國 立 交 通 大 學

電控工程研究所

碩 士 論 文

使用目標與干擾比之語音活動偵測
建構雜訊能量估測器

Noise Power Estimator Using Target to Jammer Ratio based Voice Activity Detection

研 究 生： 趙 學 文

指導教授： 胡 竹 生 博士

中 華 民 國 一百 年 九 月

使用目標與干擾比之語音活動偵測
建構雜訊能量估測器

Noise Power Estimator Using Target to Jammer Ratio based Voice Activity Detection

研 究 生：趙學文　　Student：Hsueh-Wen Chao

指導教授：胡竹生　博士　　Advisor：Prof. Jwu-Sheng Hu

國 立 交 通 大 學

電 控 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Electrical and Control Engineering
College of Electrical and Computer Engineering
National Chiao-Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master
In

Electrical and Control Engineering

September 2011

Hsinchu, Taiwan, Republic of China

中 華 民 國 一百 年 九 月

使用目標與干擾比之語音活動偵測建構雜訊能量估測器

研究生：趙　學　文　　指導教授：胡　竹　生　博士

國立交通大學

電控工程研究所碩士班

摘　要

本論文提出一套可適用於非穩態雜訊環境的雜訊能量估測器，並使用此估測結果進行語音純化。雜訊能量估測器使用目標阻斷器和方向性補償器來估測雜訊能量，並且為了調整適應性的方向補償器，採用目標與干擾比進行語音活動偵測，之後根據語音活動偵測結果結合頻譜遮罩進行語音純化。為了確認此演算法的效能，在各種不同的雜訊環境下進行實驗，並透過訊雜比與知覺語音評價兩項客觀標準進行評量，由實驗證明，此法在非穩態雜訊的環境中，可以有效的壓抑干擾雜訊，目標聲源也能同時保有不錯的語音品質。

Noise Power Estimator Using Target to Jammer Ratio based Voice Activity Detection

Student： Hsueh-Wen Chao　　Advisor： Prof. Jwu-Sheng Hu

Institute of Electrical and Control Engineering
National Chiao-Tung University

ABSTRACT

This thesis proposes a noise power estimator for nonstationary noisy environment, and uses the result for speech purification. The noise power estimator combines target blocker and directivity compensator to estimate noise power. For updating the adaptive directivity compensator, we use target to jammer ratio (TJR) based voice activity detection (VAD). The VAD result is also combined with spectrum mask for speech purification. To verify the effectiveness of the proposed algorithm, we experimented in different noise environments, and verified it by signal-to-noise ratio (SNR) and perceptual evaluation of speech quality (PESQ). The experimental results show that the algorithm is able to reduce interference and also save the speech quality.

誌　謝

本論文的完成，首先最感謝的是我的指導教授胡竹生老師，感謝老師願意收我為 X-Lab 的一份子，感謝老師教導我做研究所該有的積極態度和正確的觀念，而當我在研究的路上遇到困難與迷惘的時候，感謝老師總是能給我正確的指引，由衷感謝老師兩年來的悉心指導。

接著要感謝實驗室的各位學長姐、同學與學弟們，特別感謝明唐學長總是能為我解決各種難題，讓我在研究上可以順利前進，也感謝 Simon 學長帶我進入聲音的領域，是你啟發我對於研究的興趣；還有感謝一起在研究之路上努力的同學們，做事有條有理的育成、各方面學識能力都很紮實的建安、擅長鋼琴的昀軒、幽默風趣的偉庭、擅長寫程式的新文、喜歡點心的湘筑、做事認真的耕維、很 MAN 的昭男，感謝你們的陪伴讓我在研究路上不孤單，也感謝各位學弟們的幫忙，助你們研究生涯順利。

更要感謝我的父母，願意包容我各式各樣的任性，一直在我的背後支持我、給我鼓勵，沒有你們就不會有現在的我，對於多年來的養育之恩，在此獻上最誠摯的謝意。

十幾年的求學生涯在此暫時告一段落了，在此希望能將自己所學好好的回饋社會，為了更進步的世界貢獻自己的一份力量。

目　　錄

摘　要 ..... **i**

ABSTRACT ..... **ii**

誌　謝 ..... **iii**

目　錄 ..... **iv**

表　列 ..... **v**

圖　列 ..... **vi**

第一章　緒論 **1**

1.1　研究動機 1
1.2　文獻回顧 1
1.4　論文架構 2

第二章　雜訊能量估測器 **4**

2.1　固定式方向補償器 5
2.2　適應性方向補償器 7
2.3　多麥克風時的參數調整 8
2.3.1　Wiener filter 求參數 8
2.3.2　NLMS 求參數 11

第三章　語音活動偵測 **13**

3.1　Target to Jammer Ratio 13

第四章　實驗結果與分析 **21**

4.1　實驗環境 21
4.2　不同雜訊環境下之實驗結果 25
4.3　實驗結果與分析 35

第五章　研究成果與未來展望 **37**

參考文獻 **38**

表　列

表 4-1：測試用模擬環境 22
表 4-2：不同雜訊環境下 SNR (Wiener filter).....26
表 4-3：不同雜訊環境下 PESQ (Wiener filter).....27
表 4-4：不同雜訊環境下 SNR (NLMS).....31
表 4-5：不同雜訊環境下 PESQ (NLMS).....32

圖　列

圖 2-1：雜訊能量估測器架構圖 4
圖 2-2：環境示意圖 4
圖 2-3：到達時間差示意圖 5
圖 2-4：Target Blocker 之方向增益 6
圖 3-1：Generalized Sidelobe Canceller 架構示意圖 13
圖 3-2：Adaptive Noise Canceller 輸出理論增益 15
圖 3-3：TJR 之方向增益圖形 16
圖 3-4：TJR 之方向增益圖形(寬間距) 17
圖 3-5：各頻率門檻值 18
圖 3-6，VAD 判斷圖 18
圖 3-7，VAD 判斷圖(8 顆麥克風) 19
圖 4-1：各聲源方位示意圖 21
圖 4-2：目標聲源 Man 22
圖 4-3：雜訊 Babble 23
圖 4-4：雜訊 F16 23
圖 4-5：干擾聲源 Woman 1 24
圖 4-6：干擾聲源 Woman 2 24
圖 4-7：Case 1 處理結果(Wiener filter) 28
圖 4-8：Case 2 處理結果(Wiener filter) 28
圖 4-9：Case 3 處理結果(Wiener filter) 29
圖 4-10：Case 4 處理結果(Wiener filter) 29
圖 4-11：Case 1 處理結果(NLMS) 33
圖 4-12：Case 2 處理結果(NLMS) 33
圖 4-13：Case 3 處理結果(NLMS) 34
圖 4-14：Case 4 處理結果(NLMS) 34

第一章 緒論

1.1 研究動機

我們身處的周遭環境總是充滿了許多聲音方面的干擾，包括各種家電用品、交通工具、人們講話的聲音，這些干擾在我們收集聲音訊號的時候都會造成妨礙。為了避免這些干擾雜訊所造成的影響，希望能設計一套方法，可以估計周遭環境的干擾雜訊或是得到干擾雜訊的特性，在收集聲音訊號時，便可以有效將干擾雜訊所造成的影響去除，以達到語音純化的效果。這樣的做法有兩個主要研究目標，其一是最大程度上的壓抑干擾雜訊，其二是消減干擾雜訊的同時又能保持目標聲源的不失真。

干擾雜訊大致可以分為兩類：穩態雜訊(stationary noise)或非穩態雜訊(non-stationary noise)，在穩態雜訊的壓抑部分，Wiener 濾波器及其各種延伸變型可以有相當好的表現，而當干擾雜訊屬於非穩態雜訊的情況，如何在盡量不損害目標聲源的狀況下消除其他干擾雜訊，是一個具挑戰性的研究課題。

麥克風陣列比起單聲道的麥克風，多了空間上的資訊，可以利用聲音的來源方位不同建立空間上的濾波器，將干擾雜訊與目標聲源區分開來，使估計干擾雜訊的效果更好。

1.2 文獻回顧

語音純化的技術已經發展了很長一段時間，在單顆麥克風裝置的強況，常見的做法是對雜訊估測，例如估測雜訊的能量頻譜密度(Power Spectral density, PSD)以建立增益函數的 SS(Spectral Subtraction)[9]，或是估計雜訊的統計特性，以建立雜訊 Gaussian 模型，而使用麥克風陣列主要的概念為建立空間濾波器(Spatial Filter)，接收目標聲源而來的訊號，並壓抑其他干擾聲源，一般被稱為波束形成

器(Beamformer)，最早的 Beamformer 是 DS (Delay-and-Sum) Beamformer，依照目標聲源的角度計算到達各麥克風的時間差，將延遲補償回去之後相加起來即為 DS Beamformer，此方法在計算上較簡單，但是如果想要形成更為尖銳的波束，進而壓抑其他角度干擾聲源的影響，就需要更大的麥克風陣列才能達到較好的效果，因此之後有人提出了 MVDR (Minimum Variance Distortionless Response) Beamformer [1]，相較 DS Beamformer，MVDR 的波束更為尖銳，也對其他角度的干擾聲音進行壓抑。

1969 年 Griffiths 提出了使用 MMSE (Minimum Mean Square Error)的適應性波束形成器(Adaptive Beamformer)[2]，適應性波束形成器具有可調整的權重，對於各種環境更具穩健性，Griffiths 和 Jim 在之後提出知名的 GSC (Generalized Sidelobe Canceller)演算法[3]，GSC 對於各種雜訊的壓抑有很好的表現，但是需要兩顆以上的麥克風，以及較大的麥克風間距，之後 Aarabi 提出了使用相位差資訊來進行濾波的 PBF(Phase-error Based Filter)[4]，在雙麥克風裝置上有很好的效果，Kim 和 Jeong 等人利用相位差資訊與方向性的補償器來進行雜訊能量的估計[5]，在很小間距的雙麥克風裝置，如手機上也能有很好的效果。

另外也有使用預錄資料或資訊進行語音純化的方式，Gannot *et al.*提出使用相對轉移函數(RTF, relative transfer function)的方式進行 GSC[8]，此方法為事先利用乾淨聲源訓練出整體環境相對轉移函數，這樣的方式得到的空間濾波器會比單純利用角度延遲資訊計算得到的更精確。Dahl 提出使用預錄參考訊號的 RSAB (Reference Signal based Adaptive Beamformer)[6]，預先錄好乾淨的聲源，將其和雜訊混合後，把乾淨聲源當參考訊號進行 LMS (Least Mean Square)調整 Adaptive Beamformer 的參數，在各種環境也有很好的語音純化效果。

1.3 論文架構

本論文為利用目標聲源阻斷器(Target Blocker)與方向性補償器，結合 TJR

(Target to Jammer Ratio)資訊進行估測雜訊，並使用頻譜遮罩(Mask)將雜訊從輸入音源中減去以達到語音純化的目的，之後再於各種模擬環境下進行實驗，且以 SNR 和 PESQ 兩項標準進行客觀的效能評比，各章主要內容如下：

第二章：雜訊能量估測器架構介紹。

第三章：以 TJR 資訊進行 VAD，並建立頻譜遮罩。

第四章：實驗結果與分析。

第五章：研究成果與未來展望。

第二章 雜訊能量估測器(Noise Power estimator)

本雜訊估測能量器的架構來自於 Kim 和 Jeong 等人的雜訊能量估測器[5]，並對其做一些修改，分為三個部分，目標聲源阻斷器(Target Blocker)、固定式方向補償器(Fixed Directivity Compensator)和適應性方向補償器(Adaptive Directivity Compensator)，如圖 2-1 所示：

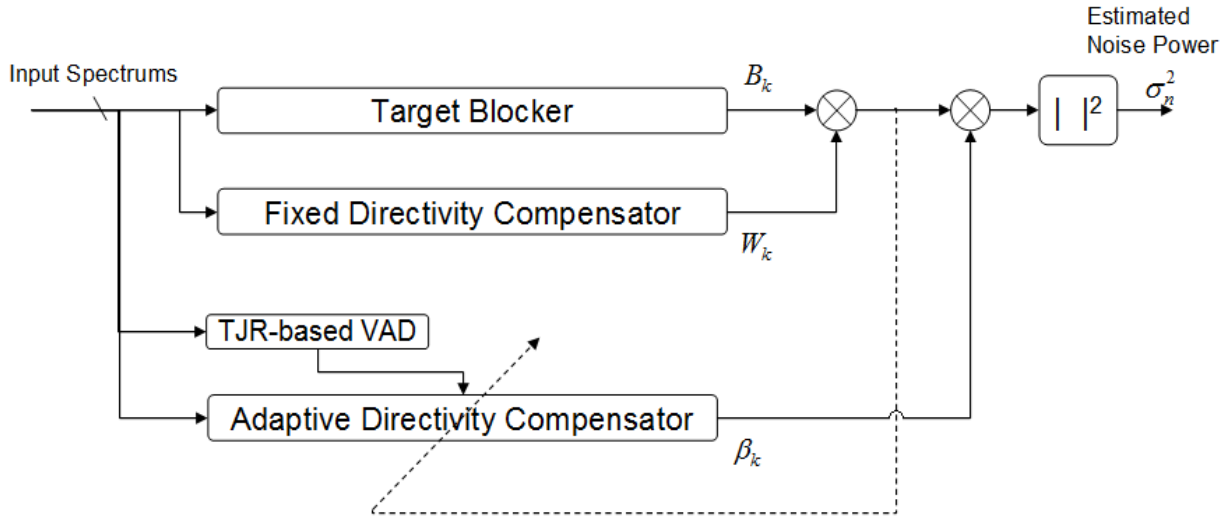


圖 2-1：雜訊能量估測器架構圖

假設目標聲源位於麥克風陣列的中間正前方，如下圖表示：

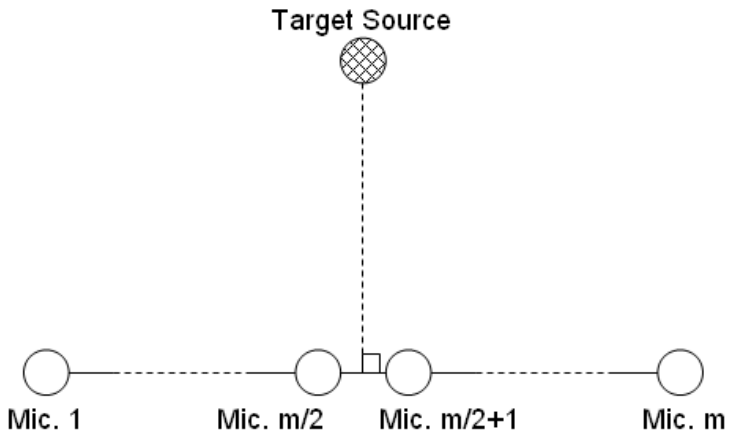


圖 2-2：環境示意圖

在雙麥克風的場合，將 Target Blocker 定義為將兩麥克風收的頻譜訊號相減，經過 Target Blocker 後的訊號如下：

$$B(t,k) = X^{(1)}(t,k) - X^{(2)}(t,k) \qquad (2\text{-}1)$$

其中 $X^{(i)}(k)$ 表示在第 t 個框架(frame)，經過 STFT (Short Time Fourier Transform) 之後，第 i 顆麥克風在第 k 個頻帶所接收到的頻譜訊號。此 Target Blocker 為對正中間方向的零波束形成器(Nullformer)，可以將目標聲源方向作相當大程度的壓抑。

2.1 固定式方向補償器(Fixed Directivity Compensator)[5]

假設一雙麥克風裝置，在遠場假設下，聲音的波形近似於平面波，如果聲音來源方向並非正前方，此音波到達兩麥克風的所耗時間會不同，我們可由聲音的來源方向計算出此音波到達兩麥克風的時間差（IDT, Interaural Time Difference）。

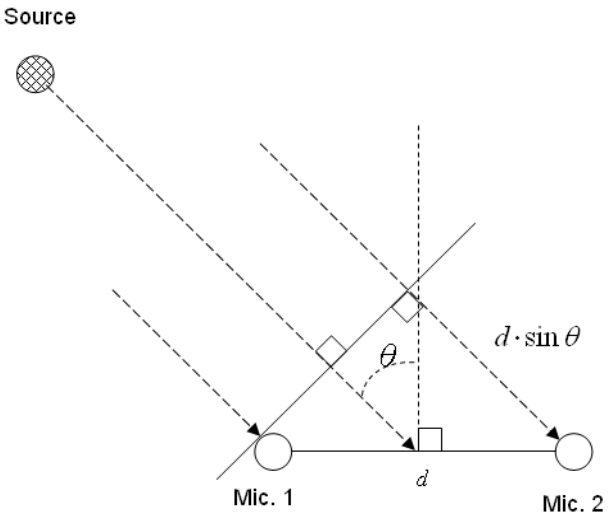


圖 2-3：到達時間差示意圖

兩顆麥克風收到 θ 方向來的訊號可分別表示如下：

$$x^{(1)}[n] = x_{\theta}[n],$$

$$x^{(2)}[n] = x_{\theta}[n - \frac{d \cdot f_S}{C} sin(\theta)] \qquad (2\text{-}2)$$

