

在零值音框嵌入法中使用諧波特性調適峰值頻譜鎖定長度於語音增強之研究

吳威例

陸清達

沈俊宏

亞洲大學 資訊傳播學系 研究生

亞洲大學 資訊傳播學系 教授

亞洲大學 資訊傳播學系 副教授

e-mail: 5az147852@gmail.com

e-mail: lucas1@ms26.hinet.net

e-mail: shenj@asia.edu.tw

摘要

語音在接收過程中會與背景雜訊混合，因此需要將語音中的雜訊移除，讓收話端聽得清楚發話端的講話內容。雜訊強度估計準確度，會影響增強的語音品質，所以本文嘗試透過零值音框嵌入法，加入諧波特性調適峰值頻譜鎖定法估測雜訊，提升音框中的雜訊估測位準。由於零值音框嵌入處沒有語音信號，因此接收端可以利用該處估測雜訊頻譜強度。為了提升雜訊頻譜估測位準，可以偵測母音音框的強健頻譜，做為更新雜訊估測音段長度的參考依據，在沒有母音的區段中，藉由增加音段長度與峰值頻譜鎖定的方式，適當高估雜訊頻譜強度，讓語音增強處理時，可以有效移除背景雜訊；相對的，在有母音的音段中，即時更新零值音框嵌入處的雜訊估測值，避免峰值頻譜鎖定的機制高估雜訊，造成嚴重的語音失真。實驗結果證明：本文所提出的方法，在無語音的區段可以有效的移除背景雜訊，降低殘留雜訊，讓增強語音聽起來更清晰與舒適。

關鍵詞：雜訊估測、語音增強、零值音框嵌入、諧波特性調適、峰值頻譜鎖定。

Abstract

Speech and noise signals are mixed in the same channel in speech communication. A speech enhancement system is employed to remove corruption noise, enabling a listener to understand the meaning of received speech. The accuracy of noise estimation significantly affects the performance of enhanced speech. This paper proposes using a frame-zero padding (FZP) and peak-spectrum-holding methods adapted by harmonic properties to improve the accuracy of noise estimation. Because speech signals are absent during the zero-padded frames, we can estimate the magnitude of noise spectrum during these periods. In order to improve the performance of frame-zero padding method,

robust harmonics in a vowel frame are estimated and employed to adapt the segment length for noise estimation. In the case of a non-vowel frame, the segment length is increased to adequately over-estimate the noise magnitude by the peak-spectrum holding method. So the residual noise can be significantly reduced in enhanced speech. On the contrary, the noise estimate is updated instantaneously during a vowel period, so the noise estimate can be prevented from over-estimation obtained by the peak-spectrum holding method. Accordingly, enhanced speech does not suffer from serious speech distortion. Experimental results show that the proposed method can efficiently remove background and residual noise during speech-pause regions, enabling enhanced speech to sound distinct and comfortable.

Keywords: noise estimation, speech enhancement, frame-zero padding, harmonic adaptation, peak spectrum holding.

1. 前言

語音增強處理主要是將混合著雜訊的語音信號中，盡可能的提取乾淨的原始語音，讓增強語音聽起來更清晰，語音增強處理是被多媒體相關領域廣泛應用的重要技術，日常生活中所使用的通訊設備都會應用到語音增強處理技術，例如：手機、助聽器、麥克風，以及語音通信網路應用等，都是需要使用語音增強處理才能發揮更好的效能。由於雜訊干擾通常都是隨機的，想從混合著雜訊的語音中，提取近乎於乾淨的語音，是一項極具挑戰的研究主題，而本研究的主要目的就是在受干擾的語音信號中有效的移除背景雜訊，降低殘留雜訊，讓增強語音聽起來更清晰。

為了提升增強語音的品質，有許多相關研究提出許多先進的語音增強演算法[1]-[15]，在文獻[7]中利用感知的雜訊遮蔽閾值(noise masking threshold)來調適語音增強系統的參數，可以有效防止音樂型殘留雜訊的產生。有些方

法則是使用後處理的技術，將音樂型殘留雜訊降低，文獻[10]提出對原始語音作預先處理，在語音取樣點與取樣點之間嵌入零值取樣點(zero-padding)，藉由在零值嵌入處的信號來估測語音中的雜訊振幅，此方法在取樣點變化緩慢的雜訊中，可以準確的估計雜訊振幅，使得語音增強的效果得以提升；雖然該方法可以有有效的估測變化緩慢的雜訊取樣值，但是卻無法準確估測變化快速的雜訊，例如：白雜訊(white noise)、F16 戰機座艙雜訊等等，使得語音增強效果變差。文獻[1]根據每個音框的功率頻譜，動態調整頻譜刪減的參數，然後利用多種方法抑制音樂型殘留雜訊，使頻譜刪減效果保持較高的訊雜比，也可以有效避免音樂型雜訊的產生；而張秀珍等人[2]提出結合頻譜刪減及人耳遮蔽特性的卡爾曼(Kalman)語音增強方法，找到感知差異上的語音失真和雜訊去除的最佳平衡點，使得語音失真不嚴重的情況下，可以降低音樂型殘留雜訊的產生。

