Time-domain audio features

Valerio Velardo

Join the community!

thesoundofai.slack.com

Time-domain features

- Amplitude envelope (AE)
- Root-mean-square energy (RMS)
- Zero-crossing rate (ZCR)
- ...

Amplitude envelope

- Max amplitude value of all samples in a frame

Amplitude envelope

- Max amplitude value of all samples in a frame

$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Amplitude envelope

- Max amplitude value of all samples in a frame

$$AE_t = \max_{k=t \cdot K}^{(t+1) \cdot K - 1} s(k)$$

Amplitude envelope at frame *t*

Amplitude envelope

- Max amplitude value of all samples in a frame

$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Amplitude envelope at frame *t*

Amplitude of *k*th sample

Amplitude envelope

- Max amplitude value of all samples in a frame

$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Amplitude envelope at frame *t*

Frame size

Amplitude of *k*th sample

Amplitude envelope

- Max amplitude value of all samples in a frame

$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Amplitude envelope at frame *t*

First sample of frame *t*

Amplitude of *k*th sample

Amplitude envelope

- Max amplitude value of all samples in a frame

Last sample of frame *t*

Amplitude of *k*th sample

$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Amplitude envelope at frame *t*

First sample of frame *t*

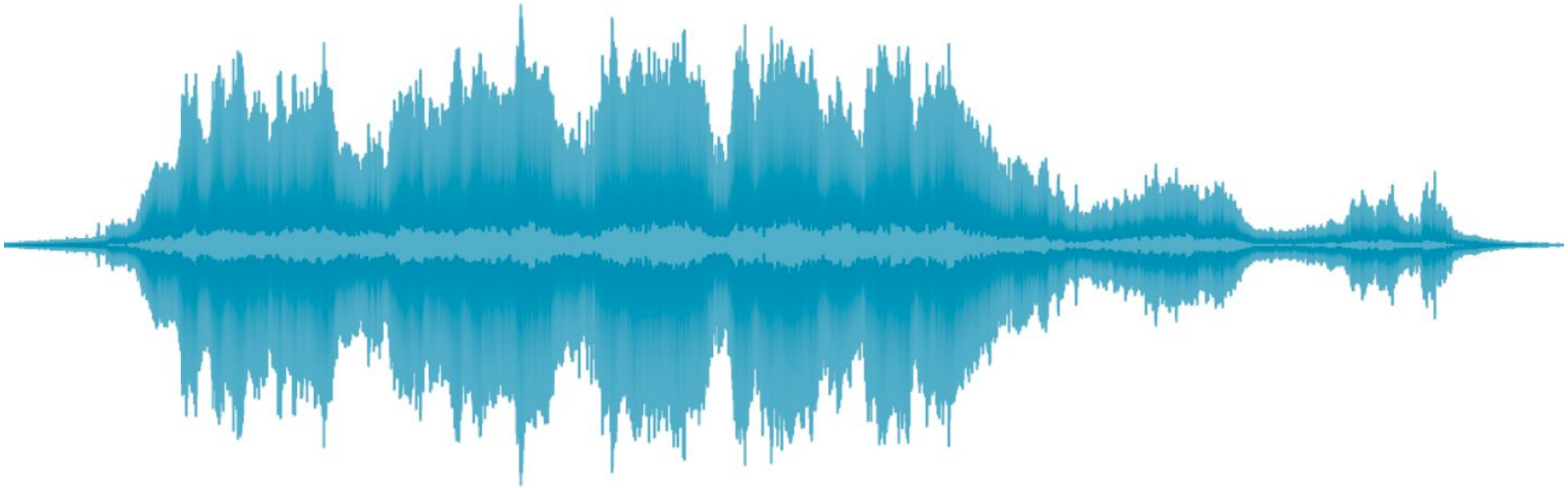
Amplitude envelope

- Max amplitude value of all samples in a frame

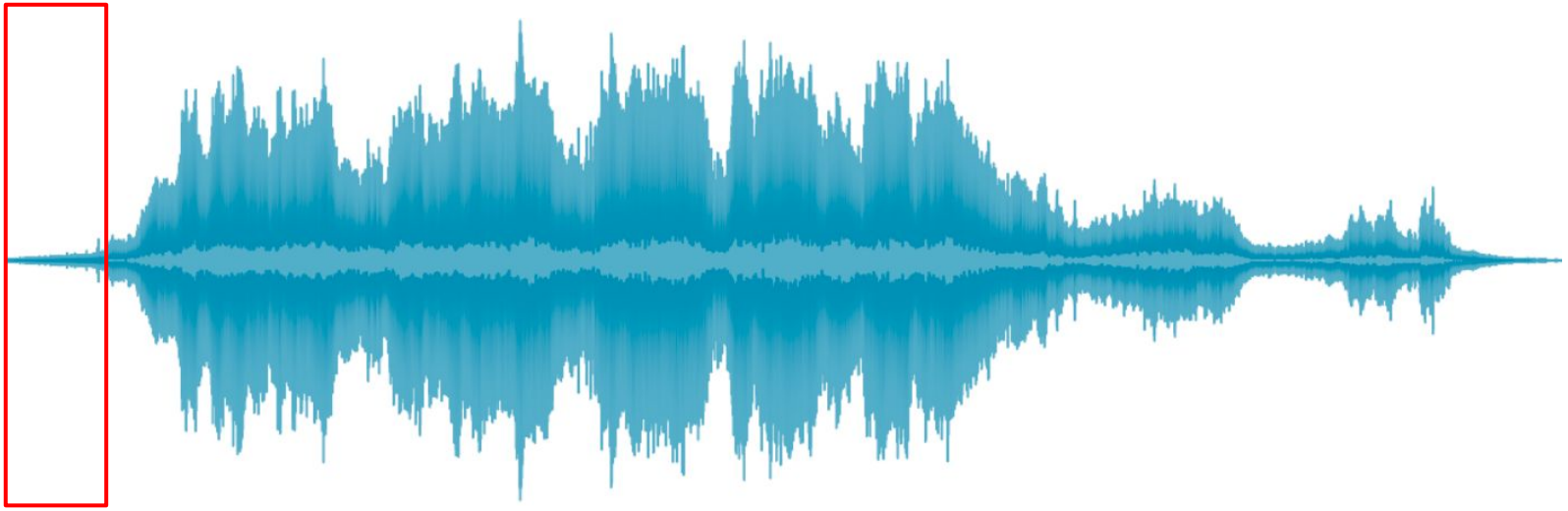
$$AE_t = \max_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)$$

