



**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 miền tần số

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Nội dung

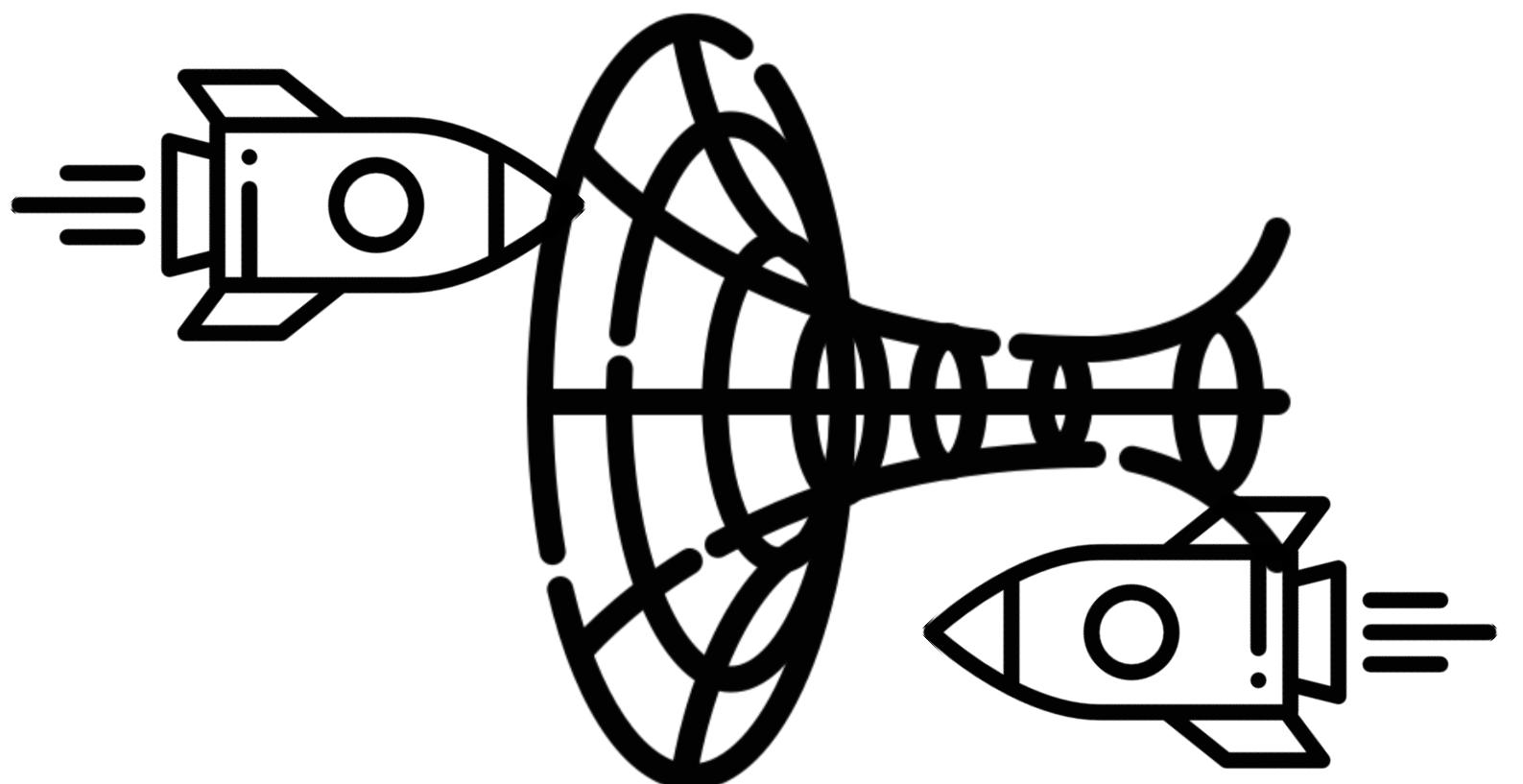
- Short-Time Fourier Transform
- Mel-spectrograms
- Mel-Frequency Cepstral Coefficients
- Band energy ratio (BER)
- Spectral Centroid (Tâm phẩy)
- Bandwidth (Băng thông)

Chuyển đổi Fourier số 1D

Tóm tắt chuyển đổi Fourier 1 chiều

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i2\pi k \frac{n}{N}}$$

- *time t*
- *Real number \mathbb{R}*
- $\sum_{n=0}^{\infty} x(nT_s)\delta(t - nT_s)$
- Rời rạc + không tuần hoàn
- Rời rạc + tuần hoàn



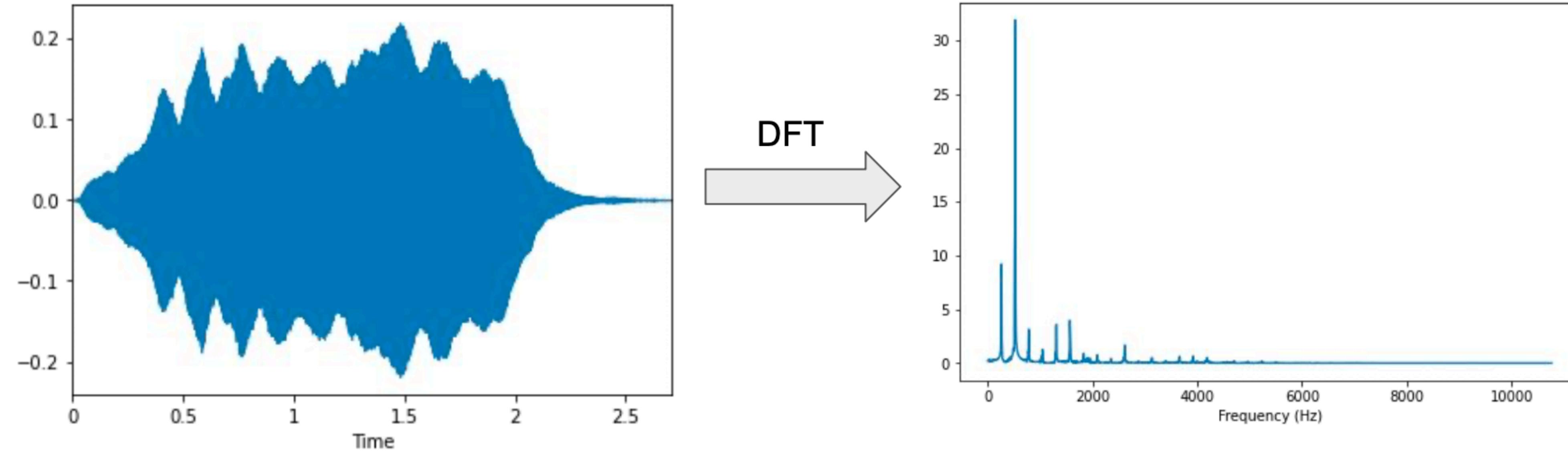
$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot e^{i2\pi k \frac{n}{N}}$$

- *frequency ν*
- *Complex number \mathbb{C}*
- chồng chập về biên độ + pha của phổ xung Dirac
- Liên tục không tuần hoàn
- Hội tụ về rời rạc

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

Không biết được nhiều, đặc tính khác... xảy ra tại thời điểm nào.



