

國立交通大學
電機與控制工程研究所

碩士論文

以轉移函數比值之方法

實現雙聲道麥克風聲源分離系統



A Dual-channel Sound Sources
Separation System Using Transfer Function Ratio
Approach

研究生： 陳 治 宏

指導教授： 胡 竹 生 博士

中華民國九十八年六月

以轉移函數比值之方法
實現雙聲道麥克風聲源分離系統

A Dual-channel Sound Sources
Separation System Using Transfer Function Ratio
Approach

研究生：陳治宏

Student : Chin-Hung, Chen

指導教授：胡竹生教授

Advisor : Prof. Jwu-Sheng, Hu

國立交通大學
電機與控制工程學系
碩士論文



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

Electrical and Control Engineering

May 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

以轉移函數比值之方法 實現雙聲道麥克風聲源分離系統

研究生：陳 治 宏

指導教授：胡 竹 生 教授

國立交通大學電機與控制工程研究所碩士班



摘 要

本論文針對環境中的語音訊號干擾源：如喇叭回音、背景音樂、或非使用者的語音等，使用雙聲道麥克風訊號處理，抑制干擾源的影響，保留使用者的語音資料而達到聲源分離的目的。本系統在硬體上使用 Microchip 推出的 PIC18F4550 來收集雙聲道類比麥克風資料並將其轉換為數位訊號，再透過 USB2.0 介面將聲音資料傳遞至電腦端，最後利用轉移函數比值演算法(Transfer Function Ratio)做出空間濾波器，達到聲源分離的效果。可以應用在吵雜環境中的語音純化等方面。

A Dual-channel Sound Sources Separation System Using Transfer Function Ratio Approach

Student : Chin-Hung, Chen

Advisor : Prof. Jwu-Sheng, Hu

Institute of Electrical and Control Engineering



ABSTRACT

The objective of this thesis is to suppress the interference signals such as loudspeaker echo, background music, or competing speech. We design an array signal processing technique on a dual-channel microphone for sound sources separation, which minimizes the influence of interference sources and retain the target speech. This system collects the audio data from two analog microphones via Microchip PIC18F4550, transforming the audio data into digital data, and transmits to PC through USB 2.0 interface. We implement the spatial filter by transfer function ratio approach for sound sources separation. Experiments for different types of interferences are conducted to show the effectiveness of the proposed system.

誌謝

呼呼，論文終於完成了～真是歷經千辛萬苦。這篇論文的誕生，當然首先最感謝的就是我的指導教授胡竹生老師。每當我研究上遇到問題，停滯不前時，老師總是會指導我前進的方向，讓我能夠繼續往前走，不會迷惘。對老師多年來的悉心指導，在此獻上最誠摯的謝意。

接著要感謝的，是實驗室裡的每一個伙伴，感謝永融學長，從東元比賽、PMC 比賽以來一直都帶著我往前進，有了他的帶領，比賽通通都拿冠軍啦～感謝興哥(佳興學長)在研究學習上常常給我指導，著實給我很大的幫助。感謝明唐，不只在研究上常給我建議，也常找大家去打球，維持體魄健壯，感謝我的前夥伴們：papa、hcy、瓊文，俊宇、還有 PLE 和 dowind，陪我度過了研究生生活的前兩年，常陪我去唱歌狂歡，心情苦悶時去喝兩杯，讓我這幾年來過得很開心～還有冰箱美人鏗元學姊、聯誼大師兄勁源學長、鐵拳無敵 judo、球場扛壩子肉鬆、時尚達人 lundy、英文很強嘟嘟、巴西帥哥活多福、沛錡、小蔡、聖翔、阿厂一ㄣ，賽門、多虧了有大家的陪伴，讓我的研究生生活過得多采多姿，留下了美好的回憶。

可惜的是，我的媽媽侯芳綺等不及看到我畢業，等不及看到我穿著碩士服的樣子，就先走了一步。媽，我完成我的畢業論文了，您有看到了嗎？感謝我的姊姊陳婉庭，媽媽離開後家中所有大小事情都是我們兩個人承擔了，我有很多事情不懂，姊姊每天辛苦上班之餘，還要一邊處理家裡很多事情，真的是很辛苦。最後感謝我的爸爸陳子鈞，雖然你還是每天想著要當總統，要上電視演講，但至少姊姊在上班，我在新竹做研究時爸爸都可以照顧自己，不會到處亂跑，希望爸爸以後也要聽話喔～

目 錄

摘 要.....	ii
ABSTRACT.....	iii
誌 謝.....	iv
目 錄.....	v
表 列.....	vii
圖 列.....	viii
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目標.....	2
1.3 論文架構.....	2
第二章 系統原理分析.....	3
2.1 陣列式訊號處理.....	3
2.1.1 陣列式訊號處理簡介.....	3
2.1.2 波束形成(Beamforming).....	3
2.1.3 理想均勻線性陣列模型.....	4
2.1.4 非理想均勻線性陣列模型.....	5
2.2 聲源分離.....	8
2.2.1 雙聲源假設.....	8
2.2.2 轉移函數比值演算法.....	10
2.2.3 聲源分離使用轉移函數比值演算法.....	12
第三章 系統軟硬體設計及實現.....	13
3.1 實驗平台.....	13
3.1.1 麥克風前端濾波及前端放大電路.....	13
3.1.2 參考電壓選擇電路.....	16
3.1.2 單晶片 PIC18F4550.....	16
3.2 USB 傳輸裝置.....	19
3.2.1 USB 通訊協定.....	19
3.2.2 USB 傳輸模式.....	21
3.4.3 USB 裝置韌體.....	22
3.3 電腦端軟體實現.....	24
3.3.1 軟體流程.....	24
3.3.2 使用流程.....	25
第四章 實驗結果與結論.....	26
4.1 在不同訊號雜訊比下的實驗結果分析.....	26
4.2 不同雜訊種類下的實驗結果分析.....	31
4.2.1 雜訊來源為非使用者的語音.....	31

4.2.2 雜訊來源為單一頻率的噪音.....	33
4.3 結論與未來展望.....	37
參考文獻.....	38



表 列

表 3-1. USB 四種傳輸模式比較.....	21
表 4-1. 背景音樂極大時的實驗結果.....	27
表 4-2. 背景音樂較小時的實驗結果.....	29
表 4-3. 背景音樂極小時的實驗結果.....	30
表 4-4. 干擾源為人聲時的實驗結果.....	32
表 4-5. 干擾源分別為 300Hz、500Hz 之單頻訊號時的實驗結果.....	34
表 4-6. 干擾源為單一頻率 2200Hz 時的實驗結果.....	35



圖 列

圖 2-1. 典型 MVDR 波束型成圖.....	3
圖 2-2. 理想均勻線性陣列模型.....	4
圖 2-3. 非理想均勻線性陣列模型.....	5
圖 2-4. 近場效應能量逸散.....	6
圖 2-5. 雙聲源系統模型.....	8
圖 3-1. 全系統硬體圖.....	13
圖 3-2. 麥克風前端濾波及放大電路架構圖.....	13
圖 3-3. 放大器及濾波器電路圖.....	14
圖 3-4. 放大器及濾波器模擬電路的頻率響應圖.....	15
圖 3-5. 麥克風及放大濾波器電路實體圖.....	15
圖 3-6. 參考電壓選擇電路圖.....	16
圖 3-7. 參考電壓選擇電路實體圖.....	16
圖 3-8. PIC18F4550 實驗板外觀.....	16
圖 3-9. PIC18F4550 硬體方塊圖.....	17
圖 3-10. USB 封包類型.....	20
圖 3-11. USB 裝置韌體流程圖.....	22
圖 3-12. 軟體實現流程圖.....	24
圖 3-13. 使用流程圖.....	25
圖 4-1. SNR 比與雜訊大小對純化效果影響測試實驗環境示意圖.....	26
圖 4-2. 純化結果比較圖(背景音樂極大).....	28
圖 4-3. 純化結果比較圖(背景音樂較小).....	29
圖 4-4. 純化結果比較圖(背景音樂極小).....	30
圖 4-5. 以人聲為雜訊來源實驗環境示意圖.....	31
圖 4-6. 干擾源為人聲之純化結果比較圖.....	32
圖 4-7. 語音訊號之波形圖、頻譜圖以及頻譜圖之局部放大圖.....	33
圖 4-8. 以 300Hz 之單頻訊號為雜訊源之純化結果比較圖.....	34
圖 4-9. 以 2200Hz 之單頻訊號為雜訊源之純化結果比較圖.....	36

第一章 緒論

1.1 研究動機

生活中，我們幾乎在任何時間都會聽見一個以上的聲源同時發出聲音，如在背景音樂中講話，在吵雜的工地講話，或甚至是兩個人同時講話等。但是通常我們所關心的都只有其中一個聲源，也就是我們的語音訊號。如果我們可以排除掉週遭的干擾，留下語音訊號，就可以大幅提升語音辨識率，達到語音純化的效果。

以往要消除干擾源的影響，往往要知道干擾源的特性，一般的處理方式是由時域、頻域去分析干擾源的特性，如果目前的干擾源在時域、頻域上沒有什麼固定的特徵，如電腦不斷的播放不同的音樂，那麼藉由時域、頻域去消除干擾源的影響，效能往往不理想。在本論文中我們擷取了訊號在空間中的特徵來做處理。桌面環境下的干擾源位置經常是固定的，如喇叭放好之後就很少會去更動喇叭的位置，只要針對當時的位置環境做空間濾波(Spatial Filter)，由空間上的特徵來對訊號處理，如此即可將與使用者不同角度的干擾源影響去除掉，達到增加語音訊號的訊噪比(SNR)，如此即可純化語音訊號，提高語音訊號辨識率。

然而，傳統的空間濾波方法受限於遠聲場平面波的假設，無法應用在語音訊號的環境中。在語音輸入訊號上，使用者與麥克風的距離是相當近的，一般情況介於 0.1~1m 之間，因此必須對傳統的空間濾波加以改良。本論文所使用的轉移函數比值(Transfer Function Ratio)演算法即為改良式的空間濾波，將訊號的時域、頻域特徵綜合考慮，設計出一改良的空間濾波器，有效達到增加語音訊號的訊噪比(SNR)，純化語音訊號的目的。

1.2 研究目標

本論文研究目標如下：

1. 利用 PIC18F4550 擷取雙聲道之類比麥克風陣列訊號，並將其轉換為數位訊號。
2. 使用 USB2.0 介面，將已經轉換為數位訊號的聲音訊號傳遞至電腦端，解決資料傳輸的問題。
3. 使用轉移函數比值演算法達到聲源分離的目的。

1.3 論文架構

整篇論文大致上可以切分為兩個部份，分別是第二章的系統原理分析與第三章的系統軟硬體設計與實現。系統原理分析這章節會對系統所使用的相關演算法的原理做介紹，系統軟硬體設計與實現這章節則會針對實驗平台的軟硬體開發環境、系統的軟硬體架構與 USB 資料傳輸架構做一個說明。最後的第四章節會呈現系統各項功能的測試結果並對研究成果做個結論。

第二章 系統原理分析

2.1 陣列式訊號處理

2.1.1 陣列式訊號處理簡介

陣列式訊號，是指數個相同的感應器排成特定的形狀，接收空間中傳遞的訊號，經過處理之後，達到空間濾波(Spatial Filter)的功能。空間濾波是指訊號在空間中從不同的角度的輸入，分別給予不同的增益。當不同的訊號在頻譜上重疊的部分太多，如語音訊號與 while noise，一般的頻域濾波器並無法將兩個訊號分開，但如果語音訊號與雜訊源對於陣列式感應器的輸入方向不同，具有不同的空間上的資訊，即可利用這樣的資訊設計空間濾波器，將不需要的訊號濾除，達到只接收特定方向的訊號。

2.1.2 波束形成(Beamforming)

波束形成的功能是為了接收空間中特定角度的訊號，同時壓抑其他方向的干擾。典型的 MVDR 波束形成圖如下：

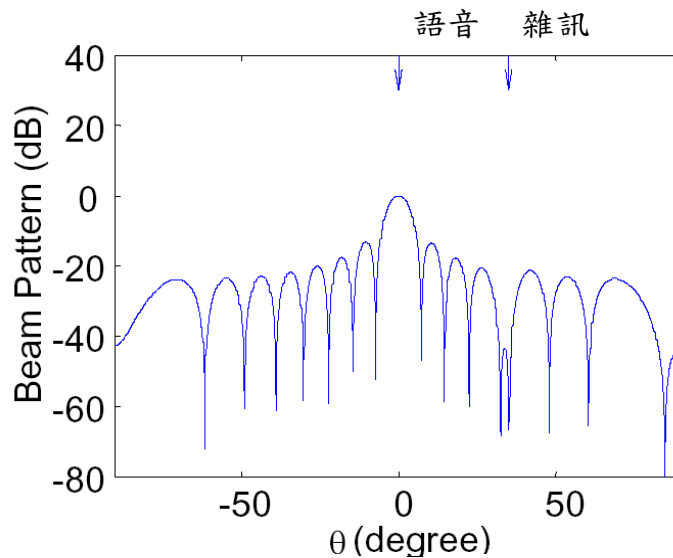


圖 2-1. 典型 MVDR 波束形成圖

利用輸入訊號具有空間上的資訊，將語音訊號輸入方向的增益固定在 0dB，其他方向的增益被壓低在 -10dB 以下，特別是雜訊輸入方向的增益被壓低到 -40dB。如此當輸入訊號經過處理之後，只會保留住語音訊號，其他方向的訊號變小，特別是雜訊輸入方向的訊號會變的更小。輸入輸出的 SNR 即可大為提高。