其中 C 代表聲速，d 代表兩顆麥克風間距，f_s 為取樣頻率，經過 Fourier Transform 之後，兩顆麥克風收到 θ 方向來的頻譜訊號可分別表示為：

$$X^{(1)}(t,k) = X_{\theta}(t,k),$$
$$X^{(2)}(t,k) = e^{-j\frac{2\pi}{C}\frac{k\cdot f_S}{N_{FFT}}d\cdot\sin(\theta)} X_{\theta}(t,k) \qquad (2\text{-}3)$$

其中 N_{FFT} 為作 FFT(Fast Fourier Transform)時所使用的點數，在時域上的延遲經過 Fourier Transform 之後成為頻域上相位差，利用 2-3 式計算前述 Target Blocker 在各頻帶所造成的各方向增益函數如下：

$$\mathrm{D_{TB}}(k,\theta) = \left|\frac{X^{(1)}(t,k) - X^{(2)}(t,k)}{X_{\theta}(t,k)}\right| = \left|1 - e^{-j\frac{2\pi}{C}\frac{k\cdot f_s}{N_{FFT}}d\cdot\sin(\theta)}\right| \qquad (2\text{-}4)$$

將此增益函數繪出的圖形即為下圖：

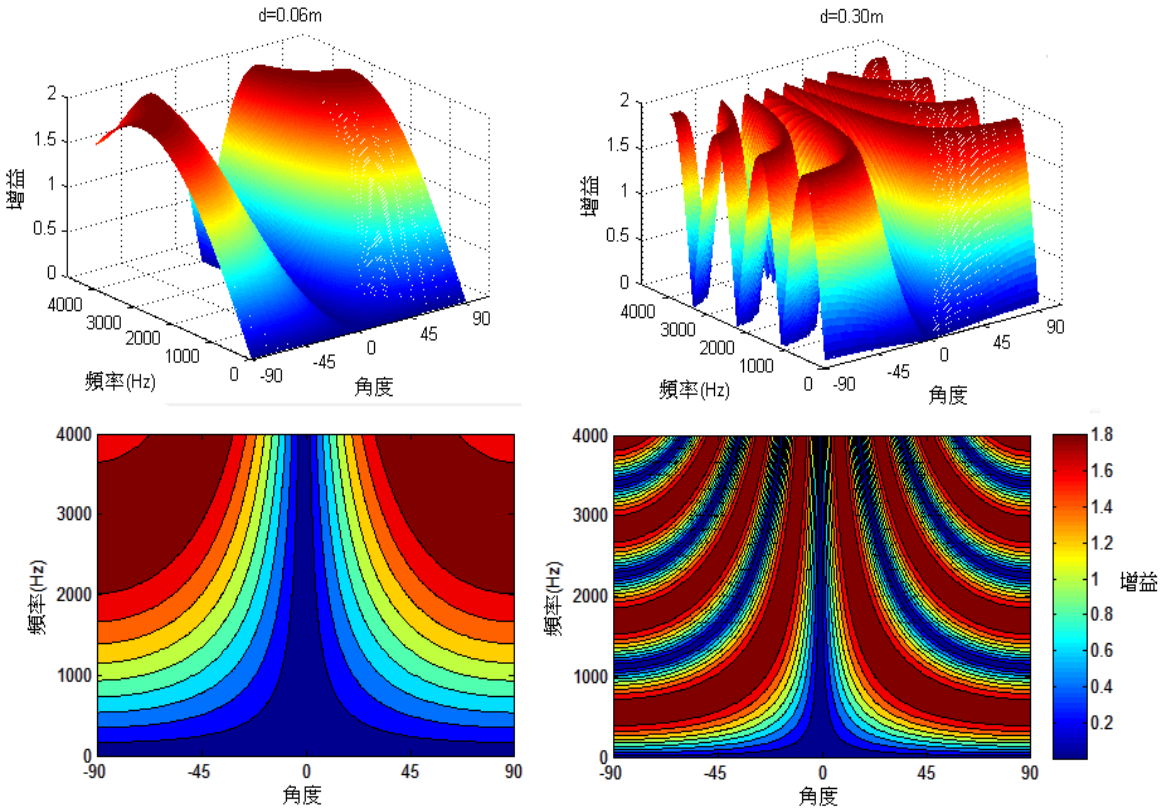


圖 2-4：Target Blocker 之方向增益，麥克風間距為 0.06 公尺(左)與 0.3 公尺(右)，上圖為立體圖，下圖為等高線圖。

圖 2-4 所示，Target Blocker 不但把 0° 方向的目標聲源隔絕掉，在各頻率上

也造成不同的增益，尤其在低頻的部分被壓抑較多，高頻的部分，也會隨麥克風間距不同，有著空間上的混疊(Aliasing)失真，這些部分都會造成噪音估測上的失真，而方向補償器則是將此增益在各角度取平均後除回去：

$$W(k)=\left(\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}|D(k,\theta)|d\theta+\delta_w\right)^{-1} \tag{2-5}$$

δ_w 為一微小的數，目的是避免因為增益 D_{TB} 過小，造成 W 趨近於無窮大。將經過 Target Blocker 的訊號 $B(t,k)$ 乘上此固定補償增益，即可補償在各頻帶所造成的失真。

2.2 適應性方向補償器(Adaptive Directivity Compensator)

再上一章節固定式方向補償器的部分，我們利用ITD計算Target blocker在各頻帶所造成之失真，並將其補償回去，其他還有許多難以計算的造成失真原因，例如錄音環境所造成的影響，錄音設備的誤差之類，則在本章節利用適應性的方式進行補償。

適應性方向補償器為利用語音活動偵測(VAD，Voice activity detection)的做法，當VAD判斷此時沒有目標聲源，理論上此時最好的雜訊估測即為目前收到的語音訊號，因此建立可調整的適應性參數，讓雜訊估測的結果可以逼近接收到的雜訊訊號，此參數只在沒有目標聲源時進行更新，表示如下：

$$\beta(t,k)=\begin{cases}\alpha\dfrac{R(t,k)}{|B(t,k)\cdot W(t,k)|}+(1-\alpha)\,\beta(t-1,k), & \text{if VAD}(t,k)=0\\ \beta(t-1,k), & \text{if VAD}(t,k)=1\end{cases} \tag{2-6}$$

其中 α 為遺忘因子(Forgetting Factor)，控制 β 值的更新速度， R為兩麥克風收到頻譜訊號振福的平均如下：

$$R(t,k)=\frac{|X^{(1)}(t,k)|+|X^{(2)}(t,k)|}{2} \tag{2-7}$$

而VAD的部分使用的是TJR資訊，將於第三章介紹。

最後得到的雜訊能量估測結果如下：

$$\sigma_N^2(t,k) = \left|B(t,k)\cdot W(k)\cdot \beta(t,k)\right|^2 \qquad (2\text{-}8)$$

將原始輸入訊號的振幅減去估測的雜訊能量開根號，即為估測的目標聲源頻譜訊號振幅，再將原始輸入訊號的相位角合上可得語音純化的輸出 $\hat{S}$，以下面式子表示：

$$\hat{S}(t,k) = \left|\frac{\left|X^{(1)}(t,k)\right| + \left|X^{(2)}(t,k)\right|}{2} - \sqrt{\sigma_N^2(t,k)}\right| \exp(\mathrm{j}\frac{\angle \mathrm{X}^{(1)}(t,k) + \angle \mathrm{X(t,k)}^{(2)}}{2}) \qquad (2\text{-}9)$$

2.3 多麥克風時的參數調整

在使用多麥克風裝置的場合，可以有更多的參數來調整適應式方向補償器，理論上不同間距的麥克風對在不同頻帶都會有效果好壞之差，間距較小的麥克風對會在高頻部分有較好的表現，反之亦然，為了得到更好的雜訊估測結果， 不同間距的麥克風對分別估計雜訊，並調整各結果在各頻帶所佔的比例，將各麥克風對的估測結果，整合成一最接近實際雜訊的估測結果，本文利用兩種方法來求整合參數，並將於後面章節透過實驗比較其優缺點。

2.3.1 Wiener filter求參數

此為使用反矩陣的解法，可以得到在所選取的範圍內最小均方差(MMSE)的權重，是在理論上最佳的解法。

假設M顆麥克風的裝置，可以得到M-1個不同間距的雙麥克風，將各麥克風對所接收的輸入訊號，分別通過Target Blocker之後，原本2-1式改寫成為以下向量形式：

$$\boldsymbol{b(t,k)} = \left[B^{(1)}(t,k) \quad B^{(2)}(t,k) \quad B^{(3)}(t,k) \quad \cdots \quad B^{(M-1)}(t,k)\right]^{\mathrm{T}} \qquad (2\text{-}10)$$

其中麥克風對的挑選原則，為了避免角度偏差，盡量挑選左右對稱的麥克風為一

對：

$$B^{(m)}(t,k)=\begin{cases}X^{(m/2+1)}(t,k)-X^{(M-m/2+1)}(t,k) & ,\text{m is even}\\ X^{(m/2+0.5)}(t,k)-X^{(M-m/2+0.5)}(t,k), & \text{m is odd}\end{cases} \tag{2-11}$$

固定式方向補償器的計算如2.1節所描述，因為麥克風間距不同所以W值也不同，表示如下：

$$\boldsymbol{w(t,k)}=\left[W^{(1)}(t,k)\quad W^{(2)}(t,k)\quad W^{(3)}(t,k)\quad \cdots\quad W^{(M-1)}(t,k)\right]^{\mathrm{T}} \tag{2-12}$$

將Target Blocker和固定式方向補償器對應項相乘，此為各麥克風對尚未進行適應性調整的估計雜訊：

$$\boldsymbol{v(t,k)}=\left[W^{(1)}(t,k)\cdot B^{(1)}(t,k)\quad W^{(2)}(t,k)\cdot B^{(2)}(t,k)\quad \cdots\quad W^{(M-1)}(t,k)\cdot B^{(M-1)}(t,k)\right]^{\mathrm{T}} \tag{2-13}$$

也將權重$\boldsymbol{\beta}$表示為向量形式：

$$\boldsymbol{\beta(t,k)}=\left[\beta^{(1)}(t,k)\quad \beta^{(2)}(t,k)\quad \cdots\quad \beta^{(M-1)}(t,k)\right]^{\mathrm{T}} \tag{2-14}$$

讓輸出的雜訊能量估測結果如下：

$$\sigma_{\mathrm{N}}^{2}(t,k)=|\,\boldsymbol{v(t,k)^{T}\,\beta(t,k)}|^{2} \tag{2-15}$$

為了解出M-1個β值，需M-1個方程式，故取前M-1個音框的向量$\boldsymbol{v}$，組成$(\mathrm{M}-1)\times(\mathrm{M}-1)$的方陣V：

$$\boldsymbol{V(t,k)}=\left[\boldsymbol{v^{T}(t,k)}\quad \boldsymbol{v^{T}(t-1,k)}\quad \cdots\quad \boldsymbol{v^{T}(t-(M-2),k)}\right]^{\mathrm{T}} \tag{2-16}$$

而$\boldsymbol{r}$為各音框之R所組成的向量表示如下：

$$\boldsymbol{r(t,k)}=\left[R(t,k)\quad R(t-1,k)\quad \cdots\quad R(t-(M-2),k)\right]^{\mathrm{T}} \tag{2-17}$$

在VAD=0時，輸出的最佳估測雜訊能量即為目前收到訊號的能量：

$$\boldsymbol{V(t,k)\,\beta_{best}(t,k)=r(t,k)} \tag{2-18}$$

故$\boldsymbol{\beta}$之更新方式改寫為下式：

$$\boldsymbol{\beta(t,k)}=\begin{cases}\alpha\boldsymbol{V^{-1}(t,k)r(t,k)}+(1-\alpha)\,\boldsymbol{\beta(t-1,k)}, & \text{if VAD(t,k)}=0\\ \boldsymbol{\beta(t-1,k)}, & \text{if VAD(t,k)}=1\end{cases} \tag{2-19}$$

此時的α(Forgetting Factor)可以取較大的值或是直接設為 1，因為取 M-1 個音框作反矩陣計算的動作，就包含取這 M-1 個音框之平均的意義。

另外，使用反函數的動作必需考慮穩定性的問題，如果方陣 V 是不可逆的或是很接近不可逆的，rank 數不足 M-1 會造成無法正確的解出各β值，因為聲音為連續性的訊號，這種狀況很容易發生，為了解決這個問題，取向量$\boldsymbol{v}$時需多取幾組以增加穩定性，而由於 V 不再是方陣，所以解$\boldsymbol{\beta}$時改為使用虛擬反矩陣(pseudo inverse)，將以上求解式子改寫如下：

$$\boldsymbol{\beta(t,k)} = \begin{cases} \boldsymbol{(V^H(t,k)V(t,k))^{-1}V^H(t,k)r(t,k)}, & \text{if VAD(t,k)} = 0 \\ \boldsymbol{\beta(t-1,k)}, & \text{if VAD(t,k)} = 1 \end{cases} \quad (2-20)$$

其中

$$\boldsymbol{V(t,k)} = \left[\boldsymbol{v^T(t,k)} \quad \boldsymbol{v^T(t-1,k)} \quad \cdots \quad \boldsymbol{v^T(t-(M-2)+d,k)}\right]^{\mathrm{T}} \quad (2-21)$$

$$\boldsymbol{r(t,k)} = \left[R(t,k) \quad R(t-1,k) \quad \cdots \quad R(t-(M-2)+d,k)\right]^{\mathrm{T}} \quad (2-22)$$

其中 d 即為多取的組數，可以根據目前雜訊的性質來做設定，當目前雜訊屬於較非穩態雜訊時，例如人聲，鄰近音框之間相似度較低，反矩陣運算時穩定性較高，d 就可以取小一點，反之亦然。最後的雜訊能量估測輸出為下：

$$\sigma_N^2(t,k) = (\boldsymbol{v^T(t,k)\,\beta(t,k)})^2 \quad (2-23)$$

由 2-9 式同樣方法可得語音純化的輸出$\hat{S}$。

透過此 Wiener filter 的解法可以得到理論上最佳的值，但是此法存在兩個問題造成估測成果降低，其一是因為使用了反矩陣運算，需要消耗較大的計算時間，其二是為了保證反矩陣運算穩定且有最佳解，要確保有足夠 M-1 個自由度，因此需使用較長時間的資料一起計算，而且因為只在 VAD 判斷無目標聲源時更新參數，所取各音框的相隔時間也被不確定性的拉長，在每一筆資料重要性(權重)都相同的情況下，雖然得到的解為統計意義上長時間的最佳解，但是對於突發性的變動就不能追蹤得好，所以 Wiener filter 在穩態雜訊時有很好的表現，在非穩

態雜訊時就比較沒有辦法估計的完整。

2.3.2 NLMS求參數

另外一種解法為使用正規化最小均方法(Normalized Least Mean Square, NLMS)，在 VAD 判斷目前只有雜訊的時候，進行雜訊追蹤， NLMS 演算法各步驟描述如下：

NLMS 的輸出為估測的雜訊 $\hat{N}$ ：

$$\hat{N}(t,k) = \boldsymbol{\beta}^{H}(t,k)\boldsymbol{v}(t,k) \tag{2-24}$$

其中 $\boldsymbol{\beta}$ 和 $\boldsymbol{v}$ 定義和2.3.1相同：

$$\boldsymbol{\beta}(t,k) = \left[\beta^{(1)}(t,k) \quad \beta^{(2)}(t,k) \quad \cdots \quad \beta^{(M-1)}(t,k)\right]^{T} \tag{2-25}$$

$$\boldsymbol{v}(t,k) = \left[W^{(1)}(t,k)\cdot B^{(1)}(t,k) \quad W^{(2)}(t,k)\cdot B^{(2)}(t,k) \quad \cdots \quad W^{(M-1)}(t,k)\cdot B^{(M-1)}(t,k)\right]^{T} \tag{2-26}$$