為了有效估測變化快速的雜訊信號，文獻[3]提出一個新的雜訊估計方法，藉此改善零值嵌入法[6]中，無法對抗變化快速雜訊的缺點，將原本的零值嵌入法，修改成零值音框嵌入法，藉由在零值音框嵌入處估測雜訊頻譜，達到雜訊頻譜估計的目的；然後利用估計出來的雜訊頻譜值，代入通用型頻譜刪減法，從事雜訊消除處理；並藉由調整雜訊刪減量的控制因子，改善零值嵌入法中無法對抗變化快速雜訊的問題，達到提升語音增強的目標，但是將零值嵌入音框處估計到的雜訊，直接套用到語音區段使用，並無法準確估計受干擾語音中的雜訊，導致雜訊移除效果不佳。在文獻[4]中加入多重音框內插法來估測雜訊頻譜強度，提升雜訊估計準確度；接著使用知覺增益因子[8][9]從事語音增強；由於這類方法[4][5]的雜訊估測位準偏低，導致雜訊移除效果不佳。

Li 等人[11]提出使用非負矩陣分解法(nonnegative matrix factorization) 與 Kullback-Leibler 距離為測度，將矩陣分解為稀疏與低秩(low-rank)空間的矩陣，將受干擾語音信號的聲譜信號分類為語音(sparse, 稀疏矩陣)空間及雜訊(low-rank, 低秩矩陣)，該方法不需要任何的語音及雜訊的預先(priori)資訊便可以将語音聲譜解析，達到語音增強的效果。Yu 等人[12]提出使用矩形濾波矩陣在時域中實現語音增強，透過增強語音與濾波信號的 Pearson 相關係數平方值最大化，可以得到縮小秩(reduced rank)的溫納濾波器矩陣，並直接將該

矩陣與受干擾語音信號相乘，便可以得到增強語音。Lu[13]提出了使用梳型濾波器重建諧波頻譜，並利用中值濾波器消除音樂型殘留雜訊。在眾多語音增強的演算法中，頻譜刪減演算法具有簡單與高彈性的優點，其做法是將雜訊頻譜從受干擾的語音頻譜中刪除；而具彈性的特點是可以透過各種參數的介入，因應各種不同的情形去調整過度刪減因子，讓經過頻譜刪減演算法處理後的增強語音能夠有較少的語音失真，以及更佳的雜訊移除效果。Sukanya 與 Ananthakrishna[14]使用訊雜比值調適的過度刪減因子，移除背景雜訊，並且在靜音區段偵測殘留雜訊，調整其頻譜強度，防止音樂型殘留雜訊產生。

經由以上分析可得知：先進的雜訊估測演算法在估測雜訊時，都沒有考慮到語音變化的特性，導致雜訊估測的效能不佳；為了更精準的追蹤雜訊頻譜的強度，本文所提出的諧波特性和調適峰值頻譜鎖定法將考慮語音的特性；首先在零值音框嵌入處估測雜訊頻譜，並偵測母音音框的強健頻譜，做為雜訊頻譜強度更新音段的參考依據，在有母音的音段中，即時更新零值音框嵌入處的雜訊估測值，避免峰值頻譜鎖定的機制高估雜訊，造成嚴重的語音失真；相對的，在沒有母音的區段中，藉由增加音段長度與峰值頻譜鎖定的方式，適當高估雜訊頻譜，確保語音增強處理時，可以有效移除背景雜訊。

本文的其餘各節安排如下：在第二節介紹零值音框嵌入法，第三節說明本文提出的零值音框嵌入與諧波特性和調適峰值頻譜鎖定法，第四節呈現實驗結果與分析，第五節為本文的結論。

2. 零值音框嵌入法

零值音框嵌入法是將零值音框嵌入到語音的取樣點之間，並且在嵌入的零值音框取樣點中，估計雜訊頻譜值。此方法是一種單通道語音增強技術與零值嵌入法不同，零值嵌入法是將零值嵌入到語音取樣點與取樣點之間，並且在嵌入的零值取樣點中估計雜訊；這種方法對於變化較緩慢的雜訊，例如：汽車(car)雜訊、直昇機座艙(helicopter-cockpit)雜訊之類的雜訊估測準確度較高，但是對於快速變化的雜訊，如：白(white)雜訊，戰鬥機座艙(F16-cockpit)之類的雜訊估測準確度較低，零值音框嵌入法主要目的為改善零值嵌入法無法準確估測變化快速的雜訊。

藉由嵌入的零值音框取樣點的位置，即可取得雜訊通道的估計值，將估計出來的雜訊振幅，直接跟受干擾語音信號做刪減處理。零值音框嵌入法的示意圖如圖 1，將零值音框嵌入到乾淨的語音中。例如：取樣點 0 到 256、512 到 768、1024 到 1279 等均為嵌入的零值音框；取樣點 236 到 511、768 到 103、1280 到 1535 為乾淨的語音信號，餘此類推。

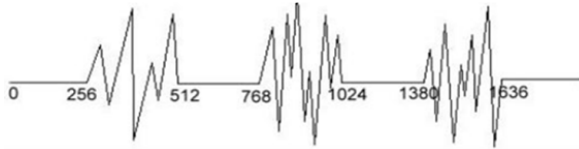


圖 1 零值音框嵌入示意圖。

嵌入的零值音框為估計雜訊頻譜的位置，零值音框嵌入法的表示式如下：

$$\tilde{s}[n+(2m-1)N] = \begin{cases} s[n+(m-1)N], & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

其中： $\tilde{s}[n]$ 表示嵌入零值音框之後的語音信號； $s[n]$ 為原始語音信號， N 為音框長度 ($N=256$)， m 表示音框序列編號。