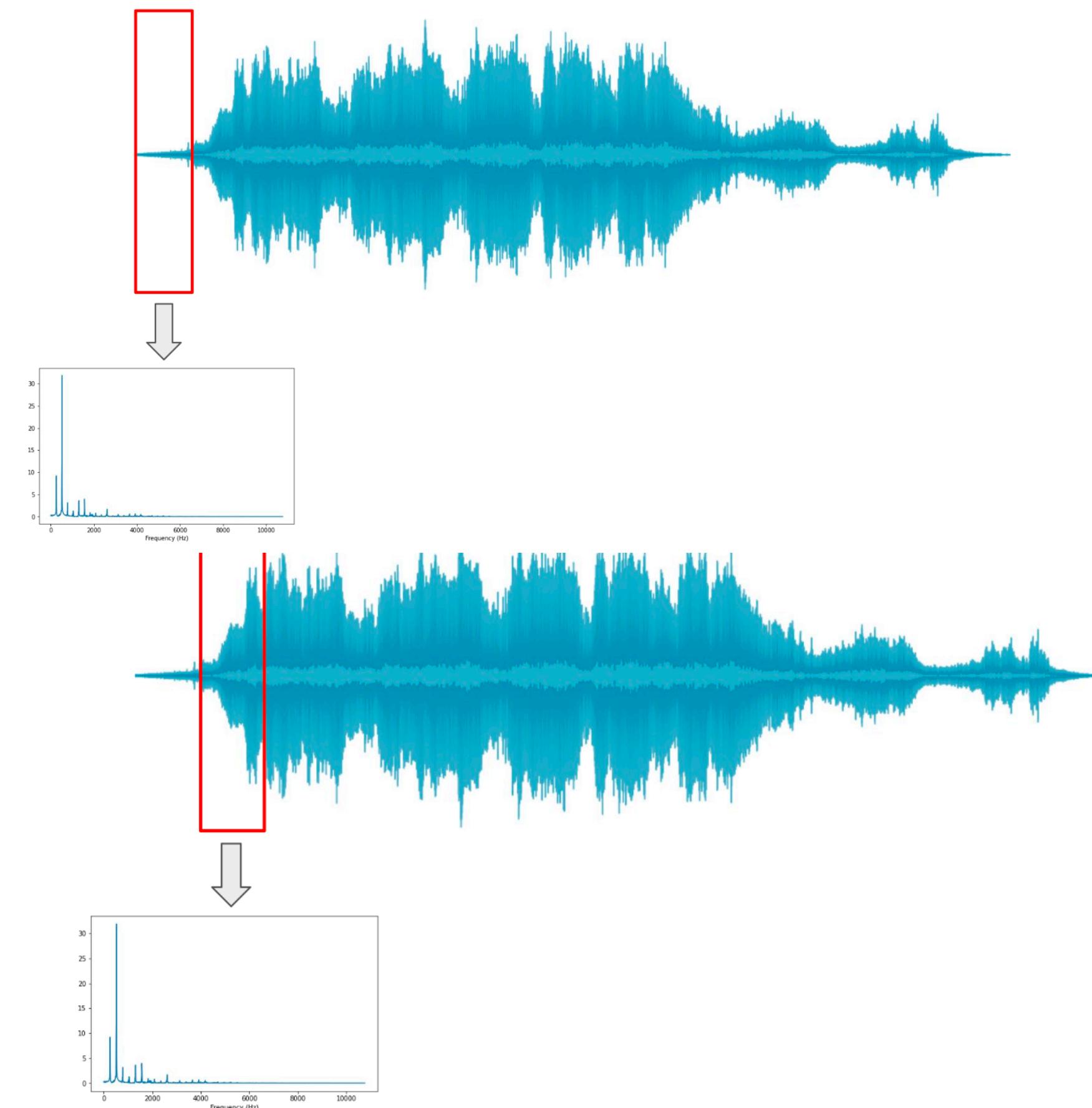
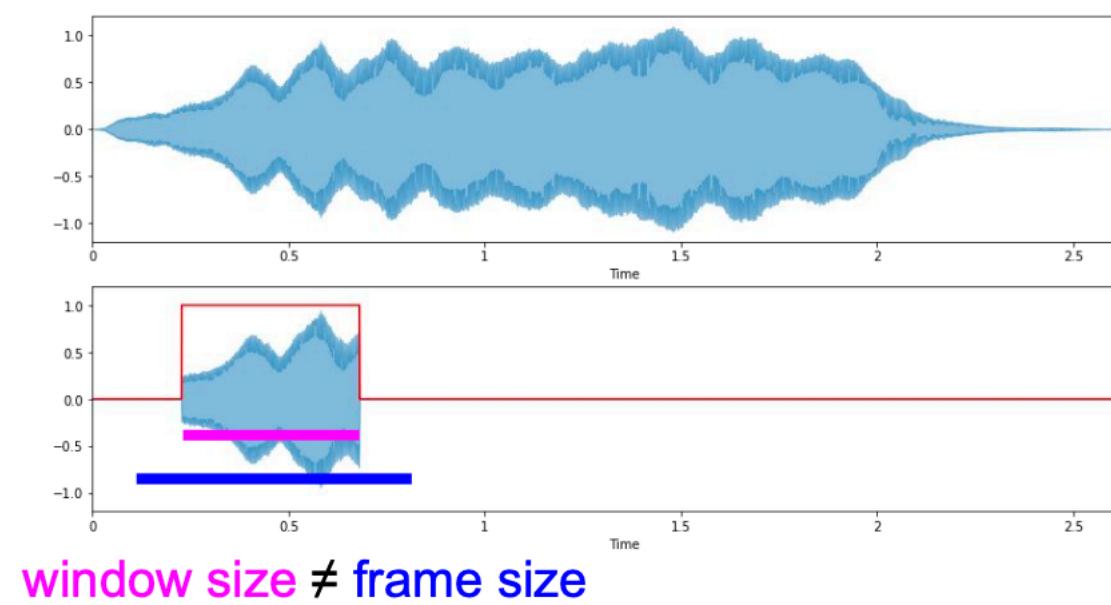
Từ FFT, Làm sao biết được, ví dụ nhiều, xảy ra ở thời điểm nào trong bản nhạc?

