

國立交通大學
電機與控制工程研究所

碩士論文

使用適應波束形成與增益衰減後濾波器
之殘響消除方法



Dereverberation of Sound Signal Using Adaptive
Beamforming and Gain Suppression Post-filtering

研究生：楊宗翰

指導教授：胡竹生 博士

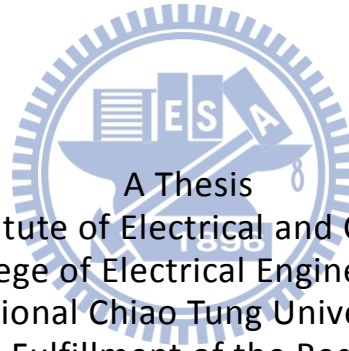
中華民國一百零一年九月

使用適應波束形成與增益衰減後濾波器之殘響消除方法

Dereverberation of Sound Signal Using Adaptive
Beamforming and Gain Suppression Post-filtering

研究生：楊宗翰 Student : Zong-Han Yang
指導教授：胡竹生 博士 Advisor : Jwu-Sheng Hu

國立交通大學
電機與控制工程研究所
碩士論文



A Thesis
Submitted to Institute of Electrical and Control Engineering
College of Electrical Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Master
in

Electrical and Control Engineering

September 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年九月

使用適應波束形成與增益衰減後濾波器 之殘響消除方法

研究生：楊 宗 翰

指導教授：胡 竹 生 博士

國立交通大學電機與控制工程研究所碩士班



本論文提出一套依照使用者需求調整訊號衰減率(Signal Decay Rate)來抑制聲音迴響的方法。本方法利用麥克風陣列接收到的訊號中有空間資訊的特性，首先將此空間資訊經由波束形成的方法，先濾除非目標訊號方向的迴響，接著依照訊號衰減的特性來估測訊號的衰減率，並對聲音的迴響能量做估算。結合能量衰減估測與麥克風陣列得到的空間資訊，本論文提出一套調整增益函式的後置濾波器。本方法不需要知道在實際應用中難以觀察的聲音脈衝響應(Acoustic Impulse Response)的資訊，實驗結果將證明本方法的效果。

Dereverberation of Sound Signal Using Adaptive Beamforming and Gain Suppression Post-filtering

Student : Zong-Han Yang

Advisor : Prof. Jwu-Sheng Hu

Institute of Electrical and Control Engineering



ABSTRACT

This thesis proposes a method of de-reverberation by adjusting signal decay rate. The method utilizes spatial information in the received signals of microphone array. First, by beamforming operation, the direct-path and reverberation of non-target signals are filtered. Then, we estimate the signal decay rate (the energy of reverberation) according to the characteristic of signal decay. By combining estimation of energy decay and the spatial information from the microphone array, a method to adjust the gain function during post-filtering is proposed. The proposed algorithm does not need to know the information of acoustic impulse response which is often difficult to obtain in real practice. Experimental results show the effectiveness of the proposed method.

誌 謝

兩年的碩士生活，在不知不覺中很快的就飛逝而去了，我的碩士生涯也即將畫下句點。在這兩年中，我學習到很多，無論是研究或是處事的態度。首先要感謝的是我的指導教授，除了在我的研究中指導我之外，還會告訴我們研究應該有的態度。除此之外，另外要感謝的就是我的奶奶、爸爸、媽媽、哥哥、姐姐，他們總是會鼓勵我，並讓我沒有後顧之憂的學習。

而對於實驗室的大家，感謝幫助我最多的唐哥，祝你也能早日畢業。也感謝知道很多美食資訊的阿吉，還有很強壯的 JUDO，以及很會做甜點的阿法。另外感謝耕博，唐哥以外幫我最多的就是你了。什麼運動都很威的男哥，幫我投履歷的沛錡學長，在 CIC 教我很多東西的育成，APPLE 迷 MACACA，很幽默的昀軒，很有趣的湘筑，希望各位學長姐都能達到自己的理想。

此外，很 A 的汶建，常常會一起喇賽，中文很好的丹尼爾，還有既是同學又是室友的鳴哥，有你們這些同學真的很棒。還有碩一的室友智凱，大家一起出去玩，真的很開心，碩二的室友小莫、彥樺，總是會大家一起喝酒、一起出去吃大餐，一起聊天。你們都是很棒的室友，讓我可以任何情況下回寢室，心情都能變得很好。還要感謝美食一姐小紅豆，常常陪我出去吃好料的。感謝納豆，常常可以一起宅一起混。感謝喇叭，可以一起打球又可以揪吃飯的，超讚。還有畢業前常常陪我去買東西的阿花。在這兩年的碩士生活中，你們大家都是功不可沒的，希望大家都能一切順心。

目 錄

摘 要	I
ABSTRACT	II
誌 謝	III
目 錄	IV
表 列	VI
圖 列	VII
第一章 緒論	1
1.1 研究動機.....	1
1.2 研究目標.....	2
1.3 文獻回顧.....	3
1.4 論文貢獻.....	4
1.5 論文架構.....	5
第二章 背景技術介紹.....	6
2.1 陣列訊號處理.....	6
2.2 封閉空間的迴響與迴響時間	8
2.2.1 封閉空間的迴響	8
2.2.2 迴響時間.....	10
2.3 DAHL 演算法.....	11
2.4 後混響譜方差.....	14
第三章 論文方法.....	22
3.1 系統架構說明.....	22
3.2 波束形成的比較與選擇	23
3.3 利用可變訊號衰減率消除迴響	30
第四章 實驗結果與分析.....	39
4.1 實驗環境介紹.....	39
4.2 系統消除迴響效果.....	41

第五章 結論	50
5.1 研究成果.....	50
5.2 未來展望.....	50
REFERENCE	51



表 列

表 3-1: 知覺語音評價分數表-----	27
表 3-2: 訊號殘響比對照表-----	38
表 3-3: 對數譜失真對照表-----	38
表 4-1: 波束形成輸出結果比較表-----	43
表 4-2: 系統輸出結果比較表-----	44



圖 列

圖 2-1: 均勻線性陣列架構圖	8
圖 2-2: 聲源直接到達路徑與反射路徑	9
圖 2-3: 聲音脈衝響應概要圖	10
圖 2-4: Dahl 演算法訊號擷取圖	12
圖 2-5: Dahl 演算法架構圖	14
圖 3-1: 系統架構	22
圖 3-2: 波束形成之訊號殘響比	25
圖 3-3: 波束形成之對數譜失真	26
圖 3-4: $\sigma_v=100, \rho_v=0.001$ 卡曼濾波器頻譜圖	29
圖 3-5: $\sigma_v=100, \rho_v=10$ 卡曼濾波器頻譜圖	29
圖 3-6: $\sigma_v=100, \rho_v=1000$ 卡曼濾波器頻譜圖	29
圖 3-7: 訊號衰減圖	30
圖 3-8: 能量衰減曲線	31
圖 3-9: 最小平方方法逼近曲線	32
圖 3-10: 後置濾波器之訊號殘響比	33
圖 3-11: 後置濾波器之對數失真比	33
圖 3-12: 後置濾波器之訊號殘響比	34
圖 3-13: 後置濾波器之對數失真比	34
圖 3-14: 純化訊號之訊號殘響比	36
圖 3-15: 純化訊號之對數譜失真	36
圖 3-16: 兩階段輸出之訊號殘響比	37
圖 3-17: 兩階段輸出之對數譜失真	37
圖 4-1: 線性麥克風陣列平台	40
圖 4-2: 無響室聲源與麥克風陣列關係	40
圖 4-3: 室內空間聲源與麥克風陣列關係	41
圖 4-4: Dahl 演算法之波束方向圖	42
圖 4-5: 時間延遲相加波束之波束方向圖	42
圖 4-6: TF-GSC 之波束方向圖	42

圖 4-7: 卡曼濾波器之波束方向圖	-----43
圖 4-8: 輸入訊號	-----45
圖 4-9: Dahl 演算法波束形成輸出	-----45
圖 4-10: 時間延遲相加波束輸出	-----45
圖 4-11: TF-GSC 波束形成輸出	-----46
圖 4-12: 輸入訊號時頻圖	-----46
圖 4-13: Dahl 演算法波束形成時頻圖	-----46
圖 4-14: 時間延遲相加波束時頻圖	-----47
圖 4-15: TF-GSC 波束形成時頻圖	-----47
圖 4-16: Dahl 演算法全系統輸出	-----48
圖 4-17: 時間延遲相加波束全系統輸出	-----48
圖 4-18: TF-GSC 全系統輸出	-----48
圖 4-19: Dahl 演算法全系統時頻圖	-----49
圖 4-20: 時間延遲相加波束全系統時頻圖	-----49
圖 4-21: TF-GSC 全系統時頻圖	-----49



第一章 緒論

1.1 研究動機

隨著科技的發展，手提電話已經成為人們的生活必需品了，為了因應各種場合的需求，手提電話也發展出不同的接聽方式，例如：耳機或藍芽接聽、手持聽筒以及免持聽筒。其中免持聽筒除了使用者的聲音，還會把一些環境中的噪音也一起接收，即使是在安靜的室內，也會因為聲音的反彈造成的迴響而使語音有失真的現象，所以希望能藉由收到的聲音，來抑制空間中的迴響，使語音訊號的品質上升。



在目前的應用中，除了上述的手提電話免持聽筒模式，助聽器也是其中一項重要的應用，若能藉由抑制空間中的迴響，使助聽器的使用者能聽到更清楚的聲音，以利於與別人的交談與互動。

若希望能抑制訊號源以外方向的聲音，使用空間濾波器是一個不錯的方式。由於在一般的使用環境下，聲音的迴響會從四面八方反射過來，但是主要聲源的方向只有一個，因此可利用空間濾波器(Spatial Filter)先做初步的抑制。在麥克風陣列的技術中，空間濾波器在聲音訊號處理中是個頗為熱門的領域，利用訊號的空間關係，對不同方向的訊號做出不同增益值，以達到空間濾波的效果，藉以分離空間中不同方向聲源的訊號。

利用估測聲音訊號的能量，通常只能就估計的能量來當做基準，若能量估計的非常不準確的話，會使後面一連串處理的效果大打折扣。因此若能有一個可以調整參數的機制加入估測能量的過程中，就能依現有的情況來做調整，使最後得到的訊號，更能符合不同使用者的各種需求。

在本論文中，利用 Dahl 演算法的方法先找出一組權重(Weight)，讓收到的訊號使用這組權重來達到波束形成(Beamformer)，也就是空間濾波的功能。由於對於消除迴響來說，迴響時間(Reverberation Time,RT60)是一個很重要的指標參數，所以我們先藉由訊號衰減的情況來計算迴響時間以及訊號的衰減率(Delay Rate)。接下來利用已估測的訊號衰減率來估計後段迴響能量(Late Reverberation Energy)，再依照不同頻帶(Frequency bin)的能量比來產生出增益函式(Gain Function)，經過空間濾波器的訊號再以不同頻帶的增益函式去抑制迴響，可以使處理過後的訊號的迴響明顯改善。此外，由於估測出不同的訊號衰減率，會使估計後段迴響能量的不同，造成消除迴響以及失真的狀況不同。本論文提供一個方法，可以依使用者的需求，藉由調整不同的訊號衰減率達到不同的消除迴響以及失真的權衡。

1.2 研究目標

在此將本論文的目標分為：

1. 探討適應性空間濾波器之演算法。

2. 探討估測訊號衰減速度之演算法。
3. 探討與比較不同訊號衰減速度與消除迴響的相關性。

1.3 文獻回顧

一般的情況下，錄製各種不同情況下的迴響聲音，是很不方便的，例如不同的迴響聲音通常是隨著房間大小以及裡面的擺設、材質來決定的，所以 J.B. Allen 和 D.A. Berkley 研究了一套理論[1]，利用這套理論就可以輕易的模擬出各種不同大小、材質的房間內迴響的情況，讓聲音迴響的研究者們，能更輕易的取得需要的聲音資料。




消除迴響的方法可以分為很多種類，一般將消除迴響分成兩大類，第一類為需要估計聲音脈衝響應以消除回響(Reverberation cancellation)為目的，第二類為不用估計聲音脈衝響應以抑制回響(Reverberation suppression)為目的。

消除迴響有不少的方法，例如 Marc Delcroix 等人使用 Linear-predictive Multi-input Equalization(LIME)的 Blind deconvolution 方法[2-5]，利用聲音濾波器模型(Sound filter model)的原理，來還原目標聲源。例如 J.D.Polack 發展出可以套用到回響情況的一組模型[6]，利用估測出的迴響時間(Reverberation Time,RT60)來套入模型中，利用模型來估算迴響的能量，將其估測的能量使用 postfilter 來抑制聲音的迴響[7-8]。由於此模型需要先估測迴響時間，所以 M.R. Schroeder 等人[9-10]也對迴響時間的估測作了研究。也有人把兩者結合

在一起，例如 K. Furuya 等人[11]將經過 Blind deconvolution 的訊號再接著利用 postfilter 再做一次處理。另外 Lin.X 等人[12]從極點(pole)和零點(zero)的角度來做消除迴響，除此之外，由於聲音的迴響是來自四面八方的，所以也有人使用空間濾波器來消除迴響，例如 K.Kumatani 等人[13]就利用此一特性，使用空間濾波器來消除迴響。

即使在消除迴響的研究領域中，已經有了一段很長時間的發展了，但是要能完全不失真的把聲音的迴響部分消除掉，依然是一件非常困難且有挑戰性的事情，若要達到此一目標，還是非常需要所有相關研究的研究者的努力。

1.4 論文貢獻



本研究的創新部分為相較於一般消除迴響的方法，增加了可以基於使用者的不同使用情況之下，可以調整訊號衰減率，來達到不同的目的需求。以往消除迴響的方法，主要都是以固定的形式來消除迴響，通常都無法調整輸出的結果，可能會導致無法有一個較通用的方法。

本研究首先使用麥克風陣列接收訊號，利用其空間資訊經由 Dahl 演算法的波束形成技術，先消除聲源以外方向的迴響，再接著使用可調整增益衰減後濾波器對訊號聲源方向的迴響作壓抑的動作，如此不但能提高訊號殘響比，也可以依照不同的使用需求，調整對應的參數，讓此方法能更為通用。

1.5 論文架構

本論文包含了三個主要的部分，分別為麥克風陣列技術、封閉空間迴響的介紹、論文提出的演算法與方法的實驗與分析。以下描述各章節的內容：

第二章：技術背景介紹

麥克風陣列技術、聲音迴響與迴響時間、Dahl 演算法以及後混響譜方差的介紹。

第三章：論文方法

介紹本論文的架構與演算法。利用改變訊號衰減率來估算迴響部分的能量，並依照能量的比例壓抑迴響。



第四章：實驗的結果與分析

分析實驗結果，測試論文方法的效果。

第五章：結論

對論文方法與測試結果進行評估與總結。

第二章 技術背景介紹

2.1 陣列訊號處理

在一般的訊號處理中，多半著重於如何利用訊號的時域或頻域的特徵，來對訊號進行處理。而陣列訊號處理的技術，就是除了在原先的訊號即擁有的時域或頻域資訊中，還能同時多得到空間資訊，使得訊號處理的資訊更為豐富。

陣列訊號處理，顧名思義便可知，是利用排列成線性或是環狀等特定形狀的多個感測器來接收訊號，並進行處理的技術。由於在空間中任一點聲源發出訊號後，經過在空間中的傳遞，不同感測器接收到時會產生許多差異，如接收能量不同或接收時間的延遲。利用不同感測器接收到同一聲源訊號時所產生的差異，進行處理與分析，藉此來獲得空間的資訊。

在陣列訊號處理中，依使用目的與研究方向的不同，大致可將研究領域分為兩大類：第一類著重在估測訊號的數量或是方位，此類稱為到達角估測 (Direction of arrival estimation)。第二類為利用訊號的空間資訊，對不同方向的訊號做出不同的增益，以達到空間濾波的效果，藉此來分離不同方向的訊號，這一類稱為波束形成(Beamformer)，也就是空間濾波器(Spatial Filter)。

然而，在陣列訊號處理的理論中，都是基於兩個假設，使得理論更為精簡：
第一個假設是窄頻訊號（Narrow Band Signal），而第二個假設為遠場平面波
(Far Field Plane Wave)。

假設陣列感測器的位置如圖 2-1 所示， $s(t)$ 為原始訊號， $n(t)$ 為雜訊，則 M
個感測器輸出可表示成向量形式：

$$\begin{aligned} \mathbf{z}(t) &= \begin{bmatrix} z_1(t) \\ \vdots \\ z_M(t) \end{bmatrix} = \begin{bmatrix} s(t)e^{jw\frac{\bar{z}_1 \cdot \bar{r}}{c}} \\ \vdots \\ s(t)e^{jw\frac{\bar{z}_M \cdot \bar{r}}{c}} \end{bmatrix} + \begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix} \\ &= \begin{bmatrix} e^{jk\bar{z}_1 \cdot \bar{r}} \\ \vdots \\ e^{jk\bar{z}_M \cdot \bar{r}} \end{bmatrix} s(t) + \begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix} = \mathbf{a}(\bar{r})s(t) + \mathbf{n}(t) \end{aligned} \quad (2.1.1)$$

其中， $\mathbf{a}(\bar{r})$ 稱為 array manifold vector，代表訊號聲源到各個感測器間的時間關係，可表示成：

$$\mathbf{a}^T(\theta) = [1 \quad e^{jkd\sin\theta} \quad \dots \quad e^{jk(M-1)d\sin\theta}] \quad (2.1.2)$$

