

# 基於稀疏成份分析之旋積盲訊號源分離方法

## Convolutional Blind Source Separation Based on Sparse Component Analysis

莊祥龍 Hsiang-Lung Chuang  
國立中央大學資訊工程學系

Department of Computer Science and Information Engineering  
National Central University  
pee0402@hotmail.com

謝宇勳 Yu-Shiun Shie  
國立中央大學資訊工程學系

Department of Computer Science and Information Engineering  
National Central University  
e8010232002@gmail.com

林昶宏 Chang-Hong Lin  
國立中央大學資訊工程學系

Department of Computer Science and Information Engineering  
National Central University  
fredlin1017@gmail.com

王家慶 Jia-Ching Wang  
國立中央大學資訊工程學系

Department of Computer Science and Information Engineering  
National Central University  
jcw@csie.ncu.edu.tw

### 摘要

本論文針對的是在不知道源訊號個數的情況下，一個稀疏欠定的旋積盲訊號源分離。我們的演算法分為兩個階段，先估計混合矩陣然後才利用此矩陣分離源訊號。在估計混合矩陣上，首先定義了兩個特徵參數，包括了 Level-Ratio 以及 Phase-Difference，我們藉由 KNN Graph 方式，去除資料中的離群樣本，並用 K-Means 分群演算法對其餘的資料分群，然後應用 DOA 解決不同頻率間的排列問題，以達到估計混合矩陣的目的。此外，我們對此混合矩陣進行相位之補償，以獲得更精確之混合矩陣估計。本方法是建立於最大後驗機率方法上，在求得混合矩陣之後，利用最小 L1 範數去解一個欠定的線性最佳化問題。此外，對於未知的源訊號個數，我們利用 K-Means 演算法和貝氏資訊準則作結合，並對所有頻帶的結果做整體考量，以達到估測源訊號個數的目的。在實驗模擬的部分，會將我們提出的方法與參考文獻作比較，也證實了此演算法在分離訊號效能之優越性。

關鍵詞: 盲訊號源分離，最大概似估計，稀疏成份分析，貝氏資訊準則

## 一、簡介

近年來盲訊號源分離蓬勃發展，相關的論文如雨後春筍般出現。如同前面所敘述，我們所面對的是一個「雞尾酒會問題」。換句話說，我們希望在源訊號以及混合過程的資訊未知的條件之下，單憑混合訊號就能達到重建源訊號的目的。可想而知，在前提和變數如此多的情況下，想得到源訊號並沒有這麼簡單。所以盲訊號源分離絕對是一個極富挑戰性的研究課題。

盲訊號源分離依混合模型的型態可分成兩類，一個是瞬時混合模型(Instantaneous Mixing Model)[1],[2]另一個則為旋積混合模型(Convolution Mixing Mode)[3],[5]-[10]。令  $X$  是一個  $M \times T$  的混合訊號矩陣，每一列向量都代表一個混合訊號； $S$  表一個  $N \times T$  的源訊號矩陣，它的列向量分別代表著某一個源訊號； $H$  指的是混合矩陣，其中  $M$  是麥克風數， $N$  則為語者個數，而  $T$  表示時間域上的樣本長度。依照上述代號，我們可將瞬時混合模型表示成：

$$X = H \times S \quad (1)$$

然而在這篇論文當中，我們所提出的方法是建構在旋積混合模型之上，接下來會在後面的章節詳細說明此模型以及運用和推導的過程。

盲訊號源分離問題可依麥克風和源訊號的個數區分為兩種案例，分別為  $M \geq N$  的過定(Over-Determined)問題和  $M < N$  的欠定(Under-Determined)問題[6]-[9]。典型的盲訊號源分離就被歸類為過定的問題，而到目前為止，最常被用來解決這類問題的方法就是獨立成份分析(Independent Component Analysis, ICA)[1]-[4]。但是在現實生活中，我們也經常會碰到麥克風數量少於語者的情形。在這種狀態下我們是無法藉由獨立成份分析解決問題的。原因在於獨立成份分析是以迭代的方式不斷對混合矩陣做更新，使得混合矩陣具有分離訊號的能力，所以在獨立成份分析底下，混合矩陣必需為一個方陣(Square Matrix)，這與欠定盲訊號源方離背道而馳，因為當我們的問題具備欠定的特質時，混合矩陣會是一個矩形矩陣(Rectangular Matrix)。基於這個原因，這幾年稀疏成份分析(Sparse Component Analysis, SCA)逐漸流行，這種分析方式大多建立在資料具備稀疏性的假設前提下，然後再利用一些統計方法達到目的。稀疏成份分析至今已在訊號分離領域中佔有很重要的地位[6]-[9]。