Calculate AE for all the frames

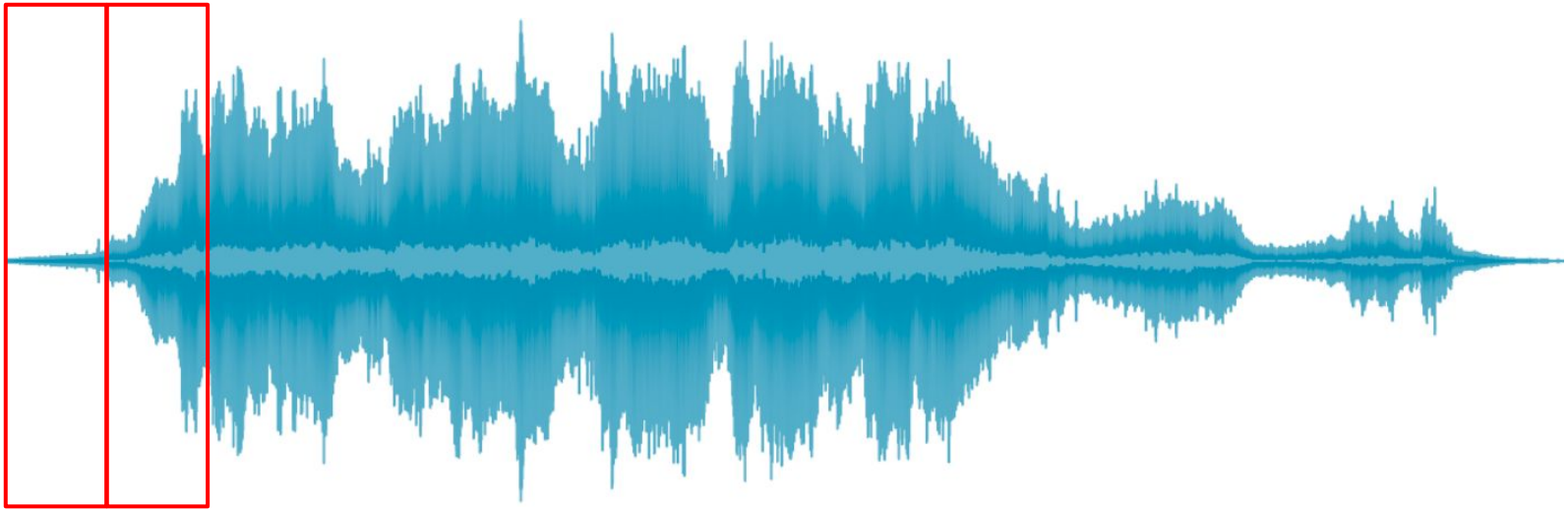
Amplitude envelope



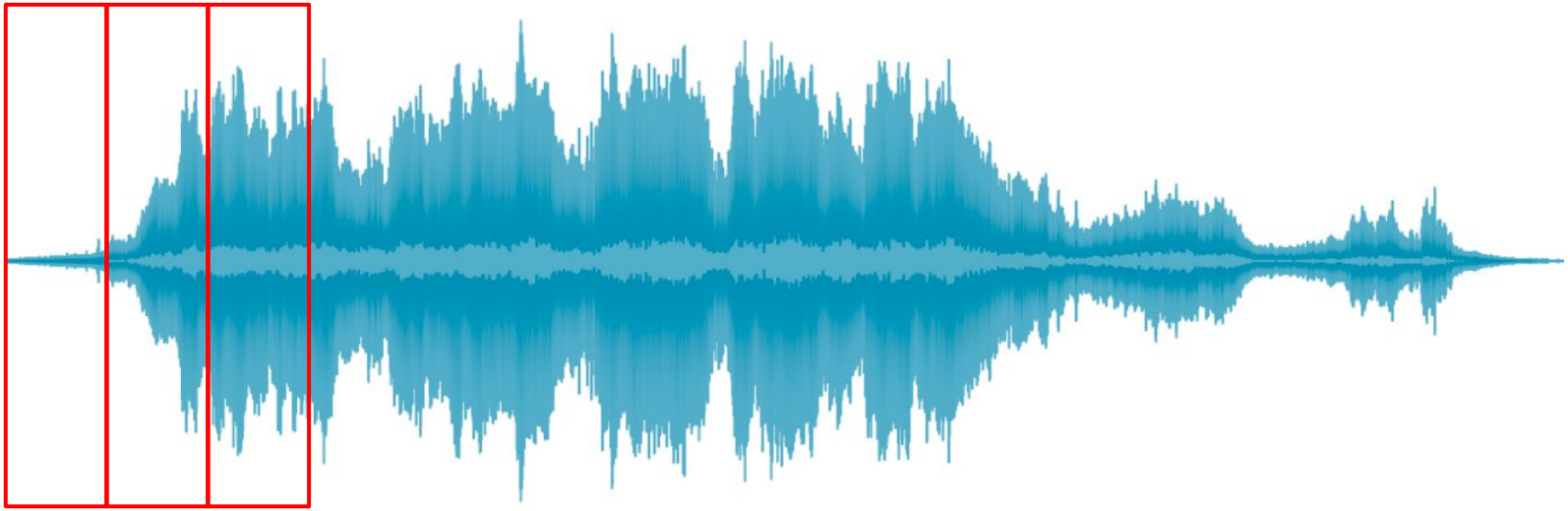
Amplitude envelope



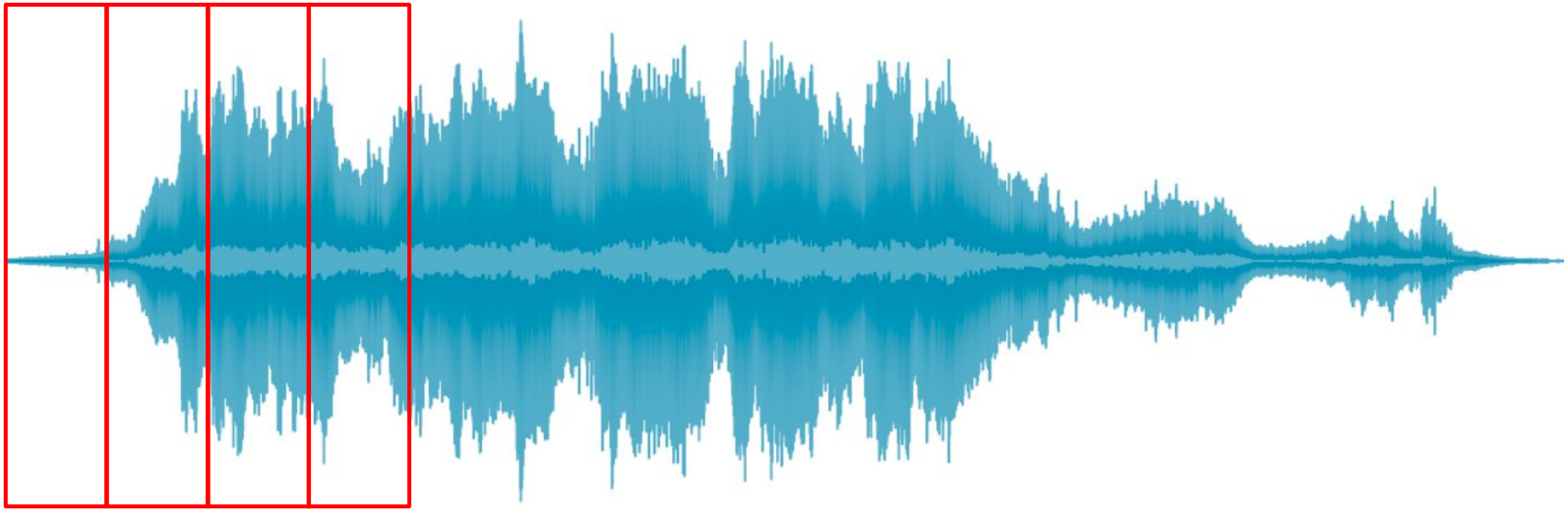
Amplitude envelope



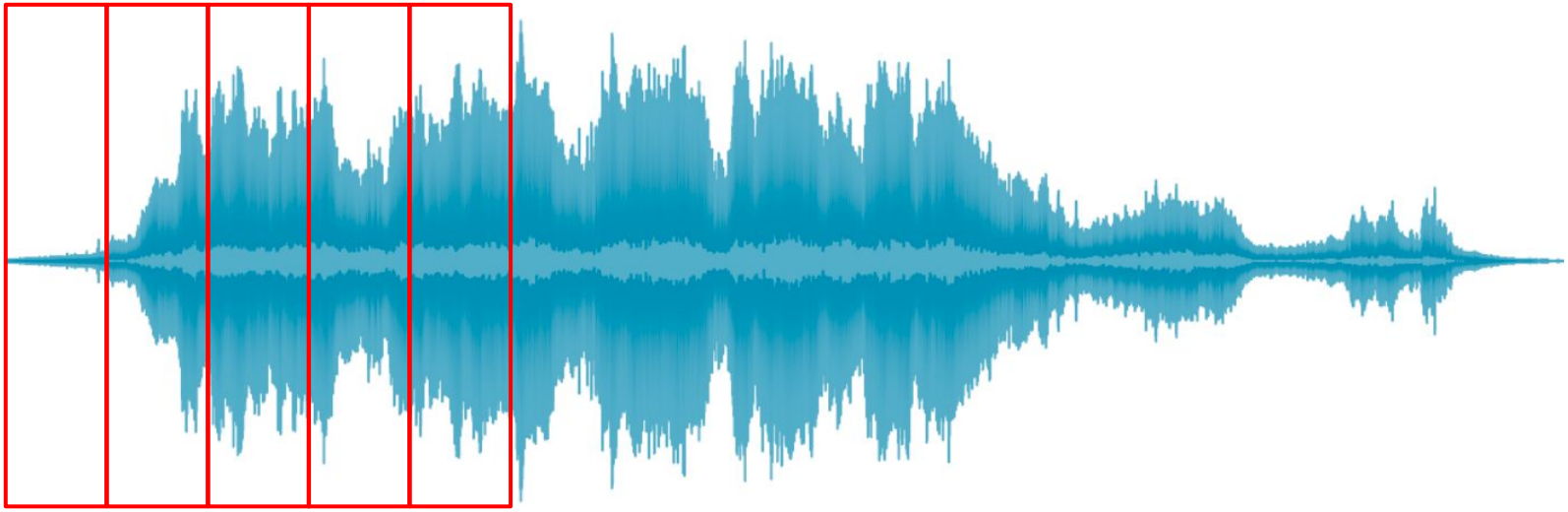
Amplitude envelope



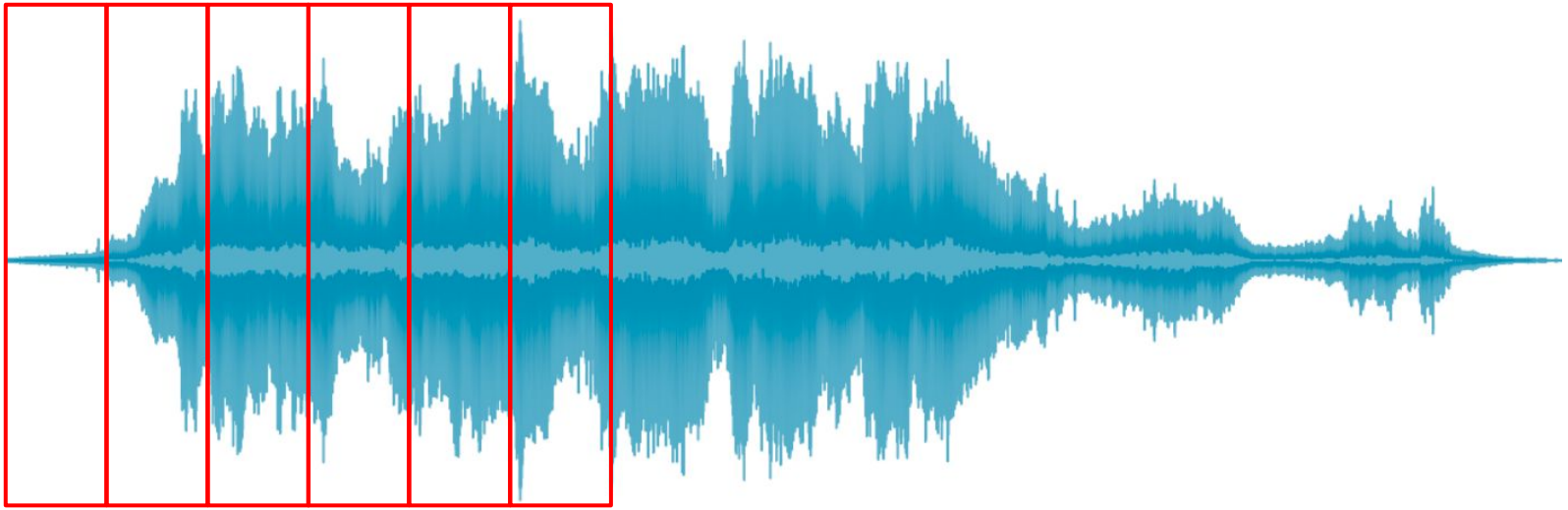