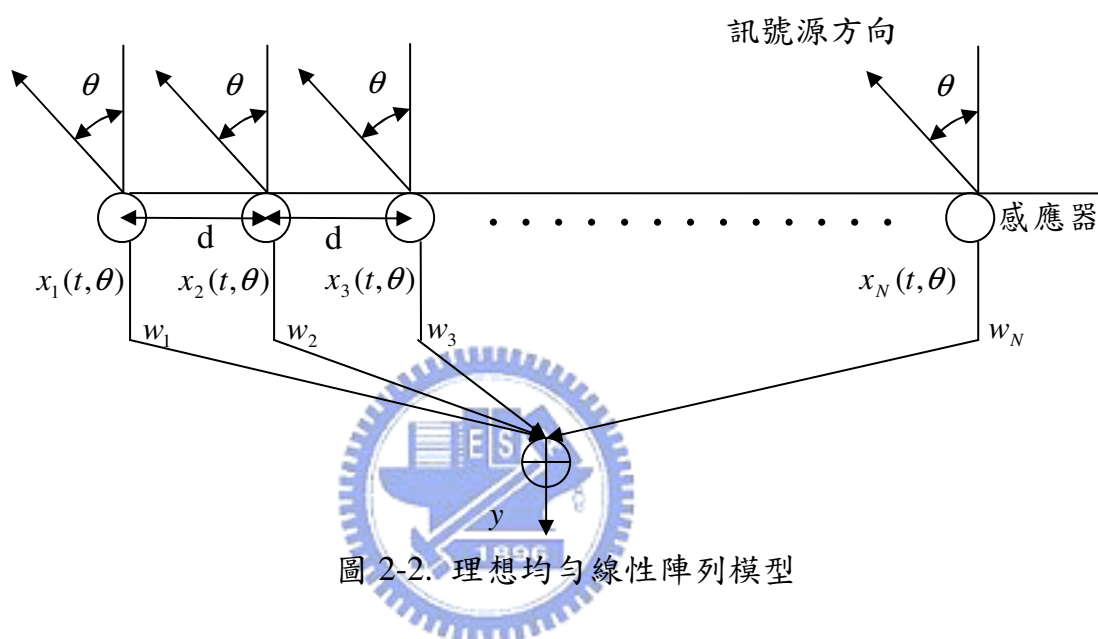


圖 2-2. 理想均勻線性陣列模型

2.1.3 理想均勻線性陣列模型

上圖 2-2 所示為均勻線性陣列模型的排列方式，是一般最常用的模型。訊號源設定為遠場平面波，訊號平面波在空間中傳播從 x_1 到 x_N 之間的能量損失忽略不計，所以每個感應器對於訊號源的相對角度均為 θ 。若訊號傳遞速度為 c ，第 1 個感應器收到的訊號 $x_1(t, \theta)$ 為 $s(t, \theta)$ ，那麼第 N 個感應器收到的訊號 $x_N(t, \theta)$ 就會是 $s(t - (N - 1) \times d \times \sin \theta / c, \theta)$ ，即為第 1 個感應器收到的訊號延遲 $(N - 1) \times d \times \sin \theta / c$ 時間。每個感應器接收到的訊號有不同的延遲，這些不同的延遲即是陣列式訊號處理中最重要的空間資訊。

當陣列式訊號環境滿足下列的假設：

- 1、 窄頻訊號(Narrow Band)。
- 2、 遠場平面波(Far Field Plane Wave)。

輸出訊號 y 就可以表示為：

$$y = \sum_{i=1}^N w_i x_i(t, \theta) \quad (2-1-1)$$

$$x(t, \theta) = \begin{bmatrix} x_1(t, \theta) \\ x_2(t, \theta) \\ \vdots \\ x_N(t, \theta) \end{bmatrix} = s(t, \theta) \begin{bmatrix} 1 \\ e^{jk_c d \sin \theta} \\ \vdots \\ e^{jk_c (N-1)d \sin \theta} \end{bmatrix} \quad (2-1-2)$$

$$k_c = \frac{2\pi}{\lambda_c}, \quad \lambda_c \text{ 窄頻訊號載波波長。}$$

理想狀況之下列出的式子非常的具有規律性，大部分的波束形成演算法都是在這樣的基礎之下推導出有規則的數學式，方便分析與使用。

2.1.4 非理想均勻線性陣列模型

理想狀況之下所推導的式子，無法應用在語音訊號的環境中，在語音輸入訊號上，使用者與麥克風的距離是相當近的，一般情況介於 0.1~1m 之間，如下圖 2-3 所示：

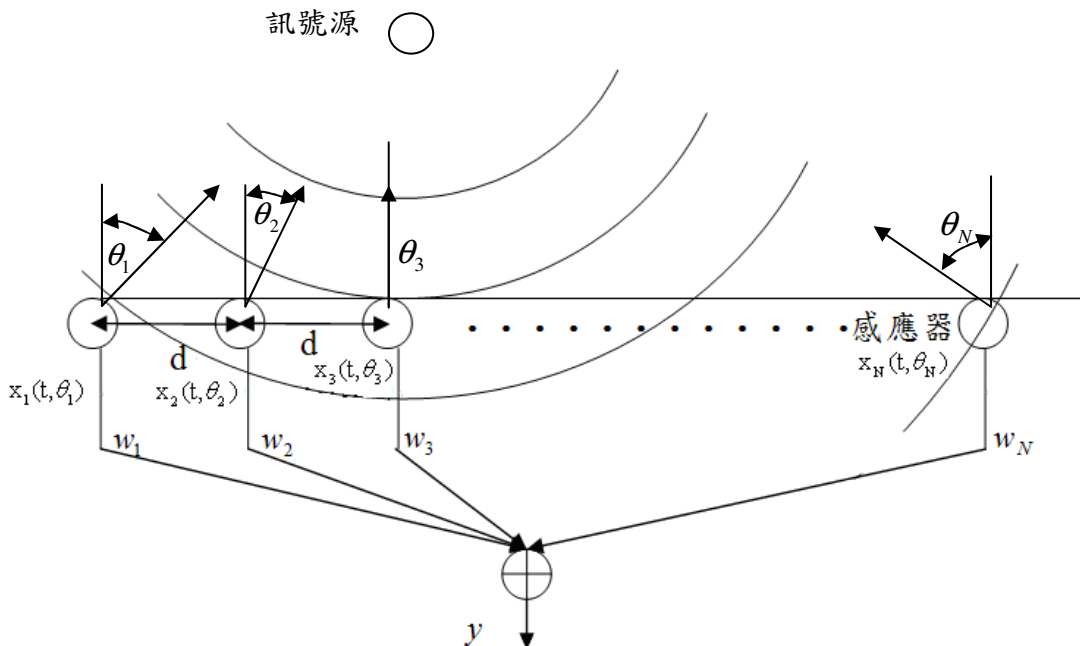


圖 2-3. 非理想均勻線性陣列模型

當使用者在近距離使用麥克風錄音的時候，屬於近場(Near Field)球形波，語音訊號亦非窄頻訊號，而每個麥克風與放大電路的特性都不大相同，造成的影響主要有下列幾項：

- 1、 元件不匹配：每個麥克風對於頻率 ω 與方向 θ 的增益不同，放大電路也因為元件本身有誤差而造成頻率響應不同，因此收到的訊號要再乘以個別的麥克風與放大電路的頻率響應與方向響應 $g_i(\omega, \theta_i)$ 。
- 2、 麥克風指向訊號源的方向：每一個麥克風指向訊號源的角度均不同，不再是單一的 θ ，因此式子並無法化簡成只有 t 、 θ 兩個變數。
- 3、 能量散失無法忽略：因為距離非常接近，一般情況介於 0.1~1m 之間，因此球形波達到第 1 個麥克風與到第 N 個麥克風的能量逸散相當嚴重，以下面的情況為例子：

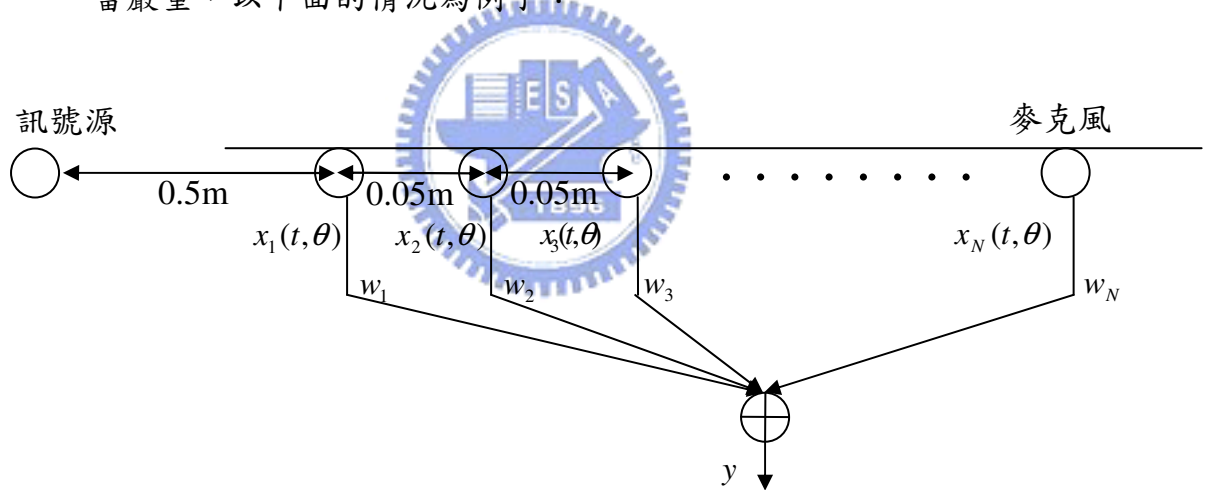


圖 2-4. 近場效應能量逸散

如圖 2-4 所示，訊號源與所有麥克風在同一直線上，且在所有麥克風的左邊，距離第 1 個麥克風為 0.5m，每個麥克風間距為 0.05m，陣列式麥克風個數為 8 個，語音訊號能量集中在球型波的表面。

則第 1 個麥克風收到的能量:第 8 個麥克風收到的能量

$$= \frac{1}{(0.5)^2} : \frac{1}{(0.5 + 7 \times 0.05)^2} = 2.89 : 1$$

能量逸失高達 2.89 倍，無法將這樣的效應忽略不計。

將每個麥克風各別的能量逸失對增益造成的影響表示 k_i 。

所以近場寬頻的環境下，陣列式麥克風處理完的訊號只能以下面的式子表示

$$y = \sum_{i=1}^N w_i x_i(t, \theta_i) g(\omega, \theta_i) k_i \quad (2-1-3)$$

式子化簡將會變得相當的困難，在這樣子的基礎之下，大部分的波束形成演算法，也會變得相當複雜，分析與使用將變得相當不容易。



2.2 聲源分離

2.2.1 雙聲源假設

首先考慮環境中有兩個聲源， N 個麥克風，兩個聲源 s_1 、 s_2 分別經空間轉移函數 $A_{11}(\omega) \sim A_{1N}(\omega)$ 、 $A_{21}(\omega) \sim A_{2N}(\omega)$ 後，由 N 個麥克風收集聲音，錄到的聲音為 $x_1(k, e^{j\omega}) \sim x_N(k, e^{j\omega})$ ，系統如下圖所示：

