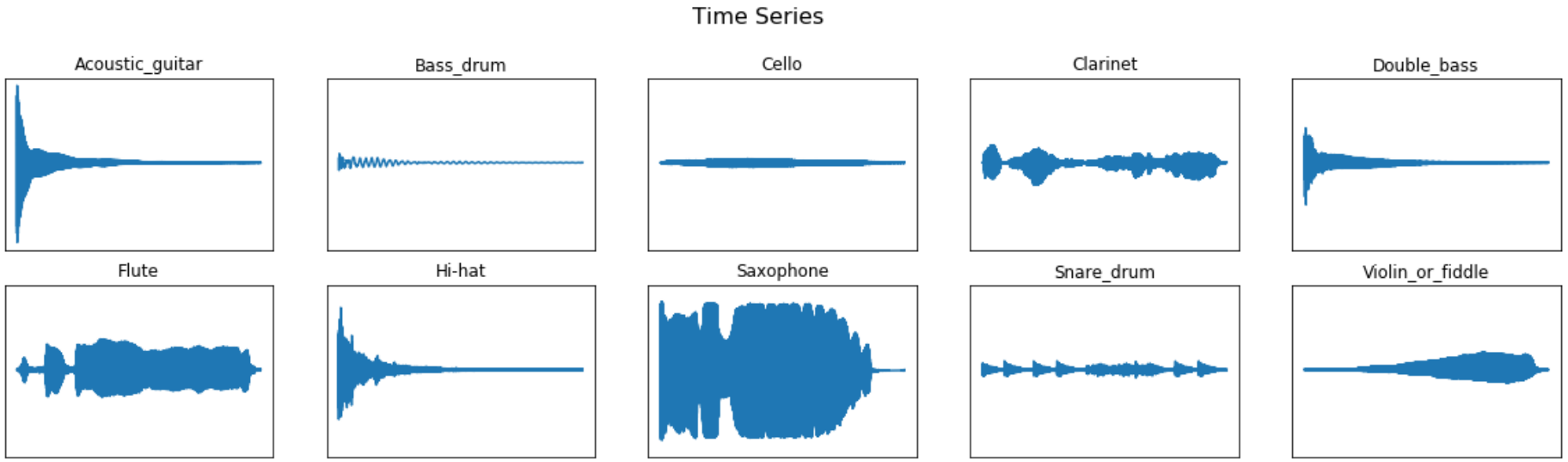
Audio Classification: Learning the Mel Scale

Amitangshu Pal

Audio Classification

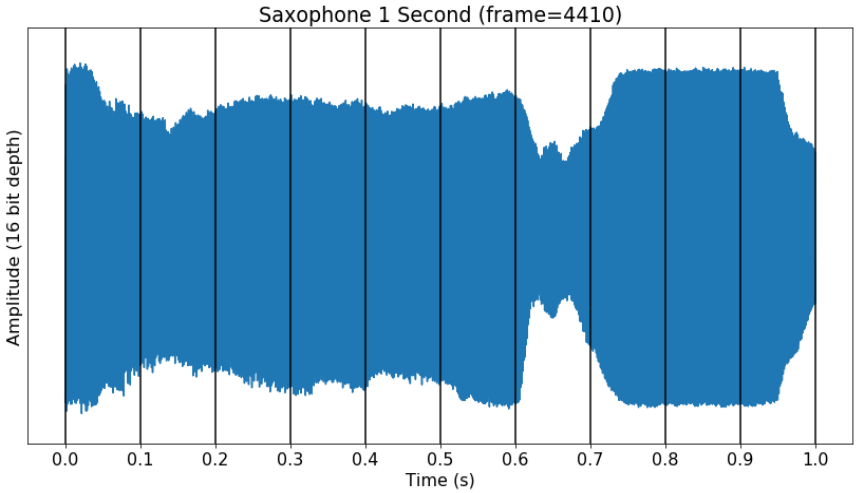
- There are a bunch of sounds coming from different musical instruments

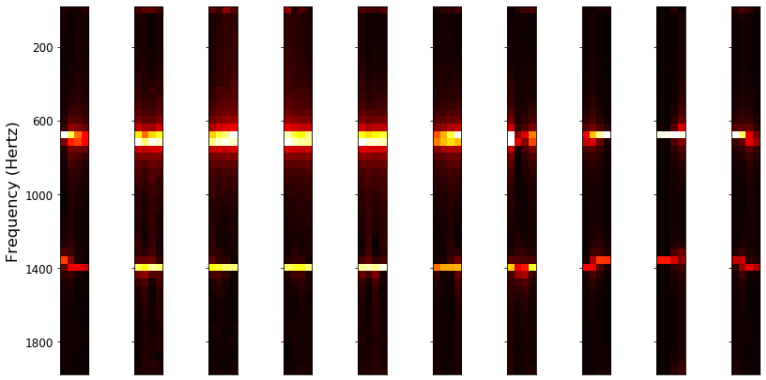
- Can your smart devices (say alexa or your smartphone) distinguish between them?
  - Human can do this distinction pretty accurately
  - How can we include this feature in smart devices?