經過零值音框嵌入的信號就可以發送到接收端，信號在傳送的過程當中，將會受到通道雜訊的干擾，所以接收端收到的信號的表示式如下：

$$y[n] = \tilde{s}[n] + d[n] \quad (2)$$

其中： $y[n]$ 為接收端收到的受干擾語音信號， $d[n]$ 為通道中的雜訊干擾信號。

在接收端中，為了抑制通道中的雜訊，必須在受干擾語音信號中估計雜訊；由於在發送端已經嵌入零值音框，所以可以在零值音框中，估計雜訊頻譜；然後將受干擾語音信號的頻譜刪減雜訊頻譜估計值，便可以得到增強語音信號頻譜。

為了估測雜訊頻譜強度，首先將嵌入零值音框的語音序列進行傅立葉轉換 (Fourier transform, FT) 得到受干擾語音信號的頻譜，表示式如下：

$$|\hat{S}_m(\omega)|^2 = |Y_m(\omega)|^2 - |\hat{D}_m(\omega)|^2 \quad (3)$$

其中： $\hat{S}_m(\omega)$ 為乾淨語音頻譜估測值， $Y_m(\omega)$ 和 $\hat{D}_m(\omega)$ 分別為受干擾語音頻譜與雜訊頻譜估測值。

由於嵌入零值音框中沒有語音，所以全部都是雜訊，故我們可以在受干擾語音信號的零值音框嵌入處，估測通道雜訊頻譜，並且沿用到後續音段中的所有音框，直到下一個零值音

框嵌入出現為止。這種方法雖然可以不斷的更新雜訊頻譜估測值，但是增強後的語音信號中，殘留雜訊量非常多，主因是雜訊頻譜強度估測值偏低的關係，因此本研究嘗試使用諧波特性和調適峰值頻譜鎖定法，提升雜訊估測位準，使殘留雜訊可以大幅降低，讓增強語音聽起來更清晰。

3. 諧波特性和調適之峰值頻譜鎖定法

諧波特性和調適峰值頻譜鎖定法，就是在有母音的音段中，即時更新零值音框嵌入處的雜訊估測值，避免峰值頻譜鎖定的機制高估雜訊，造成嚴重的語音失真；在沒有母音的區段中，藉由增加音段長度與峰值頻譜鎖定的方式，適當高估雜訊頻譜，確保語音增強處理時，可以有效移除背景雜訊。

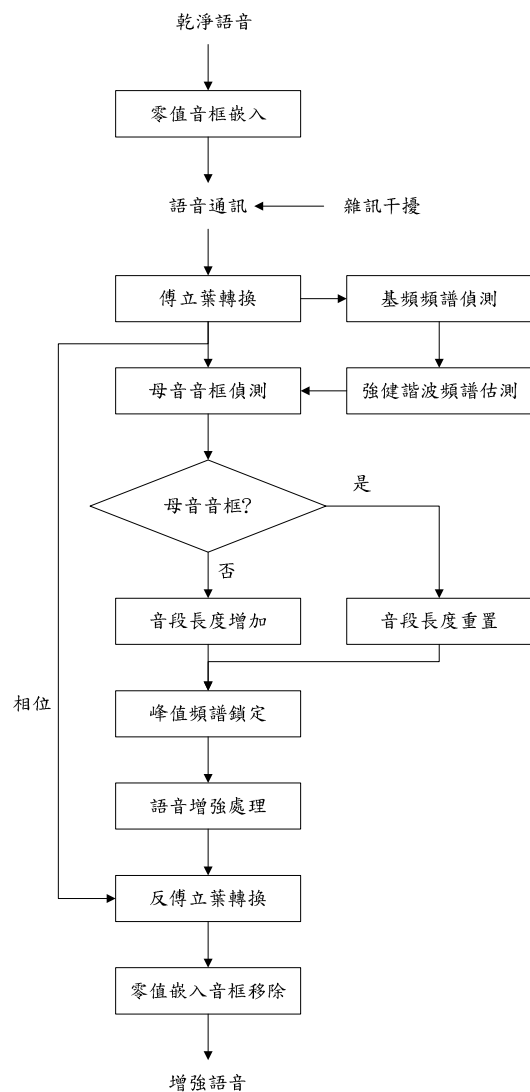


圖 2 諧波頻譜調適之峰值頻譜鎖定法流程圖。

本文的諧波特性和適峰值頻譜鎖定法之流程圖如圖2，首先在乾淨語音中嵌入零值音框，然後發話端將語音信號藉由通道傳送到受話端。在傳送過程中，將受到雜訊干擾，受話端先透過傅立葉轉換，取得受干擾語音信號之頻譜，先藉由接收訊號估測基頻頻率與強健諧波頻譜，做為判斷音框中是否有母音的參考依據，再根據每個音段母音是否存在，調整雜訊的更新音段長度；若母音存在，就使用零值音框嵌入處的頻譜當作雜訊估測值，確保雜訊的估測準確性；相對的，若母音不存在，則增加音段長度，進行峰值頻譜鎖定，亦即選取雜訊頻譜強度的累積最大值，當作雜訊估測值，使得估測雜訊強度不會偏低。然後將受干擾語音信號的頻譜刪減雜訊頻譜估測值(如式(3))，便可以得到增強語音信號頻譜。最後將訊號經由反傅立葉轉換，將訊號轉回時域，再將傳送端嵌入之零值音框移除，便可以得到增強語音信號。

3.1 諧波頻譜偵測法

本節將介紹母音音框的偵測方法，首先將受干擾語音透過低通濾波器來取得 50~500Hz 基頻(fundamental frequency)範圍的低通信號，接著使用式(4)計算該低通信號的自相關函數：

$$R_\phi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \phi(n) \cdot \phi(n+|\tau|) \quad (4)$$

