

國立交通大學

工學院聲音與音樂創意科技
碩士學位學程

碩士論文



語音強化與立體聲迴聲消除於智慧型
電視之應用

Speech Enhancement and
Stereophonic Acoustic Echo Cancellation
for Voice Interface of Smart TV

研究生：劉淵瀚
指導教授：胡竹生 博士

中華民國 一百零二 年 十月

語音強化與立體聲迴聲消除
於智慧型電視之應用

Speech Enhancement and Stereophonic Acoustic Echo
Cancellation for Voice Interface of Smart TV

研究生：劉淵瀚

Student : Yuan-Han Liu

指導教授：胡竹生 博士

Advisor : Jwu-Sheng Hu

國立交通大學

工學院聲音與音樂創意科技碩士學位學程

碩士論文



A Thesis
Submitted to Master Program of Sound and Music Innovative Technologies
College of Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Master
in

Engineering

October 2013

Hsinchu, Taiwan, Republic of China

中華民國 一 百 零 二 年 十 月

語音強化與立體聲迴聲消除 於智慧型電視之應用

研究生：劉淵瀚

指導教授：胡竹生 博士

國立交通大學工學院聲音與音樂創意科技碩士學位學程

摘要

本論文提出一套利用麥克風陣列結合適應性空間濾波器及聲學迴聲消除的方法進行語音強化。智慧型電視的興起，使遠距語音訊號處理應用再度受到重視，例如語音命令控制和視訊會議都屬於遠距語音的應用，利用這些功能的同時第一個面對的問題即是如何抑制電視播放的聲音、迴聲和環境噪音對目標聲源之影響，這將是語音介面是否能夠成功運用於智慧型電視的關鍵。

本論文設計一套能消除立體聲迴聲並且抑制干擾源和剩餘迴聲之適應性波束形成器。藉由實際智慧型電視的環境配置下進行實驗模擬，麥克風收到訊號分別經過聲學迴聲消除處理後在通過最小方差無失真響應的波束形成器，配合適應性噪音消除達到語音強化的效果；其中空間前處理和噪音估測所需的角位資訊，事先運用轉移函式比值法求得目標聲源在空間中相對轉移函式。

Speech Enhancement and Stereophonic Acoustic Echo Cancellation for Voice Interface of Smart TV

Student : Yuan-Han Liu

Advisor : Prof. Jwu-Sheng Hu

Master Program of Sound and Music Innovative Technologies
National Chiao Tung University

ABSTRACT

An approach proposed in this thesis combines adaptive spatial filtering of microphone array and acoustic echo cancellation for speech enhancement. The growing trend of smart TV users has reintroduced the value of using long distance speech signal process application again. For example, voice control and video conference both are long distance speech applications. When you use those applications, the one of the problems you may encounter is how to minimize the TV sound, echo, and environment noise all of which affects the target source. This issue is the key to whether or not smart TV's speech interface will be successful.

This thesis presents a stereophonic acoustic echo cancellation (SAEC) after using an adaptive beamformer for noise reduction and residue echo cancellation. In the real smart TV environment setting, this thesis simulated and experiment with, microphones which received voice signals that passed through a stereophonic acoustic echo cancellation process and then a minimum variance distortionless response (MVDR) beamformer combine adaptive noise cancellation (ANC) for speech enhancement. Spatially pre-process and noise estimate both need to know the directional information between target source and microphones in any given space. To find out this data, one must calculate relative transfer function (RTF) using the transfer function ratio (TFR) method.

誌謝

本論文順利的完成同時碩士生涯也劃下句點，給自己在學習的道路上立下一個新的里程碑。首先感謝胡竹生老師的循循善誘和諄諄教誨，教會了我成為工程師以及做研究的態度，用嚴謹的思考解決問題用鍥而不捨的精神找出解答；也感謝老師在研究上給予的幫助，提供一個充滿能量和資源充沛的研究環境，讓我有幸在學生涯中得到這些難得的經驗。

兩年一眨眼就過去了，在實驗室的日子就在家裡一樣溫馨，不論實驗室聚餐大家開心吃飯喝酒或是一般日子去學生餐廳閒話家常都是很美好的回憶。最感謝的一定非明唐學長莫屬，在研究上碰到的瓶頸或是問題經過開導如同得到一盞明燈般有方向繼續進行下去。感謝碩士論文的內容中很多問題一起討論的耕維學長細心的解答或幫助，每天都會話家常的昭男學長希望畢業後還能繼續保持我們的動態，每天都很開心的阿吉學長無論什麼問題都能細心傾聽給出中肯的建議，沉穩的男人勁源學長多次幫助過我，兼具鐵漢及柔情於一身的 Judo 學長答應要罩我，對甜點非常堅持的阿法學長分享手工製作的蛋糕給我吃。也感謝鳴哥在一進交大什麼都不懂的時候給我的幫助告訴我很多學程及實驗室的資訊，大學就認識的翰哥食量驚人講話中肯，搞笑幽默的建廷中文超好的 Daniel 很開心認識你們。不論音樂體育都熟悉也是個日韓通的期元開心當你半年的室友，好奇可愛的哲宇有你可以討論聲音處理上的程式真好，搞笑但專業的健身教練阿文在我眼中你是個毅力非凡的人，我的地下室好夥伴鳴遠在你身上我學會強者的謙虛低調，認真的佑軒謝謝你每次無怨無悔的幫忙我做些雜工，帥氣的鬍子、酷酷的小樂、和藹的小綜、老闆樣的凱翔有你們這些學弟點綴了我碩士生活中美好記憶。聲音組明日之星知琬和錢丹、實驗室新活寶 Winnie、嘴上功夫一流的哈魯，還有來不及認識的你希望能在未來成為實驗室的驕傲。最後在碩士生涯中遇見的好朋友罐頭，未來在台北奮鬥工作的日子裡還能時常保持聯繫也不會忘了我們這兩年的運動與擔任實驗室管理員的時光；在研究所能遇到足智多謀的大夢很開心一起修課、玩樂、練團、暢聊、出遊、甚至到寫論文口試彼此形同手足同甘共苦有你真好，讓我的兩年時光增色不少。

感謝善於交際的王可柔、吉他之神許峻豪、沉穩內斂的蔡坤廷、成熟幽默的許大哥、攝影好手黃品睿、漫畫天才燴飯、歌王小高、甜美的小容、女神 EG、凡事講求專業的蔡鈺群、老實的鄭中皓、有學姊風範的欣儀，開心和你們一起吃飯聊天逛街上課，不會

忘記 WOCMAT 研討會和 COLCK 樂團演出的一同付出的點點滴滴，真的很開心能在聲音學程遇到這群對聲音有相同興趣的朋友；也感謝從大學一直陪伴我到研究所的朋友，分享生活中一切的翊綺、和每個人都麻吉的阿哲、又帥又多金的大鴻、憨厚又甜蜜的嘎銘和無話不聊的小玥讓我在一個新的環境中還能有故友相伴相助真是三生有幸。感謝因調酒認識的一群愛好者給我很多歡樂的時光，特別感謝我的啟蒙者小莫，交大林志玲小彥，跳舞女王喬安娜因酒緣成為了好友。這兩年裡感謝 Jen 的交換語言讓我更了解美國希望妳中文越來越好，樂天的 Maiko 也希望在台灣生活多彩多姿，有趣的 MM 很開心在研究所又能與妳再度成為同學，感謝交通大學多給了我了兩年時間讓我當學生的機會，校園生活最後能在此劃下句點已心滿意足。

最後感謝我背後默默支持我的父母和家人，爸爸劉岳冬、媽媽戴錫卿和哥哥劉淵浩，提供給我安穩幸福的依靠，不論發生什麼事都能給予我支持與鼓勵，讓我在快樂中成長自由中學習，也給了我很多人生的啟發與開導，我愛你們。

謹以此論文獻給我生命中最摯愛的父母



目 錄

摘 要.....	I
ABSTRACT	II
誌 謝.....	III
圖 目 錄.....	VI
第一章 緒論	1
1.1 研究動機	1
1.2 研究目標	1
1.3 相關文獻探討	2
1.4 論文架構	3
第二章 適應性陣列訊號處理	4
2.1 陣列訊號處理	4
2.2 適應性訊號處理	9
2.3 適應性空間濾波器	16
第三章 系統架構與相關技術	24
3.1 應用說明與系統架構	24
3.2 轉移函式比值演算法	27
3.3 立體聲迴聲消除演算法	29
第四章 實驗結果與分析	32
4.1 AEC 效能評估分析	34
4.2 SAEC 結合 Beamformer 效能評估	59
第五章 結論	67
5.1 研究成果	67
5.2 未來展望	67
參考文獻.....	68

圖目錄

圖 2-1	陣列模型.....	4
圖 2-2	空間濾波器架構.....	6
圖 2-3	均勻線性陣列空間響應(M=8, frequency=3450 Hz, d=0.05 m).....	7
圖 2-4	Grating Lobe 示意圖.....	8
圖 2-5	適應性濾波器處理架構.....	9
圖 2-6	空間聲學反射產生的聲學迴聲.....	9
圖 2-7	AEC 主體架構.....	10
圖 2-8	維納濾波器架構.....	11
圖 2-9	LMS 運算架構.....	13
圖 2-10	Wiener Filter in Frequency Domain.....	14
圖 2-11	Matrix processor.....	17
圖 2-12	麥克風與聲源方向關係圖.....	18
圖 2-13	MVDR 輸入訊號.....	18
圖 2-14	MVDR Beam Pattern.....	18
圖 2-15	MVDR 輸出訊號.....	19
圖 2-16	MVDR 結合雜訊估測.....	20
圖 2-17	MVDR+Noise Estimate 輸入訊號.....	21
圖 2-18	FBF Beam Pattern.....	22
圖 2-19	BM Beam Pattern.....	22
圖 2-20	經過 MVDR 處理之 Y_{FBF} 訊號.....	22
圖 2-21	經過 MVDR+Noise Estimate 處理之輸出 Y 訊號.....	23
圖 3-1	AEC 結合 BF 架構圖.....	24
圖 3-2	AEC+BF 架構圖.....	25
圖 3-3	AEC+BF+TFR 架構圖.....	25
圖 3-4	智慧型電視應用之系統架構圖.....	26
圖 3-5	單聲源系統模型.....	27
圖 3-6	SAEC 架構圖.....	29
圖 3-7	SAEC Wiener 架構圖.....	31
圖 4-1	類比麥克風陣列裝置.....	32
圖 4-2	錄製器材與裝置.....	33
圖 4-3	目標聲源與干擾源擺放位置.....	33
圖 4-4	麥克風與單聲道喇叭空間配置圖.....	34
圖 4-5	模擬之 RIR.....	35
圖 4-6	RIR 說明圖.....	36
圖 4-7	TDLMS learning curve during single-talk.....	39

圖 4- 8	FDLMS learning curve during single-talk	39
圖 4- 9	FDWiener learning curve during single-talk.....	40
圖 4- 10	FDWiener learning curve during double-talk	40
圖 4- 11	比較各演算法求得之 RIR	41
圖 4- 12	比較各演算法迴聲消除成果.....	42
圖 4- 13	空間配置與響應.....	43
圖 4- 14	AEC Coefficient error norm.....	43
圖 4- 15	比較估測和模擬之 RIR	44
圖 4- 16	模擬 MSE 結果.....	45
圖 4- 17	真實 MSE 結果.....	45
圖 4- 18	比較時域和頻域運算真實 RIR	45
圖 4- 19	模擬 Double-Talk 輸入訊號	47
圖 4- 20	模擬 Double-Talk 結果	47
圖 4- 21	真實 Double-Talk 輸入訊號	48
圖 4- 22	真實 Double-Talk 結果	48
圖 4- 23	PESQ Improvement of AEC.....	49
圖 4- 24	麥克風與雙聲道喇叭空間配置圖	50
圖 4- 25	喇叭(Left)和麥克風(Mic1)之間的 Cross-Correlation 關係.....	51
圖 4- 26	左右聲道對各麥克風之間的延遲關係圖	52
圖 4- 27	SAEC 在模擬環境之 MSE	53
圖 4- 28	SAEC 在真實環境之 MSE	53
圖 4- 29	SAEC Coefficient error norm	54
圖 4- 30	不同演算法對各聲道估測的 RIR	55
圖 4- 31	真實環境左右聲道之 RIR	56
圖 4- 32	模擬環境參考訊號對輸出的影響.....	58
圖 4- 33	實際環境及參考訊號對輸出的影響.....	58
圖 4- 34	Beamformer SIRI	60
圖 4- 35	Beamformer SERI.....	61
圖 4- 36	Beamformer interference suppression.....	61
圖 4- 37	Beamformer echo suppression	62
圖 4- 38	System interference and echo suppression.....	63
圖 4- 39	Log spectrum distortion (LSD)	64
圖 4- 40	PESQ improvement of system	65
圖 4- 41	original speech.....	65
圖 4- 42	各演算法在 SIR0(dB),SER0(dB)下運算之結果.....	66

表目錄

表 2-1	MVDR 模擬使用參數	18
表 4-1	RIR 環境設定	34
表 4-2	各演算法對 Filter size 的 ERLE 評估	37
表 4-3	各演算法對時間延遲的 ERLE 評估	38
表 4-4	模擬真實之 RIR 環境設定	42
表 4-5	模擬與實際之 ERLE 數值	46
表 4-6	模擬單聲道回聲消除(Double Talk)	48
表 4-7	實際單聲道回聲消除(Double Talk)	48
表 4-8	SAEC 之 RIR 環境設定	50
表 4-9	SAEC 實驗參數	50
表 4-10	左右聲道對個麥克風之間的延遲關係	51
表 4-11	輸入音檔混合情況	61
表 4-12	Mean Opinion Score	64



第一章 緒論

1.1 研究動機

行動通訊數位化之後，由單純的語音通話大量且快速的擴展成智慧型手機和平板電腦等行動通訊產品的發展，電視數位化成為必然趨勢，而此時智慧型電視(Smart TV)也就此誕生。Smart TV 已經不在是只有單向傳輸顯示器之功能，未來應用中會發展出互動平台而在眾多功能內語音命令控制和視訊會議等這些功能是需要先透過遠距離語音訊號處理，提供純化過的語音給遠端通訊者或是語音辨識器做處理。

Smart TV 語音命令控制中噪音源主要來至於電視所播放出來的聲音，於是希望提出解決且排除此噪音之方法，期望留下目標語音訊號壓低噪音源的影響達到純化效果。消除噪音源需先知其特性，在 Smart TV 應用中噪音分為兩類：第一類為電視喇叭播放出之聲音稱為迴聲，已知參考訊號的情形將可以運用聲學迴聲消除演算法降低語音中迴聲成分；第二類為環境中存在的方向性干擾源，可以針對合理範圍需求的环境位置設計空間濾波器，由空間上的特徵來增加語音訊號的訊噪比，達到語音純化的效果。

在遠距離目標聲源可以假設為遠場平面波的傳遞方式處理，當聲源和麥克風距離不夠遠時或接收端有失配現象，必須對此傳統空間濾波器加以改良，在此利用轉移函式比值演算法作為聲源和麥克風之間系統鑑別重要方法，找出兩者之間的相對轉移函式導入最小方差無失真響應波束型成器作為空間濾波器之演算法。

1.2 研究目標

本論文研究目標如下：

1. 探討適應性訊號處理以及研究雙聲道迴聲消除演算法。
2. 探討空間濾波器之演算法。
3. 使用轉移函式比值演算法完成系統鑑別改進空間濾波器。

1.3 相關文獻探討

智慧型電視在語音介面是否成功運用的關鍵在於如何抑制電視播放的聲音、迴聲與環境噪音對目標聲源之影響。目前技術所討論的語音純化應用於智慧型電視方法 J. Park et al.[1]，利用遙控器和智慧型電視裝設麥克風分別裝配主要及次要麥克風用於接收目標聲源及估測環境中之干擾訊號進行抑制，前提是主要麥克風和使用者間距離需近距操作方能有良好的語音強化表現。而本論文提出具有空間特徵抑制電視播放聲源之方法，需要具有適應性及陣列訊號處理[2][3]觀念；另外使用聲學迴聲消除法，降低由喇叭播放的聲音對語音品質的影響。本論文利用頻域維納濾波器的架構[4][5]設計 AEC，而在 AEC 的運算中時間延遲問題[6]是很實際且重要的，絕對要避免參考訊號與接收訊號兩者時間延遲大於 FIR 長度，否則將會無法準確估測出迴聲路徑導致消除迴聲失效。因此透過相關係數方法找出兩訊號間延遲來解決此問題。

麥克風陣列可達到空間濾波的功能，一般而言稱之為波束形成器 Beamformer[7]，Beamformer 用於麥克風陣列早用於第二次世界大戰，接著慢慢衍生出諸如 Fourier Beamformer、LCMV[8]、MVDR(Minimum Variance Distortionless Response Beamformer)[9][10]、GSC (Generalized Sidelobe Canceller) [11][12]、TF-GSC (Transfer Function-GSC)[13]、DTF-GSC(Dual-source TF-GSC)[14][15]等。在各種 Beamformer 中最簡單實現的技術為 Fourier Beamformer，但是它需要較大的麥克風陣列才可以達到較好的效果，這是因為越多的麥克風可以形成較尖銳的 Beam Pattern，進而減少其他非聲源角度之干擾源影響。這樣的缺點會造成為了增加效果而必須一直擴大麥克風陣列的體積，Capon[9]提出了一種可以自動消除干擾源的 MVDR，它除了可以將所量測出之聲源角度作完整聲音之接收，並且還可讓非聲源角度之聲音接收達到最低，相對於 Fourier Beamformer 增加了抑制干擾源的效果。另外 Frost[8]提出 LCMV 的方法由，這個方法需先計算出目標聲源的角度以及干擾源的角度，Beamformer 的技術針對目標聲源收音並且濾除其他方向之雜訊，則此系統將會變得更為實用，缺點是計算量提高，優點是提供固定的方向對干擾進行壓抑。Griffiths 和 Jim[11]根據 Frost 的演算法修改成 GSC 架構。GSC 架構分成三個部分：第一部分固定波束形成(Fixed Beamformer, FBF)目的為滿足設計的限制。第二個部分阻擋陣列(Blocking Matrix, BM)目的為產生只有包含雜訊的訊號。第三個部分沒有限制的 LMS 演算法，主要的目的是消除固定波束形成(FBF)的雜訊。LCMV 和 GSC 方法經 Buckley [12]證明是具有等價的效果，而 GSC 把約束問題轉為非

約束型問題簡化了運算量。Griffiths 和 Jim 的演算法的缺點為假設麥克風陣列與聲源的關係只有簡單的時間差，因此 Sharon Gannot[13]於 2001 年提出以估測轉移函式比值取代轉移函式，由於轉移函式不易取得但相對之間比值關係是較容易求得，因此麥克風陣列與目標聲源之間的關係取代單純只有假設時間差的關係。

相對轉移函式在多聲道免提式通訊裝置應用中為重要的演算法，由於它可以將系統鑑別出麥克風與目標聲源間的關係，分別有 O. Shalvi[16]提出利用非穩態訊號和 I. Cohen [17]提出利用語音訊號求出 RTF 的估測，經系統鑑別後的資訊可以應用於音源分離[18]或是空間濾波器的設計[13][14]。最後在 AEC 的設計上為了符合真實電視雙聲道的不同也利用立體聲迴聲消除(Stereophonic Acoustic Echo Cancellation, SAEC)演算法[21] [29] [30][31]來實現本論文中迴聲消除架構。

1.4 論文架構

本論文所使用的原理及技術分成兩大部份，第一部份適應性訊號處理，其中 AEC 理論是適應性處理中很經典的應用，不僅介紹時域 LMS 基本演算法基礎，更研究及實現頻域維納濾波器，甚至為了符合智慧型電視應用將單聲道迴聲消除推廣為利用立體聲迴聲消除演算法。第二部分是說明本論文所使用的陣列訊號處理，選用的適應性空間濾波器為 MVDR 波束形成器結合適應性雜訊消除演算法。最後將轉移函式比值演算法鑑別出目標聲源與麥克風陣列間的相對關係，用來改善 MVDR 的效能。論文最後會呈現系統各項功能結果透過客觀評估標準測試及分析並對研究成果作出結論。論文章節架構如下：

第二章、適應性陣列訊號處理

第三章、系統架構與相關技術介紹

第四章、實驗結果與分析

第五章、結論

第二章 適應性陣列訊號處理

2.1 陣列訊號處理

數個感應器排成特定的形狀，接收來自空間中所傳遞的訊號，並經過訊號處理，此技術稱為陣列訊號處理。在陣列訊號處理領域中，依照其目的不同，大致可以將其研究領域分為兩大類，第一種類的研究著重於估測訊號的數量或在空間中的方位，此類研究一般來說稱為到達角估測 (Direction of arrival estimation)。而另一種類的研究則是利用訊號的空間關係，希望能夠對不同方向的訊號作出不同的增益，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，這一類的研究一般稱之為波束形成(Beamforming)，也就是一種空間濾波器(Spatial Filter)。

在陣列訊號處理理論中，基於兩個假設

1. 窄頻訊號 (Narrow band signal)
2. 遠場平面波 (Far field plane wave)

假設一陣列感應器擺置如圖 2-1 所示， $s(t)$ 為原始訊號， $n(t)$ 為雜訊

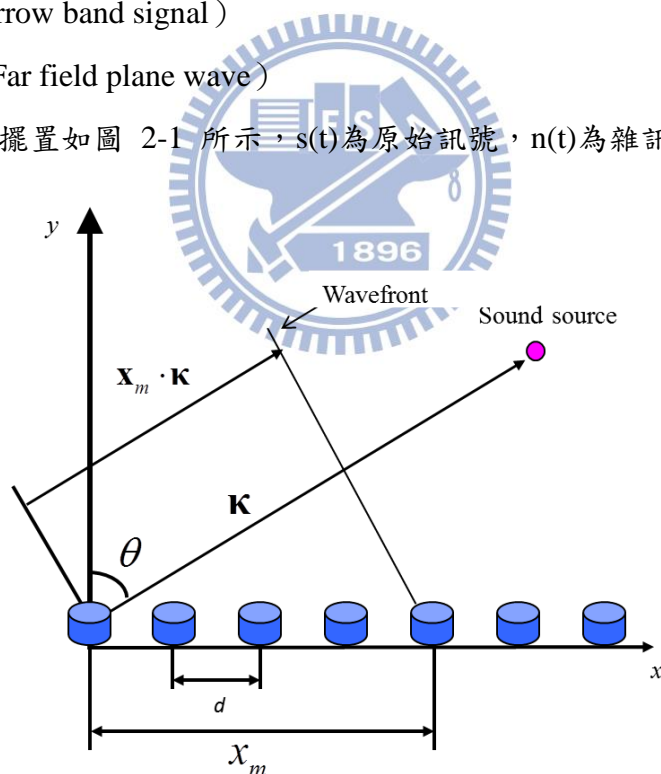


圖 2-1 陣列模型

則 M 個感應器輸出可寫成下列向量形式

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{j\omega_c \frac{\bar{x}_1 \cdot \boldsymbol{\kappa}}{c}} \\ \vdots \\ e^{j\omega_c \frac{\bar{x}_M \cdot \boldsymbol{\kappa}}{c}} \end{bmatrix} s(t)e^{j\omega_c t} + \begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix} \quad (2.1.1)$$

$$= \mathbf{a}(\boldsymbol{\kappa})r(t) + \mathbf{n}(t)$$

$k_c = \frac{\omega_c}{c} = \frac{2\pi}{\lambda_c}$ k_c 稱為波數(Wavenumber)而 λ_c 為波長， c 為波速

\bar{x}_m 接收器位置坐標 $m = 1, 2, \dots, M$

$\boldsymbol{\kappa} = (\sin \theta, \cos \theta)$ 單位向量

$\mathbf{a}(\boldsymbol{\kappa})$ 稱為陣列拓樸向量(Array Manifold Vector or Steering Vector)，期中包含了訊號傳遞到感應器之間時間關係

$r(t) = s(t)e^{j\omega_c t}$ 聲源窄頻波形(Narrowband Source Waveform)

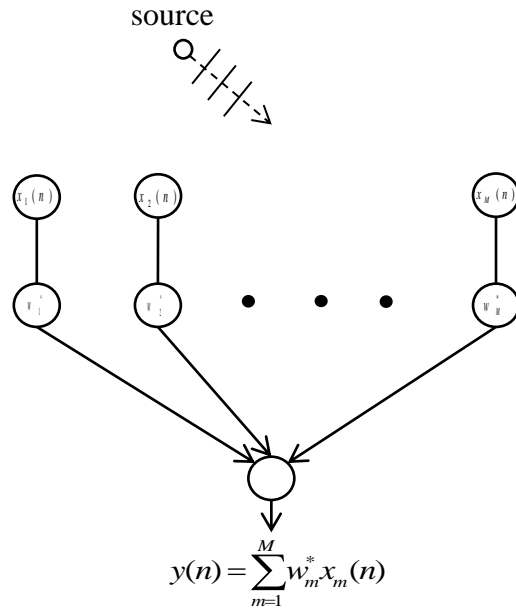
不同的陣列型態會造成不同的空間響應，並會決定陣列的空間解析度，舉例來說，一維的陣列只能解析一維的空間維度，而二維的陣列就可解析二維的空間維度，論文中所實現的陣列型態屬於一維陣列的一部分，因此本章節將介紹屬一維陣列的均勻線性陣列。

均勻線性陣列 (Uniform Linear Array)，是指一組陣列感應器以線性方式排列，並且感應器之間的距離相等，圖 2-1 其實就是表示一個均勻線性陣列。若以第一個感應器當作參考點，每個感應器對於訊號源相對角度皆為 θ ，感應器的坐標表示為 $\bar{x}_m = ((m-1)d, 0)$ ，波程差表示為 $\bar{x}_m \cdot \boldsymbol{\kappa} = (m-1)d \sin \theta$ 則第 M 個感應器收到的時間為訊號到達第一個感應器後延遲 $\frac{(M-1) \cdot d \cdot \sin \theta}{c}$ ，因此均勻線性陣列的陣列拓樸向量可寫成如

(2.1.2) 式，均勻線性陣列的優點是容易實現且公式容易推導，運算量較其它多維陣列型態低，但缺點為只能對一維空間作解析。

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{j\omega_c \frac{d \sin \theta}{c}} & \dots & e^{j\omega_c \frac{(M-1)d \sin \theta}{c}} \end{bmatrix}^T = \begin{bmatrix} 1 & e^{jk_d \sin \theta} & \dots & e^{j(M-1)k_d \sin \theta} \end{bmatrix}^T \quad (2.1.2)$$

使用數位訊號處理分析，空間濾波器 (Spatial Filter) 指的就是將感應器輸出乘上各自加權值的線性組合，如圖 2-2(a)所示



(a) 線性權重組合



(b) 等價 SISO 系統

圖 2-2 空間濾波器架構

因此均勻線性陣列的總輸出可寫成如下形式：

$$y(n) = \sum_{m=1}^M w_m^* x_m(n) = \mathbf{w}^H \mathbf{x}(n) \quad (2.1.3)$$

其中 w_m 是陣列加權係數， $*$ 和 H 分別代表共軛複數和共軛轉置矩陣。假設只有一個目標聲源的情況，可以得到

$$y(n) = \mathbf{w}^H \mathbf{x}(n) = \mathbf{w}^H [\mathbf{a}(\boldsymbol{\kappa})r(n)] + \mathbf{w}^H \mathbf{n}(n) \quad (2.1.4)$$

可以等價為一個線性系統 $r(n)$ 與 $y(n)$ 分別為系統的輸入與輸出，圖 2-2(b)所示。而此陣列響應方程式 $\mathbf{w}^H \mathbf{a}(\boldsymbol{\kappa})$ 稱之為 Array Pattern 或 Beam Pattern 表示如下

$$p(\theta) = \mathbf{w}^H \mathbf{a}(\boldsymbol{\kappa}) = \sum_{m=1}^M w_m^* e^{jk_c(m-1)d \sin \theta} \quad (2.1.5)$$

若將以均勻加權線性陣列(Uniformly Weighted Linear Arrays)為例，均勻加權的權重為

$$w_m = \frac{1}{M}, \quad m = 1, 2, \dots, M \quad (2.1.6)$$

寫為矩陣型式

$$\mathbf{w} = \frac{1}{M} \mathbf{1} \quad (2.1.7)$$

則 $p(\theta)$ 可化簡成如下所示：

$$p(\theta) = \frac{1}{M} \sum_{m=1}^M e^{jk_c(m-1)d \sin \theta} = \frac{1}{M} \frac{e^{jk_c M d \sin \theta} - 1}{e^{jk_c d \sin \theta} - 1} = \frac{1}{M} e^{j \frac{k_c(M-1)d}{2} \sin \theta} \frac{\sin\left(\frac{k_c M d}{2} \sin \theta\right)}{\sin\left(\frac{k_c d}{2} \sin \theta\right)} \quad (2.1.8)$$

若將 $p(\theta)$ 取 Magnitude 可得其 Beam Pattern，如圖 2-3 所示。

從圖 2-3 可看出，不同角度入射的訊號會有不同的增益，而角度和增益的關係是由陣列的加權值所決定，因此波束形成就可達到空間濾波的效果，而在波束形成理論中，就是用適當的方法去計算出加權值，將訊號作空間濾波，就可得到想要的訊號。

