

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

巫明諺
亞洲大學資訊傳播學系
研究生
rodpo99@hotmail.com

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

洪國樑
亞洲大學資訊傳播學系
研究生
horizon.3111@gmail.com

摘要

在背景雜訊抑制處理中，雜訊強度估計的準確度，直接影響到雜訊抑制後的音訊品質，所以本文嘗試將多重音框內插法加入零值音框嵌入法中，提升語音音框中雜訊頻譜估計的準確度。零值音框嵌入法是在語音信號中，插入零值音框，然後將信號受到通道雜訊的干擾，由於零值嵌入處沒有語音信號，因此利用該處估計雜訊頻譜強度；為了提升語音音框的雜訊頻譜估計準確度，我們將最接近的前後兩個零值音框嵌入處的雜訊頻譜，施以多重音框內插法估計雜訊頻譜強度，之後再由雜訊頻譜與聽覺遮蔽值共同調適知覺增益因子，處理受干擾語音信號，達到抑制雜訊的目的；經過實驗證實本文所提出來的可以大幅改善了零值嵌入法與零值音框嵌入法的效果。

關鍵詞：語音增強、雜訊抑制、雜訊估計、零值音框嵌入法、知覺因子。

Abstract

The estimation of noise spectrum is an important task for the application of noise reduction. It is attributed to the fact that the accuracy of noise estimation can dominate the performance of enhanced speech. The idea of frame-zero-padding method is to periodically insert zero frames into a clean speech signal. Hence this zero padded signal is transmit through a communication channel with the corruption of additive noise. Because speech signals are absent at the zero-padded frames. Thus we can estimate the magnitude of background in the zero-padded frames. In order to improve the estimation accuracy for channel noise, the noise spectra are estimated by interpolating

the noise spectra between two successive zero-padding frames. Hence, both the estimated noise magnitude and the noise masking threshold are employed to adapt a perceptual gain factor for noise reduction. Experimental results show that the proposed method significantly outperforms the zero-padding and the frame-zero-padding methods.

Keywords: Keywords: speech enhancement, noise reduction, noise estimation, frame-zero-padding, perceptual gain factor.

1. 前言

語音增強處理是一種被多媒體相關領用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、電話、錄音機等都需要使用語音增強處理才能發揮更好的效能；而語音增強系統可以分為單通道語音增強處理與多通道語音增強處理，其中單通道語音增強處理較為困難，因為當語音只由一個麥克風所接收的時候，語音與雜訊會混合在一起，要如何把語音信號中的雜訊成分移除，是一項極具挑戰與有趣的研究主題，而本研究的主要目的就是要移除單通道中的背景雜訊信號。

在單通道的噪音環境中，作語音增強的研究有很多，文獻[1]中利用感知的噪聲遮蔽閾值(noise masking threshold)來調適語音增強系統的參數，比較可以防止音樂型噪音的產生；知覺增益因子是一種使用人耳聽覺感知特性[13][14]的語音增強技術，使用人耳聽覺感知特性可以將雜訊抑制到人耳聽不見的程度，且不會產生過度刪減的問題，所以將刪減參數經由知覺增益因子調適後可以大幅改善刪減過度的問題。有些方法則是使用後處理的技術，將音樂型雜訊降低，文獻[7]提出對原始語音作預

處理的語音增強技術，該方法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的振幅，這樣的方法在取樣點變化緩慢的雜訊中，可以較為準確的估計雜訊強度，使得語音增強的效果得以提升；雖然該方法可以有效的估測變化緩慢的雜訊取樣值，但是卻無法估測取樣點變化快速的雜訊，例如：白色雜訊(white noise)、F16 雜訊等等，使得語音增強效果變差；文獻[12]根據每個音框的功率頻譜，動態調整頻譜刪減的參數，然後利用多種方法壓抑音樂型噪音，使頻譜刪減效果保持較高的訊雜比，也使音樂型雜訊得到了抑制；文獻[14]利用模擬聲音沿著人耳到大腦傳輸路徑的聽覺感知模型，以此模型中頻譜估計的初期階段來抽取語音特徵參數，並透過隱藏式馬可夫模型套件(HTK)來訓練出連續數字串的語音辨認系統並測試此辨識器的效能；文獻[16]採用最小平方誤差法，他們使用先驗訊雜比的方法來做語音增強，讓語音達到更好品質；文獻[17]提出結合頻譜刪減及人耳遮蔽特性的卡爾曼增強方法，找到感知差異上的語音失真和雜訊去除的最佳均衡點，使得語音失真不嚴重的情況下，消除殘留的音樂型噪音。