在許多文獻中，源訊號的個數往往被當成是已知。但實際上，在很多時候，語者的個數是無法預知的。所以如何得知或估測  $N$  就變得格外重要。通常是藉由判定資料分佈中群聚個數的方式來取得  $N$  值，目前有許多方法都已經有不錯的成效。而且不少學者都開始在他們所提出來的演算法中考量源訊號個數是未知的情形。

我們的目的就是利用盲訊號源分離有效的將涵蓋於觀察訊號中的語音給取出。除了準確的估測語者個數之外，也希望在欠定的限制下，當混合訊號中有其餘的干擾成份(噪音、殘響等等)時，仍然可以達到分離之效用。

本論文的組織如下：第 2 部分為盲訊號源分離模型；第 3 部分為特徵參數選取以及源訊號數之估計；第 4 部分為混合係數矩陣估測以及相位補償技術；第 5 部分為實驗結果；最後，第 6 部分為結論。

## 二、盲訊號源分離模型

這篇論文考量到一個旋積混合型態的模型。我們利用下面的數學表示方式來描述此模型。

$$x_q(t) = \sum_{k=1}^N \sum_{l=0}^{L-1} h_{qk}(l) s_k(t-l) \quad (2)$$

其中  $x_q$  是感測器  $q$  對應的麥克風混合訊號(Mixing Signal)， $q=1,2,\dots,M$ ， $s_k$  為第  $k$  個語者的源訊號(Source Signal)， $k=1,2,\dots,N$ ， $h_{qk}$  則是語者  $k$  到麥克風  $q$  的脈衝響應，並且令這個濾波器(Filter)的型式為一個  $L$  階( $L$  - Tap)的有限脈衝響應(Finite Impulse Response, FIR)濾波器。由於語音在時間域上的稀疏特性並不明顯，所以我們採用短時傅利葉轉換(Short Time Fourier Transform, STFT)，以取樣頻率  $f_s$  將時間域上的混合訊號  $x_q(t)$  轉換成頻率域上的時間序列  $x_q(f_i, \tau_j)$ ，並且在時頻域上做訊號的觀察和處理：

$$x_q(f_i, \tau_j) \leftarrow \sum_{t=-T/2}^{T/2} x_q(t + \tau_j S) \text{win}(t) e^{-j2\pi f_i t}$$

, where  $f_i \in \{0, (1/T)f_s, \dots, ((T-1)/T)f_s\}$  (3)

其中  $f_i$  是某個頻帶， $\tau_j$  為短時傅利葉轉換音窗的指標(Frame Index)， $S$  為窗的位移量。至於這篇論文，我們所使用的是一個漢寧窗(Hanning Window)。然而，在時頻域上執行盲訊號源分離的另一個好處是我們可以將旋積混合過程單純視為各個頻帶的瞬時混合型式，即如同以下之敘述。

$$X(f_i, \tau_j) = H(f_i) S(f_i, \tau_j) = \sum_{k=1}^N H_k(f_i) S_k(f_i, \tau_j) \quad (4)$$

, where  $X(f_i, \tau_j) \in C^{M \times 1}$ ,  $S(f_i, \tau_j) \in C^{N \times 1}$ ,  $H(f_i) \in C^{M \times N}$

其中  $X(f_i, \tau_j)$  和  $S(f_i, \tau_j)$  分別代表混合訊號以及源訊號在時頻域上的成份。 $H(f_i)$  則是某一個頻帶的混合矩陣。然而，假設在一個時頻點上，只有一個源訊號在活動，我們令  $H_k(f_i)$  是  $H(f_i)$  的第  $k$  個行向量，則可將式子(4)簡化為：

$$X(f_i, \tau_j) = H_k(f_i) S_k(f_i, \tau_j) \quad , k \in \{1, \dots, N\} \quad (5)$$