誤差函數定為參考訊號與NLMS的輸出之差，參考訊號為當VAD=0時，麥克風接收到的輸入訊號之絕對值：

$$e(t,k) = R(t,k) - \hat{N}(t,k) \tag{2-27}$$

權重更新如下：

$$\boldsymbol{\beta}(t,k) = \begin{cases} \boldsymbol{\beta}(t-1,k) + \mu\dfrac{e(t-1,k)\boldsymbol{v}^{*}(t,k)}{\gamma_{NLMS} + \boldsymbol{v}^{H}(t,k)\boldsymbol{v}(t,k)}, & VAD(t,k) = 0 \\ \boldsymbol{\beta}(t-1,k), & VAD(t,k) = 1 \end{cases} \tag{2-28}$$

其中 μ 決定NLMS的權重更新速度，$0 < \mu < 2$，γ_{NLMS} 為一微小的值，只是為了使分母不為零，限制權重更新的速度不會太快，如此可以確保NLMS收斂。使用NLMS所需要的計算時間會比使用反矩陣來得少很多，調整 μ 可以讓雜訊估測器適合穩態雜訊或是不穩態雜訊之間作權衡。

而雜訊能量估測器的輸出為NLMS的輸出之能量：

$$\sigma_N^2(t,k) = \hat{N}^*(t,k)\,\hat{N}(t,k) \qquad (2\text{-}29)$$

由 2-9 式同樣方法可得語音純化的輸出 $\hat{S}$。實驗結果將在第四章呈現，並進行比較與討論。

第三章 語音活動偵測(VAD, Voice activity detection)

VAD 在語音訊號處理上有很重要的幫助，例如在語音辨識中，將輸入語料經過 VAD 切割之後再進行辨識，可以提升辨識率，在作語音純化處理時，也可以將不屬於語音的雜訊刪除。VAD 有許多不同的做法，在單麥克風裝置上，較常見使用的判斷特徵包含能量、過零率、亂度或長時間語音資訊(long-term speech information)等，而在多麥克風裝置上，則多了空間資訊可供判斷，可以根據估計訊號來源方向(DOA, direction of arrival)來推測該訊號是否由目標聲源所發出。本文所使用特徵為接下來章節所介紹的 TJR。

3.1 Target to Jammer Ratio (TJR)

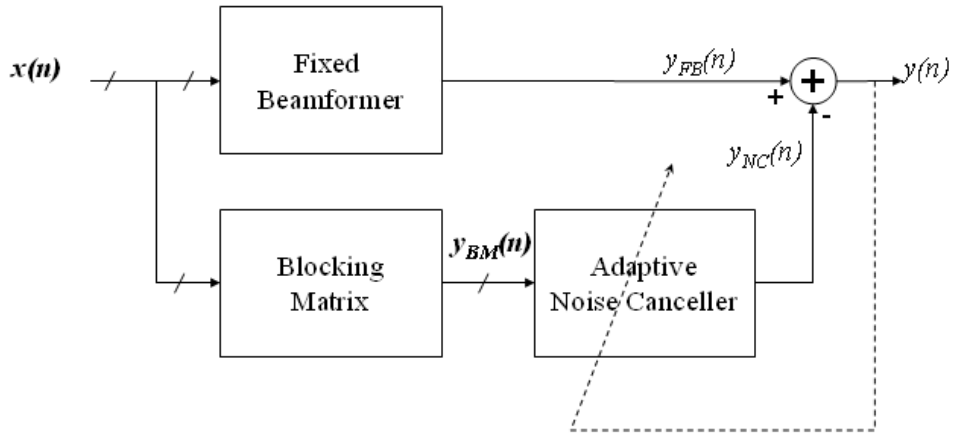


圖 3-1：Generalized Sidelobe Canceller 架構示意圖。

原始 TJR 是定義是於 GSC 的架構下[7]，圖 3-1 為 GSC 的架構簡圖，主要分為三個部分，Fixed Beamformer、Blocking Matrix 和 Adaptive Noise Canceller，Fixed Beamformer 為對目標聲源方向的 Delay-and-Sum Beamformer，輸出 y_{FB} 包含了各種雜訊和增強後的目標聲源，Blocking Matrix 則是將目標聲源去除，只留下雜訊部份 $\mathbf{y}_{\mathbf{BW}}$，Adaptive Noise Canceller 以將輸出 y 的能量最小化為目的調整權重，將 $\mathbf{y}_{\mathbf{BW}}$ 結合成估計的雜訊 y_{NC}，輸出 $y = y_{FB} - y_{NC}$。

而TJR為代表目標聲源的GSC輸出 y，與代表雜訊的Adaptive Noise Canceller輸出 y_{NC} 之能量比值，代表意義類似於能量的訊噪比(Signal-to-noise ratio, SNR)，表示如下式：

$$P_{t,o} = \{y^2(n)\}$$
$$P_{j,o} = \{y_{NC}{}^2(n)\} \quad (3\text{-}1)$$

$$TJR_o = 10log_{10} P_t - 10\log_{10} P_j \quad (3\text{-}2)$$

而在本文中所使用的TJR定義跟原始TJR不同，首先為了簡化計算，將Adaptive Noise Canceller的部分省去，將TJR定義為Fixed Beamformer輸出 y_{FB} 與Target Blocker輸出 y_{TB} 之能量比值，並於頻譜上計算，由於 y_{FB} 相較 y 少了消去雜訊的步驟，此時的TJR為有偏差的訊噪比(Biased SNR)，表示如下：

$$P_t(t,k) = \{Y_{FB}{}^2(t,k)\}$$
$$P_j(t,k) = \{Y_{TB}{}^2(t,k)\} \quad (3\text{-}3)$$

$$TJR(t,k) = 10log_{10} P_t(t,k) - 10\log_{10} P_j(t,k) \quad (3\text{-}4)$$

在雙麥克風裝置，且目標聲源在正前方的場合，Target Blocker部分即為第二章中所提到的：

$$Y_{TB}(t,k) = B(t,k) = X^{(1)}(t,k) - X^{(2)}(t,k) \quad (3\text{-}5)$$

Fixed Beamformer則使用對正前方的Delay-and-Sum Beamformer：

$$Y_{FB}(t,k) = X^{(1)}(t,k) + X^{(2)}(t,k) \quad (3\text{-}6)$$

而在本文中以取絕對值的方式計算：

$$TJR(t,k) = 10log_{10} \frac{|Y_{FB}(t,k)|}{|Y_{TB}(t,k)|} = 10log_{10} \frac{|X^{(1)}(t,k) + X^{(2)}(t,k)|}{|X^{(1)}(t,k) - X^{(2)}(t,k)|} \quad (3\text{-}7)$$

如果目標聲源方位並不是在麥克風陣列的正前方，則先估算延遲時間後，將頻譜上延遲造成的相角反乘回去，再計算 Y_{FB} 和 Y_{TB} 。

將 TJR 定義如此簡化最大的好處在於容易計算其理論增益，進而設定正確的門檻值(threshold)，如果訊號經過 Adaptive Noise Canceller 之後，其在頻譜上各角度的增益會變得比較複雜，正負角度的增益會有些不對稱，各頻帶之間也差異很大，如果包含 Adaptive Noise Canceller 一起計算理論增益以設定門檻值，則容易使誤差累積，VAD 判斷錯誤導致參數更新錯誤，進而錯誤設定門檻值，又造成 VAD 判斷錯誤，如果不使用理論增益，則需平均、Hang over 機制等長時間的資訊，需要付出輸出時間延遲的代價。簡化 TJR，也可以將門檻值 Offline 計算，節省許多運算時間。

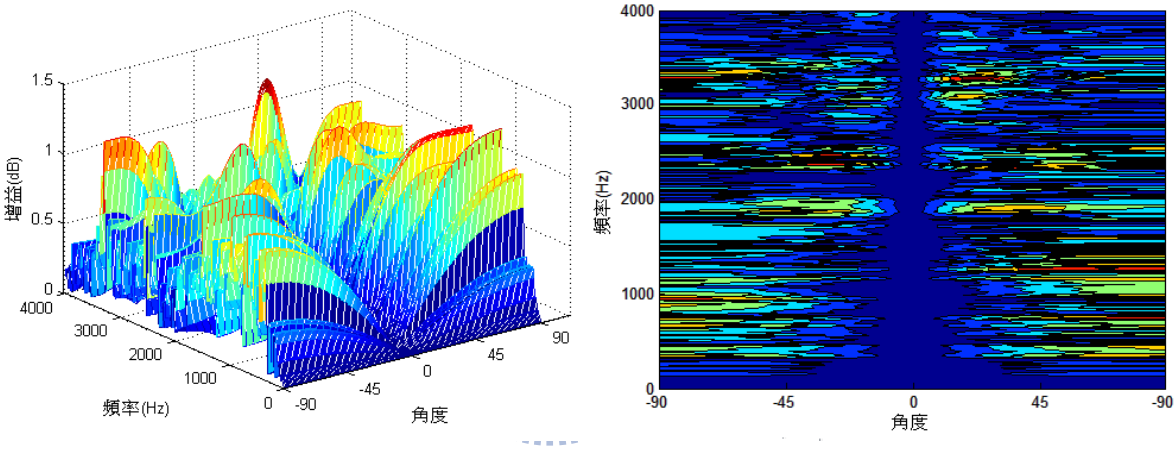


圖 3-2：Adaptive Noise Canceller 輸出理論增益，八顆麥克風 GSC。

由理論上來估計 TJR 在各頻帶與各角度所造成的增益函數：

$$D_{TJR}(k,\theta) = log_{10}\frac{\mathrm{D_{FB}}(k,\theta)}{\mathrm{D_{TB}}(k,\theta)+\gamma} = log_{10}\frac{\left|1+e^{-j\frac{2\pi}{C}\frac{k\cdot f_s}{N_{FFT}}d\cdot sin(\theta)}\right|}{\left|1-e^{-j\frac{2\pi}{C}\frac{k\cdot f_s}{N_{FFT}}d\cdot sin(\theta)}\right|+\gamma} \tag{3-8}$$

γ 為一較小的值，作用只是為了避免分母為零，$\mathrm{D_{FB}}$ 和 $\mathrm{D_{TB}}$ 分別為 Fixed Beamformer 和 Targer Blocker 的增益函數，將上式增益函數畫出可得下圖 3-3：

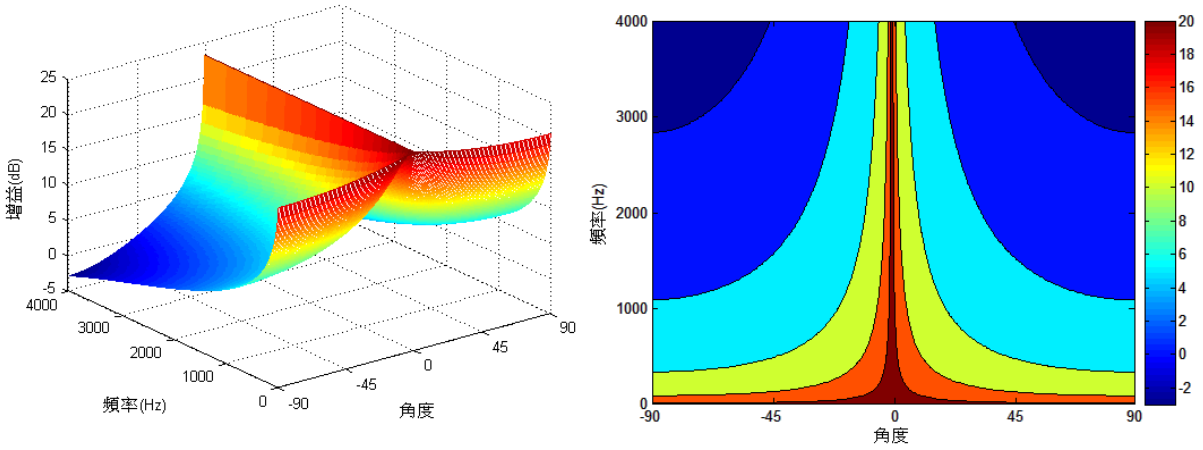


圖 3-3：TJR 之方向增益圖形，左圖為立體圖，右圖為等高線圖，d=0.03m。

由圖3-3，可以看出TJR的增益圖型在目標方向是非常尖銳的，因此它對於目標聲源角度偵測有不錯的效果，但是在較低頻部分各角度都有較高的增益，表示在極低頻部分增益的鑑別度較低，作VAD的時候低頻分辨率會下降。如果當加大兩麥克風間距時，低頻的增益鑑別度會比較好，但是在高頻的部分會產生空間上的混疊(Aliasing)失真，由圖3-4，即使不是來自於目標聲源方位的高頻訊號也會有很大的增益，如果以TJR資訊作VAD時即會在此處發生錯誤，因此須避免用較大麥克風間距的裝置來判斷。根據頻率=聲速/波長，如果以聲速340公尺/秒來計算，本文所使用的取樣頻率為8000Hz，想要完全不造成高頻混疊失真的話，麥克風間距需在0.0425公尺以下，但是實際進行實驗時，覺得在兩倍左右，大約0.08公尺以內都在可以接受的偏差範圍。如果為多顆麥克風裝置，可以讓不同間距的麥克風對，分別負責判斷不同的頻帶，較寬間距的麥克風對判斷較低頻，較窄間距的判斷較高頻以互補優缺點。

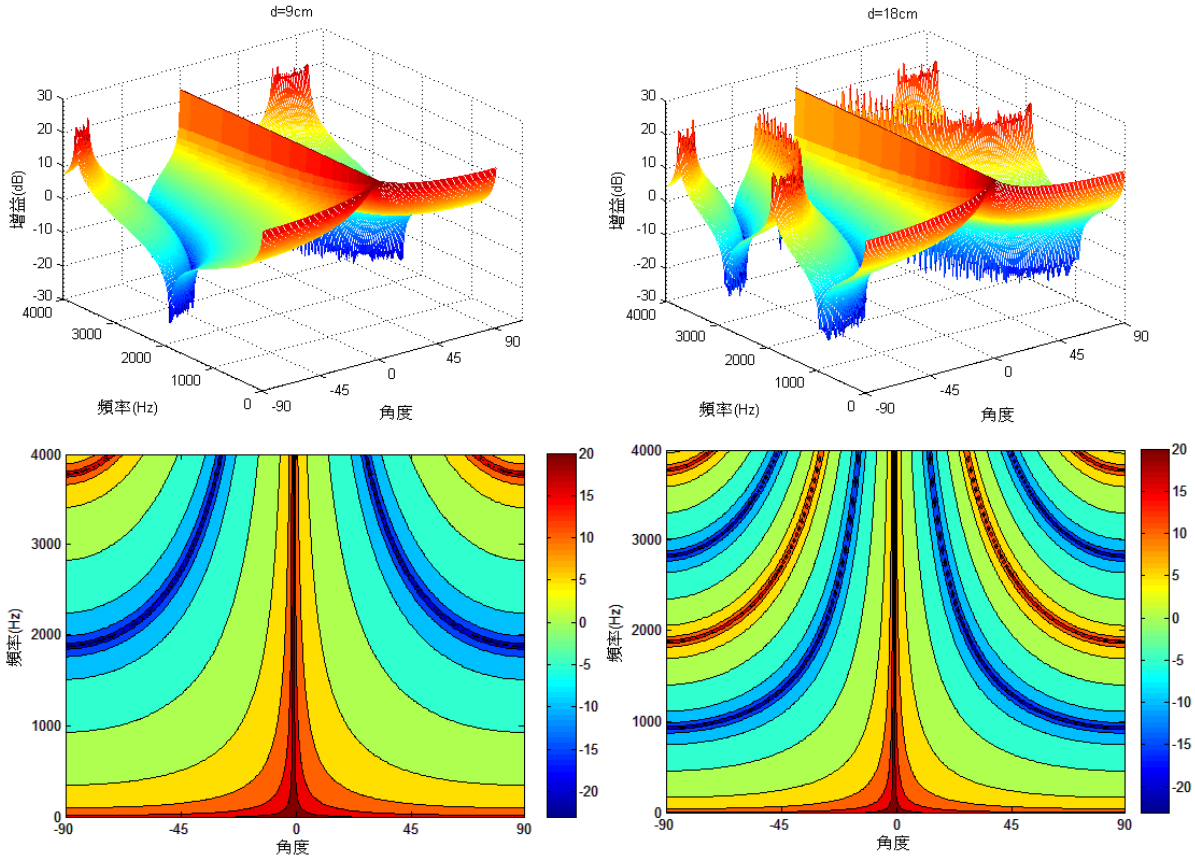


圖 3-4：TJR 之方向增益圖形(寬間距)，麥克風間距為 0.09 公尺(左)與 0.18 公尺(右)，上圖為立體圖，下圖為等高線圖。

因為TJR在各頻率都有不同增益，所以門檻值須在各頻率分開設定，分別定為增益函數於各頻率的平均，如下式表示：

$$\overline{D}_{TJR}(k)=\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} D_{TJR}(k,\theta)d\theta \tag{3-9}$$