其中： τ 是時間的偏移量，用以計算音高週期的位置。

由式(4)可以得知：當低通受干擾語音信號 $\phi(n)$ 作偏移相乘時，自相關函數的值會由高峰到低谷，再回到高峰處，而第二個高峰處即為擬尋找的音高週期位置。

為了準確取得音高週期，也可以透過平均強度變化函數(average magnitude difference function, AMDF)求得，如式(16)所示：

$$AMDF(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} |\phi(n) - \phi(n+|\tau|)| \quad (5)$$

由式(5)可以得知：當低通受干擾語音信號 $\phi(n)$ 做偏移相減時，平均強度變化函數值 AMDF 會由谷底往高峰處變化，再回到谷底處，而第二個谷底處即為擬尋找的音高週期。

為了提升音高週期的鑑別度，可以將 $R_\phi(\tau)$ 與 $AMDF(\tau)$ 相除得到權重自相關函數(weighted autocorrelation function, WAF)，提升音高週期的準確性，如式(6)：

$$WAF(\tau) = \frac{R_\phi(\tau)}{AMDF(\tau) + \varepsilon} \quad (6)$$

其中： ε 是一個非常小的數值，用來防止分母為零。

在音高週期處，自相關函數出現極大值；相對的，平均強度變化函數則出現極小值，將兩個函數分別置於分子和分母(如式(6))，利用分子和分母成反比的特性，在音高週期位置會出現極大值，而非音高週期所在位置的權重自相關函數值會很小，故可以提高音高週期的鑑別度。

取得音高週期之後，代入式(7)，即可得到基頻 F_0 的頻率估測值：

$$F_0 = \frac{N}{T_0} \quad (7)$$

其中： T_0 代表音高週期。

使用式(7)求得的基頻頻率與實際的基頻頻率有些微的差異，因此我們透過式(8)微調基頻頻率位置到 F_0^* [9]：

$$F_0^* = F_0 - F_0^{Bias} \quad (8)$$

其中： F_0^{Bias} 為 F_0 與實際頻譜峰值的頻率偏差量，可由式(9)計算取得[9]：

$$F_0^{Bias}(l) = \frac{1}{t_e - t_i} \cdot \sum_{m=t_i}^{t_e-1} F_0(m, l) - F_0'(m, l) \quad (9)$$

其中： l 表示母音音段索引； t_i 表示一個母音音段的初始音框索引； t_e 表示該母音音段的結束音框索引。

透過式(4)-(9)，以音段為單位修正基頻頻率，可以準確的估測基頻頻率的變化軌跡，並且適當修正偏差量。雖然母音音段的高頻諧波頻譜會被背景雜訊破壞，但是低頻的諧波頻譜卻能在雜訊干擾下，依然不受影響，主要原因就是該低頻諧波頻譜的能量很大，具有極佳的強健性，也能遮蔽背景雜訊，讓較低能量的雜訊雖然存在，但人耳依然聽不到。接著可以從低頻到高頻之間，尋找強健的諧波頻譜，當頻率逐漸提高時，如果相鄰的諧波頻率變化量太大，表示諧波頻譜開始不穩定，因此就可以標示強健諧波頻譜的數量 r ，亦即頻率在 $\omega_c(k\omega_0^*)$ 與 $\omega_c((k+1)\omega_0^*)$ 兩者之間變化劇烈時，就可以標示強健諧波頻譜的邊界，而 r 的值为 k 的最大值， $\omega_c(r\omega_0^*)$ 即為最後一個強健諧波頻譜。具有足夠強健諧波的音框就是屬於母音音框，其表示式如式(10)：

$$F^v(m) = \begin{cases} 1, & \text{if } r(m) \geq \delta_v \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

其中： $F^v(m)$ 為母音音框旗號； $r(m)$ 為強健諧波頻譜數量； δ_v 為母音音框的強健諧波頻譜數量門檻值。

在式(10)中，當一個音框具有足夠的強健諧波頻譜數量 δ_v 時，則將該音框歸類為母音音框，並且將母音音框旗號 $F^v(m)$ 設為“1”；相對的，若某一音框不是母音音框，亦即強健諧波頻譜數量太少(低於 δ_v)，則將母音音框旗號 $F^v(m)$ 設為“0”。

3.2 諧波特性和調適之峰值頻譜鎖定

雜訊峰值頻譜鎖定(Peak-spectrum-holding)法[5]，就是在沒有語音訊號的零值嵌入處估測雜訊頻譜強度，並且施予極大值鎖定，使得估測雜訊為實際背景雜訊的極大值，提升語音增強處理系統移除背景雜訊的效能。

雜訊峰值頻譜鎖定法是在零值嵌入處，偵測雜訊頻譜的峰值 $N_{\max}(m, \omega)$ ，並沿用到後續的音框中，如式(11)：

$$|\hat{N}(4\ell + m, \omega)| = |N_{\max}(4\ell, \omega)| \quad (11)$$

其中： ℓ ， m 分別表示音段索引及在音段中的音框索引(0-3)； $N_{\max}(\ell, \omega)$ 表示零值音框嵌入處取得的受干擾語音訊號的峰值頻譜。

在雜訊估測過程中，我們將會鎖定雜訊峰值頻譜，如式(12)。雜訊峰值頻譜鎖定法在估測背景雜訊時，合理高估雜訊頻譜強度，讓語音增強處理時(如式(3))，可以有效的移除雜訊，但卻造成較嚴重的語音失真。為了更精準的追蹤雜訊頻譜的強度，本文所提出的諧波特性和調適峰值頻譜鎖定法將考慮語音的特性。首先在零值音框嵌入處估測雜訊頻譜，並尋找母音音

