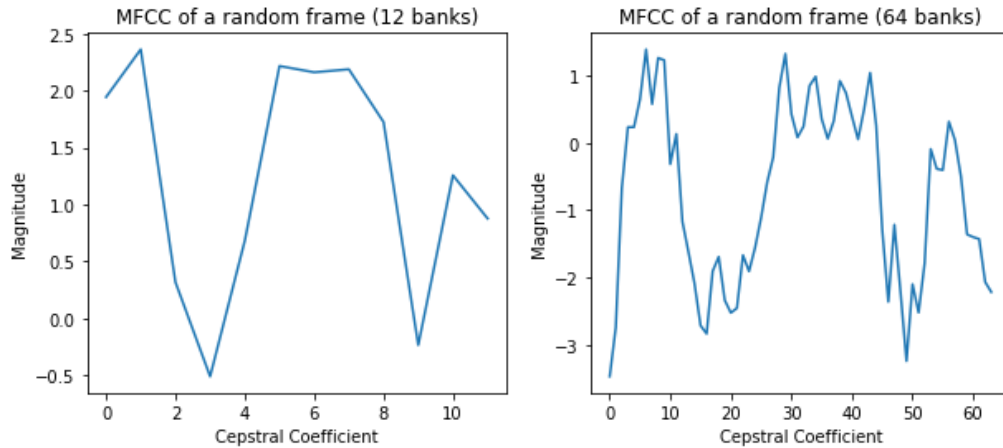


Digital Signal Processing Laboratory

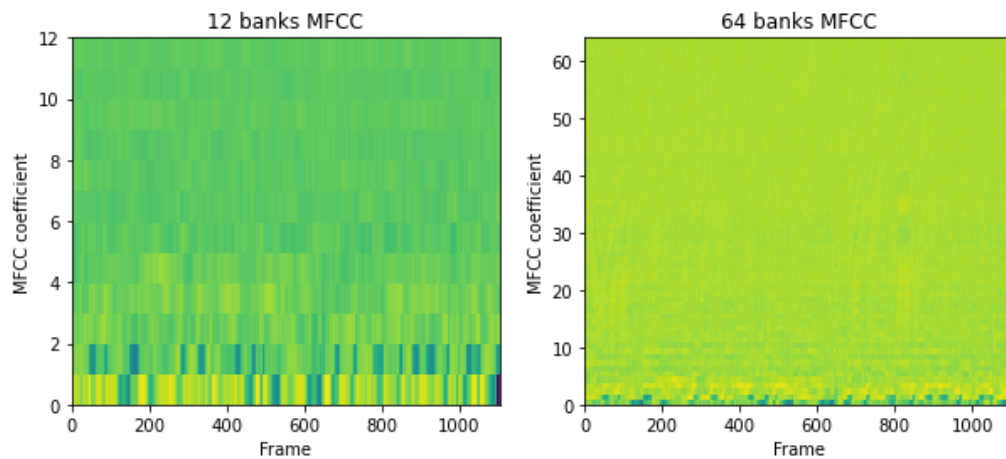
Lab 9 Audio Reconstruction from MFCC

Results

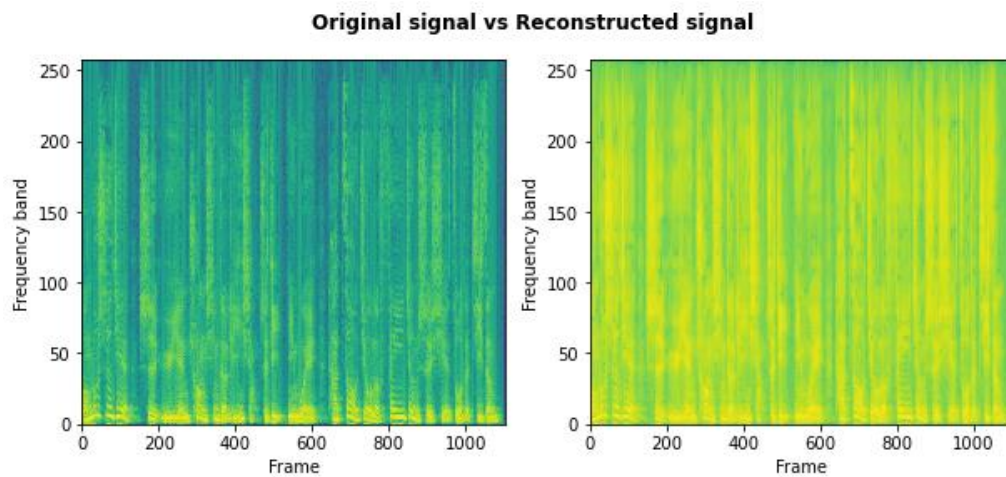
Demo 1 :



Demo 2 :

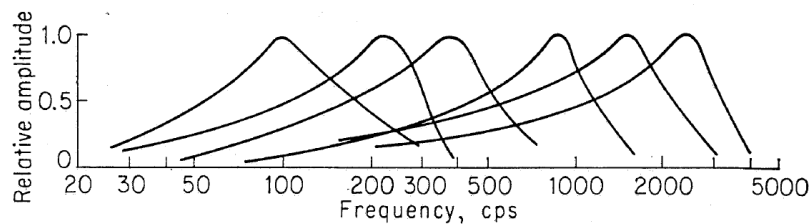


Demo 3 :



1. Why not using rectangular filters for “energy” calculation?

因為經過實驗後發現，人耳的構造對於某幾個特定頻率的聲音較為敏感，而其程度會隨著距離此頻率的距離愈遠而類似線性的下降，如下圖所示，因此我們利用 triangular filter 來近似此特性。



2. Why should mel-filters be overlapping?

因為人耳並非只對某些特定頻率的聲音敏感，而是其敏感程度會隨著頻率而有所起伏，因此利用不同 filters 互相 overlapping 的方式，可以達成近似於人耳的聽覺。

3. Why are high-frequency MFCCs usually abandoned when doing speech recognition?

因為人耳對於低頻率的訊號較高頻率敏感，且人類說話的聲音也都落在低頻的區域，因此高頻的 MFCC 通常不會被留下以降低計算量。

4. Do you think MFCC is good for speaker identification purposes? (How about baby-cry detection?)

我認為 MFCC 比較適合用來做語音辨識而不適合直接用來做語者辨識，原因在於它主要是將每一小段時間的頻率成分做特徵萃取，因此他可以保留不同頻率隨時間變化的資訊，而我們在說不同的文字時，就是在發出不同頻率成分的聲音，所以 MFCC 適合做 speech recognition。但若不同語者說了相同的文字，他們發出聲音的頻率成分其實是相近的，差別只有在音色上，而這樣的差別較難直接用 MFCC 分辨出來，可能需要再搭配一些機器學習或是深度學習的分類器才有辦法進行 speaker recognition。至於 baby-cry detection 我認為是可以利用 MFCC 達成的，因為 baby-cry 的聲音比起正常沒有 baby-cry 時會有特定的頻率成分，而這樣的成分可以用 MFCC 特徵萃取出來，因此可以做 detection。

5. If two sounds have similar MFCCs, does that imply they sound similar to our ears?

由於 MFCC 描述了一段聲音訊號在每段不同時間所含的頻率成分，它記錄了時間與頻率 domain 的訊號，而我們在辨別聲音時雖然也會考慮到其 time domain 與 frequency domain 上的資訊，但最重要的音色卻是 MFCC 沒辦法得知的，因此我認為就算兩個聲音訊號有著相似的 MFCC，並不代表這兩個聲音在我們聽起來會相似。