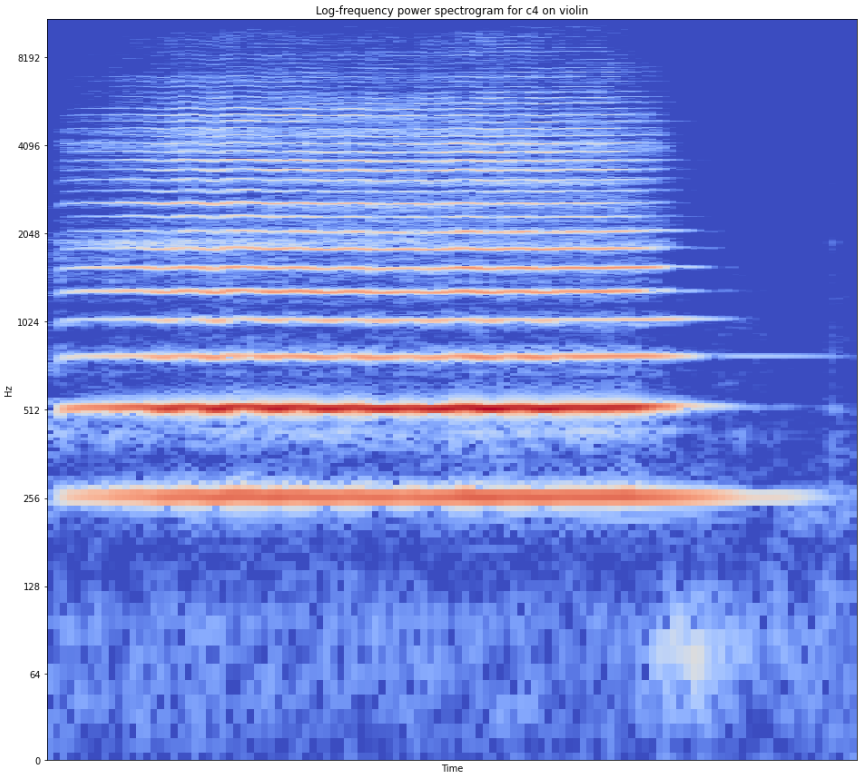
Src: Seth Adams

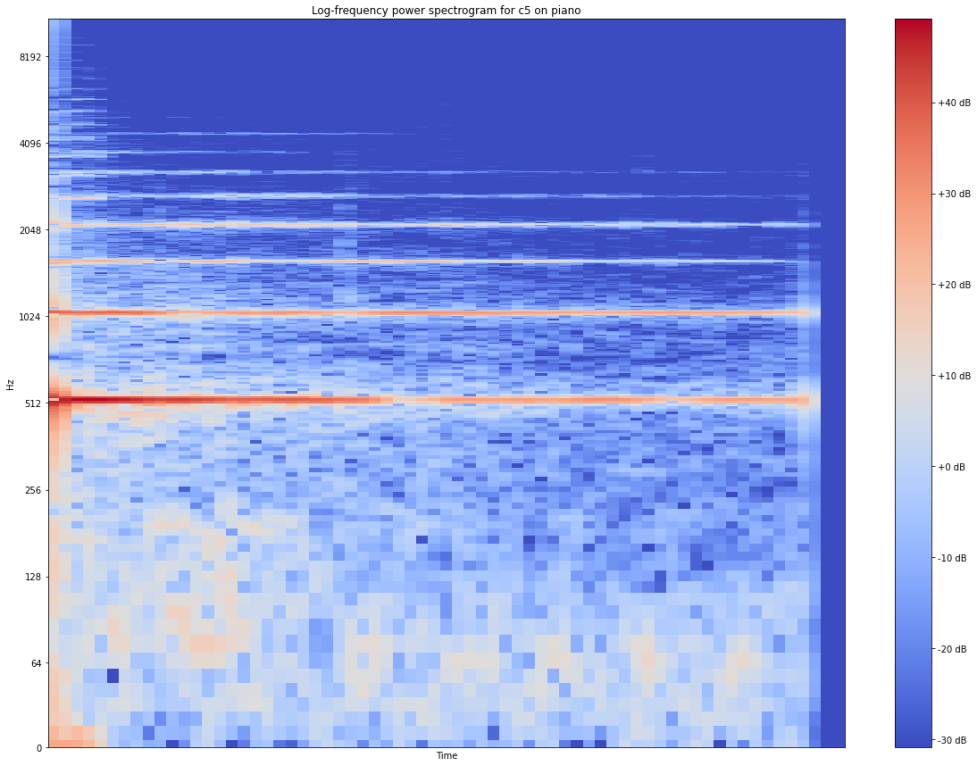
Spectrogram





Spectrogram



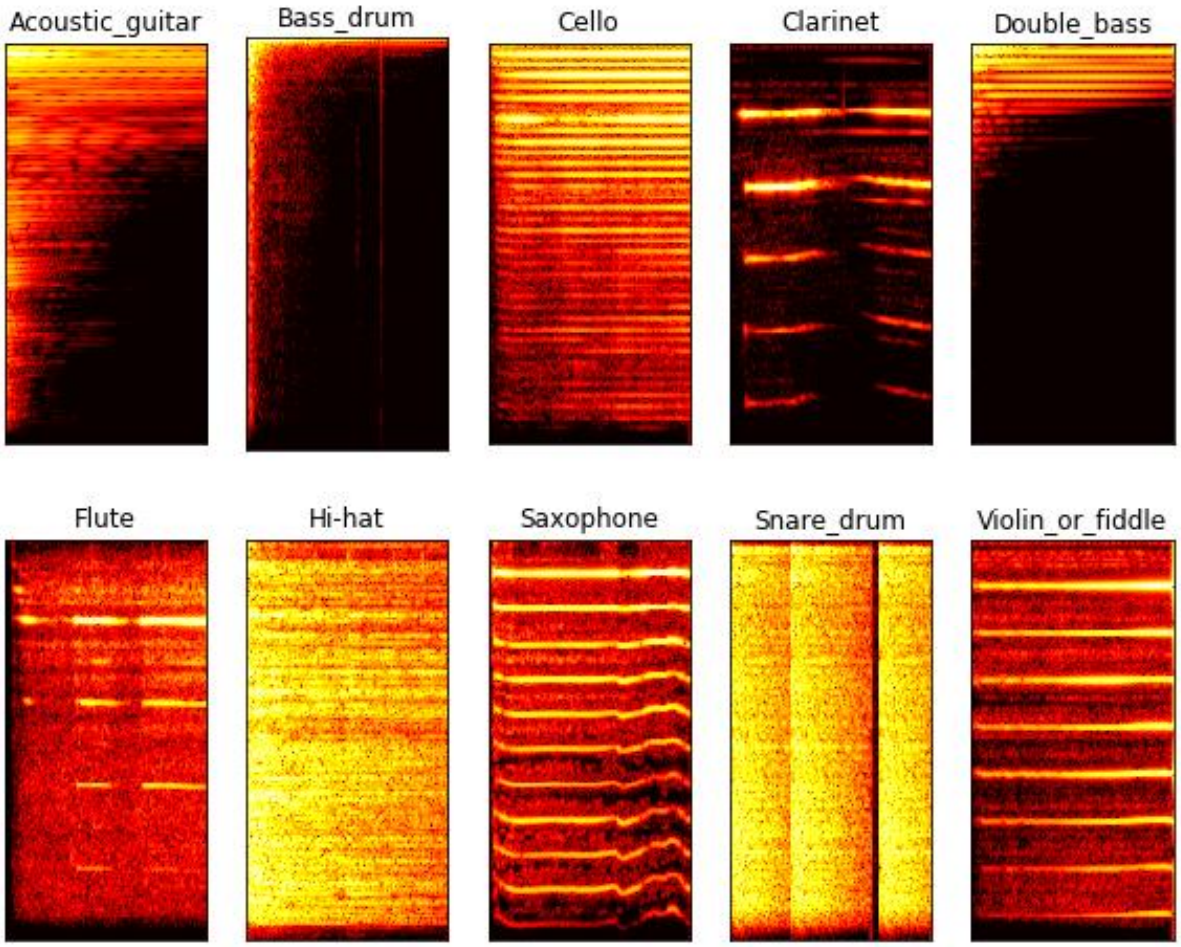


Violin

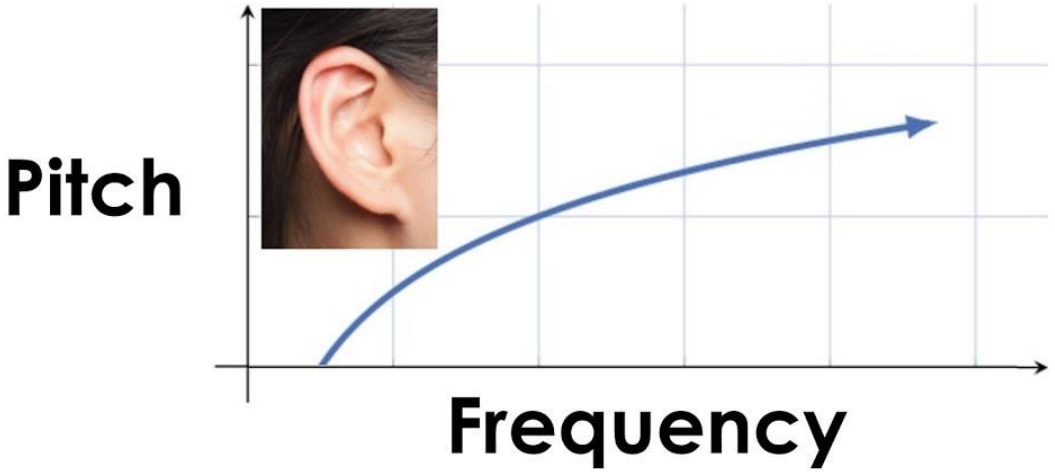
Piano

https://github.com/musikalkemist/AudioSignalProcessingForML/blob/master/3-%20Intensity%2C%20loudness%2C%20and%20timbre/intensity_and_timbre.ipynb

Spectrogram



Logarithmic Perception of Frequency



Perception of frequency

Human perceives frequencies logarithmically

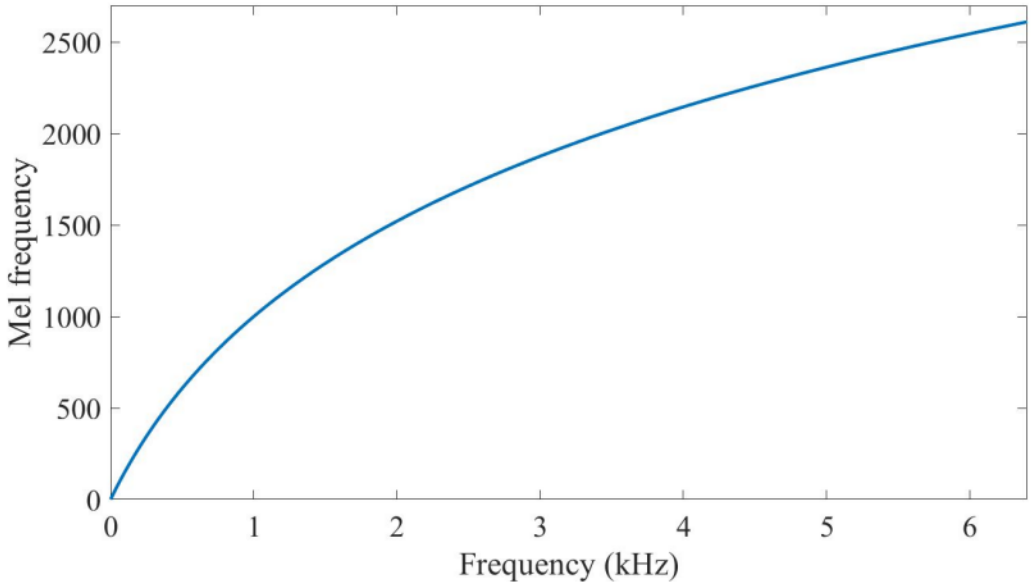
https://www.youtube.com/watch?v=i4c60ebnGc8

