



**FPT SCHOOL OF BUSINESS
& TECHNOLOGY**

Digital Signal Processing

Các thuộc tính của tín hiệu Audio

Phd. Trần Thanh Trúc

Lý thuyết: Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Nội dung

- Mel-spectrograms
- Mel-Frequency Cepstral

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Thử nghiệm cảm nhận tâm lý về âm thanh

- Mẫu thứ 1: C2 - C4 (65 - 262 Hz)
- Mẫu thứ 2: G6 - A6 (1568 - 1765 Hz)

200 Hz

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-spectrograms > Thú nghiệm cảm nhận tâm lý về âm thanh

Con người cảm nhận tần số theo bậc logarithm

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Đặc tính audio lý tưởng

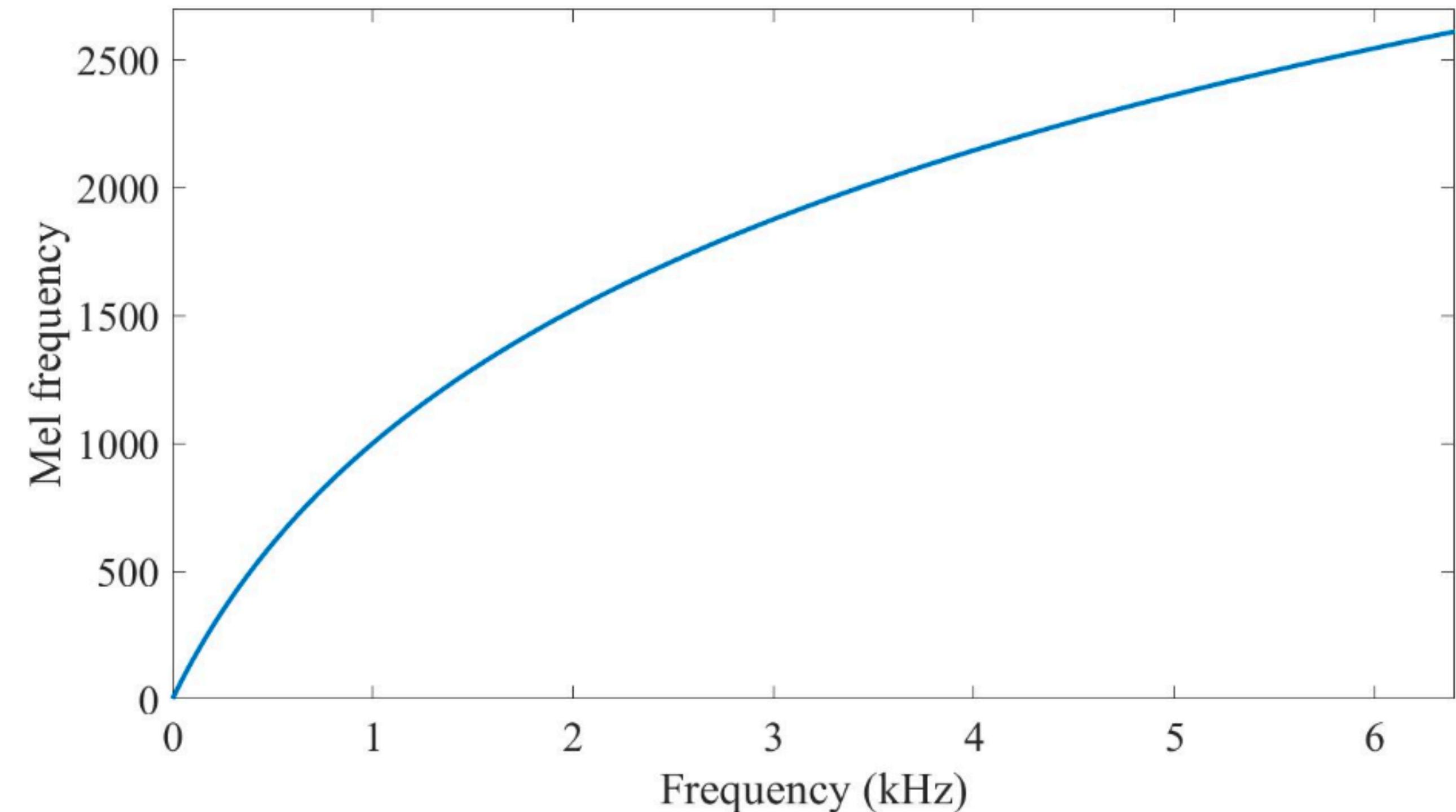
- Có thể biểu diễn dạng Time-Frequency
- Biểu diễn biên độ theo cảm nhận âm
- Biểu diễn tần số theo cảm nhận âm

Mel spectrograms

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

- Mel-scale:
 - Thang logarithm
 - Khoảng cách giữa 2 tần số Mel, thể hiện khoảng cách về cảm nhận
 - $1000 \text{ Hz} = 1000 \text{ Mel}$

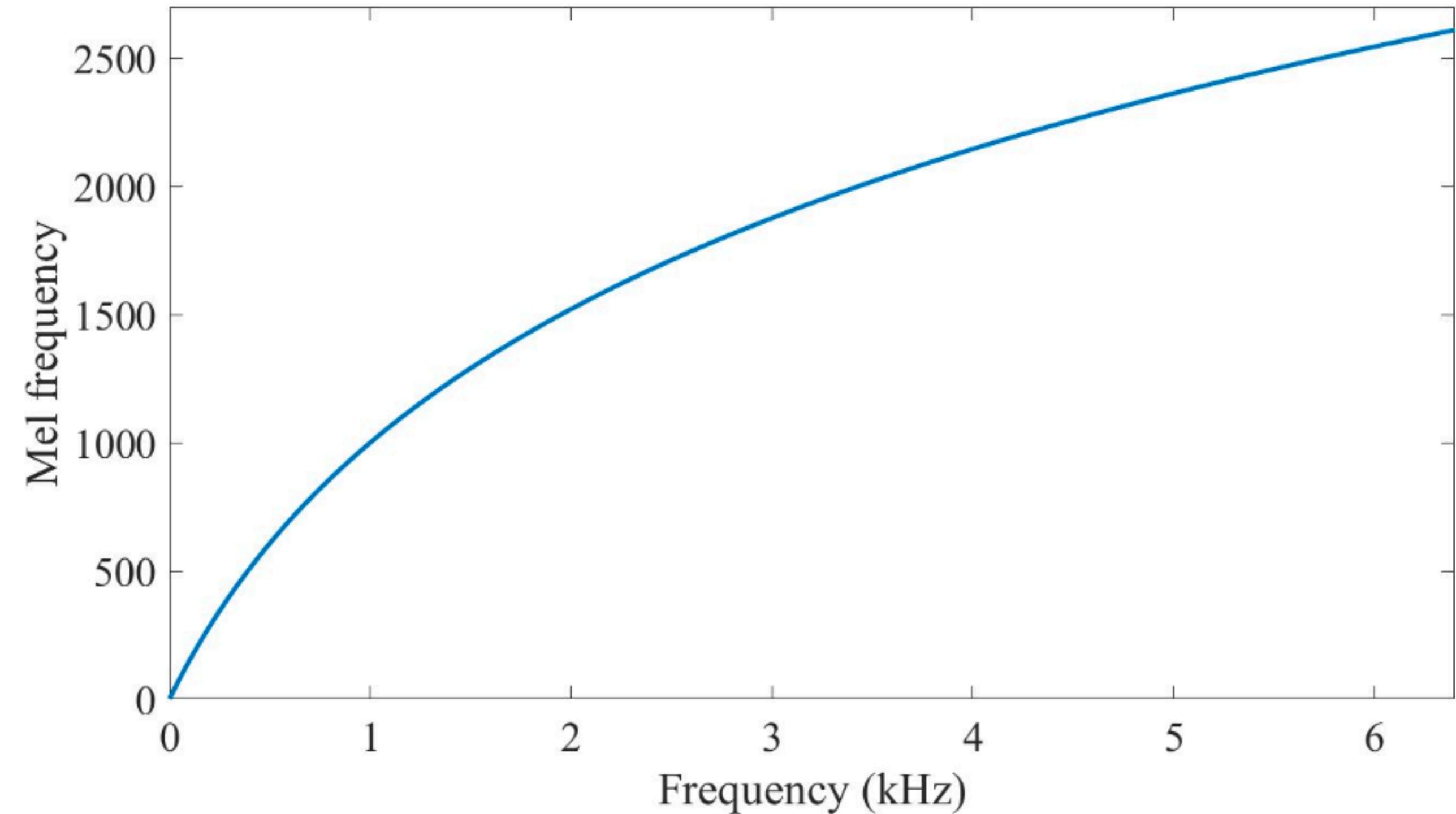


Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

$$m = 2595 \cdot \log\left(1 + \frac{f}{500}\right)$$

$$f = 700\left(10^{m/2595} - 1\right)$$



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

- Trích xuất biểu đồ Mel-Spectrogram
 1. Chuyển đổi STFT
 2. Chuyển biến độ sang thang dB
 3. Chuyển tần số sang tần số Mel

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

- Chuyển tần số sang tần số Mel
 - Chọn số lượng băng Mels
 - Lựa chọn dãy các bộ lọc Mels
 - Áp dụng bộ lọc Mel vào trong spectrogram

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

- Số lượng băng Mels

40 128
 60 90

Số lượng băng Mels phụ thuộc vào nhu cầu và bài toán cụ thể.

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

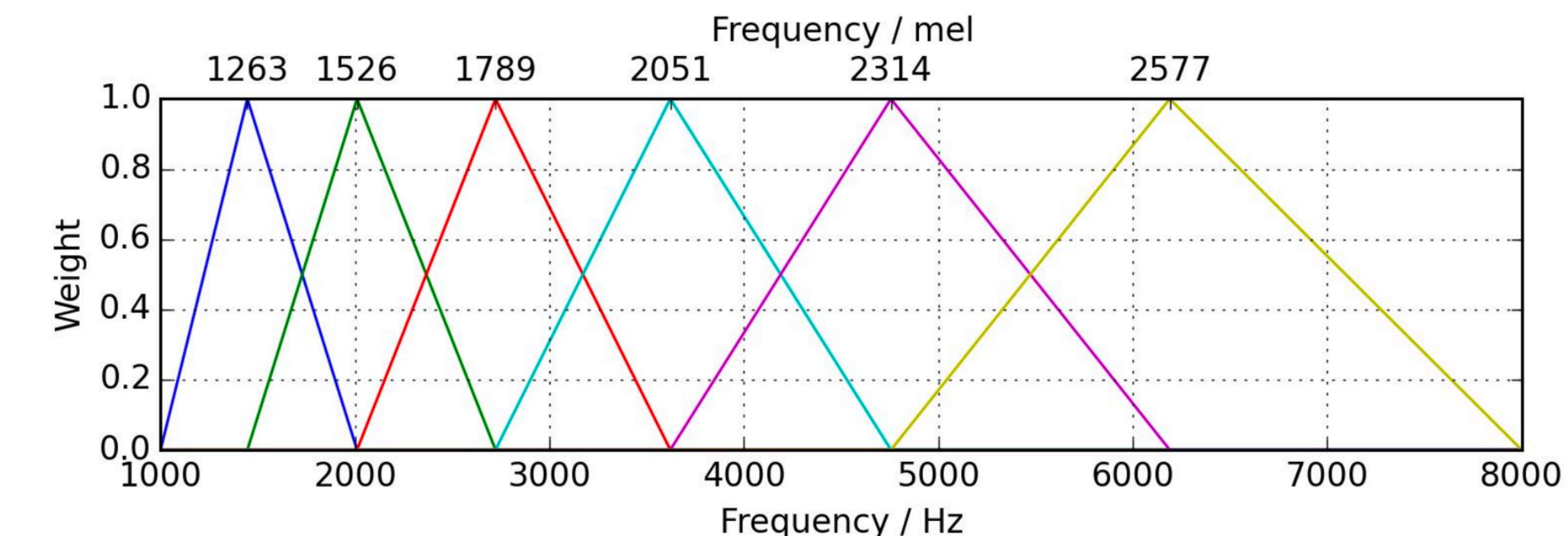
Mel-Spectrogram

- Tạo dãy các bộ lọc Mels
 - Chuyển đổi tần số thấp nhất/ cao nhất sang tần số Mel
 - Khởi tạo số #bands đều đặn
 - Chuyển ngược thành tần số Hz
 - Làm tròn đến bin tần số gần nhất
 - Tạo các bộ lọc hình tam giác

