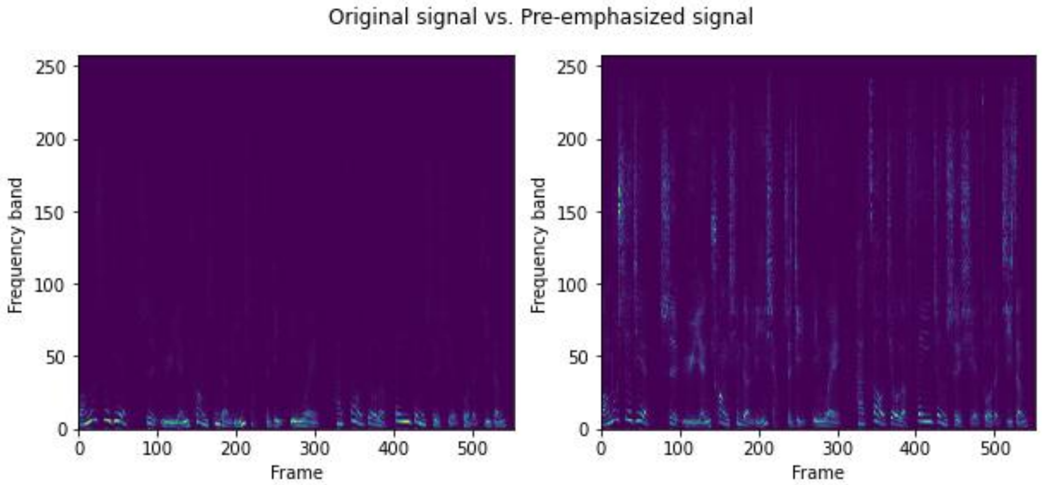
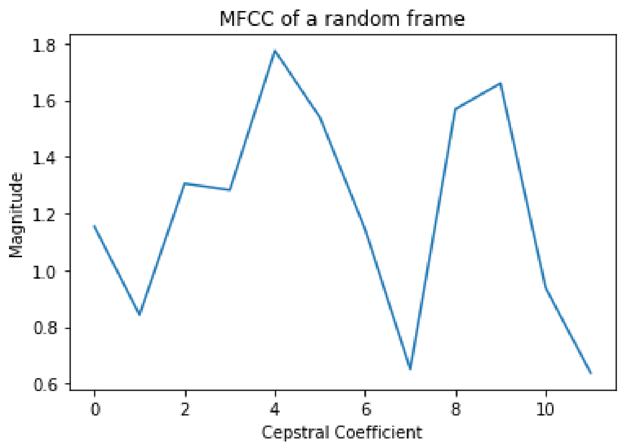
Digital Signal Processing Laboratory
Lab 9 Audio & Speech: MFCC

