藉由疊代權重知覺因子於語音增強之研究

洪國樑　　陸清達　　巫明諺
亞洲大學資訊傳播學系　亞洲大學資訊傳播學系　亞洲大學資訊傳播學系
研究生　　副教授　　研究生
horizon.3111@gmail.com　lucas1@ms26.hinet.net　rodpoy99@hotmail.com

摘要

殘留雜訊是大部分語音增強演算法所面臨的問題，所有演算法也都致力於抑制殘留雜訊的強度，並且讓語音信號保持在可接受的程度。其中以聽覺遮蔽閾值調適的演算法，是目前認為最理想的方法，該類方法是讓人耳聽不見的雜訊保留，藉由保留更多的受干擾語音信號，使得語音失真降低；但是雜訊頻譜估計不準確，使得增強語音的殘留雜訊依然很吵雜，因此本研究的目的就是為了找出效果更佳的降噪方法，使抑制噪音的效果更為理想。本研究是利用聽覺遮蔽閾值來控制雜訊抑制的刪減量，進而達到保留語音與抑制雜訊的效果，並且也施以多次疊代語音增益因子更新的方式來提升殘留雜訊抑制效果，實驗結果顯示：本研究提出的方法，對雜訊抑制有顯著的改善效果。

關鍵詞：聽覺遮蔽閾值、疊代、語音增強、雜訊抑制、頻譜刪減。

Abstract

Most speech enhancement systems suffer from annoying residual noise. Many studies attempt to suppress the magnitude of residual noise, enabling the speech quality to be kept at an acceptable level. An auditory masking-property adapted speech enhancement is the best method among the algorithms. It is due to the fact that this method preserves inaudible residual noise by preserving more amounts of noisy speech components, yielding the reduction of speech distortion. However, the residual noise is still annoying to the human ear in the enhanced speech. The major reason is the inaccurate estimation on the magnitude of background noise. In this study, we aim at improving the performance of a perceptual speech enhancement system by reducing more amount of residual noise. This can be obtained by iteratively modifying the noise masking threshold in the perceptual gain factor. Experimental results show that the proposed method can significantly improve the performance of the perceptual gain in noise reduction.

Keywords: auditory masking threshold, iteration, speech enhancement, noise reduction, spectral subtraction.

1. 前言

隨著科技的進步，語音錄製與語音傳遞早已是屢見不鮮的電子技術；但隨著語音錄製與傳遞工具的不同，不僅會產生雜訊，也會將週遭環境中的背景雜訊聲音一併錄製，造成語音可辨識度與舒適度降低。為了有效抑制背景雜訊，已有許多方法陸續提出[1]-[9]，[11]，[13]。雖然大部分的語音增強方法都可以針對

雜訊做有效的刪減與抑制，但在處理過的輸出信號中，仍會隨機產生令人難以忍受的音樂型殘留雜訊。有許多研究致力於削減音樂型殘留雜訊的相關研究[1]-[4]，[9]，[11]，[13]，但是因為音樂性殘留雜訊效應，使得處理後的語音依然很吵雜。為避免雜訊抑制後，殘留雜訊會干擾聲音的舒適程度，因此有許多方法將雜訊抑制在雜訊遮蔽閾值(noise masking threshold)之下，或是與知覺因子(perceptual gain factor)結合的方法[3]，使雜訊成為不被人耳察覺的低度能量，即使殘留雜訊存在，也不會覺得吵雜。

在進行雜訊與語音能量估測時，為增強估測能量的精確值，有研究提出使用疊代的方式去增強雜訊抑制的效果[5]；在雜訊消除的過程中，往往會先行測試信噪比，再來調整參數。有論文提出二步驟決策導向的方法改進預先估計信噪比的方法，實驗結果顯示了該方法有顯著的改善[6]。雜訊估計方面，有人提出調適時間-頻率方塊閾值的演算法來增強聲音信號。這個方法藉著最小化雜訊估計的估測來校正信號特徵的參數，且實驗結果也顯示可以改進聲音信號的品質[7]。也有方法提出了混合使用溫納濾波器壓制雜訊及利用頻譜二維矩陣來維持語音品質，與刪減語音降噪後所產生的殘留雜訊[8]。Virag[9]對語音增強加入了聽覺遮蔽閾值模型的研究，這是一個可以即時自動調適的語音增強系統，還可以在人耳知覺上找到最佳的調適，避免增強後的語音產生不自然的聽覺效果，因此這些加入聽覺遮蔽閾值調適的語音增強方法，已廣泛的被大家認為是理想的語音增強方法。

