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

具陣列拓樸向量校正之多重訊號分類演算法

於多聲源切音與分離

Multiple Source Segmentation and Separation Using

MUSIC Algorithm with Calibrated Array Manifold Vector

研 究 生： 呂 孟 瑋

指導教授： 胡 竹 生 博士

中 華 民 國 一 百 零 二 年 十 月

具陣列拓樸向量校正之多重訊號分類演算法於多聲源切音與分離

Multiple Source Segmentation and Separation Using MUSIC Algorithm with Calibrated Array Manifold Vector

研 究 生：呂 孟 瑋　　Student：Meng-Wei Lu

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

具陣列拓樸向量校正之多重訊號分類演算法於多聲源切音與分離

研究生：呂　孟　瑋　　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校正過之陣列拓樸向量(Array Manifold Vector)，提升多重訊號分類演算法(Multiple Signal Classification)效果在寬頻估測時的準確度，並實現多聲源切音與分離的方法。本方法結合了聲源頻譜與空間分佈資訊，利用機率決策對未知數量聲源方位進行分類，並將不同聲源語音利用波束形成原理進行切音與分離。本方法由於進行了陣列拓樸向量的完善校正，保證在低訊噪比下對多聲源的方位保有相當程度的正確率，且可排除錯誤偵測的聲源方位。

Multiple Source Segmentation and Separation Using MUSIC Algorithm with Calibrated Array Manifold Vector

Student： Ryan Lu Advisor： Prof. Jwu-Sheng Hu

Master Program of Sound and Music Innovative Technologies

National Chiao Tung University

ABSTRACT

This thesis proposes a system structure for multiple sound sources segmentation and separation using MUSIC (Multiple Signal Classification) algorithm. Using a calibrated array manifold vector, the proposed calibration method improves the accuracy of the MUSIC algorithm for wide-band detections, hence providing high accuracy source segmentation and separation results. The system structure uses a multiple signal classification algorithm to detect the location of sound sources and estimate their spectrum distributions. The multiple sources tracking method is implemented by a probability decision method regarding spatial and spectrum distributions. Using the estimated directivity, multiple sources were extracted from the array signals using beamforming methods. This proposed system structure can track and separate multiple sources at the same time and maintain high detection rate under very low SNR conditions.

誌　謝

看著眼下的自己，會有種兩年頓時就過去的錯覺，快如一次的朝夕。可是只要停步稍作回首，就會見到來路上不斷造就現在自己的，那時光的風刻雨雕。在所有遇到的磨鍊與引導中最為感謝的，是胡竹生博士這兩年來給予的耐心教誨，老師不僅教會了我用嚴謹的思考去處理問題，更以深廣的視野與不竭的熱誠帶給了我各種啟發，引領我用正確的態度去面對世界。

過去兩年，實驗室裡的大家就猶如自己在新竹的家人，朝夕共處，同窗研究。其中最感謝的當然是超厲害的唐哥，帶給我各式各樣的幫助與建議，讓我在聲音工程領域的學習過程順利許多。再來也同樣感謝的是鳴哥，還記得自己剛進實驗室的時候在你桌上看到了一台 uDAC2，那時候就發現我們有許多共同的興趣，是你帶領我熟悉實驗室以及學程的一切，並給予了剛踏入研究所生活的我許多鼓勵。感謝時時給予精闢建議的阿吉學長，常常在游泳池碰到的 Judo 學長，帥氣幽默的勁源學長，深夜才會出現的 Alpha 學長，聲音組未來的大哥大耕維學長，每天都會準時提醒大家吃飯的昭男學長，愛喝調酒中所謂「男人的苦味」又吃不胖的翰哥，在實驗室好像一直在看影集的建廷，還有帥氣會畫畫的Daniel。私底下其實很愛玩遊戲的鳴遠，人生勝利哥一直跑日本都不會膩的期元，人很親切又可愛的哲宇，同樣是蘋果粉的凱翔，現在沒什麼鬍子的鬍子，活潑的學妹 Winnie 跟錢丹，還有聲音組的超級新秀知琬。大家一起買飯一起出遊一起東跑跑西跑跑，陪我走過了冬的凜冽跟夏的狂熱，我會記得這些日子的。

這邊要特別感謝佑軒，我像個控制欲強的女友要你隨傳隨到，要你在精神時光屋裡幫我做好多次實驗，真是太感謝你了哈哈，也順便感謝你現在翻開了這本論文讓我沒有白印，還認真讀了我的誌謝讓我沒有白寫。特別感謝罐頭，因為對 3C 產品的熱誠讓我們很有話聊，我不會忘記在每周一次約定好的游泳時程裡，我們對著激灩的池水暢談夢想與人生。當然還要感謝小山東，從學程的課程選擇，挑戰 CLORK 筆電樂團，到做專題，趕論文，我們幾乎一直是同病相憐著，也不知道在空無一人的實驗室裡一起熬過了多少夜晚，閒扯過多少入流與不入流的淡，醒著的，醉過的，總之我會記得這些的。

除了實驗室的大家庭，聲音學程跟音樂所的朋友，也讓一直鑽研理工科的自己見到了很不一樣的風景。首先要感謝的當然是學程一姊可柔，愛唱歌愛揪團吃吃喝喝看電影買衣服的好夥伴，兩年來真受過妳不少照顧了！還有Tina，坤熊，峻豪，鄭博，許大哥，鈺群，Play，我會記得一起同台演出CLORK筆電樂團以及辦WOCMAT的那些日子。平日常打打鬧鬧說說笑笑的EG，愛畫漫畫的燴飯，交男朋友就脫離游泳戰隊的小容，對後搖非常熱愛的文青小涼，還有愛做健康料理懂我文字一起創作畢業歌曲還得了第二名的欣儀，因為有了你們，我的碩士生活充滿了各式各樣的驚喜。

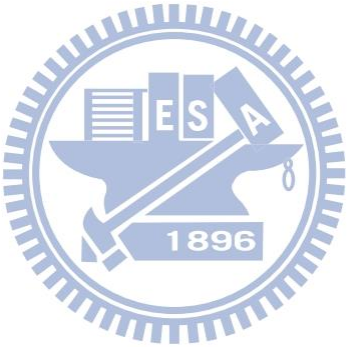
當然還有很多朋友也一直都與我不計時不計量地共享彼此的生命：深刻卻短暫的，溫和而長久的，一個月，三個月，四年，或終至十年的。去美國的文旭，去德國的玟媗，去法國的玳薇，需要時常不在的志豪，常讓志豪不在的小百合，湊在一塊就歡笑不斷的梳子跟詩茹，晚上愛跑步的花花，晚上不愛睡覺的郁祺，一直很小范的多年室友小范，總挑咪挺時段揪團去健身房的魚丸・・・。背對時光前進的這條路上，我的勇氣來自於你們與我並行，謝謝你們。不論爾後將前往什麼地方什麼時間，你們身上都會有我緊隨的思念。

最後，在此僅以本論文向永遠支持我的家人獻上最誠摯的謝意，我愛你們。

目　錄

表 列

圖 列



第一章　緒論

1.1 研究動機

耳朵是種遠超乎人想像的一種精密儀器。人類於日常生活中，在在面臨著需要依靠耳朵接受環境中的聲音以做出判斷的情況。在喧囂的街頭，人類可以僅憑一聲叫喊就辨別出是否是熟人，是從哪個方向叫喊的，甚至是與己身大概離的距離等等資訊。在四面八方擁來人聲的嘈雜餐廳中，人也可以憑意念，專注聆聽眼前的人說的每字每句。這種能辨別聲音來向，並從此方向取得最多資訊量的能力，正是多麥克風陣列的各種研究與應用中，一直希望能夠實現的。

在常見的語音系統中，大部分的情況都是對單一聲源訊號進行語音純化。但是當使用情境是屬於如會議對話之類的多講者應用，就不能只針對單一聲源做處理而忽視其他發話者，必須要對當前訊號進行反應且同時對多訊號進行處理。對於這樣的需求，於是有了多麥克風陣列的研究。

利用多麥克風陣列的物理分布特性，可以經由一些幾何運算對聲源方位進行估測，這些方法稱之為(Direction of Arrival estimation，DOA)。聲源方位的估測方式有許多不同種類，如對不同聲源方位估測可能機率的 ML (Maximum Likelihood)法，利用聲源到不同麥克風時間差異的 TDE (Time Delay Estimation)，利用聲源能量差異進行的聲源估測，以及對麥克風訊號進行特徵分解的 Eigenstructure Method。其中，Eigenstructure Method 的方法在估測多聲源訊號與對抗雜訊的穩健程度都有著不錯的效果，本論文即是使用 Eigenstructure Method 中的多重訊號分類演算法 MUSIC (Multiple Signal Classification) 來估算聲源方向。

1.2 研究目標

本論文提出一個實現高準度之多聲源切音與分離的系統。系統中首先是利用 MUSIC 演算法估測數個聲源方向，並同時估量各個聲源在空間中的分佈情形。利用此一能量的空間分布資訊，再結合機率的決策法，使得系統能對變動的聲源方位做連續追蹤。追蹤的結果再次利用機率的決策法來分類出不同聲源各自的方位變化，最後再利用束波器(Beamformer)做純化分離，切出各個聲源的語音紀錄。另外，本論文亦提出了一套陣列拓樸向量(Array Manifold Vector)的校正方法，此一校正過之 Array Manifold Vector 能為前端的 MUSIC 演算法有效提高估測聲源方位的準確性，同時也能為後端的 Beamformer 提升語音純化與分離的效果。

於是在此將本論文的目標分為：

1. 利用選定的聲源方位估測方法，進行多聲源方位的估測與聲源追蹤。
2. 對於追蹤的數個聲源進行聲源分離與切音。
3. 使用校正過之陣列拓樸向量以提高聲源的方位估測與分離的效果。
4. 建立一個完整可擴充的多聲源估測與分離系統架構。

1.3 文獻回顧

實行多麥克風陣列的訊號處理時，麥克風間因為製程而造成的增益與相位差異，以及陣列本身在實際空間中與理論上的差異都會對採集到的訊號特徵造成影響。要取得實際的陣列拓樸向量，勢必要對陣列的各個角度各個頻帶進行繁瑣的估測來求得，為了使估測更有效率，多數人使用的做法是量測少數幾個方向的陣列拓樸向量，再用 least-squares estimation 與 interpolation 的做法求出校正的陣列拓樸向量[1-1] [1-2]。本論文提出一套同樣是進行少數估測並利用幾何資訊與 interpolation 來求得陣列拓樸向量的方法。

利用多麥克風陣列做多聲源方位估測的技術，最早是對雷達相關的研究中被提出，當時是利用能量與相位關係對訊號方位進行估測。後來研究的發展主要有三種類型聲源方位進行估測：一種是利用聲音在空間中傳遞時間差來

做方位估計的 TDE (Time Delay Estimation)[1-3]，如 GCC (Generalized cross-correlation)的方法。第二種是使用波束形成器對指定方向專注(focus)的效果來估測能量以決定聲源方位，如 SBF (Steered beamformer)的方法。第三種則是利用不同訊號間特徵向量的分布差異作互相投影來做方位估計的 Eigenstructure Method[1-4]，如本論文使用的 MUSIC 方法[1-5]。而對於已知的方位估測進行追蹤，多數的研究採用 Particle filter 對個別聲源進行追蹤。利用 Particle filter 估測當前聲源狀態的機率分佈，再對資料關聯做處理使得個別聲源的追蹤能推展到多聲源的情況，以此估測觀測值間的對應關係[1-6]。

有了追蹤的資訊，最後則是利用基本的波束形成器對對各個聲源進行分離。波束形成器主要根據其計算得到權重的方法，技術上大致分成兩類：一種是不依賴輸入訊號的狀況，同樣給予一對任何訊號皆可使用的資料獨立解 (Data Independent)，如 LS (Least Square)解[1-7]；另一種則是依照輸入訊號的狀況，經由限制條件的約束來求得的統計最佳解 (Statistically Optimum) ，如 LCMV(Linearly Constrained Minimum-Variance)解[1-8]。

1.4 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的介紹與機率模型在決策上的應用、論文提出的系統架構與方法的實驗與分析。以下描述各章節的內容：

第二章：背景技術介紹

麥克風陣列相關技術原理的介紹。包含聲源方位偵測方法 MUSIC 的理論與推導，利用機率模型進行決策的方法與基礎理論，波束形成器的基礎理論與應用。

第三章：論文方法

介紹本論文的系統架構與相關演算法。包含陣列拓樸向量的校正步驟，利用 MUSIC 方法的結果找出多聲源方位的演算法，對各聲源估測其頻率響應再利用聲源的頻率特性與空間分佈關係對各聲源進行機率決策而求得追蹤資訊的流程，分離各聲源語音的波束形成器與其為了求得最佳解而調整相關參數。

的應用。

第四章：實驗的結果與分析

對實際環境中錄製的多聲源，驗證論文系統架構中各級的效果，並對校正過之陣列拓樸向量的效果提升進行探討。

第五章：研究成果與未來展望

對論文提出的系統架構與測試結果做評估與總結，並提出此架構未來的發展性與擴充性。

第二章　背景技術介紹

2.1 麥克風陣列訊號處理

傳統的數位訊號處理技術中，不外乎是藉由對訊號的時域及頻域資料狀態做分析與處理，來進行不同應用的相關研究與發展。

多麥克風的陣列訊號處理，是將多個麥克風以特定的幾何形狀排列在一起，並同時接收訊號。因為在空間中有著位置上的差異，每個麥克風對空間中的同一點聲源接收到的訊號，會有時間延遲與能量衰減等等差異。對多麥克風所接收到的多訊號進行處理與分析，就能從中獲得有利的空間資訊，並藉此進行語音品質的純化與提升。

在多麥克風的陣列訊號處理的應用中，研究領域大致上分為兩大類別：

波束形成理論(Beamformer)：

使用多訊號間的空間資訊，對空間中不同方位的訊號進行分離與濾波。這個技術能將多麥克風訊號以不同權重的相位與增益作疊加，對空間中指定方位的訊號產生增強或減弱的效果，一般視為一種空間濾波器(Spatial Filter)，或稱之為波束形成器(Beamformer)。

訊號到達方位估測(Direction of Arrivals Estimation，DOA)：

利用陣列感測器間的差異，對空間中聲源的個數或方位進行估測。技術層面常見的有三種，一種是利用聲音在空間中傳遞時間差來做方位估計的 TDE (Time Delay Estimation)，第二種是利用不同訊號間特徵向量的分布差異作互相投影來做方位估計的 Eigenstructure Method，第三種則是使用波束形成器對指定方向專注(focus)的效果來估測能量以決定聲源方位。

無論屬於哪一類，陣列訊號處理的理論通常都會先有以下兩個基本假設，使得理論推導更為精簡：

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

基於這兩個假設，隨應用有了各種不同的陣列架構。不同的陣列架構對麥克風間訊號的差異也會有不同的規則，看應用的訴求而有不同的功能(不同角度判定範圍、仰角估測等等)。以下列舉兩種基本常用的陣列結構。

均勻線性陣列(Uniform Linear Array，ULA)：

均勻線性陣列是最基本的陣列結構，其架構圖如圖 2.1 所式。

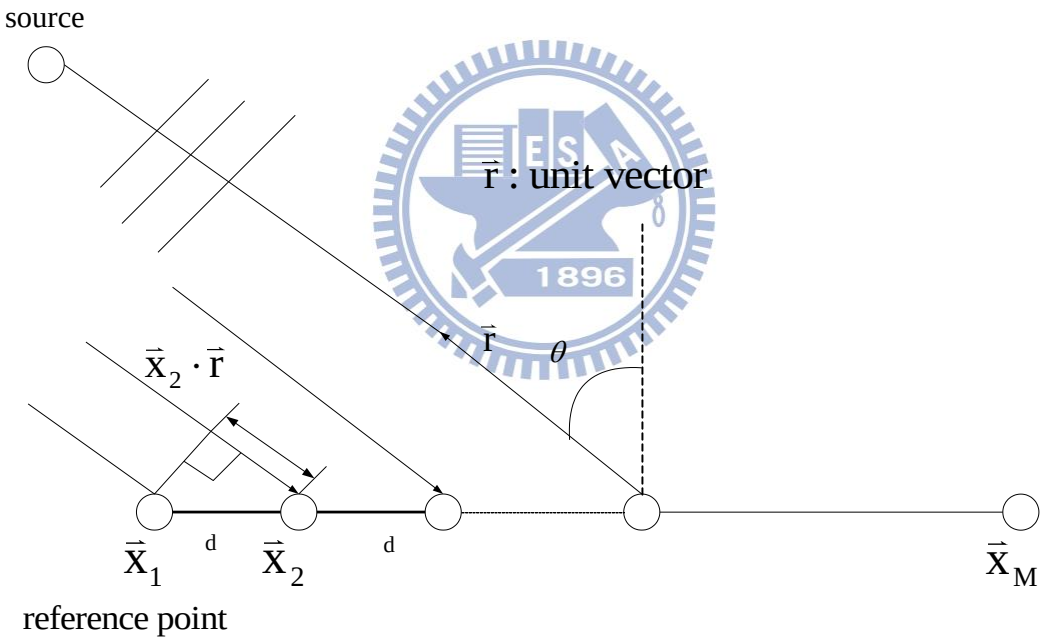


圖 2.1 均勻線性陣列架構圖

d 代表陣列的麥克風間距，M 為陣列的麥克風個數。基於聲源為遠場平面波的假設，將能推導陣列對訊號的向量表示。令 s(t)代表訊號來源，n(t)代表雜訊，則具有 M 個麥克風的均勻線性陣列可以表示為以下形式：

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

$$=\begin{bmatrix} e^{jk_c\vec{x}_1\cdot\vec{r}} \\ \vdots \\ e^{jk_c\vec{x}_M\cdot\vec{r}} \end{bmatrix}s(t)+\begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix}=\mathbf{a}(\vec{r})s(t)+\mathbf{n}(t)$$

$k_c=\frac{w_c}{c}=\frac{2\pi}{\lambda_c}$ ，k_c 稱為 wavenumber，而 λ_c 為波長，c 為波速。

$\mathbf{a}(\vec{r})$被稱為陣列拓樸向量(array manifold vector)，代表了由訊號來源到各麥克風的空間差異造成的時間轉換關係，可以將其表示為：

$$\mathbf{a}^T(\theta)=[1 \quad e^{jk_c d\sin\theta} \quad \cdots \quad e^{jk_c(M-1)d\sin\theta}] \qquad (2.1.2)$$

均勻環型陣列(Uniform Circle Array, UCA)：

均勻環型陣列，或簡稱環狀陣列(Ring Array)是基本的二維陣列結構，如圖 2.2 所式。UCA 擁有兩個維度的空間估測能力，本論文所使用的陣列結構即為 UCA 的結構。