框的強健諧波頻譜，做為更新雜訊峰值頻譜鎖定音段的參考依據，在有母音的音段中，即時更新零值音框嵌入處的雜訊估測值，避免峰值頻譜鎖定的機制高估雜訊，造成嚴重的語音失真；相對的，在沒有母音的區段中，藉由增加音段長度與峰值頻譜鎖定的方式，適當高估雜訊頻譜，確保語音增強處理時，可以有效移除背景雜訊，使增強後的語音聽起來更清晰。經由諧波頻譜特性調適之峰值頻譜鎖定之雜訊頻譜估測值如式(13)。由式(13)可以得知：當零值音框嵌入處的語音音框不是母音時($F^v(4\ell + 2) = 0$)，進行峰值頻譜鎖定；相對的，若零值音框嵌入處的語音音框是母音時($F^v(4\ell + 2) = 1$)，則使用零值音框的頻譜強度 $|N[4\ell, \omega]|$ ，當做雜訊頻譜的估測值。

圖 3 為雜訊頻譜估測軌跡圖，由圖 3(a)可以看出：在白雜訊干擾的環境中，諧波頻譜特性調適之峰值頻譜鎖定法，在有母音存在的區段中，讓估測雜訊強度維持在低位準，避免語音增強處理時，因刪減過多受干擾語音信號，保留更多母音成分，確保增強語音的品質；而在沒有母音的區段中(尤其是靜音區段)，藉由增加音段長度，以較長的時間維持峰值頻譜鎖定方式，適當高估雜訊頻譜，讓語音增強處理時，可以有效移除背景雜訊，提高增強語音聽覺的舒適度。由圖 3(b)可以看出：強健諧波頻譜在靜音區段會因為雜訊干擾產生估測錯誤的情形，因此可以將式(10)的強健諧波頻譜數量門檻值 δ_v 設為 10，確保偵測母音音框的正確性。

4. 實驗結果

在實驗中，我們使用中文女生語音訊號，加入 NOISEX_92 雜訊語料庫中的白色(white)雜訊、F16 戰機座艙(F16-cockpit) 雜訊、工廠

$$|N_{\max}(4\ell, \omega)| = \begin{cases} |N(4\ell, \omega)|, & \text{if } |N(4\ell, \omega)| \geq |N_{\max}[4(\ell-1), \omega]| \\ |N_{\max}[4(\ell-1), \omega]|, & \text{otherwise} \end{cases} \quad (12)$$

$$|N_{\max}^{\text{harmonic}}(4\ell, \omega)| = \begin{cases} |N(4\ell, \omega)|, & \text{if } |N(4\ell, \omega)| \geq |N_{\max}^{\text{harmonic}}[4(\ell-1), \omega]| \& F^v(4\ell + 2) = 0 \\ |N_{\max}^{\text{harmonic}}[4(\ell-1), \omega]|, & \text{if } |N(4\ell, \omega)| < |N_{\max}^{\text{harmonic}}[4(\ell-1), \omega]| \& F^v(4\ell + 2) = 0 \\ |N[4\ell, \omega]|, & \text{otherwise} \end{cases} \quad (13)$$

其中： F^v 為母音音框旗號，其值如式(10)。

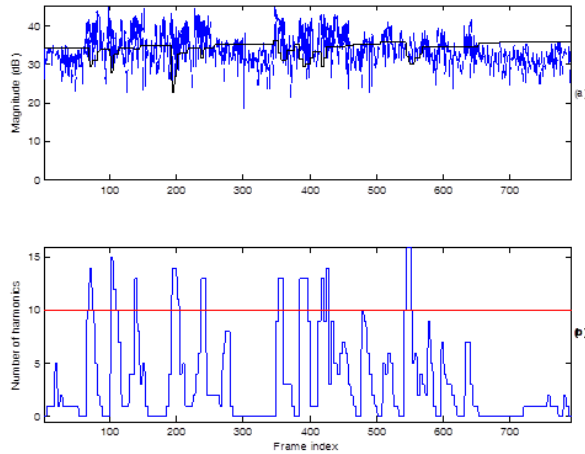


圖 3 雜訊頻譜估測軌跡圖；(a)女音信號在白雜訊干擾的環境中，某一次頻帶(次頻帶為 12)的受干擾語音頻譜軌跡圖(藍色)與雜訊頻譜估測軌跡圖(黑色)；(b)強健諧波頻譜估測數量軌跡圖。

表 1 在各種不同雜訊的環境使用各種雜訊估測法的增強語音訊雜比值改善度之比較表

Noise type	SNR (dB)	Enhancement method		
		ZP [6]	FZP [8]	Proposed
white	0	-9.84	-1.51	6.83
	5	-5.06	1.61	5.94
	10	0.39	4.06	4.89
F16	0	-4.64	-1.52	5.83
	5	0.31	1.63	4.92
	10	5.31	3.89	4.09
factory	0	-3.21	-2.23	5.57
	5	1.80	1.09	4.70
	10	7.09	3.71	3.89
helicopter	0	0.06	-0.72	7.18
	5	5.11	2.14	6.34
	10	9.97	4.07	5.29
babble	0	-0.36	-1.60	4.60
	5	4.61	1.42	4.76
	10	9.57	3.61	2.91
car	0	18.98	-1.04	7.93
	5	23.95	1.79	7.68
	10	29.20	3.98	7.22