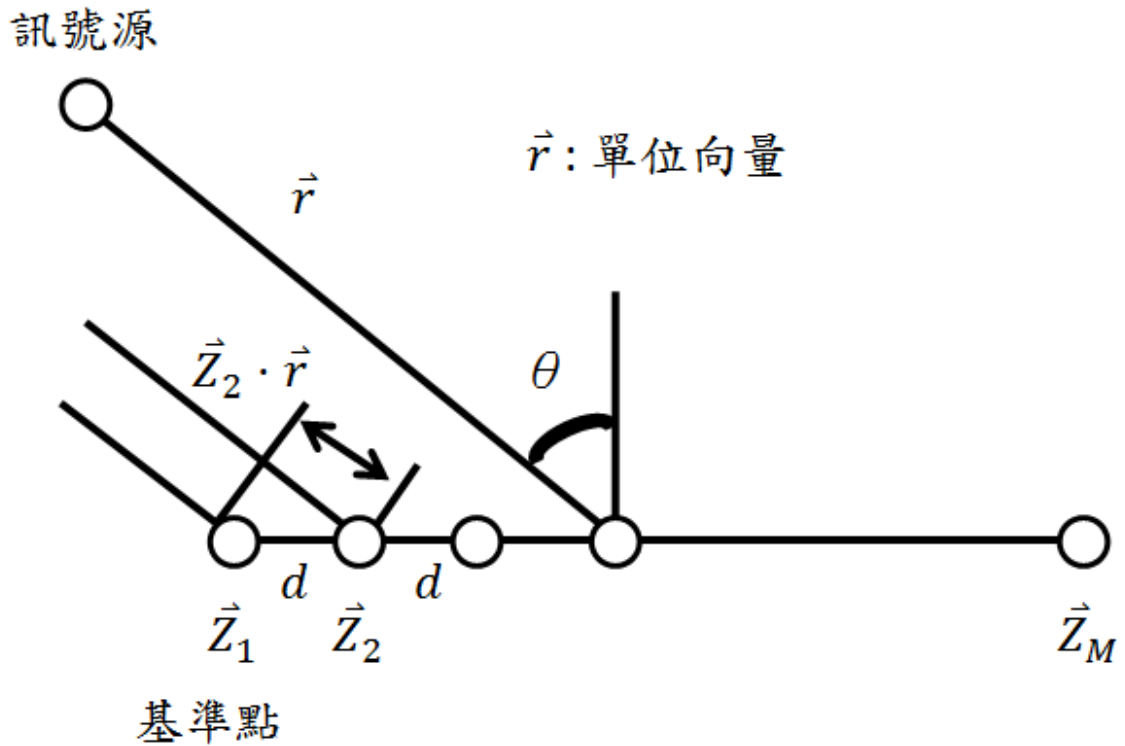


圖 2-1 均勻線性陣列架構圖

2.2 封閉空間的迴響與迴響時間

2.2.1 封閉空間的迴響

迴響是本論文的主題，並能用直觀的反射來描述。目標訊號產生聲波，此聲波會被房間內的牆壁反射，被麥克風接收。圖 2-2 是一個包含直接接收部分 (Direct Path) 和一次反射部分 (Reflection) 的例子。由於不同的路徑長度以及被牆壁吸收的聲音能量，導致聲波到達麥克風位置時，會有不同的振幅與相位。接收訊號中，聲音訊號的延遲和衰減的情形即稱為迴響。

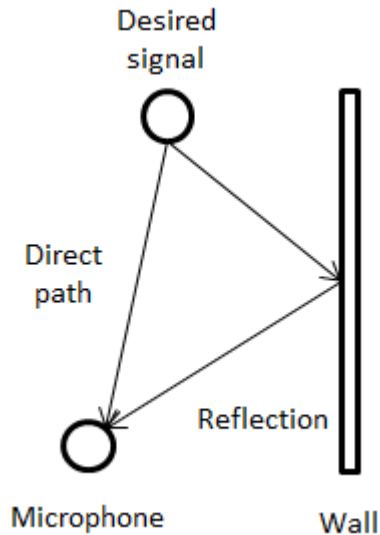


圖 2-2 聲源直接到達路徑與反射路徑

迴響是目標訊號與麥克風之間許多的傳播路徑產生的，接收訊號一般包含直接接收部分和反射部分。反射部分又分成緊接在直接接收後面的早迴響 (Early Reflection)，以及早反射之後的晚反射 (Late Reflection)，如圖 2-3 所示。

以下為詳細的定義：

直接接收聲音 (Direct Sound)： 最先接收到的聲波，此聲波沒經過任何反射即稱為直接接收聲音。此聲音會由於聲速和距離造成延遲。

早迴響 (Early Reverberation)： 接收到的時間比直接接收聲音略晚一點，且有經過反射而接收到的聲音即稱為早迴響。通常早迴響指在直接接收聲音之後 80-100 毫秒 (ms) 之內接收到的聲音。早迴響會造成頻率的失真，但是能提高語音品質。

後迴響 (Late Reverberation)： 在早迴響之後接收到的聲音都稱為後迴響。

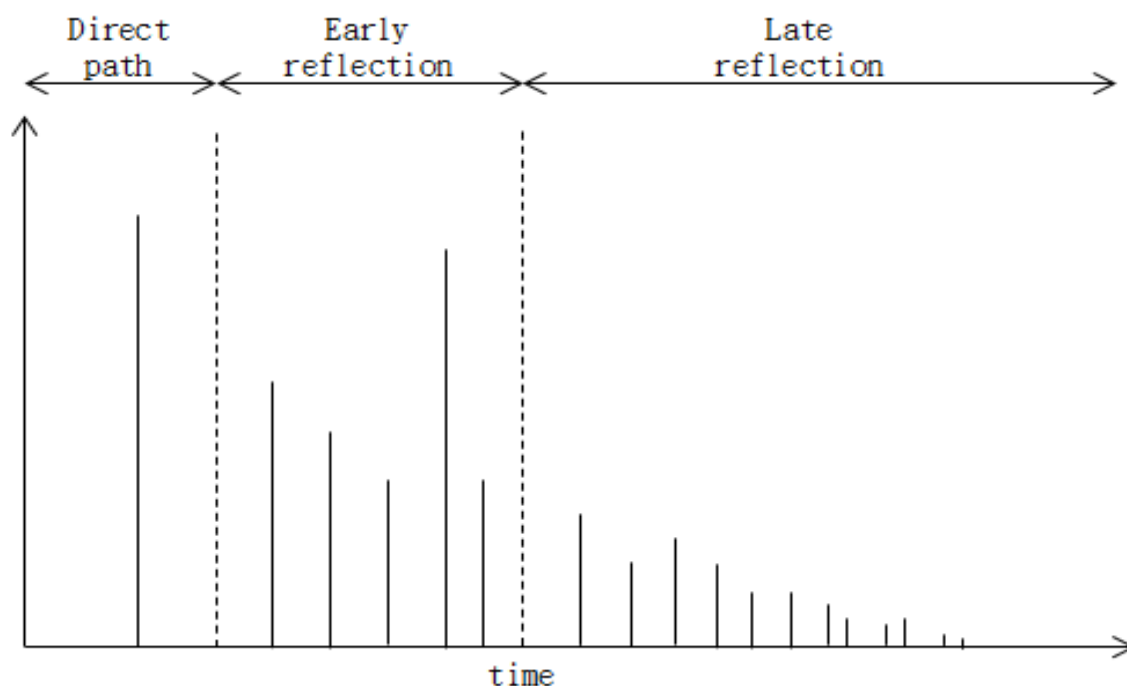


圖 2-3 聲音脈衝響應概要圖



2.2.2 迴響時間

迴響時間(Reverberation Time, RT60)在研究迴響的領域中，是一個非常重要的參數。迴響時間的定義是聲音在空間中下降 60 分貝(dB)所需要的時間。會選定 60 分貝的原因是，因為聲音下降了 60 分貝之後，人耳幾乎聽不到聲音。本論文中，估測迴響時間的方法，主要是根據 M.R.Schroeder 的方法來估算迴響時間。

此理論是根據能量下降曲線(Energy Decay Curve, EDC)下降 60 分貝來定義迴響時間，能量下降曲線可以藉由觀察房間脈衝響應的積分得到，表示成下式

$$EDC(t) = \int_t^{\infty} h^2(\tau) d\tau \quad (2.2.1)$$

$h(t)$ 為房間的脈衝響應。上式的積分通常稱為 Schroeder 積分，用來計算經過時間 t 之後剩餘的脈衝響應能量。迴響時間的計算方法如下：

$$\widehat{RT}_{60} = \frac{60}{qf_s} \quad (2.2.2)$$

q 為能量下降曲線的斜率， f_s 為取樣頻率。

2.3 Dahl 演算法

一般而言，濾波器的係數都是固定的，並不會隨著環境的變化做適當的調整。若能根據輸入的訊號，使用訊號處理的方式來讓濾波器自動調整適合當下環境的係數，即為適應性濾波器。同理，能自動調整係數的空間濾波器即為適應性空間濾波器。

本章節將介紹的適應性空間濾波器，稱為 Dahl 演算法[14]。依照適應性訊號處理的觀念，必須先得到希望達到的聲音訊號的特性，而 Dahl 演算法的訊號擷取圖如圖 2-5 所示

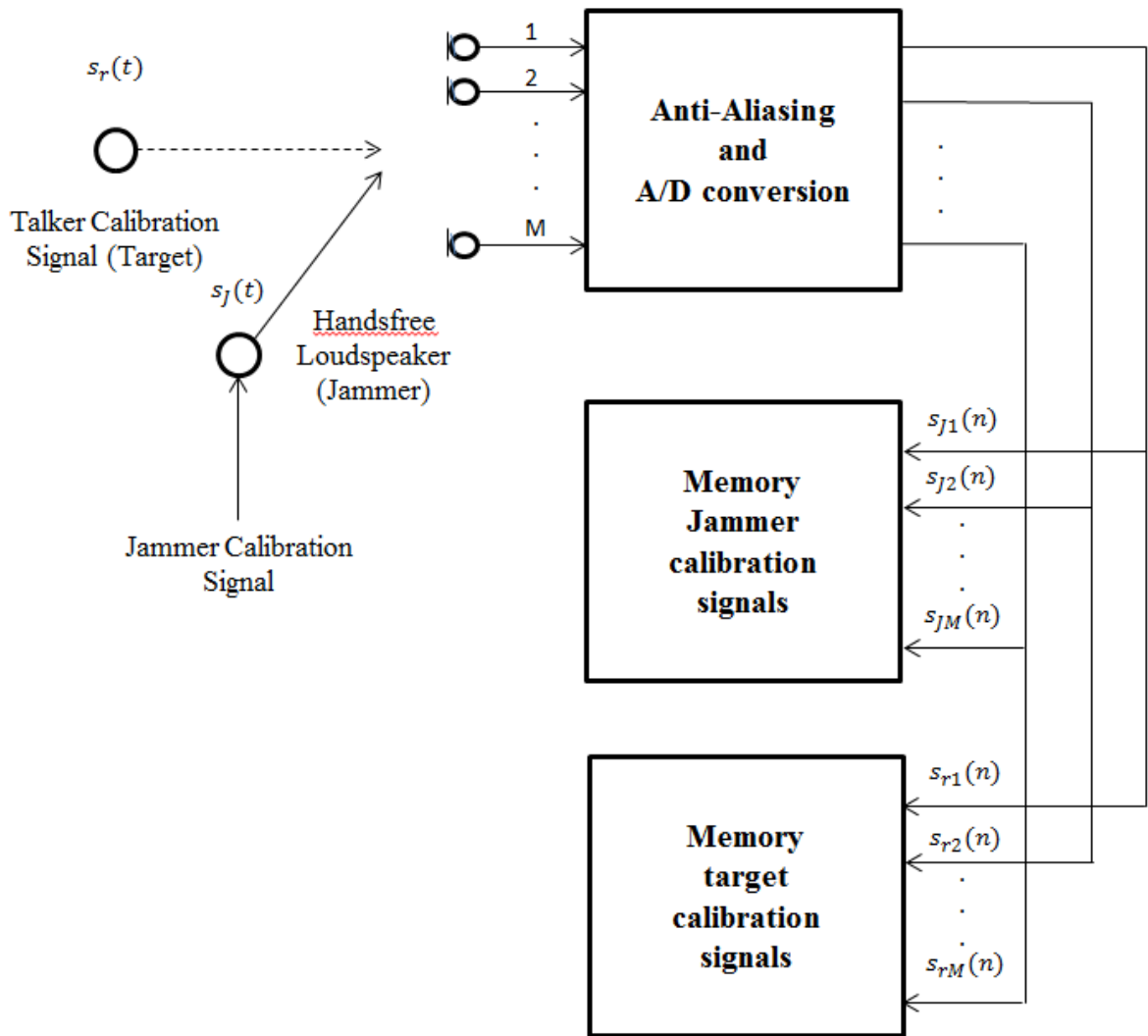


圖 2-4 Dahl 演算法訊號擷取圖

Dahl 演算法的訊號擷取圖要分成兩個部分來操作，首先利用 M 個麥克風的麥克風陣列，在安靜的環境下錄製希望達到的目標訊號，就是目標聲源方向的訊號。接下來是錄製希望消除的固定干擾聲源，亦是希望空間濾波器能夠濾掉的訊號。例如，如果環境中有人的說話聲以及喇叭播放的音樂聲，則由上述的兩個部分的操作方法來說，就是先使用麥克風陣列在安靜的環境之下錄製一段人講話的聲音，接下來一樣是在安靜的環境之下錄製一段喇叭播

放的音樂聲，如此，就完成 Dahl 演算法預錄的部分。

而 Dahl 演算法的架構圖如圖 2-6 所示，此架構可分為上下兩個部分，上面的部分是將麥克風陣列收到的訊號乘上一組空間濾波器的係數來當做輸出的訊號。下面的部分主要的功能是用來更新空間濾波器的係數，更新空間濾波器係數的方法，就是把麥克風即時收到的訊號與預錄的目標聲源以及固定干擾聲源相加，把此一相加的結果當成是 LMS 演算法的輸入，再利用 LMS 演算法來調整空間濾波器的係數，經過調整的係數會不斷的變動，一直到收斂至某一範圍，如此一來，適應性空間濾波器的輸出訊號就會與目標訊號的誤差達到最小，也就是指空間濾波器在目標訊號方向的增益最大，在固定干擾聲源方向的增益會被降低，如此就能達到濾除固定干擾聲源的效果。

在 Dahl 演算法中，上半部分的空間濾波器的訊號處理與下半部分的空間濾波器的係數調適不可同時進行，若在空間濾波器處理訊號的過程中，干擾聲源移動導致方向改變，則必須重新啟動下半部分的空間濾波器係數調適的功能，來調整出一組適合當下情況的係數。

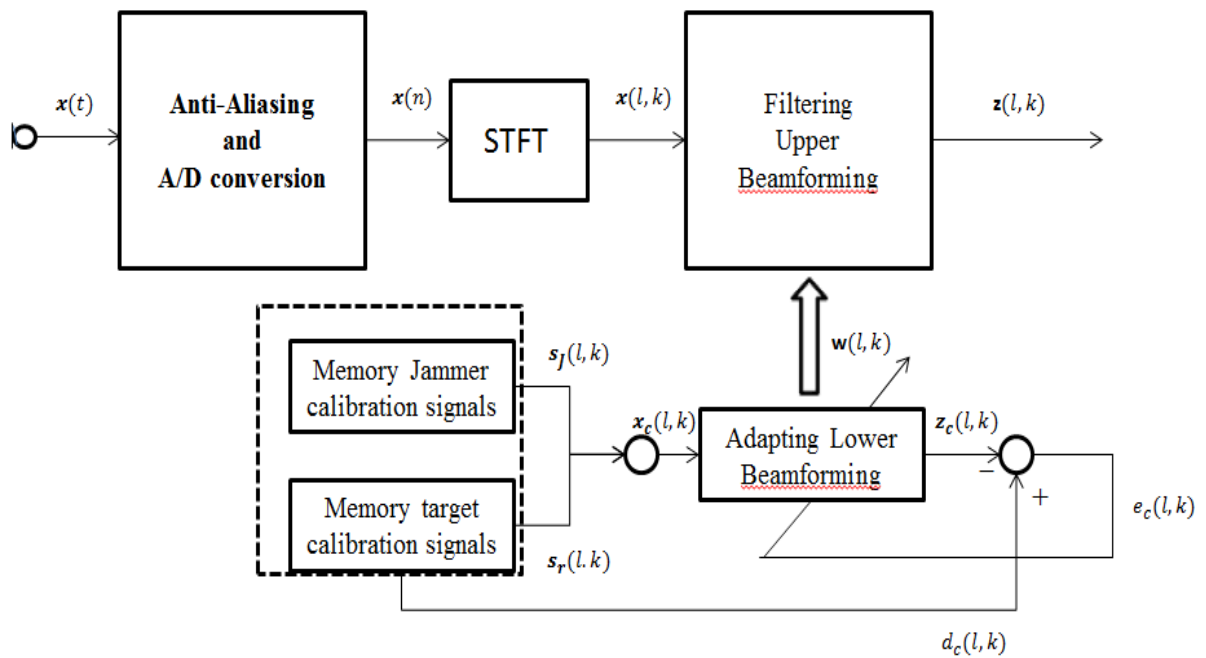


圖 2-5 Dahl 演算法架構圖

2.4 後混響譜方差(Late Reverberant Spectral Variance, LRSV) 估計

本章節將介紹利用 J.D.Polack 的迴響統計模型來估計迴響部分的能量。

以下將詳細介紹迴響統計模型的內容：

迴響訊號的結果，是聲音訊號 $s(t)$ 和因果非時變的聲音脈衝響應(Acoustic Impulse Response, AIR)迴旋積分(convolution)的產物，將聲音脈衝響應函式表示成 $h(t)$ ，可得

$$z(t) = \int_{-\infty}^t s(\theta) h(t-\theta) d\theta \quad (2.4.1)$$

因為我們的目標是抑制後迴響，所以聲音脈衝響應可以分成兩部分 $h_e(t)$ 和 $h_l(t)$ ，因此可寫成

$$h(t) = \begin{cases} h_e(t), & \text{for } 0 \leq t \leq T_l \\ h_l(t), & \text{for } t \geq T_l \\ 0, & \text{otherwise} \end{cases} \quad (2.4.2)$$

T_l 的選擇，是要使得 $h_e(t)$ 包含了直接路徑(Direct path)和一些的早反射(Early reflection)，通常的選擇範圍在 40-80 毫秒(ms)之間。

由(2.4.1)和(2.4.2)式可寫成

$$z(t) = \int_{-\infty}^t s(\theta) h_e(t-\theta) d\theta + \int_{-\infty}^t s(\theta) h_l(t-\theta) d\theta \quad (2.4.3)$$

此理論的主要目的就是要估測 LRSV $\lambda_{z_l}(l, k)$ 。

$$\lambda_{z_l}(l, k) = E\{|Z_l(l, k)|^2\}$$

其中 $Z_l(l, k)$ 是 $z_l(t)$ 的短時間傅立葉轉換(Discrete Short Time Fourier Transform, STFT)。

因為在實際的房間中，聲音的行為非常複雜，以致於很難模擬，所以通常都會使用統計房間聲學(Statistical Room Acoustics, SRA)來模擬。統計房間聲學通常利用聲源與麥克風之間的距離、房間體積和迴響時間等來描述聲源到麥克風之間的轉移函數。

