

使用母音諧波頻譜調適之最小控制遞回平均雜訊估測法於語音增強之研究

Estimation of Noise Magnitude Using Minima-Controlled-Recursive-Averaging Algorithm Adapted by Vowel Harmonic for Speech Enhancement

沈于中
亞洲大學資訊傳播研究所
Email:cashshen@gmail.com

陸清達
亞洲大學資訊傳播學系
Email:lucas1@ms26.hinet.net

摘要

準確的估測雜訊頻譜強度，可以有效改善語音增強系統的效能，而雜訊估測演算法經常遭遇兩種問題：分別是雜訊頻譜強度低估，導致語音增強語音的殘留雜訊太多，語音仍然吵雜而不夠清晰；另一個問題是雜訊頻譜強度高估，導致增強後的語音失真過大，致使語音訊息的可理解度下降。因此如何準確而有效的估測雜訊頻譜強度，對於改善語音增強處理系統的效能而言，是非常重要的。本文嘗試透過母音諧波頻譜調適(vowel harmonic spectrum adaptation, VHSA)最小控制遞回平均雜訊估測法(minima-controlled-recursive-averaging, MCRA)，根據不同的雜訊類型、頻譜中雜訊的多寡、及分佈情形不同，將語音存在的機率依據諧波強度作調適，提高語音存在機率的估測準確度，避免產生語音失真，確保增強語音的品質及清晰度；另一方面，也能有效的移除背景雜訊，達到提升語音增強效能的目的。實驗結果證明：本文提出的方法可以有效的提高雜訊強度估測的準確性，而且效能優於最小控制遞回平均雜訊估測法。

關鍵詞：雜訊估測、諧波調適、語音存在機率、語音增強、最小控制遞回平均估測法。

Abstract

Accurately estimating noise magnitude can improve the performance of a speech enhancement system. However, most of noise estimators suffer from either overestimation or

underestimation on the noise level. An overestimate on noise will cause serious speech distortion. On the contrary, a great quantity of residual noise will be introduced when noise magnitude is underestimated. Accordingly, how to accurately estimate noise magnitude is important for speech enhancement. In this study, we employ a minima-controlled-recursive-averaging (MCRA) algorithm adapted by vowel harmonics to estimate noise level. A speech-presence probability is adapted by the number of robust harmonics, enabling a vowel spectrum to obtain the value of speech-presence probability approaching unity. The vowel spectra can be well preserved. Consequently, the enhanced speech quality is improved while background noise is efficiently reduced. Experimental results show that the proposed method can accurately estimate noise magnitude and can improve the performance of the MCRA algorithm.

Keywords: noise estimation, harmonic adaptation, speech-presence probability, speech enhancement, minimum-recursive-controlled averaging.

1. 前言

語音增強處理系統需要一個高效能的雜訊估測器，才可以有效的抑制背景雜訊，進而降低語音失真[1]-[7]；傳統的雜訊估測方法是以音框分析為基本單位，然後在語音暫停時，估測背景雜訊的頻譜強度[1]，該方法對於穩定變化的雜訊，有不錯的估測效能，然而對於不穩定變化的雜訊，估測效果變差。

有些雜訊估測的方法在受干擾語音頻譜的次頻帶或音段中估測雜訊頻譜[8]-[14]，該方法受限於音的清音部分或微弱的母音的信號影響，雜訊估測的準確度效能不佳。此外，因雜訊在隨著雜訊類型、頻譜分佈、雜訊多寡不同，追蹤估測雜訊及判別語音能力的困難度較高，導致效能不佳[2][9][10]。

馬丁[11][12]提出了一個雜訊頻譜強度估測法，使用最小統計演算法做基礎，並使用平滑的功率來追蹤估計雜訊的最小值，以獲得雜訊頻譜音強估測值，其優點是在劇烈變化的頻譜中增強估計；而其缺點是其法必須施予補償偏移的運算，因它的離群值[10]，造成約兩倍大的方差[11]；為避免雜訊音強低估，必須透過偏移量調整方式，提高雜訊頻譜估測的準確度。此外，傳統的方法可能因為低能量的因素影響估測的正確性，或者是最小搜索窗口太短，影響音高週期的鑑別性[3]。另一個比較高效率的最小值的追蹤方法[13]，其優點是在一定時間內更新頻譜最小值；缺點是雜訊頻譜估計時，更新率太緩慢，雜訊的能量若急遽上升的情況下，無法有效的追蹤雜訊頻譜的變化[4][14]而影響估測準確度。Cohen 與 Berdugo[14]提出最小控制遞回平均(MCRA)雜訊估測法，該方法在均化與平滑化的頻譜中估測雜訊頻譜，其中平滑因子是透過語音存在機率做調適，其優點是利用語音存在機率，區分語音與非語音區域，做不同的平滑化處理，降低雜訊估測值在雜訊區域的變化程度，而且計算複雜度低，並且可以有效追蹤即時變化的雜訊；但是該方法的語音存在閾值為固定不變，並沒有考慮語音變化的特性，導致雜訊估測的準確性會隨著語音受到不同環境雜訊的干擾而改變，故其強韌性(robustness)可以再提升。