圖 3-5 為門檻值的表示圖，一般來說，為低頻較高、高頻較低的遞減曲線，右圖為麥克風間距較大，發生空間上混疊失真實的情況，在某些頻帶增益較高，導致 VAD 判斷錯誤，因此間距較大的裝置只建議採用較低頻的判斷結果。

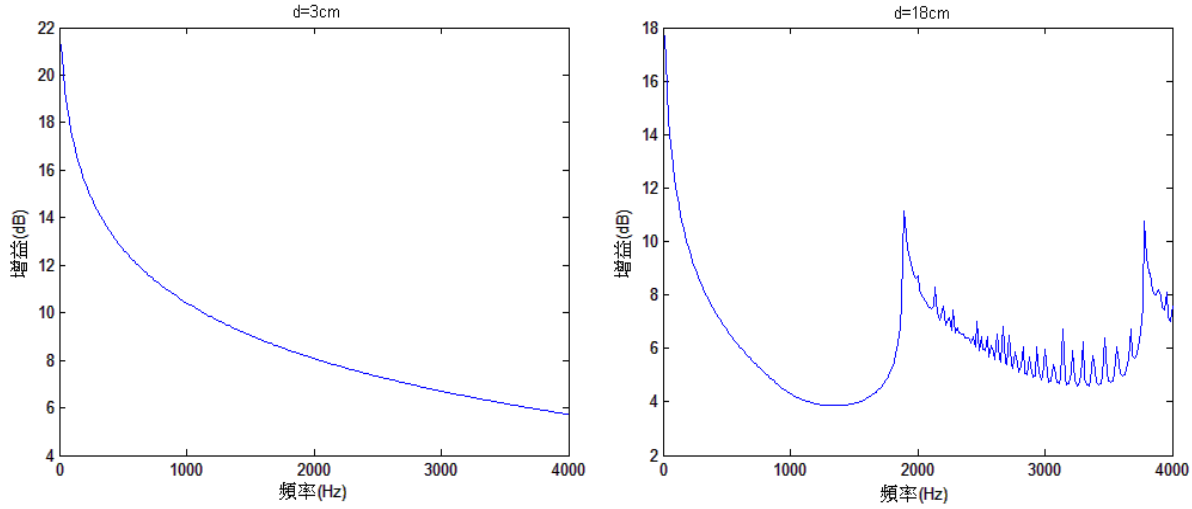


圖 3-5：各頻率門檻值，麥克風間距 0.03 公尺(左)與 0.18 公尺(右)。

而 VAD 的判斷方式如下：

$$VAD(t,k)=\begin{cases}1, \text{if TJR(t,k)} > \overline{D}_{TJR}(k) \\ 0, \text{otherwises}\end{cases} \tag{3-10}$$

理論上的判斷結果為圖 3-6 所示，以左圖 3 公分間距來說，判定為目標聲源的角度大約為正負 20 度之間，高頻的部分會稍微小一些，而在極低頻的部分則會快速擴大到正負 35 度左右，因此在低頻的判斷較容易發生錯誤，而右圖為麥克風間隔較大的情況，高頻混疊失真的部分須捨棄，相較小間隔來說，低頻的容許角度更窄更精確，極低頻的擴大情況也相較好一些。

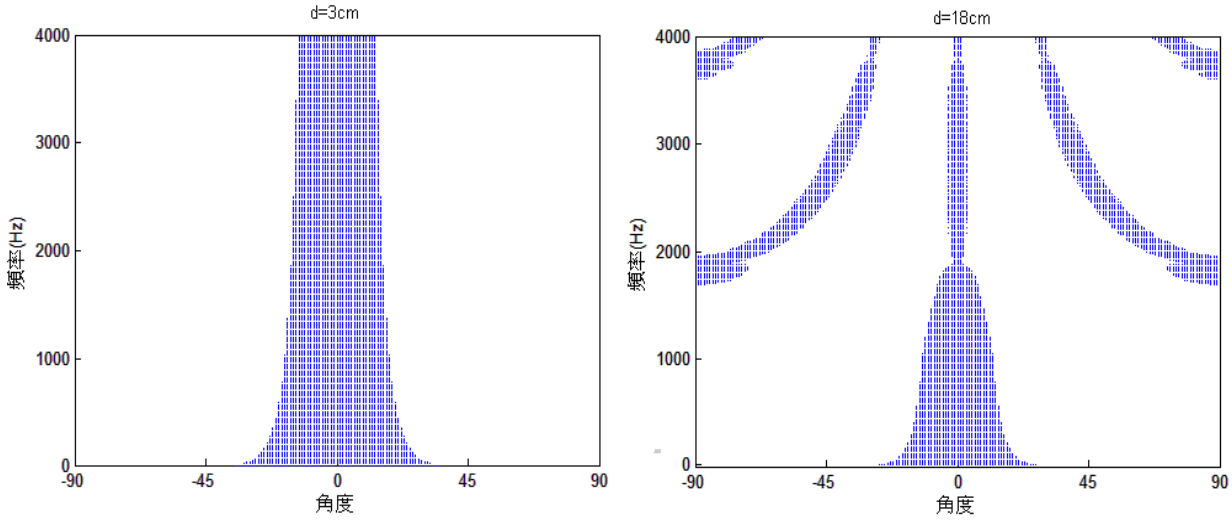


圖 3-6，VAD 判斷圖，0.03 公尺(左)與 0.18 公尺(右)。

在多顆麥克風的場合，不同間隔的麥克風對，分別對應不同頻帶，避免高頻混疊失真，並且盡量壓抑低頻快速擴大現象，圖 3-7 為使用八顆麥克風的情況，左圖為門檻值，右圖為理論上 VAD 判斷圖，可以被判斷為目標聲源的聲源，大約都在正負 15 度之間。

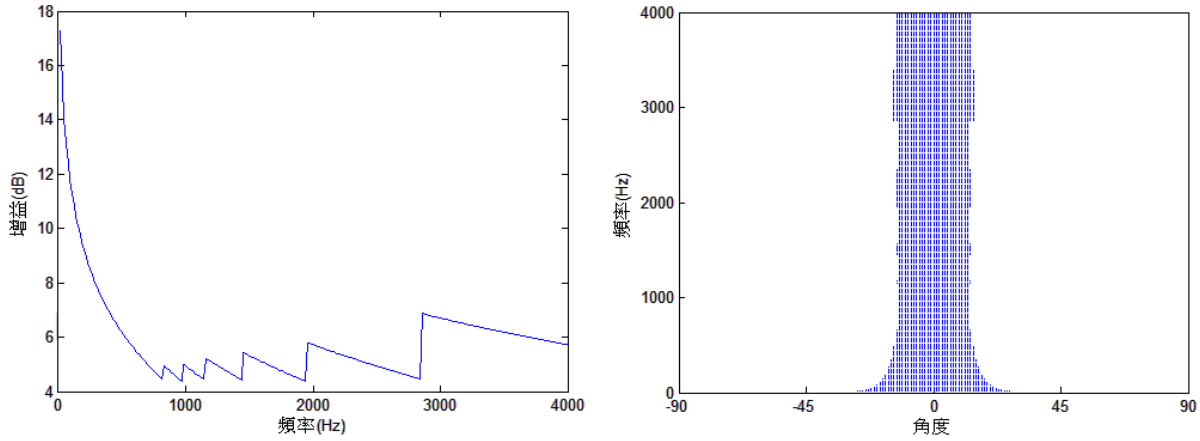


圖 3-7，VAD 判斷圖(8 顆麥克風)，間距 0.03 公尺，各頻率門檻值(左)與 VAD 判斷圖(右)。

此處所使用的 VAD 不同於一般以音框為單位的 VAD，而是各音框中每頻帶也都作判斷，以單元(Cell)為單位的 VAD，代表此音框中這一頻帶的訊號比較像目標聲源，或是比較像干擾雜訊。根據心理聲學的觀念，在極短的時間之內，很鄰近的頻率中，一般人耳沒辦法解析度很高的分辨兩個以上的聲音，只會有能量較高的那個聲音留下來，因此我們可以上述的 VAD 製作頻譜遮罩(Mask)。估測的雜訊能量改寫為：

$$\sigma_N^2(t,k)=\begin{cases}|\boldsymbol{v}^T(t,k)\,\boldsymbol{\beta}(t,k)|^2, & VAD(t,k)=1\\ R^2(t,k), & VAD(t,k)=0\end{cases} \tag{3-11}$$

如此最後的純化輸出會只留下判斷屬於目標聲源的部分，屬於雜訊的成分直接捨棄，這樣的做法可以最大幅度的加強 SNR，但是作為代價，當 VAD 不是極度精準時，會造成純化後語音的失真上昇，頻譜上的不連續也會造成處理後的聲音變得比較尖銳，因此如果以語音品質為目標時，並不建議直接刪去雜訊部分，而是預設一最小值(spectral floor) ，表示如下：

$$\sigma_N{}^2(t,k)=\begin{cases}|\,\boldsymbol{v}^T(t,k)\,\boldsymbol{\beta}(t,k)\,|^2, & VAD(t,k)=1\\ \beta_f\cdot R^2(t,k), & VAD(t,k)=0\end{cases} \tag{3-12}$$

β_f 即為 SNR 和語音品質的權衡(trade-off)因子，會設定為一較大的數，本文設定為 0.95，之後由 2-9 式以同樣方法可得語音純化的輸出 $\hat{S}$ 。

如果後端有需要使用以音框為單位的 VAD，例如純化後的聲音送給後端為語音辨識器，則可以將各音框中每一頻帶的 VAD 判斷結果統計起來，再取適當之門檻值：

$$\mu_S=\frac{1}{\mathrm{k}_2-\mathrm{k}_1+1}\sum_{k=k1}^{k2}VAD(t,k) \tag{3-13}$$

其中 k_1 和 k_2 分別捨棄鑑別度不高的低頻，和容易包含空間上混疊失真的高頻，只使用中頻較可靠的部分判斷：

$$VAD_{frame}(t)=\begin{cases}1, \text{if} > threshold\\ 0, \text{otherwises}\end{cases} \tag{3-14}$$

之後可再適當的加上指數移動平均(Exponential Moving Average, EMA)與Hang over機制，這一部分本論文沒有使用到，就不在此詳細介紹。

第四章 實驗結果與分析

4.1 實驗環境

本章將前面章節的演算法進行實驗，使用的錄音裝置為線性八顆麥克風陣列，各麥克風間距為 3 公分，圖 4-1 為聲源位置示意圖，為了比較在各種噪音下演算法的效果，將各聲源分別單獨錄音，之後再用後製混和，認定 Target Source 為欲純化之目標聲源，位於正前方 0°，而 Interference 為干擾聲源，來自於 45°或-45°，Interference 有兩個方位是為了模擬同時兩不同方向的不同干擾聲源環境，而 F16 noise 屬於穩態雜訊，背對麥克風裝置對牆壁播放，為了模擬沒有明顯方向性之雜訊。

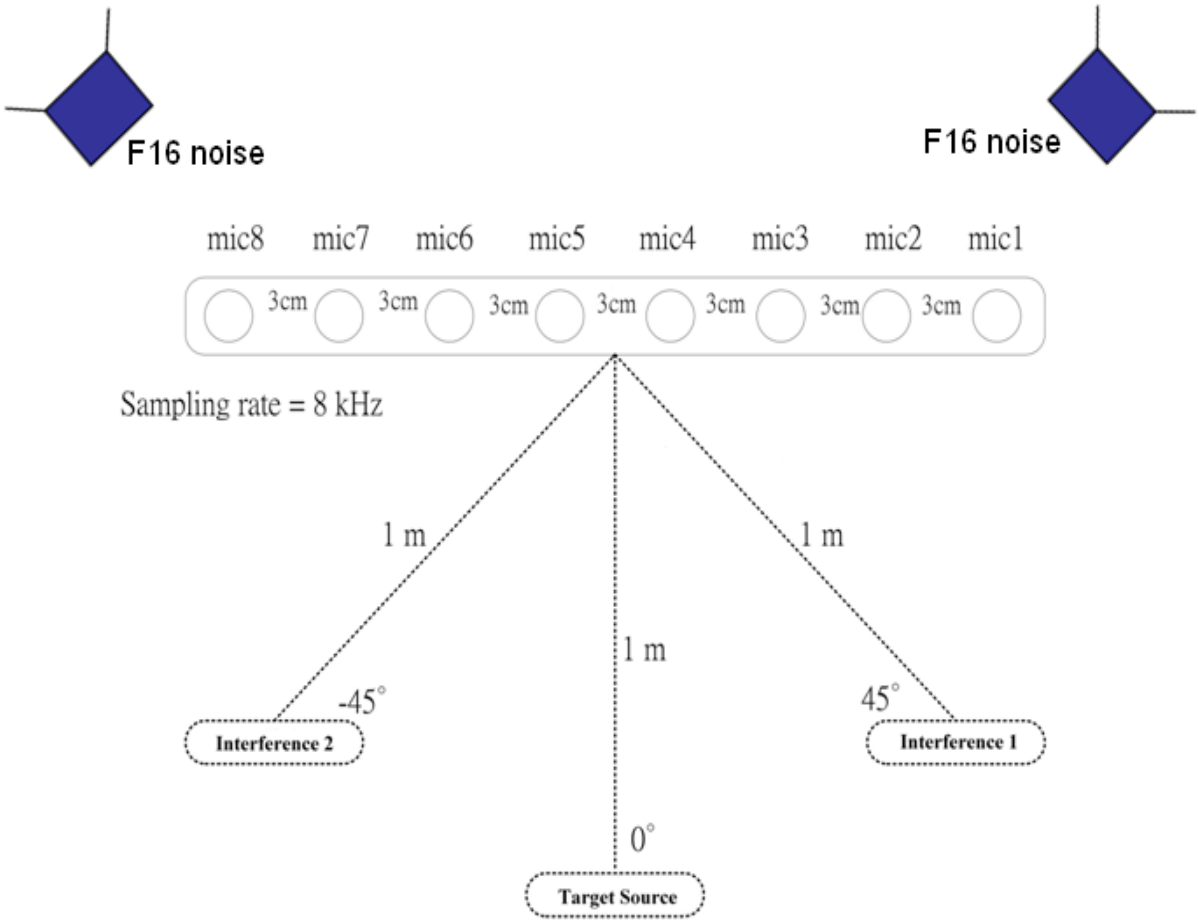


圖 4-1：各聲源方位示意圖。

模擬環境有四種情況，所包含之語料列於表 4-1，Case 1 為基本表現演算法估測雜訊的效果，Target Source 固定為一男性的人聲，此時雜訊為沒有明顯特性的 Babble noise，Case 2 表現當干擾聲源為不穩態雜訊時的效果，Woman 1 為干擾聲源，是一女性的人聲，Case 3 表現當有兩不同的干擾聲源，分別來自不同之方向時，演算法估測雜訊的效果，Woman 2 為另一女性的人聲，Case 4 加入沒有方向性的雜訊來評估效果。

環境	Target Source	Interference 1	Interference 2	F16 noise
Case 1	Man	Babble		
Case 2	Man	Woman 1		
Case 3	Man	Woman 1	Woman 2	
Case 4	Man	Woman 1	Woman 2	F16

