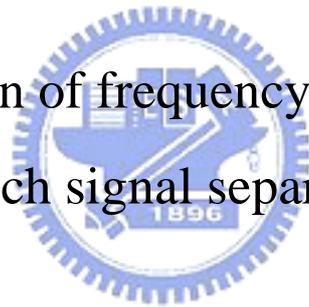


國立交通大學
電機與控制工程研究所

碩士論文

頻域獨立成分分析法於語音訊號分離之研究

An investigation of frequency domain ICA for
speech signal separation



研究生：連憶如

指導教授：胡竹生博士

中華民國九十三年七月

頻域獨立成分分析法於語音訊號分離之研究
An investigation of frequency domain ICA for
speech signal separation

研究生：連憶如

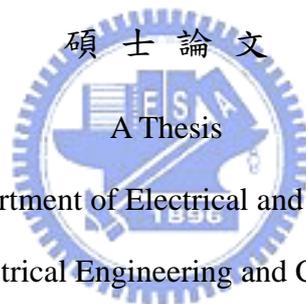
Student: Yi-Ru Lian

指導教授：胡竹生

Advisor: Dr. Jwu-Shen Hu

國立交通大學

電機與控制工程研究所



Submitted to Department 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

July 2004

Hsinchu, Taiwan, Republic of China

中華民國九十三年七月

頻域獨立成分分析法於語音訊號分離之研究

研究生：連憶如

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



在本論文中，針對由聲源與 channel 捲積(convolute)而來之混和語音訊號在頻域利用獨立成分分析法做訊號分離。將對於頻域獨立成分分析法所延伸出之兩大問題- permutation problem 與 dilation problem 加以探討。由於混和訊號需滿足複數之 FastICA 演算法對於其訊號源之假設，而一般語音訊號無法符合其假設，因此設計出一組適合 FastICA 分離之訊號源使得 FastICA 能夠成功地完成訊號分離以找出其對應之 demixing matrix，而達成訊號分離之目的。除此之外，當訊號源為一固定噪音與一移動語音時，亦可利用 noise 之 channel 效應不變之特性將語音分離出來。

An investigation of frequency domain ICA for speech signal separation

Student : Yi-Ru Lian Advisor : Dr. Jwu-Shen Hu

Institute of Electrical and Control Engineering
National Chiao-Tung University



Abstract

In this thesis, we use the frequency domain independent component analysis to separate the speech signals convoluted by the sources and the channels. We will discuss two well-known problems in frequency domain ICA – permutation problem and dilation problem. It is necessary for the sources of mixed signals to satisfy the assumptions of FastICA to separate the mixtures well. However, most speech signals do not satisfy the assumptions. Therefore, a design of the source signals satisfying the assumptions in order to use the FastICA separation is proposed. As a result, we can obtain the correct demixing matrix to separate the speech signals. Secondly, we can also extract the speech when the speech is not in a fixed location because of the invariant of the channel of the noise.

目錄

摘要.....	ii
Abstract.....	iii
目錄.....	iv
圖目錄.....	vi
第一章 序論.....	1
1.1 研究動機.....	1
1.2 章節概要.....	2
第二章 獨立成分分析法(Independent Component Analysis ,ICA)之原理.....	3
2.1 問題描述.....	4
2.2 ICA 之限制.....	5
2.2.1 對聲源之限制.....	5
2.2.2 對感測器的限制.....	7
2.2.3 不確定性.....	7
2.3 ICA 之前處理.....	9
2.3.1 Centering.....	9
2.3.2 Whitening.....	9
2.4 ICA 之演算法.....	11
第三章 頻域獨立分析法 (Frequency domain ICA).....	19
3.1 頻域獨立分析法之架構.....	20
3.1.1 STFT.....	21
3.1.2 針對複數之 FastICA.....	22
3.1.3 dilation problem 與 permutation problem.....	29
3.2 針對 dilation 與 permutation problem 之演算法.....	29
3.3 討論.....	35
第四章 以 ICA 估計 demixing matrix.....	41
4.1 利用預先估計之 demixing matrix 執行聲音分離.....	41
4.1.1 演算法架構.....	41
4.1.2 training signal 之設計.....	42
4.1.3 permutation 與 dilation problem 之處理.....	46
4.1.4 分離結果.....	46
4.2 利用空間預先估計之 demixing matrix 萃取移動中音源.....	53
4.2.1 原理.....	53
4.2.2 分離結果.....	53
第五章 結論與未來工作.....	56
5.1 結論.....	56

5.2 未來工作.....	56
參考文獻.....	57



圖目錄

圖 1- 1 MIMO System	1
圖 2- 1 BSS 問題之圖示	4
圖 2- 2 super-Gaussian 之統計分佈圖	7
圖 2- 3 ICA 演算法之結構	12
圖 2- 4 subgaussian 之統計分佈圖	14
圖 2- 5 聲源訊號	18
圖 3- 1 frames 之重疊關係	21
圖 3- 2 分別對每個頻率做 ICA 之圖示	22
圖 3- 3 G 之函數圖形	23
圖 3- 4 用以測試 FastICA 之五個訊號源	27
圖 3- 5 混合之訊號	27
圖 3- 6 目標函數之收斂	28
圖 3- 7 分離之結果	28
圖 3- 8 dilation and permutation problem	29
圖 3- 9 microphone array 與 source 之關係	30
圖 3- 10 directivity pattern	32
圖 3- 11 模擬位置圖	36
圖 3- 12 聲源訊號 s_1	36
圖 3- 13 聲源訊號 s_2	36
圖 3- 14 混和訊號 x_1	36
圖 3- 15 混和訊號 x_2	36
圖 3- 16 經過 FastICA 分離後之訊號	37
圖 3- 17 directive pattern	38
圖 3- 18 在低頻與高頻之 directivity pattern	38
圖 3- 19 分離結果之 envelop	39
圖 3- 20 s_1 與 s_2 之 spectrogram	40
圖 3- 21 以 Noboru Murata 方法之分離結果	40
圖 4- 1 演算法架構	41
圖 4- 2 FastICA 可分離之訊號	43
圖 4- 3 語音之 spectrum	43
圖 4- 4 產生 training signal 之示意圖	44
圖 4- 5 training signal	44
圖 4- 6 聲源訊號 s_1	47
圖 4- 7 聲源訊號 s_2	47
圖 4- 8 混和訊號 x_1	47

圖 4- 9 混和訊號 x_2	48
圖 4- 10 分離結果 y_1 , SNR=12.7863 db	48
圖 4- 11 分離結果 y_2 , SNR=20.8392 db	48
圖 4- 12 聲源訊號 s_1	49
圖 4- 13 聲源訊號 s_2	49
圖 4- 14 混和訊號 x_1	49
圖 4- 15 混和訊號 x_2	50
圖 4- 16 分離結果 y_1 , SNR=13.3790 db	50
圖 4- 17 分離結果 y_2 , SNR=18.9306 db	50
圖 4- 18 聲源訊號 s_1	51
圖 4- 19 雜訊 s_2	51
圖 4- 20 混和訊號 x_1	51
圖 4- 21 混和訊號 x_2	52
圖 4- 22 分離訊號 y_1 , SNR=18.1304 db	52
圖 4- 23 分離訊號 y_2 , SNR=21.1537 db	52
圖 4- 24 聲源與麥克風關係圖	54
圖 4- 25 混和訊號 x_1	54
圖 4- 26 混和訊號 x_2	54
圖 4- 27 分離訊號 y_1	55
圖 4- 28 分離訊號 y_2	55
圖 4- 29 SNR 與 s_1 角度之關係	55

第一章 序論

1.1 研究動機

訊號分離為將 source 訊號從混和訊號分離出來的一項技術，此技術在語音辨識(speech recognition)、行動通訊、圖形識別(pattern recognition)等領域皆有其應用。

對於語音而言，聲源訊號與麥克風所接收到的混和訊號之關係可視為一 MIMO(Multiple Input Multiple Output) System。為便於描述，在此討論 two input two output system



圖 1-1 MIMO System

接收到的訊號 $X(t)$ 實際上是聲源訊號 $S(t)$ 經過空間之 channel $A(t)$ (mixing filter matrix) 之 convolution 效應(可描述聲源訊號之延遲(delay)、衰減及反射之效應)而產生的結果

$$X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} * \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} = A(t) * S(t) \quad (1.1)$$

若欲將聲源訊號由混和訊號還原出來，則需找出可將此 channel 效應消除之 demixing filter matrix。而對於這種混和效應為 convolution 之問題，可利用 STFT(Short-Time Fourier Transform)將混和訊號轉至 frequency domain 使得 convolution 效應變為較簡化之相乘關係

$$X(w) = A(w)S(w) \quad (1.2)$$

於是 demixing matrix 之估測是在 frequency domain 針對每個頻率所進行的。針對

此一問題，有一較直覺之做法是一次只發一個聲源，利用接收到的訊號與原訊號相除即可得知此聲源與麥克風間的 channel 關係，然而這種看似簡單的方法並不容易實行。第一，做 STFT 時接收到的訊號要與聲源訊號之起始時間點要相同，否則所取之 frame 無法正確對應，於是所計算出來的 demixing matrix 並不正確，而在實作上，要對兩訊號取一模一樣之起始點並不容易。第二，此方法對於 channel 之找尋一次只能針對一個聲源，若有兩個聲源則需要重複相同的動作尋找兩次。

ICA (Independent Component Analysis) 的方法就是在訊號源彼此互相獨立的假設之下，去找一個 demixing matrix 來消除 mixing process 而分離出彼此互相獨立的訊號。由於 ICA 找尋 demixing matrix 時並不需要訊號源之資訊且不需要將聲源訊號分別發聲，因此並沒有上述兩大問題，然而在 frequency domain ICA 中，延伸出來了 permutation problem 以及 dilation problem 且由於一般之語音訊號無法在每個頻率之下皆滿足 FastICA 之假設，故分離效果並不理想，於是本論文設計了適合 frequency domain ICA 分離之訊號來找尋正確之 demixing matrix 以達成訊號分離之目的。

1.2 章節概要

本論文的組織架構簡述如下

第二章：對於 ICA 的原理、使用上的限制、及其演算法的推導做一個詳細的說明。

第三章：針對 frequency domain ICA 之由來及架構，以及其延伸而出之問題做說明並且對於已有之解決方法加以探討。

第四章：設計出一個適合 frequency domain ICA 分離之訊號以找出正確之 demixing matrix，並且探討 demixing matrix 對於移動中聲源之抽離效果。

第五章：對於本論文做一個總結。

第二章 獨立成分分析法(Independent Component Analysis ,ICA)之原理

Independent Component Analysis (以下簡稱ICA)是一種利用統計和計算的方法，在多變數的資料中找出其中的獨立成分，此方法最早由 Ans、Herault 與 Jutten 在 1983 年提出[1]，一開始是用以解神經生理學方面之問題[2]，而在訊號處理方面較早的研究是 blind signal deconvolution 之相關問題[3]，然而一直到 90 年代中期，ICA 的研究成果仍然受到侷限，當時所提出之演算法通常只能在某些限制之下使用。直到 1995 年 A.J. Bell 與 T.J. Sejnowski 提出基於 informax principle[4][5]之方法之後才受到重視，而 1997 年 Aapo Hyvärinen 提出 FastICA[6][7]之後，由於其計算效率高，ICA 才開始大規模的被應用在各個領域。

在許多應用中(例如：資料分析、語音分離[8]，醫學上之血流參數分析[9]....等等)，我們也許會量測到一些資料(mixed data)，而將其中的獨立元素抽取出來是必須的。例如在語音辨識中，也許利用麥克風量測到好幾個聲源所組合起來的聲音，而語音辨識卻只能對乾淨的語音訊號做辨識，因此做辨識之前必須將麥克風所收到的聲音中獨立的聲源抽取出來。這種語音訊號的訊號分離(Blind Signal Separation)就是有名的雞尾酒派對問題(cocktail party problem)，所謂雞尾酒派對問題就是在一個空間中，同時有很多聲音一起出現，可能有很多人在講話，甚至夾雜背景音樂，雖然在一個吵雜的環境下，我們仍然能專心的聽到想聽的聲音，並且與人交談。這個機制牽涉到大腦的構造與神經的傳導，但在訊號處理的領域中也是一個有趣的問題。而 ICA 對於這樣的一個問題，在訊號源是獨立的前提下提供了一個解決的方式。

2.1 問題描述

ICA 所處理的問題：Blind Source Separation 的問題如圖一，為便於描述，在此僅考慮兩個聲源訊號與兩個麥克風之情形，這兩個聲源經過空間中混合 (mix) 過程之後，由麥克風接收到混合之訊號，在此問題中，我們對聲源與混合過程皆無法事先知道，可擁有的資訊只有麥克風所量到的訊號。

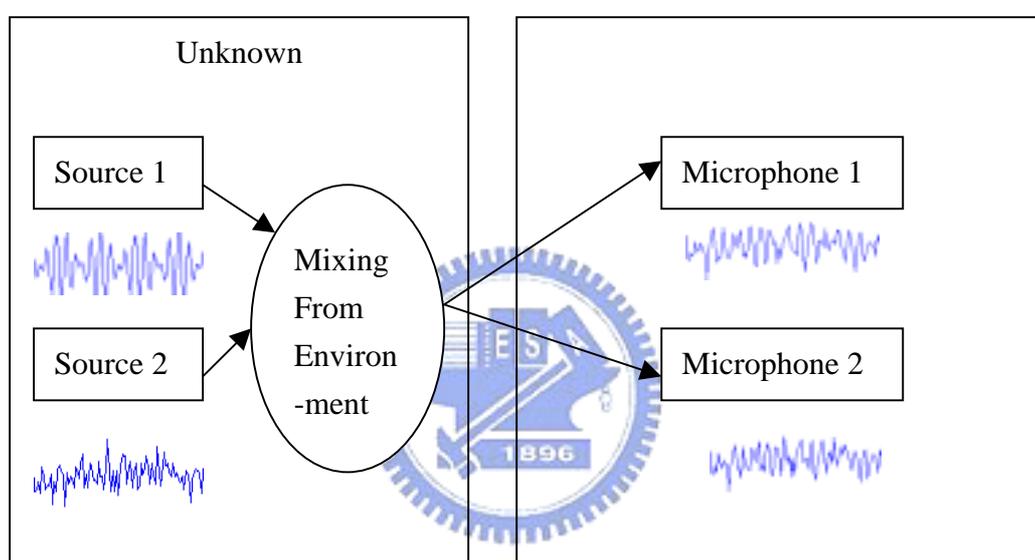


圖 2-1 BSS 問題之圖示

在這些量測到的混和訊號當中，包含了一些來自聲源的資訊，而所量測的訊號之個數與聲源之個數也許有很多個。可將量測的訊號以 $x_j(t)$ 來表示，其中 $j = 1, \dots, m$ ，也就是有 m 個量測到的訊號；而聲源訊號以 $s_i(t)$ 來表示，其中 $i = 1, \dots, n$ ，有 n 個聲源訊號。在大部分的情況下為了簡化問題，聲源與所量測到的訊號間的關係，只考慮線性組合，如(2.1)， a_{ij} 代表第 j 個聲源到第 i 個麥克風的放大倍數。其中，混合矩陣 A (mixing matrix) 與聲源訊號 S 皆未知。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = AS \quad (2.1)$$

因此，解 Blind Signal Separation 也就相當於去找一個矩陣 W (demixing matrix)，使得接收到的訊號經過 W 的轉換能夠得到原來的聲源訊號。

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \approx S \quad (2.2)$$

至於如何決定這個 W 則可以利用訊號源與訊號源之間獨立的關係。由於訊號源彼此是獨立的，所以接收到的訊號經過 W 的轉換，如果有將混合的效應消除，則轉換出來的 Y 應該是彼此獨立的，也就是原來彼此獨立的訊號源，這就是 ICA 最基本的觀念。

2.2 ICA 之限制

ICA 可以用以解決 Blind Signal Separation 的問題，但 ICA 的演算法本身有一些限制，在使用之前，必須確定問題本身符合這些限制[10]。

2.2.1 對聲源之限制

1. 聲源本身在統計上獨立(independent)

由於 ICA 的演算法就是用聲源彼此獨立的特性來發展出來的，因此所有的聲源必須滿足彼此獨立的特性，而在大部分訊號分離的問題中，聲源也的確是彼此獨立的。

在統計上，若是隨機變數(random variables) s_1, s_2, \dots, s_n 是彼此獨立的，則其 joint probability density function 可以拆為 marginal probability density function 的乘積。

$$p(s_1, s_2, \dots, s_n) = p_1(s_1)p_2(s_2)\dots p_n(s_n) \quad (2.3)$$

$p_i(s_i)$ 為 s_i 之 marginal probability density function

2.獨立元素必須為 Nongaussian distributions

由於 ICA 尋找獨立元素的方法來自中央極限定理(central limit theorem)，這個定理告訴我們：Nongaussian 分佈且互相獨立的隨機變數之和的分佈會比其中任意隨機變數的分佈更加接近 gaussian。因此，在尋找獨立元素時，可利用此一特性，讓找出來的隨機變數的分佈盡可能的不是 gaussian 分佈，當我們找到一組 W 可以使得計算出來的結果最不接近 gaussian 分佈時，也就找出原來的獨立聲源了。如 2.4 式，希望算出之 y_i 盡可能不是 gaussian 分佈。

$$y_i = \sum_{j=1}^n w_{ij} x_j \quad (2.4)$$

基於這個觀念，若是原來的聲源就是 gaussian 分佈，則不管 W 如何尋找，所計算出來的結果的分佈皆還是 gaussian 分佈，因此就無法找到讓結果最 nongaussian 的 W，也就無法找出原來的獨立聲源了。

雖然對於聲源有如此之限制，但由於語音本身並非 gaussian，因此 ICA 在實用上並不會因此受到侷限。語音訊號在統計上是 super-Gaussian 如圖所示(圖中灰線是 gaussian 分佈以做比較)，其分佈比 gaussian 較狹窄一些。

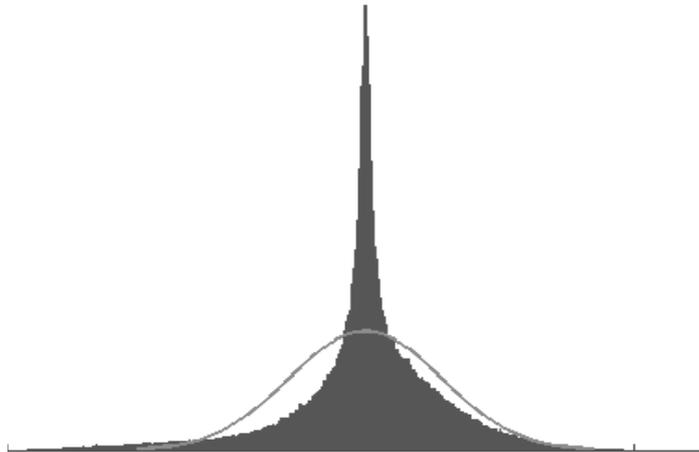


圖 2- 2 super-Gaussian 之統計分佈圖

2.2.2 對感測器的限制

考慮收到 m 個麥克風訊號 $X = AS$ ，其中 A 是 mixing matrix； S 是聲源訊號，有 n 個。以線性方程式的角度來看：若 $m=n$ ，則 $S = A^{-1}X$ ，有解；若 $m>n$ ，則此方程式為 over-determined linear equations，也就是方程式比未知數多，在這種情形之下，可先將維度(dimension)降回 n ，再執行 ICA；若是 $m<n$ ，方程式比未知數少，則缺乏足夠的資訊找到原來的獨立元素。