(factory)雜訊、汽車(car)雜訊、雞尾酒會之多重人聲(babble)雜訊、直升機座艙(helicopter-cockpit)雜訊，分別在 0 dB、5 dB 及 10 dB 的雜訊干擾的環境下，測試各種演算法的效能，

我們也使用訊雜比值(SNR)改善度、語音波形圖及語音聲譜圖來評估增強語音的品質。

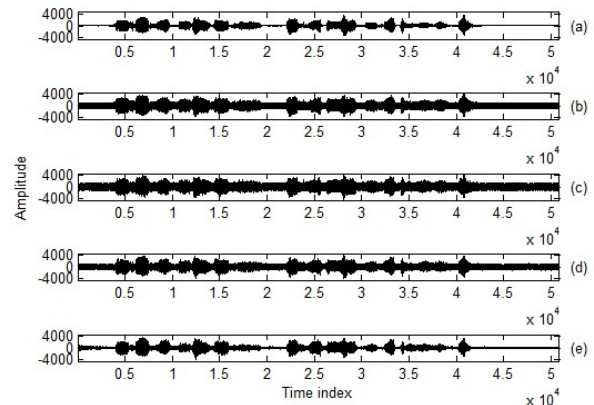


圖 4 語音波形圖；(a)乾淨語音；(b)受白色雜訊干擾的語音，訊雜比值為 0 dB；(c)使用零值嵌入法增強後的語音；(d)使用零值音框嵌入法增強後的語音；(e)使用本文方法增強後的語音。

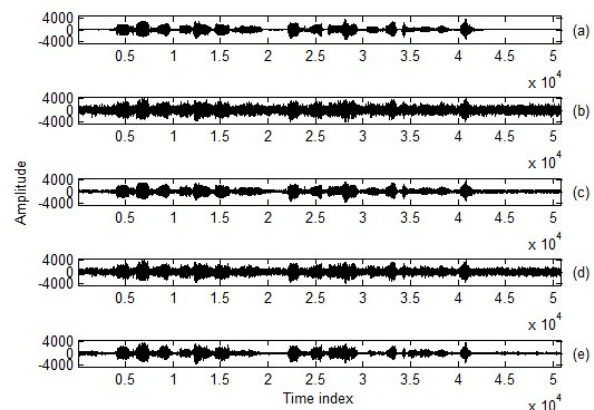


圖 5 語音波形圖；(a)乾淨語音；(b)受 F16 戰機座艙雜訊干擾的語音，訊雜比值為 0 dB；(c)使用零值嵌入法增強後的語音；(d)使用零值音框嵌入法增強後的語音；(e)使用本文方法增強後的語音。

使用各種雜訊估測演算法處理語音訊號在各種不同雜訊環境下的訊雜比改善度如表 1 所示，在表 1 中，ZP (zero padding)為零值嵌入法[6]，FZP (frame-zero padding)為零值音框嵌入法[8]；由表 1 可以看出：在汽車(car)雜訊干擾的部份，零值嵌入法的效果優於 FZP 與本文所提出的方法，主要因為零值嵌入法是藉由零值嵌入處的取樣點信號來估測語音中的雜訊的振幅，該方法在取樣點變化緩慢的雜訊中，前後兩個相鄰取樣點的雜訊振幅變化很小，所以使用前一個取樣點的振幅來估測目前取樣點的振幅，可以很準確的估計雜訊振幅大

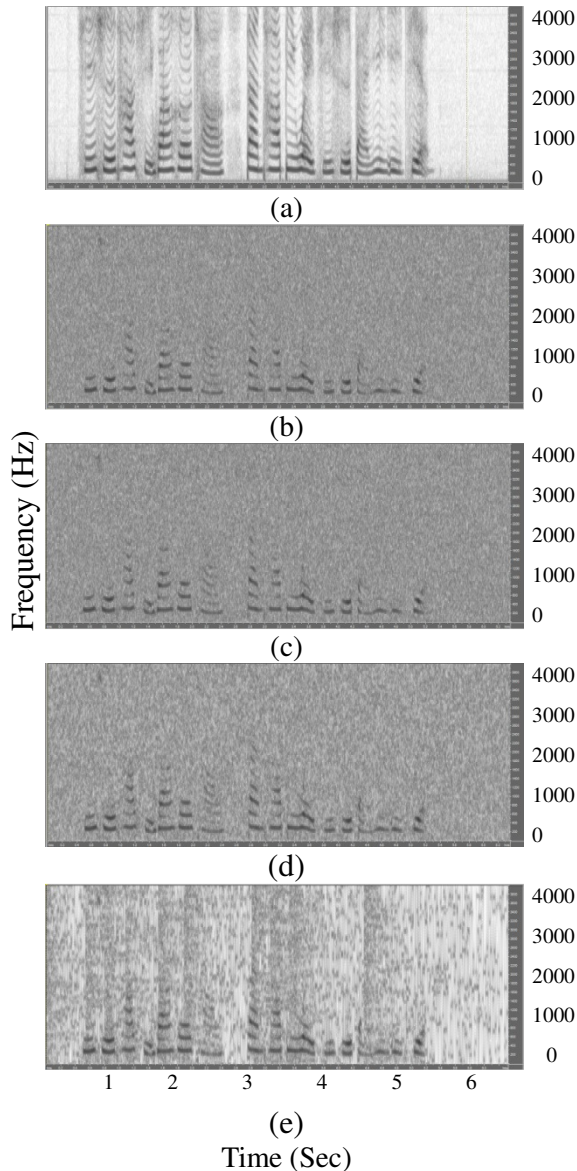


圖 6 語音聲譜圖；(a)乾淨語音；(b)受白色雜訊干擾的語音，訊雜比為 0 dB；(c)使用零值嵌入法增強後的語音；(d)使用零值音框嵌入法增強後的語音；(e)使用本文方法增強後的語音。

