國立交通大學

電信工程研究所

碩士論文

使用廣義高斯模型於未知聲源數之訊號分離

Using Generalized Gaussian Mixture Model to Detect Sound Locations of Unknown Number of Sources for Sound Segregation

研 究 生：何立文　　Student: Li-Wen Ho

指導教授：冀泰石 博士　　Advisor: Dr. Tai-Shih Chi

中華民國一百零一年八月

使用廣義高斯模型於未知聲源數之訊號分離

Using Generalized Gaussian Mixture Model to Detect Sound Locations of Unknown Number of Sources for Sound Segregation

研 究 生：何立文　　Student: Li-Wen Ho

指導教授：冀泰石 博士　　Advisor: Dr. Tai-Shih Chi

國立交通大學

電信工程研究所

碩士論文

A Thesis
Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao-Tung University
In Partial Fulfillment of the Requirements
for the Degree of
Master of Science in
Communication Engineering
August 2012
Hsin-Chu, Taiwan, Republic of China

中華民國一百零一年八月

使用廣義高斯模型於未知聲源數之訊號分離

學生：何立文　　指導教授：冀泰石 博士

國立交通大學電信工程研究所

感知訊號處理實驗室

摘要

本論文主要探討一種在未知聲源數目下可分離語音的方法。近年來，利用遮蔽來分離訊號的方法已有許多的研究，但是大部分分離的方法都必須在已知聲源數目的情形下才能運行，這是很不實際的做法。在我們的方法中，我們使用廣義高斯混合模型(generalized Gaussian mixture model)來估測聲源方向(direction of arrival, DOA)的統計直方圖分布情形，進而獲得混合訊號中的聲源數目以及聲音源方向。而在計算廣義高斯模型的相關參數中，我們採用了 E-M 演算法來求取所需的參數。在分離語音的部分，是利用 DOA 權重遮蔽分離法，使用的語音特徵是聲音源來向(DOA)。根據每個 T-F unit 的 DOA 給予各聲源遮蔽相對應位置不同的權重比例，比例就代表著該 unit 中各聲源佔有的成分多寡。

在我們的模擬中，給予兩個麥克風不同延遲與不同衰減的訊號，此舉是為了模擬聲音來自不同的方位，再利用兩個麥克風所收到的混合訊號之間的差異來做分離。論文中除了使用廣義高斯分布，也使用了高斯分布和拉普拉斯分布來估測聲源數，並且與 NPCM 以及 DOA-NPCM 比較。在分離語音部分，也測試了二元遮蔽法、DOA 權重遮蔽法、NPCM 以及 DOA-NPCM。實驗結果顯示，五種方法在空間解析度的表現大同小異，但在正確率上，NPCM、DOA-NPCM 和廣義高斯模型有較好的效果。在分離的比較上， 無論任何情形，DOA 權重遮蔽法有最好的 SDR。

Using Generalized Gaussian Mixture Model to Detect Sound Locations of Unknown Number of Sources for Sound Segregation

Student: Li-Wen Ho Advisor: Dr. Tai-Shih Chi

Institute of Communication Engineering

National Chiao-Tung University

Perception Signal Processing Laboratory

Abstract

In this thesis, we propose a method to separate speech signals from spectrograms of sound mixtures with unknown number of sources. Recently, many sparse source separation algorithms using time-frequency masking have been proposed. However, most of these algorithms demand a known number of mixed sources in advance, which is not convenient in practice. In our proposed method, we first model the histogram of estimated angles of the direction of arrival (DOA) with a generalized Gaussian mixture model (GGMM) for detecting the number of sources and sound locations. The GGMM parameters are estimated using the expectation-maximization (EM) algorithm. Based on DOA information of each time-frequency (T-F) unit of the mixed spectrogram, a DOA-weight mask is estimated for each speech signal. The spectrogram of each speech signal is then extracted using the corresponding mask. In our simulations, speech signals are given different delays and amplitude at two microphones to produce DOA information for different locations. In addition to the generalized Gaussian distribution, the Gaussian distribution and the Laplace distribution are also investigated in modeling the DOA histogram. Two kinds of masks, the binary mask and the DOA-weight mask, are investigated in segregating signals from the mixture. Simulation results are compared with outputs of NPCM and DOA-NPCM. Results show that all methods perform equivalently in tests of spatial resolution. On the other hand, the NPCM, DOA-NPCM and GGMM have higher accuracy in estimating the DOA. For segregation, DOA-weight mask performs the best in most test conditions.

誌　　謝

歲月如梭，轉眼兩年研究生涯匆匆走過，首先我要感謝我的指導教授冀泰石博士，除了教導我碩士論文的研究與方法，更在平時相處討論中，讓我了解面對困難必須勇往直前，而不是逃避退縮，凡事都要有積極進取的心，不該懷有得過且過的人生態度，在此再次致上最誠摯的謝意。

接著，我要感謝實驗室的全體學長姊、同學以及學弟妹，面對我無數次擾人的請教與討論，仍熱心地詳細說明與講解，並提供相關的參考資料，讓我對語音處理有更進一步的了解，並且能夠順利完成碩士論文。

最後要感謝我的家人，除了提供生活上的需求，讓我可以專心於研究上面之外，更常常以電話的方式，關心我的生活起居，鼓勵我不要放棄，讓獨自在新竹的我，不會感到孤獨寂寞。另外，我要特地感謝我的女友，在我失落想放棄時，給予安慰與鼓勵，並給我最大的支持。

要感謝的人實在太多，無法一一表達，在此一併致上最深的感恩。

目　　錄

中文摘要……i

英文摘要…… ii

誌謝……iii

目錄……iv

表目錄……vi

圖目錄……viii

第一章 緒論……1

1.1 研究背景……1

1.2 盲訊號分離法簡介……2

1.3 研究方法簡介……4

1.4 章節大綱……4

第二章 演算法與相關方法介紹……5

2.1 E-M 演算法(expectation-maximization algorithm)……5

2.2 廣義高斯分布(generalized Gaussian distribution)……6

2.3 nonlinear projection and column masking (NPCM)……7

2.4 遮蔽分離法簡介……10

第三章 研究方法……11

3.1 系統架構圖……11

3.2 計算 DOA……..…..11

3.3 機率模型……....13

3.4 E-M 演算法……..……14

3.5 聲源數目與 DOA 估計……..……..…16

3.6 DOA-NPCM……19

3.7 語音分離……20

第四章 模擬結果……23

4.1 實驗設置……....25

4.2 模擬結果……27

4.2.1 模擬一……27

4.2.2 模擬二……30

4.2.3 模擬三……33

4.2.4 模擬四……37

第五章 結論與外來展望……38

參考文獻……...39

表 目 錄

表 3-1　經過(a)15 次遞迴的參數(b)25 次遞迴的參數……...18

表 3-2　經過遞迴後的參數值……..19

表 4-1　時間延遲、強度衰減與 DOA 對照表……..……..25

表 4-2　高斯混合模型初始參數……..…….26

表 4-3　拉普拉斯混合模型初始參數……..26

表 4-4　廣義高斯混合模型初始參數……..26

表 4-5　DOA-NPCM 以及 NPCM 參數設定……..27

表 4-6　(84.5, 78.9)高斯模型估計結果……..28

表 4-7　(84.5, 78.9)拉普拉斯模型估計結果……..28

表 4-8　(84.5, 78.9)廣義高斯模型估計結果……..28

表 4-9　(84.5, 78.9)NPCM 與 DOA-NPCM 估計結果…..……..28

表 4-10　(78.9, 73.2)高斯模型估計結果……..29

表 4-11　(78.9, 73.2)拉普拉斯模型估計結果……..29

表 4-12　(78.9, 73.2)廣義高斯模型估計結果……..29

表 4-13　(78.9, 73.2)NPCM 與 DOA-NPCM 估計結果…..……..29

表 4-14　各種模型估計結果，N_S=1……..30

表 4-15　各種模型估計結果，N_S =2……..31

表 4-16　各種模型估計結果，N_S =3……..31

表 4-17　各種模型估計結果，N_S =4…….....31

表 4-18　各種分離法的分離結果，N_S =2…….....35

表 4-19　各種分離法的分離結果，N_S =3…….....37

表 4-20　各種分離法的分離結果，N_S =4…….....37

圖 目 錄

圖 2-1　廣義高斯分布圖……..…….6

圖 2-2　PCA 與 NPCM 比較圖……..……7

圖 2-3　NPCM 估測聲源數流程圖……..……9

圖 3-1　系統架構圖……11

圖 3-2　DOA 示意圖……...…….12

圖 3-3　DOA 直方統計圖……...…….13

圖 3-4　DOA 直方統計圖……17

圖 3-5　DOA 直方統計圖……17

圖 3-6　DOA 直方統計圖……17

圖 3-7　NPCM 散點圖與各方向加總曲線圖……20

圖 3-8　能量比例與 DOA 轉換曲線圖……21

圖 3-8　實際聲源能量比例與 DOA 散點圖……22

圖 4-1　聲源與麥克風示意圖……23

圖 4-2　聲源到兩麥克風的訊號……23

圖 4-3　訊號來源方向示意圖……25

圖 4-4　高斯混合模型初始分布圖……26

圖 4-5　拉普拉斯混合模型初始分布圖……26

圖 4-6　廣義高斯混合模型初始分布圖……26

圖 4-7　(84.5, 78.9)高斯模型估測分布圖……..……28

圖 4-8　(84.5, 78.9)拉普拉斯模型估測分布圖……...……28

圖 4-9　(84.5, 78.9)廣義高斯模型估測分布圖……...……28

圖 4-10　(78.9, 73.2)高斯模型估測分布圖……..……29

圖 4-11　(78.9, 73.2)拉普拉斯模型估測分布圖……..……29

圖 4-12　(78.9, 73.2)廣義高斯模型估測分布圖……..……29

圖 4-13　DOA 統計圖……31

圖 4-14　整體正確率趨勢圖……..……32

圖 4-15　聲源聲譜圖……...……33

圖 4-16　混合訊號聲譜圖……...…33

圖 4-17　二元分離法分離後的聲源聲譜圖……...…34

圖 4-18　權重分離法分離後的聲源聲譜圖……..……34

圖 4-19　NPCM 分離後的聲源聲譜圖……...…34

圖 4-20　DOA-NPCM 分離後的聲源聲譜圖……...……35

圖 4-21　實際聲源能量比例與DOA 散點圖……...……...…36

圖 4-22　實際聲源能量比例與DOA 散點圖……...……...……36

第一章　緒論

1.1 研究背景

在日常生活中，常常會遇到多個聲源同時出現的情況，例如在一個吵嘈的環境中，語音會夾雜著各式的噪音如手機鈴聲、汽機車噪音等，或者是在同一個空間中，有很多人同時說話，但我們通常都只想要集中在某個人的聲音上面，要如何從吵雜的混合訊號中抽取出我們想要的訊號源，這就是所謂的雞尾酒派對問題(cocktail-party problem)。

為了解決上述的問題，盲訊號分離法(blind source separation, BSS)在近十幾年來成為很熱門的一個研究主題，所謂「盲」指的就是我們只有收到的混合訊號(mixtures)，而訊號源(sources)和混合的過程(mixing process)皆為未知，此方法的目標就是在只有混合訊號的情況下，分離出原本的訊號源。因此盲訊號分離法廣泛地應用於未知訊號的處理，如在生醫訊號處理(biomedical signal processing)的應用，在量測到的腦電波訊號(electroencephalogram, EEG)中，可能混合著肌肉運動、眼球活動、心臟跳動等訊號源，不同位置所量測到的 EEG 訊號即為混合訊號，BSS 的目的便是從這些混合訊號中分離出原本的訊號源。其他應用還有特徵擷取(feature extraction)、通訊(telecommunication)、金融序列分析(financial time series analysis)等[1][2]，在語音訊號處理(audio signal processing)上的應用大部分為語音分離(audio separation)，也就是用來解決前段所提的雞尾酒派對問題，在許多聲源的混合音訊中，個別分離出每個聲音。

此研究主題又可分成兩部分，有些研究是在假設聲源數目為已知的情形下進行，另一研究是將聲源數目當作未知，因此在分離前必須先預估聲源數，再進行音源分離。

1.2 盲訊號分離法簡介

假設有 N_S 個聲音源，N_m 個麥克風，則我們可以將混合的訊號寫成下式：

$$x_j(t) = \sum_{i=1}^{N_s} \sum_l h_{ji}(l) s_i(t-l),\ j = 1, \ldots, N_m \qquad (1\text{-}1)$$

s_i 為各個訊號源，$h_{ji}(l)$代表聲源 i 到麥克風 j 的脈衝響應，我們的目標就是在未知聲源數 N_S、聲音源 s_i 和脈衝響應 h_{ji} 的情形下，只從兩個麥克風收到的聲音來分離出原來的聲音。

因為是在聲譜圖上做處理，因此將(式 1-1)作 short-time Fourier transform (STFT)，可得到如下的結果：

$$X_j(f,\tau) = \sum_{i=1}^{N_s} H_{ji}(f) S_i(f,\tau) \qquad (1\text{-}2)$$

$H_{ji}(f)$代表聲源 i 到麥克風 j 的頻率響應，$S_i(f,\tau)$代表各個聲源的 STFT，τ 是 time-frame index。

在盲訊號分離法中，有兩個方法被廣泛的探討：一是獨立成分分析法[1][2] [18-24] (independent component analysis, ICA)，其立論基礎是假設每個聲源都是互相獨立的，經過混合後不會影響聲音的本質，因此可以估測得到混合反矩陣，也就是分離矩陣，將混合訊號乘上分離矩陣就可以得到分離後的聲音；另一個稱作稀疏成分分析法[4][25-28] (sparse component analysis, SCA)，此方法是假設聲源訊號在某些 domain 是很稀疏的，稀疏的意思是聲音訊號大部分的值都接近 0，也就是說混合訊號中每一個成份點，通常只有一個主要的聲音源存在。

就稀疏成分分析法而言，聲譜圖可視為語音頻率成分隨時間的變化，而不同的人講話會有不同的基頻與倍頻，說話速度與斷句也不同，所以不同音源的聲譜圖交集是很少的(disjoint)，因此混合訊號聲譜圖的每個 T-F unit 都只來自於其中一

