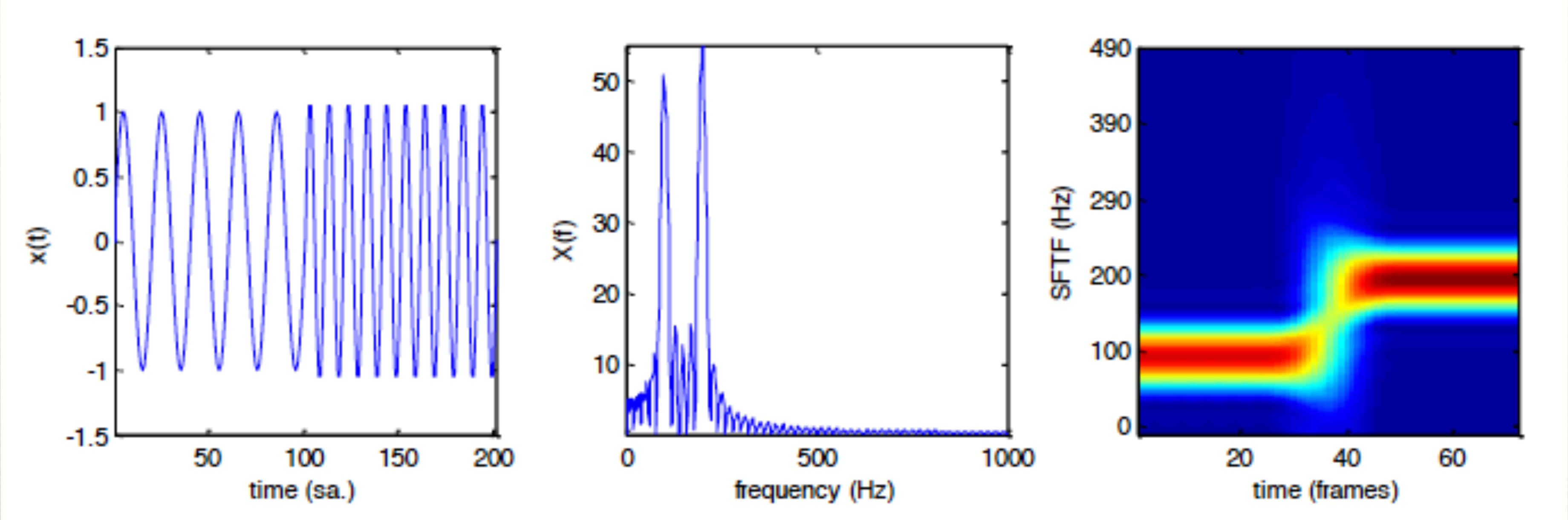
Sound waves in a computer



- Analog continuous signal from the microphone
  - Discretized in time - sampling.
  - Digitized in values - quantization.

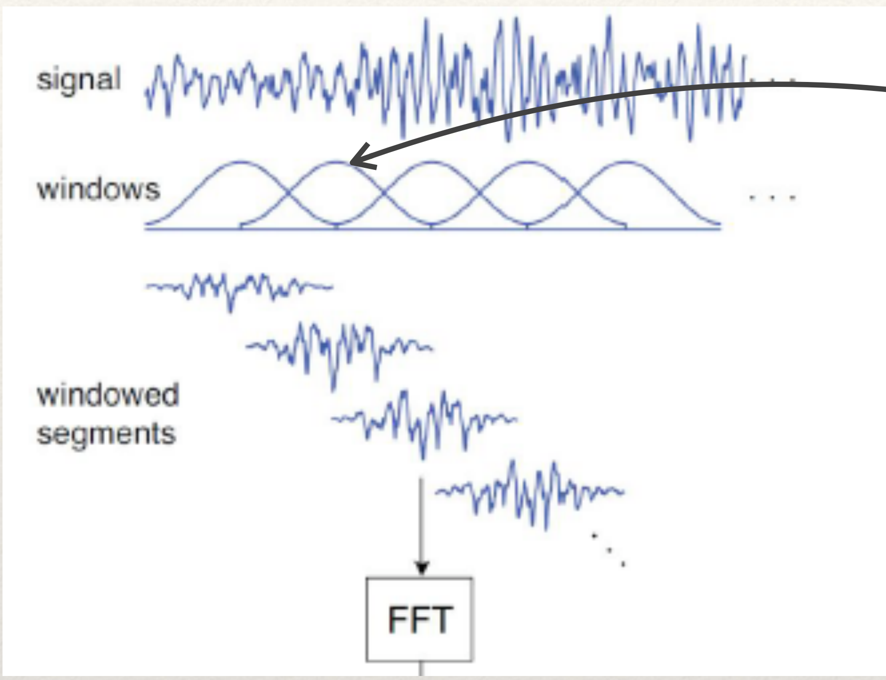
http://mlsp.cs.cmu.edu/courses/fall2014/lectures/slides/Class1.Introduction.pdf

