利用小波聽覺分頻處理與訊號子空間分解於車內噪音消除*

王駿發[1] 楊宗憲[2] 張凱行[3]

國立成功大學電機工程研所

wangjf@csie.ncku.edu.tw[1] chyang@icwang.ee.ncku.edu.tw[2]

casey019@ms55.hinet.net[3]

摘要 在傳統的訊號子空間語音強化方法(Signal Subspace Speech Enhancement Method)中，其主要是利用噪音能量是均勻分佈於訊號所在的向量空間而語音訊號能量則是分佈於某一子空間的特性，藉由特徵分解(Eigen-Decomposition)來分析出語音訊號及背景噪音，來進行噪音消除。而在車內噪音環境中，噪音能量的分佈在低頻帶為最多延伸到高頻則逐漸較少，單一的訊號子空間的語音強化方法已不能更有效的消除位在低頻帶的背景噪音。本論文提出一個基於人耳聽覺特性的分頻處理，並結合訊號子空間強化方法來克服此一問題。實驗的驗證，則是採用 TAICAR 車內語音資料庫來進行，實驗結果說明本文所提出的方法比起傳統訊號子空間強化法，更適用於車內噪音的消除，低頻噪音的消除也更明顯。

1. 前言

隨著汽車導航系統的日漸普及，除了提供汽車行車資料及娛樂外，藉由結合行動電話的無線通訊功能，更讓汽車儼然已經變成隨時可獲知各種生活資訊的行動中心。在汽車內傳統的人機介面是採用觸碰式螢幕，在行車的狀況下，這樣的介面是不夠安全的，而隨著即時語音辨識技術的日趨成熟，人機介面必定是朝著語音對話的操控方式改進。在行車環境中，充斥各種噪音，對於語音辨識系統而言，這些背景噪音會嚴重地影響辨識結果。因此，一般的辨識系統都需使用手持式或頭戴式麥克風，來促成近距離的錄音，以避免背景噪音的干擾。然而，使用這些錄音設備會對駕駛或者乘客造成不便，所以提供一個在行車環境下能實行遠距離錄音並具有抗噪音能力的麥克風系統，是有其需求。本論文提出一個利用小波聽覺分頻處理與訊號子空間分解來達成車內背景噪音消除的目的。

Ephraim 和 Van-Trees 於 1995 年提出一套基於訊號子空間分解的語音強化系統 [1]，利用噪音能量是均勻分佈於訊號所在的向量空間而語音訊號能量則是分佈於某一子空間的特性，藉由特徵分解來分析出語音訊號及背景噪音，並進一步用一線性估測器來處理得到強化後的語音。由於特徵分解的運算複雜度高，在本論文中採用子空間追蹤(Subspace Tracking)的方式來做特徵分解，這個演算法稱為 PAST (Projection Approximation Subspace Tracking, PAST) [2]，以期能符合即時(Real-Time)的應用。而在車內噪音環境中，噪音能量的分佈在低頻帶為最多延伸到高頻則逐漸較少，在實驗過程中發現，單一的訊號子空間的語音強化方法已不能更有效的消除位在低頻帶的背景噪音。因此，本論文提出一個基於人耳聽覺特性的分頻處理，並結合訊號子空間