Amplitude envelope



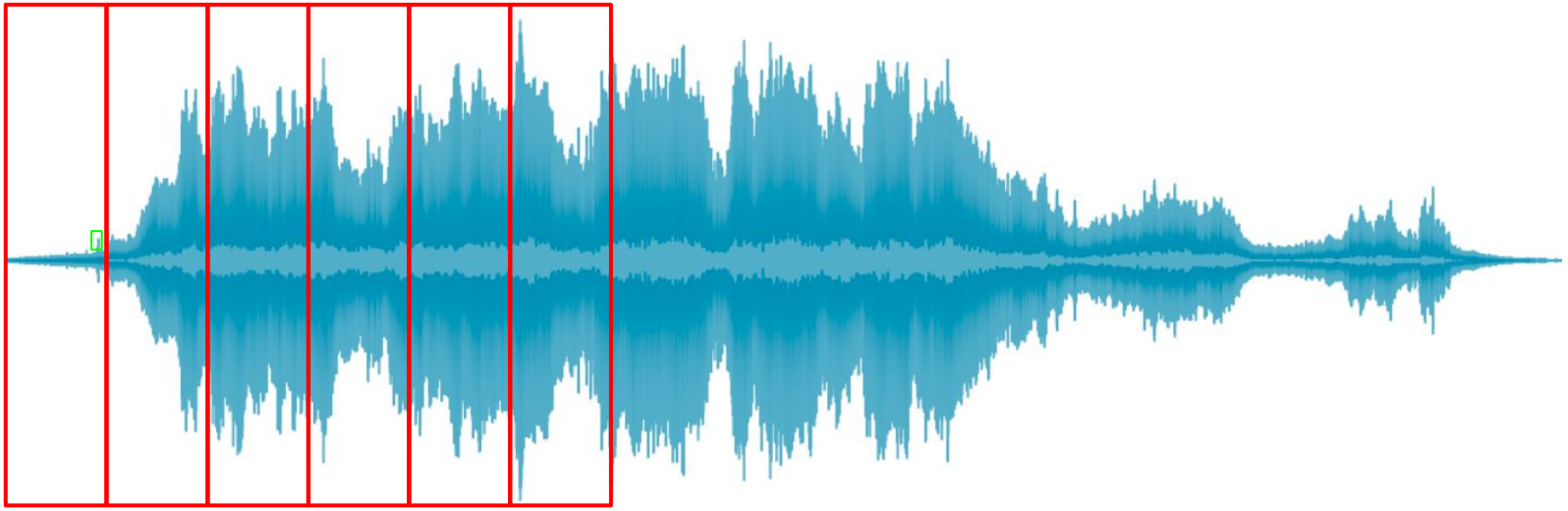
Amplitude envelope



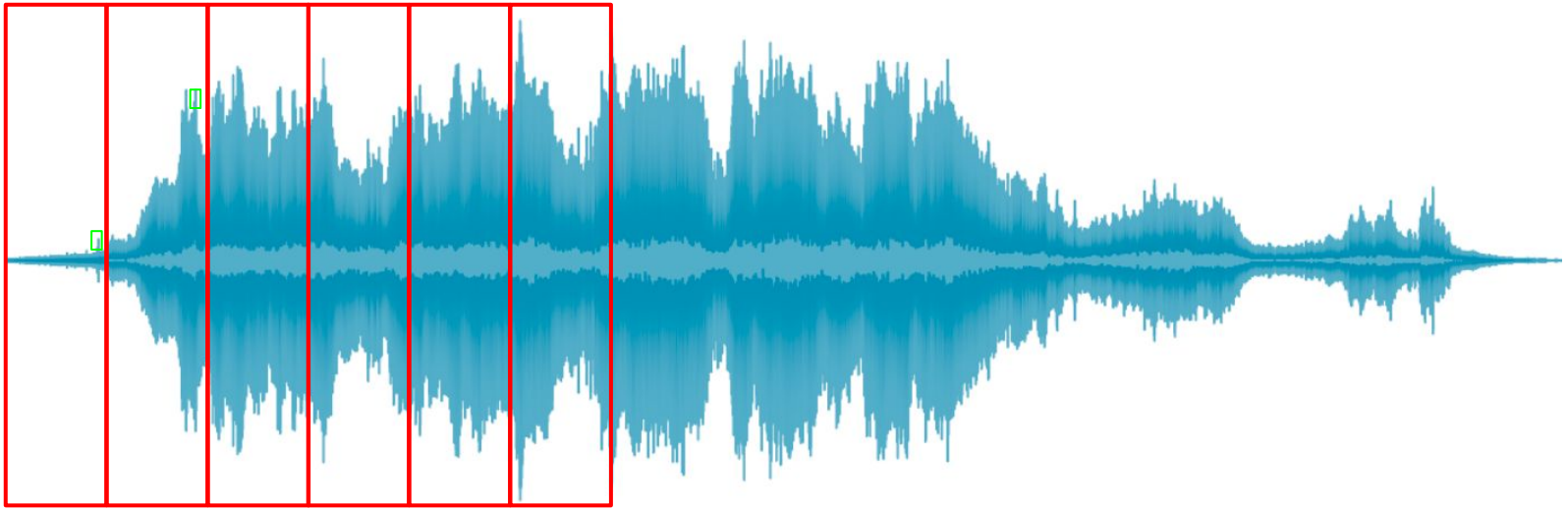
Amplitude envelope



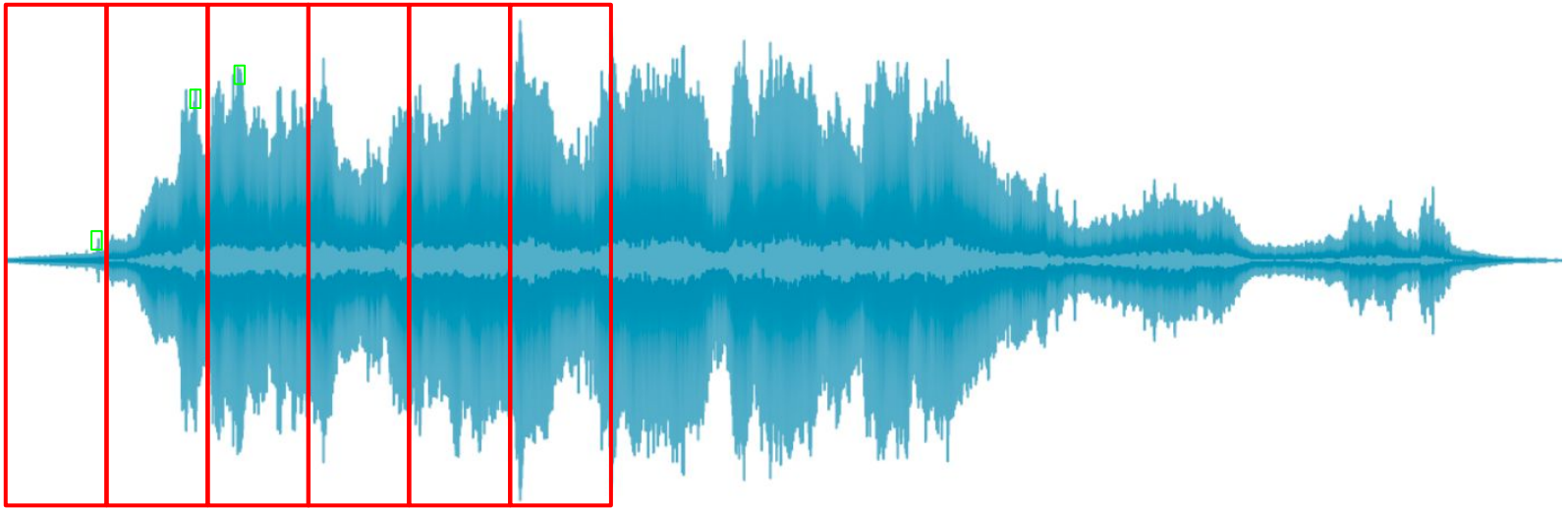
Amplitude envelope



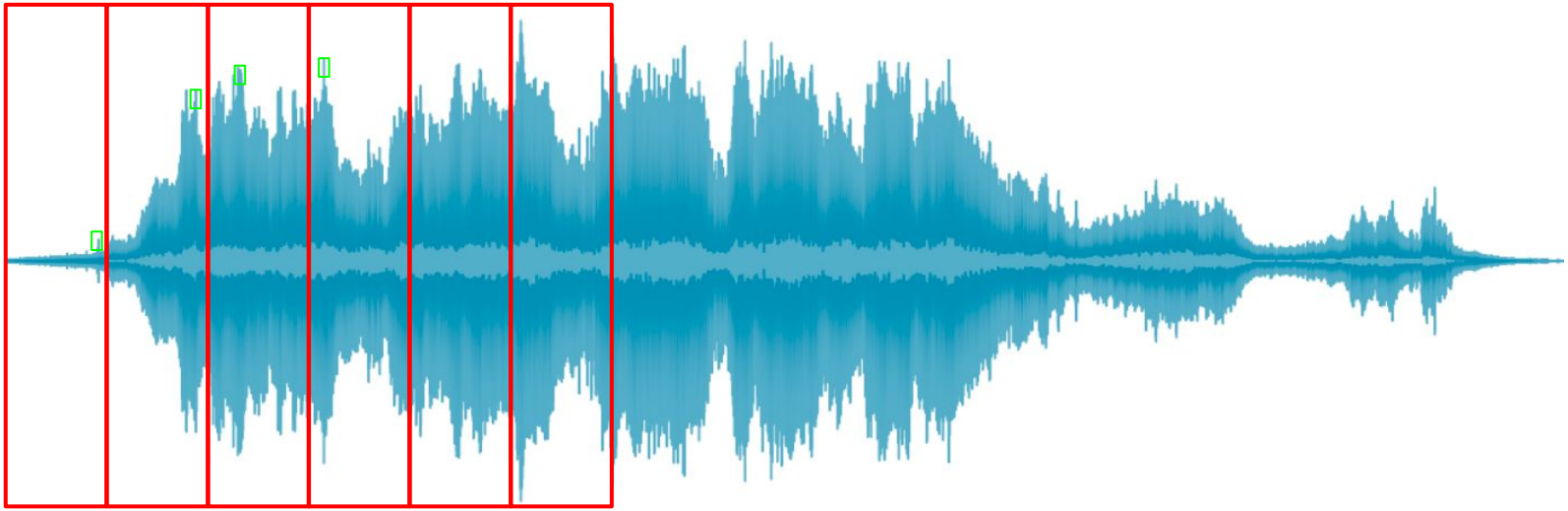
Amplitude envelope



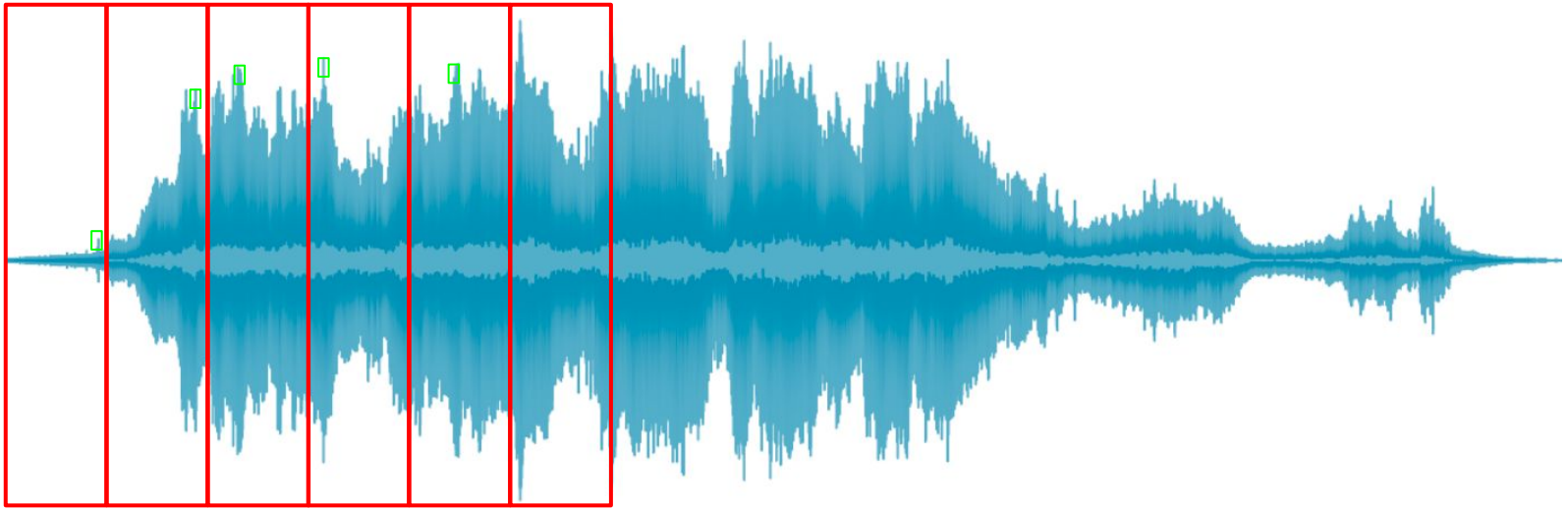
Amplitude envelope