由於雜訊頻譜估計並不準確，使得雜訊抑制的效果仍有缺失。因此本研究提出的方法，是在聽覺遮蔽閾值中，加入權重來調適知覺因子的演算法；並經由多次疊代，計算每一個次頻帶中的增益因子，來加強對雜訊的抑制效果，進而達到有效的抑制雜訊與降低語音失真的目的。實驗結果顯示：經本研究提出的方法增強處理後的語音信號，在聲譜圖與波形圖中都可以看出對抑制殘留雜訊有明顯的改善；主觀聽覺測試也證明抑制雜訊的能力有明顯效果，也能有效的防止語音失真；實驗最後也將本研究提出的方法與知覺因子[3]作比較，同時使用客觀的最小統計演算法[10]來測量信噪比，證明了本研究的方法對於抑制雜訊有顯著的改善。

2. 疊代知覺語音增強法

2.1 系統方塊圖

本研究提出的方法流程如下圖 1 所示：首先，將一個受雜訊干擾的語音信號音框化，接著使用傅立葉轉換將信號轉換到頻域，然後使用最小統計量法估計雜訊能量[10]，以及聽覺遮蔽閾值的能量預估。由於聽覺遮蔽閾值的特性，估計到高能量的位置，通常說明著有語音存在，相反的，低能量的位置，則較少有明顯的語音，因此利用估計到的聽覺遮蔽閾值能量，進行知覺增益因子的權重計算，以決定雜訊能量抑制與語音能量保留的多寡。但因為雜訊頻譜能量估計並不會完全準確，導致增強後的語音信號中，殘留雜訊含量依然很多，聽起來也非常吵雜；因此我們施以疊代，重複計算增益因子，使雜訊能量可以有效抑制，語音的部分可以保留避免因誤刪而引起更多的語音失真，提升語音增強的效能。

2.2 聽覺遮蔽閾值

Virag[9]在語音增強中，加入了聽覺遮蔽閾值模型的相關研究；這聽覺遮蔽閾值反應了人耳的生理反應，當人耳在專心聽一段語音時，往往會將周遭不明顯的聲音自動忽略，這

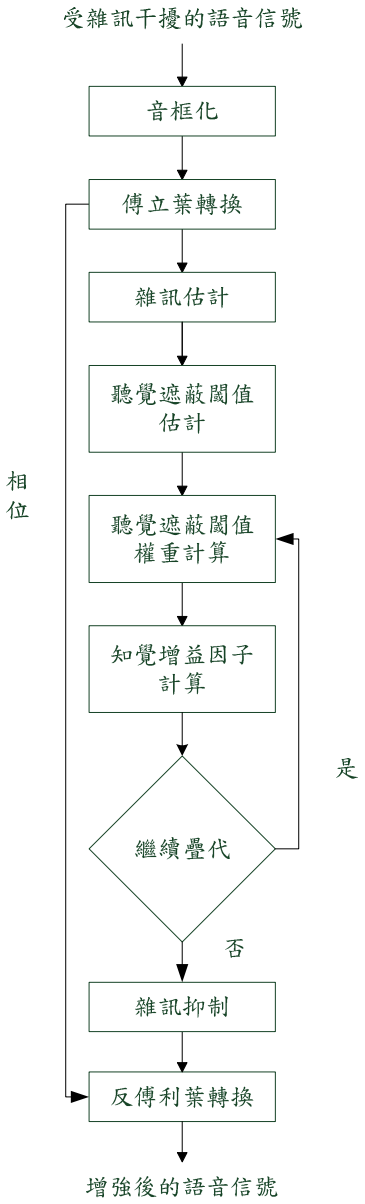


圖 1 疊代知覺語音增強系統流程圖

時候語音的能量，便成為控制聽覺遮蔽閾值能量的指標了。

對一個語音增強系統來說，因為現今仍沒辦法很明確的對語音與雜訊做自動化區別，或多或少會有些誤判，通常會將影響力較薄弱的語音信號判斷為雜訊，使得語音增強系統將該較弱語音特徵刪除，造成語音增強後會產生極為不自然的語音失真。所以只是純粹將雜訊剃除，那麼人耳聽起來的語音增強效果會顯得非常不自然。既然沒辦法將一段語音很明確的判斷是語音特徵或雜訊區段，Virag[9]就提出結合聽覺遮蔽閾值的語音增強系統，對雜訊區段保留些許不影響語音舒適度的雜訊，以避免產生類似語音失真這種不自然的處理效果，且這些加入聽覺遮蔽閾值調適的語音增強方法，已廣泛的被大家認為是理想的語音增強方法。

