Speech Recognition (DSAI 456)

Lecture 3

Mohamed Ghalwash
mghalwash@zewailcity.edu.eg

Lecture 2 Recap

- Waveform is combined of several sinusoidal waveforms
- Frequency Spectrum (Freq VS Amplitude)
- Spectrogram (Time vs Freq vs Amplitude)
- Mel

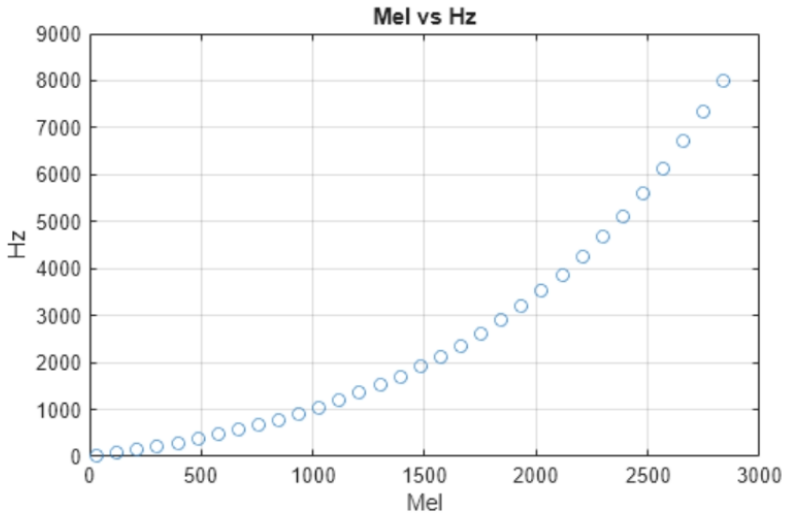
What is

- Mel Spectrum
- MFCC
- DFT

Mel to the Rescue

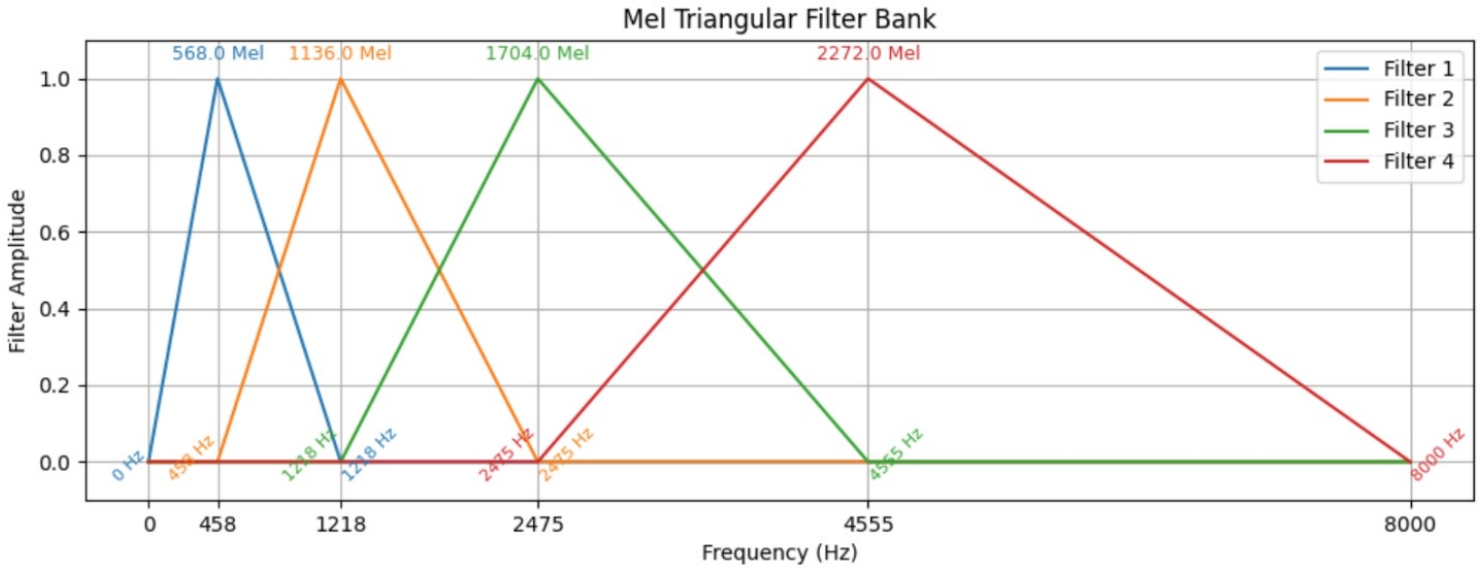
- Designed to match human perception of pitch: how humans *perceive* sound at different frequencies
- Equal intervals in mels represent equal perceived distances between pitches to a human listener
- The scale is anchored at 1000 Hz being equal to 1000 mels
- Below approximately 500 Hz, the mel and Hz scales are roughly equivalent

