國 立 交 通 大 學

工學院聲音與音樂創意科技

碩士學位學程

碩 士 論 文

暫態噪音聲源方位追蹤

Transient Noise Source Tracking

研 究 生： 張 哲 鳴

指導教授： 胡 竹 生 博士

中 華 民 國 一 百 零 一 年 九 月

暫態噪音聲源方位追蹤

Transient Noise Source Tracking

研 究 生：張　哲　鳴　　Student：Che-Ming Chang

指導教授：胡　竹　生　博士　Advisor：Jwu-Sheng Hu

國 立 交 通 大 學
工學院聲音與音樂創意科技碩士學位學程
碩 士 論 文

A Thesis
Submitted to Master Program of Sound and Music Innovative Technologies
College of Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Engineering

September 2012

Hsinchu, Taiwan, Republic of China

中 華 民 國 一 百 零 一 年 九 月

暫態噪音聲源方位追蹤

研究生：張　哲　鳴　　指導教授：胡　竹　生　博士

國立交通大學工學院聲音與音樂創意科技碩士學位學程

摘　要

本論文提出了一套偵測暫態噪音並使用麥克風陣列追蹤暫態噪音聲源方位的方法。麥克風接收的訊號經過時域振幅刪減法處理之後，可以非常有效的消除麥克風接收到的穩態與非穩態的非暫態噪音訊號對於偵測準確度的影響，對於語音也有一定的抑制效果。使得本方法在環境不理想時也有相當可靠的辨識率，可取代關鍵字做為另一種聲音喚醒機制的選擇。本方法對低維度的矩陣做運算，只需要幾個接收訊號的音框，同時從中辨識出只存在暫態噪音的音框，並只針對這些音框進行聲源方位追蹤，這樣運算量低的特性很適合應用於即時系統上。

Transient Noise Source Tracking

Student： Che-Ming Chang Advisor： Prof. Jwu-Sheng Hu

Master Program of Sound and Music Innovative Technologies

National Chiao Tung University

ABSTRACT

This thesis presents a method of detecting transient noise using microphone array so the transient noise source orientation can be computed. Through the time-domain amplitude subtraction, the effect of both stationary and non-stationary noise in the microphone signal can be effectively eliminated. This includes signals such as voice. It is shown that this method is reliable when the environment is not ideal. This makes the method a better candidate to be a sound cue than key word based mechanism. This method operates on a low-dimensional matrix and needs only a few window frames of receive signals. Meanwhile the source location can also be determined from those frames. This low computational requirement makes it ideal for real-time applications.

誌　謝

兩年的生活說長不長，可是要說短的話，有時候其實又還頗漫長的。但是我終於完成這篇論文了!最要感謝的是我的指導教授胡竹生老師，謝謝老師兩年前給我機會進來 XLAB，讓我可以和這麼優秀的大家一起學習與合作。老師淵博的學識和創新的思維總是能夠啟發我讓我向上，可以微笑自信的走出交大校門，這真的非常謝謝老師一直以來耐心的指導。

接著要感謝的是我的父母親，有很多時候碰到困難，我感到灰心無助，你們總是可以靜靜的聽我鬼打牆般的煩惱，然後細心的給予我建議或是鼓勵我讓我好過一點，你們是我最愛的爸媽也是我最好的朋友。

再來要感謝 XLAB 聲音組的夥伴們：一直在學業上很照顧我的唐哥；聲音組的明日之星耕維，祝你事業與愛情一切順利；還有兩年來不管是一起出去聯誼、熬夜寫作業或是在實驗室裡掙扎生論文的室友兼戰友的宗翰，我們終於順利畢業了!還有哲宇、大夢和小山東，如果可以早點認識你們就好了。還有實驗室漂亮的助理淑伶，謝謝妳在我當實驗室管理員各項事務上的幫忙。另外要感謝實驗室的學長同學們：永融學長、勁源學長、阿吉學長、JUDO 學長、振華學長、德洋學長、男哥、建廷、Daniel、罐頭、阿文、期元、明遠、還有已經畢業的學長姐們，有太多的話想說，但是誌謝還是寫一面就好，所以私底下再說吧。和你們相處很開心，很不想就這樣離開，真的很高興認識你們，很謝謝對我的照顧與包容。

最後也要感謝聲音學程的學弟妹還有輔大的朋友們。找好吃的東西、練吉他上台表演、第一次自己辦聯誼、規劃旅遊行程一起出去玩…。進來交大前在苦悶兵營裡的我從來沒想過之後還能夠再增加這些只屬於學生的美好回憶，以及除了研究生活以外另一種的酸甜苦辣。