求得聽覺遮蔽閾值的方法分成三個步驟，首先將被雜訊干擾的語音信號切割成每個音框長度為 256 個取樣點，取樣頻率為 8kHz，藉由快速傅立葉轉換成頻域的信號，再根據巴克頻帶表(Bark scale, 表 1)，將信號劃分為 18 個臨界頻帶。接著對每個不同的臨界頻帶，使用擴散函數（如圖 2）做摺積(convolution)計算。最後，根據雜訊與語音型態不同的，減去

表 1 巴克頻帶(Bark scale)與線性頻帶對照表。[9]

關鍵頻帶編號	間隔	間隔數	對應頻率
1	1-3	3	0-94
2	4-6	3	94-187
3	7-10	4	187-132
4	11-13	3	312-406
5	14-16	3	406-500
6	17-20	4	500-625
7	21-25	5	625-781
8	26-29	4	781-906
9	30-35	6	906-1094
10	36-41	6	1094-1281
11	42-47	6	1281-1469
12	48-55	8	1469-1719
13	56-64	9	1719-2000
14	65-74	10	2000-2312
15	75-86	12	2312-2687
16	87-100	14	2687-3125
17	101-118	18	3125-3687
18	119-128		3687-4000

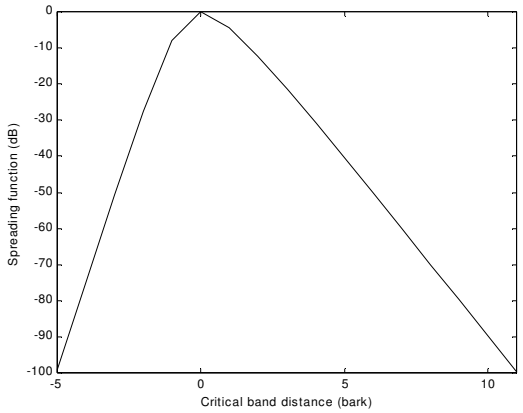


圖 2 擴散函數表示圖。

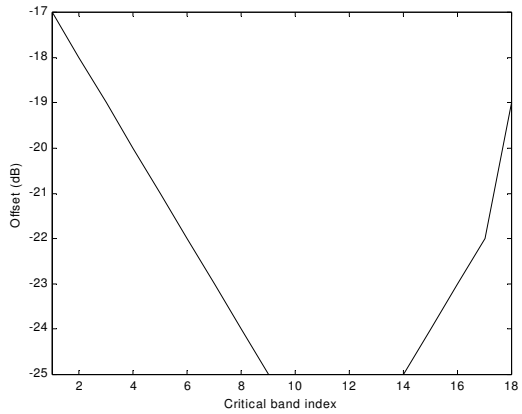


圖 3 相對閾值偏移表示圖，本論文計算聽覺遮蔽閾值的閾值偏移量。

相對閾值偏移(圖 3)後即可計算出聽覺遮蔽閾值。

由於計算聽覺遮蔽閾值，必須使用語音增強後的信號，因此在估算聽覺遮蔽閾值之前，會先估計雜訊能量，並做一次簡單的語音增強，但這樣卻會增加了音樂型雜訊，尤其是在低信噪比的區段；而那些音樂型雜訊，出現在乾淨語音信號的區段時，會影響到聽覺遮蔽閾值的估測，尤其是在臨界頻帶大於 15 的高頻區域，因此降低高頻區域的相對閾值偏移量，以取得更準確的聽覺遮蔽閾值。[9]

2.3 聽覺遮蔽閾值權重

一段受雜訊干擾的語音信號 $y(m,n)$ 可以用一段乾淨語音 $s(m,n)$ 再加上雜訊 $d(m,n)$ 來表示，如式(1)，其中 m 為音框索引，n 為取樣點索引：

$$y(m,n)=s(m,n)+d(m,n) \tag{1}$$

經傅立葉轉換後，受雜訊干擾的語音信號 $Y(m,\omega)$ 經每個次頻帶的增益因子 $g(m,\omega)$ 做語音增強處理後，可以得到增強後的語音信號頻譜 $\hat{S}(m,\omega)$。

$$\hat{S}(m,\omega)=g(m,\omega)\cdot Y(m,\omega) \tag{2}$$

經過語音增強後的語音信號頻譜 $\hat{S}(m,\omega)$，並不會完全等於乾淨的語音信號 $S(m,\omega)$，這之間的差異，就稱為頻譜失真 $E(m,\omega)$。

$$E(m,\omega)=\hat{S}(m,\omega)-S(m,\omega)=E_S(m,\omega)+E_R(m,\omega) \tag{3}$$