$$m = 2595 \cdot \log \left(1 + \frac{f}{500} \right)$$



$$f = 700 \cdot \left(10^{\frac{m}{2595}} - 1 \right)$$

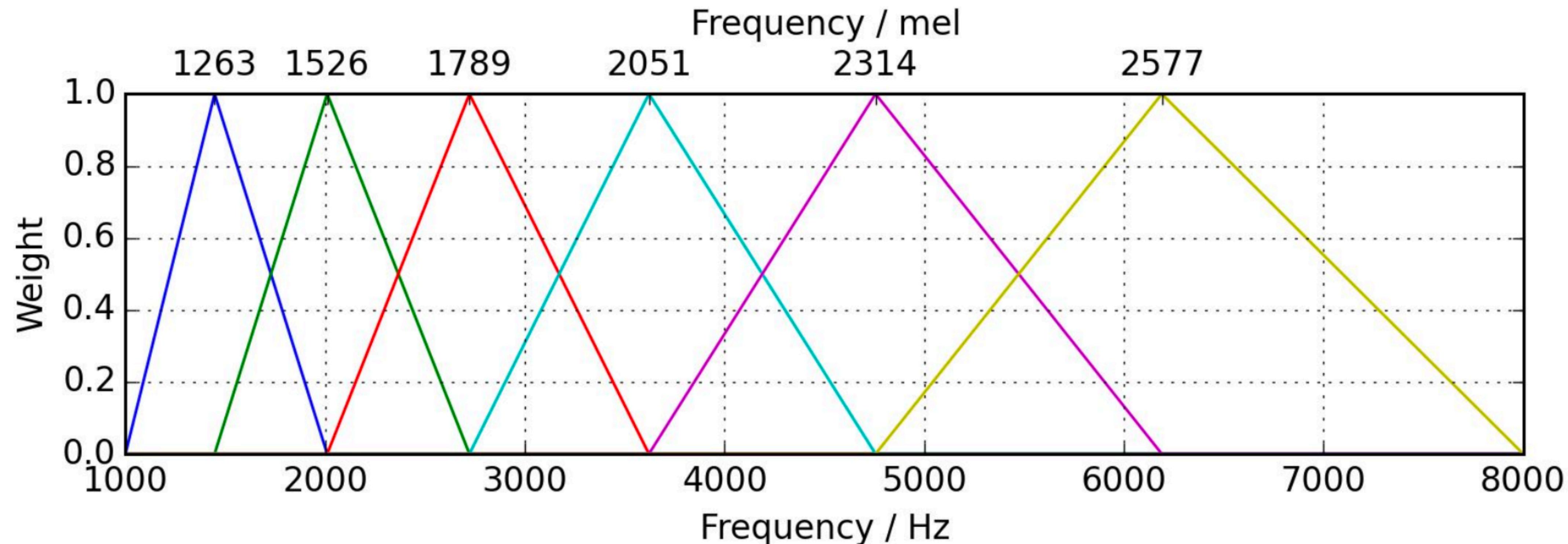


Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

$$M = \left(\#bands, \frac{\text{framesize}}{2} + 1 \right)$$

$$Y = \left(\frac{\text{framesize}}{2} + 1, \#frame \right)$$



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

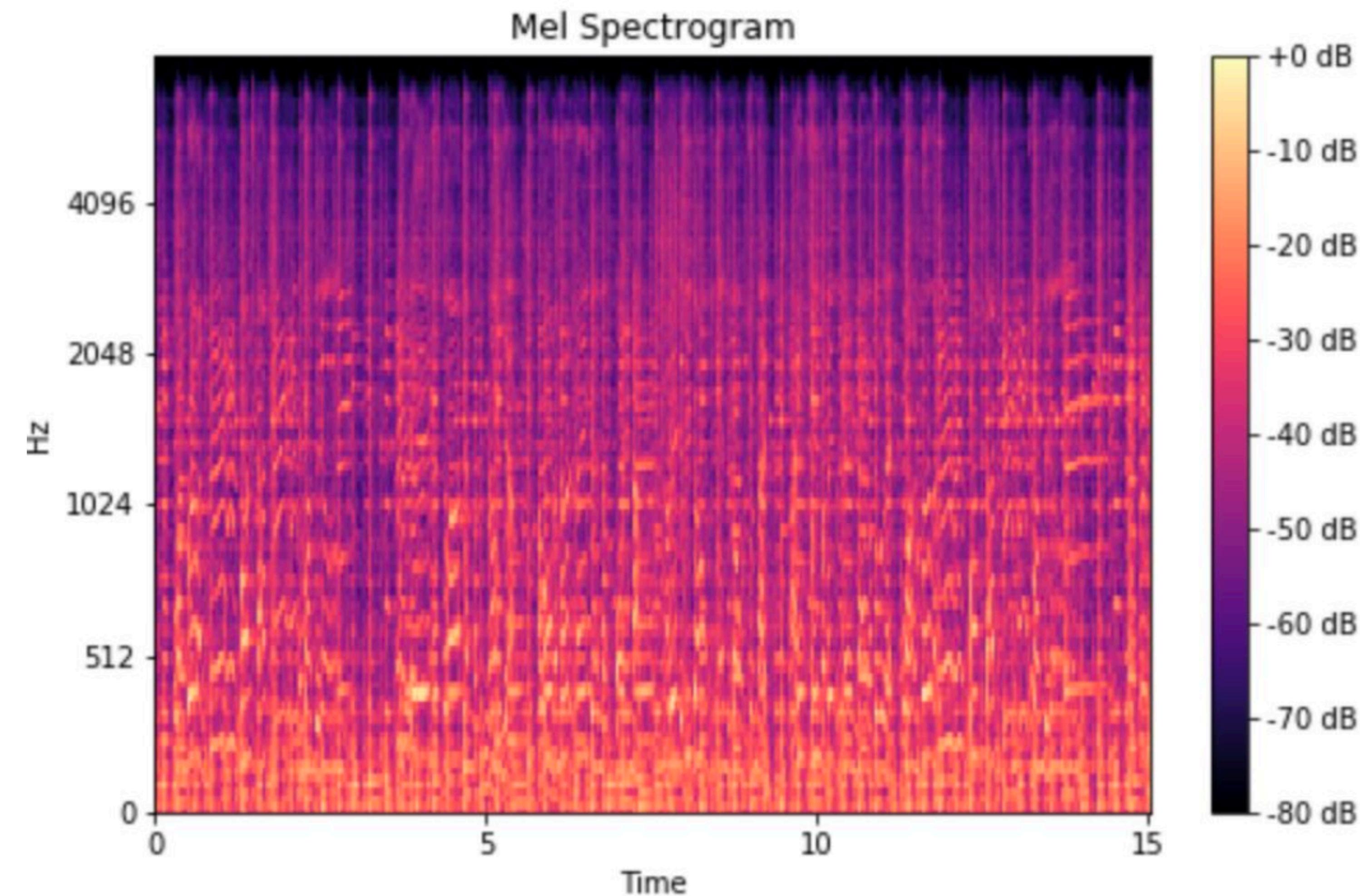
Mel-Spectrogram

Mel spectrogram = M \mathbf{Y}

(# bands, # frames)

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Spectrogram

- Ứng dụng của Mel-Spectrogram
 - Phân loại audio
 - Nhận dạng mood tự động
 - Phân loại thể loại âm nhạc
 - Phân loại nhạc cụ
 - ...

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral Mel-Frequency Cepstral

Cepstrum
 Spectrum

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Lịch sử của Cepstrum
 - Được hình thành trong quá trình nghiên cứu các tín hiệu rung chấn (seismic) từ thập kỷ 1960
 - Được ứng dụng vào làm đặc trưng của audio để nhận dạng giọng nói từ từ thập kỷ 1970
 - Được ứng dụng để xử lý âm nhạc từ thập kỷ 2000

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Định nghĩa Cepstrum

Time-domain signal

$$C(x(t)) = F^{-1} [\log(F[x(t)])]$$

Spectrum

Log spectrum

Cepstrum

The diagram shows the mathematical definition of the Cepstrum. It consists of a large green rectangular box containing the equation $C(x(t)) = F^{-1} [\log(F[x(t)])]$. Inside this green box, there is a smaller blue rectangular box containing the term $\log(F[x(t)])$, which is labeled "Spectrum". Below the green box, the word "Log spectrum" is written in orange. At the bottom of the green box, the word "Cepstrum" is written in green.