以下為此模型的三個假設：

1. 房間的大小要遠大於波長。
2. 平均的共振頻率間隔要小於頻寬的三分之一。假設一個房間的體積是 V 立方公尺，迴響時間是 RT_{60} 秒，這些情況下的頻率要大於 Schroeder 頻率：

$$f_g = 2000 \sqrt{RT_{60} / V} \text{。}$$

3. 聲源、麥克風與牆壁的距離，至少要大於半波長。

Polack 利用非穩態隨機程序來描述聲音脈衝響應(Acoustic Impulse Response, AIR)。

$$h(t) = \begin{cases} b(t)e^{-\bar{\delta}t}, & \text{for } t \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.4.5)$$

$b(t)$ 是平均為零的高斯穩態白噪音訊號，而平均阻尼常數(damping constant) $\bar{\delta}$ 則是關係著迴響時間，2.4.6 式為平均阻尼常數與迴響時間的關係式：

$$\bar{\delta} = \frac{3 \log_e(10)}{RT_{60}} \quad (2.4.6)$$

此段落將介紹廣義的迴響統計模型(Generalized Statistical Reverberation Model)。

Polack 的統計模型在直達混響聲能比(Direct to Reverberation Ratio, DRR)小於 0 分貝(dB)時對於後混響譜方差的估測式很有幫助的。但是在直達混響聲能比大於 0 分貝(dB)的時候，會造成後混響譜方差過估測(Overestimate)的現象，這將會導致消除迴響後的訊號嚴重的失真。與 Polack 的統計模型相同，將聲音脈衝響應 $h(t)$ 拆解成 $h_d(t)$ 和 $h_r(t)$ 兩部分：

$$h(t) = \begin{cases} h_d(t), & \text{for } 0 \leq t < T_r \\ h_r(t), & \text{for } t \geq T_r \\ 0, & \text{otherwise} \end{cases} \quad (2.4.7)$$

T_r 的選擇，是要使得 $h_d(t)$ 包含直接路徑(Direct path)， $h_r(t)$ 包含直接路徑之後的所有反射。稍後會根據時頻(Time-frequency)轉換的音框率(Frame rate)定義 T_r 。為了能在統計模型中模擬出能量與直接路徑的關係，提出了以下的模型：

$$h_d(t) = \begin{cases} b_d(t)e^{-\bar{\delta}t}, & \text{for } 0 \leq t \leq T_r \\ 0, & \text{otherwise} \end{cases} \quad (2.4.8)$$

$b_d(t)$ 是平均為零的高斯穩態白噪音訊號，而平均阻尼常數(damping constant) $\bar{\delta}$ 則是關係著迴響時間。在迴響成分(Reverberant component) $h_r(t)$ 部分，則用以下模型來描述：

$$h_r(t) = \begin{cases} b_r(t)e^{-\bar{\delta}t}, & \text{for } t \geq T_r \\ 0, & \text{otherwise} \end{cases} \quad (2.4.9)$$

$b_r(t)$ 是平均為零的高斯穩態白噪音訊號，而平均阻尼常數(damping constant) $\bar{\delta}$ 則是關係著迴響時間。

在統計房間聲學的情況下，直接路徑部分與反射路徑部分的聲音脈衝響應是不相關的(Uncorrelate)，基於這個原因，更進一步假設 $b_d(t)$ 與 $b_r(t)$ 是不相關的，

$$E\{b_d(t)b_r(t+\tau)\} = 0 \quad \circ$$

$h(t)$ 的能量包絡線(Energy envelope)可以表示成

$$E_h\{h^2(t)\} = \begin{cases} \sigma_d^2 e^{-2\bar{\delta}t}, & \text{for } 0 \leq t < T_r \\ \sigma_r^2 e^{-2\bar{\delta}t}, & \text{for } t \geq T_r \\ 0, & \text{otherwise} \end{cases} \quad (2.4.10)$$

σ_d^2 與 σ_r^2 分別是 $b_d(t)$ 與 $b_r(t)$ 的變異數(Variance)。

以下將詳細介紹後混響譜方差(Late Reverberant Spectral Variance, LRSV)的估測。

麥克風收到的訊號，可以表示成

$$\begin{aligned} z(t) &= \int_{-\infty}^t s(\theta)h(t-\theta)d\theta \\ &= \int_{t-T_r}^t s(\theta)h_d(t-\theta)d\theta + \int_{-\infty}^{t-T_r} s(\theta)h_r(t-\theta)d\theta \end{aligned} \quad (2.4.11)$$

固定聲源與麥克風的設定，迴響訊號在時間 t 和延遲 τ 的自相關運算可定義為

$$r_{zz}(t, t+\tau) = E_z\{z(t)z(t+\tau)\} \quad (2.4.12)$$

在固定 h 的條件下，可得

$$\begin{aligned} r_{zz}(t, t+\tau; h) &= \int_{t-T_r}^t \int_{t-T_r+\tau}^{t+\tau} E_s\{s(\theta)s(\theta')\} h_d(t-\theta)h_d(t+\tau-\theta') d\theta d\theta' \\ &+ \int_{-\infty}^{t-T_r} \int_{-\infty}^{t-T_r+\tau} E_s\{s(\theta)s(\theta')\} h_r(t-\theta)h_r(t+\tau-\theta') d\theta d\theta' \end{aligned} \quad (2.4.13)$$

根據 2.4.8、2.4.9、2.4.10 三式，可得

$$E_h\{h_d(t-\theta)h_d(t+\tau-\theta')\} = \sigma_d^2 e^{-2\bar{\delta}t} e^{\bar{\delta}(\theta+\theta'-\tau)} \delta(\theta-\theta'+\tau)$$

$$E_h\{h_r(t-\theta)h_r(t+\tau-\theta')\} = \sigma_r^2 e^{-2\bar{\delta}t} e^{\bar{\delta}(\theta+\theta'-\tau)} \delta(\theta-\theta'+\tau)$$

因為隨機程序 h 和 s 之間沒有實質關係，所以可以假設兩者為統計上獨立。

$E\{b_d(t)b_r(t+\tau)\} = 0$ 等同於 $E\{h_d(t)h_r(t+\tau)\} = 0$ 。則空間上的平均自相關函式為

$$r_{zz}(t, t+\tau) = E_h\{r_{zz}(t, t+\tau; h)\} = r_{z_d z_d}(t, t+\tau) + r_{z_r z_r}(t, t+\tau) \quad (2.4.14)$$

其中

$$r_{z_d z_d}(t, t+\tau) = \sigma_d^2 e^{-2\bar{\delta}t} \int_{t-T_r}^t E\{s(\theta)s(\theta+\tau)\} e^{2\bar{\delta}\theta} d\theta \quad (2.4.15)$$

與

$$\begin{aligned} r_{z_r z_r}(t, t+\tau) &= \sigma_r^2 e^{-2\bar{\delta}t} \int_{-\infty}^{t-T_r} E\{s(\theta)s(\theta+\tau)\} e^{2\bar{\delta}\theta} d\theta \\ &= \sigma_r^2 e^{-2\bar{\delta}t} \int_{t-2T_r}^{t-T_r} E\{s(\theta)s(\theta+\tau)\} e^{2\bar{\delta}\theta} d\theta \\ &+ \sigma_r^2 e^{-2\bar{\delta}t} \int_{-\infty}^{t-2T_r} E\{s(\theta)s(\theta+\tau)\} e^{2\bar{\delta}\theta} d\theta \end{aligned} \quad (2.4.16)$$

根據 2.4.14 式，第一項與直接路徑的訊號相關，作用時間在 $t-T_l$ 到 t 之間。

第二項則與迴響訊號相關，是重疊遮蔽的主要原因。首先考慮在時間 $t - T_l$ 的自相關函式

$$r_{zz}(t - T_r, t - T_r + \tau) = r_{z_d z_d}(t - T_r, t - T_r + \tau) + r_{z_r z_r}(t - T_r, t - T_r + \tau) \quad (2.4.17)$$

其中

$$r_{z_d z_d}(t - T_r, t - T_r + \tau) = \sigma_d^2 e^{-2\bar{\delta}(t - T_r)} \int_{t - 2T_r}^{t - T_r} E\{s(\theta)s(\theta + \tau)\} e^{2\bar{\delta}\theta} d\theta \quad (2.4.18)$$

與

$$r_{z_r z_r}(t - T_r, t - T_r + \tau) = \sigma_r^2 e^{-2\bar{\delta}(t - T_r)} \int_{-\infty}^{t - 2T_r} E\{s(\theta)s(\theta + \tau)\} e^{2\bar{\delta}\theta} d\theta \quad (2.4.19)$$

根據 2.4.16 式， $r_{z_r z_r}(t, t + \tau)$ 項可表示為

$$r_{z_r z_r}(t, t + \tau) = \kappa e^{-2\bar{\delta}T_r} r_{z_d z_d}(t - T_r, t - T_r + \tau) + e^{-2\bar{\delta}T_r} r_{z_r z_r}(t - T_r, t - T_r + \tau) \quad (2.4.20)$$

$\kappa = \sigma_r^2 / \sigma_d^2$ ，這邊因為假設 $\sigma_d^2 \geq \sigma_r^2$ ，所以 $\kappa \leq 1$ 。依照 2.4.17 式，2.4.20 可以重新寫成

$$r_{z_r z_r}(t, t + \tau) = e^{-2\bar{\delta}T_r} (1 - \kappa) r_{z_r z_r}(t - T_r, t - T_r + \tau) + \kappa e^{-2\bar{\delta}T_r} r_{zz}(t - T_r, t - T_r + \tau) \quad (2.4.21)$$

後迴響成分則可以利用觀察下式得到

$$r_{z_l z_l}(t, t + \tau) = e^{-2\bar{\delta}(T_l - T_r)} r_{z_r z_r}(t - T_l + T_r, t - T_l + T_r + \tau) \quad (2.4.22)$$

若利用短時間內的譜密度(Power Spectral Density, PSD)的觀點觀察，2.4.22 以及 2.4.21 式分別可寫成

$$P_{z_l z_l}(t, f) = e^{-2\bar{\delta}(T_l - T_r)} P_{z_r z_r}(t - T_l + T_r, f) \quad (2.4.23)$$

$$P_{z_r z_r}(t, f) = e^{-2\bar{\delta}T_r} (1 - \kappa) P_{z_r z_r}(t - T_r, f) + \kappa e^{-2\bar{\delta}T_r} P_{zz}(t - T_r, f) \quad (2.4.24)$$

在此定義 $T_r f_s = R$ ， R 為短時傅立葉轉換(Short Time Fourier Transfer, STFT)的音框率(Frame rate)，從短時傅立葉轉換的觀點，2.4.23 式可寫成

$$\lambda_{z_l}(l, k) = e^{-2\bar{\delta}\left(T_r - \frac{R}{f_s}\right)} \lambda_{z_r}(l - N_l + 1, k) \quad (2.4.25)$$

$$\lambda_{z_r}(l, k) = e^{-2\bar{\delta}\frac{R}{f_s}} (1 - \kappa) \lambda_{z_r}(l - 1, k) + \kappa e^{-2\bar{\delta}\frac{R}{f_s}} \lambda_{z_z}(l - 1, k) \quad (2.4.26)$$

若能得到 $\hat{\lambda}_z(l, k)$ 、 $\hat{\kappa}$ 和 $\hat{\delta}$ 這些參數，就能利用 2.4.25 以及 2.4.26 二式計算出後混響譜方差。 $\hat{\lambda}_z(l, k)$ 與 $\hat{\delta}$ 在前面已經介紹過了，接下來的部分為介紹直達混響聲能比(Direct to Reverberation Ratio, DRR) κ 的定義：

$$\kappa = \frac{\sigma_r^2}{\sigma_d^2} = \frac{1 - e^{-2\bar{\delta}T_r}}{e^{-2\bar{\delta}T_r}} \frac{E_r}{E_d} \quad (2.4.27)$$

其中

$$\frac{E_d}{E_r} = \frac{\int_0^{T_r} h^2(t) dt}{\int_{T_r}^{\infty} h^2(t) dt} \quad (2.4.28)$$

在實際環境應用下，聲音脈衝響應都不是事前可先知道的訊息，所以我們只能靠盲估計(Blind estimate)來得到直達混響聲能比。

在很多實際的情況中，麥克風跟訊號源的距離是會變動的。直達混響聲能比是依據在麥克風與訊號源之間的距離改變的，所以估測直達混響聲能比是要能根據不同的距離而改變的。當 κ 太大的時候， $\hat{\lambda}_z(l, k)$ 會大於 $|Z(l, k)|^2$ ，會導致過估測的產生，所以應該要讓 κ 值降低。在無噪音的環境中，迴響部分

的能量在有語音訊號的時候會小於接收訊號的能量，在無語音訊號的情況下會大致等於接收訊號的能量。所以 κ 至少需滿足 $|Z(l,k)|^2 - \hat{\lambda}_z(l,k) \geq 0$ 。

當語音訊號被偵測到且滿足 $|Z(l,k)|^2 - \hat{\lambda}_z(l,k) < 0$ 時， κ 值會降低。當 $|Z(l,k)|^2 - \hat{\lambda}_z(l,k) > 0$ 時， κ 值則緩慢上升。若 $|Z(l,k)|^2 - \hat{\lambda}_z(l,k) = 0$ ，就假設 κ 值式正確的，下式為 κ 更新的機制

$$\hat{\kappa}(l+1) = \begin{cases} \hat{\kappa}(l) + \mu_k \left(1 - \frac{\sum_{k=0}^{\frac{K}{2}-1} \hat{\lambda}_z(l,k)}{\sum_{k=0}^{\frac{K}{2}-1} |Z(l,k)|^2} \right), & \text{speech present} \\ \hat{\kappa}(l), & \text{otherwise} \end{cases} \quad (2.4.29)$$

μ_k 為步長(Step size)，範圍在 1 到 0 之間。在 $\hat{\kappa}$ 更新時，也將範圍限制在 1 到 0 之間。



第三章 論文方法

一般利用迴響時間估測後迴響譜方差來抑制迴響的理論，都是使用一些方法估測迴響時間，所以通常只要估測出迴響時間之後，其消除迴響的效果就大致上固定了。本論文提出的方法，即是能依照使用者的需求來改變迴響時間的方法，如此可以適合在不同需求下的應用。

3.1 系統架構說明

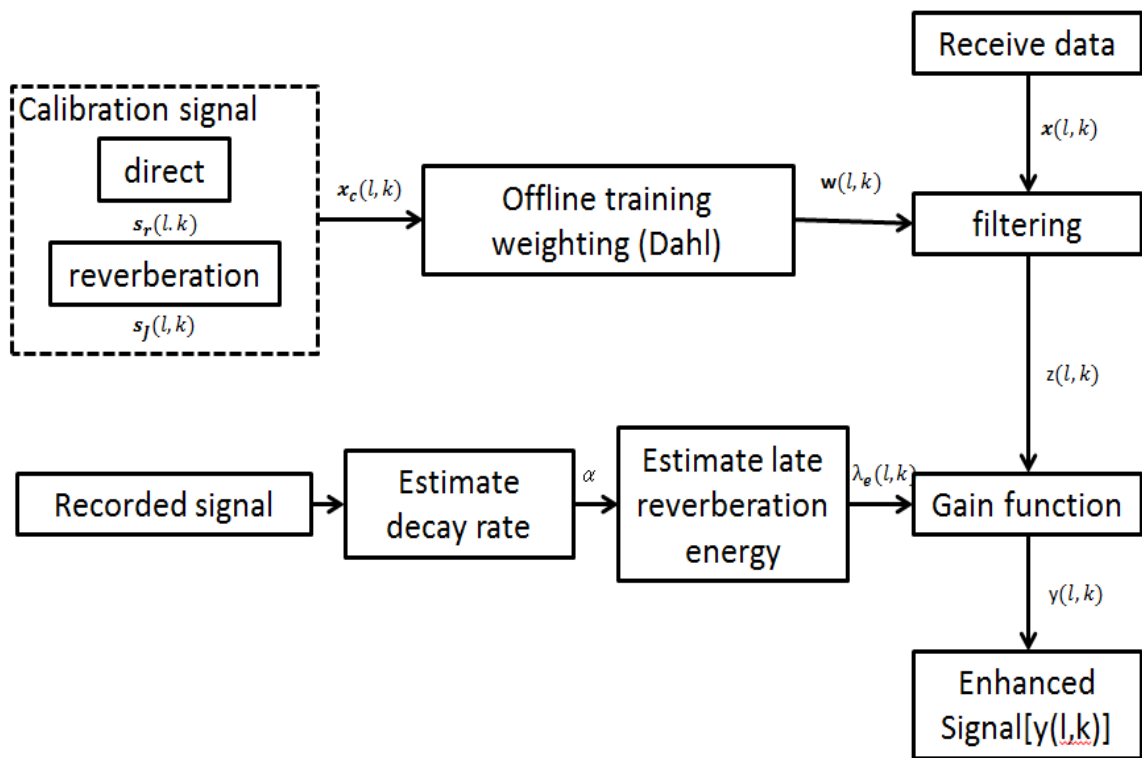


圖 3-1 系統架構

系統架構演算法說明：

1. 將麥克風陣列接收的聲音訊號，乘上一組事先訓練(Training)好波束形成的權重(Weight)當作第一階段的輸出。

2. 使用第一階段的輸出與事先調整好的訊號衰減率，估測出後迴響譜方差。
3. 利用後迴響譜方差運算出的增益函式，壓抑第一階段輸出訊號的迴響，當作最後的輸出訊號。

在本論文的系統架構中，主要分成兩個部分。第一個部分為波束形成，這個部分主要是利用波束形成可以壓抑方向聲源以外訊號的特性，來濾除聲源方向以外的迴響，此部分我們選擇了需要先預錄的 Dahl 演算法，詳細的內容以及步驟將會在後面的章節介紹。第二個部分是利用估測後迴響能量的方式，來得到後置濾波器，以用來消除迴響。這個部分我們在估測訊號衰減率的步驟中，加入了變數 δ 來讓使用者可以依照各自不同的條件，來選擇適合的 δ 來達到較好的消除迴響效果。在估測訊號衰減率的部分加入可供調整的變數，可以讓本方法更適合在各種情況下消除迴響，且能得到較好消除迴響的效果。

3.2 波束形成的比較與選擇

波束形成的方法有很多種，以下將列舉較常見的波束形成方法比較。

以下將列舉比較的方法：

1. 時間延遲相加波束(Delay and Sum Beamformer)[15]
2. Transfer Function Generalized Sidelobe Canceler(TF-GSC)[16]