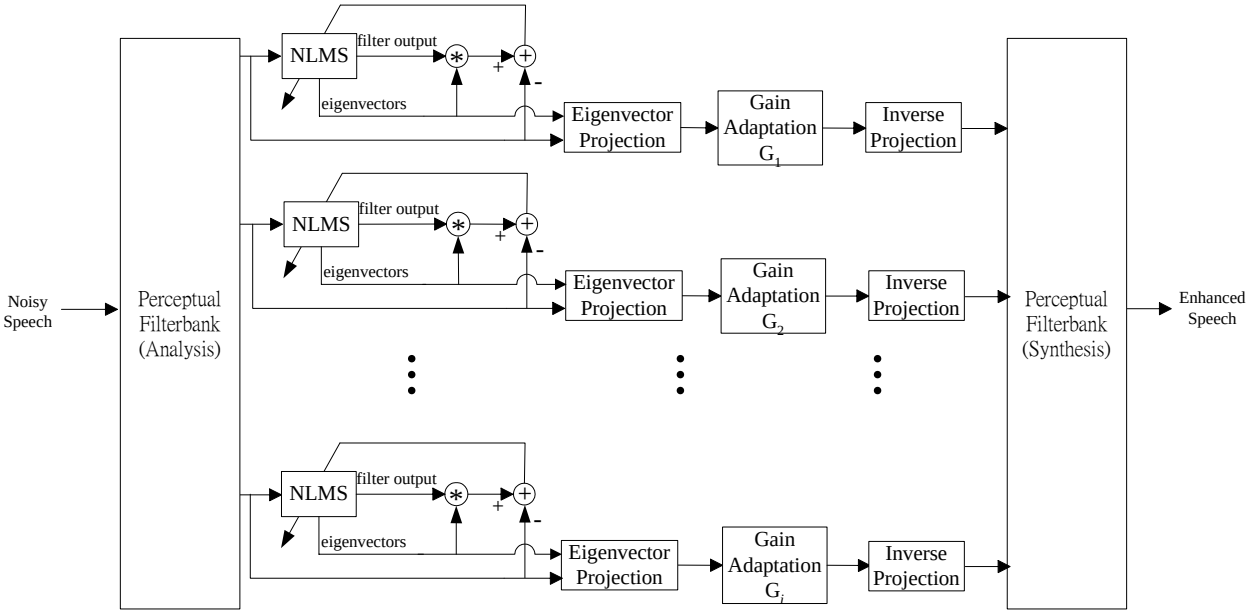
* This work was supported by the National Science Council of the Republic of China, Taiwan, Contract Nos. NSC 92-2213-E-006-022

強化方法來克服此一問題。此一聽覺分頻處理係利用小波轉換(Wavelet Transform)來實現，藉由小波將聲音分解成多個頻帶，而各個子頻帶的分佈則符合人耳聽覺響應的特性，各子頻帶的訊號再經由子空間方法進行噪音消除，再由小波反轉換合成各子頻帶的訊號，進而得到強化後的語音。實驗的驗證，則是採用 TAICAR 車內語音資料庫來進行，實驗結果說明本文所提出的方法比起傳統訊號子空間強化法，更適用於車內噪音的消除，低頻噪音的消除也更明顯。

本論文的章節結構如下：第二節是所提出來的噪音消除系統架構，包含小波聽覺分頻處理、訊號子空間語音強化以及子空間追蹤法之描述；第三節說明實驗結果，包含 TAICAR 車內語音資料庫的介紹以及本文所提出之方法跟其它訊號子空間語音強化法之比較；最後，第四節則是結論與討論。

2. 系統架構

本論文所提出的車內噪音消除系統，如圖一所示。在系統前端，麥克風所錄到的雜訊語音，經由小波聽覺濾波組(Perceptual Wavelet Filterbank)分成數個子頻帶訊號，各個子頻帶則由訊號子空間語音強化來進行噪音消除的處理，而訊號子空間的拆解則是由子空間追蹤法來完成。由子空間追蹤法所估算出來的特徵值(Eigenvalue)，則用以計算各個子頻帶訊號的增益值。語音強化的處理為將子頻帶訊號經過特徵向量(Eigenvector)投影轉換後，由增益值來調整其訊號大小，再經過反轉換來得到強化後的語音訊號。以下各小節則對小波聽覺分頻處理、訊號子空間語音強化以及子空間追蹤法做一描述。