個訊號源，也就是有很稀疏(sparse)的特性[5]。利用聲音訊號擁有的稀疏特性，因此(式 1-2)可改寫如下[4] [6]：

$$X_j(f,\tau) \approx H_{ji}(f) S_i(f,\tau) \tag{1-3}$$

而遮蔽分離法，就是找出各個聲源的遮蔽，與混合訊號相乘得到分離後的訊號 Y_i，

$$Y_i(f,\tau) = X_1(f,\tau) M_i(f,\tau) \tag{1-4}$$

M_i 代表每個聲源的遮蔽，最後再將 Y_i 經過 ISTFT 得到分離後的聲音。

盲訊號分離法中還存在一些問題，一個是混合訊號中訊號源的數目，許多分離法都必須要在已知訊號源數目的情形下才能使用，例如 k-means 演算法，就必須先給定一個已知的聲源數 N_S。但是在現實情形中，這是一個困難且有違常理的要求。另一個問題，也就是訊號源的數量 N_S 與麥克風的數量 N_m 可能會不相等。一般最簡單的假設是 $N_m = N_S$，稱為 even-determined problem，此時可利用線性代數的反矩陣特性得到分離後的聲音；當 $N_m > N_S$ 時稱為 over-determined problem，雖然無法直接求得反矩陣，但可使用 pseudo-inversion，依然可以得到分離後的聲音；但是當訊號源數量大於混和訊號數量，也就是 $N_m < N_S$ 時，稱為 under-determined problem，此時就無法直接用矩陣計算將訊號源解回來，因此有一些研究致力於解決這個問題[6-9]。

在最近的研究中，有學者提出一個方法，稱作 nonlinear projection column masking, NPCM [3]，此方法可在未知聲源數的情形下估測聲源數。而在之前的研究中，我們推廣 NPCM 的想法，提出了 frequency bin-wise nonlinear masking algorithm[10]來做語音的分離，但是當時我們假設聲源數目是已知，且只利用了聲譜圖中強度的資訊，導致在聲源很靠近的情況下，分離結果不如預期。因此在

這篇論文中，我們想探討的是，在聲源數目未知的條件下，僅考慮 DOA 資訊的空間解析度的極限為何？在使用 DOA 當作語音特性時，音源分離會有什麼限制或者需要符合的條件？

1.3 研究方法簡介

本論文主要針對如何在分離聲音前，先估計出訊號源的數目，之後再做語音的分離。我們先利用廣義高斯模型來估計 DOA 在直方統計圖的分布，從中得出訊號源的數目，之後再應用聲音源的方向來做語音的分離。在估計廣義高斯模型中，我們採用 E-M 演算法來取得所需的相關參數。我們希望得到的結果是，一個廣義高斯分布就代表一個聲音源，如此一來，我們就可以根據最後得到的廣義高斯分布的數目來得到混合訊號中的聲音源數目。另外也將 NPCM 作調整，使其想法也能應用於 DOA 資訊上，提出 DOA-NPCM 的方法，並與前面提出的廣義高斯模型方法做音源數目預估之比較。

在分離訊號的部分，是根據前面估計的聲源數目與位置，再利用遮蔽方式作分離。在每個 T-F unit 中，各個訊號源成分的強弱會影響其 DOA 偏移的量，因此可以從每個 T-F unit 得出的 DOA，來推得其各個訊號源佔有的成分比例，進而估計出各自的遮蔽，最後與混合訊號相乘得到分離後的聲音。

1.4 章節大綱

本論文的其餘各章內容如下：第二章介紹 E-M 演算法、廣義高斯分布、NPCM，還有常用的遮蔽分離法：二元遮蔽法以及權重遮蔽法。第三章為研究方法的說明，主要在介紹估計聲源數的方法、更新參數使用的 E-M 演算法、DOA-NPCM 以及分離聲音的方。第四章為實驗結果與相關討論，最後第五章是結論以及未來展望。

第二章　演算法與相關方法介紹

2.1 E-M 演算法(expectation-maximization algorithm)

在統計學上，E-M 演算法是為了得到最大似然率(maximum likelihood)或者最大事後機率(maximum a posteriori, MAP)，而利用遞迴式找尋統計模型參數的方法，此模型中還有隱變數(latent variables)。E-M 演算法在 E-step (expectation)以及 M-step (maximization)交替遞迴更新模型參數，E-step 中，是根據當下的模型參數，計算出對數似然率(log likelihood)的期望值；在 M-step 中，是計算出新的參數使得對數似然率的期望值有最大值。在作法上，是將似然函數個別對模型參數微分並令其為零，接著找出個別式子的解，得到更新參數的數學式。

E-M 演算法是由 Arthur Dempster、Nan Laird 和 Donald Rubin[11]在他們 1977 年發表的論文中提出且命名，他們也指出其實此方法之前已經被很多作者在他們特定的研究領域中多次提出過。除此之外，Arthur Dempster 等人在 1977 年的論文中，他們針對更廣泛的問題提出解決方法以及收斂的分析。儘管此方法在更早之前就已被提出，但 Arthur Dempster 等人的該篇論文在 Royal Statistical Society 期刊上仍獲得熱烈的討論。他們建立的 E-M 演算法，在統計分析上已成為非常重要的工具。然而在 1983 年，C. F. Jeff Wu 指出 Arthur Dempster 等人提出的收斂分析是有缺陷的，C. F. Jeff Wu 並且也提出了修正後的收斂分析。

E-M 演算法經常被使用在機器學習和電腦視覺的資料分群。在自然語言處理中，E-M 演算法有兩個重要的實例，分別是 Baum-Welch algorithm (也稱作 forward-backward) 以及 inside-outside algorithm 。在心理學，E-M 演算法是必不可少的工具，用來估計項目反應理論(item response theory)模型中的項目參數以及潛在能力。因為擁有處理 missing data 以及觀察未知變數的能力，E-M 逐漸變成理財投資組合的實用工具。E-M 演算法以及其衍伸出的變種演算法，也被廣泛應用

於醫學影像重建，特別是在正電子發射斷層掃描(positron emission tomography)和單光子發射電腦斷層掃描(single photon emission computed tomography)。

因為 E-M 演算法擁有處理潛在變數以及 missing data 的能力，所以在許多領域都可看見 E-M 的身影，但是其本身也有一些缺點，例如在有許多資料是不可觀察的情形之下，收斂速度就會變慢許多。另一個缺點是經過 E-M 演算法得到的最終結果，可能只是區域性的最佳解(local maximum)。

另外，一些方法也被提出來為了加速 E-M 演算法收斂的速度，例如共軛梯度法(conjugate gradient)和改進的牛頓-拉夫森技術(Newton–Raphson techniques)。期望條件式最大化(expectation conditional maximization, ECM)用條件式最大化步階取代每個 M-step，也就是說每個參數是個別地被最大化，但其他參數保持固定的情形。

2.2 廣義高斯分布(generalized Gaussian distribution)

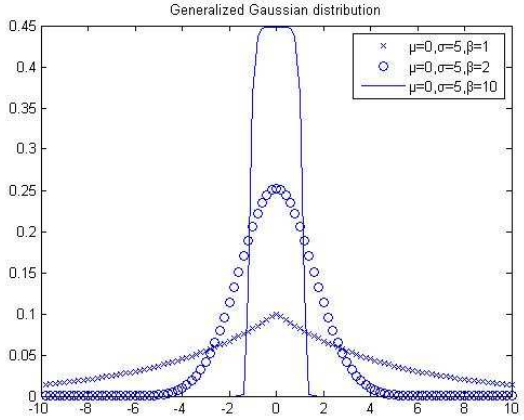


圖 2-1　廣義高斯分布圖

$$f(x)=\frac{\beta}{2\sigma\Gamma(1/\beta)}e^{-\left(\frac{|x-\mu|}{\sigma}\right)^{\beta}} \tag{2-1}$$

$$\Gamma(z)=\int_0^{\infty}e^{-t}t^{z-1}dt \tag{2-2}$$

$$\text{var} = \frac{\sigma^2 \Gamma(3/\beta)}{\Gamma(1/\beta)} \tag{2-3}$$

其中 μ 是平均值，σ 是規模參數(scale parameter)，β 是形狀參數(shape parameter)，$\Gamma(\cdot)$是 Gamma 函數，變異數如(式 2-3)所示。廣義高斯分佈又名廣義常態分布(generalized normal distribution)，高斯分布與拉普拉斯分布是是其中 2 個特殊的例子，也就是說，當 $\beta = 2$ 的時候，就是常態分布，當 $\beta = 1$ 的時候，就是拉普拉斯分布。而當 $\beta = \infty$的時候，它會趨近於平均分布，因此可以將它看作是多種分布的集合。也因為它的自由度比高斯和拉普拉斯兩種分布多一維，因此在估計資料分布時，最終得到的結果，更有機會接近真實的情形。

高斯分佈是一個在數學、物理及工程等領域都非常重要的機率分佈，在統計學的許多方面有著重大的影響力。另外在機率論與統計學中，拉普拉斯分布是以皮埃爾-西蒙·拉普拉斯的名字命名的一種連續機率分布。由於它可以看作是兩個不同位置的指數分布背靠背拼接在一起，所以它也叫作雙指數分布。拉普拉斯分布函數讓我們聯想到常態分佈，但是，常態分佈是用相對於平均值的差的平方來表示，而拉普拉斯用相對於平均值的差的絕對值來表示。因此，拉普拉斯分布的尾部比常態分佈更加平坦，但在峰值的位置比較尖銳。

2.3 nonlinear projection and column masking (NPCM)

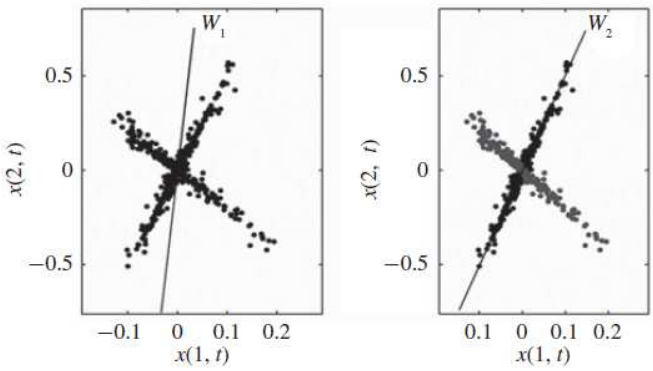


圖 2-2 左圖是 PCA 估計得到的主要方向
右圖是 NPCM 估計得到的主要方向

圖 2-2 是兩個聲源經過一個 2x2 矩陣之後的混合訊號散點圖，從左右兩圖可以看出 NPCM 與 principal component analysis (PCA)的差別，PCA 是將所有取樣點都一起考慮，因此容易受到偏離較大的點的影響，導致最終得到的方向 w_1 有所偏移；但 NPCM 是將偏離較大的點遮蔽後，只將較精準的取樣點進行運算，減少偏移較大點所造成的影響，所以估計的方向 w_2 會更精準。

從(式 2-4)可明顯得知，PCA 的投影量計算是一個餘弦函數：

$$y_t = \|x_t\| * \left|\cos(\widehat{w, x_t})\right| \tag{2-4}$$

目的是要找到一個　　方向 w 使得 $E[y_t^2]$擁有最大值，E[・]是代表期望值。得到的 w 就代表是最主要的方向。

而 NPCM 使用的投影量計算式如(式 2-5)：

$$\max J(\mathbf{w}) = \sum_t \|\mathbf{x}_t\| \exp\left(-\rho \sin^4\left(\widehat{\mathbf{w}, \mathbf{x}_t}\right)\right) \tag{2-5}$$

可以看出投影計算是經過一個非線性指數函數之後的加總，此舉是為了降低偏移大的點影響。而最終目標是找到使得 J(w)有最大值的方向 w，此 w 就是最主要的方向。

延伸以上的想法，加上遞迴的方式，就可以用來估計混合訊號中的聲源數目，以下先介紹 NPCM 的演算法：

步驟一：給定 ρ 、k 和 ε 的初始值，ρ 和 k 是控制遮蔽的範圍，ε 是演算法終止的條件，並令遞迴的次數 p 為 0。

步驟二：根據(式 2-6)更新遮蔽向量 u 的值：

$$u_t^{(p)} = \begin{cases} 1 & p = 0 \\ \left(1 - \exp\left(-\kappa \sin^4\left(\widehat{\mathbf{w}_i, \mathbf{x}_t}\right)\right)\right) u_t^{(p-1)} & p \geq 1 \end{cases} \tag{2-6}$$

步驟三：找到使得 H(w)有最大值的 w：

$$\max H(\mathbf{w}) = \sum_t u_t^{(p)} ||\mathbf{x}_t|| \exp\left(-\rho \sin^4\left(\widehat{\mathbf{w}, \mathbf{x}_t}\right)\right) \tag{2-7}$$

重複步驟二和三直到 H_i / H_{max} 小於 ε 則終止進行，得到的 w 數量就是聲源的數量。

以下舉一個例子說明：ρ =10^6、k=10^4 和 ε =0.4

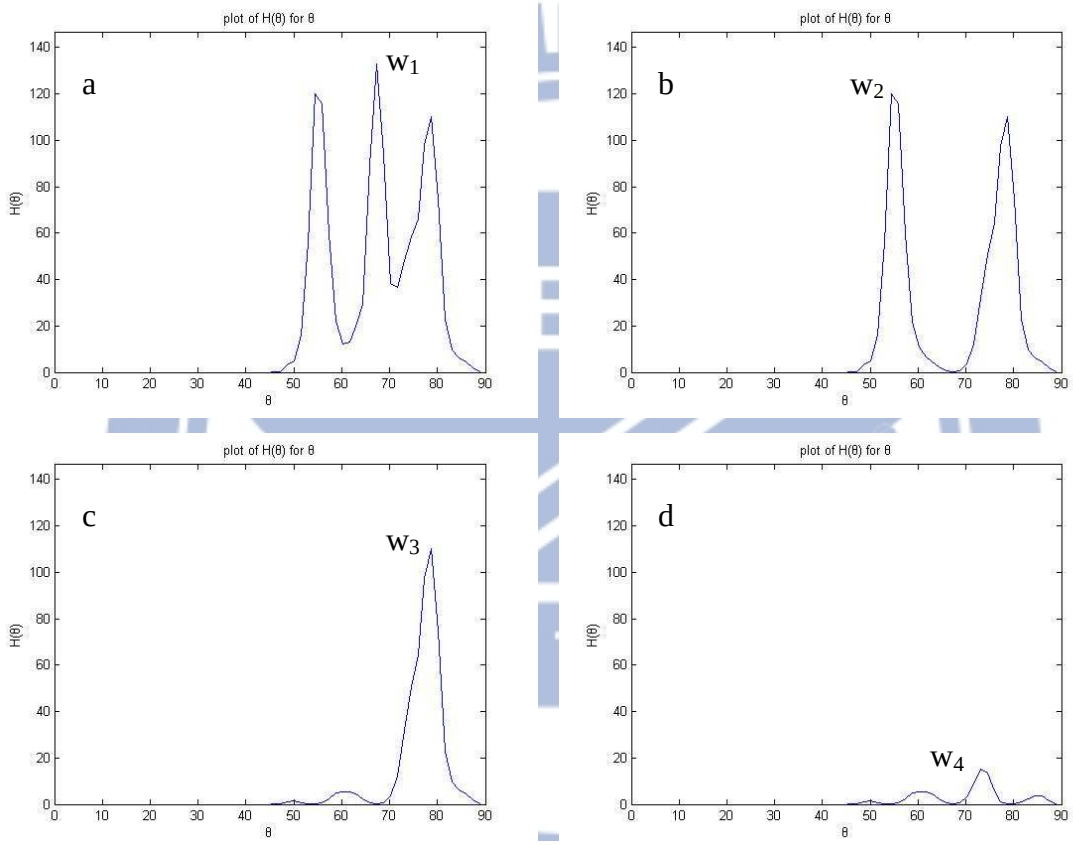


圖 2-3 NPCM 估測聲源數流程圖

圖 2-3 是使用 NPCM 估測混合訊號中聲源數的流程圖，圖 2-3 (a)可以看出有 3 個峰值，先找到最大峰值的位置 w_1，接者更新遮蔽向量 u 將 w_1 附近的取樣點遮蔽，再把剩下的取樣點作下一個方向的估計，以此類推得到 w_2 如圖 2-3 (b)和 w_3 如圖 2-3 (c)，此 3 個峰值分別得到 H_i / H_{max}=1, 0.9 和 0.83，而在圖 2-3 (d)中，得到 w_4 的 H_i / H_{max}=0.11，已小於 0.4 符合演算法終止的條件，因此最後可以得到聲源數目為 3。