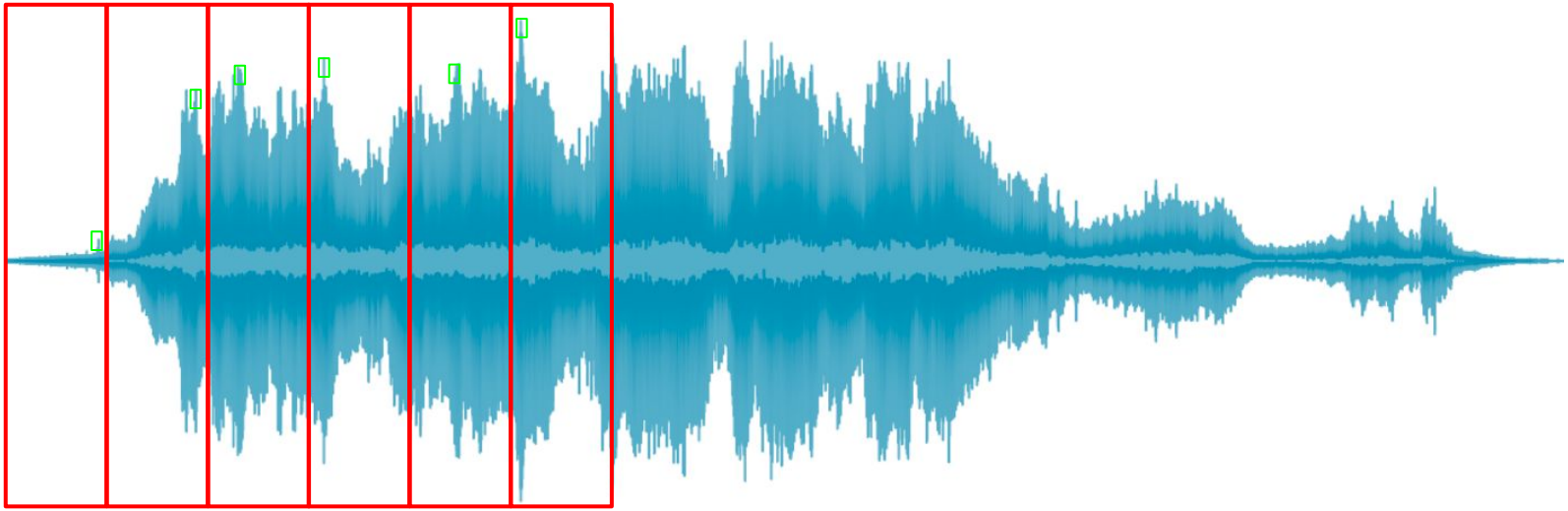
Amplitude envelope



Amplitude envelope



Amplitude envelope



Amplitude envelope

- Max amplitude value of all samples in a frame
- Gives rough idea of loudness

Amplitude envelope

- Max amplitude value of all samples in a frame
- Gives rough idea of loudness
- Sensitive to outliers

Amplitude envelope

- Max amplitude value of all samples in a frame
- Gives rough idea of loudness
- Sensitive to outliers
- Onset detection, music genre classification

Root-mean-square energy

- RMS of all samples in a frame

Root-mean-square energy

- RMS of all samples in a frame

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

Root-mean-square energy

- RMS of all samples in a frame

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

Energy of *k*th sample

Root-mean-square energy

- RMS of all samples in a frame

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

Sum of energy for all samples in frame *t*

Root-mean-square energy

- RMS of all samples in a frame

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

Mean of sum of energy

Root-mean-square energy

- RMS of all samples in a frame
- Indicator of loudness

Root-mean-square energy

- RMS of all samples in a frame
- Indicator of loudness
- Less sensitive to outliers than AE