其中 $E_S(m,\omega)$ 為語音失真，而 $E_R(m,\omega)$ 則為殘留雜訊。

$$E_S(m,\omega)=[g(m,\omega)-1]\cdot S(m,\omega) \tag{4}$$

$$E_R(m,\omega)=g(m,\omega)\cdot D(m,\omega) \tag{5}$$

$S(m,\omega)$ 與 $D(m,\omega)$ 分別表示為乾淨語音信號的頻譜與雜訊頻譜。如果雜訊是與語音統計特性不相關的加入型雜訊，那麼增益因子 $g(m,\omega)$ 可以經由限制殘留雜訊低於聽覺遮蔽閾值[3]的情形下，讓語音失真最小化的方式來得到。

$$\min_{g(m,\omega)}\left\{E_S^2(m,\omega)\right\} \quad \text{subject to the constraint } E_R^2(m,\omega)\le T(m,\omega) \tag{6}$$

知覺增益因子可由下式表示[3]：

$$g_P(m,\omega)=\frac{1}{1+\max\left(\sqrt{\frac{|D(m,\omega)|^2}{T(m,\omega)}}-1,\ 0\right)} \tag{7}$$

其中 $T(m,\omega)$ 與 ω 分別表示聽覺遮蔽閾值與對應的頻帶：每個臨界頻帶裡的聽覺遮蔽閾值能量都相同。而本研究提出的知覺增益因子則做了下列的修改：

$$g_P(m,\omega)=\frac{1}{1+\max\left(\sqrt{\frac{|D(m,\omega)|^2}{\gamma(m,\omega)\cdot T(m,\omega)}}-1,\ 0\right)} \tag{8}$$

$$\gamma(m,\omega)=\begin{cases}\alpha\cdot\dfrac{10\cdot\log_{10}(T(m,\omega))}{T_{ref}} & \text{，若} 10\cdot\log_{10}(T(m,\omega))<T_{ref}\\ 1 & \text{，其他}\end{cases} \tag{9}$$

其中α為權重因子，T_{ref}為聽覺遮蔽閾值的判斷值。

在式(9)中，當聽覺遮蔽閾值小於T_{ref}值，那麼將該臨界頻帶的信號歸類為雜訊為主區，故增加抑制的權重；相反的，如果聽覺遮蔽閾值大值T_{ref}值，那麼將視為語音為主區，$\gamma(m,\omega)$值取 1，不予處理該頻帶的頻譜，保留所有語音部分。而權重因子α則可以控制雜訊與語音的刪減量，有效的輔助與控制知覺因子，使增強後的語音不會剩下過多的殘留雜訊或過度的失真。將式(8)得到的增益因子代入式(2)，即可得到語音增強後的前處理信號。接著針對這前處理信號的殘留雜訊進行刪減，讓語音增強的效果更好，因此本研究加入了疊代以及權重因子。為了判別語音增強後的信號，是否需要施以疊代，或是更改疊代次數以及權重因子α，因此對前處理過的信號做後信噪比測試。