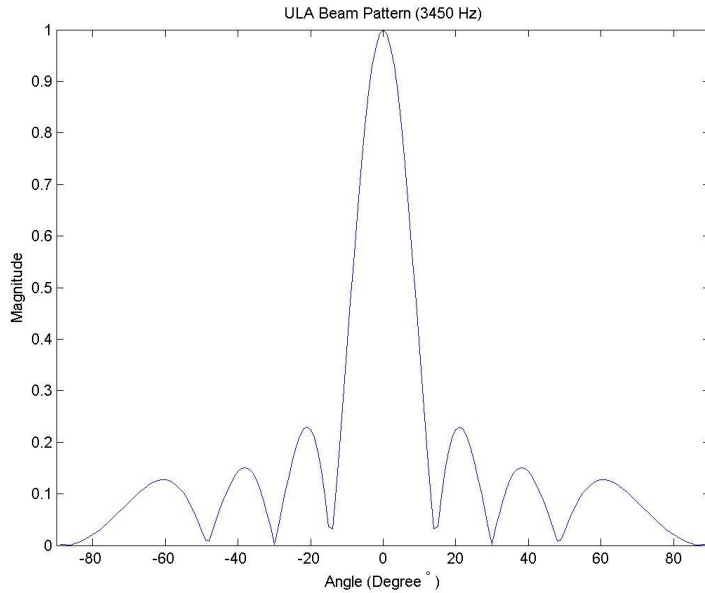


圖 2-3 均勻線性陣列空間響應(M=8, frequency=3450 Hz, d=0.05 m)

將 (2.1.8) 式取絕對值可得

$$|p(\theta)| = \frac{1}{M} \left| \frac{\sin\left(\frac{k_c M d}{2} \sin \theta\right)}{\sin\left(\frac{k_c d}{2} \sin \theta\right)} \right| \quad (2.1.9)$$

由 (2.1.9) 式可以看出 $|p(\theta)|$ 對 $\sin \theta$ 是一週期為 λ_c/d 的週期性的函式，關係圖如圖 2-4 所示。

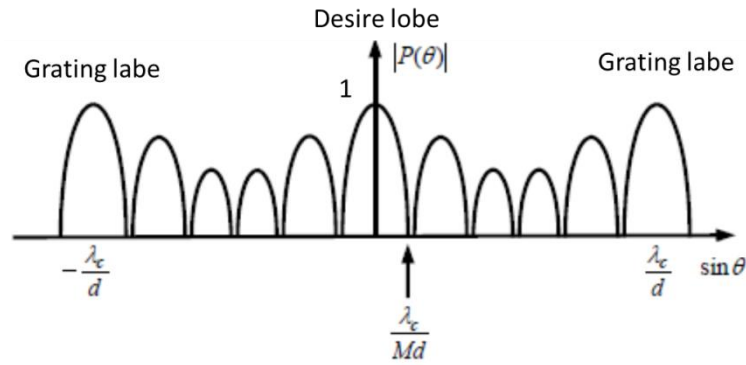


圖 2-4 Grating Lobe 示意圖

在均勻線性陣列中，預期訊號的角度在 $\pm 90^\circ$ 間，而在這角度之間我們希望 Main lobe 只會出現一次，如果 Main lobe 出現兩次以上，則會造成非預期的訊號被接收近來。從圖 2-4 得知，Grating Lobe 發生在 $\sin \theta = \lambda_c/d$ 的時候，因此若讓 $\lambda_c/d > \sin 90^\circ = 1$ ，則可避免在 $\pm 90^\circ$ 間出現兩個以上的 Main lobe。在空間的取樣也需要遵守 Nyquist spatial sampling criterion 避免 Spatial aliasing 發生，取樣頻率必須是訊號頻率的兩倍以上。而通常我們都會選取 $d \leq \lambda_c/2$ ，可以同時避免 Grating Lobe 和 Spatial aliasing 的問題。麥克風陣列的孔距(Aperture)會影響波束寬度(Beam Width, BW) $BW = \frac{2\lambda_c}{Md}$ ，因此孔距與波束寬度的關係為反比關係 $Aperture \uparrow \Leftrightarrow BW \downarrow$ ，可得麥克風數目越多產生越大的孔距和越窄的波束寬度獲得較高的解析度。

2.2 適應性訊號處理

一般而言，濾波器的係數設計出來後都是固定的，並不會自動的變動。而適應性濾波器指的是能根據輸入信號，用訊號處理的技巧來適應性地調整濾波器係數，讓濾波效果更能適應現在環境，以完成某些特定的需要。

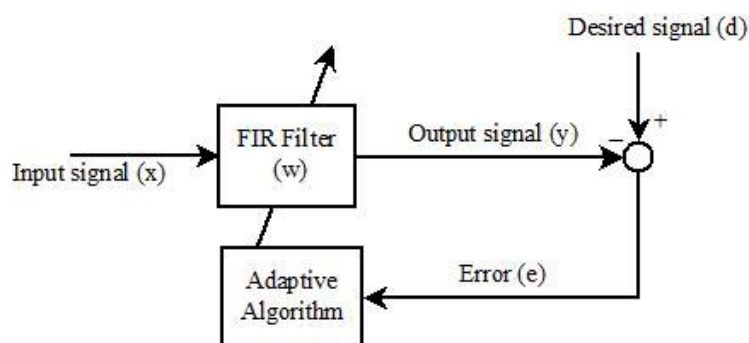


圖 2-5 適應性濾波器處理架構

適應性濾波器處理架構圖如圖 2-5 所示，當訊號(x)輸入適應性濾波器處理之後，輸出訊號(y)與希望達成的訊號(d)之間的差產生誤差訊號(e)，將誤差訊號代入適應性演算法做為調整適應性濾波器係數重要資訊，如此經由誤差訊號及適應性演算法不斷的調整適應性濾波器的係數，係數會不斷的變動，最後達到某個穩定的值，此時系統輸出訊號與希望達成的訊號就會非常接近。

從通訊迴聲產生原因看可分為聲學迴聲(Acoustic Echo)和線路迴聲(Line Echo)，相對應的消除技術稱為聲學迴聲消除(Acoustic Echo Cancellation, AEC)和線路迴聲消除(Line Echo Cancellation, LEC)，而本論文所會探討的問題是應用在免持或是視訊會議的裝置，麥克風接受到經由喇叭所發出的聲音引起的聲學迴聲的系統稱之為 LEM (loudspeaker-enclosure-microphone) system，如圖 2-6 即為此系統，當雙方通話的情況下 speech 1 透過喇叭播放同時會讓對方麥克風接收到，發生這種迴授情況對於語音品質來說不好的，因此將會介紹如何利用適應性訊號處理的方法來解決聲學迴聲的問題。

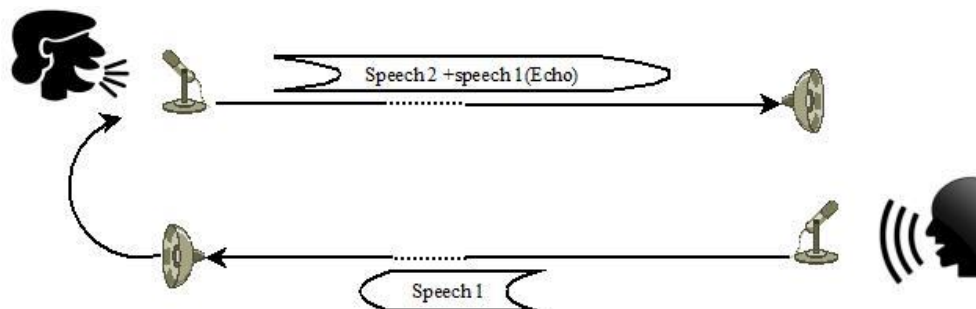


圖 2-6 空間聲學反射產生的聲學迴聲

關於如何解決 AEC 會介紹維納解(Wiener Solution)以及計算量較小的最小平均平方方法(Least-Mean-Square, LMS)。圖 2-7 為 AEC 的主體，透過估測喇叭聲源與麥克風收到訊號之間的關係來達到消除迴聲的功能。換句話說，利用喇叭聲源與麥克風訊號就可以估測出迴聲路徑(echo path)。 $x(n)$ 為喇叭欲播放的聲源、 W 為真實的迴聲路徑、 $v(n)$ 環境雜音、 $s(n)$ 與 $y(n)$ 分別為目標聲源與麥克風收到的訊號。 \hat{W} 即為估測的迴聲路徑。

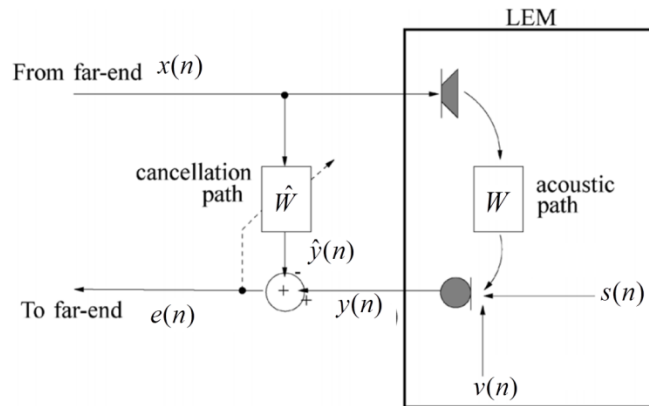


圖 2-7 AEC 主體架構

透過維納濾波器估測迴聲路徑，假設估測的迴聲路徑為一 FIR 濾波器 \hat{W}

$$\hat{W} = [w_0, w_1, \dots, w_{N-1}]^T \quad (2.2.1)$$

其中 N 為 FIR 濾波器長度。喇叭欲播出訊號 $x(n)$ 為往前 N 筆的訊號向量

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^T \quad (2.2.2)$$

因此濾波器的輸出可寫成 \hat{W} 與 $\mathbf{x}(n)$ 的內積

$$\hat{y}(n) = \hat{W}^H \mathbf{x}(n) \quad (2.2.3)$$

估測迴聲與麥克風訊號的誤差

$$e(n) = y(n) - \hat{y}(n) \quad (2.2.4)$$

並且維納濾波器的最佳化標準為 Minimum mean square error(MMSE)，定義目標函數最

小值 ($\min_{\hat{W}} J(\hat{W})$)

$$\begin{aligned}
J(\hat{W}) &\equiv E\{|e(n)|^2\} \\
&= E\{e(n)e^H(n)\} \\
&= E\{(y(n) - \hat{y}(n))(y(n) - \hat{y}(n))^H\} \\
&= E\{(y(n) - \hat{W}^H \mathbf{x}(n))(y(n) - \hat{W}^H \mathbf{x}(n))^H\} \\
&= E\{y^2(n)\} - E\{y(n)\mathbf{x}^H(n)\hat{W}\} - E\{\hat{W}^H \mathbf{x}(n)y^H(n)\} + E\{\hat{W}^H \mathbf{x}(n)\mathbf{x}^H(n)\hat{W}\} \\
&= E\{y^2(n)\} - \mathbf{r}_{yx} \hat{W} - \hat{W}^H \mathbf{r}_{xy} + \hat{W}^H \mathbf{R}_{xx} \hat{W}
\end{aligned} \tag{2.2.5}$$

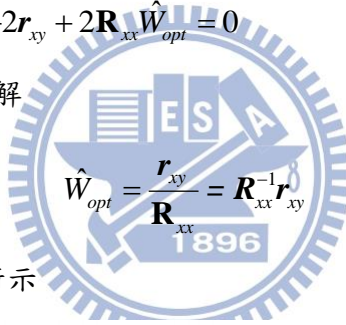
其中 $\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ 為訊號 $\mathbf{x}(n)$ 的自相關矩陣(Auto-correlation matrix) ,

$\mathbf{r}_{xy} = E\{\mathbf{x}(n)y^H(n)\}$ 為訊號 $\mathbf{x}(n)$ 與 $y(n)$ 的互相關向量(Cross-correlation vector)。

求 $\min_{\hat{W}} J(n)$ 的解，由(2.2.5)式對 \hat{W} 微分求極值方法可得

$$\begin{aligned}
\nabla_{\hat{W}} J(n) &= \nabla_{\hat{W}} E\{|e(n)|^2\} \\
&= \nabla_{\hat{W}} \left\langle E\{y^2(n)\} - \mathbf{r}_{yx} \hat{W} - \hat{W}^H \mathbf{r}_{xy} + \hat{W}^H \mathbf{R}_{xx} \hat{W} \right\rangle \\
&= -2\mathbf{r}_{xy} + 2\mathbf{R}_{xx} \hat{W}_{opt} = 0
\end{aligned}$$

最佳解 \hat{W}_{opt} 即為維納濾波器的解



$$\hat{W}_{opt} = \frac{\mathbf{r}_{xy}}{\mathbf{R}_{xx}} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xy} \tag{2.2.6}$$

維納濾波器架構如下圖 2-8 所示

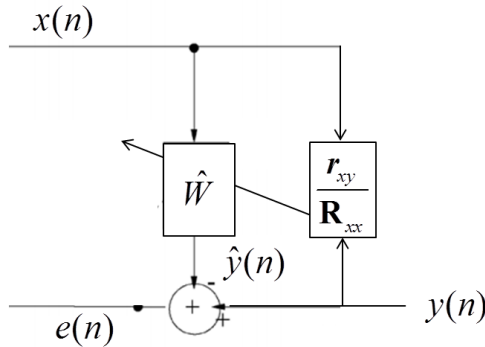


圖 2-8 維納濾波器架構

想要實現維納濾波器可以使用 Sample Matrix Inversion, SMI 方法透過時間平均得到 \mathbf{R}_{xx}^{-1} 和 \mathbf{r}_{xy} 的估計值使得 FIR 濾波器的解為最佳解，由於 FIR 濾波器的係數通常長度的挑選是因應系統中延遲速度和迴聲路徑響應所決定，如果 FIR 長度太短很容易因為些微延遲而導致無法算出正確的 \hat{W} ，FIR 長度為 M 時 \mathbf{R}_{xx} 為 $M \times M$ 矩陣此時計算 \mathbf{R}_{xx}^{-1} 的時候會出現相當龐大的運算量，所以接下來介紹計算量較小且較容易實現的 Time Domain

LMS 演算法以及 Frequency Domain Wiener Filter 演算法。

2.2.1 Least-Mean-Square Adaptive Filter in Time Domain

同樣要解圖 2-7 AEC 架構下的問題，如果要在時域實現最佳解的維納濾波器是很困難的。因此，Widrow 和 Hoff [19]提出了另一種可疊代的維納 FIR 濾波器實現方法為 LMS 演算法，LMS 指的是找出一組權重 \hat{W} 使得誤差平方項最小。

目標函數如(2.2.7)式所示

$$\min_{\hat{W}} J(\hat{W}) \equiv \min_{\hat{W}} E\{|e(n)|^2\} = \min_{\hat{W}} E\{(y(n) - \hat{W}^H \mathbf{x}(n))(y(n) - \hat{W}^H \mathbf{x}(n))^H\} \quad (2.2.7)$$

(2.2.7)式的意義就是找出一組 \hat{W} 使誤差平方項最小，而 \hat{W} 的找法則需用 Steepest-Descend Method，如(2.2.8)式，

$$\hat{W}(n+1) = \hat{W}(n) + \mu p, \quad n \geq 0 \quad (2.2.8)$$

其中(2.2.8)式意義為從 $\hat{W}(n)$ 出發，並前進 μp 的距離， μ 為一個比重稱為步長(Step Size)。而 p 的選取必須從(2.2.7)式開始推導，將(2.2.7)式展開可得

$$J(\hat{W}) = E\{y^2(n)\} - \mathbf{r}_{xy}^H \hat{W}(n) - \hat{W}^H(n) \mathbf{r}_{xy} + \hat{W}(n)^H \mathbf{R}_{xx} \hat{W}(n) \quad (2.2.9)$$

為了找一組 \hat{W} 使 $J(w)$ 最小，對(2.2.9)式取 $\nabla_{\hat{W}}$ 得

$$\nabla_{\hat{W}} J(\hat{W}) = -2\mathbf{r}_{xy} + 2\mathbf{R}_{xx} \hat{W}(n) \quad (2.2.10)$$

因此，為了讓 $\hat{W}(n)$ 往 $J(w)$ 最低處的方向與強度前進，我們取

$$p = -\nabla_{\hat{W}} J(\hat{W}) = 2\mathbf{r}_{xy} - 2\mathbf{R}_{xx} \hat{W}(n) \quad (2.2.11)$$

故(2-7)式可寫為

$$\hat{W}(n+1) = \hat{W}(n) + 2\mu \left[\mathbf{r}_{xy} - \mathbf{R}_{xx} \hat{W}(n) \right] \quad n \geq 0 \quad (2.2.12)$$

在實做上， \mathbf{r}_{xy} 和 \mathbf{R}_{xx} 估計值可用離散形式近似於瞬間值：

$$\mathbf{r}_{xy} = \mathbf{x}(n)y^H(n), \quad \mathbf{R}_{xx} = \mathbf{x}(n)\mathbf{x}^H(n) \quad (2.2.13)$$

所以(2.2.13)式代入(2.2.12)式可寫為：

$$\begin{aligned} \hat{W}(n+1) &= \hat{W}(n) + 2\mu \left[\mathbf{x}(n)y^H(n) - \mathbf{x}(n)\mathbf{x}^H(n)\hat{W}(n) \right] \\ &= \hat{W}(n) + 2\mu \mathbf{x}(n) \left[y^H(n) - \mathbf{x}^H(n)\hat{W}(n) \right] \quad n \geq 0 \end{aligned} \quad (2.2.14)$$

因此，LMS Algorithm 可整理如下：

Filter out :

$$\hat{y}(n) = \hat{W}^H(n)x(n) \quad (2.2.15)$$

Error function :

$$e(n) = y(n) - \hat{y}(n) \quad (2.2.16)$$

Update weight :

$$\hat{W}(n+1) = \hat{W}(n) + 2\mu x(n)e^H(n) \quad n \geq 0 \quad (2.2.17)$$

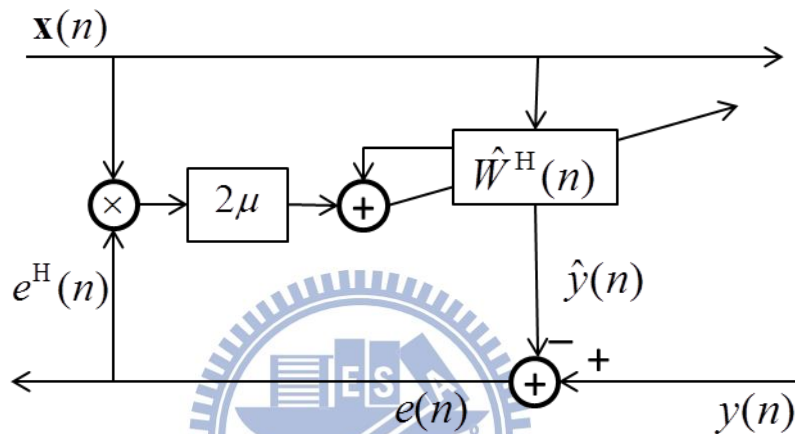


圖 2-9 LMS 運算架構

圖 2-9 LMS 運算架構利用訊號流程圖的形式呈現，在 LMS 演算法中，為了確保其收斂， μ 的範圍必須為 $0 < \mu < \frac{1}{\lambda_{\max}}$ ，其中 λ_{\max} 表示為自相關矩陣 \mathbf{R}_{xx} 的最大特徵值。

然而在實際應用中， \mathbf{R}_{xx} 的具體值是不知道的，參數 μ 的值也需要試探性的選擇。若 μ 取值小，能保證收斂，但需要注意的是如果值取的過小時，收斂速度將會非常慢；若 μ 取值大，可以提高收斂速度，但取值過大時會以訊號變動過快失真為代價。

在實際的應用中使用時域 LMS 解決 AEC 的問題會遇到的困難有收斂速度不夠快的問題，因此，也有在頻域運算根據數位訊號處理理論可以透過重疊儲存相加方法當 50% 重疊時運算效率達到最高的優點來解決 AEC 的問題，J. Shynk [20] 提出 Fast Block LMS (FBLMS) 準確的描述了 Block LMS 算法的頻域實現，本論文利用頻域估測迴聲路徑的方法在 2.2.2 章節中詳加說明。

2.2.2 Wiener Filter in Frequency Domain

此章節介紹維納濾波器在頻域實現 AEC 的效果，由 F. Yang et al. [21]、J.S. Hu et al. [22] 文獻了解到利用維納濾波器在頻域可做各式應用，利用頻域的維納濾波器比起 LMS 和 FBLMS 在環境變異較大的應用中有較優異的效能，因為它能直接利用開迴路的最佳近似解權重估測不必透過 Steepest-Descend Method 閉迴路逼近最佳解。圖 2-10 為頻域維納濾波器的架構圖。

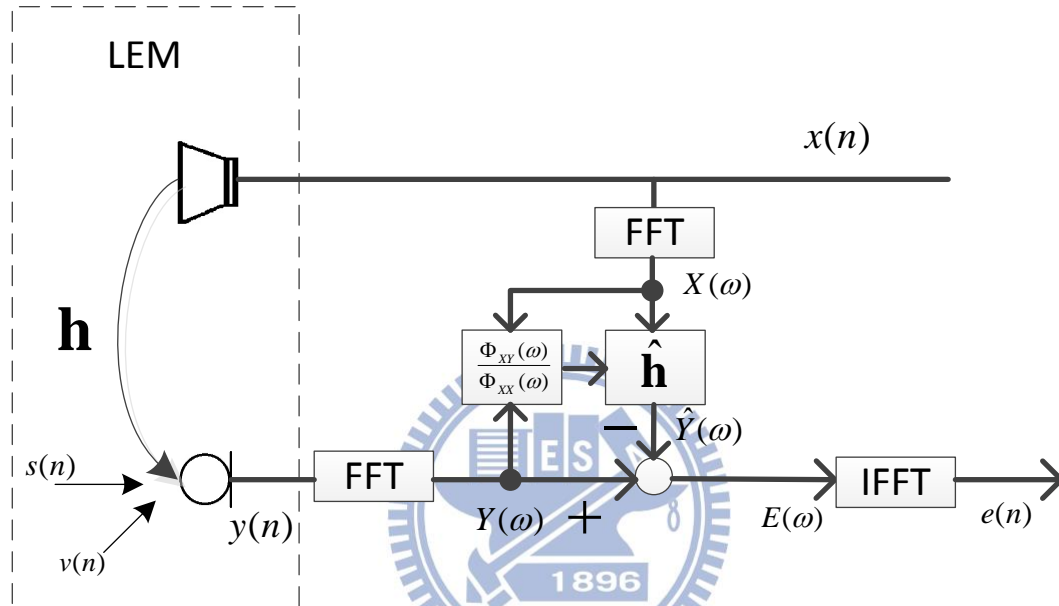


圖 2-10 Wiener Filter in Frequency Domain

由圖 2-10 LEM 環境中麥克風收到訊號的訊號 $y(n)$ 可以表示為

$$y(n) = x(n) * \mathbf{h} + s(n) + v(n) \quad (2.2.18)$$

在此 $*$ 代表的是摺積運算、 $x(n)$ 代表喇叭播放訊號、 \mathbf{h} 為迴聲路徑或稱為房間脈衝響應 (Room Impulse Response, RIR) $v(n)$ 環境雜音和 $s(n)$ 在 LEM 環境中的目標聲源。接下來將推導頻域維納濾波器，先將(2.2.18)式兩邊都做傅利葉轉換可得

$$Y(\omega) = X(\omega)H(\omega) + S(\omega) + V(\omega) \quad (2.2.19)$$

其中 $Y(\omega)$, $X(\omega)$, $H(\omega)$, $S(\omega)$, $V(\omega)$ 分別是 $y(n)$, $x(n)$, \mathbf{h} , $s(n)$, $v(n)$ 的傅利葉轉換。透過頻域維納濾波器估測迴聲路徑 $\hat{\mathbf{h}}(\omega)$ ，可將濾波器的輸出可寫成 $\hat{\mathbf{h}}(\omega)$ 與 $X(\omega)$ 的內積

$$\hat{Y}(\omega) = \hat{\mathbf{h}}^*(\omega)X(\omega) \quad (2.2.20)$$

其中 $X(\omega)$ 是喇叭欲播出訊號， $*$ 代表共軛複數。維納濾波器的最佳化標準為 MMSE，定義目標函數最小值

$$\min_{\hat{\mathbf{h}}(\omega)} J \equiv \min_{\hat{\mathbf{h}}(\omega)} E\{|E(\omega)|^2\} = \min_{\hat{\mathbf{h}}(\omega)} E\{|Y(\omega) - \hat{\mathbf{h}}^*(\omega)X(\omega)|^2\} \quad (2.2.21)$$

可以經由先前(2.2.6)推導得到(2.2.21)結果為

$$\hat{\mathbf{h}}(\omega) = \frac{\Phi_{XY}(\omega)}{\Phi_{XX}(\omega)} \quad (2.2.22)$$

其中 $\Phi_{XX}(\omega)$ 代表 $x(n)$ 最自己本身的功率頻譜密度 Auto-Power Spectral Density (PSD) 且 $\Phi_{XY}(\omega)$ 代表著 $y(n)$ 和 $x(n)$ 之間的 Cross-PSD

$$\begin{aligned} \Phi_{XY}(\omega) &= \sum_{\tau=-\infty}^{\infty} E\{x(n)y(n-\tau)\}e^{-j\omega\tau} \\ &= \sum_{\tau=-\infty}^{\infty} R_{xy}(\tau)e^{-j\omega\tau} \end{aligned} \quad (2.2.23)$$

想要估計 $\hat{\mathbf{h}}(\omega)$ 時需要先得到 $\Phi_{XY}(\omega)$ 及 $\Phi_{XX}(\omega)$ ，以下為 Cross(Auto)-PSD 的估計

$$\Phi_{XY}(i, \omega_k) = \alpha \Phi_{XY}(i-1, \omega_k) + (1-\alpha)X(i, \omega_k)Y^*(i, \omega_k) \quad (2.2.24)$$

$$\Phi_{XX}(i, \omega_k) = \alpha \Phi_{XX}(i-1, \omega_k) + (1-\alpha)X(i, \omega_k)X^*(i, \omega_k) \quad (2.2.25)$$

在此參數 α 是一個平滑參數或遺忘因子，用來控制疊代過程的有效記憶；每一個音框 (frame) 的指標用 i 表示，頻率由連續 ω 變成離散 $\omega_k = 2\pi k / N$ ， $k \in \{0, 1, \dots, N-1\}$ 而其中 N 根據 FFT 的音框大小 (frame size) 所決定。當運算環境為廣義平穩時 $\Phi_{XY}(i, \omega_k)$ 和 $\Phi_{XX}(i-1, \omega_k)$ 可以直接使用 $X(i, \omega_k)Y^*(i, \omega_k)$ 和 $X(i, \omega_k)X^*(i, \omega_k)$ 當作 PSD 的估測，然而運算環境為非平穩時就要採用簡單的一階遞迴平均 $X(i, \omega_k)Y^*(i, \omega_k)$ 和 $X(i, \omega_k)X^*(i, \omega_k)$ 而得， α 功用在於可以調整平滑程度取用範圍為 $0 < \alpha < 1$ 的常數，舉例來說 α 取 0.95 意味著每二十筆資料做一次平均。經由得知 $\Phi_{XY}(\omega)$ $\Phi_{XX}(\omega)$ 可以估計出(2.2.22)式，因此，經過下列運算就能迴聲消除

$$E(\omega) = Y(\omega) - \hat{Y}(\omega) \quad (2.2.26)$$

最後再透過 IFFT 轉換即可求得時域訊號。

2.3 適應性空間濾波器

麥克風陣列可以利用訊號的空間特徵，對空間內的訊號進行濾波處理，這種濾波行為是運用角度或波數的相關性進行描述。從頻域層面分析空間濾波器類似於 FIR 架構，利用複數增益組合陣列接收器的輸出，根據訊號的空間相關性對訊號進行增強或抑制。通常，我們想進行空間濾波，使得從一個或一組特定的角度來的訊號通過有效的組合得到增強，而使從其它角度到來的雜訊通過相消性的組合得到抑制。

設計陣列，並達到某種性能準則，需要在陣列結構、陣列數目、SNR 以及很多因素之間進行折衷考量才能完成。適應性空間濾波器應用方式很多種類，最簡單的稱為 Fixed Beamforming，它是一組經過特殊設計的固定增益，不論任何訊號都與此組增益組合，例如 DAS、Hamming、Hann、Dolph-Chebyshev、LS[3][7]設計出的增益都屬與此種範疇。而另一種稱為 Adaptive Beamforming，它的增益設計與訊號有關，隨著不同的增益條件和輸入訊號做適當的調整，例如 LCMV[8]、MVDR[9]、GSC[11]等架構都是此技術的應用。

