Audio & Speech: MFCC

DSP Lab 2022 autumn Audio and Speech week 1

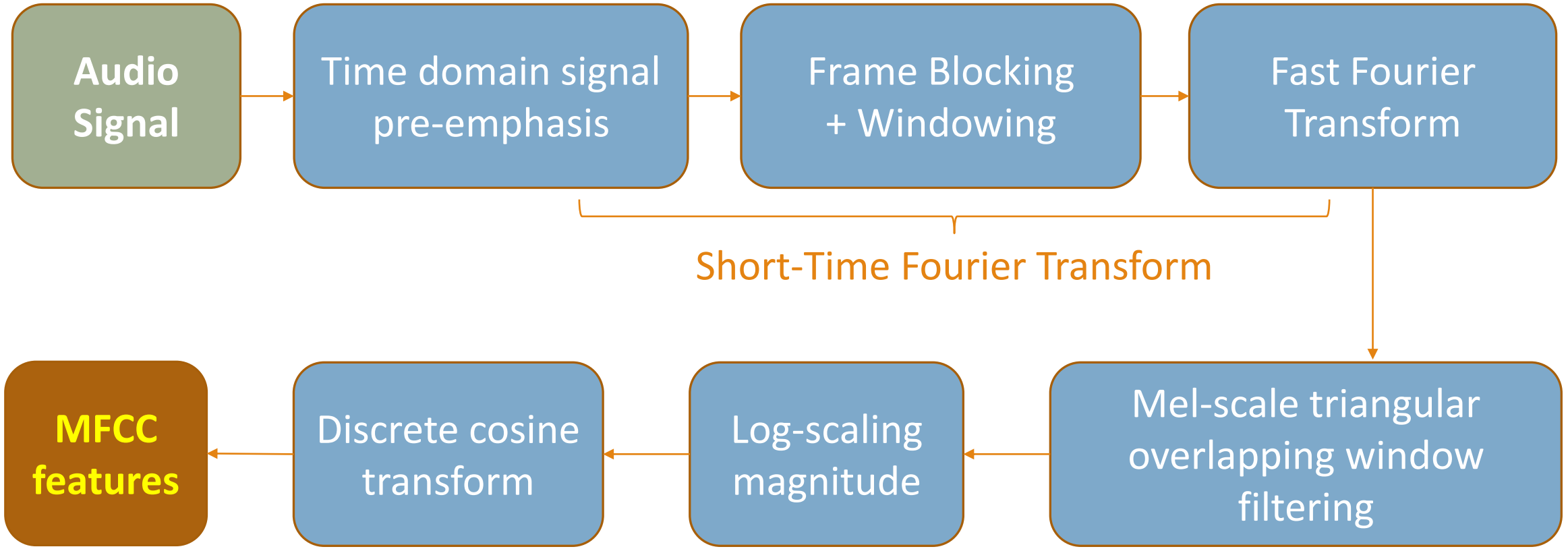
授課老師：劉奕汶教授, 李祈均教授

助教：張蘭云、鄭語芳、楊晶宇、林蔭澤

Outline:

- MFCC feature extraction overview.
- Elaboration on some components.
- Your lab demo tasks this week.
- Report questions.

MFCC extraction flow chart



Component(1): **Pre-emphasis**

- On time domain: $y[n] = x[n] - 0.95\, x[n-1]$

- On frequency domain: $H(z) = 1 - 0.95z^{-1}$

Bonus question 1:

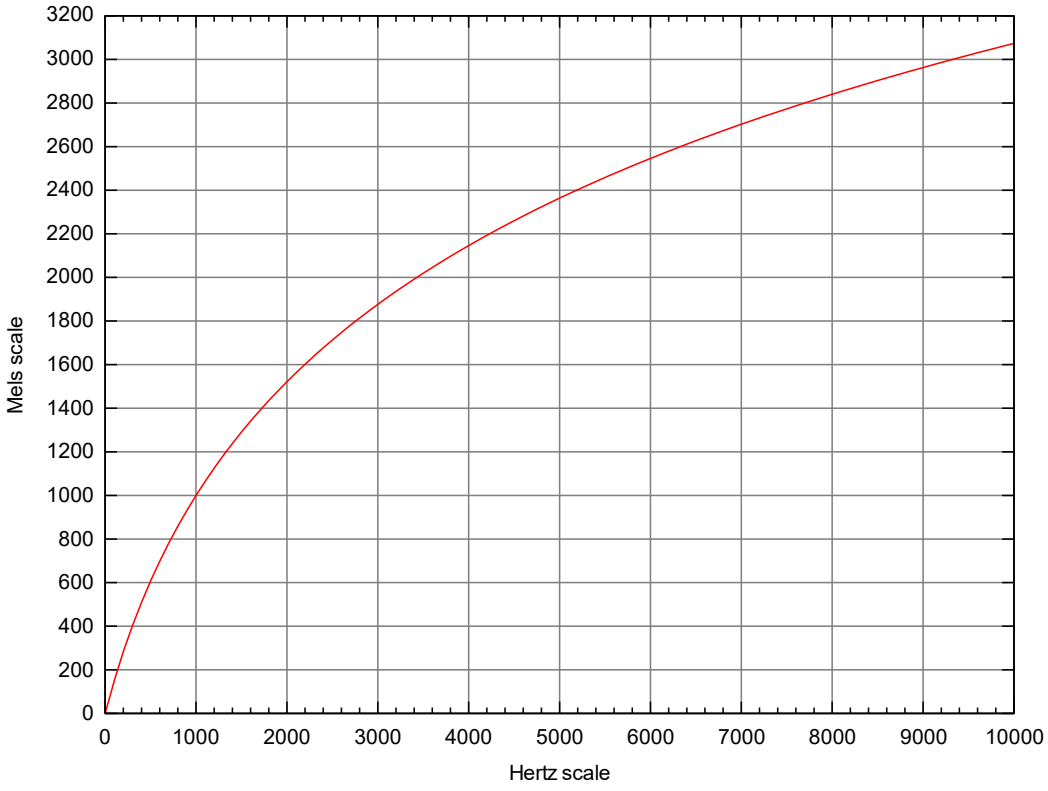
Plot the magnitude curve of H(z) with coefficients 0.95, 0.99, and 0.65 in MATLAB. Change the pre-emphasis coefficient in your lab code correspondingly and observe the difference on the spectrogram. How does the coefficient influence the results?

Component(2): **Mel-scale**

- Conversion:

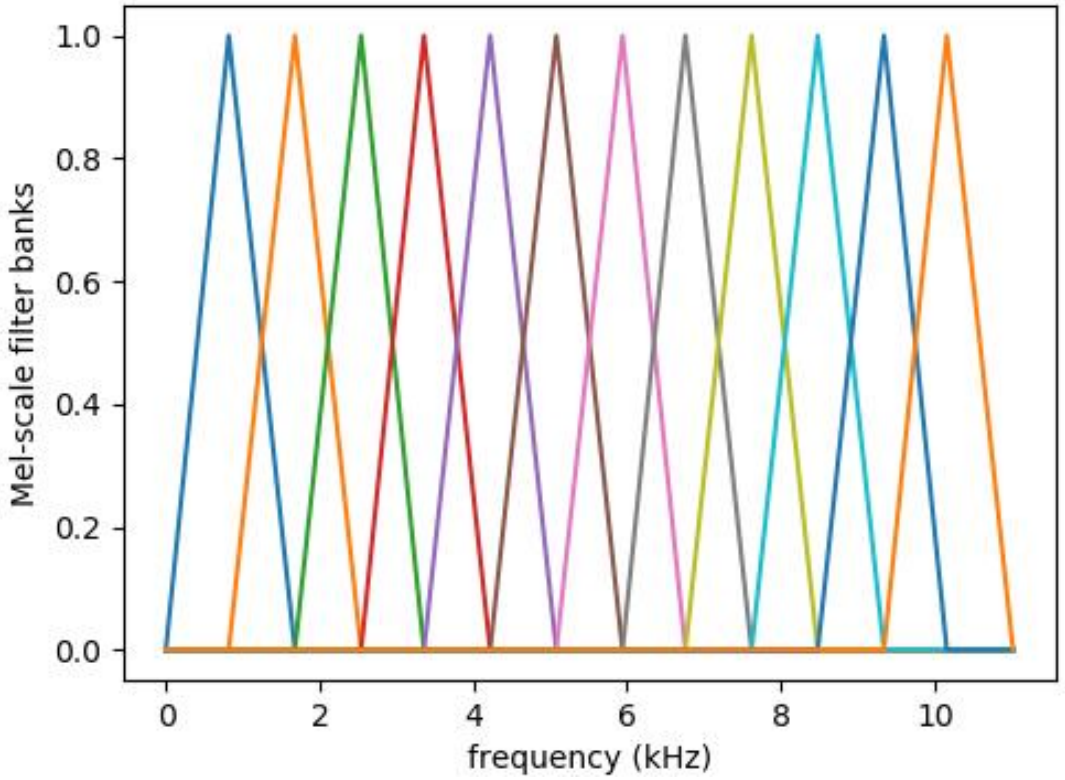
$$Mel = 2595 \times log_{10}(1 + \frac{f}{700\ Hz})$$