Results

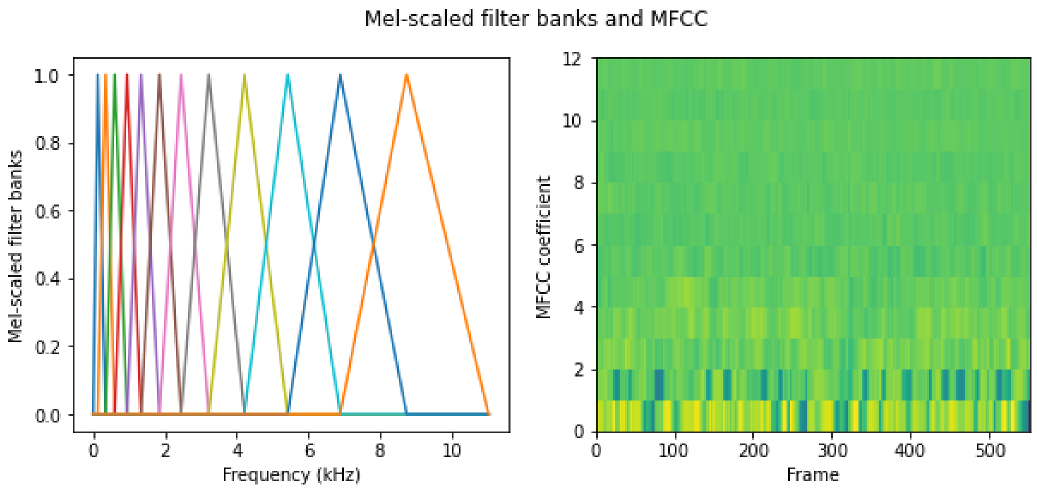
Demo 1：



Demo 2：



Demo 3：

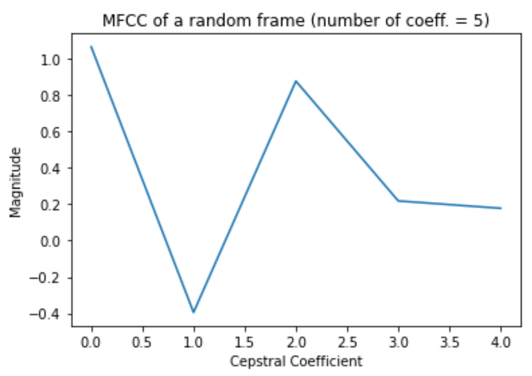


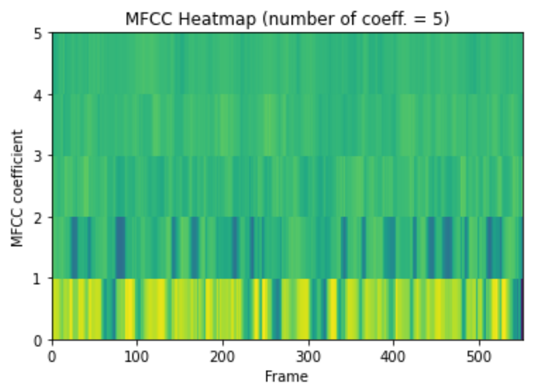
1. **Give an intuitive and a mathematical explanation of time-domain pre-emphasis; suggest why this step is beneficial for audio analysis.**

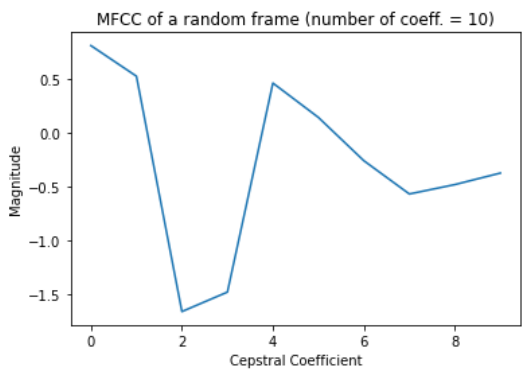
$H(z) = 1 - 0.95z^{-1}$ 為一高通濾波器，由於人耳對於部分頻率的訊號較為敏銳，且為了消除發聲過程中聲帶和嘴唇的效應，此 pre-emphasis 步驟可用來補償語音信號受到發音系統所壓抑的高頻部分。而在經過了此步驟之後，聲音會變的比較尖銳清脆，但是音量也會變小。

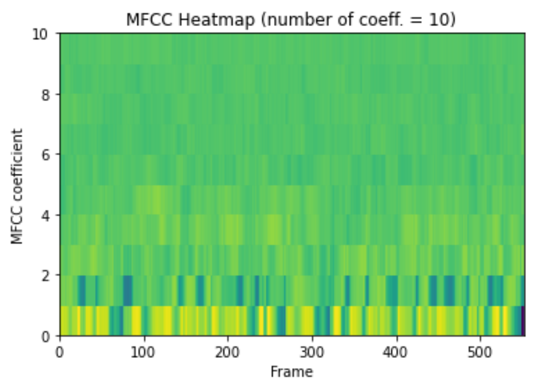
2. **Compare the MFCC of a chosen frame under different number of banks/coefficients. Compare the MFCC heatmap under different number of banks/coefficients. Is it the more the merrier? Why or why not? In case of this audio file, what is the number of banks you would choose? Why?**