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

Giải pháp: chuyển đổi FFT từng đoạn thay vì cả bài nhạc

Sử dụng phương pháp windowing

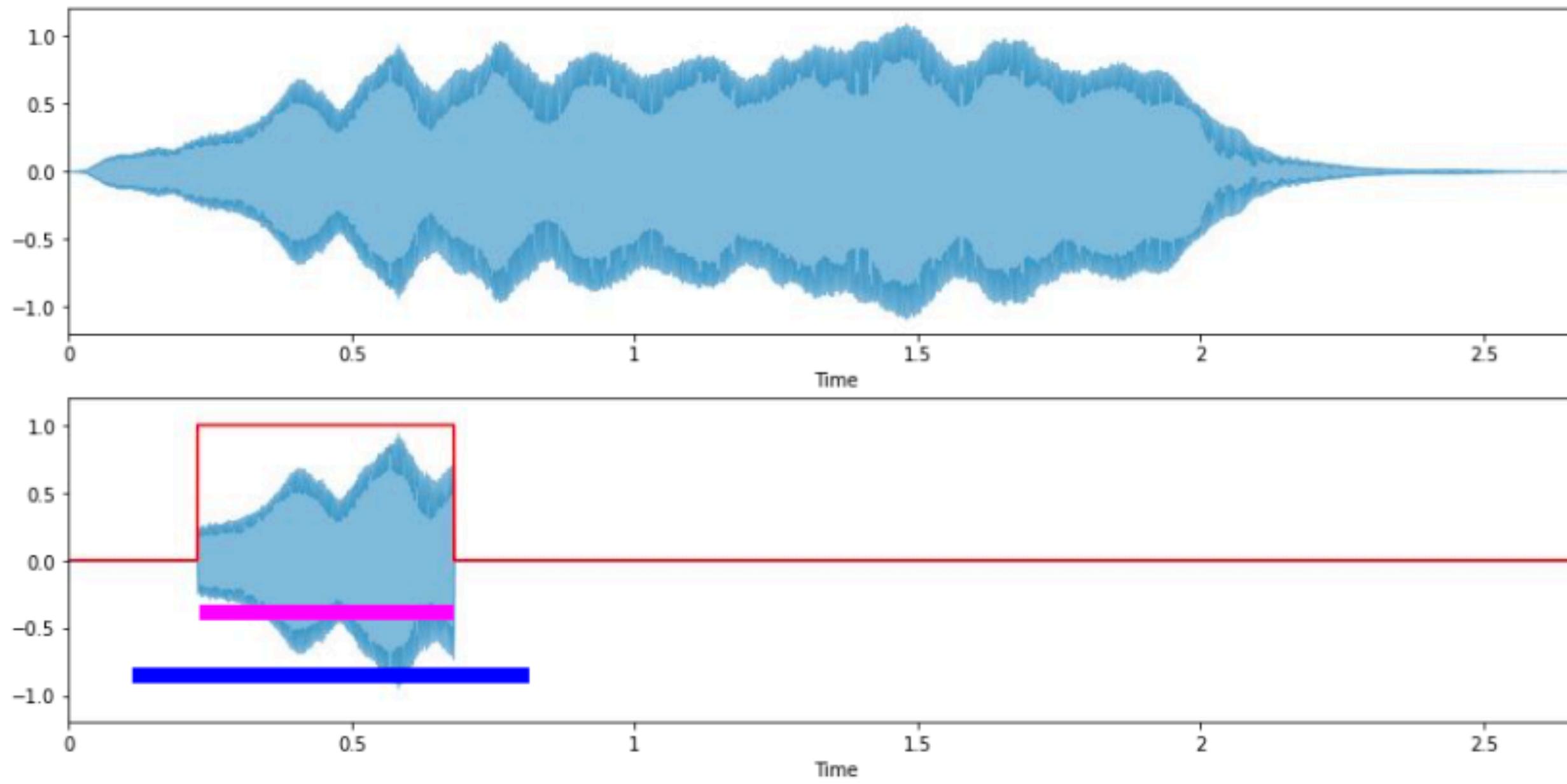


Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

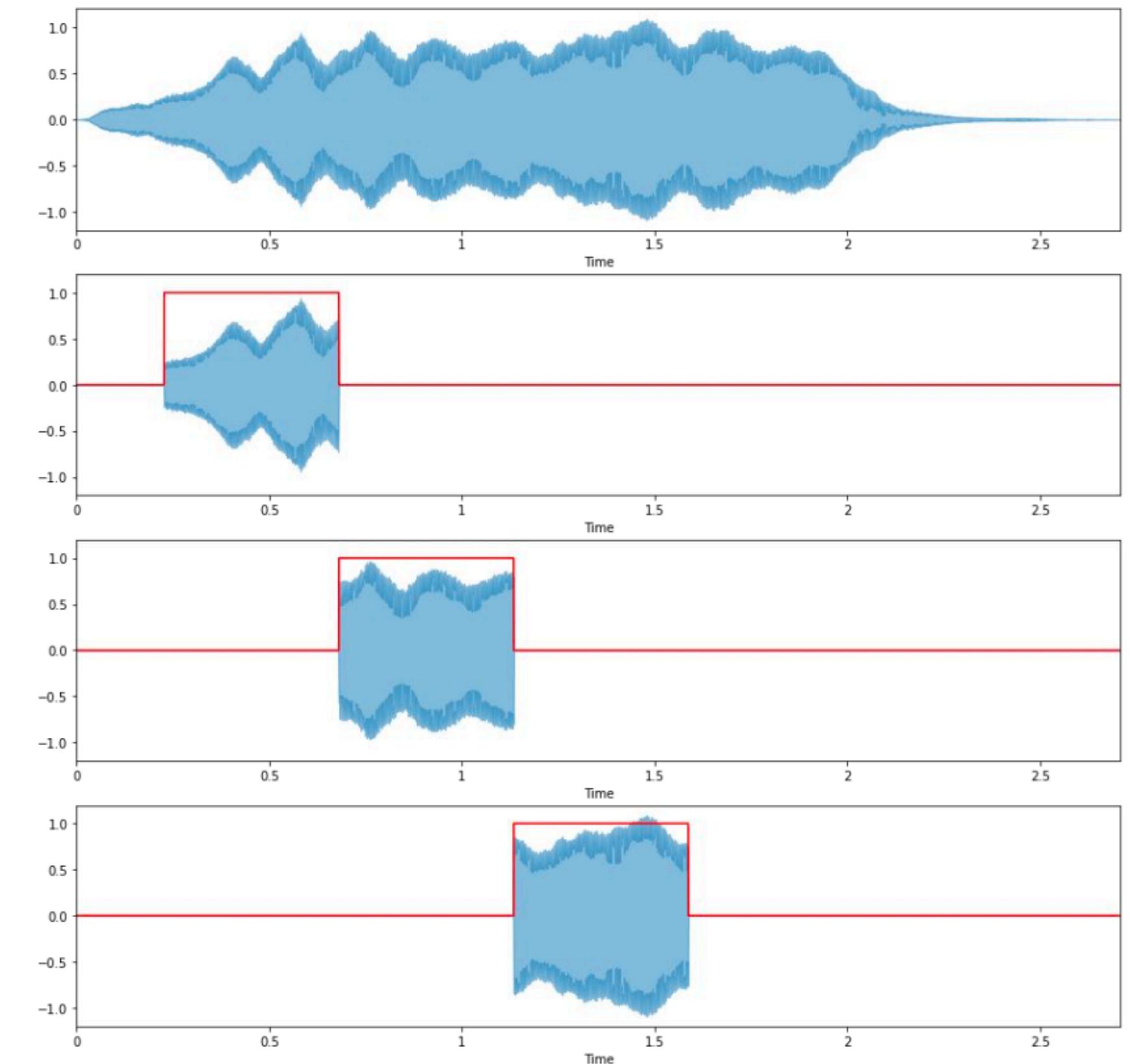
Chuyển đổi STFT (Short-Time Fourier Transform)

Giải pháp: chuyển đổi FFT từng đoạn thay vì cả bài nhạc

Sử dụng phương pháp windowing



window size ≠ frame size

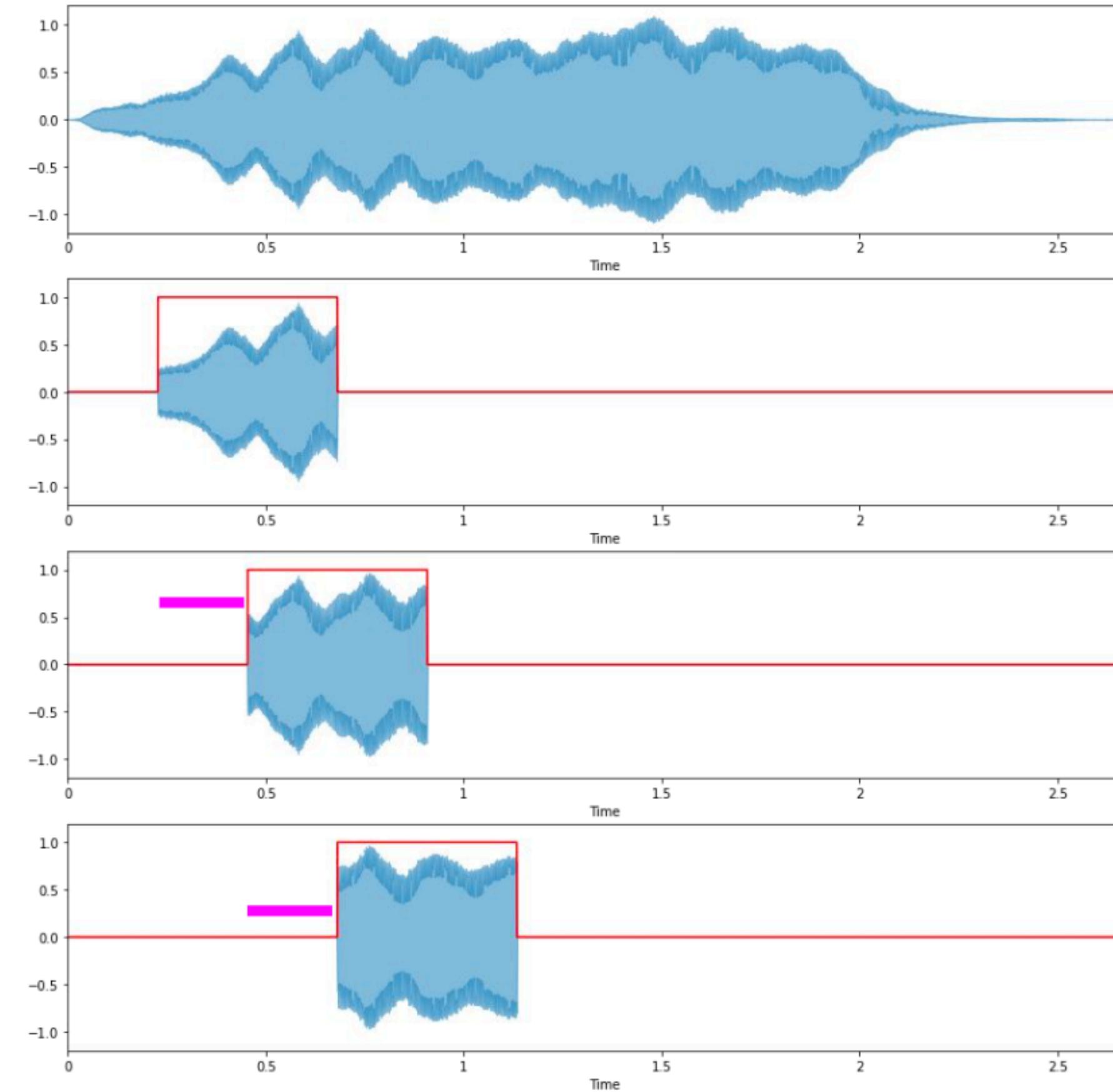


Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

Lưu ý: chúng ta sử dụng phương pháp overlapping frame

hop size (H)