$$mel(f) = 2595 \log(1 + \frac{f}{700})$$



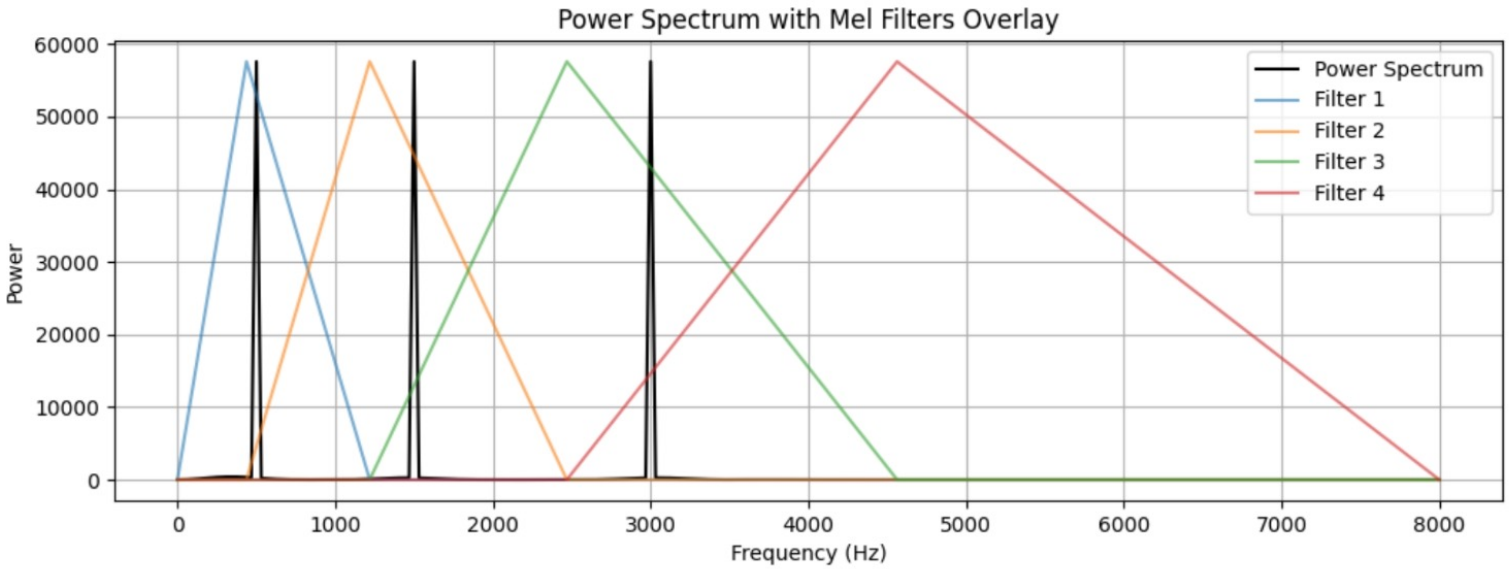
Frequency Spectrum to Mel Spectrum

- Mel channels (bands) are equally spaced points between lower/upper mels
- A series of triangular filters are created, with each filter centered at a point on the Mel scale
- Each filter captures the energy within its respective frequency band



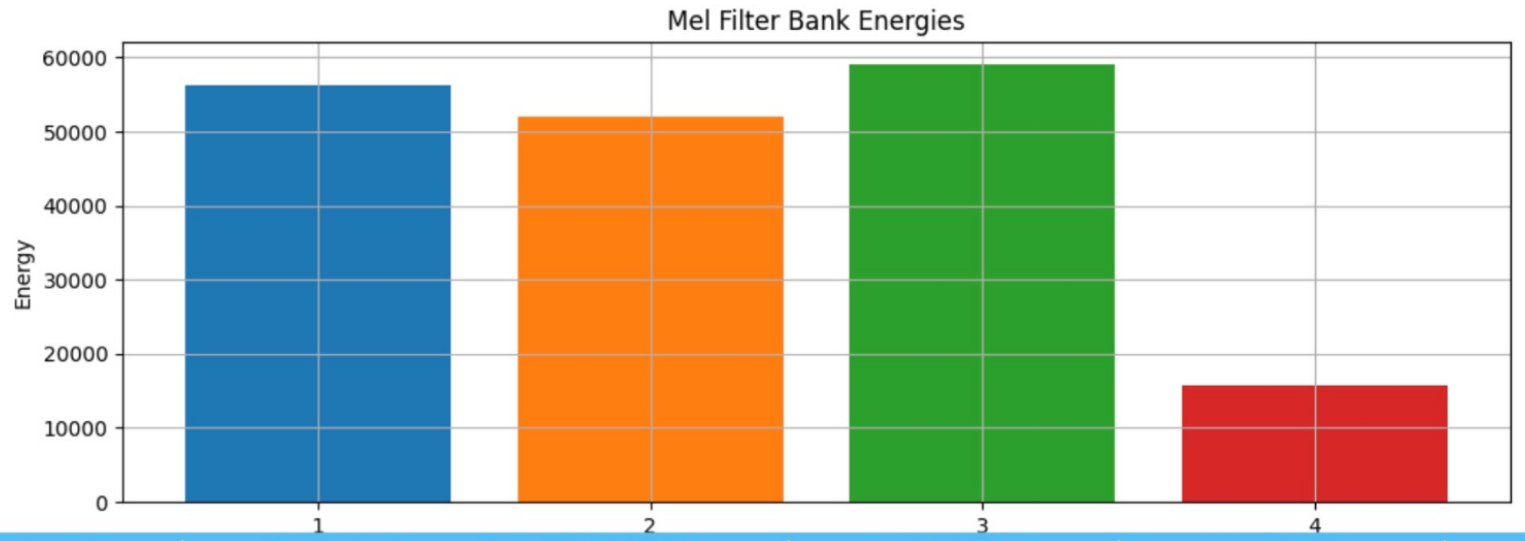
Frequency Spectrum to Mel Spectrum

- Multiply (overlay) filters by the frequency spectrum



Frequency Spectrum to Mel Spectrum

- Each filter captures the energy within its respective frequency band
- The scalar output from each filter is called a *channel (band)*
- The output for each input frame from the filter bank is a vector, represents the log energy of (mel-spaced) frequency bands



Learn More

Slidev · Course Homepage