2.3.1 Minimum Variance Distortionless Response (MVDR) Beamformer

在智慧型電視的應用只需要針對一個目標聲源方向增強和未知的干擾方向，所以選用 MVDR 做為波束形成器是很適合的方法。空間濾波器如圖 2-2 (a)所示，感應器接收到的訊號 $x(n)$ 中有包含目標聲源 $s(n)$ 和雜訊 $n(n)$ ，由(2.1.1)式可以寫成頻域快拍 (Snapshot)

$$\mathbf{X}(\omega) = \mathbf{X}_s(\omega) + \mathbf{N}(\omega) \quad (2.3.1)$$

$\mathbf{X}_s(\omega)$ 為目標聲源向量， $\mathbf{N}(\omega)$ 為雜訊向量。其中目標聲源向量可以寫為

$$\mathbf{X}_s(\omega) = R(\omega)\mathbf{a}(\omega, k) \quad (2.3.2)$$

$R(\omega)$ 為 $s(n)$ 的頻域快拍， $\mathbf{a}(\omega, k)$ 是一個波數為 k 的平面波對應的陣列拓樸向量，假設雜訊是零平均且和目標聲源訊號是非相關。圖 2-11 所示，利用矩陣操作 $\mathbf{W}^H(\omega)$ 對 $\mathbf{X}(\omega)$ 進行處理。 $\mathbf{W}^H(\omega)$ 的維度為 $1 \times M$ 。我們感興趣的第一個準則稱為無失真準則 (Distortionless Criterion)，這個準則要求在沒有雜訊的情況下，對於任意 $R(\omega)$ ，有

$$Y(\omega) = R(\omega) \quad (2.3.3)$$

接下來透過模擬了解 MVDR 的效果

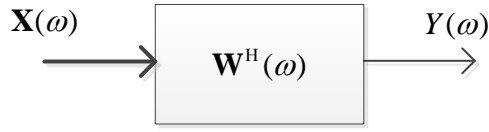


圖 2-11 Matrix processor

在這個約束條件底下，希望在有雜訊的存在的情況下，使得 $Y(\omega)$ 有最小方差 (minimize variance)。因此，

$$Y(\omega) = R(\omega) + Y_n(\omega) \quad (2.3.4)$$

目標是 minimize $E[|Y_n(\omega)|^2]$ 。

無失真約束條件由(2.3.3)式可說明 $Y(\omega) = \mathbf{W}^H(\omega)\mathbf{X}_s(\omega) = \mathbf{W}^H(\omega)R(\omega)\mathbf{a}(\omega, k) = R(\omega)$

意味著目標聲源方向維持原始大小，表示式為

$$\mathbf{W}^H(\omega)\mathbf{a}(\omega, k) = 1 \quad (2.3.5)$$

輸出雜訊的均方值為

$$\begin{aligned} E[|Y_n(\omega)|^2] &= E[|\mathbf{W}^H(\omega)\mathbf{N}(\omega)|^2] = E[\mathbf{W}^H(\omega)\mathbf{N}(\omega)\mathbf{N}^H(\omega)\mathbf{W}(\omega)] \\ &= \mathbf{W}^H(\omega)E[\mathbf{N}(\omega)\mathbf{N}^H(\omega)]\mathbf{W}(\omega) = \mathbf{W}^H(\omega)\mathbf{S}_{NN}(\omega)\mathbf{W}(\omega) \end{aligned} \quad (2.3.6)$$

所以我們得到一個最小化問題

$$\min_{\mathbf{w}} E[|Y_n(\omega)|^2] = \min_{\mathbf{w}} \mathbf{W}^H(\omega)\mathbf{S}_{NN}(\omega)\mathbf{W}(\omega)$$

$$\text{Subject to } \mathbf{W}^H(\omega)\mathbf{a}(\omega, k) = 1$$

利用 Lagrange multiplier 的方式求得最佳解，寫出目標函數(cost function)

$$F = \mathbf{W}^H(\omega)\mathbf{S}_{NN}(\omega)\mathbf{W}(\omega) + \lambda(\omega)[\mathbf{W}^H(\omega)\mathbf{a}(\omega, k) - 1] + \lambda^*(\omega)[\mathbf{a}^H(\omega, k)\mathbf{W}(\omega) - 1] \quad (2.3.7)$$

對 $\mathbf{W}^H(\omega)$ 取 complex gradient 成為 $\nabla_{\mathbf{w}^H} F = 0$ 得

$$\mathbf{W}_o^H(\omega) = -\lambda(\omega)\mathbf{a}^H(\omega, k)\mathbf{S}_{NN}^{-1}(\omega) \quad (2.3.8)$$

假設 $\mathbf{S}_{NN}(\omega)$ 為非奇異性矩陣(Nonsingular Matrix)，利用(2.3.5)式約束條件求得

$$\lambda(\omega) = -[\mathbf{a}^H(\omega, k)\mathbf{S}_{NN}^{-1}(\omega)\mathbf{a}(\omega, k)]^{-1} \quad (2.3.9)$$

把(2.3.9)式代入(2.3.8)式得到 MVDR 最佳解為

$$\mathbf{W}_o^H(\omega) = \frac{\mathbf{a}^H(\omega, k)\mathbf{S}_{NN}^{-1}(\omega)}{\mathbf{a}^H(\omega, k)\mathbf{S}_{NN}^{-1}(\omega)\mathbf{a}(\omega, k)} \quad (2.3.10)$$

MVDR 波束形成器最初由 Capon 推導而來有時也稱成為 Capon 波束形成器。

藉由模擬環境說明 MVDR

Mic number	4	NFFT	256 samples
Overlap	128 samples	Input SIR	0 dB
Distance	0.07 m	Desired signal (woman)	0°
Sample frequency	8k Hz	Interference (White Noise)	45°

表 2-1 MVDR 模擬使用參數

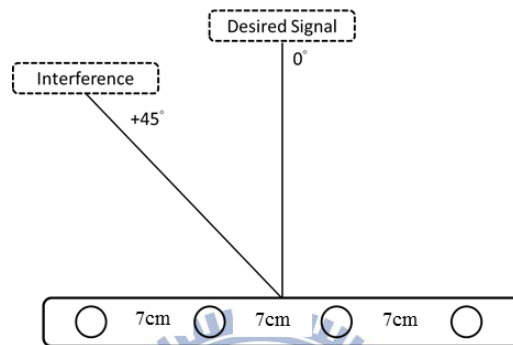


圖 2-12 麥克風與聲源方向關係圖

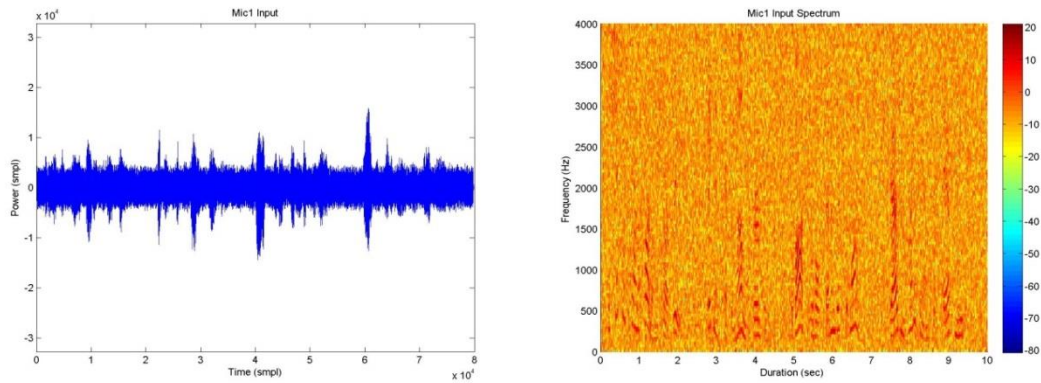
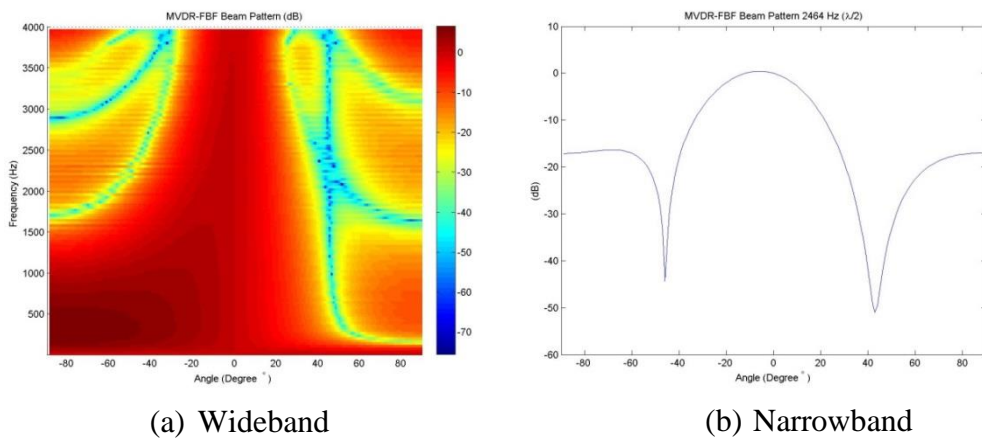


圖 2-13 MVDR 輸入訊號



(a) Wideband

(b) Narrowband

圖 2-14 MVDR Beam Patten

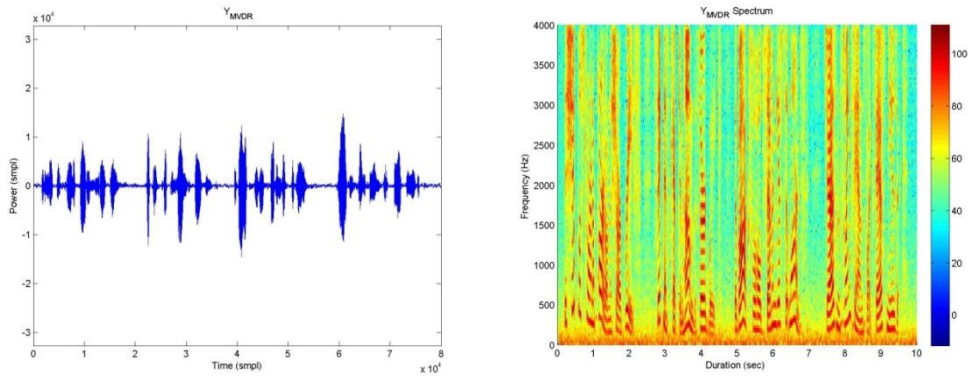


圖 2-15 MVDR 輸出訊號

圖 2-12 為 0° 方向目標聲源和 45° 方向的干擾源對麥克風關係圖，模擬出圖 2-13 輸入訊號。麥克風陣列訊號處理和數位訊號處理在於取樣上有懸殊的差異，在時間域取樣能有很大量充沛的資料量使用，而空間域的取樣只能透過麥克風個數，因此，要透過稀少的空間資料量找出最好的解，藉由 MVDR、LCMV、GSC 等演算法的幫助，對空間訊號處理得出所期望的結果。在本論文使用的 MVDR 將會利用 SMI(Sample Matrix Inversion)的方式求出 $\mathbf{S}_{NV}^{-1}(\omega)$ ，而設計 ULA 的 $\mathbf{a}(\omega, k)$ 陣列拓樸向量有兩種方式可以獲得，第一種方式，聲源特性接近遠場平面波時可以使用時間延遲模型如(2.1.2)式，陣列拓樸向量寫為 $\mathbf{a}(\theta) = [1 \quad e^{jkd \sin \theta} \quad \dots \quad e^{j(M-1)kd \sin \theta}]^T$ ，把希望設計的角度 θ 以及麥克風間距 d 麥克風個數 M 代入即可以得到 MVDR 使用的 $\mathbf{a}(\omega, k)$ 。第二種方式，當聲源和麥克風距離較近或是目標聲源不再固定角度上的時候轉移函式比值法作為聲源和麥克風之間系統鑑別重要方法，找出兩者之間的相對轉移函式(Relative Transfer Function, RTF)。在模擬環境中使用已經準確的設計目標聲源的來向，因此利用時間延遲模型的 $\mathbf{a}(\omega, k)$ 代入(2.3.10)式經由 MVDR 運算可得圖 2-14 (a)為各頻帶之表現(b)為根據麥克風間距所對應波長之表現，在低頻表現較差的原因為波長較長不具方向性的關係。圖 2-15 經過 MVDR 處理過後之訊號，明顯看出干擾源訊號有受到 MVDR 壓抑。

2.3.2 MVDR 結合雜訊估測

理論上 MVDR 會針對目標聲源以外的訊號會進行壓抑，但在本論文提出的方法中， $\mathbf{S}_{NV}^{-1}(\omega)$ 的更新並不會隨時在更新，而是透過使用前的事先訓練(Training)得知，這樣設計是因為在使用環境中一定會有某些固定干擾源存在，如電扇和冷氣等，希望透過 MVDR 的空間前處理可以把比較強的干擾訊號先抑制，希望透過結合雜訊估測適應性演算法可以再針對 MVDR 無法抑制的訊號進行消除。此想法是由 GSC 架構而得，Griffiths 和 Jim[11]根據 Frost 的演算法修改成 GSC 架構。GSC 架構分成三個部分：第一部分固定波束形成器(Fixed Beamformer, FBF)目的為滿足設計的限制。第二個部分阻擋陣列(Blocking Matrix, BM)目的為產生只有包含雜訊的訊號。第三個部分利用 NLMS 演算法消除固定波束形成器輸出雜訊部分稱為適應性雜訊消除器(Adaptive Noise Canceler, ANC)。固定波束形成器設計方式為 2.3.1 章節所介紹的 MVDR，此章節將介紹設計阻擋陣列以及如何透過適應性雜訊消除器消除環境中的雜訊。圖 2- 16 為 MVDR 結合雜訊估測架構。

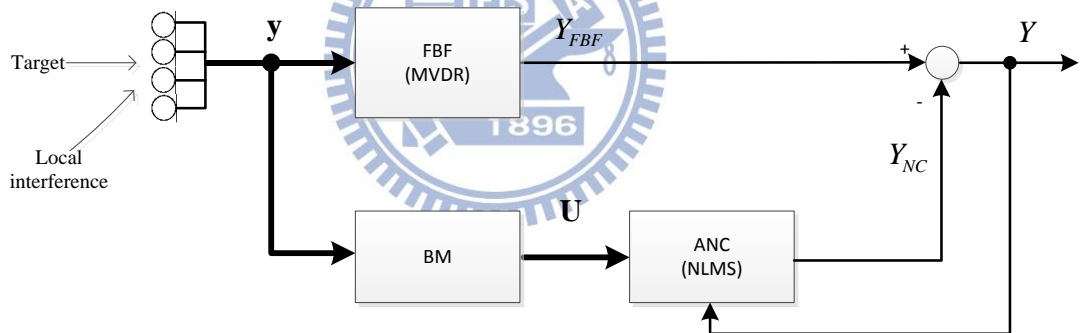


圖 2-16 MVDR 結合雜訊估測

圖 2-16 中，麥克風陣列收到的訊號 \mathbf{y} 為多聲道訊號，經過 MVDR 空間前處理後得到 $Y_{FBF}(\omega)$ 可表示為

$$Y_{FBF}(i, \omega) = \mathbf{W}_o^H(\omega) \mathbf{y}(i, \omega) \quad (2.3.11)$$

其中 $\mathbf{W}_o^H(\omega)$ 為(2.3.10)式 MVDR 求出的結果，每一個音框指標用 i 表示，雜訊參考訊號 (Noise reference signal) 為

$$\mathbf{U}(i, \omega) = \mathcal{H}^H(\omega) \mathbf{y}(i, \omega) \quad (2.3.12)$$

BM 設計方法和 GSC 相似，希望把目標聲源的方向阻擋，讓 \mathbf{y} 經過 BM 後的訊號只剩下雜訊部分，而 BM 的設計由文獻 Sharon Gannot[13]之方法而來，目標聲源方向的陣列拓

樸向量由(2.1.2)式寫為

$$A(\omega) = [A_1(\omega) A_2(\omega) \cdots A_M(\omega)]^T \quad (2.3.13)$$

阻擋矩陣的設計為

$$\mathcal{H}(\omega) = \begin{bmatrix} -\frac{A_2^*(\omega)}{A_1^*(\omega)} & -\frac{A_3^*(\omega)}{A_1^*(\omega)} & \cdots & -\frac{A_M^*(\omega)}{A_1^*(\omega)} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (2.3.14)$$

$\mathcal{H}(\omega)$ 為 $M \times (M-1)$ 的矩陣且 $\mathcal{H}(\omega)$ 的行空間為 $A(\omega)$ 的零空間，可透過下式驗證

$$A^H(\omega)\mathcal{H}(\omega) = 0 \quad (2.3.15)$$

介紹完 BM 的設計後，接著要利用 NLMS 的方法實現 ANC。系統輸出訊號可寫為

$$\mathbf{Y}(i, \omega) = \mathbf{Y}_{FBF}(i, \omega) - \mathbf{G}^H(i, \omega)\mathbf{U}(i, \omega) \quad (2.3.16)$$

$\mathbf{G}(i, \omega)$ 為 NLMS 演算法中的時變參數，可寫為

$$\mathbf{G}_m(i+1, \omega) = \mathbf{G}_m(i, \omega) + \mu \frac{\mathbf{U}_m(i, \omega)\mathbf{Y}_m^*(i, \omega)}{P_{est}(i, \omega)} \quad \text{for } m = 2, \dots, M \quad (2.3.17)$$

$$P_{est}(i, \omega) = \rho P_{est}(i-1, \omega) + (1-\rho) \sum_m |\mathbf{y}_m(i, \omega)|^2 \quad (2.3.18)$$

μ 為步長(Step Size)， ρ 為估測輸入能量所使用的 Forgetting Factor。以上就是 MVDR 結合雜訊估測的適應性演算法。

模擬環境除了表 2- 1 條件以外多加了一個環境中變動的干擾源訊號在 -45° 位置。

圖 2-17 為混合後的訊號 -45° 的干擾源，能量會由小而大的變化。

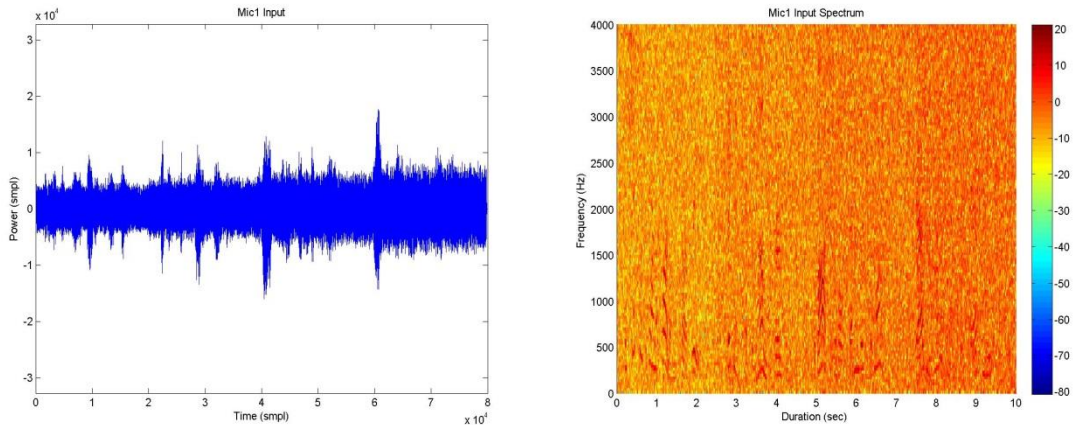


圖 2- 17 MVDR+Noise Estimate 輸入訊號

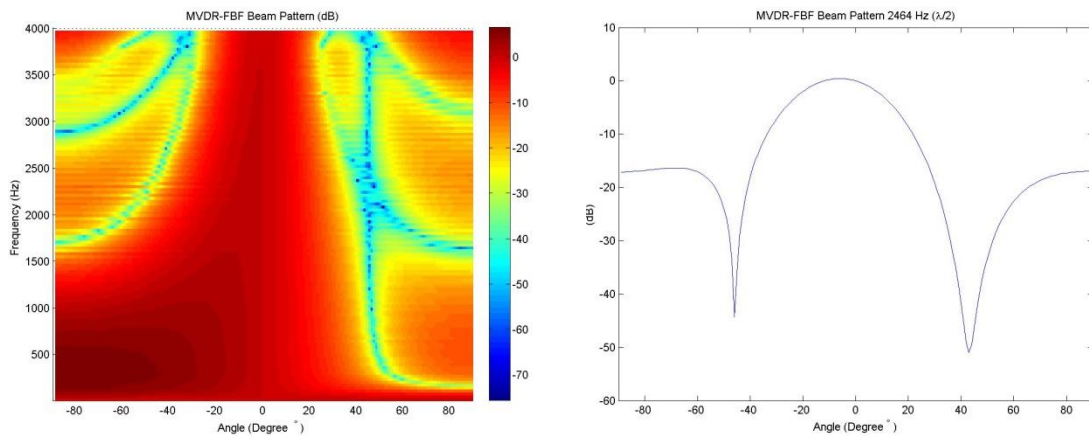


圖 2-18 FBF Beam Pattern

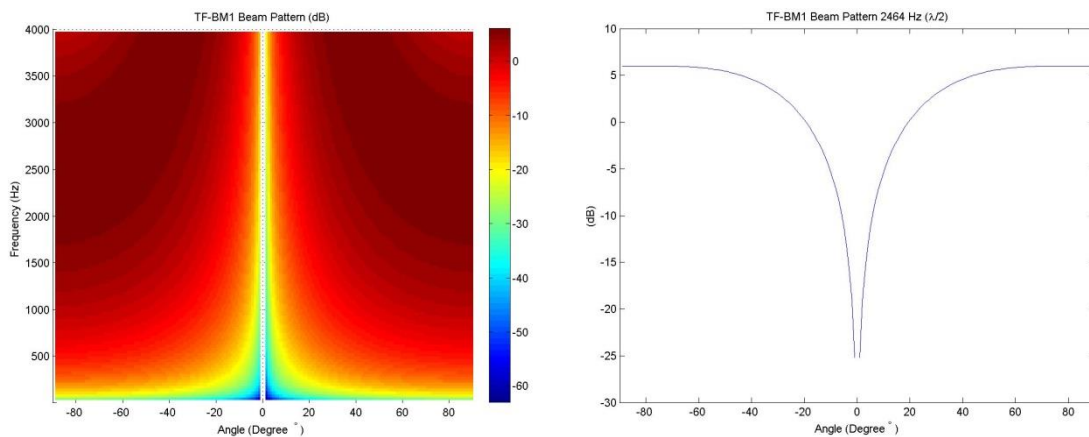


圖 2-19 BM Beam Pattern

圖 2-18、圖 2-19 分別代表為 FBF 和 BM 的 Beam Pattern，都是固定的權重不隨著訊號改變。圖 2-20 為輸入訊號通過 FBF 權重後的訊號， -45° 干擾源並沒有包含在 MVDR 事前訓練的權重內，當其能量越大破壞輸出 SIR 也隨之嚴重。

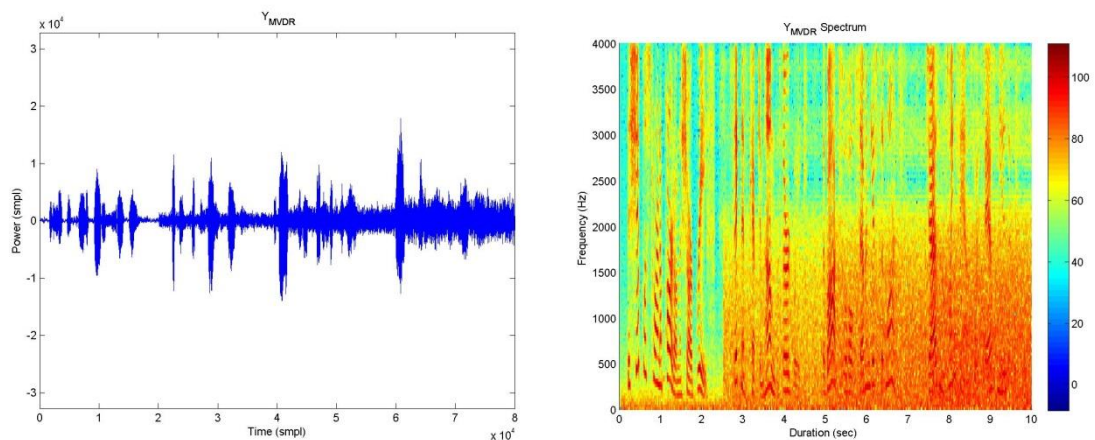


圖 2-20 經過 MVDR 處理之 Y_{FBF} 訊號

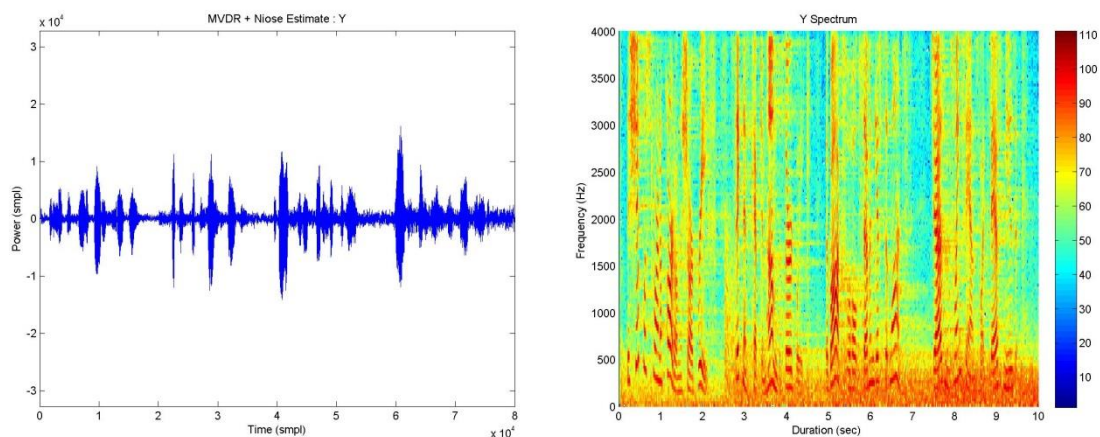


圖 2- 21 經過 MVDR+Noise Estimate 處理之輸出 Y 訊號

透過雜訊估測的適應性以算法能針對 MVDR 未估測出的雜訊提供了抑制功能，圖 2- 21 看出當 -45° 干擾源出現時有壓抑效果。論文架構中主要設計雜訊估測是因為希望除了透過 MVDR 消除環境中固定存在的干擾源外，也會存在 MVDR 訓練時未包含的雜訊。此方法改變 TF-GSC 原本的架構中 FBF 部分利用空間前處理的概念導入 MVDR 把環境中固定干擾源壓抑但 BM 中雜訊估測卻依舊會存在。



第三章 系統架構與相關技術

3.1 應用說明與系統架構

結合聲學迴聲消除與空間濾波器的方法有很多種，W. Kellermann et al.[23][24][25][26]有研究 AEC-BF 和 BF-AEC 前後處理的影響以及利用 GSC 的特性讓 AEC 嵌入其中成為 GSAEC 一個相當特別的架構。AEC-BF 架構如圖 3-1 (a)，在每一個麥克風接收到的訊號分別做 AEC 處理優點為有效解決迴聲問題以及達到空間濾波的方法，不過缺點是當麥克風陣列數目增加時 AEC 演算法同時也需要增加。BF-AEC 架構如圖 3-1 (b)，優點在於可以減少 AEC 的運算量但在此情況下使用 BF 為適應性演算法會導致 AEC 無法有效的收斂以致於效能下降。圖 3-1 (c)運用 GSC 的特性進行設計稱為 GSAEC，其中 FBF 部分為固定權重所以在此後面接上了 AEC 除了不會影響收斂速度外也減少了 AEC 的使用次數，不受麥克風陣列數量增加有所影響，缺點是 AEC 設計只有上路徑具有迴聲消除功能，迴聲訊號有可能由下路徑漏過造成效能降低。在 I. Cohen 團隊提出了 TF-GSC 後也陸續出現結合迴聲消除的相關研究，其中 G. Reuven[27][28]比較了不同組合的影響，也提出了一個新的 ETF-GSC 的架構，在這些架構下面各有優缺點也有針對不同的使用情境設計。