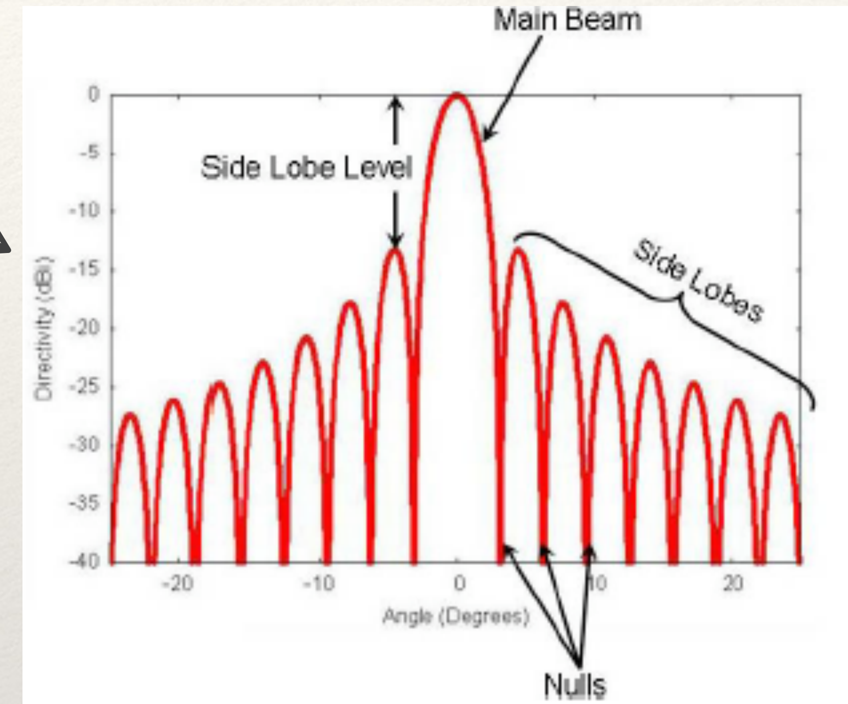
Why do we need time varying Fourier Transform



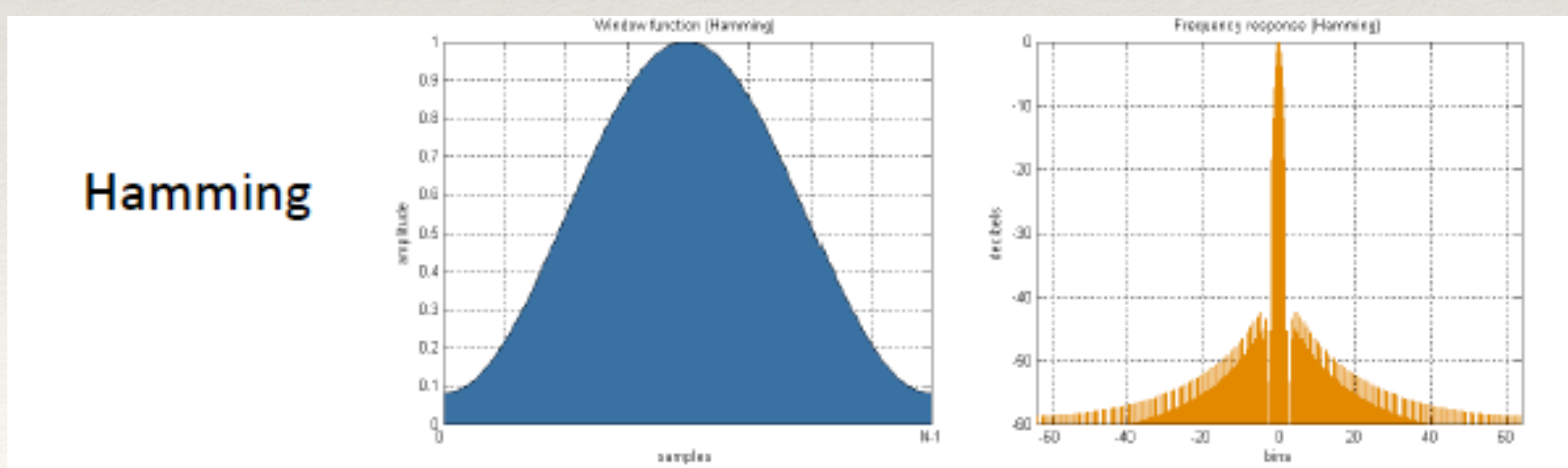
- When the signal properties change in time
  - DFT will only capture the average spectral character
  - Short-window analysis can indicate the change in spectrum.

