



**FPT SCHOOL OF BUSINESS
& TECHNOLOGY**

Digital Signal Processing

Các thuộc tính của tín hiệu Audio

Phd. Trần Thanh Trúc

Lý thuyết: Trích xuất một số thuộc tính của tín hiệu Audio

Trích xuất một số thuộc tính của tín hiệu Audio

Nội dung

- Phần 1:
 - Trích chọn các đặc trưng miền thời gian
- Phần 2:
 - Trích chọn các đặc trưng miền tần số
- Phần 3:
 - Trích chọn các đặc trưng miền time-frequency

Trích xuất một số thuộc tính của tín hiệu Audio

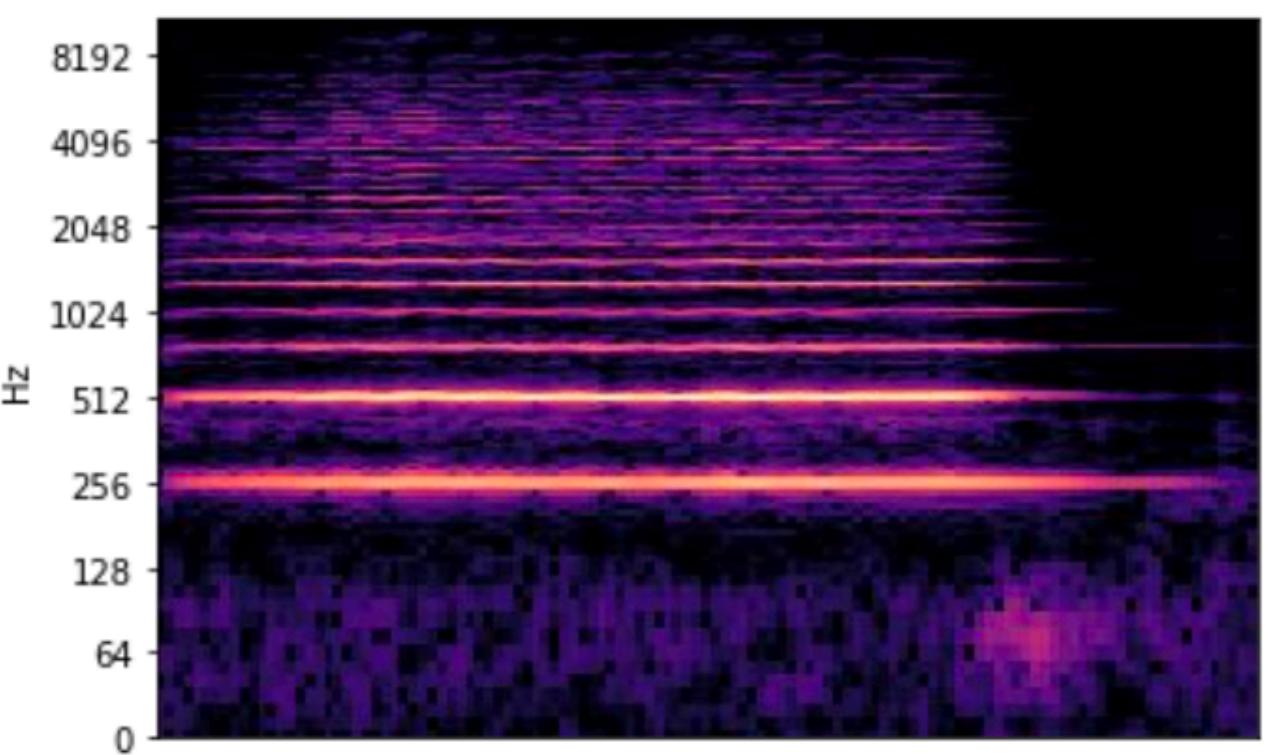
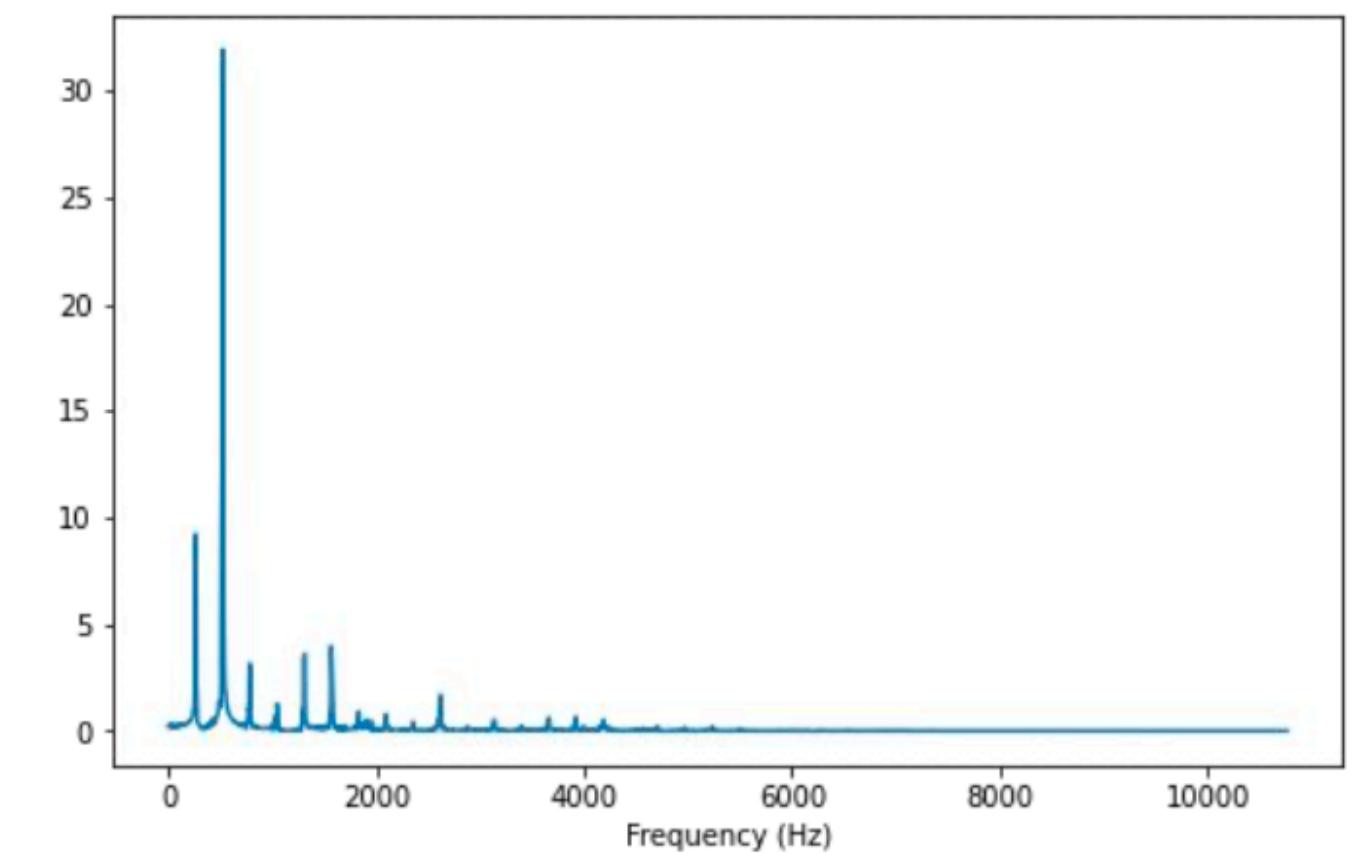
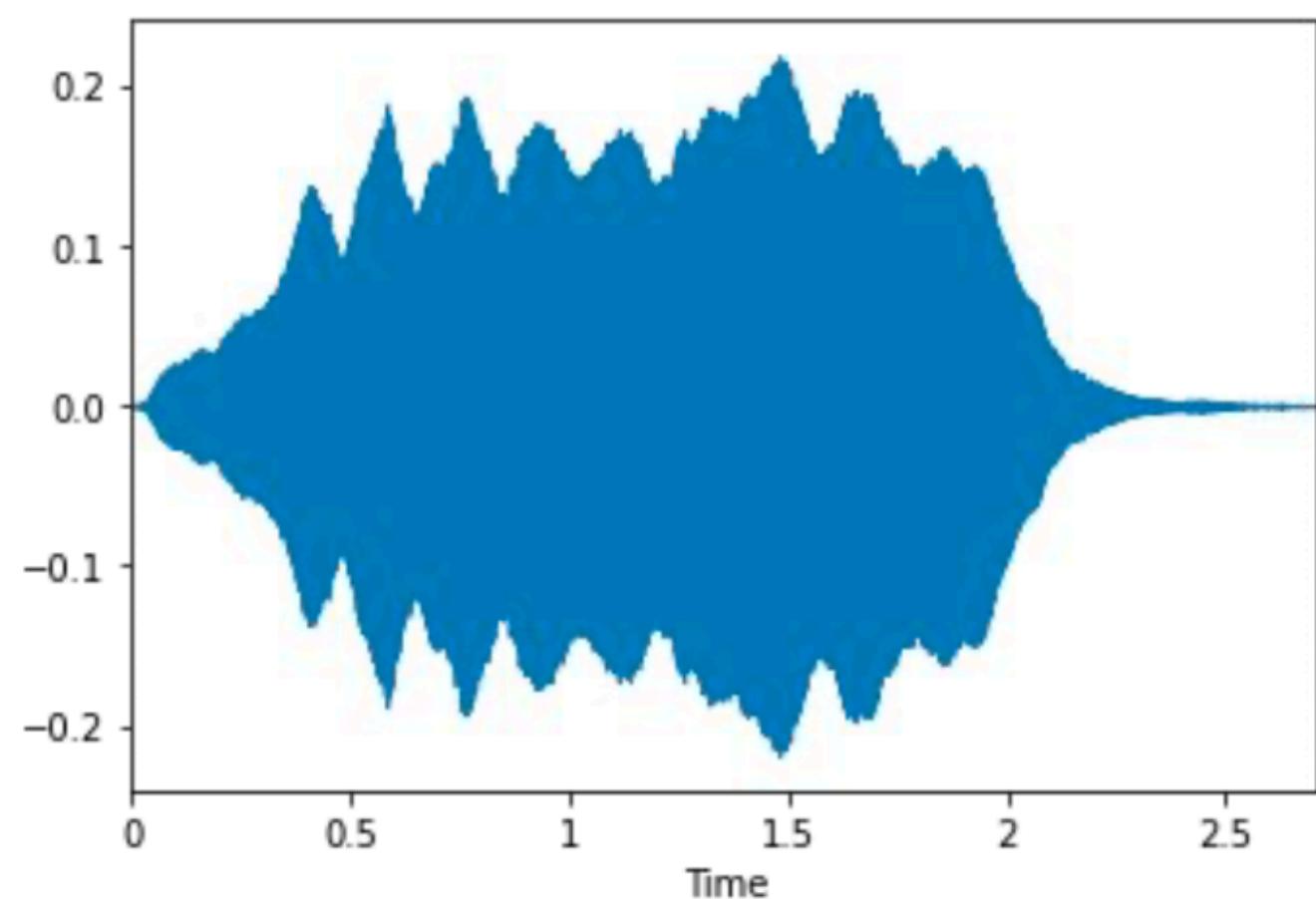
Mục tiêu

