# Low level Audio Features

by Nimesh Gautam

# Dimensions of Types of Audio Features

- <u>Level of abstraction</u>
- Temporal Scope
- Musical aspect
- <u>Signal Domain</u>
- ML approach

# LEVEL OF ABSTRACTION

1. HIGH LEVEL FEATURES

   CHORDS, RHYTHM, MELODY, GENRE, ETC

2. MEDIUM LEVEL FEATURES

   PITCH, NOTE ONSETS, MFCC, ETC

3. <u>LOW LEVEL FEATURES</u>

Low Level Audio Features

# LOW LEVEL AUDIO  FEATURES

- **Typically Computed directly from raw audio waveform**

- **Simple statistical summaries of waveform**

Low Level Audio Features

# LOW LEVEL AUDIO FEATURES

1. Amplitude Envelope (TD)

2. RMS Energy (TD)

3. Zero- Crossing Rate  (TD)          TD = Time Domain

4. Band energy ratio  (FD)           FD = Frequency Domain

5. Spectral Centroid (FD)

6. Spectral Spread/ Bandwidth (FD)

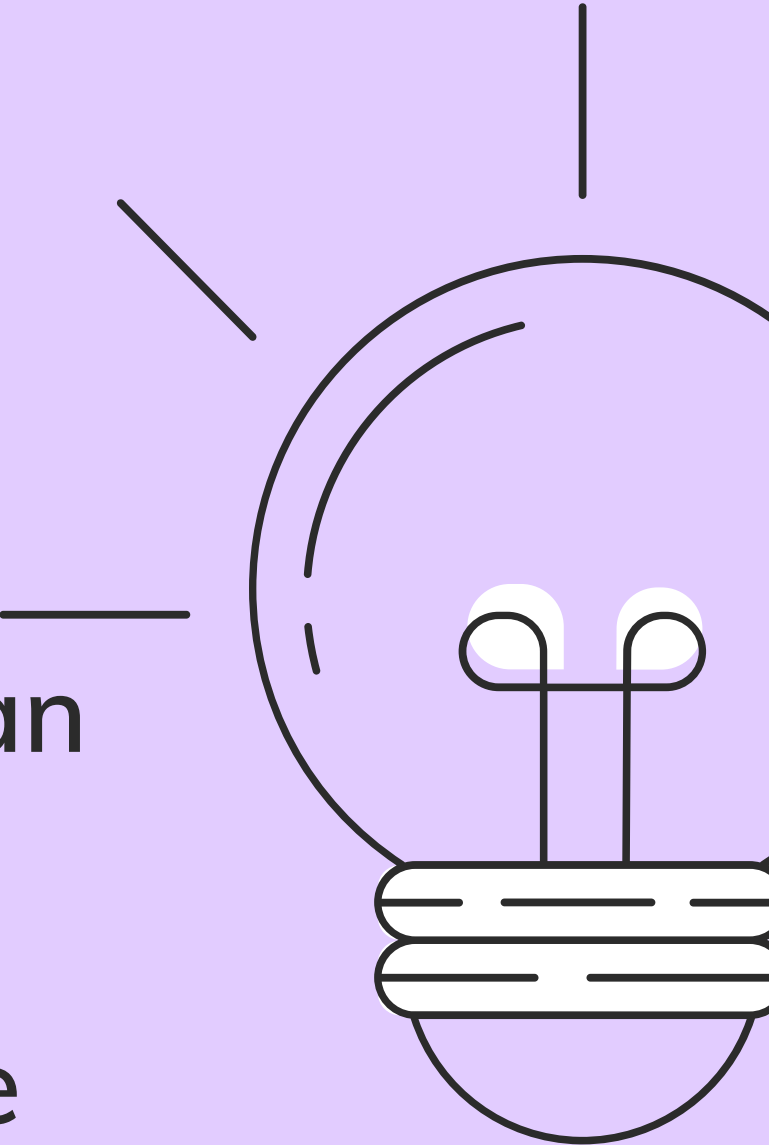7. Spectral flux, etc. (FD)

Low Level Audio Features

# Amplitude Envelope

INTRODUCTION:

- **Envelope is a smooth curve outlining extremes of an oscillating signal**

- **Maximum amplitude among all samples in a frame**



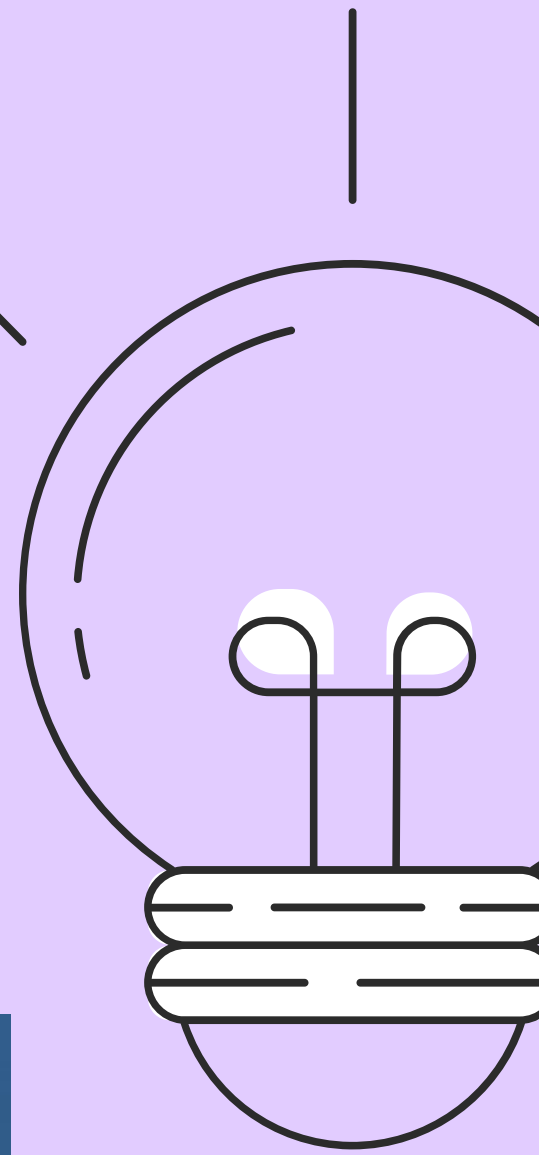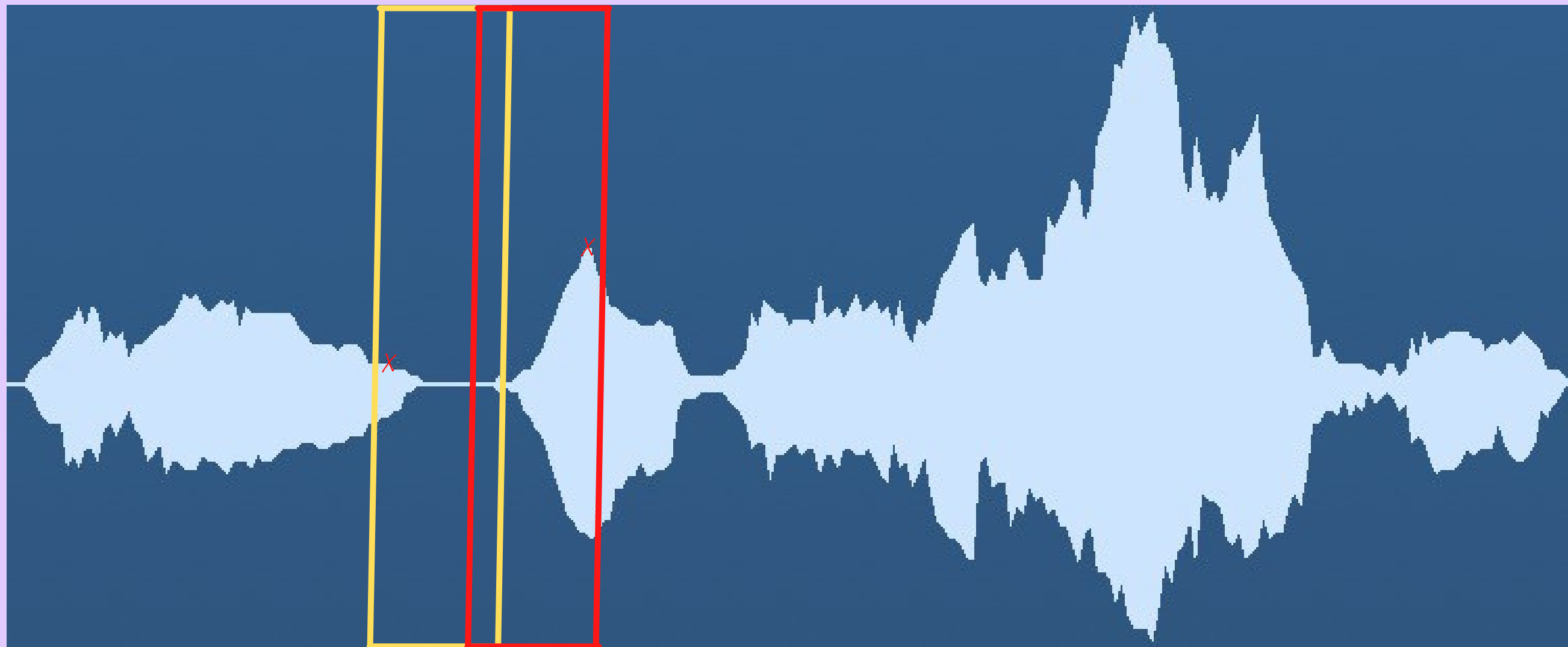Source: Wikipedia