為了更精確的估測雜訊的頻譜，本文利用基頻附近的頻譜峰值，來估測母音諧波頻譜的位置，並且根據相鄰頻帶的諧波頻譜位置變化量，選定強健諧波(robust harmonic)，然後依照諧波頻譜的數量調適語音存在的機率，提高最小控制遞回平均(MCRA)雜訊估測法的準確性。若是一個音框的母音諧波頻譜較多，我們降低語音門檻值，該音框大部分的頻譜會歸類為語音頻譜，所對應的語音存在機率提升；如果一個音框的諧波頻譜數量很少，可能是微弱母音諧波頻譜，稍微提升語音門檻，使該音框大部分的頻譜會歸類為類似語音的頻譜，所對應的語音存在機率稍高；如果一個音框完全沒有母音諧波頻譜，而且音強相對偏弱，調高語音門

檻值，使該音框大部分的頻譜歸類為雜訊，讓所對應的語音存在機率降低。依據語音存在機率，調適 MCRA 演算法的雜訊的估測參數，達到提升雜訊估測效能的目的，實驗結果證明：本文提出來的的方法，可以準確的估測雜訊，並且能大幅改善 MCRA 雜訊估測法的效能，使得語音增強處理系統可以有效的移除背景雜訊，而語音失真也維持在很低的位準。

本文的其他章節安排如下：第二節介紹最小控制遞回平均(MCRA)雜訊估測法，第三節介紹本文提出的母音諧波頻譜調適之最小控制遞回平均雜訊估測法，第四節呈現實驗結果，第五節為本文的結論。

2. 最小控制遞回平均(MCRA)雜訊估測法

一個受到雜訊干擾的語音信號 $y(n)$ ，可以表示為乾淨語音 $s(n)$ 和加性雜訊 $d(n)$ 的和，如式(1)：

$$y(n) = s(n) + d(n) \quad (1)$$

其中： n 是離散時間索引。

將 $y(n)$ 施以短時傅立葉轉換(short-time Fourier transform, STFT)，則受干擾語音信號的頻譜如式(2)：

$$Y(k, l) = \sum_{n=0}^{N-1} y(n + lM)h(n)e^{-j(2\pi/N)nk} \quad (2)$$

其中： k 表示次頻帶索引， l 表示音框索引， M 是音框索引， h 表示分析視窗函數， N 為音框的長度。

為了分別在語音存在和不存在區域做不同程度的平滑化處理，以達到準確估測雜訊的效果，故設立兩個假定： $H_0(k, l)$ 和 $H_1(k, l)$ ，它們分別表示第 k 個頻帶的第 l 個音框中，語音不存在和存在的假定，其定義如式(3)：

$$\begin{aligned} H_0(k, l): Y(k, l) &= D(k, l) \\ H_1(k, l): Y(k, l) &= S(k, l) + D(k, l) \end{aligned} \quad (3)$$

其中 $S(k, l)$ 和 $D(k, l)$ 分別表示 STFT 中，乾淨的語音和雜訊訊號，令 $\lambda_d(k, l) = |D(k, l)|^2$ 表示雜訊的功率頻譜，為了避免估測雜訊變化過度急遽，在雜訊估測前，加入遞回平均雜訊估測法

[14]，其作法是在語音不存在區域做平滑處理，語音存在區域不做處理，如式(4)：

$$\begin{aligned} H'_0(k, l): \hat{\lambda}_d(k, l+1) &= \\ \alpha_d \hat{\lambda}_d(k, l) + (1 - \alpha_d) |Y(k, l)|^2 & \\ H'_1(k, l): \hat{\lambda}_d(k, l+1) &= \hat{\lambda}_d(k, l) \end{aligned} \quad (4)$$

其中 $\alpha_d (0 < \alpha_d < 1)$ 是一個平滑參數，其值設定為0.95， H'_0 表示語音不存在假定， H'_1 表示語音存在假定。

使用平滑因子在前後音框做不同的處理，可以防止估測不準確的雜訊強度，在代入語音增強處理時，產生音樂型殘留雜訊。

為了提升雜訊估測頻譜平滑化的正確性，Cohen 等人[14]使用語音存在機率調整平滑化因子，如式(5)：

$$\begin{aligned} \hat{\lambda}_d(k, l+1) &= \\ \hat{\lambda}_d(k, l) p'(k, l) + [\alpha_d \hat{\lambda}_d(k, l) & \\ + (1 - \alpha_d) |Y(k, l)|^2 \times (1 - p'(k, l))] & \\ \tilde{\alpha}_d(k, l) \hat{\lambda}_d(k, l) + [1 - \tilde{\alpha}_d(k, l)] |Y(k, l)|^2 & \end{aligned} \quad (5)$$

其中： $p'(k, l) \triangleq p(H'_1(k, l) | Y(k, l))$ 表示頻譜在第 k 個次頻帶中第 l 個音框，語音存在的歸類($p'(k, l)=1$ 或 0)，決定於 $Y(k, l)$ 中語音存在機率。在機率閾值以下語音存在的低機率區域，歸類為語音不存在區域，因此次頻帶音框中的平滑度可以由語音存在機率來決定，如式(6)：[14]

$$\tilde{\alpha}_d(k, l) \triangleq \alpha_d + (1 - \alpha_d) p'(k, l) \quad (6)$$