2.4 遮蔽分離法簡介

一般而言，使用遮蔽分離法之前，都會先將語音經過 STFT，之後再針對轉換後的訊號做處理。當每個聲源的遮蔽都估計好之後，就將混合語音的轉換後訊號直接跟個別的遮蔽相乘，如此一來就完成了分離的工作，最後再將訊號個別作 ISTFT，轉回時域的訊號就是分離後的聲音。

一般常用的分離遮蔽法分兩種，第一種稱作 hard mask，也稱作二元遮蔽分離法，顧名思義，就是最終得到的遮蔽，其中的數值不是 0 就是 1，這是基於語音的稀疏特性所推論而得。也就是說當混合語音轉到頻域時，每個 T-F uint 中只會有一個主要的聲源成分，而這成分就佔了此 unit 大部分的能量，因此就把這個 unit 分配給該聲源。在作法上，就是在屬於此聲源的遮蔽中，找到相對應的 unit 位置並令其值為 1，而其餘聲源的遮蔽，在這個位置上的值都是 0。此作法雖然簡單快速，但是因為遮蔽中的值只有 0 和 1，所以會有時頻圖不連續的問題，當轉回到時域後，就會有較嚴重的失真(musical noise)。

另一種方法是與二元遮蔽分離法相對應的，稱作 soft mask，又稱權重遮蔽分離法，差異在於，估計得到的各個遮蔽，其中的數值不再只有 0 和 1 而已。遮蔽裡的每個 unit，會是一個介於 0 到 1 的數值，而所有遮蔽同位置 unit 相加的總和等於 1，如下式子所示：

$$0 \le M_i(f,\tau) \le 1 \tag{2-8}$$

$$\sum_{i=1}^{N_s} M_i(f,\tau) = 1 \tag{2-9}$$

數值大小的決定會隨著使用的語音特徵(feature)差異而有不同，在我們的研究中，我們是根據該位置的 DOA 來決定該 unit 中每個聲源的比重。使用這樣的方式是希望藉由權重遮蔽的方式，來改善二元遮蔽法遇到的失真問題。

第三章　研究方法

3.1 系統架構圖

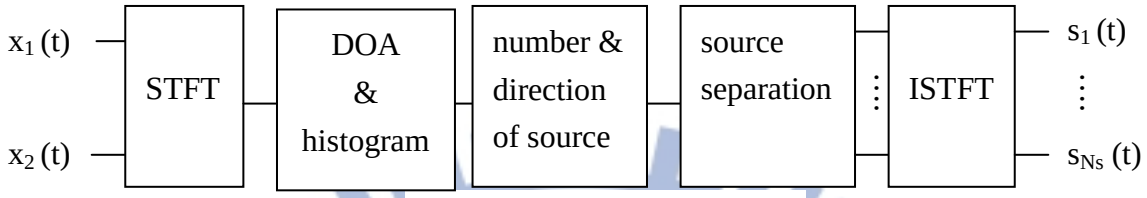


圖 3-1 系統架構圖

首先將 2 個麥克風收到的混合訊號作 STFT，之後利用 2 個聲譜圖的相位差算出每個 T-F unit 的 DOA，並將 DOA 作直方統計圖；接著用 generalized Gaussian mixture model 加上 E-M 演算法估測聲源數以及聲源方位；然後根據前一階估計的結果來求取分離聲源的遮蔽，混合訊號乘上各個遮蔽後就是分離後的聲譜圖；最後再將這些聲譜圖作 ISTFT 回到時域，就得到分離後的語音了。

3.2 計算 DOA

根據[12]這篇論文，我們利用 $X_1(f,\tau)$與 $X_2(f,\tau)$之間的相位差，進而求得 DOA 的資訊：

$$\varphi(f,\tau)=\angle\frac{X_1(f,\tau)}{X_2(f,\tau)} \tag{3-1}$$

$$d(f,\tau)=\cos^{-1}\frac{\varphi(f,\tau)c}{D2\pi f} \tag{3-2}$$

其中的 c 是聲音的速度，在此設定為每秒 340 公尺；D 是兩個麥克風之間的距離，定為 8 公分。下圖是計算 DOA 的示意圖：

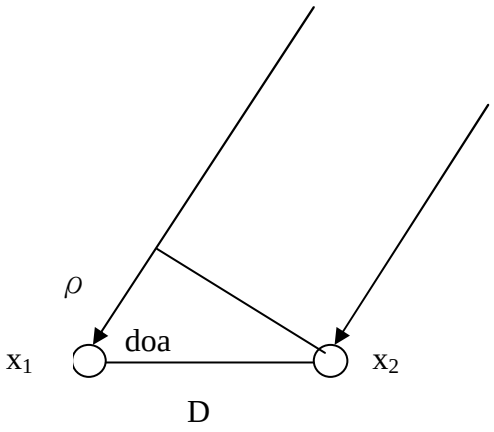


圖 3-2　DOA 示意圖

因為聲源到麥克風的距離遠大於兩個麥克風的間距，因此我們假設聲源到麥克風的路徑是平行線，接著從圖 3-2 可以得知：

$$\cos(doa) = \frac{\rho}{D} \tag{3-3}$$

而 ρ 的距離可以從聲速 c 與兩個麥克風訊號的相位差 φ 計算得到，我們知道相位差可以表示成

$$\varphi = 2\pi ft \tag{3-4}$$

因此

$$\rho = ct = \frac{ct2\pi f}{2\pi f} = \frac{\varphi c}{2\pi f} \tag{3-5}$$

再將(式 3-5)代入(式 3-3)得到

$$\cos(doa) = \frac{\varphi c}{D2\pi f} \tag{3-6}$$

最後再對(式 3-6)取 $\cos^{-1}$，就可以得到(式 3-2)的結果。

以下是一個模擬的例子，設定 3 個聲源分別位於 78.9、61.2 和 47.6 度，根據上述的方法計算 DOA，最後得到如下圖的直方統計圖：

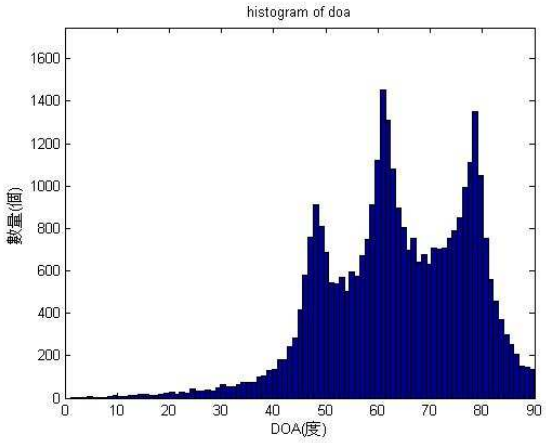


圖 3-3　DOA 直方統計圖

3.3 機率模型

我們假設一個聲源的來向只有一個，因此每個聲源會有各自的 DOA。我們視每個 time-frequency slot 為一個觀察點 d(f , τ)，因此我們會有 N = T x F 個觀察點 {d_1, …, d_n, …, d_N }，T 是 frame 的數目，F 是 frequency bin 的數目。每個觀察點的能量為 a(f, τ) = |X(f, τ)|^2。一般常見的分布情形有高斯分布以及拉普拉斯分布，在此，我們假設每個聲源的 DOA 在統計直方圖上是一個廣義高斯的分布(GGD)，此分布與前述兩者的差異在於多了一個形狀參數 β_m，如此一來，分布情形的自由度更高，更能接近真實的統計直方圖分布，其分布的平均值就代表該聲源的方向。

$$G(d_n \mid \mu_m, \sigma_m, \beta_m) = \frac{\beta_m}{2\sigma_m \Gamma(1/\beta_m)} e^{-(|d_n-\mu_m|/\sigma_m)^{\beta_m}} \tag{3-7}$$

(式 3-7)中的

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt. \tag{3-8}$$

μ_m 為平均值，σ_m 為規模參數，$\Gamma(\cdot)$是 Gamma 函數。

在我們觀察的混合訊號中，我們假設每個聲源的 DOA 都有足夠的觀察點來形成各自的分布，其中有些受到其他聲源的干擾較小，因此 DOA 會接近真實的聲源方向，我們視這些 unit 為 dominant unit，其他受到較大干擾的 unit 會偏移較

大的量，我們視為 less dominant unit。

根據以上的假設，我們得出一個廣義高斯混合模型(GGMM)如下：

$$p(d)=\sum_{m=1}^{M}\alpha_m G(d\mid\mu_m,\sigma_m,\beta_m) \tag{3-9}$$

我們準備足夠數量(M)個廣義高斯分布(G(・))來估計訊號源數目，每個分布佔有的比重是 α_m，滿足 $0 \leqq \alpha_m \leqq 1$ 且每個 α_m 總和為 1。

根據以上所述的模型，我們可以得到如下的對數似然函數(log likelihood function)：

$$\ln p(d\mid\alpha,\mu,\sigma,\beta)=\sum_{n=1}^{N}\ln\left\{\sum_{m=1}^{M}\alpha_m G(d_n\mid\mu_m,\sigma_m,\beta_m)\right\} \tag{3-10}$$

3.4 EM 演算法

我們要得到最接近直方圖的廣義高斯分布，就必須最大化對數似然函數(式 3-10)，在此我們利用 E-M 演算法[13]。E-M 演算法分成兩個階段，在 expectation 和 maximization 之間遞迴更新模型參數。在 E-step 中，我們計算(式 3-11)的值：

$$\gamma(d_{nm})=\frac{\alpha_m G(d_n\mid\mu_m,\sigma_m,\beta_m)}{\sum_{j=1}^{M}\alpha_j G(d_n\mid\mu_j,\sigma_j,\beta_j)} \tag{3-11}$$

在 M-step 中，我們要更新參數以讓對數似然函數有最大值，因此我們將(式 3-10)分別對 μ_m、σ_m 以及 β_m 微分並令其等於 0，首先是將(式 3-10)對 μ_m 微分得到下列的式子：

$$g(\mu_m)=\sum_{n=1}^{N}\gamma(d_{nm})\eta(d_n)\mid d_n-\mu_m\mid^{\beta_m-1}=0 \tag{3-12}$$

(式 3-12)中的 $\eta(d_n)=\begin{cases}1 & \text{if } d_n-\mu_m<0\\ -1 & \text{if } d_n-\mu_m\geq 0\end{cases}$ (3-13)

因為(式 3-12)無法直接求解，因此引入牛頓法，得到下列式子：

$$\mu_m^{(t+1)}=\mu_m^{(t)}-\frac{g(u_m^{(t)})}{g'(u_m^{(t)})} \tag{3-14}$$

(式 3-14)中的

$$g'(u_m^{(t)})=(\beta_m-1)\sum_{n=1}^{N}\gamma(d_{nm})\,|d_n-u_m^{(t)}|^{\beta_m-2} \tag{3-15}$$

但又因為當 β_m 小於 1 時，會有問題，而根據[14]，我們可使用下列式子來更新 μ_m：

$$\mu_m^{new}=\frac{\sum_{n=1}^{N}\gamma(d_{nm})d_n}{\sum_{n=1}^{N}\gamma(d_{nm})} \tag{3-16}$$

接著換將(式 3-10)對 σ_m 微分得到下列的式子：

$$\sum_{n=1}^{N}\gamma(d_{nm})\left[-\frac{1}{\sigma_m}+\beta_m\sigma_m^{-\beta_m-1}\,|d_n-\mu_m|^{\beta_m}\right]=0 \tag{3-17}$$

因此

$$\sigma_m^{new}=\left[\frac{\beta_m\sum_{n=1}^{N}\gamma(d_{nm})\,|d_n-\mu_m^{new}|^{\beta_m}}{\sum_{n=1}^{N}\gamma(d_{nm})}\right]^{1/\beta_m} \tag{3-18}$$

接著將(式 3-10)對 β_m 微分得到下列的式子：

$$\phi(\beta_m)=\sum_{n=1}^{N}\gamma(d_{nm})\left[\frac{1}{\beta_m}+\frac{1}{\beta_m^2}\psi(1/\beta_m)-\left(\frac{|d_n-\mu_m^{new}|}{\sigma_m^{new}}\right)^{\beta_m}\ln\left(\frac{|d_n-\mu_m^{new}|}{\sigma_m^{new}}\right)\right]=0 \tag{3-19}$$

(式 3-19)中的 $\psi(\upsilon)=\frac{\Gamma'(\upsilon)}{\Gamma(\upsilon)}$ 稱作 digamma function [15] (3-20)

因為(式 3-19)無法直接求解，因此引入牛頓法，得到下列式子：

$$\beta_m^{(t+1)}=\beta_m^{(t)}-\frac{\varphi(\beta_m^{(t)})}{\varphi'(\beta_m^{(t)})} \tag{3-21}$$