Low Level Audio Features
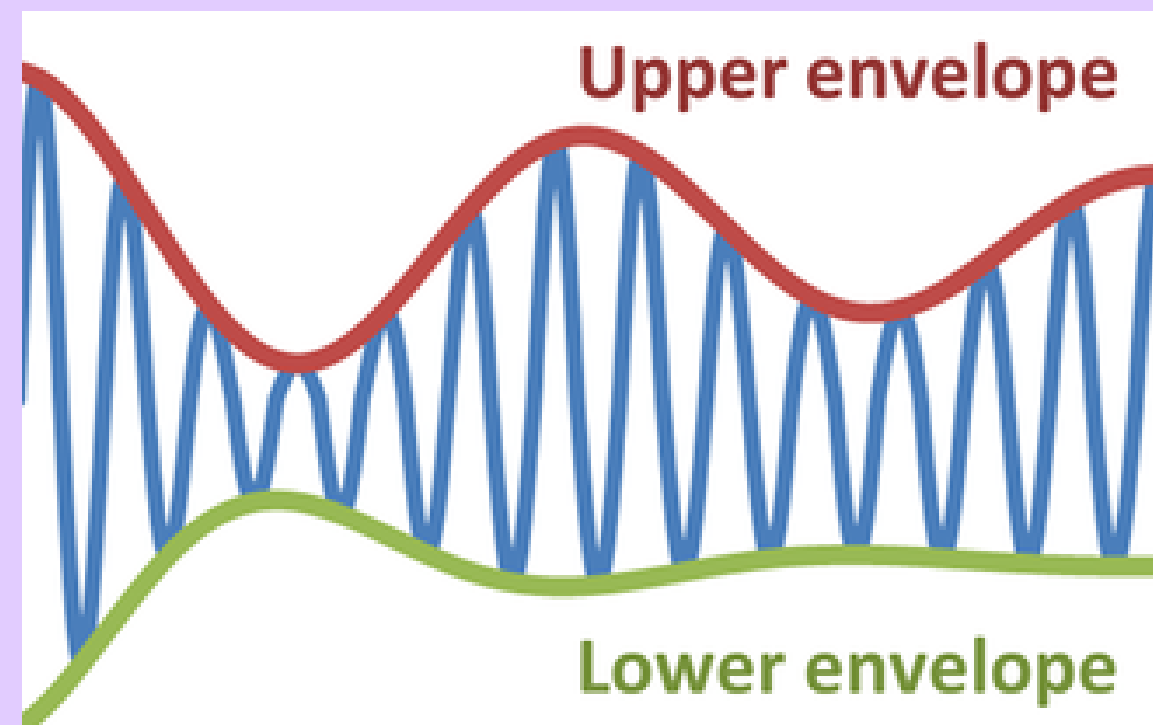
# Amplitude Envelope

COMPUTATION

$$AE(t) = \max_{k=t.K}^{(t+1).K-1} s(k)$$
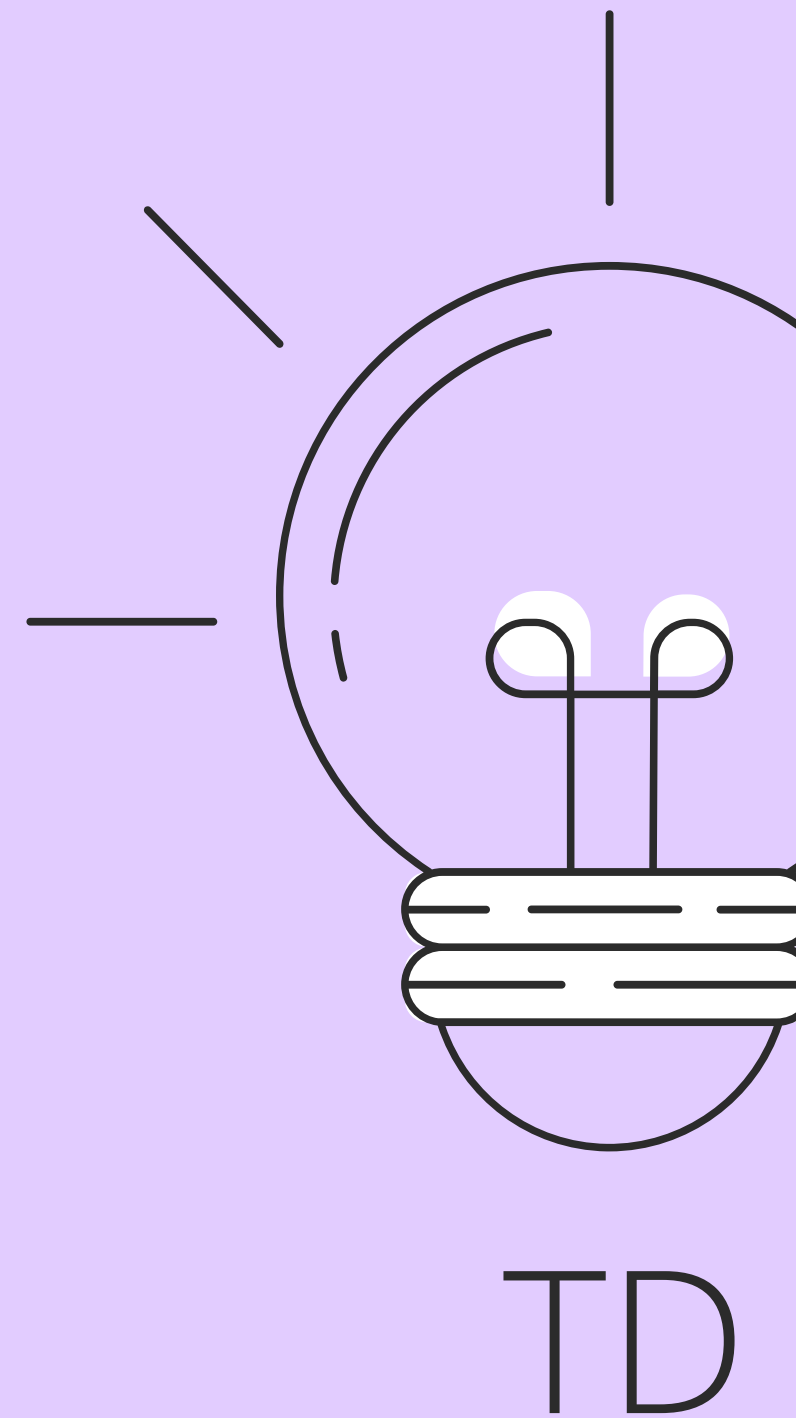
K = Frame size

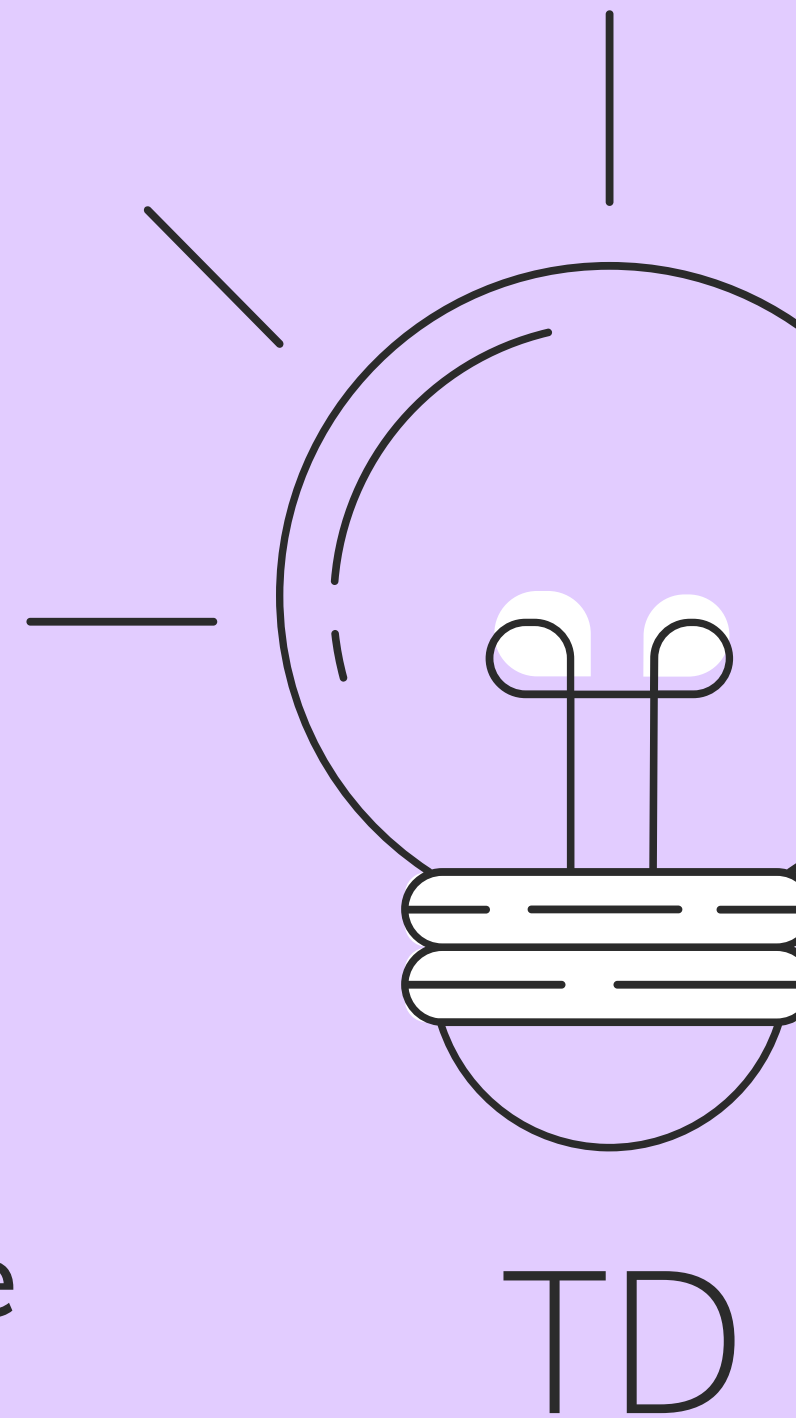s(k) = Amplitude of kth sample

# Amplitude Envelope

USES:

- Onset Detection

- Music Genre classification

- Gives rough idea about loudness

TD

# RMS Energy/ Power/ Level

INTRODUCTION:

- Root Mean Square Energy

- Relates to perceived sound intensity

- Indicator of loudness

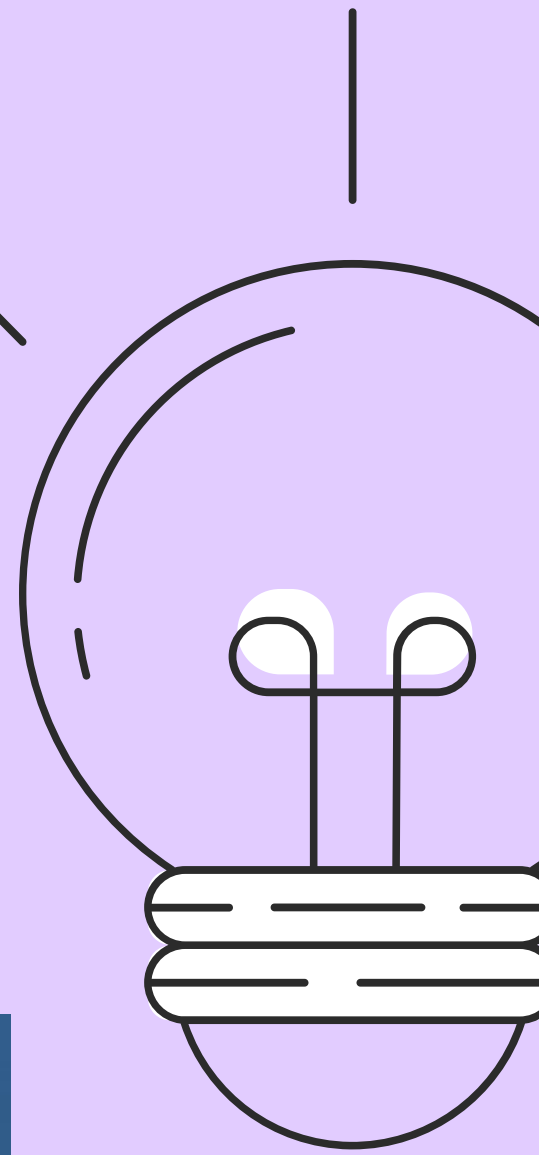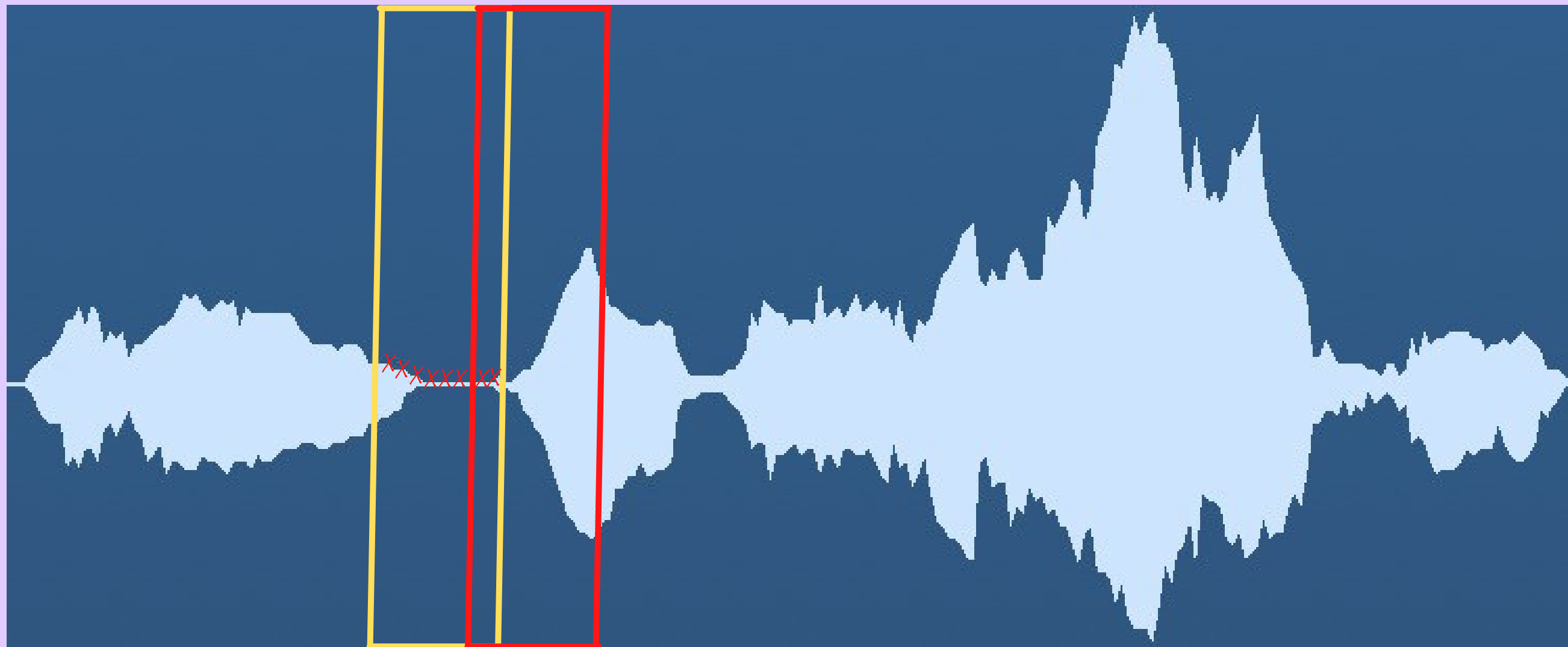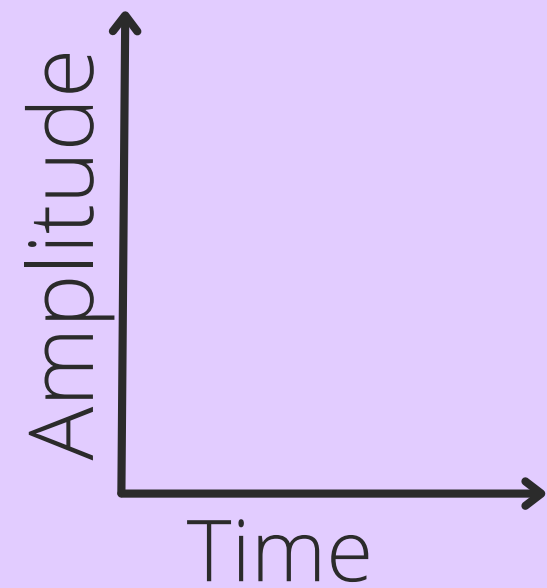- Less sensitive to outliers than Amplitude Envelope

TD

# RMS Energy

COMPUTATION

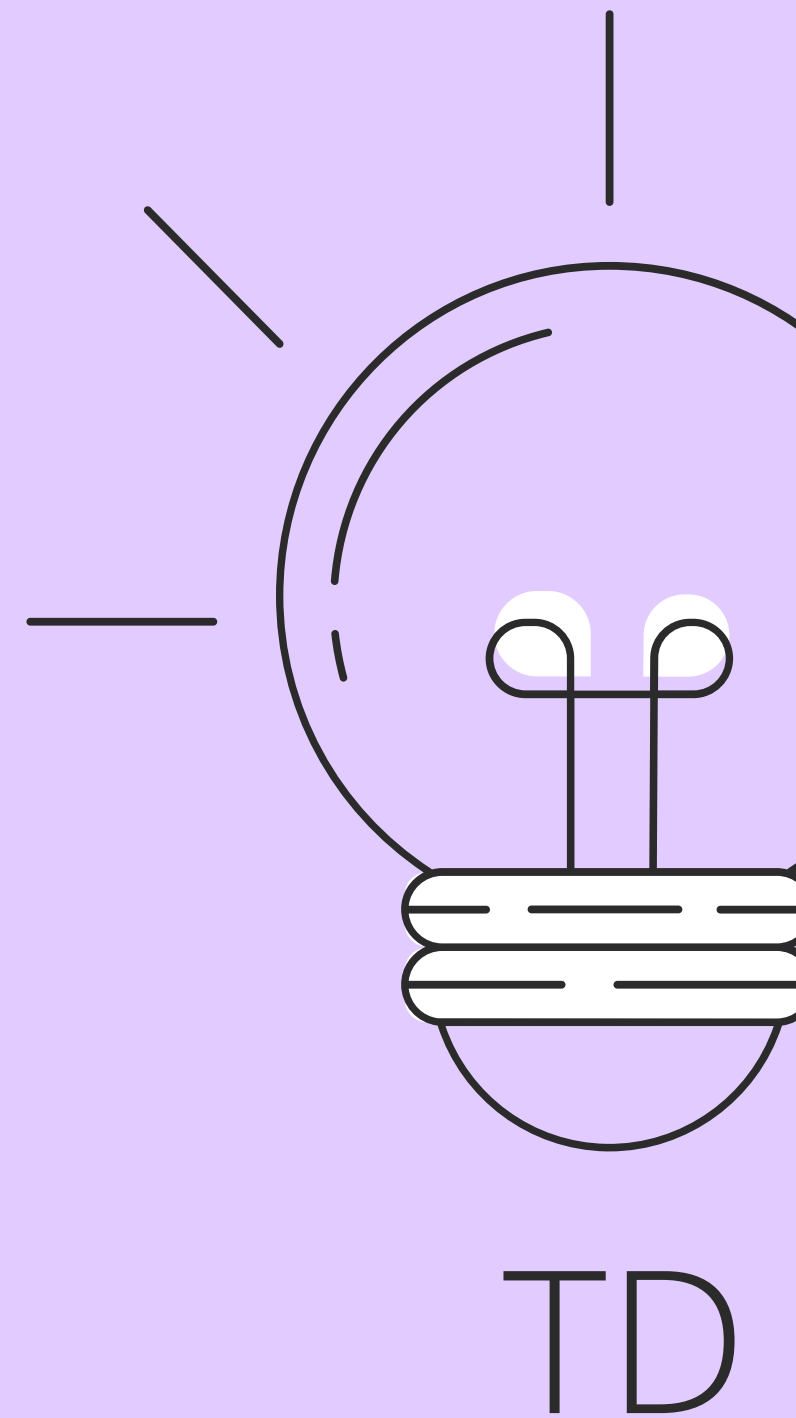$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} s(k)^2}$$