為了提高語音存在機率 $p'(k, l)$ 估測準確性，將 $p'(k, l)$ 在頻率域做平滑化處理，如式(7)：

$$\begin{aligned} \hat{p}'(l, \omega) &= \\ \alpha_p \cdot \hat{p}'(l-1, \omega) + (1 - \alpha_p) \cdot I(l, \omega) & \end{aligned} \quad (7)$$

其中： $\hat{p}'(l, \omega)$ 表示經計算後增強的語音存在機率頻譜， l 表示音框索引，而 $\alpha_p (\alpha_p=0.2)$ 是一個平滑參數， $I(l, \omega)$ 表示語音存在標示函數，以語音存在閾值 $\delta_\gamma(l)$ 做界定，標示是否為語音存在區域，如式(8)：

$$I(\omega, l) = \begin{cases} 1, & \text{if } \gamma(l, \omega) > \delta_\gamma(l) \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

其中： $r(l, \omega)$ 定義表示平滑化區域頻譜功率 $P_{Local}^S(l, \omega)$ 與其最小功率 $P_{min}(l, \omega)$ 之比值，如

式(9)：

$$\gamma(l, \omega) = P_{Local}^S(l, \omega) / P_{min}(l, \omega) \quad (9)$$

其中：區域頻譜功率 $P_{Local}^S(l, \omega)$ 的計算方式，如式(10)：

$$P_{Local}(l, \omega) = \sum_{i=-\omega_1}^{i=\omega_1} b(i) \cdot |Y(l, \omega - i)|^2 \quad (10)$$

其中： b 表示一個視窗函數其長度為 $2\omega_1 + 1 (\omega_1 = 1)$ 。

為了降低區域頻譜功率在各個音框間的變化程度，因此也對它施予平滑化處理，如式(11)：

$$P_{Local}^S(l, \omega) = \alpha_s P_{Local}^S(l-1, \omega) + (1 - \alpha_s) P_{Local}(l, \omega) \quad (11)$$

其中： $\alpha_s (\alpha_s=0.8)$ 是平滑參數。

接著將利用平滑化區域頻譜功率，求取其最小頻譜功率，首先將平滑化區域頻譜的前一個音框能量最小值和當前平滑化區域音框能量作比較，取其最小值作為當前音框最小值頻譜功率 $P_{min}(l, \omega)$ ，如式(12)：

$$P_{min}(l, \omega) = \min \{P_{min}(l-1, \omega), P_{Local}^S(l, \omega)\} \quad (12)$$

為了避免雜訊頻譜過度低估，設置參考功率頻譜 $P_{tmp}(l, \omega)$ ，在適當時機將平滑化區域功率頻譜的最小值提升，其計算方式，如式(13)：

$$P_{tmp}(l, \omega) = \min \{P_{tmp}(l-1, \omega), P_{Local}^S(l, \omega)\} \quad (13)$$

使用式(12)和(13)持續更新平滑化區域頻譜最小值，當更新音框數到達設定的音段長度 $L (0.5 \sim 1.5$ 秒)時，將 $P_{tmp}(l, \omega)$ 更新為目前平滑化區域頻譜的功率值，如式(14)：

$$P_{tmp}(l, \omega) = P_{LOCAL}^S(l, \omega) \quad (14)$$

平滑化區域頻譜的最小值 $P_{min}(l, \omega)$ 可以由式(15)得到，該式可以合理提升估測平滑化區域頻譜最小值的能量。

$$P_{min}(l, \omega) = \min \{P_{tmp}(l-1, \omega), P_{Local}^S(l, \omega)\} \quad (15)$$

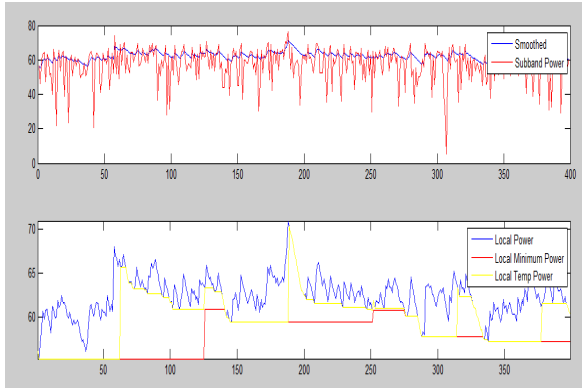


圖 1 雜訊估測頻譜強度軌跡圖:(上圖)平滑化功率頻譜軌跡圖；紅色表示含語音的雜訊 $Y(l, \omega)$ ；藍色曲線表示將 $Y(l, \omega)$ 平滑化後的頻譜 $P_{Local}^S(l, \omega)$ 。(下圖)平滑化區域頻譜功率最小值估測軌跡圖；紅色曲線表示平滑化頻譜最小值功率軌跡圖 $P_{min}(l, \omega)$ ；黃色曲線表示於平滑化最小功率頻譜之上調參考訊號軌跡圖 $P_{tmp}(l, \omega)$ 。