(式 3-21)中的

$$\varphi'(\beta_m^{(t)})=\sum_{n=1}^{N}\gamma(d_{nm})\left[-\frac{1}{(\beta_m^{(t)})^2}-\frac{\psi'(1/\beta_m^{(t)})}{(\beta_m^{(t)})^4}-\frac{2\psi(1/\beta_m^{(t)})}{(\beta_m^{(t)})^3}-\left(\frac{|d_n-\mu_m^{new}|}{\sigma_m^{new}}\right)^{\beta_m^{(t)}}\left[\ln\left(\frac{|d_n-\mu_m^{new}|}{\sigma_m^{new}}\right)\right]^2\right] \tag{3-22}$$

(式 3-22)中的

$$\psi'(\upsilon)=\frac{d^2}{d\upsilon^2}\ln\Gamma(\upsilon)$$ 稱作 trigamma function [15] (3-23)

最後是 α_m 的更新：

$$\alpha_m^{new}=\frac{\sum_{n=1}^{N}\gamma(d_{nm})}{N} \tag{3-24}$$

完成 4 個參數的更新後，回到 E-step 的步驟，繼續下一次的遞迴。

3.5 聲源數目與 DOA 估計

在估計聲源數及方位時，我們並不是將聲譜圖上所有 unit 的 DOA 都拿來作直方統計圖，我們只取大於聲譜圖中最大能量值 0.3 倍的 unit 來作 DOA 統計。此舉是因為能量大的 unit 為 dominant unit 的機率比較高，如此一來，就可以更突顯統計圖上的峰值，提升估計的效能。下圖是一個例子，可以輕易看出選擇能量較大的 unit 來做統計圖，會比全部 unit 來做統計圖，更突顯峰值；選擇 0.3 倍是比較適合的，因為倍數太小，突顯的效果還是不夠明顯，如果倍數太大，則取得的觀察點太少。

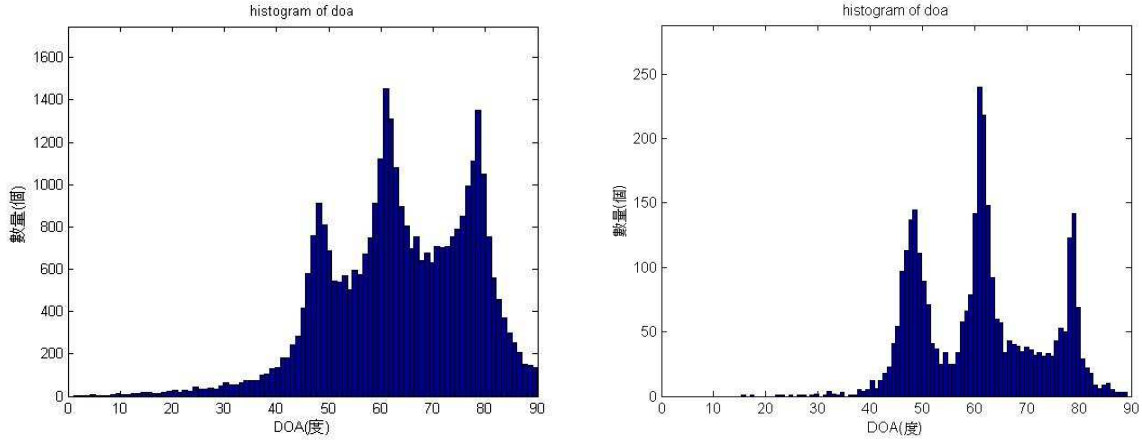


圖 3-4　左圖是聲譜圖上所有 unit 的 DOA 統計圖
右圖是能量大於 0.1 倍最大能量的 unit DOA 統計圖

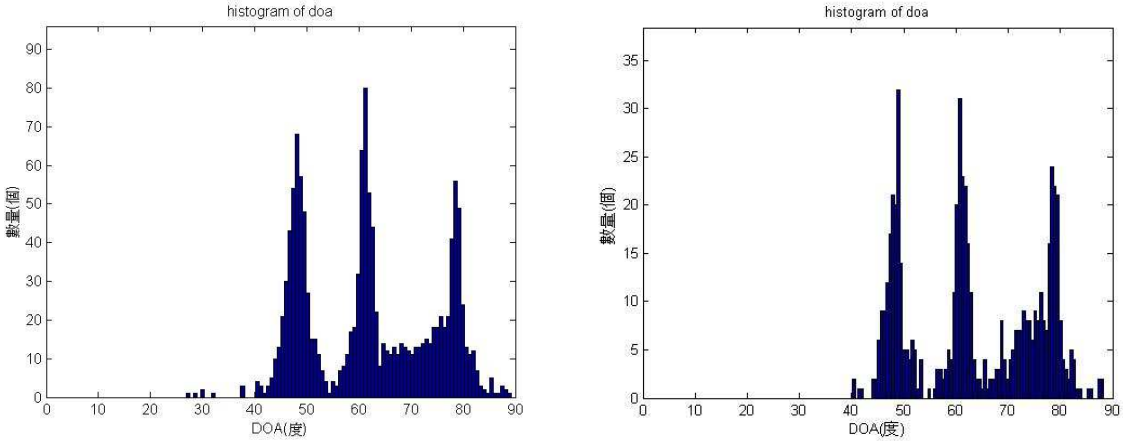


圖 3-5　左圖是能量大於 0.2 倍最大能量的 unit DOA 統計圖
右圖是能量大於 0.3 倍最大能量的 unit DOA 統計圖

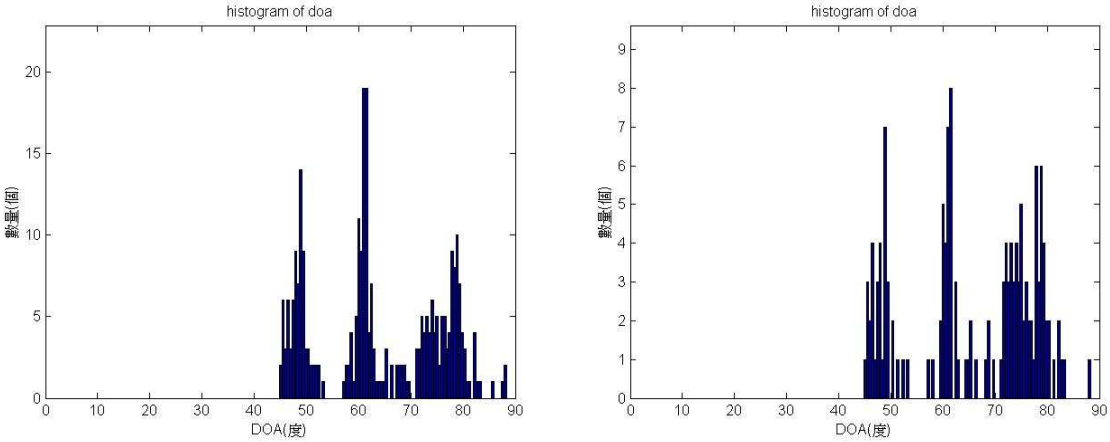


圖 3-6　左圖是能量大於 0.4 倍最大能量的 unit DOA 統計圖
右圖是能量大於 0.5 倍最大能量的 unit DOA 統計圖

表 3-1　經過(a)15 次遞迴的參數(b)25 次遞迴的參數

(a) 經過 15 次遞迴								
m	1	2	3	4	5	6	7	8
μ_m	13.4	**29.7**	32.0	41.2	**47.2**	**61.7**	66.0	**76.8**
σ_m	0	4.23	1.52	4.83	3.81	2.38	9.24	4.32
β_m	2.17	1.41	19.1	4.62	0.97	0.90	1.10	1.26
α_m	~0	**0.15**	0.02	~0	**0.33**	**0.24**	0.05	**0.22**
(b) 經過 25 次遞迴								
μ_m	13.4	**30.4**	32.0	41.1	**47.5**	**61.1**	65.2	**76.1**
σ_m	0	4.55	1.31	5.37	3.33	1.64	6.60	5.18
β_m	2.17	1.33	33.4	4.11	1.00	0.87	1.05	1.30
α_m	~0	**0.17**	0.01	~0	**0.32**	**0.21**	0.04	**0.25**

表 3-1 顯示一個 $N_S = 4$，角度組合為(78.9, 61.2, 47.6, 29.9)的例子，表格中的數值是廣義高斯混合模型參數經過(a) 15 次與(b) 25 次遞迴之後的結果。如果我們把 α_m 的門檻定為 0.1，α_m 大於 0.1 的分布視為有效聲源，忽略 α_m 小於 0.1 的分布，則從表格中可看出 15 次遞迴與 25 次遞迴得到的結果是一樣的，都可判斷出聲源數是 4。另外，15 次遞迴與 25 次遞迴得到的其他參數數值差異也不大，因此我們認為 15 次遞迴已是足夠的了，且判定該分布是否為聲源的標準就是 α_m 是否大於 0.1。

但在測試中，有時會遇到 α_m 大於 0.1 的分布中，其各自的平均值(μ_m)很相近(差距小於 3.5 以內)，也就是代表同一個有效聲源，因此在遇到此類情形時，我們會將其視為同一個聲源，而該聲源的 DOA 為那些代表同一聲源的分布，其各自 μ_m 的平均值。以下就是一個例子，$N_S = 2$，角度組合為(78.9 , 47.6)：

表 3-2　經過遞迴後的參數值

m	1	2	3	4	5	6	7	8
μ_m	0	0	44.5	**46.1**	**48.6**	56.6	68.3	**78.5**
σ_m	0	0	0.99	1.19	1.32	8.05	7.73	0.92
β_m	2	1.72	1.45	2.21	22.5	1.67	1.09	0.72
α_m	0	0	0	**0.15**	**0.18**	0.07	0.07	**0.52**

從表 3-2 可以看出，第 4 個和第 5 個分布的 α_m 都大於 0.1，我們視其為有效聲源，但兩者的 μ_m 相差很小，符合差距小於 3.5 以內的條件，因此我們將其視為同一個聲源，且該聲源的 DOA 以兩個 μ_m 的加權平均代表：

$$\frac{0.18*48.6}{0.18+0.15}+\frac{0.15*46.1}{0.18+0.15}=47.46$$

3.6 DOA-NPCM

從前面篇章的介紹中，可以得知 NPCM 是利用聲譜圖中強度資訊作為估計混合訊號中聲源數目的依據，在此，我們根據 NPCM 的想法，提出了可以使用 DOA 資訊來作為估計依據的方法，得出的結果除了聲源數目外，還能獲得聲源方位。

從 NPCM 的數學式(式 3-25)中，我們發現 x_t 是一個 2 維向量，可以與 w 形成一個夾角，根據夾角大小給定該取樣點的權重。

$$\max J(\mathbf{w})=\sum_t \|\mathbf{x}_t\| \exp\left(-\rho \sin^4\left(\widehat{\mathbf{w},\mathbf{x}_t}\right)\right) \qquad (3\text{-}25)$$

但是 DOA 的資訊是一個純量，無法與 w 形成夾角進行之後的計算，因此我們將每個 T-F unit 中的 DOA 取餘弦值(cosine)和正弦值(sine)分別為橫軸和縱軸座標，也就是將原本為純量的 DOA 轉為一向量，如此一來，就能與 w 形成夾角，

進行之後的計算與估計。

以下舉一個例子說明：左圖為 3 個不同方位聲源混合成的訊號經過餘弦和正弦轉換後所畫成的散點圖；右圖為該混合訊號經過 DOA-NPCM 計算後畫出的每個方向加總曲線圖。

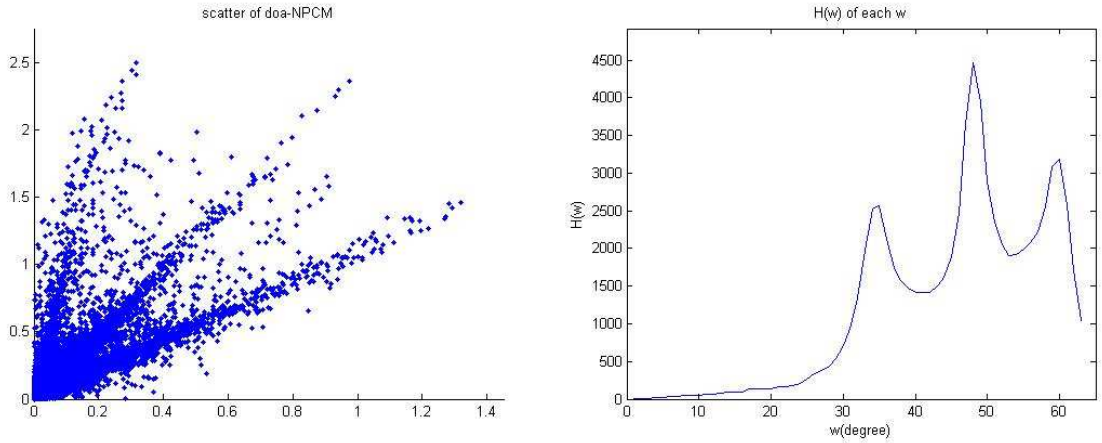


圖 3-7 左圖是散點圖
右圖是每個方向的加總曲線圖

3.7 語音分離

根據[16]，當有兩個聲源的時候，混合訊號的 ITD (interaural time differences) 與兩個聲源的振幅會有下列式子的關係：

$$\delta = \frac{d_1 + d_2}{2} + \frac{1}{f}\arctan\left[\frac{(A_2^2 - A_1^2)}{(A_2^2 + A_1^2)}\tan\phi\right] \qquad (3\text{-}26)$$

其中 d_1 和 d_2 分別是兩個聲源到達兩個麥克風的時間差，A_1 和 A_2 分別是兩個聲源的振幅，f 是頻率，δ 是 ITD，而 ϕ 的定義如下：

$$\phi = f\frac{(d_2 - d_1)}{2} \qquad (3\text{-}27)$$

另外，我們令

$$R = \frac{A_1}{A_1 + A_2} \tag{3-28}$$

(式 3-26)得到的 ITD 經過類似 3-2 介紹的 DOA 計算，就可得到以下的結果：

$$DOA = \cos^{-1}\frac{\delta c}{D} \tag{3-29}$$

接著我們將以上的關係式延伸應用到 T-F unit 上，也就是說，每個 T-F unit 都可以根據 d(f, τ)的值，找到相對應的 R，進而得出兩個聲源的強度比例，因此估計出各自的遮蔽，最後將混合的語音分離，此為權重遮蔽分離法。