為了有效估測變化快速的雜訊信好，文獻[10]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，改善了零值嵌入法中，無法對抗變化快速雜訊的問題，達到提升語音增強的目標，但是將空白音框處估計到的雜訊直接套用到語音部分使用會較為不準確，所以本研究嘗試加入多重音框內插法來估測雜訊頻譜強度，提升零值音框嵌入法的估測準確度，並且在最後加入知覺增益因子[2]-[5]做調適，經多重音框內插法處理後的語音估計準確度較原本的直接帶入準確，之後再使用知覺增益因子將刪減參數作調適，實驗結果證實：本文所提出的方法對於改善零值音框嵌入法有顯著的效果，當干擾程度越大效果越明顯，而知覺增益因子對於改善殘留噪音的效果也相當明顯。

本文的其餘各節安排如下：在第二節介紹零值音框嵌入法，第三節說明多重框內插法，第四節呈現實驗結果與分析，第五節為本文的

結論。

2. 零值音框嵌入與多重音框內插法

2.1 零值音框嵌入法

零值音框嵌入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。

零值音框嵌入法是將原本的零值取樣點嵌入，修改成嵌入零值的音框，藉由零值音框嵌入的位置來估計雜訊的頻譜強度，零值音框嵌入法的示意圖如圖 1，將零值音框嵌入乾淨的語音中，0 到 256 為零值音框；256 到 512 為乾淨的語音訊號，依此類推。嵌入的零值音框視為我們要估計的雜訊頻譜值的位置，用它來估計受干擾訊號的雜訊頻譜強度。

零值音框嵌入法的數學表示式如下式所示：

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

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，訊號在傳送的过程當中，將會受到通道雜訊的干擾，所以接收端收到的信號如下式所示：

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

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，我們必須在受干擾語音信號中估計雜訊；由於我們在發送端已經嵌入零值音框，所以我們可以在零值音框中，估計雜訊頻譜；然後將受干擾語音信號的頻譜刪減頻譜估計值，便可以得到增強語音信號頻譜。

2.2 多重音框內插法

零值音框嵌入法[5]是由語音前的零值音框來估計出雜訊，然後套用到後面四個音框作

為估計到的雜訊，而本文所提出的方法是多重音框內插法，多重音框內插法可以改善零值音框嵌入法估計雜訊頻譜強度的準確度，當雜訊變化量較大時，由於語音變動快速所以只參考一個音框的話無法準確的估計到雜訊，而本文所提出的方法為同時參考前後音框並且做內插法，這樣所得到的估計值會比起參考單一音框的估計值準確。

多重音框嵌入法的實現流程如下：首先在連續兩個相鄰的零值音框嵌入處取得雜訊實際頻譜，然後做內插法取得語音音框部份的噪音估計值，如圖 2 所示，音框 1 與音框 5 表示零值音框嵌入處；將音框 1 與音框 5 的雜訊取得後做內插法得到音框 2、音框 3、音框 4 的雜訊估計值，多重音框內插法的數學式如下式所示：

$$|\hat{D}(l,m)|^2 = 0.25[(5-m)|D(l,1)|^2 + (m-1)|D(l+1,1)|^2] \quad (3)$$

$\hat{D}(l,m)$ 表示嵌入零值音框之後的語音信號， $D(l,m)$ 為實際雜訊信號， m 表示音框序列編號，當 m 為 2 時為第二個音框取前零值音框嵌入處雜訊頻譜之權重為 0.75，後零值音框嵌入處之雜訊頻譜權重為 0.25，便可得到第二個音框，因為距離前音框嵌入處較近所以參考前音框嵌入處的雜訊頻譜較多，距離後音框嵌入處較遠所以參考後音框嵌入處的雜訊頻譜較少，當 m 為 3 時，為第三個音框，雜訊參考值為 0.5 倍權重的前零值音框嵌入處的雜訊頻譜，加上 0.5 倍權重的後零值音框嵌入處的雜訊頻譜，由於距離兩個音框嵌入處的距離一樣所以各取 0.5 倍之權重；當 m 為 4 時為第四個