Mel Scale

- Logarithmic scale → equal distances have same perceptual distance
- 1000 Hz = 1000 Mel

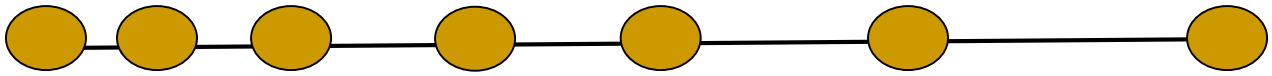
$$m = 2595 \log\left(1 + \frac{f}{700}\right)$$

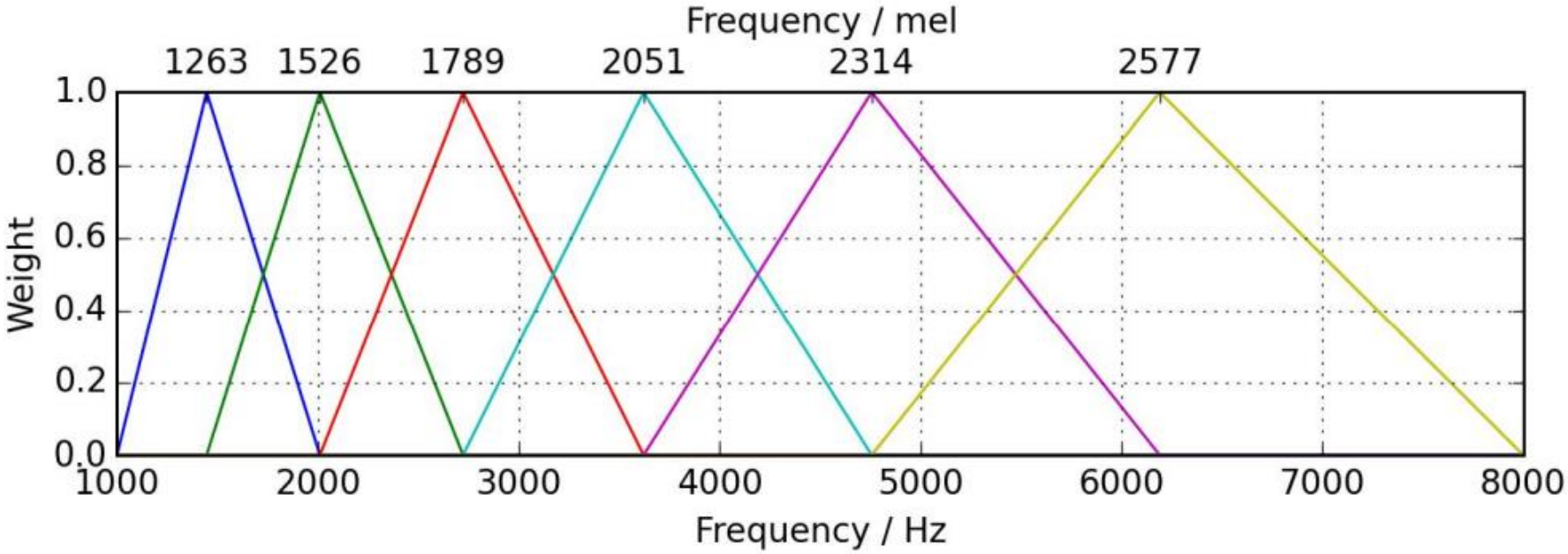
$$f = 700\left(10^{\frac{m}{2595}} - 1\right)$$



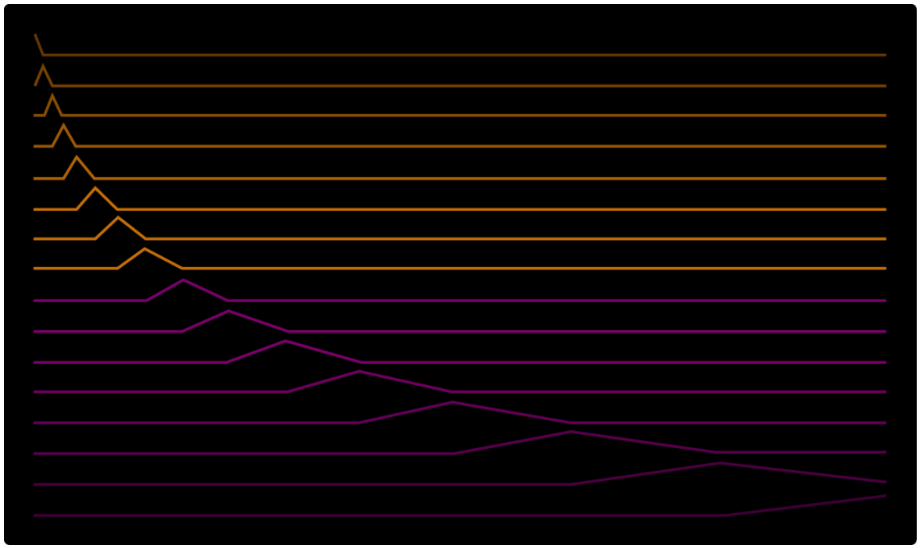
Mel Filter Banks

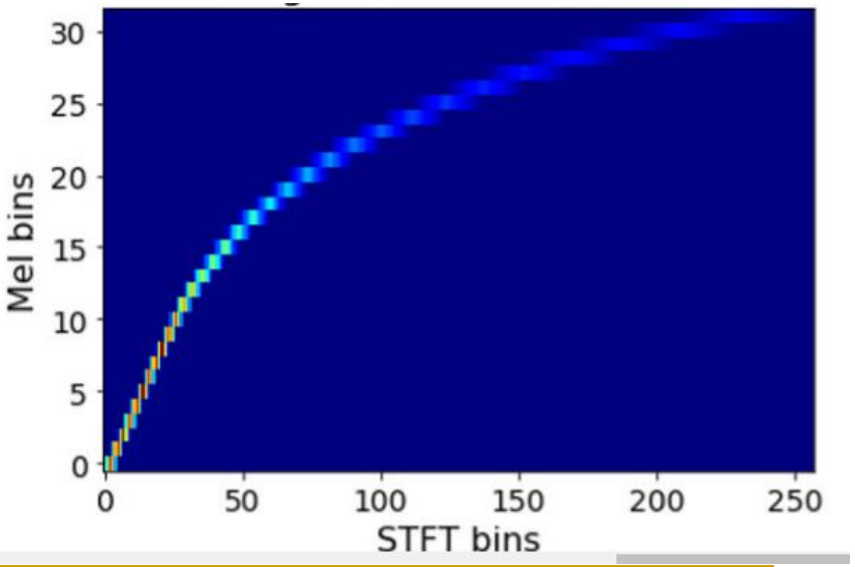
$$f = 700\left(10^{\frac{m}{2595}} - 1\right)$$