表 4-1：測試用模擬環境

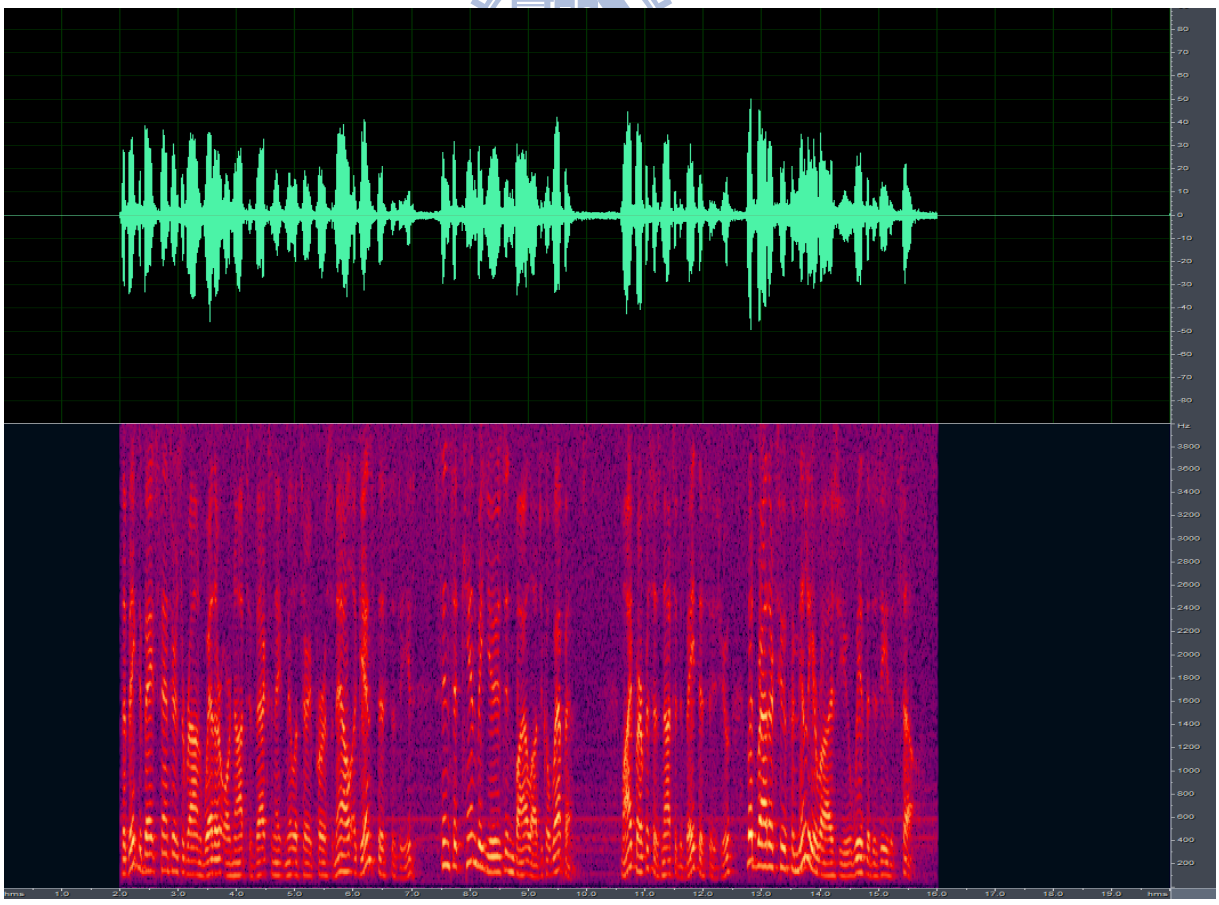


圖 4-2：目標聲源 Man，上圖為時域訊號，下圖為頻域訊號。

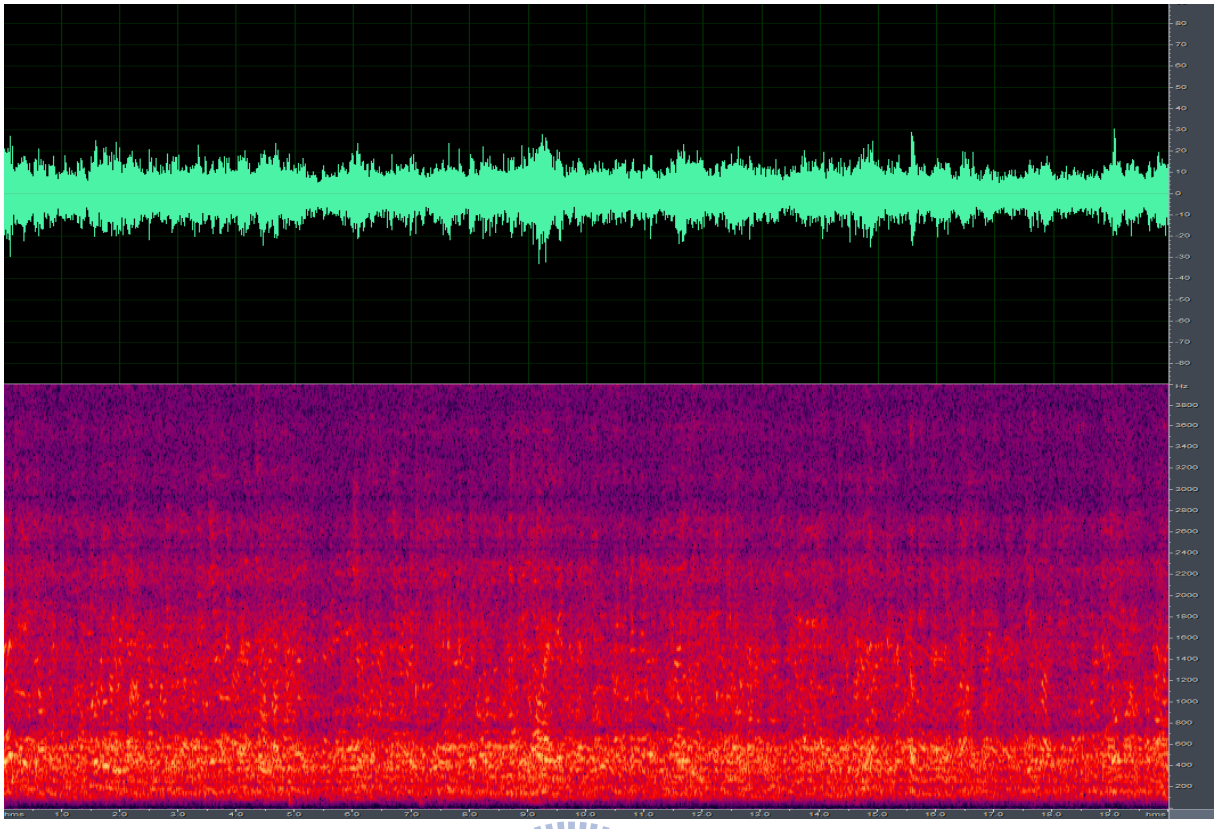


圖 4-3：雜訊 Babble，上圖為時域訊號，下圖為頻域訊號。

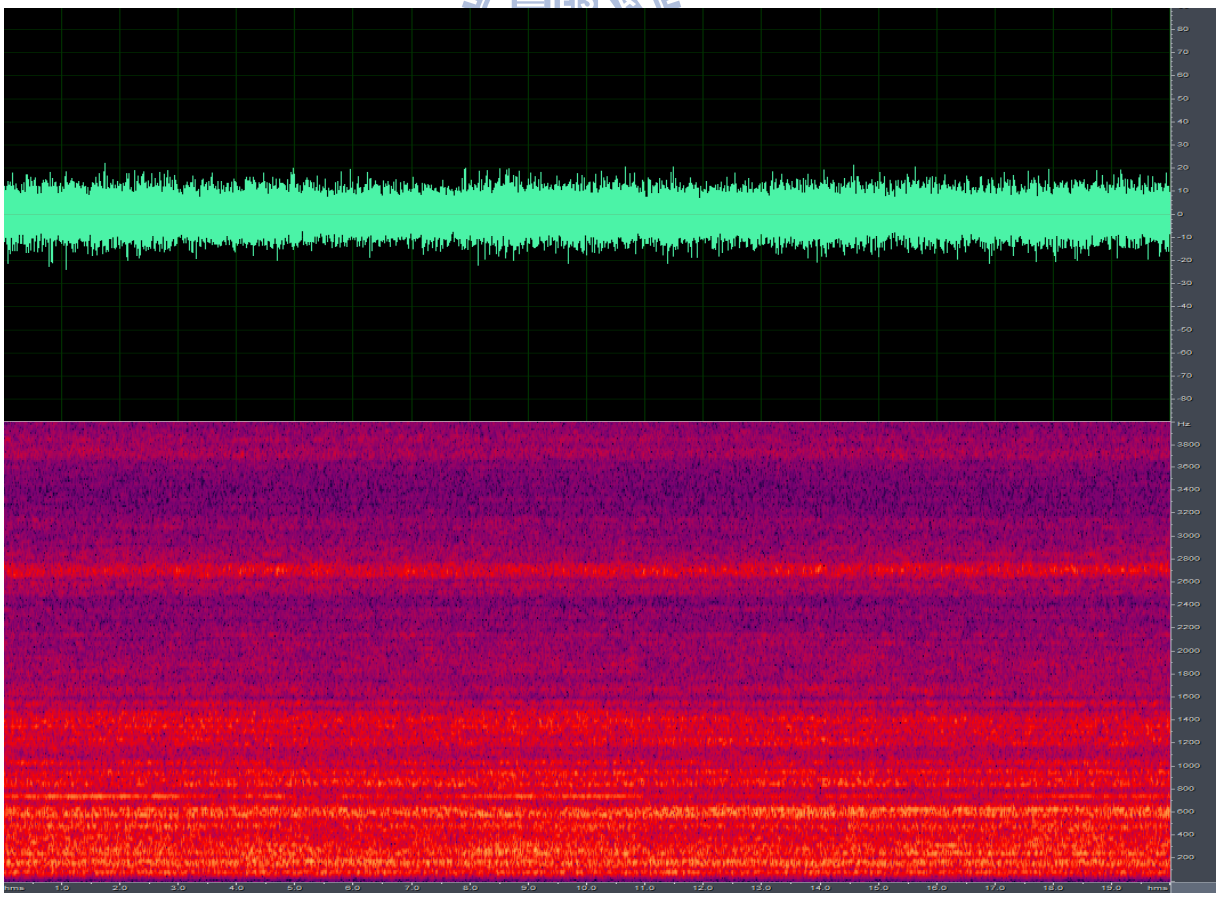


圖 4-4：雜訊 F16，上圖為時域訊號，下圖為頻域訊號。

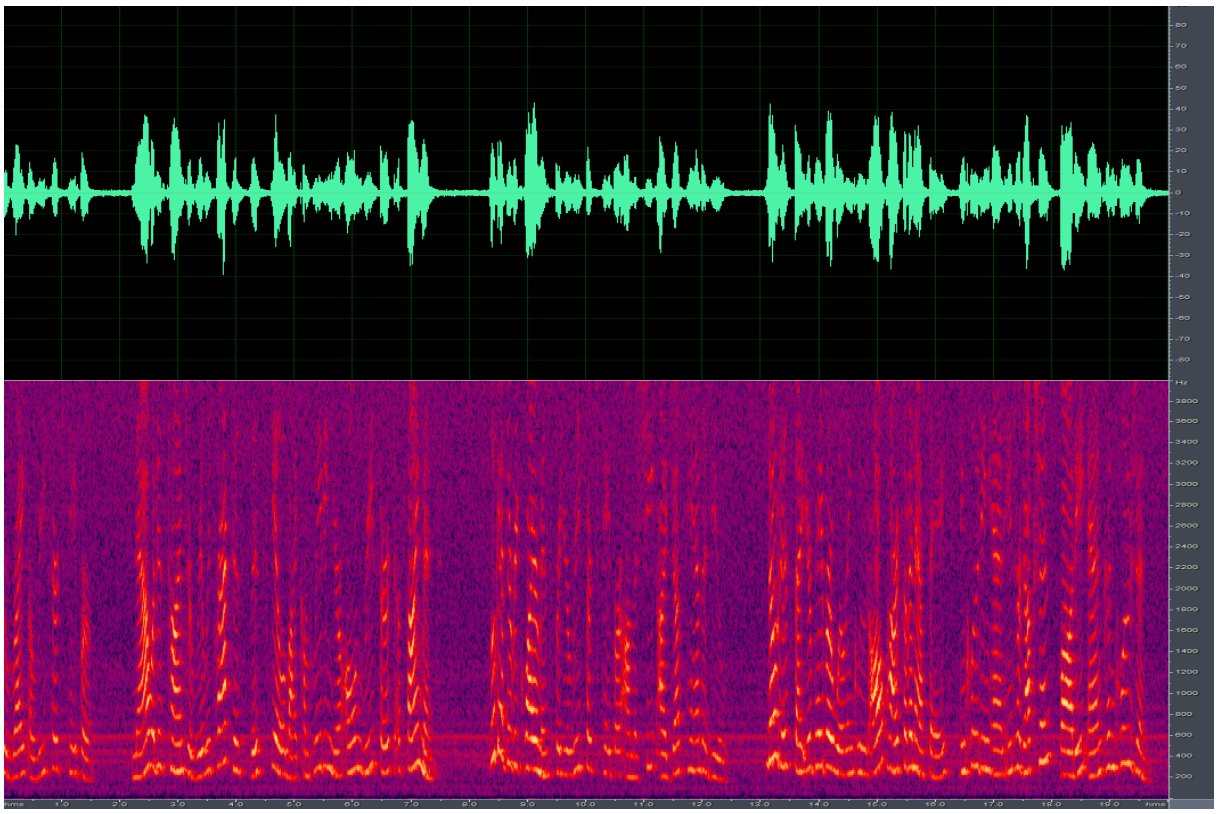


圖 4-5：干擾聲源 Woman 1，上圖為時域訊號，下圖為頻域訊號。

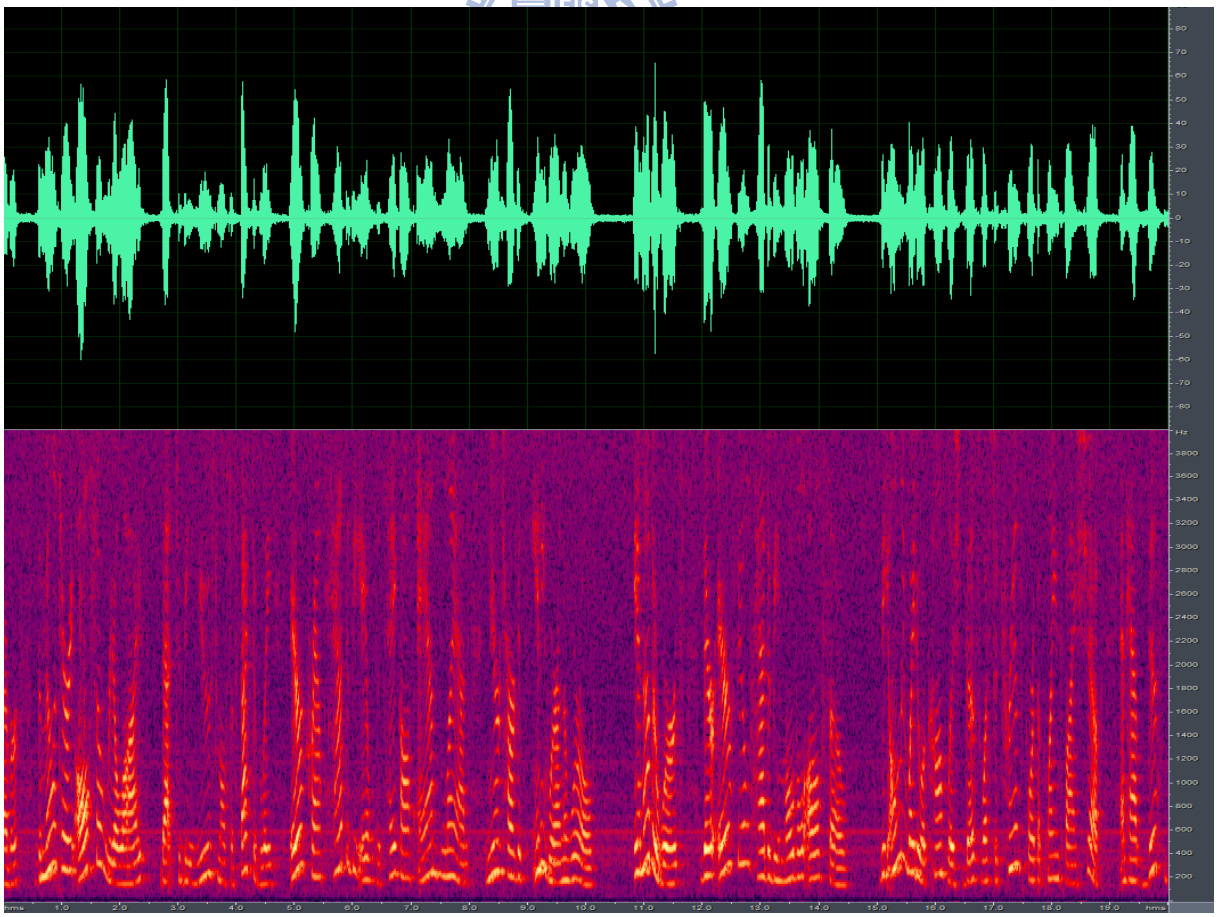


圖 4-6：干擾聲源 Woman 2，上圖為時域訊號，下圖為頻域訊號。

為了比較演算法的各種結果表現，本文採用兩種客觀性的評估標準，其一為 SNR，在實驗中的 SNR 計算方式如下：

$$10\log_{10}\left(\frac{\sum_{i=M}^{N}x^2(i)}{N-M+1}\right) \tag{4-1}$$

假設干擾雜訊為第 M_1 到第 N_1 筆，而目標聲源加干擾雜訊為 M_2 到 N_2 筆，其 SNR 為：

$$10\log_{10}\left(\frac{\sum_{i=M_2}^{N_2}x^2(i)}{N_2-M_2+1}\right)-10\log_{10}\left(\frac{\sum_{i=M_1}^{N_1}x^2(i)}{N_1-M_1+1}\right) \tag{4-2}$$