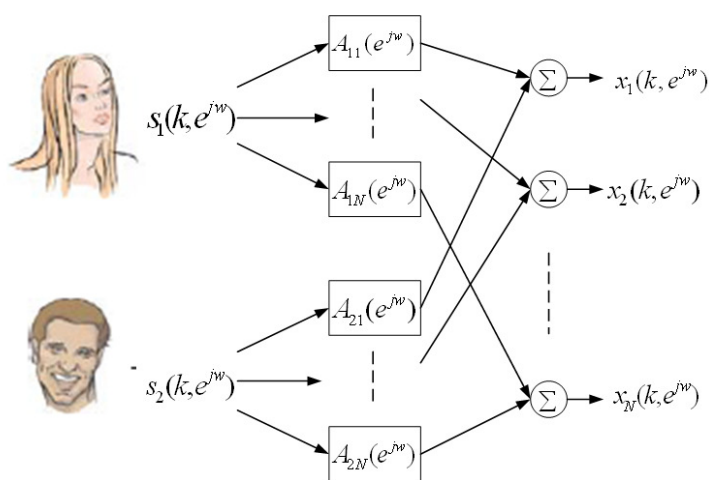


圖 2-5. 雙聲源系統模型

如圖 2-5，兩個聲源 s_1 、 s_2 經過空間中的某個轉移函數 $A_{11} \sim A_{1N}$ 、 $A_{21} \sim A_{2N}$ ，由麥克風陣列 $x_1 \sim x_N$ 所接收。麥克風陣列所收到的訊號可表示為

$$\begin{bmatrix} x_1(k, \omega) \\ \vdots \\ x_N(k, \omega) \end{bmatrix} = \begin{bmatrix} A_{11}(\omega) & A_{21}(\omega) \\ \vdots & \vdots \\ A_{1N}(\omega) & A_{2N}(\omega) \end{bmatrix} \begin{bmatrix} s_1(k, \omega) \\ s_2(k, \omega) \end{bmatrix} \quad (2-2-1)$$

假設 y_1 是由收到的麥克風資料 $x_1(k, \omega) \sim x_N(k, \omega)$ ，經過某種聲源分離估測 $w_1(k, \omega) \sim w_N(k, \omega)$ 後所估測出的 s_1 ，則 y_1 可表示為

$$\begin{aligned} y_1 &= w_1(k, \omega)x_1(k, \omega) + w_2(k, \omega)x_2(k, \omega) + \dots + w_N(k, \omega)x_N(k, \omega) \\ &= w_1(k, \omega)[A_{11}(k, \omega)s_1(k, \omega) + A_{21}(k, \omega)s_2(k, \omega)] + \dots \\ &\quad + w_N(k, \omega)[A_{1N}(k, \omega)s_1(k, \omega) + A_{2N}(k, \omega)s_2(k, \omega)] \\ &= [w_1(k, \omega)A_{11}(k, \omega) + \dots + w_N(k, \omega)A_{1N}(k, \omega)]s_1(k, \omega) \\ &\quad + [w_1(k, \omega)A_{21}(k, \omega) + \dots + w_N(k, \omega)A_{2N}(k, \omega)]s_2(k, \omega) \end{aligned}$$

經過整理之後可以得到下面的式子

$$\begin{bmatrix} A_{11}(k, \omega) & A_{12}(k, \omega) & \dots & A_{1N}(k, \omega) \\ A_{21}(k, \omega) & A_{22}(k, \omega) & \dots & A_{2N}(k, \omega) \end{bmatrix} \begin{bmatrix} w_1(k, \omega) \\ w_2(k, \omega) \\ \vdots \\ w_N(k, \omega) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2-2-2)$$

我們可以發現，式(2-2-2)是一個 $AX=B$ 的形式，可以使用最小平方方法 (Least Square) 來得到估測方法的矩陣 W ：

$$W = (A^T A)^{-1} A^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2-2-3)$$

其中 $A = \begin{bmatrix} A_{11}(k, \omega) & A_{12}(k, \omega) & \dots & A_{1N}(k, \omega) \\ A_{21}(k, \omega) & A_{22}(k, \omega) & \dots & A_{2N}(k, \omega) \end{bmatrix}$

由(2-2-3)式，可以得到估測方法的矩陣 W ，但是在此式中，空間轉移函數之脈衝響應矩陣 A 是必要資訊。但在現實環境下，此脈衝響應矩陣 A 是不易得到的。

2.2.2 轉移函數比值演算法

為了估測空間轉移函數，在這裡假設只有一個聲源 s ， N 個麥克風 $x_1 \sim x_N$ ，則麥克風接收到的聲音資訊可表示為

$$x_n(k, \omega) = A_n s(k, \omega) + v_n(k, \omega) \quad (2-2-4)$$

其中 A_n 為空間轉移函數， v_n 為環境中的雜訊。

空間中的轉移函數可能較難得到，但或許可以得到轉移函數的比值，而這個數據可能會是有用的，這就是轉移函數比值的概念。由(2-2-4)式，套入轉移函數比值的概念，可得以下的式子

$$x_n(k, \omega) - \frac{A_n(\omega)}{A_1(\omega)} x_1(k, \omega) = U_n(k, \omega) \quad (2-2-5)$$

觀察(2-2-5)式， $x_n(k, \omega)$ 及 $x_1(k, \omega)$ 都是已知的麥克風資料，如果能夠得到轉移函數比值 $\frac{A_n(\omega)}{A_1(\omega)}$ 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 $\frac{A_n(\omega)}{A_1(\omega)}$ ，即所謂的轉移函數比值矩陣命名為 $H_n(\omega)$ ，再經過移項整理後可得

$$x_n(k, \omega) = H_n(\omega) x_1(k, \omega) + U_n(k, \omega) \quad (2-2-6)$$

去計算(2-2-6)式 K 個 frame 的交功率頻譜密度(Cross-power Spectral Density)，可得以下的式子

$$\begin{aligned} \Phi_{x_n x_1}(k, \omega) &= H_n(\omega) \Phi_{x_1 x_1}(k, \omega) + \Phi_{U_n x_1}(k, \omega) \\ k &= 1, 2, \dots, K \end{aligned} \quad (2-2-7)$$

將式(2-2-7)以矩陣型式表示得以下的式子

$$\begin{array}{c} \left[\begin{array}{c} \Phi_{x_n x_1}(1, \omega) \\ \Phi_{x_n x_1}(2, \omega) \\ \vdots \\ \Phi_{x_n x_1}(K, \omega) \end{array} \right] \\ \text{B} \end{array} = \begin{array}{c} \left[\begin{array}{ccc} \Phi_{x_1 x_1}(1, \omega) & 1 \\ \Phi_{x_1 x_1}(2, \omega) & 1 \\ \vdots & 1 \\ \Phi_{x_1 x_1}(K, \omega) & 1 \end{array} \right] \\ \text{A} \end{array} \begin{array}{c} \left[\begin{array}{c} H_n(\omega) \\ \Phi_{U_n x_1}(k, \omega) \end{array} \right] \\ \text{X} \end{array} \quad (2-2-8)$$

觀察(2-2-8)式可以發現，(2-2-8)式也是一個 $AX=B$ 的形式，轉移函數比值矩陣 $H_n(\omega)$ 的話可以利用最小平方法 $X=(A^T A)^{-1} A^T B$ 來求得，推導如下：

$$\begin{aligned} \left[\begin{array}{c} H_n(\omega) \\ \Phi_{U_n x_1}(k, \omega) \end{array} \right] &= \left\{ \left[\begin{array}{cccc} \Phi_{x_1 x_1}(1, \omega) & \Phi_{x_1 x_1}(2, \omega) & \cdots & \Phi_{x_1 x_1}(K, \omega) \\ 1 & 1 & 1 & 1 \end{array} \right] \left[\begin{array}{c} \Phi_{x_1 x_1}(1, \omega) & 1 \\ \Phi_{x_1 x_1}(2, \omega) & 1 \\ \vdots & 1 \\ \Phi_{x_1 x_1}(K, \omega) & 1 \end{array} \right] \right\}^{-1} \\ &\quad \left[\begin{array}{c} \Phi_{x_n x_1}(1, \omega) \\ \Phi_{x_n x_1}(2, \omega) \\ \vdots \\ \Phi_{x_n x_1}(K, \omega) \end{array} \right] \\ &= \frac{1}{\left\{ \Phi_{x_1 x_1}^2(K, \omega) \right\} - \left\{ \Phi_{x_1 x_1}(K, \omega) \right\}^2} \\ &\quad \left[\begin{array}{cc} 1 & -\left\{ \Phi_{x_1 x_1}(K, \omega) \right\} \\ -\left\{ \Phi_{x_1 x_1}(K, \omega) \right\} & \left\{ \Phi_{x_1 x_1}^2(K, \omega) \right\} \end{array} \right] \left[\begin{array}{c} \left\{ \Phi_{x_1 x_1}(K, \omega) \Phi_{x_n x_1}(K, \omega) \right\} \\ \left\{ \Phi_{x_n x_1}(K, \omega) \right\} \end{array} \right] \end{aligned}$$

由上式可以求得轉移函數比值矩陣 $H_n(\omega)$ ：

$$H_n(\omega) = \frac{\left\langle \Phi_{x_1 x_1}(\omega) \Phi_{x_n x_1}(\omega) \right\rangle - \left\langle \Phi_{x_1 x_1}(\omega) \right\rangle \left\langle \Phi_{x_n x_1}(\omega) \right\rangle}{\left\langle \Phi_{x_1 x_1}^2(\omega) \right\rangle - \left\langle \Phi_{x_1 x_1}(\omega) \right\rangle^2} \quad (2-2-9)$$

其中 $\left\langle \beta(\omega) \right\rangle \equiv \sum_{k=1}^K \beta(k, \omega)$

2.2.3 聲源分離使用轉移函數比值演算法

考慮一個兩個聲源的系統：聲源 s_1, s_2 ， N 個麥克風 $x_1 \sim x_N$ ，則麥克風

$x_1 \sim x_N$ 接收到的聲音資訊可表示為：

$$\begin{bmatrix} x_1(k, \omega) \\ \vdots \\ x_N(k, \omega) \end{bmatrix} = \begin{bmatrix} A_{11}(\omega) & A_{21}(\omega) \\ \vdots & \vdots \\ A_{1N}(\omega) & A_{2N}(\omega) \end{bmatrix} \begin{bmatrix} s_1(k, \omega) \\ s_2(k, \omega) \end{bmatrix} \quad (2-2-10)$$

如果轉移函數比值 $\frac{A_{11}(\omega)}{A_{1n}(\omega)}$ 、 $\frac{A_{21}(\omega)}{A_{2n}(\omega)}$ $n=2,3,\dots,N$ 為已知，將上式代入轉移函數比值的概念可得：

$$\begin{aligned} x_1(k, \omega) - \frac{A_{11}(\omega)}{A_{1n}(\omega)} x_n(k, \omega) &= A_{21}(\omega) s_2(k, \omega) - A_{2n}(\omega) \frac{A_{11}(\omega)}{A_{1n}(\omega)} s_2(k, \omega) \\ &= A_{2n}(\omega) \left(\frac{A_{21}(\omega)}{A_{2n}(\omega)} - \frac{A_{11}(\omega)}{A_{1n}(\omega)} \right) s_2(k, \omega) \end{aligned} \quad (2-2-11)$$

觀察(2-2-11)式，若已知轉移函數比值 $\frac{A_{11}(\omega)}{A_{1n}(\omega)}$ 、 $\frac{A_{21}(\omega)}{A_{2n}(\omega)}$ $n=2,3,\dots,N$ ，

則可以經過簡單的乘法及減法運算，完全消除掉聲源 s_1 的部份。

由(2-2-11)式，如果 $\left(\frac{A_{21}(\omega)}{A_{2n}(\omega)} - \frac{A_{11}(\omega)}{A_{1n}(\omega)} \right)^{-1}$ 存在，則可得

$$\left(x_1(k, \omega) - \frac{A_{11}(\omega)}{A_{1n}(\omega)} x_n(k, \omega) \right) \left(\frac{A_{21}(\omega)}{A_{2n}(\omega)} - \frac{A_{11}(\omega)}{A_{1n}(\omega)} \right)^{-1} = A_{2n}(\omega) s_2(k, \omega) \quad (2-2-12)$$

由(2-2-12)，可得 $A_{2n}(\omega) s_2(k, \omega)$ ， $n=1,2,\dots,N$ 。同理亦可得 $A_{1n}(\omega) s_1(k, \omega)$ ， $n=1,2,\dots,N$ 。如此即可達成聲源分離的效果。