Mel Filter Banks

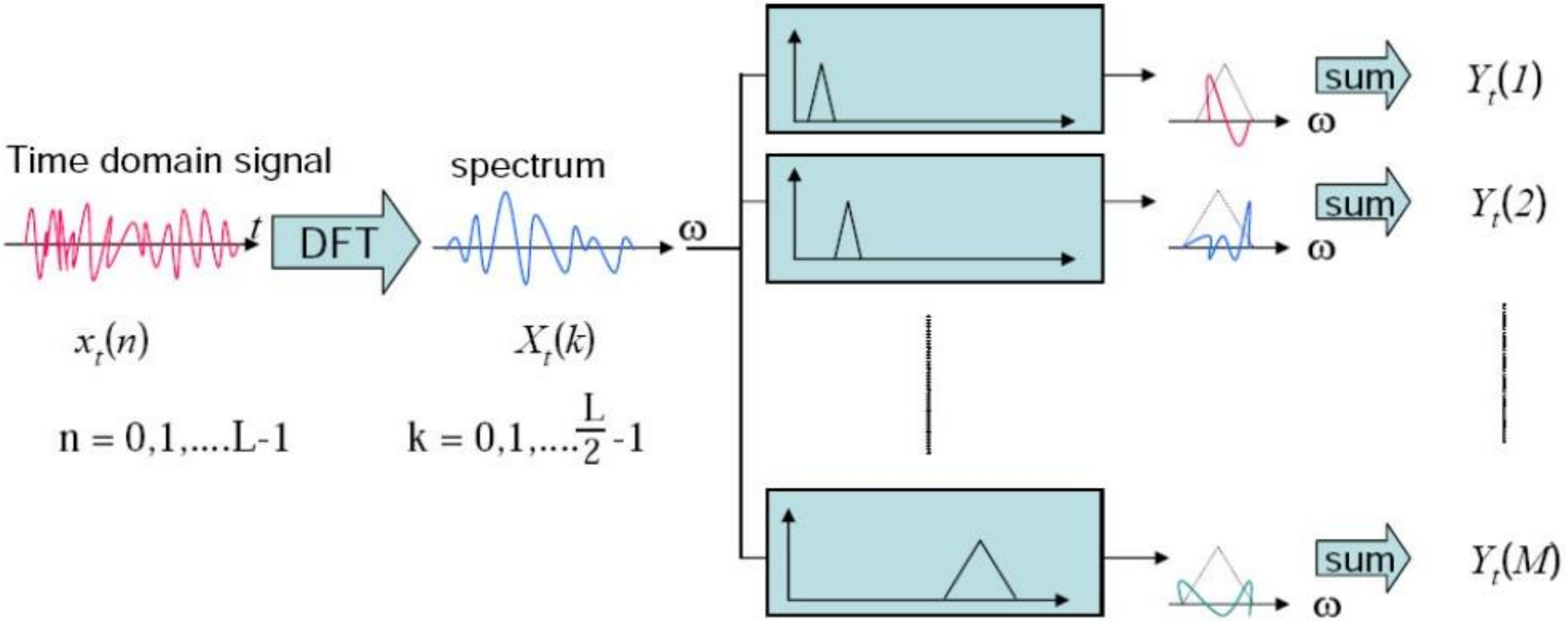




https://developer.apple.com/documentation/accelerate/computing_the_mel_spectrum_using_linear_algebra

https://ieeexplore.ieee.org/document/9174990?denied=

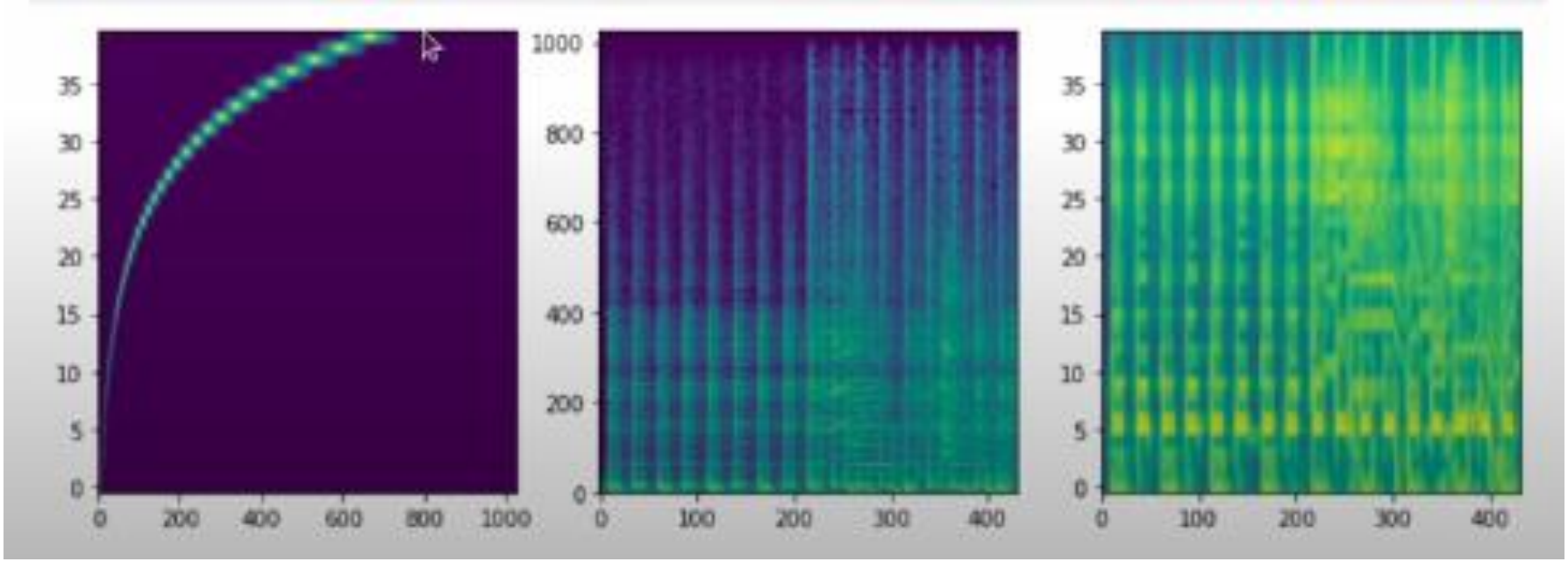
Mel Spectrogram Steps



https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9

Mel Filter Banks

Filter bank matrix | Spectrogram matrix | = | Mel Spectrogram matrix

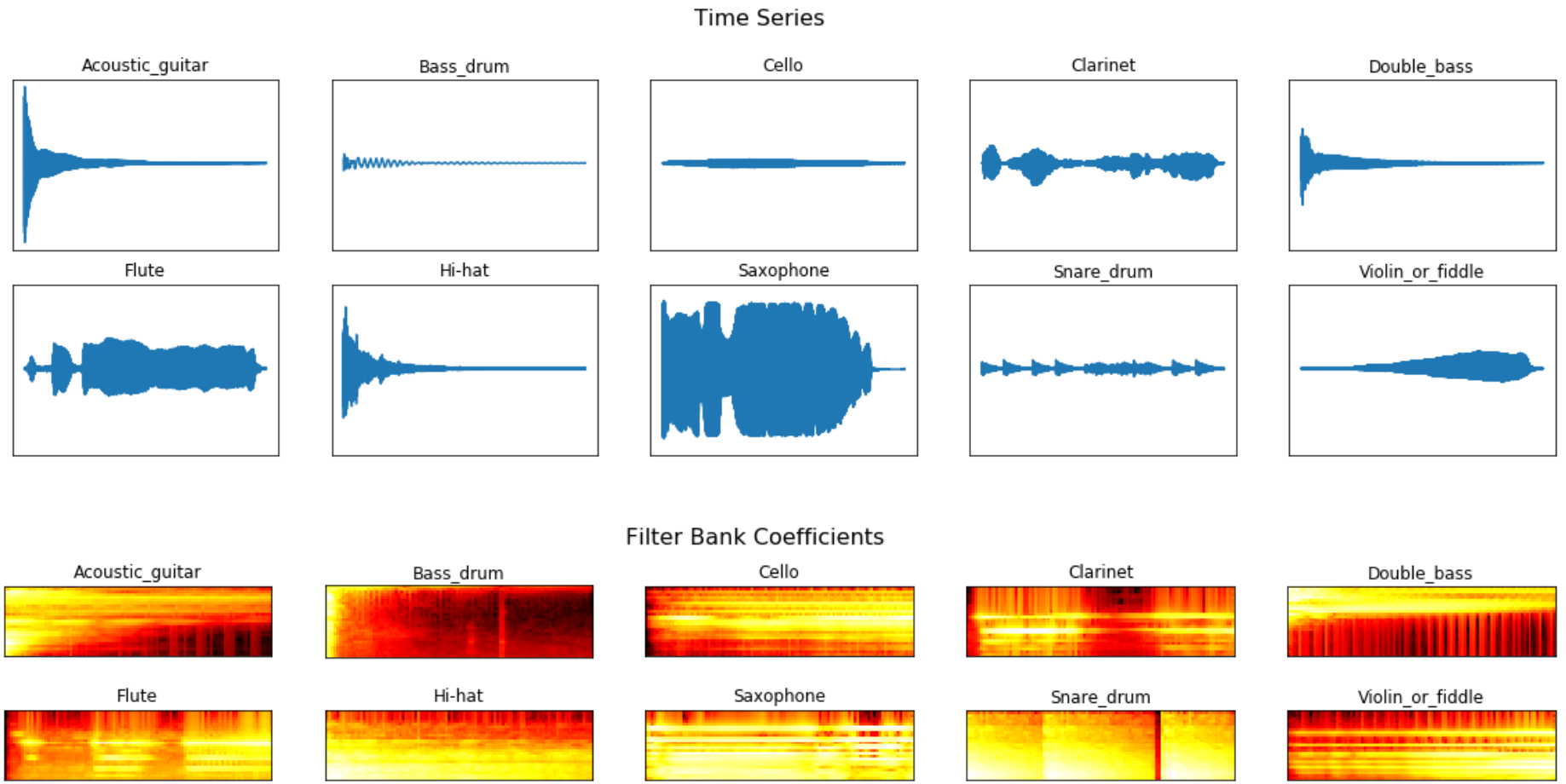


40×1024 1024×450 40×450

Also can be looked as **dimension reduction** or **lossy compression** of spectrogram