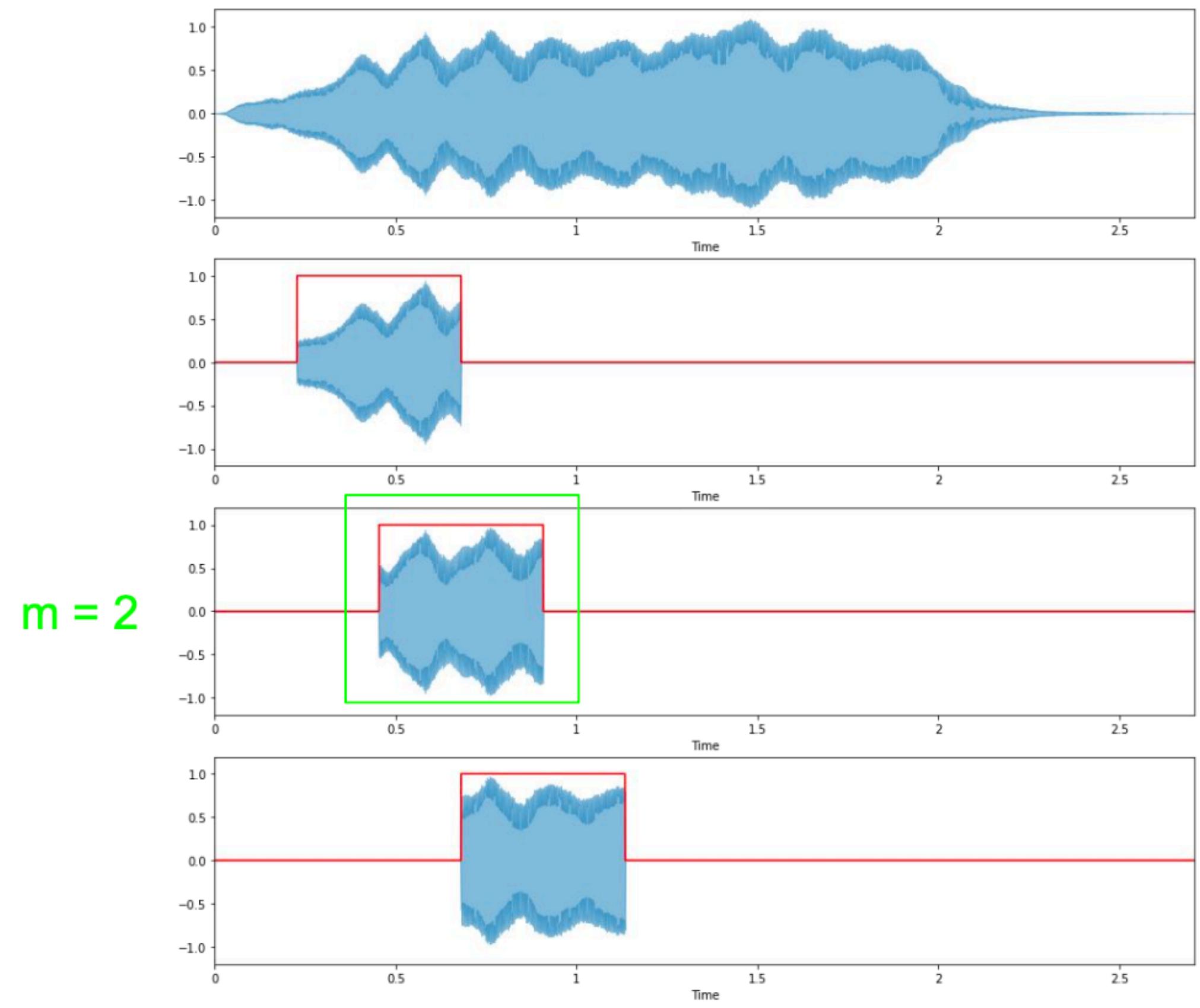
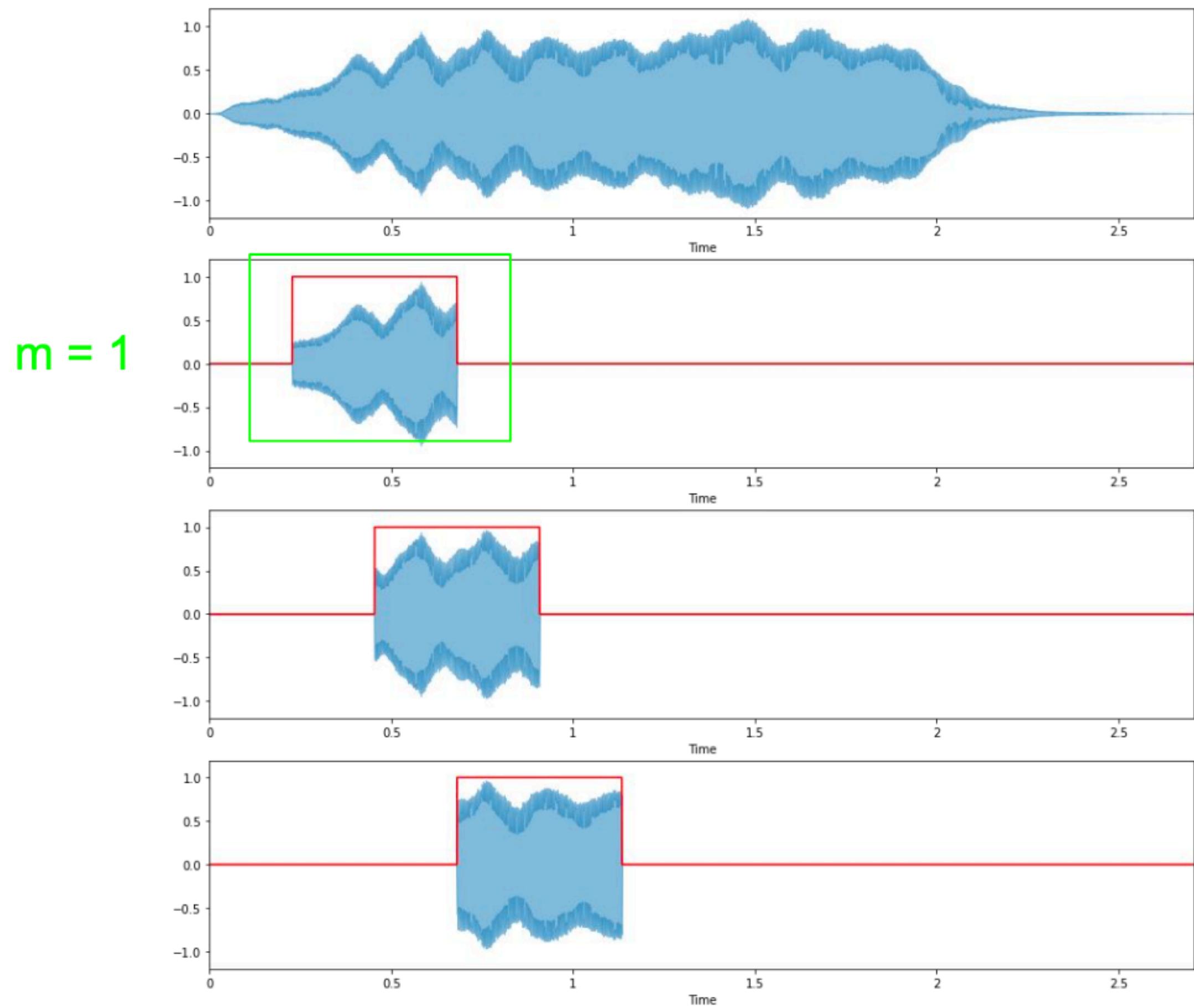