由圖 1 之上圖可以得知：使用平滑因子適度將功率頻譜軌跡平滑化，降減低受干擾語音訊號急遽變化特性，再利用該平滑化的功率頻譜估測雜訊，可以避免估測時，雜訊頻譜變動過度劇烈。在圖 1 下圖中，因 $P_{min}(l, \omega)$ 尋找平滑化頻譜最小值功率，雜訊位準明顯低估，因此透過 $P_{tmp}(l, \omega)$ ，在每一個音段上調至 $P_{Local}^S(l, \omega)$ 中，可以有效提高雜訊估測位準，最後取 $P_{tmp}(l, \omega)$ 與 $P_{Local}^S(l, \omega)$ 之最小值作為雜訊估測位準 $P_{min}(l, \omega)$ ，使得雜訊估測位準提升，準確度也更高。由於雜訊估測值 $P_{min}(l, \omega)$ 在許多區域仍然跟不上實際雜訊的變化，因此本論文嘗試改善 MCRA 演算法的雜訊估測位準之準確性。

3. 母音諧波頻譜調適之最小控制遞回平均(MCRA)雜訊估測法

由圖 1 $P_{min}(l, \omega)$ 的變化軌跡可以得知：雖然最小控制遞回平均雜訊估測法可以即時的追蹤雜訊頻譜強度的變化，估測效果也不錯；但是卻沒有考慮語音變化的特性，使得利用該估測雜訊頻譜從事語音增強後，會造成語音失真或殘留雜訊過多的現象，因此本文嘗試利用母音諧波頻譜特性，來調適最小控制遞回平均

雜訊估測法，提升增強語音的品質。由式(8)可以發現：在最小控制遞回平均雜訊估測法中，語音存在門檻值 $\delta_\gamma(l)$ 為固定值，並沒有考慮語音相關特性，這將會影響分類語音存在機率的準確度。因此我們使用母音諧波頻譜特性來調適 $\delta_\gamma(l)$ 的最適位準，讓使用母音諧波調適之最小控制遞回平均雜訊估測法 (vowel harmonic adapted MCRA, VH-MCRA) 估測法可以有效偵測微弱母音的區域，進而保留這些微弱母音，提升增強語音品質；其作法為降低語音存在門檻值 $\delta_\gamma(l)$ ；為了避免將雜訊區域歸類為有語音存在區域，我們會提高語音存在門檻值 $\delta_\gamma(l)$ ，確保語音存在機率估測的準確性。

首先我們先將受干擾語音透過低通濾波器來獲得 50~500Hz 諧波範圍的低通信號 $\phi(n)$ ，接著使用式(16)計算該低通信號的自相關函數 $R_\phi(\tau)$ ：

$$R_\phi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \phi(n) \cdot \phi(n + |\tau|) \quad (16)$$

其中： τ 是時間偏移量， N 是音框大小， n 是取樣點。

由式(16)可以得知：當低通受干擾語音信號 $\phi(n)$ 做偏移相乘時，自相關函數值會由高峰到谷底，再回到高峰處，而第二個高峰處即為擬尋找的音高週期。

為了準確取得音高週期自相關信號的峰值，可以透過平均強度變化函數 (average magnitude difference function, AMDF) 來求得，如式(17)所示：

$$AMDF(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} |\phi(n) - \phi(n + |\tau|)| \quad (17)$$

由式(17)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)，提升音高週期的準確性，如式(18)：

$$WAC(\tau) = \frac{R_\phi(\tau)}{AMDF(\tau) + \varepsilon} \quad (18)$$

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(AMDF)則出現極小值，將兩個函數分別置於分子和分母(如式(18))，利用分子與分母成反比的特性，將自相關函數與平均強度變化函數相除，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，可以提高音高週期的鑑別度。

取得音高週期之後，代入式(19)，即可得到基頻 F_0 的頻率估測值

$$F_0 = \frac{N}{T_0} \quad (19)$$

其中： T_0 代表週期。

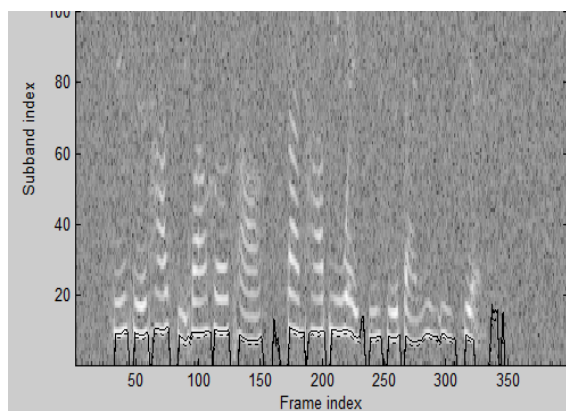


圖 2 基頻頻率估測軌跡圖。白色表示諧波頻譜；實線表示頻譜峰值 $F_0^*(l)$ ；虛線表示透過音高週期估測所得的基頻頻率軌跡 $F_0(l)$ 。

由於式(19)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(20)微調基頻頻率位置到 F_0^* ：[7]

$$F_0^* = F_0 - F_0^{Bias} \quad (20)$$

其中： F_0^{Bias} 為 F_0 與實際頻譜峰值的頻率偏差量，可由式(21)計算取得：

$$F_0^{Bias}(t) = \frac{1}{t_e - t_i} \cdot \sum_{l=t_i}^{t_e-1} F_0(l) - F_0'(l) \quad (21)$$