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Định nghĩa Cepstrum

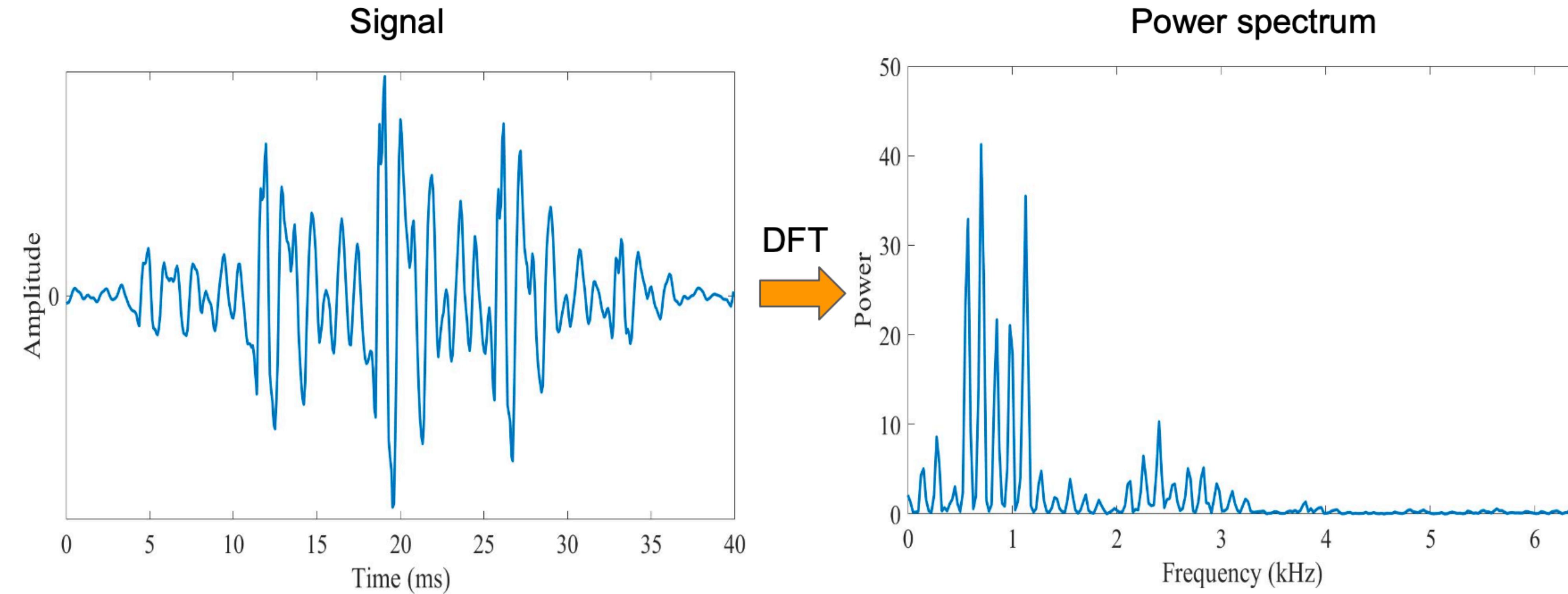


Spectrum
of
a spectrum

Cepstrum

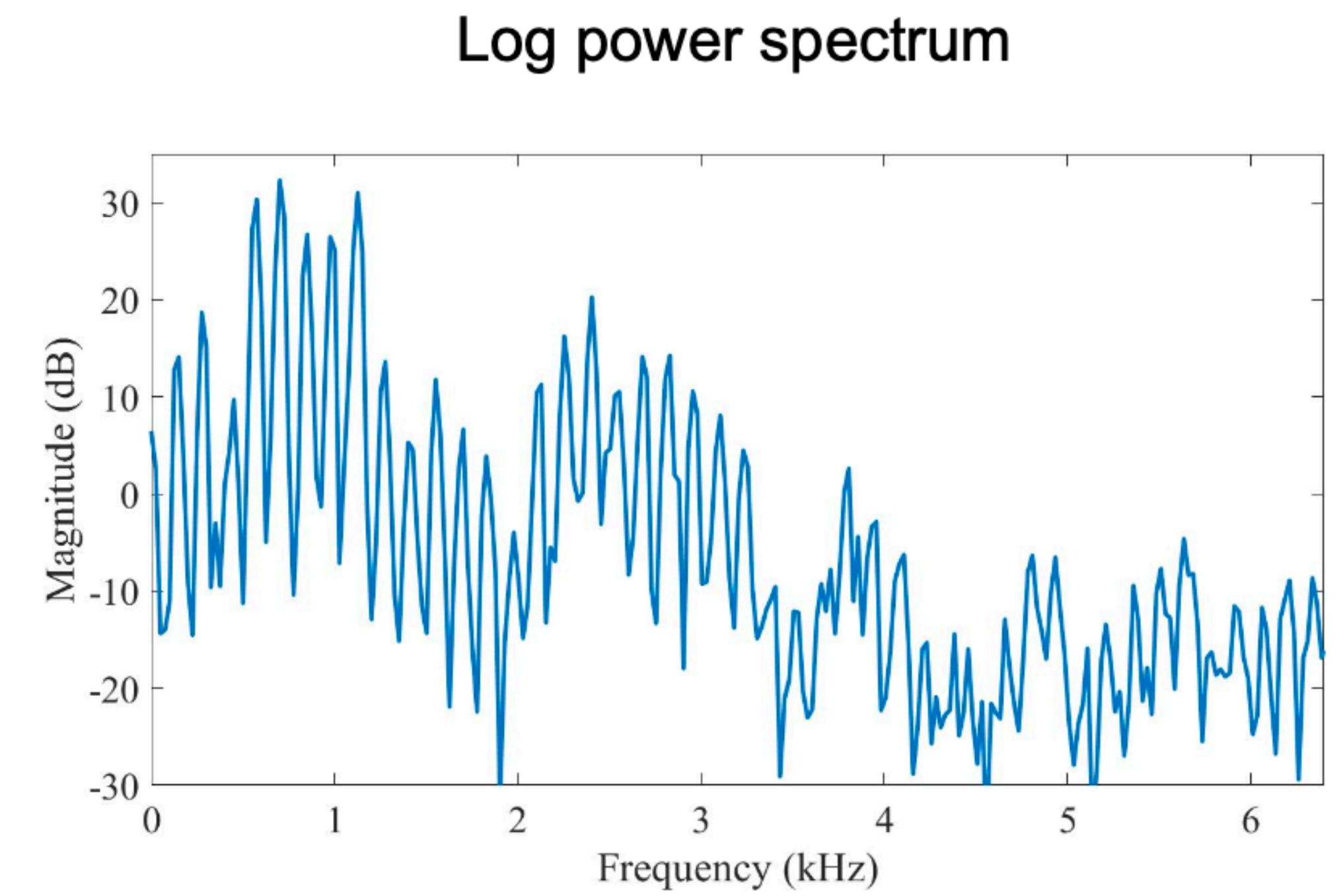
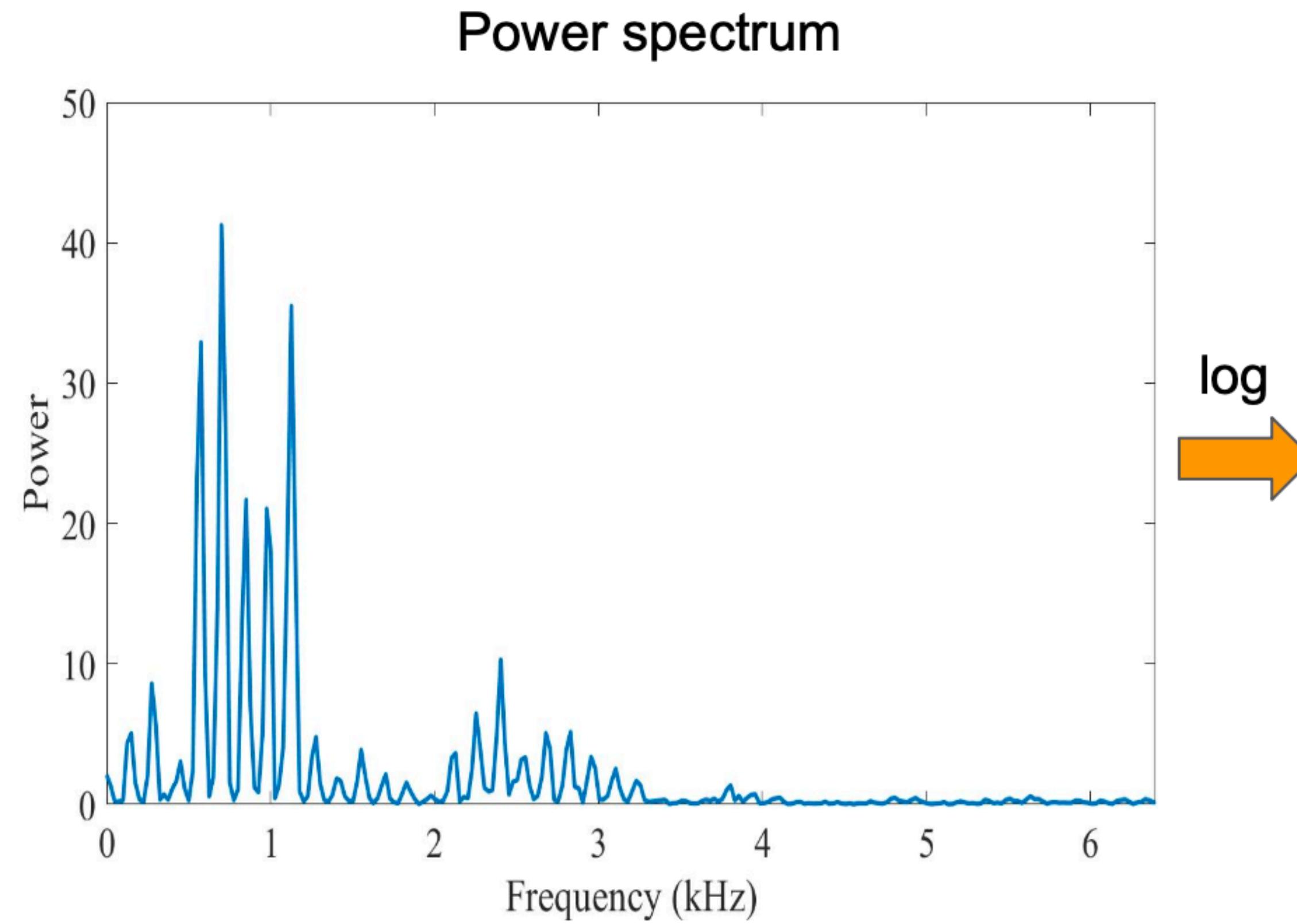
Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