$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

$$S(m, k) = \sum_{n=0}^{N-1} x(n + mH) \cdot w(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)



Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

$$S(m, k) = \sum_{n=0}^{N-1} x(n + mH) \cdot w(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

Starting sample of
 current frame

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Chuyển đổi FFT
 - Chuyển đổi từ miền thời gian —> miền tần số toàn bộ tín hiệu
 - Biểu diễn dưới dạng vectơ các mẫu tần số
 - Giá trị của mẫu là số phức
- Chuyển đổi STFT
 - Chuyển đổi từ miền thời gian —> tần số từng frame tín hiệu
 - Biểu diễn dạng ma trận mẫu các tần số và frame thời gian (# frequency-bin, #frame)
 - Giá trị của mẫu là số phức

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

Loại bỏ thành phần tần số âm do FFT

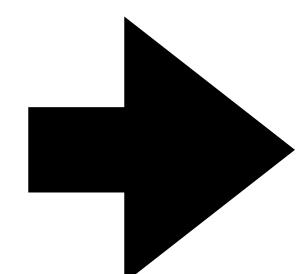
$$\# \text{ frequency bins} = \frac{\textit{framesize}}{2} + 1$$

$$\# \text{ frames} = \frac{\textit{samples} - \textit{framesize}}{\textit{hopsize}} + 1$$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Ví dụ:
 - Tín hiệu: 10000 samples
 - Frame size = 1000
 - Hop size = 500



STFT → (501,19)

▶ Frequency bins = $\frac{1000}{2} + 1 = 501 \rightarrow \left(0, \frac{s_r}{2}\right)$

▶ frames = $\frac{10000 - 1000}{500} + 1 = 19$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Lưu ý 1:

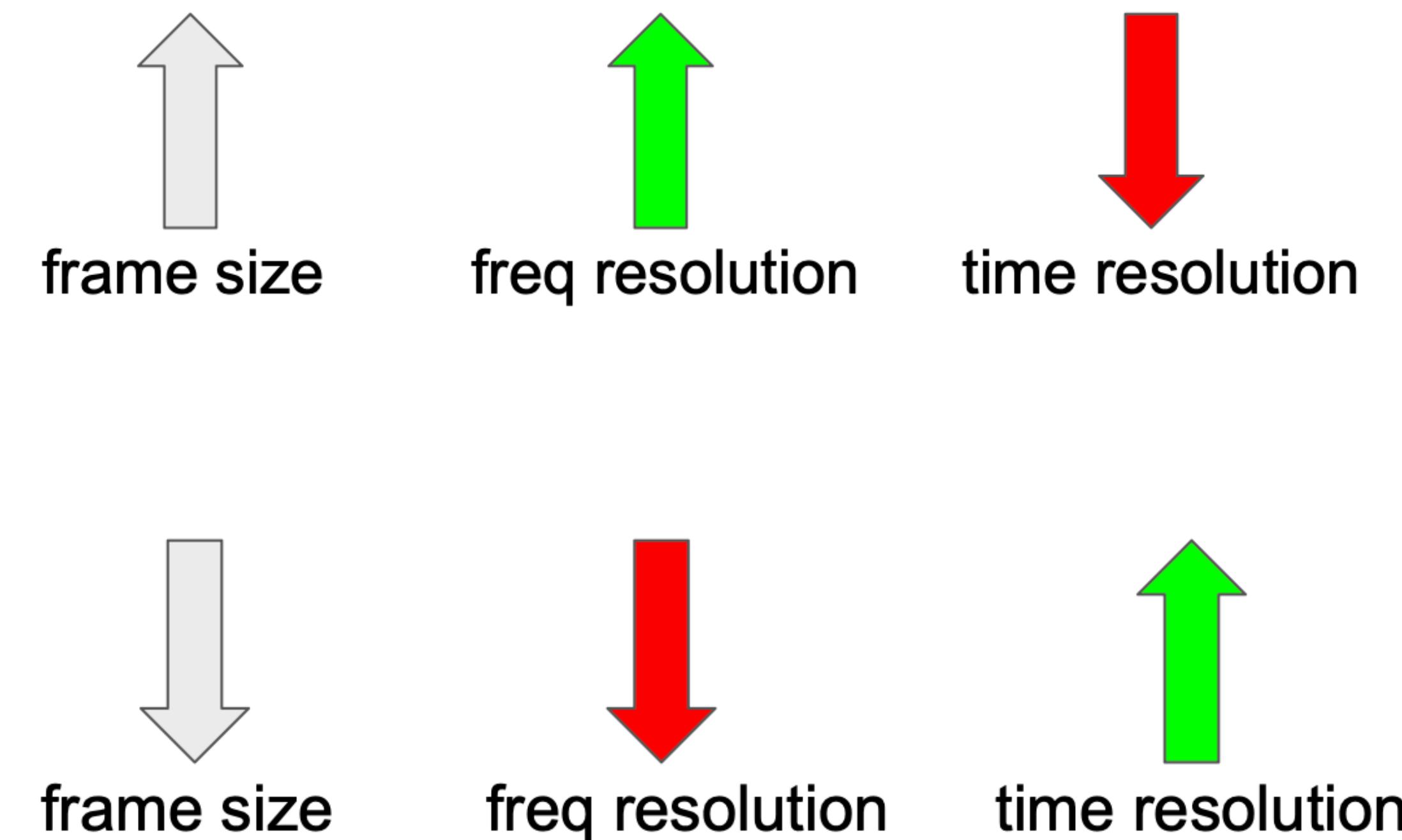
frame size $\in \{512, 1024, 2048, 4096, 8192\}$

Các bạn hãy giải thích tại sao?

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Các yếu tố cần cân đối khi xây dựng STFT



Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Lưu ý 2: Kích thước các Hợp

frame size $\in \{512, 1024, 2048, 4096, 8192\}$

256, 512, 1024, 2048, 4096

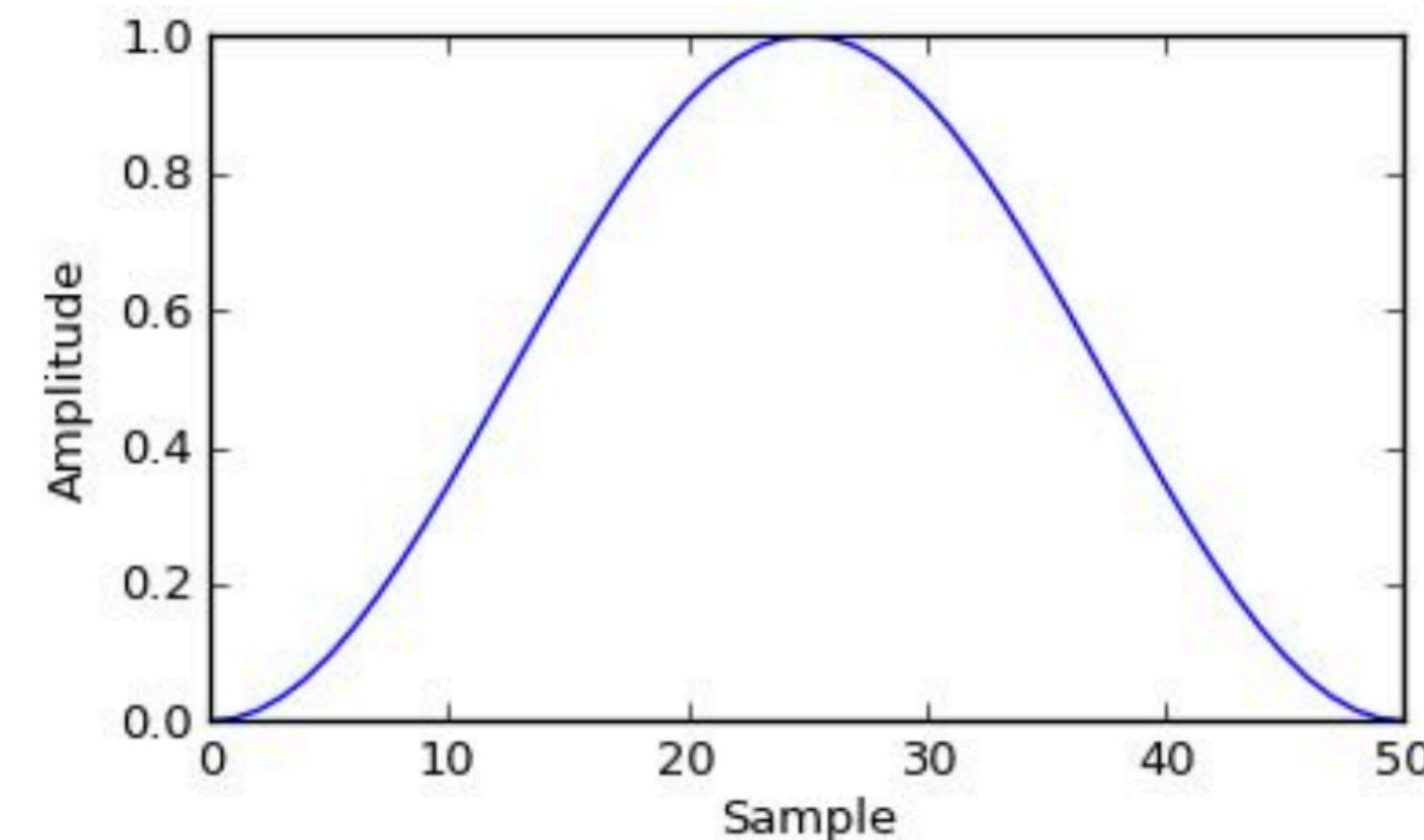
$\frac{1}{2}K, \frac{1}{4}K, \frac{1}{8}K$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

- Lưu ý 3: phải sử dụng Hanning window

$$w(k) = 0.5 \cdot (1 - \cos\left(\frac{2\pi k}{K-1}\right)), k = 1 \dots K$$



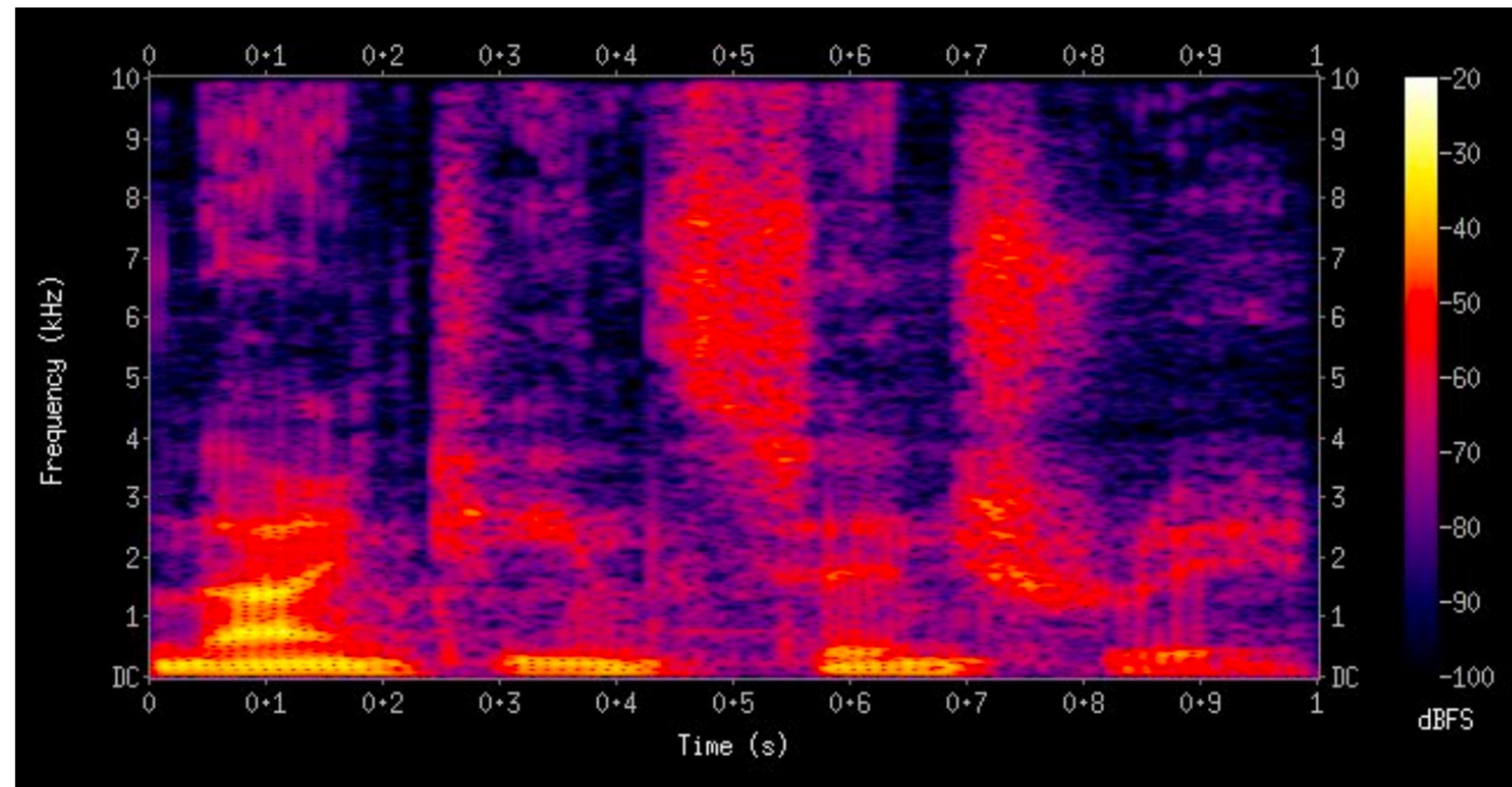
Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)

$$Y(m, k) = |S(m, k)|^2$$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Chuyển đổi STFT (Short-Time Fourier Transform)



Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

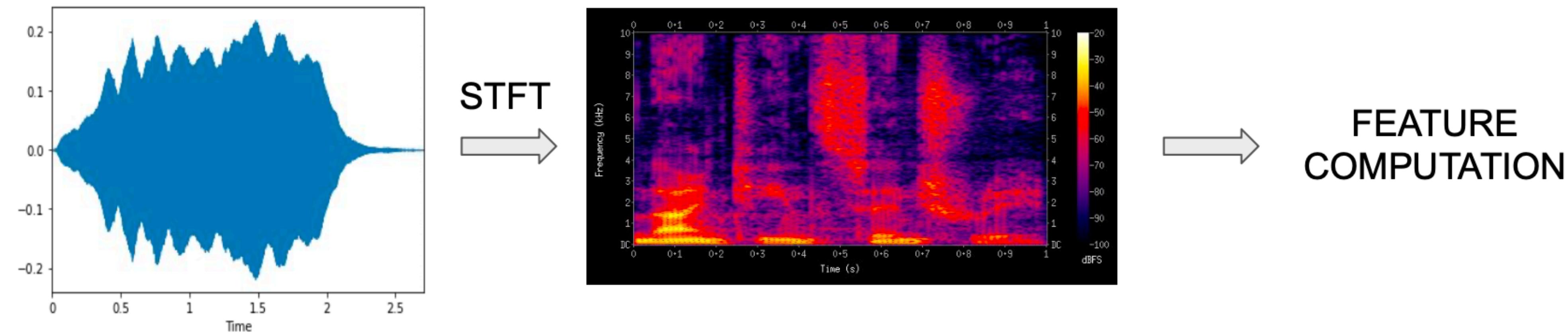
Band Energy Ratio (BER)