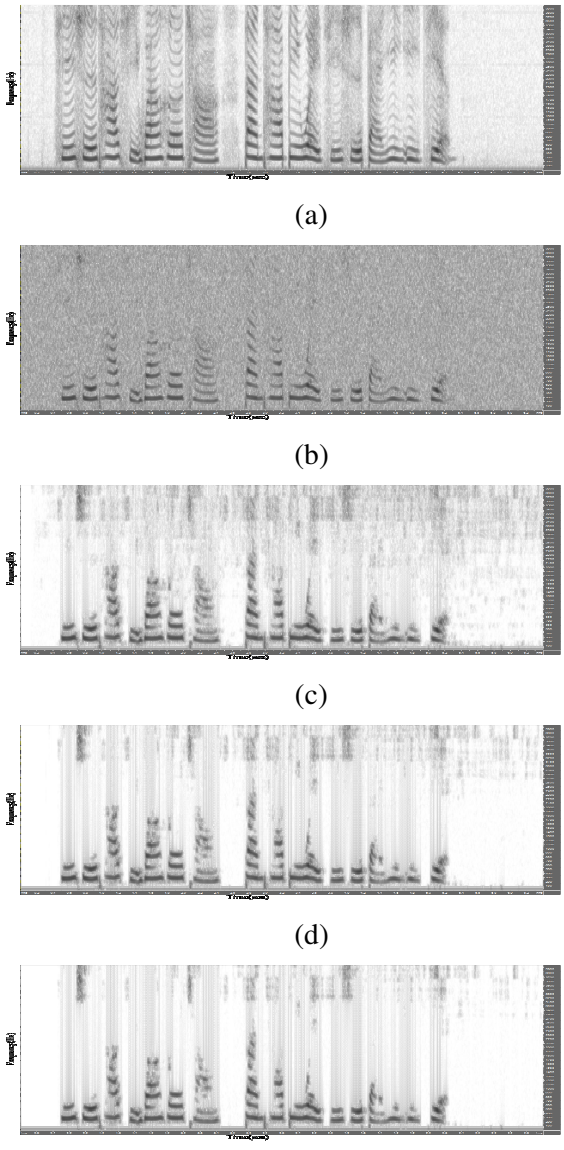
$$\xi(m,\omega)=20\cdot\log_{10}\left(\frac{|\tilde{S}(m,\omega)|}{|\tilde{R}(m,\omega)|}\right) \qquad (10)$$

$\tilde{S}(m,\omega)$與$\tilde{R}(m,\omega)$分別為前處理信號中估計到的語音頻譜與殘留雜訊能量。得到了後信噪比，接著施以疊代、改變疊代次數或更改權重因子α；隨著疊代的次數增加與更改權重因子α後，殘留雜訊的能量就會越小，且再次測得的後信噪比也隨著增加，當測量出的後信噪比到達最大值時，此時就是客觀測試中，效果最好的時候。

最後，經過數次疊代與調整權重因子α後，將得到的增益因子式(8)代入式(2)，接著使用反傅立葉轉換，將在頻域處理後的信號反轉回時域，即可得到語音增強後的語音增強信號。

2.4 知覺因子疊代

準確的估計雜訊區段與能量，是語音增強系統中極為困難的一環，當語音部分被系統誤判為雜訊時，處理後的效果便會產生語音失真；而雜訊量估計的太低，又會導致整個降噪系統的效果不佳；因此使用疊代的方法，來重複計算增益因子，使雜訊量能夠有效刪除，並且必須保留語音部分，避免造成語音失真，補強整個語音增強系統的效能。



(a)

(b)

(c)

(d)

(e)

圖 4 使用本論文方法語音增強後的語音聲譜圖，(a)乾淨語音的聲譜圖，(b)被白色雜訊干擾的語音(訊雜比值為 5dB)，(c)疊代 3 次，(d)疊代 5 次，(e)疊代 7 次。

圖 4 為使用本論文提出的方法語音增強後的聲音頻譜圖，使用白色雜訊，訊噪比為 5dB，從上而下的疊代次數依序為 3、5、7，從圖中可以明顯看出，使用了七次疊代的語音增強效果，雜訊的頻譜量較五次與三次明顯少了許多，但因為刪減量過大，會將部分語音特徵也抑制了，因此聽起來聲音的失真程度較大，信噪比也比五次疊代低；而經由多次後信噪比測試後，訊噪比值為 5dB 時，疊代次數設定為 5，語音增強後的信噪比會最高，語音舒適度也最佳。

3. 實驗結果

本研究實驗所使用的雜訊信號是由 Noisex-92 語音資料庫提供。受雜訊干擾的語音信號，由乾淨語音分別加入 6 種的雜訊，分別為白雜訊、F16 機艙雜訊、工廠雜訊、直升機座艙雜訊、雞尾酒會的人聲雜訊以及汽車雜訊。而乾淨語音則由一位女生，講一段國語錄製而成。加入雜訊後的信噪比分別為 0、5 以及 10 分貝。在這實驗結果中，使用了客觀的最小統計演算法[10]來測量信噪比，同時也將本研究提出的方法與知覺因子[3]作比較。