第三章 系統軟硬體設計及實現

3.1 實驗平台

本系統的實際成品圖如下所示，硬體上包括了雙聲道類比麥克風、放大電路、參考電壓選擇電路及單晶片 PIC18F4550。以下小節會對其分別介紹。

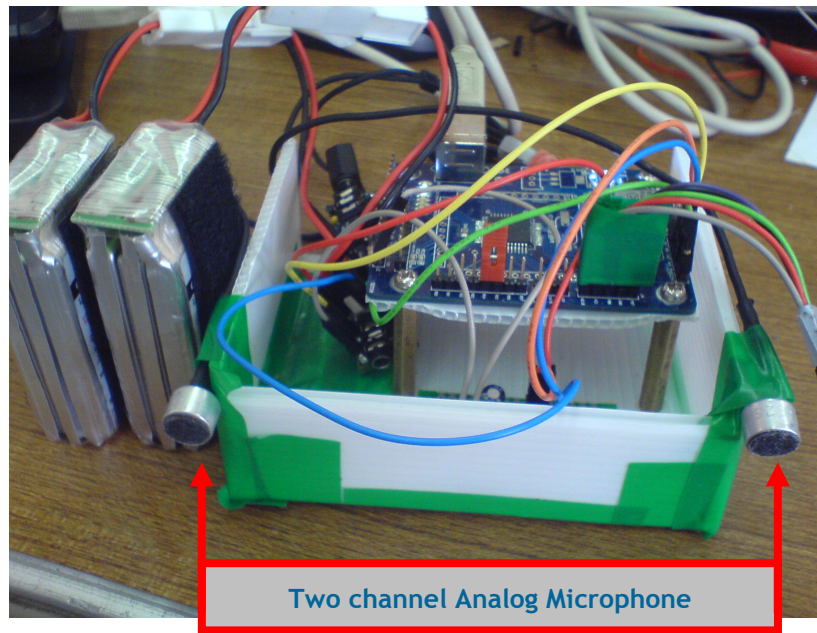


圖 3-1. 全系統硬體圖

3.1.1 麥克風前端濾波及前端放大電路

方塊圖：

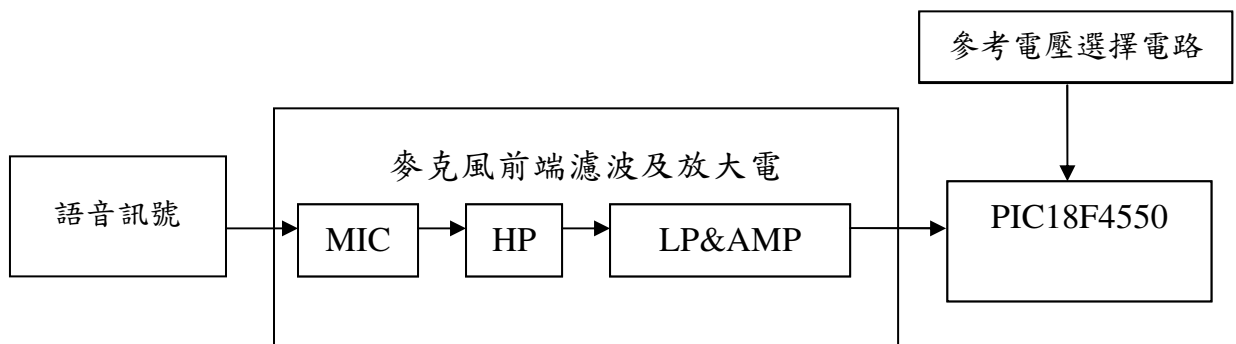


圖 3-2. 麥克風前端濾波及放大電路架構圖

電容式麥克風將語音訊號轉換為電壓訊號後，必須先通過 High-Pass Filter，主要原因有二：

- 1、 麥克風輸出的訊號除了語音訊號之外，另外帶有 5V 的直流訊號。為了避免放大器將直流訊號也一起放大，必須先通過 High-Pass Filter，濾除直流電壓部分。
- 2、 麥克風對於低頻的聲音相當的敏感，如空氣的擾動，而這樣的雜訊也往往使得語音訊號失真。

在此，我們將 High-Pass Filter 的 3dB 點定在 40Hz，濾除掉大部分的低頻雜訊，並保留住大部分的語音資訊。

訊號通過 High-Pass Filter 之後的訊號非常小，peck-to-peck 電壓值只有幾十個 mV，因此利用放大器將訊號放大到 -1V~1V 之間。語音訊號在後段 A/D 的 Sampling Rate 為 16KHz，根據 Nyquist Sampling Rate，Sampling Rate 至少要是訊號頻帶的兩倍，訊號才不會有 Aliasing 的問題，因此要再加上一個 Low-Pass Filter 將 8KHz 以上的訊號濾除。

經過濾波與放大電路後，在安靜的實驗室的環境下的輸出訊號 peck-to-peck 值約 10~15mV。

放大器及濾波器電路圖：

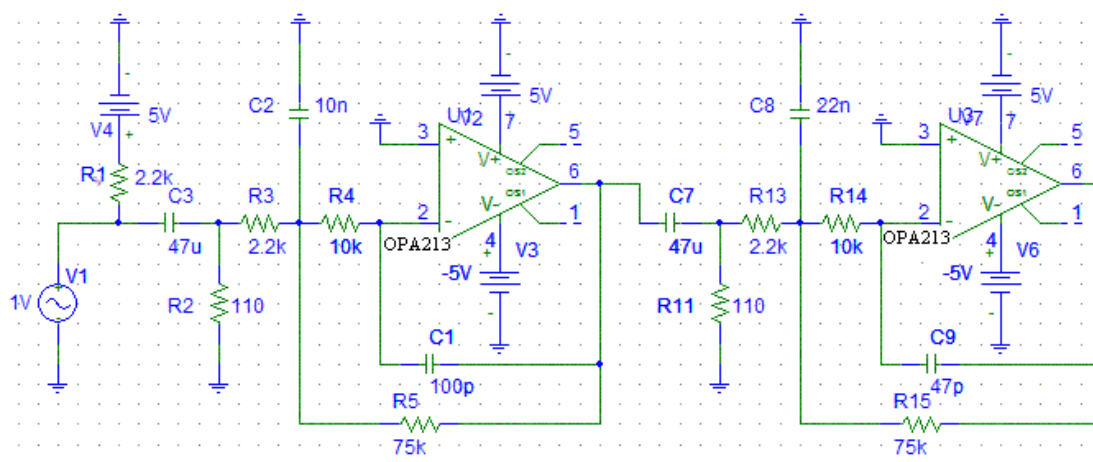


圖 3-3. 放大器及濾波器電路圖

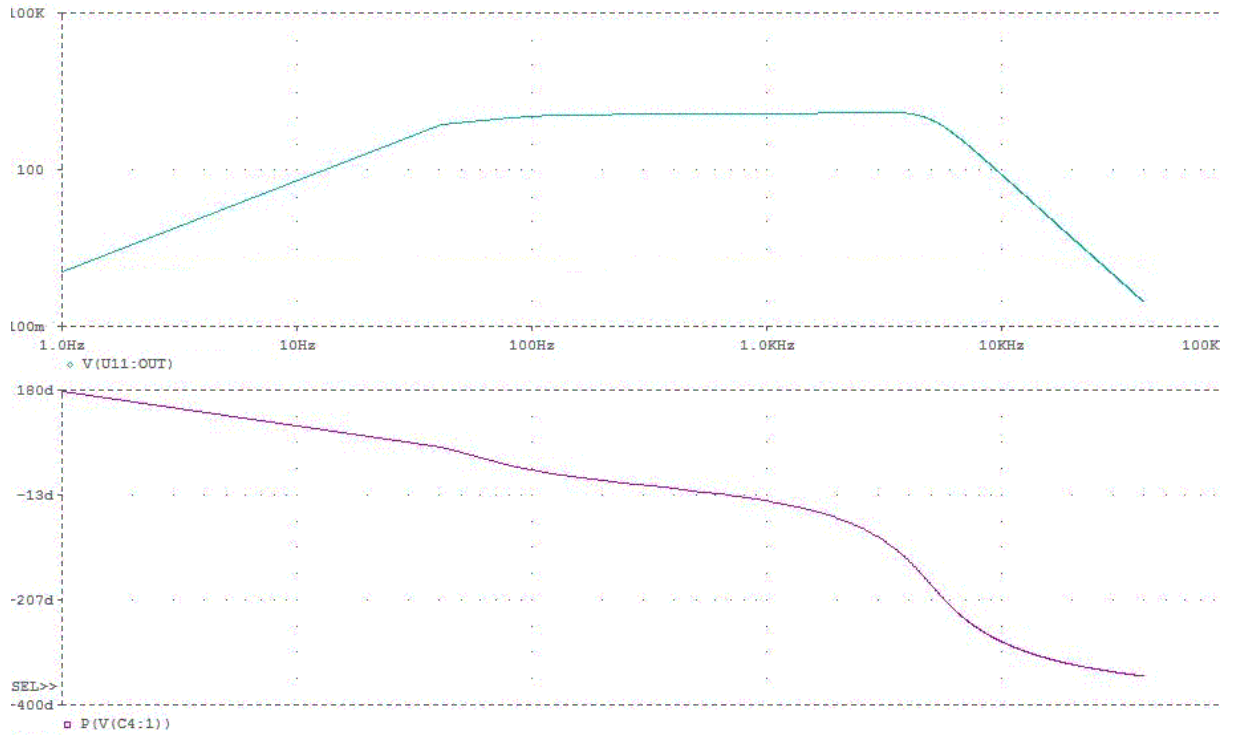


圖 3-4. 放大器及濾波器模擬電路的頻率響應圖



圖 3-5. 麥克風及放大濾波器電路實體圖

3.1.2 參考電壓選擇電路

訊號經過放大之後，要經由 PIC18F4550 的 A/D 埠轉換為數位訊號。因為後端用來作 A/D 轉換的單晶片 PIC18F4550 原本預設的 AD 範圍為 0~5V，但我們輸入的是有正負半周的聲音訊號，所以需要輸入適當的參考電壓，這邊使用前端放大電路的直流正負 12V 供電，利用齊納二極體作出分壓電路，拉出約+2V 及-2V 的參考電壓。其電路及實體圖如下：

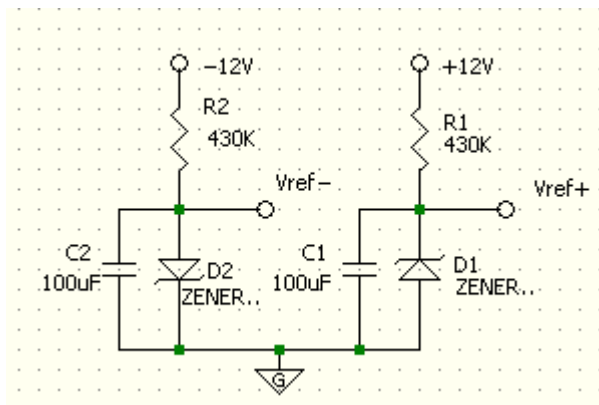
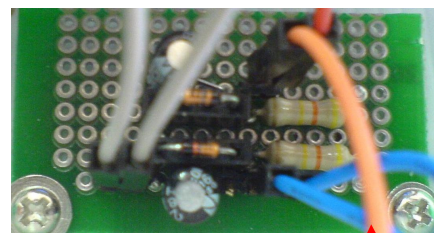


圖 3-6. 參考電壓選擇電路圖



Vref select board

圖 3-7. 參考電壓選擇電路實體圖

3.1.2 單晶片 PIC18F4550

訊號經過前端的放大及濾波之後，我們使用 Microchip 推出的單晶片 PIC18F4550，瑞昱科技所開發的實驗板來做訊號的收集，A/D 轉換，以及 USB 資料傳輸。