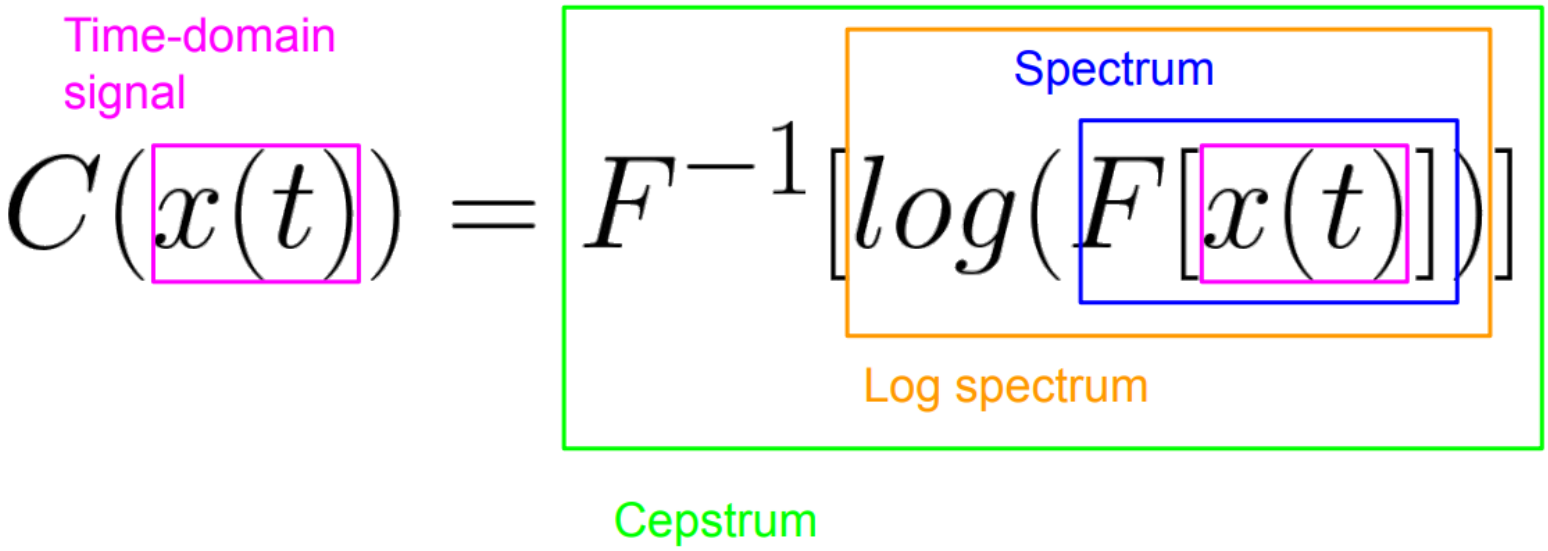
https://www.youtube.com/watch?v=WNB9JUiHAMg