本研究所提出的方法，有兩個可控制因子，分別為權重因子α與疊代次數。經由信噪比輔以測試的結果，分別得出受雜訊干擾的語音信號信噪比為 0 分貝時，權重因子設定為 1.9 倍，疊代次數為 9 次；信噪比為 5 分貝時，權重因子設定為 0.8 倍，疊代次數為 5 次；信噪比為 10 分貝時，權重因子設定為 0.2 倍，疊代次數為 2 次，可以讓增強後的語音信號成為較理想的結果。

表 2 為被雜訊干擾的語音，使用知覺因子方法抑制雜訊後的語音與使用本論文提出的方法抑制雜訊後的信噪比改進值。表中可以明顯看出，不管語音被何種雜訊干擾，經本研究的方法語音增強後，即便是被雜訊量更大的噪音破壞，處理雜訊的效能都較知覺因子方法優異。

表 2 被雜訊干擾的語音與使用不同方法抑制雜訊後的 SNR 改進值。

noise type	SNR (dB)	enhancement method: perceptual	enhancement method: proposed
white	0	2.58	**7.39**
	5	1.48	**4.60**
	10	0.30	**2.25**
F16-cockpit	0	1.77	**4.20**
	5	0.97	**3.08**
	10	-0.12	**0.48**
factory	0	0.35	**2.10**
	5	-0.33	**1.29**
	10	-1.09	**-0.58**
helicopter-cockpit	0	2.28	**5.74**
	5	1.49	**3.74**
	10	0.36	**0.82**
babble	0	0.38	**2.50**
	5	-0.36	**1.43**
	10	-1.01	**-0.31**
car	0	2.31	**9.39**
	5	1.75	**6.24**
	10	0.88	**1.54**

表 3 為被雜訊干擾的語音，使用知覺因子方法抑制雜訊後的語音與使用本論文提出的方法抑制雜訊後的改良版巴克頻譜失真測量法 (MBSD) 分數值。在客觀的測量法中，以巴克頻譜失真測量法(bark spectral distortion, BSD)的接受度最高[12]。而 Yang 等人[14]後來又提出了改良版的巴克頻譜失真測量法(modified BSD , MBSD)，其原理是在巴克頻譜失真測量過程上，多考量聽覺遮蔽閾值的效應。由於聽覺遮蔽閾值的特性，人耳往往會將低於聽覺遮蔽閾值的能量自動忽略，原因

表 3 使用不同方法抑制雜訊後的改良型巴克頻譜失真測量法 (MBSD) 分數值。

noise type	SNR (dB)	enhancement method perceptual	proposed
white	0	5.55	7.50
	5	2.93	2.63
	10	1.23	0.63
F16-cockpit	0	3.31	0.78
	5	1.17	0.57
	10	0.29	0.19
factory	0	1.39	5.74
	5	0.27	1.19
	10	0.07	0.16
helicopter-cockpit	0	0.96	1.27
	5	0.20	0.45
	10	0.03	0.13
babble	0	2.71	0.57
	5	1.13	0.20
	10	0.41	0.15
car	0	0.25	0.35
	5	0.18	0.23
	10	0.13	0.01

是它不會對人耳造成聽覺上的影響，藉此理由將原本的 BSD 方法中，任何低於聽覺遮蔽閾值的語音失真能量刪除，使語音品質評估的結果更接近人耳的感受度。而 MBSD 方法的計算公式可由下式表示[14]：

$$MBSD = \frac{1}{M}\sum_{m\in\{1\}}\sum_{\xi=0}^{K-1}\eta(\xi)\cdot\left|L_S(m,\xi)-L_{\hat{S}}(m,\xi)\right| \quad (11)$$

其中 M 與 K 分別代表有語音分佈的音框數量與一個音框中的臨界頻帶數量；m 為音框索引；ξ 為臨界頻帶索引；$L_S(m,\xi)$ 與 $L_{\hat{S}}(m,\xi)$ 分別表示原始語音及語音增強後的語音臨界頻帶之頻譜強度。$\eta(\xi)$ 表示失真考量旗號，若語音失真大於聽覺遮蔽閾值，人耳感受得到該語音，故該語音失真必須考慮，所以 $\eta(\xi)$ 設為 1；相反的，若語音失真低於聽覺遮蔽閾值，則該失真人耳聽不到，所以失真度不需要考慮，故 $\eta(\xi)$ 設為 0。

