Mel Frequency Cepstral Coefficient

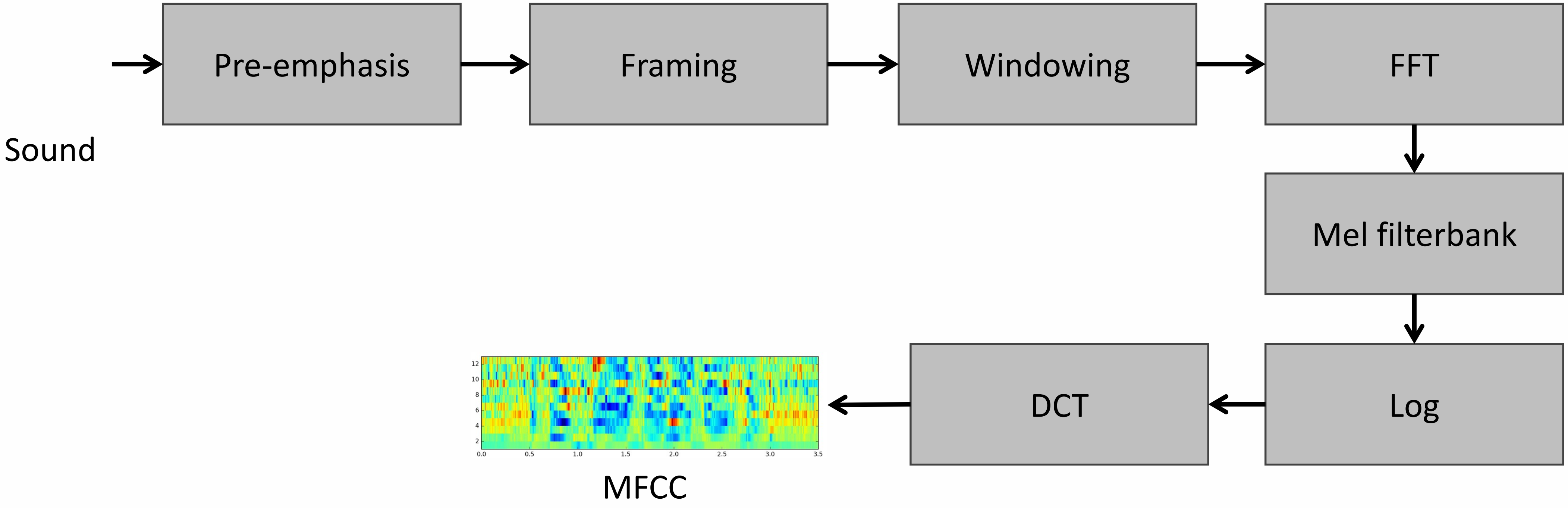
MFCC

Papangkorn Inkeaw, PhD

Department of Computer Science
Faculty of Science, Chiang Mai University

Mel Frequency Cepstral Coefficient

MFCC Process



Mel Frequency Cepstral Coefficient

Step 1: Pre-Emphasis

- Amplify the high frequencies for
  - Balance the frequency spectrum since high frequencies usually have smaller magnitudes compared to lower frequencies
  - Avoid numerical problems during the Fourier transform operation
  - Improve the Signal-to-Noise Ratio (SNR)
- Applied to a signal *x* using the first-order filter:

$$y(t) = x(t) - \alpha x(t-1)$$

- The filter coefficient α may be 0.95 or 0.97

Mel Frequency Cepstral Coefficient

Step 2: Framming

- Frequencies in a signal change over time
  - Apply Fourier transform across the entire signal → lose the frequency contours
- Frequencies in a signal are stationary over a very short period of time
  - Apply Fourier transform over this short-time frame → good approximation of the frequency contours
- Split the signal into short-time frames
  - Frame sizes in speech processing range from 20 ms to 40 ms
  - 50% (+/-10%) overlap between consecutive frames.