Mel Spectrogram



Cepstrum

MFCC → Mel Frequency **Cepstral** Coefficients



Cepstrum

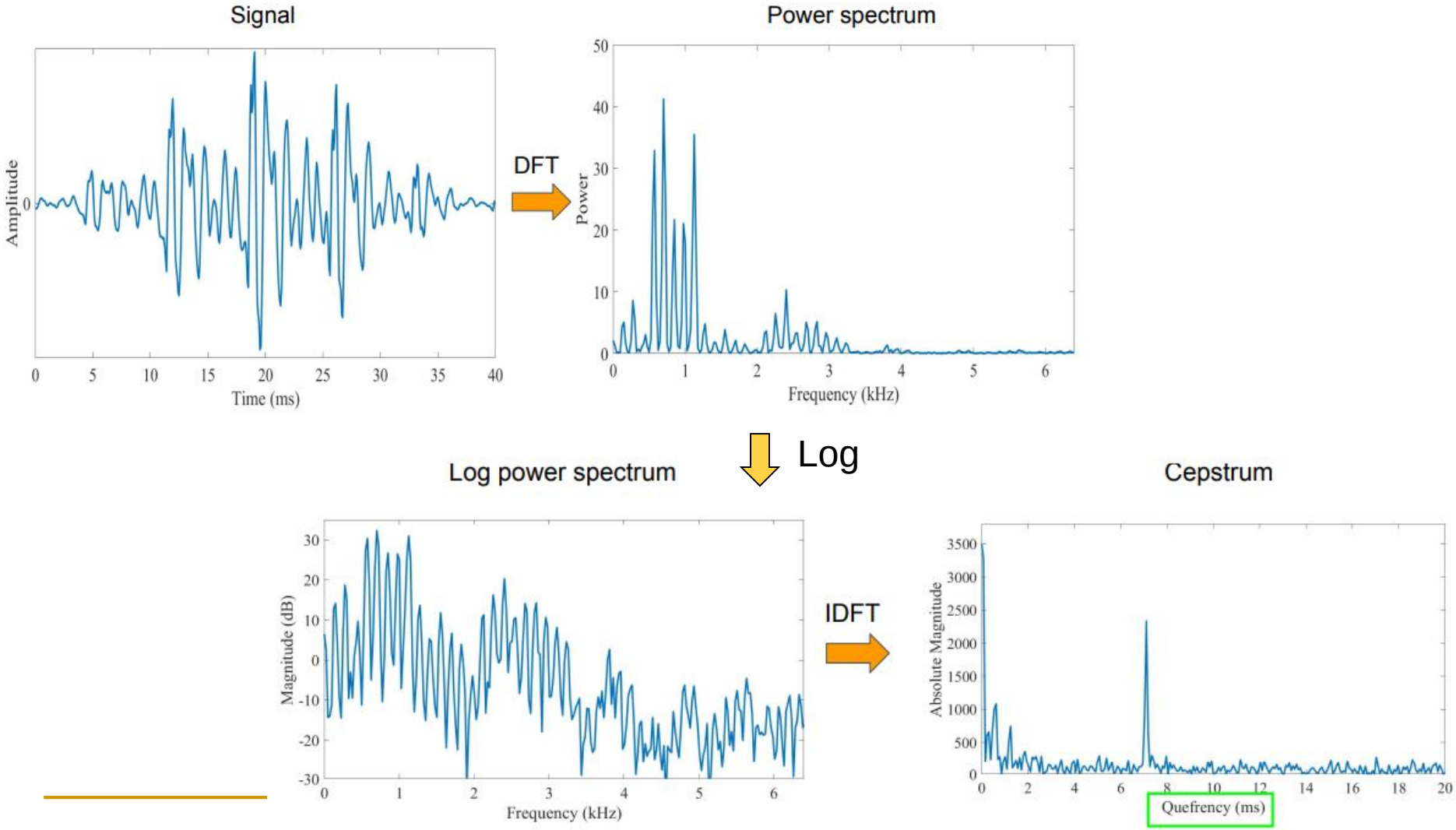
$$C(x(t)) = F^{-1}[log(F[x(t)])]$$

Time-domain signal

Spectrum

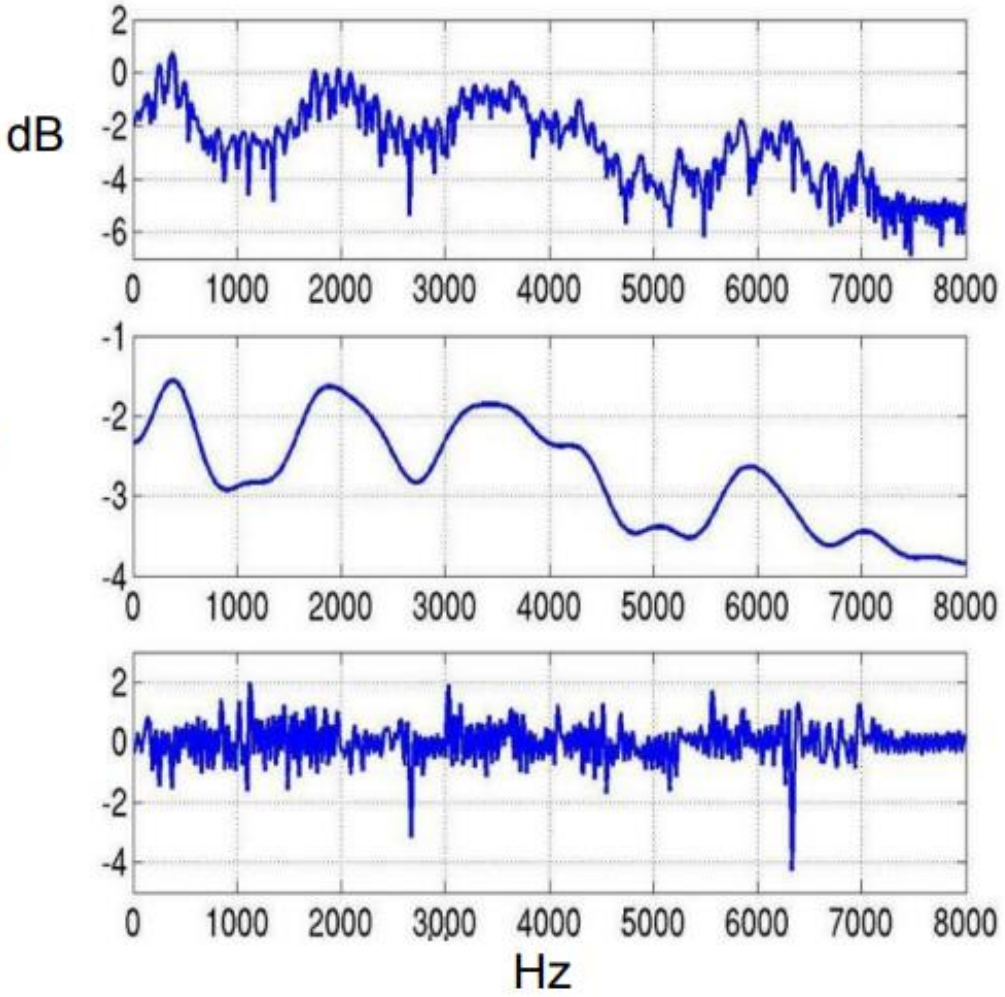
Log spectrum

Cepstrum



Understanding Cepstrum

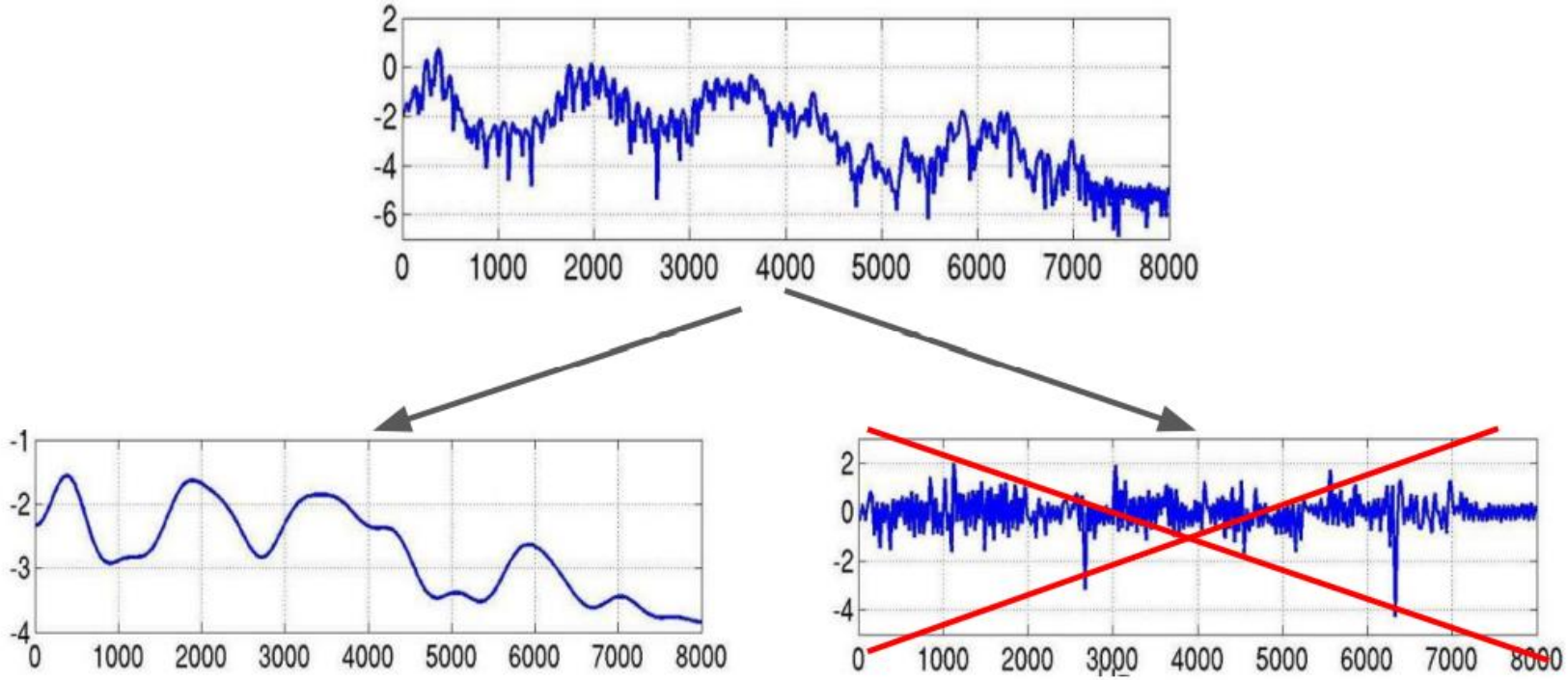
Log-spectrum



Spectral envelope

Carry identity of sound

Understanding Cepstrum



An FFT on spectrum referred to as Inverse FFT (IFFT)