2.2.3 不確定性

在 ICA 的模型中 (2.1) 我們可以看出有以下兩點不確定性

a) 無法確定獨立元素的 variance：

第 j 個麥克風所接收到的訊號可以表示成

$$x_j = \sum_i a_{ji} s_i \quad (2.5)$$

其中由於 a 與 s 皆未知，若是 s_i 被放大 α_i 倍，則 a_{ji} 只要除以 α_i 就可以互相抵銷

$$x_j = \sum_i (a_{ji} / \alpha_i) (s_i \alpha_i) \quad (2.6)$$

因此假設聲源的 variance 假設為一。

$$E\{s_i^2\} = 1 \quad (2.7)$$

b) 無法決定獨立變數的順序：

這點不確定性也跟未知的 A 與 S 有關，由於未知的 A 與 S，我們無法定義哪一個訊號源是”第幾個”。因此，當我們找到 demixing matrix W 時，希望經過 W 轉換的結果完全等於原來的訊號是不可能的

$$Y = WAS = S \quad (\text{impossible}) \quad (2.8)$$



由於不知聲源有哪些，更不可能將解出的訊號與原來的聲源做對應而使得 $Y=S$

($Y = [y_1 y_2 \dots y_n]^T, S = [s_1 s_2 \dots s_n]^T$)。

綜合以上兩點不確定性，對於 Y 與 S，我們可以整理出一個比較合理的關係

$$Y = WAS = PDS \quad (2.9)$$

其中 P 是交換矩陣(permutation matrix)，在此交換矩陣裡的每一行(column)與每一列(row)只有一個 element 是 1，其餘皆為 0。D 是一個 diagonal matrix，其中只有對角線上有值，其餘皆為 0，而對角線上的值相當於分出來的結果與原來訊號之倍數。因此，分出來的 Y 應該是原來的 S 重新排過，且大小也改變了。

2.3 ICA 之前處理

一般來說，在 ICA 的演算法真正執行之前，為了簡化問題，會先做以下兩個前處理。

2.3.1 Centering

若是接收到的混合訊號與獨立聲源的平均為零(zero-mean)，則在演算法的推導過程中可以簡化許多；若不為零，則我們可將接收到的訊號減去其平均值，而使其平均為零。

$$x_{centered} = x - E\{x\} \quad (2.10)$$

雖然僅對接收到的訊號置中，其實也對聲源訊號做了 centering

$$S = A^{-1}X \Rightarrow E\{S\} = A^{-1}E\{X\} = A^{-1} \cdot 0 = 0 \quad (2.11)$$

2.3.2 Whitening

以統計的觀點來看，比 independent 弱一點的形式是 uncorrelatedness。因此，在做 ICA 之前，若先利用前處理將訊號變為 uncorrelated，可使得訊號更加接近 independent，而介於 independent 與 uncorrelated 之間的一個關係是 whiteness，其除了為 uncorrelated 之外，variance 的值等於一。

換句話說，若 Z (為行向量)是 white，其 covariance matrix 的值為單位矩陣。

$$E\{ZZ^T\} = I \quad (2.12)$$

whitening 的動作就是去找一個 whitening matrix V 而將接收到的訊號做線性轉換

$$Z=VX \quad (2.13)$$

較普遍的一個方法就是對訊號的 covariance matrix 做 eigenvalue decomposition (EVD)

$$E\{XX^T\} = EDE^T \quad (2.14)$$

其中 E 是由 eigenvector 所構成的正交矩陣且 $E^T = E^{-1}$ ， D 是由其相對應的 eigenvalue 所組成的對角線矩陣。whitening matrix 為

$$V = ED^{-1/2}E^T \quad (2.15)$$

可藉由檢查 covariance matrix 是否為單位矩陣以確定經過 V 轉換過之 Z 是否已為 white

$$\begin{aligned} E\{ZZ^T\} &= E\{VXX^TV^T\} \\ &= E\{ED^{-1/2}E^TXX^TED^{-1/2}E^T\} \\ &= ED^{-1/2}E^TE\{XX^T\}ED^{-1/2}E^T \\ &= ED^{-1/2}E^TEDE^TED^{-1/2}E^T \\ &= ED^{-1/2}E^{-1}EDE^{-1}ED^{-1/2}E^T \\ &= ED^{-1/2}DD^{-1/2}E^T \\ &= EE^T \\ &= I \end{aligned} \quad (2.16)$$

經過 whitening 後的訊號由(2.13)可再進一步表示為

$$Z = VX = VAS = \tilde{A}S \quad (2.17)$$

whitening 的動作，亦可視為將 mixing matrix A 做一個線性轉換，假設轉換過後的 mixing matrix 為 \tilde{A} 。若將 whitening 後的訊號之 covariance matrix 以(2.17)展開可得

$$E\{ZZ^T\} = E\{\tilde{A}SS^T\tilde{A}^T\} = \tilde{A}E\{SS^T\}\tilde{A}^T \quad (2.18)$$

從(2.7)與(2.11)式可得知聲源訊號的平均值等於零且 variance 為一，又聲源訊號

間彼此獨立因此

$$E\{SS^T\} = I \quad (2.19)$$

所以

$$E\{ZZ^T\} = \tilde{A}\tilde{A}^T = I \quad (2.20)$$

這也就表示經過 whitening matrix 轉換過後之 mixing matrix 是一個正交矩陣。這也就表示當我們在尋找 demixing matrix 時，只要尋找互相垂直的向量即可，這對整個演算法的簡化有相當大的幫助。

2.4 ICA 之演算法

在 ICA 的演算法中，通常會先定義一個目標函數(Objective function)，然後對此目標函數算最大值或最小值。而在目標函數的定義上，我們必須定義一個我們可以藉由對他算最大值或最小值而得到獨立元素(Independent Component)的函數。因此，我們可以將 ICA 演算法分為兩大部分

$$\text{ICA method} = \text{Objective Function} + \text{Optimization Algorithm} \quad (2.21)$$

而目標函數與最佳化演算法的選擇也決定了 ICA 演算法的性質，其中目標函數的選擇決定了 ICA 演算法是否強健(robustness)，而最佳化演算法的選擇與 ICA 演算法的收斂速度有關。

配合 2.3 所提到的前處理，可將 ICA 整體架構描述如圖 2-3。有 correlation 的混合訊號經過前處理後，訊號彼此為 uncorrelated，再透過目標函數之最佳化，即可得到獨立元素[11]。

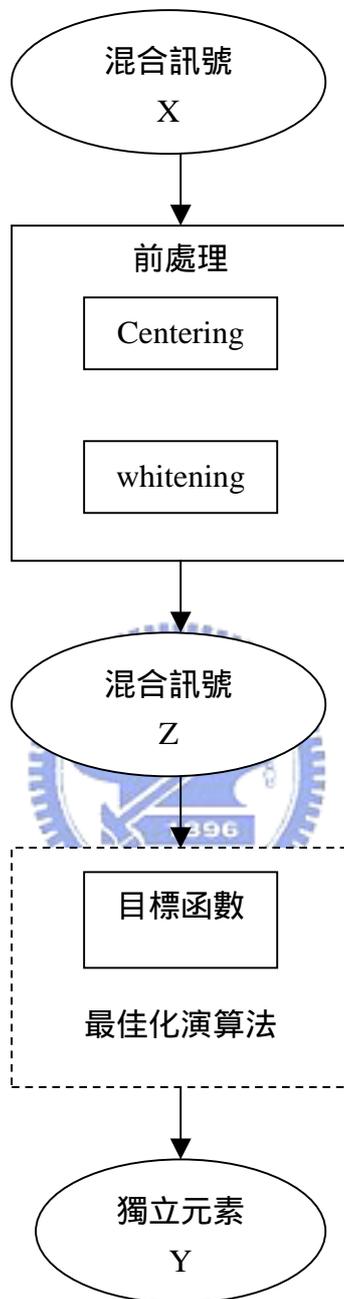


圖 2-3 ICA 演算法之結構

2.4.1 目標函數

如 2.2.1 所提到的，由於中央極限定理，我們可以藉由尋找一個使結果最不 gaussian 的 W (Maximization of Non-gaussianity) 來分出獨立元素。所以，若我們想要找到一個獨立元素 y 我們可以表示為

$$y = w^T z \quad (2.22)$$

其中 w 為行向量。由於我們想要尋找一個使結果最 nongaussian 的 W ，對於隨機變數是否 gaussian 就需要一個量化的量測，也就是去測量(2.22)中 y 的 gaussian 程度。而一般對於 gaussian 程度典型的量測就是 kurtosis 與 neg-entropy。

kurtosis

一隨機變數 y 之 kurtosis 表示為 $kurt(y)$ ，其定義如下

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (2.23)$$

若要簡化此式，我們可以假設 y 已經過調整使得其 variance 等於一，於是等式右邊可簡化為 $E\{y^4\} - 3$ 。對於 gaussian 的隨機變數 y ，其 $E\{y^4\}$ 等於 $3(E\{y^2\})^2$ ，所以 kurtosis 為零。Kurtosis 的值可能為正或為負，當一隨機變數之 kurtosis 為正，我們稱之為 supergaussian；若 kurtosis 為負，稱為 subgaussian[12]，其中一典型的例子為均勻分佈(uniform distribution)其分佈如圖所示(圖中灰線是 gaussian 分佈以做比較)

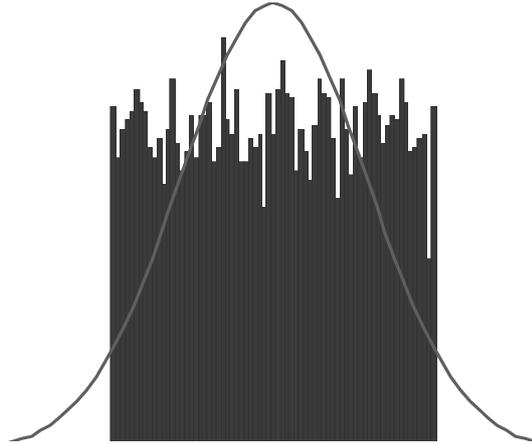


圖 2- 4 subgaussian 之統計分佈圖

因此我們可以利用 kurtosis 之絕對值或平方當作目標函數，此函數值越小代表越接近 gaussian，相反地若此函數值越大也就越 nongaussian。

Neg-entropy

雖然我們可以藉由 kurtosis 的計算得知隨機變數的 gaussian 程度，但是 kurtosis 是四次方的量測，在如此高次方的計算之下，所計算出的值很容易受到某些少數的值的影響，所以用 kurtosis 來量測 gaussian 程度並不穩健(robust)[13]。

另個測量 gaussian 的方法是 neg-entropy，這是由資訊理論(information theory)發展而來之方法。一個 probability density function 為 $p_y(\eta)$ 的隨機向量 y 之 entropy

H 定義如下

$$H(y) = -\int p_y(\eta) \log p_y(\eta) d\eta \quad (2.24)$$

而所有相同 variance 的隨機變數當中，gaussian 變數之 entropy 最大，這也就意味著 entropy 可以用來量測一隨機變數是否 gaussian。

爲了得到一個非 gaussian 之量測，我們希望對於一個 gaussian 的隨機變數，

其函數值為零，且對於所有隨機變數之函數值皆大於零，於是定義了 neg-entropy J ，對於隨機向量 y 之 neg-entropy 為

$$J(y) = H(y_{gauss}) - H(y) \quad (2.25)$$

其中 $y_{gaussian}$ 是與 y 有相同 covariance matrix 之 gaussian 隨機向量。由此式可看出當 y 為 gaussian，則函數值為零，又 gaussian 變數之 entropy 最大，所以對所有隨機變數函數值皆大於零。於是尋找 neg-entropy 之最大值即為尋找最 nongaussian 之隨機變數。

雖然 neg-entropy 可用以量測 gaussian 程度，但在計算上卻十分複雜，因此在實際應用上，需採用對於 neg-entropy 較簡化的近似[14]

$$J(y) \propto [E\{G(y)\} - E\{G(v)\}]^2 \quad (2.26)$$

其中 G 為 non-quadratic function，而 v 為 gaussian random variable。實際上，若 G 函數越平緩則可得到越 robust 之估計， G 函數有以下三種選擇：

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \quad (2.27)$$

$$G_2(y) = -\exp(-y^2 / 2) \quad (2.28)$$

$$G_3(y) = y^4 \quad (2.29)$$

其中 $1 \leq a_1 \leq 2$ 。

2.4.2 針對實數之 FastICA

FastICA 是由 Aapo Hyvärinen 所提出之演算法[15]。此演算法利用 neg-entropy 做為 non-gaussian 之量測，而欲量測之 $y = w^T z$ ，故 2.26 式可表為

$$J(w) \propto [E\{G(w^T z)\} - E\{G(v)\}]^2 \quad (2.30)$$

而 neg-entropy 之最大值可藉由尋找 $E\{G(w^T z)\}$ 之極值得到。根據 Lagrange conditions，在 $E\{(w^T z)^2\} = \|w\|^2 = 1$ 之 constrain 之下， $E\{G(w^T z)\}$ 發生極值之 w 須滿足

$$E\{zg(w^T z)\} + \beta w = 0 \quad (2.31)$$

其中 g 為 G 之微分，利用牛頓法可找出滿足(2.31)之解。將(2.31)左式表為 F ，其 gradient 為

$$\frac{\partial F}{\partial w} = E\{zz^T g'(w^T z)\} + \beta I \quad (2.32)$$

而對於 whitening 後之訊號，可做以下近似

$E\{zz^T g'(w^T z)\} \approx E\{zz^T\}E\{g'(w^T z)\} = E\{g'(w^T z)\}I$ ，因此根據牛頓法，可推出如下之 iteration

$$\begin{aligned} w &\leftarrow w - [E\{zg(w^T z)\} + \beta w] / [E\{g'(w^T z)\} + \beta] \\ w &\leftarrow w / \|w\| \end{aligned} \quad (2.33)$$

最後將等式乘上 $E\{g'(w^T z)\} + \beta$ 可得

$$\begin{aligned}
 w &\leftarrow E\{zg(w^T z)\} - E\{g'(w^T z)\}w \\
 w &\leftarrow w/\|w\|
 \end{aligned}
 \tag{2.34}$$

其收斂條件為 w 與 update 之前同方向。當有超過一個 independent component 時，在 2.3.2 曾提到過，經過 whitening 後，在尋找 demixing matrix 時，只要尋找互相垂直的向量即可，所以在找第 p 個 w 時，須將前 $p-1$ 個找出之 w 之方向減去

$$w_p \leftarrow w_p - \sum_{j=1}^{p-1} (w_p^T w_j) w_j
 \tag{2.35}$$

綜合以上所述，FastICA 演算法之流程如下：

1. Centering $\rightarrow x_c = x - E\{x\}$
2. Whitening $\rightarrow z$
3. 設定 counter $p=1$ ， m 為獨立元素之個數
4. 隨機選擇一初始之 w_p
5. $w_p \leftarrow E\{zg(w_p^T z)\} - E\{g'(w_p^T z)\}w_p$



$$w_p \leftarrow w_p / \|w_p\|$$

6. 減去先前已找到之方向

$$w_p = w_p - \sum_{j=1}^{p-1} w_j w_j^H w_p$$

$$w_p = \frac{w_p}{\|w_p\|}$$

7. 若 w_p 尚未收斂，回到第 5
8. 設定 $p=p+1$ 。若 $p < m$ ，回到第 4

以下利用 FastICA 分離兩個語音，聲源訊號為

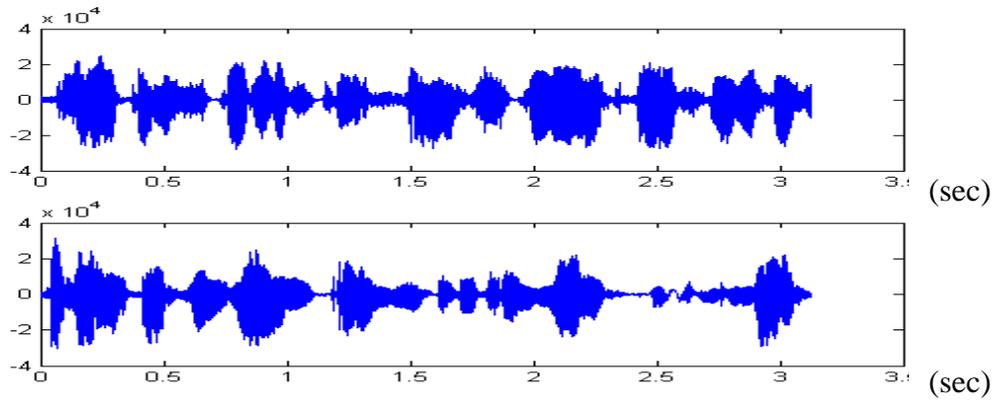
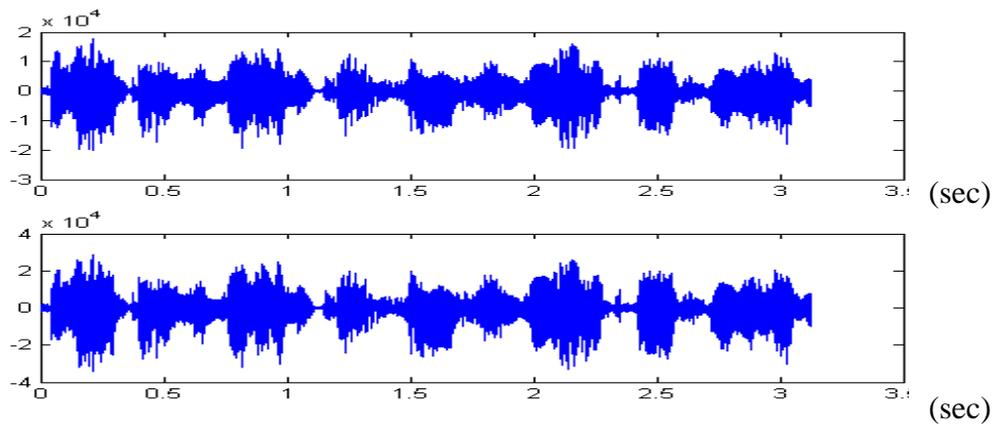
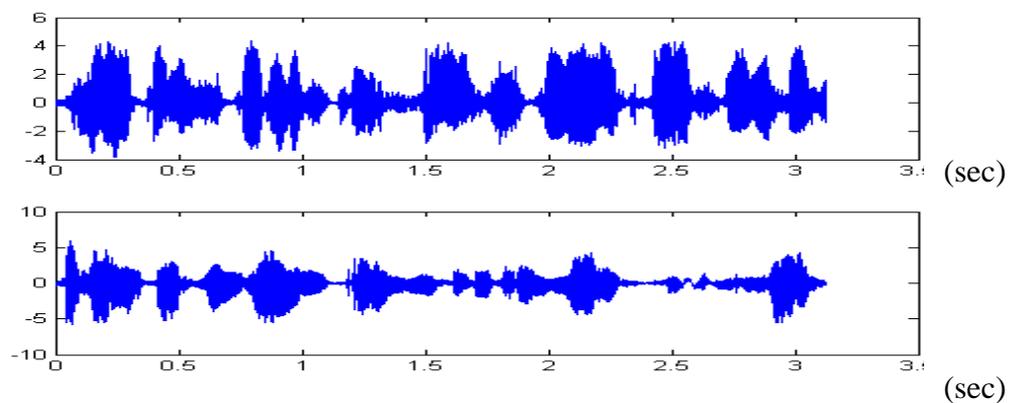


圖 2-5 聲源訊號

經過 mixing matrix $A = \begin{bmatrix} 0.5090 & 0.4047 \\ 0.9557 & 0.5525 \end{bmatrix}$ 之混和，混和訊號為



利用 FastICA 所分離出之結果為



由圖中可看出分離效果佳且 FastICA 收斂速度快，本次分離僅 20 個 steps 即可收斂。

第三章 頻域獨立分析法 (Frequency domain ICA)

雖然時域 FastICA 對於混合語音有不錯的分離效果，但卻只能處理 instantaneous mixture，也就是

1. mixing matrix 以實數表示
2. 對於實際上聲源對麥克風的 time-delay 與空間之反射沒有任何處理

由於聲源之 time-delay 與空間之反射，因此 instantaneous mixture 無法完整的描述聲源與麥克風之間的關係，所以對於真實之訊號此演算法無法真正執行訊號分離。

實際上由於空間效應，若有 n 個聲源，第 i 個麥克風所接收到的訊號應為


$$\begin{aligned}x_i(t) &= \sum_{j=1}^n \sum_{\tau} a_{ij}(\tau) s_j(t - \tau) \\ &= \sum_{j=1}^n a_{ij}(t) * s_j(t)\end{aligned}\quad (3.1)$$

其中 $*$ 代表 convolution，而 a_{ij} 為第 j 個聲源到第 i 個麥克風之 impulse response。

這個關係可以寫成矩陣形式如下

$$x(t) = A(t) * s(t) \quad (3.2)$$

其中 $x(t)$ 為由麥克風所收到的訊號所組成之向量， $s(t)$ 為聲源訊號所組成之向量，而 $A(t)$ 為 filter matrix

$$A(t) = \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{bmatrix} \quad (3.3)$$