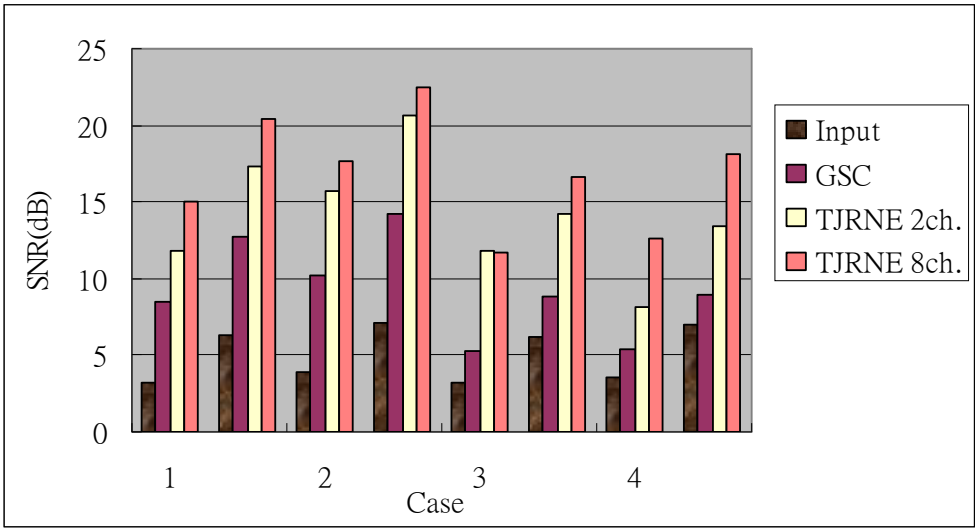
另一評估的標準為 PESQ (Perceptual Evaluation of Speech Quality, 知覺語音評價)，此標準為國際電信聯盟遠端通訊標準化組(ITU-T)所建議客觀評量 mean opinion scores (MOS)的方式，用以測量原始聲音和接受到的聲音之間的失真程度，此標準的 MOS 得分在 1.0(糟)到 4.5(無失真)之間，很嚴重失真時也可能會小於 1.0，大約 3.8 代表付費電話可以接受的範圍，更詳細的計算方式，可以參考維基百科中 PESQ 的條目: http://en.wikipedia.org/wiki/PESQ，或是 ITU-T 的建議 P.862：http://www.itu.int/rec/T-REC-P.862/en。

4.2 不同雜訊環境下之實驗結果

本節將本文所提出算法在各種環境中進行模擬，另外也調整不同之 SNR 分別進行實驗，並且為了比較本演算法的效能，取具有代表性的 GSC 演算法作為對照組，實驗結果如表 4-2、4-3 所示，表 4-2 以 SNR 評估演算法對於雜訊的削減能力，表 4-3 則以 PESQ 評估純化後語音的失真程度，其中 Input 代表麥克風裝置中間一顆所收到的原始訊號，這裡 GSC 為使用 Griffith Jim 提出之 GSC 演算法，使用 8 顆麥克風，TJRNE 即為本論文所提出之演算法，分別以不同的麥克風個數作實驗比較。

Wiener filter求參數

Case	Input	GSC	TJRNE(Wiener)			
			2	4	6	8
1	3.2080	8.5099	11.8589	14.5168	14.6715	15.0265
	6.3432	12.6914	17.3251	20.1416	20.2180	20.3987
2	3.8982	10.2566	15.7224	17.3756	17.4926	17.6575
	7.0627	14.2143	20.5872	22.2795	22.6164	22.4679
3	3.1853	5.2738	11.8162	11.0976	11.6834	11.7267
	6.2416	8.7897	14.1998	15.7502	16.5656	16.6257
4	3.5545	5.3510	8.1535	10.1629	11.2930	12.6609
	6.9525	8.9989	13.4261	15.1512	16.8727	18.1140



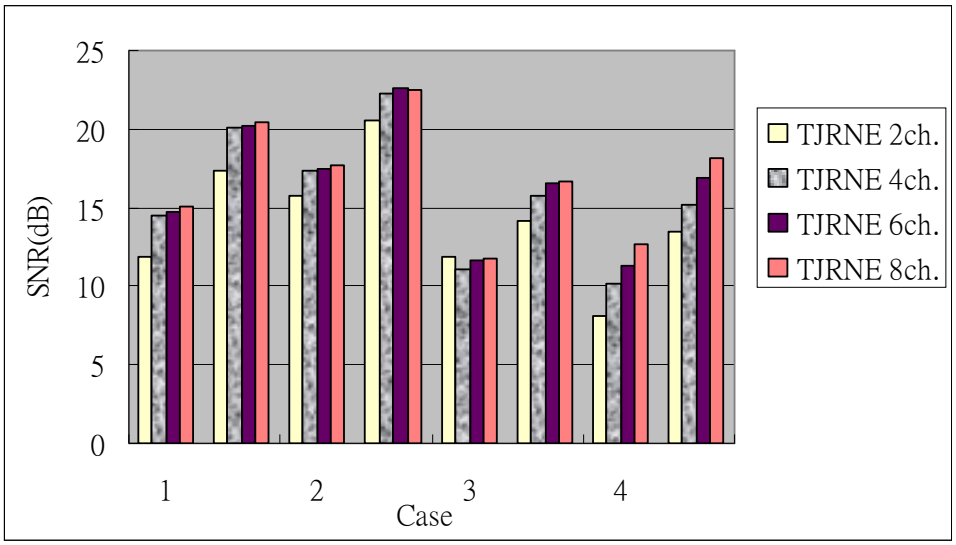
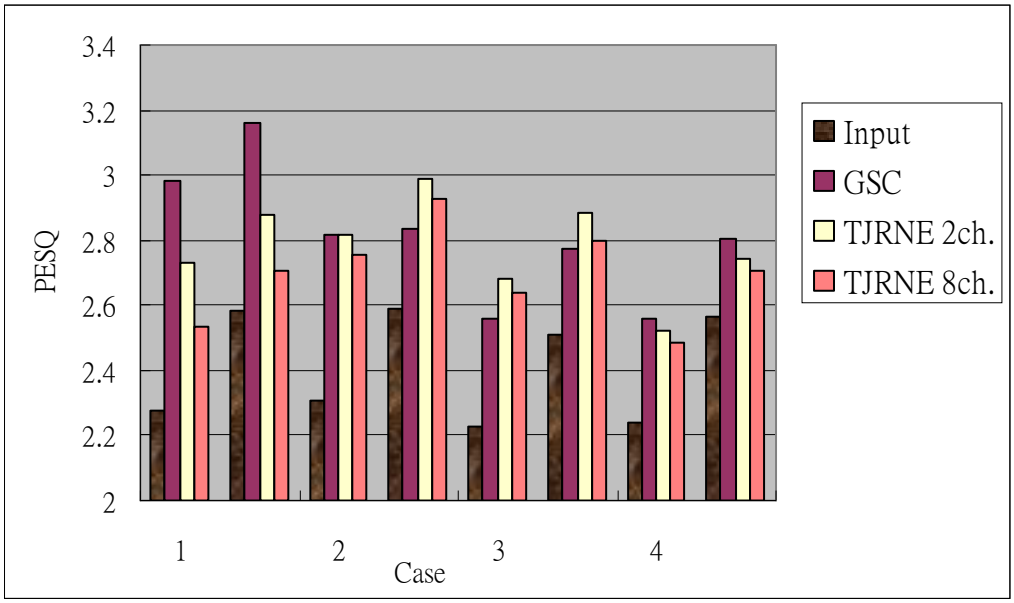


表 4-2：不同雜訊環境下 SNR (Wiener filter)，中圖為純化效果比較，下圖為多麥克風效果比較。

Case	Input	GSC	TJRNE(Wiener)			
			2	4	6	8
1	2.2768	2.9853	2.7282	2.6720	2.5786	2.5353
	2.5852	3.1626	2.8763	2.8661	2.7609	2.7051
2	2.3088	2.8176	2.8155	2.8367	2.7869	2.7526
	2.5925	2.8351	2.9876	2.9843	2.9430	2.9293
3	2.2262	2.5615	2.6790	2.6810	2.6484	2.6415
	2.5080	2.7739	2.8812	2.8530	2.8224	2.8001
4	2.2407	2.5568	2.5250	2.5228	2.5175	2.4842
	2.5621	2.8033	2.7401	2.7558	2.7387	2.7083



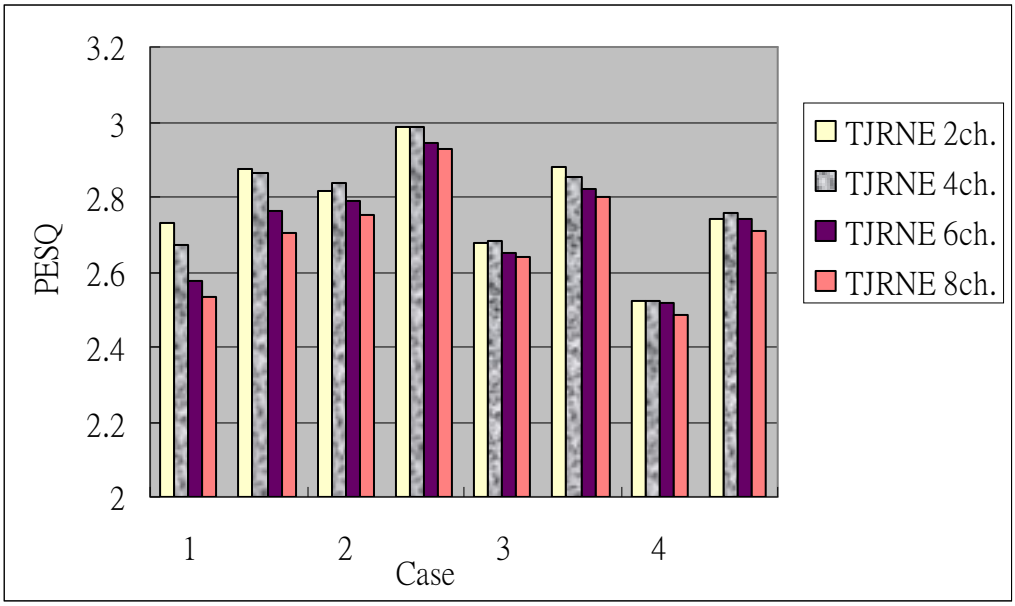


表 4-3：不同雜訊環境下 PESQ (Wiener filter) ，中圖為純化效果比較，下圖為多麥克風效果比較。

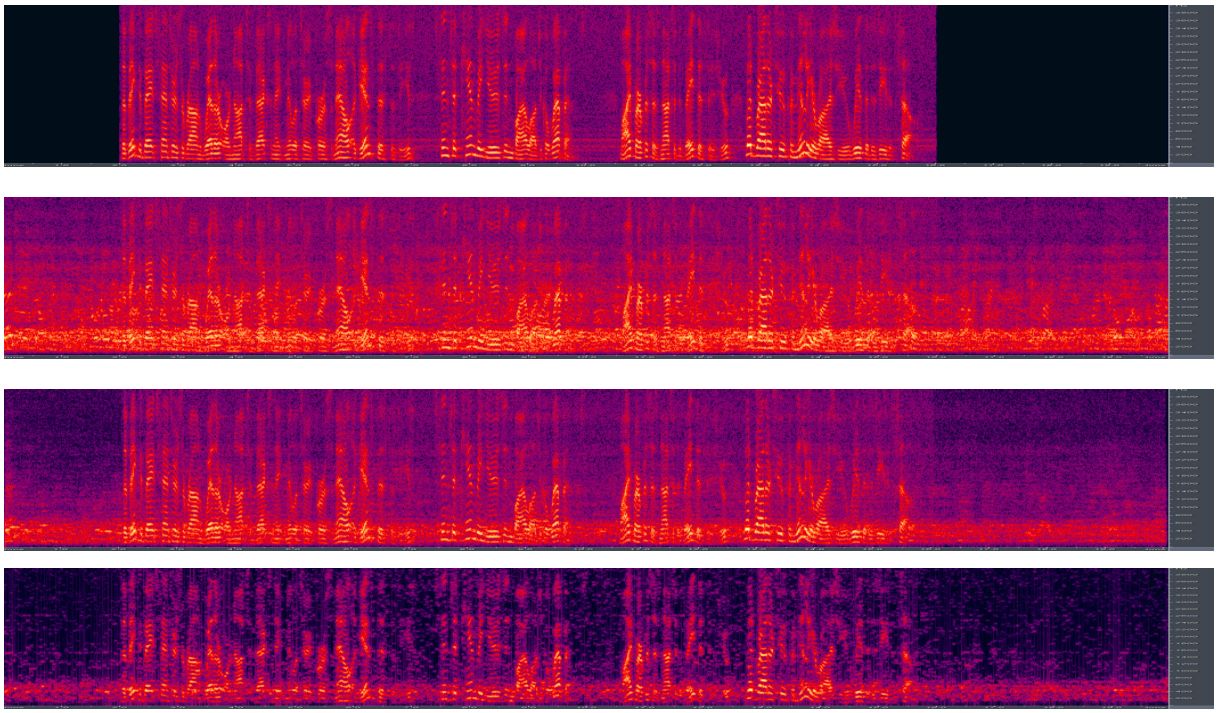


圖 4-7：Case 1 處理結果(Wiener filter)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

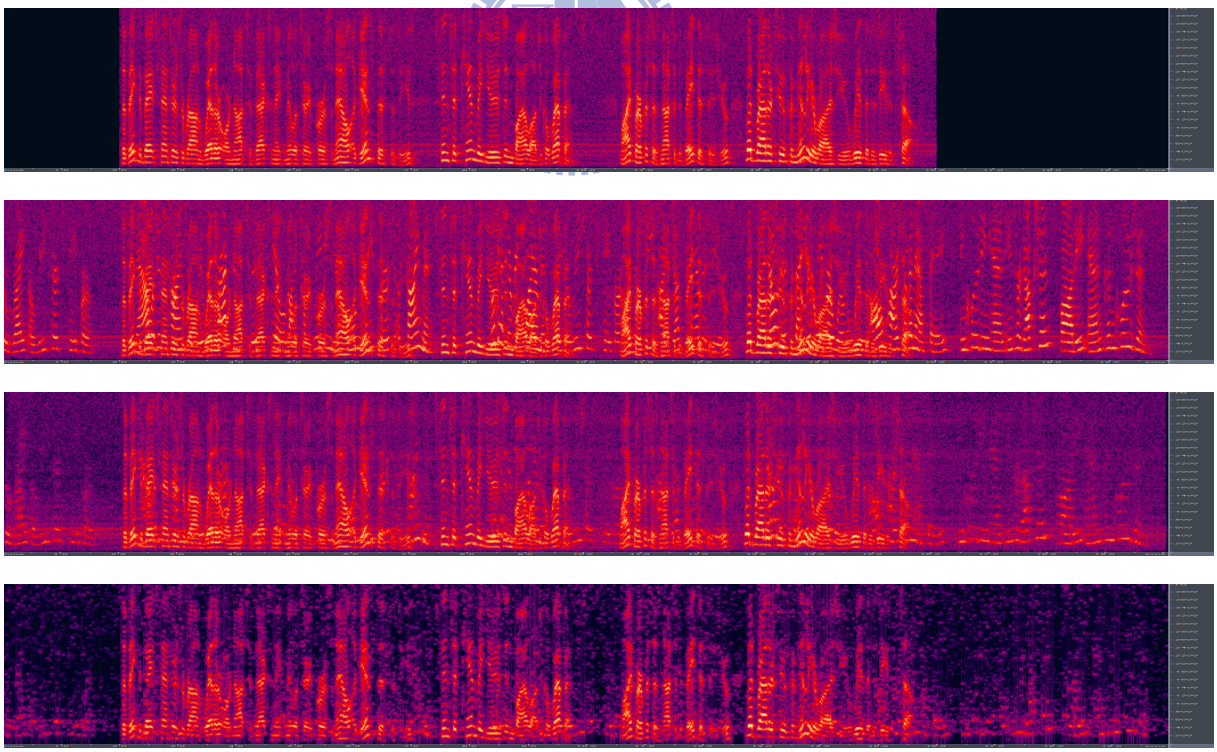


圖 4-8：Case 2 處理結果(Wiener filter)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