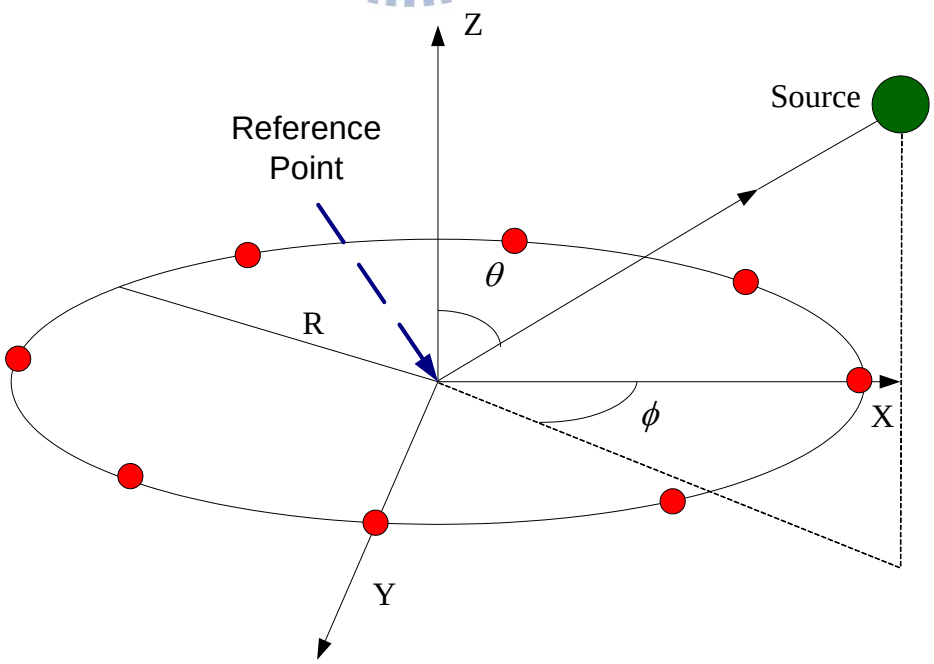


圖 2.2 均勻環型陣列架構圖

若設圖 2.2 的圓心為參考點，可以將 UCA 結構的 array manifold vector 表示為：

$$\mathbf{a}(\phi)=\begin{bmatrix} 1 \\ e^{jk_c\cdot R\sin\theta\cos\phi} \\ e^{jk_c\cdot R\sin\theta\cos(\phi-2\pi/M)} \\ \vdots \\ e^{jk_c\cdot R\sin\theta\cos(\phi-2(M-1)\pi/M)} \end{bmatrix} \qquad (2.1.3)$$

R 代表陣列的環形半徑，M 為陣列的感測器個數。

2.2 訊號到達角度估測(DOA)

如前所述，訊號到達角度的估測是陣列訊號處理技術中一個重要的研究方向。若依照對訊號的處理方式來進行分類，巨觀上可分為下列兩大類。第一類為利用訊號到不同麥克風間的時間延遲，來估測訊號的到達方位，被稱為 TDE (Time Delay Estimation)，常見的方法如 GCC (Generalized Cross-Correlation)，。第二種，也是本論文所使用的方式，被稱為特徵結構法 (Eigenstructure Method)，常見的方法如 MUSIC 與 ESPRIT。

特徵結構法是拿各麥克風擷取到的訊號來計算其資料相關矩陣 (Correlation Matrix)，並將資料相關矩陣進行特徵值分解(Eigenvalue Decomposition)，分解後可得到兩個子空間：訊號子空間(Signal Subspace)與雜訊子空間(Noise Subspace)。由於對應訊號來源方向的指向向量(Steering Vectors)必會與雜訊子空間正交，而雜訊子空間與訊號子空間亦為正交關係，由此可知指向向量其實被包含於訊號子空間中。利用其對應的關係，便可估測出訊號的到達角度。

2.2.1 Multiple Signals Classification Method (MUSIC)

利用特徵結構估測訊號到達角度的方法中，多重訊號分類演算法 MUSIC (Multiple Signal Classification)由於其演算法精簡且估測效果不錯，在特徵結構法中是一種常用的估測方法。

利用 MUSIC 演算法時，有兩個基本假設必須被滿足：

1. 訊號相關矩陣(Source Correlation Matrix)必須是滿秩(Full Rank)且需等於訊號來源的數目 D。
2. Array manifold vector $\mathbf{a}(\theta_i)$，$i=1,\cdots,D$彼此間必須式線性獨立，滿足 Array manifold Array 是滿秩，而秩等也必須等於訊號來源數目 D。

假設訊號來源數目為 D，感測器數目為 M，則將陣列所接收的訊號表示為：

$$\mathbf{x}(t)=\sum_{i=1}^{D}\mathbf{a}(\theta_i)s_i(t)+\mathbf{n}(t)=\mathbf{A}\mathbf{s}(t)+\mathbf{n}(t)$$
$$\mathbf{A}=\left[a(\theta_1)\quad\cdots\quad a(\theta_D)\right],\mathbf{s}^T(t)=\left[s_1(t)\quad\cdots\quad s_D(t)\right] \tag{2.2.1}$$

利用 STFT 將其轉換至頻域：

$$\mathbf{X}(\omega_f,k)=\mathbf{A}(\omega_f)\mathbf{S}(\omega_f,k)+\mathbf{N}(\omega_f,k),\quad f=1\cdots F \tag{2.2.2}$$

k 代表不同的時間間隔，F 代表 FFT size。

假設訊號與雜訊彼此不相關，則資料相關矩陣(Data Correlation Matrix)$\mathbf{R}_{XX}$：

$$\begin{aligned}\mathbf{R}_{XX}(\omega_f,k)&=E[\mathbf{X}(\omega_f,k),\mathbf{X}(\omega_f,k)^H]\\&=\mathbf{A}(\omega_f)\mathbf{R}_{SS}(\omega_f,k)\mathbf{A}(\omega_f)^H+\sigma_N^2(\omega_f)\mathbf{I}\end{aligned} \tag{2.2.3}$$

將資料相關矩陣特徵分解(Eigenvalue Decomposition，EVD)：

$$\mathbf{R}_{XX}(\omega_f)=\sum_{i=1}^{M}\lambda_i(\omega_f)\mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f) \tag{2.2.4}$$

其中，特徵值的大小關係為$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_M$。

雜訊相關矩陣(Noise Correlation Matrix)可以表示為：

$$\sigma_N^2(\omega_f)\mathbf{I}=\sum_{i=1}^{M}\sigma_N^2(\omega_f)\mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f) \tag{2.2.5}$$

則純訊號相關矩陣(Signal-only Correlation Matrix)可以表示為：

$$\begin{aligned}\mathbf{C}_{XX}(\omega_f)&=\mathbf{A}(\omega_f)\mathbf{R}_{SS}(\omega_f,k)\mathbf{A}(\omega_f)^H\\&=\sum_{i=1}^{M}[\lambda_i(\omega_f)-\sigma_N^2(\omega_f)]\mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f)\end{aligned} \tag{2.2.6}$$

由於$\mathbf{C}_{XX}(\omega_f)$的序為 D，可由(2.2.6)中推得一些結果：

- $\lambda_{D+1} = \lambda_{D+2} = \cdots = \lambda_{M-1} = \lambda_M = \sigma_N^2(\omega_f)$
- $R(\mathbf{C}_{XX}) = span\{\mathbf{V}_1(\omega_f), \mathbf{V}_2(\omega_f), ..., \mathbf{V}_D(\omega_f)\}$，也就是 $\mathbf{C}_{XX}$ 的 Range space 是由前 D 個特徵向量所組成。
- $R(\mathbf{A}) = span\{a(\theta_1), \cdots, a(\theta_D)\} = span\{\mathbf{V}_1(\omega_f), \mathbf{V}_2(\omega_f), ..., \mathbf{V}_D(\omega_f)\}$，表示 **A** 的 Range space 也可由前 D 個特徵向量所組成。
- $R^{\perp}(\mathbf{A}) = span\{\mathbf{V}_{D+1}(\omega_f), \mathbf{V}_{D+2}(\omega_f), ..., \mathbf{V}_M(\omega_f)\}$，表示 **A** 的 Null space 可由剩下的 M-D 個特徵向量組成。

經由以上推得的結果，可以定義訊號與雜訊子空間：

1. $\mathbf{R}_S(\omega_f) = span\{\mathbf{V}_1(\omega_f), \mathbf{V}_2(\omega_f), ..., \mathbf{V}_D(\omega_f)\}$，訊號子空間由前 D 個特徵向量所組成。
2. $\mathbf{R}_N(\omega_f) = span\{\mathbf{V}_{D+1}(\omega_f), \mathbf{V}_{D+2}(\omega_f), ..., \mathbf{V}_M(\omega_f)\}$，雜訊子空間由剩下的 M-D 個特徵向量所構成。

利用訊號子空間與雜訊子空間的正交關係，可以推得：

$$\mathbf{V}_j^H(\omega_f)\mathbf{a}(\theta_i) = 0 \quad , \text{i} = 1 \sim \text{D}, \text{j} = \text{D+1} \sim \text{M} \qquad (2.2.7)$$

建立一個投影到雜訊子空間的投影矩陣 $\mathbf{P}_N$：

$$\mathbf{P}_N(\omega_f) = \sum_{i=D+1}^{M} \mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f) \qquad (2.2.8)$$

利用雜訊子空間的投影矩陣 $\mathbf{P}_N$ 與訊號到達角度 $\theta_1, \cdots, \theta_D$，可以得到：

$$\mathbf{P}_N(\omega_f)\mathbf{a}(\theta, \omega_f) = 0 \qquad (2.2.9)$$

$$P_N(\omega_f)\hat{a}(\theta_i, \omega_f) = 0$$

為了能更容易找到訊號到達角度 θ，取(2.2.9)的大小：

$$\left\|\mathbf{P}_N(\omega_f)\mathbf{a}(\theta, \omega_f)\right\|_2^2 = \mathbf{a}^H(\theta, \omega_f)\mathbf{P}_N(\omega_f)\mathbf{a}(\theta, \omega_f) = 0 \qquad (2.2.10)$$

$$S_{MUSIC}(\theta, \omega_f) = \frac{1}{\mathbf{a}^H(\theta, \omega_f)\mathbf{P}_N(\omega_f)\mathbf{a}(\theta, \omega_f)} \qquad (2.2.11)$$

利用(2.2.11)便可得出 MUSIC spectrum，搜尋 MUSIC spectrum 中無限大處，便能找到估測的訊號到達角度。現實的計算結果，並無法看到無限高的

spectrum 存在，但在訊號來源角度處的 spectrum 依然會大於相鄰的其他角度，因此可以藉由搜尋局部最大值的方式找到訊號來源角度。

經由以上過程，便可以利用 MUSIC 演算法來估測出訊號來源角度。另外，仔細觀察其運算(2.2.11)會發現，array manifold vector $\mathbf{a}(\theta_i)$ 的準確度對估算結果有很大的影響，這部分留到 2.5 再做說明。

2.2.2 未知數量寬頻訊號的訊號來源角度估測

上述的演算法可以藉由找到 spectrum 的局部最大值看出聲源的可能數量與方位，不過這是建立在訊號數目已知的情形下。當訊號個數未知時，便需要利用如 AIC(akaike information criterion)[2-1]或 MDL (minimum description length)等方式觀察系統內部參數以估測出訊號個數。

從一資料相關矩陣中決定訊號子空間與雜訊子空間，除了利用估測到的訊號個數來決定之外，也可以把特徵向量對應的特徵值大小當作條件來區分：運用類似主成分分析(Principle Component Analysis，PCA)的方法，特徵值的大小象徵了對應特徵向量的重要性，對 MUSIC 演算法來說，特徵值的大小代表對應特徵向量的能量大小。若某一特徵向量擁有較多能量則可認定此特徵向量代表著訊號，屬於訊號子空間。因此可以藉特徵值大小做為鑑定，對訊號與雜訊子空間進行區分。

使用 MUSIC 估測訊號時，需要統計與分析各個頻帶底下的 MUSIC spectrum 以求得寬頻訊號的訊號來源角度。MUSIC spectrum 是一種倒數的算法，而 spectrum 的大小與分布又會因為所在頻帶不同而有所差異，在統計時往往會造成由某些 spectrum 呈現較大能量的頻帶主導估測結果的現象。因此在統計不同頻帶的 MUSIC spectrum 前，需要對不同頻帶的 MUSIC spectrum 進行均一化(Normalize)以及給上相對權重(Weighting)的動作。除了需斟酌不同頻帶間 spectrum 的大小，能否選擇一段具代表性的頻帶也影響了對寬頻訊號的訊號來源角度估測結果。

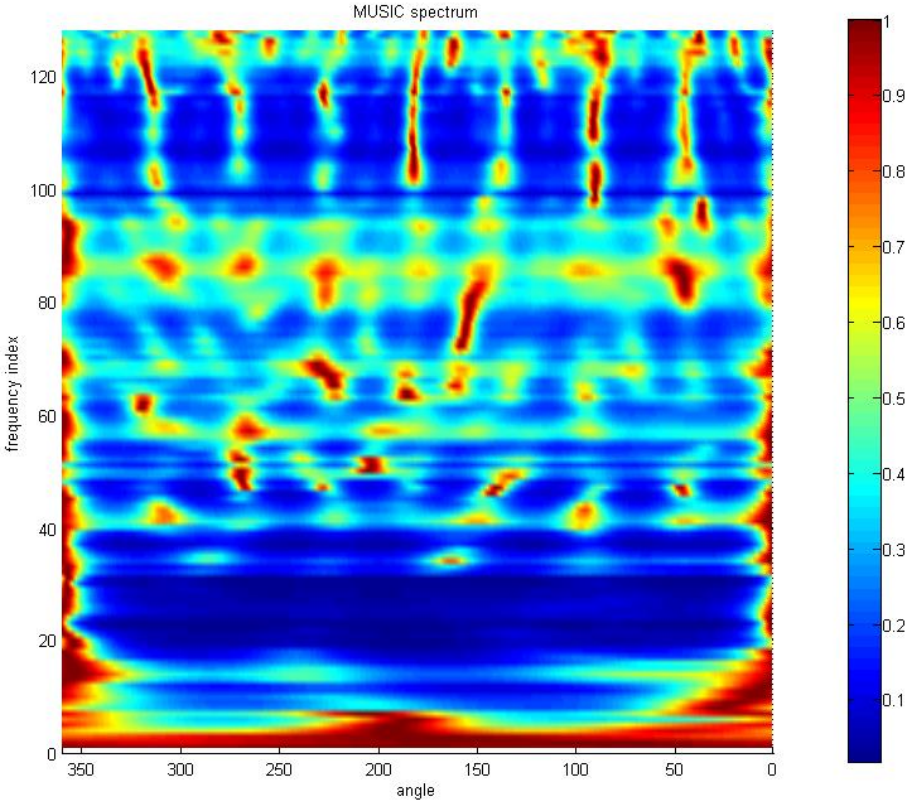


圖 2.3 各頻帶的 MUSIC spectrum

圖 2.3 為當一個聲源位於 0° (360°) 時，各頻帶的 MUSIC spectrum。由圖中可以發現，在某些頻帶中，spectrum 的最大值集中於聲源所在角度，但在某些頻帶中，spectrum 的最大值與聲源所在角度完全無關。在陣列訊號處理中，低頻時會受到 spectrum coherence 的影響，導致聲源方位估計錯誤。在高頻時，又會因 space spectrum aliasing 的關係[2-2]導致在非訊號來源的角度上估測到訊號。所以如果能避開過高或過低的頻帶，選擇中間的頻帶來進行 MUSIC spectrum 的統計可以獲得比較準確的角度估算結果。

本論文選擇將各頻帶的 MUSIC spectrum 以當頻帶最大值作為 normalize factor 來做均一化的動作，使得統計時各頻帶的最大值均等於一，進而可以對不同頻帶施加不同比重來控制估測結果的導向。選擇觀察頻帶 Ω，均一化過的 MUSIC spectrum 可表示為：

$$S_{WB-MUSIC}(\theta)=\sum_{\omega\in\Omega}\frac{S_{MUSIC}(\theta,\omega)}{\arg\max_{\theta} S_{MUSIC}(\theta,\omega)} \tag{2.2.12}$$

2.3 利用機率模型進行聲源與角度決策

貝式定理 (Bayes' Rule) 是目前使用的機率模型決策方法中，極為重要的一項推導。若 D 代表目前已知的觀測而 w 代表欲估測得到的參數，則貝式定理可表示為：

$$P(w|D)=\frac{P(D|w)P(w)}{P(D)} \tag{2.3.1}$$

其中，$P(w)$代表估測參數 w 的事前機率(prior probability)，描述參數 w 的機率分佈，$P(w|D)$代表已知觀察值 D 時參數 w 的事後機率(posterior probability)，而$P(D|w)$被稱為 likelihood function，描述已知參數 w 而會有觀察值 D 的機率關係，。$P(D)$代表觀察值的機率分佈，由於不同的估測值其觀測值皆相同，因此可以視為相同而忽略不算。由於需要確保事後機率的機率定義正確，將$P(D)$視為均一化常數，用以確保所有可能 w 的事後機率總合為一。

貝式定理衍伸出了兩種不同對參數進行估測的方式，分別為 ML (Maximum Likelihood) Estimation 與 MAP(Maximum A Posteriori) Estimation，可表示為：

$$\arg\max_{w} P(D|w) \tag{2.3.2}$$

$$\arg\max_{w} P(w|D)=\arg\max_{w} P(D|w)P(w) \tag{2.3.3}$$

其中，ML 估測的結果只與當下的觀測值有關，可視為利用當前觀測值所估得的最佳化結果。當問題是定義明確的，或 likelihood function 對所有觀測值都成立時，ML 的方法即可得到最佳解。但並非所有的問題都可以舉出所有可能的觀測值，而 likelihood function 也不可能完全正確。在此種情況下，利用 MAP 的估測方法，考慮事前機率的影響，即能修正 likelihood function 的誤差，提升估測結果的正確性。

2.4 利用波束形成器進行聲源切音與分離

波束形成是將多麥克風訊號 $x_i(n)$，經過複數的權重 w_i 疊加而形成一個空間濾波後的結果 $y(n)$。其目的為接收空間中某特定方向來的入射能量，並針對應用排除其它特定方向的干擾。

