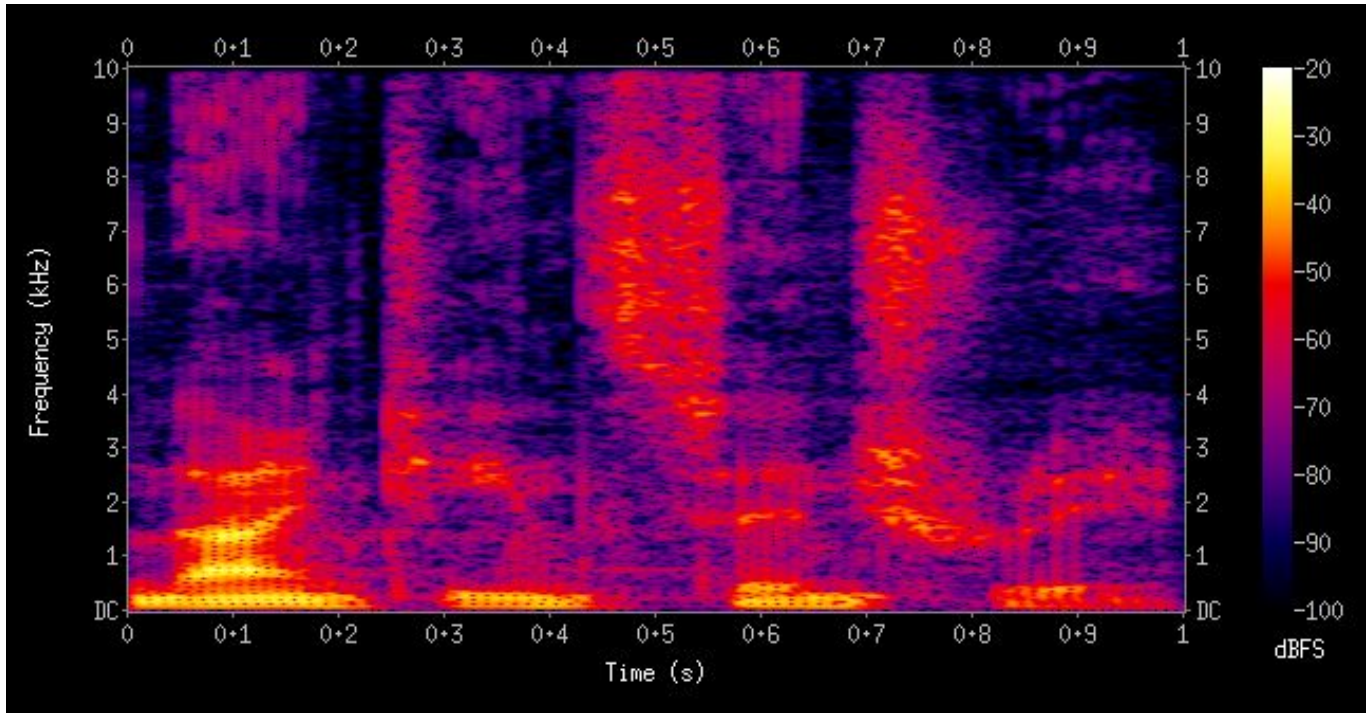
Previously...


0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1
10
9
8
7
6
5
4
3
2
1
DC
Frequency (kHz)
Time (s)
-20
-30
-40
-50
-60
-70
-80
-90
-100
dBFS

Psychoacoustic experiment

Psychoacoustic experiment

- 1st sample: C2 - C4 -> (65 - 262Hz)
- 2nd sample: G6 - A6 -> (1568 - 1760Hz)

Psychoacoustic experiment

- 1st sample: C2 - C4 -> (65 - 262Hz)
- 2nd sample: G6 - A6 -> (1568 - 1760Hz)

200 Hz

We have a problem!