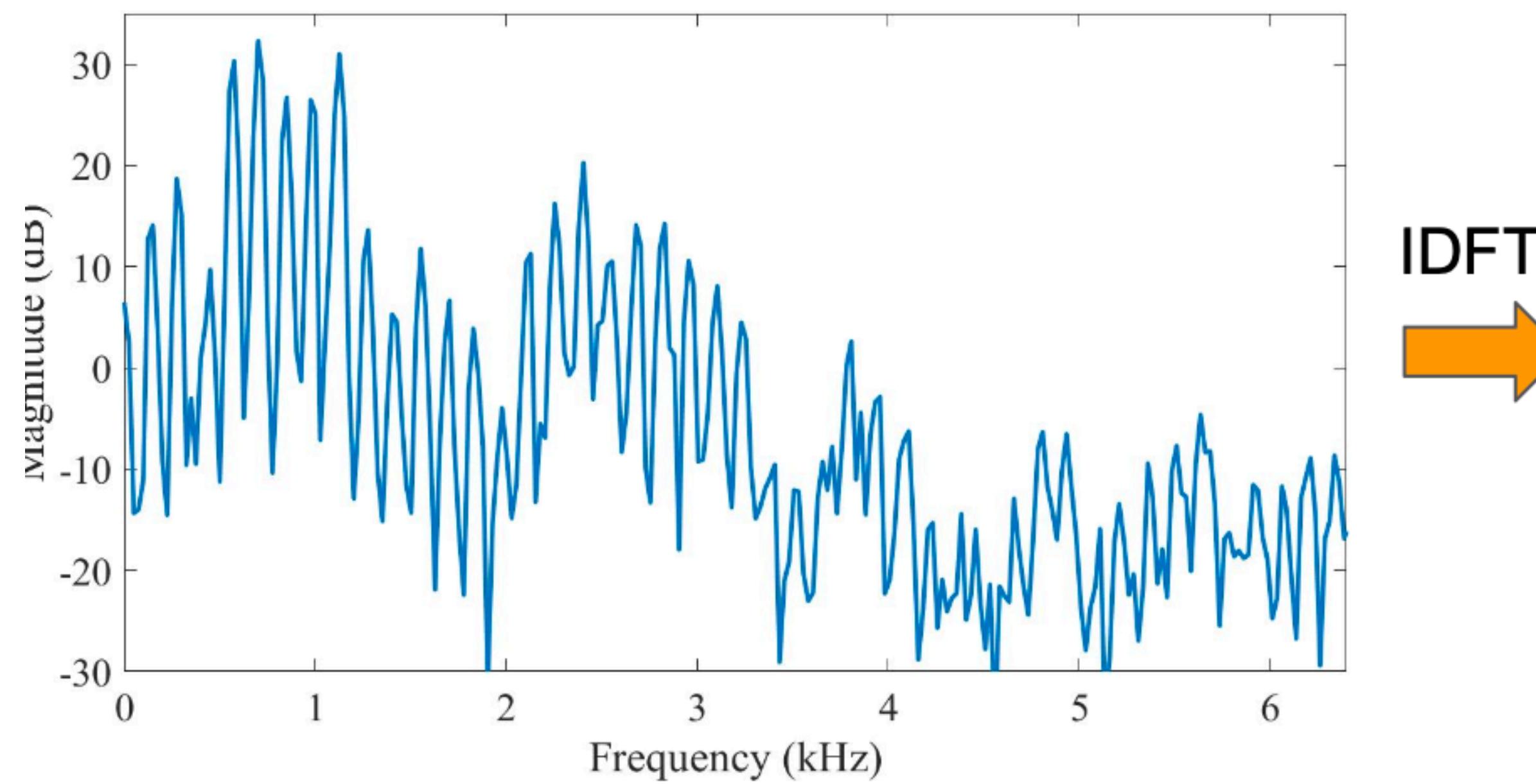
Mel-Frequency Cepstral



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

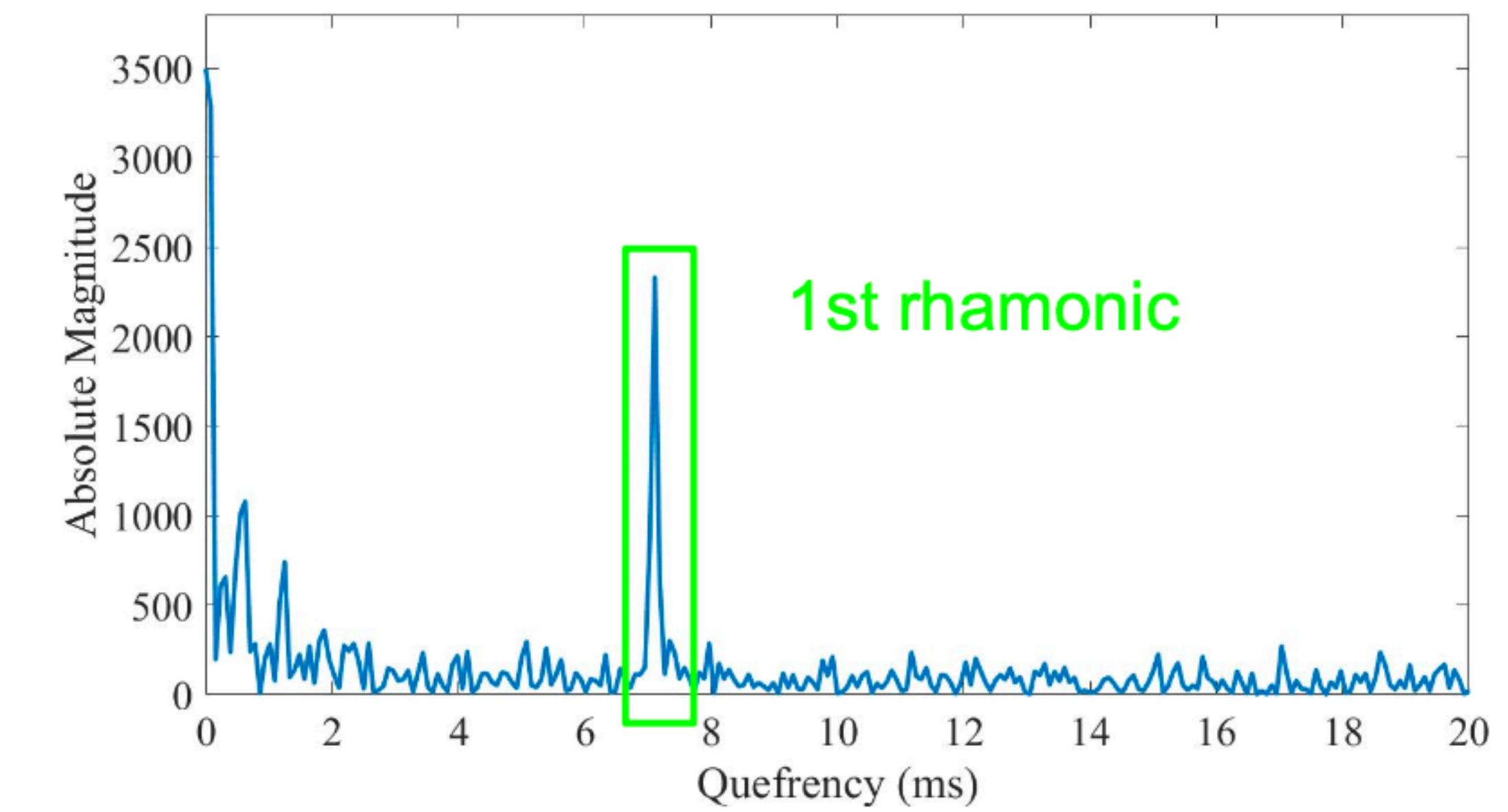
Mel-Frequency Cepstral

Log power spectrum



IDFT
→

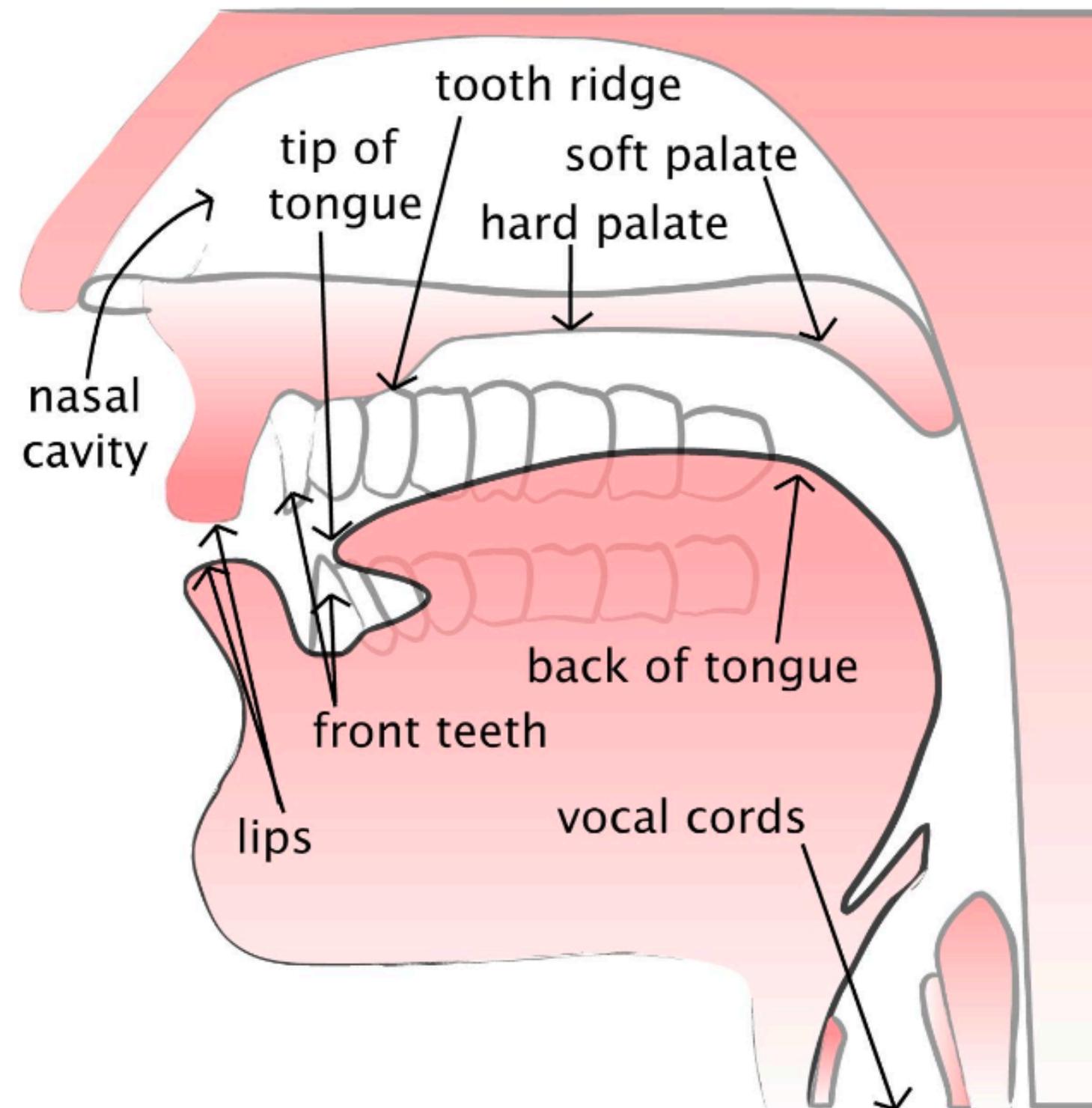
Cepstrum



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum

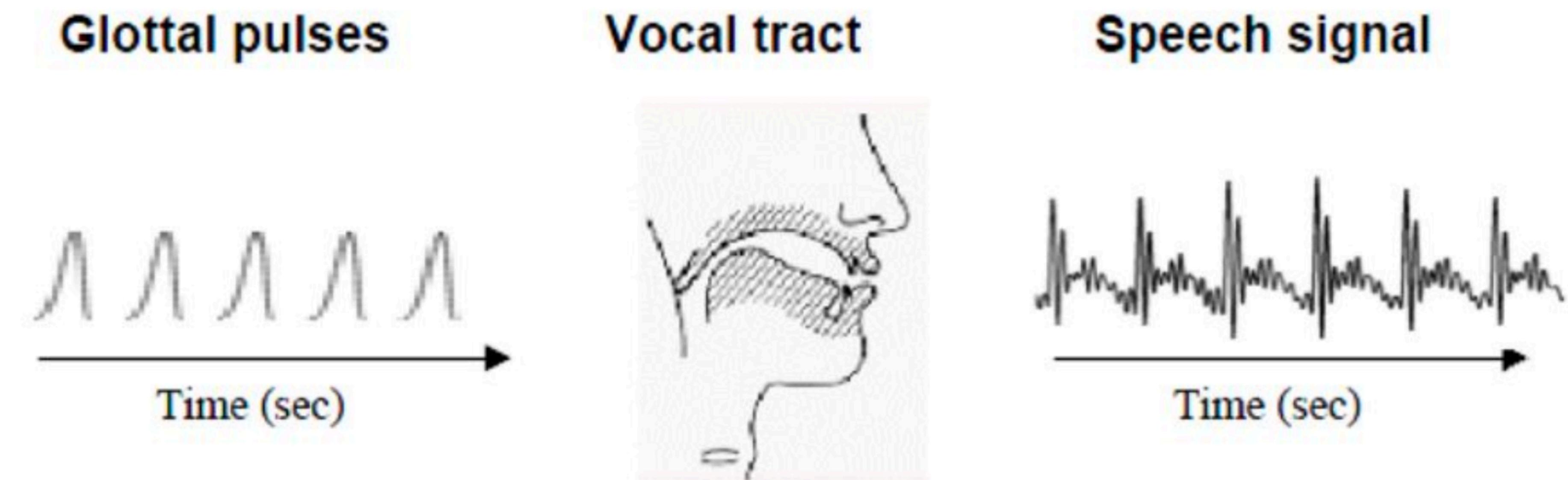


Thanh quản có tác dụng như bộ lọc

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

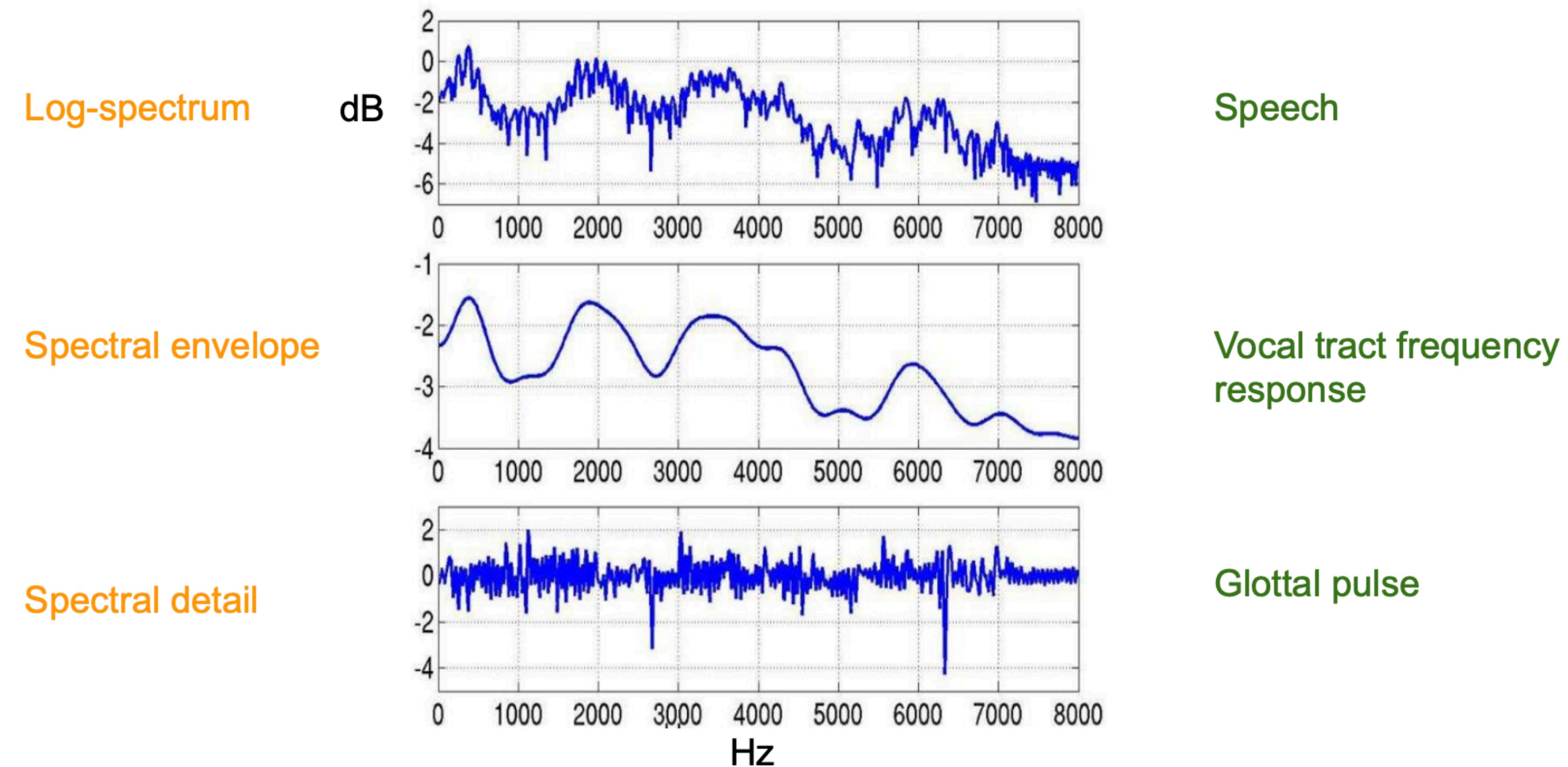
- Ý nghĩa của Cepstrum