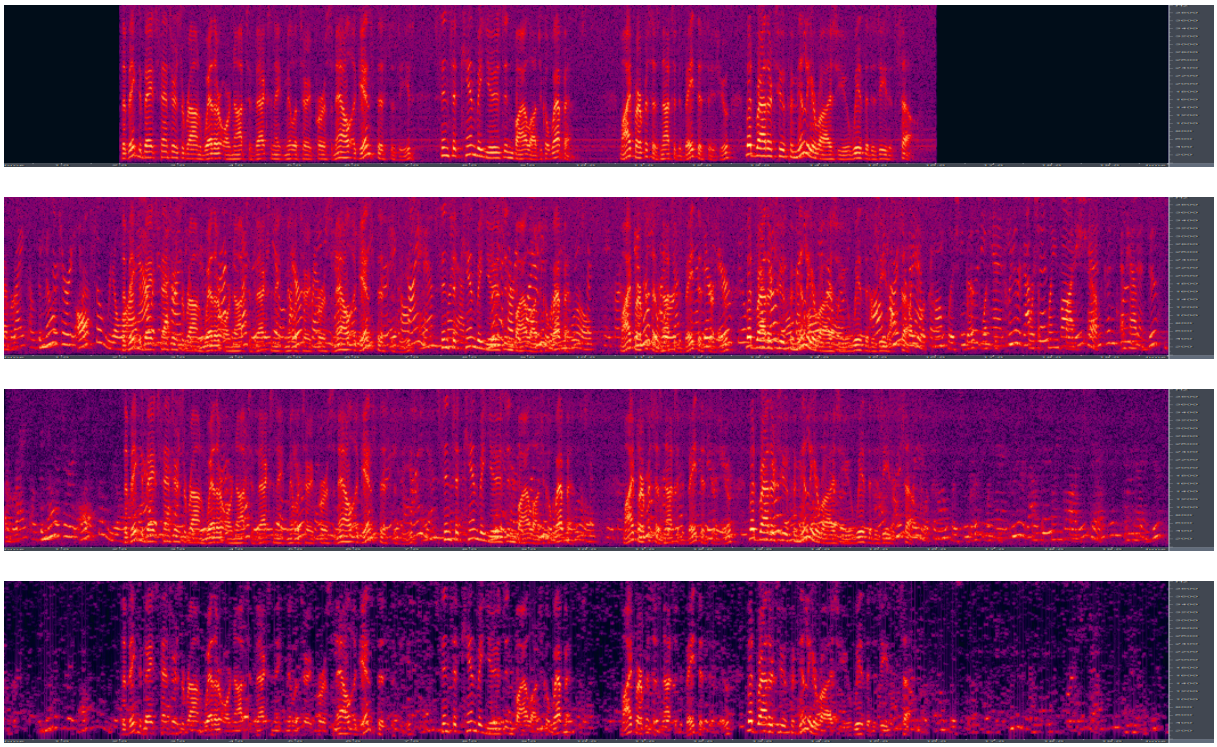


圖 4-9：Case 3 處理結果(Wiener filter)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

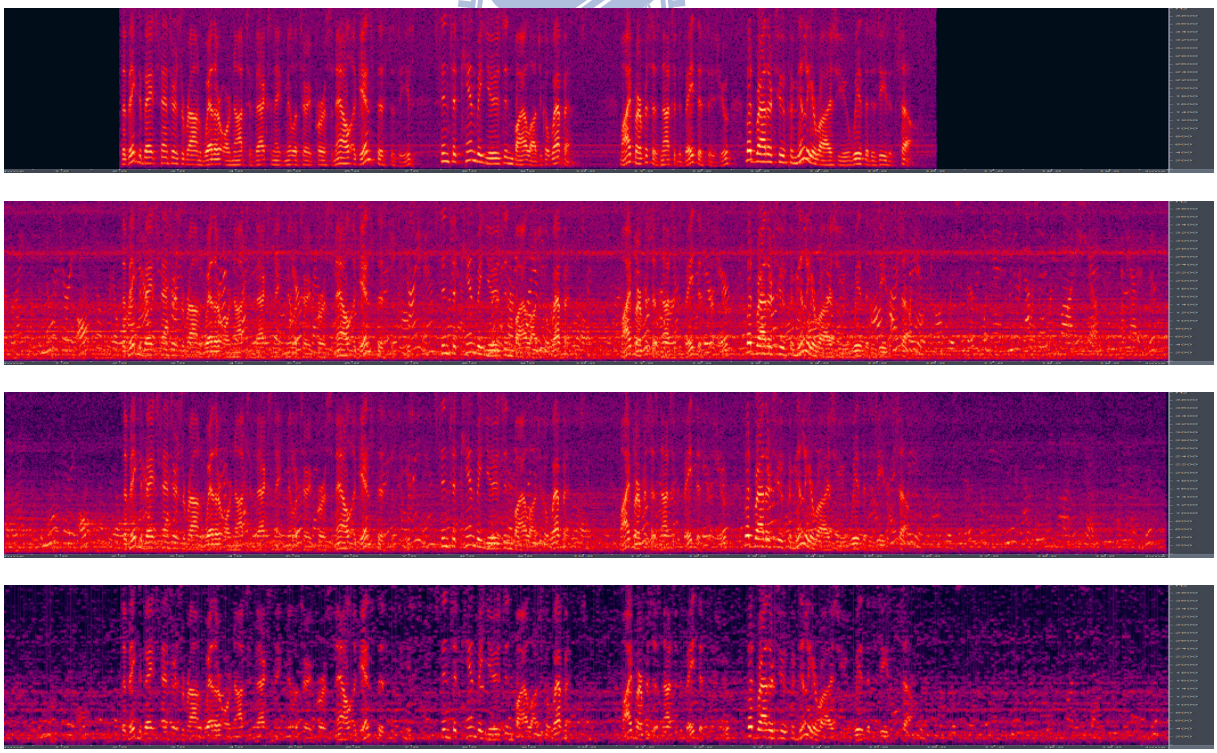


圖 4-10：Case 4 處理結果(Wiener filter)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

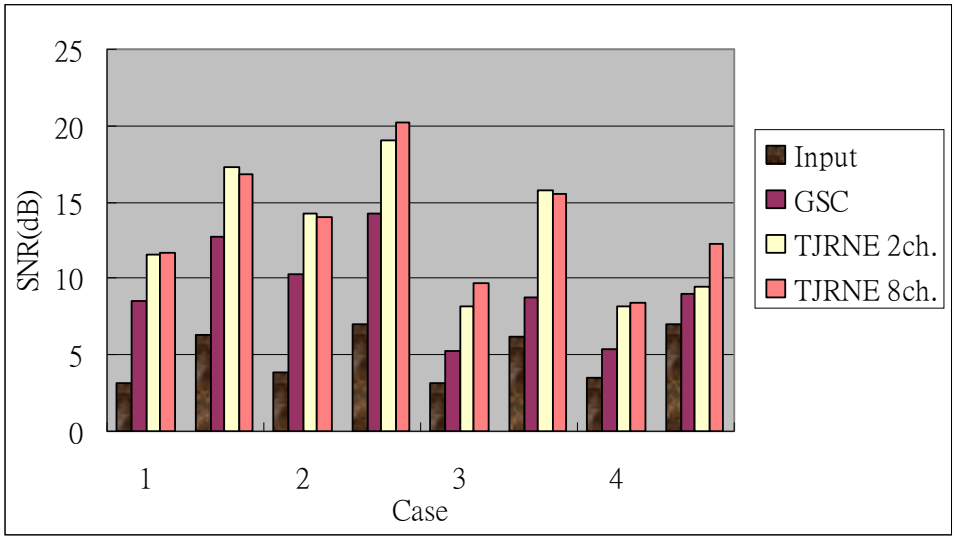
Wiener filter 小結

Case 1 中，雜訊為來自單一方向的穩態雜訊，較容易估計雜訊的特性以進行純化，將本文提出之方法與 GSC 演算法作比較，在此環境中雖然本演算法 SNR 上升較多，但是 GSC 的結果語音品質較好。Case 2 中，干擾為來自單一方向的不穩態雜訊，因為雜訊不斷的在變動，如果想要將雜訊精確的消去，Nullformer 的參數必須隨時進行更新。Case 3 中，加入了另一方向的干擾雜訊，如此容易造成雜訊的來源方向不固定，Nullformer 的參數更新將更為困難，Case 4 中加入了沒有明顯方向性的雜訊，因為也包含和目標聲源相同方位，此為單純使用空間特性很難處理的雜訊。

在使用 Wiener filter 進行整合時，麥克風個數的上升對於 SNR 的提升有正相關的幫助，而對於語音品質方面則沒有太多幫助，甚至麥克風個數少的比較好，此問題為反函數運算需長時間音框，所造成的估測不準確。

NLMS求參數

Case	Input	GSC	TJRNE(NLMS)			
			2	4	6	8
1	3.2080	8.5099	11.5556	11.7602	11.7412	11.7032
	6.3432	12.6914	17.2668	16.8929	16.8109	16.7891
2	3.8982	10.2566	14.2440	14.3565	14.0795	14.0179
	7.0627	14.2143	19.0645	20.4125	20.2757	20.1581
3	3.1853	5.2738	8.1679	9.8024	9.8297	9.7029
	6.2416	8.7897	15.7649	15.6928	15.6441	15.5553
4	3.5545	5.3510	8.1550	8.4176	8.4845	8.4447
	6.9525	8.9989	9.4227	12.1517	12.2646	12.2233



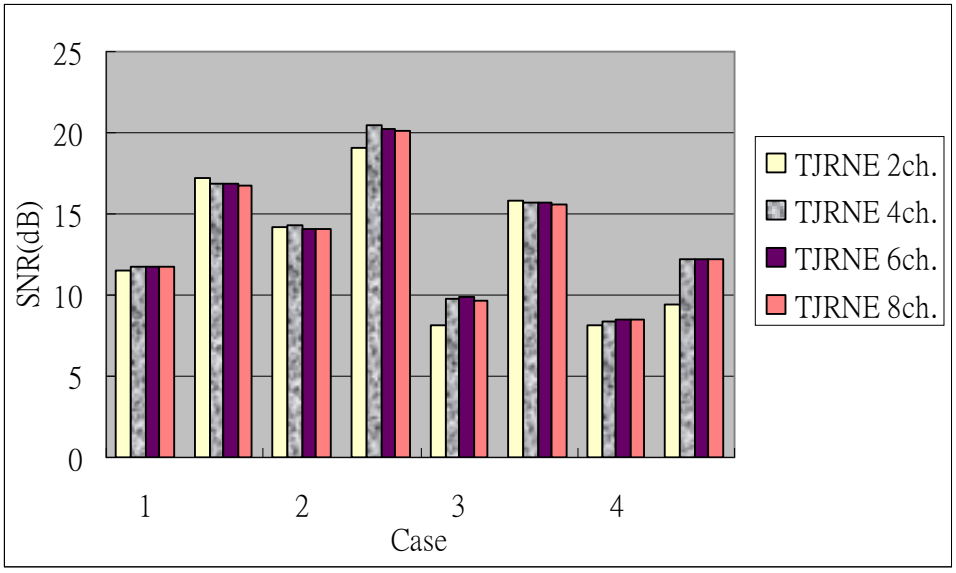
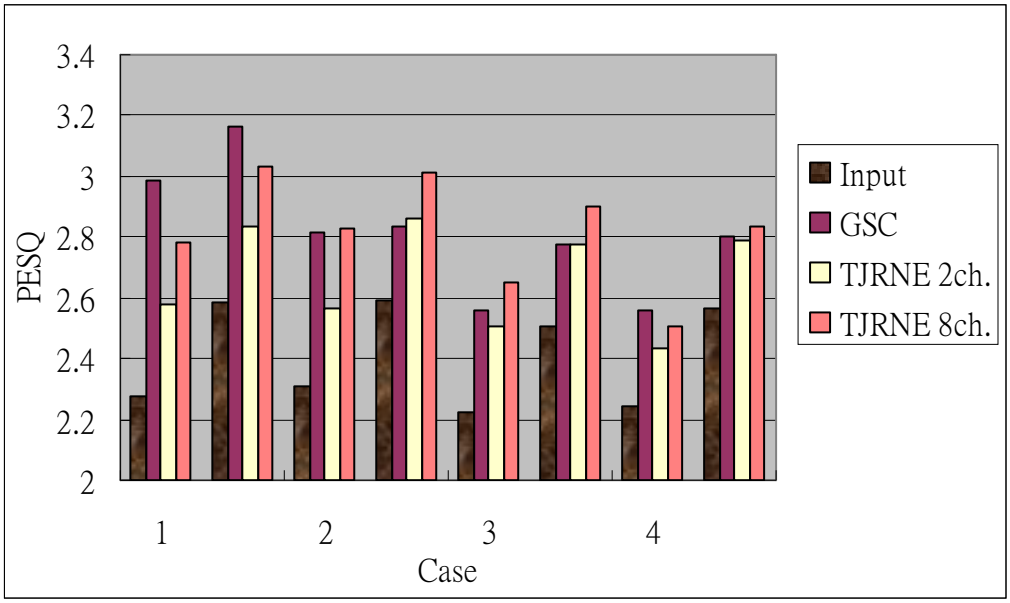


表 4-4：不同雜訊環境下 SNR (NLMS) ，中圖為純化效果比較，下圖為多麥克風效果比較。

Case	Input	GSC	TJRNE(NLMS)			
			2	4	6	8
1	2.2768	2.9853	2.5767	2.7711	2.7806	2.7807
	2.5852	3.1626	2.8372	3.0250	3.0275	3.0337
2	2.3088	2.8176	2.5655	2.6500	2.7993	2.8295
	2.5925	2.8351	2.8618	2.8599	3.0102	3.0127
3	2.2262	2.5615	2.5034	2.3842	2.6384	2.6479
	2.5080	2.7739	2.7779	2.8906	2.8928	2.8972
4	2.2407	2.5568	2.4334	2.4196	2.4973	2.5077
	2.5621	2.8033	2.7884	2.8310	2.8346	2.8351