由(11)式可以看出，改良版的巴克頻譜失真(MBSD)所產生的分數值，越低代表語音品質越佳。

由表三使用 MBSD 方法的數據分析可得知：本論文提出的方法，可將背景雜訊幾乎完全抑制，使得本研究提出的方法所產生的 MBSD 數值優於知覺因子；而在處理嚴重的白色雜訊干擾(0dB)環境中，也將較微弱的語音部份一併刪除，導致 MBSD 分數值略為遜色；在工廠雜訊與直升機座艙雜訊干擾的情形下，也有相似的問題。而對於極難處理的多重人聲雜訊(babble)干擾的環境中，本文提出的方法亦優於知覺因子的效能。

圖 5 為波形圖的比較，這是一段女性的錄音，且語音受到加入型白雜訊干擾(訊雜比值為 0dB)，由圖 4(C)中明顯可以看出，雖然知覺因子的方法對雜訊抑制已有不錯的效果，但本研究提出的方法較知覺因子更有能力抑制背景雜訊(如圖 4(d)所示)，尤其是在靜音

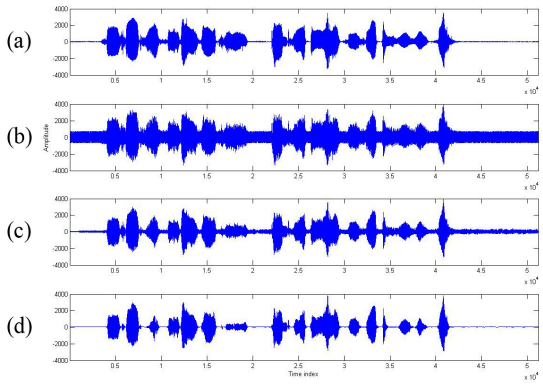
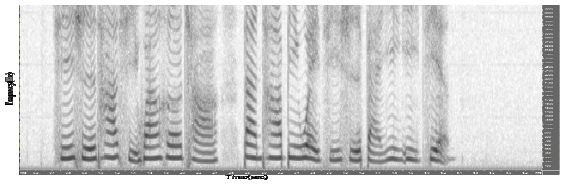


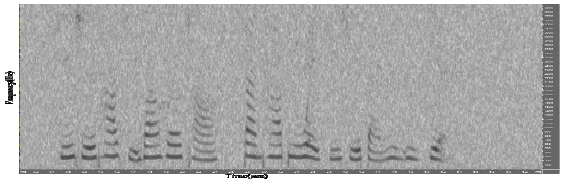
圖 5 聲音波形圖，(a)乾淨語音，(b)被白色雜訊干擾的語音(訊雜比值為 0dB)，(c)使用知覺因子增強後的語音，(d)使用本文提出方法增強後的語音。

區段；相對的，在語音的區段，本研究依然能夠保存語音信號，而不至於過度刪減，導致語音失真聽不清楚。

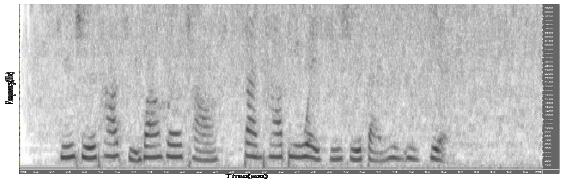
圖 6 為語音聲譜圖，語音信號為女性的說話的聲音，且將語音加入白雜訊干擾，其中訊雜比值為 5dB(如圖 6(b)所示)。從圖中也可明顯的看出知覺因子的方法以及本研究提出之方法的優劣。從圖 6(d)可以發現，使用本研究提出的方法處理後，幾乎把人耳易查覺的低頻噪音都刪減了，且濁音中的諧波頻譜也明顯有保留住，說明了語音失真不會很嚴重。經由主觀的聽覺測試，在不同信噪比的干擾環境中，使用本研究的演算法來抑制背景雜訊的效果



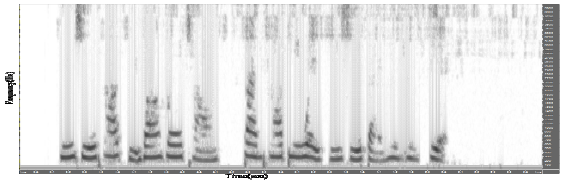
(a)



(b)



(c)



(d)

圖 6 語音聲譜圖，(a)乾淨語音，(b) 被白色雜訊干擾的語音(訊雜比值為 5dB)，(c)使用知覺因子增強後的語音，(d)使用本論文提出的方法增強後的語音。

(如圖 6(d))，皆明顯優於知覺因子的方法(如圖 6(c))，且處理後的語音聽起來也比知覺因子方法舒適，由此實驗結果證明：本研究所提出的藉由疊代權重知覺因子方法對於語音增強，可以明顯改善知覺因子的效能，主要的原因就是本方法可以有效的抑制背景雜訊，以及避免對語音信號誤判刪減，確保增強後的語音信號品質。