謝謝大家，因為有你們和妳們，讓我除了這篇碩士論文之外還有這兩年精彩的碩士生活為我的學生生涯畫下美好的句點。

目　錄

表 列

圖 列

第一章 緒論

1.1 研究動機

生活環境中存在著各式各樣的聲音，大致上可以粗分為人類想要聽到的或是不想聽到的聲音。想要聽到的聲音包括語音、或是有規律振動令人感覺悅耳愉快的樂音，而不想聽到的聲音則像是無意義的聲音或是一般人類認定的噪音,可以區分為非暫態以及暫態:非暫態噪音像是引擎聲、展場吵雜的低沉人聲、...等持續性的聲音；暫態噪音是初始有能量很強的峰值然後迅速在短時間內衰減的聲音。在我們生活周遭隨時可以聽到它，例如：敲門聲、敲擊鍵盤聲、拍手聲、... 等，都是屬於暫態噪音。

通常我們會希望能夠在完整保留想要聽到的聲音的情況下將無關的聲音去除，如語音增強技術(speech enhancement)就是要抑制和語音無關的噪音。一般要抑制非暫態噪音可以先估算噪音的頻譜，最常見的是利用頻譜刪減法[1],然後在聲音頻譜中減去噪音頻譜。如果想要消除的是暫態噪音，可以估測暫態噪音在訊號中出現的位置及頻譜，再將它刪減掉[2]。那既然有辦法從聲音或是其它非暫態的聲音之中將暫態噪音的位置和成分萃取出來，代表暫態噪音相較於其它聲音辨識度較高，藉由暫態噪音偵測也可以將暫態噪音做為一種指標，那是否可以利用這個特性做為其它用途呢？

一般聲源角度估測或是聲源追蹤的演算法[3][4][5],最後得到的聲源位置通常是訊號能量最大或是訊號相關程度最高的聲源。在這些演算法中我們無法預先設定要估測的聲源種類，在必須估測特定聲源方位的情況下，這些演算法就無法符合需求。因此聲源方位估測演算法如果能夠只針對特定聲源進行處理，就能夠做更多元的運用。

裝置喚醒(wake-up)就是偵測特定聲源並估測方位的演算法[6]的應用：我們事先設定某些關鍵字來做為控制裝置的開關，當偵測到空間中存在關鍵字時，就會估算出關鍵字的位置，並啟動事先設定的程序並反應。這樣的技術可以很有效的在訊噪比不理想的環境下達到遠距控制裝置的訴求。但是使用關鍵字做為啟動裝置的機制還是有它使用不方便的地方：如果使用者無法說話，或是關鍵字為使用者不熟悉的語言而使用者無法正確說出標準的關鍵字。所以我們希望找出某種辨識度高而且是使用者很容易就能製造的聲音，在聲控系統中輔助或是取代關鍵字喚醒技術。暫態噪音是一種符合以上需求的聲音，使用者可以簡單的藉由拍手或是敲擊物體製造暫態噪音。因此本論文嘗試利用暫態噪音達到關鍵字喚醒裝置的效果，並讓裝置知道使用者的位置資訊以便於後續與使用者互動。

1.2 研究目標

由於以上目的，在此將本論文的目標分為：

1. 壓抑非暫態噪音的聲音成分以提升暫態噪音活動偵測的穩健度。

2. 從一段聲音訊號中，準確偵測暫態噪音出現的位置。

3. 根據暫態噪音活動偵測的結果，估測暫態噪音的聲源方位。

1.3 本研究創新說明

本研究提出一套偵測環境中暫態噪音活動以及追蹤暫態噪音聲源位置的演算法。一般聲源方位估測演算法無法針對特定聲源做處理，本論文的方法可以單獨對暫態訊號聲源的位置進行估測。另外相較於一般偵測暫態訊號的演算法，本論文利用時域振幅刪減的方法來增加暫態噪音活動偵測的辨識準確度，對於降低非暫態噪音以及語音對於辨識的干擾都有相當好的效果。在偵測方法中設計不需要訓練資料(training data)低維度矩陣運算的演算法，只需 5 個音框就能從接收訊號中辨識出只存在暫態噪音的音框，因此可達到即時追蹤暫態噪音聲源的需求。

1.4 論文架構

本論文包含了三個主要的部分，分別為演算法的背景相關技術、論文提出的演算法以及方法的實驗與分析。以下描述各章節的內容：

第二章：　麥克風陣列技術的介紹，聲源方位估測方法與利用 non-localdiffusion filter 分群的方法。

第三章：　介紹本論文的架構與演算法，如何利用暫態噪音聲源估測從一段麥克風陣列接收到的訊號中找出存在暫態噪音的音框再進行聲源估測。

第四章：　對不同性質的非暫態干擾聲源測試論文方法的強健度。

第五章：　對論文方法與測試結果進行總結，提出可改進的部分。

第二章 背景技術介紹

2.1 Non-local diffusion filter

在一般訊號處理的相關研究領域裡，不論是影像或是聲音，常需利用向量的形式來描述所欲探討之資料物件。將物件的特徵以向量之數學化形式呈現後，一個物件即可被視為向量空間中的一點，而一群物件將會在空間中形成某種分佈，利用這樣的特性可以在實際應用上的進行分析，如對於物件的分群(clustering)、分類(classification)、特徵萃取(feature extraction)或辨識（recognition）。