K = Frame size
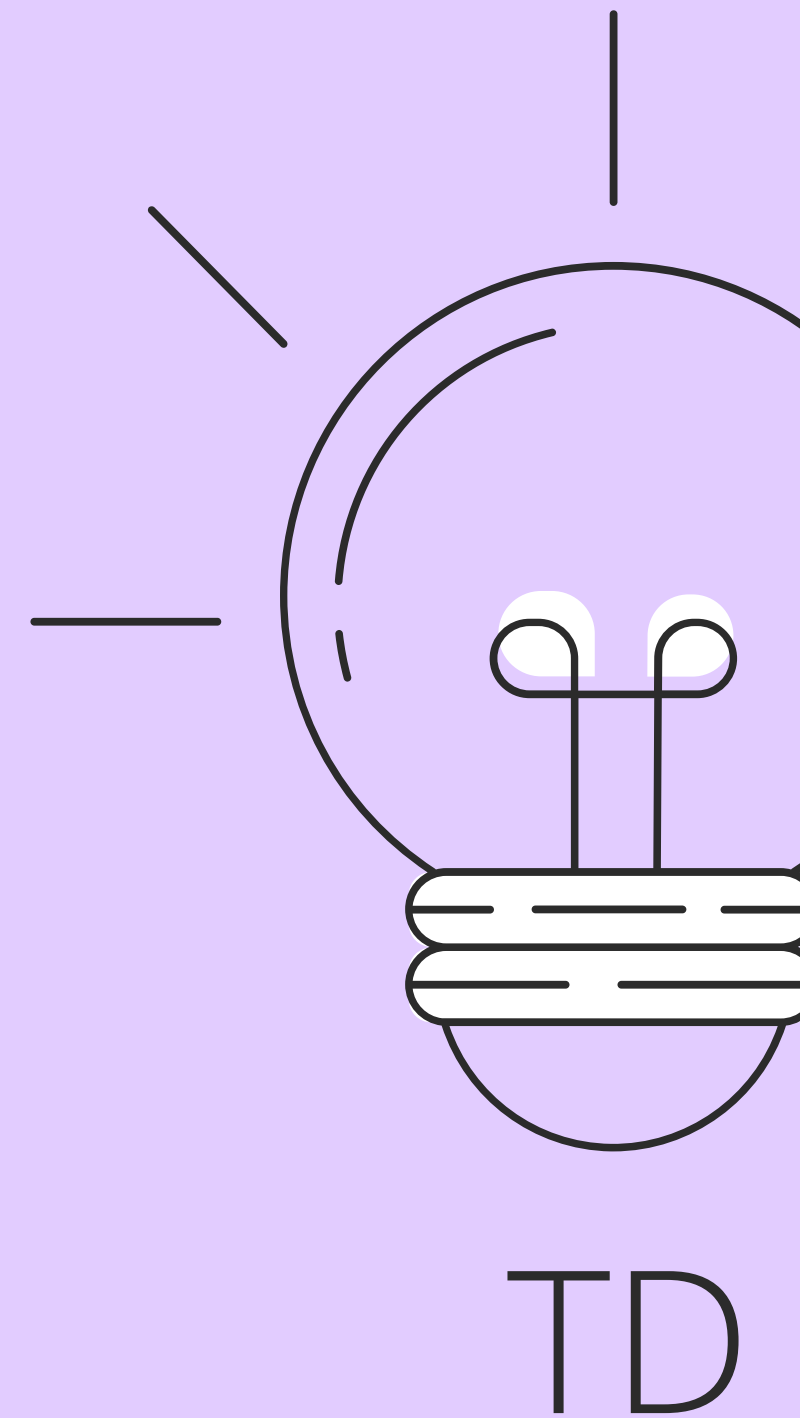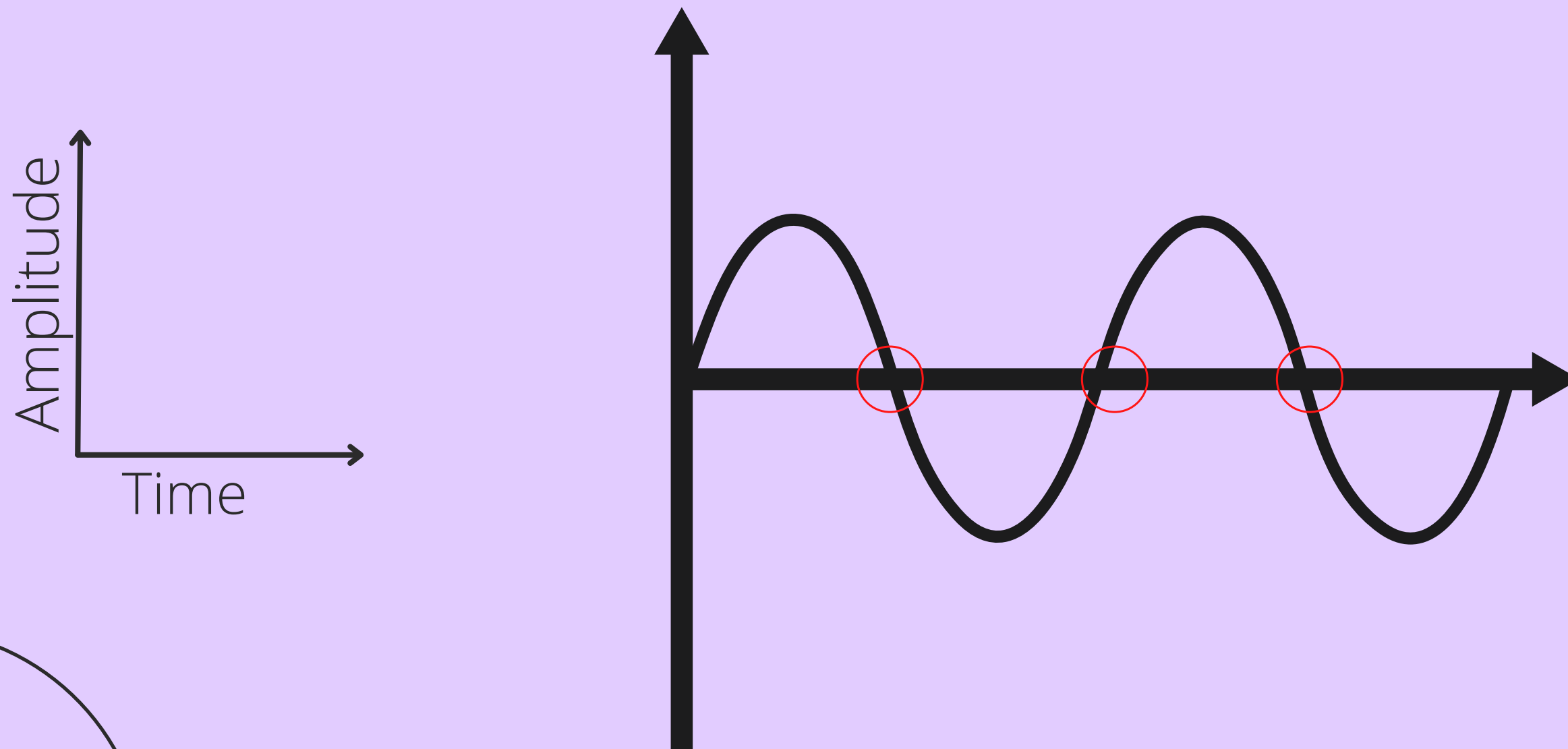s(k) = Amplitude of kth sample

# RMS Energy

USES:

- **Indicator of loudness**

- **Audio segmentation**

- **Music Genre classification**

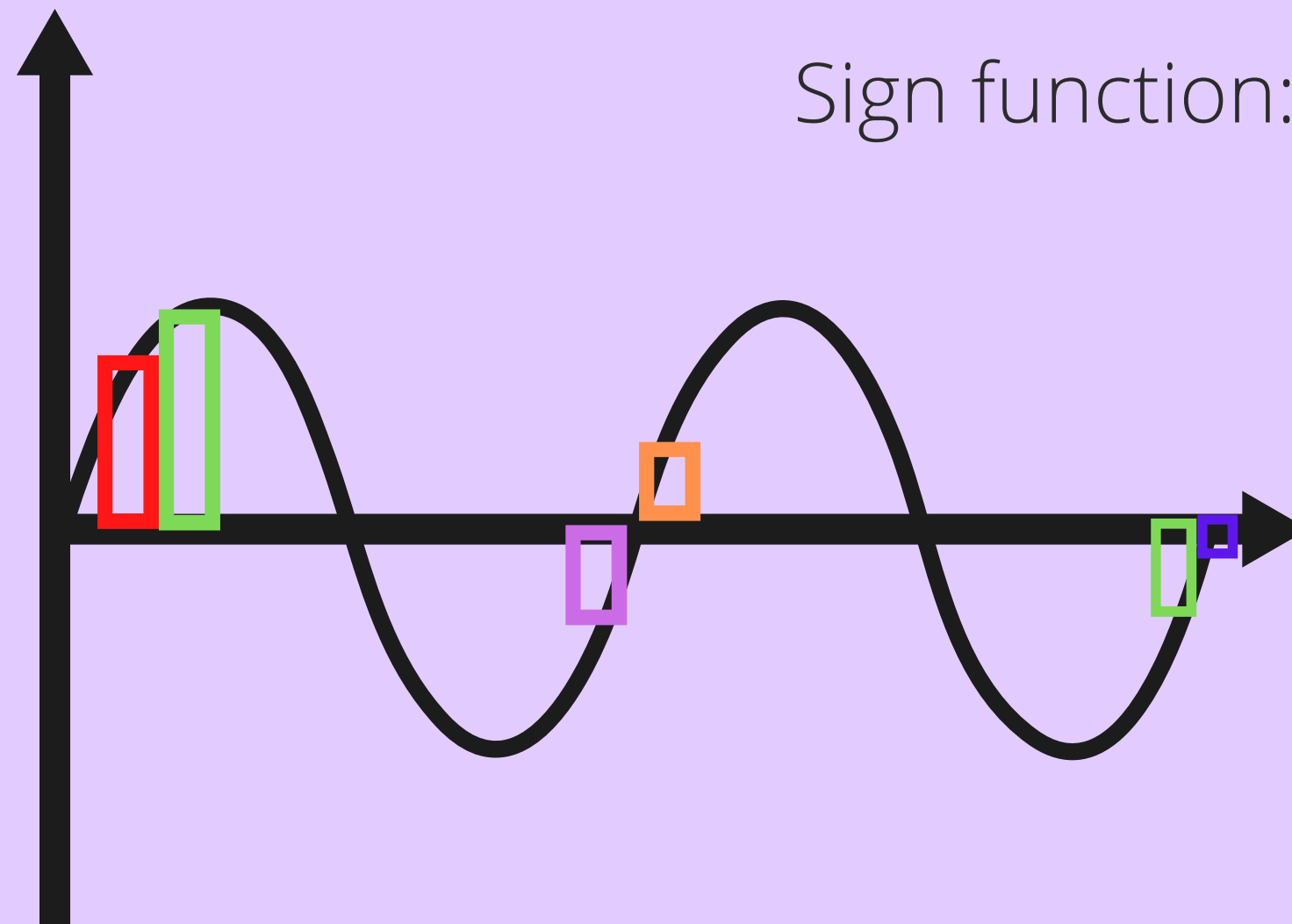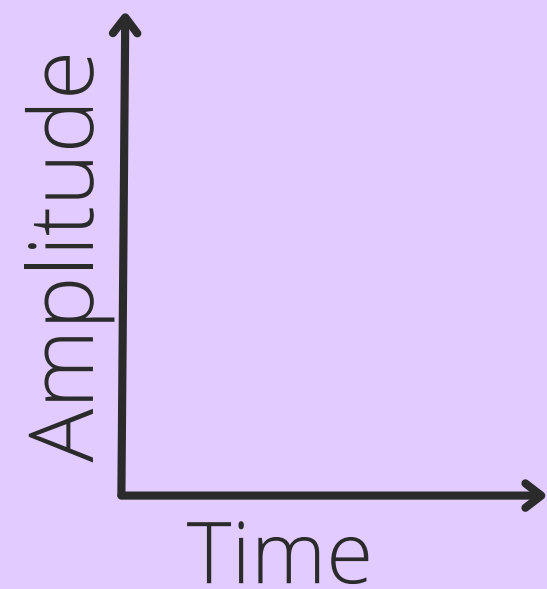TD