若是 d(f, τ)的值不在轉換曲線的 DOA 上下限之內，則不適用以上的關係式，改成將該 unit 歸類給其中一個聲源。以圖 3-8 作一個例子說明，圖中是兩個聲源位於 DOA 為 78.9 度和 61.2 度的轉換曲線圖，假如 d(f, τ)大於 78.9 度，則我們令 78.9 度聲源的遮蔽該 unit 為 1，另一個遮蔽的 unit 為 0；如果 d(f, τ)小於 61.2 度，則我們令 61.2 度聲源的遮蔽該 unit 為 1，另一個遮蔽的 unit 為 0。

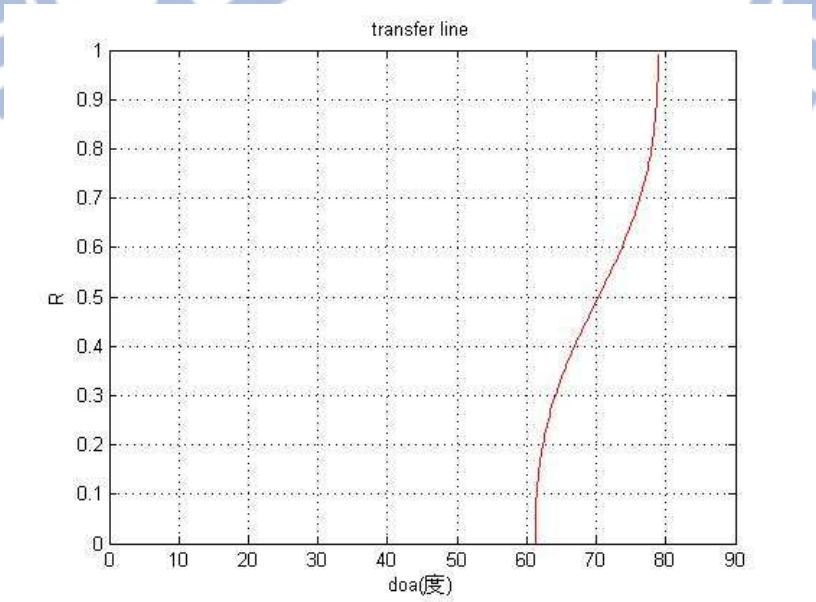


圖 3-8　聲源位於 78.9 度和 61.2 度的能量比例與 DOA 轉換曲線圖

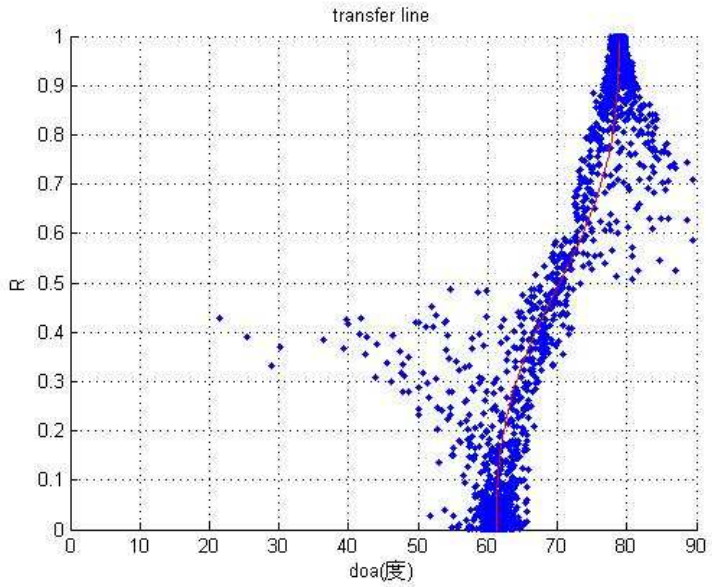


圖 3-9　實際聲源能量比例與 DOA 的散點圖，圖中曲線是(式 3-29)

另外是二元遮蔽分離法的說明，得知兩個聲源的 DOA 之後，可進一步算出兩個 DOA 的平均值，我們根據 d(f, τ)大於或小於該平均值，來判斷該 unit 是屬於哪個聲源。假如已知聲源 1 位於 78.9 度和聲源 2 位於 61.2 度，則可以得到平均值是 70.05 度，因此對於 d(f, τ)大於 70.05 度的 unit，我們將其歸類為聲源 1，所以聲源 1 的遮蔽該 unit 為 1，另一個遮蔽為 0；而 d(f, τ)小於 70.05 度的 unit，我們將其歸類為聲源 2，所以聲源 2 的遮蔽該 unit 為 1，另一個遮蔽為 0。此做法得到的遮蔽，其中的值只有 0 和 1，是為二元分離法。

第四章　模擬結果

我們使用時間延遲不同與強度衰減不同的語音當做聲音源，也就是模擬當聲音源來自某個方位角時，兩個麥克風所收到聲音的時間差與強度差。

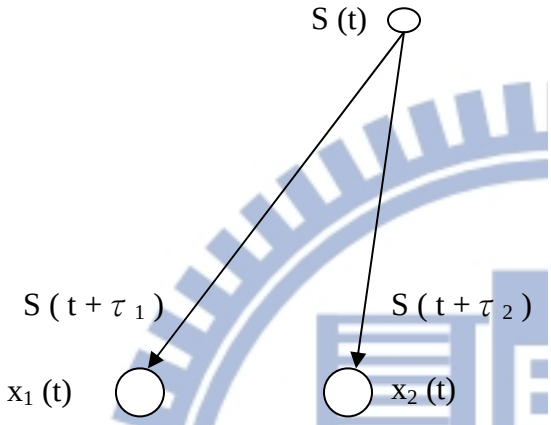


圖 4-1 聲源與麥克風示意圖

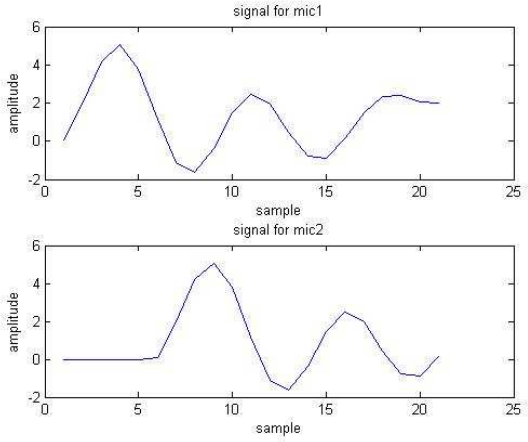


圖 4-2 聲源到兩麥克風的訊號

當聲源位於不同的位置時，對於兩個麥克風會有不同的時間延遲和強度衰減，因此我們的混合訊號可寫成(式 4-1)：

$$x_1(t) = s_1(t+\tau_{11}) + s_2(t+\tau_{12}) + \cdots + s_N(t+\tau_{1N})$$
$$x_2(t) = a_1 s_1(t+\tau_{21}) + a_2 s_2(t+\tau_{22}) + \cdots + a_N s_N(t+\tau_{2N}) \quad (4\text{-}1)$$

針對時間延遲和強度衰減不同的混合訊號，我們嘗試了前面章節所提過的三種機率分布模型來估計訊號源的數目：第一種是高斯分布，第二種是拉普拉斯分布，第三種是廣義高斯分布，並且與 NPCM 和 DOA-NPCM 比較。

在估計聲源數能力的評比上，我們從兩個面向來作比較：首先是可分辨出不同聲源的最小角度差，能分辨的角度差越小，就表示在空間上的解析度越高；最後是正確率的比較，正確率高的表示其估測得到的結果越可靠。

在正確率的計算上，我們採取全對或全錯的標準，也就是說，估測得到的結果，聲源數目一定要符合，多一個或少一個都算錯；估測得到的 DOA 也必須每個都正確，就算錯一個還是算全錯，DOA 有正負 3 度的容許值。

$$\text{正確率}(\%)=\frac{\text{正確次數}}{\text{全部測試次數}}*100\% \tag{4-2}$$

經過前面估計得到的聲源數以及聲源方位後，再用兩種遮蔽分離法進行語音分離：分別是二元遮蔽分離法以及權重遮蔽分離法，並且與 NPCM 以及 DOA-NPCM 的結果作比較。

在盲訊號分離法中，大多會用 SDR(signal to distortion ratio)、SAR(signals to artifact ratio)、SIR(signal to interference ratio)作為效能的評比，經過演算法所分離出的訊號 $\hat{s}(t)$可分解成(式 4-3)：

$$\hat{s}(t)=s_{target}(t)+s_{interf}(t)+s_{noise}(t)+s_{artif}(t) \tag{4-3}$$

其中 $s_{target}(t)$為希望分離出的目標音源，$e_{interf}(t)$為目標音源以外的音源，也就是沒有分離乾淨所產生的干擾，$e_{noise}(t)$為音源以外的雜訊干擾，$e_{artif}(t)$為經由演算法所產生的人為干擾，如 musical noise。我們使用 C.F´evotte 所提供的 BSS_EVAL Toolbox[17]將 $\hat{s}(t)$分解後，SDR、SAR、SIR 的定義如下：

$$SDR=10\log_{10}\frac{\left\|s_{target}\right\|^2}{\left\|s_{interf}+s_{noise}+s_{artif}\right\|^2} \tag{4-4}$$

$$SAR=10\log_{10}\frac{\left\|s_{target}+s_{interf}+s_{noise}\right\|^2}{\left\|s_{artif}\right\|^2} \tag{4-5}$$

$$SIR=10\log_{10}\frac{\left\|s_{target}\right\|^2}{\left\|s_{interf}\right\|^2} \tag{4-6}$$

因為我們的目標是改善語音失真的情形，所以主要都會以 SDR 作為最重要的指標，SDR 值越高，表示還原後語音失真的程度越小。

4.1 實驗設置

在各個模擬中，我們會分別針對音源來自不同角度的混合情況來做估計和分離，為了簡化複雜度，我們只測試 DOA = 0 到 90 度的方位，音源方位角與麥克風的位置關係如下圖所示：

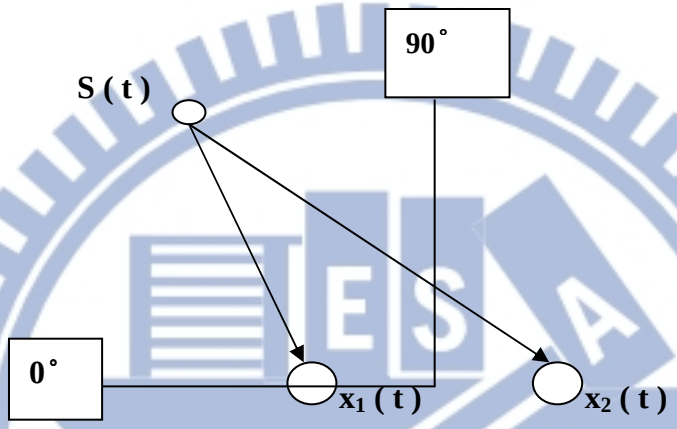


圖 4-3　訊號來源方向示意圖

利用兩個麥克風收到訊號的時間差($\tau = \tau_2 - \tau_1$)與強度不同模擬聲源方位，下表是時間差、強度衰減與 DOA 的對照表：

表 4-1 時間延遲、強度衰減與 DOA 對照表

τ (samples)	1	2	3	4	5	6	7	8	9	10
麥克風 2 衰減	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5
DOA(度)	84.5	78.9	73.2	67.3	61.2	54. 7	47.6	39.6	29.9	15.5

測試的兩個麥克風相距 8 公分；使用 TIMIT 的語料，包括男生女生，每句話大約 2.5 秒；為了增加可探討的聲源方位，我們將原本取樣頻率為 16KHz 的語料升頻到 44.1KHz；Frame 長度為 1000 點(23 ms)，Overlap 是 750 點(17 ms)；STFT 的點數為 1024 點。在三種機率混合模型中，我們皆使用 8 個分布來估計聲源數；下面列出它們各自的初始值及初始分布圖：

表 4-2 高斯混合模型初始參數

高斯混合模型				遞迴 40 次				
M	1	2	3	4	5	6	7	8
μ_m	10	20	30	40	50	60	70	80
σ_m	4	4	4	4	4	4	4	4
α_m	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

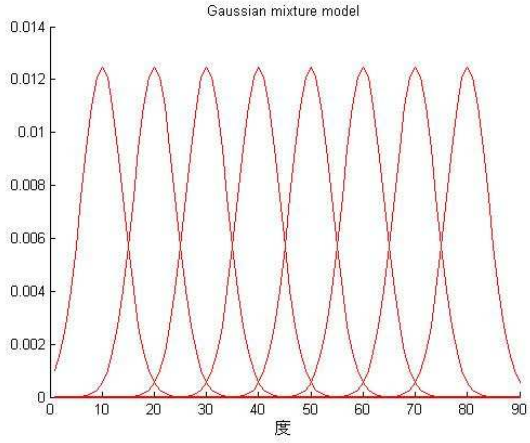


圖 4-4 高斯混合模型初始分布圖

表 4-3 拉普拉斯混合模型初始參數

拉普拉斯混合模型					遞迴 40 次			
M	1	2	3	4	5	6	7	8
μ_m	10	20	30	40	50	60	70	80
b_m	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
α_m	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

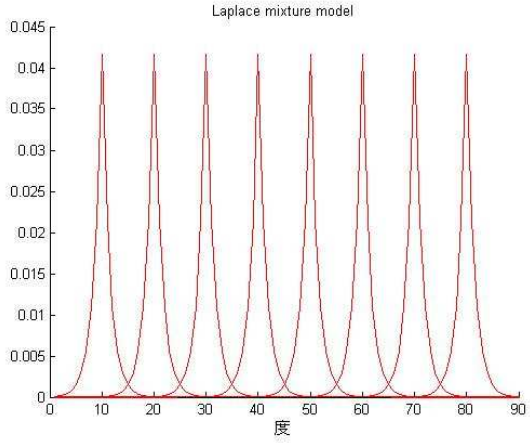


圖 4-5 拉普拉斯混合模型初始分布圖

表 4-4 廣義高斯混合模型初始參數

廣義高斯混合模型					遞迴 15 次			
M	1	2	3	4	5	6	7	8
μ_m	10	20	30	40	50	60	70	80
σ_m	4	4	4	4	4	4	4	4
β_m	2	2	2	2	2	2	2	2
α_m	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