為了提升問題分析的正確率，必須將資料物件描述詳盡，提供更多的特徵於向量表示中。但是採用較複雜的物件特徵，也相對提高了向量空間的維度，同時也需要更多的樣本數維持穩定的資料分析正確率，因此高維度會導致計算上的複雜度以及樣本數不足的問題。為了解決這個問題，我們可以透過適用於資料分類或分群的降維（dimensionality reduction）演算法來萃取或保持其在高維度空間裡所隱含的重要性質。

在近年來的研究中，根據接收資料的幾何特性做為資料分析基礎的演算法越來越多。這些幾何方法或是流形學習（manifold learning）的方法，都是為了獲取資料的結構資訊做為可作為其他進階應用的前處理[7] [8] [9]。流形學習(manifold learning)是一種降維分析的方法，流形(manifold)在數學上是低維度曲面，在定義上為一拓撲空間，其區域性（local）子空間可被視為為歐氏空間（Euclidean space），就是在它的二相鄰座標點間之距離可用歐氏距離（Euclidean distance）來衡量。流形學習的方法是假設所要分析的物件資料在高維度空間中有平滑流形分佈，再利用轉置

（re-embed）的方法將物件資料映到較低維度的歐氏空間，並區域性（locally）保持其原有在流形上的分佈，例如Isomap（ISOmetric feature MAPping）[10]以及LLE（Locally Linear Embedding）[11]。透過Isomap 與LLE 的使用，可將原先分佈於高維空間的資料點，在低維度的空間中呈現，並保持這些資料點在高維度空間中的重要分佈結構。Local Discriminant Embedding（LDE）[12]是改良後的流形學習演算法，針對具有標示類別（class label）的物件資料，進行以分類為目標的最佳化處理，建構於二個物件資料鄰近關係圖（neighborhood graph），分別用於記錄物件間的類別關係，與鄰近幾何關係，並可將最佳化問題推導至解特徵值問題，因此最佳降維空間可經由簡易的特徵值與特徵向量的計算而獲得。

Non-local diffusion filter是基於鄰近關係圖發展出來的演算法，利用在非相鄰區域的資料間建立的核心方程式，可以同時對不同區域的資料進行處理，常用於影像處理上[13] [14] [15]，目前在聲音的應用上較少，是對於音訊處理來說是很有發展性的技術，可用於語音強化上[16] [17]。

以下介紹non-local diffusion filter定義方式和應用：給定 $\mathbf{\Gamma} = \{x_i\}_{i=1}^{L}$ 代表一組由 L 個取樣點組成高維度的資料集合，為了能夠獲得資料集合 $\mathbf{\Gamma}$ 的幾何特性，我們定義核心方程式 k_σ ，並由方程式規模參數 σ 控制方程式適當的大小，再由核心方程式建構隨機行走機率轉移矩陣，最後將 $\mathbf{P}$ 乘上資料矩陣，就能夠獲得幾何特性更明顯的資料矩陣。

核心方程式表示不同資料點的相關性是由事先定義的資料幾何關係並依照資料特性所建構的，因此核心方程式可以獲取資料集合的幾何資訊。核心方程式有下列特性：

(1) 對稱性(symmetry)：

$$k_\sigma\left(x_i, x_j\right) = k_\sigma\left(x_j, x_i\right) \tag{2.1.1}$$

(2) 非負值(non - negativity) :

$$k_\sigma\left(x_i, x_j\right) \geq 0 \tag{2.1.2}$$

(3) 快速遞減(Fast decay) :

$$\sigma > 0\,,\ k_\sigma(x_i\,, x_j) \to 1 \ for\ \|x_i - x_j\| \ll \sigma$$

$$k_\sigma(x_i\,, x_j) \to 0 \ for\ \|x_i - x_j\| \gg \sigma \tag{2.1.3}$$

我們使用高斯核心(Gaussian kernel)方程式以符合以上特性：

$$k_\sigma\left(x_i, x_j\right) = e^{\left\{-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right\}} \tag{2.1.4}$$

將核心方程式正規化之後，建立非對稱的矩陣：

$$p\left(x_i, x_j\right) = \frac{k\left(x_i, x_j\right)}{d(x_i)} \tag{2.1.5}$$

$$d(x_i) = \sum_{j=1}^{M} k(x_i, x_j) \tag{2.1.6}$$

因為核心方程式一定為正的性質，$p(x_i, x_j) > 0$ 並且 $\sum_{j=1}^{M} p(x_i, x_j) = 1$ ，正規化之後的核心方程式可以表示為描述資料集合 $\mathbf{\Gamma} = \{x_i\}_{i=1}^{L}$ 中資料點之間關聯性的馬可夫鏈轉移機率方程式。資料集合 $\{x_i\}_{i=1}^{L}$ 代表馬可夫鏈的狀態空間，方程式 $p(x_i, x_j)$ 代表從資料點 x_i 到另一個資料點 x_j 一階隨機行走轉移機率。