$m_t(t) \rightarrow$ biên độ tại bin tần số n và frame t

$N \rightarrow$ #số bin tần số

BER : so sánh năng lượng ở băng tần thấp hay cao

Đo tần số thấp chiếm ưu thế hay tần số cao chiếm ưu thế

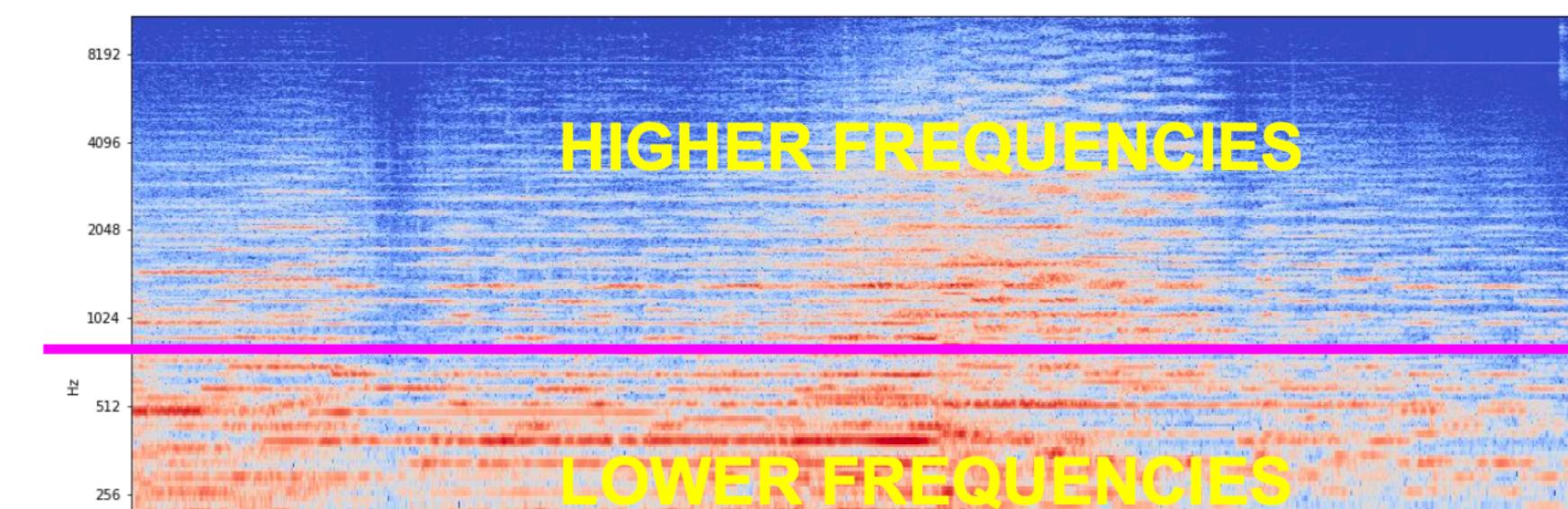


Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Band Energy Ratio (BER)

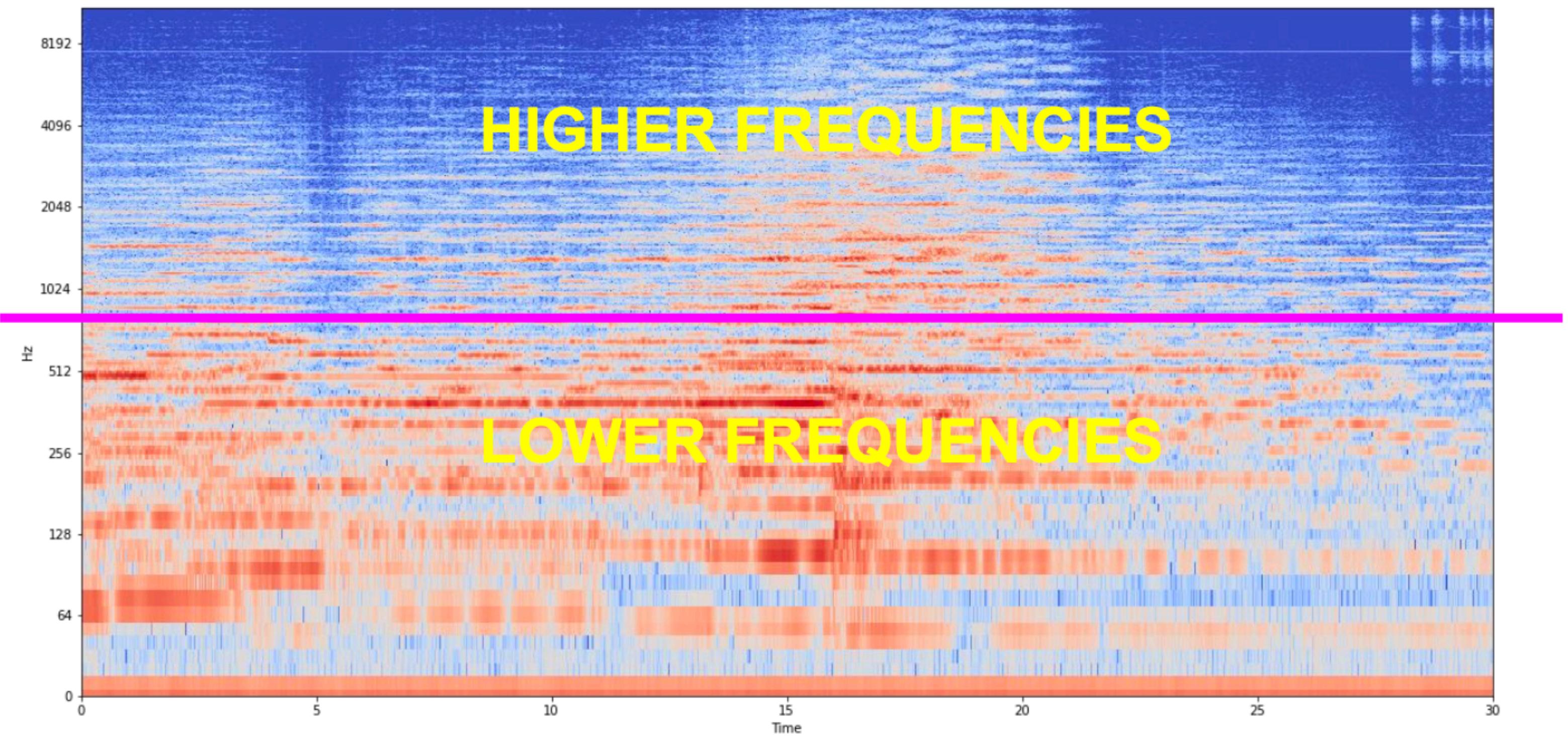
$$BER_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^N m_t(n)^2}$$

F
Split frequency
Power at t, n



Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Band Energy Ratio (BER)