圖一：車內噪音消除系統架構。

2.1. 小波聽覺分頻處理

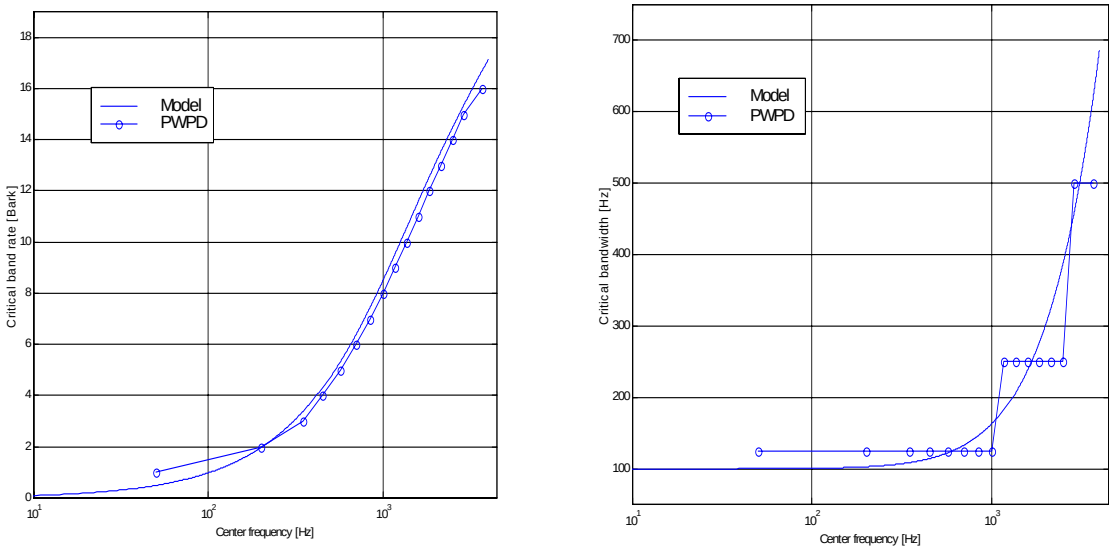
具聽覺感知的小波轉換(Perceptual Wavelet Packet Transform, PWPT)是改良自傳統小波轉換，使語音信號經 PWPT 分解後的各個子頻帶信號的頻寬接近人耳的聽覺響應 [3]，描述人耳聽覺響應的參數主要有巴克頻譜（Bark）以及關鍵頻寬（Critical Bandwidth），表一為人耳聽覺關鍵頻寬的分佈情形。圖二(a)及圖二(b)分別是在 4KHz 內，人耳的聽覺的巴克頻譜及關鍵頻寬曲線圖 [4, 5]。因此，本論文所設計的聽覺分頻處理即朝

此二曲線設計，圖二(a)及圖二(b)內亦標示了利用小轉換逼近巴克頻譜及關鍵頻寬的曲線圖。

由小轉換逼近巴克頻譜及關鍵頻寬是藉由調整小波轉換的樹狀結構來達成。依據表一的關鍵頻寬分佈情形，適當對訊號做高低頻的分解，使得子頻帶訊號的頻率分佈跟關鍵頻寬近似。圖三為所使用的具聽覺感知的小波轉換分解架構圖，其中輸入訊號經五個階段，共 16 次的高低頻分解。

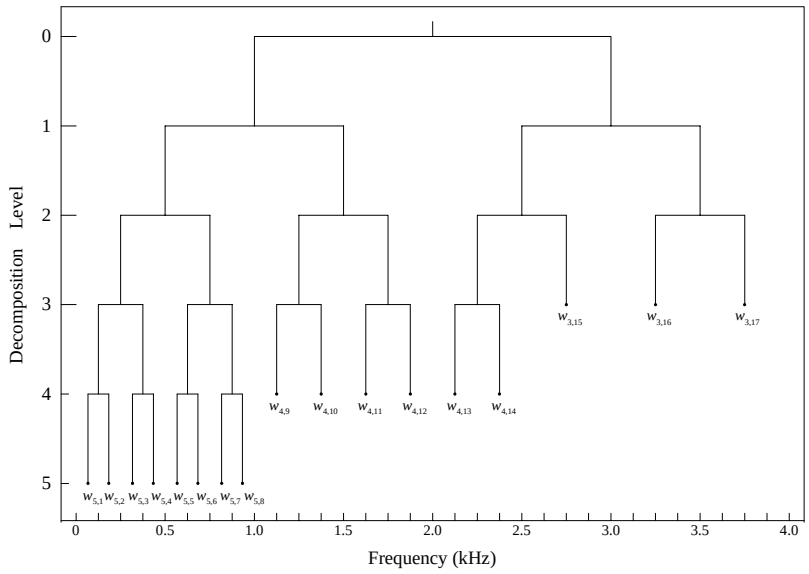
表一：關鍵頻寬分佈。

Critical Band Number	Center Frequency (Hz)	CBW	Lower Cutoff frequency (Hz)	Upper Cutoff Frequency (Hz)
1	50	-	-	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700