(2.3.5) 式可以改寫為矩陣的形式 $\mathbf{P} = \mathbf{D^{-1}K}$ ，其中 $\mathbf{D}$ 是 $\mathbf{D}_{ii} = d(x_i) = \sum_{j=1}^{M} k(x_i, x_j)$ 組成的對角矩陣，$\mathbf{X} = [x_1, x_2, \ldots, x_M]^T$ 是從 $\mathbf{\Gamma} = \{x_i\}_{i=1}^{L}$ 中取樣 M 點的資料點矩陣，將一階隨機行走(a single random-walk step)轉移機率矩陣運作在資料點矩陣上，可以寫成 $\mathbf{PX}$ ，若隨機行走為 t 階運作在資料點矩陣上時，寫成 $\mathbf{P}^t\mathbf{X}$ 。

舉例來說 $[\mathbf{PX}]_\mathbf{i}$ 為矩陣 $\mathbf{PX}$ 的第 i 列，表示 x_i 在一階隨機行走後的期望值：

$$[\mathbf{PX}]_\mathbf{i} = \sum_{j=1}^{M} \mathbf{P}_{ij} x_j$$

$$= \sum_{j=1}^{M} \Pr\{\mathcal{X}_{\tau+1} = x_j \,|\, \mathcal{X}_\tau = x_i\} x_j = \mathbb{E}[\mathcal{X}_{\tau+1} | \mathcal{X}_\tau = x_i] \qquad (2.1.7)$$

$\mathcal{X}_\tau$ 是轉移機率矩陣定義的馬可夫程序(Markovian process)，其中時間索引參數是 τ 。

這邊用一個實際的簡單例子解釋 non-local diffusion filter 是如何將數列中與其它元素數值距離差異較大的元素分離開來，並將距離差異較小的元素分為同一類。

假設有一數列為 $\mathbf{X} = [10, 200, 5, 50, 30]^{\boldsymbol{T}}$ ，由(2.1.4)、(2.1.5)、(2.1.6)

可以得到核心矩陣 $\mathbf{K} = \begin{bmatrix} 1 & 0 & 0.97 & 0.14 & 0.61 \\ 0 & 1 & 0 & 0 & 0 \\ 0.97 & 0 & 1 & 0.08 & 0.46 \\ 0.14 & 0 & 0.08 & 1 & 0.61 \\ 0.61 & 0 & 0.46 & 0.61 & 1 \end{bmatrix}$ ，

以及一階轉移機率矩陣 $\mathbf{P} = \begin{bmatrix} 0.37 & 0 & 0.36 & 0.05 & 0.22 \\ 0 & 1 & 0 & 0 & 0 \\ 0.38 & 0 & 0.4 & 0.03 & 0.18 \\ 0.07 & 0 & 0.04 & 0.55 & 0.33 \\ 0.23 & 0 & 0.17 & 0.23 & 0.37 \end{bmatrix}$ ，

$\mathbf{X}' = \mathbf{PX} = [14.7, 200, 12.9, 38.4, 25.7]^{\boldsymbol{T}}$

若為 t 階轉移機率運作在資料點矩陣，如 $t = 100$

100 階轉移機率矩陣 $\mathbf{P}^{100} = \begin{bmatrix} 0.28 & 0 & 0.26 & 0.19 & 0.27 \\ 0 & 1 & 0 & 0 & 0 \\ 0.28 & 0 & 0.26 & 0.19 & 0.27 \\ 0.28 & 0 & 0.26 & 0.19 & 0.27 \\ 0.28 & 0 & 0.26 & 0.19 & 0.27 \end{bmatrix}$

$\mathbf{X}'' = \mathbf{P}^{100}\mathbf{X} = [21.7, 200, 21.7, 21.7, 21.7]^{\boldsymbol{T}}$

2.2 麥克風陣列訊號處理

陣列訊號處理，是利用個感測器排列成特定形狀接收訊號並進行處理的技術。一般的訊號處理基本上是由訊號時域或頻域的資訊中找出訊號的特性對訊號進行處理。而在陣列訊號處理中，空間中任一點聲源發出訊號後，經過在空間中的傳遞，到達陣列中不同位置的感測器時會產生許多差異，如接收能量不同或接收時間的延遲，不同位置的感測器會接收到彼此有空間關係聯結著的相異訊號，因此我們可以利用空間資訊進行更多元的應用與分析。為了讓陣列訊號處理理論更精簡，通常有兩個基本假設：

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

陣列處理技術中主要是對感測器接收訊號的差異進行處理，但是在不同的陣列架構下，感測器間的關係也不相同，而對於不同形狀陣列結構的研究也是陣列訊號處理的重點之一，以下列舉兩種常用的陣列結構。