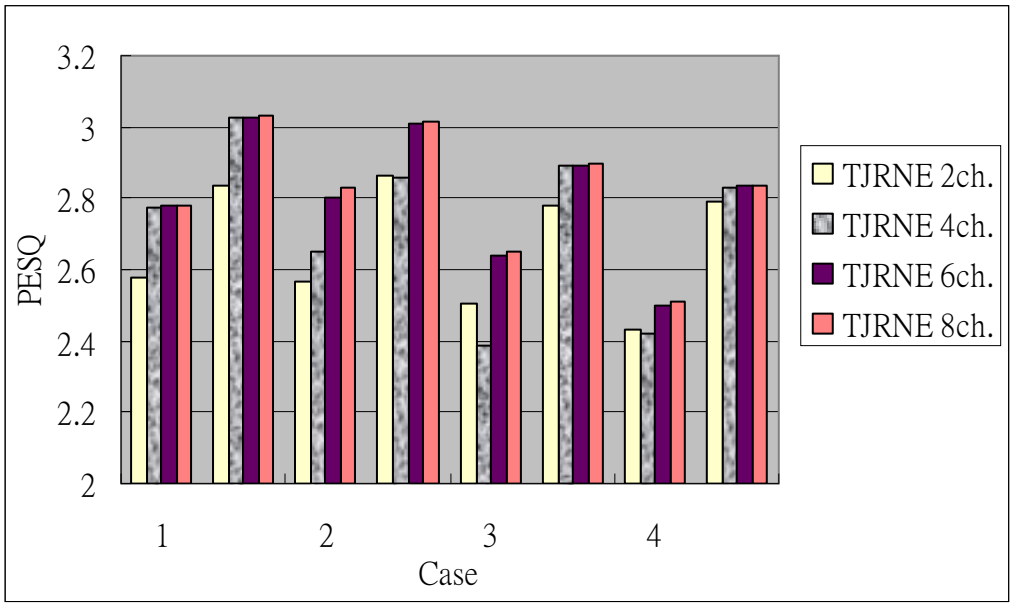


表 4-5：不同雜訊環境下 PESQ (NLMS)， 中圖為純化效果比較，下圖為多麥克風效果比較。

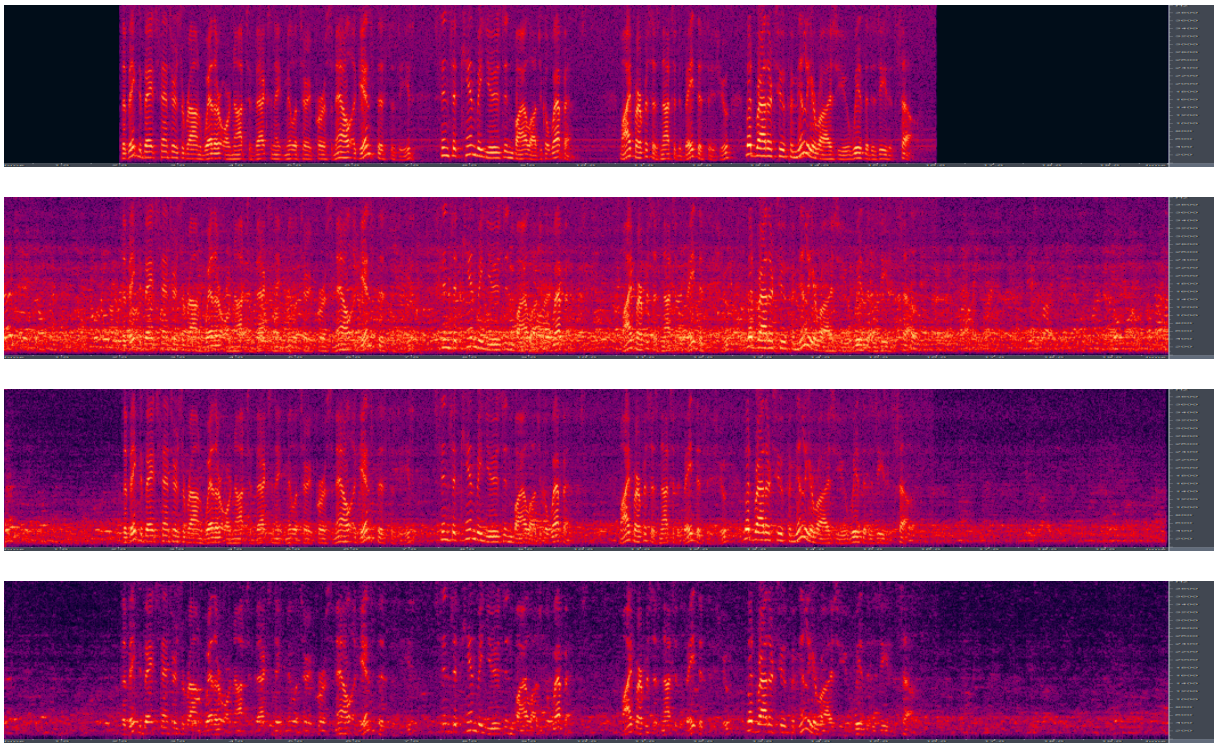


圖 4-11：Case 1 處理結果(NLMS)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

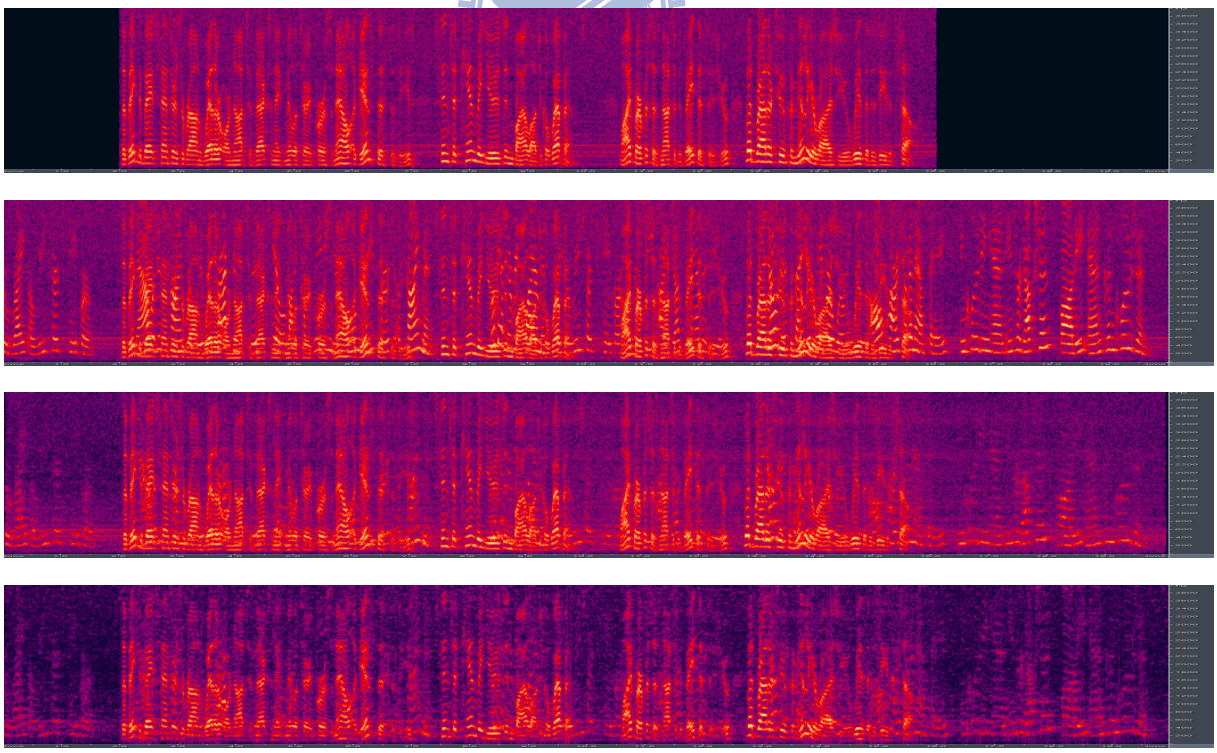


圖 4-12：Case 2 處理結果(NLMS)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

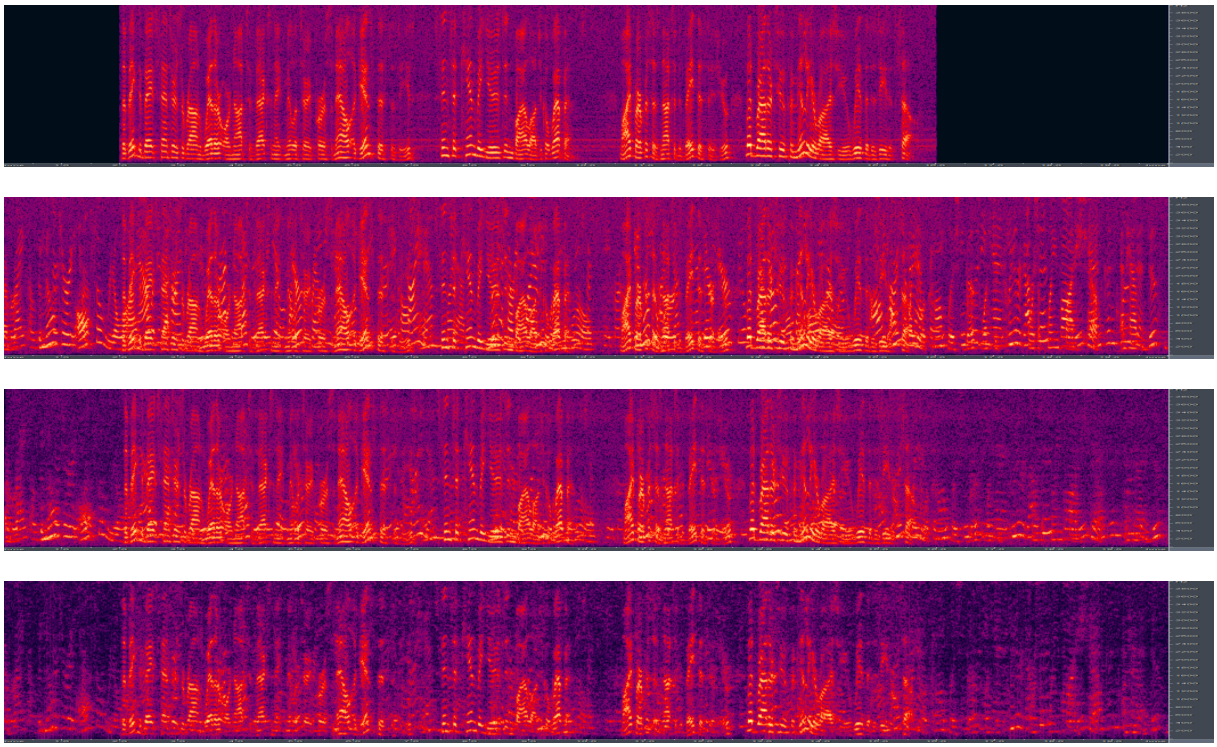


圖 4-13：Case 3 處理結果(NLMS)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

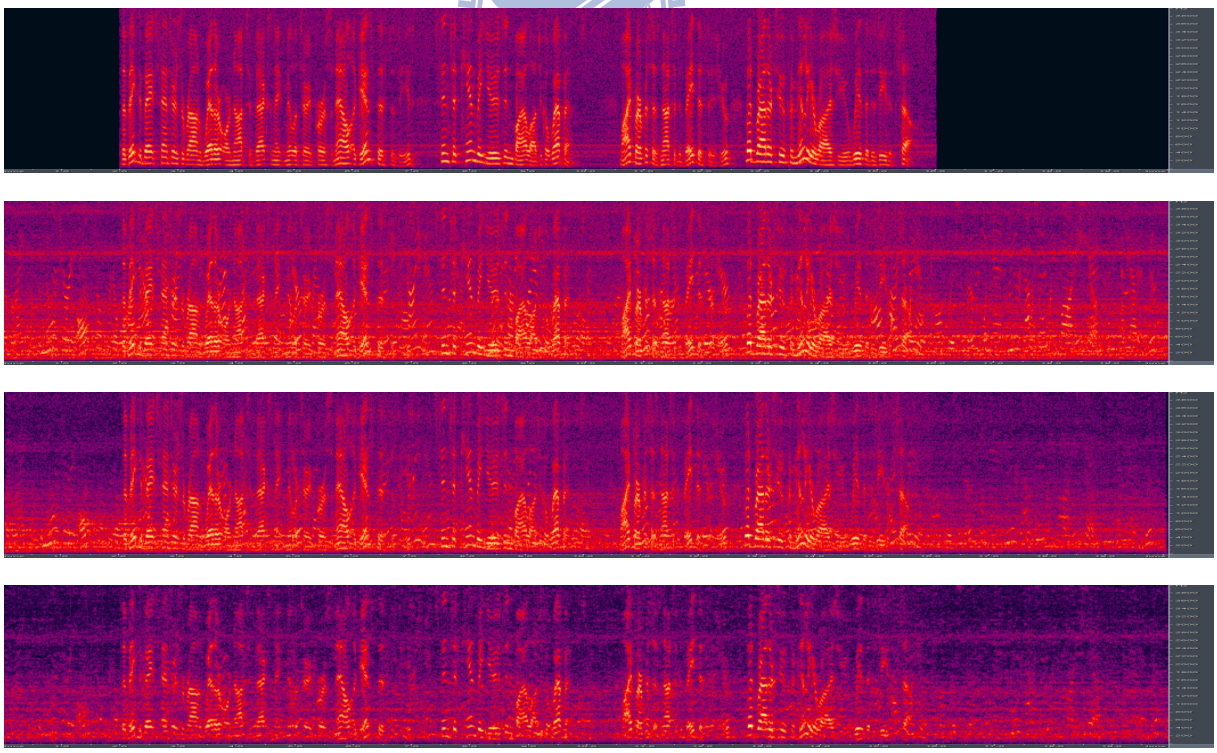


圖 4-14：Case 4 處理結果(NLMS)，(a)目標聲源，(b)麥克風收到的訊號，(c)GSC 純化結果，(d)本文提出之演算法純化結果。

NLMS 小結

與使用 Wiener filter 進行整合時相反，麥克風個數的上升對於語音品質的提升有正相關的幫助，而對於 SNR 提升方面則沒有太多幫助。各種情況綜合比較下可以得到此結論：使用 Wiener filter 對 SNR 提升較有效果，NLMS 對語音品質提升比較好。

相較於 GSC 演算法，本文所提出的演算法在各種環境下都可以得到更好的 SNR 提升，而在語音品質方面，純穩態雜訊的 Case 1 中，GSC 有較好的純化後語音品質，而在非穩態的環境，本文提出的演算法都有較好之語音品質，但在綜合環境中，使用 Wiener filter 進行整合的語音品質相較差一點。

4.3 實驗結果分析

透過實驗結果可以看出本文提出之雜訊估測演算法是有效的，使用此雜訊估測器進行語音純化，可提升目標聲源與環境干擾雜訊之 SNR，並透過語音品質評量，確保不造成目標聲源過多的失真。

在各麥克風對整合部分，Wiener filter 所得到的解是一較長時間平均最好的解，在這一段時間內此解可以消去最多雜訊，但是不見得為目前音框最佳的解，所以 Wiener filter 的方法，其表現為對 SNR 的提升有不錯的幫助，但是容易使目標聲源的失真變大，而 NLMS 為對於最佳解的追蹤，相較於 Wiener filter，NLMS 較注重於目前瞬間的資訊，因此比較能夠正確的消去目前的雜訊，所以相較不容易造成目標聲源的失真。

因此在調整各權衡的部分，進行 Wiener filter 時多取的組數，在保證穩定性的前提下，組數越少越好，而進行 NLMS 所使用的 step size，可以稍微慢一些，大約 0.7 左右在各環境都有不錯的表現。

頻譜遮罩對 SNR 的提升有很大的作用，但是也容易對目標聲源造成失真，除了使用 spectral floor 降低失真，也可以將 VAD 的 threshold 稍微調小一點，降

低被判斷成有聲音的機率。

在各種噪音環境的部分，本文演算法對於非穩態的干擾聲原或雜訊，有較好的穩健性，對於穩態雜訊或是沒有方向性的雜訊，也有不差的估測表現。

第五章 研究成果與未來展望

本論文提出一雜訊與干擾聲源估測器，以 Target Blocker 估測來自非目標聲源方向的訊號，並配合頻譜遮罩將其訊號從接收到的聲音中消去以進行語音的純化，而從實驗中可以看出，適應性補償器的部分利用 Wiener filter 或 NLMS 更新權重，能有效的提升對非穩態雜訊的穩健性。

但是此估測器也存在一些缺點，如麥克風間距不可以太大以避免混疊失真，TJR 資訊對於沒有方向性雜訊(diffused noise)估測能力較低，還有 Wiener filter 所使用間寬度、NLMS 的更新速度對於純化效果影響很大，更新速度越快代表越能跟上非穩態雜訊的變化，更新速度越慢代表擁有較長時間的統計資訊，對於穩態雜訊比較能估測的好，但卻需要較長時間來調整權重，這個權衡問題可以為未來發展的一個方向，是否可以設計一個機制，觀察估測到的雜訊特性，進而調整更新速度。

未來研究的發展方向可以考慮結合一些有事先資訊的方法，例如利用事前訓練好的 RTF(relative transfer function)來定義 TJR，對整個環境的空間資訊可以掌握的更好。

參考文獻

[1] Byung- Chul Kim; I-Tai Lu, ”High resolution broadband beamformingbased on the MVDR method,” OCEANS 2000 MTS/IEEE Conferenceand Exhibition, Volume: 3, pp. 1673 -1676, 2000.

[2] Griffiths, L.J., “A simple adaptive algorithm for real-time processing in antenna arrays,” Proceedings of the IEEE, 1969. 57(10): p. 1696-1704

[3] Griffiths, L. and C. Jim, “An alternative approach to linearly constrained adaptive beamforming,” Antennas and Propagation, IEEE Transactions on, 1982.30(1): p. 27-34.

[4] P. Aarabi, G. Shi, ”Phase-Based Dual-Microphone Robust Speech Enhancement,” Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, Volume: 34, pp.1763-1773, 2004.

[5] Kyuhong Kim, So-Young Jeong, Jae-Hoon Jeong, Kwang-Cheol Oh, Jeongsu Kim, ” Dual channel noise reduction method using phase difference-based spectral amplitude estimation,” Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference, 14-19 March 2010, p.217-220, 2010.

[6] Dahl, M. and Claesson I., “Acoustic noise and echo cancelling with microphone array”. Vehicular Technology, IEEE Transactions on, 1999. 48(5): p. 1518-1526

[7] M. W. Hoffman, Z. Li and D. Khataniar, “GSC-Based Spatial Voice Activity Detection for Enhanced Speech Coding in the Presence of Competing Speech”, IEEE Trans. Speech Audio Process., vol. 9, no. 2, pp.175-178, Feb 2001

[8] Cannot, S. and Cohen I., “Speech enhancement based on the general transfer function GSC and postfiltering”, IEEE Trans. Speech Audio Process., vol. 12, p. 516-571, Nov. 2004

[9] Berouti M., Schwartz R., Makhoul J., “Enhancement of speech corrupted by acoustic noise ,” Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79., Volume:4, pp. 208-211, 1979.