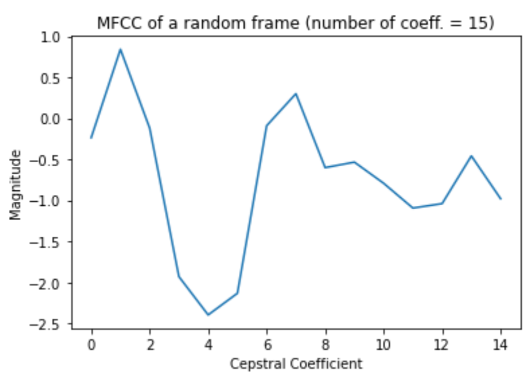
由下表結果可以發現，當 coefficient 數量愈多時，可以保留愈多每個 frame 的資料，因此每個 Cepstral coefficient 之間愈連續，且其 spectrogram 也較連續。但若是 coefficient 數量取愈高時，其增加的效果會愈來愈有限，且徒增計算量，因此我會將 coefficient 數量設在 10 到 15 之間，詳細會視音訊而異。

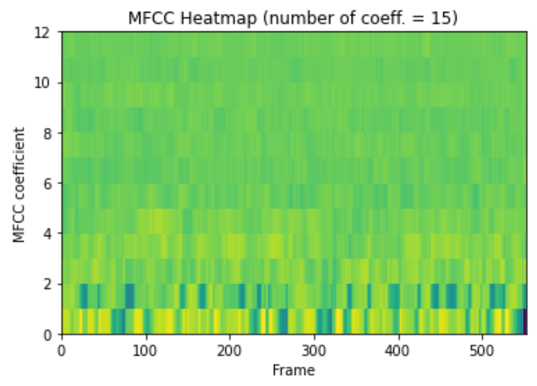


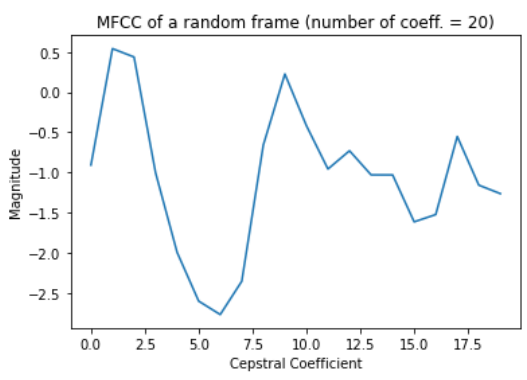


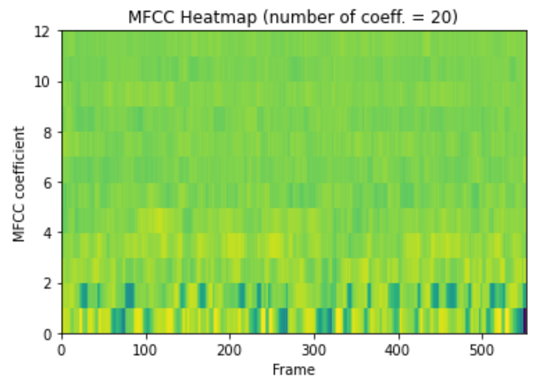






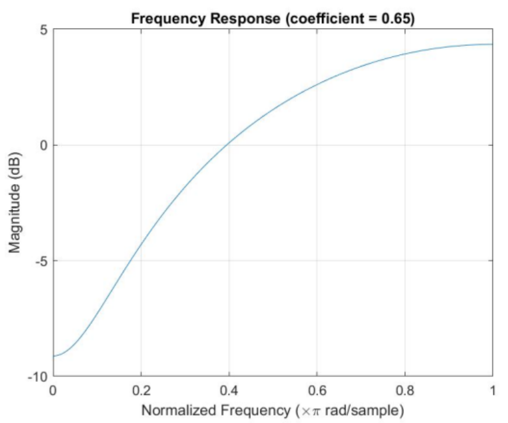
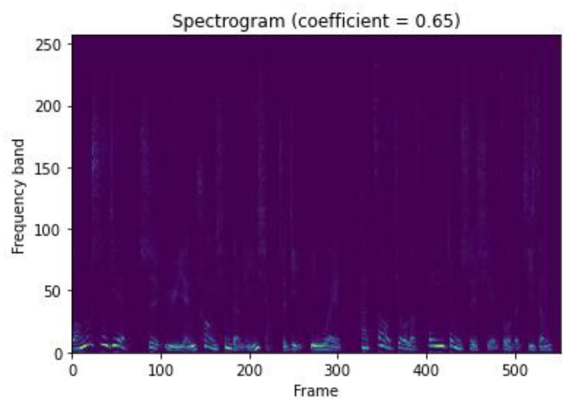