- Trích các đặc trưng có thể sử dụng cho học máy
- Sử dụng thư viện librosa

Trích xuất một số thuộc tính của tín hiệu Audio

Ôn lại: Các đặc trưng theo miền tín hiệu

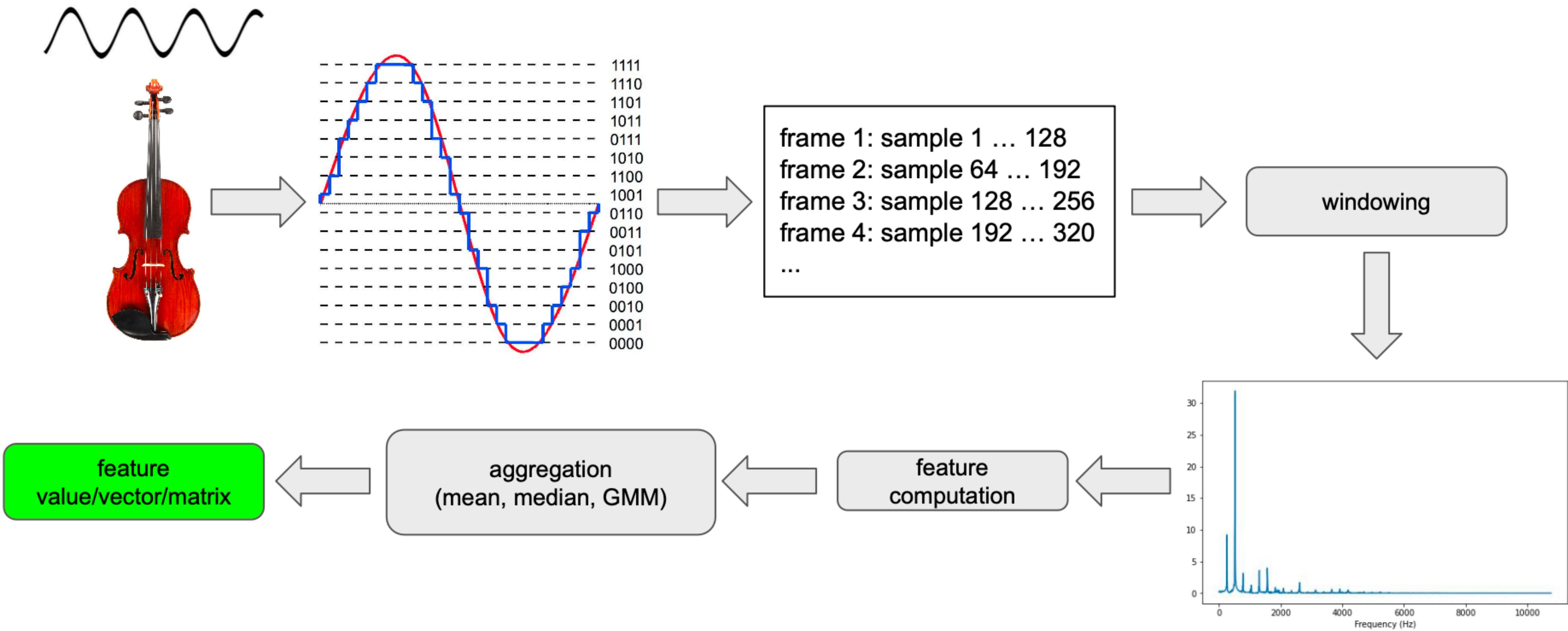
- Miền thời gian
- Miền tần số
 - Công suất in band
 - Tâm phổ
 - Dòng phổ
- Biểu diễn dạng time-frequency
 - Spectrogram
 - Mel-Spectrogram
 - Constant-Q-Transform