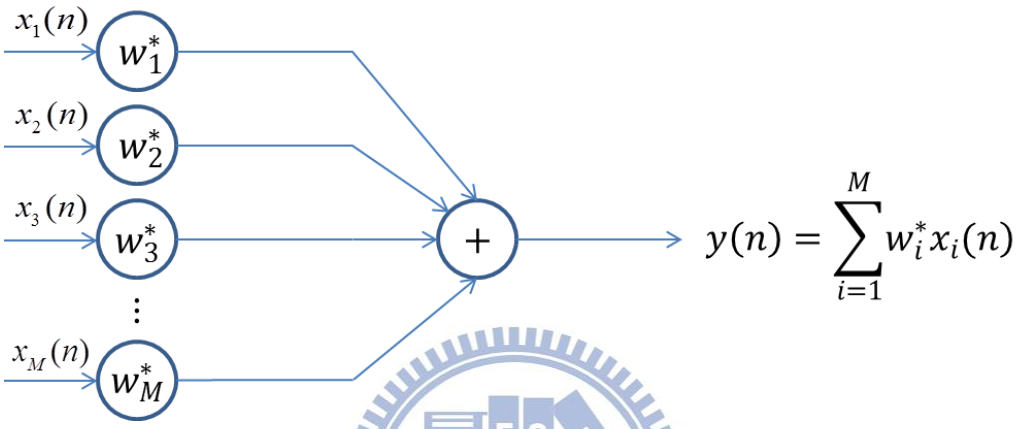


圖 2.4 波束形成示意圖

當波束形成的目標為寬頻訊號時，權重疊加的動作也常被拿到頻域去做，此一動作可以用矩陣表示為：

$$\mathrm{Y}(\omega_f, k) = \mathrm{w}(\omega_f)^{\mathrm{H}} \mathrm{X}(\omega_f, k) \tag{2.4.1}$$

另外，波束形成器根據其計算得到權重的方法，大致分成兩類：

資料獨立解 (Data Independent)：

資料獨立解就如同字面上的意義，並不依賴輸入訊號的狀況，同樣給予一對任何訊號皆可使用的解。

統計最佳解 (Statistically Optimum)：

統計最佳解則是依照輸入訊號的狀況，經由限制條件的約束來求得最佳解。大致上限制的條件皆為最小化周遭的干擾並保留住目標方向的訊號。

2.4.1 Least Square Solution

LS (Least Square)解是資料獨立解法中很基本的做法，其目的即為對下面的假設做最佳化：

$$\mathrm{A}w = \mathrm{g} \tag{2.4.2}$$

其中，$\mathrm{A} = \begin{bmatrix} a(\theta_1)^H \\ a(\theta_2)^H \\ \vdots \\ a(\theta_C)^H \end{bmatrix}$ 為針對 C 個目標方向相對應的 array manifold vector 矩陣，

$\mathrm{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_C \end{bmatrix}$ 則為對此C個目標方向給的限制條件。找出一個最能滿足式子(2.4.2)的 w，意義上便能找出對各角度產生符合限制條件響應的波束形成器。為了解出最佳化的 w，我們將問題轉化為一 Minimization Problem：

$$\hat{w} = \arg\min_{w} \left\| \mathrm{g} - \mathrm{A}w \right\|^2 \tag{2.4.3}$$

將式子(2.4.3)展開如下

$$\left\| \mathrm{g} - w\mathrm{A} \right\|^2 = \mathrm{g}^H\mathrm{g} - w^H\mathrm{A}^H\mathrm{g} - \mathrm{g}^H\mathrm{A}w + w^H\mathrm{A}^H\mathrm{A}w \tag{2.4.4}$$

為了求其最小值，將式子(2.4.4)對 w 做微分並取零：

$$-\mathrm{A}^H\mathrm{g} + (\mathrm{A}^H\mathrm{A})\hat{w} = 0 \tag{2.4.5}$$

最後可以得出最佳解 $\hat{w}$ 來當作波束形成器：

$$\hat{w} = (\mathrm{A}^H\mathrm{A})^{-1}\mathrm{A}^H\mathrm{g} \tag{2.4.6}$$

2.4.2 Linearly Constrained Minimum-Variance Beamformer

LCMV(Linearly Constrained Minimum-Variance)解是統計最佳解法中最常見的一種做法，與 LS 解法不同的是除了達成限制條件外，還希望能同時最小化波束形成器輸出結果中干擾源的功率。

經由(2.4.1)式可以導出輸出功率為：

$$E\left\{\left|Y^2\right|\right\}=E\left\{\left|w^H X\right|^2\right\}=w^H R_{XX} w \tag{2.4.7}$$

其中 R_{XX} 為不包含目標訊號的資料相關矩陣。(亦即只包含干擾源或不重要的訊號)

於是 LCMV 即為找出可以符合下式最佳化的解 w：

$$\hat{w}=\begin{cases} w^H R_{XX} w \rightarrow \text{minimize} \\ w\mathrm{A}=\mathrm{g} \rightarrow \text{constraint} \end{cases} \tag{2.4.8}$$

(為了運算方便，此處 A 的矩陣排列為 $\mathrm{A}=\left[a(\theta_1)\ \ a(\theta_2)\ \ \cdots\ \ a(\theta_C)\right]$)

藉由拉格朗日方法(Lagrange Multiplier)，可以把 Cost function 訂如下式：

$$J=w^H R_{XX} w+\mathrm{Re}\left[\lambda^*\left(w\mathrm{A}-\mathrm{g}\right)\right] \tag{2.4.9}$$

其中 λ 為 Lagrange multiplier。

將 J 微分取零後得到：

$$\nabla J=0=2R_{XX}\hat{w}+\lambda\mathrm{A} \tag{2.4.10}$$

$$\hat{w}=-\frac{\lambda}{2}R_{XX}{}^{-1}A \tag{2.4.11}$$

(2.4.11)聯合(2.4.8)式的 constraint 即可得到 Lagrange multiplier λ：

$$\lambda=\frac{-2g}{\mathrm{A}^H R_{XX}{}^{-1}\mathrm{A}} \tag{2.4.12}$$

最後將之帶入(2.4.11)可以得出最佳解 $\hat{w}$ 當作波束形成器：

$$\hat{w}=\frac{g^* R_{XX}{}^{-1}\mathrm{A}}{\mathrm{A}^H R_{XX}{}^{-1}\mathrm{A}} \tag{2.4.13}$$

2.5 陣列拓樸向量校正

縱觀麥克風陣列訊號處理的各種技術背景,不難發現對於多麥克風的訊號估測,陣列拓樸向量(array manifold vector)其準確與否對各種演算法都影響很大，不論是MUSIC spectrum 的結果(2.2.11)或是波束形成器的最佳解(2.4.6)、(2.4.13)，都會因為 array manifold vector 的失真而產生大量誤差。

陣列拓樸向量的理論算法是單純根據麥克風在空間中的擺放位置，利用幾何關係推算出時間差而得來的。可是在現實中的情形，此一算法未必能確實表現麥克風間訊號濾波的差異。硬體上來說，每個麥克風所能收到聲音的頻率響應本來就不可能完全一致，再加上擺放角度、擺放距離也並不可能完全如同設計上來的工整無誤差。此時接收到的訊號若是使用理論的陣列拓樸向量來做演算法運算，必定會因為相位誤差與增益誤差而大大降低演算的準確度。

對於陣列拓樸向量的校正最常見的做法是利用 LS(Least Square)[2-3]解。由於不可能對於所有角度都做實際估測來校正陣列拓樸向量，所以對陣列拓樸向量的校正往往是利用估計出部分角度，再類推至全角度。用 LS 解校正陣列拓樸向量，首先要對下式做最佳化：

$$\mathrm{A}_{true} = \mathrm{C}\mathrm{A}_{theo} \tag{2.5.1}$$

其中 $\mathrm{A}_{true} = \left[a_{true}(\theta_1) \quad a_{true}(\theta_2) \quad \cdots \quad a_{true}(\theta_C)\right]$ 為實際去估測的 c 個方向的陣列拓樸向量實際值，而 $\mathrm{A}_{theo} = \left[a_{theo}(\theta_1) \quad a_{theo}(\theta_2) \quad \cdots \quad a_{theo}(\theta_C)\right]$ 則為這 c 個方向的陣列拓樸向量理論值。C 為可以對理論陣列拓樸向量進行修正的校正矩陣。

利用 LS 解便可以得到校正矩陣：

$$\mathrm{C} = \mathrm{A}_{true}\mathrm{A}_{theo}^{H}(\mathrm{A}_{theo}\mathrm{A}_{theo}^{H})^{-1} \tag{2.5.2}$$

將 C 代入(2.5.1)全角度，即可得到校正完的陣列拓樸向量。這種估計出部分角度，用 LS 解校正陣列拓樸向量的方法會隨著估測角度數量 c 的增加而增加陣列拓樸向量的正確性。

本論文中提出了另一套同樣估測數量，卻擁有更高正確率的校正方法。這種方法能有效校正並取得真正符合現實硬體架構的陣列拓樸向量，而系統架構底下的各種相關演算法也確實因此而提升了準確率與泛用性。

第三章 系統架構與實作

3.1 系統架構說明

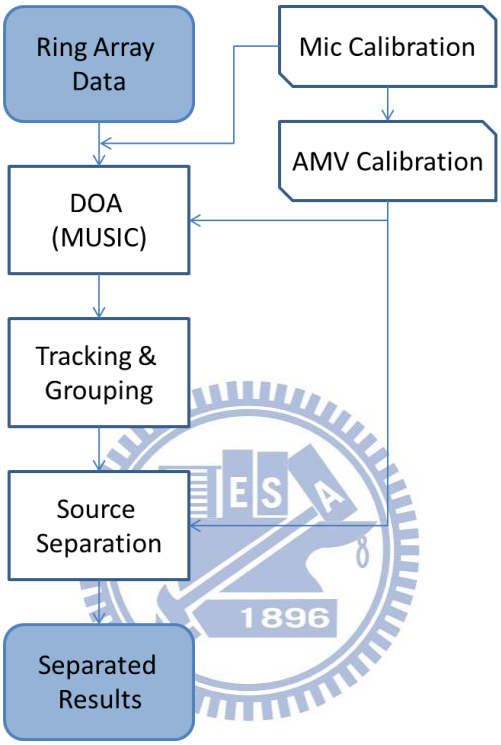


圖 3.1 系統架構

圖 3.1 為本論文提出的系統架構，經由處理環狀陣列取得的麥克風訊號，獲取聲源的空間分佈資訊與頻率特性(MUSIC)，對多聲源的到達方位進行追蹤，並使用機率決策系統來分類聲源與追蹤方位，以及排除錯誤的角度估測與聲源，最後利用波束形成器進行語音分離與切音。

另外，本論文亦提出一校正陣列拓樸向量之方法，藉此來提高方法的穩健性與準確性。預先校正好的陣列拓樸向量只適用於該環狀陣列，能有效提升 DOA 以及 Beamformer 的效果。

3.2 陣列拓樸向量校正

陣列拓樸向量的校正分為兩個部分，第一步驟是估測出兩兩麥克風間的訊號誤差，第二步驟是利用此誤差，把對陣列拓樸向量進行的部分方向估量類推至全方向。

3.2.1 雙麥克風校正

圖 3.2 雙麥克風校正

每一顆麥克風在接收訊號時，由於從硬體來看麥克風振體、電容、供電上都擁有不可避免的差異，會導致在同樣的聲源與距離之下，接收到的訊號在各頻段有不同的差異。此種差異往往會造成估測來源角度或是進行波束形成時的誤差。首先我們必須測量出各個麥克風間的差異。

要測量兩顆麥克風間某一頻率的差異，實做上會先定其中一顆麥克風當作 reference (本論文將 Mic 1 定做 reference)。如圖 3.2，先將兩顆麥克風以極近的距離並排著，指向一公尺外的喇叭。喇叭播放某一定頻的 sine 波，再將兩顆麥克風接受到的訊號取來做運算，估測出其差異。為了精確度，在運算時頻域轉換時使用了 DFT(Discrete Fourier transform)而非一般的 STFT(Short Time Fourier transform)：

$$X_1^{(k)} = \sum_{n=0}^{N-1} x_1^{(n)} \cdot e^{-i2\pi kn/N} \qquad (3.2.1)$$

將其展開可以發現$X_1^{(k)}$的實部與虛部可以藉由$x_1^{(n)}$以下列運算得到，由於$x_1^{(n)}$為一固定頻率的 sine 波，所以算出來的$X_1^{(k)}$將能直接代表在該頻率下的頻域資訊。

$$real\,(X_1^{(k)}) = \sum_{n=0}^{N-1} x_1^{(n)} \cdot \cos(2\pi k\, n/N) \tag{3.2.2}$$

$$imag\,(X_1^{(k)}) = \sum_{n=0}^{N-1} x_1^{(n)} \cdot \sin(2\pi k\, n/N) \tag{3.2.3}$$

有了$X_1^{(k)}$與$X_{\mathrm{m}}^{(k)}$的實部與虛部，便可以算出與 M 顆麥克風之間的雙麥克風(Dual Microphone)增益與相位差異：

$$|DM_m| = \frac{|X_1|}{|X_m|},\ \angle DM_m = \angle X_1 - \angle X_m,\ m = 1,2,...,M \tag{3.2.4}$$

只要掃頻打入各個頻帶的 sine 波，經過運算即能獲取所有需求頻帶的增益與相位差異。

3.2.2 陣列拓樸向量校正

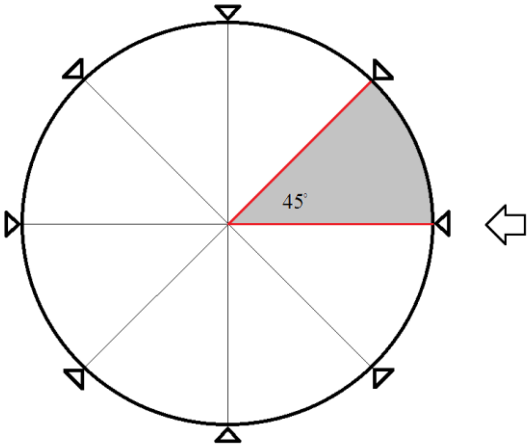


圖 3.3 陣列拓樸向量之部分校正

陣列拓譜向量的量測方法將如同 3.2.1 的做法，如圖 3.3 所示在一固定角度打入掃頻的 sine 波，將八顆麥克風間的頻響差異結果，當作該角度指向的陣列拓譜向量。而實際上，如果希望取得真實的陣列拓譜向量，勢必要從環狀陣列周圍的 360 度都打入掃頻訊號並做計算才能求得。不過這種作法太過

費時費力了，而且當麥克風陣列上只要有一顆麥克風損壞需要更換，整個測量過程就必須重新來過一次。為了簡化量測過程以及方便日後的系統硬體更換或擴充，本論文提出一套進行部分方向的陣列拓樸向量估測以類推全方向的做法。

利用式子(3.2.4)的校正結果，此後將每顆麥克風接收到的訊號 Xr_m 各自乘上相對應的 DM_m，將可以假設此陣列上的每顆麥克風接受訊號的頻響是一致的。此時若將同樣的掃頻 sine 波訊號從圖 3.3 的任一角度打入麥克風陣列，再計算麥克風訊號乘上 DM_m 後的相對關係，便可得出多麥克風間的增益差：

$$RM_m = \frac{|Xr_m| * |DM_m|}{|Xr_1|} \qquad m = 1, 2, M. \tag{3.2.5}$$

而對於相位來說，不論是雙麥克風還是陣列拓樸向量的校正結果，多麥克風之間理論上的相位差異與校正後的相位差幾乎是一樣的。

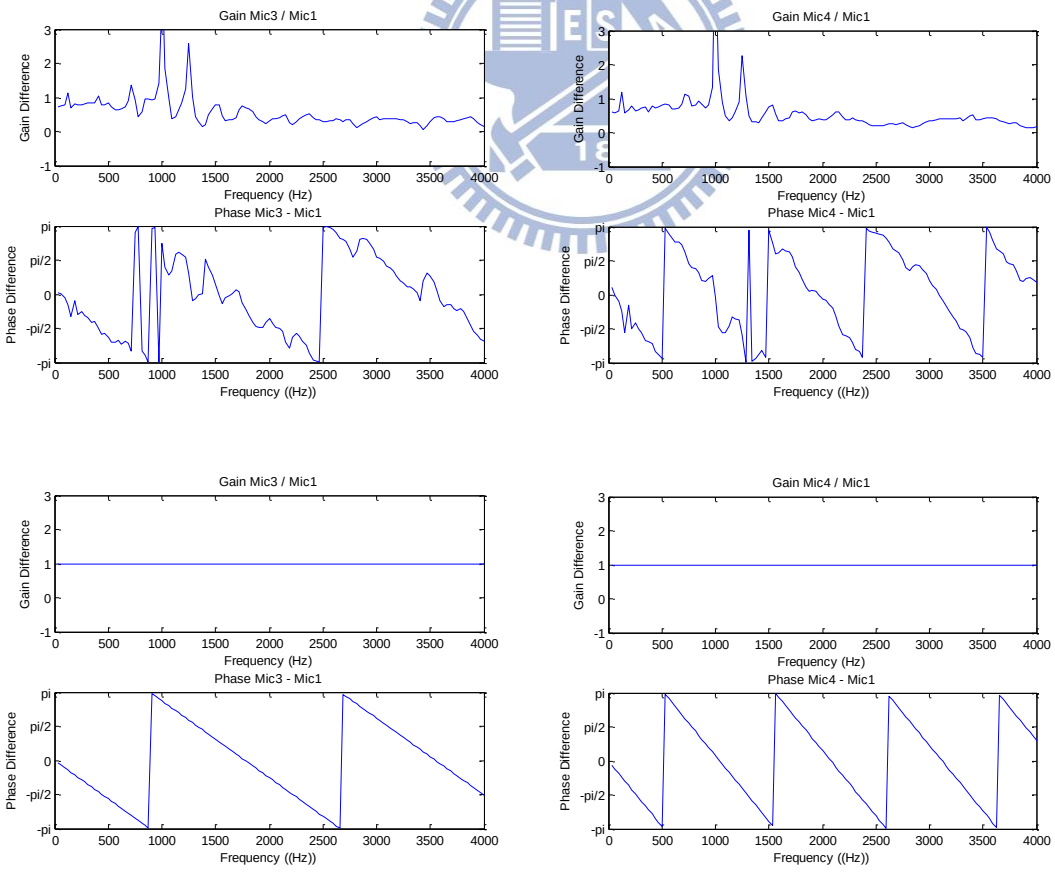


圖 3.4 多麥克風間的增益與相位差

（上列為經由校正得到的增益與相位差資訊，下列則為用理論陣列拓樸向量求得）

如圖 3.4 所示，校正過程由於雜訊或是空間中的干擾，獲取的頻域上增益與相位差資訊並非平順的。對於相位差資訊來說，不論因為干擾造成的不平順曲線，單以 least-squares 計算其斜率，校正的相位資訊斜率與理論值的斜率是幾乎一樣的，如此以來使用校正的相位資訊對於求得更接近真實的陣列拓樸向量是沒有幫助的，更何況校正的相位資訊充滿了未知的干擾，反而降低了其使用上的正確性。反觀校正過後的增益差資訊，在不同頻帶皆有不規則的增減變化，只要挑出 outlier，並經過多次實驗取得平均，並在各頻帶下通過 smoothing 的動作，即可獲得較為客觀可用的增益差資訊。於是在進行陣列拓樸向量的校正時，只會利用到各麥克風間的增益差資訊 RM_m。