Humans perceive frequency logarithmically

Ideal audio feature

- Time-frequency representation

Ideal audio feature

- Time-frequency representation
- Perceptually-relevant amplitude representation

Ideal audio feature

- Time-frequency representation
- Perceptually-relevant amplitude representation
- Perceptually-relevant frequency representation

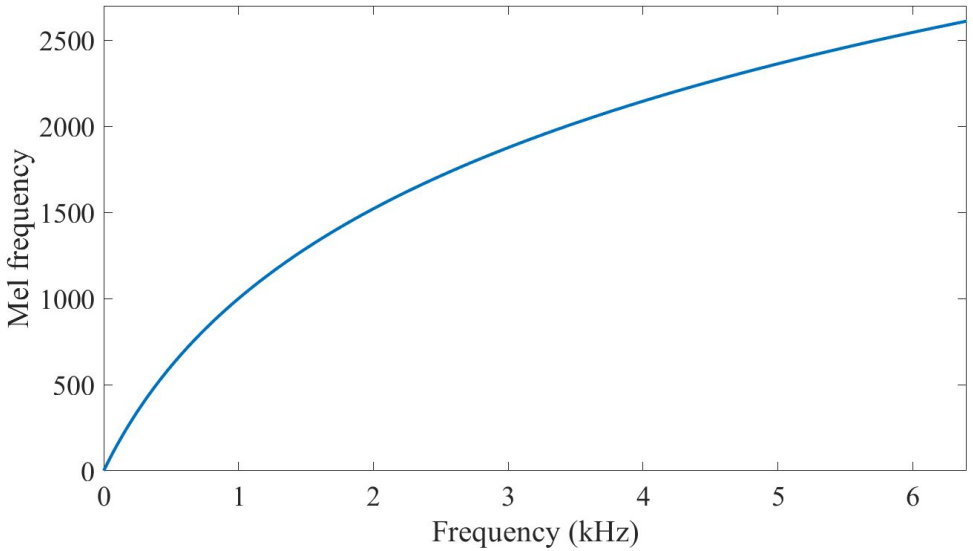
Ideal audio feature

- Time-frequency representation
- Perceptually-relevant amplitude representation
- Perceptually-relevant frequency representation

Mel spectrograms

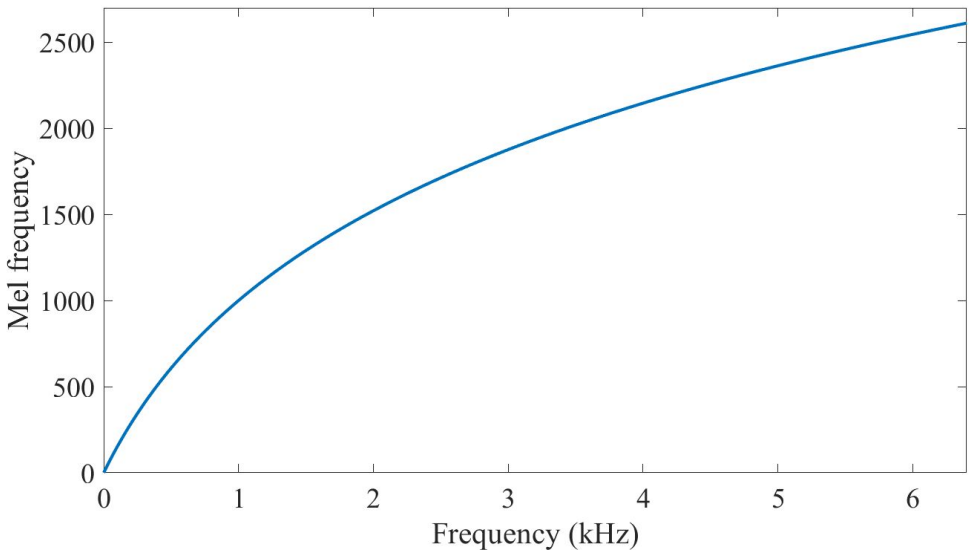
Mel-scale

- Logarithmic scale
- Equal distances on the scale have same “perceptual” distance
- 1000 Hz = 1000 Mel