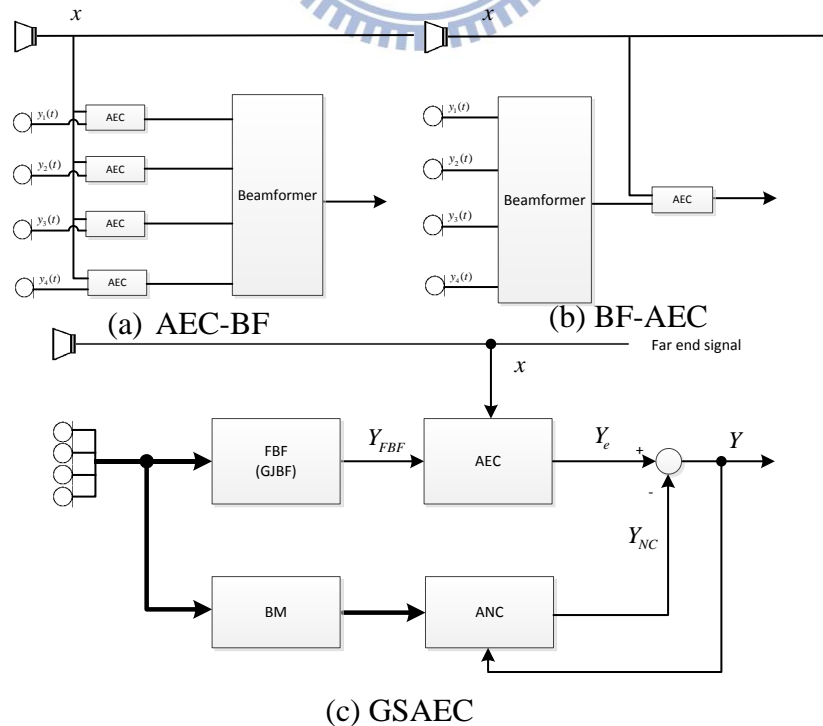


圖 3-1 AEC 結合 BF 架構圖

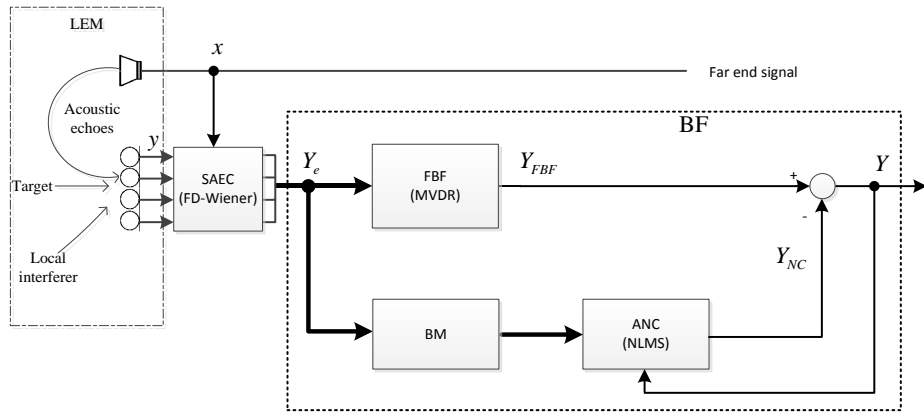


圖 3-2 AEC+BF 架構圖

說明智慧型電視的語音控制遠距使用情境，假設目標聲源為 0° 方向且遠場平面波，在此利用時間延遲模型建構陣列拓撲向量，使用者需要在準確之方向操作才會有好的表現，如果目標聲源沒有在設計方向會因為 MVDR 壓抑導致 SNR 下降。圖 3-2 之架構中需要對環境雜音偵測進行數秒鐘的訓練時間，此步驟為了讓 MVDR 把固定干擾源的資訊包含於 $S_{NV}(\omega)$ 中。SAEC 使用頻域維納濾波器方法，不需事先訓練動作，藉由適應性更新來達到消除電視喇叭造成的影響。而適應性雜訊消除方法主要針對沒有受到 MVDR 壓抑的雜訊和剩餘迴聲訊號進行改善。

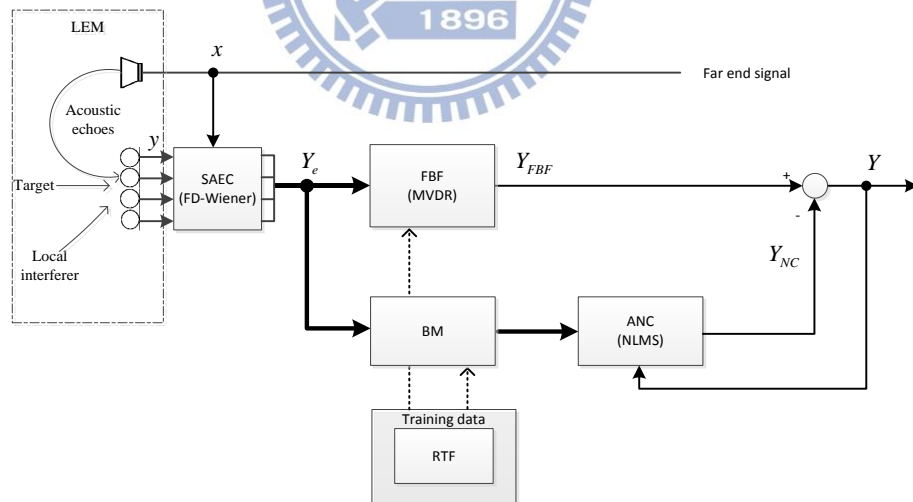


圖 3-3 AEC+BF+TFR 架構圖

由於圖 3-2 有假設了使用者聲源為遠場平面波和使用方向的限制，造成使用上有所侷限，本論文再利用 TFR 進行改良使得輸出效果提昇。圖 3-3 架構可以因應不同的使用方向所設計，利用轉移函式比值演算法來鑑別聲源所在的方向，好處在於可以找出

使用者和麥克風的相對位置不必侷限於 0° 方向，但利用此方法所要訓練的項目就多出了單獨目標聲源方向的相對轉移函式估測，在 3.2 章節中會說明如何估測。

了解不同架構演算法優缺點後整理出智慧型電視應用之系統架構圖，如圖 3-4。

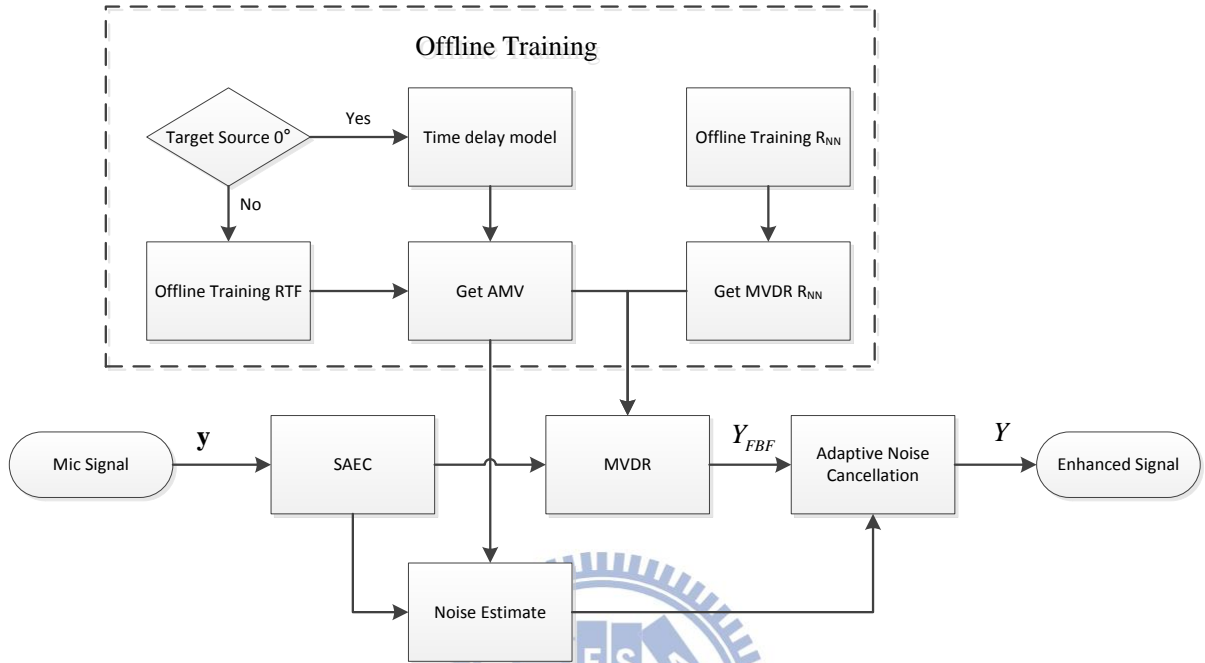


圖 3-4 智慧型電視應用之系統架構圖

系統架構演算法步驟說明：

1. 分別對環境干擾源和目標聲源做事前訓練(Training)，得到 Noise Estimate 和 MVDR 所需的資訊，經運算後各自得到一組權重。
2. 麥克風陣列接收的聲音訊號先各別做 SAEC 的處理。
3. 使用步驟 2 輸出訊號分別乘上步驟 1 所求得之權重，利用這兩組訊號做 ANC 運算當做最後的輸出訊號。

本論文的系統架構中，主要分成兩個部分，第一個部分為立體聲迴聲消除，使用頻域維納濾波器的方法做適應性的運算，消除電視所播放的訊號。第二個部分為空間濾波器，由 TF-GSC[13]架構變形而得，將空間前處理概念導入其中讓 MVDR 權重代替 FBF 權重，在 BM 和 ANC 之設計上則維持相同形式。

3.2 轉移函式比值演算法

轉移函式比值法能解決在實際聲源和麥克風之間的關係，每顆麥克風的增益不匹配或是聲源方向不同所組成的轉移函式也會有差異，轉移函式不易求得希望藉由比值找出之間的關係。O. Shalvi[16]提出利用非穩態訊號和 I. Cohen[17]提出利用語音訊號求出 RTF 的估測，經系統鑑別後的資訊空間濾波器的設計[13][14]，當做陣列拓樸向量使用。

為了估測空間轉移函式，在這裡假設只有一個聲源 s ， M 個麥克風 $y_1 \sim y_M$

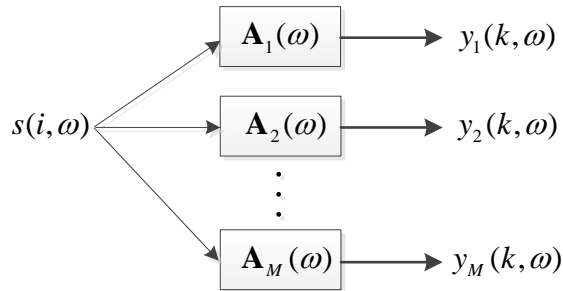


圖 3-5 單聲源系統模型

把圖 3-5 單聲源系統模型中麥克風接收到的聲音資訊表示為

$$y_m(i, \omega) = A_m s(i, \omega) + v_m(i, \omega) \quad (3.2.1)$$

其中 A_m 為空間轉移函式， v_m 為環境中的雜訊。

空間中的轉移函式可能較難得到，但或許可以得到轉移函式的比值，而此數據可能會是有用的，這就是轉移函式比值的概念。由(3.2.1)式，套入轉移函式比值的概念，代入可得以下的式子

$$y_m(i, \omega) - \frac{A_m(\omega)}{A_1(\omega)} y_1(i, \omega) = U_m(i, \omega) \quad (3.2.2)$$

觀察(3.2.2)式， $y_m(i, \omega)$ 及 $y_1(i, \omega)$ 都是已知的麥克風資料，如果能夠得到轉移函式比值

$\frac{A_m(\omega)}{A_1(\omega)}$ 的話，就可以利用簡單的乘法及減法運算完全消除掉聲源 s 的影響，只留下雜訊

$U_n(k, \omega)$ ，我們將 $\frac{A_m(\omega)}{A_1(\omega)}$ ，即所謂的轉移函式比值矩陣命名為 $H_m(\omega)$ ，再經過移項整理

後可得

$$y_m(i, \omega) = H_m(\omega) y_1(i, \omega) + U_m(i, \omega) \quad (3.2.3)$$

想獲得 $H_m(\omega)$ 利用 Cross-PSD 運算(3.2.3)式兩邊同乘上 $y_1(i, \omega)$ 每次計算第 i 個音框的

Cross-PSD，可得以下的式子

$$\Phi_{y_m y_1}(i, \omega) = H_m(\omega) \Phi_{y_1 y_1}(i, \omega) + \Phi_{U_m y_1}(i, \omega) \quad i = 1, 2, \dots, K \quad (3.2.4)$$

將式(3.2.4)以矩陣型式表示得以下的式子

$$\begin{matrix} \begin{bmatrix} \Phi_{y_m y_1}(1, \omega) \\ \Phi_{y_m y_1}(2, \omega) \\ \vdots \\ \Phi_{y_m y_1}(K, \omega) \end{bmatrix} \\ \mathbf{B} \end{matrix} = \begin{matrix} \begin{bmatrix} \Phi_{y_1 y_1}(1, \omega) & 1 \\ \Phi_{y_1 y_1}(2, \omega) & 1 \\ \vdots & \vdots \\ \Phi_{y_1 y_1}(K, \omega) & 1 \end{bmatrix} \\ \mathbf{A} \end{matrix} \begin{matrix} \begin{bmatrix} H_m(\omega) \\ \Phi_{U_m y_1}(i, \omega) \end{bmatrix} \\ \mathbf{X} \end{matrix} \quad (3.2.5)$$

觀察(3.2.5)式可以發現，(3.2.5)式也是一個 $\mathbf{AX}=\mathbf{B}$ 的形式，轉移函式比值矩陣 $H_m(\omega)$ 可以利用最小平方方法 $\mathbf{X}=(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B}$ 形式求得，推導如下：

$$\begin{aligned} \begin{bmatrix} H_m(\omega) \\ \Phi_{U_m y_1}(i, \omega) \end{bmatrix} &= \left\{ \begin{bmatrix} \Phi_{y_1 y_1}(1, \omega) & \Phi_{y_1 y_1}(2, \omega) & \dots & \Phi_{y_1 y_1}(K, \omega) \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \Phi_{y_1 y_1}(1, \omega) & 1 \\ \Phi_{y_1 y_1}(2, \omega) & 1 \\ \vdots & 1 \\ \Phi_{y_1 y_1}(K, \omega) & 1 \end{bmatrix} \right\}^{-1} \\ &= \frac{\begin{bmatrix} \Phi_{y_1 y_1}(1, \omega) & \Phi_{y_1 y_1}(2, \omega) & \dots & \Phi_{y_1 y_1}(K, \omega) \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \Phi_{y_m y_1}(1, \omega) \\ \Phi_{y_m y_1}(2, \omega) \\ \vdots \\ \Phi_{y_m y_1}(K, \omega) \end{bmatrix}}{\begin{bmatrix} 1 & -\{\Phi_{y_1 y_1}(K, \omega)\} \\ -\{\Phi_{y_1 y_1}(K, \omega)\} & \{\Phi_{y_1 y_1}^2(K, \omega)\} \end{bmatrix} \begin{bmatrix} \{\Phi_{y_1 y_1}(K, \omega) \Phi_{y_m y_1}(K, \omega)\} \\ \{\Phi_{y_m y_1}(K, \omega)\} \end{bmatrix}} \\ &= \frac{\{\Phi_{y_1 y_1}^2(K, \omega)\} - \{\Phi_{y_1 y_1}(K, \omega)\}^2}{\{\Phi_{y_1 y_1}^2(K, \omega)\} - \{\Phi_{y_1 y_1}(K, \omega)\}^2} \end{aligned} \quad (3.2.6)$$

由(3.2.6)式可以求得轉移函式比值矩陣 $H_m(\omega)$ ：

$$H_m(\omega) = \frac{\langle \Phi_{y_1 y_1}(\omega) \Phi_{y_m y_1}(\omega) \rangle - \langle \Phi_{y_1 y_1}(\omega) \rangle \langle \Phi_{y_m y_1}(\omega) \rangle}{\langle \Phi_{y_1 y_1}^2(\omega) \rangle - \langle \Phi_{y_1 y_1}(\omega) \rangle^2} \quad (3.2.7)$$

其中 $\langle \beta(\omega) \rangle \equiv \sum_{i=1}^K \beta(i, \omega)$ 定義為平均運算子。

3.3 立體聲迴聲消除演算法

立體聲迴聲消除由多聲道迴聲消除法中衍生而來，A. Hirano et al. [29] 提出使用單一參考聲源形成的立體聲藉由延遲特性減少對估計雙聲道迴聲路徑時所需要使用的濾波器在硬體上運算次數，M. M. Sondhi et al. [30]介紹使用在視訊會議上雙聲道麥克風產生的立體聲藉由對遠端使用者(Far-end)的空間系統鑑別減少對近端使用者(Near-end)估計雙聲道迴聲路徑時所需要使用的濾波器運算次數、T.Gänsler et al. [31]整理了多種適應性濾波器處理立體迴聲消除方法、F. Yang et al. [21]提出立體聲迴聲抑制基於頻域下使用維納濾波器及增益函數的演算法。

在本論文所使用的情境底下選用雙聲道迴聲消除法來做說明，智慧型電視應用很廣除了視訊會議也會透過麥克風陣列做語音控制的功能，此時的電視喇叭聲源播出的雙聲道訊號並不一定具有特定性質有可能兩個喇叭播出的是單聲道訊號、立體聲訊號或雙聲道訊號此時各聲道具有不同的聲音，在這較複雜的情況下選擇使用雙聲道迴聲消除方法來實現 SAEC 上的架構。

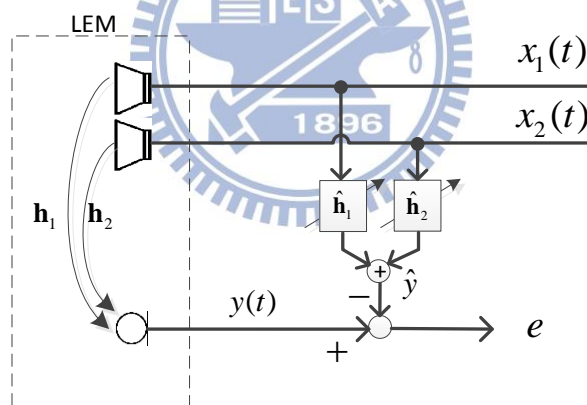


圖 3-6 SAEC 架構圖

估測 $\hat{\mathbf{h}}_1$ 和 $\hat{\mathbf{h}}_2$ 時是利用誤差訊號作為更新的參考依據，如 LMS 或 Fast Block LMS 等演算法，圖 3-6 為架構設計是可以滿足立體聲道迴聲消除，而消除的迴聲 $x_1(t)$ 和 $x_2(t)$ 可以互為相關性或非相關性的訊號都能處理。SAEC 的架構是分別對於預播出喇叭聲音 $x_1(t)$ 和 $x_2(t)$ 做 AEC 的運算， $\hat{\mathbf{h}}_1$ 和 $\hat{\mathbf{h}}_2$ 估計出在 LEM 環境中的迴聲路徑 (\mathbf{h}_1 和 \mathbf{h}_2) 可以透過各種方法來設計這兩組 FIR 濾波器， $x_1(t)$ 和 $x_2(t)$ 通過這兩個濾波器估計出 \hat{y} 從訊號 y 減去可得誤差訊號 e ，此訊號會是近端迴聲最小的值。

先寫出迴聲訊號由(3.3.1)所示

$$y(t) = \mathbf{h}_1(t) * x_1(t) + \mathbf{h}_2(t) * x_2(t) \quad (3.3.1)$$

其中*代表摺積， \mathbf{h}_1 和 \mathbf{h}_2 代表在 LEM 中喇叭到麥克風之間的脈衝響應， x_1 和 x_2 為喇叭訊號。(取樣訊號均假設為時間 t 的正整數指標)。誤差訊號可以寫成

$$e(t) = y(t) - (\hat{\mathbf{h}}_1^T \mathbf{x}_1 + \hat{\mathbf{h}}_2^T \mathbf{x}_2) \quad (3.3.2)$$

$\hat{\mathbf{h}}_1$ 和 $\hat{\mathbf{h}}_2$ 為 FIR 濾波器係數均為 N 維的向量，定義 $\mathbf{x}_1 \equiv [x_1(t) \ x_1(t-1) \ \cdots \ x_1(t-N+1)]^T$ 和 $\mathbf{x}_2 \equiv [x_2(t) \ x_2(t-1) \ \cdots \ x_2(t-N+1)]^T$ 為最近 N 筆資料組成的麥克風訊號取樣， T 代表轉置。誤差訊號可以寫成矩陣形式簡化為

$$e(t) = y(t) - \hat{\mathbf{h}}^T \mathbf{x} \quad (3.3.3)$$

把 $\hat{\mathbf{h}}_1$ 和 $\hat{\mathbf{h}}_2$ 串為行向量 $\hat{\mathbf{h}} \equiv [\hat{\mathbf{h}}_1^T \ \hat{\mathbf{h}}_2^T]^T$ ，同樣的將 \mathbf{x}_1 和 \mathbf{x}_2 串為行向量 $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T]^T$ 。假設 N 的值夠大， $y(t)$ 可以寫為

$$y(t) = \mathbf{h}_1^T \mathbf{x}_1 + \mathbf{h}_2^T \mathbf{x}_2 = \mathbf{h}^T \mathbf{x} \quad (3.3.4)$$

\mathbf{h}_1 和 \mathbf{h}_2 代表真實的房間脈衝響應向量並且把 \mathbf{h}_1 和 \mathbf{h}_2 串為行向量 $\mathbf{h} \equiv [\mathbf{h}_1^T \ \mathbf{h}_2^T]^T$ 。將(3.3.4)代入(3.3.3)可寫為

$$e(t) = (\mathbf{h} - \hat{\mathbf{h}})^T \mathbf{x} = \tilde{\mathbf{h}}^T \mathbf{x} \quad (3.3.5)$$

在此定義

$$\tilde{\mathbf{h}} \equiv \mathbf{h} - \hat{\mathbf{h}} \quad (3.3.6)$$

稱之為錯位向量(misalignment vector)。以 LMS 為例，更新 $\hat{\mathbf{h}}$ 時利用(2.2.17)所示

$$\hat{\mathbf{h}}(t+1) = \hat{\mathbf{h}}(t) + 2\mu \mathbf{x}(t)e(t) \quad (3.3.7)$$

等價於

$$\tilde{\mathbf{h}}(t+1) = \tilde{\mathbf{h}}(t) + 2\mu \mathbf{x}(t)e(t) \quad (3.3.8)$$

在此 μ 為步長。當希望消除迴聲 $e(t) \rightarrow 0$ 並不總是成立，應為 $\min\{e(t)\}$ 意味著 $\|\tilde{\mathbf{h}}\| \rightarrow 0$ (i.e., $\hat{\mathbf{h}} \rightarrow \mathbf{h}$)才是適應性調整的目標，所以重要的是最早收斂得到 $\hat{\mathbf{h}} \rightarrow \mathbf{h}$ 確保各聲道的迴聲消除。

接著透過頻域維納濾波器來求解 SAEC 問題把架構圖修改為圖 3-7，配合本論文 2.2.2 章節中所使用的迴聲消除方法所以訊號皆以頻域狀態表示。維納濾波器由於是開迴路運算，如果使用圖 3-6 SAEC 架構圖作為相同架構，當出現參考訊號為兩相關訊號時因為開迴路運算會讓參考訊號多算入誤差訊號中，所以必須避免這個問題故提出圖 3-7 SAEC Weiner 架構圖，利用兩階段的運算分別算出各階段的最佳解得到誤差訊

號。

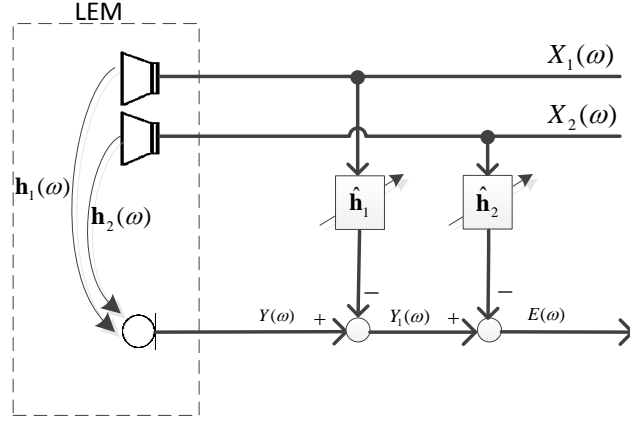


圖 3-7 SAEC Wiener 架構圖

推導利用頻域維納濾波器演算法求解 SAEC，先寫出迴聲訊號由(3.3.1)所示

$$Y(\omega) = \mathbf{h}_1^H(\omega)X_1(\omega) + \mathbf{h}_2^H(\omega)X_2(\omega) = \mathbf{h}^H(\omega)\mathbf{X}(\omega) \quad (3.4.1)$$

其中^H是共軛轉置運算， \mathbf{h}_1 和 \mathbf{h}_2 代表在 LEM 中喇叭到麥克風之間的脈衝響應， $X_1(\omega)$ 和 $X_2(\omega)$ 為喇叭訊號，定義 $\mathbf{h}(\omega) \equiv [\mathbf{h}_1^T(\omega) \quad \mathbf{h}_2^T(\omega)]^T$ 和 $\mathbf{X}(\omega) = [X_1^T(\omega) \quad X_2^T(\omega)]^T$ 。第一級訊號運算，先對 $X_1(\omega)$ 訊號進行迴聲消除可得 $Y_1(\omega)$ 如下所示

$$Y_1(\omega) = Y(\omega) - \hat{\mathbf{h}}_1^H(\omega)X_1(\omega) \quad (3.4.2)$$

其中 $\hat{\mathbf{h}}_1$ 為維納濾波器係數，利用頻域維納濾波器(2.2.22)式方法求得

$$\hat{\mathbf{h}}_1(\omega) = \frac{\Phi_{X_1 Y}(\omega)}{\Phi_{X_1 X_1}(\omega)} \quad (3.4.3)$$

第一級運算利用(3.4.2)式讓 $Y(\omega)$ 訊號迴聲消除，再進行第二級運算，對參考訊號 $X_2(\omega)$ 的迴聲消除運算可得最後輸出的誤差訊號如下所示

$$E(\omega) = Y_1(\omega) - \hat{\mathbf{h}}_2^H(\omega)X_2(\omega) \quad (3.4.4)$$

其中 $\hat{\mathbf{h}}_2$ 為維納濾波器係數，利用頻域維納濾波器(2.2.22)式方法求得

$$\hat{\mathbf{h}}_2(\omega) = \frac{\Phi_{X_2 Y_1}(\omega)}{\Phi_{X_2 X_2}(\omega)} \quad (3.4.5)$$

(3.4.3)和(3.4.5)中所用到的 Cross(Auto)-PSD 均是透過(2.2.24)式和(2.2.25)式的一階遞迴方法求出。圖 3-7SAEC 架構下運算可以對 $X_1(\omega)$ 和 $X_2(\omega)$ 兩個喇叭訊號做雙聲道迴聲消除，不論 $X_1(\omega)$ 和 $X_2(\omega)$ 為相關性或非相關性的訊號都能同時處理。

第四章 實驗結果與分析

本章節音檔分為模擬環境和真實環境，在 AEC 的評估章節當中利用 RIR[32]模擬環境可以清楚了解所設計的空間響應，接著再對此條件下的訊號進行不同迴聲消除演算法之比較與分析。最後在利用真實環境錄製的音檔做驗證，接下來將介紹麥克風陣列平台以及所用於智慧型電視應用之器材。

實驗環境為室內環境，實驗測試將針對 4 顆類比麥克風陣列輸入，模擬智慧型電視中收音裝置的情形。圖 4- 1 為類比麥克陣列輸入裝置照片，麥克風陣列間距為 7cm 共四顆置於智慧型電視正上方。以下 4.1 節與 4.2 節均以此實驗錄製環境。



圖 4- 1 類比麥克風陣列裝置

圖 4- 2 介紹錄製器材與裝置，類比麥克風陣列透過 NI USB-6210 裝置錄製音檔，16 位元解析度、8k 取樣率，錄製 4 顆類比麥克風訊號以及 2 個雙聲道喇叭輸出訊號。電視使用 TOSHIBA-40CL20S，主要利用電視上左右聲道喇叭撥放聲音當作真實迴聲路徑。撥放音檔及錄製音檔的功能透過筆記型電腦 ACER TravelMate 6293。圖 4- 3 為目標聲源 0° 與干擾源 45° 距離麥克風 2 公尺的架設位置。



圖 4-2 錄製器材與裝置



圖 4-3 目標聲源與干擾源擺放位置

4.1 AEC 效能評估分析

寬頻分析 AEC 利用 RIR(Room Impulse Response) 模擬環境。

Room Size (Width, Depth, Height)	10, 6, 2.5 (m)
RT60	0.4
Filter Size	2048
Sampling frequency	8k
Loudspeaker (x,y,z) coordinates	5.5465, 6, 0.952 (m)
Microphone (x,y,z) coordinates	5.105, 6, 1.5 (m)
Input	White Noise
Microphone orientation	Omnidirectional

表 4-1 RIR 環境設定

RIR 主要可以模擬房間中的迴聲與空間環境，和房間大小、反射係數、接收端與發射端擺放位置不同會有不同的空間響應，透過 RIR 模擬產生的一組脈衝響應和發射端所要播放出來的訊號做摺積運算當作接收端的訊號。表 4-1 為本次實驗分析所使用的環境設定按照實際環境中麥克風與喇叭擺設相對位置所做的模擬，圖 4-4 為 RIR 麥克風與喇叭空間分布，根據不同的分佈情形會有不同的響應，利用和實際應用中類似的擺放位置進行模擬。