首先，測量出圖 3.3 中的 45 度角裡的部分陣列拓樸向量增益差：

$$G(\phi)=\begin{bmatrix} RM_1(\phi) \\ RM_2(\phi) \\ RM_3(\phi) \\ \vdots \\ RM_M(\phi) \end{bmatrix}, \phi = 0^\circ, 5^\circ, \ldots, 40^\circ \tag{3.2.6}$$

在量測此 45 度角的區間時，本論文選用每五度做一次量測(0，5，10，15，…，40)。有了 45 度角裡的部分陣列拓樸向量增益差，即可藉由陣列麥克風的對稱性質，類推出全方向的陣列拓樸向量增益差：

$$G(\phi+45^\circ)=\begin{bmatrix} RM_M(\phi) \\ RM_1(\phi) \\ RM_2(\phi) \\ \vdots \\ RM_{M-1}(\phi) \end{bmatrix}, G(\phi+90^\circ)=\begin{bmatrix} RM_{M-1}(\phi) \\ RM_M(\phi) \\ RM_1(\phi) \\ \vdots \\ RM_{M-2}(\phi) \end{bmatrix}, \ldots, G(\phi+315^\circ)=\begin{bmatrix} RM_2(\phi) \\ RM_3(\phi) \\ RM_4(\phi) \\ \vdots \\ RM_1(\phi) \end{bmatrix} \tag{3.2.7}$$

每 45 度角裡共量測了 9 個點(3.2.6)，以此 9 個點可以推算出環狀陣列 72 個角度的陣列拓樸向量增益差(3.2.7)。最後，再藉由 Spline 內插法擴展此 72 個點成為 360 個點，亦即完整的陣列拓樸向量增益差。

為了結合理論陣列拓樸向量的相位資訊與校正過的增益差資訊，需要將理論的陣列拓樸向量拿來使用。理論的陣列拓樸向量，每顆麥克風的權重皆為 1(增益差為零)：

$$a(\phi)=\begin{bmatrix} a(\phi)_1 \\ a(\phi)_2 \\ \vdots \\ a(\phi)_M \end{bmatrix},\quad |a(\phi)_1|=|a(\phi)_2|=\cdots=|a(\phi)_M|=1 \tag{3.2.8}$$

只要將校正後的增益差資訊加到理論陣列拓樸向量中，便可以求得完整的陣列拓樸向量 $\hat{a}(\phi)$：

$$\hat{a}(\phi)_m = G(\phi)_m a(\phi)_m \ ,\ m=1,2,...,M\ ,\ \phi=0^\circ,1^\circ,\ldots,360^\circ \tag{3.2.9}$$

利用此方法來做測量，就算日後環狀陣列上的任一顆麥克風需要做更換，只要再重新量測一次雙麥克風訊號差值 DM_m（3.2.4），便能繼續套用原先估測好的陣列拓樸向量來使用。

3.3 主要聲源方位估測與特徵值分解的聲源分離

使用 MUSIC 演算法估測多聲源方位時，首先需要藉著 MUSIC spectrum 的大小判斷聲源數。在 MUSIC spectrum 上，只要有 local maximum 就表示該處可能存在聲源。由於在現實中受到背景雜訊、空間響應等等影響，MUSIC spectrum 會有多個 local maximum，這時候就必須設定門檻值來排除由雜訊干擾或空間造成的聲源，篩選出真正的聲源資訊。

門檻值的設定，首先從 MUSIC spectrum 裡找到最大值與最小值：

$$s_{\max}=\arg\max_{\theta}\mathbf{S}_{MUSIC}(\theta) \tag{3.3.1}$$

$$s_{\min}=\arg\min_{\theta}\mathbf{S}_{MUSIC}(\theta) \tag{3.3.2}$$

再藉由 MUSIC spectrum 中的最大最小值決定初始門檻值：

$$GTH=(1-\frac{\beta}{N})*(s_{\max}-s_{\min})+s_{\min} \tag{3.3.3}$$

(3.3.3)中的 ß 是一個介於 0 到 1 之間比值，代表介於最大最小值間的搜尋範圍，而 N 預設為所有搜尋選項個數的遞減迴圈變數。

由 MUSIC spectrum 中找出所有 local maximum：

$$\{s(\theta_1), s(\theta_2), ..., s(\theta_M)\} = \arg\underset{\theta}{Max}_{Local}\{\mathbf{S}_{MUSIC}(\theta)\} \quad (3.3.4)$$

其中，$s(\theta_1) \geq s(\theta_2) \geq \cdots \geq s(\theta_M)$ 由大到小排序。

當 $s(\theta_i)$ 大於 GTH 成立時，更新 $s_{\max} = s(\theta_i)$ 與 $N = i$，並重複(3.3.3)式更新 GTH。簡單來說利用此 GTH 搜尋的做法是將由大到小排序的 local maximum，前一個與後一個做比較，當兩個緊鄰的 local maximum 大小差異過大時，則終止搜尋。

藉著門檻值雖然可以找到具代表性的聲源來源方位，對真正聲源數量的估計還是會有些誤差。例如當不存在聲源時，估測到的聲源數量其時都是空間中的雜訊。又例如訊噪比很低時，背景雜訊也可能會被誤判為主要聲源。為了引導出較正確的聲源估測結果，在找到已知聲源之後，還必須對這些聲源的能量進行估量。

從意義上來看，特徵向量對應了各聲源的來源方位，而特徵值的大小則代表了聲源方位所投影的能量。由於已經將各聲源來向照強弱排序(3.3.4)，於是可以假設較強的聲源來向剛好是對應到擁有較大特徵值的特徵向量，並利用此對應關係估測聲源方向的能量大小。

在計算 MUSIC spectrum 時，已經將訊號相關矩陣進行了特徵分解：

$$\mathbf{R}_{XX}(\omega_f) = \sum_{i=1}^{M} \lambda_i(\omega_f)\mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f) \quad (3.3.5)$$

已知訊號子空間個數(D)，訊號與雜訊子空間分別為：

$$\mathbf{R}_S(\omega_f) = span\{\mathbf{V}_1(\omega_f), \mathbf{V}_1(\omega_f), \ldots, \mathbf{V}_j(\omega_f), \ldots, \mathbf{V}_D(\omega_f)\} \quad (3.3.6)$$

$$\mathbf{R}_N(\omega_f) = span\{\mathbf{V}_{D+1}(\omega_f), \mathbf{V}_{D+2}(\omega_f), ..., \mathbf{V}_M(\omega_f)\} \quad (3.3.7)$$

已知目前聲源到達方位：

$$\bar{\boldsymbol{\theta}} = \{\theta_1, \theta_2, ..., \theta_i, ..., \theta_D\} \quad (3.3.8)$$

要知道 $\mathbf{V}_j(\omega_f)$ 所對應的聲源方位，得先建立包含所有雜訊子空間的投影矩陣 $\mathbf{P}_N$（包含所有除了 $\mathbf{V}_j(\omega_f)$ 之外的特徵向量)：

$$\mathbf{P}_N(\omega_f) = \sum_{i \neq j} \mathbf{V}_i(\omega_f)\mathbf{V}_i^H(\omega_f) \tag{3.3.9}$$

然後利用此 $\mathbf{P}_N$ 與已知的各個角度資訊 θ_j (3.3.8)，計算 MUSIC：

$$\mathbf{S}_{MUSIC}(\theta, \omega_f) = \frac{1}{\mathbf{a}^H(\theta, \omega_f)\mathbf{P}_N(\omega_f)\mathbf{a}(\theta, \omega_f)} \tag{3.3.10}$$

以此來配對各個聲源特徵向量與對應角度：

$$\theta_i(\omega_f) = \arg\max_{\theta \in \bar{\theta}} \mathbf{S}_{MUSIC}(\theta, \omega_f) \tag{3.3.11}$$

同時若配對成功的特徵向量有大於一給定的門檻值，便將此特徵向量 $\mathbf{V}_j(\omega_f)$ 確認為主要聲源，並把其對應的特徵值當作此聲源的頻譜能量：

$$\mathbf{E}_i(\omega_f) = \lambda_j(\omega_f) \tag{3.3.12}$$

3.4 多聲源方位追蹤演算法

單純由 MUSIC Spectrum 篩選觀察到的角度追蹤資訊，除了容易產生錯估之外，也會因為難以分辨是否是聲源而將雜訊或空間干擾當作聲源。藉由機率決策法將可以排除這些錯誤的資訊，並獲得較為完整的聲源追蹤資訊。

首先需要估測暫定聲源的狀態。對聲源方位估測的觀測結果，有著三種可能狀態：第一種狀態，是將聲源判定為尚未被追蹤的新聲源；第二種狀態，是將聲源判定為錯誤偵測的結果；第三種狀態，是將聲源判定為某個追蹤中的聲源。

對於狀態的指定方法，利用 MAP 方法，估測聲源對不同狀態的事後機率值，並取事後機率最大的狀態為此聲源的狀態指定。

經由 DOA 的觀測結果，假設目前有存在 Q 個暫定聲源：

$$\mathbf{O}^{(t)} = \{O_1, O_2, ..., O_q, ..., O_Q\} \tag{3.4.1}$$

對於第 q 個暫定聲源，各狀態的事後機率可表示為：

$$P_{q,j}^{(t)} = P(f(q) = j \mid O_q) \tag{3.4.2}$$

$$P_q^{(t)}(H_0) = P(f(q) = 0 \mid O_q) \tag{3.4.3}$$

$$P_q^{(t)}(H_{-1}) = P(f(q) = -1 \mid O_q) \tag{3.4.4}$$

其中，f(q)為對暫訂聲源 O_q 的狀態指定函數。$f(q) = 0$ 代表 O_q 被判定為新聲源；$f(q) = -1$ 代表 O_q 是偵測錯誤的結果。

利用貝式定理，可以將事後機率展開：

$$P(f(q)|O_q)=\frac{P(O_q|f(q))P(f(q))}{P(O_q)} \tag{3.4.5}$$

假設每一個暫定聲源的機率相同，則 $P(O_q)$ 可以省略，將其視為均一化常數。(3.4.5)便可被加以簡化為 likelihood function 與事前機率的相乘：

$$P(f(q)|O_q)=P(O_q|f(q))P(f(q)) \tag{3.4.6}$$

對(3.4.6)中的 $P(O_q|f(q))$，利用聲源的到達角度，對於各狀態定義：

$$P(O_q|f(q))=\begin{cases}1/2\pi, & f(q)=-1\\ 1/2\pi, & f(q)=0\\ N(O_q;S_j;\sigma^2), & f(q)\geq 1\end{cases} \tag{3.4.7}$$

其中假設了新聲源與錯誤偵測兩者狀態的機率分佈皆為常態分佈。$N(O_q;S_j;\sigma^2)$ 是將追蹤聲源 S_j 作為中心，σ^2 為變異數的高斯分佈，以此作為對已追蹤聲源 S_j 與暫定聲源 O_q 在空間上的相關程度描述。

對(3.4.6)中的 $P(f(q))$，定義各狀態的事前機率：

$$P(f(q))=\begin{cases}(1-P_q)(1-P_s(O_q))P_{false} & f(q)=-1\\ P_qP_s(O_q)P_{new} & f(q)=0\\ P_qP_s(O_q)P(Obs_j^{(t)}|\mathbf{O}^{(t-1)}) & f(q)>=1\end{cases} \tag{3.4.8}$$

(3.4.8)中的 P_{new} 和 P_{false} 分別為新聲源與錯誤偵測狀態的事前機率值。

(3.4.8)中的 P_q 是 O_q 由空間分佈上觀測到的聲源存在機率，用 MUSIC spectrum 定義如下：

$$P_q=\sigma\left(\frac{\mathbf{S}_{MUSIC}(\theta)}{\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\mathbf{S}_{MUSIC}(\theta)}-th\right) \tag{3.4.9}$$

其中的 $\sigma(\bullet)$ 為 logistic sigmoid function，將 spectrum 大小調整到[0,1]的區間中，使其符合機率假設。

(3.4.8)中的 $P_S(O_q)$ 為估測暫定聲源 O_q 是否為聲源的機率，定義其為：

$$P_s(O_q)=P_{lr}P_{hr}P_{hc} \tag{3.4.10}$$

此一機率由三個特徵做評估：

(3.4.10)中的 P_{lr}, P_{hr} 兩個特徵是基於聲源是以人聲為主的假設，主要能量集中於中央頻帶，因此利用中頻與高低頻的能量比，作為偵測是否為正常聲源的特徵；而 P_{hc} 是另一個觀察錯誤偵測的特徵，對每一個聲源方向估測其訊號的共振鋒(Formant)，並在 Formant 附近計算 harmonic 的數量[3-1]，以此作為估測的特徵。結合上述特徵，可由聲源的頻譜響應估測暫定聲源 O_q 是否為聲源的機率。

(3.4.8)中，$P(Obs_j^{(t)} \mid \mathbf{O}^{(t-1)})$ 為利用過去的空間資訊，估測在目前觀察中已追蹤聲源 S_j 出現的事前機率：

$$P(Obs_j^{(t)} \mid \mathbf{O}^{(t-1)}) = P(E_j \mid \mathbf{O}^{(t-1)}) P(A_j^{(t)} \mid \mathbf{O}^{(t-1)}) \tag{3.4.11}$$

(3.4.13)中的 $P(E_j \mid \mathbf{O}^{(t-1)})$ 代表 S_j 在目前觀察中依然存在的機率：

$$P(E_j \mid \mathbf{O}^{(t-1)}) = P_j^{(t-1)} + (1 - P_j^{(t-1)}) \times \frac{P_o P(E_j \mid \mathbf{O}^{(t-2)})}{1 - (1 - P_o) P(E_j \mid \mathbf{O}^{(t-2)})} \tag{3.4.12}$$

其中，P_o 為 S_j 在此次觀測中未出現的機率，$P_j^{(t)}$ 為由此次觀測中 S_j 出現的機率，其定義為：

$$P_j^{(t)} = \sum_q P_{q,j}^{(t)} \tag{3.4.13}$$

(3.4.13)中的 $P(A_j^{(t)} \mid \mathbf{O}^{(t-1)})$ 代表 S_j 在此次觀測中依然在活動的機率，對於已追蹤聲源 S_j 活動中的機率估測，可藉由一階馬可夫程序將其展開：

$$\begin{aligned} P(A_j^{(t)} \mid \mathbf{O}^{(t-1)}) &= P(A_j^{(t)} \mid A_j^{(t-1)}) P(A_j^{(t-1)} \mid \mathbf{O}^{(t-1)}) \\ &\quad + P(A_j^{(t)} \mid \neg A_j^{(t-1)}) \times [1 - P(A_j^{(t-1)} \mid \mathbf{O}^{(t-1)})] \end{aligned} \tag{3.4.14}$$

其中，若假設 S_j 依然在動作與停止動作的機率相同的話，定義 $P(A_j^{(t)} \mid \mathbf{O}^{(t)})$ 為：

$$P(A_j^{(t)} \mid \mathbf{O}^{(t)}) = \frac{1}{1 + \dfrac{[1 - P(A_j^{(t)} \mid \mathbf{O}^{(t-1)})][1 - P(A_j^{(t)} \mid O^{(t)})]}{P(A_j^{(t)} \mid \mathbf{O}^{(t-1)}) P(A_j^{(t)} \mid O^{(t)})}} \tag{3.4.15}$$

由於 MUSIC spectrum 可視為當前觀察中，各角度在空間上的分佈特性，故可藉此定義在當前觀察中，以追蹤聲源 S_j 依然在活動的機率：

$$P(A_j^{(t)} \mid O^{(t)}) = \sigma(\bar{\mathbf{S}}_{MUSIC}(\theta_j) - th) \tag{3.4.16}$$

經過以上的計算後，可估測暫定聲源對各狀態的事後機率。此時只需要選

擇事後機率最大的狀態，便可判定此暫定聲源的狀態。

決定暫定聲源的狀態後，對於追蹤中的聲源，還必須根據觀測值來更新聲源的方位。但由於系統在低 SNR 時估測的準確度會下降，難免有誤估，為了穩定聲源方位的變化，利用聲源方位的移動方向變化來控制追蹤中聲源的更新比例。一般的聲源方位變化並不會有瞬間的大改變，而是漸進式地變化。在一個維度的角度搜尋範圍中觀察聲源變化,其移動的改變可能只有正或負。在此利用一階馬可夫程序，估測對不同移動方向的機率，並將機率值作為聲源位置的更新比例。

根據過去的移動方向，估測目前移動方向的機率，：

$$P(D^{(t)} \mid \mathbf{e}^{1:t}) = \alpha P(e^{(t)} \mid D^{(t)}) \sum_{D^{(t)} \in \{+,-\}} P(D^{(t)} \mid D^{(t-1)}) P(D^{(t)} \mid \mathbf{e}^{1:t-1}) \tag{3.4.17}$$

對於已追蹤聲源，利用目前觀測更新其聲源方位：

$$W_j^{(t)} = P(D^{(t)} \mid \mathbf{e}^{1:t}) \tag{3.4.18}$$

$$S_j^{(t)} = S_j^{(t-1)} + W_j^{(t)} \times (O_q^{(t)} - S_j^{(t-1)}) \tag{3.4.19}$$

藉由以上機率運算，可以決定對於每個暫訂聲源的當前狀態。當聲源狀態為新聲源時，若其狀態的機率值超過設定的門檻值時，便認定其為新聲源，並開始進行追蹤：

$$P_q^{(t)}(H_0) > TH_{new} \tag{3.4.20}$$

對於追蹤中的聲源，則是藉由每一次的觀察結果來計算其消失的機率，來確認追蹤中聲源的目前狀況：

$$P_{del}^{(t)}(S_j) = \alpha(1 - P_j^{(t)}) + (1-\alpha) P_{del}^{(t-1)}(S_j) \tag{3.4.21}$$