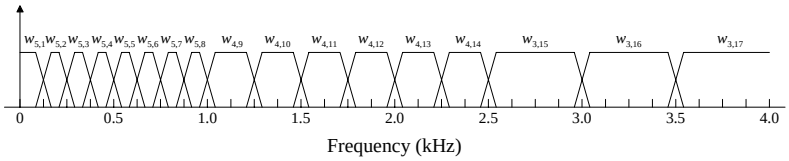


(a) (b)

圖二：4KHz 內(a)人耳的聽覺的巴克頻譜及(b)關鍵頻寬曲線圖。



(a)



(b)

圖三：(a)聽覺感知分頻的樹狀結構及(b)各子頻帶的頻寬。

2.2. 語音子空間強化

在本論文中採用的語音強化演算法為訊號子空間分析法 [1]。在子空間分析法中，從含有雜訊的語音訊號中的共變異數矩陣(Covariance Matrix)由特徵分解求出其特徵向量及特徵值，接著利用這兩個資訊和一個線性預估器將背景噪音消除。

一個 K 維語音訊號向量 y 其線性模型為

$$y = \sum_{m=1}^{M} s_m V_m \qquad , M < K \tag{1}$$

其中 $s_1, ..., s_M$ 為平均值為零的隨機變數，而 $V_1, ..., V_M$ 為基底。訊號分佈的向量空間其維度為 K，而語音訊號分佈的子空間其維度為 M，$M < K$。式子(1)可以表示為 $y = Vs$，$V \equiv [V_1, ..., V_M]$ 為 $K \times M$ 的矩陣，其秩(Rank)為 M，且 s 為一行向量表示為 $s \equiv (s_1, ..., s_M)^T$。語音訊號的共變異數矩陣 $R_y = VR_sV^T$ 其秩為 M，R_s 為 s 的共變異數矩陣並假設其為正定矩陣(Positive Definite Matrix)。$M < K$ 的性質使得在 K 維語音訊號 y 中，R_y 有 $K - M$ 個特徵值為零，這個對於在以子空間演算法做語音強健中極為重要。

令 w 為 K 維向量表示背景白色噪音，其平均值為零。其共變異數矩陣 $R_w = E\{ww^T\} = \sigma_w^2 I$。白色噪音的共變異數矩陣其秩為 K，也就是說它會佈滿整個歐式空間 R^K 中。因此，對於背景為白色噪音的雜訊語音

訊號，整個 K 維的向量空間由 M 維的訊號子空間及 K 維的噪音子空間所組合而成，其中可以將 $K-M$ 的特徵值為零所對應的子空間去除掉，而剩下的 M 維的雜訊子空間，可以用一線性預估器將其乾淨語音粹取出來。

底下將說明線性預估器的求取，假設雜訊語音為 $Z(n)=Y(n)+W(n)$，$W(n)$ 為 K 維的背景噪音向量，$Y(n)$ 為 K 維的語音向量，n 為訊號音框的索引。令 $H(n)$ 為一 $K\times K$ 的乾淨語音之線性預估器亦即

$$\hat{Y}(n)=H(n)Z(n) \tag{2}$$

則其預估錯誤訊號則為

$$\begin{aligned}\varepsilon(n)&=\hat{Y}(n)-Y(n)\\&=(H(n)-I)Y(n)+H(n)W(n)\\&=\varepsilon_{\mathrm{y}}(n)+\varepsilon_{w}(n)\end{aligned} \tag{3}$$