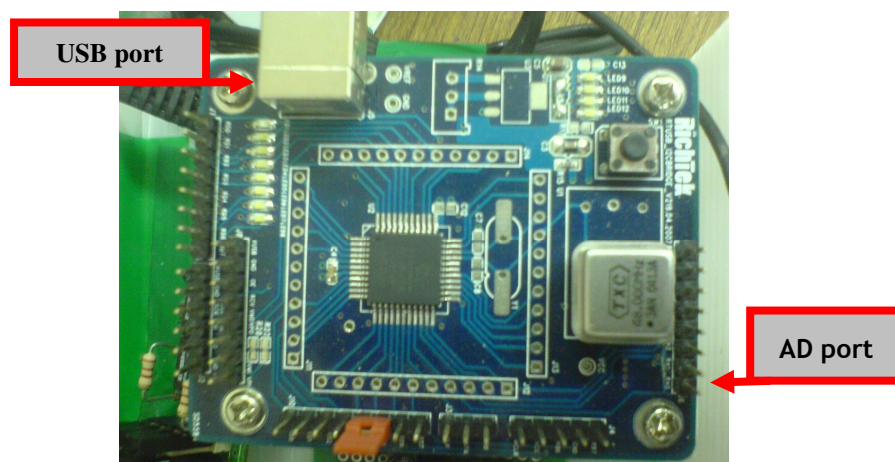


圖 3-8. PIC18F4550 實驗板外觀

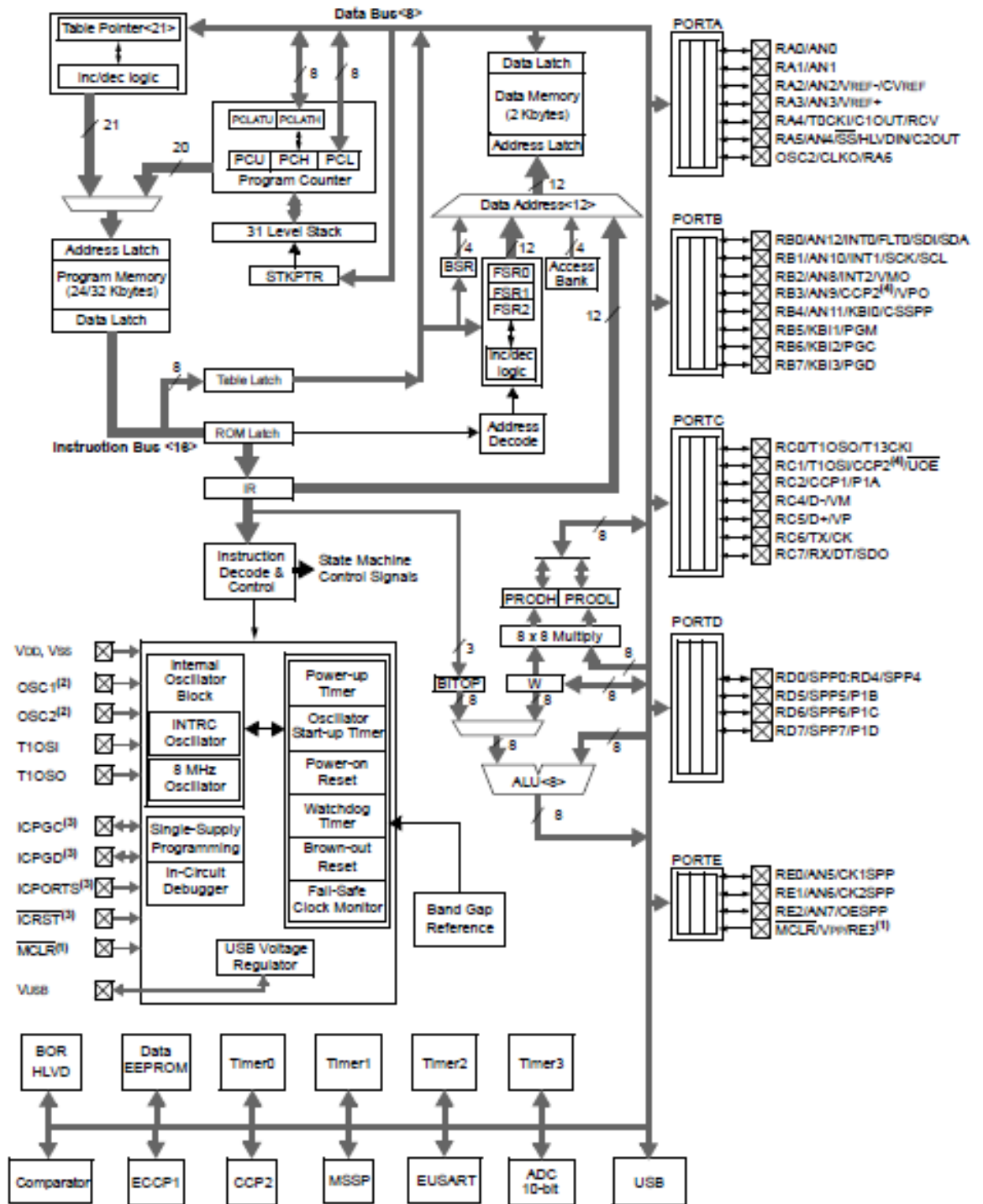


圖 3-9. PIC18F4550 硬體方塊圖

A/D port 即為圖 16 中的 RA0~RA5，經由前端電路放大後的訊號由 RA0 及 RA4 輸入，參考電壓由 RA2 及 RA3 輸入做 A/D 轉換：

每筆資料儲存位元：8-bit。

Maximum Sampling Rate：8KHz。

Sampling Range：-2V~+2V。

以下列出較為重要的 PIC 18F4550 硬體規格。

- Operating frequency up to 48MHz
- 32 Kbytes Flash program memory
- 5 bidirectional I/O ports
- USB2.0 Interface
- 10-BIT ANALOG-TO-DIGITAL CONVERTER (A/D) MODULE



3.2 USB 傳輸裝置

3.2.1 USB 通訊協定

USB 的通訊協定主要可以分為三種封包(Package)：

1、執照封包(Token)：指示型態與方向。包含了 PID、Address、Endpoint、CRC。PID 包含 IN、OUT、Setup、SOF。位址跟端點提供接收裝置的位址及接收的端點。CRC 檢測此封包是否接收正確。無論何種傳輸，每次傳輸的第一個封包一定為執照封包。

- (1)、IN：裝置送封包給主機。
- (2)、OUT：主機送封包給裝置。
- (3)、Setup：僅給控制傳輸使用，用來確認裝置要求。
- (4)、SOF：每 1mS 發生一次，用來表示 USB Frame 的抵達。

2、資料封包(Data)：實際上的傳輸資料。包含了 PID、Data、CRC。PID 包含 Data0、Data1。Data 則為實際上的傳輸資料。CRC 檢測此封包是否接收正確。

- (1)、Data0，Data1，兩種封包會交替傳送，透過主機內部的 Toggle 位元狀態與 DATA PID 相比較，可偵測到錯誤的交握封包。

3、交握封包(Hand-Shake)：USB 的傳輸狀態，有四種 PID，除了等時傳輸之外，每次傳輸的最後一個封包必為交握封包。

- (1)、ACK：成功。正確無誤的接收資料。
- (2)、NAK：忙線。再試一次。
- (3)、STALL：停滯。發生無法預料的問題，表示不瞭解裝置的需求。
- (4)、NYET：端點忙線。資料傳輸成功，但端點尚未準備好接收下一筆資料。

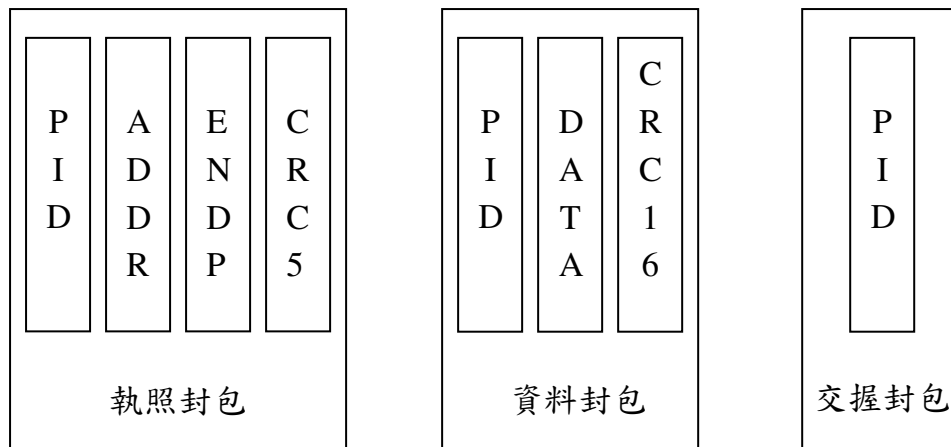


圖 3-10. USB 封包類型

封包的接收及發送由 USB 介面處理。



3.2.2 USB 傳輸模式

USB 系統提供了四種不同的傳輸模式：

- 1、 巨量傳輸(Bulk)：突發性的傳輸模式。資料封包大小為 8、16、32、64 位元組。除了資料封包之外另有交握封包(Hand-Shake Package)，及自動錯誤資料檢核機制(CRC)，如資料傳送錯誤，可要求裝置重送封包，確保資料的正確性。
- 2、 中斷傳輸(Interrupt)：類似巨量傳輸，資料封包大小為 1~64 位元組。高速的裝置中。需經由主機規則固定間隔詢問。
- 3、 等時傳輸(Isochronous)：在固定的時間傳出封包，主要使用在音頻與影像等資料流中。為了確保封包可以在固定的時間送出，無 Hand-Shaking 封包，僅具有 CRC 錯誤檢核，資料傳輸錯誤亦不再重送封包。時間是最重要的要求條件。
- 4、 控制傳輸(Control)：用來配置及傳送命令給裝置，確認裝置要求。

而在實驗平臺的架設中，只用到控制傳輸及巨量傳輸。

	封包大小(Byte)	時間	資料檢 查	應用
巨量傳輸	8、16、32、64	盡快完成	有	儲存裝置、印表機
中斷傳輸	1~64	1~255ms	有	滑鼠、鍵盤
等時傳輸	1~1024	1ms	沒有	影像
控制傳輸		盡快完成	有	命令

表 3-1. USB 四種傳輸模式比較

3.4.3 USB 裝置韌體

USB 裝置韌體流程圖：

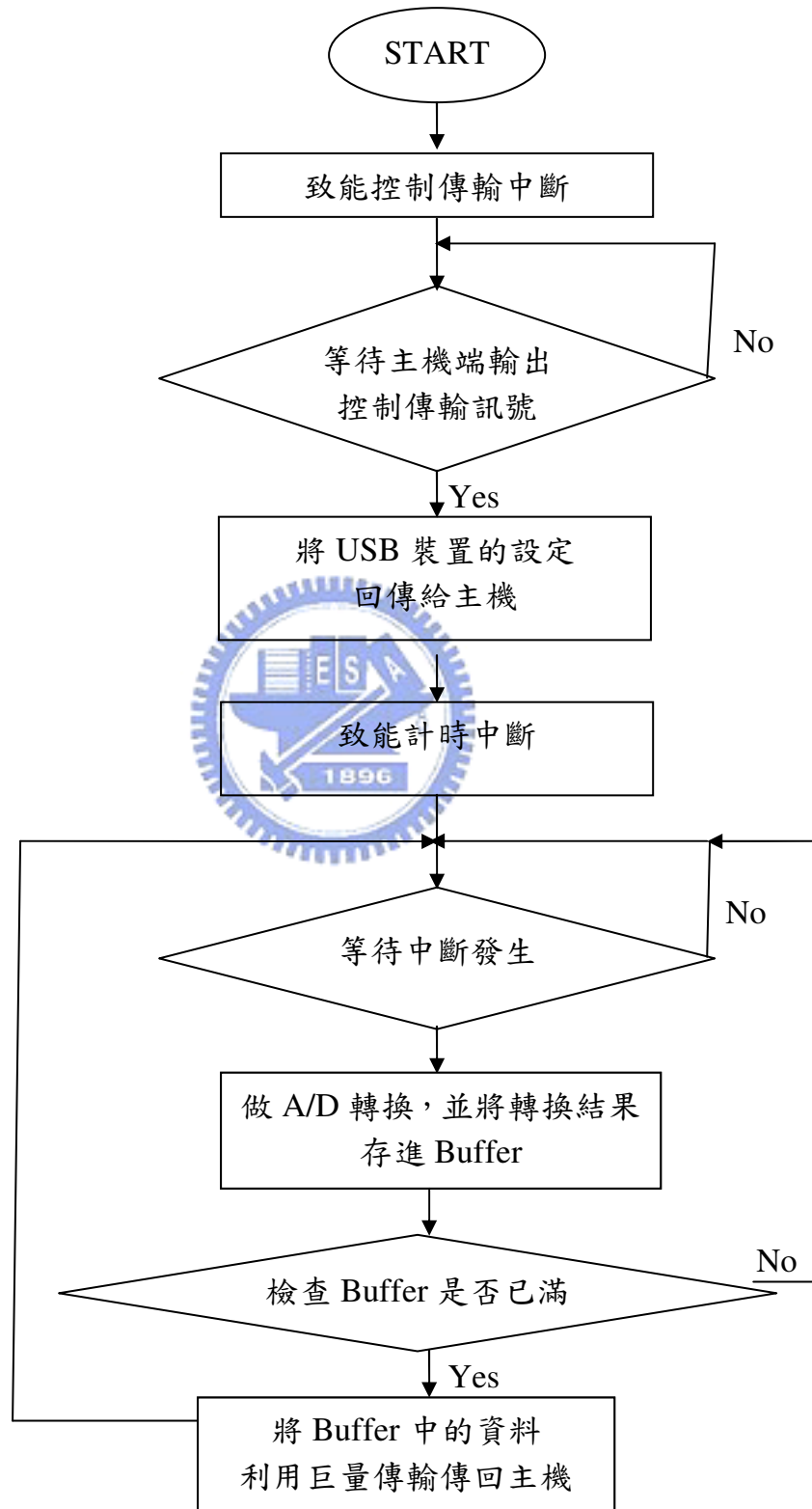


圖 3-11. USB 裝置韌體流程圖

當韌體燒入 USB 裝置的 RAM 之後，以 USB 接線與主機連接。主機經由控制傳輸要求 USB 裝置回傳裝置描述元，主機由裝置描述元的資訊，確認 USB 裝置，並使用相對應的驅動程式來驅動 USB 裝置，回傳的裝置描述元主要有下列幾項：

- 1、 Device Descriptor：規格版本、裝置群組、VID、PID、配置數目。
- 2、 Configuration Descriptor：端點長度、界面數目、界面數值、電源屬性、所需電源。
- 3、 Interface Descriptors：切換設定、端點數目、界面群組。
- 4、 Endpoint Descriptors：使用端點及方向、端點型態、最大封包大小、輪詢間隔。
- 5、 String Descriptors：公司、產品名稱。

實驗平台 PIC18F4550 經由控制傳輸回傳裝置描述元，即可成功讓主機確認 USB 裝置，使用相對應的驅動程式。主機成功確認 USB 裝置之後，啟動計時中斷，計時中斷的產生頻率為 8KHz。