Root-mean-square energy

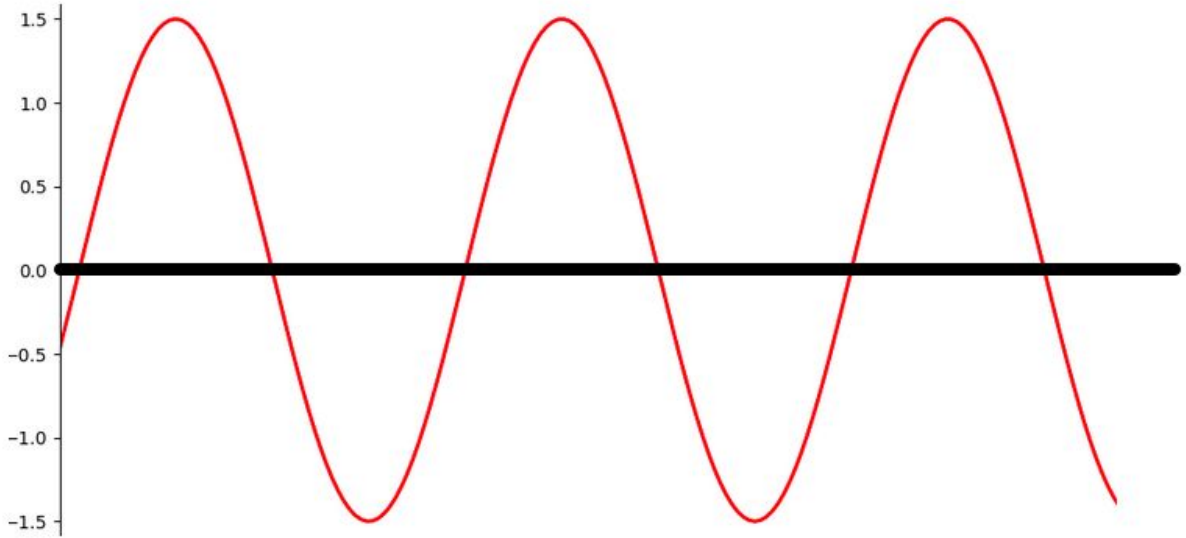
- RMS of all samples in a frame
- Indicator of loudness
- Less sensitive to outliers than AE
- Audio segmentation, music genre classification

Zero crossing rate

- Number of times a signal crosses the horizontal axis

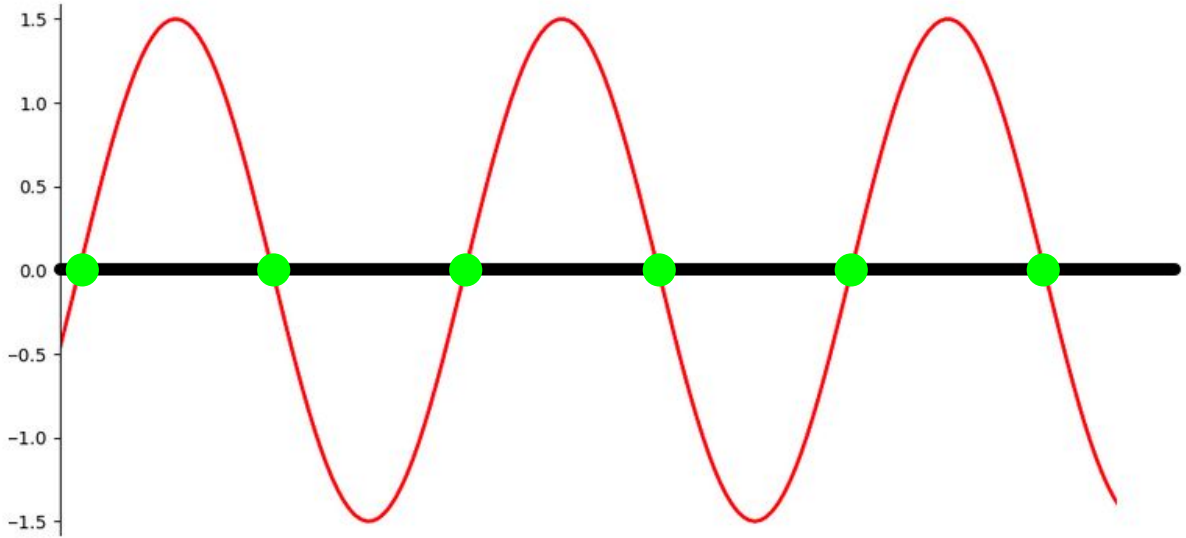
Zero crossing rate

- Number of times a signal crosses the horizontal axis



Zero crossing rate

- Number of times a signal crosses the horizontal axis



Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

Sign function:

- s(k) > 0 → +1
- s(k) < 0 → -1
- s(k) = 0 → 0

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

+ + 0

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

-

+

Zero crossing rate

- Number of times a signal crosses the horizontal axis

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

-

+

2

Zero crossing rate applications

- Recognition of percussive vs pitched sounds

Zero crossing rate applications

- Recognition of percussive vs pitched sounds
- Monophonic pitch estimation

Zero crossing rate applications

- Recognition of percussive vs pitched sounds
- Monophonic pitch estimation
- Voice/unvoiced decision for speech signals

What’s up next?

- Implement amplitude envelope
- Visualise amplitude envelope for different music genres