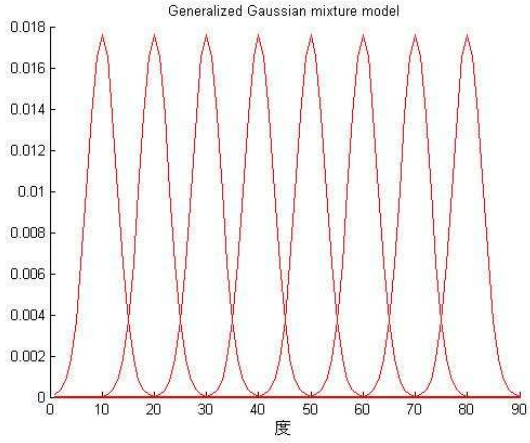


圖 4-6 廣義高斯混合模型初始分布圖

表 4-5 DOA-NPCM 以及 NPCM 參數設定

DOA-NPCM		NPCM	
ρ	10^6	ρ	10^8
k	10^4	k	10^7
ε	0.4	ε	0.4

前兩個模擬是估計聲源數以及方位角的測試：第一個模擬實驗首先探討各個方法在兩個聲源存在時，可以分辨出的最小角度差；第二個模擬實驗為測試各種方法在不同聲源數存在時，得到的估測正確率。我們會給定若干種的角度組合，再統計各種方法能夠正確估計的次數。下一節會詳細的列出估計後的結果以及遞迴後的參數值。

第三個模擬實驗是比較兩種遮蔽法在聲源數為 2 時的分離效果，我們使用 TIMIT 語料庫中的 8 句話，每 2 句話一組，一共 4 個組合，因此每種角度組合得到的結果是 4 次測試的平均。下一節會詳細的列出分離後的聲譜圖以及其 SAR、SDR、SIR 值，可以看出兩種分離法對分離結果的影響。

第四個模擬實驗為欠定問題(under-determined problem)，也就是聲源數量大於混合訊號數量，我們會模擬 3 個聲源、4 個聲源個別來自不同的方位角，並列出兩種分法分離的結果。

4.2 模擬結果

4.2.1 模擬一：最小角度差

從表 4-1 可以看出，角度差最小的組合為(84.5 , 78.9)以及(78.9 , 73.2)，因此我們就對三種模型在這兩種情況下作測試，得到以下的結果：

(84.5 , 78.9)測試結果：

表 4-6 高斯模型估計結果

高斯混合模型	(84.5 , 78.9)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	0	0	78.5	**79.0**	**84.1**
σ_m	0	0	0	0	0	0.05	0.41	1.0
α_m	0	0	0	0	0	0.02	**0.36**	**0.62**

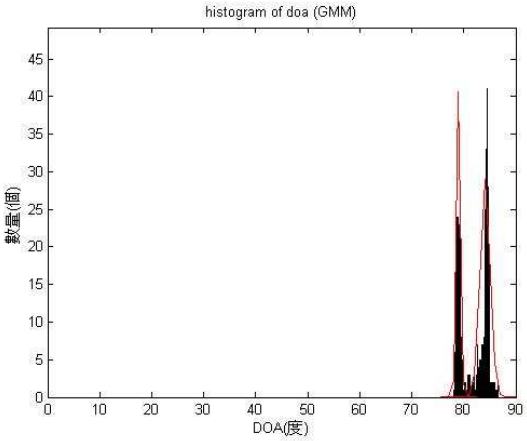


圖 4-7 高斯模型估測分布圖

表 4-7 拉普拉斯模型估計結果

拉普拉斯混合模型	(84.5 , 78.9)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	0	78.6	79.0	**79.0**	**84.2**
b_m	0	0	0	0	0.03	0.34	0.37	0.62
α_m	0	0	0	0	~0	~0	**0.39**	**0.60**

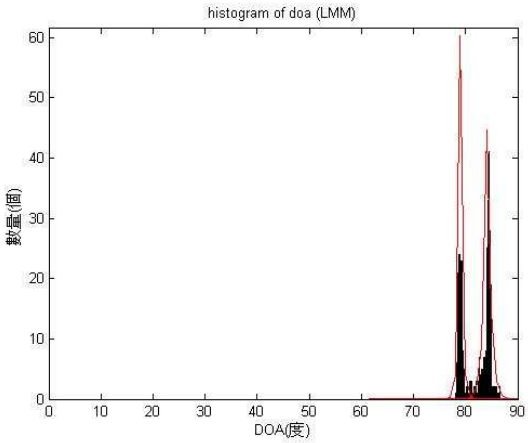


圖 4-8 拉普拉斯模型估測分布圖

表 4-8 廣義高斯模型估計結果

廣義高斯混合模型	(84.5 , 78.9)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	0	0	78.5	**78.9**	**84.0**
σ_m	0	0	0	0	0	0.16	0.54	0.83
β_m	2	2	2	2	2	2.08	2.54	1.06
α_m	0	0	0	0	0	0	**0.37**	**0.63**

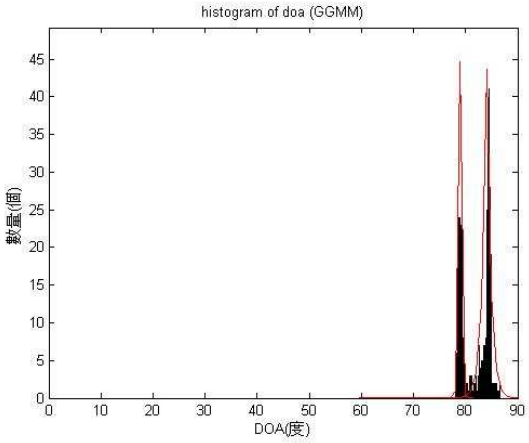


圖 4-9 廣義高斯模型估測分布圖

表 4-9 NPCM 與 DOA-NPCM 估計結果

NPCM		s_1	s_2		DOA-NPCM		s_1	s_2	
	Hi/Hmax	1	0.52	0.15		Hi/Hmax	1	0.42	~0
	w	41.5	43.0			w	84.5	78.8	

(78.9 , 73.2)測試結果：

表 4-10 高斯模型估計結果

高斯混合模型	(78.9 , 73.2)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	0	72.7	**73.1**	**73.6**	**78.6**
σ_m	0	0	0	0	0.1	0.4	0.8	0.9
α_m	0	0	0	0	0	**0.17**	**0.23**	**0.60**

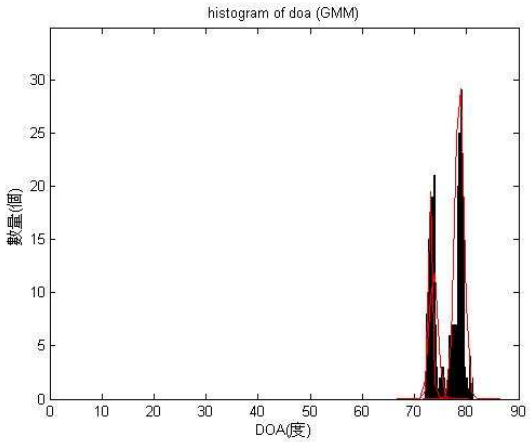


圖 4-10 高斯模型估測分布圖

表 4-11 拉普拉斯模型估計結果

拉普拉斯混合模型	(78.9 , 73.2)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	72.7	73.3	73.4	**73.4**	**78.6**
b_m	0	0	0	0.02	0.44	0.51	0.51	0.68
α_m	0	0	0	0.02	0	~0	**0.37**	**0.60**

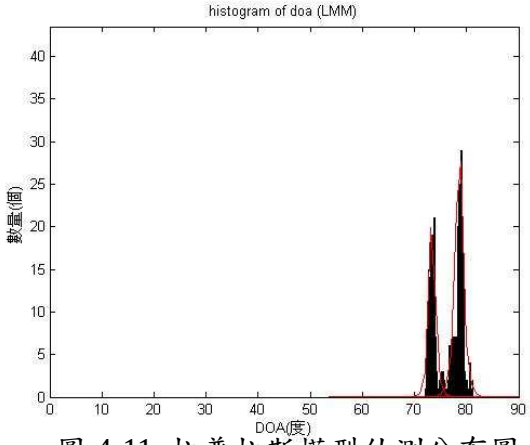


圖 4-11 拉普拉斯模型估測分布圖

表 4-12 廣義高斯模型估計結果

廣義高斯混合模型	(78.9 , 73.2)							
M	1	2	3	4	5	6	7	8
μ_m	0	0	0	0	0	72.6	**73.3**	**78.5**
σ_m	0	0	0	0	0	0.23	0.81	0.89
β_m	2	2	2	2	1.18	2.15	2.27	1.11
α_m	0	0	0	0	0	~0	**0.38**	**0.62**

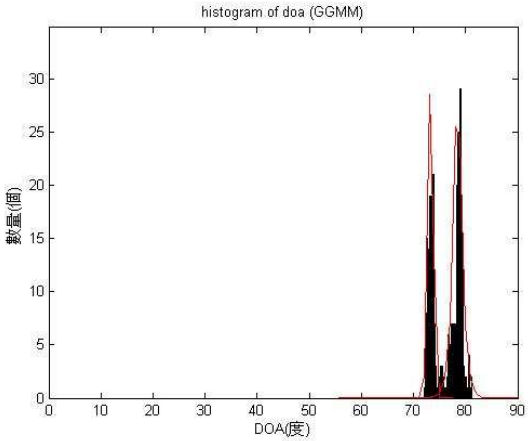


圖 4-12 廣義高斯模型估測分布圖

表 4-13 NPCM 與 DOA-NPCM 估計結果

NPCM		s_1	s_2		DOA-NPCM		s_1	s_2	
	Hi/Hmax	1	0.74	0.06		Hi/Hmax	1	0.41	~0
	w	40.1	41.5			w	78.8	73.1	

從表 4-6 到 4-13 可以看出，除了 NPCM 本身只能估計出聲源數目之外，其他每種方法對於以上 2 種角度組合都可以分辨出聲源數與聲源角度，因此可以推斷每種方法都能夠處理角度差在 5.5 度的情形。唯有表 4-10，高斯模型的結果顯示有 3 個分布，不過進一步判斷有 2 個分布 μ_m 相差 3.5 度內，視為同一聲源，因此也可視為正確的估計。

4.2.2 模擬二：正確率

本模擬中，又細分成 4 個部份，首先是當只有 1 個聲源存在時，因此會有 10 種情形，結果如下表：

表 4-14 各種方法估計結果，N_S=1

方法 DOA(度)	高斯	拉普拉斯	廣義高斯	NPCM	DOA-NPCM
84.5	O	O	O	O	O
78.9	O	O	O	O	O
73.2	O	O	O	O	O
67.3	O	O	O	O	O
61.2	O	O	O	O	O
54.7	O	O	O	O	O
47.6	O	O	O	O	O
39.6	X	O	O	O	O
29.9	X	X	X	O	O
15.5	X	X	X	O	X
正確率	**70 %**	**80 %**	**80 %**	**100 %**	**90 %**

從表 4-14 可以看出每個方法的效果都不錯，唯有當 DOA 太小時，會造成錯誤。原因是當 DOA 偏小的時候，直方統計圖會變得發散，導致不易正確估計出 DOA 位置；而 NPCM 是根據強度資訊做估測，因此不受影響，所以正確率可以達到 100%。以下兩張圖可以看出不同方位在統計直方圖上的明顯差異：