本研究的主要貢獻：藉由聽覺遮蔽閾值的權重因子來調適，並施以疊代計算知覺增益因子，達到抑制背景雜訊與保留語音信號的目的。經過實驗結果證實：，無論在客觀或是主觀的測量方法上，本研究提出的方法在雜訊殘留量與語音舒適度都明顯優於知覺因子方法。

4. 結論

本研究藉由疊代與權重知覺因子方法，來改善知覺增益因子消除雜訊的效果。聽覺遮蔽閾值受到權重因子調適之後，可以調降雜訊為主區域的聽覺遮蔽閾值，並且搭配多次疊代的方式，降低聽覺遮蔽閾值，使得所對應的知覺因子降低，達到抑制背景雜訊的目標；對於語音區域的部分，則不予處理，避免產生更多的語音失真。經過實驗結果證實：無論在客觀或是主觀的測量方法上，本研究提出的方法在雜訊殘留量與語音舒適度都明顯優於知覺因子方法。

誌謝

本研究由行政院國家科學委員會專題研究計畫之經費支助，計畫編號為 NSC 99-2221-E-468-001。

參考文獻

[1] A. Amehraye, D. Pastor, and A. Tamtaoui, “Perceptual improvement of Wiener filtering,” Proc. ***IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 08)***, 2008, pp. 2081-2084.

[2] C. –T. Lu, K. –F. Tseng, C. –T. Chen, M. –Y. Wu, K. –L. Hong, and T. –T. Chou, “Speech enhancement using three-step-decision gain factor,” in Proc. ***National Symposium on Telecom. (NST)***, 2010.

[3] C. -T. Lu and –H. C. Wang, “Speech enhancement using perceptually-constrained gain factors in critical-band-wavelet-packet transform,” ***Electron. Lett.***, vol.40, no.6, pp.394-396, 2004.

[4] C. –T. Lu and K. -F. Tseng, “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.

[5] C. –T. Lu, K. –F. Tseng, C. –T. Chen, and J. –P. Wang, “Reduction of residual noise for speech enhancement using iterative noise reduction gain factor,” ***National Symposium on Telecom. (NST)***, pp. 350-354, 2009.

[6] C. Plapous, C. Marro, and P. Scalart, “Improved singal-to-noise ratio estimation for speech enhancement,” ***IEEE Trans. Audio Speech Languge Process.***, vol. 14, no. 6, pp. 2098-2108, 2006.

[7] G. Yu, S. Mallat, and E. Bacry, “Audio denoising by time-frequency block thresholding,” ***IEEE Trans. Signal Process.***, vol. 56, no. 5, 2008.

[8] H. Ding, I. Y. Soon, S. N. Koh, and C. K. Yeo, “A spectral filtering method based on hybrid Wiener filters for speech enhancement,” ***Speech Commun.***, vol. 51, pp. 259-267, 2009.

[9] N. Virag, “Single channel speech enhancement based on masking properties of the human auditory system,” ***IEEE Trans. Speech Audio Process.***, vol. 7, no. 2, 126-137, 1999.

[10] R. Martin, “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” ***IEEE Trans. Speech Audio Process.***, vol. 9, no. 5, pp. 504-512, 2001.

[11] S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” ***IEEE Trans. Audio Speech Language Process.***, Vol. 18, No. 2, 2010.

[12] S. Wang, A. Sekey, A. Gersho, “An objective measure for predicting subjective quality of speech coder,” ***IEEE J. Select. Areas Commun.,*** vol. 10, no. 5, pp. 819-829, June 1992.

[13] W. Jin, X. Liu, M. S. Scordilis, L. Han, "Speech Enhancement Using Harmonic Emphasis and Adaptive Comb Filtering," ***IEEE Trans. Audio Speech Language Process.***, pp. 356-368, Vol. 18, No. 2, 2007.

[14] W. Yang, M. Benbouchta, and R. Yantorno, “Performance of the modified bark spectral distortion as an objective speech quality measure,” in Proc. ***Int. Conf. Acoust., Speech, Signal Processing,*** vol. 1, Washington, USA , 1998, pp. 541-544.