# Zero Crossing Rate

INTRODUCTION:

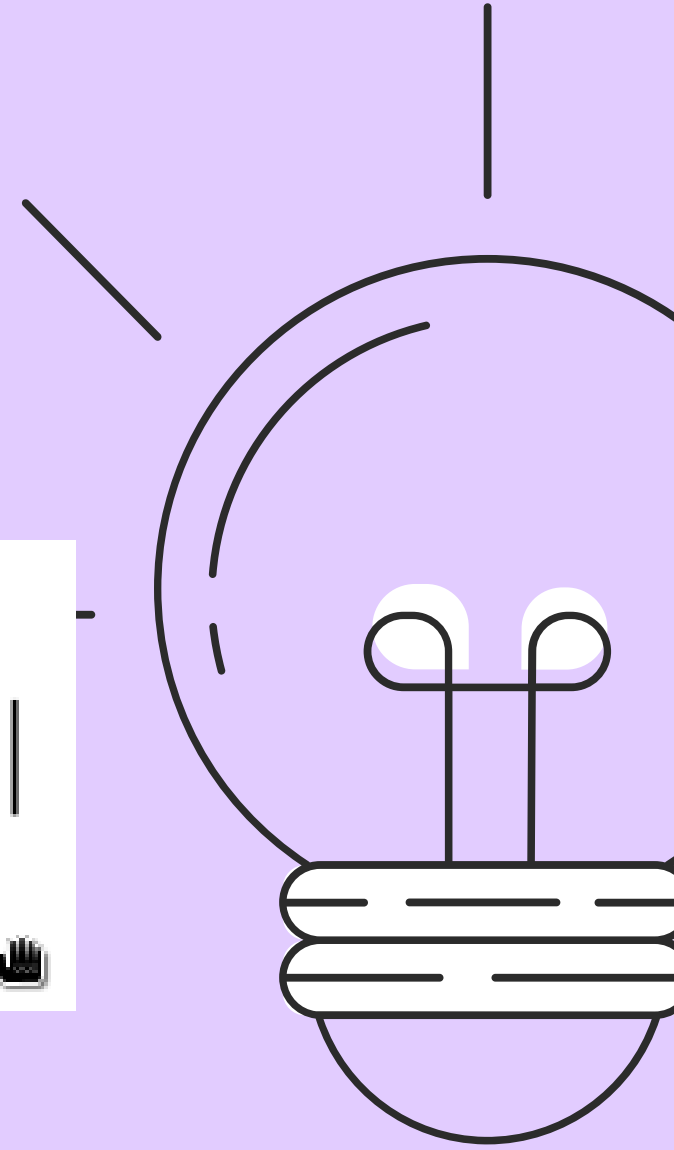- **No of times amplitude value changes its sign.**

TD

# Zero Crossing Rate

COMPUTATION

$$ZCR_t = \frac{1}{2} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} |\operatorname{sgn}(s(k)) - \operatorname{sgn}(s(k+1))|$$

Amplitude

Time

Sign function:    s(k) > 0        +1
                  s(k) < 0        -1
                  s(k) = 0         0

Low Level Audio Features
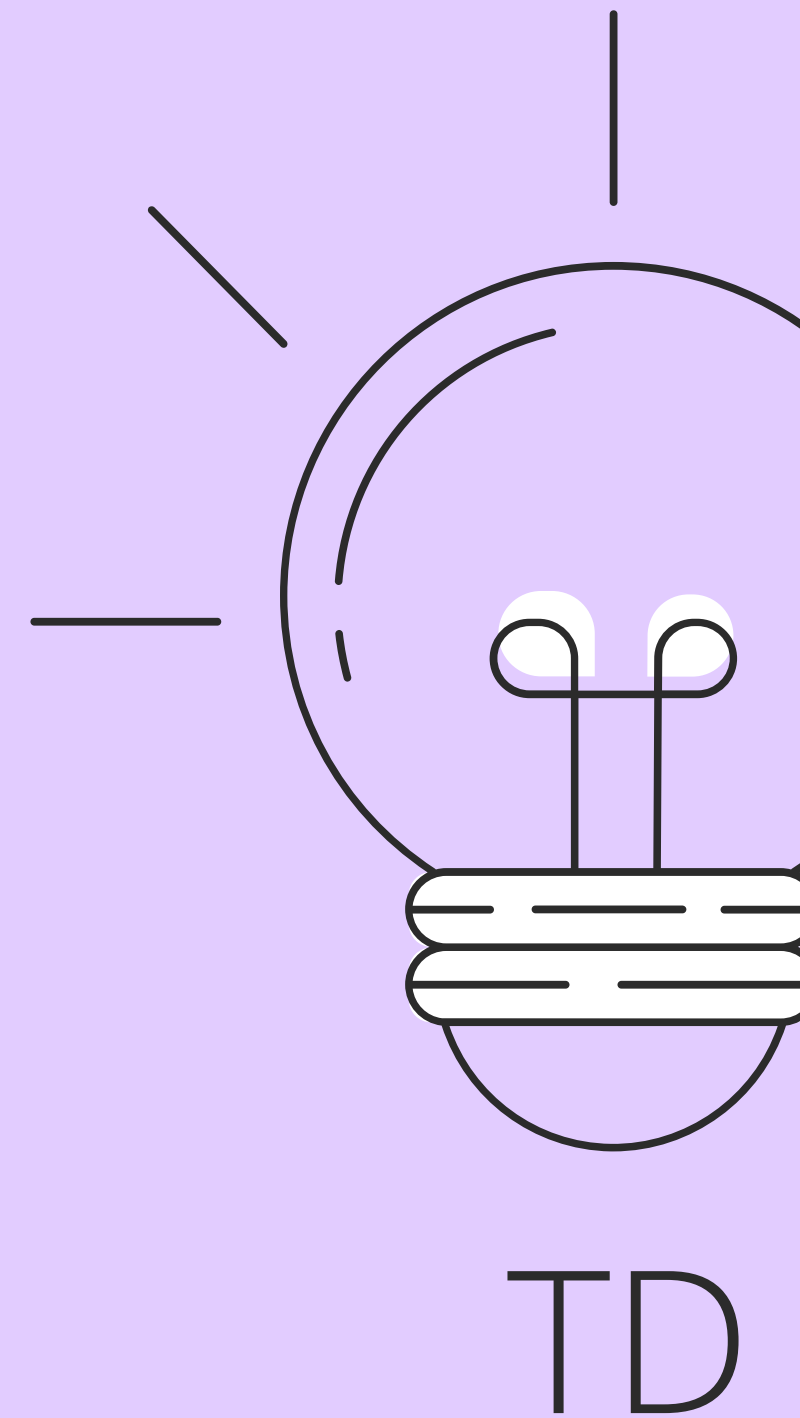
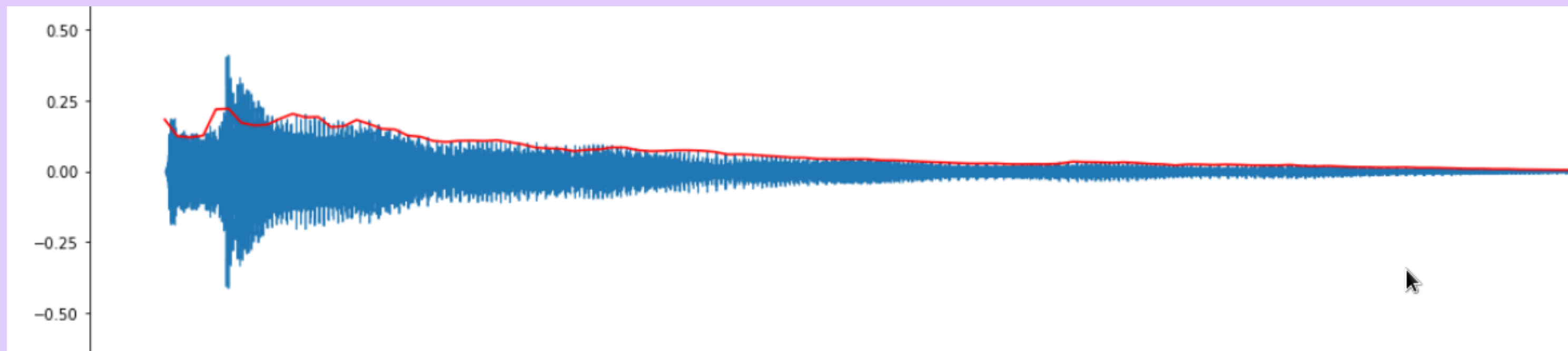# Zero Crossing Rate

USES:

- **Recognition of percussive vs pitches sounds**

- **Monophonic pitch estimation**

- **Speech Recognition**
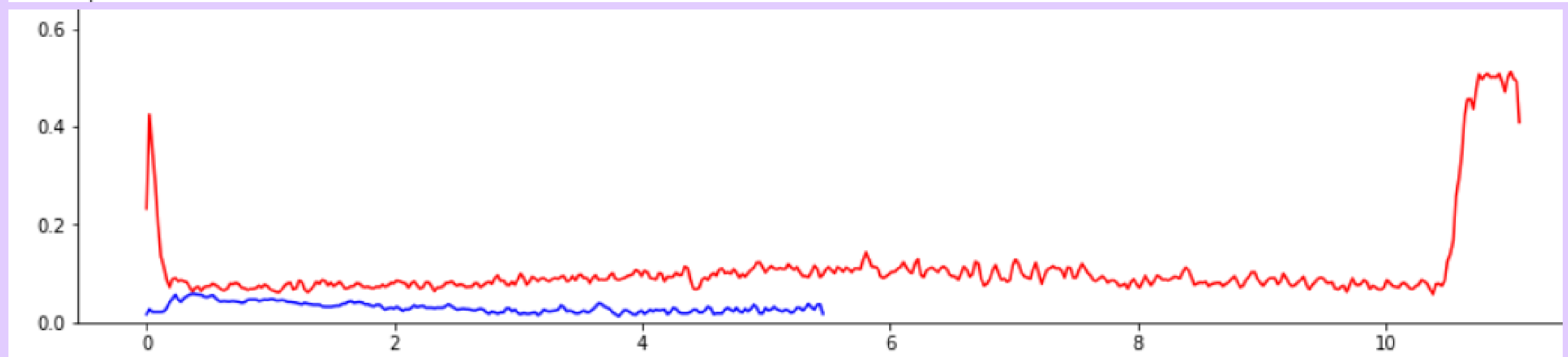
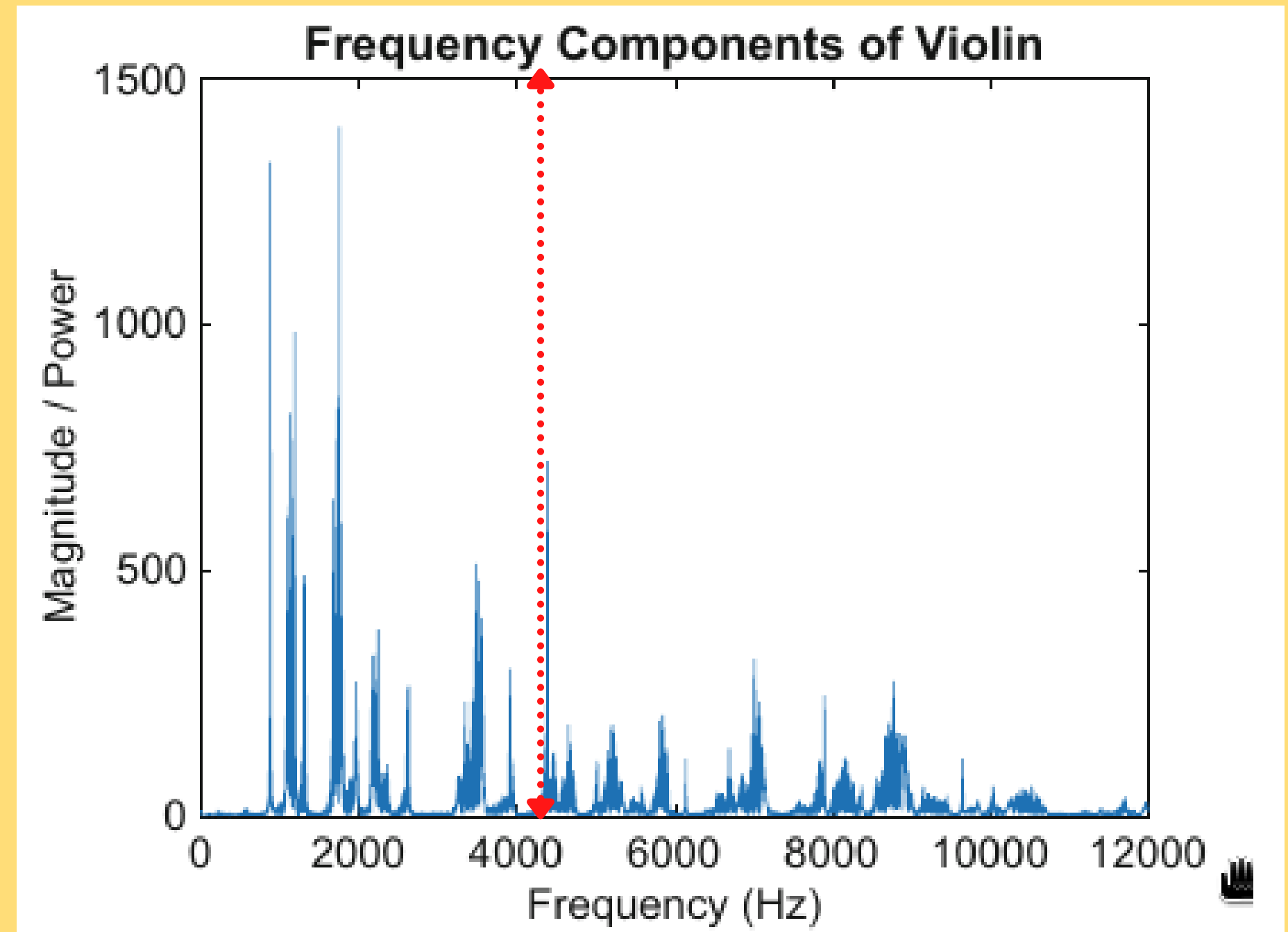Low Level Audio Features