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum

Thoại = Phép chập giữa thanh quản và
đo động rung của thanh quản

$$x(t) = e(t) \times h(t)$$

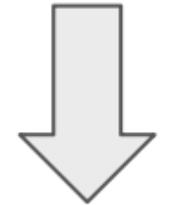
$$X(t) = E(t) \times H(t)$$

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

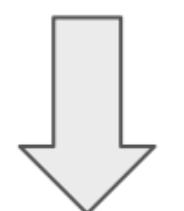
Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum

$$X(t) = E(t) \cdot H(t)$$



$$\log(X(t)) = \log(E(t) \cdot H(t))$$



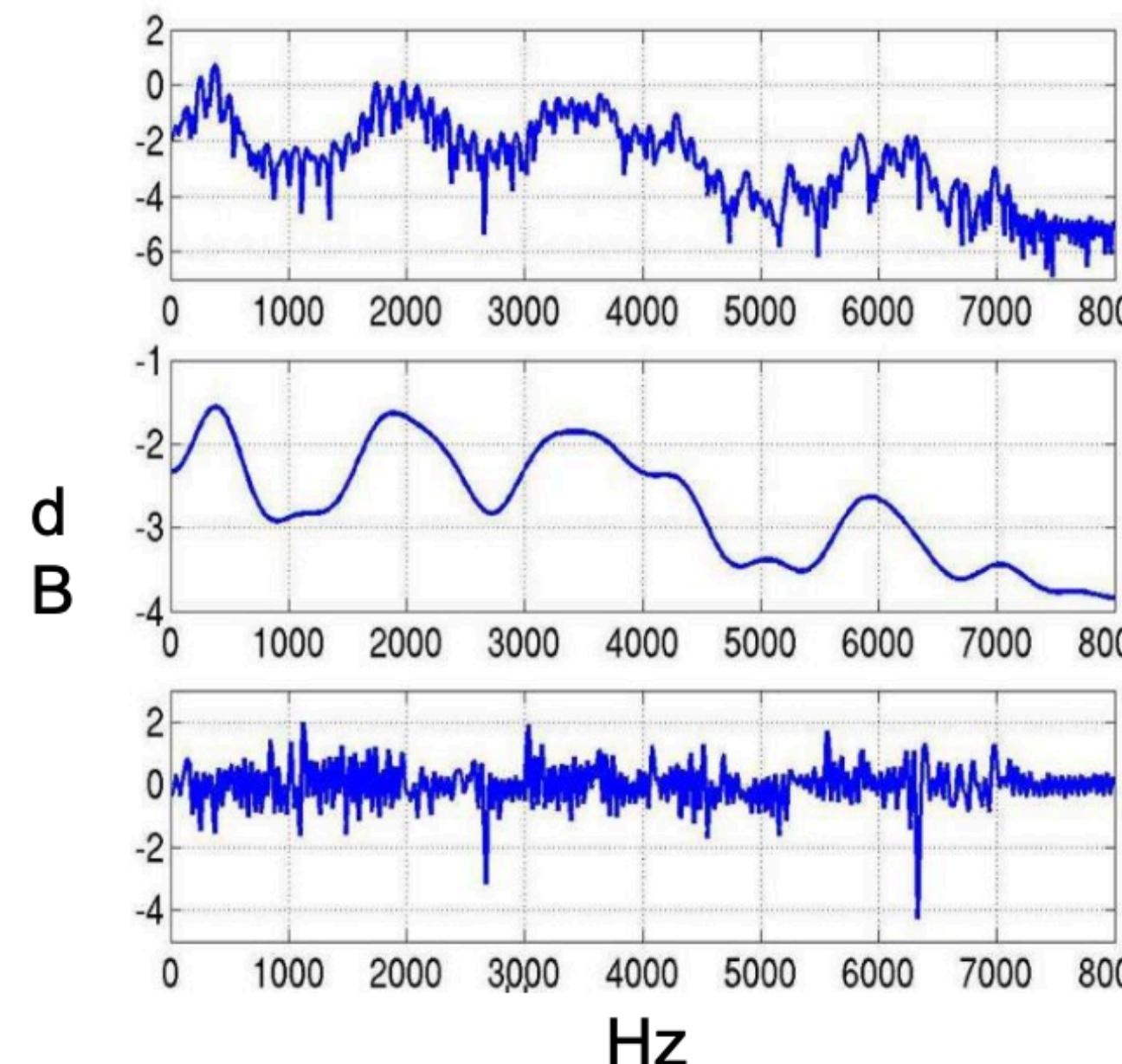
$$\log(X(t)) = \log(E(t)) + \log(H(t))$$