Phần 1: Trích chọn các đặc trưng miền thời gian

Trích chọn các đặc trưng miền thời gian

Pipeline xử lý âm thanh số miền thời gian



Trích chọn các đặc trưng miền thời gian

Frame âm thanh

- Đặc tả frame âm thanh
 - Tần số sampling: 44100Hz
 - Chu kỳ sampling: 0.0227ms
 - Số mẫu trên frame: 256 - 8192 samples/ frame

1 sample @44.1KHz = 0.0227ms

Duration 1 sample << Ear's time resolution (10ms)

$$\text{Thời gian frame: } d_f = \frac{1}{S_r} \times K$$

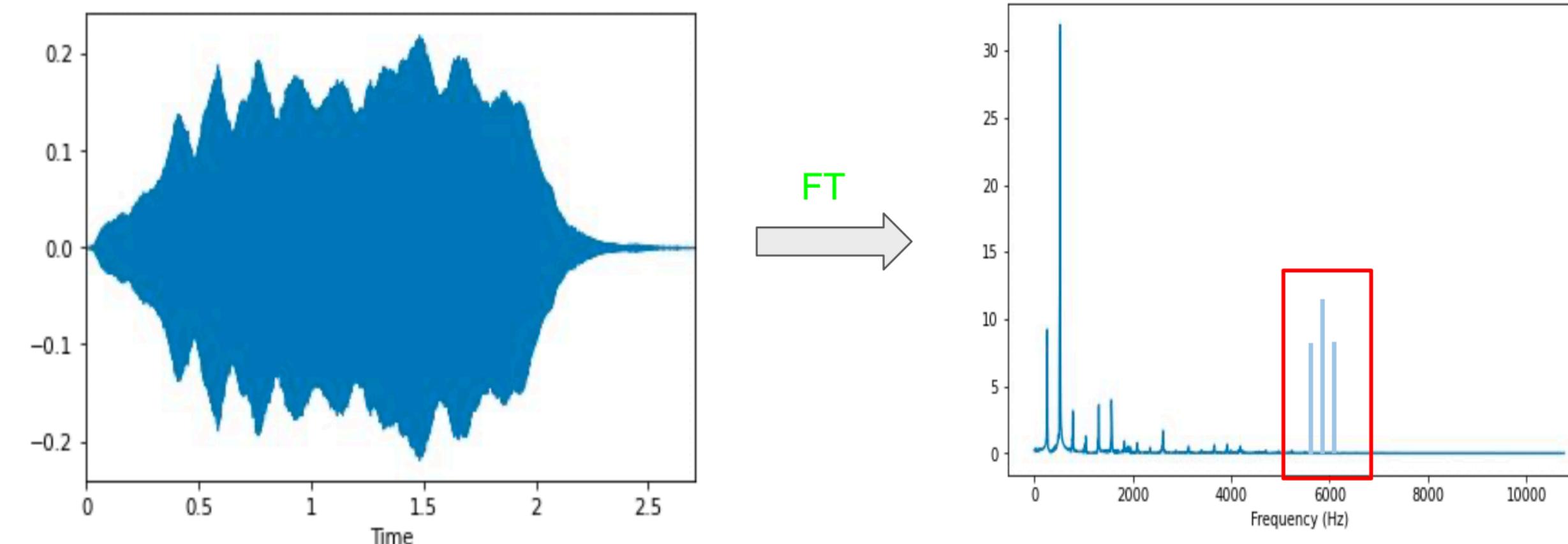
$$d_f = \frac{1}{\frac{44100}{512}} \cdot \boxed{K} = 11.6ms$$