Amplitude Envelope

RMS Energy
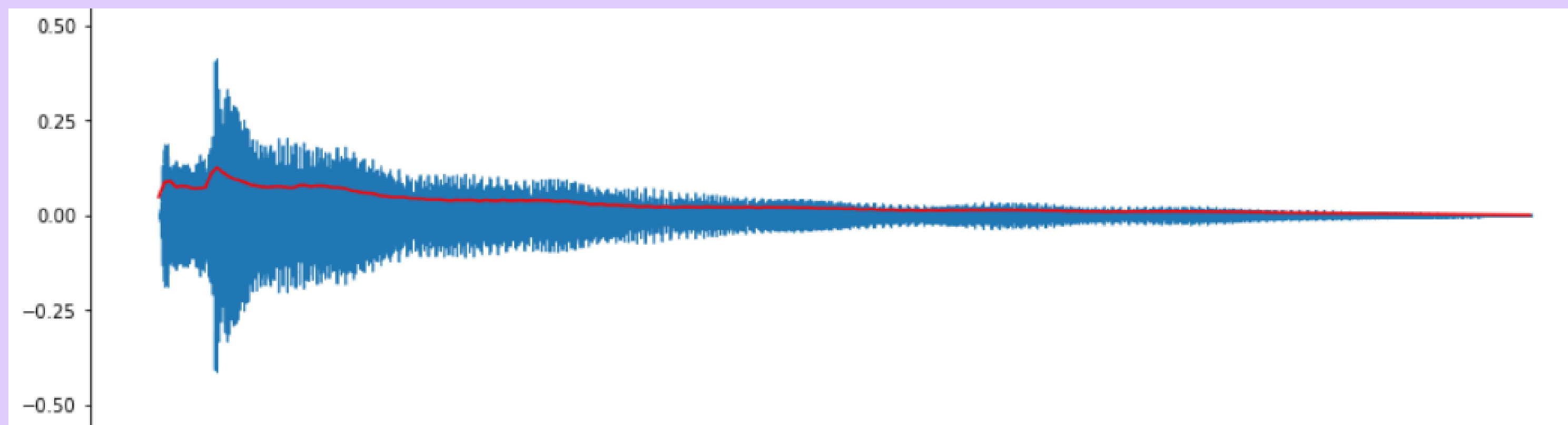
ZCR

# Frequency Domain Low level Features

Time Domain Plot of Digital Signal

Fast Fourier Transform
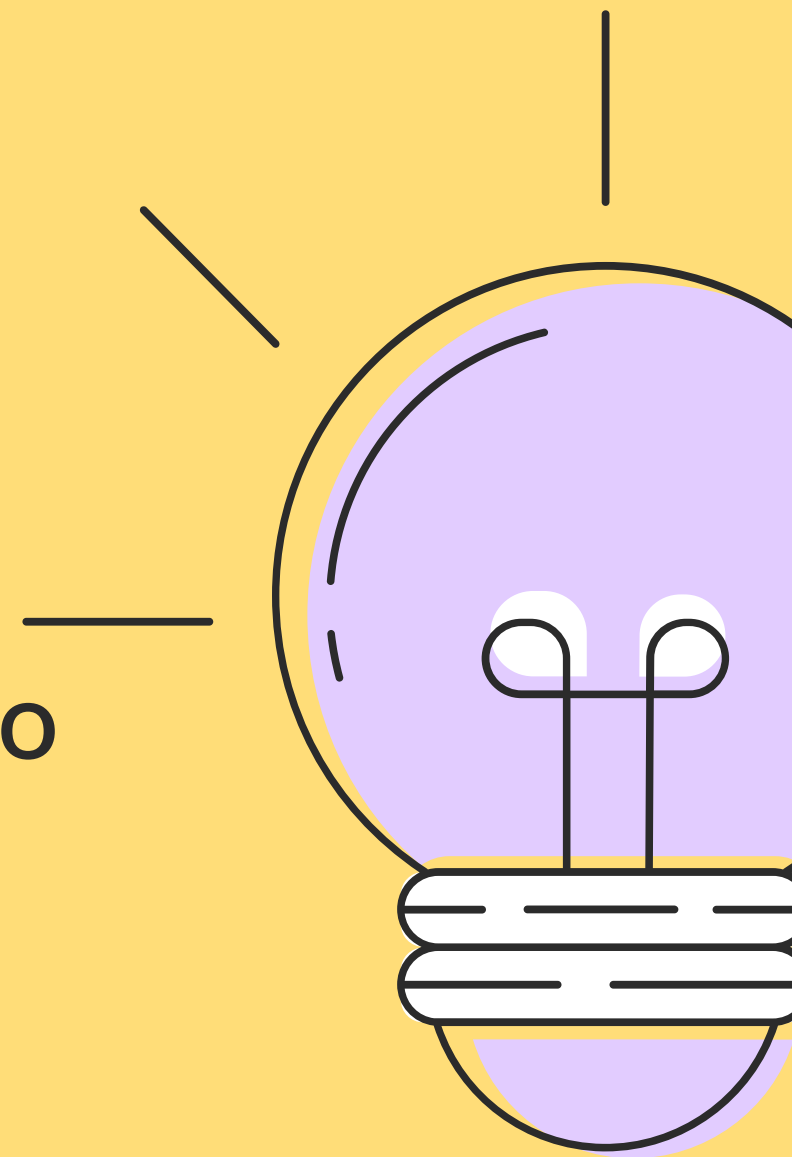


source(book):  Music SImilarity and Retreival, Peter Knees, Markus Schedl

# Band Energy Ratio

- Relates the energy in the lower frequency bands to the energy in higher bands.
- Measure of how dominant low frequency are
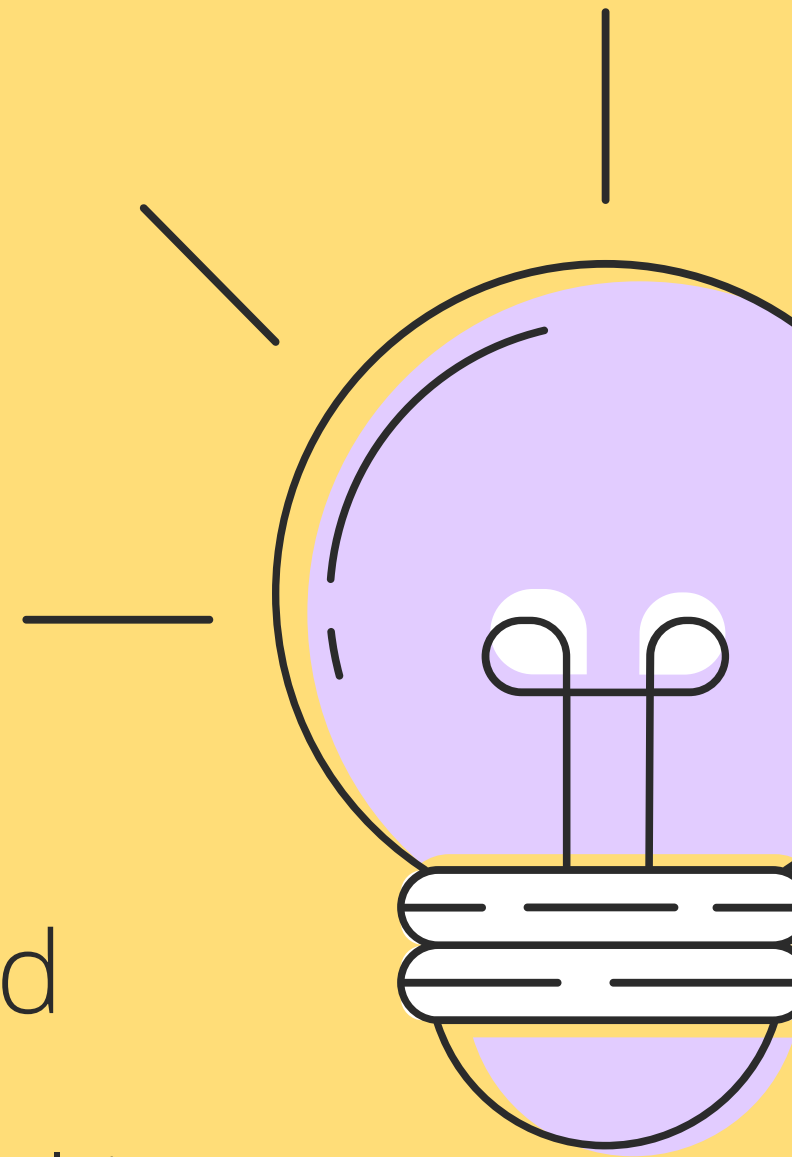
FD

Low Level Audio Features

# Band Energy Ratio

COMPUTATION

$$BER_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^{N} m_t(n)^2}$$
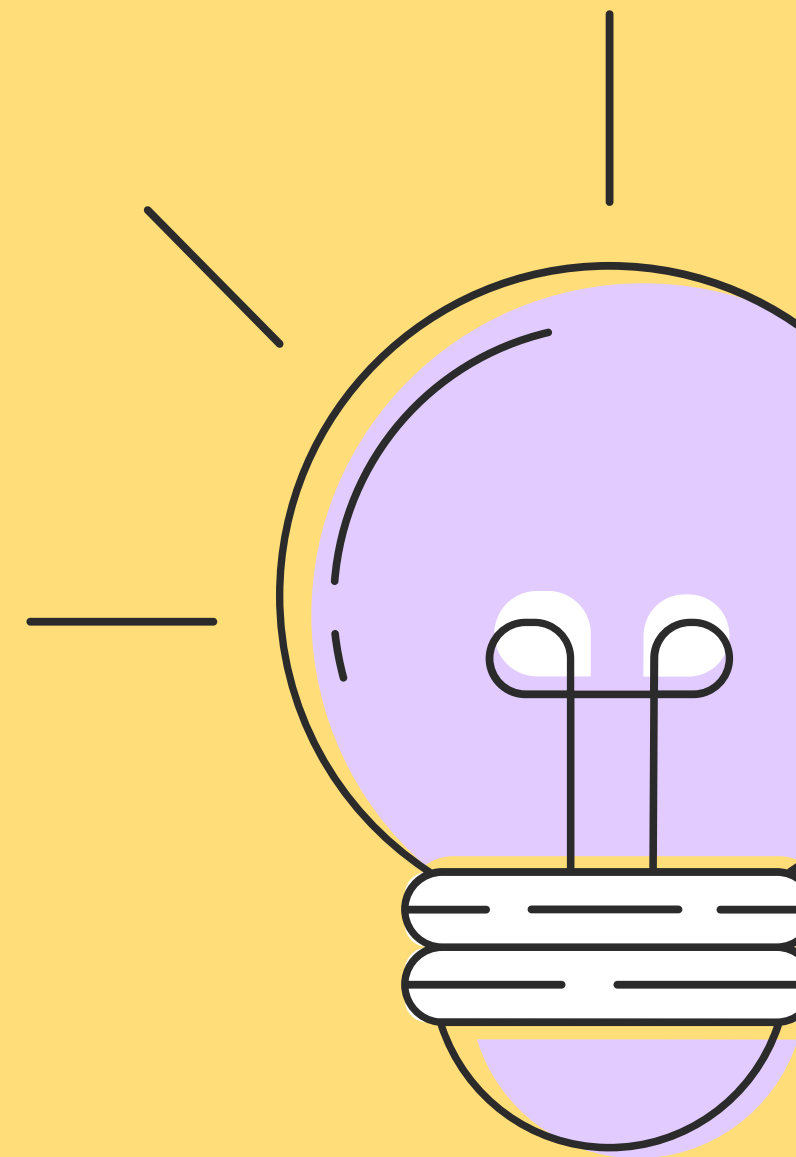
F = split frequency band

mt(n) = Magnitude of signal in frequency domain in frequency band n

# Band Energy Ratio

USES:

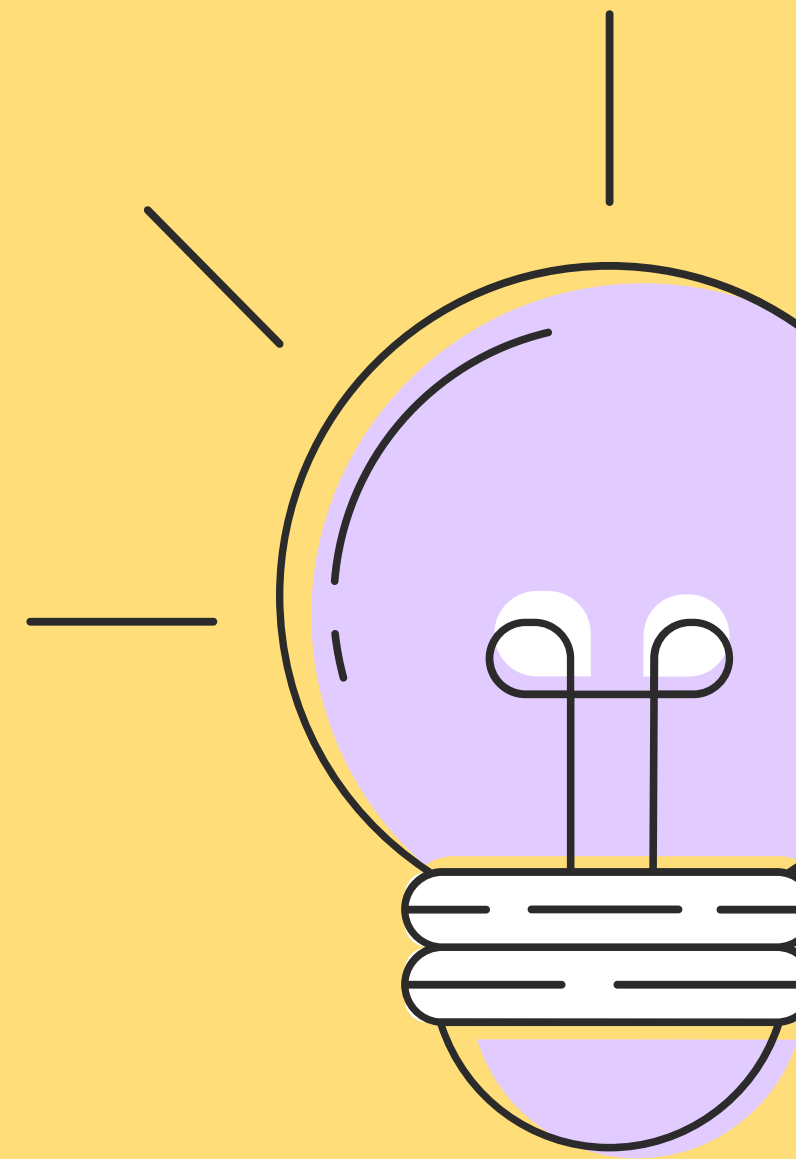- **Speech/ Music discrimination**
- **Music Genre classification**

TD

# Spectral Centroid

- Center of gravity of the magnitude spectrum i.e.,

- Frequency band where most of the energy is concentrated
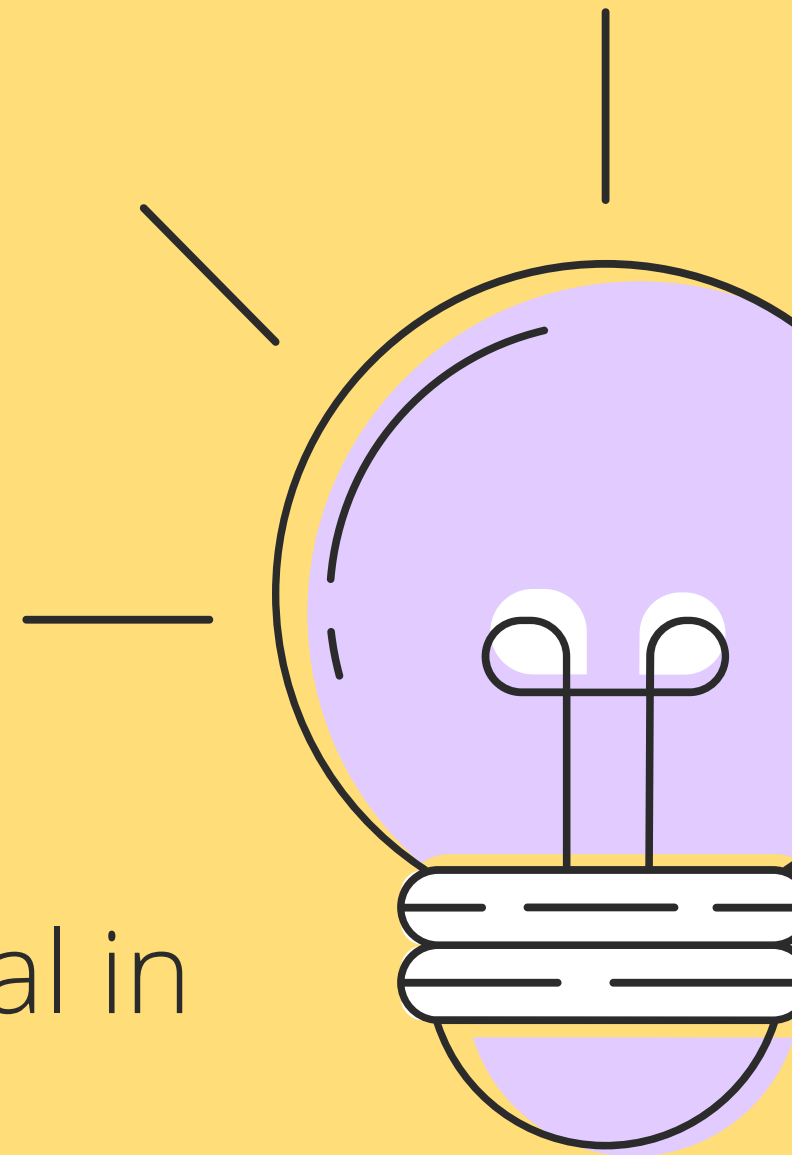
- Measure of brightness of sound