$\varepsilon_{y}(n)\equiv(H(n)-I)Y(n)$ 代表訊號的失真量，I 為單位矩陣(Identity Matrix)，$\varepsilon_{w}(n)\equiv H(n)W(n)$ 代表噪音的殘餘量。定義訊號失真能量及噪音殘餘能量分別為 $\bar{\varepsilon}_{y}^{2}(n)$ 、$\bar{\varepsilon}_{w}^{2}(n)$。則訊號失真能量表示為

$$\begin{aligned}\bar{\varepsilon}_{y}^{2}(n)&=tr(E[\varepsilon_{y}(n)\varepsilon_{y}^{T}(n)]),\\&=tr((H(n)-I)R_{y}(n)(H(n)-I)^{T})\end{aligned} \tag{4}$$

且噪音的殘餘能量為

$$\begin{aligned}\bar{\varepsilon}_{w}^{2}(n)&=tr(E[\varepsilon_{w}(n)\varepsilon_{w}^{T}(n)]),\\&=tr(H(n)R_{w}(n)H(n)^{T})\end{aligned} \tag{5}$$

$R_{y}(n)$ 及 $R_{w}(n)$ 分別為乾淨語音訊號及噪音訊號的共變異數矩陣。因要其訊號失真能量最小化，而最小化的情況要限制在噪音能量小於一個很小的常數，因此其最佳的線性預估器定義如下

$$H_{opt}(n)\equiv\arg\min_{H(n)}\bar{\varepsilon}_{y}^{2}(n),\qquad \text{Subject to}:\tfrac{1}{K}\bar{\varepsilon}_{w}^{2}(n)\le\sigma^{2} \tag{6}$$

其中 σ^{2} 是一個正的常數值。求解式子(6)，可利用拉氏乘子法(Lagrange Multiplier)，得到

$$L(H(n),\mu)\equiv\bar{\varepsilon}_{y}^{2}(n)+\mu(\bar{\varepsilon}_{w}^{2}(n)-K\sigma^{2}) \tag{7}$$

及

$$(\bar{\varepsilon}_{w}^{2}(n)-K\sigma^{2})=0,\qquad \mu\ge 0 \tag{8}$$

而 μ 為拉氏乘子。對式子(7)取梯度運算(gradient)，令其為零求解，則線性預估器可得到為

$$H_{opt}(n)=R_{y}(n)(R_{y}(n)+\mu R_{w}(n))^{-1} \tag{9}$$

由特徵值分解，式子(9)可表為

$$H_{opt}(n) = U(n)\Lambda_y(n)(\Lambda_y(n) + \mu\Lambda_w(n))^{-1}U^T(n) \tag{10}$$

特徵分解 $R_y(n) = U(n)\Lambda_y(n)U^T(n)$，$U(n)$ 為特徵向量的矩陣，$\Lambda_y(n)$ 為特徵值矩陣，$\Lambda_w(n)$ 為 $R_w(n)$ 的特徵值矩陣。令 $G(n) = \Lambda_y(n)(\Lambda_y(n) + \mu\Lambda_w(n))^{-1}$，則

$$H_{opt}(n) = U(n)G(n)U^T(n) \tag{11}$$

2.3. 子空間追蹤演算法

子空間語音強健最後的線性預估器須要雜訊語音的特徵分解，其運算複雜度高。所以在本論文中採用追蹤疊代的方式來逼近特徵值，這個演算法稱為 PAST (Projection Approximation Subspace Tracking) [2]。PAST 的演算法用來追蹤子空間的特徵值在許多文獻中被證明是準確且計算複雜度低的。若以子空間演算法，其運算複雜度為 $O(n^3)$，n 為輸入向量的維度，若子空間追蹤方法來做計算，其運算複雜度可以減少至 $O(nr)$，其中 n 為輸入向量的維度，r 為我們需要的特徵值暨特徵向量的的數目。

PAST 演算法其原理為對所給定的成本函數(Cost Function)求取最小值，成本函數的決定與共變異數矩陣有關，

$$J(u(n)) = \sum_{i=1}^{n} \beta^{n-i} \left\| Z(i) - u(n)u^T(n)Z(i) \right\|^2 \tag{11}$$