小，但是本文的方法可以大幅改善零值音框嵌入法的效能，主要原因就是合理高估雜訊頻譜強度，讓語音增強處理可以有效移除背景雜訊。而本文所提出的方法在白(white)雜訊的干擾環境下，效能優於其他兩種方法；而且在較嚴重雜訊干擾的環境中(訊雜比為 0 dB 與 5 dB)，對於 F16 戰機座艙雜訊、工廠雜訊、直升機座艙雜訊、雞尾酒會之多重人聲雜訊環境下，訊雜比改善度均優於零值嵌入法[6]與零值音框嵌入法[8]，因此可以證明本文提出的方法可以大幅改善零值音框嵌入法的效能，主要原因就是合理高估雜訊頻譜的強度，有效的移除

背景雜訊、降低殘留雜訊，讓增強語音聽起來更清晰。

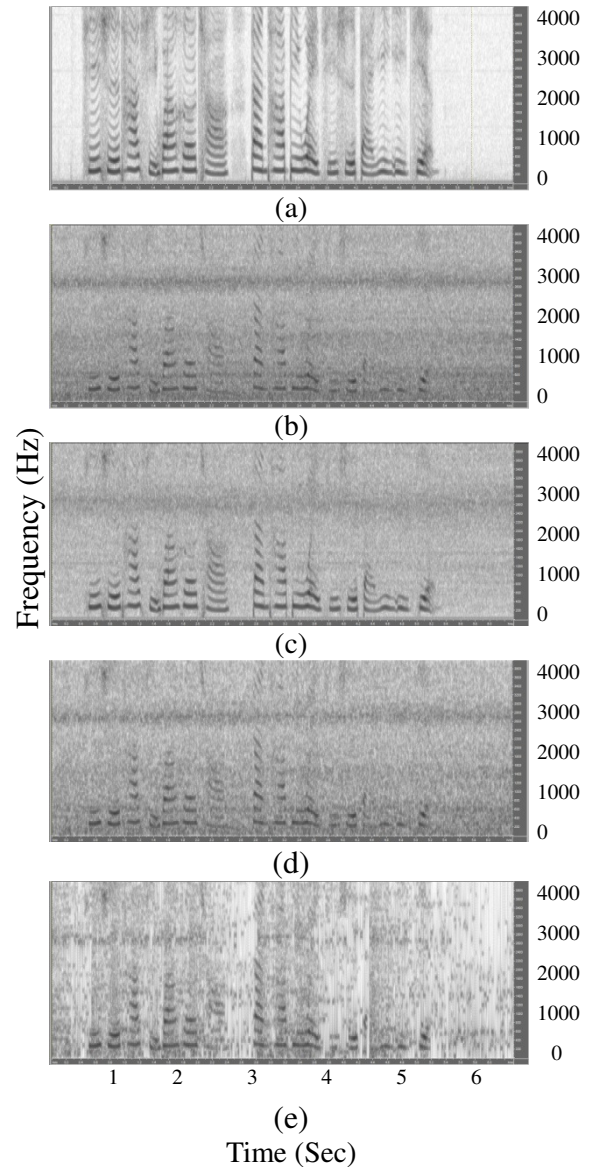


圖 7 語音聲譜圖；(a)乾淨語音；(b)受 F16 座艙雜訊干擾的語音，訊雜比為 0 dB；(c)使用零值嵌入法增強後的語音；(d)使用零值音框嵌入法增強後的語音；(e)使用本文方法增強後的語音。

圖 4 為受白色雜訊干擾的女生語音信號，訊雜比為 0 dB(如圖 4(b))，由圖 4(c)可以很明顯看出：零值嵌入法對於處理白雜訊干擾語音的效果不佳，甚至將雜訊的部份放大；而零值音框嵌入法(如圖 4(d))能將部份的雜訊抑制，效果優於零值嵌入法(如圖 4(c))；而本文所提出的方法(如圖 4(e))能有效抑制變化快速之白雜訊，尤其是在靜音的區段；相對的，在有語音的區段中，本文的方法使用諧波頻譜特性調

適，雖然具備有效移除背景雜訊的能力，但是卻不會造成嚴重的語音失真，確保增強語音的品質。

圖 5 為受 F16 戰機駕駛艙雜訊干擾的女生語音信號，訊雜比為 0 dB(如圖 5(b))，由圖中可以看出：零值嵌入法(如圖 5(c))對於移除雜訊的效果優於零值音框嵌入法(如圖 5(d))，尤其是靜音的區段；經由本文所提出的改善方法(如圖 5(e))則可以有效的抑制 F16 戰機駕駛艙雜訊，有效改善零值音框嵌入法的雜訊移除效能。