其中： t 表示音段索引； t_i 表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。

引。

基頻頻率的估測軌跡如圖 2 所示，由該圖可以看出：透過式(19)-(21)，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的估測狀況，以及修正偏差量的必要性。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；因此我們將利用這些強健諧波頻譜，調適語音存在門檻值 $\delta_\gamma(l)$ 。

強健諧波頻譜可由式(22)求得：

$$K = \{k \mid |F_0^k - F_0^{k-1}| \leq \delta_{F_0} \text{ and } |F_0^{k+1} - F_0^k| > \delta_{F_0}\} \quad (22)$$

其中： K 是強健諧波頻譜數量， F_0^k 為第 k 個諧波頻率， δ_{F_0} 為強健諧波的頻率變化閾值。

為了求得母音音框的強健諧波頻譜，我們從低頻往高頻的地方搜尋諧波頻率，若目前諧波頻率位置 F_0^k 與前一個低頻諧波頻率 F_0^{k-1} 間隔變化不大($\leq \delta_{F_0}$)，而且與下一個高頻諧波頻率 F_0^{k+1} 變化較大($> \delta_{F_0}$)，則將目前諧波頻率視為強健諧波的最高頻率，而諧波頻率索引則定義為強健諧波頻譜個數。

為了使用諧波頻譜調適式(8)的語音存在閾值 $\delta_\gamma(l)$ ，我們分別根據每個音框不同諧波頻譜數量 $N_h(l)$ 給予不同的權重值 $W(l)$ ，如式(23)所示：

$$\delta_h(l) = W(l) \cdot \delta_r(l) \quad (23)$$

其中：

$$W(l) = \begin{cases} \rho_1, & \text{if } N_h(l) \geq 3 \\ \rho_2, & \text{if } 0 < N_h(l) < 3 \\ \rho_3, & \text{otherwise} \end{cases} \quad (24)$$

依據強健諧波頻譜數量 $N_h(l)$ 的多寡程度，區分為高語音存在區權重 ρ_1 (具有3個(含)以上的強健諧波的權重值)、低語音存在區權重 ρ_2 (3個以下，0個以上的強健諧波)、以及無強健諧波權重 ρ_3 (0個諧波)三種頻譜類型，我們根據不同的雜訊類型、雜訊多寡和分佈的不同，給予不同的參數值，其實驗結果如表1所示。

表 1 各類型雜訊母音諧波最佳調適參數修正閾值

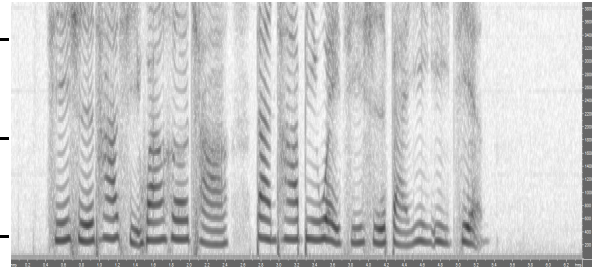
noise type	SNR (dB)	Optimum weight		
		ρ_1	ρ_2	ρ_3
White	0	0.2	0.1	1
	5	0.1	0.5	1
	10	0.1	0.7	0.4
F16	0	0.9	0.1	0.3
	5	0.8	0.3	0.4
	10	0.9	0.9	1
Factory	0	0.2	0.9	0.8
	5	0.9	0.9	0.7
	10	0.5	0.3	1.1
Helicopter	0	0.8	0.1	0.5
	5	0.7	0.9	1.3
	10	0.7	0.1	0.5
Babble	0	1	0.8	1.2
	5	0.9	0.6	0.9
	10	1.3	1.1	0.7
Car	0	0.7	0.3	0.6
	5	0.9	0.3	0.5
	10	1.3	1.1	0.7

我們對各種不同的雜訊類型在多樣污染程度的環境中，給定不同的參數調適 $\delta_r(l)$ ，以訊雜比(SNR)表現作為比較，尋得最佳參數，並且將結果呈現在表 1 中；由該表可以發現：在白色雜訊干擾的環境中， ρ_1 數值最小，代表在諧波頻譜較多的區域中，必須調降諧波頻譜調適閾值 $\delta_r(l)$ ，讓更多的頻譜歸類為語音信號；而在 F16 戰鬥機座艙雜訊、helicopter 直升機座艙雜訊、雞尾酒會(babble)人聲雜訊、汽車雜訊(car)等環境下， ρ_2 數值最小；代表在諧波頻譜稍多的區域，必須調降諧波頻譜調適閾值 $\delta_r(l)$ ，讓更多的頻譜歸類為語音信號，而具有稍多諧波頻譜的音框，所對應的 ρ_1 數值反而較大，使得更多頻譜被歸類為雜訊信號，這歸因於這四種雜訊屬於帶通彩色雜訊，因雜訊分佈在多個次頻帶，導致諧波頻譜數量被高估，為了提高雜訊移除效能，故調高 ρ_1 數值。