3. Dahl 演算法[14]

4. 參考訊號之卡曼濾波器(Reference Signal Kalman filter)[17-18]

首先介紹評斷標準：訊號殘響比(Signal to Reverberation Ratio, SRR)與對數譜失真(Log Spectral Distortion, LSD)。

訊號殘響比，第 l 個音框的瞬時部分訊號殘響比表示式為

$$SRR_{seg}(l) = 10 \log_{10} \left(\frac{\sum_{n=lR}^{lR+N-1} z_d^2(n)}{\sum_{n=lR}^{lR+N-1} (z_d(n) - \hat{z}_d(n))^2} \right) [dB] \quad (3.2.1)$$

N 為音框的長度， R 為音框率(Frame rate)， $z_d(n)$ 為直達訊號(Direct Signal)， $\hat{z}_d(n)$ 為純化後的訊號。

對數譜失真可表示為

$$LSD(l) = \left(\frac{2}{K} \sum_{k=0}^{\frac{K}{2}-1} \left| \mathcal{L}\{\hat{Z}_d(l, k)\} - \mathcal{L}\{Z_d(l, k)\} \right|^p \right)^{\frac{1}{p}} [dB] \quad (3.2.2)$$

$$\mathcal{L}\{X(l, k)\} \triangleq \max\{20 \log_{10}(|X(l, k)|), \delta\} \quad (3.2.3)$$

$$\delta = \max_{l, k} \{20 \log_{10}(|X(l, k)|)\} - 50 \quad (3.2.4)$$

3.2.2 式中的 p 通常會選擇 1、2 或是 ∞ ， $Z_d(l, k)$ 與 $\hat{Z}_d(l, k)$ 分別為 $z_d(n)$ 與 $\hat{z}_d(n)$ 經過短時間傅立葉轉換後的結果。3.2.3 式的目的是為了將對數譜失真限制在 50 分貝的範圍內。

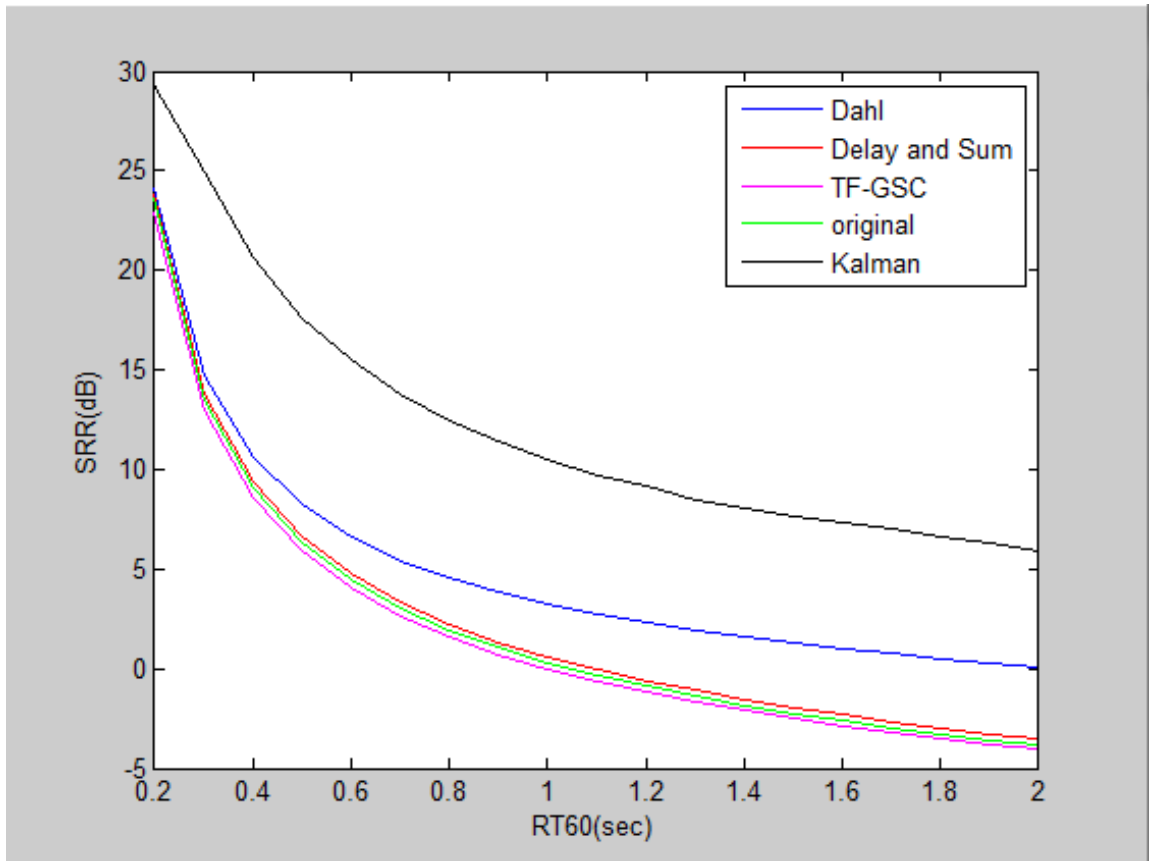


圖 3-2 波束形成之訊號殘響比



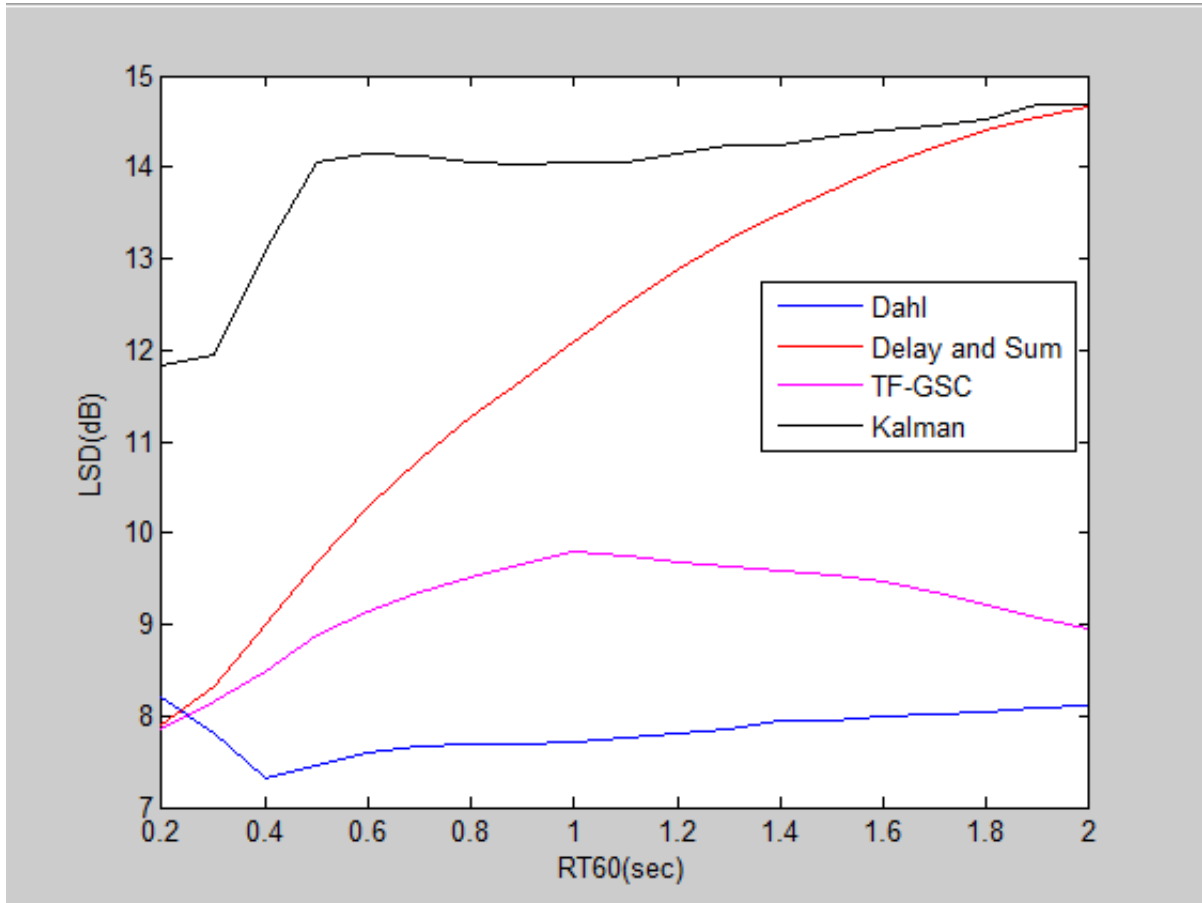


圖 3-3 波束形成之對數譜失真

其中，由於參考訊號之卡曼濾波器有參數 σ_v 與 ρ_v 需要調整，在此訂定知覺語音評價(Perceptual Evaluation of Speech Quality, PESQ)[19-20]為基準來調整 σ_v 與 ρ_v 兩個參數。知覺語音評價的分數5分為最高，表示與音品質極佳，分數越低表示與音品質越差，如表 3-1 所示。主要是選定參考訊號之卡曼濾波器的知覺語音評價與 Dahl 演算法的知覺語音評價最接近的一組 σ_v 與 ρ_v 。選擇知覺語音評價當作比較基準的原因，是聲音的迴響對知覺語音評價的影響不大，而知覺語音評價對於迴響以外的失真較為敏感[7]。而在 Dahl 演算法 TF-GSC 的部分，則是錄製一段足夠長的參考訊號，使這兩種方法可以讓其權重訓練至收斂為止。

Listening-Quality Scale :	
Quality of the speech/connection	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

表 3-1 知覺語音評價分數表

此段落將詳細介紹上述波束形成的比較。根據圖 3-2 可以清楚的發現，在經過第一階段的波束形成器之後，訊號殘響比由大至小順序為：參考訊號之卡曼濾波器、Dahl 演算法、時間延遲相加波束，最小為 TF-GSC。根據圖 3-3，也可以發現對數譜失真由大至小順序為：參考訊號之卡曼濾波器、時間延遲相加波束、TF-GSC，最小則是 Dahl 演算法。

經由圖 3-2、圖 3-3 的觀察，能得知除了 Dahl 演算法知外，其餘波束形成表現出訊號殘響比越大，對數譜失真也會越大。以下就需要事先訓練權重的 TF-GSC、Dahl 演算法以及參考訊號之卡曼濾波器分別介紹影響評比標準結果的原因：

1. TF-GSC :

TF-GSC 訓練權重的部分，主要是訓練出麥克風陣列中麥克風的相對關係，所以是以其中一顆麥克風收到的訊號為參考訊號，藉此得到其他麥克風收到的訊號與參考訊號的相對關係，因此在訓練的過程中，也會將訊號聲源到參考訊號麥克風之間的聲音訊號響應也訓練進去，所以使得 TF-GSC 消

除迴響的效果不彰，導致其訊號殘響比結果是最差的，在對數譜失真部分，由於上述的原因，可推論造成對數譜失真的主要原因為迴響。

2. Dahl 演算法：

由章節 2.3 敘述的訓練方法，首先固定麥克風陣列與訊號聲源的架構，例如麥克風與訊號聲源之間的距離、麥克風之間的距離等。以此固定架構在無響室錄製一段語音訊號，作為麥克風訊號接收的直接路徑的部分，接下來在聽得出迴響效果的室內環境，錄製一段與無響室中一樣的語音訊號，作為麥克風接收到的輸入訊號使用。接著根據 Dahl 演算法中適應性空間濾波器係數調整的部分訓練出一組權重，接著將輸入訊號與此權重相乘，即可得到第一階段的輸出。而由於訓練的過程中，把迴響部分當成是雜訊，而無響室錄製的訊號當成是目標訊號，所以能消除較多迴響部分的訊號，且造成比較少的失真。

3. 參考訊號之卡曼濾波器：

在參考訊號之卡曼濾波器調整 σ_v 與 ρ_v 兩個參數的過程中，可以發現 ρ_v 越大，則訊號的低頻部分會消除越多，由於一般麥克風陣列接收到的訊號都是低頻部分的能量較強，所以如果低頻部分消除過多，會使語音訊號嚴重失真，如圖 3-4、圖 3-5 以及圖 3-6 所示。由圖 3-4、圖 3-5 與圖 3-6 可知， ρ_v 越大，雖然可以消除較多的迴響，但是也會因為消除過多的低頻訊號導致嚴重的失真。

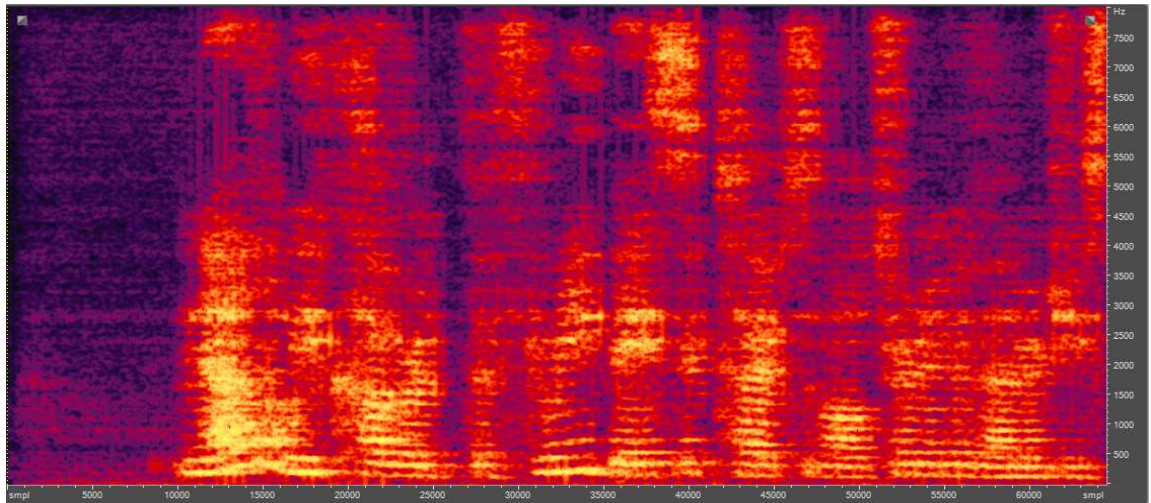


圖 3-4 $\sigma_v=100, \rho_v=0.001$ 卡曼濾波器頻譜圖

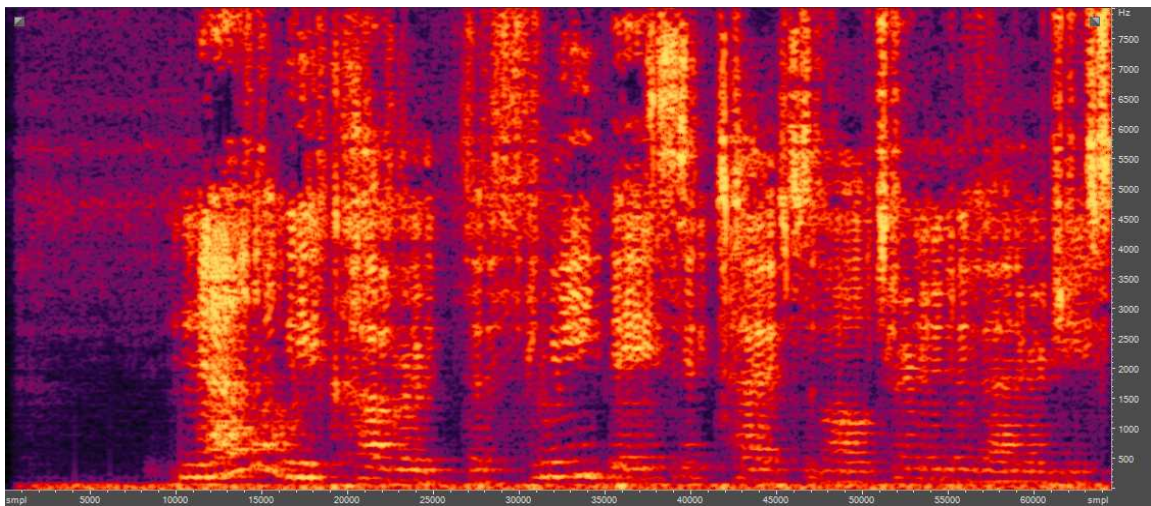


圖 3-5 $\sigma_v=100, \rho_v=10$ 卡曼濾波器頻譜圖

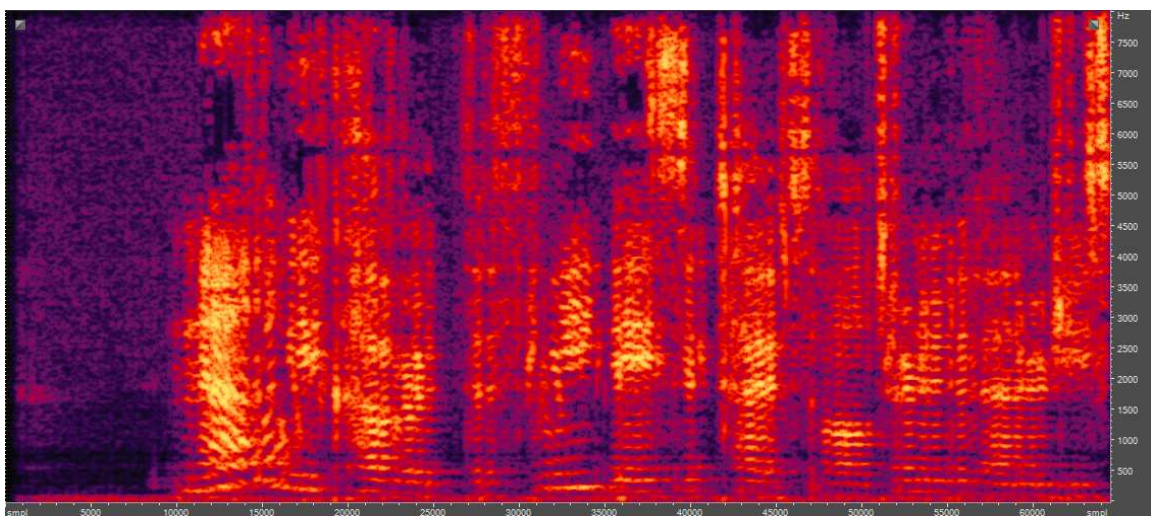


圖 3-6 $\sigma_v=100, \rho_v=1000$ 卡曼濾波器頻譜圖

根據上述的原因，加上考慮到在波束形成之後，會使用後置濾波器，依照迴響訊號的能量比再壓抑迴響一次，為了避免最後輸出的語音失真太嚴重，所以選擇訊號殘響比第二大，而對數譜失真最小的 Dahl 演算法。

