

Lab11: Linear Prediction and Time-variant Filtering

Lab objectives

1. 實作線性預測與時變濾波以產生激發訊號。
2. 消除 frame-rate artifacts，比較有無偷看，有無使用 hann window，有無消除多預測的 p 個點，這些變因對音質的影響。
3. 找出最佳的線性預測變數量(p 值)來消除語音中的母音。
4. 比較各種聲音濾出的激發訊號。
5. 嘗試在頻域上實作此線性預測方法(未完成)。

Method

這裡先對上面提到的各個方法做更詳細的說明。注意，作業生成檔案名稱會分別使用 LB(look back), OL(overlap), TR(tail removing)標註有無使用這些方法。

● 偷看法(Look back)

當初想到這個方法一開始是因為思考到，線性預測的原理是拿前 p 個值去預測現在的值，那如果現在是第一個值，那該如何計算呢？於是我去搜尋 lpc() 函式的說明，發現 Matlab 會在訊號前後都補上 p 個零來計算最小方差解。但實作上，為了讓前 p 個值的預測更為準確，我們要補上 p 個真正的訊號值(前面一個 frame 的最後 p 個值)，預測係數與生成激發函數時皆會使用到補上的這些值。

$$X = \begin{bmatrix} x(1) & 0 & \cdots & 0 \\ x(2) & x(1) & \cdots & \vdots \\ \vdots & x(2) & \cdots & 0 \\ x(m) & \vdots & \vdots & x(1) \\ 0 & x(m) & \cdots & x(2) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & x(m) \end{bmatrix}, \quad a = \begin{bmatrix} 1 \\ a(2) \\ \vdots \\ a(p+1) \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

photo from Matlab lpc function documentaion

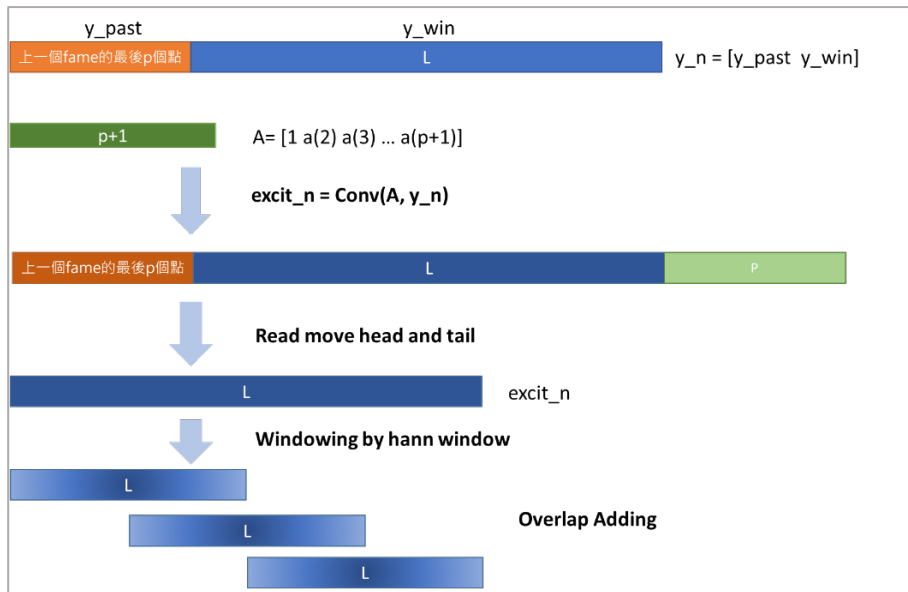
● Hann Window 50% overlap add

使用 Hann Window 且進行 50% overlap 的目的是為了將每個 frame 之間的不連續以類似 fade-in / fade-out 的方式解決。實作方式是在計算完每個 frame 的激發訊號後，將其與相同長度之 hann 窗進行內積。而平移步長只有 frame 長的一半，因此每次預測的激發訊號之間會有一半的重疊，而做法就是直接將重疊處相加。

● 消除多預測的 p 個點(Tail Removed)

由於每次預測出係數為一個長度為 p+1 的陣列，故與訊號做摺積後會比原訊號多出 p 個點，針對這 p 個點，有兩種可行的方式，第一個是將其保留，但將其與下個 frame 的開頭 p 個點做平均；第二個方式是將其直接消除。