比較(3.2)與(2.1)可看出其中差別在於 A 與 S 相乘或 convolution 的關係，若是將 (3.2)做 Fourier Transform

$$x(w) = A(w)s(w) \quad (3.4)$$

其中 $x(w)$ 、 $A(w)$ 、 $s(w)$ 分別為 $x(t)$ 、 $A(t)$ 、 $s(t)$ 之 Fourier Transform。

由於利用 Fourier Transform 可將 convolution 的問題轉回相乘的問題，於是 ICA 便可在頻域執行。本章首先將頻域 ICA 之演算法架構做詳細的解說，接著針對頻域 ICA 所延伸出之兩大問題-dilation problem 和 permutation problem 整理了兩個做法：第一個做法在 2000 年由 Satoshi KURITA[16]等人所提出，而第二個做法是在 1998 年由 Noboru Murata[17]所提出之方法，並且針對這兩個做法加以討論。

3.1 頻域獨立分析法之架構

由於整個演算法是在頻域做，所以一開始會先利用 Short-Time Fourier Transform(以下簡稱 STFT)將時域之訊號轉到頻域，接下來針對每個頻率做 ICA，由於轉換到頻域之資料為複數，因此使用之 FastICA 需針對複數處理。由於個別對每個頻率做 ICA 因而延伸出 Scaling 與 permutation 之問題，所以在將訊號做 Inverse Fourier Transform 轉回時域之前必須先處理這兩個問題。

3.1.1 STFT

首先，將接收到的訊號 $x_i(t)$ 切割成一段段 frames，而 frame 之間有 $2/3$ 的長度重疊，圖 3.1 描述出前三個 frame 的重疊關係。若訊號 $x(t)$ 可分割為 R 個 frames，每個 frame 前進了 T 個 samples，則第 r 個 frame 的值为

$x(t_r) = [0, \dots, 0, x(t_r + 1), x(t_r + 2), \dots, x(t_r + K), 0, \dots, 0]$ 其中 $tr = rT, r = 0, \dots, R - 1$ 。

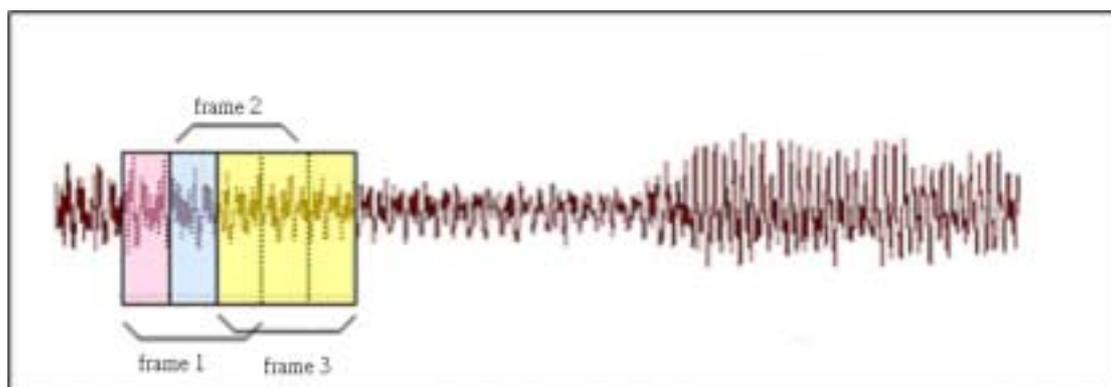


圖 3-1 frames 之重疊關係

對每個 frame 做 DFT 之後，接收之訊號與聲源訊號在第 r 個 frame 之關係為

$$\begin{bmatrix} x_1(w, t_r) \\ \vdots \\ x_n(w, t_r) \end{bmatrix} = \begin{bmatrix} a_{11}(w) & \dots & a_{1n}(w) \\ \vdots & \ddots & \vdots \\ a_{n1}(w) & \dots & a_{nn}(w) \end{bmatrix} \begin{bmatrix} s_1(w, t_r) \\ \vdots \\ s_n(w, t_r) \end{bmatrix} \quad (3.5)$$

於式中可看出對某個頻率 w 而言 mixing matrix 為 instantaneous mixture，而此形式為 FastICA 可處理的形式，因此頻域所執行之 FastICA 是對某個頻率所做的，要對每個頻率都做完 ICA 才能得到分離訊號，如圖 3-2 所示。與時域之形式的不同點在於：在時域所統計的資料為對時間之取樣點，而此處由(3.5)可看出統計之資料並非為對時間之取樣點而是每個 frame 轉至頻域後，在要做 ICA 之特定頻率下每個 frame 在此頻率之成分。

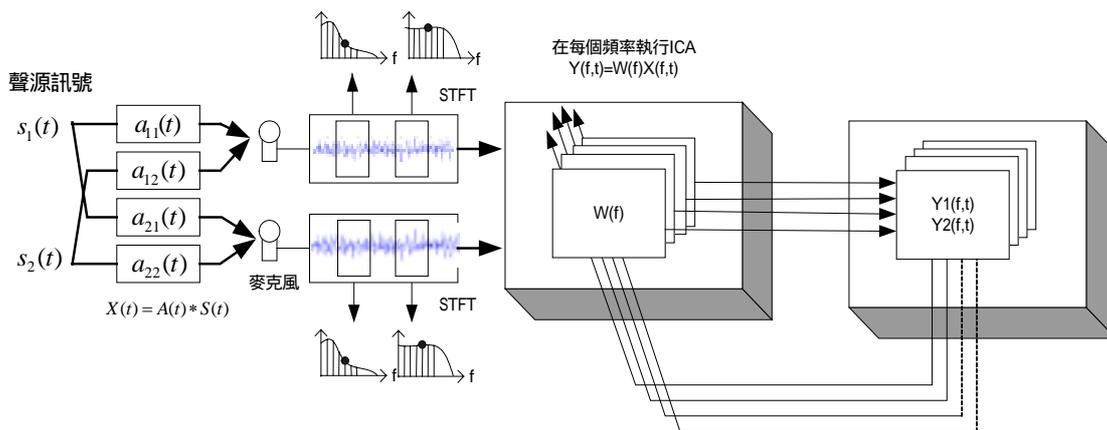


圖 3-2 分別對每個頻率做 ICA 之圖示

3.1.2 針對複數之 FastICA

針對複數之 FastICA 與實數之 FastICA 有相同的形式[18]，也是定義一個目標函數再找出其最佳解。其目標函數如下

$$J_G(w) = E\{G(|w^H x|^2)\} \quad (3.6)$$

其中 w 為 n 維之複數向量。若是 G 函數隨著變數之增加而增加的越緩慢，則演算法將越 robust。在此，有三個建議的 G 函數

$$\begin{aligned} G_1(y) &= \sqrt{a_1 + y} \\ G_2(y) &= \log(a_2 + y) \\ G_3(y) &= \frac{1}{2}y^2 \end{aligned} \quad (3.7)$$

其函數圖形如圖 3-3，其中可看出 G_1 與 G_2 變化較 G_3 緩慢，因此較 robust。

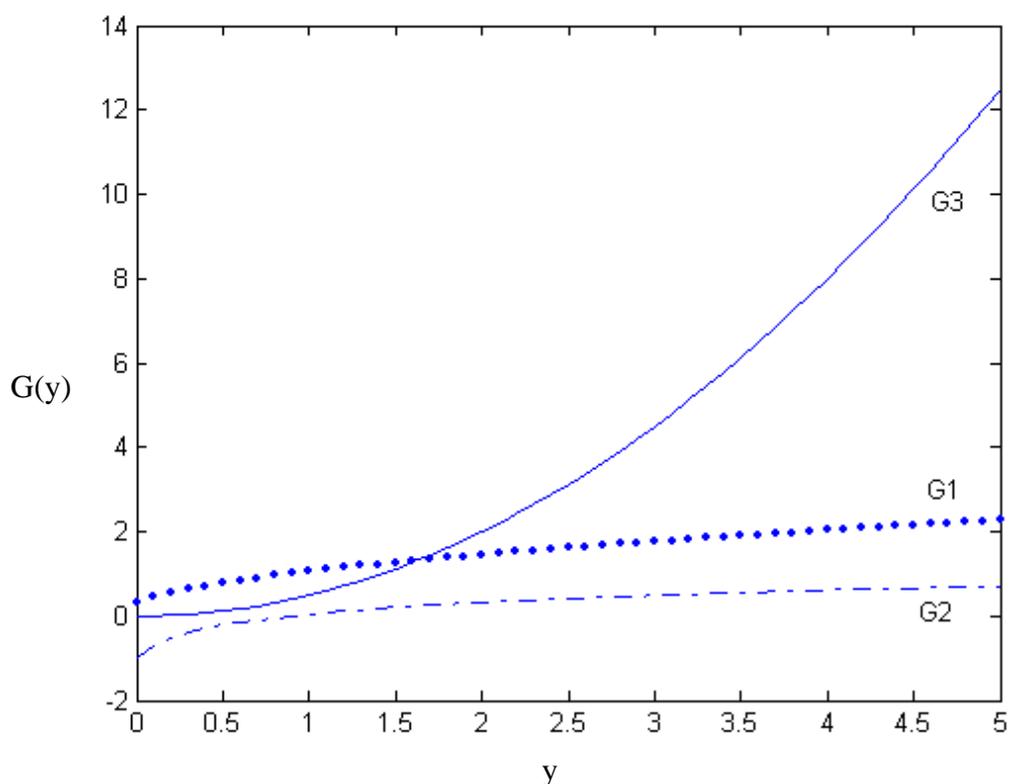


圖 3-3 G 之函數圖形

演算法之推導：

有了目標函數 J ，接下來需要找出其極值，這是一個最佳化問題：

$$\begin{aligned} & \text{optima } E\{G(|w^H x|^2)\} \\ & \text{s.t. } E\{|w^H x|^2\} = \|w\|^2 = 1 \end{aligned} \quad (3.8)$$

其中 $w = w_r + iw_i$ ， $x = x_r + ix_i$ 。在推導的過程中假設欲尋找之獨立變數之實部與虛部為 uncorrelated 且有相同之 variance 也就是 $E\{ss^T\} = o$ 。

根據 Kuhn-Tucker conditions $E\{G(|w^H x|^2)\}$ 在 $\|w\|^2 = 1$ 之限制條件底下之最佳解必須滿足

$$\nabla E\{G(|w^H x|^2)\} - \beta \nabla E\{|w^H x|^2\} = 0 \quad (3.9)$$

其中 β 為實數， ∇ 為 gradient。式中之 gradient 需分別對 w 之實部與虛部做計算，

第一項為

$$\nabla E\{G(|w^H x|^2)\} = E\{(\nabla |w^H x|^2)g(|w^H x|^2)\} \quad (3.10)$$

其中 $|w^H x|^2 = (w^H x)(w^H x)^*$ 、 g 為 G 之微分，又 $w = w_r + iw_i$

$$\begin{aligned} & |(w^H x)|^2 \\ &= (w_r^T - jw_i^T)x(w_r^T + jw_i^T)x^* \\ &= w_r^T x w_r^T x^* + jw_r^T x w_i^T x^* - jw_i^T x w_r^T x^* + w_i^T x w_i^T x^* \end{aligned} \quad (3.11)$$

所以

$$\begin{aligned} & \nabla |w^H x|^2 \\ &= \nabla_{w_r} |w^H x|^2 + \nabla_{w_i} |w^H x|^2 \\ &= \nabla_{w_r} (w_r^T x w_r^T x^* + jw_r^T x w_i^T x^* - jw_i^T x w_r^T x^* + w_i^T x w_i^T x^*) \\ & \quad + \nabla_{w_i} (w_r^T x w_r^T x^* + jw_r^T x w_i^T x^* - jw_i^T x w_r^T x^* + w_i^T x w_i^T x^*) \\ &= x w_r^T x^* + w_r^T x x^* + j(x w_i^T x^* - w_i^T x x^*) \\ & \quad + j[j(w_r^T x x^* - x w_r^T x^*) + x w_i^T x^* + w_i^T x x^*] \\ &= 2x w_r^T x^* + 2j x w_i^T x^* \\ &= 2x(w_r^T + jw_i^T)x^* \\ &= 2x w^T x^* = 2x(w^H x)^* \end{aligned} \quad (3.12)$$

而(3.9)中第二項為

$$\begin{aligned} & E\{|w^H x|^2\} \\ &= E\{(w^H x)(w^H x)^H\} \\ &= E\{w^H x x^H w\} \\ &= w^H E\{x x^H\} w \quad (\text{for white } x, E\{x x^H\} = I) \\ &= w^H w \end{aligned} \quad (3.13)$$

因此其 gradient 為

$$\nabla E\{|w^H x|^2\} = \nabla(w^H w) = 2w \quad (3.14)$$

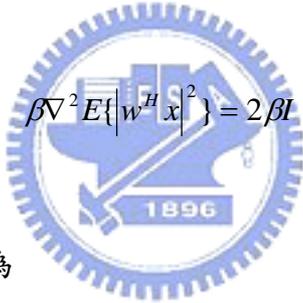
由(3.10)、(3.12)、(3.14)，可將(3.9)化簡為

$$E\{2x(w^H x)^* g(|w^H x|^2)\} - 2\beta w = 0 \quad (3.15)$$

接下來，利用牛頓法求出滿足(3.9)之 roots，因此需分別求出 $\nabla E\{G(|w^H x|^2)\}$ 與 $\beta \nabla E\{|w^H x|^2\}$ 之 Jacobian matrix

$$\begin{aligned} & \nabla^2 E\{G(|w^H x|^2)\} \\ &= E\{(\nabla^2 |w^H x|^2)g(|w^H x|^2) + (\nabla |w^H x|^2)(\nabla |w^H x|^2)^T g'(|w^H x|^2)\} \quad (3.16) \\ &\approx 2E\{g(|w^H x|^2) + |w^H x|^2 g'(|w^H x|^2)\}I \end{aligned}$$

其中之近似使用了 $E\{xx^T\} = 0$ (可由 $E\{ss^T\} = 0$ 推導出)，而 $\beta \nabla E\{|w^H x|^2\}$ 之 Jacobian matrix 為



$$\beta \nabla^2 E\{|w^H x|^2\} = 2\beta I \quad (3.17)$$

因此(3.9)之 Jacobian matrix 為

$$2E\{g(|w^H x|^2) + |w^H x|^2 g'(|w^H x|^2)\} - \beta I \quad (3.18)$$

於是 Newton iteration 為

$$\begin{aligned} w^+ &= w - \frac{E\{x(w^H x)^* g(|w^H x|^2)\} - \beta w}{E\{g(|w^H x|^2) + |w^H x|^2 g'(|w^H x|^2)\} - \beta} , \\ w_{new} &= \frac{w^+}{\|w^+\|} \end{aligned} \quad (3.19)$$

最後，在方程式兩邊乘以 $E\{g(|w^H x|^2) + |w^H x|^2 g'(|w^H x|^2)\} - \beta$ 可得

$$\begin{aligned} w^+ &= E\{x(w^H x)^* g(|w^H x|^2)\} - E\{g(|w^H x|^2) + |w^H x|^2 g'(|w^H x|^2)\}w \\ w_{new} &= \frac{w^+}{\|w^+\|} \end{aligned} \quad (3.20)$$

演算法之流程：

此複數 FastICA 之流程與實數 FastICA 之流程相似，如下

1. Centering $\rightarrow x_c = x - E\{x\}$
2. Whitening $\rightarrow z$
3. 設定 counter $p=1$ ， m 為獨立元素之個數
4. 隨機選擇一初始之 w_p
5. $w_p = E\{x(w_p^H z)^* g(|w_p^H z|^2)\} - E\{g(|w_p^H z|^2) + |w_p^H z|^2 g'(|w_p^H z|^2)\} w_p$ ，

$$w_p = \frac{w_p}{\|w_p\|}$$

6. 減去先前已找到之方向

$$w_p = w_p - \sum_{j=1}^{p-1} w_j w_j^H w_p$$

$$w_p = \frac{w_p}{\|w_p\|}$$



7. 若 w_p 尚未收斂，回到第 5
8. 設定 $p=p+1$ 。若 $p < m$ ，回到第 4

模擬

測試之訊號中第 j 個訊號源為 $s_j = r_j(\cos \phi_j + i \sin \phi_j)$ ，其中不同的 r_j 有不同的分佈，而 ϕ_j 是從 $-\pi$ 到 π 之 uniform distribution，以此方式產生之訊號源可滿足 FastICA 之限制：獨立變數之實部與虛部為 uncorrelated 且有相同之 variance。在本模擬中，以上述方式產生五個訊號源，如圖 3-4，圖中左邊五個為訊號源 $s_1 \sim s_5$ 右邊為其分別對應之分佈統計圖。將產生出之訊號源通過一個隨機產生複數之 mixing matrix 可得混合之訊號，如圖 3-5。圖 3-6 為 FastICA 目標函數之收斂，