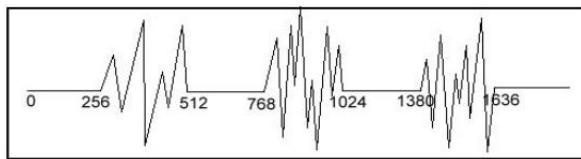


圖 1 零值音框嵌入法

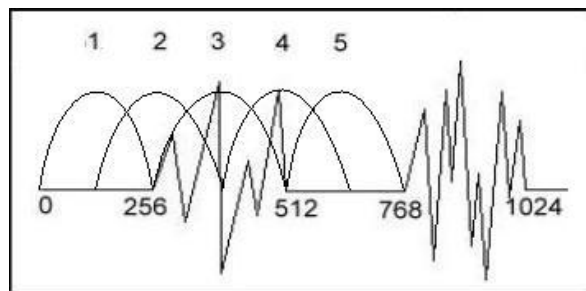


圖 2 多重音框內插法

音框，雜訊參考值為 0.25 倍權重的前零值音框嵌入處的雜訊頻譜，加上 0.75 倍權重的後零值音框嵌入處的雜訊頻譜，由於距離後音框嵌入處較近，所以參考後音框的雜訊頻譜權重為 0.75，而距離前音框嵌入處較遠，故僅參考 0.25 倍的權重。

經過前述步驟後取得雜訊的估計值後，再將取得的噪音估計值帶入公式(4)就可以得到知覺增益因子，知覺增益因子[3]是使用了人耳的聽覺特性的一種語音增強技術，使用人耳聽覺特性來做調適，可以將雜訊部份抑制到人耳聽不到的程度，又不會讓增強後的語音有過度刪減的問題產生。知覺增益因子[3]的公式如下

$$g_p(m,w) = \frac{1}{1 + \max(\sqrt{\frac{|D(m,w)|^2}{T(m,w)}} - 1, 0)} \quad (4)$$

3. 實驗結果

在實驗中，我們使用華語女生語音訊號，加入 NOISX_92 雜訊語料庫中的白色(White)雜訊、F16 戰機座艙(F16-cockpit)雜訊、工廠(factory)雜訊、汽車(car)雜訊、多重人聲(babble)雜訊、直升機座艙(helicopter-cockpit)雜訊，分別在 0dB、5dB、10dB 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紋圖來評估增強後的語音品質。

使用各種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 為零值嵌入法，FZP 為零值音框嵌入法；由表 1 可以看出：在汽車(Car)雜訊部份零值嵌入法的效果優於本文所提出的方法，主要因為零值嵌入法是藉由零值嵌入處的信號來估測語音中的雜訊的振幅，這樣的方法在取樣點變化緩慢的雜訊中，前後兩的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小。而本文所提出的方法在白色(White)雜訊與 F16 戰機座艙(F16-cockpit)雜訊環境下，改善度遠高於零值嵌入法[6]與零值音框嵌入法[5]，工廠(factory)雜訊、多重人聲(babble)雜訊直升機座艙(helicopter-cockpit)雜訊，在 0dB 的情況下也能有效改善。

圖 3 與圖 4 分別為受到白色雜訊與 F16 雜訊干擾的女聲語音，訊雜比均為 0dB，由圖 3 可以很明顯看出零值嵌入法對於處理白色雜

表 1 語音訊號改善度，表中 ZP 為零值嵌入法，FZP 為零值音框嵌入法。

Noise type	SNR (dB)	Enhancement method		
		ZP	FZP	Proposed
white	0	-3.02	5.31	6.81
	5	-3.02	3.65	4.50
	10	-7.10	-3.34	-3.05
F16	0	2.12	5.24	6.60
	5	2.11	3.43	4.14
	10	2.12	0.70	0.96
factory	0	3.97	4.95	6.15
	5	3.96	3.25	3.85
	10	3.97	0.59	0.84
helicopter	0	7.00	6.22	7.24
	5	7.01	4.04	4.60
	10	7.02	1.12	1.38
babble	0	6.21	4.97	6.60
	5	6.20	3.01	3.78
	10	6.21	0.25	0.50
car	0	26.25	6.23	7.55
	5	26.25	4.09	5.13
	10	26.25	1.03	1.70