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

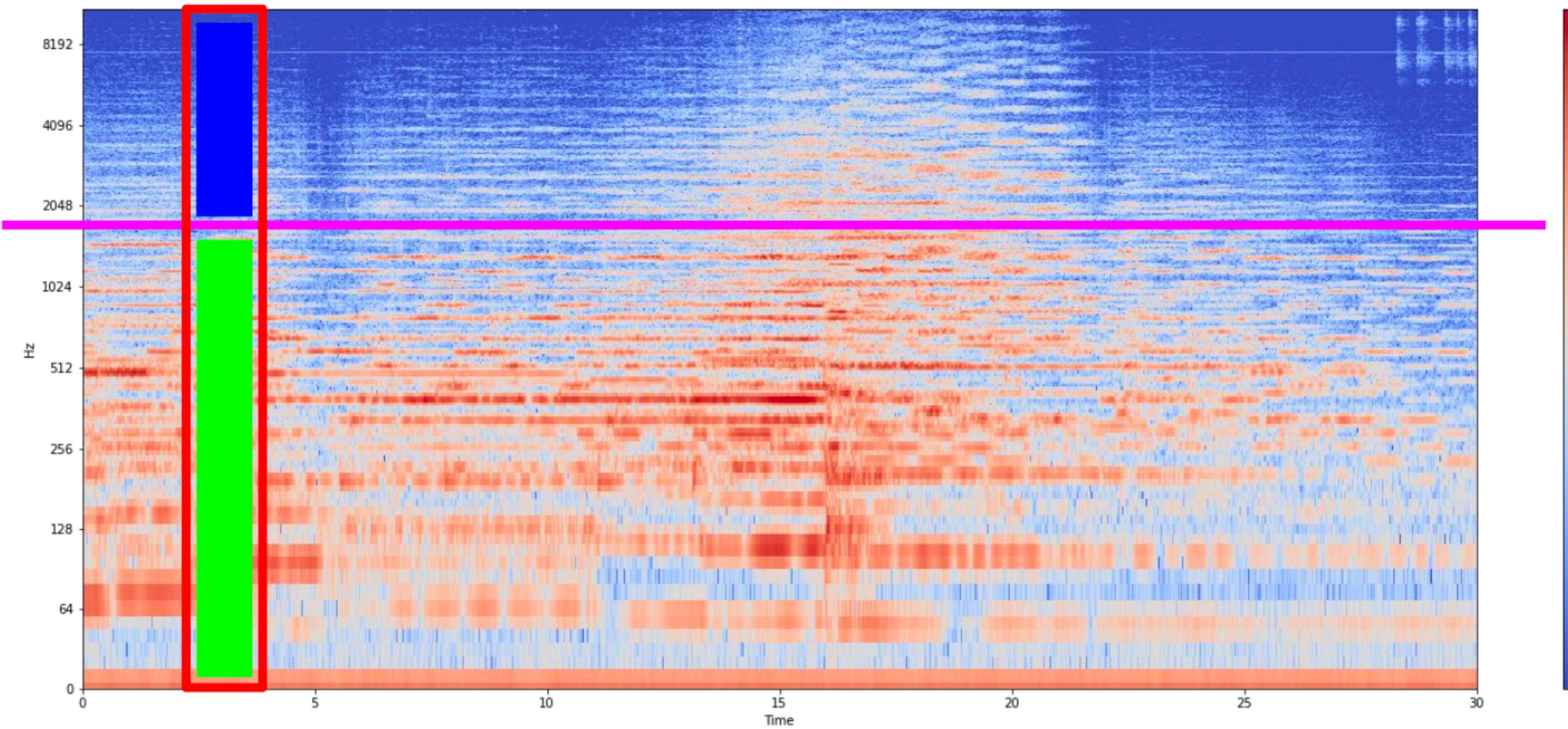
Power in the lower frequency bands

$$\sum_{n=F}^N m_t(n)^2$$

Power in the higher frequency bands

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Band Energy Ratio (BER)



Power in the lower frequency bands

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

Power in the higher frequency bands

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Band Energy Ratio (BER)

- Ứng dụng của BER:
 - Phân tách nhạc và thoại
 - Phân loại âm nhạc (ví dụ phân loại genre)

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Tâm phổ (spectral centroid)

- Spectral Centroid:
 - Là trọng số trung tâm của phổ biên độ (amplitude spectrum)
 - Là băng tần chiếm năng lượng nhiều nhất
 - Đo độ sáng (chói) của âm thanh (brightness)
 - Là trung bình trọng số của các tần số

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Tâm phẩ (spectral centroid)

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

Weight for n

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Tâm phổ (spectral centroid)

- Ứng dụng của Spectral Centroid:
 - Phân loại Audio
 - Phân loại Music

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Bảng thông

- Bảng thông:
 - Được trích xuất từ tâm phổi
 - Là dải phổi xung quanh tâm phổi
 - Các thay đổi so với tâm phổi
 - Có tác dụng mô tả cảm nhận âm sắc

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Bảng thông

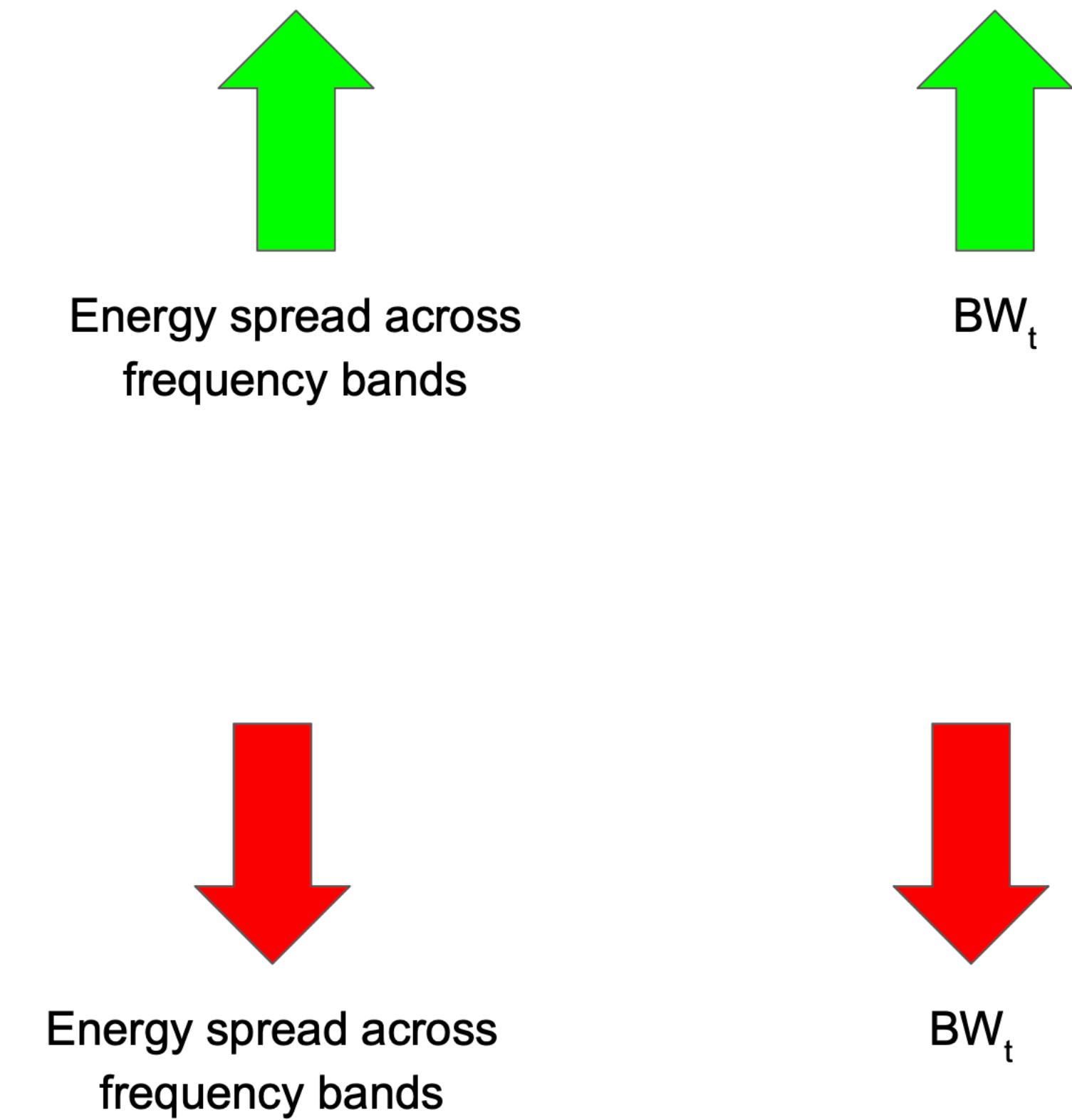
- Là trung bình các trọng số của khoảng cách so với tâm phẩ

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

Distance of frequency band from
 spectral centroid Weight for n
 $|n - SC_t|$ $m_t(n)$

Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Bảng thông



Trích xuất một số thuộc tính của tín hiệu Audio miền tần số

Băng thông

- Ứng dụng của băng thông
 - Xử lý music

Quiz và Lab

Nội dung

- Thực hành STFT
- Thực hành xác định BER
- Thực hành xác định tâm phổ (Spectral Centroid)
- Thực hành xác định băng thông
- Thực hiện phân loại tín hiệu trên tập dataset bằng machine-learning.