其中 $u(n)$ 為 K 維的向量，且 $0 \le \beta \le 1$ 為消散係數(Forgetting Factor)，β 可以使成本函數的極值所在會是下面定義的相關矩陣 $R_z(n)$ 中的特徵向量之一。

$$\hat{R}_Z(n) = \sum_{i=1}^{n} \beta^{n-i} Z(i)Z(i)^T \tag{12}$$

定義一個 $J'(u(n))$ 如下

$$J'(u(n)) = \sum_{i=1}^{n} \beta^{n-i} \left\| Z(i) - u(n)u^T(i-1)Z(i) \right\|^2 \tag{13}$$

$J(u(n))$ 和 $J'(u(n))$ 不同在於使用了 $u^T(i-1)$ 代替 $u^T(n)$，直覺的觀察，$J'(u(n))$ 可以被用來近似 $J'(u(n))$。因為語音訊號的統計特性為穩態(stationary)，也就是某個時間區段它變化的很慢所以 $u^T(i-1) \approx u^T(n)$。當 $i << n$ 時，β^{n-i} 會變得很小使得最後 $J'(u(n)) \approx J(u(n))$。我們可以用適應性梯度演算法將 $J'(u(n))$ 取梯度運算迭代求出特徵向量。其 PAST 演算法如下所示。

表二: PAST 演算法。

初始化： $d_i(0)=0, \beta=0.95$

$U(0)=[u_1(0)\,|\,u_2(0)\,|\,\ldots\,|\,u_k(0)]=I_k$

For $n=1,2,\ldots$ do

$Z_1(n)=Z(n)$

For $i=1,2,\ldots,k$ do

$vi(n)=u_i^T(n-1)Z_i(n)$

$d_i(n)=\beta d_i(n-1)+|v_i(n)|^2$

$E_i(n)=Z_i(n)-u_i(n-1)v_i(n)$

$u_i(n)=u_i(n-1)+T(n)E_i(n)\dfrac{v_i(n)}{d_i(n)}$

$Z_{i+1}(n)=Z_i(n)-u_i(n)v_i(n)$

end

end

輸出：

$U(n)=[u_1(n)\,|\,u_2(n)\,|\,\ldots\,|\,u_k(n)]$

其中 $d_i(n)$ 為子空間的特徵值。對於乾淨語音以及噪音為無相關的所以在特徵值上的分佈可以寫成

$$\Lambda_z(n)=\Lambda_y(n)+\Lambda_w(n) \tag{14}$$

所以只要求雜訊語音的特徵值且 $\Lambda_w(n)$ 為白色噪音其統計特性已知，可以將 $\Lambda_z(n)-\Lambda_w(n)$ 得到乾淨語音的 $\Lambda_y(n)$。一般來說噪音的統計特性不是穩態的。所以要完成估算 $\Lambda_w(n)$，通常都假設噪音在某段時間內變化很慢，因為實際一般環境中噪音得變化其實不大(如汽車內、室內冷氣聲…等)。所以在雜訊語音中，前一段的噪音資訊可以存起來給下一個語音段使用以求出 $\Lambda_Y(n)$。所以在用追蹤演算法估算 $\Lambda_w(n)$ 時用指數式的方式來疊加平均，以達到正確真實的 $\Lambda_w(n)$，其指數式的方式來疊加平均如下

$$\Lambda_w(n)=\beta\Lambda_w(n-1)+U^T(n)W(n) \tag{15}$$

β 為一平滑係數(smoothing factor)且 $W(n)$ 的值從之前的靜音區段選擇代用。所以可以計算噪音的能量在每個語音段之間，其計算出來的噪音能量資訊直接給下個語音段使用。

3. 實驗

對於本文所提出的方法，則是採用 TAICAR 車內語音資料庫來進行實驗的驗證。以下就對 TAICAR 資料庫做一介紹，接著對車內所蒐集的雜訊音檔進行噪音消除的實驗。

3.1. TAICAR 車內語音資料庫

在國外有很多的語音資料庫收集之方法，例如：日本的 CIAIR、歐洲的 SpeechDat 等等，然而在汽車環境下的語料收集卻是很少見，TAICAR 資料庫目標就在於收集汽車環境下的語料以方便各種語音處理技術之開發。例如：噪音補償技術、噪音下動態語音偵測技術、強健型語音辨識技術、語音調適技術等。語料的錄音參考國內執行過的大型計畫「MAT 語料收集」之作法，先由程式從 100 萬字的文字庫中挑選出能夠涵蓋所有國語基本音節的短詞、單字等，並加上英文、數字部分，總共這樣的語料有 360 份。為了實際記錄各種不同路況，錄音時分兩種路段：市區路段以及快速道路路段。市區路段下，時速為 0～50 公里；快速道路則需維持在 70～100 公里。