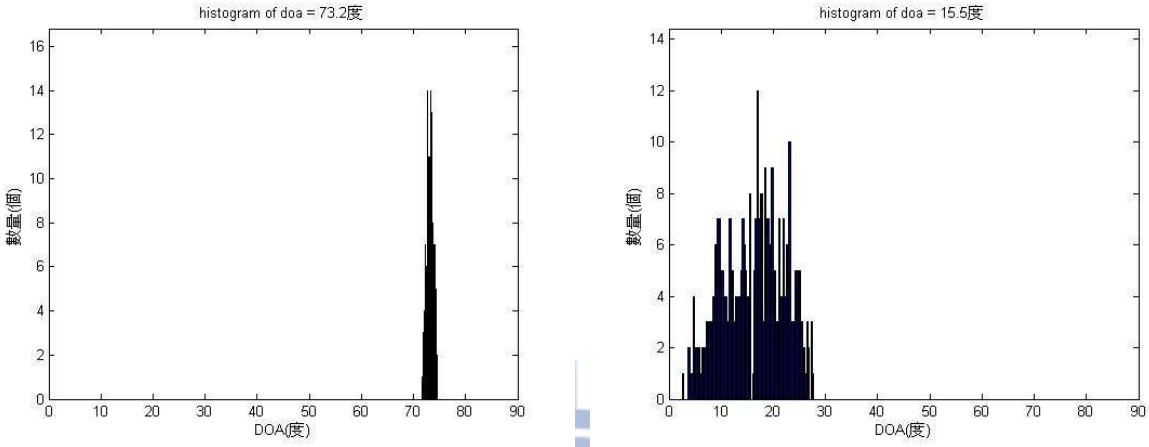


圖 4-13　左圖是 DOA=73.2 度的統計圖，較為集中
右圖是 DOA=15.5 度的統計圖，較為發散

第二個部分中，也就是當有 2 個聲源存在時，我們做了 45 組的角度組合測試，結果如下表：

表 4-15 各種方法估計結果，N_S=2

模型	高斯	拉普拉斯	廣義高斯	NPCM	DOA-NPCM
正確率	**53 %**	**39 %**	**61 %**	**88 %**	**73 %**

第三個測試是 3 個聲源存在時，一共有 120 組角度組合，結果如下表：

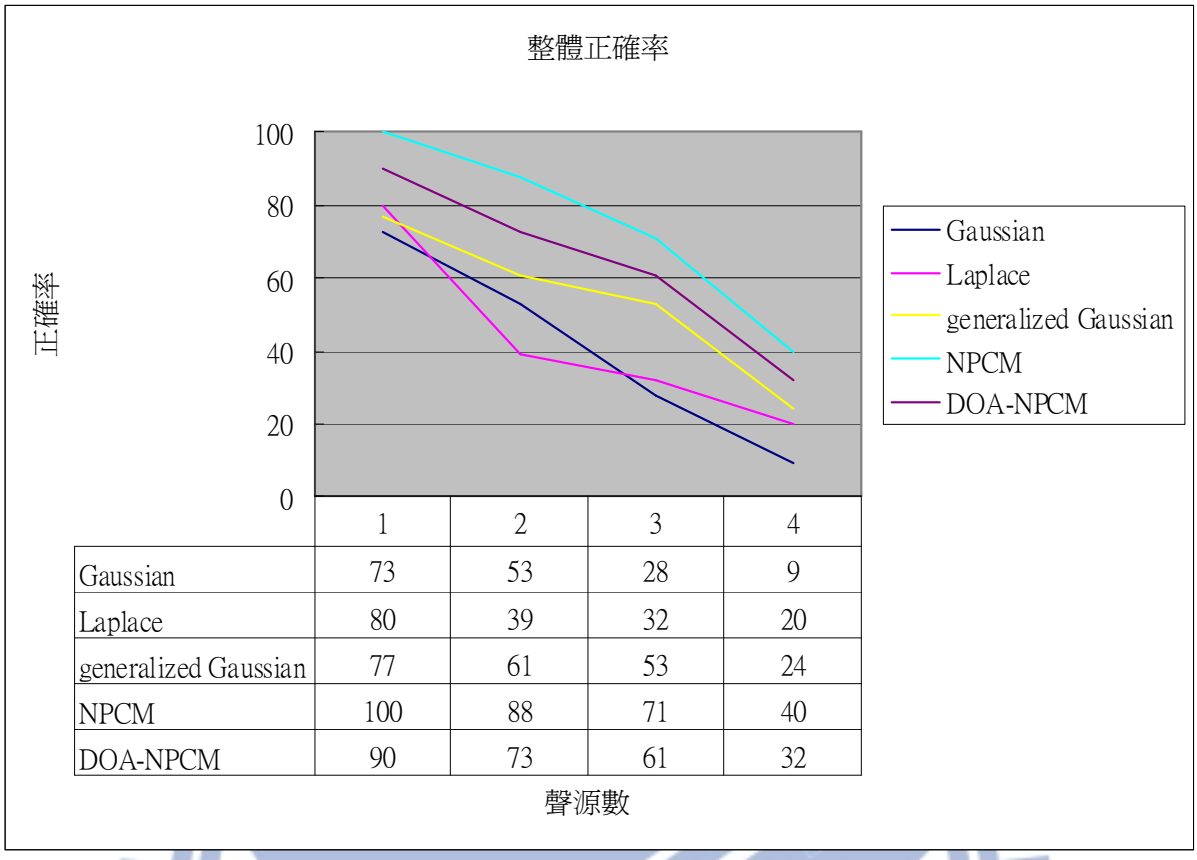
表 4-16 各種方法估計結果，N_S=3

模型	高斯	拉普拉斯	廣義高斯	NPCM	DOA-NPCM
正確率	**28 %**	**32 %**	**53 %**	**71 %**	**61 %**

最後一個測試，是 4 個聲源存在時，一共有 210 組角度組合，結果如下表：

表 4-17 各種方法估計結果，N_S=4

模型	高斯	拉普拉斯	廣義高斯	NPCM	DOA-NPCM
正確率	**9 %**	**20 %**	**24 %**	**40 %**	**32 %**



	1	2	3	4
Gaussian	73	53	28	9
Laplace	80	39	32	20
generalized Gaussian	77	61	53	24
NPCM	100	88	71	40
DOA-NPCM	90	73	61	32

圖 4-14 整體正確率趨勢圖

從前面 4 個部分的模擬中可看出，隨著聲源數增加，估計的正確率越來越差，原因是因為聲源數變多，DOA 在直方統計圖的表現會越來越混亂，導致峰值不明顯，因此無法正確的找到有效聲源的數目和方向。錯誤的情形有以下幾種：一是將非聲源誤認為是聲源，導致聲源數的高估；二是低估了聲源數，原因是把兩個聲源誤認為同一個聲源；最後一種是聲源數估計對了，但是判斷的聲源角度偏差太多，大部分是發生在 DOA 太高或太低的情形。NPCM 在各個測試中都有最好的表現，原因是其只利用聲譜圖強度的資訊，不會受到因 DOA 太低導致的統計圖混亂現象，但 NPCM 只能估計聲源的數目，無法得出聲音確實的來源方向。

在整體上的表現，可以發現廣義高斯模型、NPCM 以及 DOA-NPCM 在 3 個聲源同時存在時，還可以達到 50%以上的正確率，但在聲源數為 4 時，正確率就剩 40%以下，若從 50%的正確率做分野，我們可以說這三種方法最大可以處理的聲源數目為 3。

4.2.3 模擬三：2 個聲源

我們一共測試了 4 種角度組合，分別為(84.5, 78.9)、(73.2, 47.6)、(67.3, 29.9)以及(84.5, 15.5)，此 4 個角度組合分別有不同的角度差距，角度差從小到大都有，分別為 5.6、25.6、37.4 和 69.0 度。每個組合會有 4 次不同句話的測試，最後的結果為 4 次測試的平均。除了比較二元遮蔽與權重遮蔽的方法，我們也與 NPCM 與 DOA-NPCM 分離的結果作比較。

因為 ITD 的資訊在 4000Hz 以上會變得混亂，因此我們只針對 4000Hz 以下的部分作處理，下圖為其中一次測試(73.2, 47.6)的聲源聲譜圖、混合訊號聲譜圖以及分離後語音的聲譜圖：

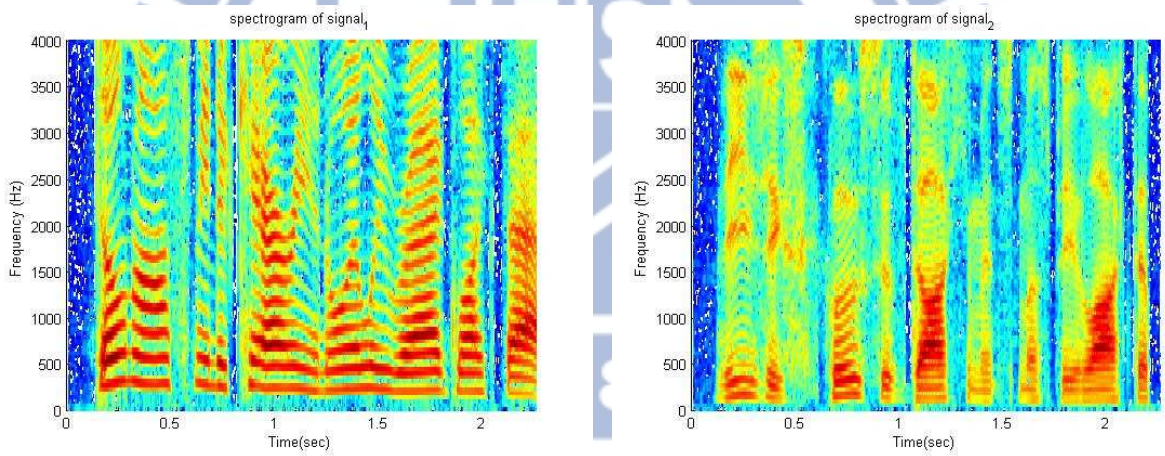


圖 4-15　左圖是聲源 1 的聲譜圖
右圖是聲源 2 的聲譜圖

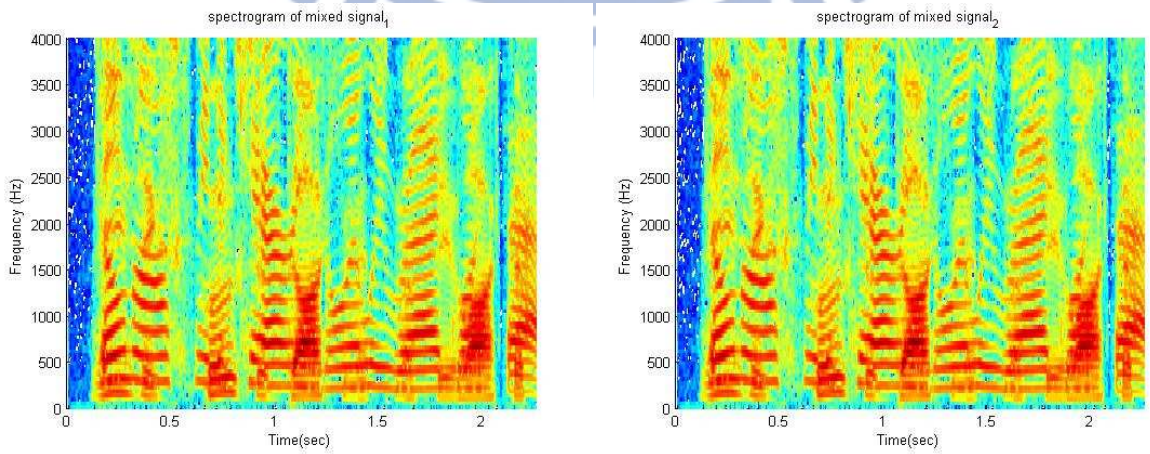


圖 4-16　左圖是混合訊號 1 的聲譜圖
右圖是混合訊號 2 的聲譜圖

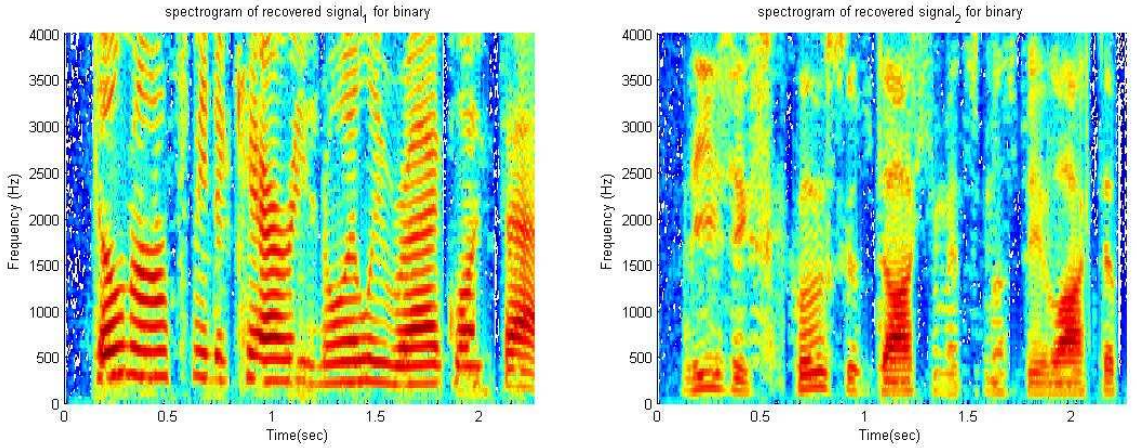


圖 4-17　左圖是用二元分離法分離後的聲源 1 聲譜圖
右圖是用二元分離法分離後的聲源 2 聲譜圖

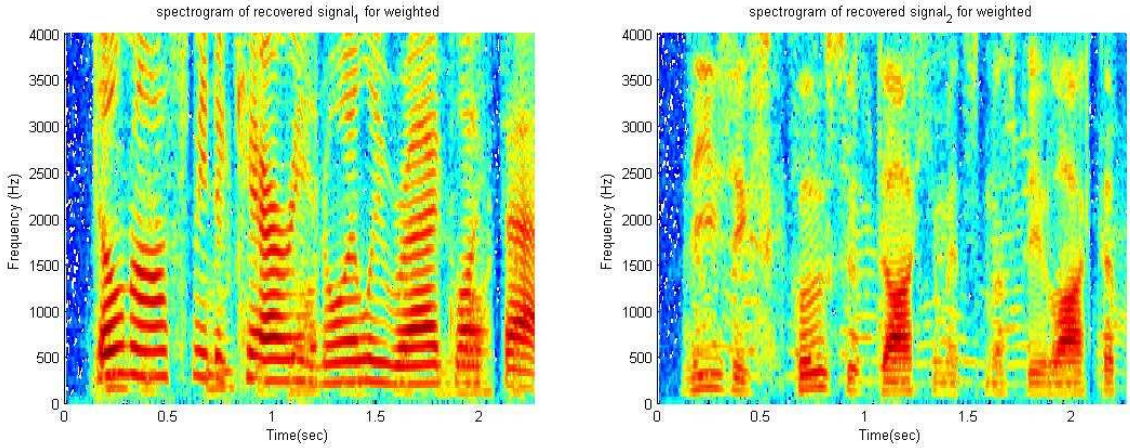


圖 4-18　左圖是用權重分離法分離後的聲源 1 聲譜圖
右圖是用權重分離法分離後的聲源 2 聲譜圖

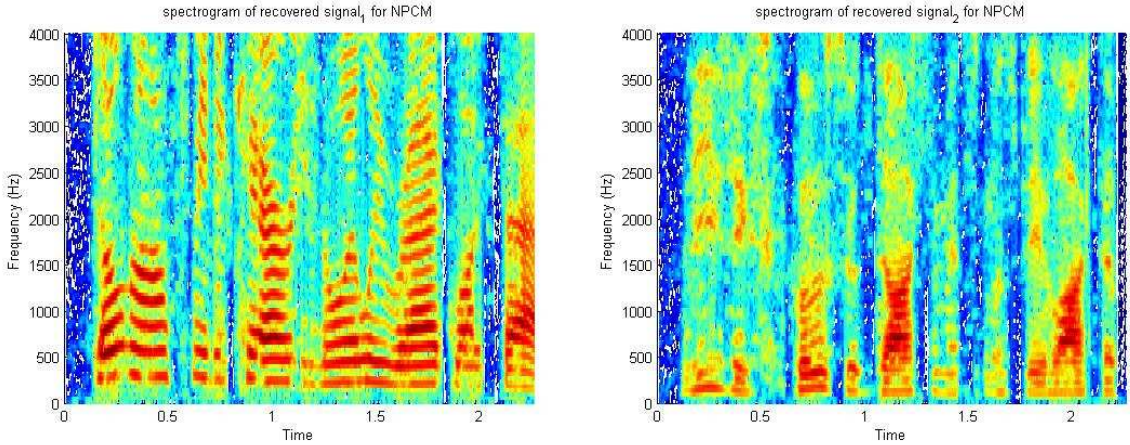


圖 4-19　左圖是用 NPCM 分離後的聲源 1 聲譜圖
右圖是用 NPCM 分離後的聲源 2 聲譜圖

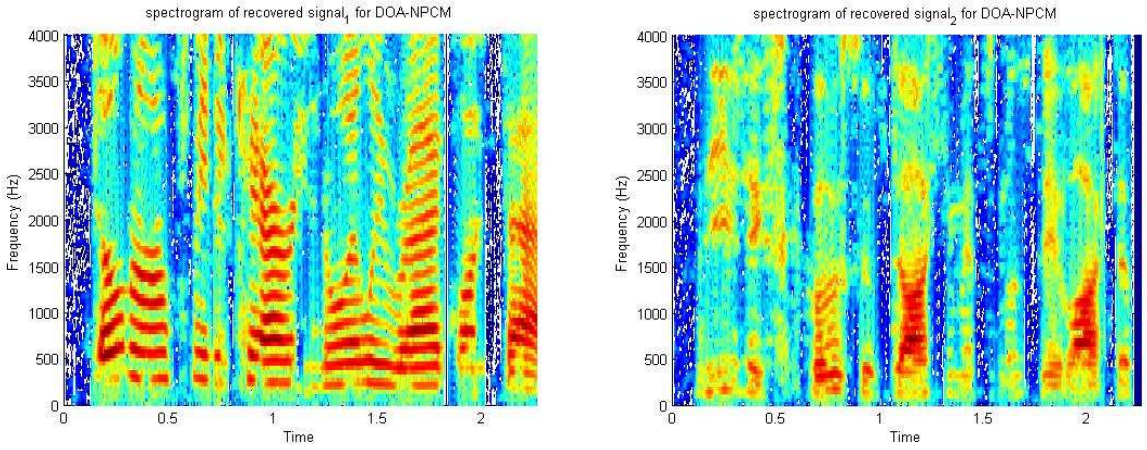


圖 4-20　左圖是用 DOA-NPCM 分離後的聲源 1 聲譜圖
右圖是用 DOA-NPCM 分離後的聲源 2 聲譜圖

表 4-18 各種分離法的分離結果，N_S=2

分離	**(84.5 , 78.9)**				**(73.2 , 47.6)**			
方法	二元	權重	NPCM	DOA-NPCM	二元	權重	NPCM	DOA-NPCM
SAR	10.10	11.52	8.96	10.84	10.09	11.37	5.41	6.91
SDR	10.03	**10.29**	5.77	9.88	10.01	**10.21**	5.41	6.91
SIR	29.39	16.87	**16.99**	17.4	29.09	17.02	**40.81**	49.47
分離	**(67.3 , 29.9)**				**(84.5 , 15.5)**			
方法	二元	權重	NPCM	DOA-NPCM	二元	權重	NPCM	DOA-NPCM
SAR	9.89	11.11	4.76	5.56	8.69	9.82	3.92	2.78
SDR	9.78	**9.94**	4.75	5.55	8.23	**8.28**	3.91	2.76
SIR	28.57	16.74	**43.27**	46.41	22.03	14.33	**46.53**	44.95