圖中可看出收斂僅需二到三個 step。圖 3-7 為經過 FastICA 分離出之結果，分離效果佳。

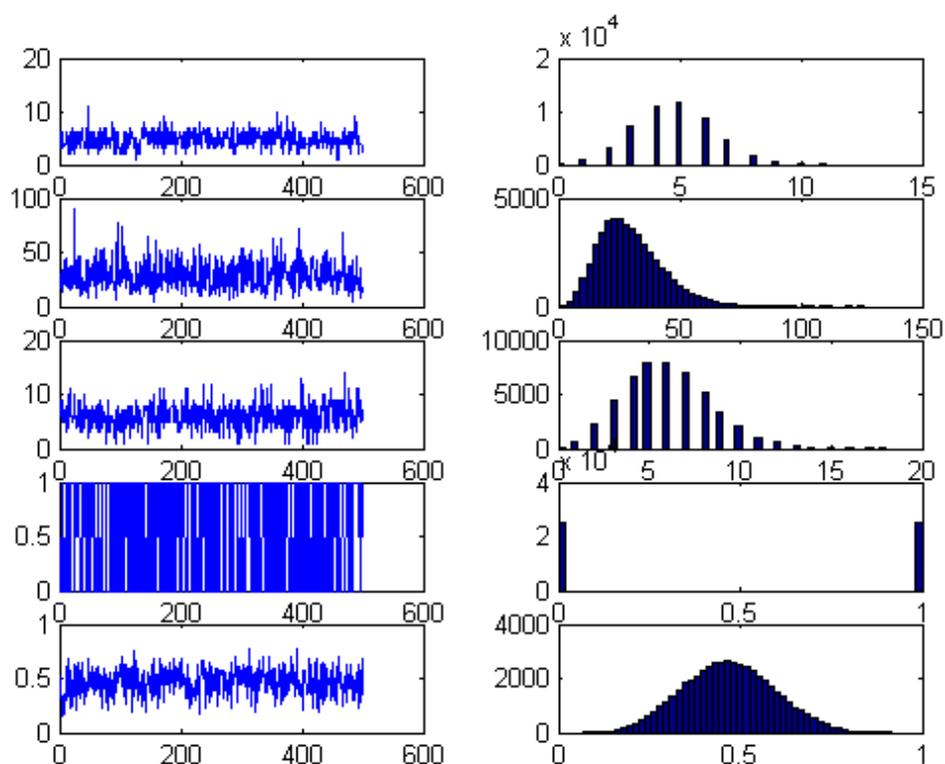


圖 3-4 用以測試 FastICA 之五個訊號源

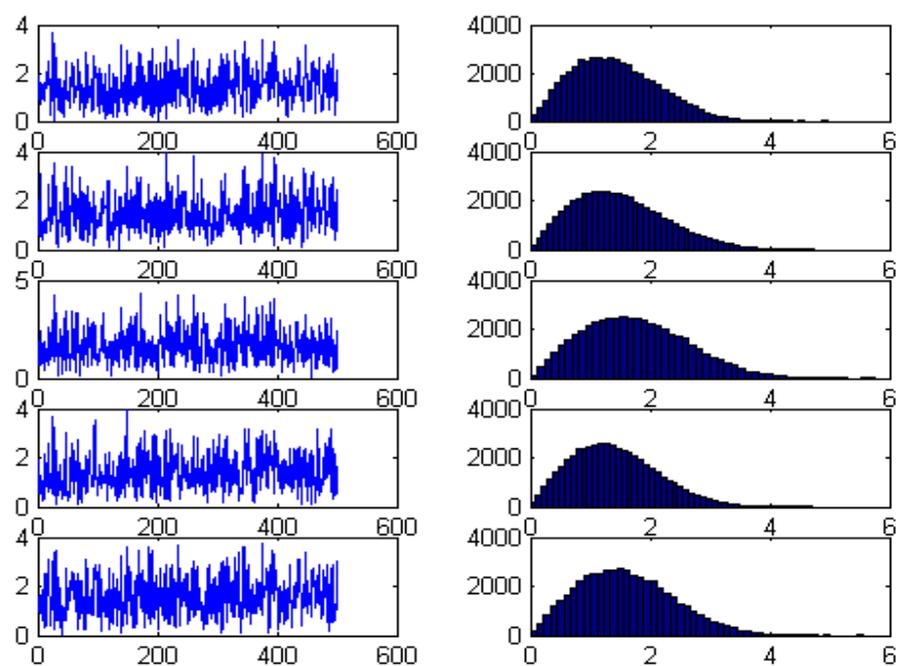


圖 3-5 混合之訊號

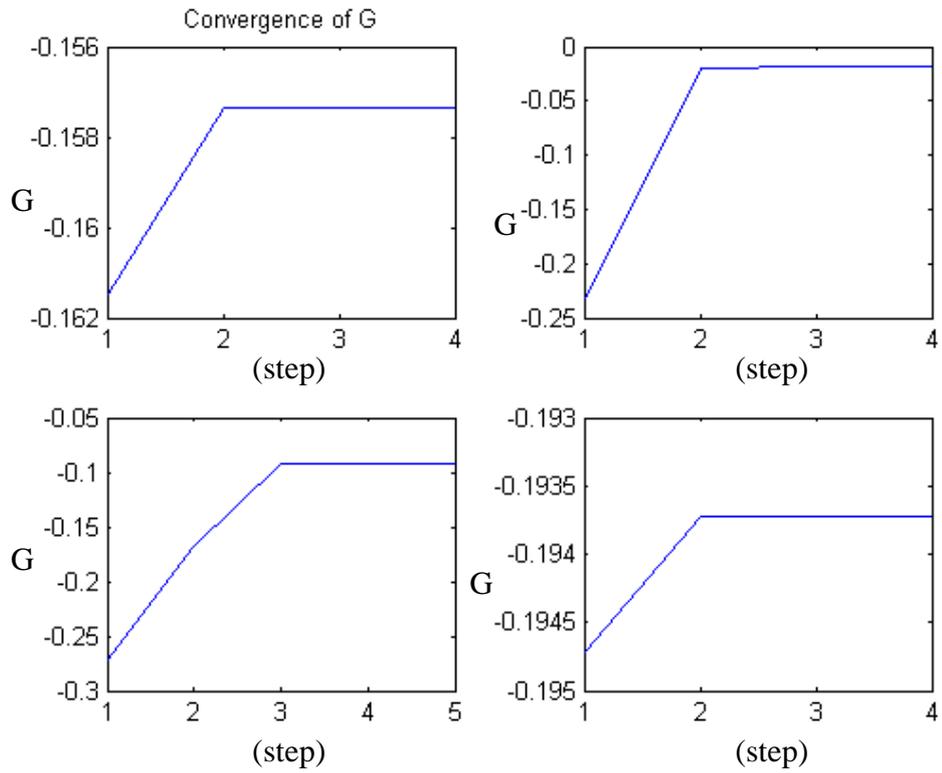


圖 3-6 目標函數之收斂

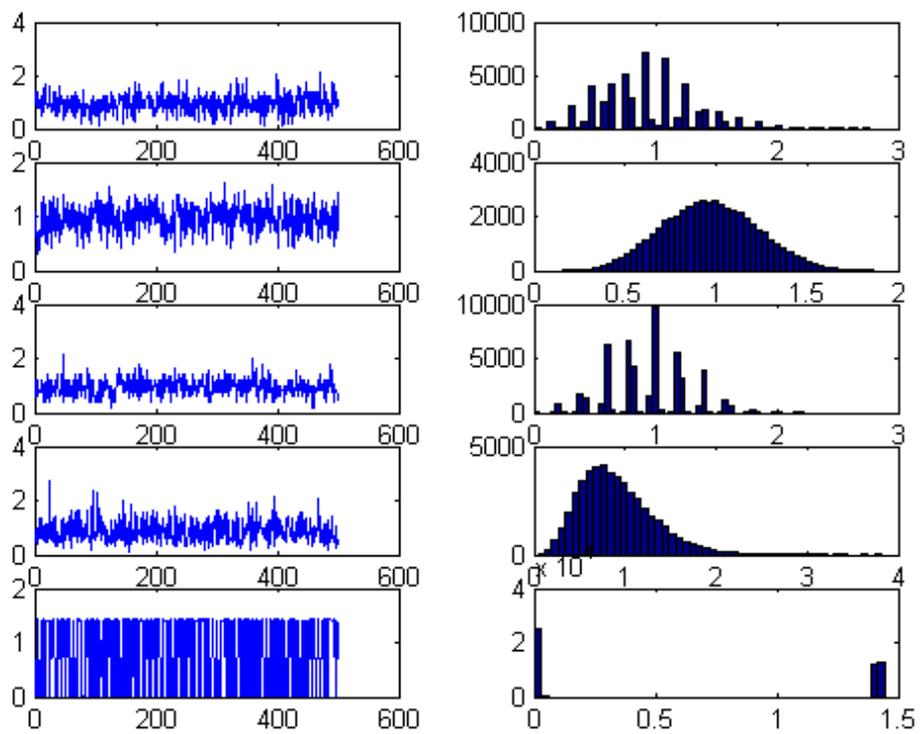


圖 3-7 分離之結果

3.1.3 dilation problem 與 permutation problem

由於針對每個頻率個別做訊號分離，因此在 2.2.3 ICA 之不確定性所提到的 ”無法確定獨立變數之大小與順序”在此產生了問題。

當我們個別對每一個頻率做 ICA 時，在每個頻率下所分離出的結果為原來之訊號乘上一未知的倍數，而造成最後將每個 frame 之頻譜重組回去時發生問題，即為 dilation problem。同理，由於分離結果順序的未知，也造成了 permutation problem，如圖 3-8。因此，在將訊號轉回 time domain 之訊號前，必須將這兩大問題克服。

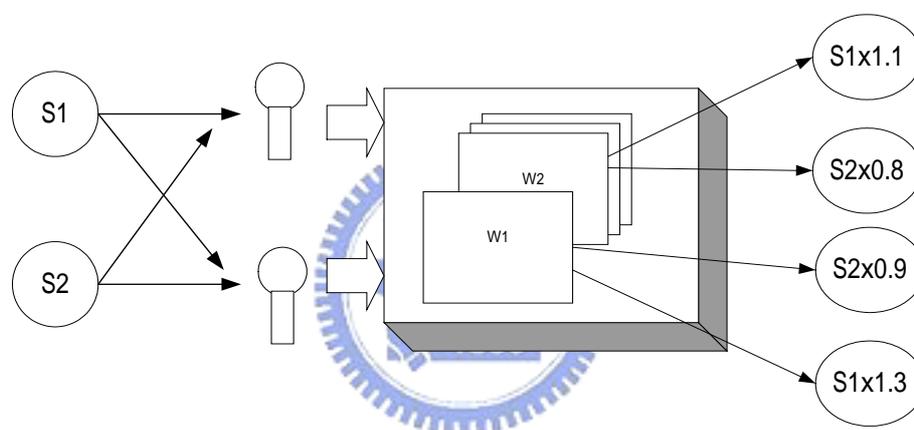


圖 3- 8 dilation and permutation problem

3.2 針對 dilation 與 permutation problem 之演算法

在此介紹 Satoshi KURITA 等人利用 array signal processing 的觀念以及 Noboru Murata 利用 demixing matrix 之 inverse 與語音之 envelope 的想法所提出的解決之道。

Satoshi KURITA 的方法

Satoshi KURITA 的方法著重於不同 source 對於 microphone array 之時間延遲 (arriving lags)，如圖 3-9，其中第 k 個麥克風與參考點 O 之距離為 d_k ，而第 m

個聲源之入射角為 θ_m ，若 $m=1\sim M$ ； $k=1\sim K$ ，則接收到的訊號為

$$X = AS$$

$$\begin{bmatrix} x_1(f) \\ \vdots \\ x_K(f) \end{bmatrix} = \begin{bmatrix} a_{11}(f) & \cdots & a_{1M}(f) \\ \vdots & \ddots & \vdots \\ a_{K1}(f) & \cdots & a_{KM}(f) \end{bmatrix} \begin{bmatrix} s_1(f) \\ \vdots \\ s_M(f) \end{bmatrix} \quad (3.21)$$

其中 S 為 source 在參考點 O 之訊號，而 mixing matrix 若只考慮時間延遲為

$$a_{km}(f) = \exp(j2\pi f \tau_{km}) \quad , \quad \tau_{km} = \frac{d_k \sin \theta_m}{c} \quad (3.22)$$

其中 c 為聲速。

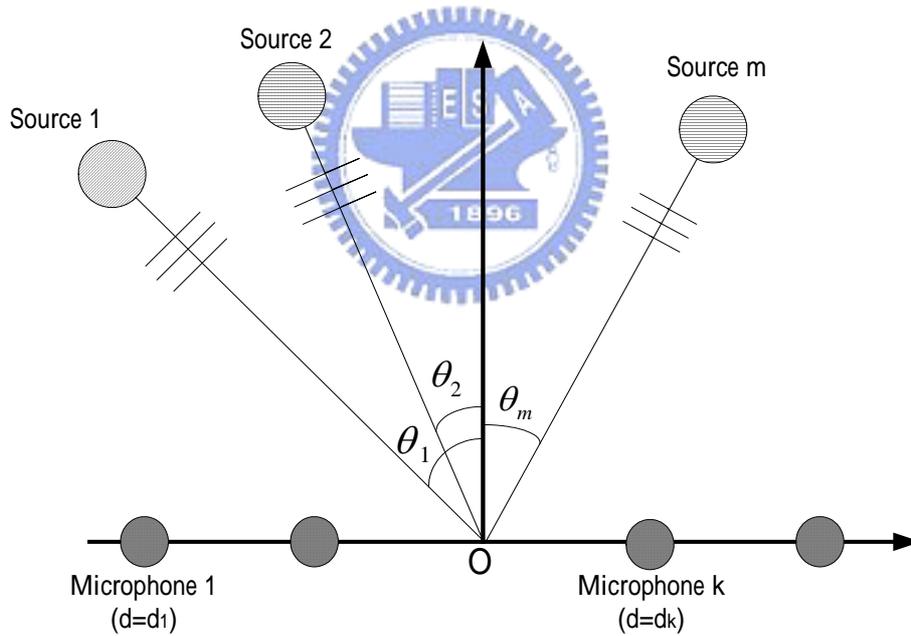


圖 3- 9 microphone array 與 source 之關係

考慮 $k=m=2$ ，分離後之訊號為

$$\begin{bmatrix} y_1(f) \\ y_2(f) \end{bmatrix} = \begin{bmatrix} w_{11}(f) & w_{12}(f) \\ w_{21}(f) & w_{22}(f) \end{bmatrix} \begin{bmatrix} x_1(f) \\ x_2(f) \end{bmatrix} \quad (3.23)$$

因此

$$y_1 = w_{11}(f)x_1(f) + w_{12}(f)x_2(f) \quad (3.24)$$

此式表示 output signal 為 array signals： x_1 與 x_2 個別乘上一個 weighting 並相加。

以 array signal processing 的觀點來看，這個動作代表著利用 directivity pattern

$$F_m(f, \theta) = \sum_{k=1}^2 w_{mk}(f) \exp[j2\pi f d_k \sin \theta / c] \quad (3.25)$$

以估計 DOA (direction-of-arrival) [19]。

將 3.23 式改寫可得

$$\begin{aligned} \begin{bmatrix} y_1(f) \\ y_2(f) \end{bmatrix} &= \begin{bmatrix} w_{11}(f) & w_{12}(f) \\ w_{21}(f) & w_{22}(f) \end{bmatrix} \begin{bmatrix} \exp[j2\pi f d_1 \sin \theta_1 / c] & \exp[j2\pi f d_1 \sin \theta_2 / c] \\ \exp[j2\pi f d_2 \sin \theta_1 / c] & \exp[j2\pi f d_2 \sin \theta_2 / c] \end{bmatrix} \begin{bmatrix} s_1(f) \\ s_2(f) \end{bmatrix} \\ &= \begin{bmatrix} F_1(f, \theta_1) & F_1(f, \theta_2) \\ F_2(f, \theta_1) & F_2(f, \theta_2) \end{bmatrix} \begin{bmatrix} s_1(f) \\ s_2(f) \end{bmatrix} \end{aligned} \quad (3.26)$$

若分離成功，則 $F_1(f, \theta_1)$ 與 $F_1(f, \theta_2)$ 其中一個為零；同理 $F_2(f, \theta_1)$ 與 $F_2(f, \theta_2)$ 其中一個為零；若 y_1 對應於 s_2 ，則當 $\theta = \theta_1$ 時 F_1 等於零，而 $\theta = \theta_2$ 時 F_2 等於零。

於是可藉由畫出 F_m 對 θ 之 pattern 以得知是否有 permutation 發生。如圖 3-10，上

半圖為 $f = f_1$ 之 directive pattern；下半圖為 $f = f_2$ 。在圖中可看出 $F_1(f_1, \theta)$ 與

$F_2(f_2, \theta)$ 在同一角度(Source2)為零，而 $F_2(f_1, \theta)$ 與 $F_1(f_2, \theta)$ 皆在 Source1 之角度為

零，故有 permutation 發生。而 dilation 的問題可藉由 Source 之角度之得知，而

算出該角度之 F 值即可得知放大之倍數，而得以解決。

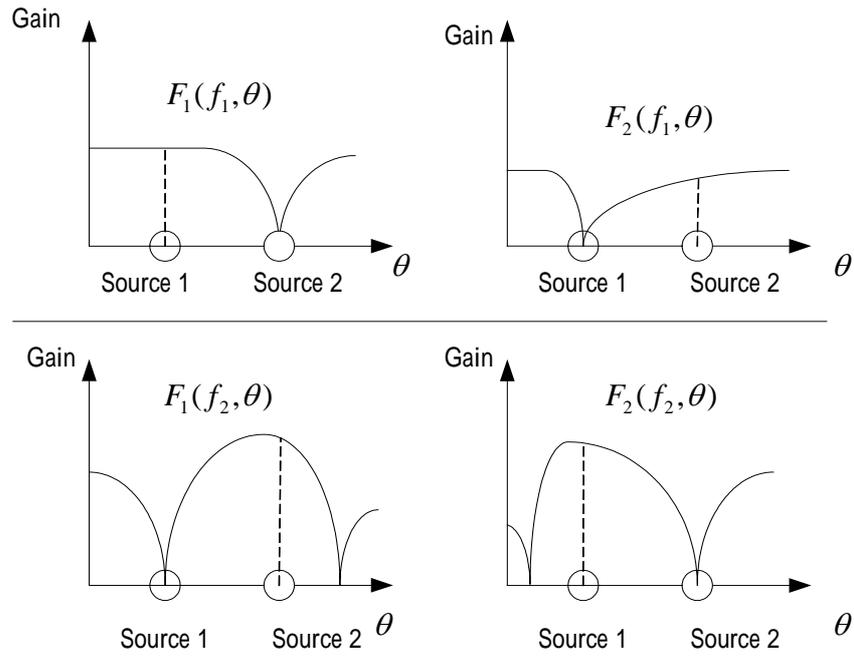


圖 3- 10 directivity pattern

Noboru Murata 的方法

Noboru Murata 提出了利用 demixing matrix 之 inverse 來消除 dilation problem，此方法又稱為 split spectrum[20]。由於做 ICA 之前會先乘上一個 whitening matrix V ，因此整體的 demixing matrix 為 WV (W 為 ICA 針對已 whitening 之訊號所找出之 demixing matrix)，由(2.9)可得 $WVA=PD$ ，所以 $(WV)^{-1} = AD^{-1}P$ 。考慮聲源與麥克風數皆為 2，且有 permutation 發生，即

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} d_2 s_2 \\ d_1 s_1 \end{bmatrix} \quad (3.27)$$

則

$$\begin{aligned}
(WV)^{-1} \begin{bmatrix} y_1 \\ 0 \end{bmatrix} &= AD^{-1}P \begin{bmatrix} y_1 \\ 0 \end{bmatrix} = AD^{-1}P \begin{bmatrix} d_2 s_2 \\ 0 \end{bmatrix} \\
&= AD^{-1} \begin{bmatrix} 0 \\ d_2 s_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \frac{1}{d_1} & 0 \\ 0 & \frac{1}{d_2} \end{bmatrix} \begin{bmatrix} 0 \\ d_2 s_2 \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 0 \\ s_2 \end{bmatrix} = \begin{bmatrix} a_{12} s_2 \\ a_{22} s_2 \end{bmatrix} = \begin{bmatrix} v_1(w, tr, 1) \\ v_2(w, tr, 1) \end{bmatrix} = \begin{bmatrix} \text{第一個麥克風收到的 } s_2 \\ \text{第二個麥克風收到的 } s_2 \end{bmatrix}
\end{aligned} \tag{3.28}$$

同理

$$(WV)^{-1} \begin{bmatrix} 0 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} s_1 \\ a_{21} s_1 \end{bmatrix} = \begin{bmatrix} v_1(w, tr, 2) \\ v_2(w, tr, 2) \end{bmatrix} = \begin{bmatrix} \text{第一個麥克風收到的 } s_1 \\ \text{第二個麥克風收到的 } s_1 \end{bmatrix} \tag{3.29}$$

$$V(w, tr, 1) = \begin{bmatrix} v_1(w, tr, 1) \\ v_2(w, tr, 1) \end{bmatrix}; \quad V(w, tr, 2) = \begin{bmatrix} v_1(w, tr, 2) \\ v_2(w, tr, 2) \end{bmatrix}$$

而沒有 permutation 發生的情形也可用相同之推導而得到個別麥克風所分別收到的聲源，而消除 dilation problem。

針對 permutation problem，可利用語音訊號之特性：語音訊號長時間來看，由於有聲音大小的 modulation，故為 non-stationary。假設同一個聲源在不同頻率在大小上有相似的 modulation，則可利用此一性質來判別分離後之訊號何者來自同一個聲源。

將頻率為 w 之第 i 個聲源訊號表為

$$s_i(w, t_r) = a_i(w, t_r) e^{j\phi_i(w, t_r)} \tag{3.30}$$

由於 non-stationary 的特性， $a_i(w, t_r)$ 隨時間而變，也就相當於 $s_i(w, tr)$ 之 envelop，當 $s_i(w, tr)$ 與 $s_j(w, tr)$ 為 independent，則不同 source 在同一頻率或不同頻率之 envelop 之 correlation 為零，而同一 source 在不同頻率之 correlation 不為零(由於大小受到相似的 modulation)。

$$\begin{aligned}
\text{corr}(a_i(w,ts), a_j(w,ts)) &= 0, \quad i \neq j \\
\text{corr}(a_i(w,ts), a_j(w',ts)) &= 0, \quad i \neq j, \quad w \neq w' \\
\text{corr}(a_i(w,ts), a_i(w',ts)) &\neq 0
\end{aligned} \tag{3.31}$$

因此利用 envelop 之 correlation coefficient 來對每個頻率做重組是合理的。

$$\gamma(a_i(w,ts), a_j(w',ts)) = \frac{\text{corr}(a_i(w,ts), a_j(w',ts))}{\sqrt{\text{corr}(a_i(w,ts), a_i(w,ts)) \text{corr}(a_j(w',ts), a_j(w',ts))}} \tag{3.32}$$

為使 envelop 更加明顯，於是利用 moving average ξ

$$\hat{\xi} v(w, tr; i) = \frac{1}{2M+1} \sum_{tr'=tr-M}^{tr'+M} \sum_{j=1}^2 |v_j(w, tr; i)| \tag{3.33}$$

頻譜重組之方法分為以下幾個步驟：

1. 對每個頻率分離出來並且已消除 dilation problem 之訊號做 correlation

$$\text{sim}(w) = r(\hat{\xi} v(w, ts, ; i), \hat{\xi} v(w, ts, ; j)) \tag{3.34}$$

再依照 correlation 之大小排序, correlation 最小表示 envelop 之特徵最明顯

$$\text{sim}(w_1) \leq \text{sim}(w_2) \leq \dots \leq \text{sim}(w_N) \tag{3.35}$$