所謂的波束形成(Beamforming)，即為一種空間上之濾波器，它利用訊號的空間關係，希望能夠對不同方向的訊號做出不同的增益，以達到空間濾波的效果，藉以分離

空間中不同方向聲源的訊號。依波束形成定理，我們靠著麥克風陣列的源訊號方向和時間延遲去近似混合過程。因此當頻率為  $f_i$  時，語者  $k$  到麥克風  $q$  的混合係數可表示為：

$$h_{qk}(f_i) = g_{qk} e^{j2\pi f_i c^{-1} d_q \cos \theta_k} \quad (6)$$

其中  $g_{qk}$  為訊號  $k$  至麥克風  $q$  的增益值， $d_q$  表感測器  $q$  與麥克風陣列中心之間的距離， $\theta_k$  是源訊號  $k$  對應到麥克風陣列的角度。我們可利用式子(6)，將混合矩陣表現成下面的形式，以下有關混合矩陣的推導過程，多數都是建立在這個預設形式之上。

$$H(f_i) = \begin{bmatrix} h_{11}(f_i) & \cdots & h_{1N}(f_i) \\ \vdots & \ddots & \vdots \\ h_{M1}(f_i) & \cdots & h_{MN}(f_i) \end{bmatrix} \quad (7)$$

### 三、特徵參數選取以及源訊號數之估計

#### (一) 樣本型態

我們定義了兩個混合訊號的特徵參數(Level-Ratio 和 Phase-Difference)[11]。利用觀察資料的二階範數對混合訊號的絕對值頻譜(Magnitude Spectrum)做正規化，我們稱之為 Level-Ratio，我們這邊用  $\psi_q^L(f_i, \tau_j)$  表示；至於 Phase-Difference 被定義成與一個指定的混合訊號之間的相位角度差，以  $\psi_q^P(f_i, \tau_j)$  來表示。它們的表示式分別顯示如下：

$$\psi_q^L(f_i, \tau_j) = \frac{|x_q(f_i, \tau_j)|}{|X(f_i, \tau_j)|_2} \quad (8)$$

$$\psi_q^P(f_i, \tau_j) = \phi[x_q(f_i, \tau_j)] - \phi[x_1(f_i, \tau_j)] \quad (9)$$

其中  $\phi$  為相位的運算子。然後利用一個複數表示式(Complex Representation)來表現這兩個特徵參數。

$$\psi_q(f_i, \tau_j) = \psi_q^L(f_i, \tau_j) \times \exp[j\psi_q^P(f_i, \tau_j)] \quad (10)$$

於是我們得到了一個新的樣本型態(Sample Form)，由  $M$  個 Level-Ratio 和 Phase-Difference 組成的複數值所構成。將原先的觀察資料轉換成這樣的資料型式後，我們即可使用這些新建立的樣本，做後續的處理和訊號分析，包括估計源訊號個數以及混合矩陣。令  $T$  為向量的轉置，則樣本型態表示如下：

$$\Psi(f_i, \tau_j) = [\psi_1(f_i, \tau_j) \cdots \psi_M(f_i, \tau_j)]^T \quad (11)$$

## (二) 源訊號數之估計

我們是藉由 K-Means 分群演算法加上貝氏資訊準則，達到估測源訊號數（語者數）的目的。利用貝氏資訊準則判斷群聚個數  $C$  是  $c$  還是  $c+1$  時，執行 K-Means 分群法後所回傳的結果，何者較能描述資料模型。下面是我們用來選擇模型的貝氏資訊準則公式。

$$BIC = \sum_{u=1}^c n_u^c \times \log |\Sigma_u^c| - \sum_{u=1}^{c+1} n_u^{c+1} \times \log |\Sigma_u^{c+1}| - \lambda \left( M + \frac{M(M+1)}{2} \right) \times \log(n) \quad (12)$$