- Human perceptual scale of pitches

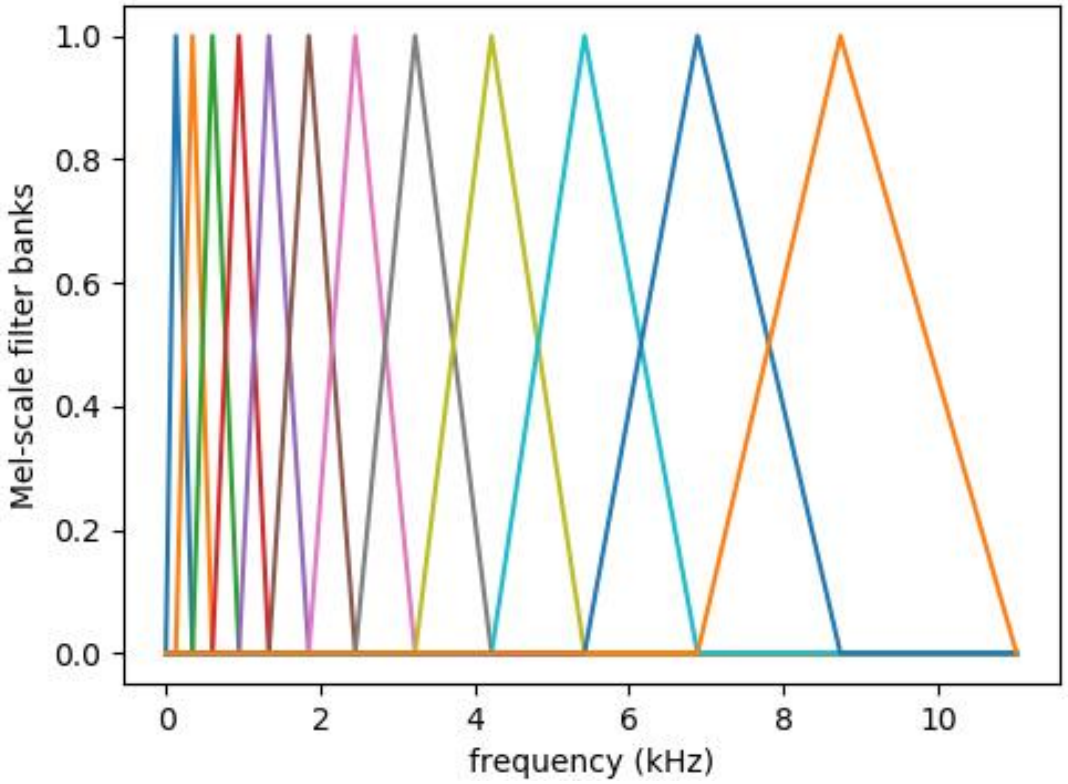


Component(3): **Mel-scale filtering**