3.3 利用可改變的訊號衰減率消除迴響

首先介紹如何估測訊號衰減率，此部分需要在空間中播放聲音，使聲音充滿整個空間，接著關掉聲源，而訊號的衰減率，可以由關掉聲源之後的訊號能量衰減曲線觀察得到。訊號的衰減狀況如圖 3-7 所示。

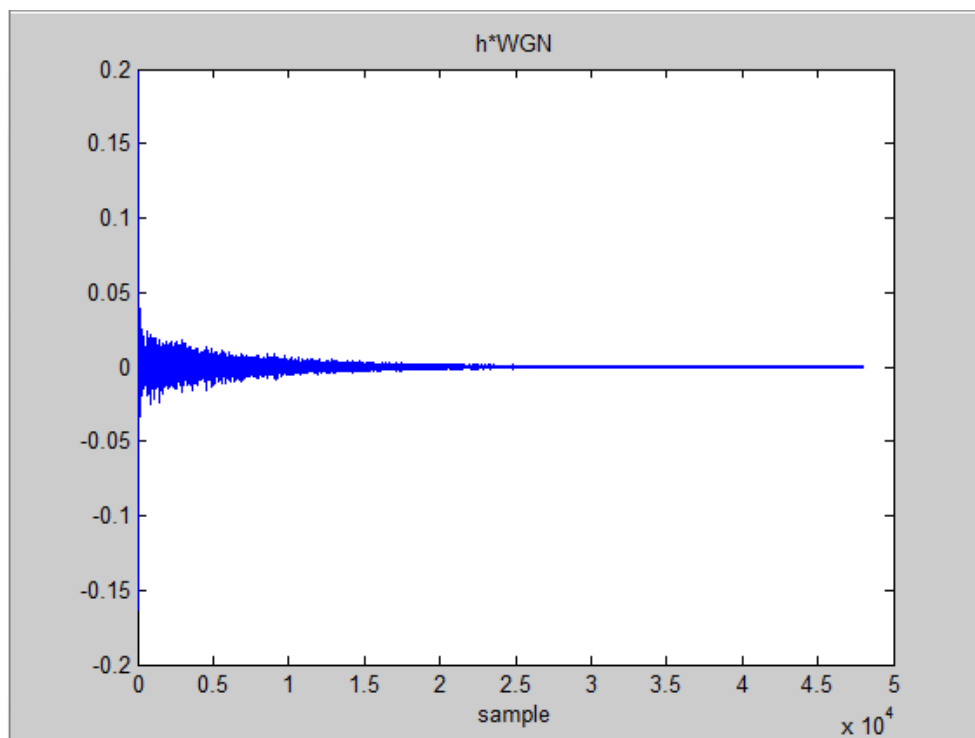


圖 3-7 訊號衰減圖

完成上述的訊號衰減預錄之後，可利用 2.2.1 式計算出訊號能量衰減的情況，並能得到如圖 3-8 的能量衰減曲線。

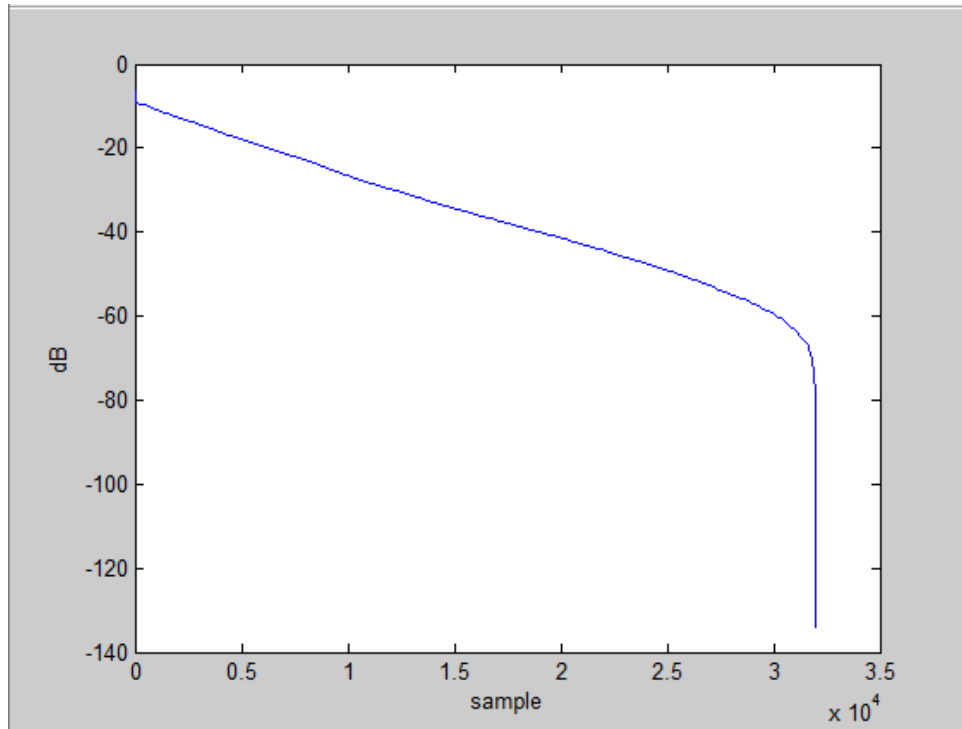


圖 3-8 能量衰減曲線

由圖 3-8 可觀察中間部分的衰減率幾乎等於定值，因此利用最小平方法 (Least Square, LS) 逼近能量衰減曲線，最小平方法定義如下：

$$J(p, q) = \min (EDC(n) - (p + \delta qn))^2 \quad (3.3.1)$$

P 為逼近線的位移，q 為逼近線的斜率， δ 為可調整的變數，調整方式稍後會介紹，能量衰減曲線與逼近線如圖 3-9 所示。

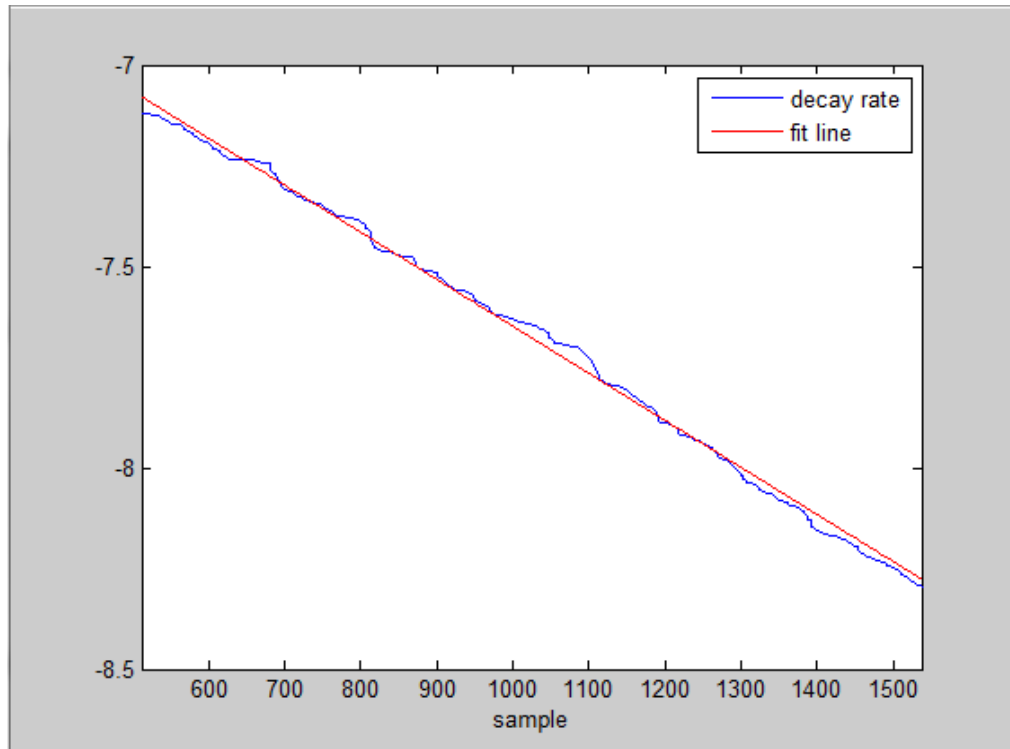


圖 3-9 最小平方法逼近曲線圖

得到衰減斜率 q 之後，可利用 2.2.2 式計算出迴響時間。將估算的迴響時間代入 2.4.25 式與 2.4.26 式，可求得後混響譜方差，再依照下式即可計算出一組後置濾波器。

$$G(l, k) = \frac{\lambda_{zd}}{\lambda_z} \quad (3.3.2)$$

λ_z 為麥克風接收訊號的能量， λ_{zd} 為直接路徑部分的訊號能量。改變 3.3.1 式中的 δ 所得到的後置濾波器的響應效果如圖 3-10、圖 3-11、圖 3-12、圖 3-13 所示。