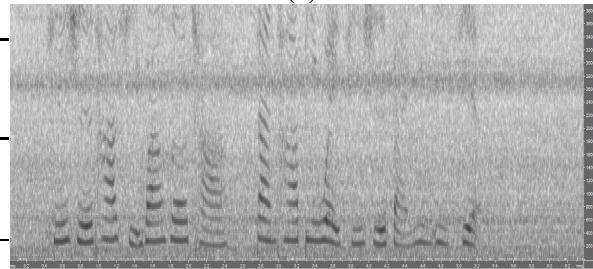
4. 實驗結果

在實驗中，語音受到各種雜訊干擾包括：白色雜訊、F16 戰機駕駛艙(F16-cockpit)、直升機駕駛艙(helicopter-cockpit)雜訊、工廠(factory)和雞尾酒會人聲(babble)雜訊、以及汽車(car)雜訊等等，這些雜訊信號乃摘錄自 Noisex-92

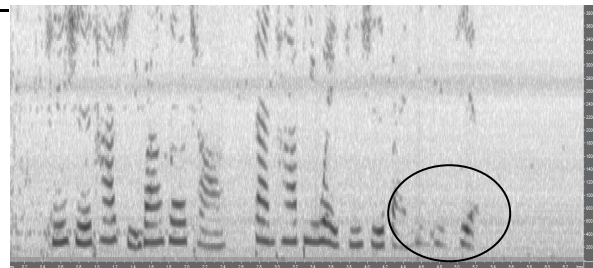
雜訊語料庫[15]，分別採用訊雜比值為 0，5，10 分貝做實驗，實驗中分別經過不同的雜訊，而以相同的語音增強方法[5]處理受干擾語音信號進行系統的評估。



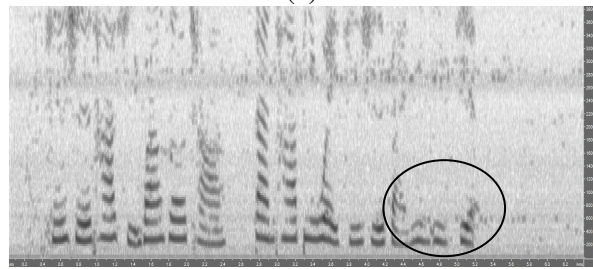
(a)



(b)



(c)



(d)

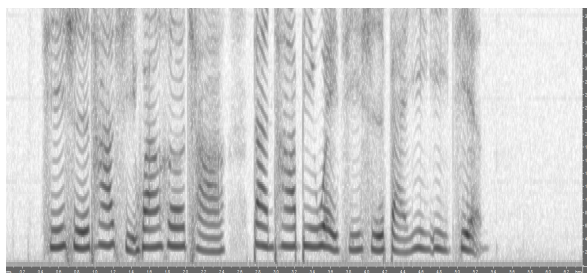
圖 3 語音聲譜圖；(a)乾淨語音；(b)受 F16 戰機座艙雜訊干擾的語音，訊雜比值為 5dB；(c)使用 MCRA 方法估測雜訊的增強語音；(d)使用本文方法估測雜訊的增強語音。

由表 2 可以發現：經本文方法處理後的訊雜比均優於 MCRA 演算法，尤其在汽車(car)雜訊，訊雜比改善度大幅提升，證明本文提出的方法可以改善 MCRA 雜訊估測法。由圖 3、4 的聲譜圖中橢圓形標示處可以看出：經本文方法處理後的諧波頻譜保留成效均優於 MCRA 的方法，據此說明本文提出的方法，可

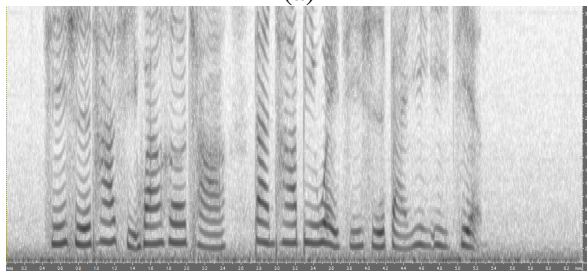
以改善增強語音的品質。由圖 5 至 10 的增強語音波形圖可以看出：經本文方法處理後的波形失真度，相較於 MCRA 而言，可以大幅的減少(如橢圓形標示處)，微弱母音的信號並沒有因為訊雜比較低，而給予過度的抑制，導致過多的語音失真，因此可以證實：母音出現的區域具有諧波頻譜，透過諧波頻譜調適語音存在機率，有效的將微弱的母音偵測出來，使其歸類為語音存在區域，提高增強語音的品質。

表 2 在各種不同雜訊的環境使用不同雜訊估測法經過語音增強後的語音訊雜比值比較表

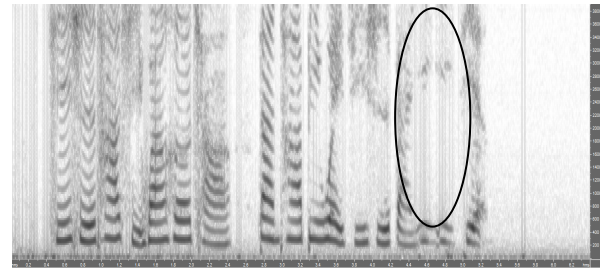
noise type	SNR (dB)	Method	
		MCRA	Proposed
White	0	7.05	8.95
	5	4.37	6.98
	10	1.25	5.04
F16	0	6.19	7.66
	5	4.06	6.40
	10	1.61	4.87
Factory	0	5.14	5.97
	5	3.51	4.82
	10	1.04	1.41
Helicopter	0	6.81	8.80
	5	4.67	6.90
	10	1.95	6.21
Babble	0	4.14	5.72
	5	2.77	4.66
	10	0.95	3.64
Car	0	8.52	14.50
	5	6.37	13.55
	10	3.66	12.17