均勻線性陣列(Uniform Linear Array，ULA)：

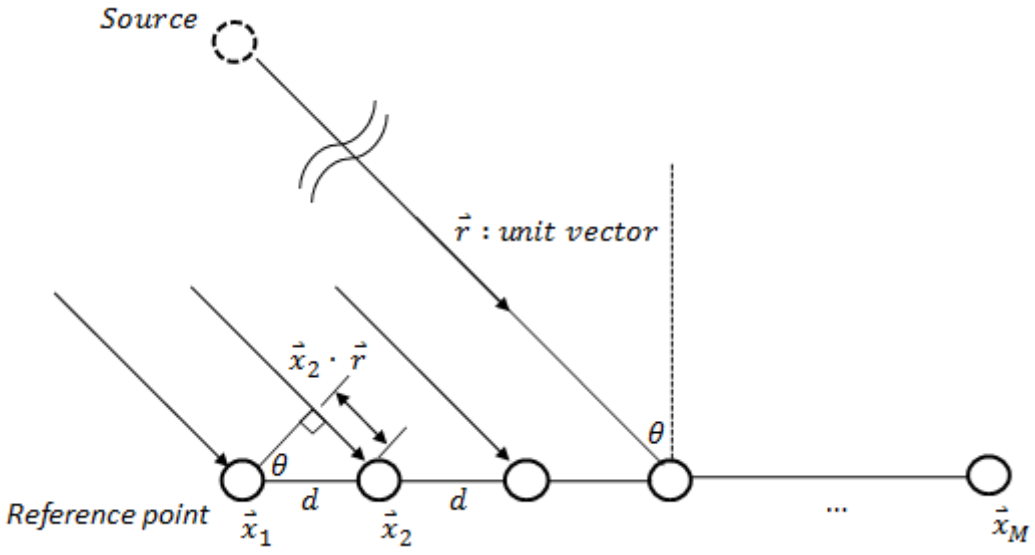

圖 2.1 均勻線性陣列架構圖

最為廣泛使用的一維陣列結構是均勻線性陣列，架構圖如圖 2.1。利用基本假設聲波為遠場平面波，可以推導陣列對訊號的向量。s(t)代表訊號來源，n(t)代表雜訊，則具有 M 個感測器的均勻線性陣列輸出可以表示為下列向量的形式：

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} s(t)e^{j\omega_c \frac{\vec{x}_1 \cdot \vec{r}}{c}} \\ s(t)e^{j\omega_c \frac{\vec{x}_2 \cdot \vec{r}}{c}} \\ \vdots \\ s(t)e^{j\omega_c \frac{\vec{x}_M \cdot \vec{r}}{c}} \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix}$$

$$= \begin{bmatrix} e^{jk_c \vec{x}_1 \cdot \vec{r}} \\ e^{jk_c \vec{x}_2 \cdot \vec{r}} \\ \vdots \\ e^{jk_c \vec{x}_M \cdot \vec{r}} \end{bmatrix} s(t) + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix} = \boldsymbol{a}(\vec{r})s(t) + \boldsymbol{n}(t) \quad (2.2.1)$$

$k_c = \frac{\omega_c}{c} = \frac{2\pi}{\lambda_c}$ ，其中 k_c 稱為波數，λ_c 為波長，c 為波速

$\boldsymbol{a}(\vec{r})$ 稱為 array manifold vector，代表由訊號來源到各感測器間的時間關係，可以將其表示為：

$$\boldsymbol{a}^T(\theta) = [1 \quad e^{jk_c d sin\theta} \quad \ldots \quad e^{jk_c (M-1) d sin\theta}] \quad (2.2.2)$$

均勻環型陣列(Uniform Circle Array, UCA)：

均勻環型陣列是一種基本的二維陣列分佈結構，如圖 2.2 感測器等距的均勻分佈於環狀結構，R 代表陣列的環形半徑，M 為陣列的感測器個數。由於 UCA 結構為二維的陣列分佈，因此可以探索 2-D 維度的空間資訊，這也是本論文主要所使用的陣列結構。