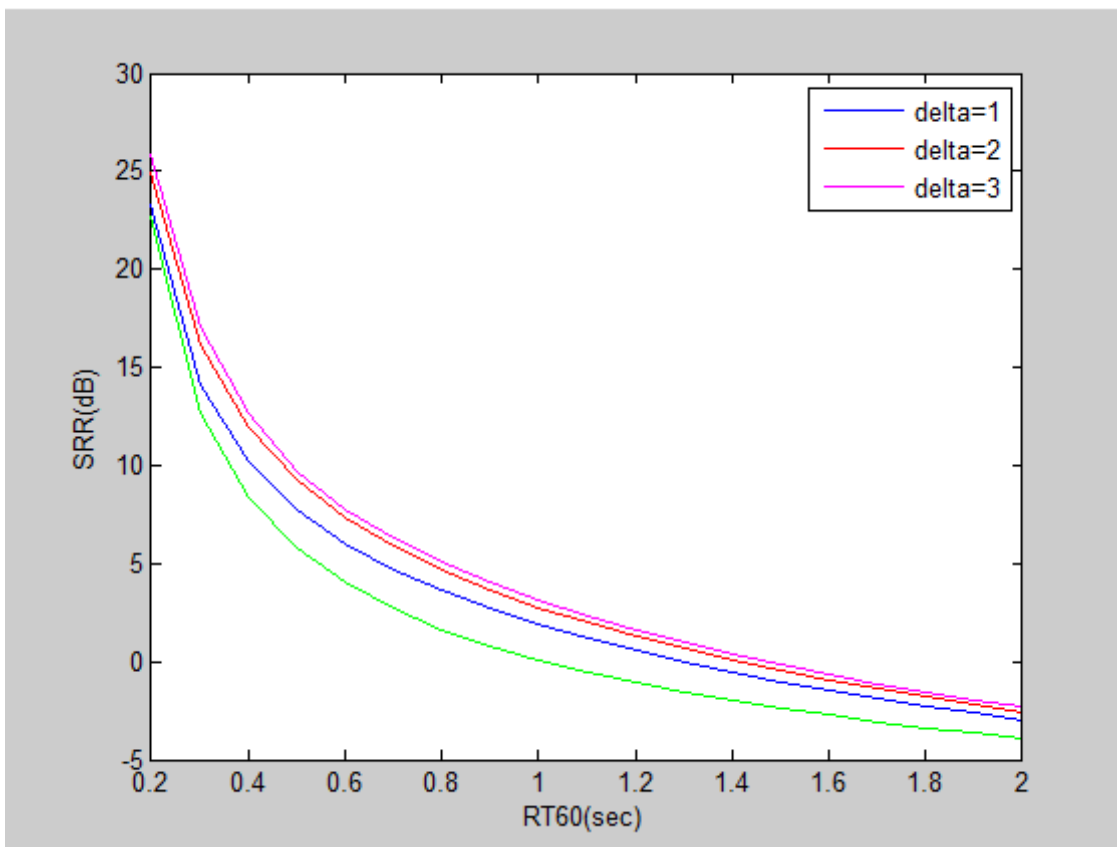


圖 3-10 後置濾波器之訊號殘響比

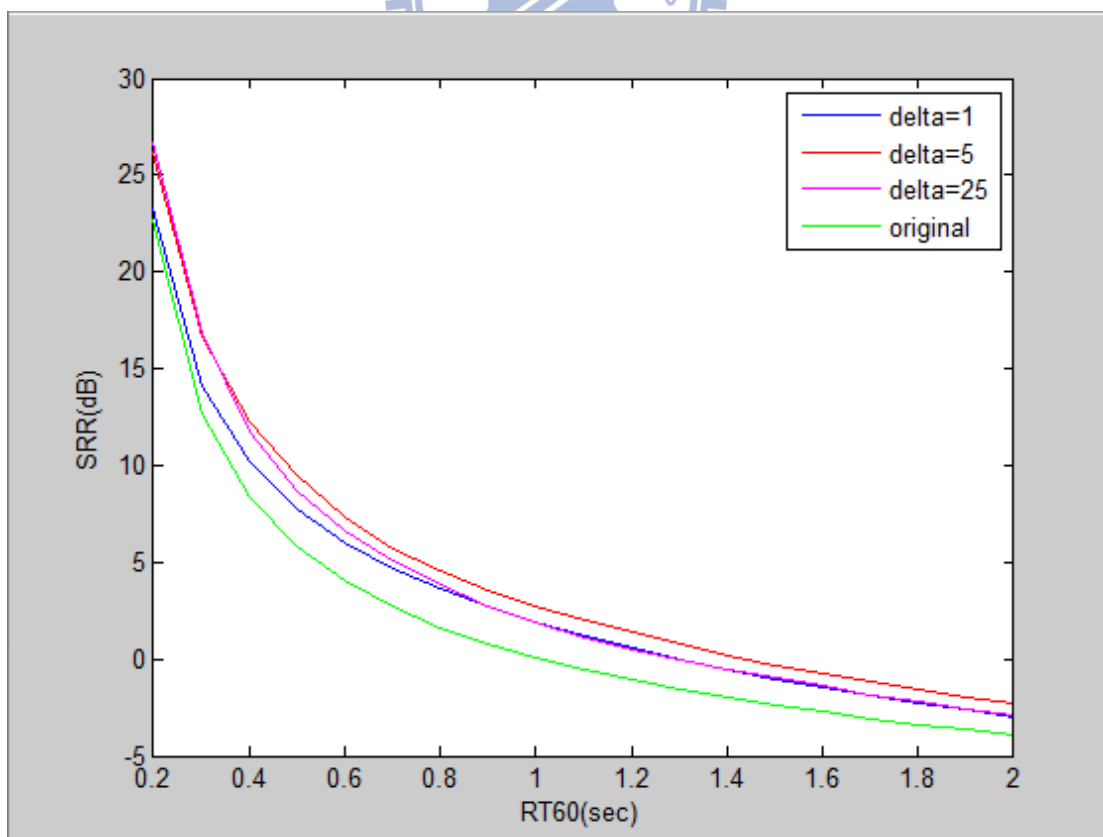


圖 3-11 後置濾波器之訊號殘響比

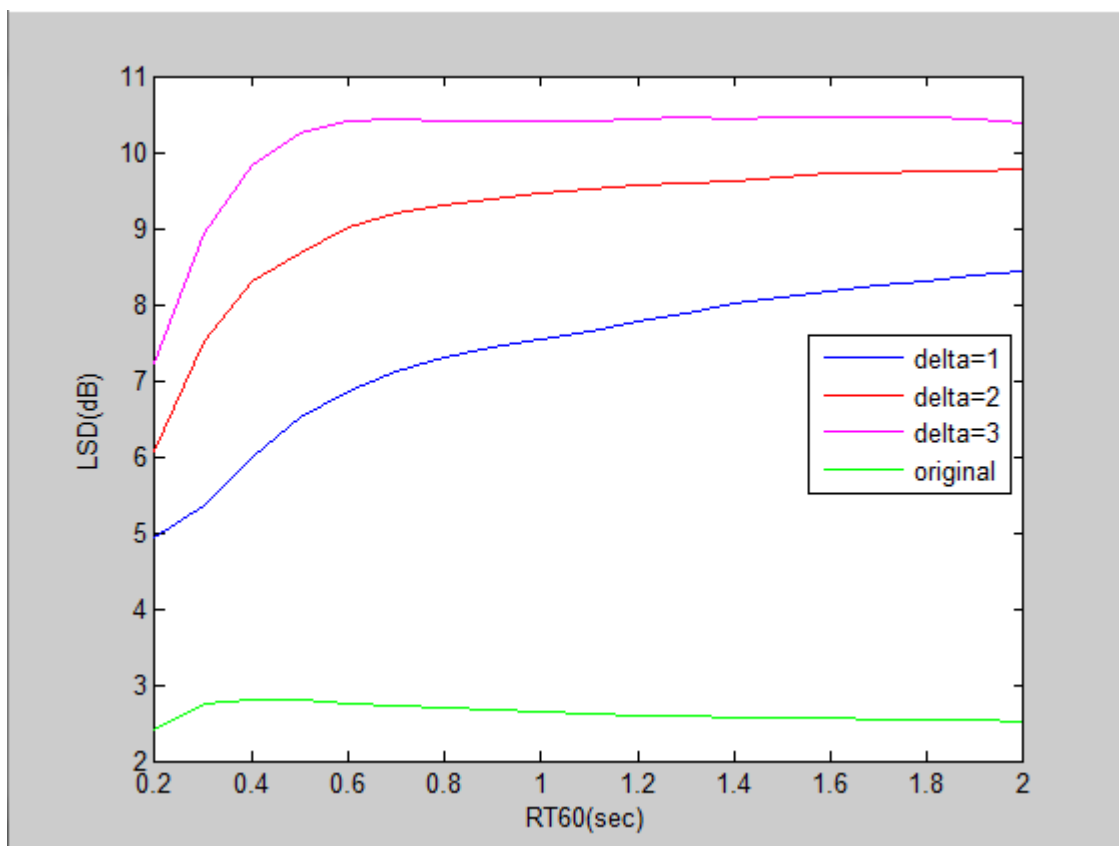


圖 3-12 後置濾波器之對數譜失真

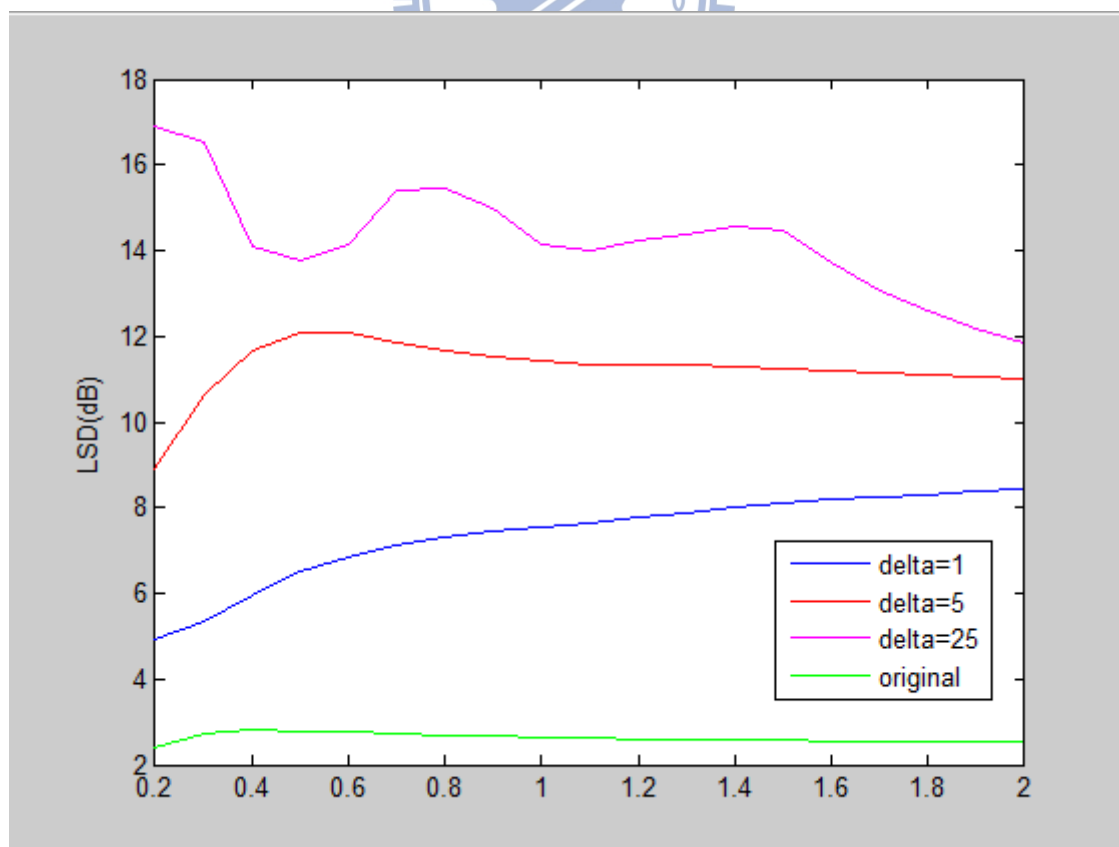


圖 3-13 後置濾波器之對數譜失真

由圖 3-10 與圖 3-12 中的 $\delta=1$ 、 $\delta=2$ 、 $\delta=3$ 可得知， δ 變大，訊號殘響比會變大，對數譜失真也會跟著變大，但是由圖 3-11 與圖 3-13 可發現如果 δ 太大，會造成後迴響的過估計，使得直接路徑部分的訊號也會被消到，所以 $\delta=25$ 時的訊號殘響比會稍為比 $\delta=5$ 小一點，而對數譜失真會變大。

純化過後輸出訊號以及兩階段各自輸出訊號的訊號殘響比與對數譜失真，如圖 3-14、圖 3-15、圖 3-16、圖 3-17 所示。由圖 3-14、圖 3-15 可觀察出，迴響時間漸漸變長時，結果會與前面波束形成介紹的輸出結果越相符。由圖 3-16 與圖 3-17 可發現，本論文架構的純化結果，能比單純只有波束形成或單通道後置濾波器在迴響時間長的情況下，越能維持效果。表 3-2 與表 3-3 是讓使用者參考的對照表，使用者可以依照不同的需求，例如需要的訊號殘響比，容忍失真的程度，依照對照表來選擇適合的 δ 。

根據圖 3-2、圖 3-3、圖 3-15 以及圖 3-16 可以觀察到，本論文的架構之輸出訊號與只經過波束形成處理之輸出訊號相比較，在迴響時間較短的情況，對數譜失真會稍微變大、訊號殘響比會稍微變大，主要是因為迴響時間較短，訊號迴響部分能量較低，造成直接路徑部分的自消除(Self cancellation)程度較高。在迴響時間較長的情況下，對數譜失真會稍微下降而訊號殘響比會較波束形成的輸出大，則顯示其直接路徑部分的自消除程度較低。

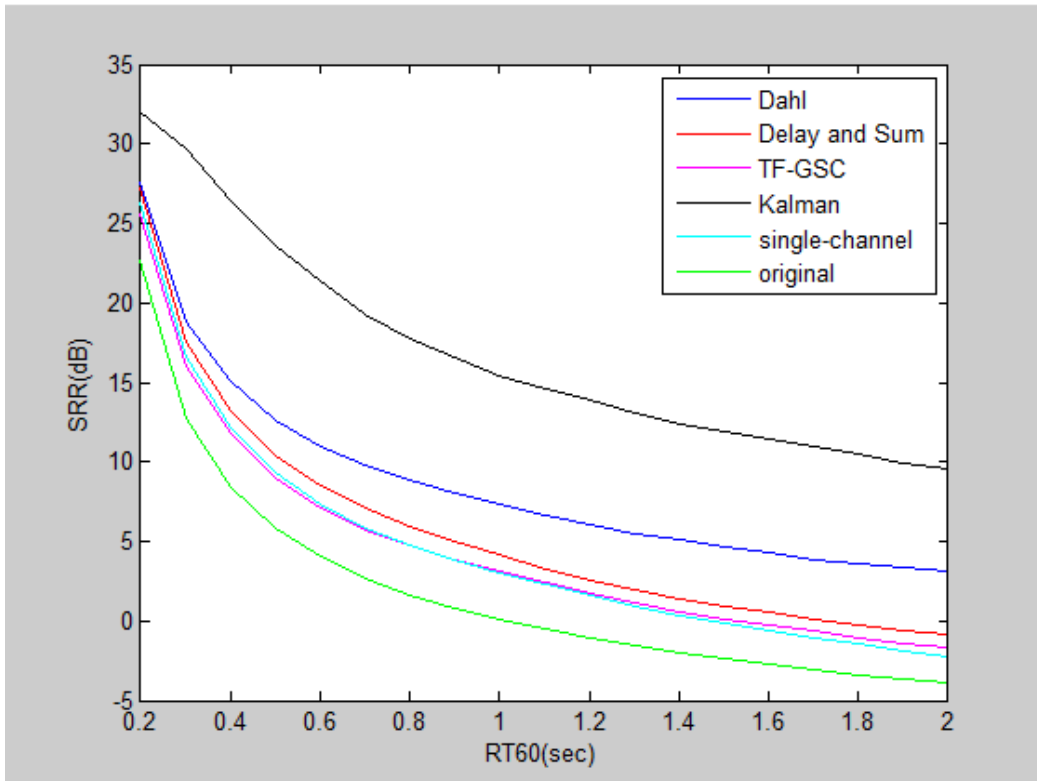


圖 3-14 純化訊號之訊號殘響比

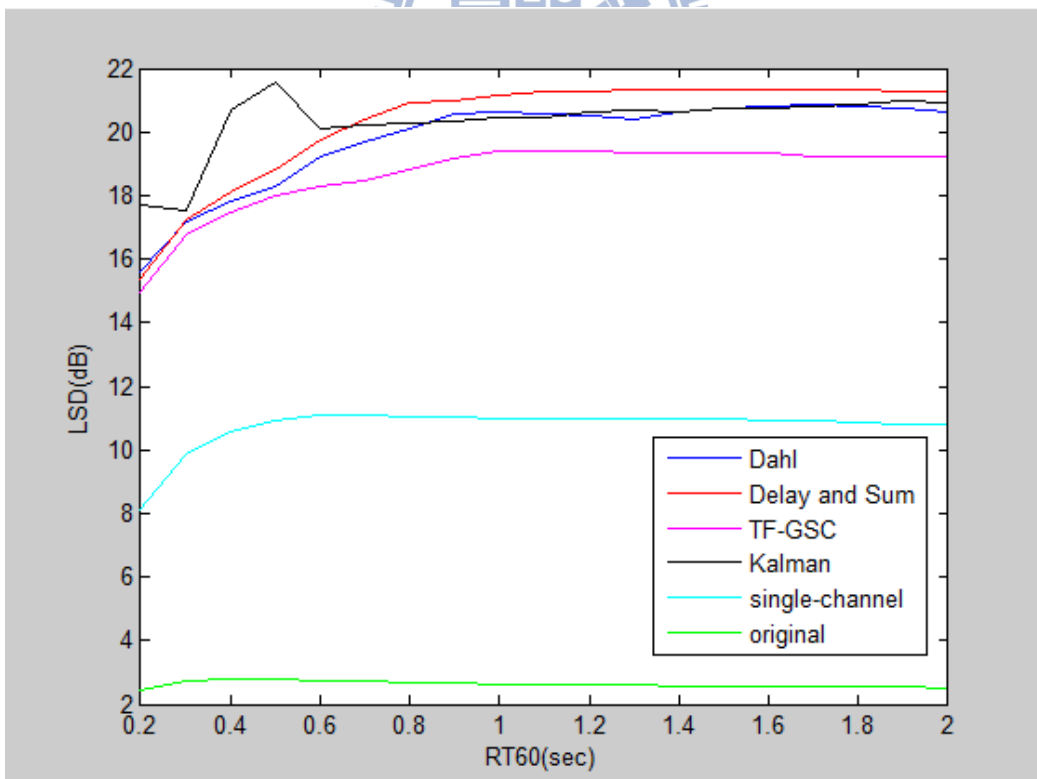


圖 3-15 純化訊號之對數譜失真

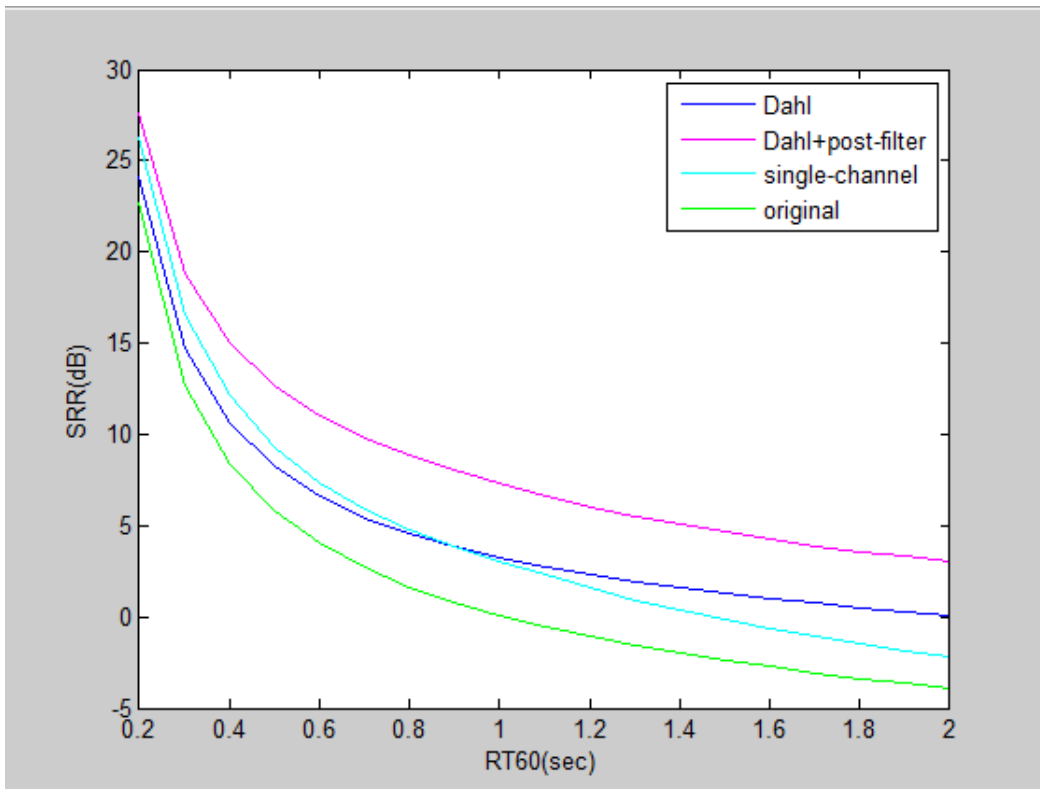


圖 3-16 兩階段輸出之訊號殘響比

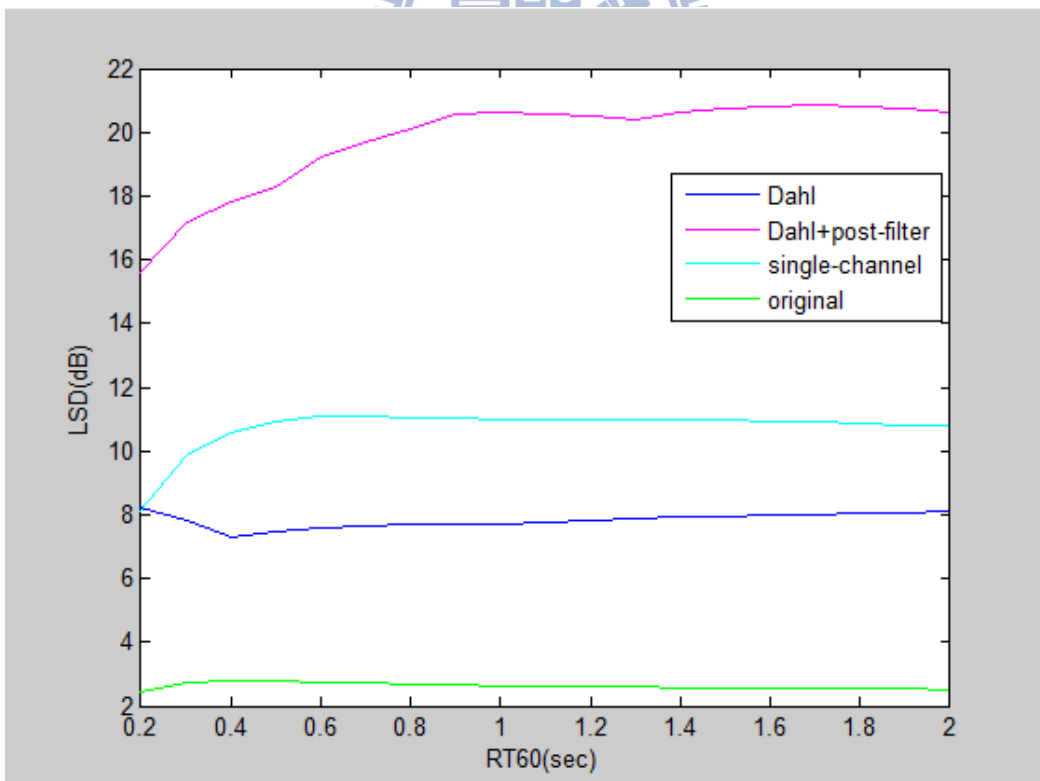


圖 3-17 兩階段輸出之對數譜失真

		RT60			
		0.5	1	1.5	2
	original	5.788241	0.073819	-2.37934	-3.90586
delta	1	10.48013	5.840099	3.701991	2.218376
	2	12.09237	6.747207	4.420009	2.847894
	3	12.76566	7.372765	4.781325	3.187679
	4	12.63602	7.357509	4.647785	3.075472
	5	12.87888	7.292633	4.556879	3.141693
	6	13.01536	7.192203	4.645835	3.179897
	7	13.04291	7.255582	4.675451	3.131081
	8	13.01453	7.273262	4.67321	3.171339

表 3-2 訊號殘響比對照表

		RT60			
		0.5	1	1.5	2
delta	1	13.94806	15.42228	16.20252	16.697
	2	16.17098	17.49553	17.96619	18.15511
	3	18.48567	20.09778	20.71019	20.45006
	4	18.27493	20.6505	20.74095	20.65159
	5	19.39956	20.28115	20.7347	20.64885
	6	20.67107	21.30488	20.67401	21.14339
	7	21.31464	21.90949	20.8894	21.52165
	8	21.65511	22.2379	21.10343	21.84882

表 3-3 對數譜失真對照表

第四章 實驗結果與分析

本章節將介紹，利用本論文方法消除迴響的效果。實驗將分為兩個部分。首先，對麥克風陣列收到的訊號先經過波束形成處理，並對其結果作分析。接下來再將波束形成的輸出經過後混響譜方差估測迴響訊號之能量，再利用後置濾波器濾除迴響，並分析其輸出訊號。

4.1 實驗環境介紹

本論文的錄音平台採用本實驗室的麥克風陣列，麥克風陣列為 8 顆線性陣列，麥克風與麥克風的距離為 3 公分。圖 4-1 為裝置的實際照片，圖 4-2 為無響室的聲源與麥克風陣列關係的實際照片，圖 4-3 為在迴響較為明顯的室內空間中聲源與麥克風陣列關係的實際照片。對於由麥克風接收的訊號，選擇 1024 個取樣點(sample)作為音框大小(Frame size)，選擇每一次移動 512 個取樣點，也就是重疊大小(Overlap size)為 512 個取樣點。

在本論文的實驗中，聲源皆與麥克風陣列成 90 度，訊號從麥克風的正前方射入。而在預錄資料的部分，則是分別在無響室以及迴響較為明顯的室內空間錄製需要的預錄資料。



圖 4-1 線性麥克風陣列平台



圖 4-2 無響室聲源與麥克風陣列關係



圖 4-3 室內空間聲源與麥克風陣列關係

4.2 系統消除迴響效果

此段落將比較各種波束形成之波束方向圖(Beam Pattern)，四種波束形成分別如以下四圖所示。由圖可以觀察到 Dahl 演算法、時間延遲相加波束與 TF-GSC 波束形成最大的能量分佈主要在目標聲源方向的角度上，而卡曼濾波器會在目標聲源的方向上有不少能量較低的部分，因此可以得知濾波器會把該方向的聲源消除，造成會有自消除的結果。造成此結果的原因是，因為錄製預錄資料時，會有人為的誤差使訊號入射角度有所偏差，導致直接路徑的聲源被當成是干擾聲源而被消除。由此可知，卡曼濾波器較不適合使用於本論文方法。