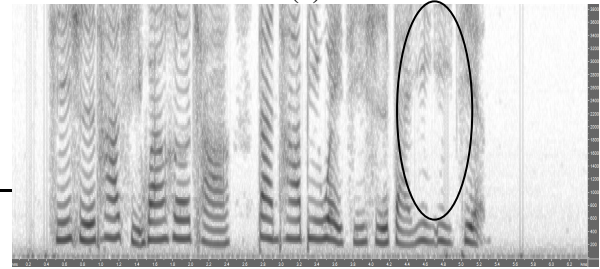
(a)



(b)



(c)



(d)

圖 4 語音聲譜圖；(a)乾淨語音；(b)受汽車(car)雜訊干擾的語音，訊雜比為 5dB；(c)使用 MCRA 方法估測雜訊的增強語音；(d)使用本文方法估測雜訊的增強語音。

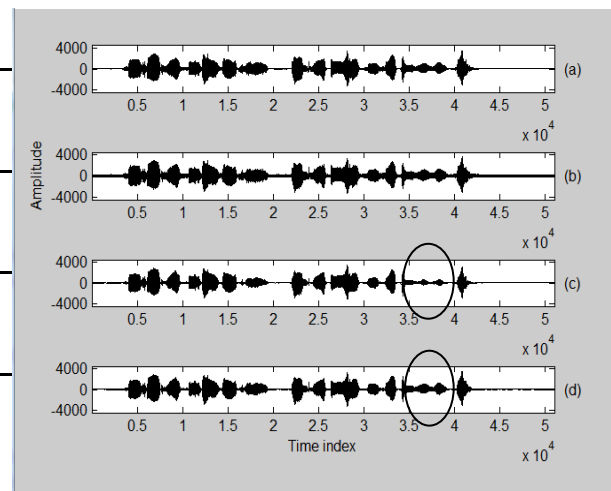


圖 5 語音波形圖；(a)乾淨語音波形；(b)受白色(white)雜訊干擾的語音波形，訊雜比為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

5. 結論

本文透過母音諧波頻譜調適最小控制遞回平均雜訊估測法中的語音存在機率門檻值，依據強健諧波頻譜數量作調適，提高語音存在機率的估測準確度，使得增強語音信號中，除了可以有效抑制背景雜訊之外，也能避免產生語

音失真，確保增強語音的品質及清晰度；實驗結果證明：本文提出的方法確實可以有效的提高雜訊強度估測的準確性，而且效能優於最小控制遞回平均雜訊估測法，使得增強語音聽起來更清晰而不吵雜。

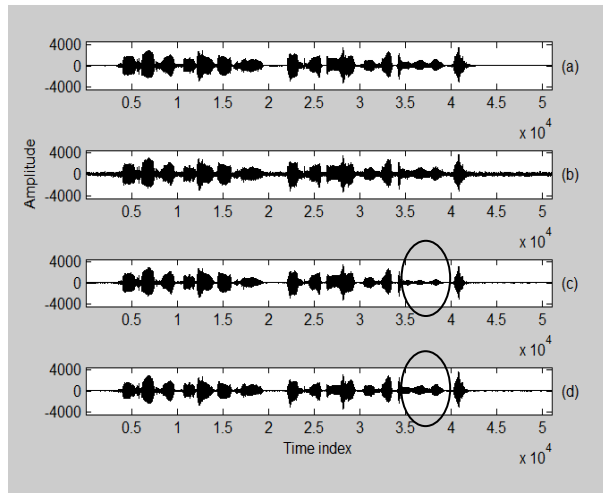


圖 6 語音波形圖；(a)乾淨語音波形；(b)受 F16 戰機座艙(F16-cockpit)雜訊干擾的語音波形，訊雜比值為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

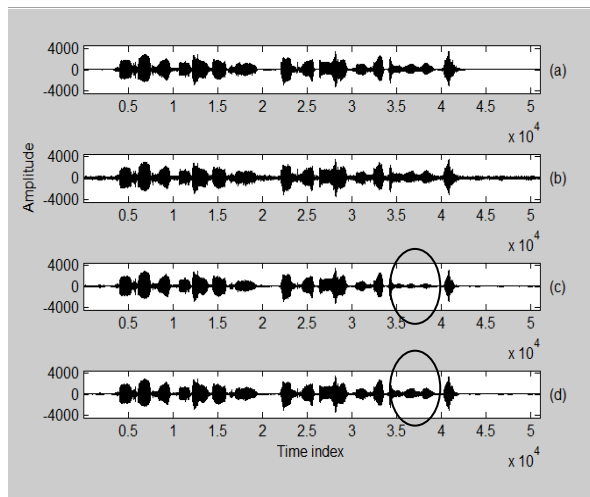


圖 7 語音波形圖；(a)乾淨語音波形；(b)受雞尾酒會的人聲(babble)雜訊干擾的語音波形，訊雜比值為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