2. 對於 w_1 ，直接使重組後之結果 $Y(w_1, tr; i)$ 等於 $V(w_1, tr; i)$
3. 對於 w_k ，將在此頻率之 envelop 與 $w_1 \sim w_{k-1}$ 重組過之 envelop 之和做 correlation，並找出可使 correlation 最大之交換方式，即

$$r(\hat{\xi}(v(w, tr;1)), \sum_{j=1}^{k-1} \xi(Y(w_j, tr;1))) + r(\hat{\xi}(v(w, tr;2)), \sum_{j=1}^{k-1} \xi(Y(w_j, tr;2))) \quad (3.36)$$

$$r(\hat{\xi}(v(w, tr;2)), \sum_{j=1}^{k-1} \xi(Y(w_j, tr;1))) + r(\hat{\xi}(v(w, tr;1)), \sum_{j=1}^{k-1} \xi(Y(w_j, tr;2))) \quad (3.37)$$

在(3.36) 與(3.37)中選擇其中較大之 correlation，用以判斷是否有 permutation。若(3.36)之值較大，表示沒有發生 permutation；若(3.37)之值較大，表示有發生 permutation。

4. 在第三步驟已得知是否有發生 permutation，故可 assign 正確之 $V(w_k, tr; i)$ 給 $Y(w_k, tr; i)$
5. 回到第三步驟，直到 $k=N$



3.3 討論

在此針對以上所提及之 FastICA 與解決 dilation 和 permutation 之兩種方法以實際模擬的方式來探討其中之特性。

在此模擬當中，聲源使用了 sampling rate 為 16k Hz 之兩段語音，將兩段聲源與 186 點之 HRTF(Head-Related Transfer Function)做 convolution 以產生混和訊號。HRTF [21] 為 MIT media Lab 所量測出聲源相對於兩耳之 transfer function，其中聲源與假人之距離為 1.4 meters；兩耳之距離為 15.2 cm，量測角度每五度量測一次，高度從-40 度(低於水平面 40 度)到 90 度(在頭頂)，每個高度皆量測 360 度(正前方為零度)。在本模擬中使用高度為零度，方位在±30 度處，如圖 3- 11

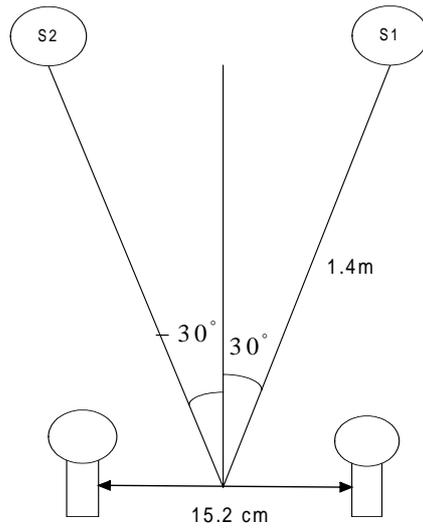


圖 3- 11 模擬位置圖

聲源訊號為



圖 3- 12 聲源訊號 s1



圖 3- 13 聲源訊號 s2

混和訊號為



圖 3- 14 混和訊號 x1



圖 3- 15 混和訊號 x2

將混和訊號每 60ms 取一個 frame，做 2048 點的 DFT，再對每個頻率做 ICA，結果如圖 3- 16，3- 16 (a)為第 30 個頻率而 3- 16 (b)為第 46 個頻率的分離情形。圖中由上而下為 s1 和 s2(聲源訊號)、x1 和 x2(混和訊號)、y1、y2(分離後的訊號)。3- 16 (a)為分離較成功之例子，由圖中可看出分離訊號與聲源訊號相接近，而 3- 16 (b)為分離失敗的例子。由於 FastICA 對於訊號源限定實部與虛部為 uncorrelated 且 variance 相同，因此在部分不符合此限制之頻率無法成功分離。

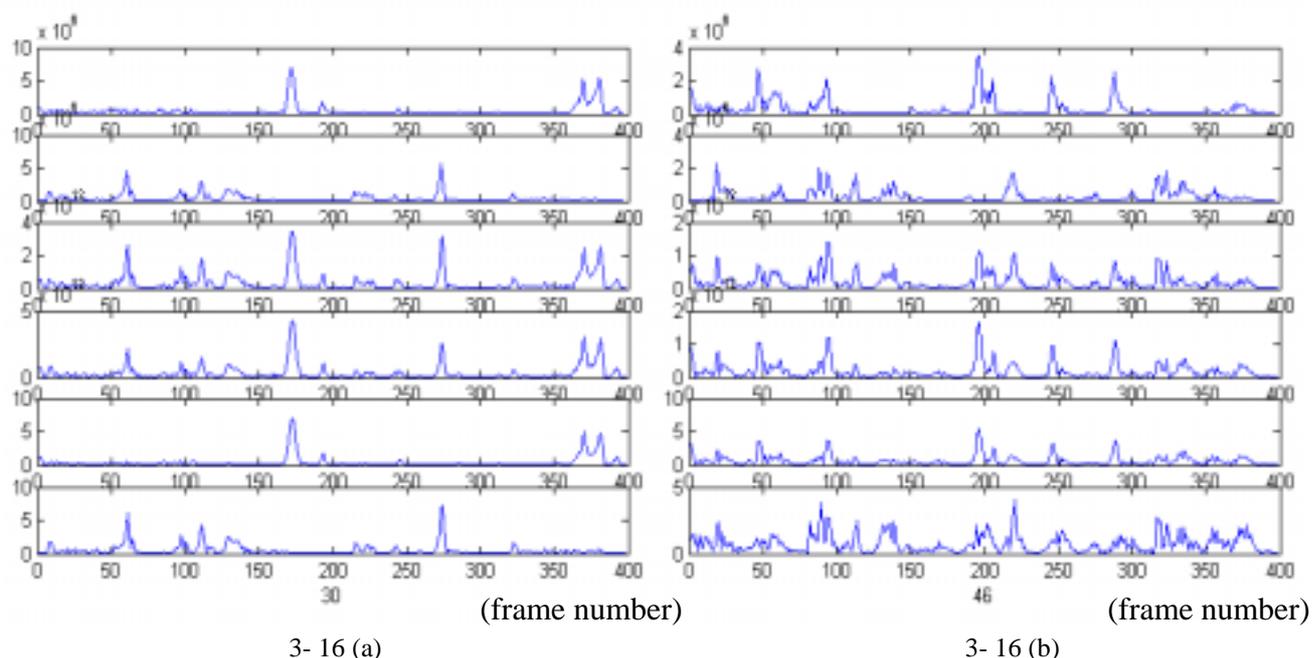
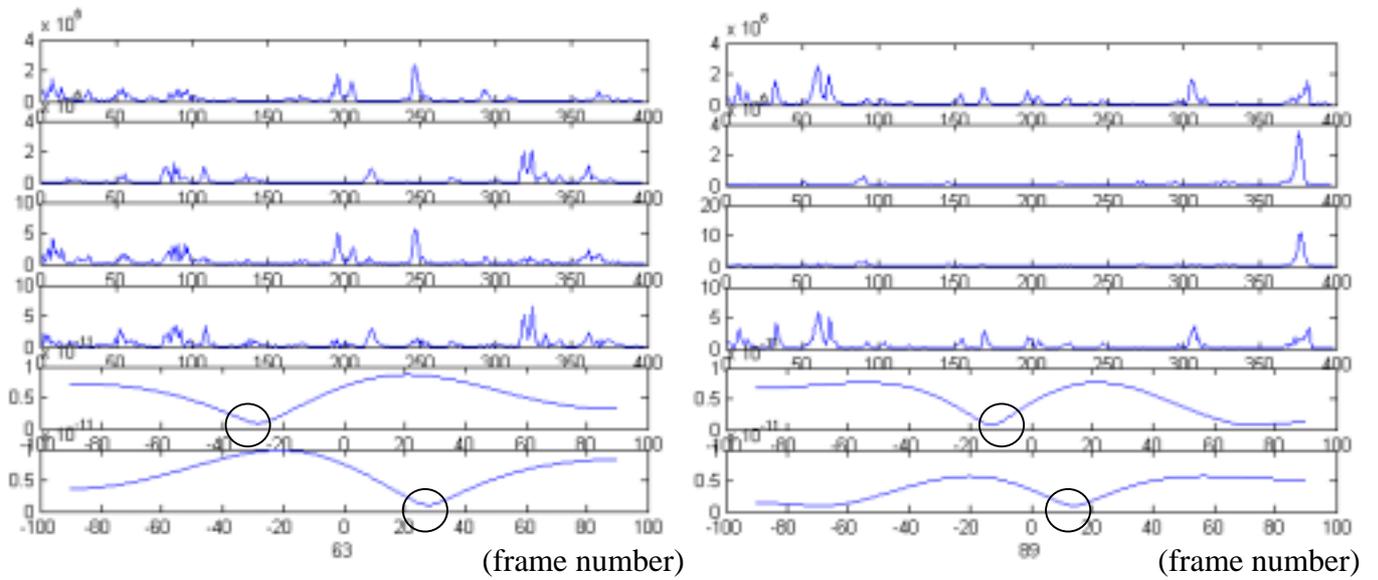


圖 3- 16 經過 FastICA 分離後之訊號

圖 3- 17 為利用 directivity pattern 來估計 DOA 之結果，3-17(a)與 3-17(b)分別為第 63 個頻率與第 89 個頻率之結果，本圖由上而下分別為 s1 和 s2、y1 和 y2、 F_1 和 F_2 ，圖中圈起之部分為估計出之 DOA，於圖中可看出估計出之方向與真正之方向(正負三十度)有相當大的誤差。造成誤差主要的原因為本方法只考慮到不同 source 對於 microphone array 之時間延遲，而未考慮整個 convolution 之效應。

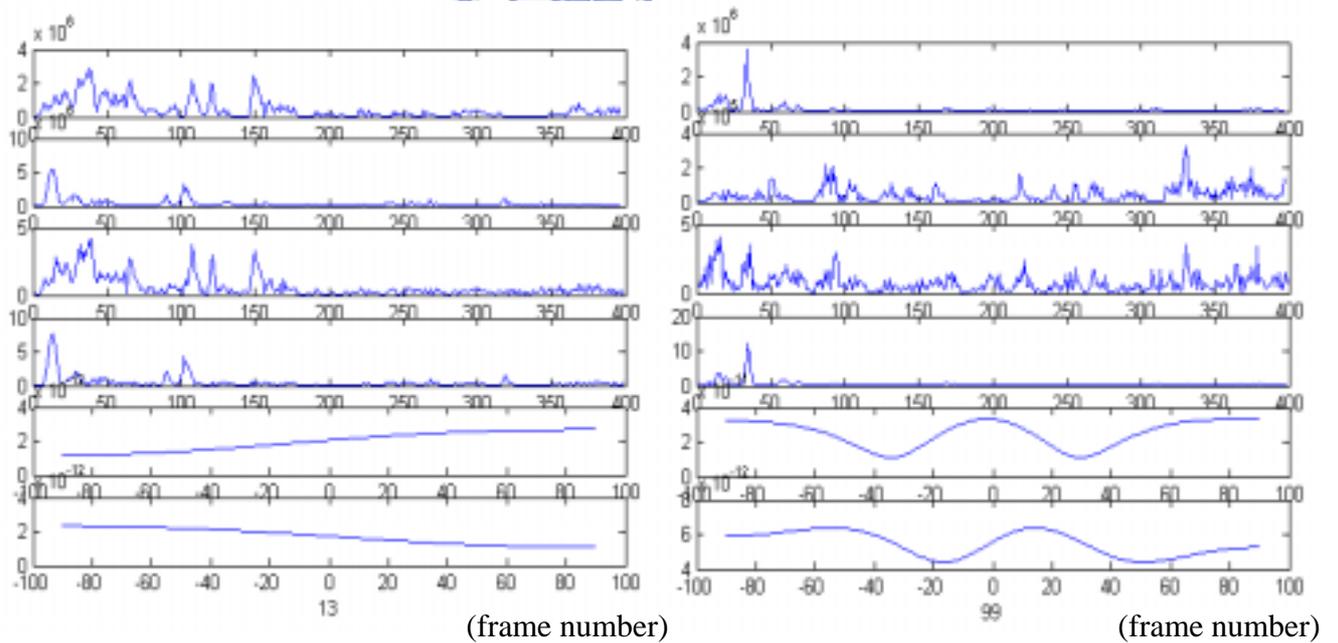


3- 17(a)

3- 17(b)

圖 3- 17 directive pattern

除此之外，directivity pattern 在低頻 DOA 之效果並不明顯(如圖 3-18(a))，而在高頻會有 aliasing 現象(如圖 3-18(b))，因此在應用上效果並不理想。



3-18(a)

3-18(b)

圖 3- 18 在低頻與高頻之 directivity pattern

Noboru Murata 的方法，在推導上即可看出確實能將 dilation problem 消除，而 permutation problem 的部分如圖 3-19，其中 3-19(a)、3-19(b)、3-19(c)分別為 $w = w_1、w_2、w_3$ 即為 envelop 特徵最明顯之三個頻率(圖中由上而下分別為 s1 和 s2、y1 和 y2、y1 之 envelop 和 y2 之 envelop)與 s1 和 s2 之 spectrogram 相對應(圖 3-20)可看出此 envelop 的確可對應到語音音量之 modulation，然而對於 envelop 較不明顯的頻率(如 3-19(d))則不易分辨。

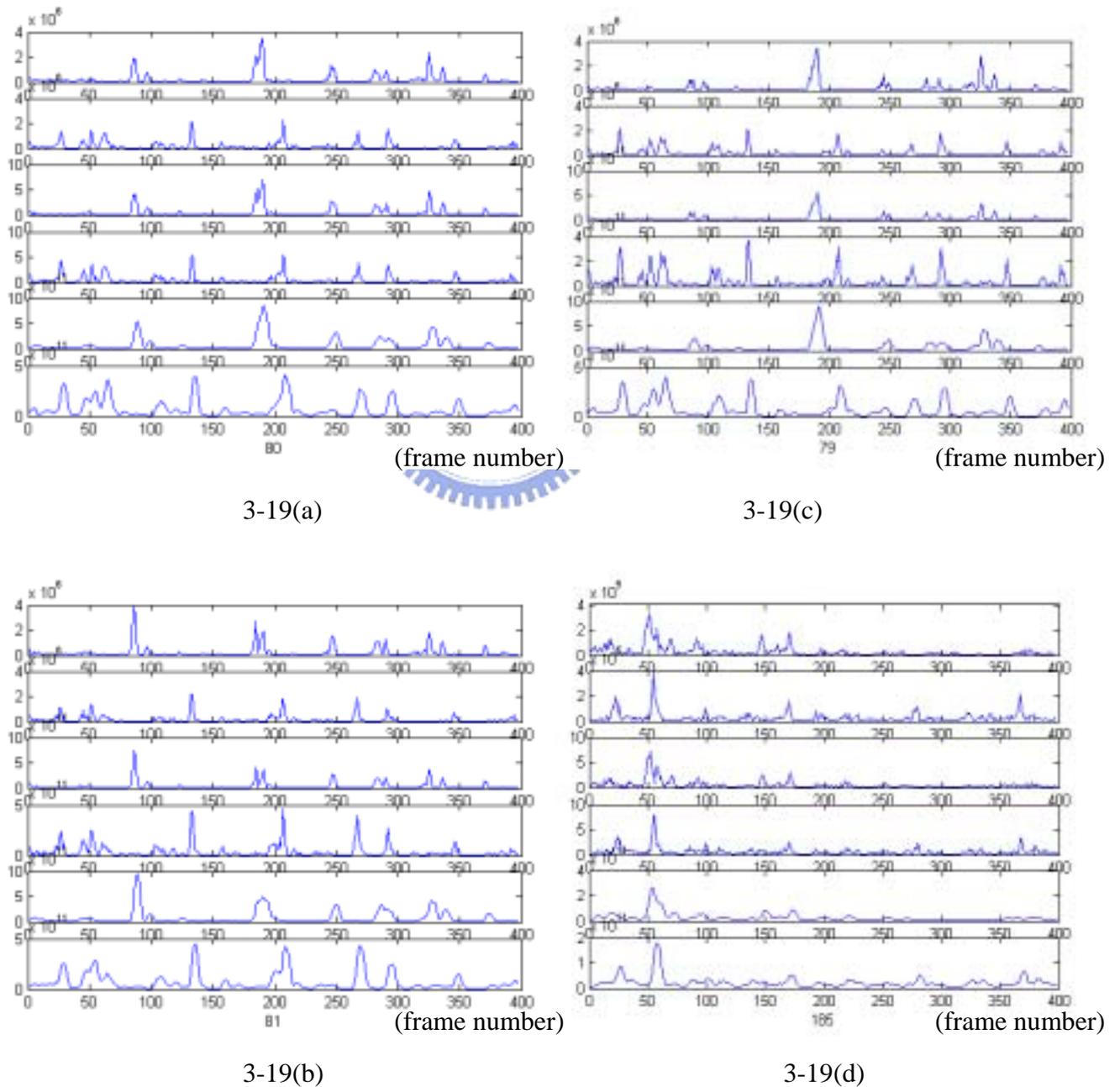


圖 3-19 分離結果之 envelop

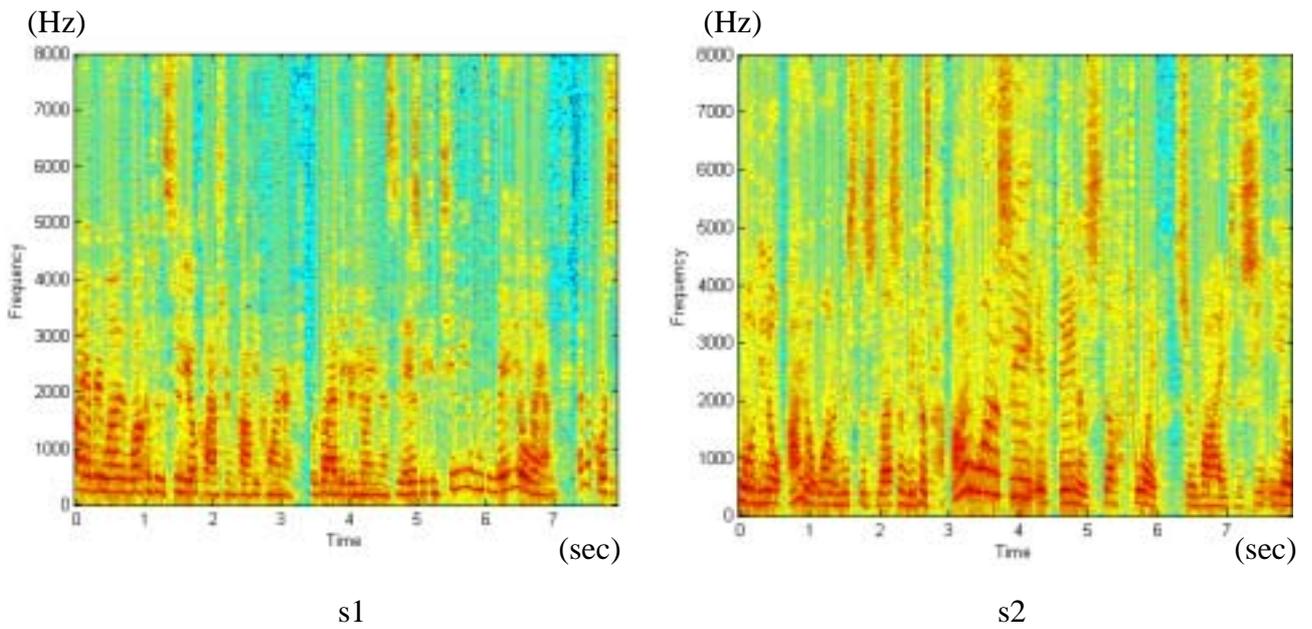


圖 3- 20 s1 與 s2 之 spectrogram

以此方法分離出之語音為



圖 3- 21 以 Noboru Murata 方法之分離結果

由於 FastICA 在部分不符合其限制之頻率無法完全將訊號分離，且 Noboru Murata 的方法雖可解決 dilation problem，但對於 envelop 不明顯之頻率無法正確判別是否有 permutation，因此分離出的語音僅能將另一聲源音量稍微壓低

第四章 以 ICA 估計 demixing matrix

4.1 利用預先估計之 demixing matrix 執行聲音分離

由於頻域 ICA 無法在每個頻率皆有良好的分離結果，且 permutation problem 尚未有效果較佳的演算法可以處理，於是在此提出了一個利用 training signal 找出 demixing matrix 的想法，由於我們可以自由的設計 training signal，為了提高在每個頻率 ICA 的分離能力，我們設計出能在每個頻率滿足 FastICA 之限制之 training signal 使 ICA 能更加正確地找出每個頻率的 demixing matrix B，以加強分離效果。