Summary of STFT Properties





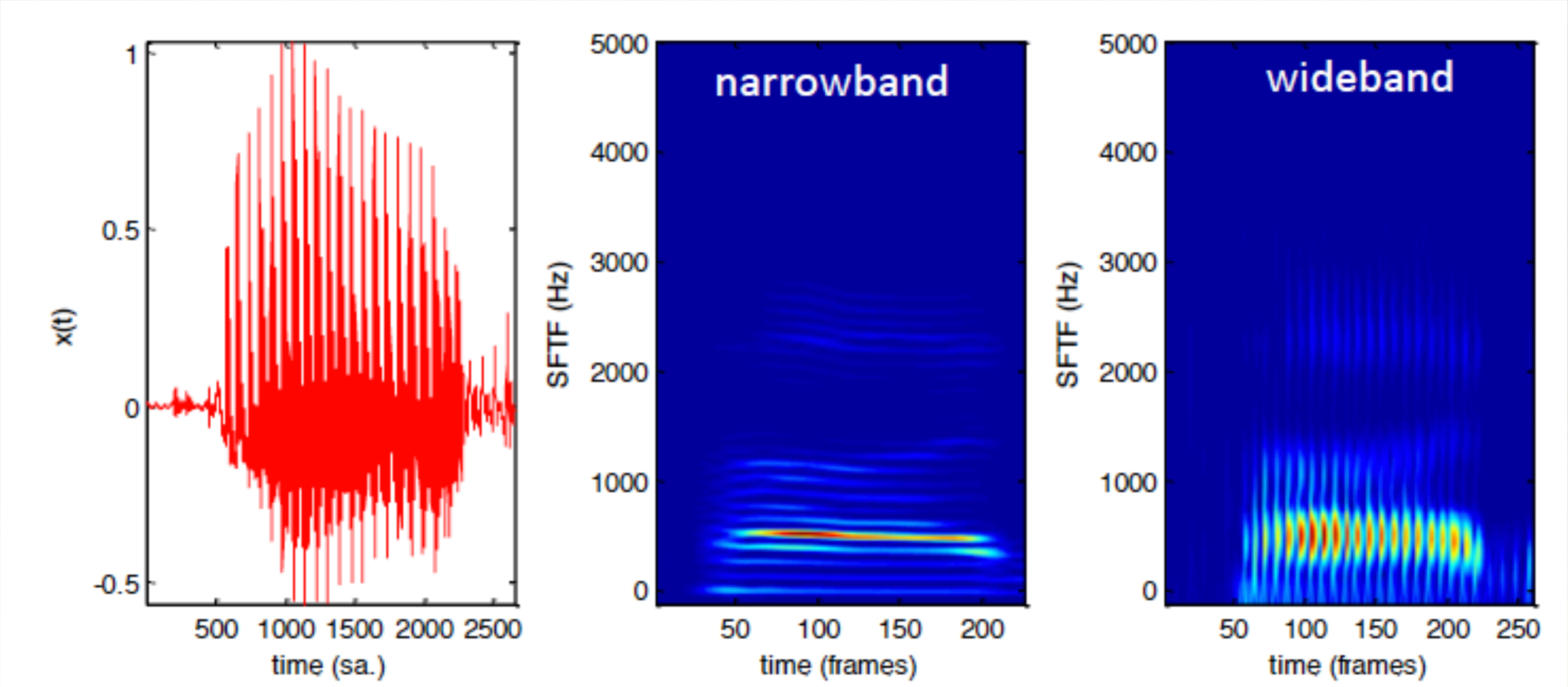
$X[k, n_0]$



http://en.wikipedia.org/wiki/Window_function

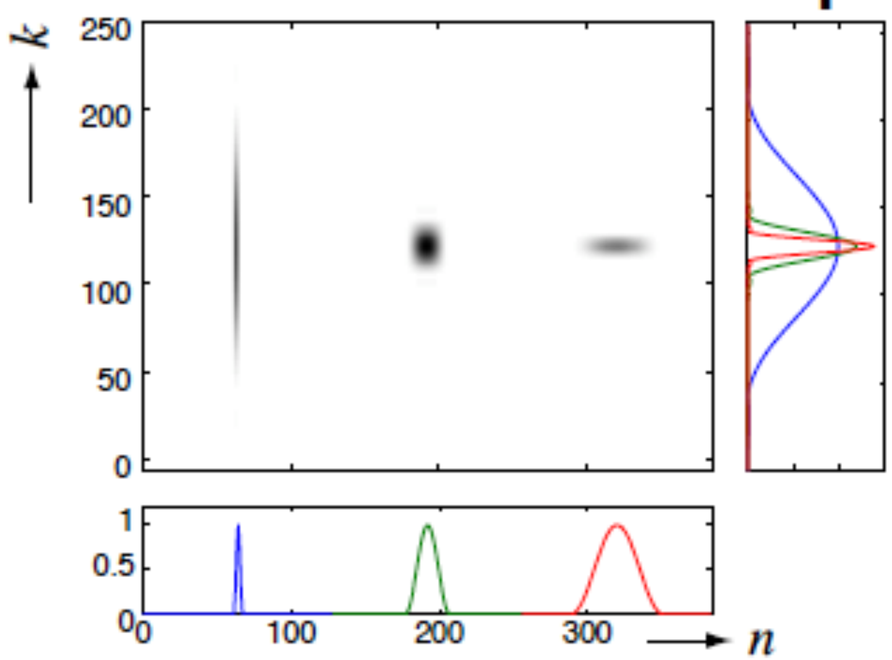
Narrowband versus Wideband

- Short windows - poor frequency resolution - wideband spectrogram
- Long windows - poor time resolution - narrowband spectrogram



Narrowband versus Wideband

- Can illustrate time-frequency tradeoff on the time-frequency plane:

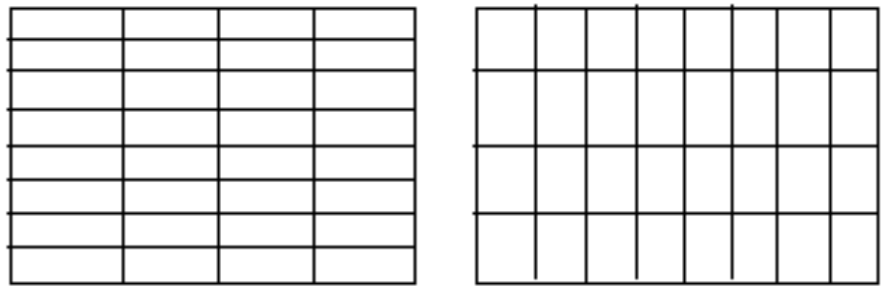


disks show 'blurring' due to window length;
***area** of disk is constant*

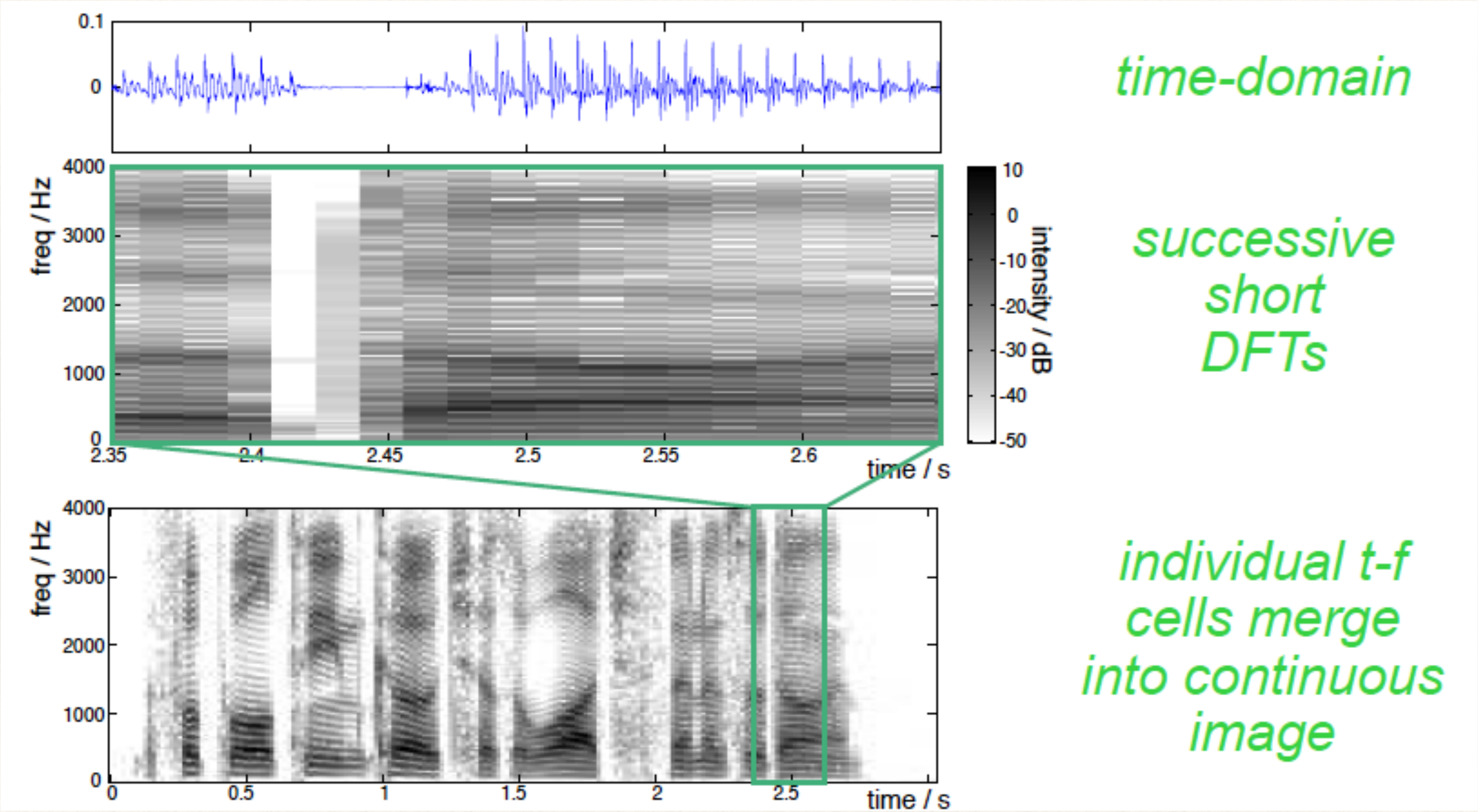
→ ***Uncertainty principle:***

$$\delta f \cdot \delta t \geq k$$

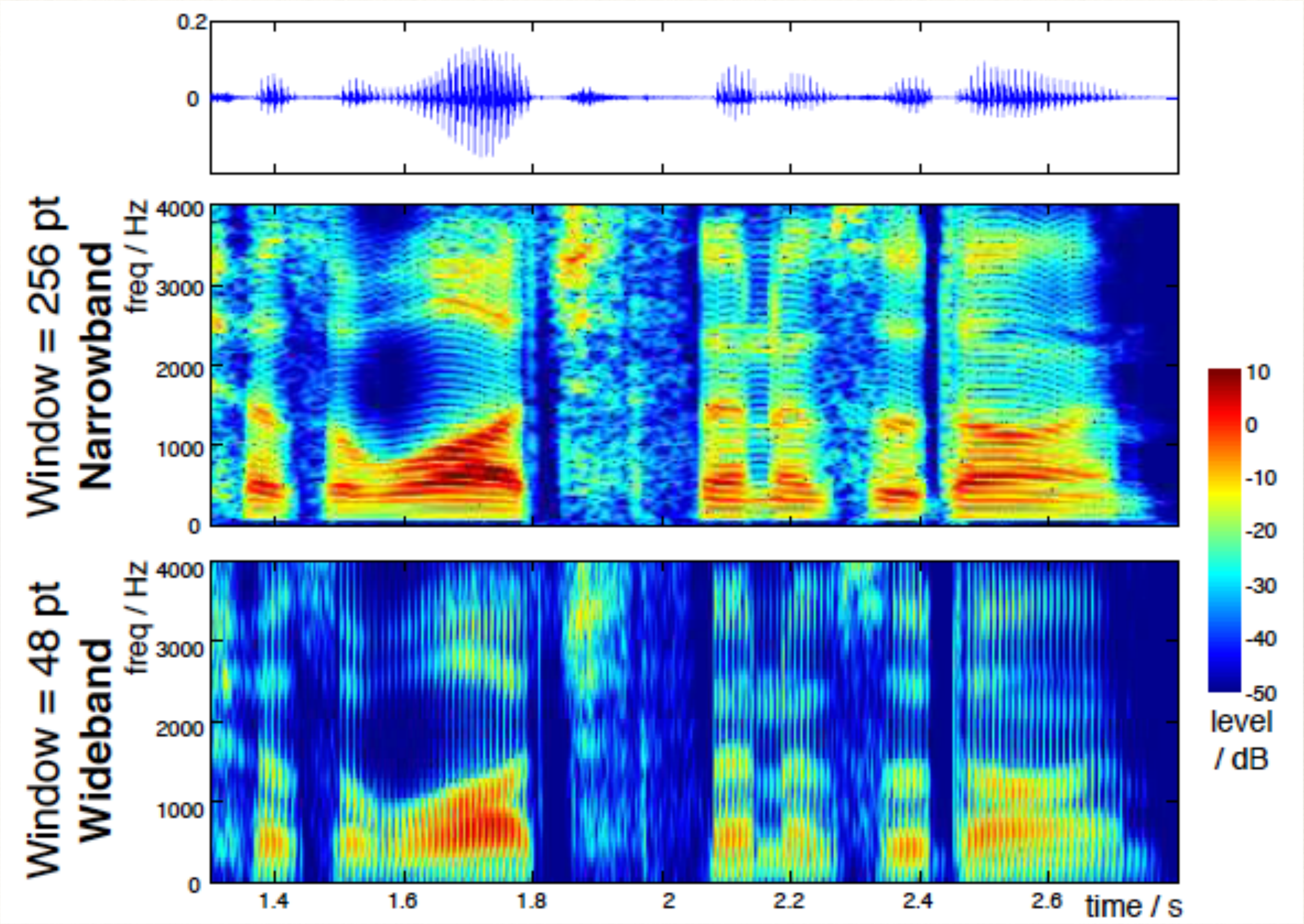
- Alternate tilings of time-freq:



Spectrogram of Real Sounds



Narrowband versus Wideband



Dan Ellis, "STFT Tutorial"