Mel-scale

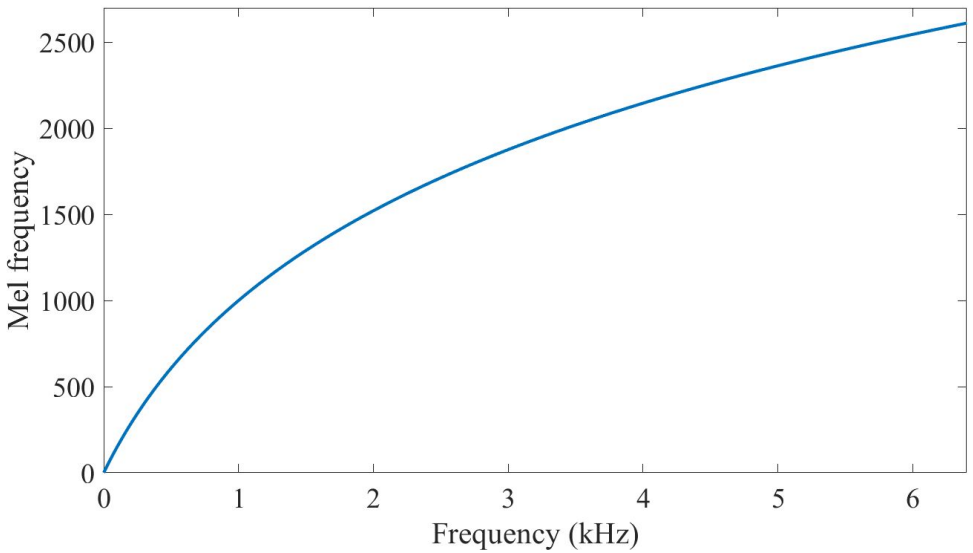
$$m = 2595 \cdot log(1 + \frac{f}{500})$$



Mel-scale

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

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



Recipe to extract Mel spectrogram

1. Extract STFT
2. Convert amplitude to DBs
3. Convert frequencies to Mel scale

Recipe to extract Mel spectrogram

1. Extract STFT
2. Convert amplitude to DBs
3. Convert frequencies to Mel scale

Convert frequencies to Mel scale

1. Choose number of mel bands
2. Construct mel filter banks
3. Apply mel filter banks to spectrogram

How many mel bands?

How many mel bands?

40

How many mel bands?

40

60

How many mel bands?

40

60

90

How many mel bands?

40

128

90

60

How many mel bands?

40

128

90

60

It depends on the problem!

Convert frequencies to Mel scale

1. Choose number of mel bands
2. Construct mel filter banks
3. Apply mel filter banks to spectrogram

Mel filter banks

1. Convert lowest / highest frequency to Mel

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

Mel filter banks

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points

Mel filter banks

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points

Mel filter banks

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points
3. Convert points back to Hertz

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

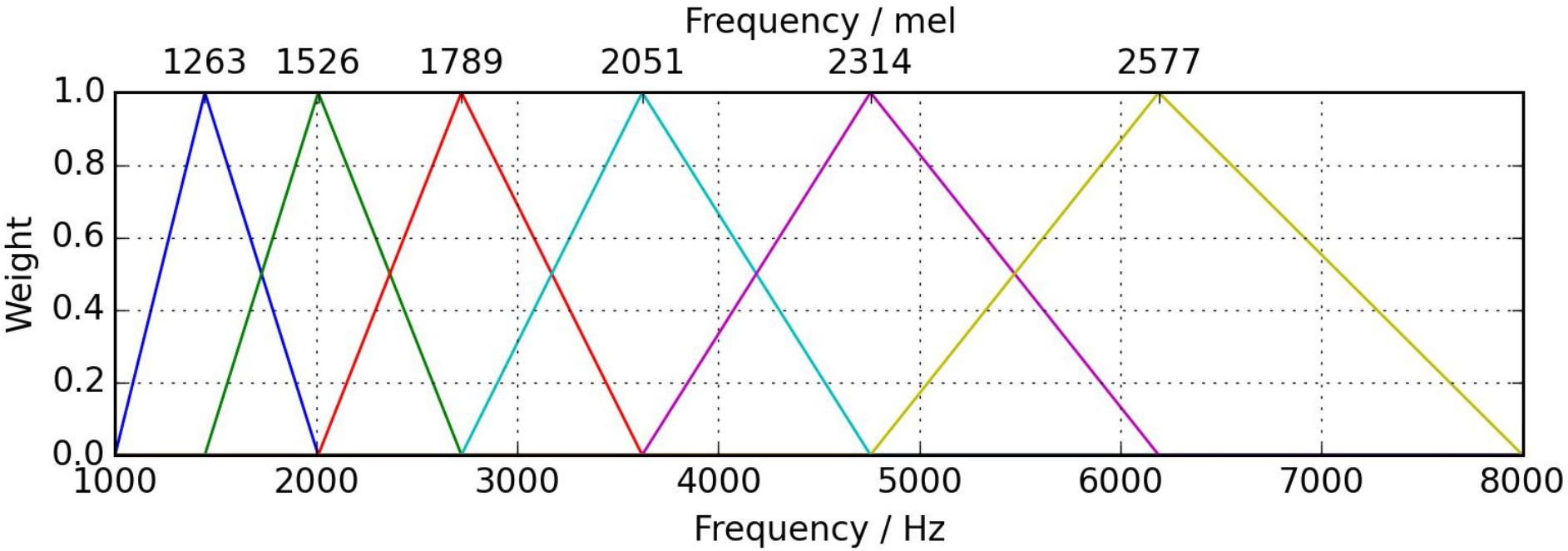
Mel filter banks

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points
3. Convert points back to Hertz
4. Round to nearest frequency bin