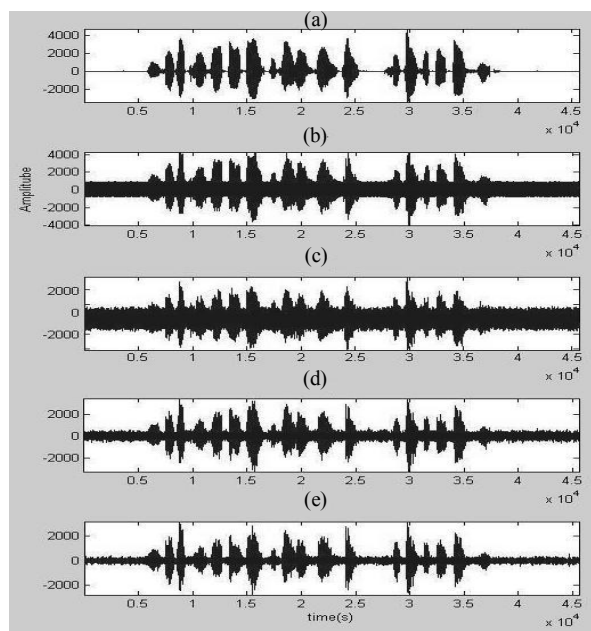


圖 3 語音波形圖(a)乾淨語音 (b)受到白色雜訊干擾的語音(0dB) (c)零值嵌入法增強後的語音 (d) 零值音框嵌入法增強後的語音 (e)零值音框嵌入法與多重音框內插法增強後的語音。

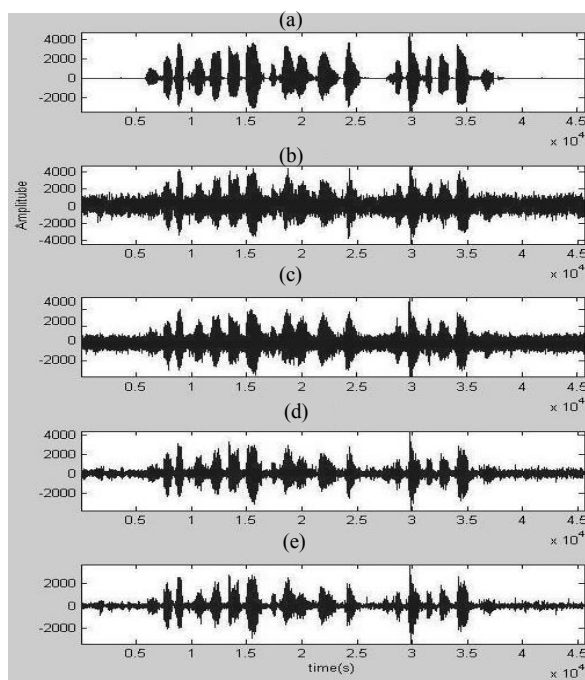


圖 4 語音波形圖(a)乾淨語音 (b)受到 F16 雜訊干擾的語音(0dB) (c)零值嵌入法增強後的語音 (d) 零值音框嵌入法增強後的語音 (e)零值音框嵌入法與多重音框內插法增強後的語音。

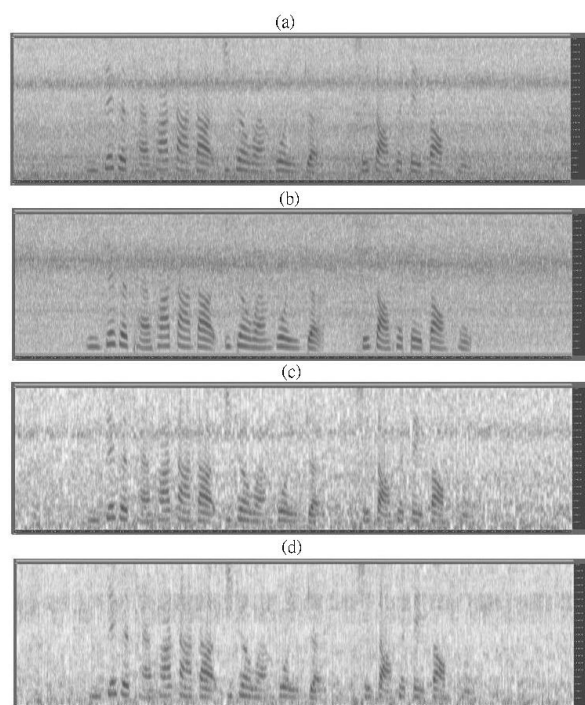


圖 6 語音聲紋圖(a)受到 F16 雜訊干擾的語音 (0dB) (b)零值嵌入法增強後的語音 (c) 零值音框嵌入法增強後的語音 (d)零值音框嵌入法與多重音框內插法增強後的語音。