http://www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf

MFCC Steps

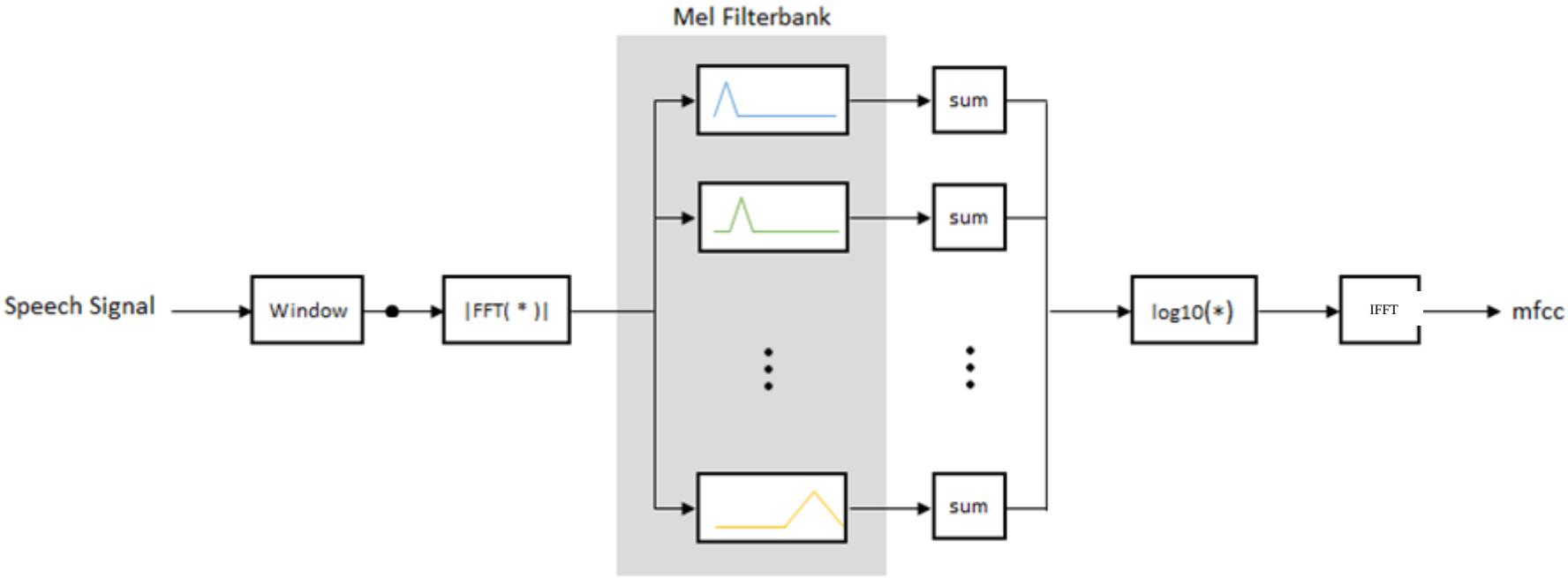
Time-domain signal

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

Spectrum

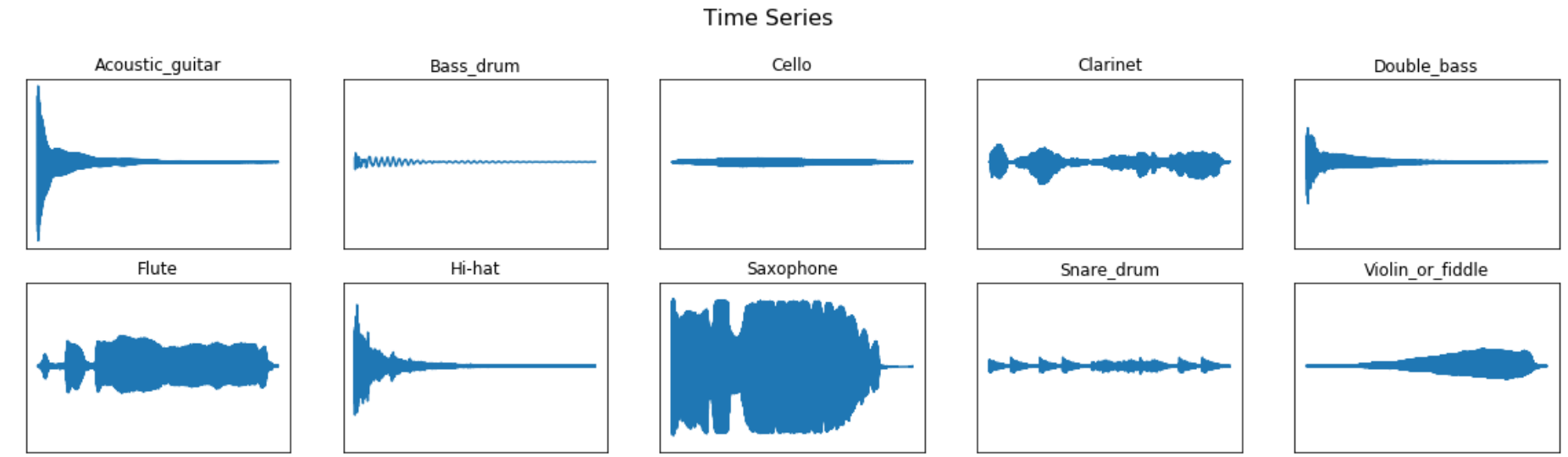
Log spectrum

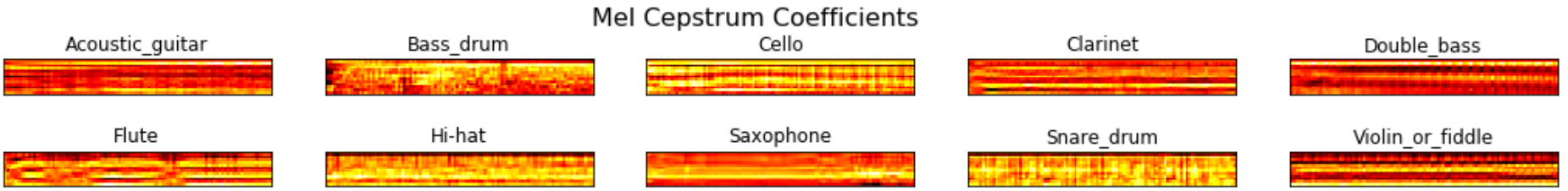
Cepstrum



https://www.mathworks.com/help/audio/ug/speaker-identification-using-pitch-and-mfcc.html