首先令  $c=2$ ，執行  $c=2$  和 3 的 K-Means 分群演算法，隨後拿這兩個分群結果去做貝氏資訊準則的判斷。若得到的值小於零，停止此程序，且把  $c$  值視為我們偵測到的源訊號數；反之如果求得的值大於零，我們接著設  $c=3$ ，並執行重複的動作，也就是做  $c=4$  的 K-Means 演算法，然後伴隨著  $c=3$  的分群結果再做一次貝氏資訊準則的判斷。K-Means 演算法會因為初始群中心設定的不同而產生明顯的誤差，造成不穩定的分群結果。有鑑於這個問題，我們經由多次執行 K-Means 演算法，並且從回傳的分群情形中選擇一個群聚變異數和(Sum of Cluster Variance) 最小的當作最終決定的結果。

針對每個頻帶，我們會接收到一個  $c$  值，之後統計各個頻帶所回傳的  $c$  值，出現頻率最高的即為最後決定的源訊號個數。

## 四、混合係數矩陣估測以及相位補償技術

我們將混合訊號的時頻點轉換成其對應的樣本型態，在這個樣本型態空間上，先將一些在資料分佈中的離群樣本捨去，使得之後在做估測時，能夠更為準確。方法是參照 KNN 演算法中，搜尋最近鄰居的方式，對每一個資料點都找出與之距離最近的  $K$  個鄰居[12]。然而每筆資料都有一個 In-Degree 值，若有一個樣本被某筆資料視為  $K$  個最近的鄰居之一的話，則該筆資料的 In-Degree 值會往上加一，示意圖如圖一所示。統計每筆資料所屬的 In-Degree 值，假使某一個資料的 In-Degree 值小於門檻值

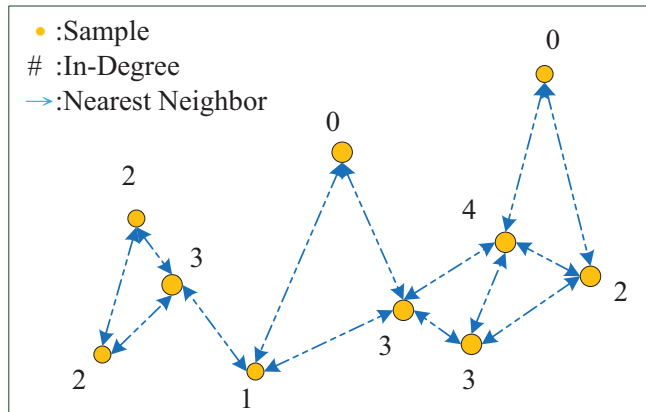


(Threshold) $V$  時，我們就將此樣本當作離群樣本。相反地，如果 In-Degree 值大於  $V$ ，這個樣本就會被保留下來，為之後估計源訊號數和混合矩陣所用。

再利用著名且應用廣泛的分群方法 K-Means 演算法，將樣本型態分割到  $N$  個群聚  $C_k, \dots, C_N$  中，並且利用下面的式子獲得混合向量：

$$h_k = \frac{1}{|C_k|} \sum_{\Psi \in C_k} \Psi, \quad k \in \{1, \dots, N\} \tag{13}$$

其中  $|C_k|$  代表第  $k$  個群聚擁有的樣本數。然而每個混合向量都會對應到一個源訊號。因為我們是根據每個頻帶上的時間序列去估測混合矩陣，所以各個頻帶執行過 K-Means 演算法後都會回傳  $N$  個群聚，並求出代表的  $h_k$ 。最後，如何確認  $h_k$  在矩陣中的位置也是一個很重要的問題。



圖一、KNN Graph ( $c=2$ )。

根據式子(6)，可以得知

$$\frac{h_k(r)}{h_k(s)} = \frac{g_{rk}}{g_{sk}} e^{j2\pi f_i c^{-1} (d_r - d_s) \cos \theta_k} \tag{14}$$

所以經推導後，DOA 可以由下式獲得

$$\theta_k = \cos^{-1} \frac{\phi\left(\frac{h_k(r)}{h_k(s)}\right)}{2\pi f_i c^{-1} (d_r - d_s)} \tag{15}$$

其中  $r$ 、 $s$  是麥克風陣列中兩支距離最近的，它們在混合向量  $h_k$  中所對應到的指標， $d_r$ 、 $d_s$  表示  $r$ 、 $s$  兩支麥克風之間的距離。因為我們對所有  $h_k(k=1, \dots, N)$  都偵測 DOA，所以共得到了  $N$  個角度值。最後根據這個結果確定  $h_k$  在混合矩陣中所對應的行索引。