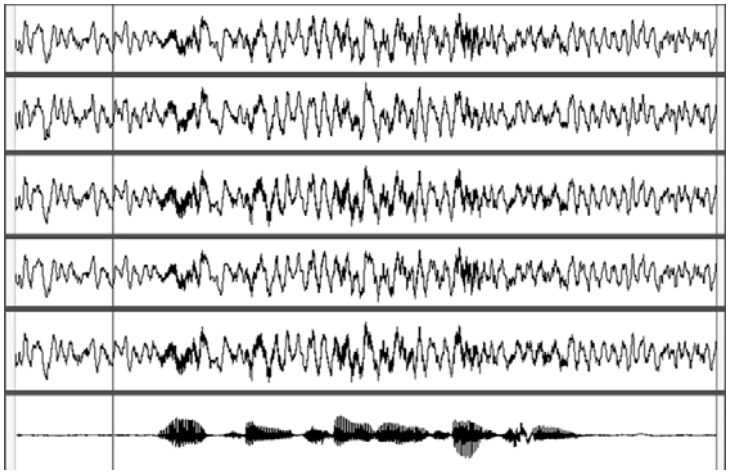
在車內錄音需考慮到便利性，因此以筆記型電腦為錄音的平台，配合上特殊硬體來進行錄音。所用錄音器材計有：

- 筆記型電腦: 負責主要的錄音程式之進行
- PCMCIA 介面之多頻道信號擷取卡：負責擷取多頻道的語音訊號
- 麥克風：1 支指向性(頭戴式，收錄乾淨語音)＋5 支全向性(收錄雜訊語音)：負責語音訊號的輸入
- 車輛：任意

車內的錄音軟體，可同時進行 6 個 channel 錄音，單音取樣: 16KHz , 16bits。在錄音之同時可標記路況、車速、語者性別、基本資料等 [6]。圖四為 TAICAR 車內音檔錄音情況，圖五則為所錄得音檔之時間波形。

(a) (b)

圖四：(a) 車內多麥克風配置及(b)語者與錄音設備。



圖五：車內六個麥克風所錄之時間波形。

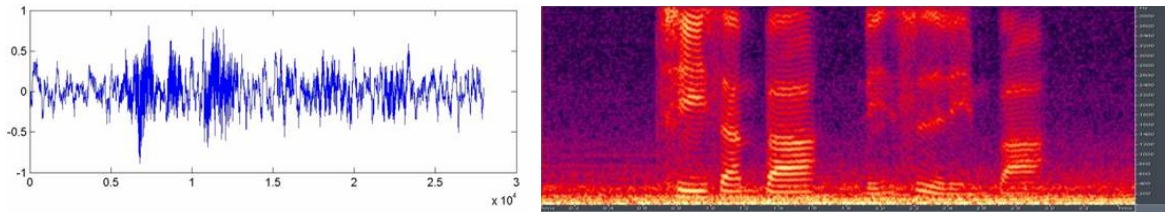
3.2. 效能評估

實驗的驗證以 TAICAR 音檔來做測試，對進行過噪音消除後之音檔進行評分。評分方式採人耳試聽為之(Mean Opinion Score, MOS)，給分等級為：5 為優，4 為好，3 為尚可，2 為略差，1 為不好。以本文所提之方法與另外兩種子空間分解方法作比較，其為子空間分解採用離散餘弦轉換(Discrete Cosine Transform, DCT)及採用 KL 轉換(Karhunen-Loeve Transform, KLT)。計有二十位試聽者給分，給分結果如表三所示。