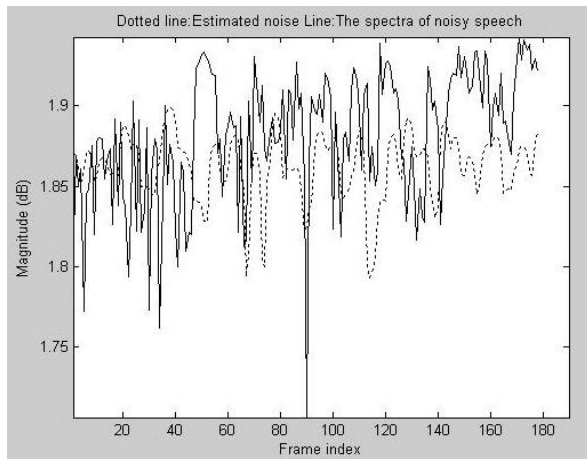


圖 7 本圖為受干擾語音訊號之完整區段;實線部份為受干擾之語音訊號，虛線部份為估計曲線圖。

訊的效果不佳，甚至將雜訊的部份放大，而零值音框嵌入法效果則較零值嵌入法好，能將部份的雜訊抑制，而本文所提出的方法則明顯抑制了白雜訊部份，由圖中可以看出本文方法在於抑制大量快速變化之雜訊部份優於兩者。圖 5 與圖 6 分別為受到白色雜訊與 F16 雜訊干擾的女聲語音，訊雜比為 0dB，聲紋圖背景部份顏色越暗雜訊就越大，由圖中可以明顯看出處理後的效果，零值嵌入法的聲紋圖部分，背景顏色較暗，不僅沒有抑制雜訊反而讓雜訊量增加了，而零值音框嵌入法則能稍微抑制雜訊，而由圖中可以看出，本文所提出的方法之聲紋圖背景部分顏色較淡，抑制掉的雜訊量較前面兩種方法多，證實本文提出的方法確實能有效抑制大量快速變化的雜訊。

圖 7 與圖 8 為噪音估計之曲線圖可以看出，原本的零值音框嵌入法是使用第一個音框的估計值當成後面四個音框的估計值，所以估計曲線不能跟上快速變化的雜訊，而加入了多重音框內插法後的估計曲線則較為平滑且對於雜訊的估計準確度也較高，且也較為接近原始噪音。

4. 結論

本研究主要貢獻為改善零值音框嵌入法估測雜訊的方法，藉由在零值音框嵌入處估測雜訊頻譜，然後加入多重音框內插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，並且在最後加入知覺增益因子，實現雜訊刪減的目的；由實驗結果證實：本文所提出的方法相較於零值嵌入法與零值音框