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum

$$\log(X(t)) = \log(E(t)) + \log(H(t))$$



Speech

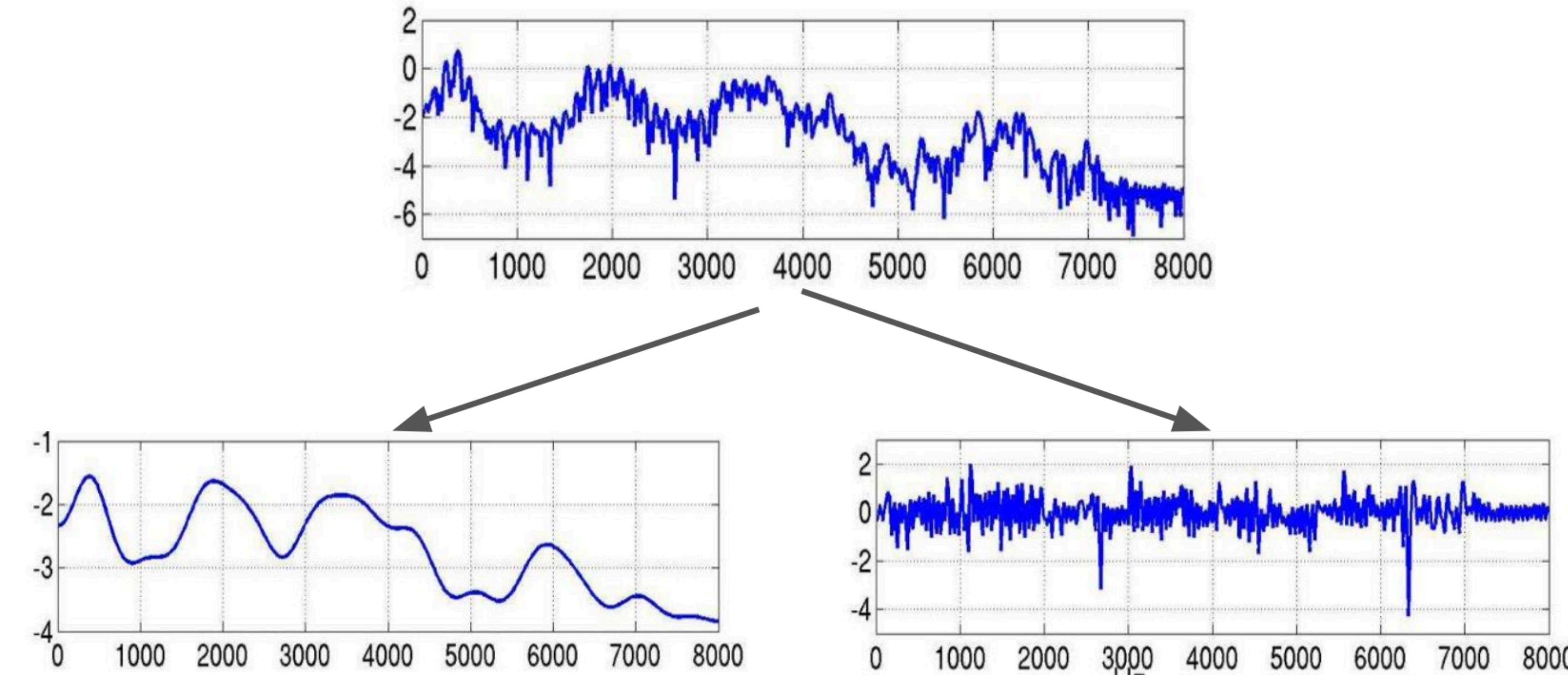
Vocal tract frequency response

Glottal pulse

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

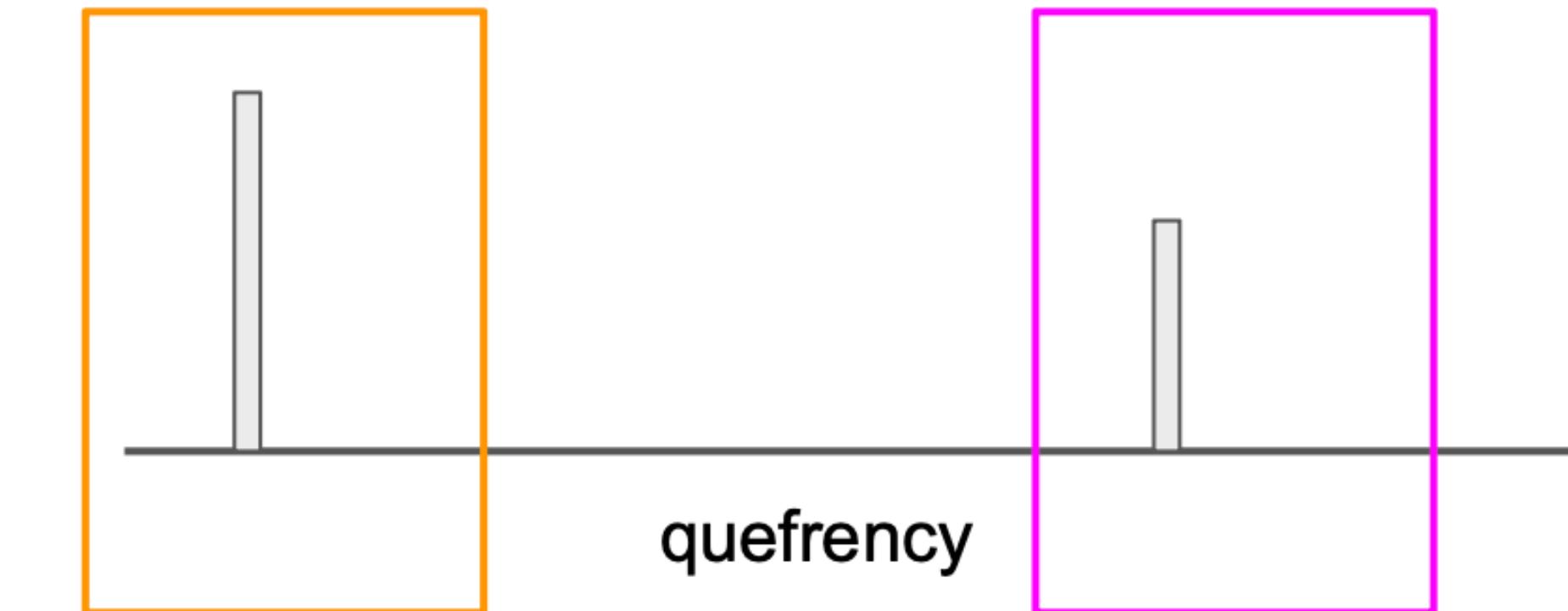
- Ý nghĩa của Cepstrum



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

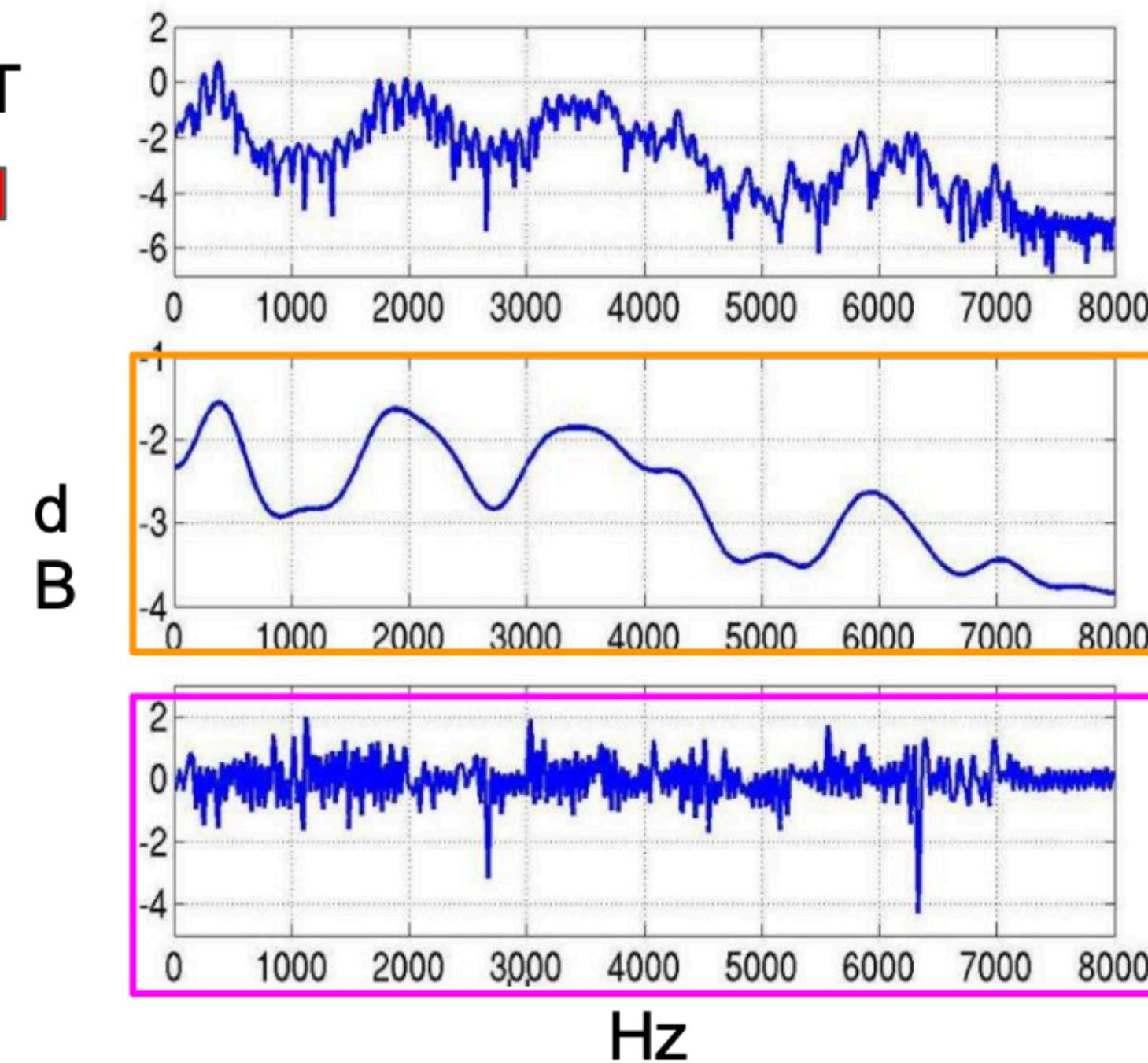
Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum



$$X(t) = \boxed{E(t)} + \boxed{H(t)}$$

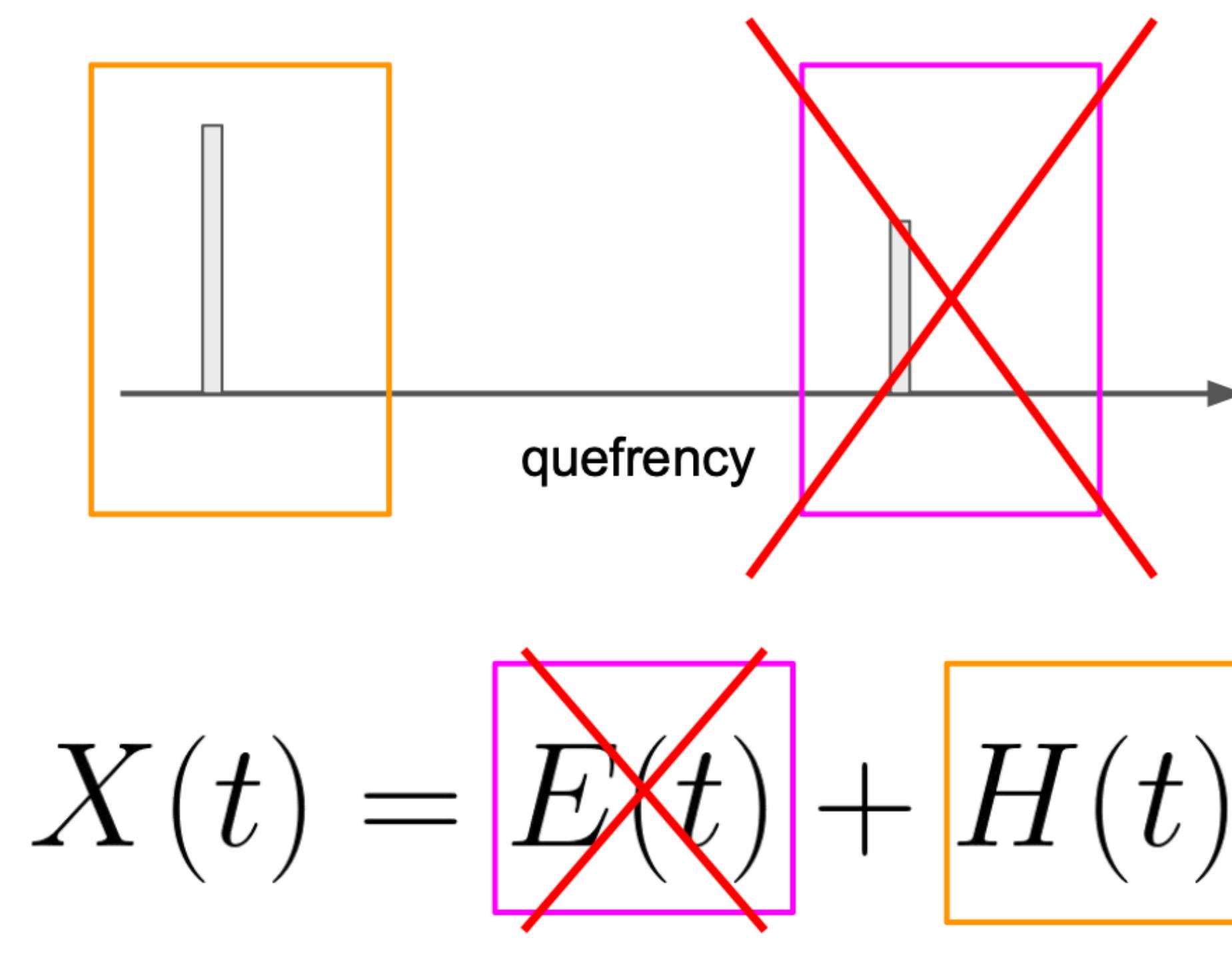
$$\log(X(t)) = \log(E(t)) + \log(H(t))$$