假設混合訊號的某個時頻點  $X(f_i, \tau_j)$ ，只有源訊號  $k$  為非零的值。透過式子(5)和式子(6)，可將  $X(f_i, \tau_j)$  可表現為：

$$\begin{aligned} X(f_i, \tau_j) &= \begin{bmatrix} g_{1k} e^{j2\pi f_i c^{-1} d_1 \cos \theta_k} \\ \vdots \\ g_{Mk} e^{j2\pi f_i c^{-1} d_M \cos \theta_k} \end{bmatrix} \times g_{s_k(f_i, \tau_j)} e^{j\phi[s_k(f_i, \tau_j)]} \\ &= \begin{bmatrix} g_{1k} g_{s_k(f_i, \tau_j)} \times e^{j(2\pi f_i c^{-1} d_1 \cos \theta_k + \phi[s_k(f_i, \tau_j)])} \\ \vdots \\ g_{Mk} g_{s_k(f_i, \tau_j)} \times e^{j(2\pi f_i c^{-1} d_M \cos \theta_k + \phi[s_k(f_i, \tau_j)])} \end{bmatrix} \end{aligned} \quad (16)$$

因為這篇論文要對由 **Level-Ratio** 以及 **Phase-Difference** 所組成的樣本作群聚分割。所以，得出了混合訊號樣本在極度稀疏的情形下表表現的型式後，我們將式子(16)代入式子(8)和式子(9)，看看若利用這種形態的樣本去定義 **Level-Ratio** 和 **Phase-Difference** 這兩種特徵參數， $\psi_q^L(f_i, \tau_j)$  和  $\psi_q^P(f_i, \tau_j)$  分別為：

$$\psi_q^L(f_i, \tau_j) = g_{qk} / \text{norm}([g_{1k} \cdots g_{Mk}]^T) \quad (17)$$

$$\begin{aligned} \psi_q^P(f_i, \tau_j) &= (2\pi f_i c^{-1} d_q \cos \theta_k + \phi[s_k(f_i, \tau_j)]) \\ &\quad - (2\pi f_i c^{-1} d_1 \cos \theta_k + \phi[s_k(f_i, \tau_j)]) \\ &= 2\pi f_i c^{-1} (d_q - d_1) \cos \theta_k \end{aligned} \quad (18)$$

然後，同樣的將上述兩個特徵參數用複數表示形態來敘述。最後，樣本  $\Psi(f_i, \tau_j)$  會以下的樣子呈現。

$$\Psi(f_i, \tau_j) = \begin{bmatrix} \psi_1^L(f_i, \tau_j) \times e^{j2\pi f_i c^{-1} (d_1 - d_1) \cos \theta_k} \\ \psi_2^L(f_i, \tau_j) \times e^{j2\pi f_i c^{-1} (d_2 - d_1) \cos \theta_k} \\ \vdots \\ \psi_M^L(f_i, \tau_j) \times e^{j2\pi f_i c^{-1} (d_M - d_1) \cos \theta_k} \end{bmatrix} \quad (19)$$

其中，第一項爲一實數值。藉由上式，我們可以說，當語音具有極度稀疏的性質時，只會因爲主導的源訊號不同造成  $\theta_k$  的改變而產生  $N$  種型式的  $\psi(f_i, \tau_j)$ 。所以當結束分群演算法估計混合矩陣之行向量的程序，並且解決了排列問題後，在最理想的情況下，也就是當極度稀疏的條件成立時，混合矩陣會成爲：

$$\begin{bmatrix} \psi_{11}^L & \cdots & \psi_{1N}^L \\ \psi_{21}^L e^{j2\pi f_i c^{-1}(d_2-d_1)\cos\theta_1} & \cdots & \psi_{2N}^L e^{j2\pi f_i c^{-1}(d_2-d_1)\cos\theta_N} \\ \vdots & \ddots & \vdots \\ \psi_{M1}^L e^{j2\pi f_i c^{-1}(d_M-d_1)\cos\theta_1} & \cdots & \psi_{MN}^L e^{j2\pi f_i c^{-1}(d_M-d_1)\cos\theta_N} \end{bmatrix} \quad (20)$$