當計時中斷發生時，PIC18F4550 控制將兩聲道之輸入訊號做 A/D 轉換，並將轉換結果存入 buffer 中；檢查是否已經將 buffer 填滿 64byte，如未填滿則繼續等待計時中斷發生，如果已經填滿 64byte 就將 buffer 內的資料以巨量傳輸的方式傳回主機。

每次完整的巨量傳輸資料，由 32 次計時中斷，將兩個聲道，每筆 1Byte 的語音訊號存至 buffer 內，因此傳輸的資料長度為 64Bytes，傳輸速率為

$$\frac{64}{\frac{1}{8K} * 32} = 16KB/s$$

3.3 電腦端軟體實現

3.3.1 軟體流程

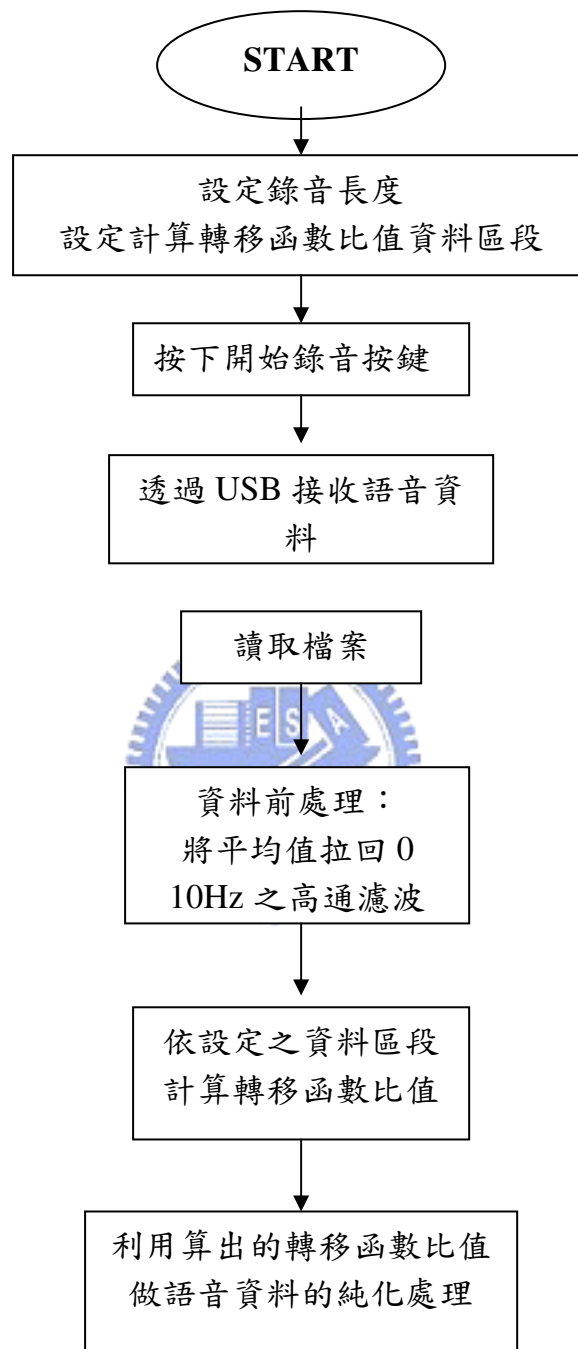


圖 3-12 軟體實現流程圖

3.3.2 使用流程

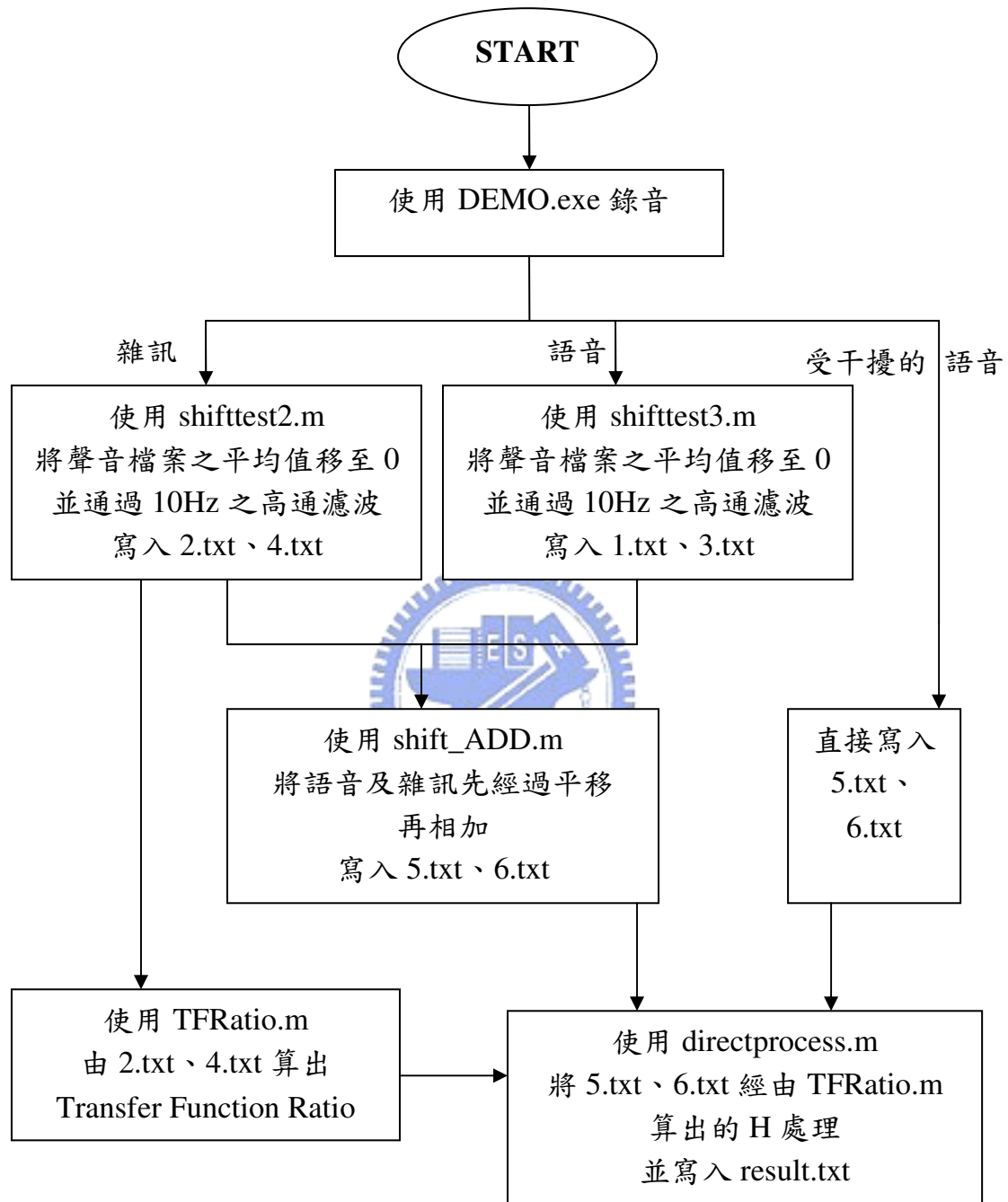


圖 3-13 使用流程圖

第四章 實驗結果與結論

在室內環境下，測試雜訊大小、訊號雜訊比、雜訊種類與聲源分離效果的关系：

4.1 在不同訊號雜訊比下的實驗結果分析

為了探討在不同的訊噪比及不同的雜訊大小對實驗效果的影響，以下在噪音極大、噪音較小、噪音極小(各相差約 5dB)三種狀況，針對各種不同的 SNR 來做實驗分析：

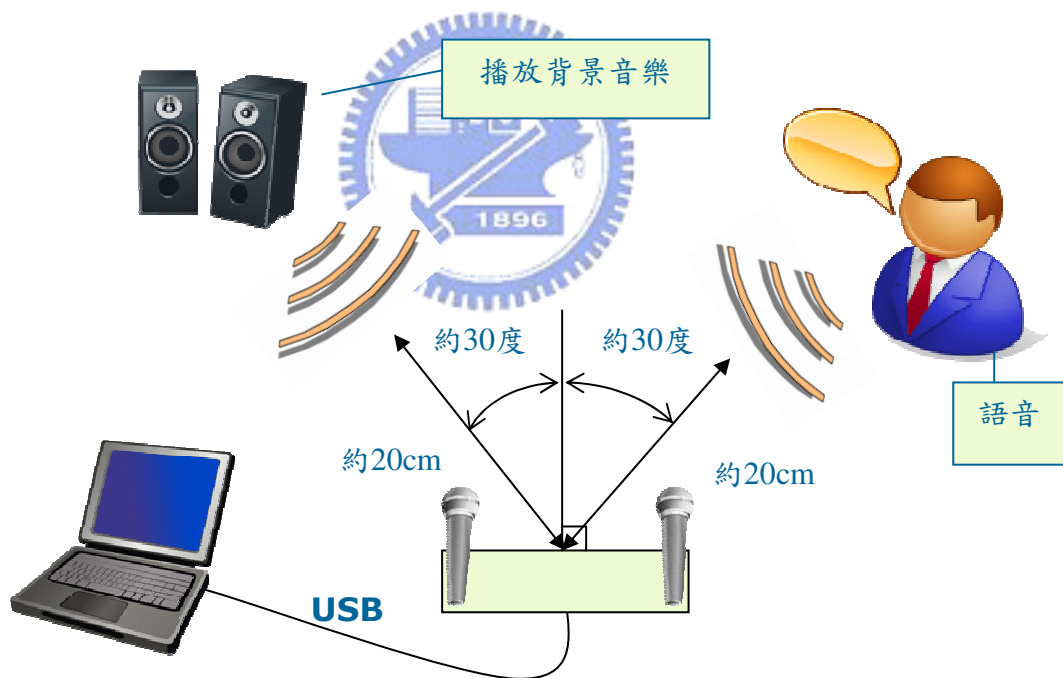


圖 4-1 SNR 比與雜訊大小對純化效果影響測試實驗環境示意圖

如上圖，在距離雙聲道麥克風約 20 公分處的左右各 30 度位置分別放上喇叭(噪音源)以及發出使用者語音，測試各種情況下的聲源分離效果，下頁表 4-1 為背景音樂極大時下的實驗結果：

實驗 編號	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
1	-8.2715	13.8894	-3.3305	9.3672	-7.5172
2	-3.2715	13.5026	-4.3306	8.3006	-6.6462

表 4-1 背景音樂極大時的實驗結果

表 4-1 為雜訊極大，約-18dB 時的實驗結果，此數位聲音訊號之能量大小是以 8-bit 之最大振幅 0~256 為 0dB 計算之。表格中使用兩個測量法：NR(noise reduction)以及 SNR(signal to noise ratio)來測量實驗的效果，其定義如下：

$$NR \triangleq \frac{\sum_{t \in T_n} z_1^2(t)}{\sum_{t \in T_n} (K \cdot y(t))^2} \quad SNR \triangleq \frac{\sum_{t \in T_s} z_{1,s}^2(t)}{\sum_{t \in T_s} (z_{1,s}(t) - K \cdot y(t))^2}$$

其中 T_n 為語音訊號不存在，只有背景雜音存在的時間， $z_1(t)$ 表示第一個麥克風所錄到的訊號， $\sum_{t \in T_n} z_1^2(t)$ 即代表第一個麥克風所錄到的背景雜音之能量， $y(t)$ 表示系統輸出的結果， $\sum_{t \in T_n} (K \cdot y(t))^2$ 即代表無語音訊號輸入下，系統對只有雜訊輸入的處理結果之能量，其中 K 為一個常數倍，代表系統對欲留下的語音訊號的增益值的倒數。

T_s 為語音訊號存在的時間， $z_{1,s}(t)$ 代表第一個麥克風所錄到的語音訊號， $\sum_{t \in T_s} z_{1,s}^2(t)$ 即代表第一個麥克風所錄到的語音訊號之能量， $\sum_{t \in T_s} (z_{1,s}(t) - K \cdot y(t))^2$ 即代表語音訊號減去 K 倍的系統輸出，即為剩下的雜訊的能量。

表 4-1 中列出了兩種不同 SNR 的輸入訊號下之程式執行結果，並一併列出 Dahl 之結果以作為對照，下圖 4-2 為實驗 1 之實際波形圖，圖中可以看出，兩種方法都可以將受到雜訊嚴重干擾的語音訊號做純化的動作，但 TFR(transfer function ratio)的方法在抑制雜訊的方面似乎效果較好。再觀察表 4-1 可以證明此一說法，TFR 的方法在 NR(noise reduction)上，效果比 Dahl 好上很多，無論輸入的 SNR 高低，皆可維持約 13dB 左右的 NR 值；在 SNR 的表現上，雖然訊號會略為失真，但因其優異的雜訊抑制能力，使得 TFR 在 SNR 的表現仍然比 Dahl 來得出色。