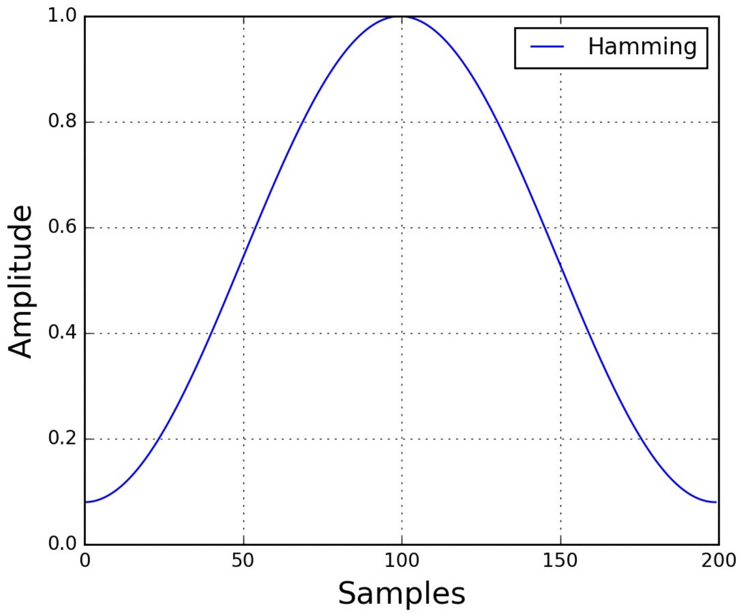
Mel Frequency Cepstral Coefficient

Step 3: Windowing

- Apply a window function such as the Hamming window to each frame.
- In order to counteract the assumption made by the FFT that the data is infinite and to reduce spectral leakage.
- A Hamming window has the following form:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

where $0 \leq n \leq N-1$, N is the window length.



Hamming Window

Source: https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html (accessed on Sep 12, 2023)

Mel Frequency Cepstral Coefficient

Step 4: Fourier-Transform and Power Spectrum

- Perform an *N*-point FFT on each frame
  - Typically, *N* is 256 or 512 (power of 2)
- Calculate the power spectrum by

$$P = \frac{|FFT(x_i)|^2}{N}$$

where x_i is the 1-th frame of signal x

Mel Frequency Cepstral Coefficient

Step 5: Mel Filter Banks

- To extract frequency bands, Apply triangular filters (typically 40 filters) on Mel-scale to the power spectrum.
- Mel-scale mimics non-linear human ear perception of sound
  - More discriminate lower frequency
  - Less discriminate higher frequency
- We can convert between Hertz (f) and Mel (m) by