嵌入法，在處理大量快速變化的雜訊部份明顯的較為有效，加入了多重音框內插法後，對於雜訊的估計準確度也更加準確，當干擾程度越大效果越明顯。

致謝

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。

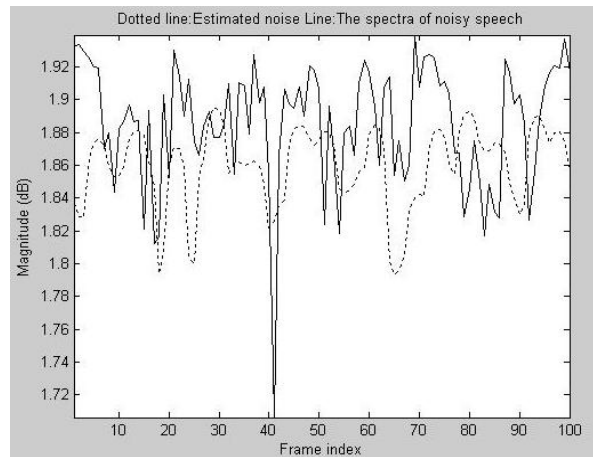


圖 8 本圖為受干擾語音訊號之部份區段;實線部份為受干擾之語音訊號，虛線部份為估計曲線圖。

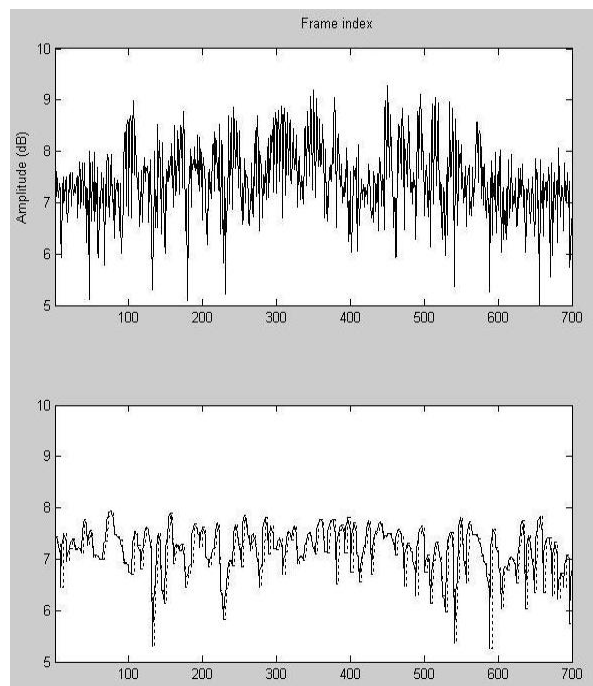


圖 9 本圖為實際受干擾語音訊號(虛線部份)與估計噪音(實線部份)之曲線圖。

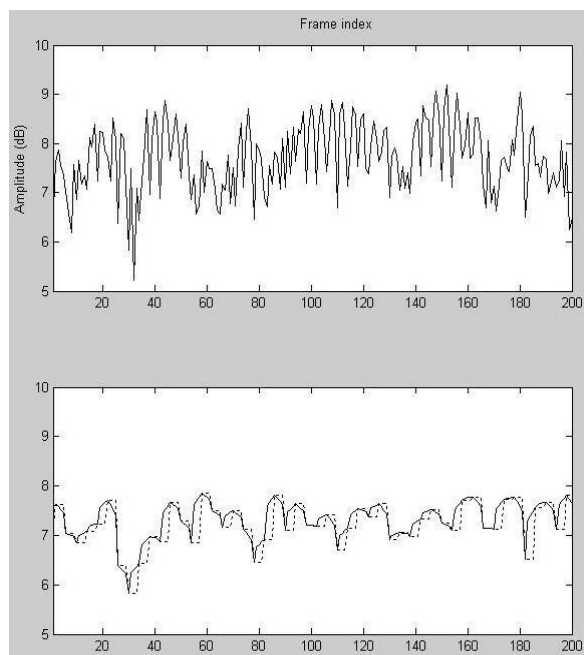


圖 10 本圖為部份區段之實際受干擾語音訊號(虛線部份)與估計噪音(實線部份)之曲線圖。

參考文獻

- [1] S.B. Jebara, "A perceptual approach to reduce musical noise phenomenon with Wiener denoising technique," *in Proc. ICASSP*, 2006, pp. 49-52.
- [2] 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.
- [3] C. T. Lu, K. F. Tseng, C. T. Chen, and J.P. Wang, "Reduction of residual noise for speech enhancement using iterative perceptual gain factor," *in Proc.NST*, pp. 350-354, 2009.
- [4] C. T. Lu, K. F. Tseng, and C. T. Chen, "Speech enhancement using perceptual-decision-directed approach," *in Proc. IEEE ICCEA*, pp. 23-27, 2010.
- [5] 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.
- [6] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *in Proc. IEEE*, vol. 6, pp. 1586-1604, 1979.
- [7] L. Singh, S. Sridharan, "Speech enhancement using pre-processing," *in Proc. SITCT*, pp. 755-758, 1997.
- [8] N. Virag, "Speech enhancement based on masking properties of the auditory system," *in Proc. IEEE ICASSP*, pp.796-799, May. 1995.
- [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] 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*, 頁 294-297, 2010/5。
- [11] 王俊評、陸清達,『使用不同長度零值音框嵌入法之語音增強處理』, *資訊傳播學術研討會*, 頁 229-234, 2010/6。
- [12] 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, 頁 201-204, 2007/4。
- [13] 林典蔚、王小川,「語音訊號中的雜訊預估與刪減方法研究」, *國立清華大學電機工程學系碩士論文*, 新竹, 2006。
- [14] 洪詠能、冀泰石,「基於聽覺感知模型的語音增強技術」, *國立交通大學電信工程學系碩士論文*, 新竹, 2007。
- [15] 黃柏凱、王小川,「可以聽覺遮蔽門檻為基礎的語音增強研究」, *國立清華大學電機工程學系碩士論文*, 新竹, 2004。
- [16] 陳俊、孫洪、董航,「基於 MMSE 先驗信雜比估計的語音增強」, *武漢大學學報*, 卷 51, 期 5, 頁 638-642, 2005/10。
- [17] 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, 頁 363-365, 2008/12。