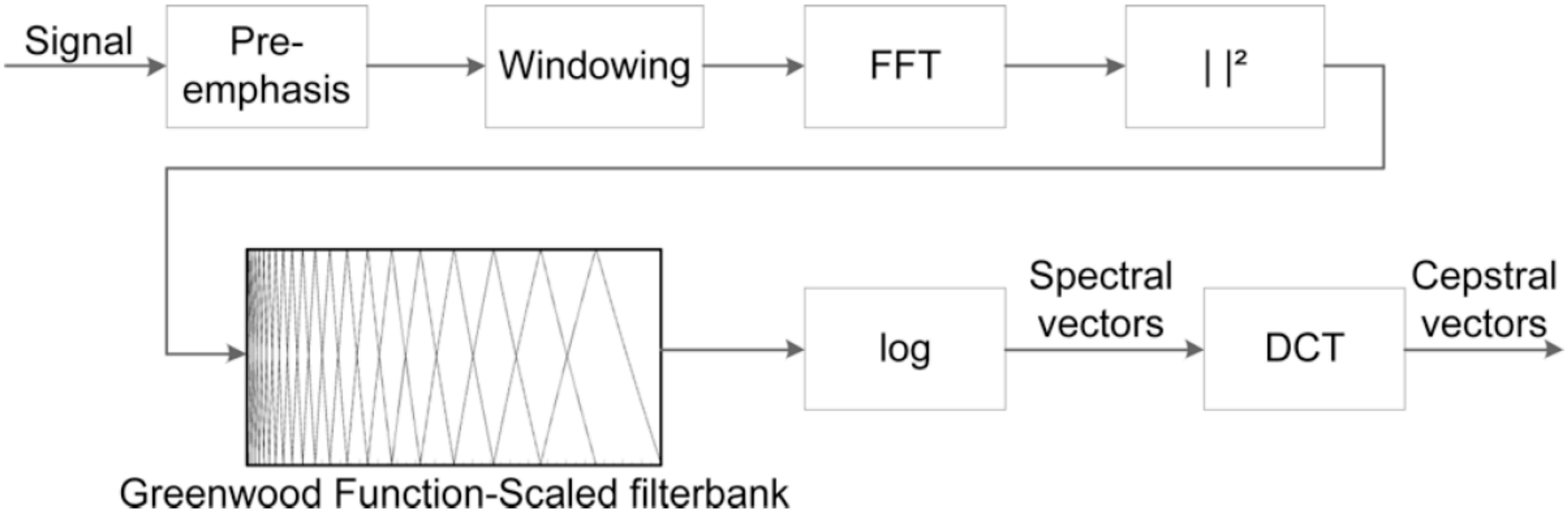
Mel Frequency Cepstral Coefficients

MFCC

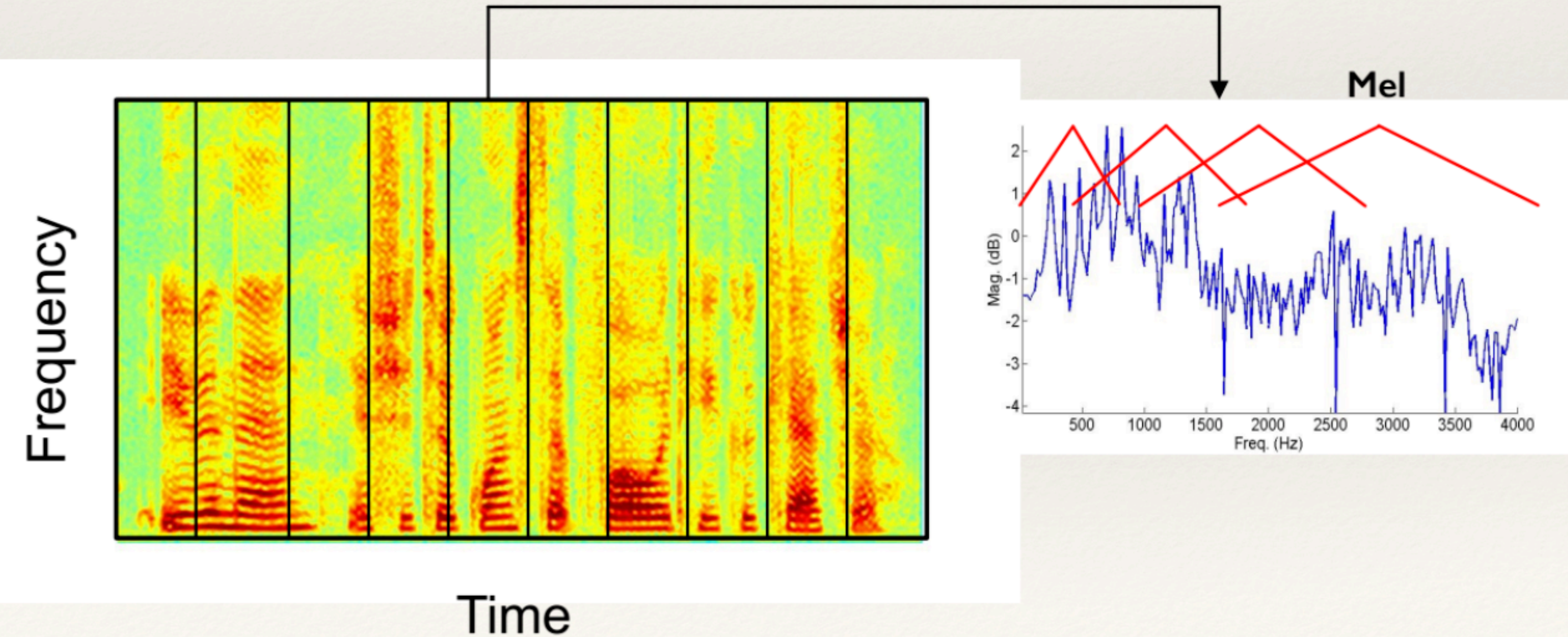
- MFCC coefficients model the spectral energy Distribution in a perceptually meaningful way
- Why do we need?
  - Automatic speech recogonition
  - Speaker Identification
  - Audio classification