FD

Low Level Audio Features

# Spectral Centroid

COMPUTATION

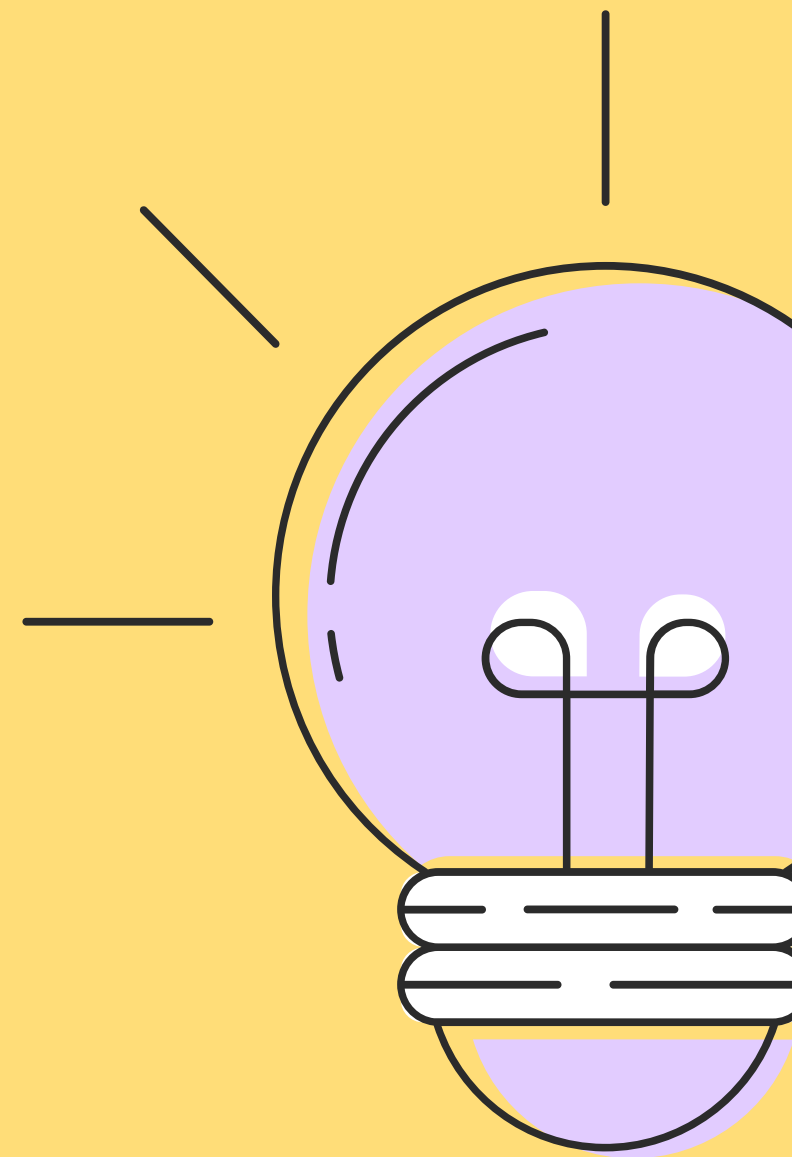$$SC_t = \frac{\displaystyle\sum_{n=1}^{N} m_t(n) \cdot n}{\displaystyle\sum_{n=1}^{N} m_t(n)}$$

mt(n) = Magnitude of signal in frequency domain in frequency band n

# Spectral Centroid
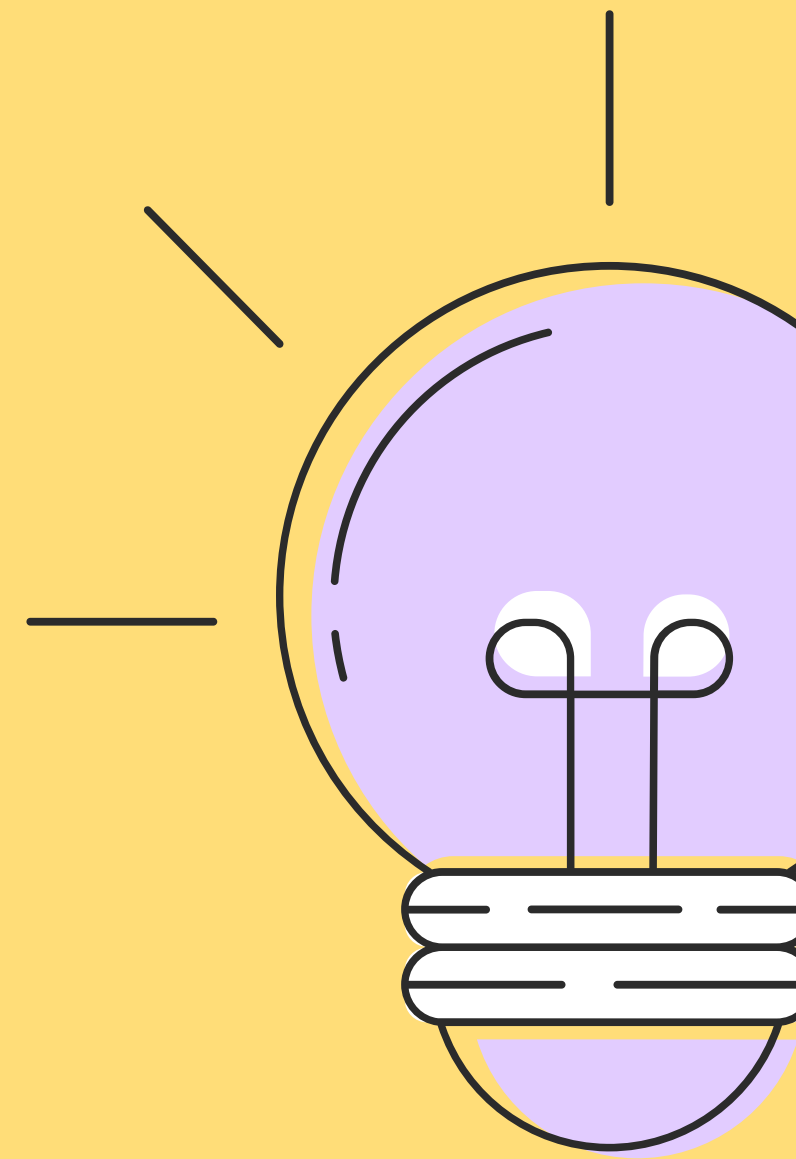
USES:

- **Audio Classification**

- **Music Classification**

TD

# Spectral Flux

INTRODUCTION:

- **Describes change in power spectrum between consecutive frames**
- **Used in onset detection, speech detector**
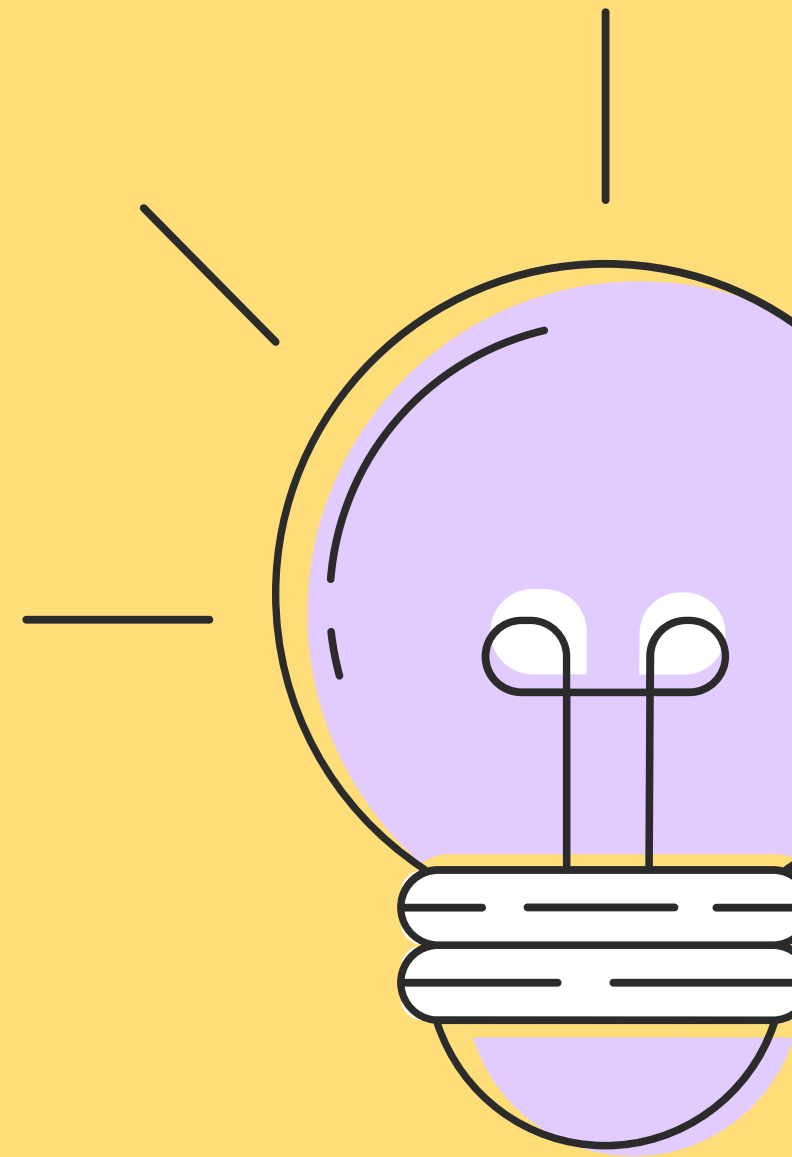- **Describing timbre of an audio signal**

FD

# Spectral Flux

$$SF_t = \sum_{n=1}^{N} (D_t(n) - D_{t-1}(n))^2$$

Dt = frame by frame normalized frequency distribution in frame t

Low Level Audio Features

# Thank You

by Nimesh Gautam