從表 4-18 的分數以及圖 4-15 到圖 4-20 中可以看出，權重分離法的 SDR 是所有方法中最高的，表示權重分離法在失真的程度上是比較輕微的，但在 SIR 上，權重分離法卻是最低的。會有如此結果的原因是，另外 3 種方法為了把語音分離，會把偏離方向較大的 unit 完全捨棄，也就是除了壓抑干擾的訊號能量，也捨棄了自己原本的訊號能量，才會導致以上的結果。而 NPCM 和 DOA-NPCM 比二元分離法失真正更嚴重的原因是二元分離法中，每個 T-F unit 都會屬於其中一個聲源，但是在 NPCM 和 DOA-NPCM 時，如果一個 T-F unit 皆不夠靠近任一聲源，

則該位置的 T-F unit 將不屬於任一聲源，也就是被丟棄的意思，因此這兩種方法遮蔽中存在的 0 個數會大於二元分離法遮蔽中的 0 個數，導致失真程度比二元法嚴重。

分離結果除了 NPCM 之外，其他方法並沒有表現出聲源相距越遠分離效果越好的趨勢，其原因是當 DOA 角度越小時，聲譜圖上的 DOA 會越不穩定（如圖 4-13），每個 unit 的 DOA 越不會落在前面介紹的 DOA 與強度比例的曲線上，導致分離結果不會因相距越遠而越好。但是 NPCM 只有利用強度資訊，不受前述的影響，所以其呈現的趨勢與相距越遠而分離效果越好的期望一致。以下的圖是以上 4 種角度組合，2 聲源強度比例與 DOA 的散點圖，可以作個簡單說明：

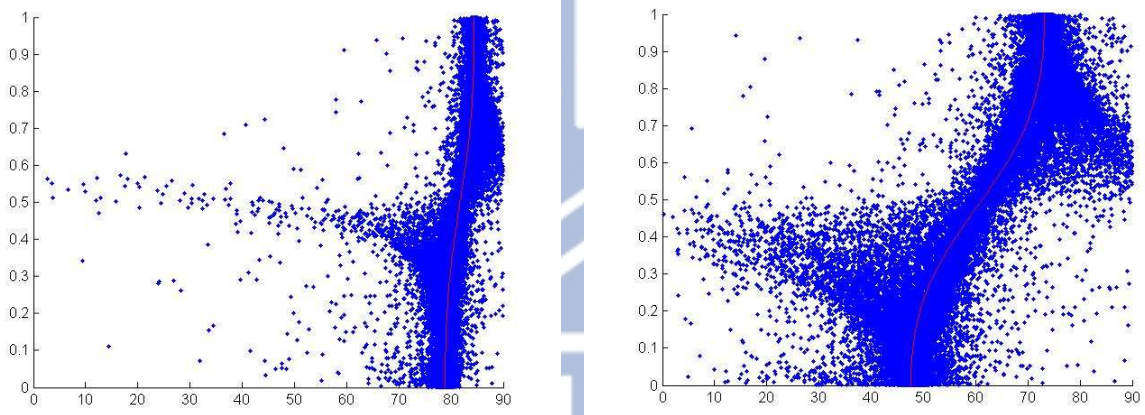


圖 4-21　左圖是(84.5, 78.9)的實際聲源能量比例與 DOA 散點圖
右圖是(73.2, 47.6)的實際聲源能量比例與 DOA 散點圖

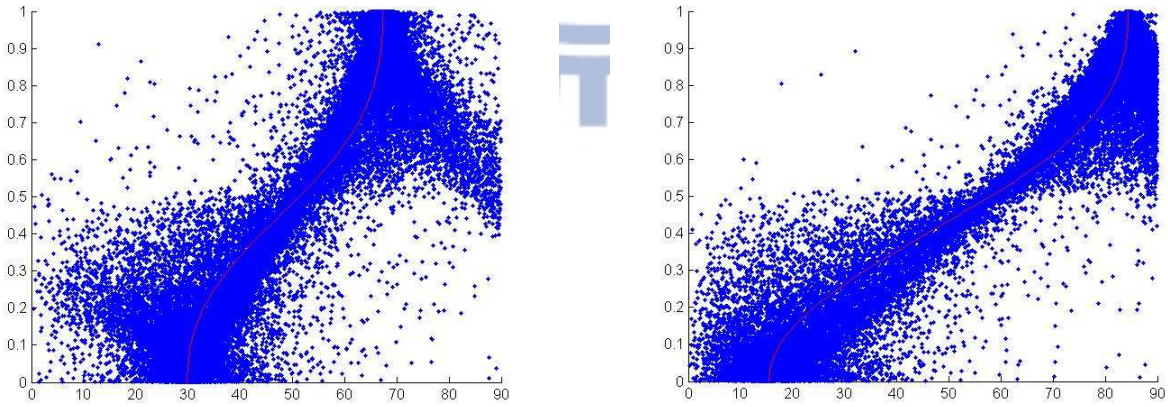


圖 4-22　左圖是(67.3, 29.9)的實際聲源能量比例與 DOA 散點圖
右圖是(84.5, 15.5)的實際聲源能量比例與 DOA 散點圖

4.2.4 模擬四：3 個 & 4 個聲源

在 3 個聲源的測試中，我們令 3 個聲源分別位於(84.5 , 73.2 , 61.2)，每個方法都會作 4 次不同句話的測試，顯示的結果是 4 次的平均。結果如下：

表 4-19 各種分離法的分離結果，N_S=3

分離方法	二元	權重	NPCM	DOA-NPCM
SAR	6.22	7.99	3.51	5.52
SDR	5.90	**6.41**	3.35	5.38
SIR	22.82	13.50	19.74	27.88

在 4 個聲源的測試中，我們令 4 個聲源分別位於(78.9 , 67.3 , 54.7 , 39.6)，每個方法都會作 4 次不同句話的測試，顯示的結果是 4 次的平均。結果如下：

表 4-20 各種分離法的分離結果，N_S=4

分離方法	二元	權重	NPCM	DOA-NPCM
SAR	4.50	6.14	2.5	2.61
SDR	4.03	**4.21**	2.32	2.35
SIR	18.05	10.64	18.92	21.3

在欠定的情況下，因為聲源變多了，單一 unit 內通常不再只有一個聲源的成分，因此 NPCM 和 DOA-NPCM 造成的失真就更嚴重了；而權重分離法還能根據 DOA 的位置，應用(式 3-29)的關係式分配 unit 的能量給不同聲源，所以會有較高的 SDR。但是隨著聲源變多，四種分離法整體的表現還是越來越差的。

第五章　結論與未來展望

本論文探討混合模型估測聲源數以及聲源方位的成效，主要是針對不同時間延遲與強度衰減的混合訊號來作測試。論文中實驗了三種不同的混合模型以及兩種 NPCM 的方法，將訊號轉到頻域後，再對每個 T-F unit 的 DOA 作統計。分別是常見的高斯混合模型和拉普拉斯混合模型，還有自由度更高的廣義高斯混合模型，加上原本的 NPCM 還有 DOA-NPCM。除了聲源數、聲源方位的估測，我們在得到前述的兩個資訊後，探討四種遮蔽分離法來分離混合訊號，分別是二元分離法、權重分離法、NPCM 以及 DOA-NPCM。

比較各種方法的結果，在空間解析度上，五個方法的表現是平分秋色，最小可分辨的角度差約為 5.5 度。在聲源數目的正確率上，雖然 NPCM 的效能是最佳的，但它無法同時提供聲源的方向。而 DOA-NPCM 和廣義高斯混合模型雖然在聲源數目的正確率表現是次佳的，但除了聲源數目外，它們還能同時獲得聲源的方向。而 NPCM、DOA-NPCM 和廣義高斯模型在聲源數為 3 時，都達到五成以上的正確率，推斷其可處理的空間最大聲源數為 3。

而在分離法的比較上，在任何的角度組合以及聲源數的情形之下，權重分離法在 SDR 的分數一直都是最高的，表示用權重的方式可以保留較多原本語音的成分，但是其他方法對於原本語音造成的破壞就較嚴重。

遮蔽分離法是在雜訊消除與語音失真程度間作取捨，當消去越多雜訊的同時，也會對原本的語音造成破壞。在此研究中，我們利用相位差資訊估計聲源數以及分離語音，但成效上會受到其本身在 DOA 太小時的限制，未來我們將結合聲音的強度資訊來進行方向估計和音源分離的研究。畢竟，從感知的觀點來看，人們也是同時使用雙耳時間差和能量差，來有效地在各種環境下進行聲源定位。

參考文獻

[1] Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley, 2001.

[2] James V. Stone, *Independent Component Analysis: a Tutorial Introduction*, MIT Press, 2004

[3] Z. Guoxu, Y. Zuyuan, X. Shengli, Y. Jun-Mei, “Mixing matrix estimation from sparse mixtures with unknown number of sources,” IEEE Transactions on Neural Networks, vol.22, pp. 211-221, 2011.

[4] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” IEEE Trans. Signal Processing, vol. 52, no.7, pp. 1830–1847, July 2004.

[5] A. Jourjine, S. Rickard, and O. Yilmaz, “Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures,” in Proc. ICASSP, vol.5, pp. 2985-2988, 2000.

[6] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” Signal Processing, vol. 77, no. 8, pp. 1833–1847, Aug. 2007.

[7] Ngoc Q.K. Duong, E. Vincent, and R. Gribonval, “Under-determined convolutive blind source separation using spatial covariance models,” in Proc. ICASSP, 2010.

[8] H. Sawada, S. Araki, and S. Makino, “A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures,” In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, October 2007.

[9] T. Xu, and W.Wang, “A block-based compressed sensing method for underdetermined blind speech separation incorporating binary mask,” in Proc. ICASSP, 2010.

[10] T. S. Chi, C. W. Huang, and W. S. Chou, “A frequency bin-wise nonlinear masking algorithm in convolutive mixtures for speech segregation, ” Journal Acoustical Society of America, vol. 131, no. 5, pp. 361-367, April 2012.

[11] Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of the Royal Statistical Society, Series B (Methodological), vol. 39, no. 1, pp. 1–38, 2007.

[12] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, “Underdetermined blind separation for speech in real environments with sparseness and ICA,” in Proc. ICASSP, 2004.

[13] Y. Bazi, L. Bruzzone, and F. Melgani, “Image thresholding based on the EM algorithm and the generalized Gaussian distribution,” Journal Pattern Recognition, vol. 40, no. 2, pp. 619 – 634, February 2007.

[14] Mohamed, M. Mohamed, and M. Jaidane-Saidane, “On the Parameters Estimation of the generalized gaussian mixture model,” In European Signal Processing Conference (EUSIPCO), pp. 2273-2277, Glasgow, Scotland, August 2009.

[15] M. Abramowitz, I.A. Stegun, *HandBook of Mathematical Tables*, Dover, New York, 1970.

[16] N. Roman, D. wang, and G.J. Brown, “Speech segregation based on sound localization,” Journal Acoustical Society of America, vol. 114, no. 4, pp. 2236-2252, October 2003.

[17] C. F´evotte, R. Gribonval and E. Vincent, BSS EVAL Toolbox User Guide, IRISA Technical Report 1706, Rennes, France, April 2005. http://www.irisa.fr/metiss/bsseval/.

[18] P. Comon, “Independent Component Analysis, a new concept?” Signal Processing, Elsevier, 36(3):287–314, April 1994.

[19] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis. Neural Computation, 9(7):1483-1492, 1997.

[20] Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE Trans. on Neural Networks, 10(3):626-634,1999.

[21] Hyvärinen, “Survey on independent component analysis,” Neural Computing Surveys, 2:94-128, 1999.

[22] E. Bingham and A. Hyvärinen, “A fast fixed-point algorithm for independent component analysis of complex-valued signals.” Int. Journal of Neural Systems, 10(1):1–8, 2000.

[23] A. J. Bell and T. J. Sejnowski, “A non-linear information maximization algorithm that performs blind separation,” In Advances in Neural Information Processing Systems 7, pages 467–474. The MIT Press, Cambridge, MA, 1995.

[24] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” Neural Computation, 7:1129-1159, 1995.

[25] P.D.O’Grady, B.A.Pearlmutter, and S.T.Rickard, “Survey of sparse and non-sparse methods in source separation,” International Journal of Imaging Systems and Technology (IJIST), vol.15, p.18–33, 2005.

[26] P. Bofill and M. Zibulevsky, “Underdetermined blind source separation using sparse representations ,” Signal Processing, vol. 81, no. 11, p.2353-2362, 2001.

[27] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, “Blind source separation via multinode sparse representation,” In Advances in Neural Information Processing Systems 14, p.1049-1056, MIT Press, 2002

[28] A. Jourjine, S. Rickard, and O. Yilmaz, “Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures,” In IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP2000), vol.5, p.2985-2988, 2000.