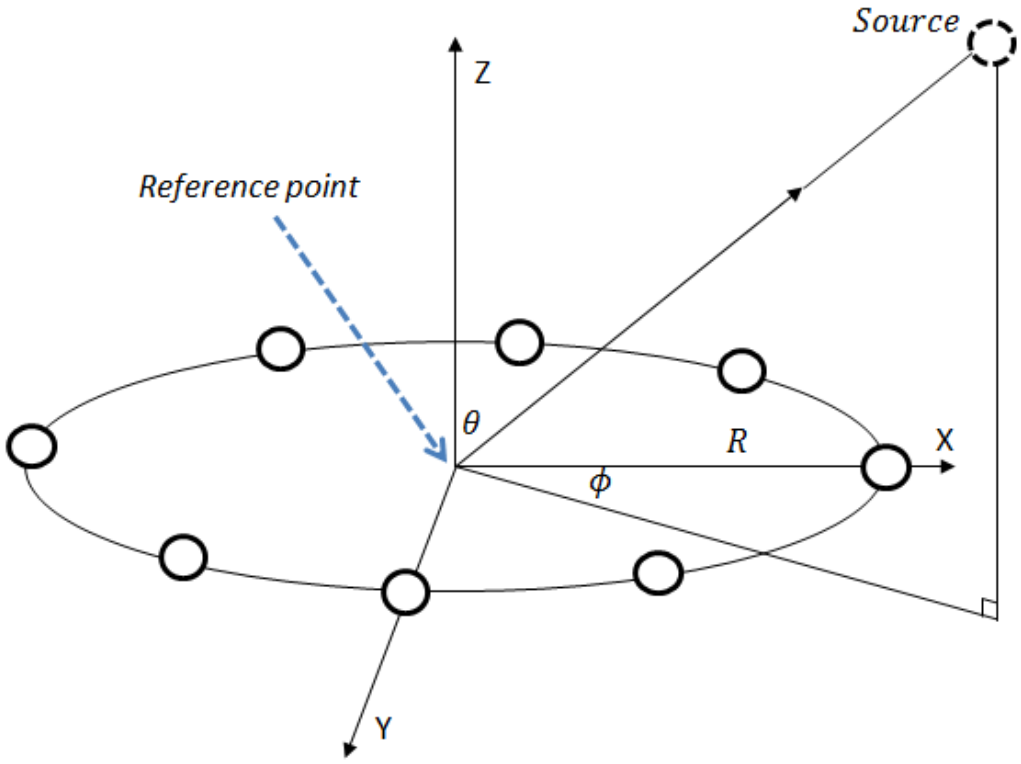

圖 2.2 均勻環型陣列架構圖

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

$$\boldsymbol{a}^T(\phi) = [1 \quad e^{jk_c R sin\theta cos\phi} \quad \ldots \quad e^{jk_c R sin\theta \cos(\phi - \frac{2(M-1)\pi}{M})}] \tag{2.2.3}$$

在陣列訊號處理的應用中，依使用目的與研究方向的不同，大致可分為兩個主要的類別：

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

著重於估測訊號的數量或在空間中的方位，利用陣列感測器間的差異對空間中聲源的個數或方位進行估測。

波束形成理論(Beamformer)：

利用訊號的空間關係，使其能對不同方位的訊號做出不同增益，以分離空間中不同方位的聲源訊號，如同空間濾波的效果。一般稱為波束形成(Beamformer)或視為一種空間濾波器(Spatial Filter)。

2.3 訊號到達角度估測

訊號到達角度(DOA)估測是列訊號處理的技術中一個重要的研究方向，依照技術面不同可以分為三大類：第一類是利用感測器接收的能量差異來偵測訊號的到達方位，常見的方法如 MVDR 與 Steered beamformer(SBF)。第二類為利用由訊號到不同感測器間的時間延遲，估測訊號的到達方位，常見的方法如 GCC(Generalized Cross-Correlation)。第三種，是利用不同訊號源間特徵向量的分佈關係，以互相投影或判斷相似度的方式估測訊號的位置，稱為特徵結構法(Eigenstructure Method)，常見的方法如 MUSIC 與 ESPRIT。

2.3.1 Multiple Signals Classification Method (MUSIC)

MUSIC 是一種利用特徵結構估測訊號到達角度的方法，在特徵結構法中是常用的估測方法[18]。所謂特徵結構法，是將各感測器接收的資料計算其相關矩陣(Correlation Matrix)，並將相關矩陣進行特徵值分解(Eigenvalue decomposition)。為了區別訊號與雜訊，依照特徵值大小將相關矩陣的特徵向量空間可分為兩個子空間，分別為訊號子空間(Signal Subspace)與雜訊子空間(Noise Subspace)。由於訊號子空間與雜訊子空間為正交關係，因此包含於訊號子空間中對應訊號來源方向的指向向量(Steering Vectors)必與雜訊子空間正交，利用對應關係尋找訊號到達角度。MUSIC 演算法必須滿足兩個基本假設：

1. 訊號相關矩陣(Source Correlation Matrix)必須是滿秩(Full Rank)且等於訊號來源的數目 D。
2. Array manifold vector $\boldsymbol{a}(\theta_i)$ ， $i = 1,2,\dots,D$ 彼此間必須線性獨立，滿足 Array manifold Array 是滿秩，而秩必須等於訊號來源數目 D。

在滿足以上兩個基本假設下，感測器接收到的時域訊號(2.2.1)可以改寫成多個訊號源的形式：

$$\mathrm{x(t)} = \sum_{i=1}^{D} \boldsymbol{a}(\theta_i) s_i(t) + \boldsymbol{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{2.3.1}$$

$$\mathbf{A} = [a(\theta_1) \cdots a(\theta_D)\,], \mathbf{s}^T(t) = [s_1(t) \cdots s_D(t)]$$