4.1.1 演算法架構

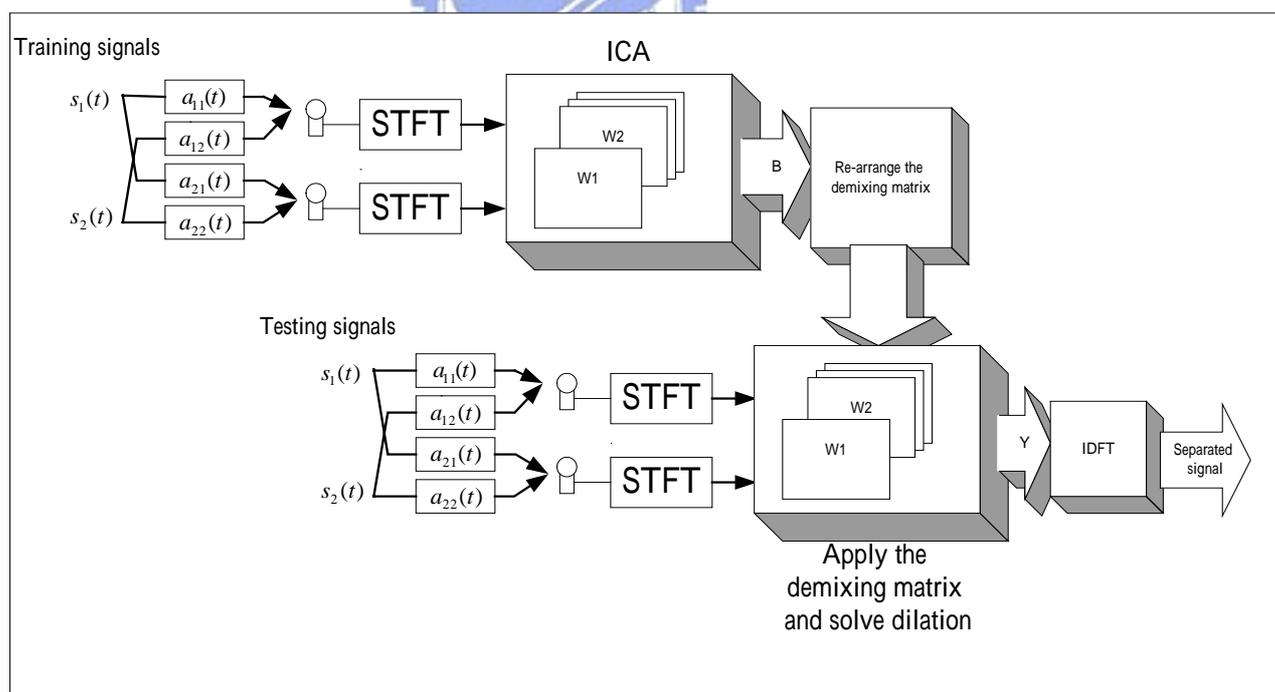


圖 4-1 演算法架構

此演算法架構如圖 4-1，圖中 training signal 為根據 ICA 可分離之條件所設計出之 training signal，將此 training signal 擺在與真正要分離之語音訊號同樣位置的地方，由於同一位置之空間效應不變，也就是 mixing matrix 相同，因此我們可以利用 training signal 將對每個頻率正確的 demixing matrix B 求出。而透過 training signal 適當之設計，也可解決 permutation problem。有了對於該位置每個頻率之 demixing matrix 之後，對於欲分離之訊號，只要用此 demixing matrix 即可將訊號分離出來。

4.1.2 training signal 之設計

由於對複數訊號做處理之 FastICA 本身對於欲分離之獨立元素有實部與虛部 uncorrelated 且 variance 相同之限制，而一般聲源訊號無法所有頻率皆滿足此限制，因此並非對每個頻率 ICA 皆能分離出獨立元素，也就造成了分離效果不佳之缺點。在此，針對這個問題，設計了在每個頻率可滿足 FastICA 限制之 training signal 使得分離效果能夠提昇，進而得到較正確之 demixing matrix。

在 3.1.2 曾提到過，FastICA 對於滿足其限制之訊號可有效分離且收斂快，因此 training signal 之設計可利用 3.1.2 之模擬訊號： $s_j = r_j(\cos \phi_j + i \sin \phi_j)$ 產生，有了一組可分離之模擬訊號即可產生一組 training signal，方法如下：

step 1. 產生一組 FastICA 可分離之訊號，如圖 4-2

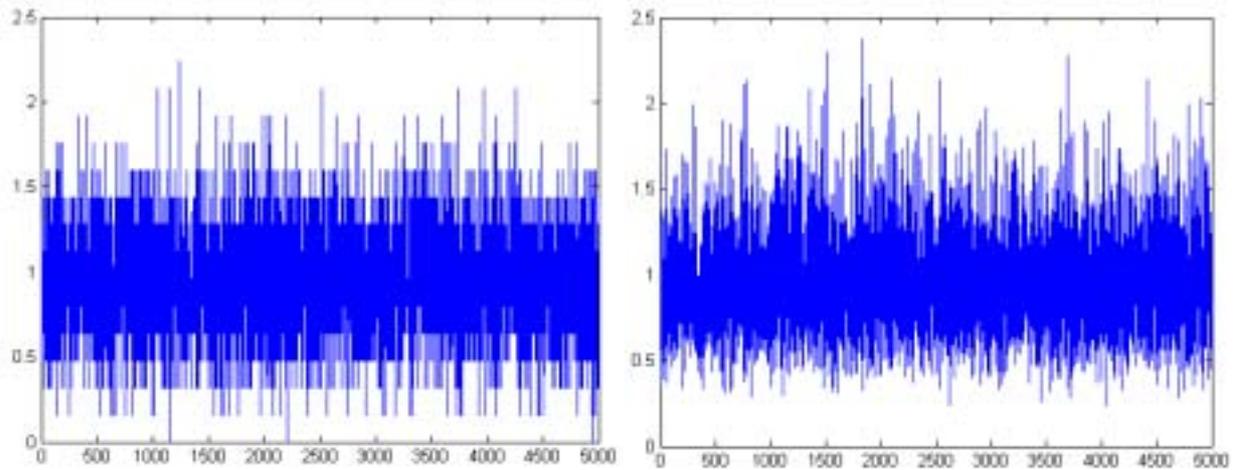


圖 4- 2 FastICA 可分離之訊號

step 2. 由於 frequency domain ICA 是對每個頻率執行訊號分離的動作，因此 step1 所產生出之訊號為在某一頻率(在 frequency domain)下之訊號，若欲產生一聲音訊號(在 time domain)則必須在每個頻率都產生一組 FastICA 可分離之訊號。為簡化起見，在每個頻率皆用同一組 step1 所產生之訊號，為了使產生之訊號接近語音，於是參考一語音之 spectrum(取樣頻率為 16k，長度 60ms，做 2048 點之 DFT 所得)如圖 4-3

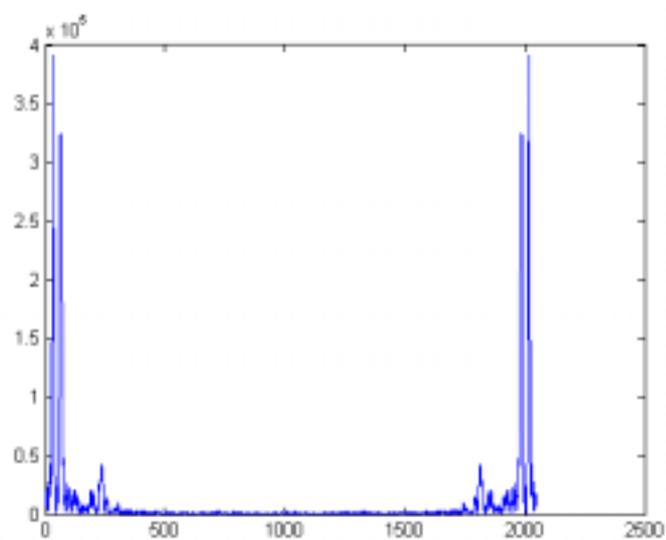


圖 4- 3 語音之 spectrum (sample number)

將 step1 所產生之訊號逐點乘上此一 spectrum 做 IDFT 回 time domain 後逐段相接
即可得到 training signal，如圖 4-4 取三點 step1 產生之訊號之示意圖。

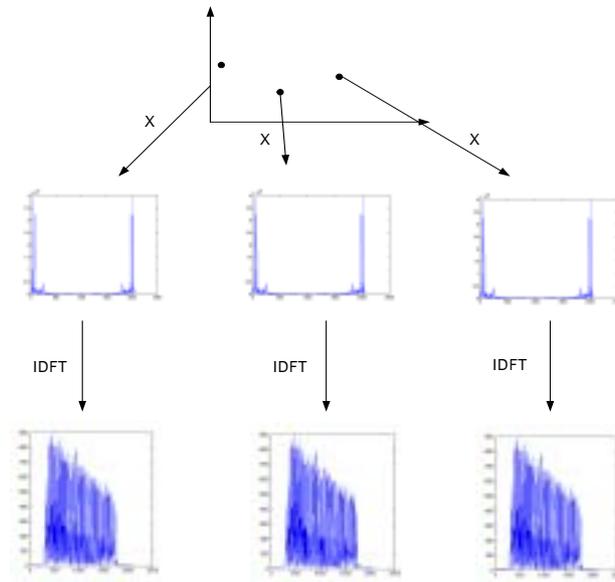


圖 4-4 產生 training signal 之示意圖

所產生之 training signal 如圖 4-5

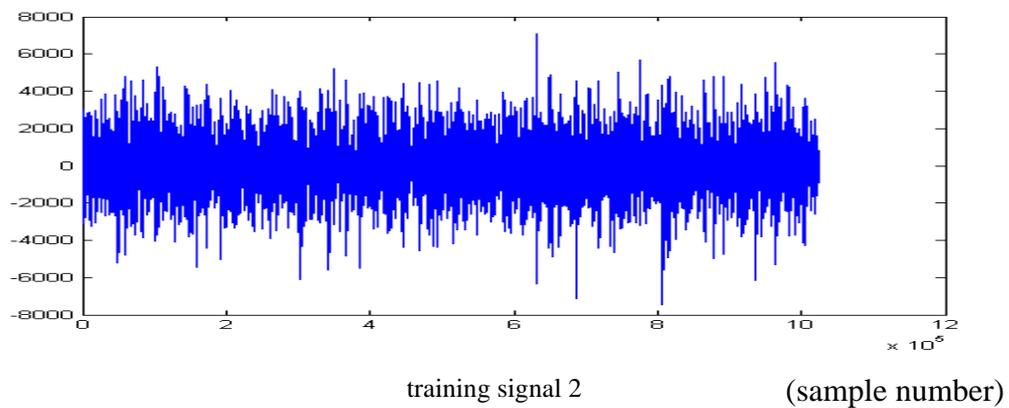
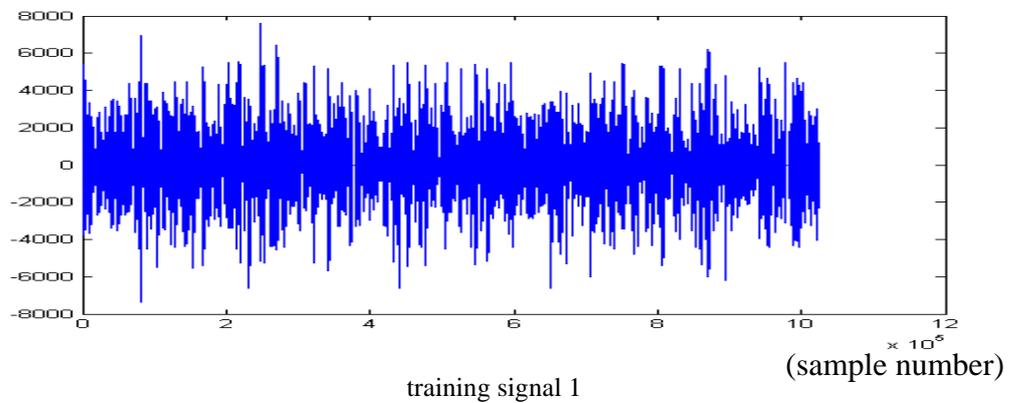


圖 4-5 training signal

以此方式所產生之訊號取 2048 點之 frame 轉到 frequency domain 之每個頻率之訊號即為 step1 所產生之 FastICA 可分離之訊號乘上一個 complex gain，而此一動作不會改變原本訊號實部與虛部 uncorrelated 且 variance 相同之特性，證明如下：

假設 step1 所產生之訊號為 $a + bj$ (為 random variable)，complex gain 為 $C + Dj$

$$(a + bj)(C + Dj) = aC - bD + (aD + bC)j \quad (4.1)$$

(1) 實部與虛部 uncorrelated

$$\begin{aligned} & E\{(aC - bD)(aD + bC)\} \\ &= E\{a^2CD + abC^2 - abD^2 - b^2CD\} \\ &= CDE\{a^2\} + C^2E\{ab\} - D^2E\{ab\} - CDE\{b^2\} \end{aligned} \quad (4.2)$$

$$\begin{aligned} &= CD(E\{a^2\} - E\{b^2\}) \\ &= 0 \end{aligned} \quad (4.3)$$

實部與虛部 variance 相同

實部之 variance：

$$\begin{aligned} & E\{(aC - bD)^2\} \\ &= E\{a^2C^2 - 2abCD + b^2D^2\} \\ &= C^2E\{a^2\} - 2CDE\{ab\} + D^2E\{b^2\} \\ &= C^2E\{a^2\} + D^2E\{b^2\} \\ &= (C^2 + D^2)E\{a^2\} = (C^2 + D^2)E\{b^2\} \end{aligned} \quad (4.4)$$

虛部之 variance：

$$\begin{aligned} & E\{(aD - bC)^2\} \\ &= E\{a^2D^2 - 2abCD + b^2C^2\} \\ &= D^2E\{a^2\} - 2CDE\{ab\} + C^2E\{b^2\} \\ &= D^2E\{a^2\} + C^2E\{b^2\} \\ &= (C^2 + D^2)E\{a^2\} = (C^2 + D^2)E\{b^2\} \end{aligned}$$

4.1.3 permutation 與 dilation problem 之處理

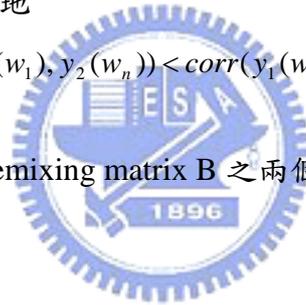
在 3.1.3 中曾提到 Frequency domain ICA 所延伸出之兩大問題：permutation problem 與 dilation problem，而在 3.2 針對此問題介紹了兩個方法，在此使用 Noboru Murata 所提出之利用 demixing matrix 之 inverse 來消除 dilation problem 的方法。而針對 permutation problem，由於 training signal 之每個頻率皆由同一訊號產生，因此只要利用第一個頻率分出之結果當作參考，將其他頻率分離出之訊號與之計算 correlation 即可得知是否有 permutation problem 發生。 $y_1(w_n)$ 與 $y_2(w_n)$ 為在第 n 個頻率分離出之結果，若

$$\text{corr}(y_1(w_1), y_1(w_n)) + \text{corr}(y_2(w_1), y_2(w_n)) > \text{corr}(y_1(w_1), y_2(w_n)) + \text{corr}(y_2(w_1), y_1(w_n)), \text{ for } n > 1$$

則表示無 permutation，相反地

$$\text{corr}(y_1(w_1), y_1(w_n)) + \text{corr}(y_2(w_1), y_2(w_n)) < \text{corr}(y_1(w_1), y_2(w_n)) + \text{corr}(y_2(w_1), y_1(w_n)), \text{ for } n > 1$$

表示有 permutation，則將 demixing matrix B 之兩個 row 交換，即可修正回來。



4.1.4 分離結果

在此模擬當中，取 30 秒之 training signal 找出各個頻率之 demixing matrix，而聲源使用了 sampling rate 為 16k Hz 8 秒之訊號，模擬環境與 3.3 之模擬相同，聲源位於 30° 與 -30° 處，麥克風與聲源之關係如圖 3-12。相同的，混和訊號為聲源訊號與 HRTF 做 convolution 之結果，在此做了三種模擬，聲源訊號分別為女聲與女聲、女聲與男聲、女聲與雜訊。

為了計算分離效果，需定義一計算效能之參數-SNR(Signal to Noise Ratio)。

已知分離結果如下

$$\begin{bmatrix} \hat{s}_1(w) \\ \hat{s}_2(w) \end{bmatrix} = \begin{bmatrix} w_{11}(w) & w_{12}(w) \\ w_{21}(w) & w_{22}(w) \end{bmatrix} \begin{bmatrix} a_{11}(w) & a_{12}(w) \\ a_{21}(w) & a_{22}(w) \end{bmatrix} \begin{bmatrix} s_1(w) \\ s_2(w) \end{bmatrix} = \begin{bmatrix} \alpha_1(w)s_1(w) + \beta_1(w)s_2(w) \\ \alpha_2(w)s_1(w) + \beta_2(w)s_2(w) \end{bmatrix} \quad (4.5)$$

因此，若使 $s_2(w)$ 為零，可得 $\alpha_1 s_1$ 與 $\alpha_2 s_1$ ；使 $s_1(w)$ 為零，可得 $\beta_1 s_2$ 與 $\beta_2 s_2$ ，於

是可得到分離結果之 SNR 定義如下：

$$SNR_1 = 10\log\left(E\left(\frac{|\alpha_1(w)s_1(w)|^2}{|\beta_1(w)s_2(w)|^2}\right)\right) ; SNR_2 = 10\log\left(E\left(\frac{|\beta_2(w)s_2(w)|^2}{|\alpha_2(w)s_1(w)|^2}\right)\right) \quad (4.6)$$