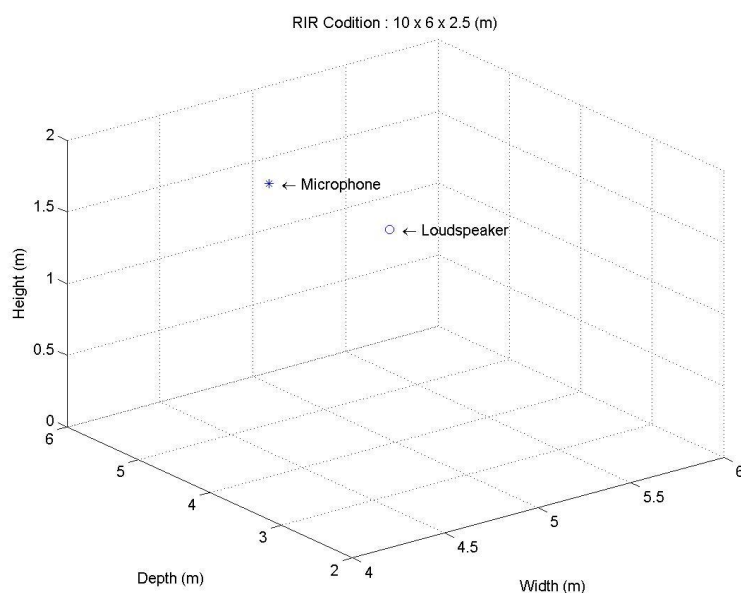


圖 4-4 麥克風與單聲道喇叭空間配置圖

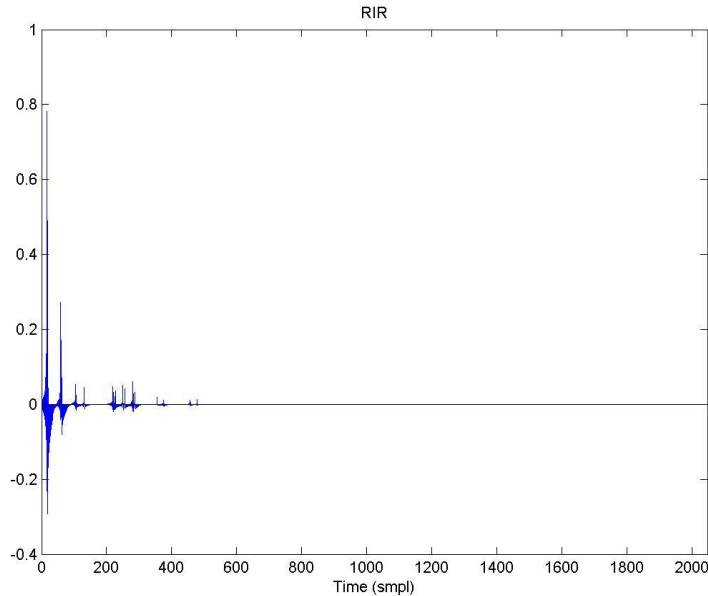


圖 4-5 模擬之 RIR

圖 4-5 是由模擬環境所產生的 RIR 是一組 FIR 係數組成，在模擬濾波器長度設定為 512 依照 RT60、空間和裝置等配置下計算出 RIR。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。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。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。圖 4-6 中反射部分又分成緊接在直接接收後面的早反射(Early Reflections)，以及早反射之後的後迴響(Subsequent Reverberation)。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。

早反射(Early Reflections)：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反射。通常早迴響指在直接接收聲音之後 50-80 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(Subsequent Reverberation)：在早反射之後接收到的聲音都稱為後迴響。

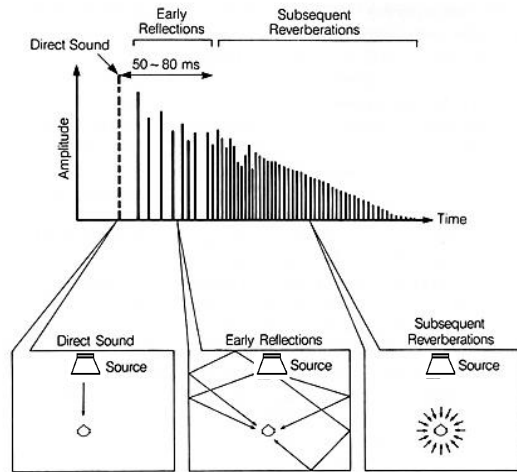


圖 4-6 RIR 說明圖

寬頻分析主要目的想說明不同 AEC 演算法之間的特性以及說明本論文選擇頻域維納濾波器做為智慧型電視應用中 AEC 的主要原因。迴聲評估方法主要分成兩項(1) 迴聲消除效能、(2)迴聲收斂速度，接下來討論這兩項效能指標。

所要評比的演算法分別如下說明：

TDLMS： 時域演算法 $\mu_{TDLMS} = 5 \times 10^{-11}$ ，可參考 2.2.1 章節。其特色實現簡單適應性系統需調整步長，需要是 sample by sample 的運算在時間解析度上相當高。

FDLMS： 頻域演算法 $\mu_{FDLMS} = 5 \times 10^{-11}$ ，可參考 J. Shynk [2][8]提出的 Fast Block LMS。其特色利用 DSP 理論中重疊儲存性質提升運算效率。

FDWiener： 頻域演算法 $\alpha=0.975$ ，可參考 2.2.2 章節。其特色除了頻域重疊儲存等特性外，優點在於非穩態環境中失真最小及開迴路運算不受參考訊號影響輸出效能。

4.1.1 迴聲參數評估

(1) 迴聲消除效能

迴聲消除成效評估方式可在時域利用信號能量之差異來評比迴聲衰減效能，先介紹 Echo Return Loss Enhancement(簡寫為 ERLE) [35]：

$$ERLE = 10 \log_{10} \frac{\overline{y^2(n)}}{\overline{(e(n) - v(n))^2}} \quad (4.1.1)$$

利用 Single-Talk 狀態(近端語音 $s(n)=0$)其中 $y(n)$ 、 $e(n)$ 和 $v(n)$ 為圖 2-7 AEC 主體架構中所示。在評估狀況時選擇 EBR(Echo-to-Background noise Ratio)約為 30dB。表 4-2 評估了三種方法在濾波器長度不同時對迴聲消除的影響，為了配合頻域運算的 AEC 選用濾波器長度和 FFT 長度有關，故連同時域的 AEC 也一併做相同濾波器長度運算做比較。可看出 LMS 濾波器長度越大時會有越好的消除迴聲表現，因為用越多係數去估計 RIR 會越接近真實迴聲路徑除此之外能容忍較多的延遲時間，但是同時也有缺點存在，運算次數會隨之增加，在選擇使用演算法的同時也需要考慮這個缺點。對不同濾波器長度整理出表 4-2 可看出 TDLMS 的 ERLE 提升非常快速因為模擬的環境除了 RIR 的影響外沒有其他雜訊干擾影響結果所以對於 RIR 估測能達到相當高的狀態，由於如圖 4-5 可以看出 RIR 在 512 (samples)內已經包含了 Direct Sound、Early Reflections 和 Reverberation 的成份，所以利用 TDLMS 是有機會能完整的估測出與模擬 RIR 相近的 FIR 係數。在此先把 FFT 長度設定為 256，由表 4-2 中可觀察在此長度下的 ERLE 也有相當程度的效果，因此爾後 AEC 架構使用的濾波器長度選擇相同長度 256 來設計。

	TDLMS	FDLMS	FDWiener
Filter size	Echo Return Loss Enhancement (dB)		
64	15.0890	8.7929	4.8481
128	16.6976	15.0666	8.9162
256	20.1558	16.5675	13.8843
512	61.1935	19.1030	17.5105

表 4-2 各演算法對 Filter size 的 ERLE 評估

選定 AEC 的 FIR 長度為 256 接下來將介紹在此條件下各演算法對接收訊號時間延遲的影響，表 4-3 中顯示各演算法對於不同時間延遲的 ERLE 情況，TDLMS 演算法在時間延遲的情況下有優異的 ERLE 表現，濾波器長度為 256 在時域來說有有充裕的取樣點數容忍延遲造成的誤差所以在小程度的延遲上還有較好的表現，但頻域維納濾波器的權重經過 FFT 轉換原本有效使用的濾波器長度其實只剩下一半所以在容忍延遲的條件上比 TDLMS 來的差。對於實際應用上時間延遲是很常發生的現象，可能因為空間距離、訊號傳遞、訊號擷取或訊號處理運算過程中造成的延遲，但落後的延遲訊號可以事先將參考訊號(喇叭欲播訊號)保留等到麥克風接收訊號時間上對齊才能使用 AEC 運算，但在 AEC 的處理訊號超前是不能發生的事情，因為訊號超前在 RIR 的解釋就失去意義，因此在處理保留參考訊號時要非常小心。

	TDLMS	FDLMS	FDWiener
Delay(smpl)	Echo Return Loss Enhancement (dB)		
128	15.8081	-0.8793	-0.0783
64	16.3328	14.579	3.6025
32	17.335	15.8387	8.3663
16	18.0546	16.5525	11.2962
0	20.1558	16.5675	13.8843

表 4-3 各演算法對時間延遲的 ERLE 評估

均方差 Mean-Square Error (MSE)也是評估迴聲衰減趨勢指標之一

$$MSE(n) = \overline{(e(n) - v(n))^2} \quad (4.1.2)$$

比較不同演算法效能利用 Single-Talk 狀態(近端語音 $s(n)=0$)其中 $e(n)$ 和 $v(n)$ 為

圖 2-7 AEC 主體架構中所示。在評估狀況時選擇 EBR(Echo-to-Background noise Ration) 約為 30dB。主要對經過演算法過後的 $e(n)$ 訊號評估,需要多次實驗得到整體平均結果。

(2) 迴聲收斂速度

迴聲的收斂速度利用 Coefficient error norm[35]

$$C(n) = \frac{\|W - \hat{W}(n)\|^2}{\|W\|^2} \quad (4.1.3)$$

利用模擬結果的 W 與演算法估測 \hat{W} 濾波器係數的差做 2-norm 運算,其中 n 為第 n 次疊代運算估測出 \hat{W} 係數。2-norm 在此的用意想找出濾波器係數更新完後與原本模擬設計的 RIR 之間相差距離的關係,因為 AEC 是透過適應性更新所以收斂後會趨近於一個定值,而步長對收斂有很大的影響所以針對步長的選擇利用收斂速度和 ERLE 做決定。圖 4-7、圖 4-8、圖 4-9 為各演算法調整不同參數的收斂情形。

由於不同參數所影響每一次調整 FIR 係數的大小都會不一樣,可以由圖 4-7、圖 4-8 收斂結果發現步長越大透過 2-norm 計算出的結果會有較大的變動存在,而收斂的時間不能以單一門檻值決定而是要以不同步長做出判斷,LMS 經由 ERLE 及收斂的評估後決定出 $\mu_{LMS} = 5 \times 10^{-11}$, FDLMS 評估過後選用 $\mu_{FBLMS} = 5 \times 10^{-11}$ 。

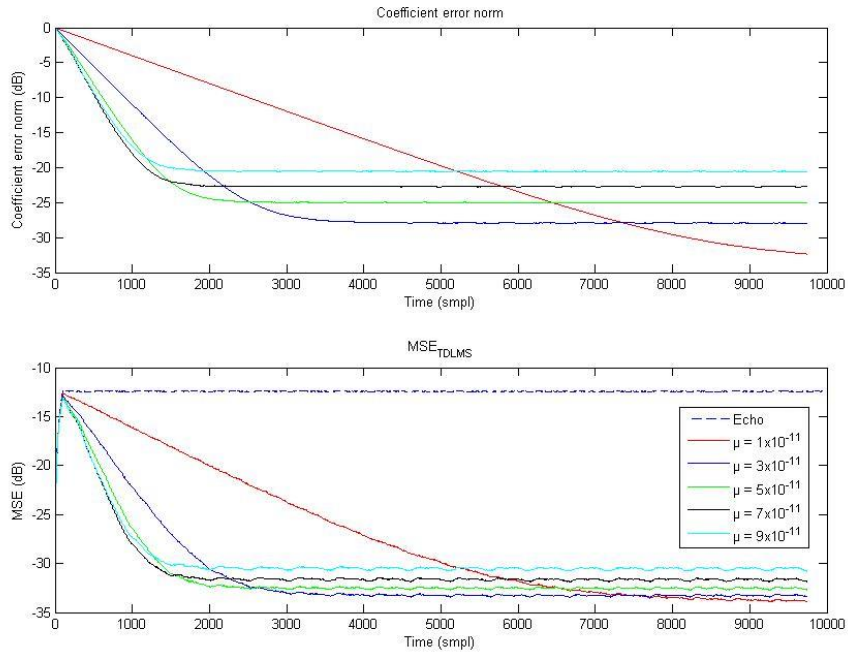


圖 4-7 TDLMS learning curve during single-talk

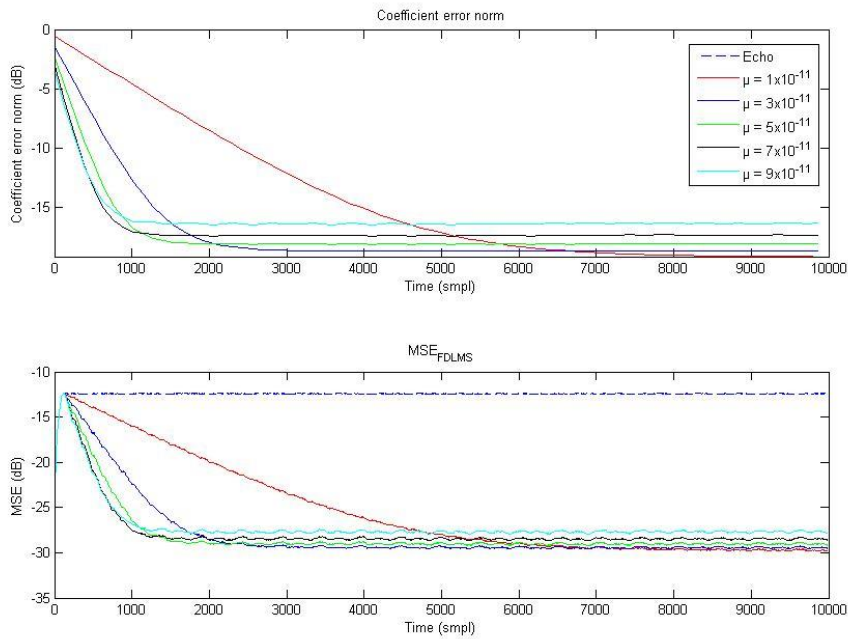


圖 4-8 FDLMS learning curve during single-talk

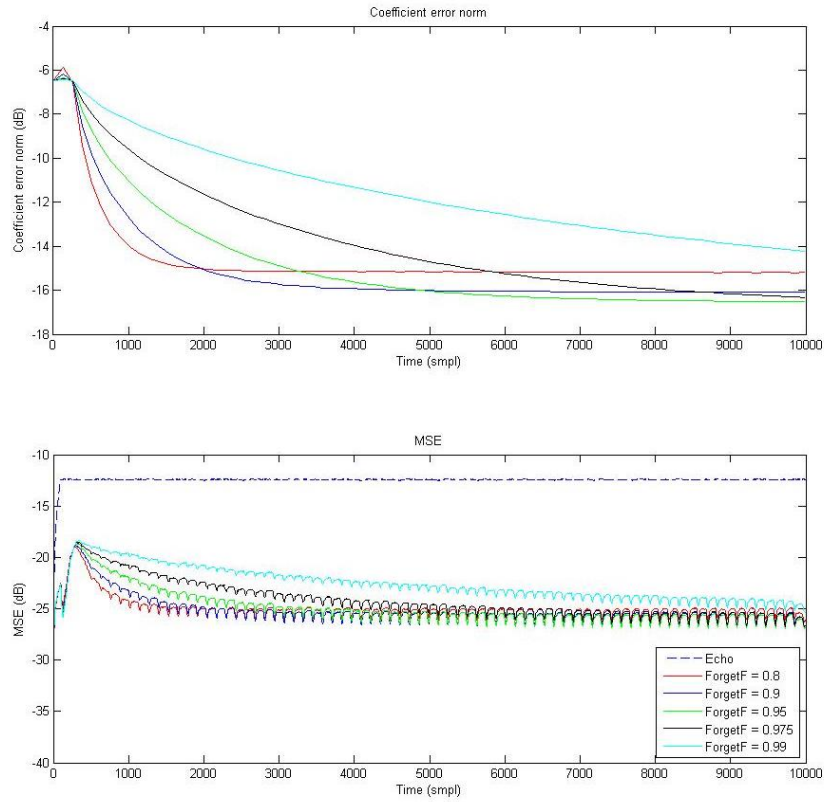


圖 4-9 FDWiener learning curve during single-talk

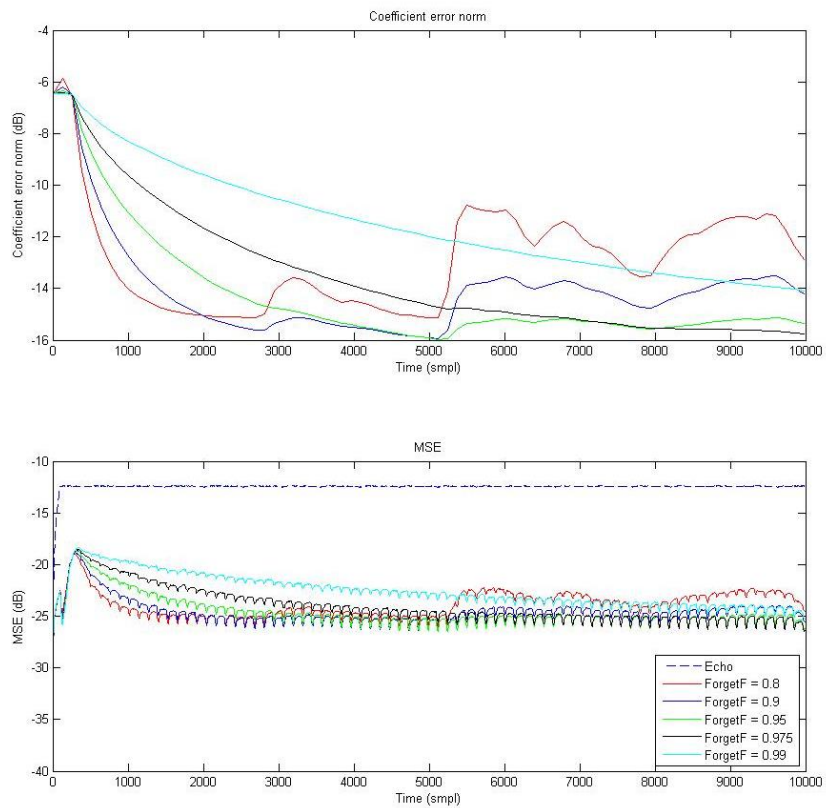


圖 4-10 FDWiener learning curve during double-talk

頻域維納濾波器和 LMS 不同的地方是不需要使用步長做收斂的動作，而所使用的方法是對訊號做 PSD 處理而其中更新 $\Phi_{XY}(\omega)$ 及 $\Phi_{XX}(\omega)$ 需使用一個參數 α 當做遺忘因子(Forgetting Factor) 而他的意義在於對於現在時刻的訊號和過去的訊號做平滑處理或解釋為用來控制疊代過程的有效記憶。由圖 4- 9 圖 4- 9 的曲線顯示 α 值取的與收斂速度和最終收斂的結果有關，當運行環境中有非平穩訊號時必須透過一階遞迴的方式估測出系統所需的功率頻譜密度，在 α 的選擇上也會和有 Double-Talk 狀態相關，利用圖 4- 10 也能看出在相同時間內(10000 samples)收斂的效能 0.95~0.975 中都呈現不錯的表現，而在一些文獻[21]中可以得到選擇 α 時可取用在 0.975 左右的數值。圖 4- 11 為各演算法的 RIR 和模擬環境的 RIR 比較，圖 4- 12 為各演算法迴聲消除成果。

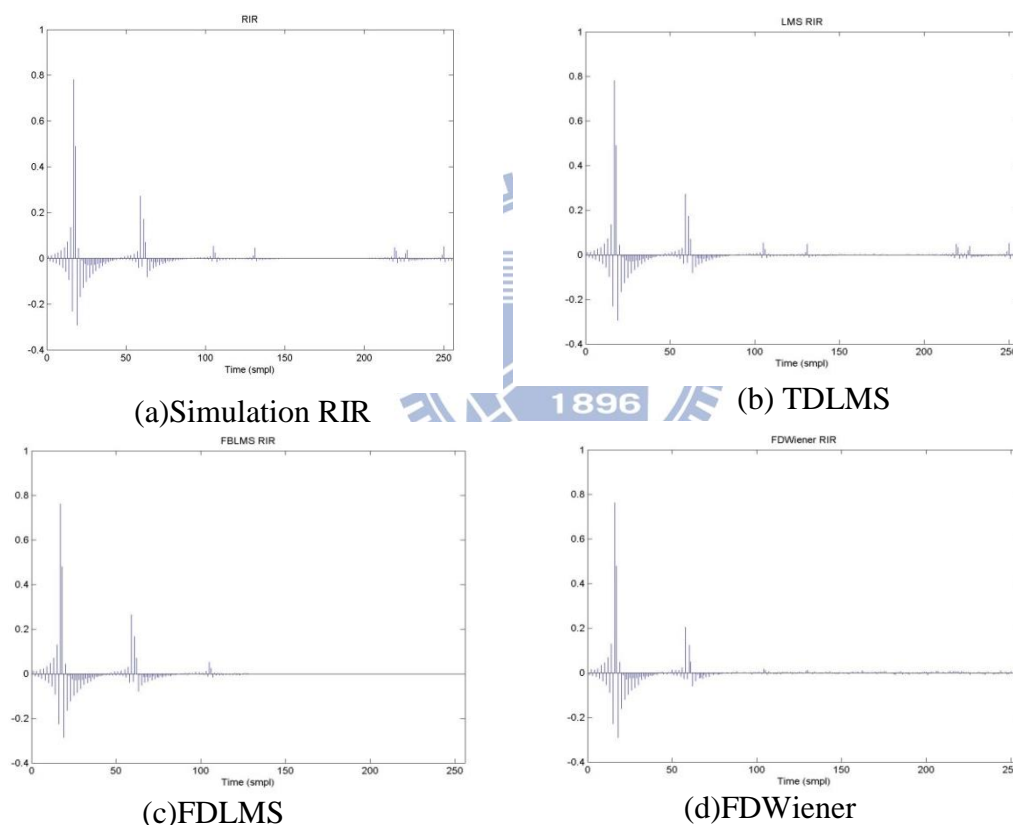


圖 4- 11 比較各演算法求得之 RIR

以上評估 ERLE 的結果而言 TDLMS 優於 FDLMS 優於 FDWiener，TDLMS 優於其它兩者的原因，在於選用 FIR 長度 N 時能完全用於估測迴音路徑而在頻域的運算對時間的解析度少了 N/2 的長度導致損失了一些估測效能，但 Direct Sound 的成份在各演算法的估測都能有效的求得。在智慧型電視的應用是針對立體聲迴聲消除設計不只考慮 ERLE 結果也必須考慮目標聲源失真程度。

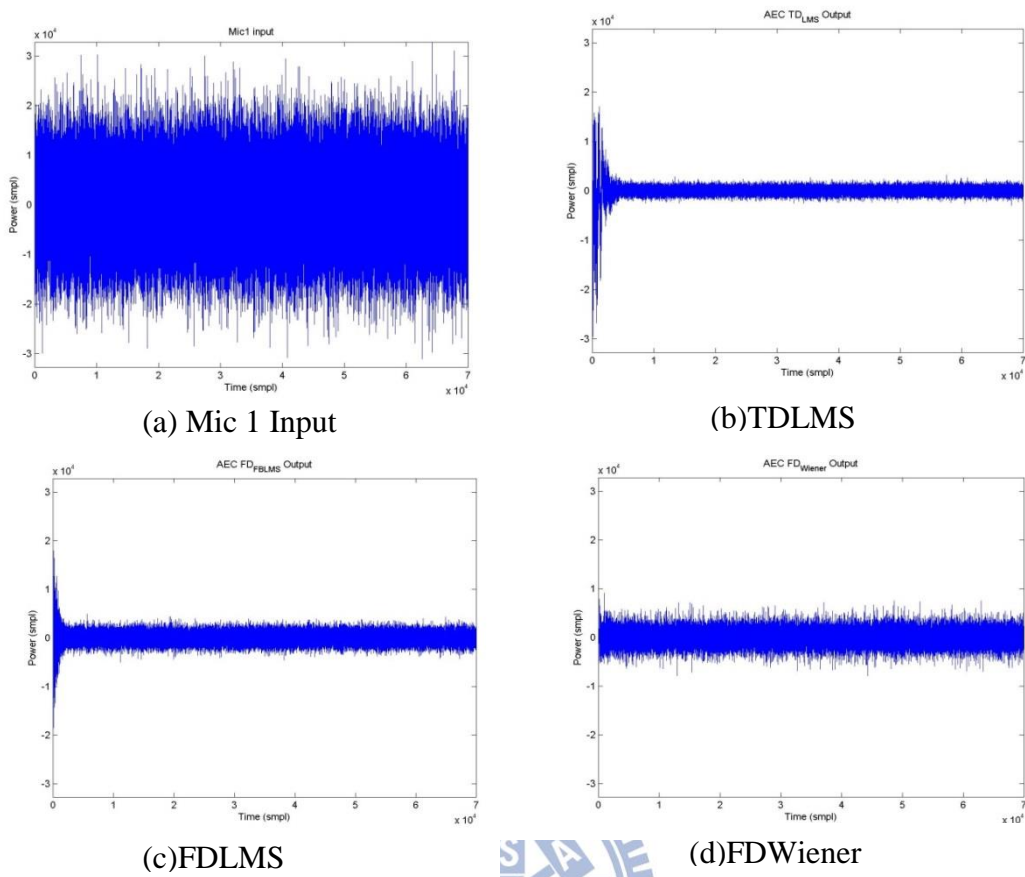


圖 4-12 比較各演算法迴聲消除成果

4.1.2 模擬迴聲真實環境

Room Size (Width, Depth, Height)	10, 5, 2 (m)
RT60	0.3
Filter Size	512
Sampling frequency	8k
Loudspeaker (x,y,z) coordinates	2.3515, 0.1, 0.08(m)
Microphone (x,y,z) coordinates	2.105, 0.1, 0.73 (m)
Input	White Noise
Microphone orientation	Omnidirectional

表 4-4 模擬真實之 RIR 環境設定

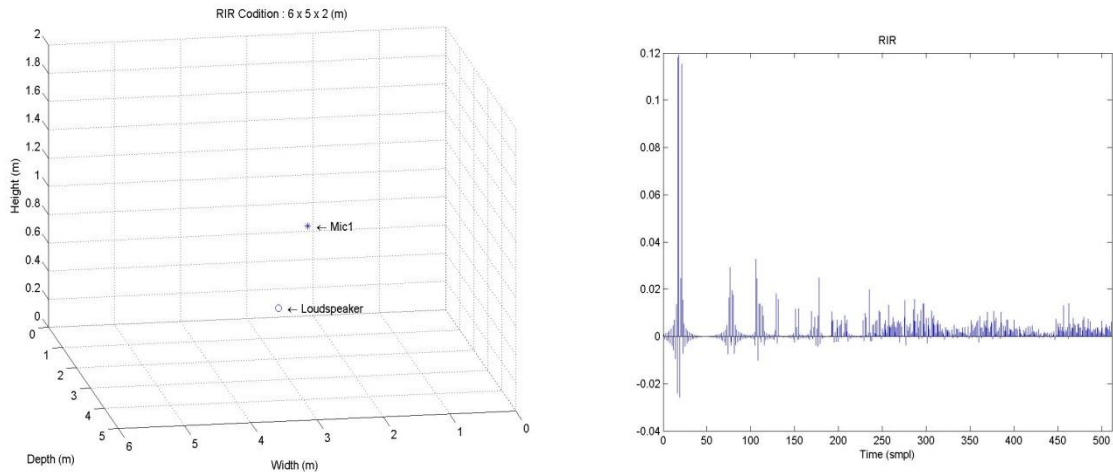


圖 4-13 空間配置與響應

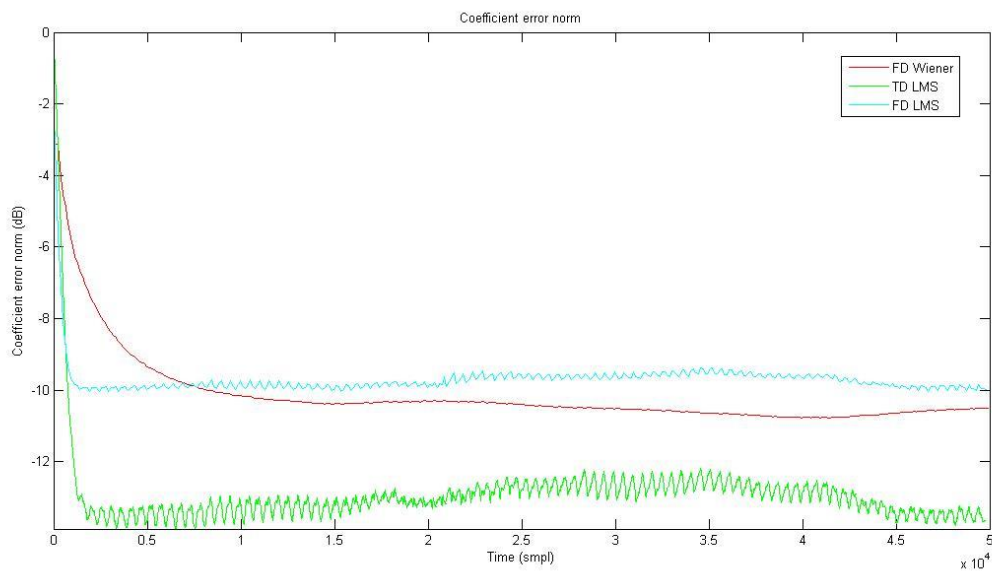
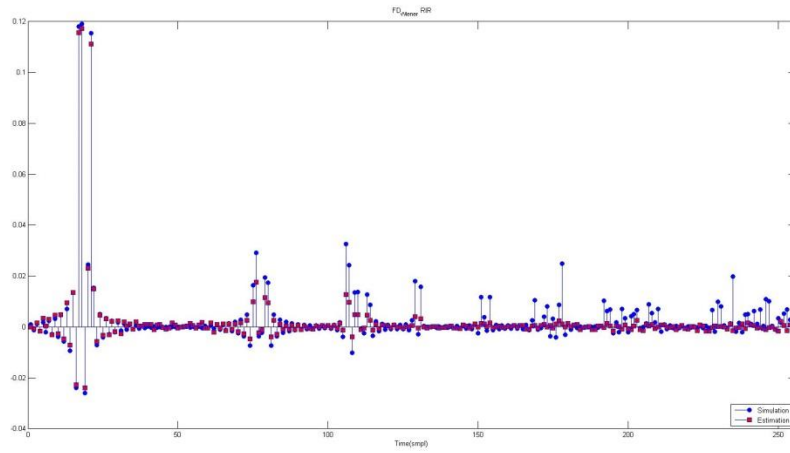