Trích chọn các đặc trưng miền thời gian

Frame âm thanh

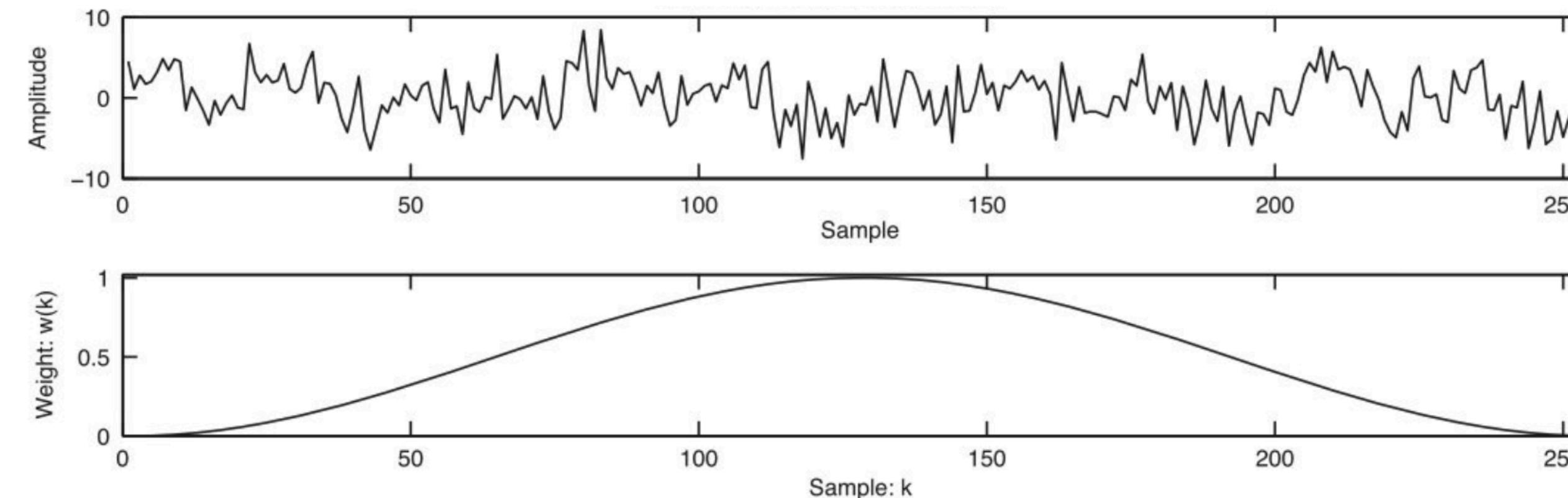
- Dùng các cửa sổ để cắt âm thanh thành từng frame
 - Nếu sử dụng các cửa sổ dạng hộp chữ nhật dẫn đến hiện tượng Spectrum Leakage do nhân tạo
 - Sử dụng cửa sổ Hanning để loại bỏ hiện tượng Spectrum Leakage do quá cắt xén đột ngột các frame.