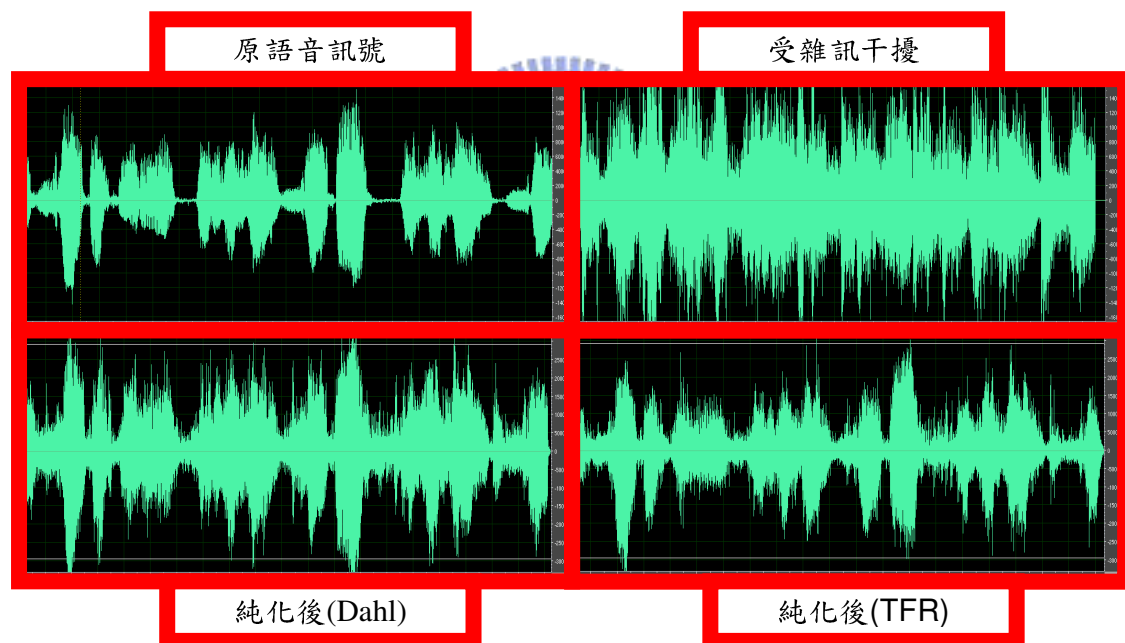


圖 4-2 純化結果比較圖(背景音樂極大)

實驗編號	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
3	-8.2671	13.8239	-5.0285	9.3798	-7.5028
4	-3.2672	13.4725	-4.3226	8.3288	-6.6131
5	1.7327	13.4756	-3.7099	7.4792	-5.8671

表 4-2 背景音樂較小時的實驗結果

上表 4-2 為背景音樂較小，約為-23dB 時的實驗結果，前面提到噪音極大的情況下，因為噪音音量已經接近錄音設備之上限，所以沒有做 SNR 為正的情況的實驗，這邊針對 SNR 約為-8.2dB、-3.2dB、+1.7dB 三種情況下討論其實驗結果：由表 4-2 可以再次看出，無論輸入訊號之 SNR 值為正值或負值，TFR 的方法都可以維持在約 13dB 的 NR 值，持續展現其優異的雜訊抑制能力，相對的 Dahl 就在輸入 SNR 值較大時，也就是輸入訊號中雜訊就比語音訊號相對較小的時候，抑制雜訊的能力就稍微比較差了，在 NR 上和 TFR 有約 6dB 的差距。

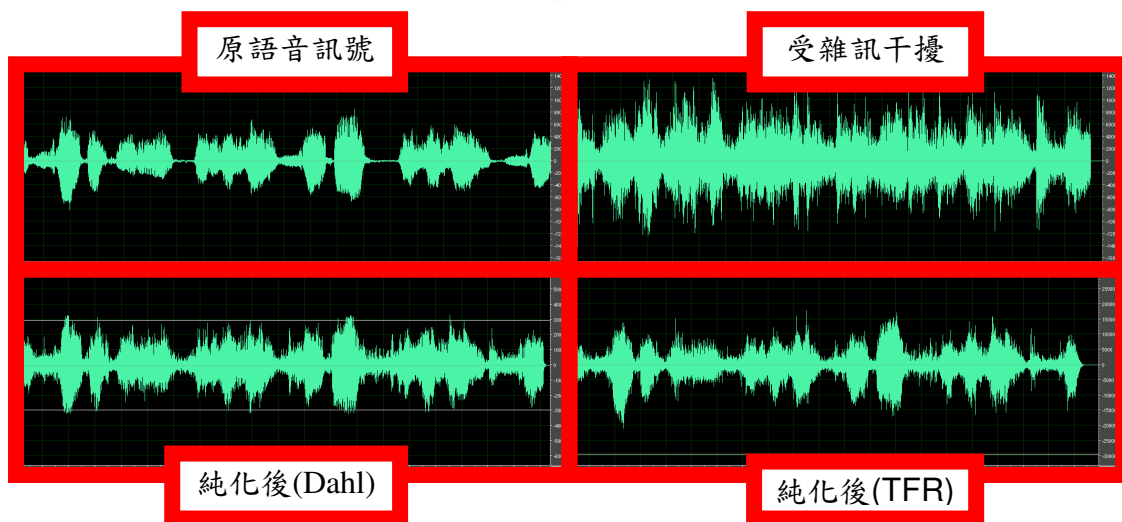


圖 4-3 純化結果比較圖(背景音樂較小)

實驗編號	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
6	-8.3086	13.4675	-5.1736	9.2411	-7.5826
7	-3.3088	13.2194	-4.3895	8.1388	-6.7312
8	1.6911	13.2576	-3.7462	7.3884	-5.9206

表 4-3 背景音樂極小時的實驗結果

上表 4-3 為背景音樂極小，約為-28dB 時的實驗數據，綜合觀察表 4-1、表 4-2、表 4-3，TFR 的方法無論背景音樂大小及輸入的 SNR 值，在雜訊抑制，也就是 NR 上都有 13dB 以上的能力，雖然因為還是會對語音訊號多少造成一點失真的情形，所以在雜訊相對於語音訊號較小，也就是輸入 SNR 值較大，約為-3.3dB 及 1.7dB 時在 SNR 方面會有反而降低的情形，但因為仍然可以大幅抑制背景雜訊，所以在 SNR 的表現上仍然是比 Dahl 要好一些的。下圖 4-4 為表 4-3 中實驗 6 的實際聲音波形圖。

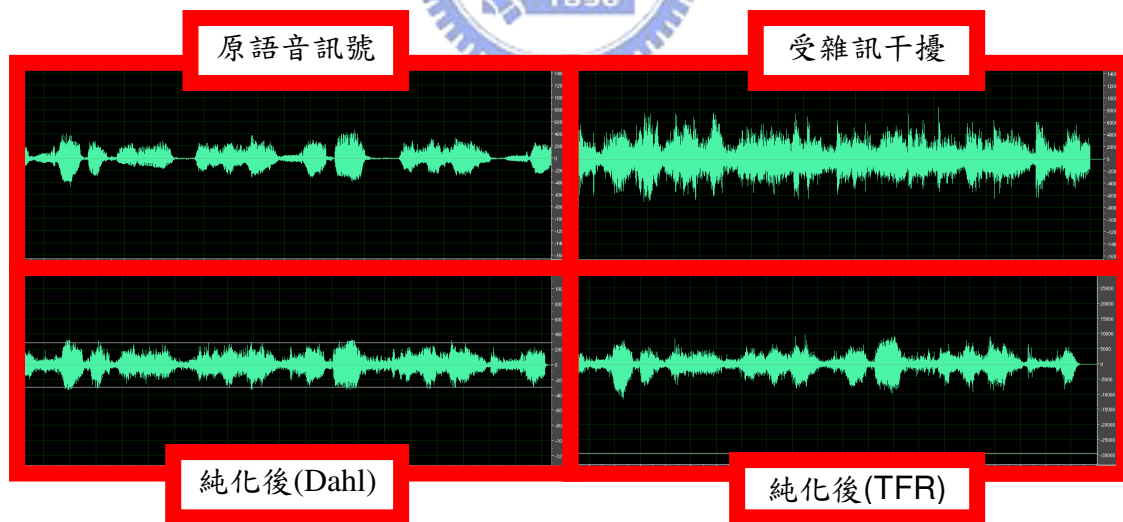


圖 4-4 純化結果比較圖(背景音樂極小)

4.2 不同雜訊種類下的實驗結果分析

4.2.1 雜訊來源為非使用者的語音

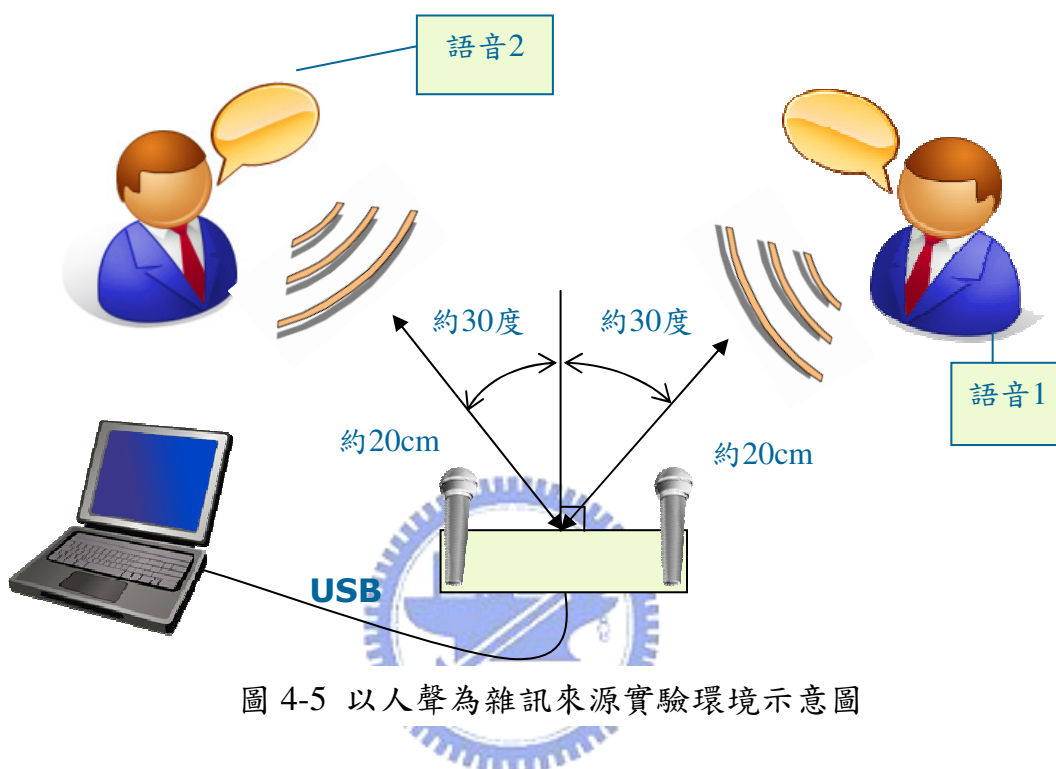


圖 4-5 以人聲為雜訊來源實驗環境示意圖

如上圖，將雜訊來源換成另一個非使用者之語音輸入，並測試聲源分離的效果是否會因而降低。

下頁圖 4-6 為此實驗之實際聲音檔案波形圖，圖 4-6 中左上圖為要留下的使用者語音，右上圖則是要抑制的語音 2，圖中可以看出，TFR 及 Dahl 兩種方法皆欲將右上的語音 2 抑制而留下左上的語音 1，但 TFR 的效果較為明顯，再觀察接下來的表 4-4，在干擾源為較單純的語音 2 下，TFR 的 NR 效果又比干擾源為背景音樂更加明顯，達到約 17dB，皆比 Dahl 好了約 8dB 以上。