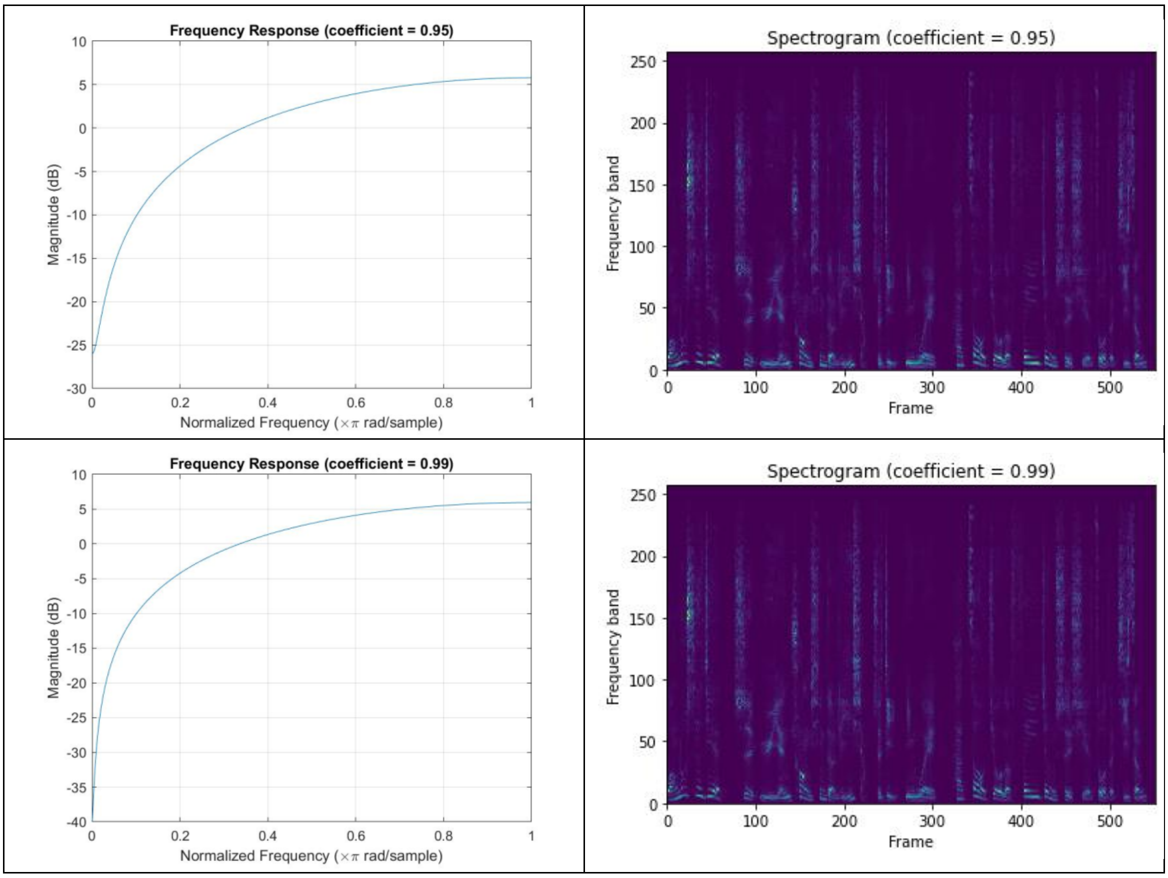




3. <u>Bonus 1:</u> **Plot the magnitude curve of H(z) with coefficients 0.95, 0.99, and 0.65 in MATLAB. Change the pre-emphasis coefficient in your lab code correspondingly and observe the difference on the spectrogram. How does the coefficient influence the results?**

　　結果如下表所示，可以發現當 coefficient 愈大時，在高頻的部分放大愈多、在低頻的部分縮小愈多，且 cutoff frequency 愈靠近低頻，因此愈接近理想的高通濾波器。而從其對應到的 spectrogram 可以看出當 coefficient 愈大時，高頻的部份愈明顯，而低頻的部分會降低一點。

Frequency response of H(z)	Spectrogram
	



4. <u>Bonus 2:</u> 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 Lab9_function_student.py)