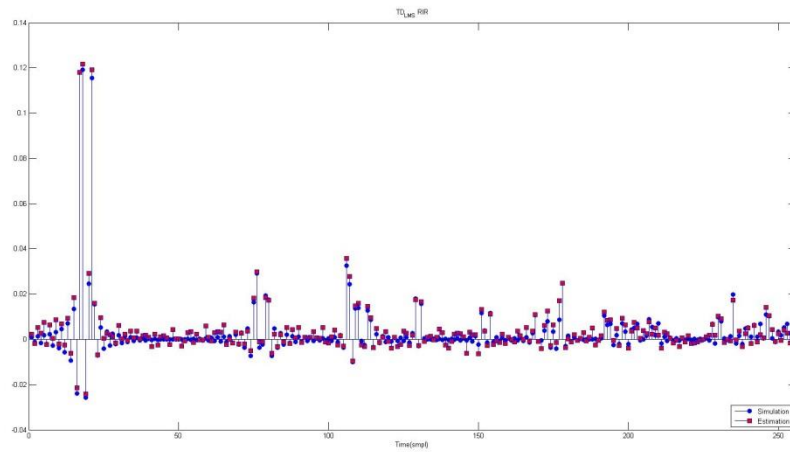
圖 4-14 AEC Coefficient error norm

評估完使用的各參數後接著討論在真實環境中的影響，由表 4-4 建構出實驗環境、圖 4-13 為空間響應。由於在真實環境中的 RIR 是較為複雜的，估測真實 RIR 所用的濾波器大小 N 越大時越能完整的讓實際 RIR 呈現出原貌，圖 4-14 可看出 TDLMS 有較小的誤差值，而 FDLMS 和 FDWiener 相差 1dB 以內單純利用 Coefficient error norm 很難判斷出迴聲消除的效果。圖 4-15 為 $N=256$ 長度下估測與模擬 RIR，透過此圖會知道不同演算法在各個 RIR 值估計值和原始之間的關係。直接接收聲音估測都能估測出來但早反射上出現了不同的響應，在早反射的估測是 TDLMS 佔了優勢，時域和頻域主要差異在於時間解析度上的不同。FDLMS 在估測迴聲路徑的步驟是在時域運算而利用頻域重疊儲存的特性減省估測的次數，缺點是只能有 $N/2$ 的估測長度，在 $N/2$ 之後估測的 RIR 為

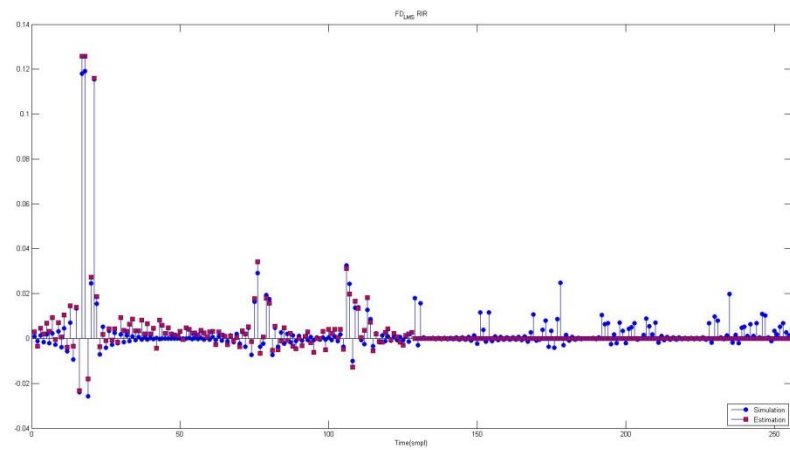
零導致 Coefficient error norm 上升的原因。FDWiener 利用 $\Phi_{XY}(\omega)$ $\Phi_{XX}(\omega)$ 的估測再算出迴聲路徑用於非穩態訊號會成為優勢。



(a) FD Wiener RIR



(b) TD LMS RIR



(c) FD LMS RIR

圖 4-15 比較估測和模擬之 RIR

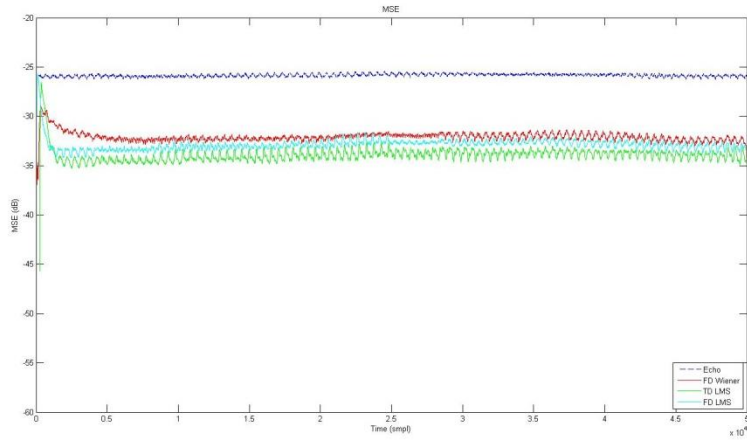


圖 4-16 模擬 MSE 結果

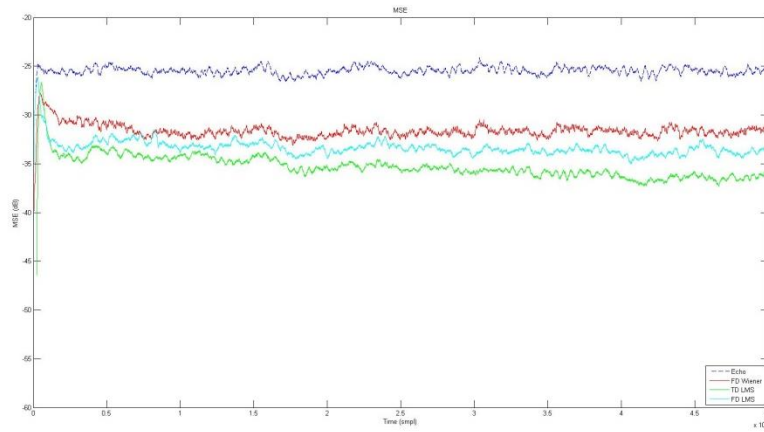
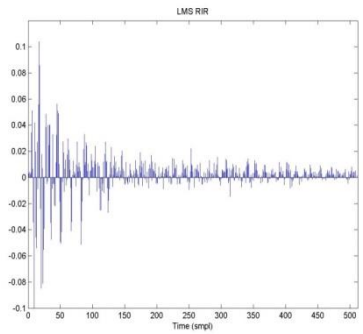
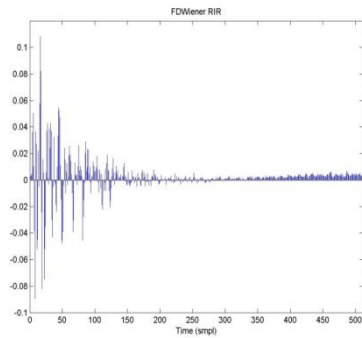


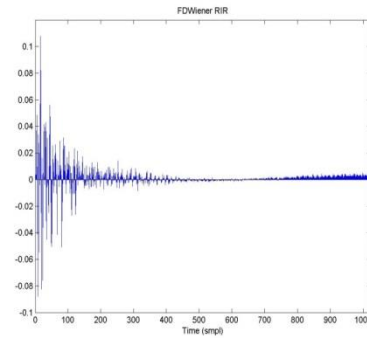
圖 4-17 真實 MSE 結果



(a) TDLMS, $N=51$



(b) FDWiener, $N=512$



(c) FDWiener, $N=1024$

圖 4-18 比較時域和頻域運算真實 RIR

	TDLMS	FDLMS	FDWiener
Filter Size 256	ERLE(dB)		
Simulation	9.78	8.12	6.66
Real Data	9.98	8.02	6.50

表 4-5 模擬與實際之 ERLE 數值

圖 4-16、圖 4-17、表 4-5 在單看迴聲消除時真實結果符合模擬結果的趨勢，圖 4-18 說明時域和頻域運算迴聲消除時 RIR 長度的差別，(a)為利用 TDLMS 估測出 512 筆響應但(b)(c)為利用 TDWiener 估測 RIR 在前半部響應可以正常估測出但後半部就無法正確估測出響應內容，因為在頻域運算根據數位訊號處理理論可以透過重疊儲存相加方法當 50%重疊時運算效率達到最高的優點但同時在估測頻域 RIR 時卻犧牲了完整性。在單看迴聲消除真實結果符合模擬的趨勢，而接下來即將討論 Double-Talk 狀態。

4.1.3 分析各演算法 Double-Talk 狀態

雙聲道迴聲消除的運算方式分別針對左聲道或右聲道的迴音進行處理，而在處理左聲道時右聲道訊號可視為雜訊會干擾左聲道迴音消除能力此時情況可以稱為 Double-talk 狀態，反之亦然。因此要做 SAEC 評估前先對 Double-talk 狀態先做分析。評估標準增加 SERI (Signal-to-Echo Ratio Improvement)和 TELR (Total Echo Level Reduction) [33]而這些定義均參考 ITU-T G.160 內容制訂。分別輸入乾淨語音、吵雜語音和處理後語音三種語料，利用乾淨語音做為判斷語音所處於的層級(分為低中高三等)，其中計算各層級中吵雜語音的 SER_{in}和處理後語音的 SER_{out}相減後得到 SERI;而 TELR 算法是透過計算在未出現人聲時吵雜語音和處理後語音的迴聲消除程度。SERI 數據的評估有目標聲源存在時刻的訊號經處理後 SER 是否有提升，如果為正數代表對迴聲抑制有效果且對目標聲源的影響較小。TELR 數據事針對沒有目標聲源存在的時刻稱為語音停頓(Speech Pauses)時計算，數值越大代表消除迴聲程度越好。

圖 4-19 模擬 Double-Talk 輸入訊號，迴聲使用高斯白雜訊加上額外的語音成分(男性人聲)就出現 Double-talk 情形而此時迴聲消除可能受到額外的語音成分牽連導致破壞人聲部分，圖 4-20 Coefficient error norm 和 MSE 部分可以清楚看到 FDWiener 受 Double talk 的影響甚小而其他兩組方法會讓人聲訊號破壞濾波器權重估測。表 4-6 模擬單聲道迴聲消除 Double-talk 狀態的處理結果，受到了 Double-talk 的影響無法直接評估迴聲消除程度透過使用 TELR，TELR 理論上應該和 ERLE 數值一樣但是由於受到了人聲的干擾效能衰減了，由於接收端擷取的訊號透過(2.2.17)式利用最速下降法更新 \hat{W} 時會把目

標聲源的成分也計算其中，在(2.2.15)式濾波輸出時出現原先參考訊號中不存在的訊號讓迴聲估測效能降低。SERI 評估在各方法都能提升比原始麥克風接收訊號來的優異，雖然說 LMS 的方法消除迴聲的能力比較顯著但是同時也會破壞了人聲訊號，在此利用 PESQ Improvement 方法評估語音失真程度之改善，可以看到 FDWiener 的優勢已經呈現出來，透過一階遞迴方式估測有非穩態訊號透過 α 可以調整對訊號的靈敏度；在智慧型電視的應用下不論是立體聲迴聲消除或是下達語音控制指令時電視聲還是持續播放這些情境都是 Double-talk 的狀態，所以為重要的一項指標。圖 4- 21、圖 4- 22 和表 4- 7 是透過實際環境中擷取的資料進行分析，和模擬結果有相互呼應。

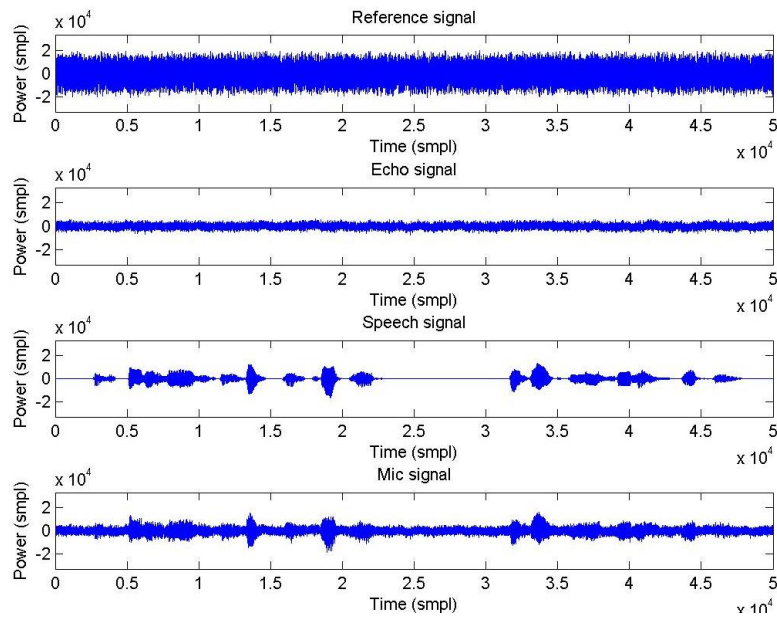


圖 4- 19 模擬 Double-Talk 輸入訊號

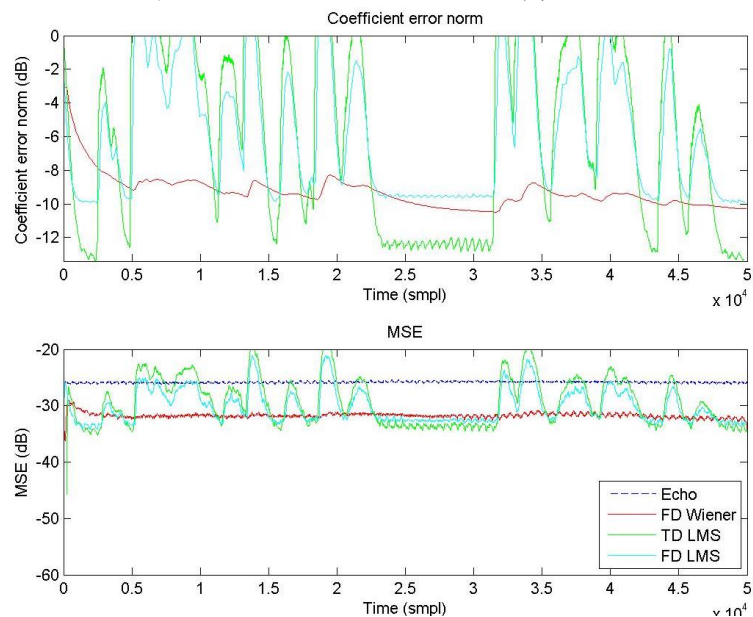


圖 4- 20 模擬 Double-Talk 結果

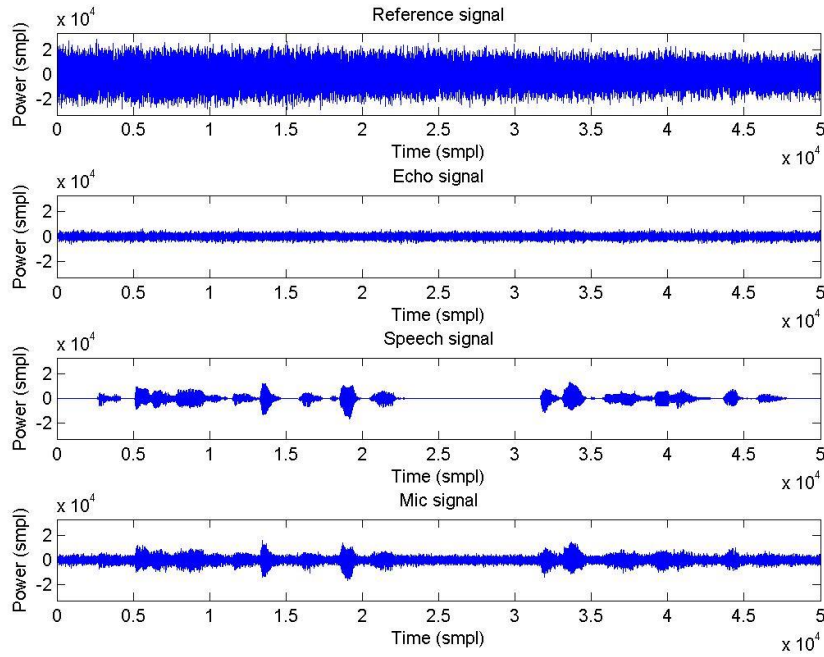


圖 4-21 真實 Double-Talk 輸入訊號

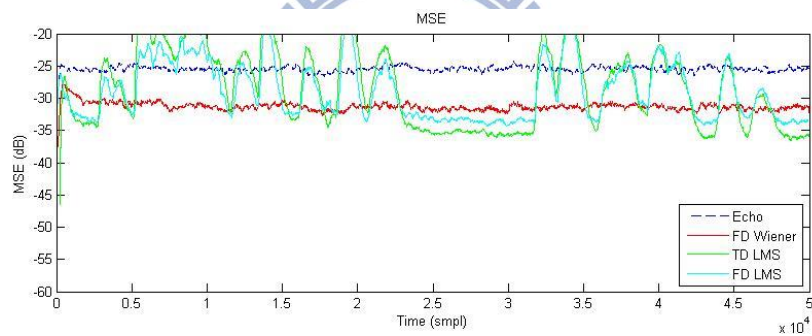


圖 4-22 真實 Double-Talk 結果

	TDLMS	FDLMS	FDWiener
ERLE(dB) (Single-talk)	9.78	8.12	6.66
SERI(dB) (Double-talk)	10.64	7.65	6.7
TELR(dB) (Double-talk)	8.813	6.79	6.31

表 4-6 模擬單聲道回聲消除(Double Talk)

	TDLMS	FDLMS	FDWiener
ERLE(dB) (Single-talk)	9.98	8.02	6.50
SERI(dB) (Double-talk)	12.61	9.14	6.48
TELR(dB) (Double-talk)	9.8	7.2	6.11

表 4-7 實際單聲道回聲消除(Double Talk)

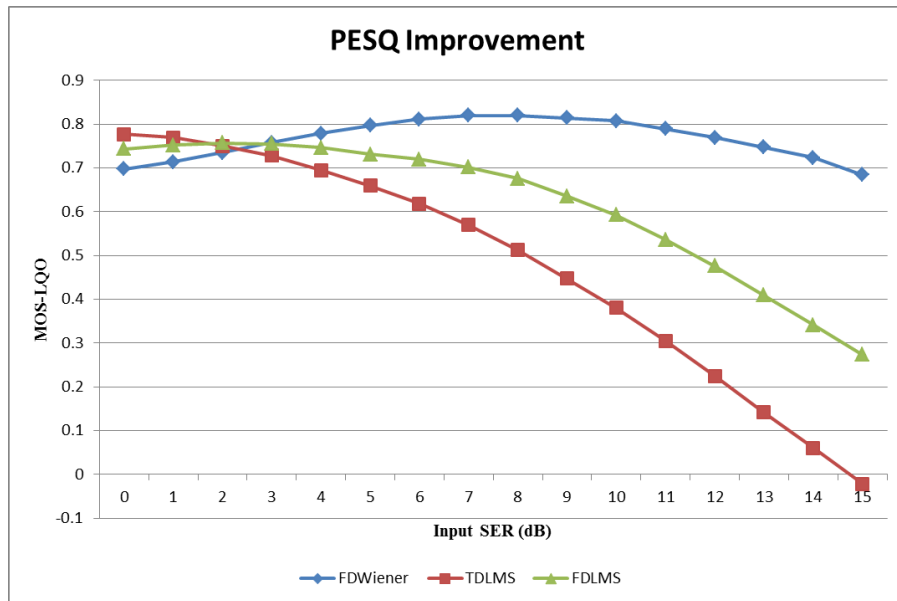


圖 4-23 PESQ Improvement of AEC

圖 4-23 為真實訊號在不同 SER 的情形下對語音品質提升之效果，可以看出在 FDWiener 提升能維持在一定的品質，但是 LMS 的方法提升的效能會有較大的變動是因為有步長的緣故，不能適應性的更新相對於訊號適當的步長會導致在誤差估測時把人聲訊號破壞或是對於迴聲消除的效能有所不同。FDWiener 利用一階遞迴更新迴聲訊號的優勢在此凸顯出來，在 Double-talk 的狀態不會讓非迴聲估測訊號有嚴重失真，如果是人聲訊號是能提升語音品質的，這是本論文 SAEC 選用 FDWiener 重要的理由。

4.1.4 SAEC 實驗與分析

接下來討論 SAEC 分別在 RIR 模擬及真實環境實驗兩種情況做分析比較，RIR 模擬環境設定如表 4-8 且雙聲道喇叭配置如圖 4-24 所示，真實環境播放和擷取資料所用的參數如表 4-9 且利用電視本身的喇叭播出訊號由麥克風陣列錄製配置如圖 4-1 所示。

Room Size (Width, Depth, Height)	10, 6, 2.5 (m)
RT60	0.3
Filter Size	512
Sampling frequency	8k
Loudspeaker (x,y,z) coordinates (m)	Left: 2.5465, 0.1, 0.08 Right: 1.4535, 0.1, 0.08
Microphone (x,y,z) coordinates (m)	Mic1: 1.895, 0.1, 0.73 Mic2: 1.965, 0.1, 0.73 Mic3: 2.035, 0.1, 0.73 Mic4: 2.105, 0.1, 0.73
Input	Independent White Noise
Microphone orientation	Omnidirectional

表 4-8 SAEC 之 RIR 環境設定

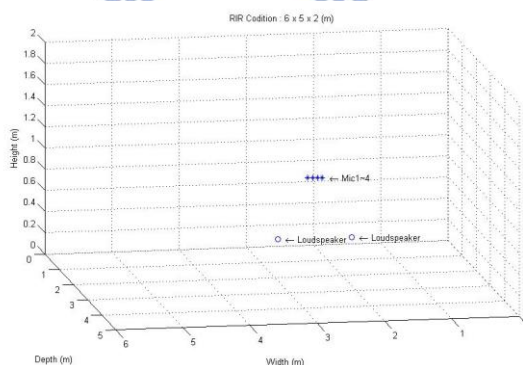


圖 4-24 麥克風與雙聲道喇叭空間配置圖

Sampling frequency	8k
Filter Size	256
Frame Size/overlap	256/128
Microphone Number	4
Play Right & Left Channel Data	Independent White Noise

表 4-9 SAEC 實驗參數

模擬環境和真實環境差別在於訊號傳遞擷取之間會有時間延遲的問題，會因每顆麥克風之間擺放位置和本身特性的差異造成不同的影響，欲利用實驗資料進行分析最先需要找出麥克風接收和喇叭撥放之間的延遲時間，利用相關係數的方式找出訊號間延遲的取樣點數。

$$C_{X,Y}(\tau, n) = \frac{E\{[X(\tau) - m_X(\tau)][Y(n+\tau) - m_Y(n+\tau)]\}}{\sqrt{E[X^2(\tau)] - E[X(\tau)]^2} \sqrt{E[Y^2(n+\tau)] - E[Y(n+\tau)]^2}} \quad (4.1.4)$$

其中 X 、 Y 分別為喇叭欲撥出訊號和麥克風接收訊號， τ 為一段取樣點數的長度 τ 越大則計算出的延遲越準確， n 為每次移動間距為一個取樣點， m_X 、 m_Y 代表期望值運算。以 Left Channel (X) 和 Mic 1 (Y) 為例，計算之間的兩者之間的 Cross-Correlation 關係如圖 4-25 所示，當 Y 訊號在 $n = 845$ (samples) 時和 X 相關性最高，換句話說當 Y 訊號延遲 845 筆取樣點時和 X 訊號最相似，利用(4.1.4)式計算喇叭和麥克風之間所有的時間延遲關係，如表 4-10。在實際環境中錄到的資料和參考訊號間的相關係數為 0.279，發生相關係數降低的原因在於 RIR 變得複雜除了環境配置影響外還存在著其他因素，第一在參考訊號擷取資料過程中有雜訊干擾，第二透過喇叭播放的訊號因為喇叭頻率響應的關係產生失真，第三麥克風頻率響應造成收到訊號的失真，由於這些原因都是可能造成接收到訊號與參考訊號間差異的原因。

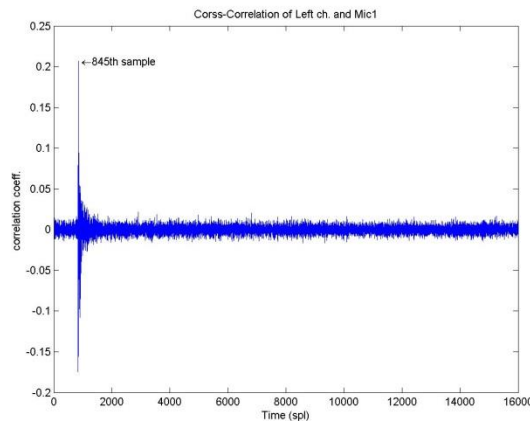


圖 4-25 喇叭(Left)和麥克風(Mic1)之間的 Cross-Correlation 關係

	Mic 1	Mic 2	Mic 3	Mic 4
Channel	Delay (samples)			
Right	845	845	844	844
Left	845	846	846	847

表 4-10 左右聲道對個麥克風之間的延遲關係

AEC 在對時間延遲的影響相當重要，參考圖 2- 7 架構說明，估測 FIR 係數時必須 $x(n)$ 和 $y(n)$ 在相同時間點或是延遲時間在 FIR 長度內可容忍範圍進行，而表 4- 10 顯示出喇叭欲播訊號 $x(n)$ 和麥克風接收到訊號 $y(n)$ 已經有了嚴重的延遲發生，所以在進行 AEC 估測前需要把這個現象解決。在這個系統中發現 847(samples) 的延遲是最嚴重的情況，進行 AEC 運算時可以接受輕微的延遲但是絕對不能有訊號超前的情況發生，故先對參考訊號 $x(n)$ 延遲 844(samples) 筆資料，為了確保各接收端訊號對於發射端訊號都不會出現訊號超前的狀態。由圖 4- 26 可以清楚說明在喇叭與麥克風之間延遲訊號的關係，而找出的這組關係就是本系統喇叭和麥克風訊號之間的延遲時間，解決這個問題如前述，把喇叭左右聲道的參考訊號延後到麥克風接收訊號能夠處理的時刻在進行 AEC 的運算。以上的考量是單純看訊號和訊號之間的關係，如此一來會把喇叭到麥克風的空間關係給破壞掉了，因此還要把喇叭到麥克風的距離所造成的時間延遲算入其中，由麥克風陣列到喇叭的最長距離約為 70 公分透過聲速公式(345m/s)和取樣頻率(8k/s)推算需要傳遞 16 個取樣點的時間訊號方才能從發射端到達接收端，故系統整體參考訊號延遲時間為 828 個取樣點的長度才會符合實際房間脈衝響應。

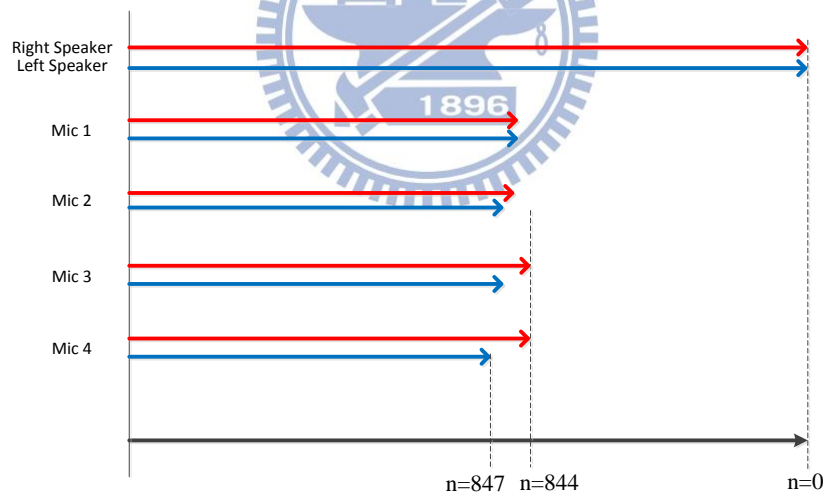


圖 4- 26 左右聲道對各麥克風之間的延遲關係圖

利用以上分析結果改進 SAEC 在真實環境實驗下產生時間延遲的問題後，分別使用 3.3 章節中 SAEC 架構設計各演算法，得出比較模擬資料和真實資料經過 SAEC 運算結果。雙聲道喇叭播出的成份最壞的狀況是兩個沒有相關性的訊號做疊加被麥克風接收到，此時各自聲道視另一聲道為雜訊而影響到估計 RIR 的效能這也就為本實驗的情境。