α 為考慮前一次機率的平滑參數。當機率超過門檻值，便認定聲源已消失：

$$P_{del}^{(t)}(S_j) > TH_{del} \tag{3.4.22}$$

3.5 利用波束形成器之多聲源切音與分離

經過前面的運算，我們將可以得到如圖 3.5 中的角度追蹤資訊。圖中的原始訊號為環狀陣列同時接收四個方向此起彼落的人聲，經過運算所得到的結果。

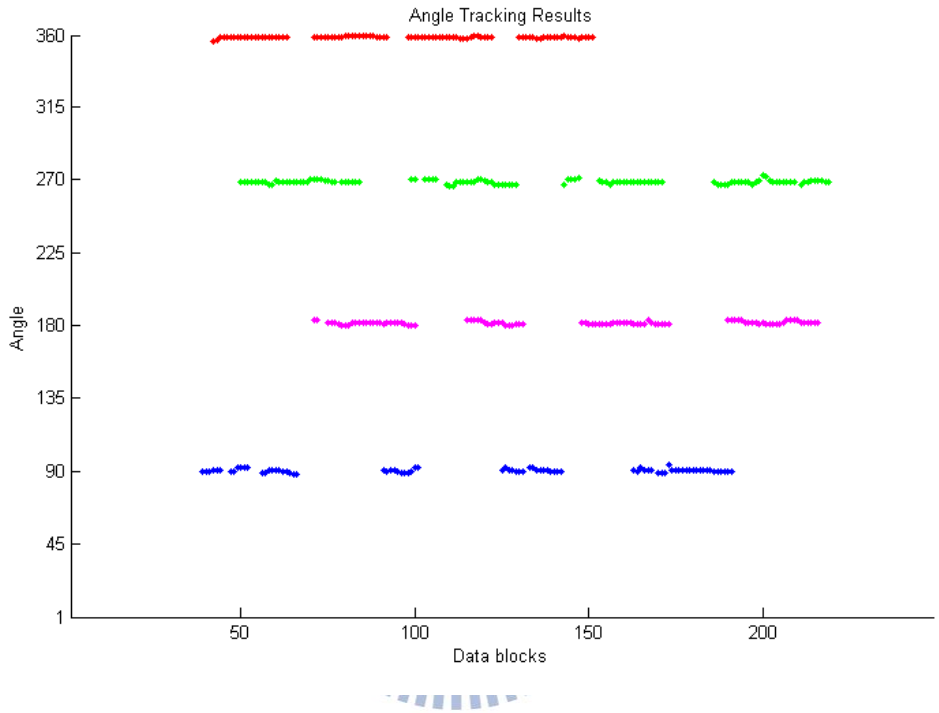


圖 3.5 四聲源角度追蹤資訊

此運算結果為每一個 Data block 增加一筆資訊。如前所述，本論文使用兩種波束形成器來進行聲源分離，接著就這兩種實作上的方法進行探討。

3.5.1 Least Square Solution

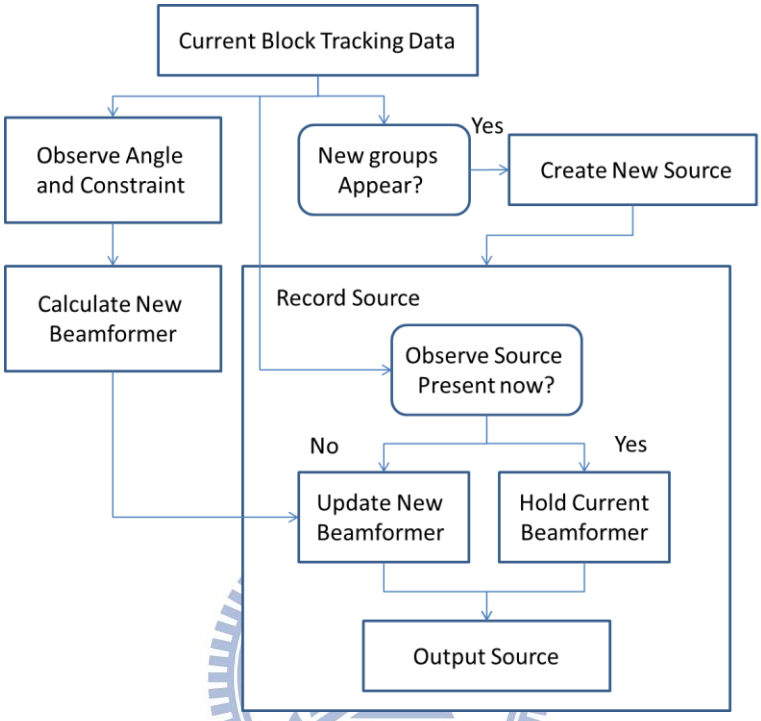


圖 3.6 LS 多聲源切音與分離架構

圖 3.6 為利用 LS 解來進行多聲源切音與分離的系統架構。首先對於當下 Data Block 的追蹤資訊會先觀察是否有新的追蹤組，有的話將會增加一筆輸出的音軌數。對於每一筆輸出的音軌，從追蹤組一出現聲源存在時就會開始每個 Block 都輸出一波束形成的結果。

為了能對變動的聲源做出最好的聲源分離效果，波束形成器的各項參數(角度、限制)會基於對角度追蹤資訊的觀察而做出相對應的調整而更新。

在 2.4 章裡可以知道，LS 解最後可以得出一最佳解 $\hat{w}$ 來當作波束形成器：

$$\hat{w} = (\mathrm{A}^H \mathrm{A})^{-1} \mathrm{A}^H \mathrm{g} \tag{3.6.1}$$

其中，$\mathrm{A} = \left[a(\theta_1)^H \quad a(\theta_2)^H \quad \cdots \quad a(\theta_C)^H \right]^T$ 為針對 C 個目標方向相對應的陣列拓樸向量矩陣，這些方向都是希望能施加限制(Constraint)的角度，而 $\mathrm{g} = \left[g_1 \quad g_2 \quad \cdots \quad g_C \right]^T$ 則代表了對此 C 個目標方向，希望施加的限制條件。

為了得到多聲源切音與分離的效果，最直覺的做法就是對於想要獲得的聲源角度，施加限制為 1，而此刻同時存在的其他聲源角度施加限制為 0。這麼一來所算出來的波束形成器將擁有去除其他聲源，並保留目標聲源語音的效用。

在實際運算(3.6.1)中的反矩陣時，會碰到因為 $A^H A$ 為奇異矩陣而造成無法計算的情況。此時需要為此矩陣添加一 Diagonal Loading：

$$\hat{w} = (A^H A + \lambda I)^{-1} A^H g \qquad (3.6.2)$$

其中 λ 為此 Loading 的大小。

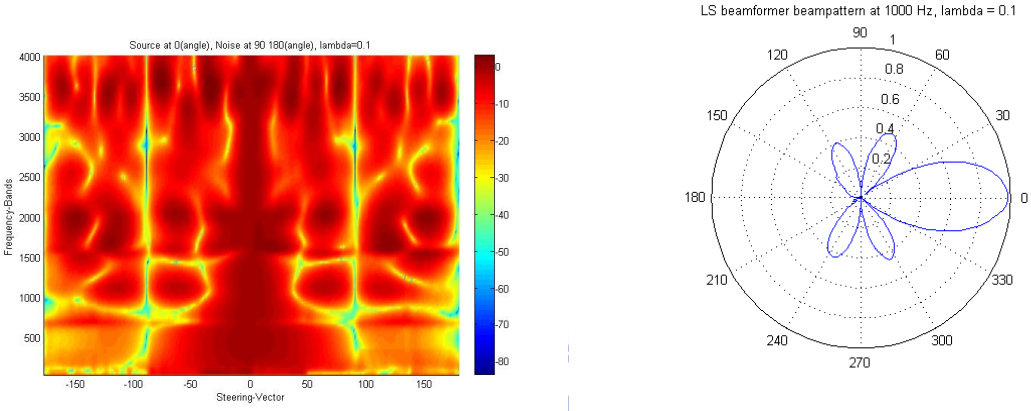


圖 3.7 LS 解得出的波束形成器，λ=0.1

（聲源於 0 度，干擾於 90 度及 180 度，右圖為 1000Hz 的情況）

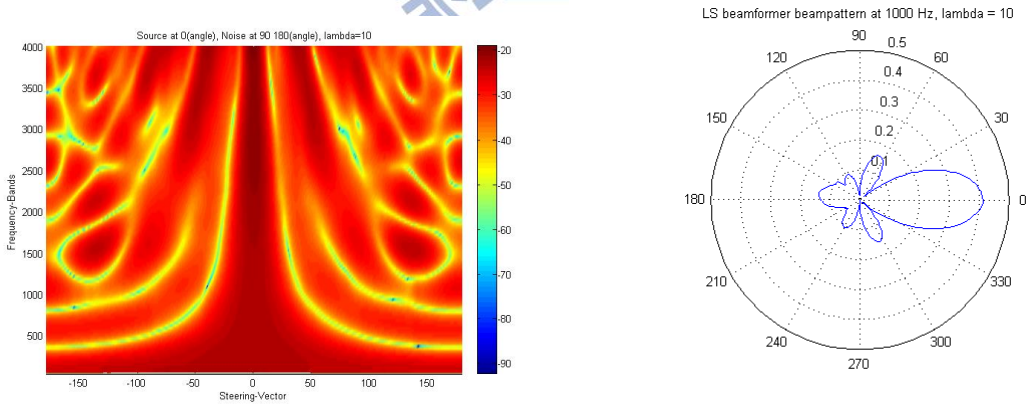


圖 3.8 LS 解得出的波束形成器，λ=10

（聲源於 0 度，干擾於 90 度及 180 度，右圖為 1000Hz 的情況）

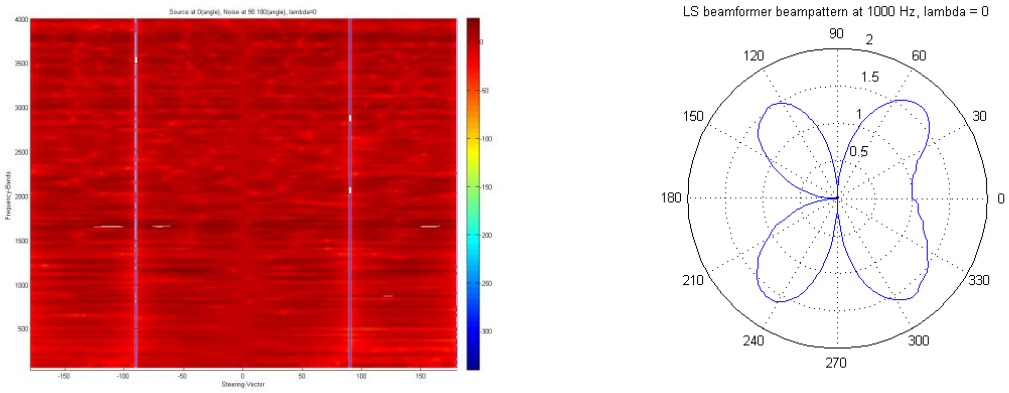


圖 3.9　LS 解得出的波束形成器，λ=0

（聲源於 0 度，干擾於 90 度及 180 度，右圖為 1000Hz 的情況）

如圖 3.7，3.8，3.9， Loading 的大小λ會對波束形成器的形狀產生影響：λ理論上越小會使得波束形成器越接近限制好的條件，當λ太大時，如圖 3.8，會使得限制條件對波束形成器結果的約束力下降，造成限制不明顯的現象。而當λ太小或是不使用 Diagonal Loading 時，雖然解出來的波束形成器完美合乎限制條件，卻因為自由度不夠反而放大了非限制條件下方向的權重。經過實驗結果，當λ介於 0.01 到 0.1 之間時能擁有最佳的效果。

除了 Loading 的大小λ之外，由角度以及限制條件個數來決定的陣列拓樸向量矩陣 A，對於算出來的波束形成器效果也有很大影響。角度的資訊會因為聲源方位變動而跟著改變，每一個 Block 變動的角度都會藉由一個設定好的 Forgetting Factor 來更新陣列拓樸向量矩陣中對應聲源做限制的角度，進而對波束形成器產生改變。

至於限制條件的個數，基本上就是當下的目標聲源以及干擾聲源的個數。而每個干擾聲源的存在與否是藉由觀察過去多個 block 的角度追蹤資訊來取得。在處理每一個 block 時，會記錄每一個建立的聲源在過去 observe_blocks 個 blocks 裡出現的次數(count)，而在為目標聲源計算波束形成器時，會藉由判斷 count 是否有達到預先給定之數值，以此來衡量聲源數目。圖 3.10 為聲源存在性觀察演算法，圖 3.11 則為其效果。

```
if group_angle ~= 0
    count = count + 1;
    if count > observe_blocks
        count = observe_blocks;
    end
elseif group_angle ~= 0
    count = count - 1;
    if count < 0
        count = 0;
    end
end
```

圖 3.10 聲源存在性觀察演算法

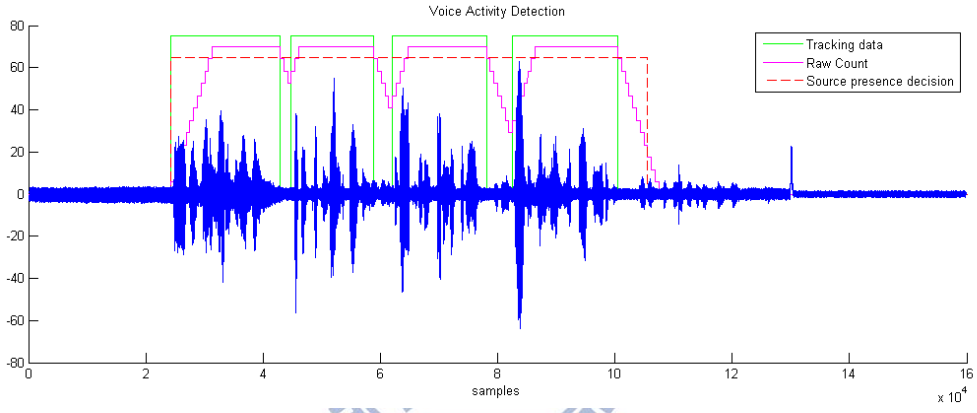


圖 3.11 聲源存在性觀察演算法效果

(藍色部分為分離的訊號，紫色此訊號的存在 count，紅色虛線為判斷存在結果)

這樣的做法是因為，是因為雖然每個 block 都會將波束形成器更新，不過為了確保音軌中的聲源語音連續性，在此聲源存在時(角度追蹤有值)，不會變動使用中的波束形成器；而當此聲源不存在時，才會以預設的 Forgetting Factor 來漸進式更新波束形成器。因為聲源存在時不能變動波束形成器，所以必須要使用上述類似 hangover 的作法來謹慎處理空間中可能還存在，卻在變更波束形成器的當下不存在的一些聲源，藉此提升分離的效果。

LS 解為一資料獨立解型的波束形成器，雖然不依賴輸入訊號的狀況就能計算出最佳解，此解對於排除它向的干擾與降噪卻有其極限。

3.5.2 Linearly Constrained Minimum-Variance Beamformer

在 2.4 章裡可以知道，LCMV 解最後可以得出最佳解 $\hat{w}$ 當作波束形成器：

$$\hat{w} = \frac{g^* R_{XX}^{-1} \mathrm{A}}{\mathrm{A}^H R_{XX}^{-1} \mathrm{A}} \tag{3.6.3}$$

與 LS 解有所不同的是，LCMV 解除了需要角度與限制條件等參數之外，還需要來源訊號的資訊才能進行最佳化估算。由於 LCMV 的算法為對輸出功率做最小化：

$$E\left\{\left|Y^2\right|\right\} = E\left\{\left|w^H X\right|^2\right\} = w^H R_{XX} w \tag{3.6.4}$$

所以運算式中的 R_{XX} 必需為不包含目標訊號的資料相關矩陣。(亦即只包含干擾源或不重要的訊號)

當使用情境為聲源不會移動的會議紀錄時，可以以圖 3.4 的追蹤資訊做事後資料相關矩陣收集，用所有目標聲源不存在的 Block 裡的 R_{XX}，計算此目標聲源要用的波束形成器：

$$R_{XX\text{source1}} = \frac{1}{N} \sum R_{XX}\,(\text{source1 not present}) \tag{3.6.5}$$

圖 3.12 為利用 LCMV 解得出的波束形成器，由於 LCMV 的算法為在已知資料相關矩陣時對輸出功率做最小化，所以能得出更適合輸入訊號的約束效果，可是這也只限於特定使用情境之下。若遇到聲源會移動，或是聲源分離必須即時完成的情況，資料相關矩陣的收集就會遇到諸多限制而使效能降低。

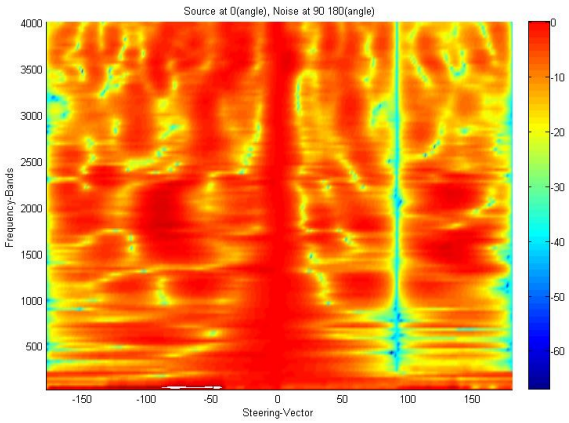


圖 3.12 LCMV 解得出的波束形成器(聲源於 0 度，干擾於 90 度及 180 度)

第四章 實驗結果與分析

本章節將討論本論文提出的系統架構實驗的結果。實驗大致上分為兩個部分。第一部分為探討校正過後的陣列拓樸向量在實際使用時會有什麼效果與助益，第二部分則探討聲音分離與切音結果的效果性與不同情況底下的穩健性。

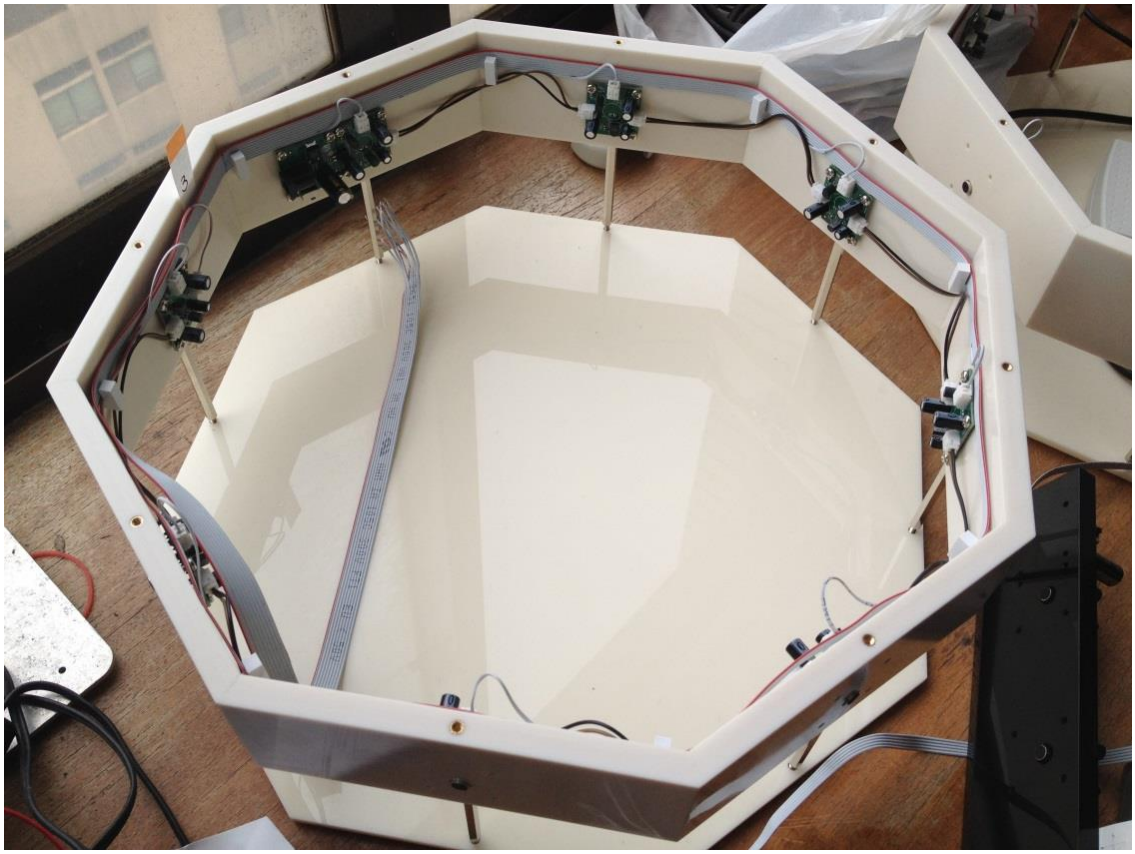


圖 4.1 環形麥克風陣列平台

本論文使用的環狀麥克風陣列平台由半徑 18 cm 的八顆類比麥克風環形陣列構成，圖 4.1 為其現實中的照片。硬體架構為多顆麥克風由 Mic1, Mic2,…,Mic8 依序排列在一八邊型麥克風架上。為了方便，這邊將 Mic1 所指的方向設為 0°，Mic2 的方向為 45°，其餘以此類推。圖 4.2 為其從上而下的平面結構圖。麥克風接收到的類比訊號經由 NI USB-6210 介面轉換為數位訊號並傳送到電腦中，錄製成 16 位元，取樣率為 8000Hz 的音檔。