當然，我們會希望被估計出的混合矩陣越逼近原始的形式越好。在經過估測各個頻帶的 DOA 後，就可以利用這些角度值對混合矩陣的頻率響應做補償，這裡是用  $\hat{\theta}_k^{f_i}$  表示頻率爲  $f_i$  時源訊號  $k$  被估計出之 DOA。觀察式子(20)，可看出相位的部分都是利用到該行向量第一項的相位，採取相位差的表現方式。在我們估計出  $\hat{\theta}_k^{f_i}$  之後，接著利用它去調整混合矩陣的相位部分，實際做法是對混合矩陣的第  $k$  個行向量乘上  $e^{j2\pi f_i c^{-1}d_1 \cos\hat{\theta}_k^{f_i}}$ 。所以修正後的混合矩陣結果爲：

$$\begin{bmatrix} \psi_{11}^L e^{j2\pi f_i c^{-1}d_1 \cos\hat{\theta}_1^{f_i}} & \cdots & \psi_{1N}^L e^{j2\pi f_i c^{-1}d_1 \cos\hat{\theta}_N^{f_i}} \\ \psi_{21}^L e^{j2\pi f_i c^{-1}R_{21}} & \cdots & \psi_{2N}^L e^{j2\pi f_i c^{-1}R_{2N}} \\ \vdots & \ddots & \vdots \\ \psi_{M1}^L e^{j2\pi f_i c^{-1}R_{M1}} & \cdots & \psi_{MN}^L e^{j2\pi f_i c^{-1}R_{MN}} \end{bmatrix} \quad (21)$$

$$R_{mn} = d_m \cos\theta_n + d_1 (\cos\hat{\theta}_n^{f_i} - \cos\theta_n) \quad (22)$$

如果 DOA 的估測夠精準，也就是  $\hat{\theta}_n^{f_i}$  等於  $\theta_n$ ，或是說兩者的差距極小，則我們令

$$R_{mn} = d_m \cos\theta_n \quad (23)$$

最後，將式子(23)代入式子(21)，並且設  $\hat{\theta}_N^{f_i} = \theta_N$ ，我們可以得到下面這種形式的混合矩陣。



$$\begin{bmatrix} \psi_{11}^L e^{j2\pi f_i c^{-1} d_1 \cos \theta_1} & \dots & \psi_{1N}^L e^{j2\pi f_i c^{-1} d_1 \cos \theta_N} \\ \psi_{21}^L e^{j2\pi f_i c^{-1} d_2 \cos \theta_1} & \dots & \psi_{2N}^L e^{j2\pi f_i c^{-1} d_2 \cos \theta_N} \\ \vdots & \ddots & \vdots \\ \psi_{M1}^L e^{j2\pi f_i c^{-1} d_M \cos \theta_1} & \dots & \psi_{MN}^L e^{j2\pi f_i c^{-1} d_M \cos \theta_N} \end{bmatrix} \quad (24)$$

盲訊號源分離在欠定的條件下，根據式子(4)，源訊號  $S$  可以有無限多個解，所以我們利用最小化 L1 範數之以及式子(1)作為限制式，此最佳化問題的解即為所求，如下列式子所示：