圖 4- 27 和圖 4- 28 為模擬和真實環境下 MSE 結果，在 SAEC 下的運算以收斂結果而言，各演算法間的差距越來越小不像單聲道迴聲消除來的明顯，是因為使用誤差訊號更新 \hat{W} 受到雙聲道間的訊號會相互影響， \hat{W} 希望在估測兩迴聲路徑中找出權衡過後最佳解。而利用維納解的方法第一級先求出左聲道的迴聲路徑，將第一級產生的誤差訊號當做下一級的輸入，在透過維納濾波器運算右聲道的迴聲路徑，不論左右聲道是有相關性或非相關訊號經由此運算都能有迴聲消除的作用。

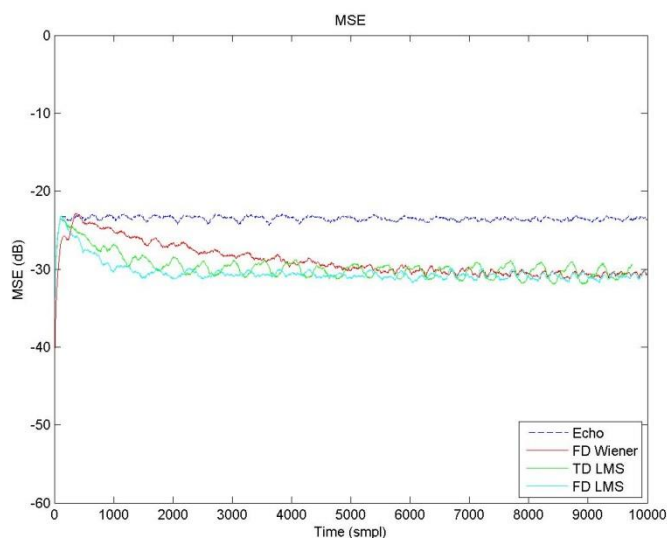


圖 4- 27 SAEC 在模擬環境之 MSE

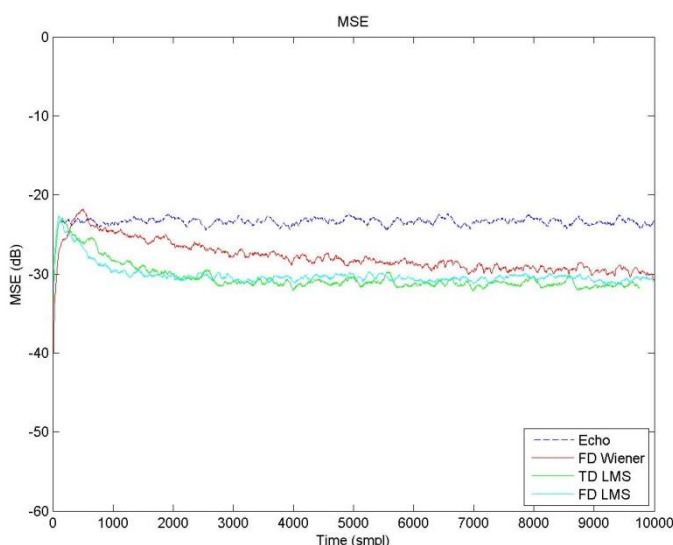


圖 4- 28 SAEC 在真實環境之 MSE

圖 4- 29 為各聲道估測與模擬 RIR 之間收斂情形，10000 筆收斂資料的時間內最後收斂的數值為 -9dB 左右，可以看出趨勢是為越來越接近正確 RIR 的形狀。其中 FDWiener 的第二級(右)聲道會受第一級(左)聲道估測時所引起失真，在第二級的 RIR 將會有所變動是必然發生的事情，在此只能利用此圖來看一些趨勢而不能當做為絕對指標。

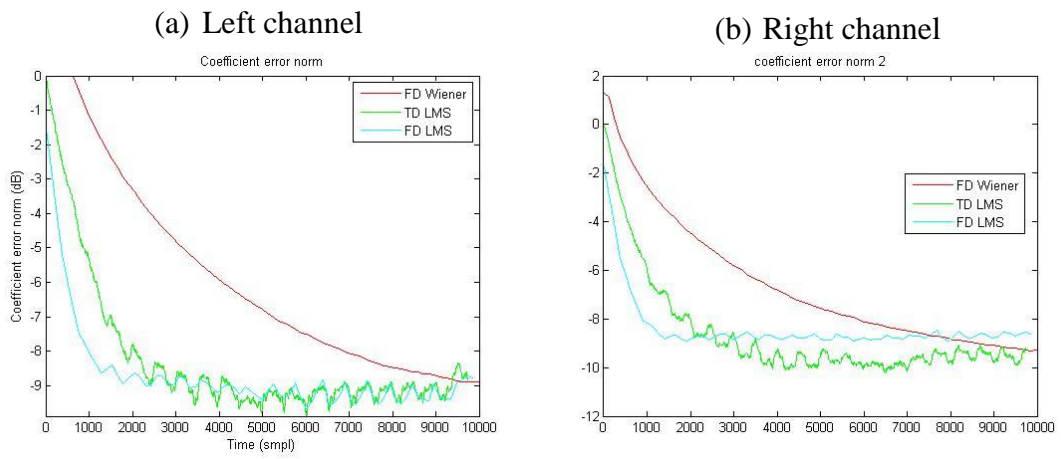
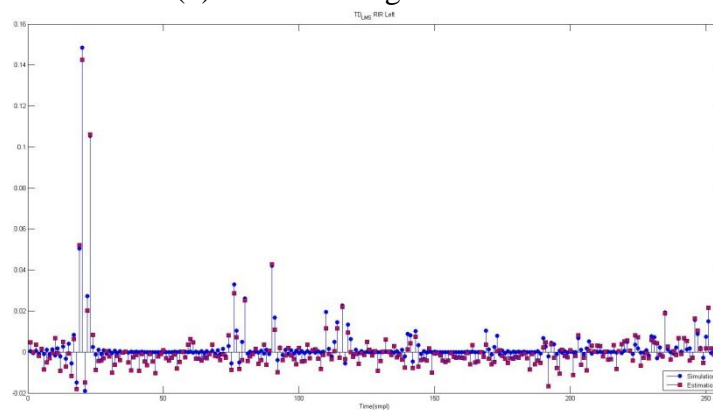
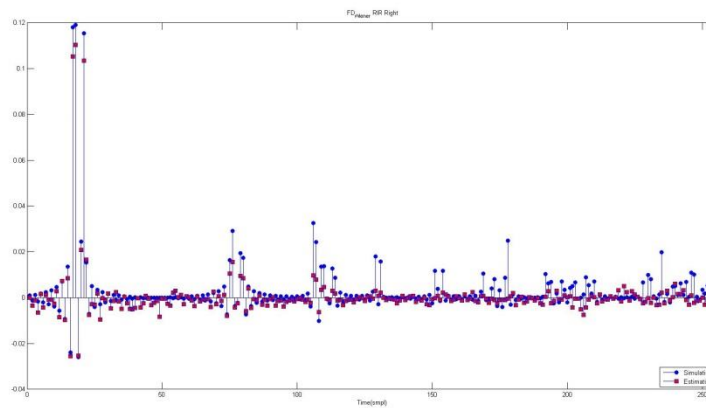
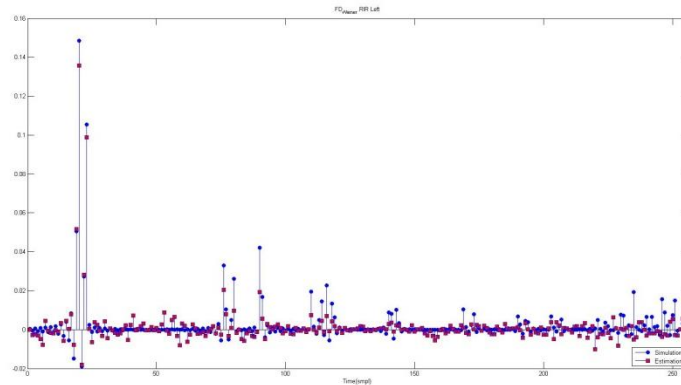
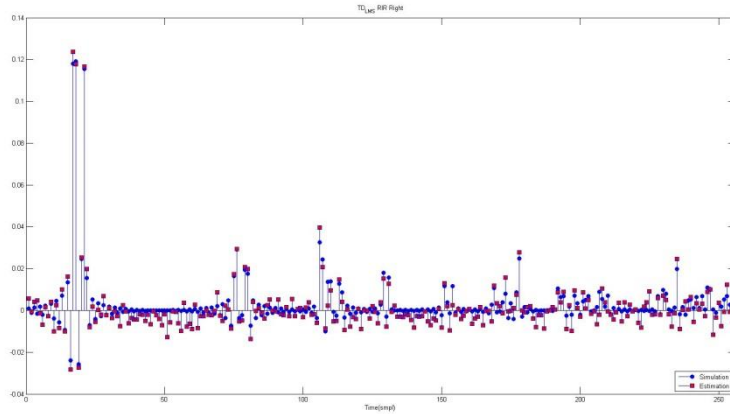
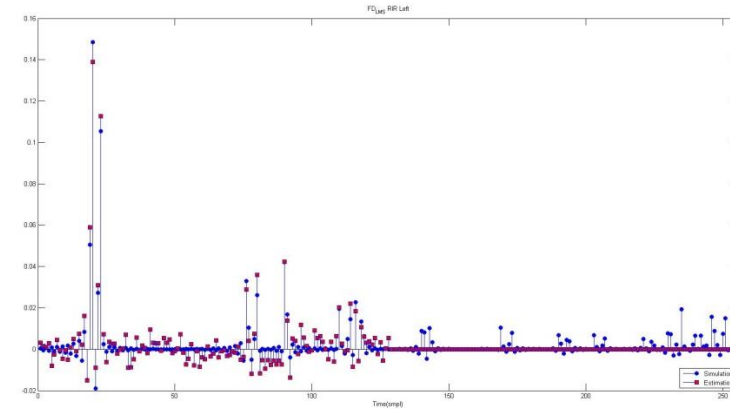


圖 4-29 SAEC Coefficient error norm

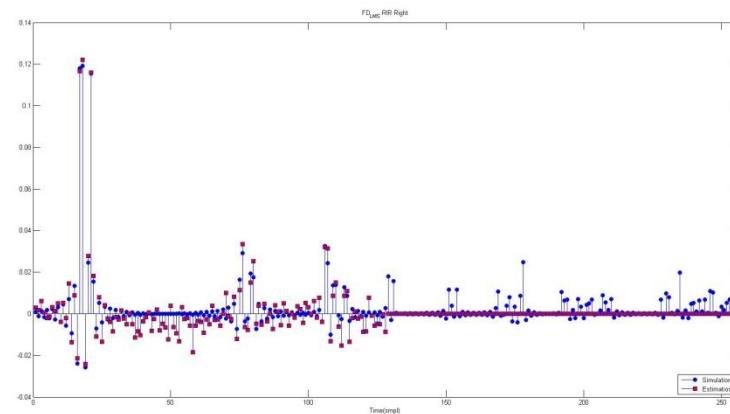




(d)TDLMS Right channel RIR



(e)FDLMS Left channel RIR



(f)FDLMS Right channel RIR

圖 4- 30 不同演算法對各聲道估測的 RIR

圖 4- 30 中除了(b)的 RIR 應該會受到些微改變外其他的估測都應該要越接近原始 RIR 越好。SAEC 的估測和 AEC 發生 Double-talk 的情況相同，(a)中的估測會完全視另一聲道為雜訊而對輸入訊號 y 做估測，在估測過程中會讓(b)中原始的 RIR 響應變為包含失真的響應。而在(c)、(d)、(e)、(f)的估測中就需要權衡兩聲道間的誤差做出更新的步驟。由於真實環境的 RIR 無法得到只能利用估測的方式獲得圖 4- 31 是不同 SAEC 演算法估測出來的結果，左右聲道差異主要不同的影響是空間特性、喇叭響應以及參考訊

號擷取過程雜訊不同所引起的。

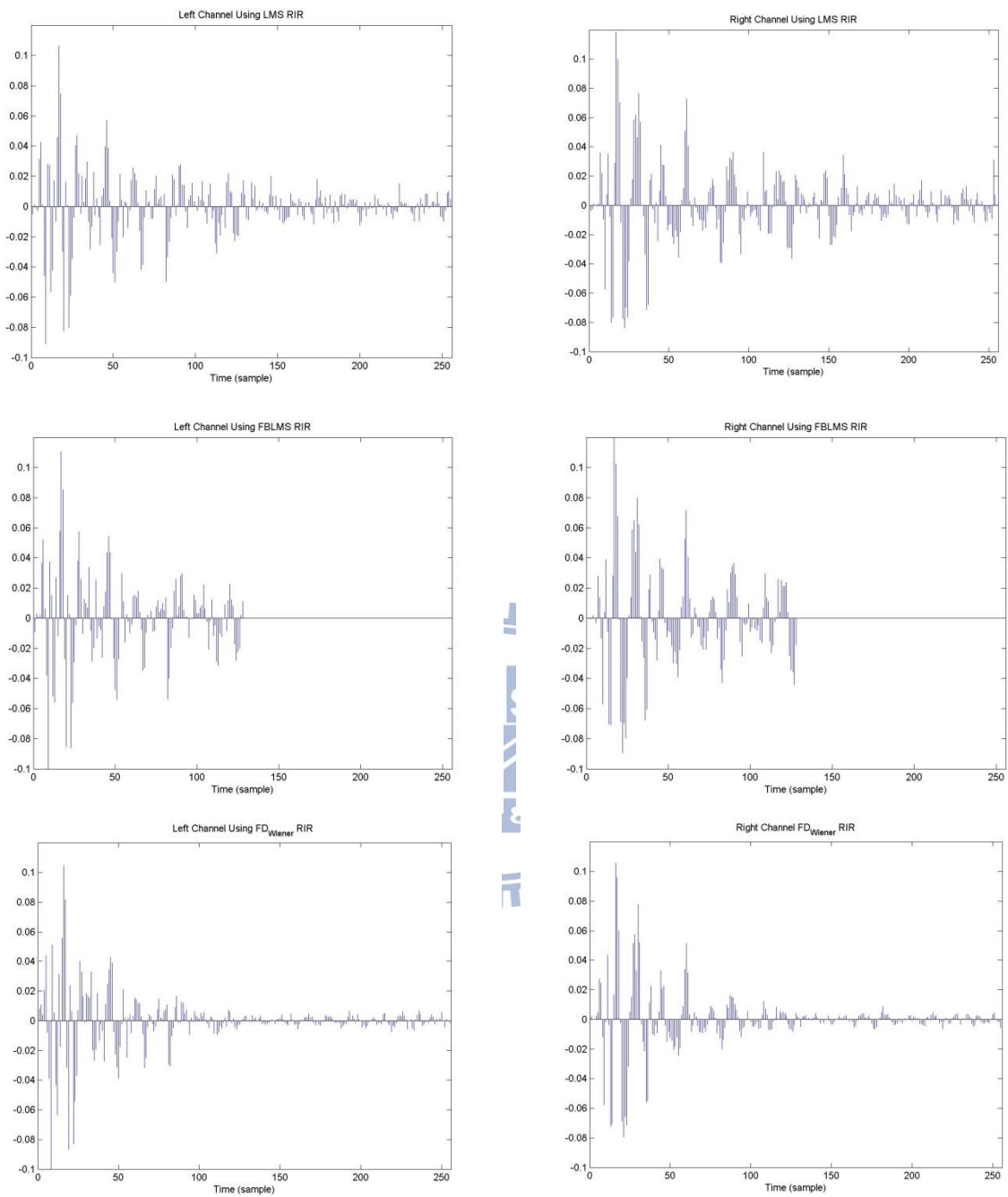


圖 4-31 真實環境左右聲道之 RIR

4.1.5 分析參考訊號對輸出訊號影響

選用頻域維納濾波器做為 AEC 演算法另一個優點是使用的遺忘因子 α 估測功率頻譜密度使用開迴路運算不受參考訊號大小的影響。LMS 運算中由於最速下降法的關係其中步長 μ 設計是和參考訊號的相關矩陣 \mathbf{R}_{xx} 中最大特徵值有關。參考訊號和播放訊號能量大小不同是很常見的問題實際狀況訊號會經過很多級音量調整的裝置，而擷取的參考訊號能量大小對有使用步長 μ 更新權重 \hat{W} 的演算法來說是具有效能影響。因此透過以下實驗可以說明參考訊號對輸出訊號的影響以及說明參考訊號能量大小的問題對頻域維納濾波器來說是優於其他演算法的優點。

圖 4-32 為模擬環境變動橫軸參考訊號由小而大產生的圖形，每一個點可以代表透過該演算法處理過後的訊號相對於輸入訊號 y 即為 ERLE。可以看到 LMS 方法在參考訊號 x 較小的時刻產生發散現象 y 訊號並沒有迴聲消除的功能，TDLMS 直到參考訊號到達 -23dBov 時才出現最佳的處理效能，FDLMS 到 -30dBov 時就能出現最佳處理效能，也就是說此時選擇的步長 μ 是符合參考訊號的相關矩陣 \mathbf{R}_{xx} 的最大特徵值相關，而選定的步長 μ 能滿足參考訊號 x 能量大小在一段範圍內的 ERLE 都能維持在相同效能，當參考訊號 x 能量逐漸增強 ERLE 效能又隨之下降。反觀在 FDWiener 的演算法卻能維持相同的 ERLE 效能，原因是在於 $\hat{Y}(\omega) = \begin{pmatrix} \Phi_{XY}(\omega) \\ \Phi_{XX}(\omega) \end{pmatrix}^{-1} X(\omega)$ 中 x 大小所影響因次會透過除法相互抵消讓運算結果維持相同效能，這算是頻域維納濾波器一項特殊的優點。

圖 4-33 為實際環境及參考訊號所構成的狀態，主要模擬和真實環境的差異影響 SAEC 結果的原因：第一點、RIR 較為複雜在 FIR 長度未能包含全部響應導致未估測成分影響迴聲消除結果。第二點、存在訊號延遲因素。第三點、參考訊號擷取過程中出現雜訊或訊號失真現象會。這些原因會導致 ERLE 的效能無法像模擬的結果一樣優異，先前利用實驗模擬討論過第一點利用增加 FIR 的長度改善估測的完整度，第二點利用相關係數找出時間延遲關係做時間對齊的步驟，第三點即為此實驗可以說明，LMS 演算法選擇步長 μ 與參考訊號的相關矩陣 \mathbf{R}_{xx} 的最大特徵值相關，但參考訊號 x 中有接收端未包含的成分會影響相關矩陣 \mathbf{R}_{xx} 也連帶影響了步長 μ 的選擇。在圖 4-32、圖 4-33 中當步長和 \mathbf{R}_{xx} 不匹配時就很快產生發散，但是使用頻域維納濾波器運算時迴聲消除能力不會有和參考訊號有關。

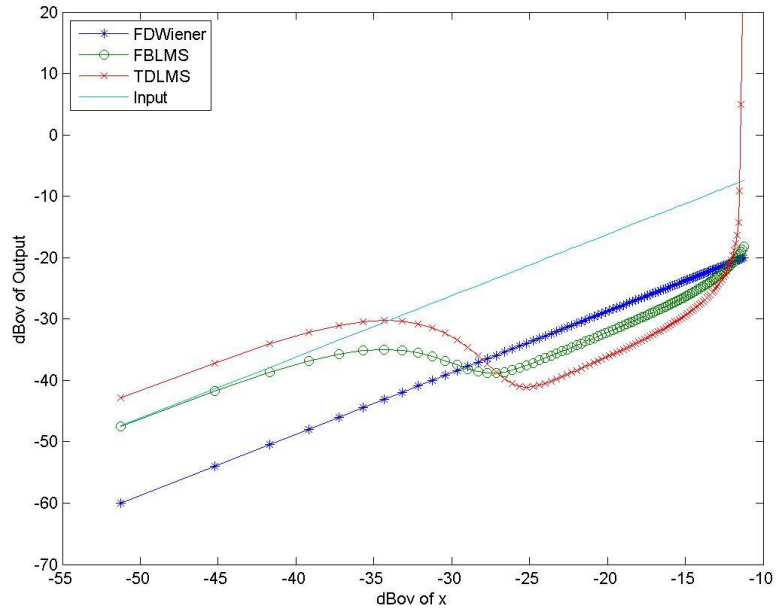


圖 4-32 模擬環境參考訊號對輸出的影響

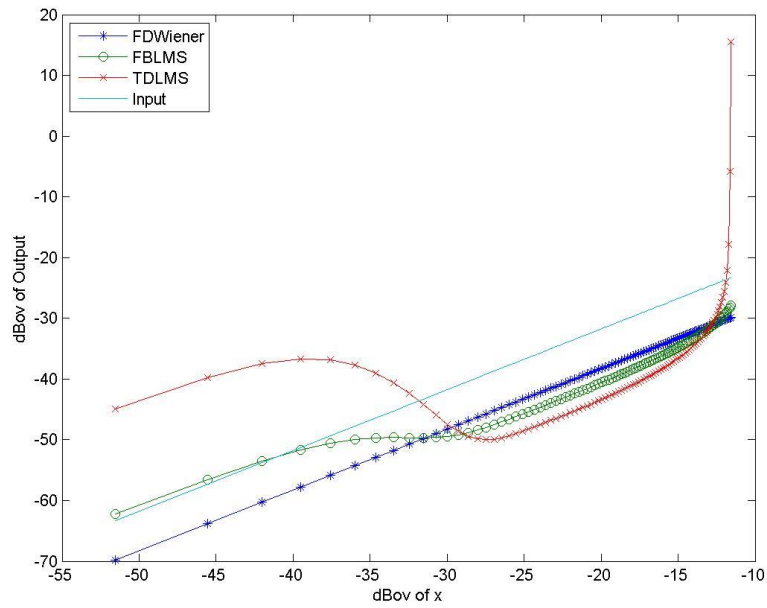


圖 4-33 實際環境及參考訊號對輸出的影響

4.2 SAEC 結合 Beamformer 效能評估

評估第三章節所提出的系統架構，如圖 3-4 智慧型電視應用之系統架構圖，波束形成的方法有很多種，以下將例舉與提出架構較類似的波束形成方法比較。

1. Generalized Sidelobe Canceller (GSC)[11]
2. Transfer Function Generalized Sidelobe Canceller (TF-GSC)[13]
3. Dual-source Transfer Function Generalized Sidelobe Canceller (DTF-GSC)[14]
4. 原架構中用時間延遲模型建構陣列拓樸向量取代轉移函式比值法(MVDR)

首先介紹評估標準為訊號干擾比之提升(Signal to Interference Ratio Improvement, SIRI)和訊號迴聲比之提升(Signal to Echo Ratio Improvement, SERI)[33]。

定義 SIR 和 SER 為

$$SIR = 10 \times \log_{10} \left\{ \frac{\sum_{n=P}^{P+Q} Y_S^2(n)}{\sum_{n=P}^{P+Q} Y_I^2(n)} \right\} \quad (dB) \quad (4.2.1)$$

$$SER = 10 \times \log_{10} \left\{ \frac{\sum_{n=P}^{P+Q} Y_S^2(n)}{\sum_{n=P}^{P+Q} Y_E^2(n)} \right\} \quad (dB) \quad (4.2.2)$$

其中 Y_S 為目標聲源經過演算法後輸出訊號， Y_I 為干擾源經過演算法後輸出訊號， Y_E 為迴聲經過演算法後輸出訊號， P 為時域訊號中目標聲源中所選取的起點而 Q 為終點。而 SIRI 和 SERI 的定義[33]為

$$SIRI = SIR_{out} - SIR_{in} \quad (4.2.3)$$

$$SERI = SER_{out} - SER_{in} \quad (4.2.4)$$

SIR_{out} 、 SER_{out} 為經過演算法後輸出求得之值， SIR_{in} 、 SER_{in} 為麥克風接收到原始訊號求得之值。

圖 4-34 為在不同輸入 SIR 訊號的情況下使用各演算法所提升 SIR 的程度，SIR 值越高代表通過此波束形成器干擾源抑制好且目標聲源失真低的結果。有 SIRI 高到低依序為，TF-MVDR+NE、DTF-GSC、TF-GSC、GSC、MVDR+NE。演算法經由 TFR 校準後的 AMV 具有較高的 SIR 提升能力，主要 MVDR+NE 效能最差的原因是經由 MVDR 訓練干擾時未能準確的讓約束條件下再目標聲源方向，因此求出的權重已經導致目標聲

源失真，在雜訊估測時也不能準確的阻擋住目標聲源方向在 ANC 的處理上有自相抵消 (Self-Cancelling) 的狀況出現，進而導致 SIR 未能提升。

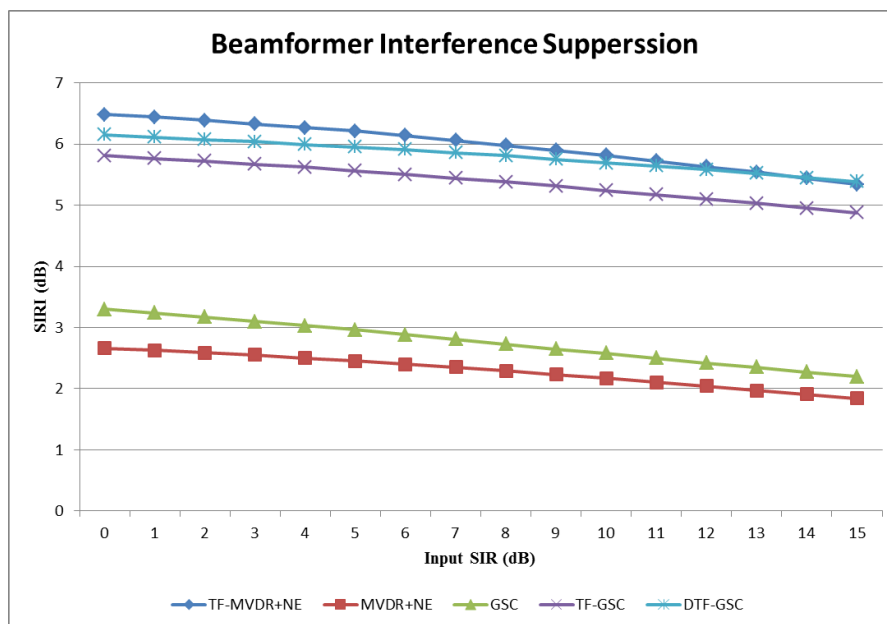


圖 4-34 Beamformer SIRI

主要提升 SER 的演算法為 SAEC，而每個演算法架構都是通過相同的 SAEC 所以迴聲抑制上都已經做了相同的運算，不同輸入的 SER 影響也不大，SERI 大約 6.54dB。SAEC 輸出訊號依舊會有剩餘迴聲未消除，此時，波束形成器會視剩餘迴聲為雜訊處理可以藉由各自演算法對雜訊之消除能力而提升輸出的 SER。SERI 值越高代表能抑制越多迴聲且保持目標聲源方向的語音。依圖 4-35 效能的優劣依序為 TF-MVDR+NE、TF-GSC、GSC、MVDR+NE、DTF-GSC。TF-MVDR+NE 和 TF-GSC 的差異在於有對干擾源方向做壓抑當迴聲也從此方向進入時會受到抑制所以會有比較好的 SERI，另外 TF-MVDR+NE 和 DTF-GSC 差異在於 BM 的不同，前者設計 BM 時維度為 M-1 除了目標聲源方向被阻擋外其餘方向都希望被估測為雜訊，但後者為維度為 M-2 比前者少會發生除了目標聲源方向外其餘方向也被阻擋意味著有某些方向的雜訊未被估測到，雜訊的估測好壞會導致 ANC 是否能成功消除雜訊的關鍵。