Hiện tượng spectral leakage framing

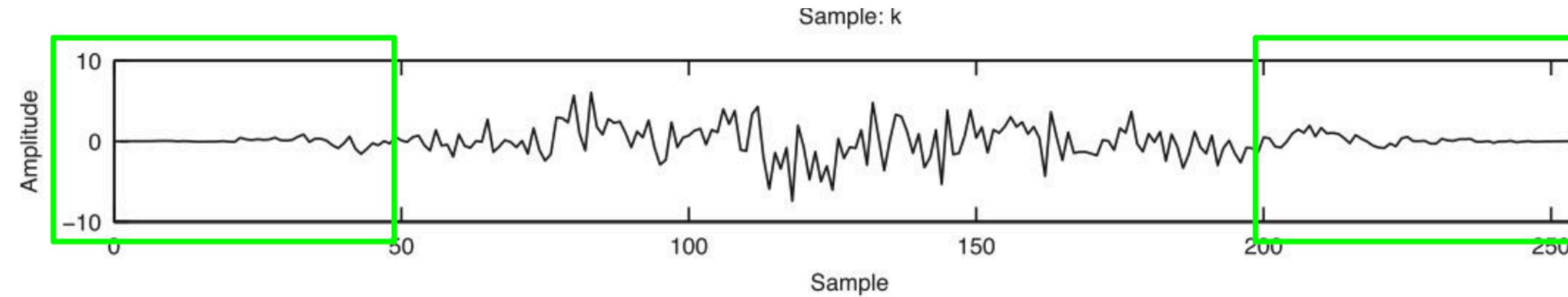


Trích chọn các đặc trưng miền thời gian

Frame âm thanh

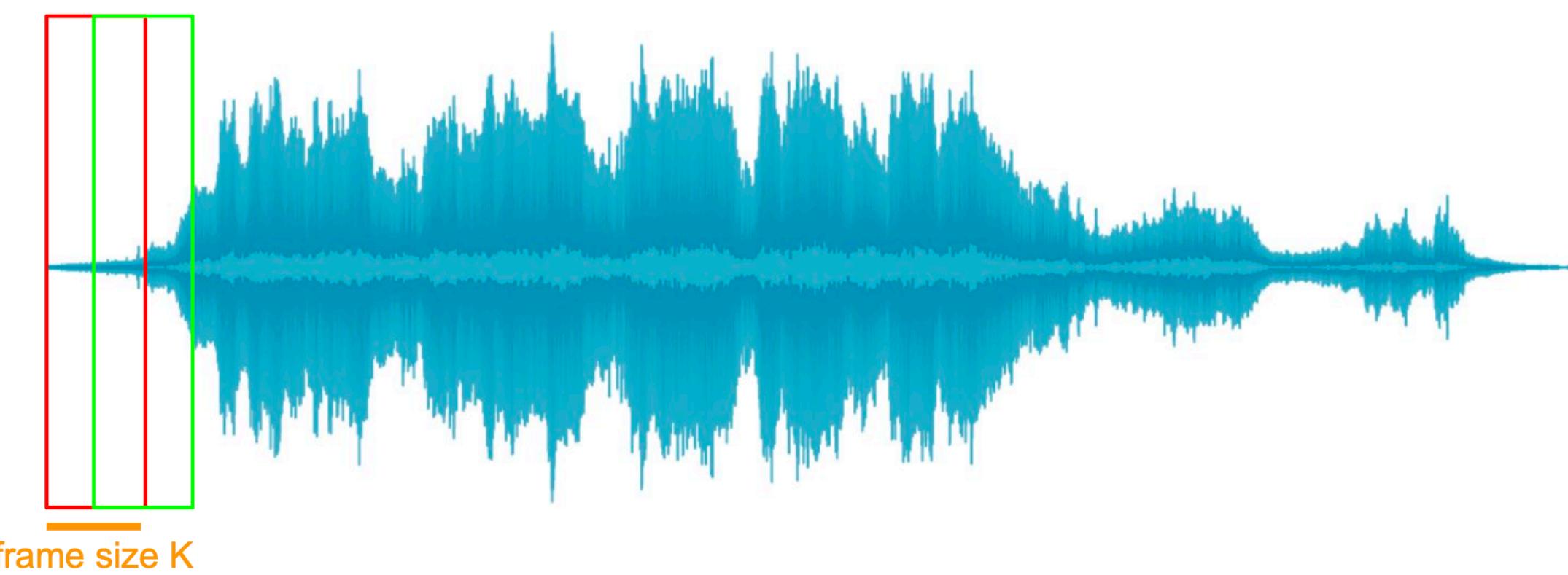
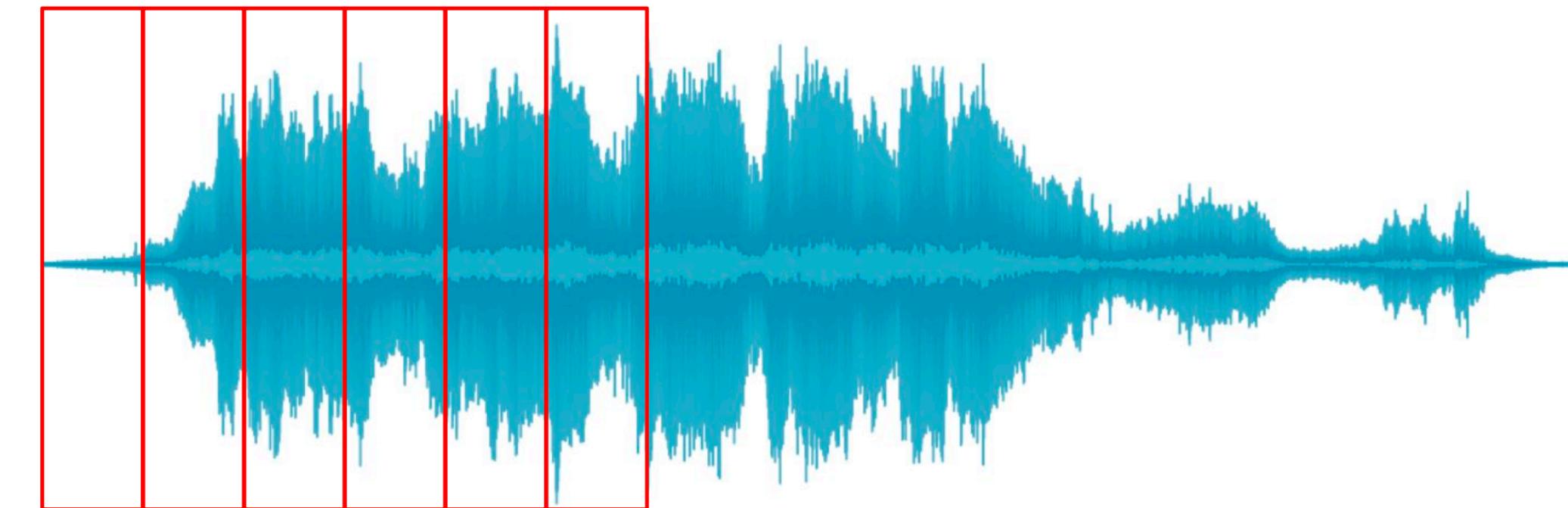
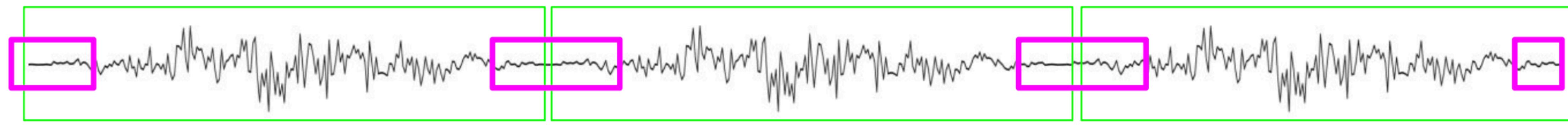