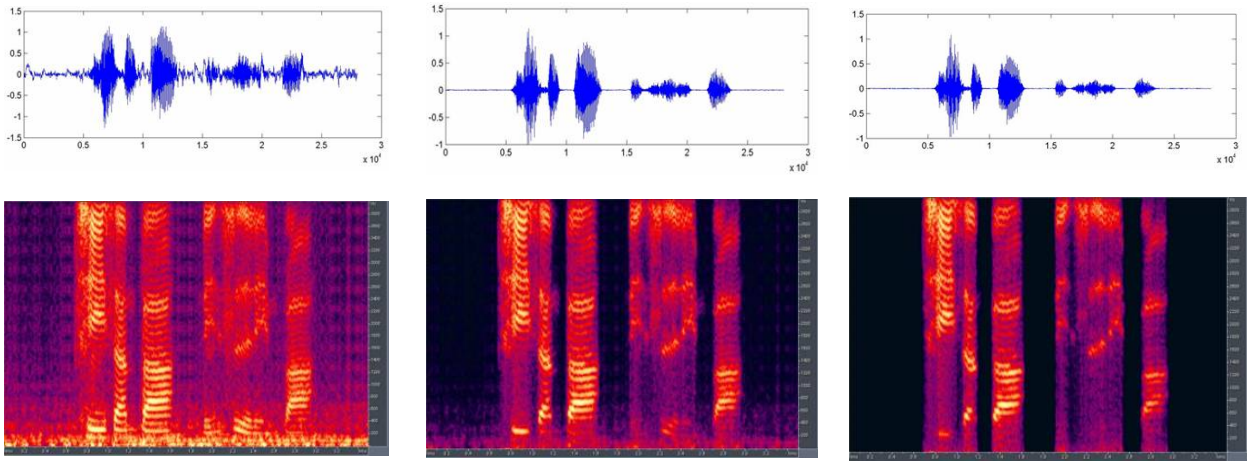
表三: MOS 測試評分。

	TAICAR 音檔		
	待速	市區路段	快速道路路段
DCT	2.6	2.1	1.9
KLT	4.4	4.1	3.8
本論文所提方法	4.2	4.0	3.9

圖六則為雜訊音檔經由上述三種方法進行噪音消除後之波形及頻譜圖。從圖六觀察噪音抑制結果，以 KLT 及本文所提方法皆優於 DCT 的效果，再從低頻帶的噪音消除來看，則是以本文所提的方法為最好。



(a)雜訊音檔波形及頻譜圖



(b) DCT　　(c) KLT　　(d)本文所提方法之結果

圖六：噪音消除結果比較之波形及頻譜圖。

4. 結論

本論文提出一個基於人耳聽覺特性的分頻處理，並結合訊號子空間強化方法來消除車內背景噪音。此一聽覺分頻處理係利用小波轉換來實現，藉由小波將聲音分解成多個頻帶，而各個子頻帶的分佈則符合人耳聽覺響應的特性，各子頻帶的訊號再經由子空間方法進行噪音消除，再由小波反轉換合成各子頻帶的訊號，進而得到強化後的語音。實驗的驗證，則是採用 TAICAR 車內語音資料庫來進行，由 MOS 評分及時間波形和頻譜圖來看，本文所提出的方法比起傳統訊號子空間採用 DCT 及 KLT 等方法，更適用於車內噪音的消除，低頻噪音的消除也更明顯。

5. 參考文獻

[1] Y. Ephraim and H. L. Van-Trees, “A signal subspace approach for speech enhancement,” *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, July 1995.

[2] B. Yang, “Projection approximation subspace tracking,” *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

[3] Shi-Huang Chen and Jhing-Fa Wang, “Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator,” accepted to appear in *The Journal of VLSI Signal Processing Systems*, Special Issue on Real World Speech Processing.

[4] O. Ghitza, “Auditory model and human performance in tasks related to speech coding and speech recognition,” *IEEE Trans. Speech and Audio Processing*, vol. 2, pp. 115-132, 1994.

[5] Lawrence Rabiner and Biing-Hwang Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993

[6] Jhing-Fa Wang, Hsien-Chang Wang and Chung-Hsien Yang, “TAICAR - A Collection of In-Car Mandarin Speech Database in Taiwan,” *O-COCOSDA2003 / PACLIC17*, Singapore.