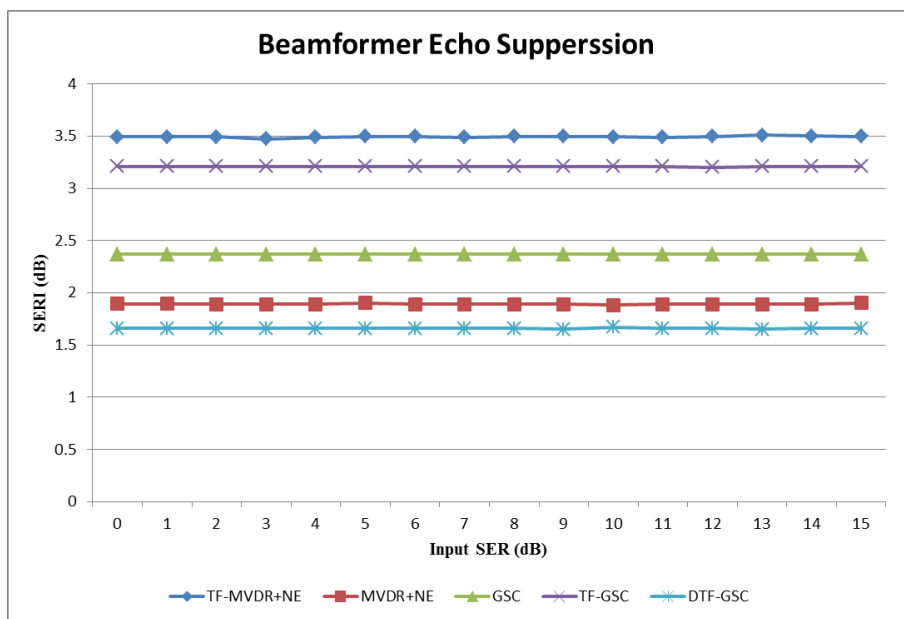


圖 4-35 Beamformer SERI

以上兩個實驗為目標聲源分別對干擾或迴聲當做輸入訊號經過波束形成器的討論，接下來實驗是波束形成器處理干擾和迴聲同時存在的情況。

Case	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
SIR	0	0	0	0	5	5	5	5	10	10	10	10	15	15	15	15
SER	0	5	10	15	0	5	10	15	0	5	10	15	0	5	10	15

表 4-11 輸入音檔混合情況

固定目標聲源的能量混入不同分量的干擾及迴聲訊號成為每個不同的音檔，指定的 SIR 及 SER 成份為表 4-11 所示，可以透過此音檔評估出系統對混合干擾及迴音成分訊號的抑制能力。

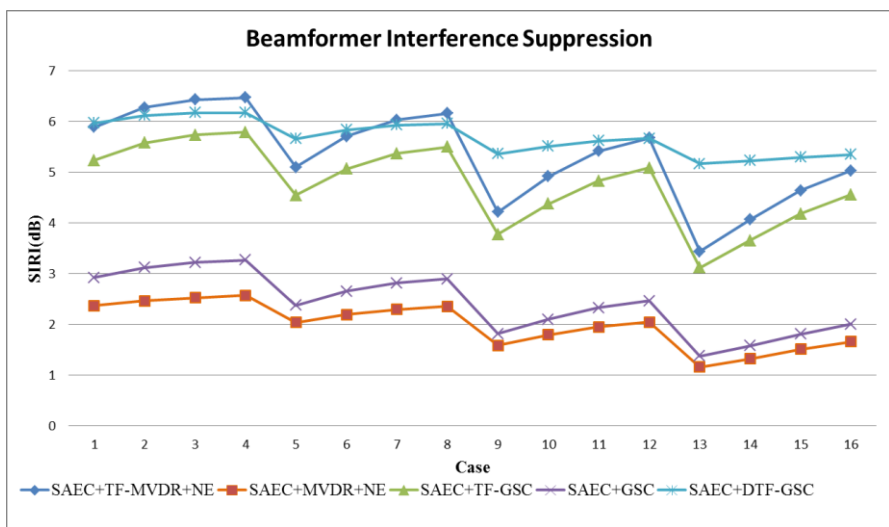


圖 4-36 Beamformer interference suppression

由於混入干擾源和迴聲的訊號通過演算法的運算會相互影響，圖 4- 36 為波束形成器對干擾壓抑程度的評估，當迴聲能量大於干擾能量增多時，在干擾的抑制會受到影響 TF-MVDR+NE、TF-GSC 可以看出很明顯的落差，在 Case5、Case9、Case13 狀況下明顯看出受到迴聲的牽連而導致干擾抑制能力下降，主要是 ANC 運算內的 Y_{NC} 為了消除存在於 Y_{FBF} 之迴聲所造成的，這些演算法中除了 DTF-GSC 在 BM 設計把干擾源方向阻擋掉不讓其成為雜訊部分，其餘都會讓該方向的訊號通過，由圖 4- 36 也能明顯看見 DTF-GSC 在此情況干擾源抑制的效能也會比較好。但因輸入是混合迴聲和干擾訊號，在干擾源有好的表現同時也會損失在迴聲的效能，圖 4- 37，抑制剩餘迴聲的情況，在提出的架構底下有較好的 SERI 的提升，由於 TF-MVDR 除了在目標聲源外保持無失真其他方向上的壓抑比 GSC 架構的方法來的好因此在剩餘迴聲的消除有較好的效果。

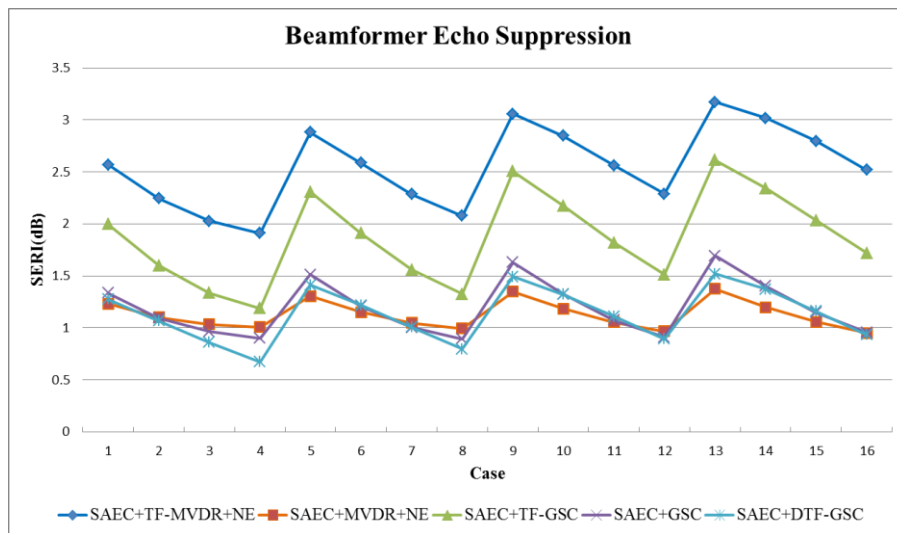


圖 4- 37 Beamformer echo suppression

各演算法都使用相同 SAEC，而 SAEC 表現優劣重要的影響是參考訊號對麥克風接收訊號間延遲關係和演算法中參數的選擇，通過迴聲消除後 SERI 維持在 6.54dB。而波束形成器可以再抑制殘餘迴聲，圖 4- 38 為不同情況下 SERI 的情況，再干擾越小時波束形成器抑制迴聲的情況越好。可明顯看出 TF-MVDR+NE 的效能優於其他演算法，而未經過 TFR 的演算法受到目標聲源的失真導致評估 SERI 效能降低。

定義 SIER(Signal to Interference-plus-Echo Ratio)

$$SIER = 10 \times \log_{10} \left\{ \frac{\sum_{n=P}^{P+Q} Y_S^2(n)}{\sum_{n=P}^{P+Q} [Y_I^2(n) + Y_E^2(n)]} \right\} \quad (dB) \quad (4.2.5)$$

其中 Y_S 為目標聲源經過演算法後輸出訊號， Y_I 為干擾源經過演算法後輸出訊號， Y_E 為迴聲經過演算法後輸出訊號， P 為時域訊號中目標聲源中所選取的起點而 Q 為終點。

SIERI(Signal to Interference-plus-Echo Ratio Improvement)定義為

$$SIERI = SIER_{out} - SIER_{in} \quad (4.2.6)$$

$SIER_{out}$ 為經過演算法後輸出求得之值， $SIER_{in}$ 為麥克風接收到原始訊號求得之值。

SIER 可以看出整體雜訊受演算法抑制效果，圖 4- 38 為系統整體提升的 SIER 趨勢。

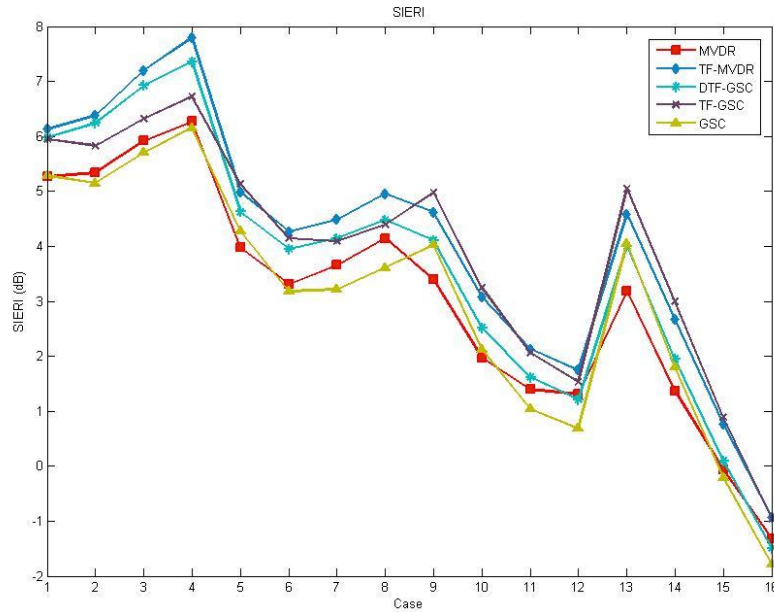


圖 4- 38 System interference and echo suppression

由圖 4- 36 和圖 4- 37 分別討論波束形成器在干擾和迴聲上的抑制，TF-MVDR+NE 的干擾抑制雖然不比 DTF-GSC 來的好，但是在迴聲的消除卻好上許多。要看整體的效能需透過圖 4- 38 說明，在 SIR 較差的狀況下 SIER 效能提升由大而小為：TF-MVDR+NE、DTF-GSC、TF-GSC、MVDR+NE、GSC。

接著介紹評估標準為、對數譜失真(Log Spectral Distortion, LSD)，計算原始語音與處理後的頻譜對數差異值，可表示為

$$LSD = \frac{1}{I} \times \sum_{i=1}^I \sqrt{\frac{1}{W} \sum_{\omega=1}^W (20 \times \log_{10} |A(\omega)S(i, \omega)| - 20 \times \log_{10} |Y(i, \omega)|)^2} \quad (4.2.7)$$

其中 $A(\omega)$ 為環境中頻域下的 RIR、 $S(i, \omega)$ 為要透過人工嘴(HATS)播放出的音檔、 $Y(i, \omega)$ 為經過演算法處理過後的輸出訊號、 ω 和 i 分別為頻帶和音框的指標，運算 LSD 時必須先選擇音檔長度範圍進行運算，計算出的對數譜失真越小表示語音品質越高。

在 LSD 所評估的音檔為相同程度的 SIR 和 SER 所混合當做輸入訊號在下列 x 坐標軸以 SIR/SER(dB)表示，圖 4- 39 和圖 4- 40 要相互配合來評估，LSD 單純利用能量差異來評估語音衰減程度，PESQ 有包含人耳聽覺模型的感受度做計算。雖然說 MVDR 在低 SIR/SER 的 LSD 較低，可是 PESQ 確很差，是因為 LSD 計算的語音區段雖然整體的能量有接近乾淨語音，但是連語音成份也一同的衰減。

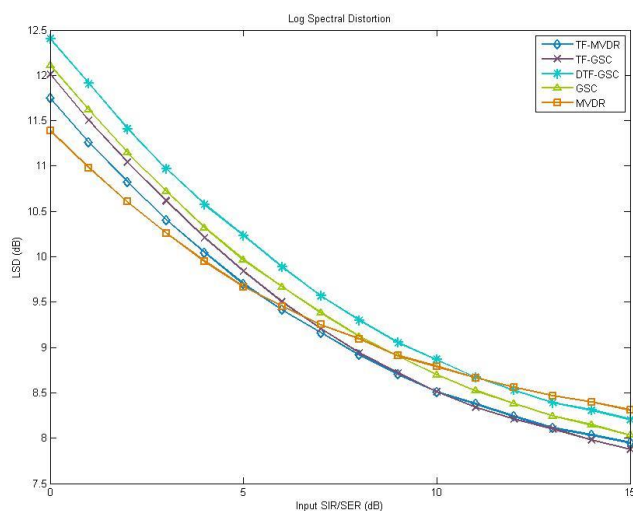


圖 4- 39 Log spectrum distortion (LSD)

最後一種評估標準為知覺聽語評估(Perceptual Evaluation of Speech Quality, PESQ) [34]由 ITU-T P.862[34]所提出，為一套自動化客觀評估語音品質的標準，其目的是以客觀儀器的量測來取代過往以人耳聽覺判定的陪審團制度，以減少語音評估時的繁瑣複雜。PESQ 基本上就是比較原始的輸入訊號及演算法處理後的輸出訊號之間的差異，而 PESQ 量測的結果以 Mean Opinion Score (MOS)的評分法來量化，如表 4- 12 所示。MOS 是以 5 分制來表示，1 分即為最差分數，而 5 分即為最佳分數。MOS 的分數來自於各類語音訊號量測結果的算術平均數。以統計上來說，若 MOS 超過 4 分，大部分的使用者會覺得滿意或非常滿意。

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

表 4- 12 Mean Opinion Score

原本的 PESQ 也會受到輸入訊號的品質所受到變動，故圖 4- 40 是利用經過演算法處理後的訊號和原本輸入受到干擾的訊號兩者之間 PESQ 之差來做評估。越大的數值越具有提升語音品質的效果，數值提升因為環境中的雜訊受到抑制或是受破壞的語音被修復。而讓數值下降的原因語音受到破壞或是環境的噪音無法壓抑。有利用 TFR 演算法校正過麥克風和目標聲源之間的空間關係的結果有不錯的表現。主要是有維持準確的目標聲源方向讓人聲不受到空間濾播器的失配問題所壓抑，能保留較完整的語音成份。

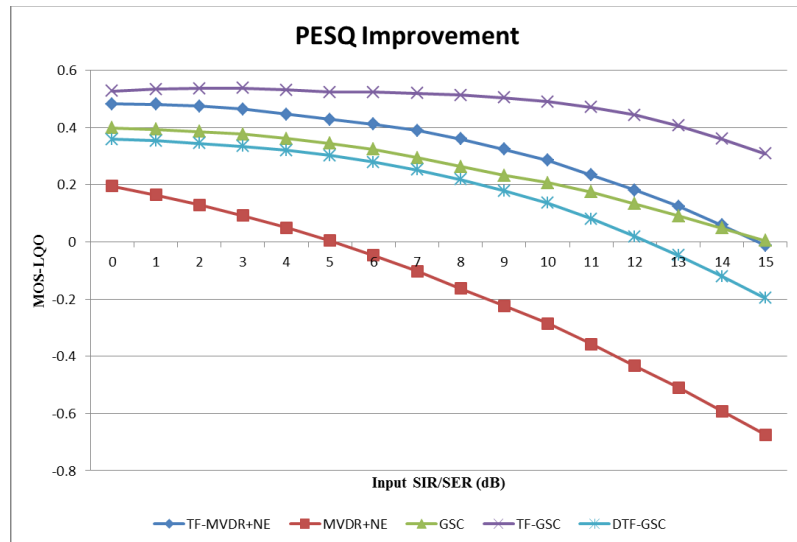


圖 4- 40 PESQ improvement of system

圖 4- 41 為第一顆麥克風所接收到的語音時域與頻域圖，圖 4- 42 (a)受到 SER=0dB 迴聲以及 SIR=0dB 干擾訊號混合語音的輸入。圖 4- 42 (b)~(f)為各演算法處理完的結果。本論文題出之方法(b)在目標聲源的維持有相當程度的表現，而(d)沒有用 TFR 的方法導致語音成份出現失真的現象影響了 SNR，(c)、(e)在干擾及迴聲抑制各有特色有不同的效果。

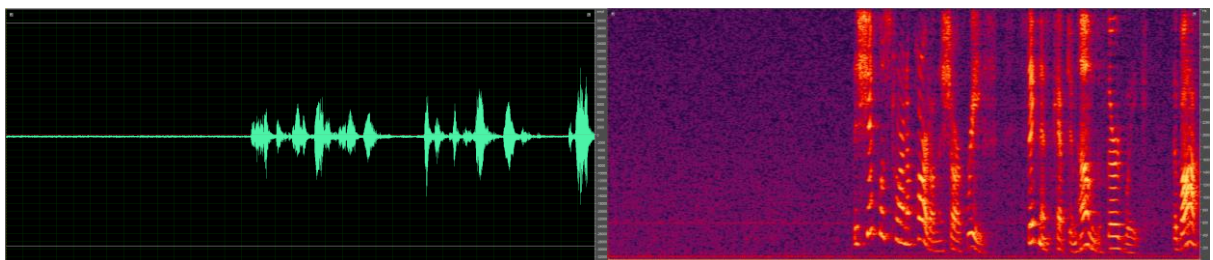
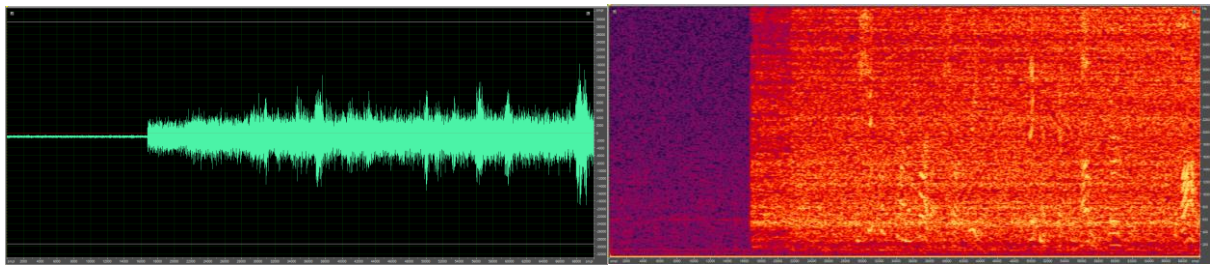
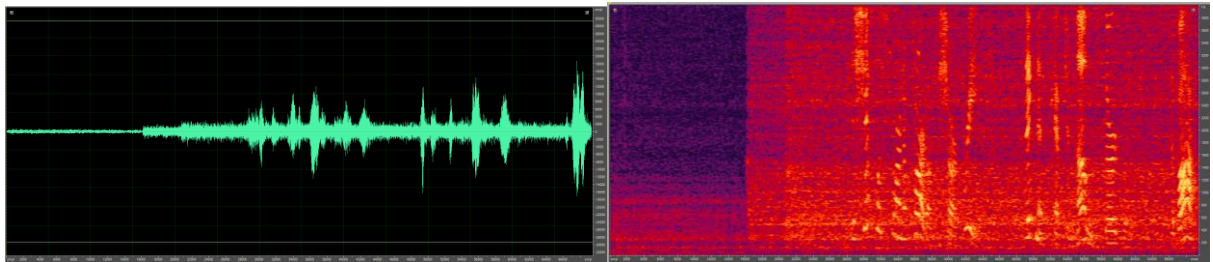


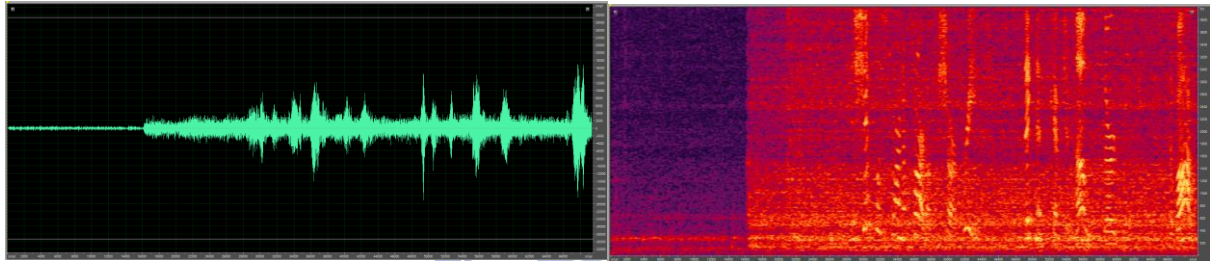
圖 4- 41 original speech



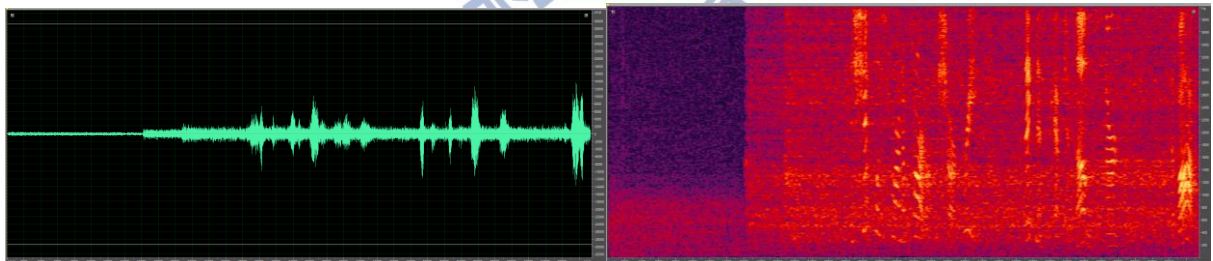
(a) Contaminated Speech



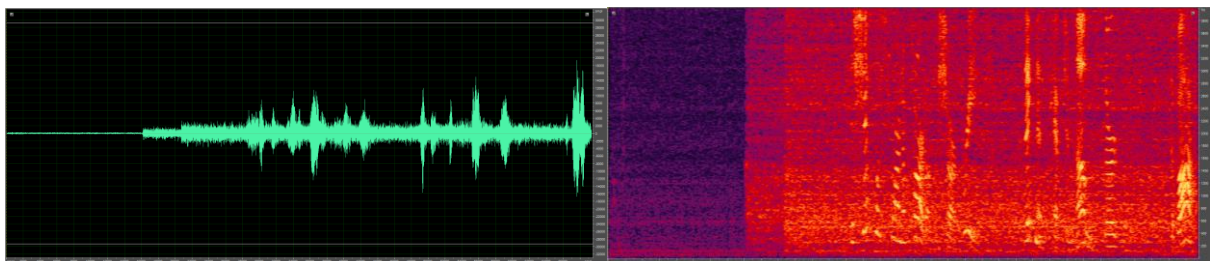
(b) TF-MVDR + NE algorithm



(c) DTF-GSC algorithm



(d) MVDR + NE algorithm



(e) TF-GSC algorithm

圖 4-42 各演算法在 $SIR0(\text{dB})$, $SER0(\text{dB})$ 下運算之結果

第五章 結論

5.1 研究成果

本論文提出一套具有立體聲迴聲消除和空間濾波效能之系統。利用頻域維納濾波器快速收斂特性設計立體聲迴聲消除提供穩定的迴聲消除能力，配合空間濾波器對干擾源和剩餘迴聲進行壓抑。4.1 章節中分析了有關 2.2 章節 AEC 內容，比較不同演算法迴聲消除程度、收斂速度效能及找出適當的參數值；也說明在 3.3 章節中頻域維納濾波器實現 SAEC 的優點，利用維納解直接透過一階遞迴計算出 PSD 求得開回路最佳解，選擇適當 α 後在迴聲消除的狀態出現非穩態訊號且有發生 Double-talk 情形可以保持迴聲估測受干擾影響較小，因為更新權重不受到誤差訊號影響也不會有參考訊號強弱導致效能的衰減的問題發生；也討論在實驗中遇到系統中訊號延遲所造成的影響利用相關係數找出延遲時間來解決此問題。4.2 章節中評估整體系統，比較不同空間濾波器於干擾和剩餘迴聲抑制能力；本論文在於修改了 TF-GSC 的架構利用 MVDR 作為 FBF 的空間前處理進而提升整體效能。論文貢獻在結合頻域維納濾波器於 SAEC 消除雙聲道訊號，不同於傳統立體聲架構下進行運算；利用空間前處理提升空間濾波器效能，當干擾強度高的時候透過 MVDR 消除干擾源，干擾強度低的時候會有較多維度用於雜訊估測得到較好的雜訊抑制能力。

5.2 未來展望

SAEC 架構中第二級聲道估測時會受第一級失真的影響，如果可以很準確的估測出第二級聲道的迴聲路徑將會使得迴聲消除上有更好的改善，在訊號時間延遲問題能也利用適應性方式求得且更細部找出麥克風間與各聲道的差異能進而改善因訊號延遲不匹配的問題；目前 MVDR 中 R_{NN} 和 AMV 以及雜訊估測中 BM 都須透過事前訓練而得，如果能利用 VAD 技術改善訓練的動作成為適應性的更新在智慧型電視的應用上將更具有智慧性。

参考文献

- [1] J. Park, G. Jang, J. Kim, S. Kim, “Acoustic interference cancellation for a voice-driven interface in smart TVs,” *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, February 2013
- [2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 4th edition, 2002.
- [3] H.L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*, John Wiley & Sons, Inc., New York, 2002.
- [4] A. Gilloire and M. Vetterli, “Adaptive Filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation,” *IEEE Transactions on Signal Processing*, vol. 40, no. 8, August 1992
- [5] E. Hansler, G. Schmidt, “Acoustic Echo and Noise Control: A Practical Approach,” John Wiley & Sons, Inc. Chapter 5 pp. 53-60, 2004.
- [6] Y. Lu, R. Fowler, W. Tian, and L. Thompson, “Enhancing echo cancellation via estimation of delay,” *IEEE Transactions on Signal Processing*, vol. 53, no. 11, November 2005.
- [7] B. V. Veen and K. Buckley, “Beamforming: A versatile approach to spatial filtering,” *IEEE Acoust., Speech, Signal Process. Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [8] O. L. Frost III, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [9] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE* vol. 57, no. 8, August 1969
- [10] Y. H. Chen and C. T. Chiang, “Adaptive beamforming using the constrained kalman filter,” *IEEE Transactions On Antennas And Propagation*, vol. 41, no. 11, November 1993.
- [11] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.
- [12] K. Buckley, “A short proof of the equivalence of LCMV and GSC beamforming,” *IEEE Trans. Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.
- [13] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

- [14] G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized Sidelobe canceller," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, may 2008.
- [15] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. 17, No. 6, August 2009
- [16] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2055–2063, Aug. 1996.
- [17] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.
- [18] C.H Chen, "A dual-channel sound sources separation system using transfer function ratio approach," *NCTU, Electrical and Control Engineering thesis*, May 2009.
- [19] B. Widrow and M. E. Hoff Jr. "Adaptive switching circuits", *IRE WESCON Conv. Rec.*, pp.96 -104 1960
- [20] J. J. Shynk, "Frequency-domain and multirate adaptive filtering", *IEEE Signal Processing Mag.*, pp.14 -37 1992
- [21] F. Yang, M. Wu, and J. Yang, "Stereophonic acoustic echo suppression based on wiener filter in the short-time Fourier transform domain," *IEEE Signal Processing Letters*, Vol. 19, No. 4, April 2012.
- [22] J.S. Hu, and M.T. Lee, "Spatially Pre-processed Target-to-Jammer Ratio Weighted Wiener Filter Using Two Microphones," *IEEE Green Circuits and Systems (ICGCS)*, pp.180-185, June 2010.
- [23] W. Kellermann, "Joint design of acoustic echo cancellation and adaptive beamforming for microphone arrays," In: *Proc. 5th Internat. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Imperial College, London, UK, pp. 81–84. 1997
- [24] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," *Proc. ICASSP1997*, pp. 219-222, May 1997
- [25] W. Herbordt, W. Kellermann, "GSAEC – acoustic echo cancellation embedded into the generalized sidelobe canceller," *Proc. EUSIPCO 2000*, vol. 3, pp. 143-1846, September 2000.
- [26] W. Herbordt, W. Kellermann, and S. Nakamura, "Combined optimization of LCMV beamforming and acoustic echo cancellation," *Proc. EURASIP European Signal Processing Conference*, 2004.

- [27] G. Reuven, S. Gannot, I. Cohen, "Joint acoustic echo cancellation and transfer function GSC in the frequency domain," In: Proc. 23rd IEEE Convention of the Electrical and Electronic Engineers in Israel, Herzlia, Israel, 6–7 pp. 412–415. September 2004.
- [28] G. Reuven, S. Gannot, and I. Cohen, "Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller," Special Iss. Speech Commun. Speech Enhancement, vol. 49, pp.623–635, Jul.–Aug. 2007.
- [29] A. Hirano and A. Sugiyama, "A compact multi-channel echo canceller with a single adaptive filter per channel," Proceedings of IEEE ICASSP. pp1922-1925. 1992.
- [30] M. M. Sondhi, D. R. Morgan and J. L. Hall, "Stereophonic acoustic echo cancellation-An overview of the fundamental problem," IEEE Signal Processing Letters, Vol. 2, No. 8, August 1995.
- [31] T. Gänsler and J. Benesty, "Stereophonic acoustic echo cancellation and two channel adaptive filtering: An overview," Int. J. Adapt. Contr. Signal Process., vol. 14, pp.565 -586. 2000.
- [32] E. Habets, "Room impulse response (RIR) generator," Jul. 2006 [Online]. Available: http://home.tiscali.nl/ehabets/rir_generator.html.
- [33] Recommendation ITU-T G.160, "Amendment 1: Revised Appendix II – Objective measures for the characterization of the basic functioning of noise reduction algorithms," International Telecommunication Union, November 2009
- [34] Recommendation ITU-T P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," International Telecommunication Union, Feb. 2001