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

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

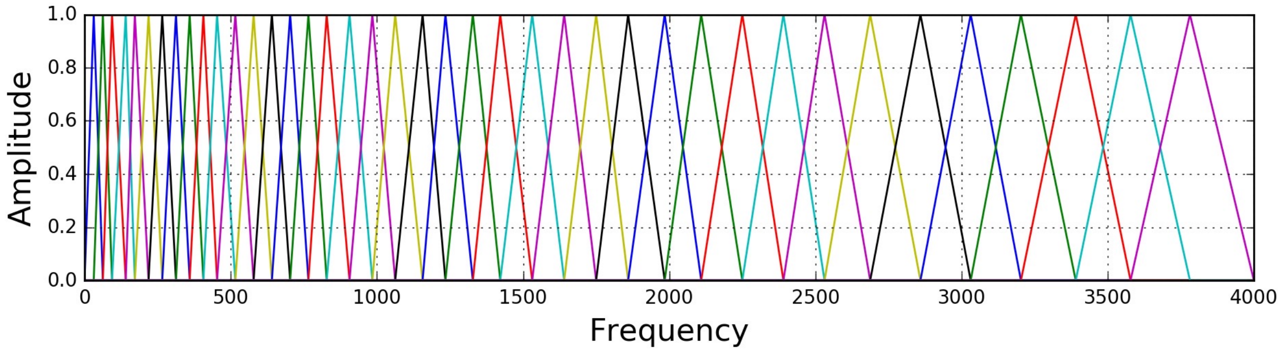
Mel Frequency Cepstral Coefficient

Step 5: Mel Filter Banks

- Model filter banks on Mel-scale by

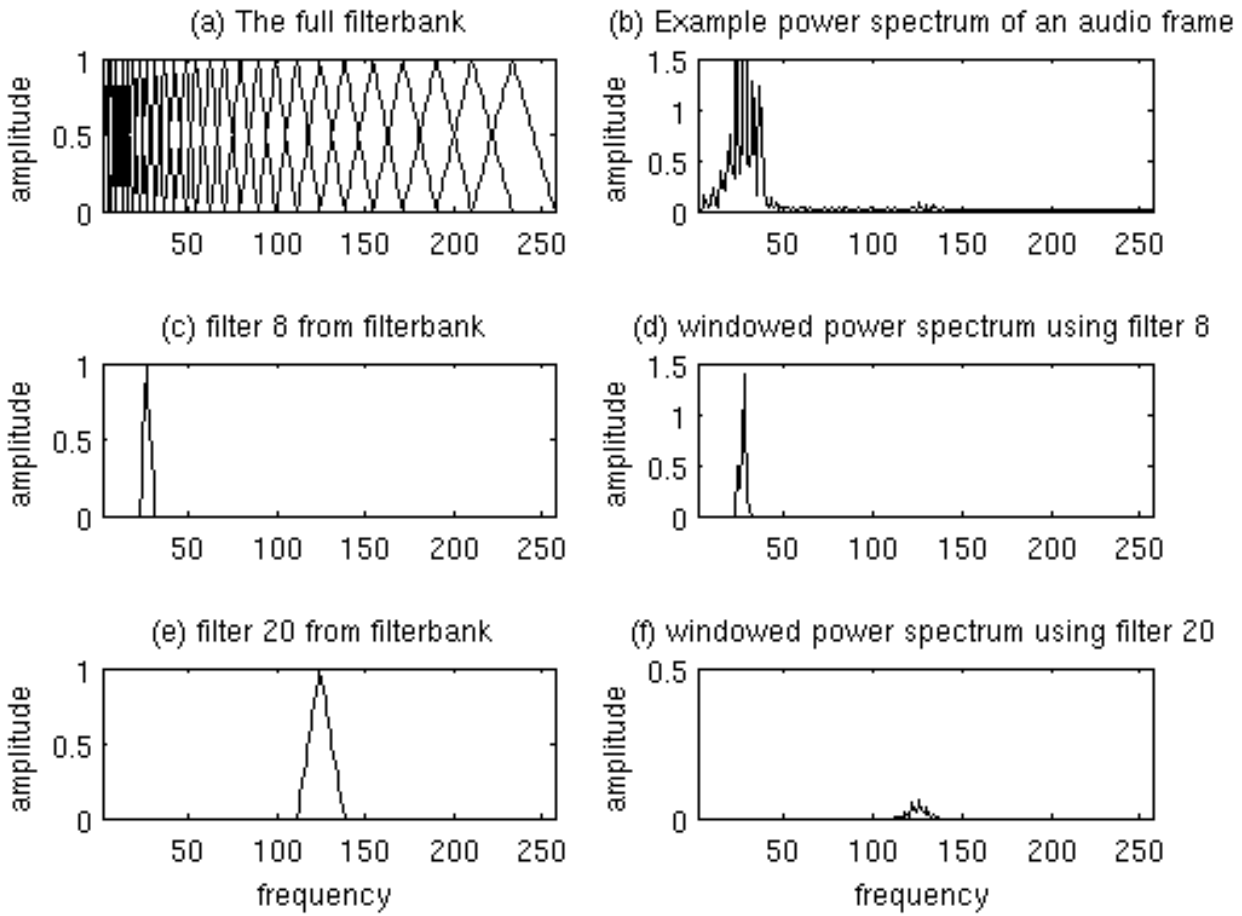
$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

where $f(m)$ is the center frequency of the triangular filter.



Mel Frequency Cepstral Coefficient

Step 5: Mel Filter Banks



Plot of Mel Filter bank and windowed power spectrum

Source: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/ (accessed on Sep 12, 2023)

Mel Frequency Cepstral Coefficient

Step 6: Log-energy

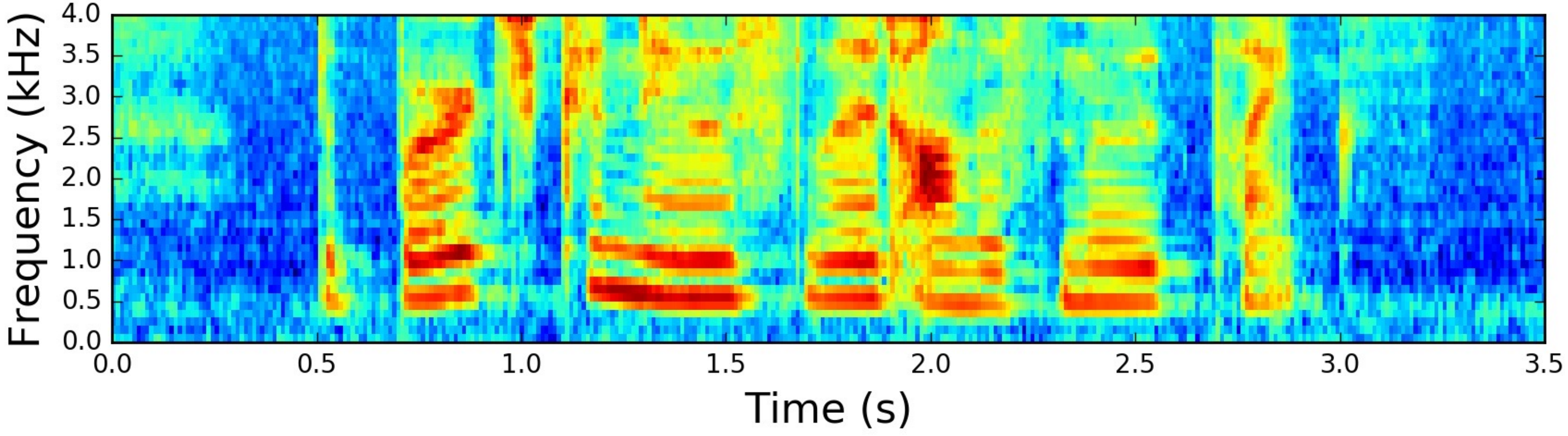
- The log-energy of each filter is calculated by

$$S(m) = \ln\left[\sum_{k=0}^{N-1} P(k)H_m(k)\right]$$

where $0 < m \leq M$ and M is the number of filters.