Mel Frequency Cepstral Coefficients

- Implementation steps



Mel Frequency Cepstral Coefficients



Mel Frequency Cepstral Coefficients

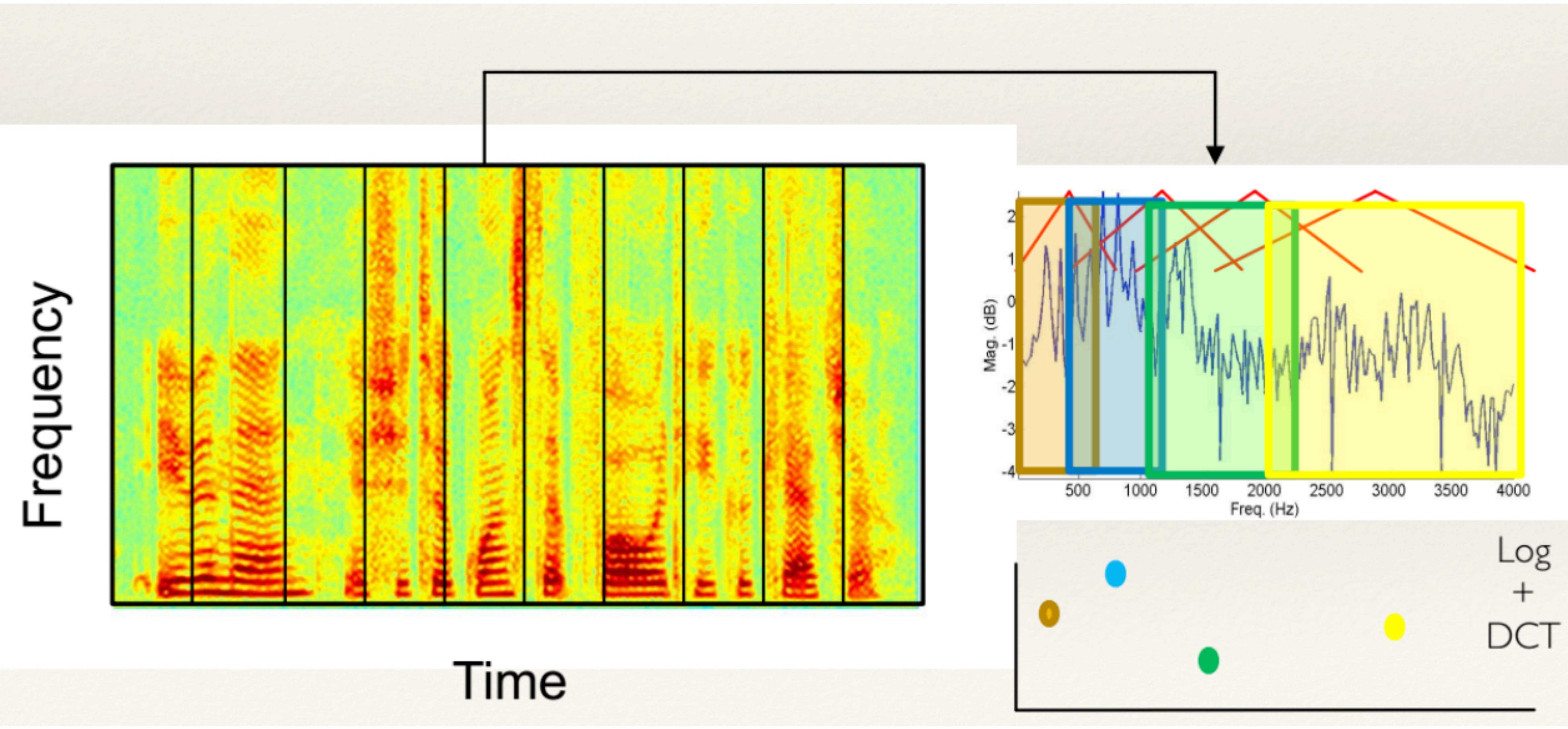


Image Processing

Image Capture and Representation

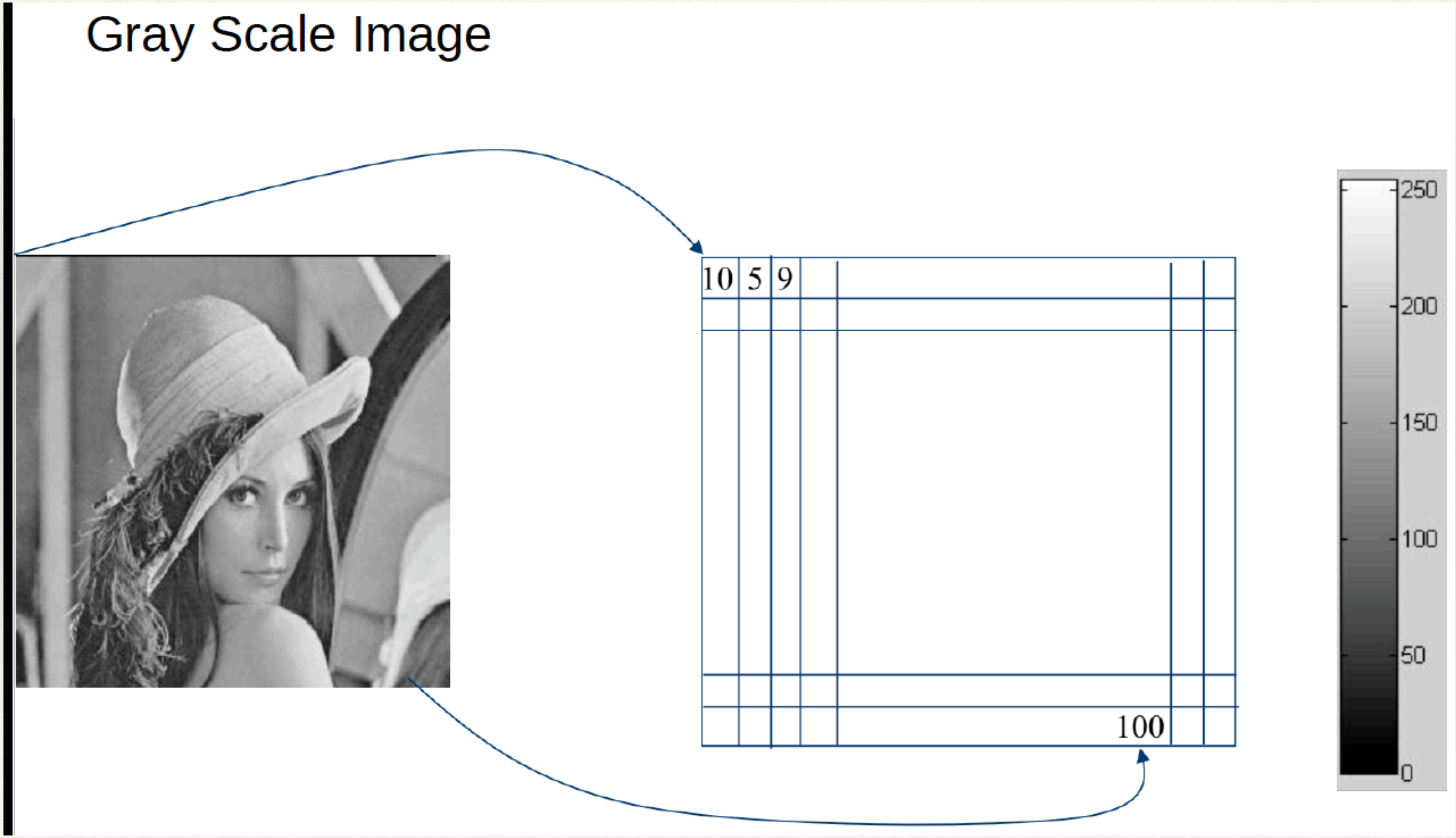
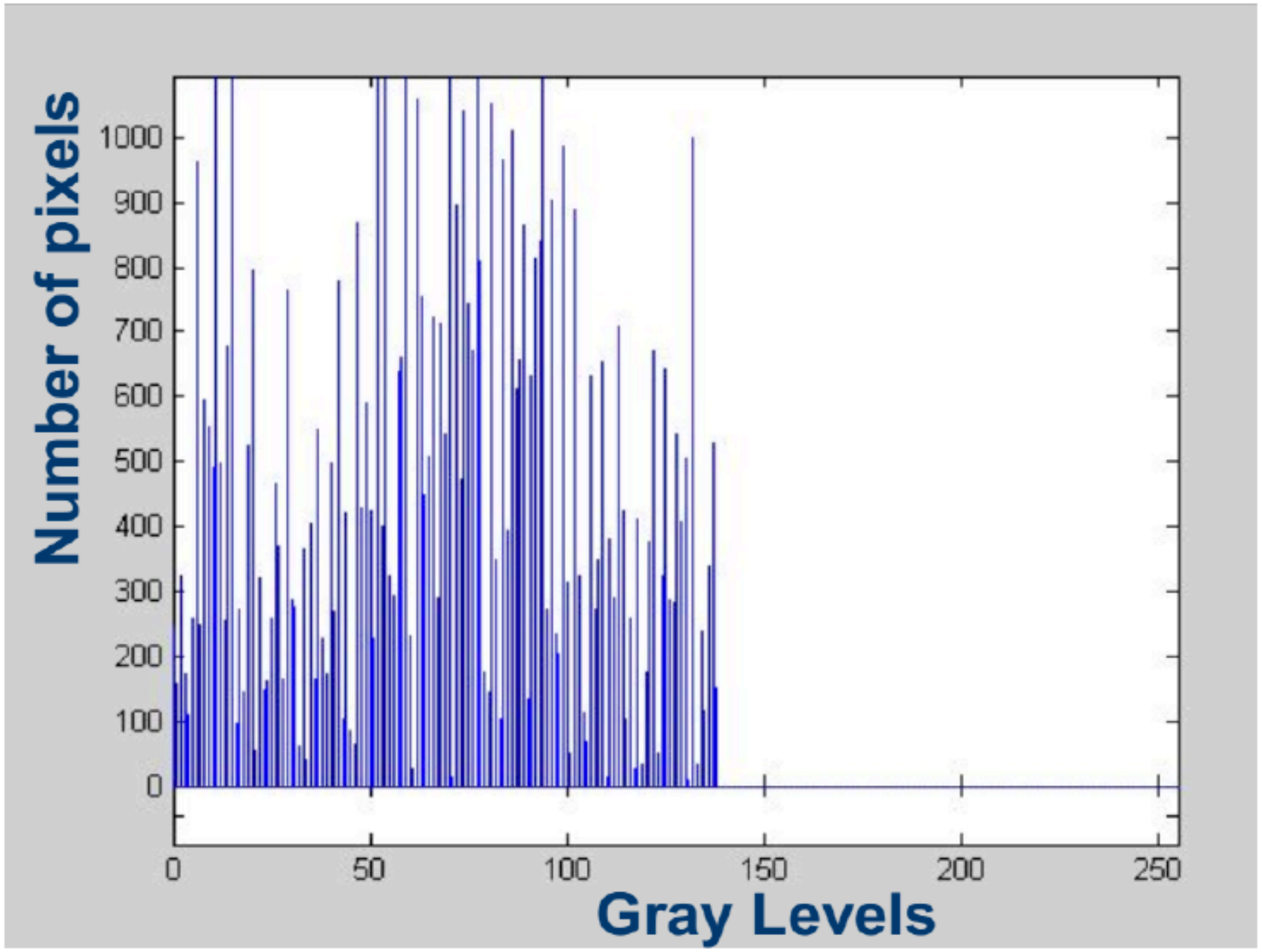