模擬一：女聲+女聲

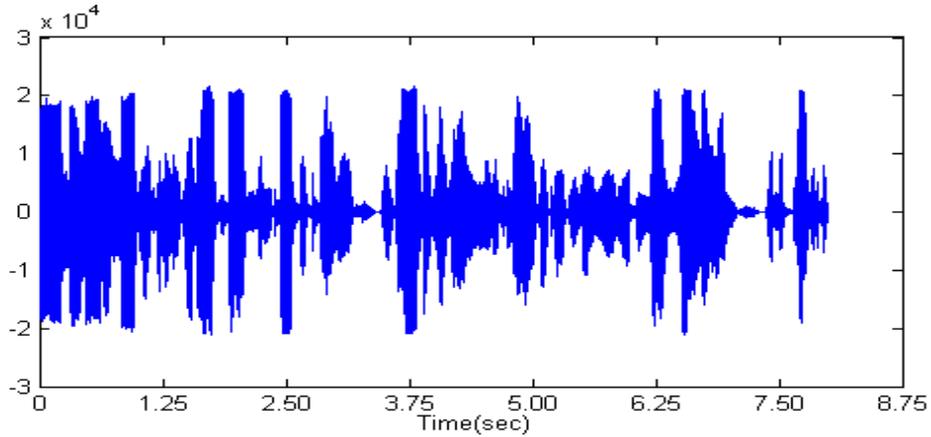


圖 4-6 聲源訊號 s1

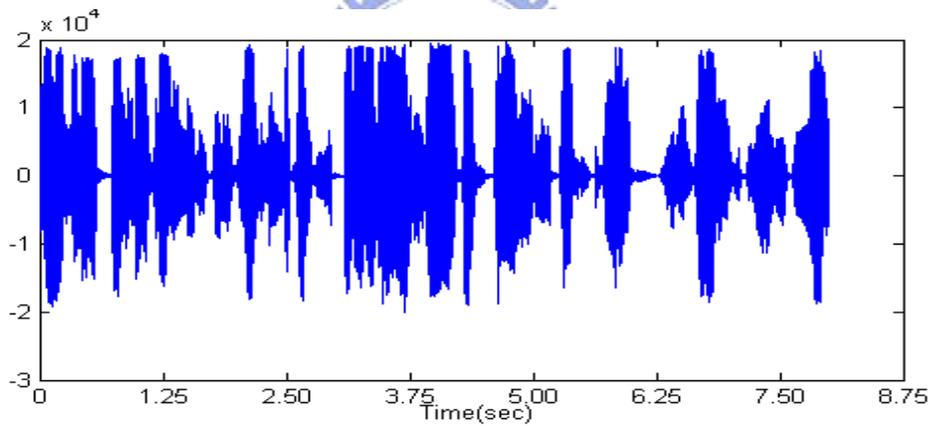


圖 4-7 聲源訊號 s2

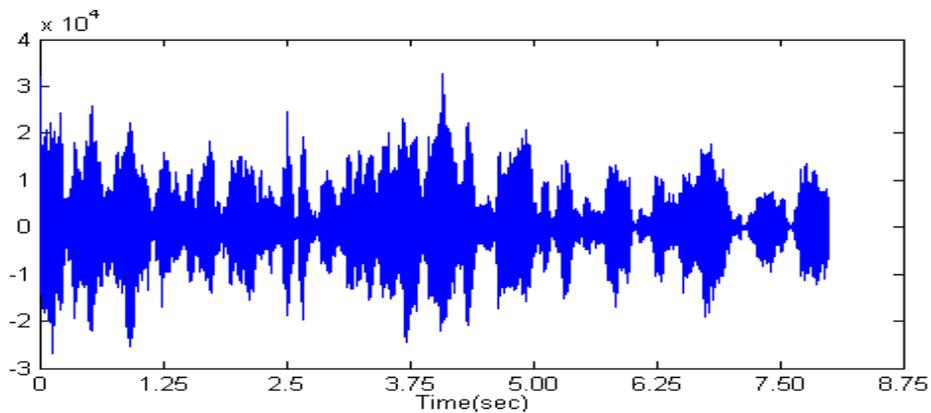


圖 4-8 混和訊號 x1

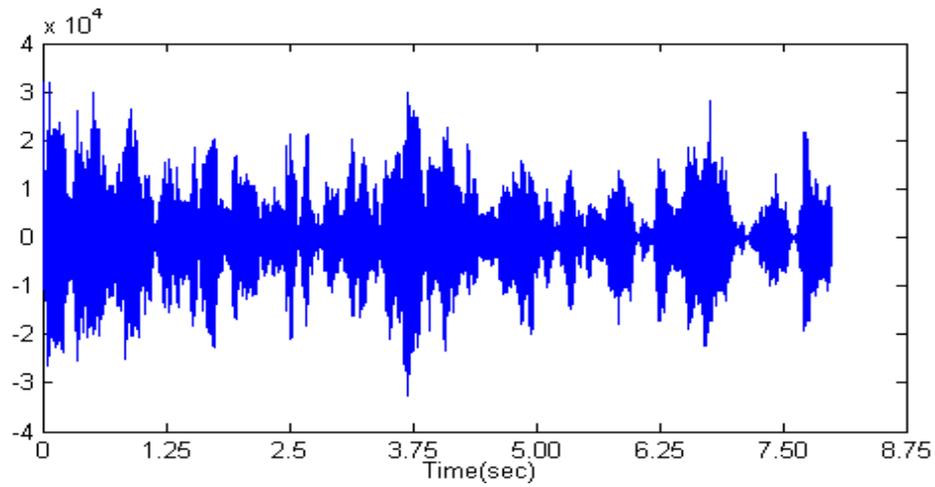


圖 4-9 混和訊號 x2

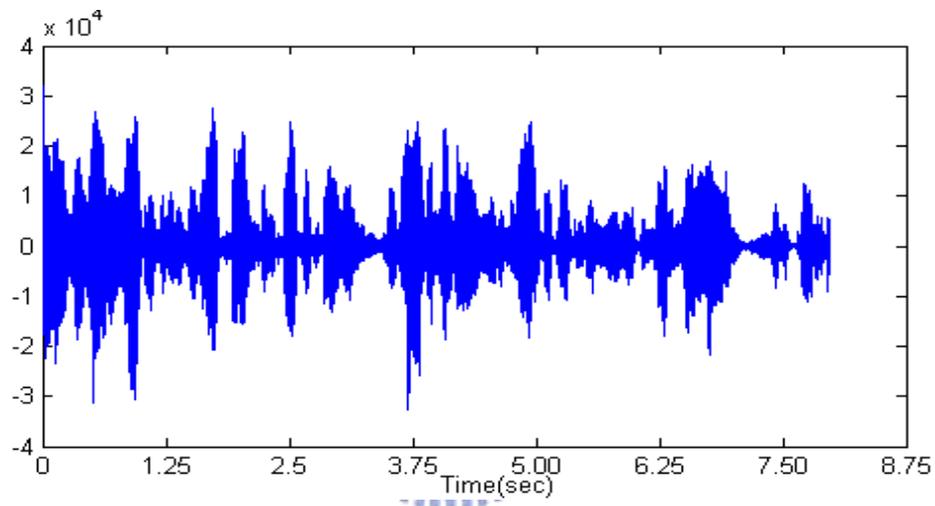


圖 4-10 分離結果 y1 , SNR=12.7863 db

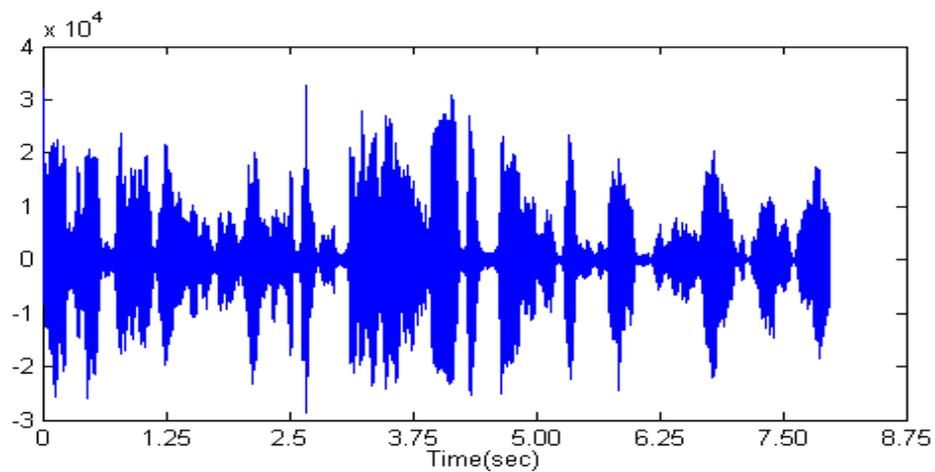


圖 4-11 分離結果 y2 , SNR=20.8392 db

模擬二：女聲+男聲

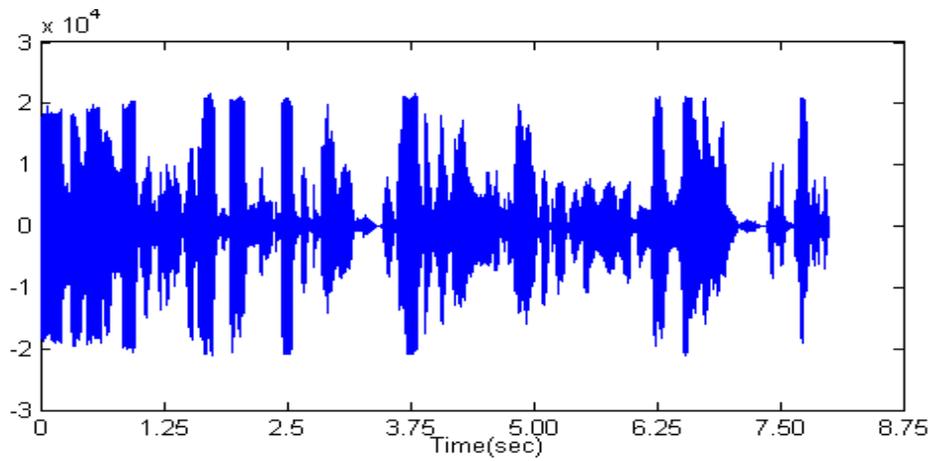


圖 4- 12 聲源訊號 s1

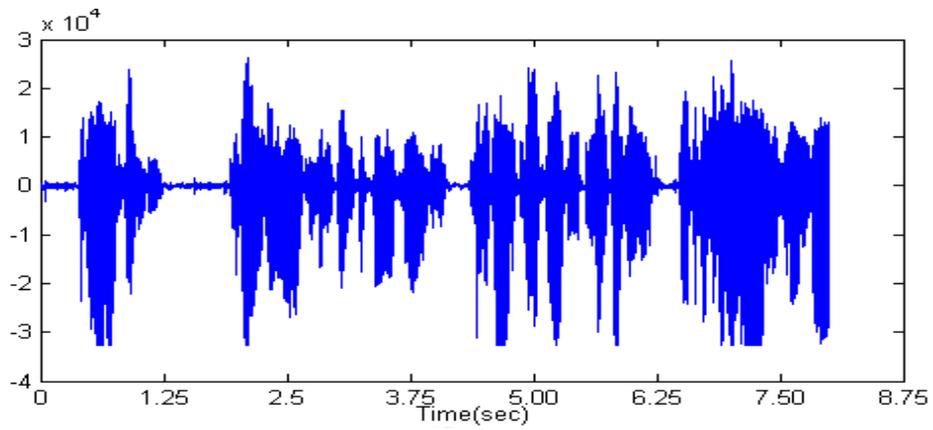


圖 4- 13 聲源訊號 s2

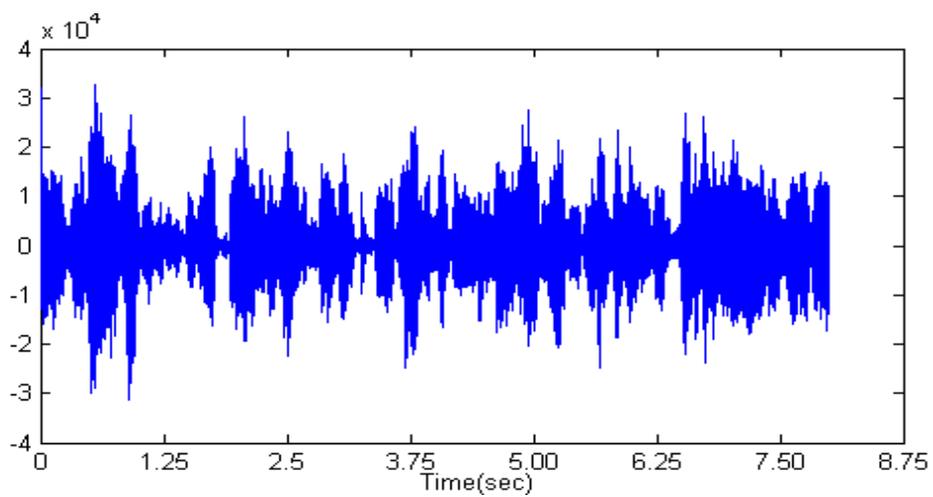


圖 4- 14 混和訊號 x1

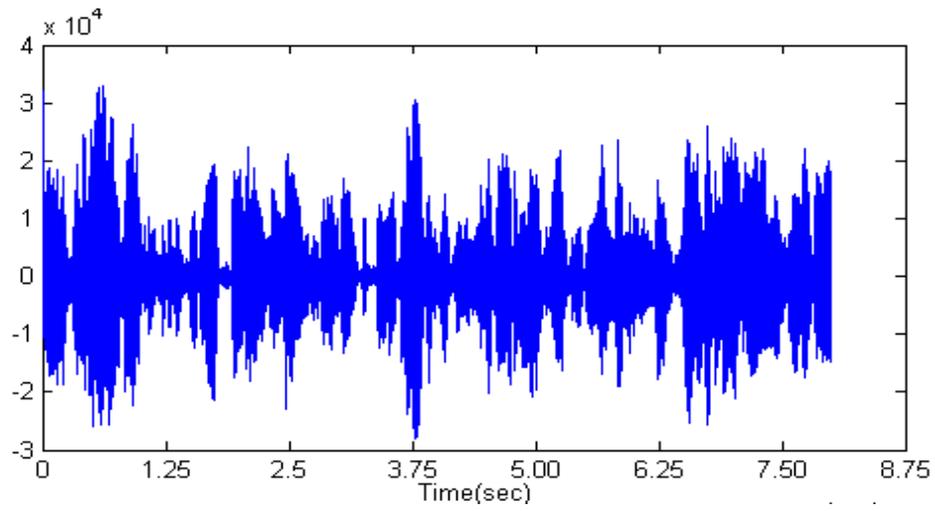


圖 4- 15 混和訊號 x2

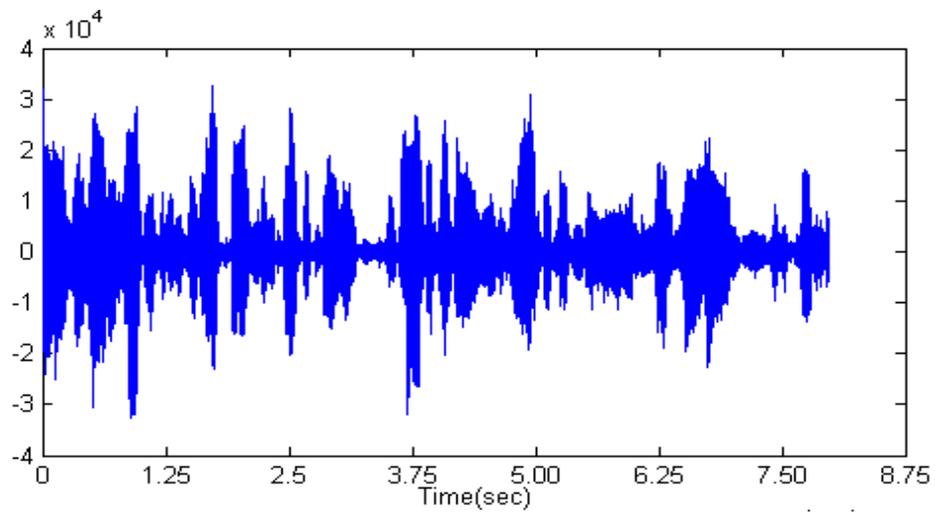


圖 4- 16 分離結果 y1 , SNR=13.3790 db

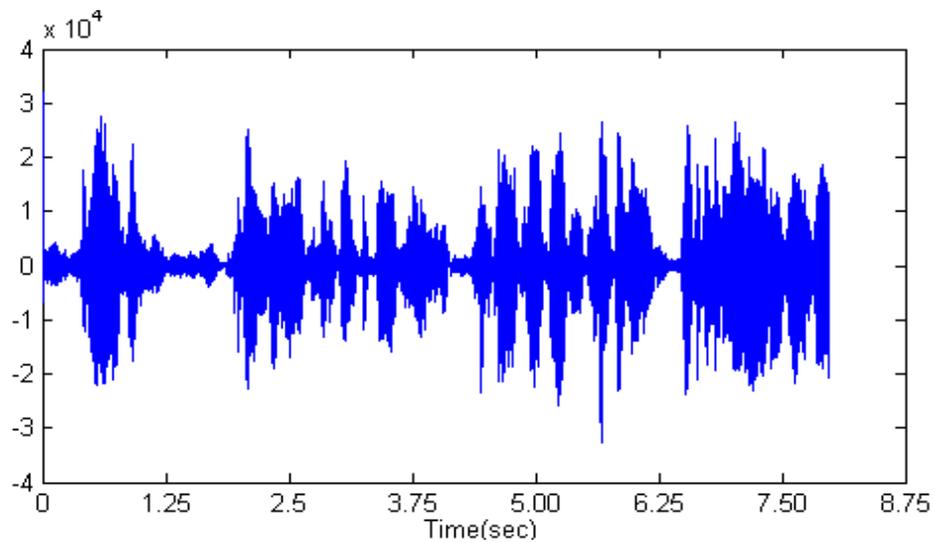


圖 4- 17 分離結果 y2 , SNR=18.9306 db

模擬三：女聲+雜訊

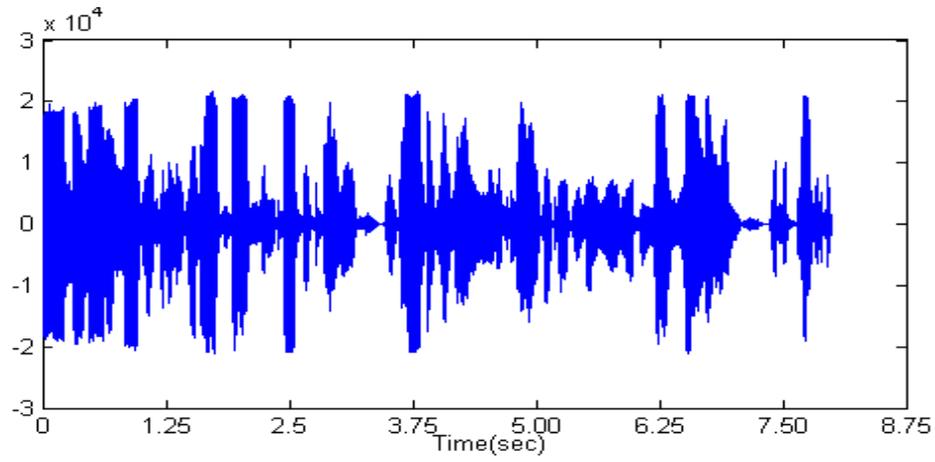


圖 4-18 聲源訊號 s_1

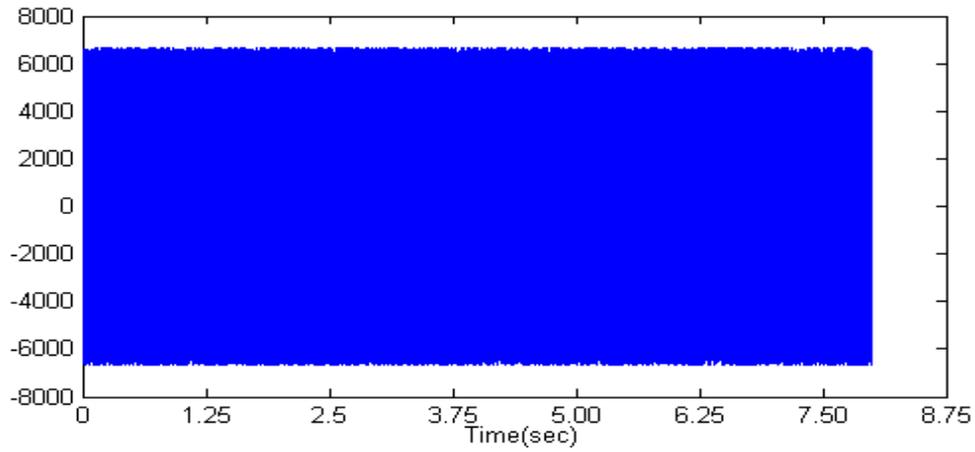


圖 4-19 雜訊 s_2

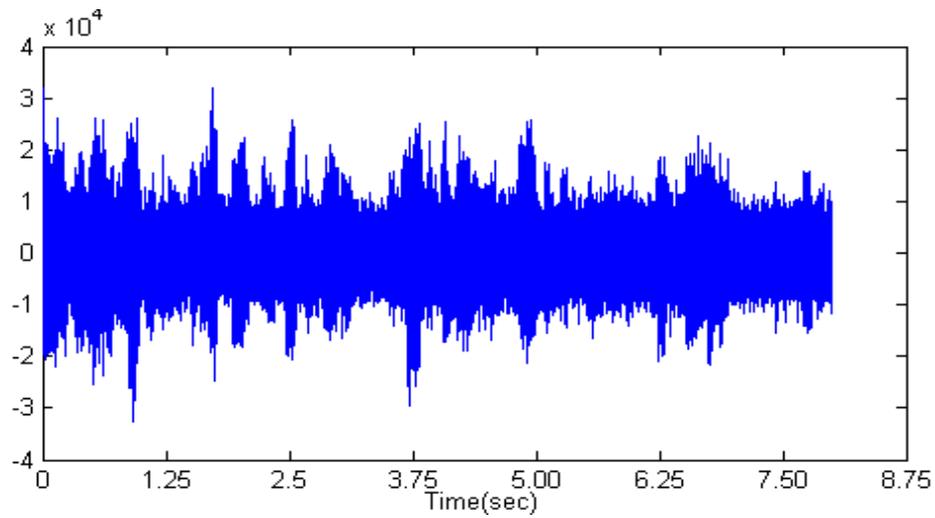


圖 4-20 混和訊號 x_1

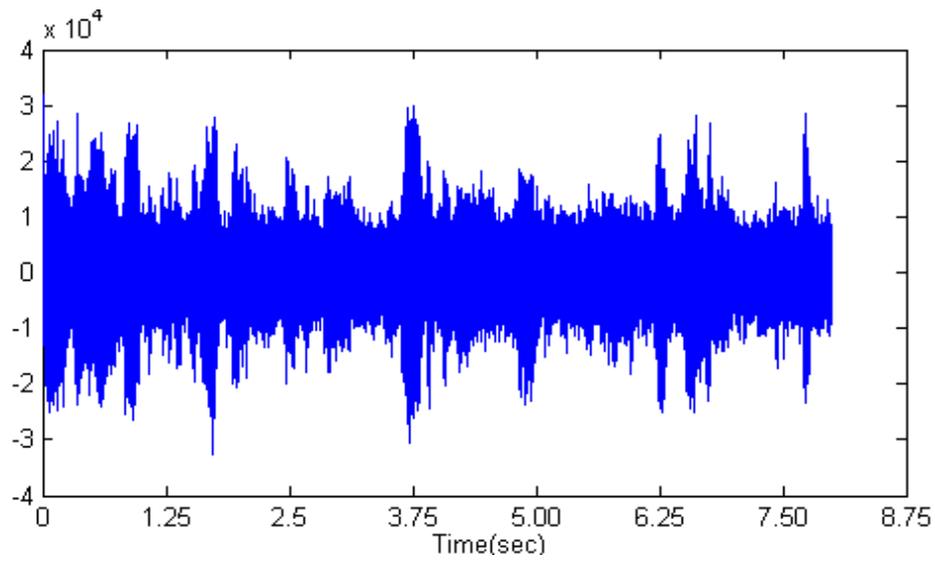


圖 4-21 混和訊號 x_2

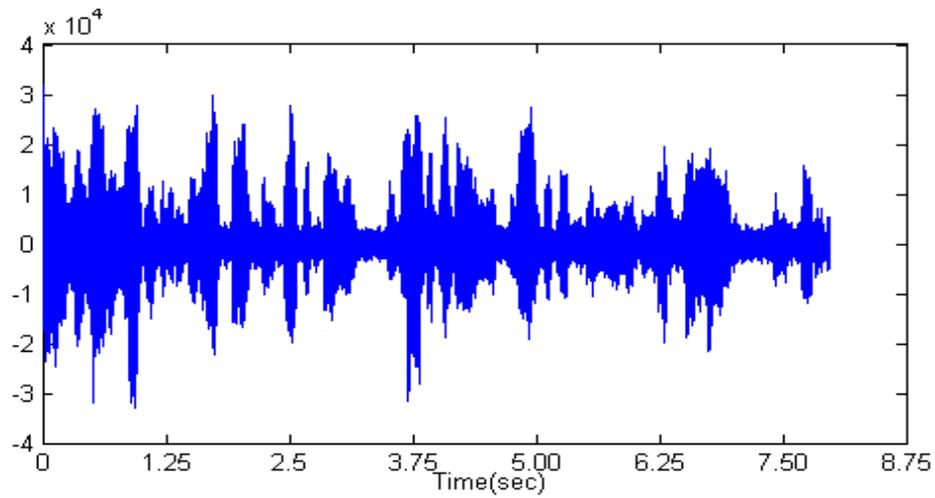


圖 4-22 分離訊號 y_1 , SNR=18.1304 db

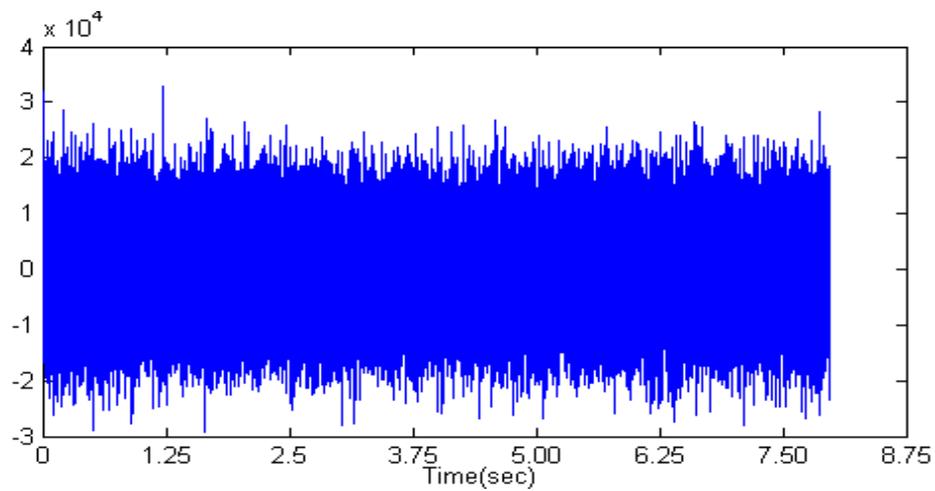


圖 4-23 分離訊號 y_2 , SNR=21.1537 db

4.2 利用空間預先估計之 demixing matrix 萃取移動中音源

4.2.1 原理

已知分離結果為

$$\begin{bmatrix} \hat{s}_1(w) \\ \hat{s}_2(w) \end{bmatrix} = \begin{bmatrix} w_{11}(w) & w_{12}(w) \\ w_{21}(w) & w_{22}(w) \end{bmatrix} \begin{bmatrix} a_{11}(w) & a_{12}(w) \\ a_{21}(w) & a_{22}(w) \end{bmatrix} \begin{bmatrix} s_1(w) \\ s_2(w) \end{bmatrix} \quad (4.7)$$

於 2.9 式曾提及 $WA=PD$ ，即

$$\begin{bmatrix} w_{11}(w) & w_{12}(w) \\ w_{21}(w) & w_{22}(w) \end{bmatrix} \begin{bmatrix} a_{11}(w) & a_{12}(w) \\ a_{21}(w) & a_{22}(w) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \text{或} \begin{bmatrix} 0 & c \\ d & 0 \end{bmatrix} \quad (4.8)$$

當 s_1 移動時，其與麥克風間之 transfer function (a_{11} 及 a_{21}) 亦會改變，因此 4.8 式可改寫為

$$\begin{bmatrix} w_{11}(w) & w_{12}(w) \\ w_{21}(w) & w_{22}(w) \end{bmatrix} \begin{bmatrix} a'_{11}(w) & a_{12}(w) \\ a_{21}(w) & a_{22}(w) \end{bmatrix} = \begin{bmatrix} a' & 0 \\ e & b \end{bmatrix} \text{或} \begin{bmatrix} f & c \\ d' & 0 \end{bmatrix} \quad (4.9)$$

而

$$\begin{bmatrix} a' & 0 \\ e & b \end{bmatrix} \begin{bmatrix} s_1(w) \\ s_2(w) \end{bmatrix} = \begin{bmatrix} a' s_1 \\ e s_1 + b s_2 \end{bmatrix} \text{或} \begin{bmatrix} f & c \\ d' & 0 \end{bmatrix} \begin{bmatrix} s_1(w) \\ s_2(w) \end{bmatrix} = \begin{bmatrix} f s_1 + c s_2 \\ d' s_1 \end{bmatrix} \quad (4.10)$$

由此可看出當 s_1 移動時，可分離出 s_1 ；同理若 s_2 移動則可分離出 s_2 。

4.2.2 分離結果

在此模擬中，聲源 s_2 仍為 noise，在 -30° 處；而 s_1 為語音，從 30° 移動到 45° ，其關係如圖 4-24。根據 4.2.1 之推論，此分離結果應可將移動之 s_1 分離出來，而 s_2 仍與 s_1 混在一起。

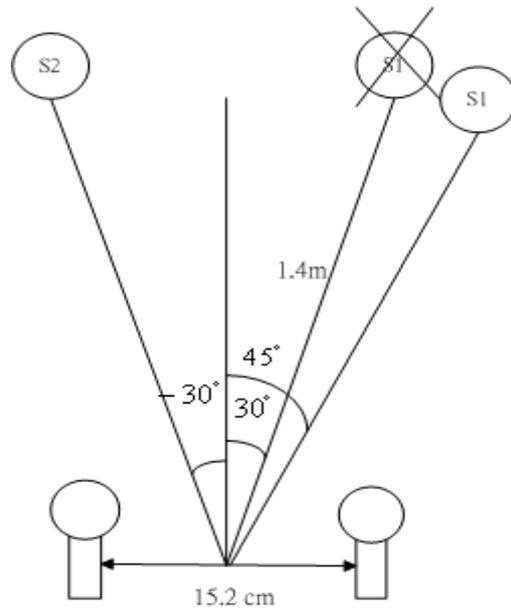


圖 4-24 聲源與麥克風關係圖

s1 如 4-18，s2 如 4-19，混和訊號如 4-25 與 4-26，分離出之訊號如 4-27 與 4-28，其中 4-27 之 SNR 為 16.6879 db；4-28 之 SNR 為 5.1589 db，由此可知對於移動之音源，此方法確實可將移動之音源分離出來。

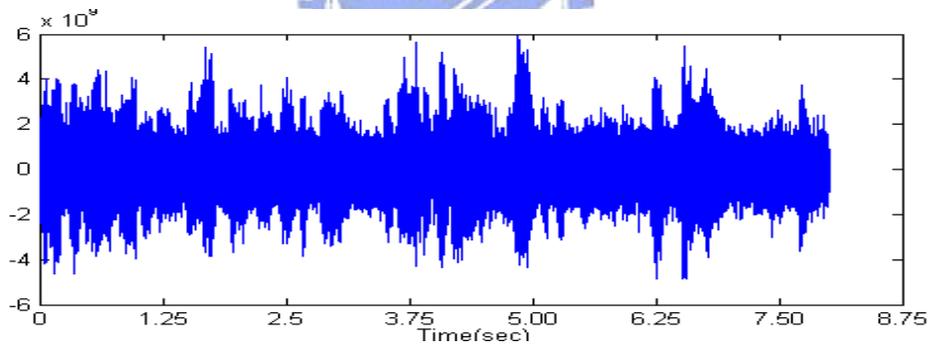


圖 4-25 混和訊號 x1

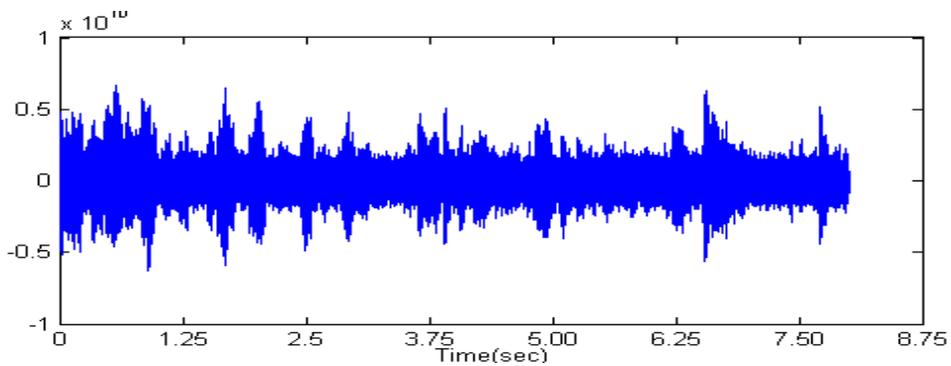


圖 4-26 混和訊號 x2

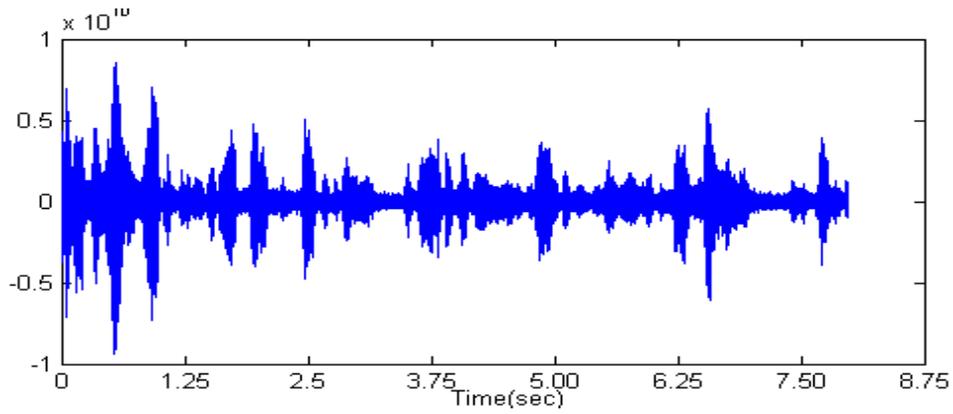


圖 4- 27 分離訊號 y1

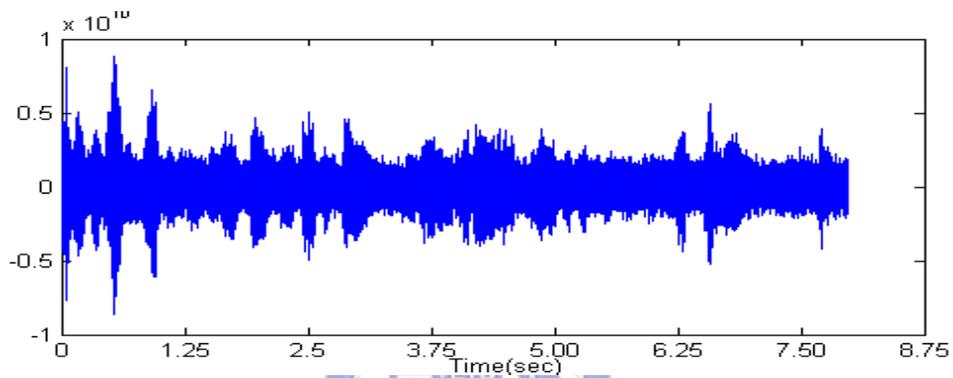


圖 4- 28 分離訊號 y2

若 noise 固定於 -30° ，s1 由 5 度移動到 360 度，每次移動五度，可將分離出之 y1 之 SNR 對於 s1 角度之關係描出如圖 4-24，其中除了 noise 之角度(330 度)之外，每個角度之 SNR 皆在 13.6615 db 以上。

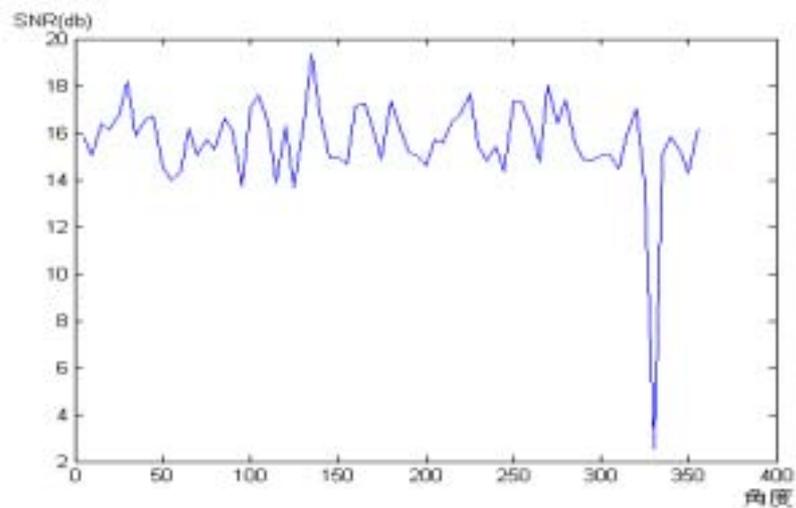


圖 4- 29 SNR 與 s1 角度之關係

第五章 結論與未來工作

5.1 結論

在本論文中，利用 ICA 來尋找 demixing matrix，以達成語音訊號分離之目標。由於訊號混和過程可視為 channel 與訊號之 convolution 效應，在此前提之下，本論文採用了 frequency domain ICA 來對每個頻率尋找其相對應之 demixing matrix，並且對於 frequency domain ICA 之兩大問題：permutation problem 與 dilation problem，整理了兩種做法並加以探討其中之優劣。