Mel Frequency Cepstral Coefficient

Step 6: Log-energy



Spectrogram of the Signal

Source: https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html (accessed on Sep 12, 2023)

Mel Frequency Cepstral Coefficient

Step 7: Discrete Cosine Transform (DCT)

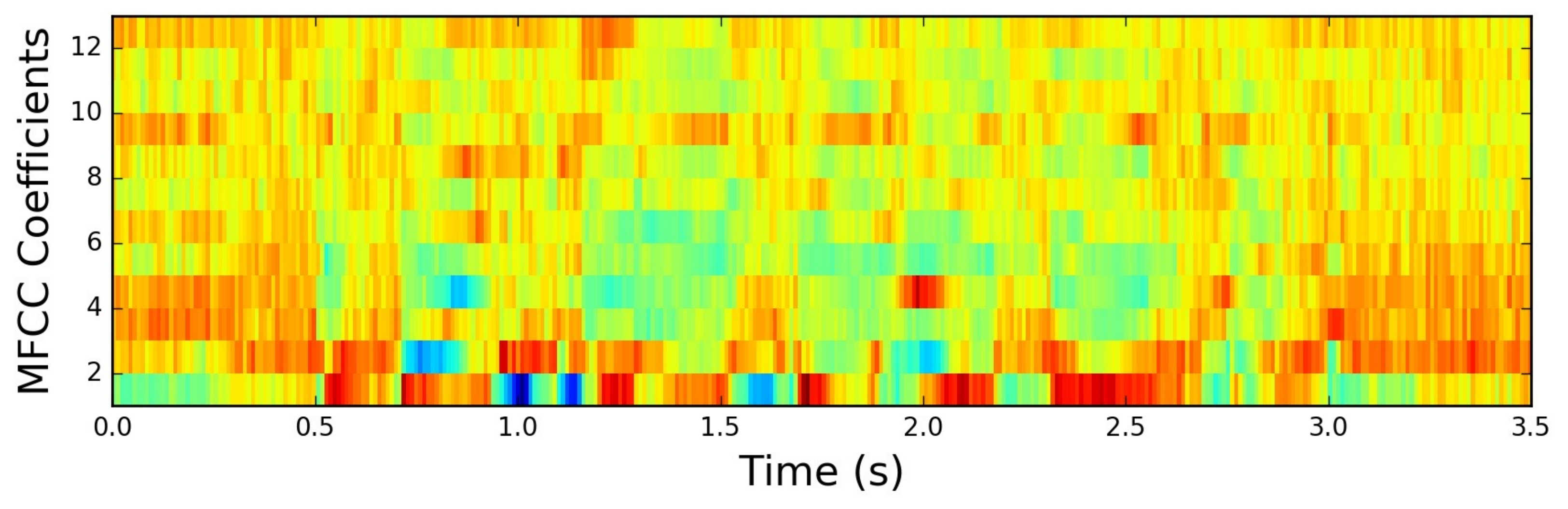
- The filter bank coefficients are highly correlated, which could be problematic in some machine learning algorithms.
- Apply Discrete Cosine Transform (DCT) to decorrelate the filter bank coefficients and yield a compressed representation of the filter banks.
- The DCT of 'M' filter outputs as Mel-frequency cepstrum coefficients is given as

$$c(q) = \sum_{m=0}^{M-1} S(m) \cos\left(\frac{\pi q \left(m - \frac{1}{2}\right)}{M}\right)$$

- Typically, the resulting cepstral coefficients (q) 2-13 are retained and the rest are discarded.

Mel Frequency Cepstral Coefficient

Step 7: Discrete Cosine Transform (DCT)



MFCCs

Source: https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html (accessed on Sep 12, 2023)

References

- Choudakkanavar, G., Mangai, J. A., & Bansal, M. (2022). MFCC based ensemble learning method for multiple fault diagnosis of roller bearing. *International Journal of Information Technology*, 14(5), 2741-2751. doi:10.1007/s41870-022-00932-x
- https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html (accessed on Sep 12, 2023)
- http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/ (accessed on Sep 12, 2023)