$$s_w(k) = s(k) \cdot w(k), k = 1 \dots K$$



Trích chọn các đặc trưng miền thời gian

Frame âm thanh



Trích chọn các đặc trưng miền thời gian

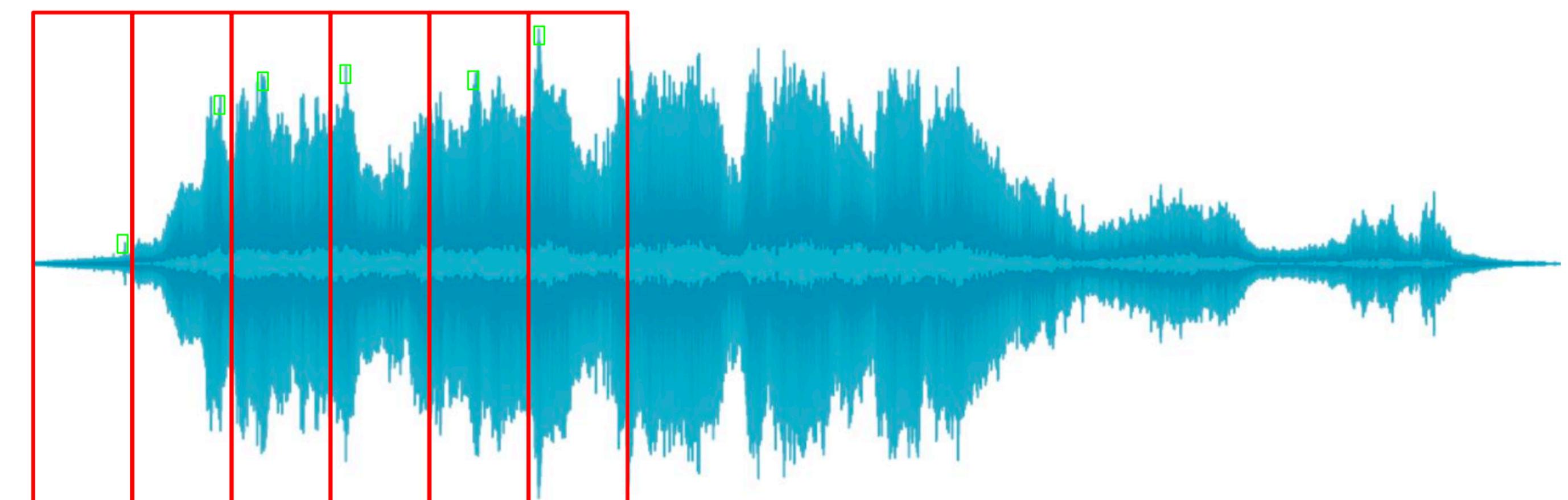
Trích chọn đường bao biên độ

Đường bao biên độ:

- Có liên hệ với độ to (loudness)
- Nhạy cảm với các âm thanh khác biệt (outliner)
- Dùng để phát hiện đoạn dạo đầu (onset), genre trong music

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

Last sample of frame t
 Amplitude envelope at frame t
 First sample of frame t
 Amplitude of k th sample