在處理訊號時，本論文選擇以一個音框(frame) 256 個 sample 當作 FFT size。由於人聲特徵音素的變化率約為 10 ms，以此長度做基準，選擇在移動 FFT 音框(frame)時以 80 個 sample 為單位以應對音素的變化。又為了使聲源方位偵測的結果較為平滑，對每個音框(frame)中的資料，選擇 16 個音框(frame)為一個 block 取平均，移動 block 時會以 8 個 frame 為單位做更新。詳細的錄音與訊號處理參數如表 4.1 所列。

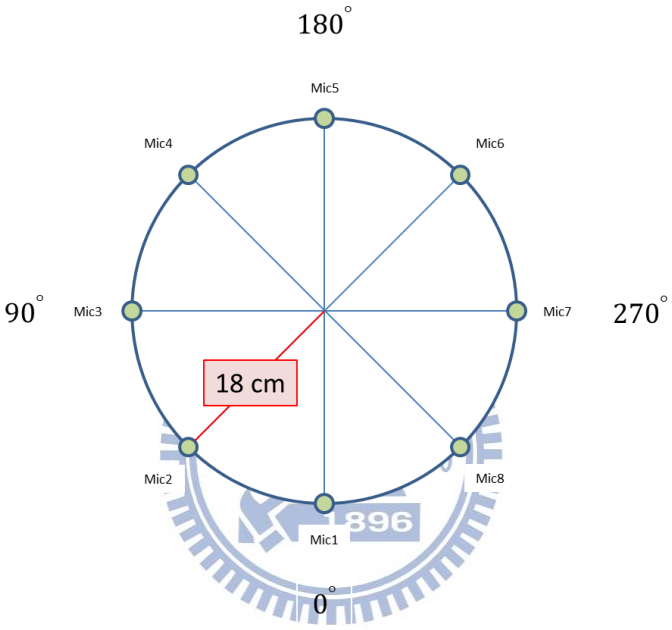


圖 4.2 環形麥克風陣列平台的平面圖

陣列架構	環形麥克風陣列
陣列半徑	18 cm
麥克風個數	8
Sampling rate	8 kHz
FFT size	256 samples
shift size	80 samples
Block size	16 frames
Block overlap size	8 frames

表 4.1 環形麥克風陣列平台錄音與訊號處理參數

4.1 陣列拓樸向量校正結果

使用 3.2 章的方法，利用環形麥克風陣列平台錄音實際在無響室中進行測試，並利用提出方法可以求得一校正過之陣列拓樸向量。若是將陣列拓樸向量當作波束形成器的話，可以畫出類似 beam pattern 的 AMV pattern：

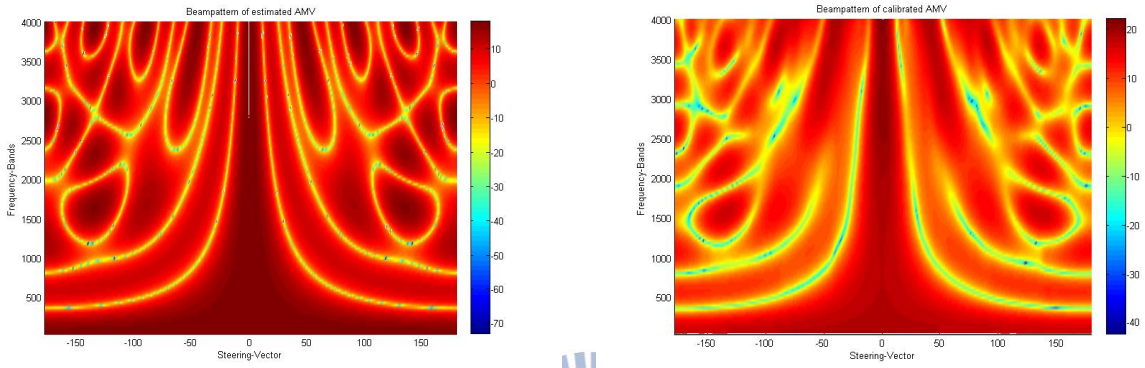


圖 4.3 陣列拓樸向量對 0 度角之 beampattern
(左為理論的陣列拓樸向量，右為校正的陣列拓樸向量)

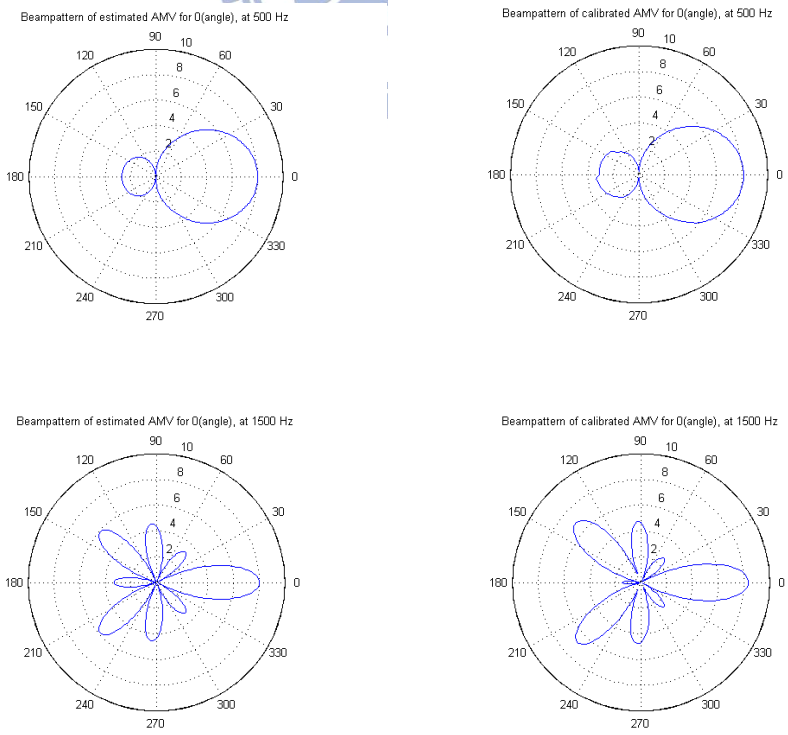


圖 4.4 陣列拓樸向量對 0 度角在 500 Hz(上)與 1500Hz(下)時之 Beampattern

由圖 4.3, 4.4 可以看出理論的陣列拓樸向量與校正的陣列拓樸向量在不同頻率與角度上權重的分配有著差異，而這些因應實際硬體與空間響應而生的差異，會使得理論的陣列拓樸向量與校正的陣列拓樸向量在實際使用時有效果上的差異。以下就以實驗進行驗證：

第一部分的實驗裡，透過對 MUSIC Spectrum 的觀察結果來比較校正陣列拓樸向量與否的差異。這邊觀察的 MUSIC spectrum 為對各個頻帶，將其最大值作為 normalize factor 來做均一化的動作，使得統計時各頻帶的最大值均等於一。選擇觀察頻帶為 Ω，均一化過的 MUSIC spectrum 可表示為：

$$S_{WB-MUSIC}(\theta)=\sum_{\omega\in\Omega}\frac{S_{MUSIC}(\theta,\omega)}{\arg\max_{\theta} S_{MUSIC}(\theta,\omega)} \tag{4.1.1}$$

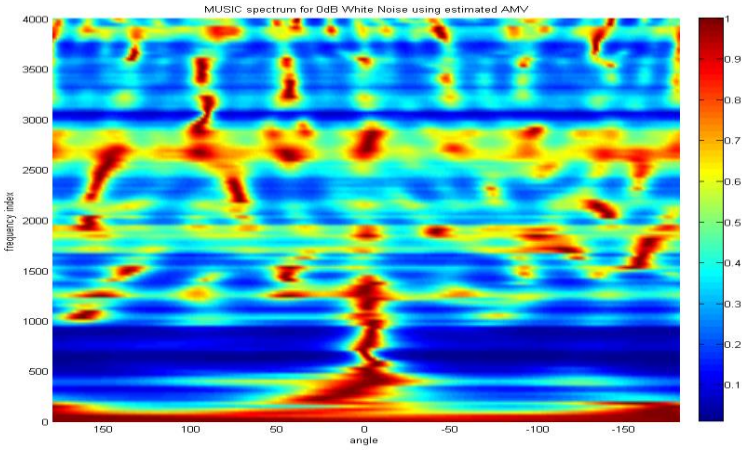
測試一：單聲源的 MUSIC Spectrum 觀察

首先進行單聲源的 MUSIC Spectrum 觀察。實驗環境為利用在無響室用環狀陣列錄製的 source 訊號與 noise 訊號。Source 訊號為利用人工頭對準第一顆麥克風(方向 0 度)定點錄製一段 20 秒的 White Noise (White Noise 1)，Noise 訊號則是利用人工頭對準八顆麥克風各錄一段 20 秒的 White Noise (White Noise 2)，並將此八組訊號合成起來，形成一個從四面八方來的 Diffused White Noise。之後再用撰寫的 SNR_Mixer 混出給定 SNR 的 Noisy Source 訊號，在此用來測試的混音訊號為 SNR 0 dB。擺置設定如表 4.2。

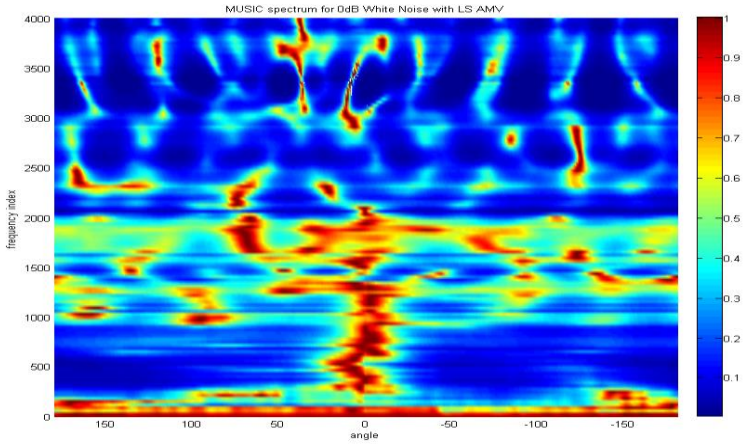
將此 Noisy Source 訊號通過本論文提出的系統，可以求得圖 4.5 之 MUSIC spectrum (多個 frame 的平均結果)。橫軸為角度(-180~179 度)，縱軸為頻率(上到下為 0~4000 Hz，Frequency Bin 為 128 個)。

	聲源方位	聲源類型
Source	0°	White Noise 1
Disturbance	Diffused	White Noise 2

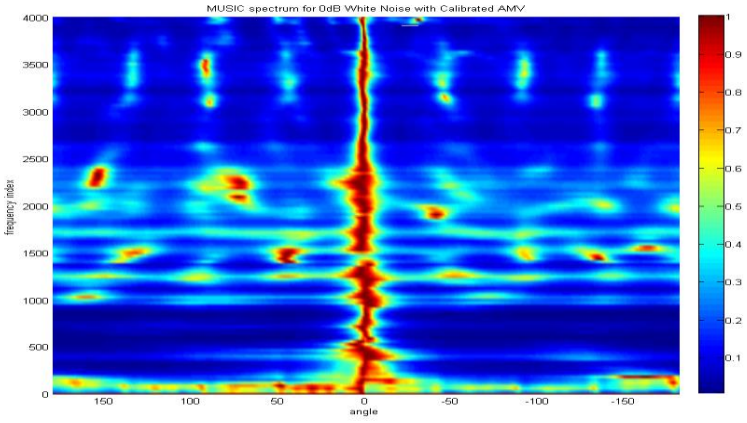
表 4.2 單聲源 MUSIC Spectrum 擺置設定



(a)使用理論的陣列拓樸向量的結果



(b)利用九個實測角度做 LS 解校正的陣列拓樸向量的結果



(c)使用本論文提出方法利用九個實測角度做校正的陣列拓樸向量的結果

圖 4.5 單聲源 MUSIC Spectrum 分布情形

利用陣列訊號處理估算聲源到達角度的各種演算法，如本論文使用的 MUSIC Spectrum，在計算 Wide-Band 的寬頻能量分部時，低頻的部分很容易受到 spectrum coherence 的影響，導致聲源方位資訊模糊；而高頻的部分，又會因 space spectrum aliasing 的關係，導致能量分散在非訊號來源的角度上，產生角度估計錯誤。

如圖 4.5(a)所呈現，這些不利的效應在使用理論陣列拓樸向量時對估測準確度有很大的影響,而當採用利用九個實測角度做 LS 解校正的陣列拓樸向量後，如圖 4.5(b)，不利效應對估測準確性的影響有稍微降低，可是在高低頻的方向特徵還是顯得模糊。而在利用本論文提出方法用九個實測角度校正陣列拓樸向量後，如圖 4.5(c)，不利的效應對估測準確性的影響明顯大幅降低，甚至在極低頻與極高頻時，MUSIC Spectrum 都還能保有相當明確的角度方向特徵。

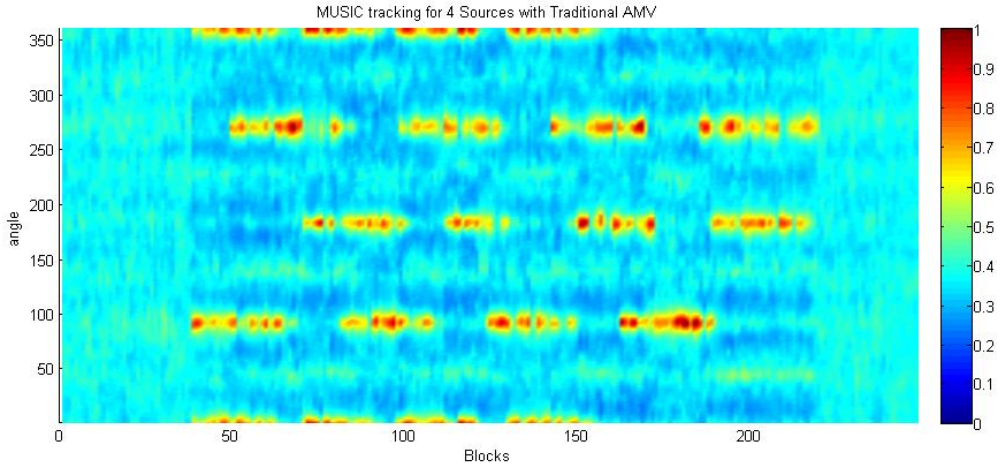
測試二：多聲源的 MUSIC Tracking 結果

再來觀察校正對於多聲源的 MUSIC Tracking 結果影響。在無響室中用環狀陣列錄製 4 組不同的 source 訊號，各自對準 0°，90°，180°，270°的方向，並將此 4 組訊號一需求合成,形成一個多聲源此起彼落的會議情境,如表 4.3。

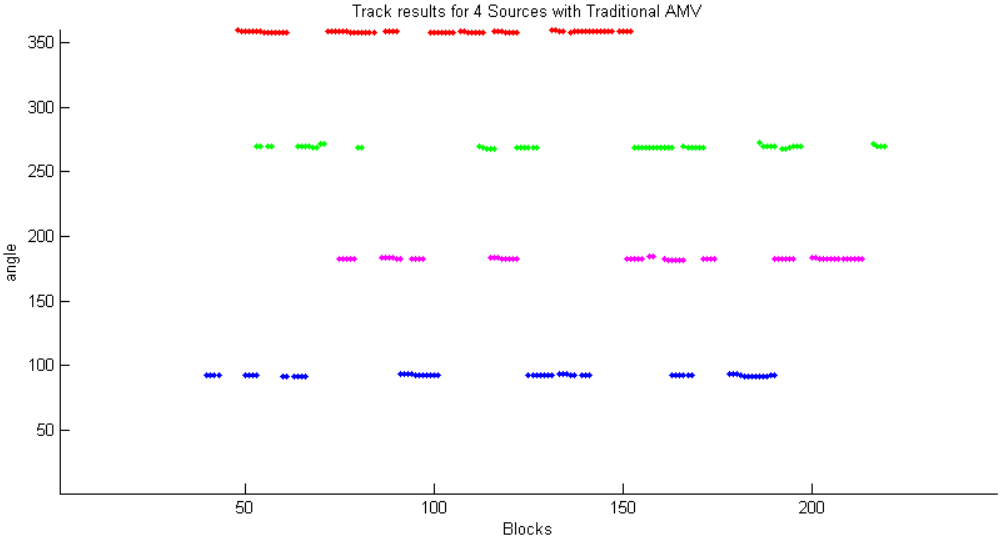
將此人聲混雜的訊號通過本論文提出的系統，可以求得圖 4.6 (a)、4.7 (a)之 MUSIC Spectrum Tracking (每一 Block 都有各角度的寬頻平均能量)，縱軸為角度(0~360 度)，橫軸為處理到的 Data Blocks。利用此 MUSIC 的能量分布紀錄，再透過特徵向量與機率決策的篩選，可以得到圖 4.6 (b)、4.7 (b)之 Tracking Results。

	聲源方位	聲源類型
Source 1	0°	Female Voice 1
Source 2	90°	Female Voice 2
Source 3	180°	Male Voice 1
Source 4	270°	Male Voice 2