$$\min_S \sum_k |S_k|, \quad k=1, \dots, N, \quad s.t. \quad HS = X \quad (25)$$

從這無限多組解中選取一個適當的答案。恢復源訊號的步驟就是依靠這個以 MAP 為基礎的方法[6]。

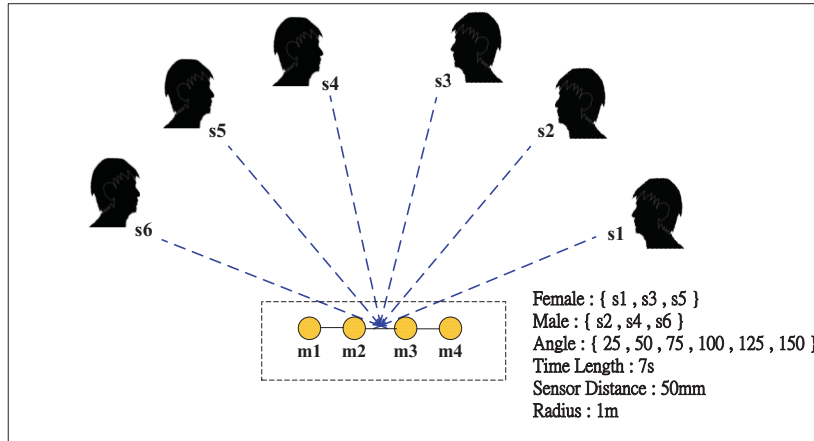
## 五、實驗結果

### (一) 實驗環境

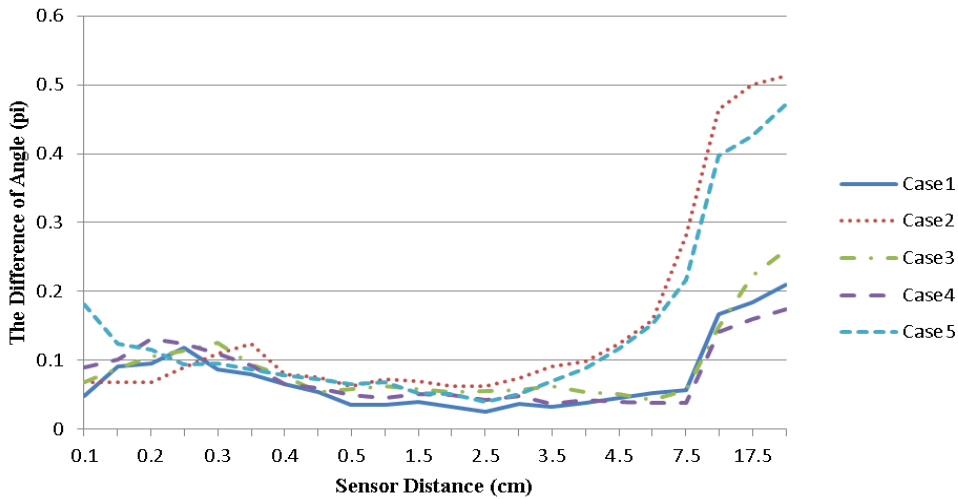
環境架設的部分，示意圖如圖二所示，我們用四支麥克風(感測器)架設一個麥克風陣列。而這個麥克風陣列中，感測器與感測器之間的距離我們設為 50 毫米。實驗情境中共存在六個源訊號，其中包含了三個男生以及三個女生的語音。並且以不同的入射角度圍繞在麥克風陣列的周圍。這些源訊號的取樣頻率(Sampling Rate)為 8000 赫茲，訊號的時間長度為 7 秒鐘。先前曾提到在執行盲訊號源分離前會利用短時傅立葉轉換將訊號轉至時頻域，而關於實驗，我們設短時傅立葉轉換裡的音框(Frame)大小為 256 個樣本點，時間位移量為 64 個樣本點，並採用漢寧窗為訊號做加權的動作。

### (二) 麥克風距離的影響

我們在測試時，發現到麥克風之間距離的長短與估計 DOA 的準確性有關連性。所以我們首先設計了一個實驗，模擬方式是固定 5 組源訊號，每一組包含了 3 個音源，然後取兩支麥克風，根據不同的麥克風距離去收錄源訊號，我們共有 22 種麥克風距離的設定，搭配 5 組源訊號，產生出了 110 種盲訊號源分離的案例。再來計算這 110 個案例中，被估計出之 DOA 與實際聲源角度的平均差距，得到的值越小，表示 DOA 估計的準確性越高。圖三即為此實驗的結果。由此圖可看出，麥克風距離大約在超過 5 公分時，距離越大，準確性越差；然而距離並不是越小越好，當它小到某種程度，估計出的 DOA 與實際音源方位角的差距反而有些微的提高。從此實驗看來，我們理想的麥克風距離大概在 0.5 至 5 公分之間。