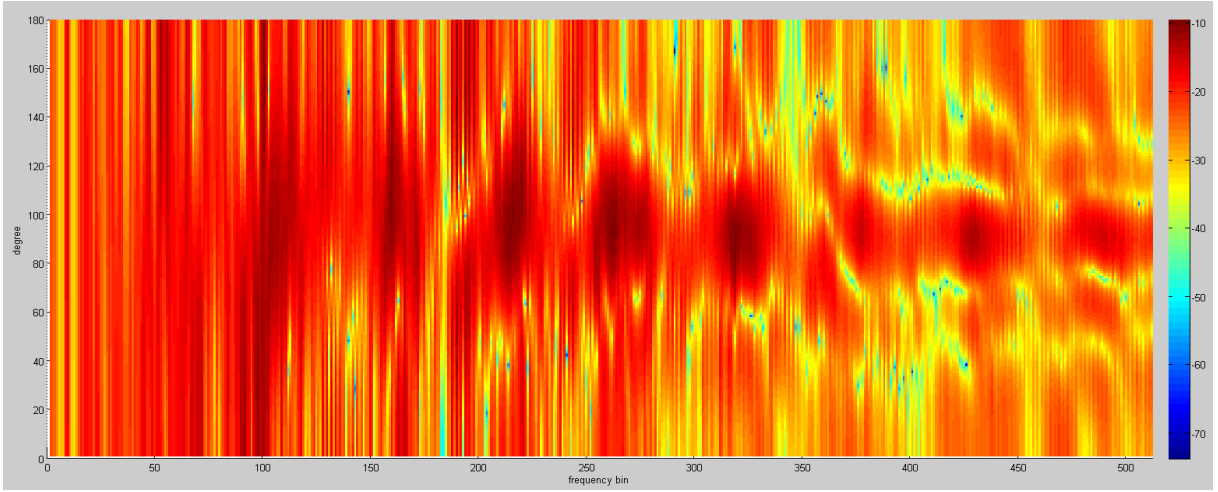


圖 4-4 Dahl 演算法之波束方向圖

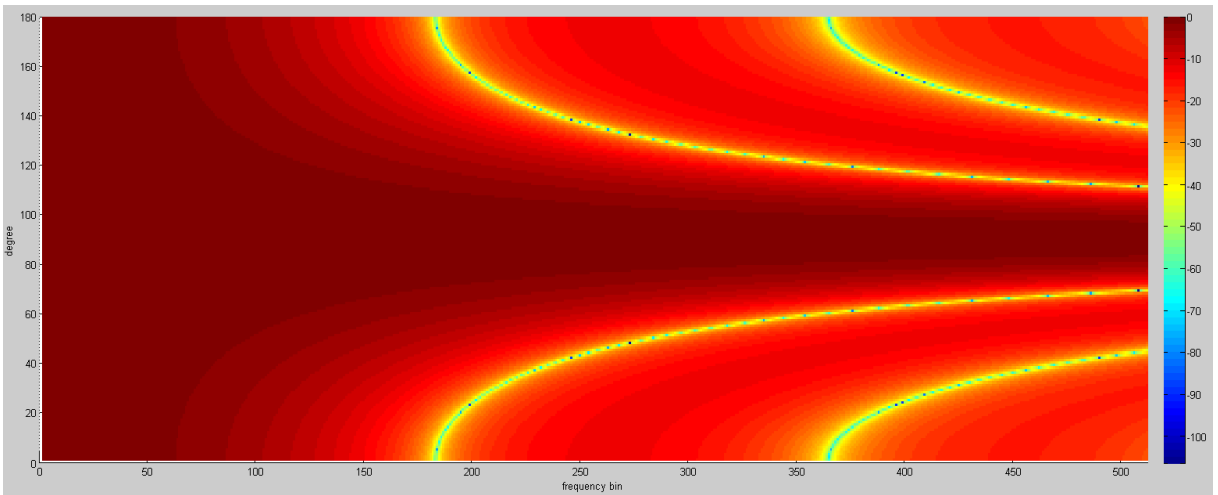


圖 4-5 時間延遲相加波束之波束方向圖

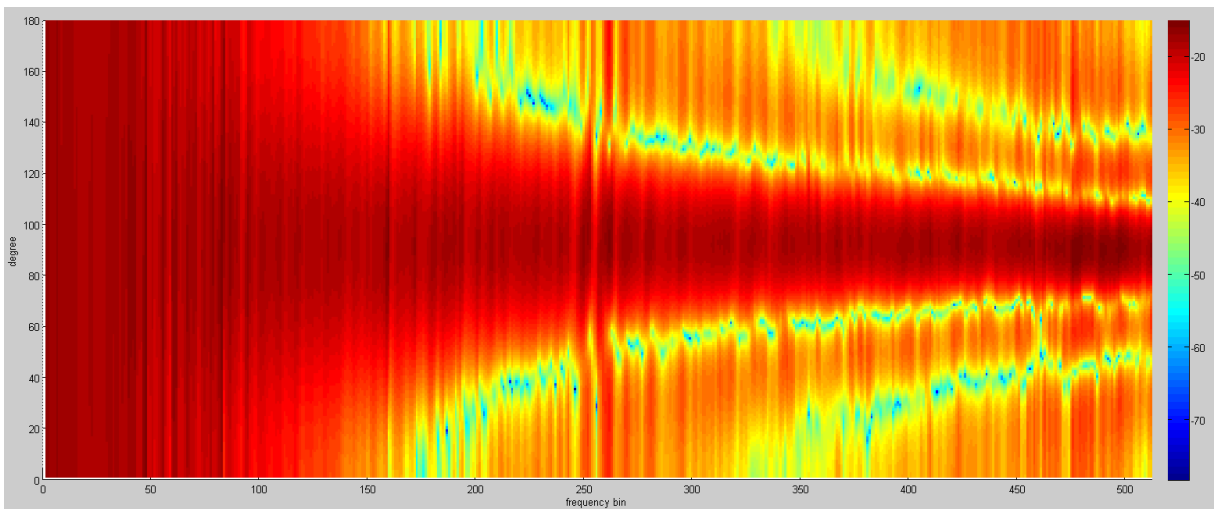


圖 4-6 TF-GSC 之波束方向圖

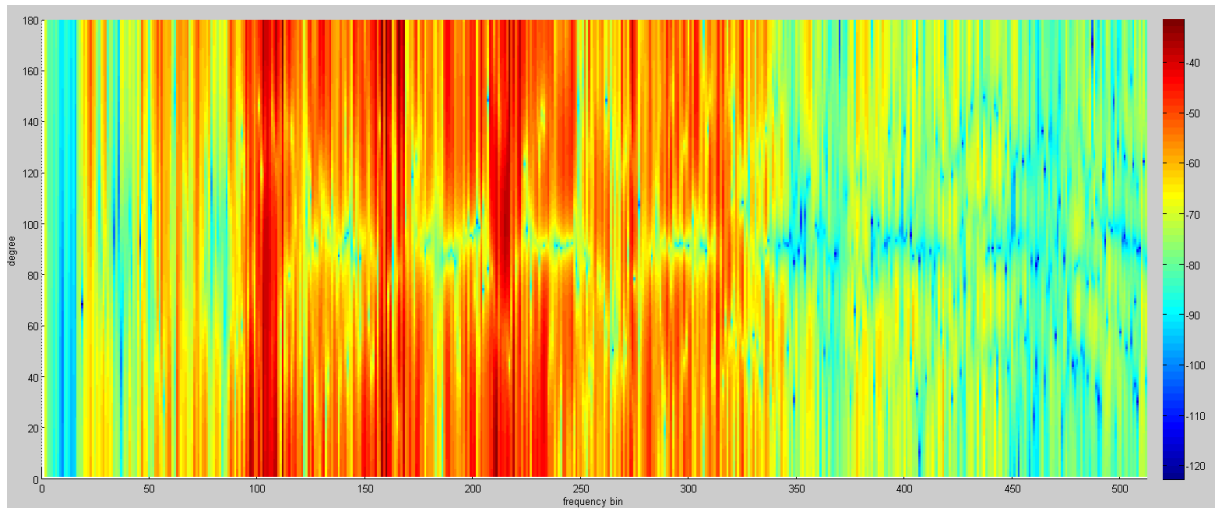


圖 4-7 卡曼濾波器之波束方向圖

本實驗估算出的迴響時間大約為 1.02 秒，不同波束形成的結果如表 4-1 所示，系統輸出結果如表 4-2 所示。

	SRR(dB)/LSD(dB)
Dahl	-1.57502/5.670427
Delay and sum	-4.36925/3.432652
TF-GSC	-4.48152/4.433173

表 4-1 波束形成輸出結果比較表

$\delta=1$	SRR(dB)/LSD(dB)
Dahl	-1.60164/10.2273
Delay and sum	-3.3789/8.282535
TF-GSC	-3.024/10.44279
$\delta=4$	
Dahl	0.112122/13.12621
Delay and sum	-2.38513/11.25157
TF-GSC	-1.13728/14.86561

表 4-2 系統輸出結果比較表

觀察表 4-1，可以觀察到訊號殘響比最大者為 Dahl 演算法，時間延遲相加波束約略比 TF-GSC 的訊號殘響比稍微大一點。由於實際的實驗，直接路徑的訊號會有雜訊的存在，所以對數譜失真會包含了雜訊的影響，導致對數譜失真的結果會稍微與模擬的部分有些出入。

觀察表 4-2，可發現 $\delta=1$ 時訊號殘響比的大小順序為 Dahl 演算法、時間延遲相加波束，最小為 TF-GSC，當 $\delta=4$ 時，訊號殘響比大小順序為 Dahl 演算法、TF-GSC，最小變為時間延遲相加波束，此結果意味著每種情況適合的參數不同。