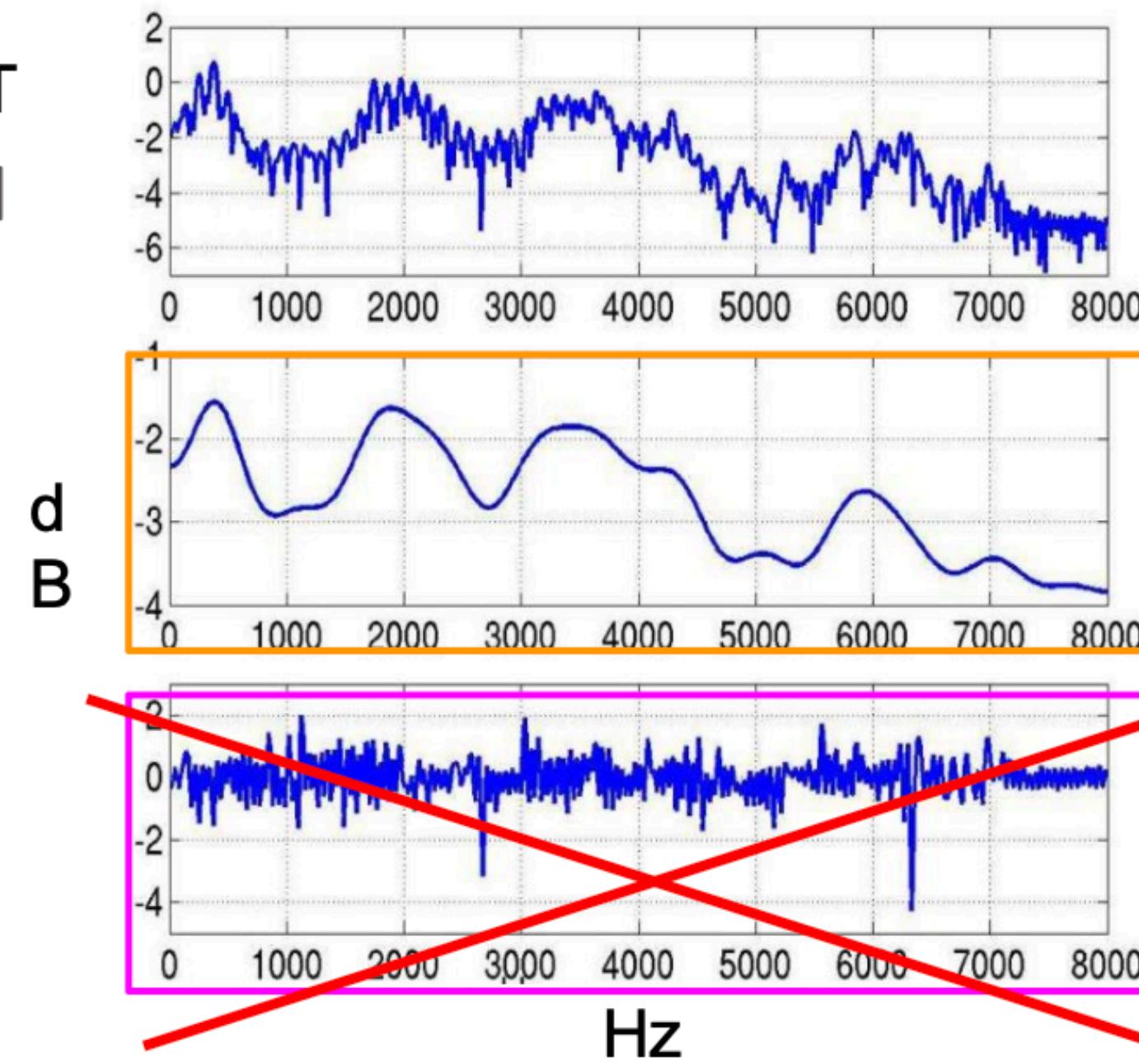
Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ý nghĩa của Cepstrum



$$\log(X(t)) = \log(E(t)) + \log(H(t))$$



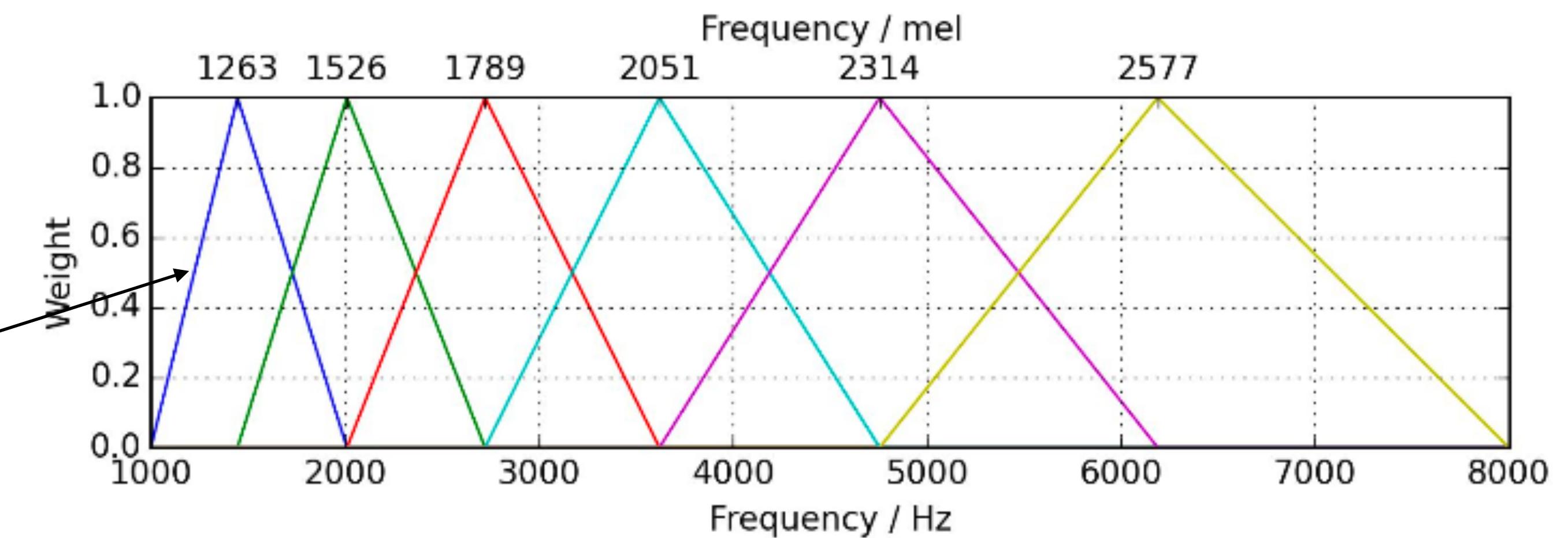
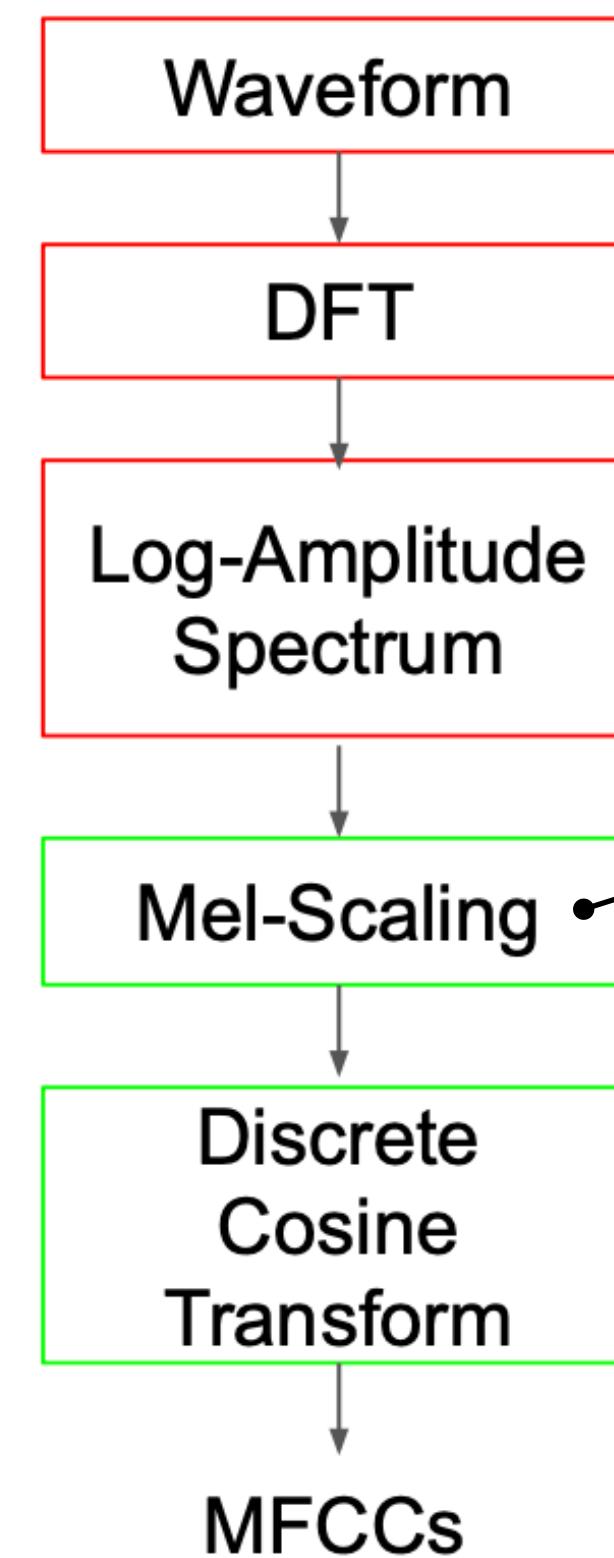
Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral Mel-Frequency Cepstral