- 方法示意圖：



Results and discussion of report questions

1. 消除 frame-rate artifact

這裡我有實做過多個方法，且有存檔以供後續比較。設備：鐵三角 m30x。

主要噪音：(1) 快速達達聲(frame-rate artifact) (2) 鞭炮聲，廣播電台調頻聲

檔名縮寫：LB: look back, 50OL: 50% overlap, TR: Tail Removed, 32ms 代表的是 framelen

a. 有無 overlap

比較檔案：

m1_sing_new/excit_p20_LB_noOL_TR_32ms , m1_sing_new/ excit_p20_LB_OL50_TR_32ms

心得：有 overlap 50%的激發訊號聽起來鞭炮聲較小，且最後幾個音音量逐漸遞增的共鳴感受有傳達出來，反之沒有 overlap 的最後音量遞增反而聲音更破。

比較檔案：

m1_sing_new/excit_p20_noLB_noOL_TR_32ms , m1_sing_new/ excit_p20_noLB_OL50_TR_32ms

心得：因為沒有 lookback 所以兩者皆有達達聲，且達達聲音量遠大於鞭炮聲。有趣的是 50% overlap 並沒有將達達聲消除，反倒讓達達聲的頻率變成了兩倍。

b. 有無 lookback

比較檔案：

m1_sing_new/excit_p20_noLB_noOL_TR_32ms , m1_sing_new/ excit_p20_LB_noOL_TR_32ms

心得：

沒有使用偷看法的聲音明顯產生出快速的「達達」聲，跟前面提到的鞭炮聲有所不同，此聲音更有規律性且聲音明顯，因此推測為目標要清除的 frame-rate artifact。後來為了驗證，我以波形顯示軟體確認，這些 impulses 的週期的確與 framelen 相同。

c. 有無 tail remove

比較檔案:

m1_sing_new/excit_p20_LB_noOL_TR_32ms , m1_sing_new/ excit_p20_LB_noOL_noTR_32ms

心得:

在有偷看的情況下，如果摺積多出來的 p 個點沒有去掉，會產生達達聲。

比較檔案:

m1_sing_new/excit_p20_noLB_noOL_TR_32ms , m1_sing_new/ excit_p20_noLB_noOL_noTR_32ms

心得:

這組比較有趣，雖然兩者皆有產生達達聲，但沒有做偷看也沒有去尾巴的音檔中達達聲竟然比沒做偷看但有做去尾的音檔的音量還小，推估可能是殘留的尾巴與預測的失準產生了一個抵銷的狀況。

2. 與第三題一起陳述

3. 調整適當的 p 值，且使用自己錄製的聲音

從處理過程中的頻譜圖可得知，p 值越大則預測出的濾波器的頻譜(包絡線)更可以貼近訊號的頻譜，聽感上則越聽不出來原本的音色。對於 a-male-singing 這份音檔，我認為 p 至少要調超過 70 才能明顯降低母音 a 的音色，但其實 p 在 20 時，濾波器已經可以充分描繪出聲音頻譜的包絡線。而對於我模仿的小喇叭音色，我發現這個音色的諧音成分非常明顯，故要使用到 p 值超過 50 才能完全濾掉它的音色，而且這個效果滿強的，產生出來的積發訊號幾乎聽不出來原本的音色。我猜測原因是因為樂器類的聲音周期性成分佔比較大，一但被濾掉就難以辨認出原本的音色。相對的，人聲有週期性成分(母音)和非周期性成分(子音，氣音等等)，雖然母音被濾掉但靠著剩下的子音我們還是有能力辨認出原本的聲音，這也許就是為甚麼我們在圖書館還能夠靠氣音溝通。

之後還有實作拿自己人聲打擊的聲音來濾出激發訊號，大鼓的聲音聽起來像是打在一片板子上，小鼓就像是打在一片紙上發出的聲音。

Conclusion

1. 偷看法可以有效清除 frame-rate artifact，在摺積產生的尾端多的值有被去除的情況下。
2. p 值越大，產生的激發訊號的殘留音色越少，但也會增加運算資源。
3. 樂器等諧音成分明顯的音色，濾出的激發訊號與原本的聲音相差甚遠。