表 4.3 多聲源 MUSIC Tracking 擺置設定

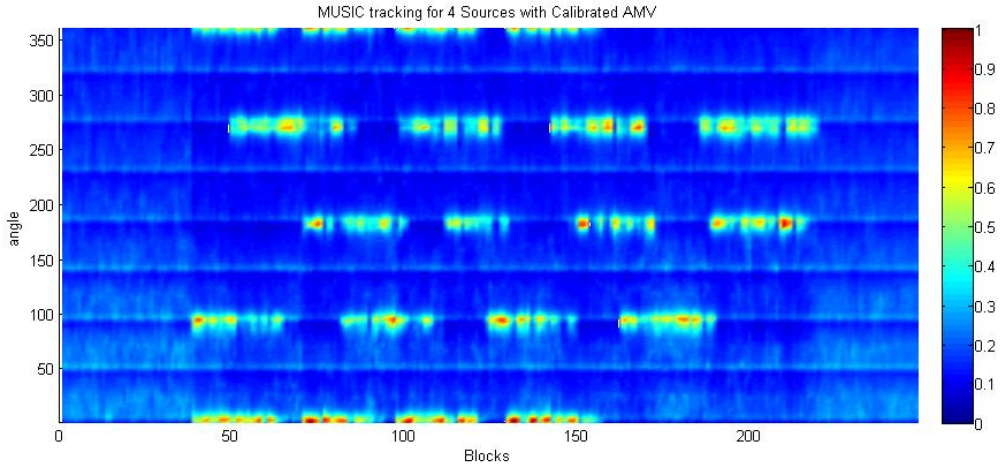


(a) Spectrum 結果

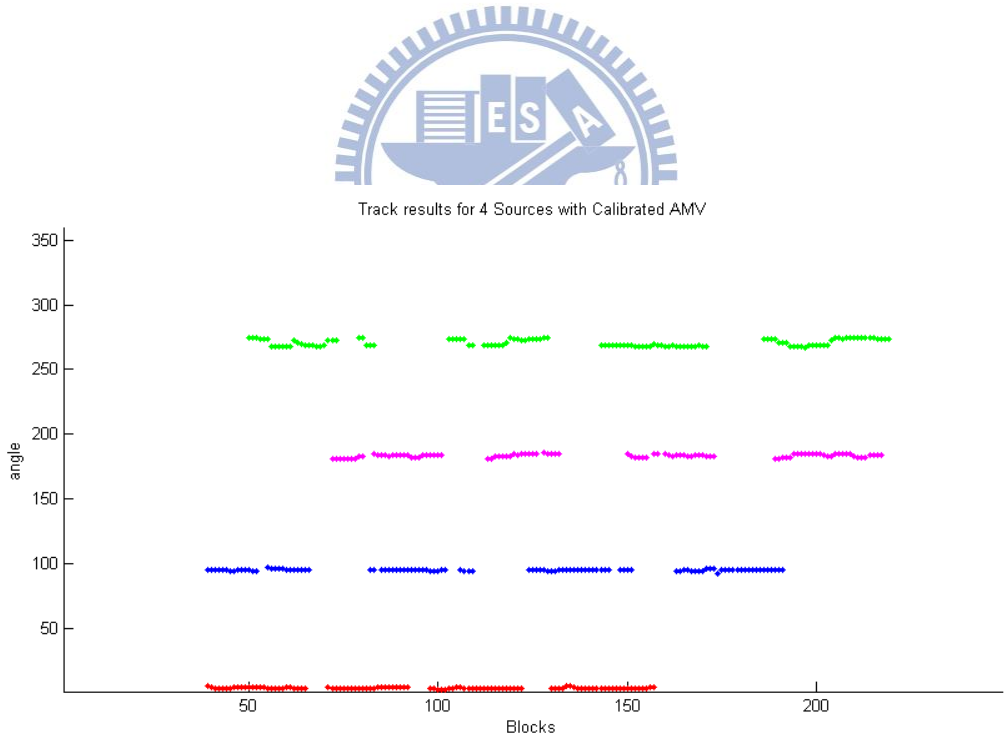


(b) Tracking 結果

圖 4.6 多聲源 MUSIC Tracking 使用理論陣列拓樸向量



(a) Spectrum 結果



(b)Tracking 結果

圖 4.7 多聲源 MUSIC Tracking 使用校正過之陣列拓樸向量

直接以肉眼觀察看來，圖 4.6 (a)與 4.7 (a)各自得到的 MUSIC Spectrum Tracking 透露的角度訊息並不會有太大差異，可是由於使用理論陣列拓樸向量算出來的 MUSIC Spectrum 在寬頻計算時低頻的 spectrum coherence 與高頻的 space spectrum aliasing 影響，能量分布無法集中於聲源的角度上，而分散在各個角度。雖然肉眼可以明顯判斷，不過當使用機率決策來獲得追蹤資訊時，如圖 4.6 (b)所見，即會造成多處聲源方向誤判或是缺漏的情形。反觀使用校正過之陣列拓樸向量的圖 4.7 (a)，能量分布的極大值都比較明顯集中在單一幾個方向上，在機率決策來獲得追蹤資訊時，如圖 4.7 (b)所見，能得到較為完整連續的角度追蹤值。

測試三：噪音情況的多聲源追蹤準確率與穩健度效能評估

圖 4.8 聲源原始訊號顯示於 Adobe Audition

為了評估追蹤值的準確性，我們做了對各個聲源估算其準確率 (Accuracy Rate)與誤報率(False Alarm Rate)的檢定。首先，如圖 4.8，先用人耳與 Adobe Audition 資料顯示的結果，挑選出人耳聽來有包含連續人聲的 sample 點，再換算出這些 Sample 點所在的 Data block，將這些 blocks 視為 True blocks，而其餘沒有人聲的 blocks 皆為 False blocks。

而對於用系統方法追蹤出來結果判定為有聲源的 blocks，將其視為 Positive blocks，而沒有追蹤到聲源的 Blocks 則視為 Negative blocks。比較人工篩選的結果與系統估算的結果，可以將所有的 Blocks 分為四類：人工篩選通過，系統也偵測到的稱為 True Positive blocks；人工篩選通過，系統卻沒有偵測到的稱為 False Negative Blocks; 人工篩選不通過，系統也偵測為沒有的稱為 True Negative blocks；人工篩選不通過，系統卻偵測到的稱為 False Positive。用這四類 blocks 的數量於是可以定義準確率(Accuracy Rate)為所有判定中，成功判定為是或是成功判定為否的機率：

$$\mathrm{AR} = \frac{\text{True Positive blocks} + \text{True Negative blocks}}{\text{Total blocks}} \tag{4.1.2}$$

其值越接近 100%代表系統估測的結果與人耳聽覺分辨的結果越相近。

而 False Alarm Rate 則定義為所有沒有人聲的 blocks 裡，被誤判為有人聲的機率：

$$\mathrm{FR} = \frac{\text{False Positive blocks}}{\text{True Negative blocks} + \text{False Positive blocks}} \tag{4.1.3}$$

另外，在測試準確率時，除了用乾淨的多人聲音檔來做檢定之外，還利用了兩種不同頻率分佈的噪音，測試本論文架構對抗噪音的穩健度。噪音的選擇中，使用 Babble noise 與 Car interior noise。其中，Babble noise 為一種非穩態(non-stationary)的噪音，而 Car interior noise 則可視為一種穩態(stationary)的噪音。

最後，為了比較出校正過的陣列拓樸向量在寬頻估測上的準確性與效果增益，所有的估測皆做了針對較窄的頻帶(500 ~ 1500 Hz)與較寬的頻帶(250 ~ 3750 Hz)的版本。

	聲源方位	聲源類型
Source 1	0°	Female Voice 1
Source 2	90°	Female Voice 2

表 4.4 多聲源追蹤準確率與穩健度效能評估擺置設定

Babble noise, narrow frequency-band selected tracking (500Hz~1500Hz):

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	94.78 %	86.75 %	0 %	1.2 %
-1 dB	95.98 %	88.35 %	0 %	0.8 %
-2 dB	92.77 %	84.74 %	3.21 %	5.22 %
-5 dB	91.57 %	72.69 %	2.01 %	14.46 %
-8 dB	83.94 %	59.84 %	0 %	15.66 %
-12 dB	69.48 %	52.61 %	2.41 %	15.66 %

(a) Using traditional AMV

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	96.79 %	89.96 %	3.61 %	0.8 %
-1 dB	96.39 %	87.15 %	2.01 %	5.22 %
-2 dB	95.58 %	81.12 %	2.01 %	16.47 %
-5 dB	93.57 %	75.1 %	0 %	25.66 %
-8 dB	87.55 %	70.68 %	0 %	31.73 %
-12 dB	66.67 %	63.45 %	0 %	24.5 %

(b) Using calibrated AMV

表 4.5 Babble noise 情況下較窄頻帶追蹤的 Accuracy Rate 與 False Alarm Rate

Babble noise, wide frequency-band selected tracking (250Hz~3750Hz):

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	90.76 %	81.93 %	0 %	0.78 %
-1 dB	89.96 %	83.94 %	0 %	0.78 %
-2 dB	89.16 %	86.35 %	1.61 %	0.8 %
-5 dB	79.52 %	65.46 %	10.04 %	16.87 %
-8 dB	61.85 %	47.39 %	13.35 %	20.08 %
-12 dB	65.46 %	48.59 %	14.06 %	21.69 %

(a) Using traditional AMV

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	97.59 %	95.58 %	1.2 %	1.2 %
-1 dB	91.16 %	92.77 %	8.03 %	4.82 %
-2 dB	92.37 %	89.96 %	6.43 %	8.43 %
-5 dB	91.16 %	81.53 %	7.23 %	17.27 %
-8 dB	92.77 %	69.88 %	1.61 %	25.7 %
-12 dB	77.11 %	64.26 %	0.8 %	30.92 %

(b) Using calibrated AMV

表 4.6 Babble noise 情況下較寬頻帶追蹤的 Accuracy Rate 與 False Alarm Rate

Car noise, narrow frequency-band selected tracking (500Hz~1500Hz):

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	94.78 %	86.75 %	0 %	0.8 %
-1 dB	95.98 %	88.35 %	0 %	0.8 %
-2 dB	95.58 %	88.76 %	0.8 %	0.8 %
-5 dB	93.98 %	87.55 %	1.61 %	0.4 %
-8 dB	86.35 %	82.73 %	7.63 %	0 %
-12 dB	83.13 %	65.06 %	0 %	4.02 %

(a) Using traditional AMV

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	96.79 %	88.76 %	0 %	0 %
-1 dB	94.78 %	86.75 %	2.41 %	4.02 %
-2 dB	92.37 %	87.95 %	2.41 %	4.02 %
-5 dB	93.98 %	85.94 %	2.81 %	2.41 %
-8 dB	91.16 %	84.74 %	4.82 %	1.61 %
-12 dB	90.36 %	73.49 %	0.8 %	0 %

(b) Using calibrated AMV

表 4.7 Car noise 情況下較窄頻帶追蹤的 Accuracy Rate 與 False Alarm Rate

Car noise, wide frequency-band selected tracking (250Hz~3750Hz):

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	91.57 %	81.93 %	0 %	0.78 %
-1 dB	91.97 %	82.33 %	0 %	0.78 %
-2 dB	91.16 %	82.33 %	0 %	0.78 %
-5 dB	90.36 %	81.12 %	1.61 %	1.61 %
-8 dB	77.11 %	71.49 %	12.73 %	0 %
-12 dB	61.45 %	57.03 %	14.46 %	0 %

(a) Using traditional AMV

SNR (dB)	Accuracy rate		False Alarm Rate	
	Source 1	Source 2	Source 1	Source 2
0 dB	97.59 %	95.58 %	1.2 %	1.2 %
-1 dB	96.79 %	91.97 %	5.45 %	4.82 %
-2 dB	94.38 %	86.75 %	3.21 %	8.43 %
-5 dB	93.98 %	90.76 %	10.91 %	12.4 %
-8 dB	92.77 %	91.16 %	14.55 %	10.85 %
-12 dB	93.98 %	83.53 %	1.61 %	2.81 %

(b) Using calibrated AMV

表 4.8 Car noise 情況下較寬頻帶追蹤的 Accuracy Rate 與 False Alarm Rate

噪音情況下的兩個聲源追蹤結果總結：

在檢定追蹤效果的時候，準確率當然是越高越好，不過若能提高準確率，在此系統架構的應用底下是可以容忍增加些許誤報率的。在使用 non-stationary 的 babble noise 當作干擾，做較窄頻帶的估測時(表 4.5)，會發現陣列拓樸向量的校正與否對於追蹤結果並沒有太多提升，甚至在一些情形下由於 babble noise 本身擁有的人聲屬性，校正過的陣列拓樸向量反而會在 MUSIC Spectrum 上放大某些特定方位的能量造成誤報，進而使追蹤準確度下降。整體來說單純就較窄頻帶的估測，兩種陣列拓樸向量的效能十分相近。

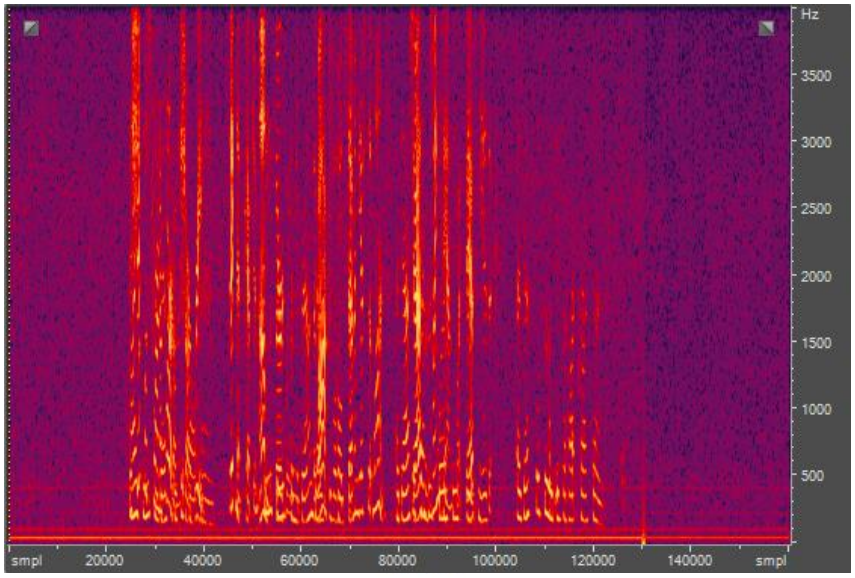


圖 4.9　乾淨聲源訊號的頻譜圖

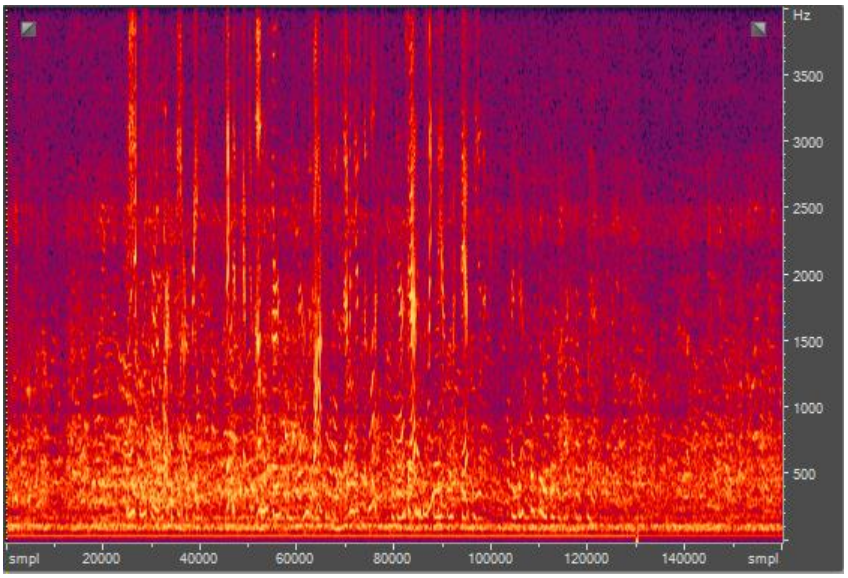


圖 4.10　聲源訊號加入 Babble noise 的頻譜圖

從圖 4.9 與 4.10 中我們可以發現，當聲源訊號中加入 babble noise 時，觀察人聲的主力頻帶(500Hz~1500Hz)幾乎被 babble noise 給蒙蔽了，只使用這一頻帶的資訊來估測聲源效果十分有限，勢必要利用目標聲源在高頻帶與低頻帶保有的明顯特徵來做估測。

而當使用較寬頻帶的 MUSIC Spectrum 來做追蹤估測時(表 4.6)，由於目標聲源擁有比 babble noise 較完整明顯的頻帶資訊，在做較寬頻帶估測時使用校正過的陣列拓樸向量能在訊噪比極低的情況下，依然保有一定程度的準確率。反觀使用理論陣列拓樸向量估測較寬頻帶資訊時，由於高低頻的特徵混亂，反而降低了其準確率。

校正過的陣列拓樸向量提升準確率的效果在使用 stationary 的 car noise 當作干擾時又更為明顯了。這兩種不同的 noise 干擾，測試的效果都是一樣的：在較窄頻帶估測時，校正過的陣列拓樸向量能維持與理論陣列拓樸向量差不多甚至更好的效果；而在做較寬頻帶估測時，校正過的陣列拓樸向量更能在訊噪比極不樂觀的條件下，如表 4.8(b)，擁有相當程度的準確率。於是可以發現陣列拓樸向量的校正對於提升追蹤結果的準確率與穩健度都是十分有效的。

4.2 多聲源語音分離效果

接著將使用已經求得的追蹤資訊，經過波束形成器來輸出結果，並測試本論文系統架構聲源分離的效果。實驗語料與配置如表 4.9 與圖 4.11 所示。

	聲源方位	聲源類型
Source 1	0°	Female Voice 1
Source 2	90°	Female Voice 2

Source 1	0°	Female Voice 1
Source 2	90°	Female Voice 2
Source 3	180°	Male Voice 1
Source 4	270°	Male Voice 2

表 4.9 多聲源追蹤準確率與穩健度效能評估擺置設定(二聲源與四聲源)

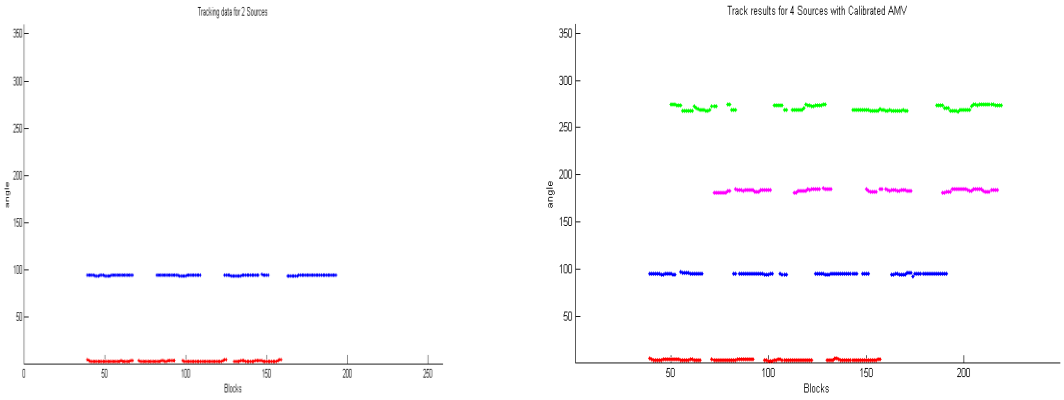


圖 4.11 多聲源語料人聲分佈圖(二聲源與四聲源)