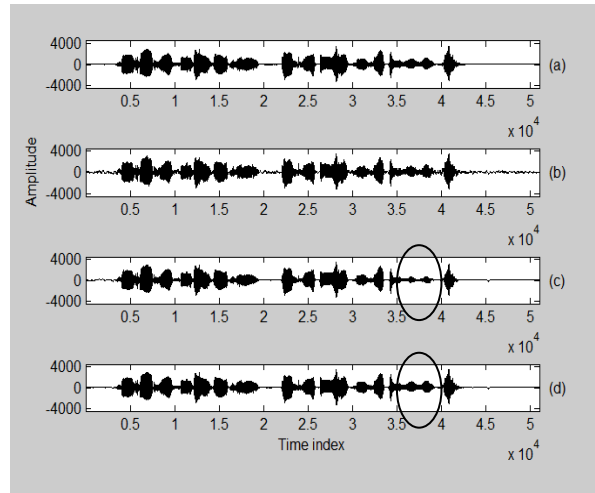


圖 8 語音波形圖；(a)乾淨語音波形；(b)受汽車(car)雜訊干擾的語音波形，訊雜比值為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

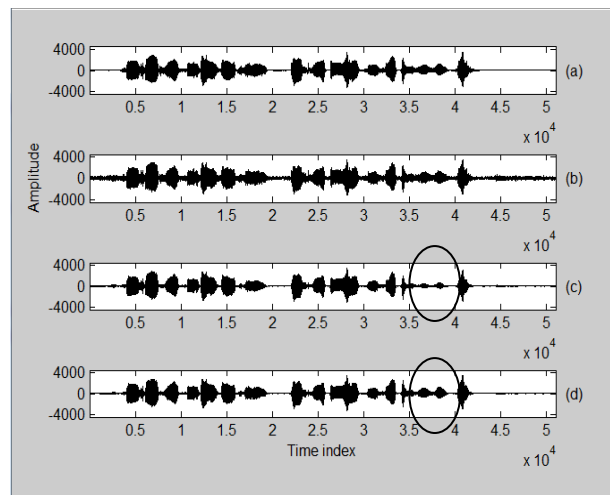


圖 9 語音波形圖；(a)乾淨語音波形；(b)受工廠(factory)雜訊干擾的語音波形，訊雜比值為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

致謝

本研究由國家科學委員會支助，計畫編號為：NSC 102-2221-E-468-004。

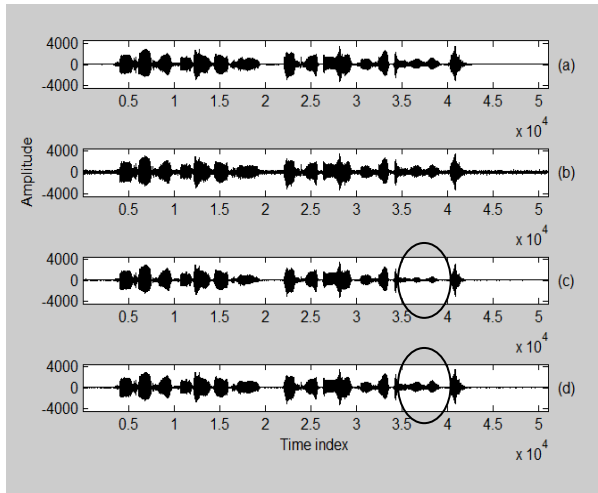


圖 10 語音波形圖；(a)乾淨語音波形；(b)受直升機座艙(helicopter-cockpit)雜訊干擾的語音波形，訊雜比為 10dB；(c)使用 MCRA 方法估測雜訊的增強語音波形；(d)使用本文方法估測雜訊的增強語音波形。

參考文獻

- [1] N. S. Kim and J.-H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, pp. 108-110, May 2000.
- [2] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft decision noise suppression filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, pp. 137-145, Apr. 1980.
- [3] I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noise environments," *Signal Process.*, vol. 81, pp. 2403-2418, Nov. 2001.
- [4] J. Meyer, K. U. Simmer, and K. D. Kammeyer, "Comparison of one-and-two-channel noise-estimation techniques," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997, pp. 137-145.
- [5] C. -T. Lu, J. -H. Shen, and K. -F. Tseng, "Speech enhancement using three-step-decision gain factor with optimal smoothing," *International Journal of Electrical Engineering*, vol. 18, no. 5, pp. 209-221, Oct. 2011.
- [6] C. -T. Lu, "Enhancement of single channel speech using perceptual-decision- directed approach," *Speech Communication*, vol. 53, no. 4, pp. 495-507, Apr. 2011.
- [7] C. -T. Lu, "Reduction of musical residual noise using block-and-directional-median filter adapted by harmonic properties," *Speech Communication*, vol. 58, pp. 35-48, Mar. 2014.
- [8] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, pp. 789-792.
- [9] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1995, pp. 153-156.
- [10] V. Stahl, A. Fischer, and R. Bippus, "Quantile based noise estimation for spectral subtraction and Wiener filtering," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 1875-1878.
- [11] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1994, pp. 1182-1185.
- [12] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech, and Audio Processing*, vol. 9, no. 5, pp. 504-512, July 2001.
- [13] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1995, pp. 1513-1516.
- [14] I. Cohen, and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12-15, Jan. 2002.
- [15] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247-251, July 1993.