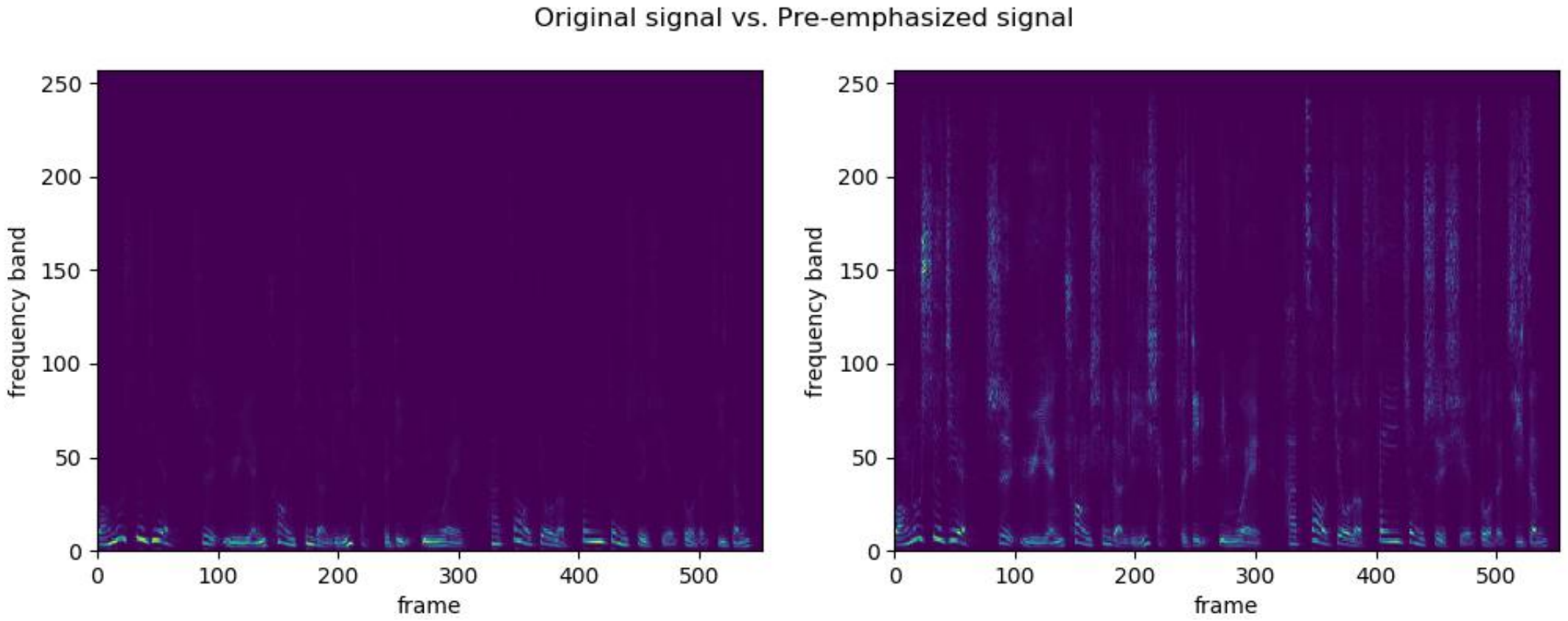
Hertz-scale Filtering Banks:



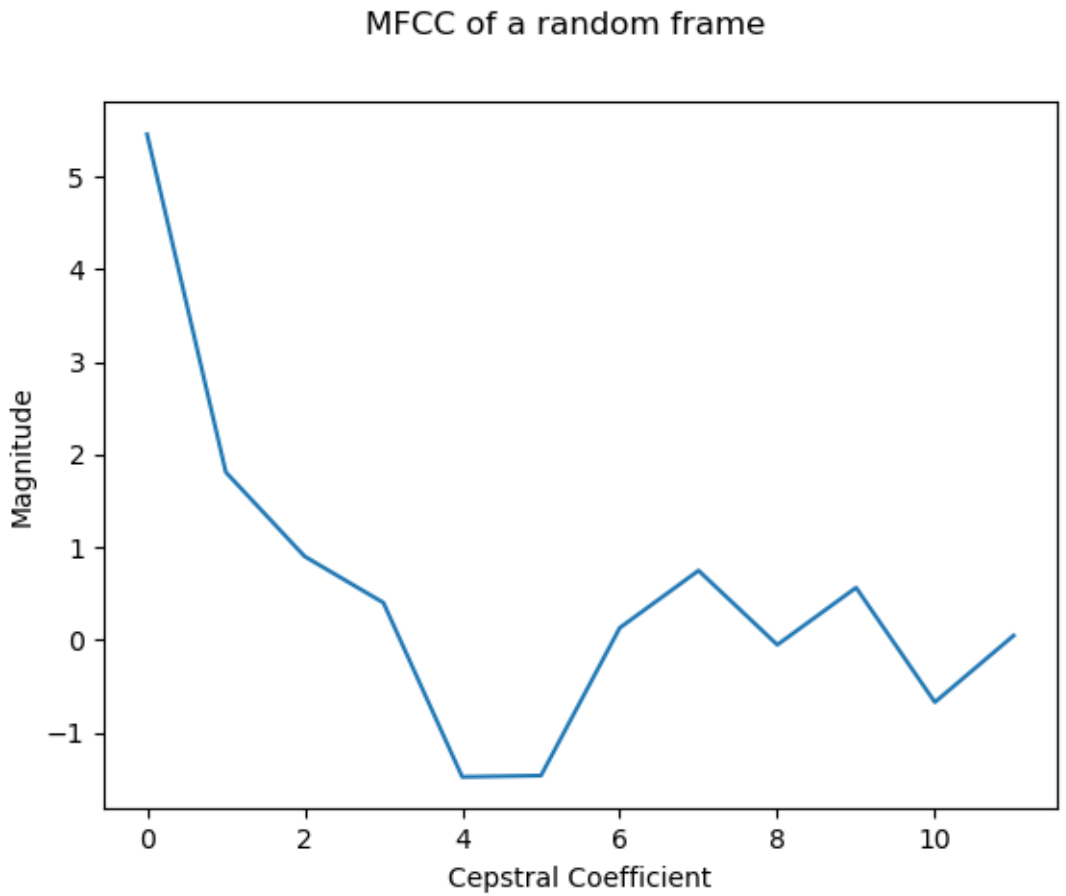
Mel-scale Filtering Banks:



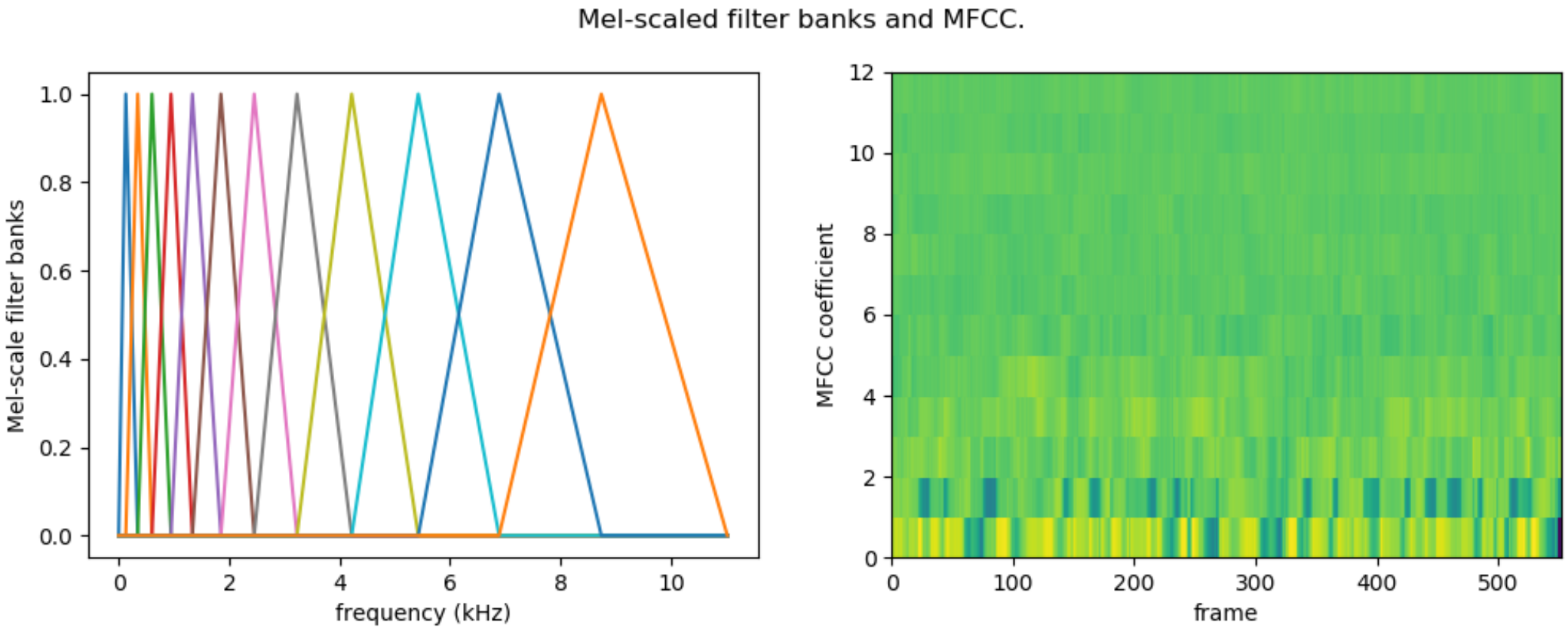
Demo(1): **Effect of pre-emphasis**



Demo(2): MFCC of a random frame



Demo(3): Filter and MFCC heatmap



Report questions:

1. Give an intuitive and a mathematical explanation of time-domain pre-emphasis; suggest why this step is beneficial for audio analysis.
2. Compare the MFCC of a chosen frame under different number of banks/coefficients. Compare the MFCC heatmap under different number of banks/coefficients. Is it the more the merrier? Why or why not? In case of this audio file, what is the number of banks you would choose? Why?
3. **Bonus 2:** notice in the code there is a discrepancy between our chosen number of FFT frequency quantization and the number of frequency channels in the output figure. Explain why the number of frequency channels halved and plot the spectrogram in which you perform conventional FFT. (hint: check out original line 22 of Lab1_function_student.py)