圖二、模擬之實驗環境。{s1、s2、s3、s4、s5、s6}代表我們的6個聲源，{m1、m2、m3、m4}則為麥克風陣列上的4個麥克風。



圖三、感測器距離與 DOA 估計之準確度的關聯性。

### (三) 分離訊號效能之比較

本實驗用最大後驗機率的基礎方法和本論文所提出的演算法做比較。根據前面所形容的實驗環境，我們藉由各種可能的案例來評估效能，包括兩支麥克風三個源訊號(2m3n)、三支麥克風四個源訊號(3m4n)以及四支麥克風五個源訊號(4m5n)。針對這三種案例，我們各別選擇了十組、十組和五組測試音檔，然後比較了兩個演算法，我們將訊號干擾比(Signal To Interference, SIR)當作效能的評估準則，實驗數據如表一所示，公式如下所示：

$$SIR = 10 \log_{10} \frac{\|y_{target}\|^2}{\|e_{interf}\|^2} \quad (26)$$

對於 2m3n 的盲訊號源分離情形，利用麥克風陣列中的 m2 和 m3 兩個感測器產生觀察訊號。在所有案例中，Baseline 方法與 Proposed 的 SIR 差距最大來到 4.89 分貝。以平均效能來講，Proposed 也高出 Baseline 方法 2.181 分貝。對於 3m4n 的盲訊號源分離情形，利用麥克風陣列中 m1、m2 和 m3 三個感測器產生觀察訊號。在所有案例中，Baseline 方法與 Proposed 的 SIR 差距最大來到 10.78 分貝。以平均效能來講，Proposed 也高出 Baseline 方法 7.321 分貝。對於 4m5n 的盲訊號源分離情形，Baseline 方法與 Proposed 的 SIR 差距最大來到 14.05 分貝。以平均效能來講，Proposed 也高出 Baseline 方法 7.27 分貝。

表一、兩支麥克風三個源訊號(2m3n)、三支麥克風四個源訊號(3m4n)以及四支麥克風五個源訊號(4m5n)實驗 SIR 比較。

Setting	Baseline [6]	Proposed
2m3n	10.705	12.886
3m4n	3.295	10.616
4m5n	-1.25	6.02

## 六、結論

本篇論文利用 KNN 的方法刪除離群資料，可以使較密集資料保留下來，以利混合矩陣之估測。此外，根據 DOA 的偵測結果，我們可以更進一步進行混和矩陣之相位補償，將其相位更精準的逼近混合矩陣的原貌，利用精準化後的混合矩陣來分離訊號，可獲得比原始方法(Baseline Method)更優異的效能。此外，K-Means 演算法和貝氏資訊準則作結合，並對所有頻帶的結果做整體考量，可達到估測源訊號個數的目的。實驗結果顯示，所有案例使用我們所提出的方法都比傳統最大後驗機率為基礎的方法要出色。

## 參考文獻

- [1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*, New York, John Wiley and Sons, 2002.
- [2] S. Roberts and R. Everson, *Independent component analysis: principles and practice*, Cambridge University Press, 2001.
- [3] S. C. Douglas, M. Gupta, H. Sawada, and S. Makino, "Spatio – temporal FastICA algorithm for the blind separation of convolutive mixtures," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 1540 – 1550, Jul. 2007.
- [4] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 666-678, Mar. 2006.
- [5] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representation," *IEEE Trans. Signal Processing*, vol. 46, pp. 2888-2898, Nov.

1998.

- [6] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and  $l_1$ -norm minimization," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 24717, 12 pages.
- [7] P. Bofill, "Underdetermined blind separation of delayed sound sources in the frequency domain," *Neurocomputing*, vol. 55, no. 3-4, pp. 627-641, 2003.
- [8] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353-2362, Jun. 2001.
- [9] Y. Li, S. I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Processing*, vol. 54, pp. 423-437, Feb. 2006.
- [10] A. Aissa-El-Bey, K. Abed-Mraim, and Y. Grenier, "Blind separation of underdetermined convolutive mixtures using their time-frequency Representation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 1540-1550, Jul. 2007.
- [11] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, pp. 1833-1847, Feb. 2007.
- [12] V. Hautamaki, I. Karkkainen, and P. Franti, "Outlier detection using k-nearest neighbor graph," *IEEE International Conference, ICPR*, pp. 430 – 433, Aug. 2004.