圖 6(b)為受到白雜訊干擾的女生語音信號，訊雜比為 0 dB，聲譜圖顏色越暗表示聲音越大聲，由圖中可以看出處理後的效果，零值嵌入法的聲譜圖(圖 6(c))中，背景雜訊所對應的顏色較暗，代表該方法不僅沒有抑制雜訊，反而讓雜訊量增加了，沒有移除雜訊的效果。零值音框嵌入法則能稍微抑制雜訊(如圖 6(d))；而由圖 6(e)中可以看出，本文所提出方法之聲譜圖的雜訊部分顏色較淡，抑制雜訊效果優於前面兩種方法，證實本文提出的方法確實能有效移除背景雜訊，讓增強語音聽起來較不吵雜；而且在母音的區段中，母音諧波頻譜保留完整，沒有過度刪除，因此沒有嚴重語音失真的情形。

圖 7 為受 F16 戰機駕駛艙雜訊干擾的女生語音信號，訊雜比為 0 dB，由圖 7(e)可以明顯看出本文方法處理後的效果，靜音區段的顏色較淡，抑制雜訊的效果明顯優於零值嵌入法(圖 7(c))與零值音框嵌入法(圖 7(d))，使得增強後的語音聽起來比較清晰，而且不吵雜。

5. 結論

本研究主要貢獻為改善零值音框嵌入法中估測雜訊的效能，藉由在零值音框嵌入處估測雜訊頻譜，然後加入諧波特性的調適峰值頻譜鎖定法來估測語音區段的雜訊頻譜強度，提升零值音框嵌入法的估測準確度。在沒有母音的區段中，藉由增加音段長度與峰值頻譜鎖定的方式，適當高估雜訊頻譜，讓語音增強處理時，可以有效移除背景雜訊；相對的，在有母音的音段中，即時更新零值音框嵌入處的雜訊估測值，避免峰值頻譜鎖定的機制高估雜訊，在語音增強處理後，造成嚴重的語音失真。經由實驗結果證實：本文所提出的方法，可以大幅改善零值音框嵌入法的雜訊估測效能，尤其是在較嚴重雜訊干擾的環境中(訊雜比值低於 5 dB)，效能更是優越。在增強語音的靜音區段中，可

以有效的降低殘留雜訊；而在有語音的區段中，因為雜訊沒有過度高估，因此增強語音不會出現嚴重的語音失真，讓增強語音聽起來更清晰與舒適。

致謝

本研究由科技部專題研究計畫之經費支助，計畫編號為 MOST 104-2221-E-468-007。

參考文獻

- [1] 王欣、羅代升、王正勇，”基於改進頻譜刪減演算法的語音增強研究”，*成都資訊工程學院學報*，卷 22，期 2，pp. 201-204，2007。
- [2] 張秀珍、伏旭暉、王霞，”一種改進卡爾曼濾波語音增強法”，*計算機應用*，卷 28，pp. 363-365，2008。
- [3] 王俊評、陸清達，”使用零值音框嵌入法之語音增強技術”，*智慧型系統工程研討會*，pp. 294-297，2010。
- [4] 巫明諺、陸清達、洪國樑，”使用零值音框嵌入與多重音框內插法於背景雜訊抑制之研究”，*資訊科技國際期刊*，卷 5，期 2，pp. 226-237，2011。
- [5] 黃苑絲、陸清達，”使用零值音框嵌入與雜訊峰值頻譜鎖定法於語音增強之研究”，*全國計算機會議*，2013。
- [6] Lim, J. S. and Oppenheim, A. V., “Enhancement and Bandwidth Compression of Noisy Speech,” in *Proceedings of IEEE*, Vol. 6, pp. 1586-1604, 1979.
- [7] Jebara, S.B., “A Perceptual Approach to Reduce Musical Noise Phenomenon with Wiener Denoising Technique,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 49-52, 2006.
- [8] Lu, C. -T. and Tseng, K. -F., “A Gain Factor Adapted by Masking Property and SNR Variation for Speech Enhancement in Colored-Noise Corruptions,” *Computer Speech and Language*, Vol. 24, No. 4, pp. 632-647, 2010.
- [9] Lu, C. -T., “Reduction of Musical Residual Noise Using Block-and-Directional-Median Filter Adapted by Harmonic Properties,” *Speech Communication*, Vol. 58, pp. 35-48, Mar 2014.
- [10] Singh, L. and Sridharan, S., “Speech Enhancement Using Pre-Processing,” in

Proceedings of IEEE Annual Conference on Speech and Image Technologies for Computing and Telecommunications, Vol. 2, pp. 755-758, 1997.

- [11] Li, M. S. Y., Gemmeke, J. H. and Zhang, X., “Speech Enhancement Under Low SNR Conditions Via Noise Estimation Using Sparse and Low-Rand NMF with Kullback-Leibler Divergence,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No. 7, pp. 1233-1242, 2015.
- [12] Yu, J., Benesty, J., Huang, G. and Chen, J., “Optimal Single-Channel Noise Reduction Filtering Matrices from the Pearson Correlation Coefficient Perspective,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 201-205, 2015.
- [13] Lu, C. -T., “Improvement of Enhanced Speech Using Harmonic Boost and Block-Directional-Median Filter,” *International Journal of Electrical Engineering*, Vol. 22, No. 3, pp. 97-113, June 2015.
- [14] Sukanya, S. M. and Ananthakrishna, T., “Dynamic Spectral Subtraction on AWGN Speech,” in *Proceedings of IEEE International Conference on Signal Processing and Integrated Networks*, pp.92-97, 2015.
- [15] Lu, C. -T., Lei, C. -L., Tseng, K. -F. and Chen, C. -T., “Estimation of Noise Magnitude Using Minima-Controlled-Recursive-Averaging Algorithm with Variable Length for Speech Enhancement,” in *Proceedings of National Symposium on Telecommunications*, 2015.