Image Capture and Representation

Histogram



- Histogram captures the distribution of gray levels in the image.
- How frequently each gray level occurs in the image

Image Filtering

- Image filtering: compute function of local neighborhood at each position

  Really important!
  - Enhance images
  - Denoise, resize, increase contrast, etc.
  - Extract information from images
  - Texture, edges, distinctive points, etc.
  - Detect patterns
  - Template matching

Image Filtering

Given function $f(x,y)$

Gradient vector $\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$

Gradient magnitude $|\nabla f(x,y)| = \sqrt{f_x^2 + f_y^2}$

Gradient direction $\theta = \tan^{-1} \frac{f_x}{f_y}$

$$\frac{df}{dx} = \lim_{\Delta x \to 0} \frac{f(x) - f(x - \Delta x)}{\Delta x} = f'(x)$$

$$\frac{df}{dx} = \frac{f(x) - f(x-1)}{1} = f'(x)$$

$$\frac{df}{dx} = f(x) - f(x-1) = f'(x)$$

Edge Detection Example

Convolution Operation in Images

Convolution

$$f * h = \sum_k \sum_l f(k,l)\,h(-k,-l)$$

f = Image

h = Kernel

h

h_1	h_2	h_3
h_4	h_5	h_6
h_7	h_8	h_9

$X - flip$

h_7	h_8	h_9
h_4	h_5	h_6
h_1	h_2	h_3

$Y - flip$

f

f_1	f_2	f_3
f_4	f_5	f_6
f_7	f_8	f_9

*

h_9	h_8	h_7
h_6	h_5	h_4
h_3	h_2	h_1

$$f * h = f_1 h_9 + f_2 h_8 + f_3 h_7 + f_4 h_6 + f_5 h_5 + f_6 h_4 + f_7 h_3 + f_8 h_2 + f_9 h_1$$