利用 STFT 將其轉換至頻域：

$$\boldsymbol{X}(\omega_f, k) = \boldsymbol{A}(\omega_f)\mathrm{S}(\omega_f, k) + \boldsymbol{N}(\omega_f, k), f = 1 \cdots \mathrm{F} \tag{2.3.2}$$

假設訊號與雜訊彼此不相關，則資料相關矩陣(Data Correlation Matrix)：

$$\begin{aligned}\mathbf{R}_{\mathrm{xx}} &= E\left(\mathbf{X}(\omega_f, k), \mathbf{X}(\omega_f, k)^H\right) \\ &= \mathbf{A}(\omega_f)\mathbf{R}_{\mathrm{ss}}(\omega_f, k)\mathbf{A}(\omega_f)^H + {\sigma_N}^2(\omega_f, k)\boldsymbol{I}\end{aligned} \tag{2.3.3}$$

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

$$\boldsymbol{R}_{xx}(\omega_f) = \sum_{i=1}^{M} \lambda_i(\omega_f) V_i(\omega_f) V_i(\omega_f)^H \tag{2.3.4}$$

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

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

$$\sigma_N{}^2(\omega_f)\boldsymbol{I} = \sum_{i=1}^{M} \sigma_N{}^2(\omega_f) V_i(\omega_f) V_i{}^H(\omega_f) \tag{2.3.5}$$

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

$$\begin{aligned}\mathbf{C}_{xx}(\omega_f) &= \mathbf{A}(\omega_f)\boldsymbol{R}_{ss}(\omega_f, k)\mathbf{A}^H(\omega_f) \\ &= \sum_{i=1}^{M} \left(\lambda_i(\omega_f) - {\sigma_N}^2(\omega_f)\right) V_i(\omega_f) V_i{}^H(\omega_f)\end{aligned} \tag{2.3.6}$$

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

$$Rs\left(\boldsymbol{C}_{xx}(\omega_f)\right) = span\{\boldsymbol{V}_1(\omega_f), \cdots, \boldsymbol{V}_D(\omega_f)\}$$

$$Rs\left(\boldsymbol{A}(\omega_f)\right) = span\{\boldsymbol{a}(\theta_1,\omega_f),\ \cdots,\boldsymbol{a}(\theta_D,\omega_f)\} = span\{\boldsymbol{V}_1(\omega_f),\cdots,\boldsymbol{V}_D(\omega_f)\}$$

$$Rs\left(\boldsymbol{A}(\omega_f)\right)^{\perp} = span\{\boldsymbol{V}_{D+1}(\omega_f),\cdots,\boldsymbol{V}_M(\omega_f)\}$$

經由以上的結果，可以定義訊號與雜訊子空間：

1. 訊號子空間由前 D 個特徵向量所組成

$$\mathbf{R}_s(\omega_f) = span\{\mathbf{V}_1(\omega_f),\cdots,\mathbf{V}_D(\omega_f)\}$$

2. 雜訊子空間由剩下的 M-D 個特徵向量所構成。

$$\mathbf{R}_N(\omega_f) = span\{\mathbf{V}_{D+1}(\omega_f),\cdots,\mathbf{V}_M(\omega_f)\}$$

利用訊號子空間與雜訊子空間的正交關係，可以推得：

$$\boldsymbol{V}_j(\omega_f)^H\boldsymbol{a}(\theta_i,\omega_f) = 0\text{ , j} = D+1,\ \cdots,\ M\text{ , }i = 1\ ,\ \cdots,\ D \qquad (2.3.7)$$

建立一個投影到雜訊子空間的投影矩陣：

$$\mathbf{P}_N(\omega_f) = \sum_{i=D+1}^{M}\mathbf{V}_z(\omega_f)\mathbf{V}_z^{\ H}(\omega_f) \qquad (2.3.8)$$

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

$$\mathbf{P}_N(\omega_f)\boldsymbol{a}(\theta,\omega_f) = 0 \qquad (2.3.9)$$

為了能更容易找到訊號到達角度 θ，取(2.2.9)的大小：

$$\left\|\mathbf{P}_N(\omega_f)\boldsymbol{a}(\theta,\omega_f)\right\|_2^2 = \boldsymbol{a}^H(\theta,\omega_f)\mathbf{P}_N(\omega_f)\boldsymbol{a}(\theta,\omega_f) = 0 \qquad (2.3.10)$$

$$\boldsymbol{S}_{MUSIC}(\theta,\omega_f) = \frac{1}{\boldsymbol{a}^H(\theta,\omega_f)\mathbf{P}_N(\omega_f)\boldsymbol{a}(\theta,\omega_f)} \qquad (2.3.11)$$

利用方程式(2.2.11)便可計算出 MUSIC spectrum，搜尋 MUSIC spectrum 中無限大處，便代表此處為訊號到達角度。但是現實環境所收到的訊號因為受到雜訊影響，並不存在無限高的 spectrum，但在訊號來源角度的 spectrum 依然會大於其他角度，因此可以藉由搜尋局部最大值的方式找到訊號來源角度。