Mel filter banks

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points
3. Convert points back to Hertz
4. Round to nearest frequency bin
5. Create triangular filters

Mel filter banks



Mel filter banks’ shape

(# bands, framesize / 2 + 1)

Convert frequencies to Mel scale

1. Choose number of mel bands
2. Construct mel filter banks
3. Apply mel filter banks to spectrogram

Applying mel filter banks to spectrogram

M = (# bands, framesize / 2 + 1)

Applying mel filter banks to spectrogram

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

$$Y = (\text{framesize} / 2 + 1, \#\ \text{frames})$$

Applying mel filter banks to spectrogram

M = (# bands, framesize / 2 + 1)

Y = (framesize / 2 + 1, # frames)

Applying mel filter banks to spectrogram

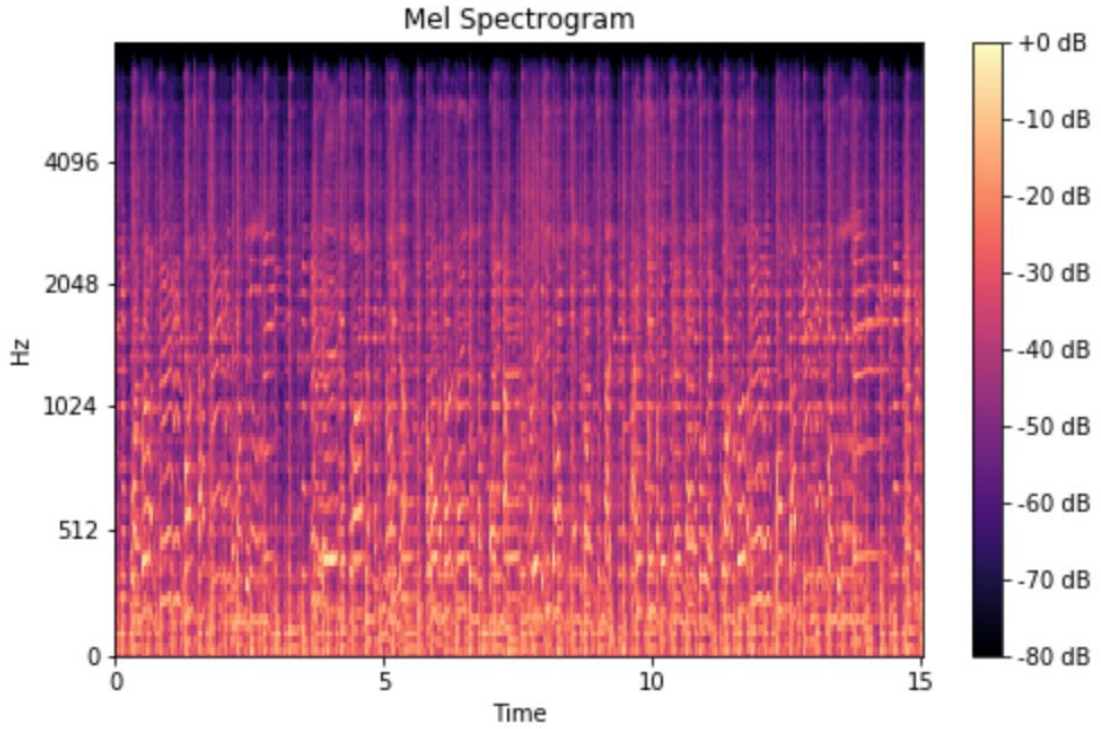
Mel spectrogram = MY

Applying mel filter banks to spectrogram

Mel spectrogram = MY

(# bands, # frames)

Applying mel filter banks to spectrogram



Mel spectrogram applications

- Audio classification
- Automatic mood recognition
- Music genre classification
- Music instrument classification
- ...

What's up next?

- Extract Mel spectrograms with Python and Librosa
- Visualise Mel spectrograms
- Extract and visualise Mel filter banks