由於一般之聲源訊號無法在每個頻率之下皆滿足 FastICA 可分離之條件(實部與虛部為相同 variance 且 uncorrelated)，造成分離效果不盡理想，於是在此論文中設計出一組滿足其條件之 training signal，由 4.1.4 之分離結果中，可看出以此 training signal 找出之 demixing matrix 對於訊號分離有不錯的效果。除此之外利用 demixing matrix，亦可分離在環境中移動之聲源，此結論可應用在當有一語者說話，而伴隨著一雜訊時，即使語者沒有固定在其原來的位置，只要雜訊位置是固定的(例如：冷氣機，風扇....等發出噪音之家電)，依舊可將語音分離出來。

5.2 未來工作

在本研究中，以模擬的方式完成了兩個聲源對兩個麥克風之語音分離。未來可基於本架構，擴充為多個聲源對多個麥克風之語音分離，並且錄實際上之混和語音，將分離效果與其他方法(例如：beamforming)做比較。

參考文獻

- [1] P. Comon. "Independent component analysis-a new concept?" ,*Signal Processing*,36:287-314,1994
- [2] C. Jutten., "Source separation : from dusk till dawn," In Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation(ICA'2000), pages 15-26, Helsinki, Finland, 2000
- [3] Shalvi, O., Weinstein, E., "New criteria for blind deconvolution of nonminimum phase systems (channels)" Information Theory, *IEEE Transactions on* , Volume: 36 , Issue: 2 , March 1990 Pages:312 – 321
- [4] A.J. Bell and T.J. Sejnowski . " A non-linear information maximization algorithm that performs blind separation." In *Advances in Neural Information Processing System 7*, pages 467-474. The MIT Press, Cambridge ,MA, 1995
- [5] A.J. Bell and T.J. Sejnowski . "An information-maximization approach to blind separation and blind deconvolution." *Neural Computation*,7:1129-1159,1995
- [6] Aapo Hyvärinen, "A family of fixed-point algorithms for independent component analysis." In Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3917-3920, Munich, Germany, 1997
- [7] Aapo Hyvärinen, E. Oja. "A fast fixed-point algorithm for independent component analysis." *Neural Computation*,9(7):1483-1492,1997
- [8] F. Sattar, M.Y. Siyal, L.C. Wee and L.C. Yen, "Blind source separation of audio signals using improved ICA method", *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on* , 6-8 Aug. 2001 , pages:452 – 455
- [9] 林巧苑, "獨立成分分析法應用於磁振腦血流灌注研究之評估," 國家圖書館, 民 90
- [10] Aapo Hyvärinen, "Independent Component Analysis", John Wiley, 2001
- [11] 張嘉芳, "以 FastICA 為基礎之時域聲音分離演算法", 交大碩士論文, 2003
- [12] Mansour, A.; Jutten, C. , " What should we say about the kurtosis?" *Signal Processing Letters, IEEE* , Volume: 6 , Issue: 12 , Dec. 1999 , Pages:321 - 322
- [13] Aapo Hyvärinen, "Survey on Independent Component Analysis ", *Neural Computing Surveys* 2:94--128, 1999.
- [14] Aapo Hyvärinen. "New approximations of differential entropy for independent component analysis and projection pursuit ," *In Advances in Neural Information Processing Systems 10*, pages 273-279. MIT Press, 1998.
- [15] Aapo Hyvärinen, "Fast and Robust Fixed-point Algorithm for independent component analysis," *IEEE Transactions on Neural Networks*, Vol.10 , No.3 , pp.626-634 , 1999

- [16] S.Kurita, H. Saruwatari, S. Kajita, K.Takeda, and F.Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP2000*,2000,pp.3140-3143.
- [17] N.Murata,S.Ikeda,andA.Ziehe, "An approach to blind source separation based on temporal structure of speech signals,"*Neurocomput.*, vol. 41,pp. 1-24,2001
- [18] Aapo Hyvärinen, "A Fast Fixed-Point Algorithm for independent component Analysis of Complex Valued Signals,"*International Journal of Neural Systems*, Vol.10, No.1,pp.1-8,2000
- [19] Hiroshi SARUWATARI, Toshiya KAWAMURA, Katsuyuki SAWAI, Atsunobu KAMINUMA, and Masao SAKATA, "Blind source separation based on fast-convergence algorithm using ICA and beamforming for real convolutive mixture," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP2002)*, pp.3097--3100, May 2002.
- [20] H. Gotanda, K. Nobu, T.Koya, K. Kaneda, T. Ishibashi and N. Haratani, "Permutation correction and speech extraction based on split spectrum through FastICA," *4th International Symposium on Independent Component Analysis and Blind source separation (ICA2003)*,April 2003,Nara,Japen
- [21] Bill Gardner and Keith Martin, "HRTF Measurement of a LEMAR Dummy-Head Microphone," MIT Media Lab, May 1994