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

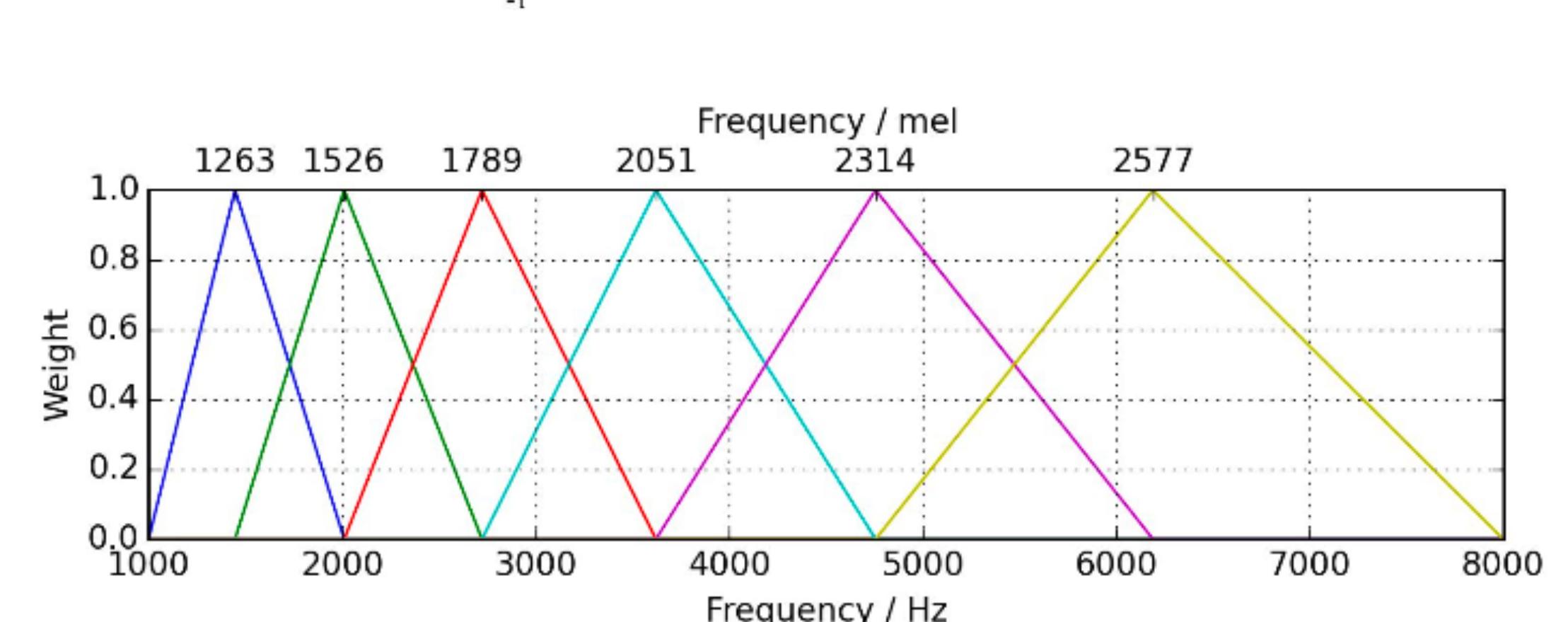
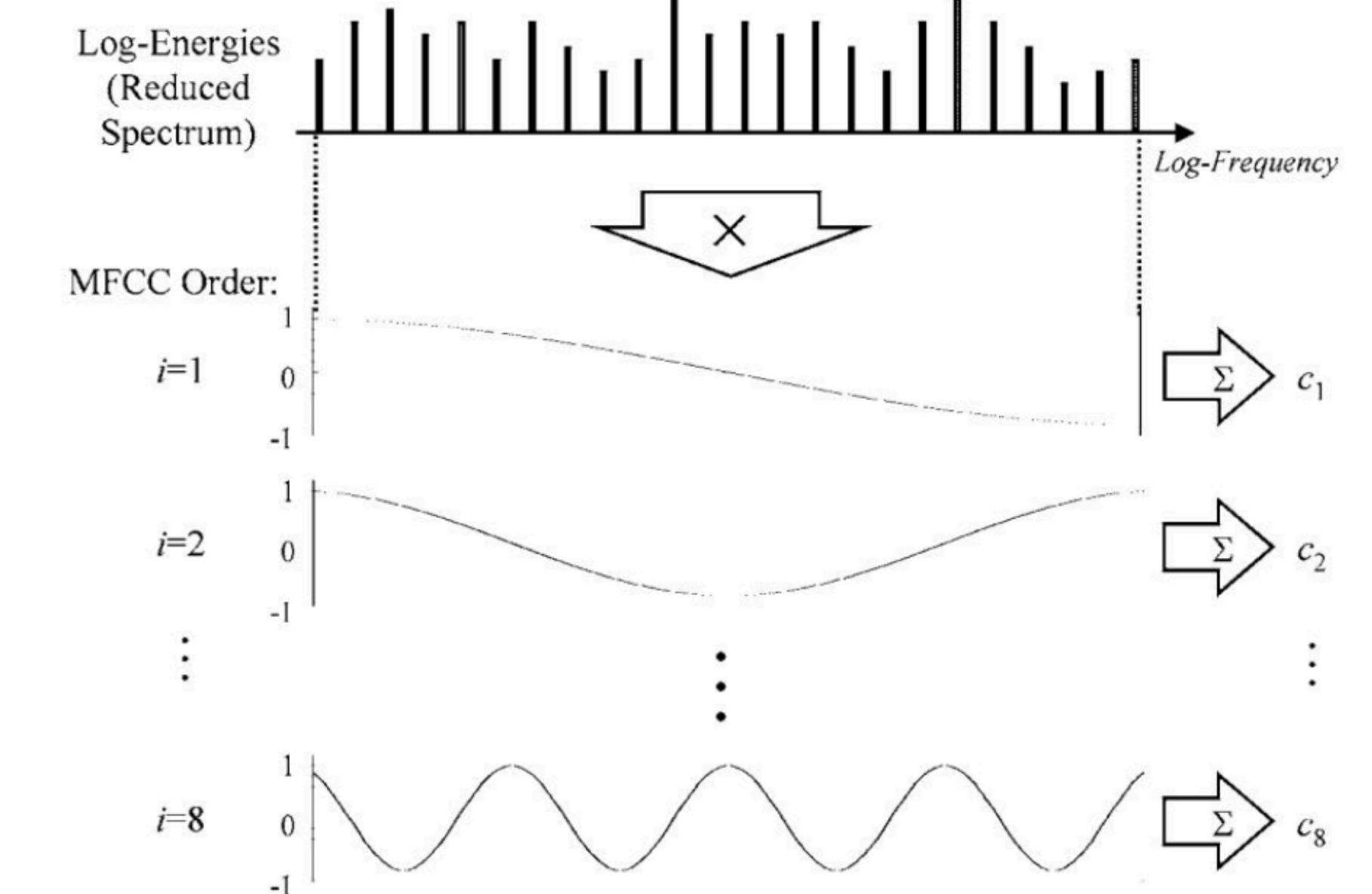
- Trích xuất tín hiệu Cepstrum (MFCC)



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Tại sao phải sử dụng chuyển đổi cosin ròng rạc để lấy giá trị của MFCC?
 - Là phiên bản đơn giản của chuyển đổi Fourier
 - Giá trị trả về là giá trị thực
 - Phân tách năng lượng của các băng Mel khác nhau



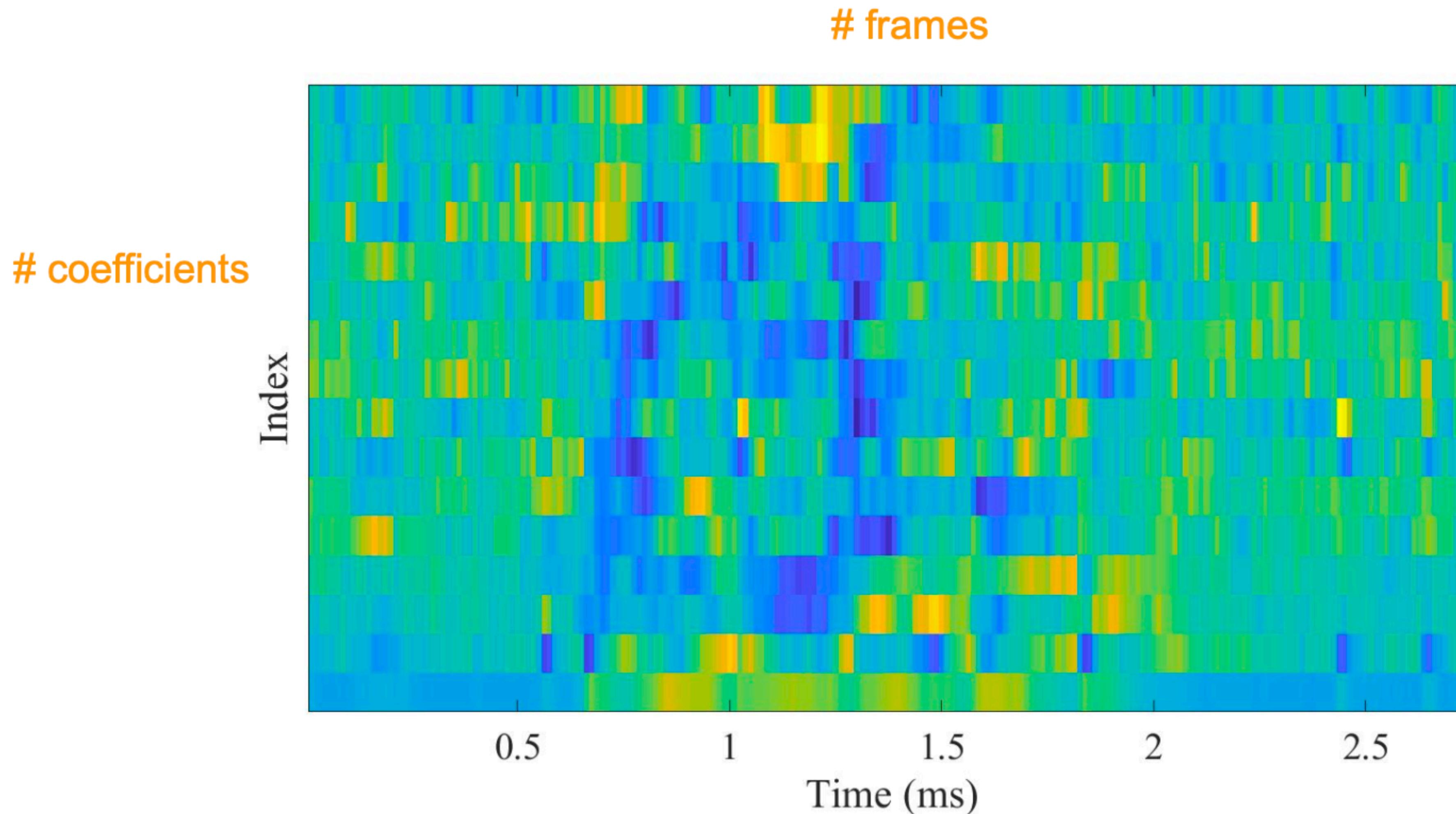
Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Bao nhiêu hệ số MFCC là phù hợp
 - Thông thường: 12 - 13 hệ số
 - Những giá trị MFCC đầu tiên sẽ chứa các thông tin rất quan trọng (đường bao tín hiệu, hình dạng âm thanh)
- Sử dụng Δ hay $\Delta\Delta$ MFCCs
- Sử dụng 39 hệ số MFCCs trên một frames

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral



Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ích lợi của MFCCs
 - Mô tả được cấu trúc chủ đạo của một spectrum
 - Bỏ qua đi các phẩn chi tiết
 - Phù hợp với xử lý lý thoại và âm nhạc
- Nhược điểm của MFCCs
 - Không có tính kháng nhiễu mạnh
 - Không phù hợp để tổng hợp âm thanh

Trích xuất tín hiệu Mel-Spectrogram, Mel-Frequency Cepstral

Mel-Frequency Cepstral

- Ứng dụng của MFCCs
 - Xử lý thoại:
 - Nhận dạng giọng nói
 - Nhận dạng người nói
 - Xử lý âm nhạc:
 - Phân loại thể loại nhạc (genre)
 - Phân loại trạng thái của âm nhạc
 - Tagging

Quiz và Lab

Nội dung

- Thực hành xác định Mel-Spectrum
- Thực hành xác định Mel Cepstrum MFCC
-