麥克風接收訊號如圖 4-8 與圖 4-12 所示，各種波束形成的輸出結果如圖

4-9 到圖 4-11 與圖 4-13 到圖 1-15 所示。



圖 4-8 輸入訊號

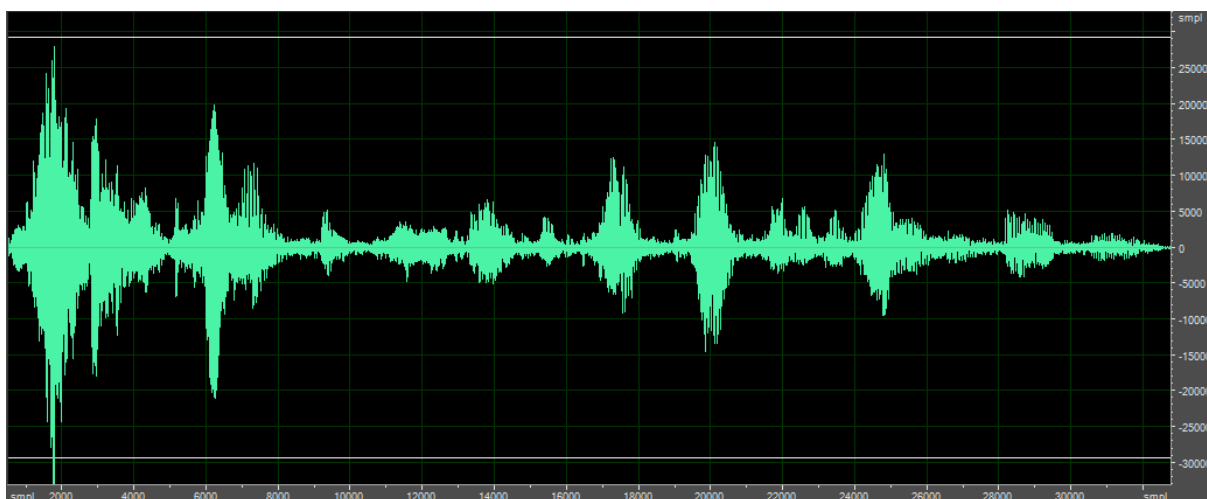


圖 4-9 Dahl 演算法波束形成輸出

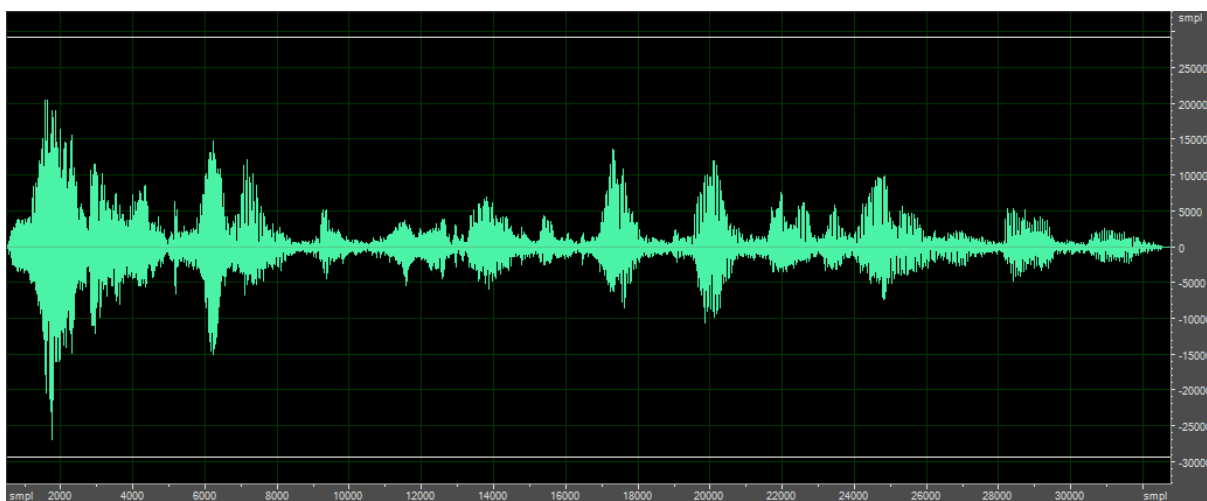


圖 4-10 時間延遲相加波束輸出

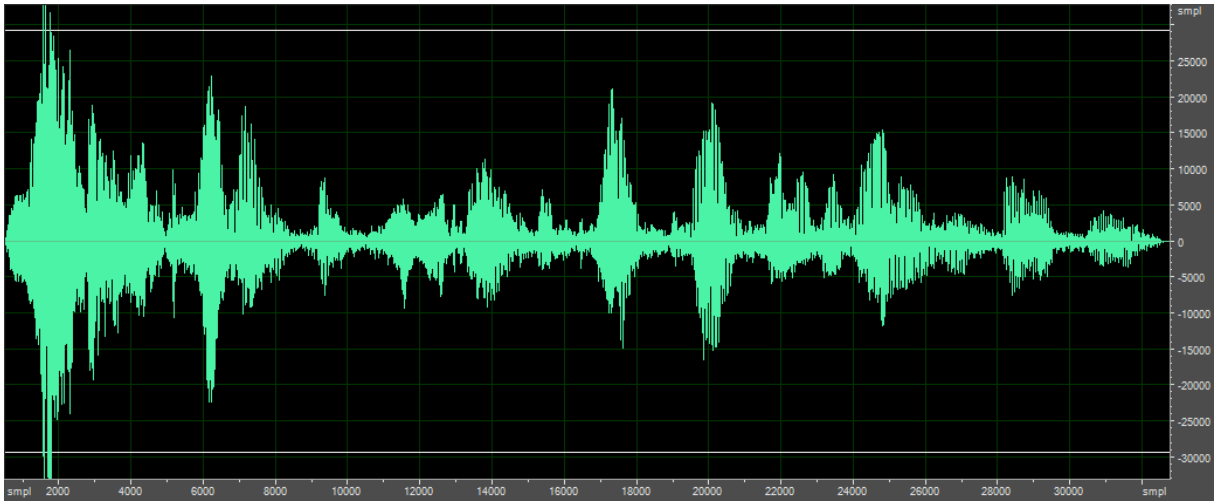


圖 4-11 TF-GSC 波束形成輸出

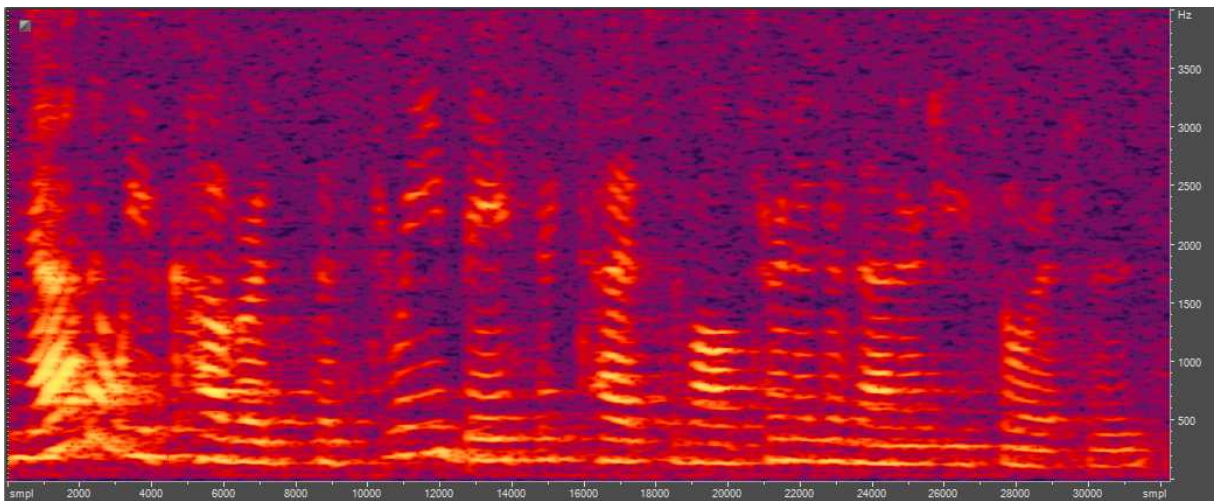


圖 4-12 輸入訊號時頻圖

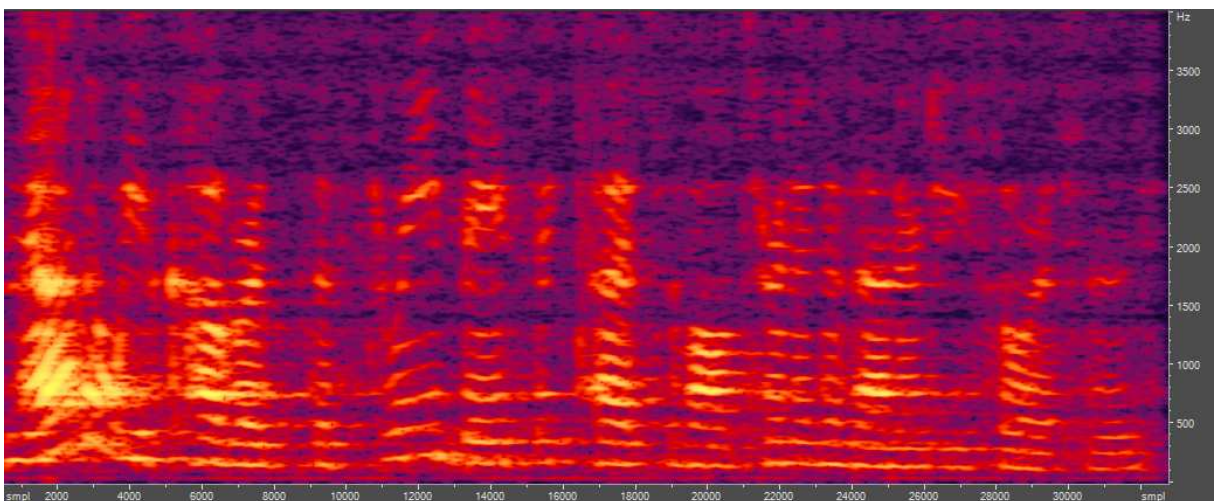


圖 4-13 Dahl 演算法波束形成時頻圖

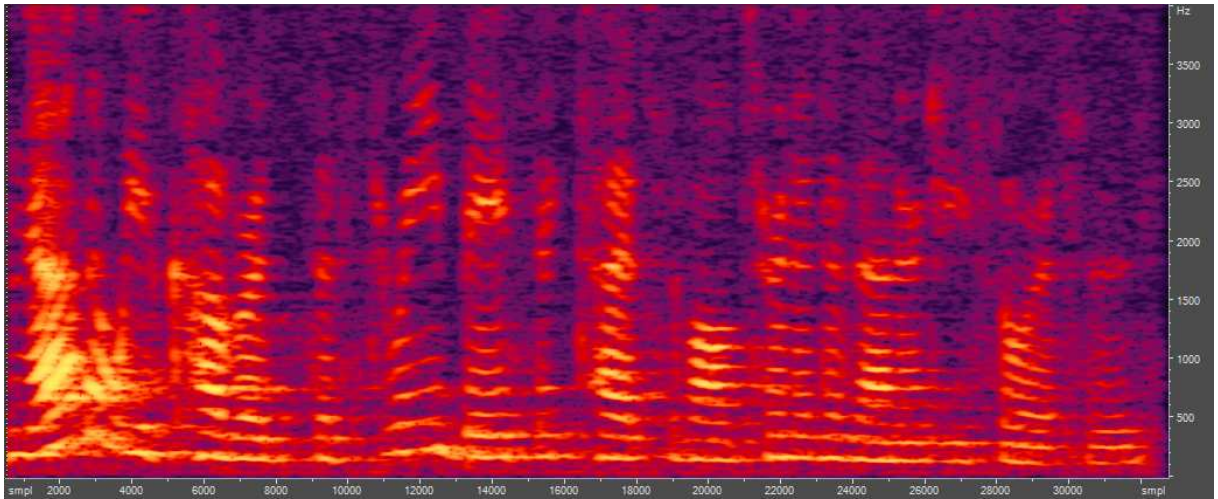


圖 4-14 時間延遲相加波束時頻圖

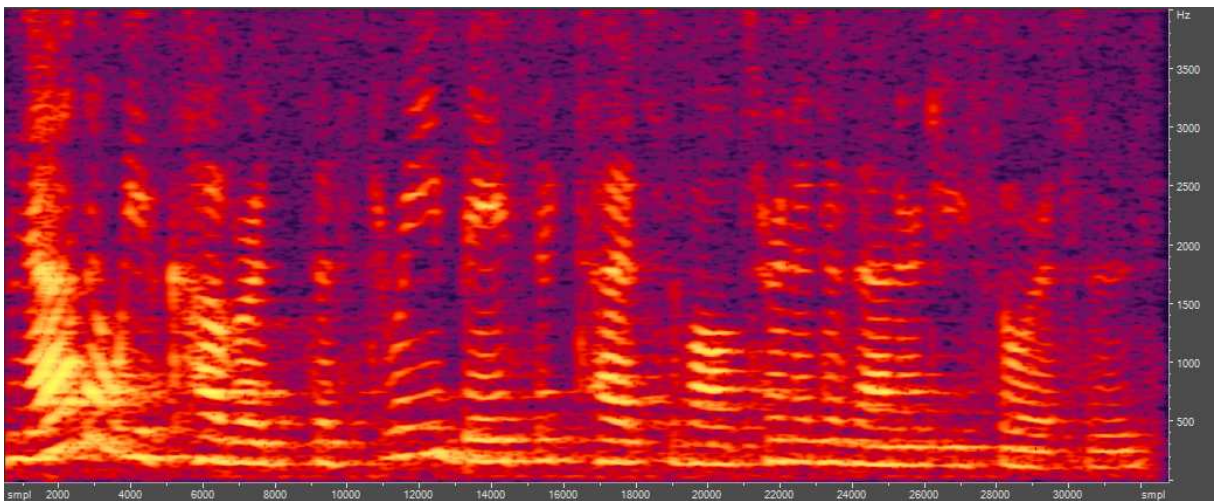


圖 4-15 TF-GSC 波束形成時頻圖

各種波束形成之全系統輸出如圖 4-16 到圖 4-21 所示。以下將固定 $\delta=4$ 進行比較。選擇 $\delta=4$ 主要是因為，在此情況下三種波束形成都能有較好的效果，亦即純化之後的訊號能有效的抑制迴響，並且不會因為失真而無法辨識內容。



圖 4-16 Dahl 演算法全系統輸出

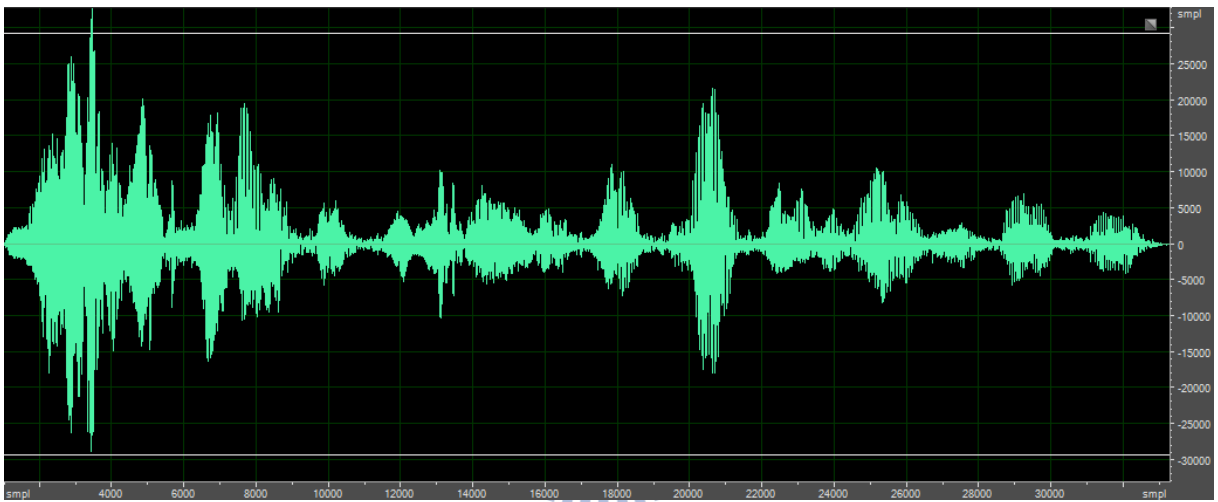


圖 4-17 時間延遲相加波束全系統輸出



圖 4-18 TF-GSC 全系統輸出

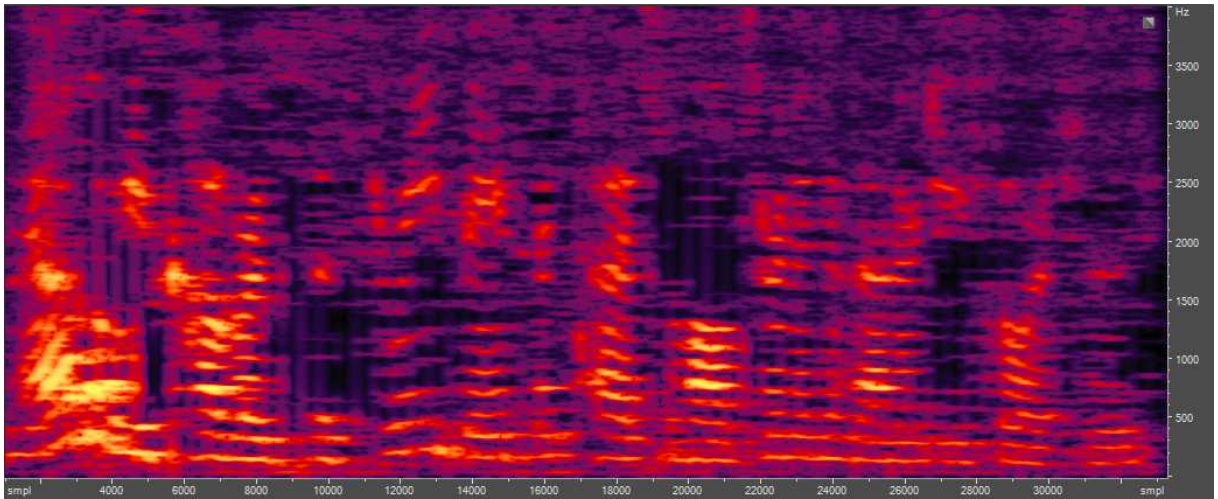


圖 4-19 Dahl 演算法全系統時頻圖

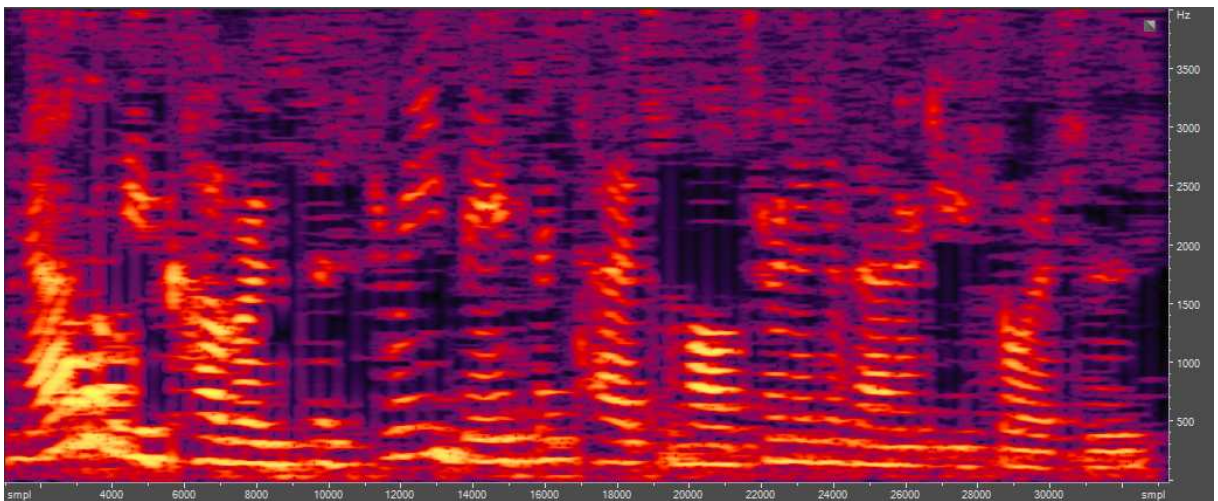


圖 4-20 時間延遲相加波束全系統時頻圖

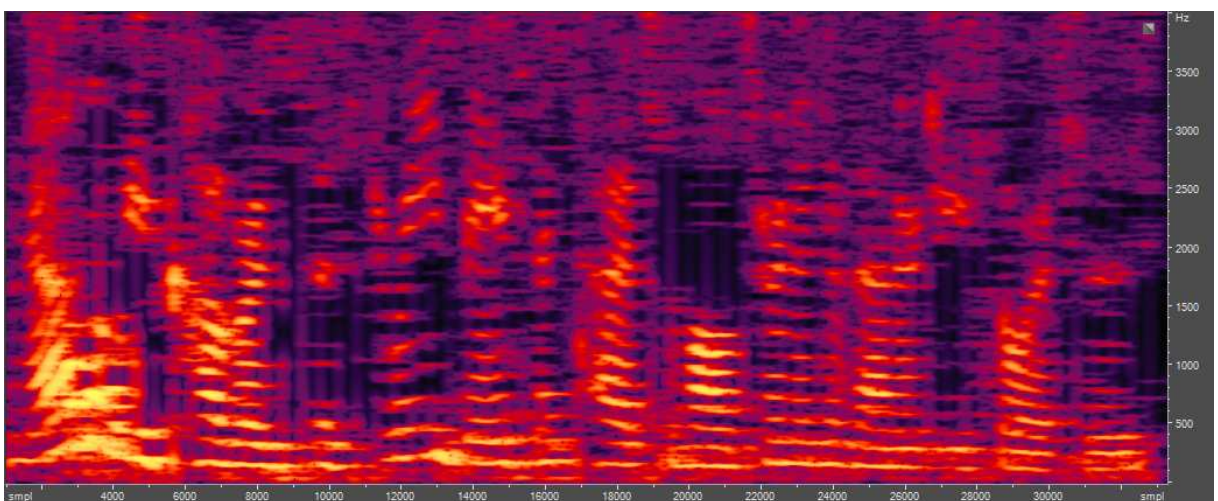


圖 4-21 TF-GSC 全系統時頻圖

第五章 結論

5.1 研究成果

本論文提出了一個可以依照使用者需求調整消除迴響效果的系統。先利用麥克風陣列的空間資訊，依照空間濾波器的特性先消除聲源方向以外的迴響，再進一步利用統計模型的特性，依照估測出的訊號衰減率，達到消除迴響的效果。由以上的實驗以及模擬結果可以得知，本論文方法在不同的需求下，皆可以利用改變參數 δ 來調整訊號的衰減率，進一步改變消除迴響的強度，在可接受失真範圍內，若 δ 上升，可有效提高訊號殘響比。由於本系統依照不同的 δ 的選擇，能有不同消除迴響的效果，所以使用者可以依照自己的情況來調整參數，使純化結果可以達到使用者的要求。

5.2 未來展望

當聲源不斷的發出聲音時，由於聲音能量的大小會不斷的改變，導致直達混響聲能比會不斷的改變。本論文中估測直達混響聲能比的部分，會因為訊號中帶有語音成分，導致直達混響聲能比會一直上升，因而無法準確的消除迴響的部分。若能發展出一套改進直達混響聲能比的估測機制，即可使得消除迴響的精準度上升，亦即使自消除的程度下降，如此可以使得在消除同樣程度迴響的情況下，讓訊號能有更少的失真，語音品質因而得到更好的改善。

Reference

- [1] J.B. Allen and D.A. Berkley, "Image Method for Efficiently Simulating Small Room Acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [2] M. Delcroix, T. Hikichi, and M. Miyoshi, "Precise dereverberation using multichannel linear prediction," *IEEE Trans. Audio, Speech, Language Processing*, vol. 15, no. 2, pp. 430–440, 2006.
- [3] M. Delcroix, T. Hikichi, and M. Miyoshi, "Dereverberation and denoising using multichannel linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 6, pp. 1791–1801, Aug. 2007.
- [4] M. Delcroix, T. Hikichi, and M. Miyoshi, "On the use of lime dereverberation algorithm in an acoustic environment with a noise source," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'06)*, 2006, vol. 1, pp. 825–828.
- [5] M. Delcroix, T. Hikichi, and M. Miyoshi, "Blind dereverberation algorithm for speech signals based on multi-channel linear prediction," *Acoustical Science and Technology*, vol. 26, no. 5, pp. 432–439, 2005.
- [6] J.D. Polack, *La transmission de l'énergie sonore dans les salles*, Thèse de doctorat d'état, Université du Maine, La mans, 1988.
- [7] E. A. P. Habets, *Single- and multi-microphone speech dereverberation using spectral enhancement*, Ph.D. dissertation, Eindhoven University, Eindhoven, The Netherlands, June 2007.
- [8] E.A.P. Habets, S. Gannot, and I. Cohen, "Dual-Microphone Speech Dereverberation in a Noisy Environment," in *Proc. of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'06)*, Vancouver, Canada, Aug. 2006.
- [9] M.R. Schroeder, "New method of measuring reverberation time," *Journal of the Acoustical Society of America*, vol. 37, no. 6, pp. 1187–1188, June 1965.
- [10]] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, "Blind estimation of

- reverberation time based on the distribution of signal decay rates,” in Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP '08), pp. 329– 332, Las Vegas, Nev, USA, March-April 2008.
- [11] K. Furuya, S. Sakauchi, and A. Kataoka, “Speech dereverberation by combining MINT-based blind deconvolution and modified spectral subtraction,” in Proc. ICASSP, May 2006, vol. 1, pp. 813–816.
- [12] Lin, X., Gaubitch, N.D., Naylor, P.A.: Blind speech dereverberation in the presence of common acoustical zeros. In: Proc. European Signal Processing Conf. (EUSIPCO), pp. 389–393. Poznań, Poland (2007)
- [13] K. Kumatani, L. Lu, J. McDonough, A. Ghoshal, and D. Klakow, “Maximum negentropy beamforming with superdirectivity,” in Proc. Eusipco, Aalborg, Denmark, 2010.
- [14] M. Dahl and I. Claesson, “Acoustic noise and echo canceling with microphone array,” IEEE Trans. Vehic. Technol., accepted for publication.
- [15] D.H Johnson and D.E Dudgeon, Array Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [16] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with application to speech,” IEEE Trans. Signal Processing, vol. 49, pp. 1614–1626, Aug. 2001.
- [17] Y. H. Chen and C. T. Chiang, “Adaptive beamforming using the constrained Kalman Filter,” IEEE Transactions on Antennas and Propagation, vol. 41, no. 11, pp. 1576–1580, 1993.
- [18] 朱育成, ” 使用卡曼濾波器追蹤參考訊號之適應性語音純化波束形成器” , 交大電控碩士論文, Sept 2011
- [19] Recommendation ITU-T P.862, Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, Int'l Telecommunication Union, Feb. 2001.
- [20]] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, “Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2001, vol. 2, pp. 749–752.

[21] “Methods for subjective determination of transmission quality,”
Recommendation P.800, International Telecommunications Union (ITU-T), Feb.
1996.