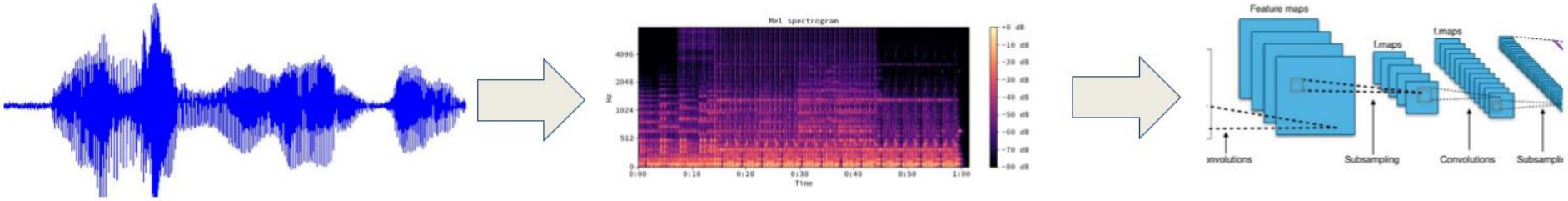
Mel Frequency Cepstrum Coefficients

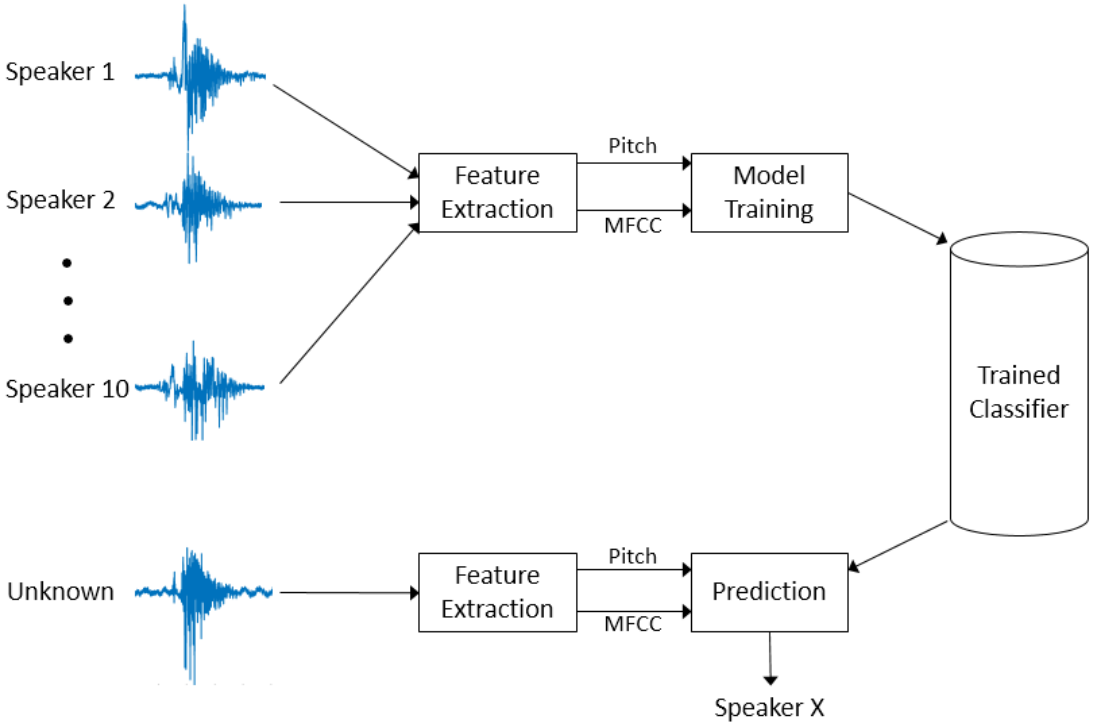




Pretty easy to classify

MFCC Use Cases

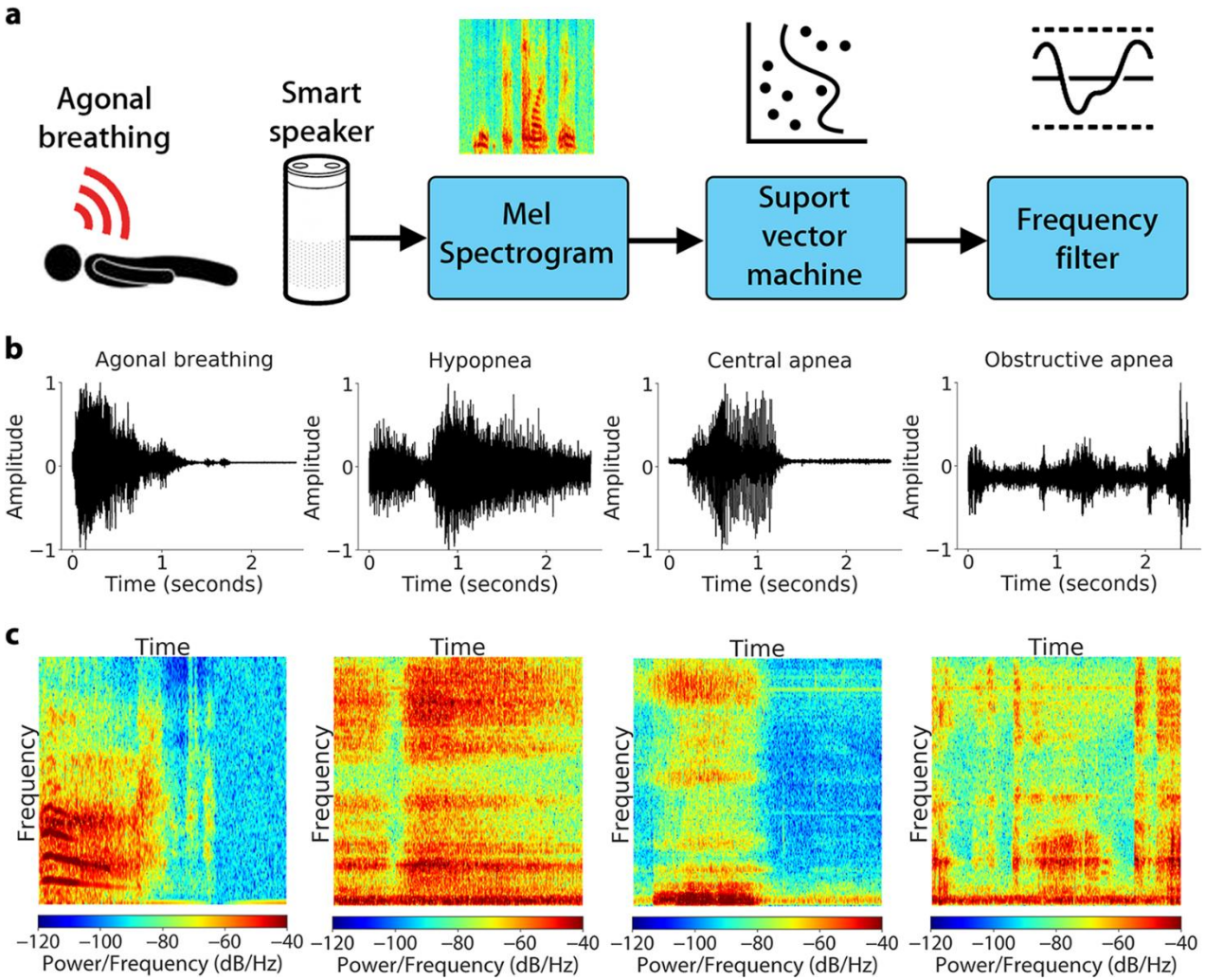




http://staff.washington.edu/dd292/Instrument.html

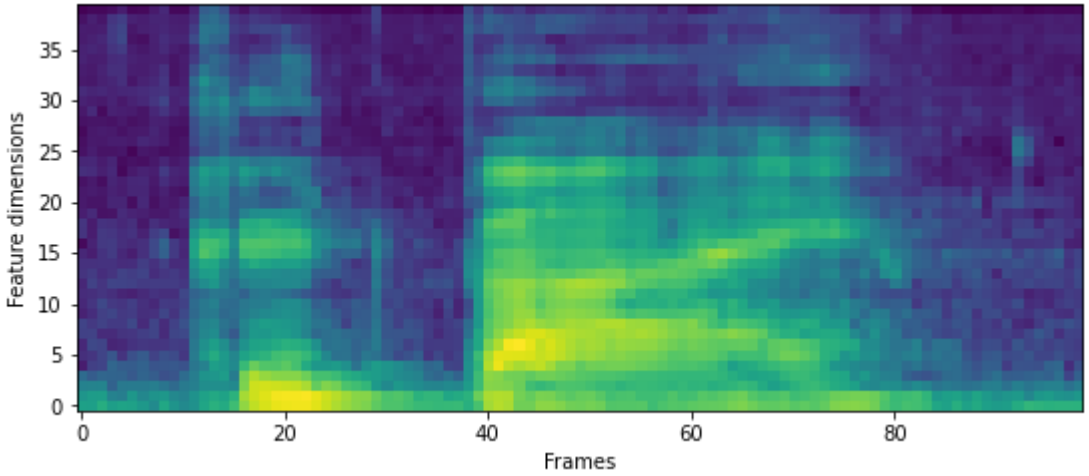
https://www.mathworks.com/help/audio/ug/speaker-identification-using-pitch-and-mfcc.html

MFCC Use Cases

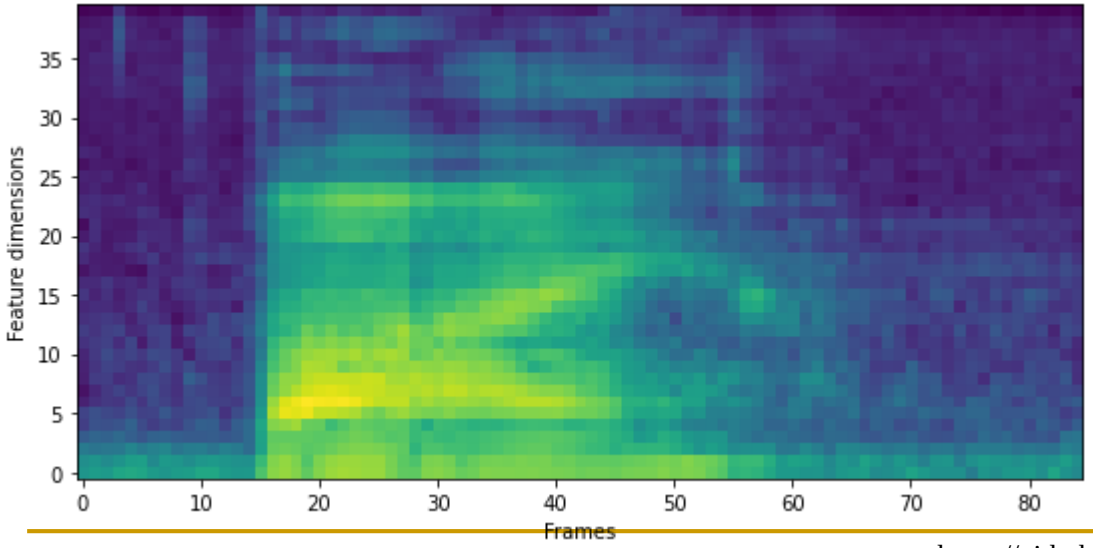


https://www.nature.com/articles/s41746-019-0128-7

DTW on MFCC



Goodbye



Bye

https://github.com/kamperh/lecture_dtw_notebook/blob/main/dtw.ipynb

DTW on MFCC