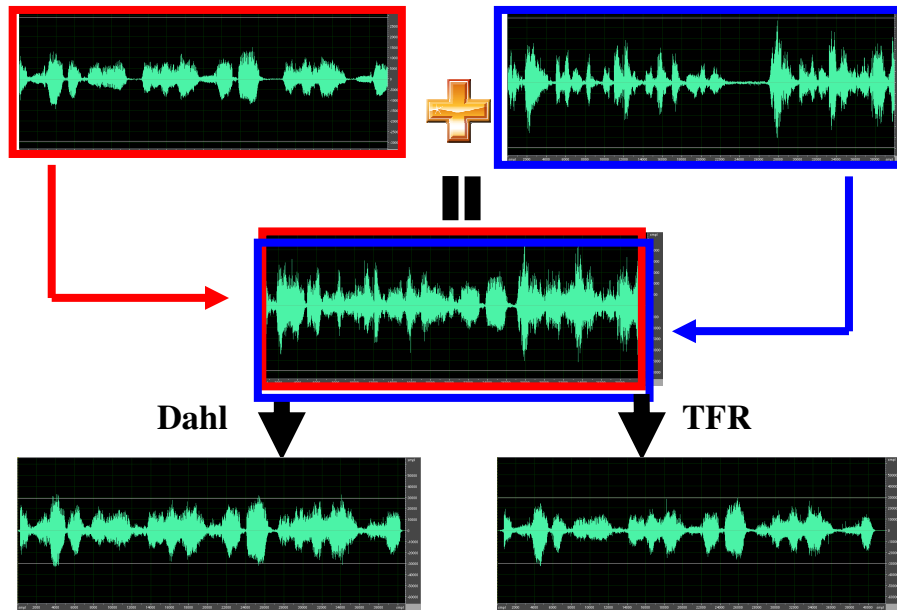


圖 4-6 干擾源為人聲之純化結果比較圖

實驗 編號	noise energy	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
1	-18dB	-3.9798	17.1904	-3.8278	9.5149	-6.3004
2		1.0201	17.191	-3.361	8.4397	-5.8099
3	-23dB	-3.9797	17.19	-3.828	9.517	-6.2992
4		1.0202	17.1893	-3.3619	8.4402	-5.8094
5		6.0201	17.1901	-3.0447	7.8476	-5.2886
6	-28dB	-3.9796	17.1855	-3.8295	9.5159	-6.2987
7		1.0202	17.1872	-3.3624	8.4387	-5.8105
8		6.0201	17.1868	-3.0457	7.8484	-5.2875

表 4-4 干擾源為人聲之實驗結果

4.2.2 雜訊來源為單一頻率的噪音

將干擾源換成單一頻率的雜訊，測試本系統之聲源分離效果是否會因而降低。在選擇頻率時，首先先觀察語音資料之頻譜圖，觀察語音訊號主要是由哪些頻率組成的：

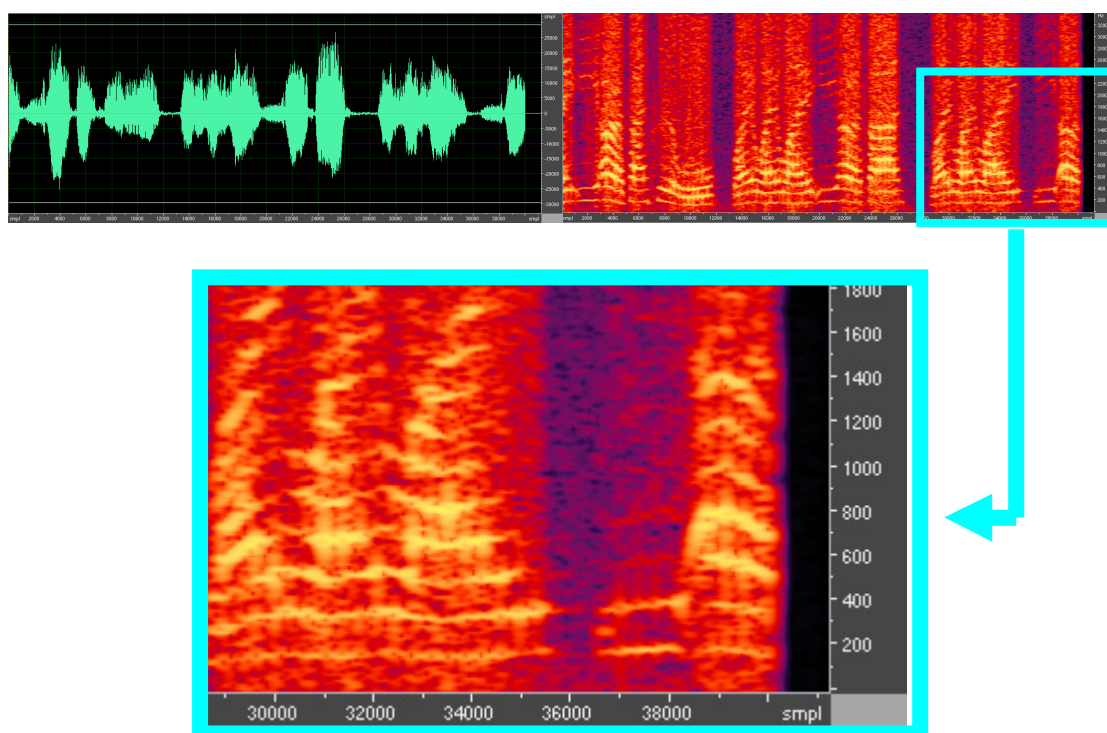


圖 4-7 語音訊號之波型圖、頻譜圖及頻譜圖之局部放大圖

圖 4-8 為語音訊號的頻譜圖以及局部放大圖，圖中可以看出，語音訊號在 300、500、700 等頻率都有一條聲紋，大於 1500 左右的頻率就沒有明顯的聲紋出現。於是我們選擇 300、500 這兩個和語音資料重疊的頻率，以及 2200 這個和語音資料不同的頻率來做干擾源，測試系統對這幾種干擾源的抑制能力。下頁表 4-5 是干擾源為 300Hz 及 500Hz 的單頻聲音之實驗結果：

實驗編號	noise frequency	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
1	300Hz	-8.8678	17.4477	-4.3897	17.001	-5.1809
2		-3.8679	17.4394	-3.723	16.2011	-4.7243
3		1.132	17.4353	-3.3007	15.3527	-4.3892
4	500Hz	-8.6766	15.0818	-4.9736	15.579	-5.678
5		-3.6767	14.8133	-4.4089	14.4926	-5.2184
6		1.3232	14.6594	-4.0404	13.4481	-4.8934

表 4-5 干擾源分別為 300Hz、500Hz 之單頻訊號的實驗結果

上表 4-5 即為分別以 300Hz、500Hz 之單頻訊號為雜訊之實驗結果數據，由表中可以發現，無論是 TFR 或 Dahl 哪種方法，用 NR 或 SNR 來測量效能，對 500Hz 的雜訊源的抑制效果都比對 300Hz 的雜訊之抑制效果稍差。推測應是因為語音訊號應該落在 500Hz 附近，再對照圖 4-7，500Hz 附近的語音聲紋的確比 300Hz 附近的聲紋明顯。下圖 4-8 為表 4-5 中實驗 1 的實驗波型圖。

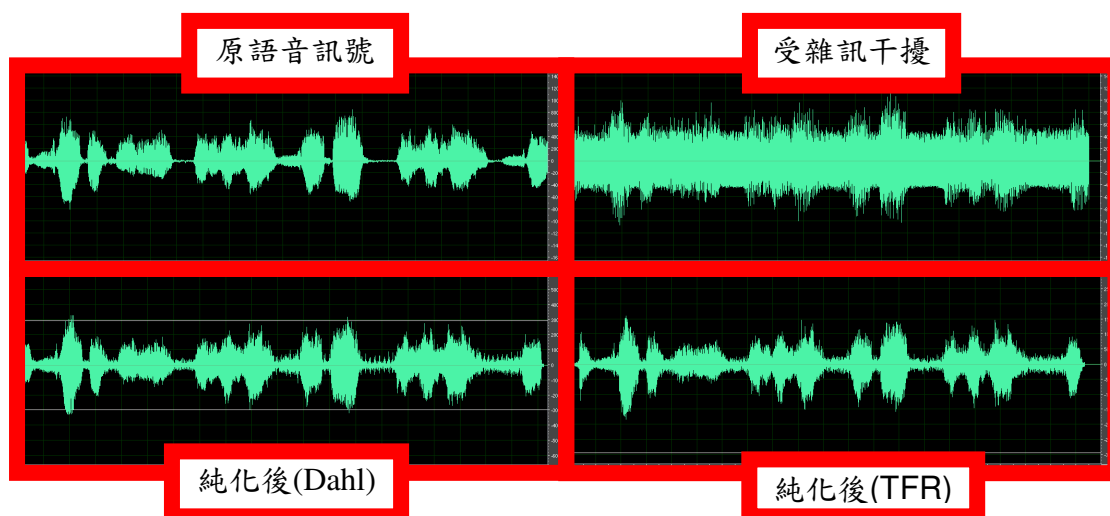


圖 4-8 以 300Hz 之單頻訊號為雜訊源之純化結果比較圖

在圖 4-7 中我們發現在頻率超過 1500Hz 之後，語音訊號就沒有明顯的聲紋了，所以我們取一個和語音訊號沒有重疊的頻率的單頻訊號來當干擾源，測試聲源分離的效果，表 4-7 為使用 2200Hz 之單頻訊號作為干擾源的實驗結果：

實驗編號	noise frequency	input SNR	TFR NR	TFR SNR	Dahl NR	Dahl SNR
7	2200Hz	-8.9297	23.7141	-4.63	30.1413	-4.614
8		-3.9298	23.634	-4.3514	29.1045	-4.4015
9		1.0701	23.5915	-4.1962	27.6524	-4.3234

表 4-6 干擾源為單一頻率 2200Hz 時的實驗結果

由表 4-6 及圖 4-9 中皆可以明顯發現，當干擾源為語音訊號並未包含的 2200Hz 之單頻訊號時，聲源分離的效果比干擾源為語音訊號包含的 300Hz、500Hz 好很多，約可以增加 5dB 以上的 NR 值。甚至在以 2200Hz 之單頻訊號做為干擾源時，出現了首次 Dahl 的效能在 NR 及 SNR 都比 TFR 來的優秀的情形，這表示在干擾源為單純訊號，且頻域和語音訊號較無重疊的時候，Dahl 的效能會明顯變好；換句話說，就是 TFR 即使在干擾源為複雜的背景音樂，頻域和欲留下的語音訊號有重疊，或是輸入之 SNR 較大等等情況，效能都不會有太明顯的降低，仍然可以抑制雜訊達 13dB 以上的 NR 值。代表此方法確實可以在各種情況之下，有效的達到抑制雜訊，並做到聲源分離的目標。

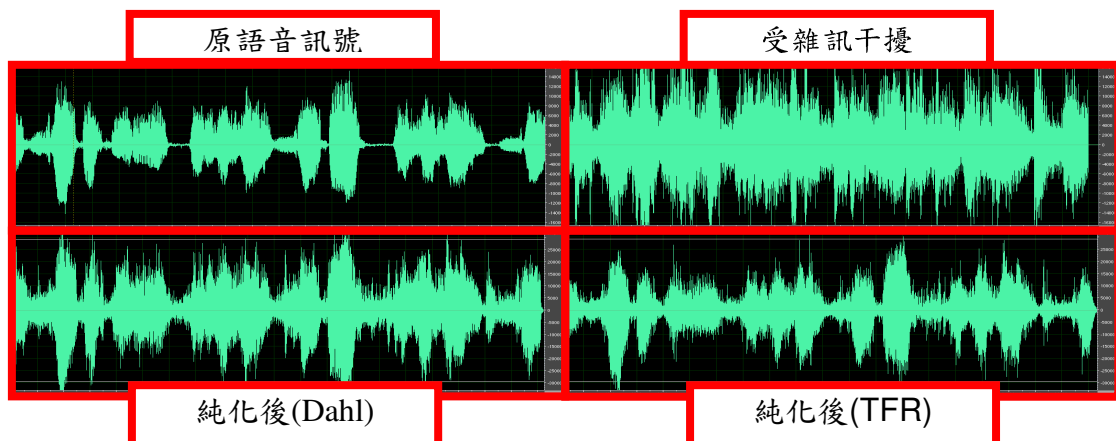


圖 4-9 干擾源為單一頻率 2200Hz 時的純化結果比較圖



4.3 結論與未來展望

本論文已實作完成利用轉移函數比值的方法實現雙聲道麥克風之聲源分離系統。經實驗證實，本系統在不同雜訊種類、不同 SNR、不同頻率下，都可以有效的抑制雜訊達 10dB 以上，強化語者語音資訊，提升語音品質。

本系統仍有許多可以再發展的空間，目前電腦端演算法的實現是先收集一段資料後將其寫入檔案，再經過讀取檔案算出轉移函數比值，最後做語音純化的動作。系統可以經過程式上的修改加強，並配合真人語音活動偵測(VAD, Voice Activity Detection)，做到即時性的聲源分離及語音純化，希望能藉由此平台建立一個具有即時聲源分離功能的麥克風模組，應用在網路即時通訊之回音消除，吵雜環境之語音純化等方面。



參考文獻

- [1] David Burshtein, Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech,2001.
- [2] 黃啟揚, 以雙核心平台實現即時影音追蹤與語音純化系統, 交大碩士論文,2008.
- [3] 康創閔, 應用於個人電腦環境之即時語音純化系統設計, 交大碩士論文,2004.
- [4] 楊佳興, 使用麥克風陣列實現即時語音純化與真人語音活動偵測系統, 交大碩士論文,2005.
- [5] Microchip, PIC18F4550 Data Sheet.
- [6] 劉維瀚, 具多重限制之最小平方法空間濾波器設計, 交大碩士論文,2002.
- [7] 葉順智, 具備 USB 介面之雙聲道全數位式音頻放大器設計, 交大碩士論文,2004.