由於系統輸入為一混音的八聲道訊號，經過波束形成器輸出為兩個聲源的訊號，這樣的情形要計算各自聲源結果的 SNR 時，將需要把其它聲源當作干擾(Noise)，而使用傳統的 SNR 算法並無法有效取得這些人聲能量當作干擾(Noise)。

為了評估波束形成器對此各聲源的訊噪比 SNR (Signal-to-noise-ratio)，在此定義一特別適用於此多聲源情況下的 SNR 算法。

首先，先算出乾淨的聲源單聲道訊號：

$$\begin{cases} S_{input1} = S_1 w_{sum} \\ N_{input1} = \sum_{m=2}^{M} S_m w_{sum} \end{cases} \quad (4.1.4)$$

其中 $S_1, S_2, ..., S_m$ 為用麥克風陣列錄得的兩組八聲道訊號(各自只包含單一聲源)。而 w_{sum} 則為一單純的 Sum beamformer：

$$w_{sum} = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]^T \quad (4.1.5)$$

以此可以得到乾淨的(只包含單一聲源)單聲道目標訊號 S_{input1}，以及被當成噪音的其他聲源訊號合 N_{input1}。

再來，將 $S_1, S_2, ..., S_m$ 的八個聲道各自進行混音，得到一組包含多個聲源的八聲道訊號 S_{mix}。以這組混聲源訊號輸入系統架構，經過角度追蹤與波束形成的運算，可以得出對該觀察聲源分離訊號的最佳波束形成解 $\hat{w}_1$。

此時將第一聲源方向的波束形成解 $\hat{w}_1$，對單聲源的訊號 $S_1, S_2, ...,$ 做波束形成，將可以得到強化的第一聲源訊號 $S_{output1}$，以及被削減的其餘聲源訊號合 $N_{output1}$：

$$\begin{cases} S_{output1} = S_1 w_1 \\ N_{output1} = \sum_{m=2}^{M} S_m w_1 \end{cases} \quad (4.1.6)$$

再來，取(4.1.4)中 S_{input1} 的人聲部分看作輸入訊號，N_{input1} 的人聲部分看作輸入噪音，可以算出聲源一的輸入訊噪比：

$$SNR_{input1} = 10\log\frac{[rms(S_{input1}(n))]^2}{[rms(N_{input1}(n))]^2} \quad (4.1.7)$$

同樣，將(4.1.6)中 $S_{output1}$ 的人聲部分看作輸出訊號，$N_{output1}$ 的人聲部分看作輸出噪音，可以算出聲源一的輸出訊噪比：

$$SNR_{output1} = 10\log\frac{[rms(S_{output1}(n))]^2}{[rms(N_{output1}(n))]^2} \quad (4.1.8)$$

最後由 $SNR_{output1}$ 與 SNR_{input1} 的差便可以求得波束形成解 $\hat{w}_1$ 對於第一聲源所產生的訊噪比增益 SNRI(Signal-to-noise-ratio improvement)：

$$SNRI_1 = SNR_{output1} - SNR_{input1} \quad (4.1.9)$$

用類似算法重複(4.1.6~4.1.8)同時也可以得到對於其他聲源所產生的訊噪比增益。雖然實際系統作聲源分離時是對混音訊號 S_{mix} 做波束形成，不過藉由分開比較乾淨訊號經過波束形成所造成的增益與削減，我們可以求得更為精確的聲源分離訊噪比增益。

測試一：使用 LS 解的二聲源分離結果

	Input SNR (dB)	Output SNR (dB)	SNR Improvement (dB)
Speaker 01	-11.0899	-0.1282	10.9617
Speaker 02	-6.2749	9.9726	16.2475

表 4.10 使用理論陣列拓樸向量計算 LS 解二聲源分離結果

	Input SNR (dB)	Output SNR (dB)	SNR Improvement (dB)
Speaker 01	-11.0899	2.0907	13.1806
Speaker 02	-6.2749	10.6737	16.9486

表 4.11 使用校正過之陣列拓樸向量計算 LS 解二聲源分離結果

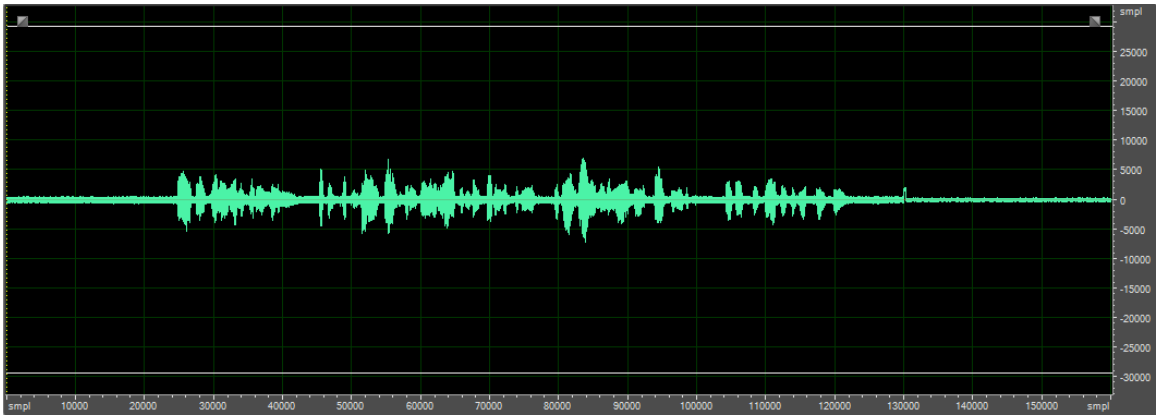


圖 4.12 使用 Sum beamformer 的合成結果(未分離狀態)

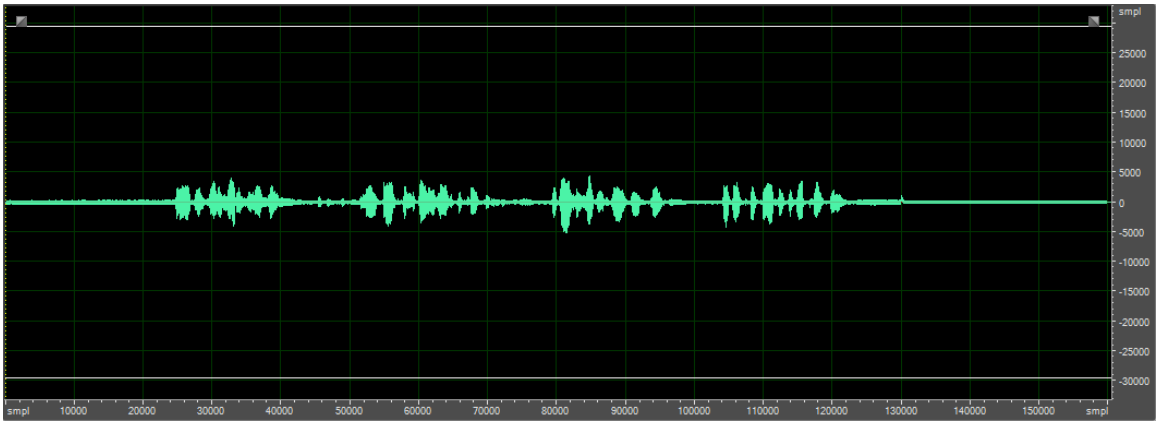


圖 4.13 使用 LS 解 beamformer 的聲源一分離結果

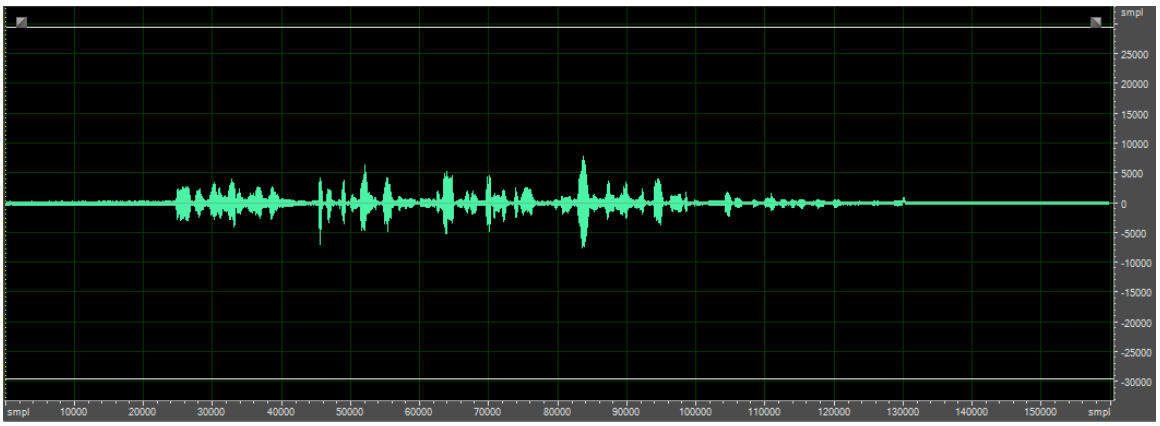


圖 4.14 使用 LS 解 beamformer 的聲源二的分離結果

測試二：使用 LS 解的四聲源分離參數效果

λ	Input SNR (dB)		Output SNR (dB)		SNR Improvement (dB)	
	Speaker 1	Speaker 2	Speaker 1	Speaker 2	Speaker 1	Speaker 2
0.01	-11.0899	-6.2749	0.0951	-2.3827	11.1850	3.8922
0.1	-11.0899	-6.2749	1.4337	0.1560	12.5235	6.4309
1	-11.0899	-6.2749	1.48920	2.4340	**12.5791**	**8.7089**

表 4.12 使用理論陣列拓樸向量計算 LS 解四聲源分離結果

λ	Input SNR (dB)		Output SNR (dB)		SNR Improvement (dB)	
	Speaker 1	Speaker 2	Speaker 1	Speaker 2	Speaker 1	Speaker 2
0.01	-11.0899	-6.2749	1.9735	2.4200	13.0634	8.6949
0.1	-11.0899	-6.2749	2.3648	3.1490	13.4547	9.4239
1	-11.0899	-6.2749	2.4980	4.2160	**13.5878**	**10.4909**

表 4.13 使用校正過之陣列拓樸向量計算 LS 解四聲源分離結果

測試三：使用 LCMV 事後解的聲源分離結果

	Input SNR (dB)	Output SNR (dB)	SNR Improvement (dB)
Speaker 01	-11.0899	-2.8458	8.2441
Speaker 02	-6.2749	17.9087	24.1836

表 4.14 使用理論陣列拓樸向量計算 LCMV 解聲源分離結果

	Input SNR (dB)	Output SNR (dB)	SNR Improvement (dB)
Speaker 01	-11.0899	6.5628	17.6527
Speaker 02	-6.2749	19.1444	25.4193

表 4.15 使用校正過之陣列拓樸向量計算 LCMV 解聲源分離結果

圖 4.15 使用 LCMV 解 beamformer 的聲源一分離結果

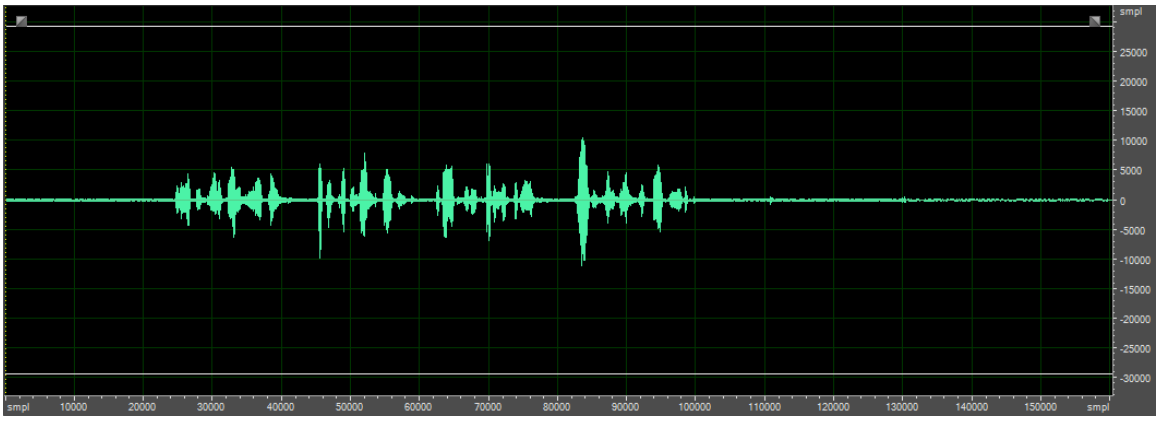


圖 4.16 使用 LCMV 解 beamformer 的聲源二分離結果

從三個測試中都可以發現，使用校正過之陣列拓樸向量在進行波束形成最佳解運算時，效果確實會有明顯的提升，輸出的聲源分離結果由人耳來做聽感評估也明顯有所提高。測試一的二聲源分離提升效果或許不明顯，可是到了測試二的四聲源分離時，由於空間干擾加重，角度探測的精密需求度也隨之提高，數據中可以明顯看出使用校正過之陣列拓樸向量的優勢。

測試二做了不同 Diagonal Loading 值的觀察，不同的聲源在不同角度中的空間特性都不盡相同，而不同的 Diagonal Loading 所建立的波束形成器權重範圍與形狀也會有差異，實驗結果發現，不同聲源 Diagonal Loading 的最佳值雖然大致分布在 0.01~1 的區間，可是當角度估測出現誤差，再與不適當的 Diagonal Loading 建立的波束形成器進行搭配時，就會在 Diagonal Loading 沒有選擇到最佳值時產生不佳的分離結果。同樣情形在使用校正過之陣列拓樸向量時，因為角度估測準確度提升，波束形成器的權重範圍與形狀也較符合實際空間特性，再調整 Diagonal Loading 就沒有出現太大幅度的效果減退。

測試三是拿 LCMV 的事後解來做測試。由於 LCMV 有將輸入訊號的空間資訊拿來當做最佳化時的考量，其排除干擾的效果是比 LS 解還要強的，可是如果在聲源分離的過程有角度上的誤差，很容易會將目標聲源也給抵消掉，表 4.14 的聲源一結果就是如此。而當聲源分離結果是使用校正過之陣列拓樸向量時，這樣子的情形就不會發生，如表 4.15

。

如果使用情境是經過限制的(例如固定聲源方向、數量的事後會議紀錄)，選擇 LCMV 事後解將能輸出較高品質的語音記錄；若是使用情境的變數很多(例如不固定聲源方位、數量、或需要即時語音處理的系統)，那選擇 LS 解將能保有較高的語音紀錄穩健度。

第五章 結論

5.1 研究成果

本論文提出一套利用校正過之陣列拓樸向量(Array Manifold Vector)，提升多重訊號分類演算法(Multiple Signal Classification)穩健度，並實現高準度之多聲源切音與分離的方法，本方法結合聲源頻譜與空間分佈資訊，利用機率決策對未知數量聲源方位進行分類，並將不同聲源語音進行切音與分離。本方法由於進行了陣列拓樸向量的完善校正，保證在低訊噪比下對多聲源的方位擁有強健的偵測率，且可排除錯誤偵測的聲源方位。

本論文方法可對未知聲源數目的連續訊號進行追蹤，並即時做聲源分離與切音的記錄。本論文提出的聲源分離架構，有不受聲源數目與聲源位置變動等影響的特性，並能有效提升輸出聲源訊號的訊噪比，輸出明顯分離後的多個聲源。由於本系統為一完整多聲源追蹤與分離之架構，而此架構為一級接著一級做資料的利用與運算，系統架構中的任一級皆可以根據研究目的來做擴充或改良。聲源方位的偵測與追蹤可提供各種語音純化的有效資訊，如語音存在估測 (VAD)、聲源分離、字詞切音等等，而聲源分離的結果亦可提供聲源特性並用於聲源類型的分類，或用於語音辨識。

5.2 未來展望

系統的架構在建立時是以即時處理著手，可是在計算多重訊號分類演算法與波束形成器時，其運算量會影響即時處理的效能。若能發展出同樣具有估量聲源豐富特徵，運算量較輕的角度估測演算法與波束形成演算法，則此系統將能更加泛用與可靠。另外，由於校正過的陣列拓樸向量在做到達角度估測運算時，能在極高與極低頻保有準確度，利用此較寬頻帶上的優勢，目標聲源的類型將不必侷限在較窄頻帶的人聲，而能有更多研究與發展的可能性。

Reference

[1-1] J. Pierre, M. Kaveh, "Experimental performance of calibration and direction-finding algorithms," *IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, May 1991, vol. 2, pp. 1365-1368.

[1-2] C.M.S. See, "Sensor array calibration in the presence of mutual coupling and unknown sensor gains and phases," *Electronics Letters*, vol. 30,. pp. 373-374, 1994.

[1-3] J. Chen, J. Benesty, Y. Huang, "Time delay estimation in room acoustic environments: an overview", *EURASIP Journal on applied signal* ,vol 2006, pp.170-188, 2006.

[1-4] C.P. Mathews, D. Zoltowski, "Eigenstructure techniques for 2-D angle estimation with uniform circular arrays", *IEEE Transactions on signal processing*, 1994.

[1-5] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Trans. Antennas and Propag.*, vol. AP-34, no. 3, pp.276-280,March 1986.

[1-6] J.M. Valin, F.Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering.," *Robotics and Autonomous Systems Journal (Elsevier)*, vol. 55, no. 3, pp. 216 – 228, 2007.

[1-7] Y. Zhao, W. Liu, and R. J. Langley, "An application of the least squares approach to fixed beamformer design with frequency-invariant constraints," *IET Signal Processing*, vol. 5, 2011.

[1-8] Barry D., Van Veen, and K.M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering." *IEEE ASSP Magazine*, pages 4-24, April 1988.

[2-1] H. WANG and M. KAVEH,".Coherent Signal-Subspace Processing for the Detection and Estimation of Angles of Arrival sf Multiple Wide-Band Sources", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-33, no. 4, AUGUST 1985.

[2-2] B. Rafaely, B. Weiss and E. Bachmat, "Spatial Aliasing in Spherical

Microphone Arrays", *IEEE Transactions on Signal Processing*, vol. 55, no. 3, MARCH 2007.

[2-3] Tao Su, Kapil Dandekar, and Hao Ling, “Simulation of mutual coupling effect in circular arrays for direction finding applications,” *Microwave and Optical Technol*. Lett., Sept. 2000

[3-1] A.G. Zeng, X.R. Liang and M.W. He, “A Novel Approach to Multimodal Function Optimization,” *Computer Engineering and Applications*, vol. 42, pp. 73-75, 2006.