Trích chọn các đặc trưng miền thời gian

Căn bậc 2 của năng lượng trong một frame

- RMS của âm
 - Chỉ dấu về độ ồn
 - Không nhạy cảm về các âm thanh khác trội như đường bao biên độ
 - Sử dụng để phân đoạn âm thanh, phân loại thể loại âm thanh

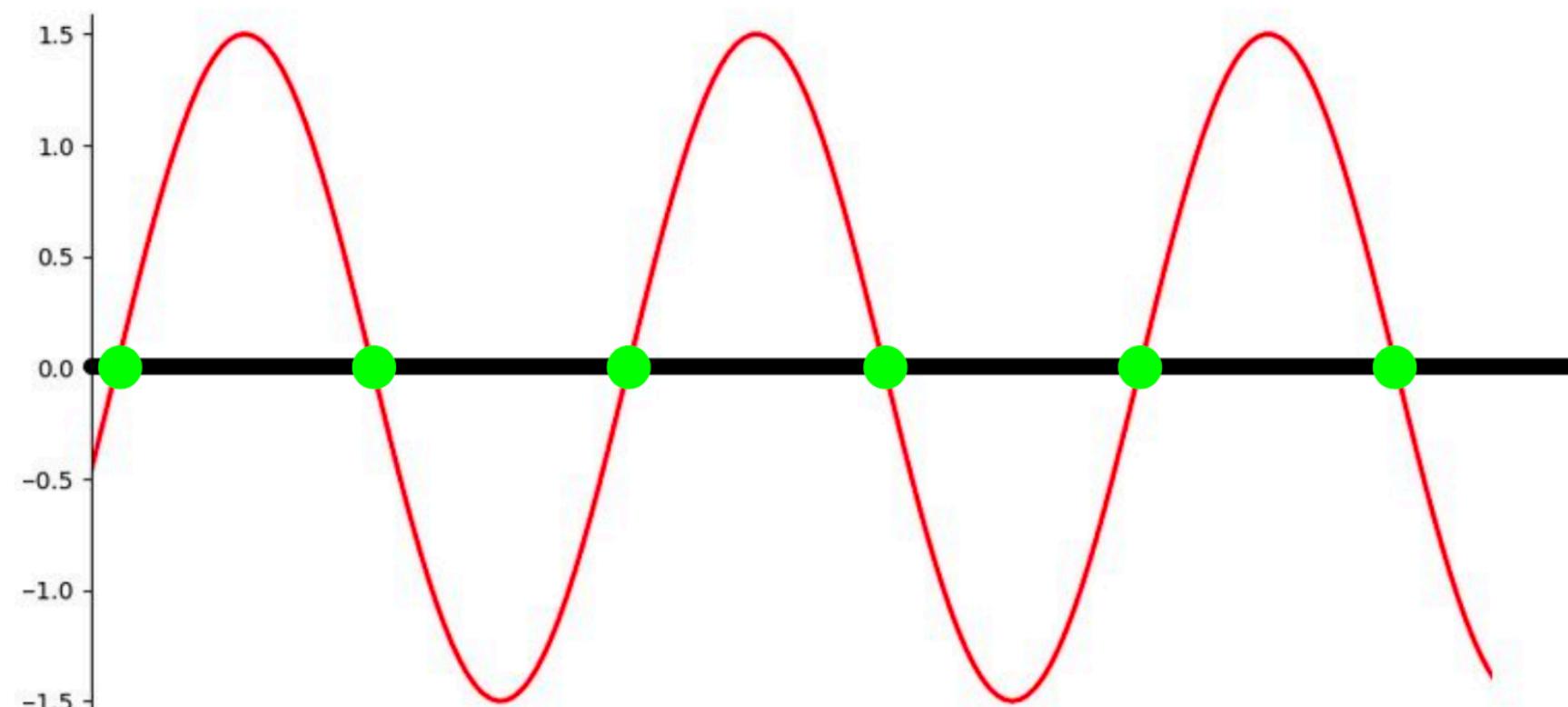
$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} s(k)^2}$$

Trích chọn các đặc trưng miền thời gian

Tần suất qua điểm 0 (zero crossing rate)

Zero crossing rate = số điểm đi qua mức 0 trong frame

- Để nhận dạng cao độ âm thanh
- Nhận dạng các âm thanh của bộ gõ
- Ước tính cao độ của đơn âm
- Xác định âm thanh thoại/ phi thoại



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

Quiz và Lab

Nội dung

- Thực hành xác định đường bao biên độ
- Thực hành xác định tìm rms của âm thanh
- Thực hành xác định tìm zero crossing rate
- Thực hiện phân loại tín hiệu trên tập dataset bằng machine-learning.