2.3.2 Steered beamformer (SBF)

Steered beamformer 是偵測各方向能量，並藉此決定聲源方位的演算法，計算簡單運算量低且容易實現[19]。如圖 2.1 均勻線性陣列架構圖，感測器接收到的時域訊號(2.2.1)經過短時距傅利葉轉換後：

$$\boldsymbol{X}(\omega_f,k)=\boldsymbol{A}(\omega_f,\theta)\mathrm{S}\,(\omega_f,k)+\boldsymbol{N}(\omega_f,k)$$

$$\boldsymbol{X}(\omega_f,k)\begin{bmatrix}X_1(\omega_f,k)\\X_2(\omega_f,k)\\\vdots\\X_m(\omega_f,k)\end{bmatrix}=\begin{bmatrix}1\\e^{j\omega_f\kappa dsin\theta}\\\vdots\\e^{j\omega_f\kappa(m-1)dsin\theta}\end{bmatrix}\mathrm{S}\,(\omega_f,k)+\boldsymbol{N}(\omega_f,k)\qquad(2.3.12)$$

利用 Array manifold vector 來估算原本的聲源訊號：

$$\boldsymbol{A}'(\omega_f,\theta')=[1\quad e^{-j\omega_f\kappa dsin\theta'}\quad\dots\quad e^{-j\omega_f\kappa(m-1)dsin\theta'}]\qquad(2.3.13)$$

$$\begin{aligned}\mathrm{S}'(\omega_f,k)&=\boldsymbol{A}'(\omega_f,\theta')\boldsymbol{X}(\omega_f,k)\\&=\boldsymbol{A}''\mathrm{S}(\omega_f,k)+\boldsymbol{A}'(\omega_f,\phi)\boldsymbol{N}(\omega_f,k)\end{aligned}\qquad(2.2.14)$$

$$\boldsymbol{A}''(\omega_f,\theta')=\boldsymbol{A}'(\omega_f,\theta')\,\boldsymbol{A}(\omega_f,\theta)$$

$$=[1\quad e^{-j\omega_f\kappa dsin\theta'}\quad\dots\quad e^{-j\omega_f\kappa(m-1)dsin\theta'}]\begin{bmatrix}1\\e^{j\omega_f\kappa dsin\theta}\\\vdots\\e^{j\omega_f\kappa(m-1)dsin\theta}\end{bmatrix}\qquad(2.3.15)$$

在(2.3.13)式中，定義 $\boldsymbol{A}'(\omega_f,\theta')$ 為 array manifold vector，從(2.3.14)，(2.3.15)可以得知，如果選定的 $\boldsymbol{A}'(\omega_f,\theta')$中，$\theta'=\theta$，$\mathrm{S}'\,(\omega_f,k)$ 會是最大的，代表選取的角度 θ' 就是聲源角度。將估算的聲源訊號大小做為 SBF spectrum，在訊號來源角度處的 spectrum 會大於其它角度，因此可以由搜尋最大值的方式找到訊號來源角度：

$$S_{SBF}(\omega_f,\theta')=\left\|\boldsymbol{A}'(\omega_f,\phi)\boldsymbol{X}(\omega_f,k)\right\|_2^2\qquad(2.3.16)$$

第三章 暫態噪音聲源方位估測演算法

暫態噪音聲源方位估測演算法主要是依賴以下兩種特性：

1. 暫態噪音不論是在頻域或時域結構上都與一般聲音訊號不一樣。從其他干擾聲源相對較小的訊號中，我們可以聽到有暫態噪音的存在，也可以從頻譜圖或是時域振幅圖上看出暫態噪音與其他聲音的差異。
2. 聲源方位估測是從一段麥克風陣列接收到的資料中，尋找主要目標成分的來源方向。而環境的噪音或是其它干擾聲源存在於資料中都會影響統計特性，進而使得估測結果不一定符合我們的需求。因此如果能夠將其它不存在目標聲源的資料移除，而只選取包含目標聲源的資料做處理，可以得到更精確的目標聲源估測結果。

本論文嘗試用這兩種特性進行暫態聲源估測：

1. 將麥克風陣列接收到的訊號利用訊號處理的方法將它們的差異加大，以增強暫態噪音活動偵測的穩健度。透過 Non-local diffusion filter 可以將頻帶能量分群，將數值分佈相似的頻帶歸類為同一類別。再從分類的結果判別音框是否存在暫態噪音。
2. 知道存在暫態噪音的音框索引參數後，就可以根據參數只針對暫態噪音做處理。由於暫態噪音是短時間內能量很強的訊號，在存在暫態噪音的音框中，暫態噪音會占音框內的訊號絕大部分成分而蓋掉其它成分的訊號，因此估測音框內訊號得到的聲源方向就是暫態噪音聲源方向。

3.1 演算法架構

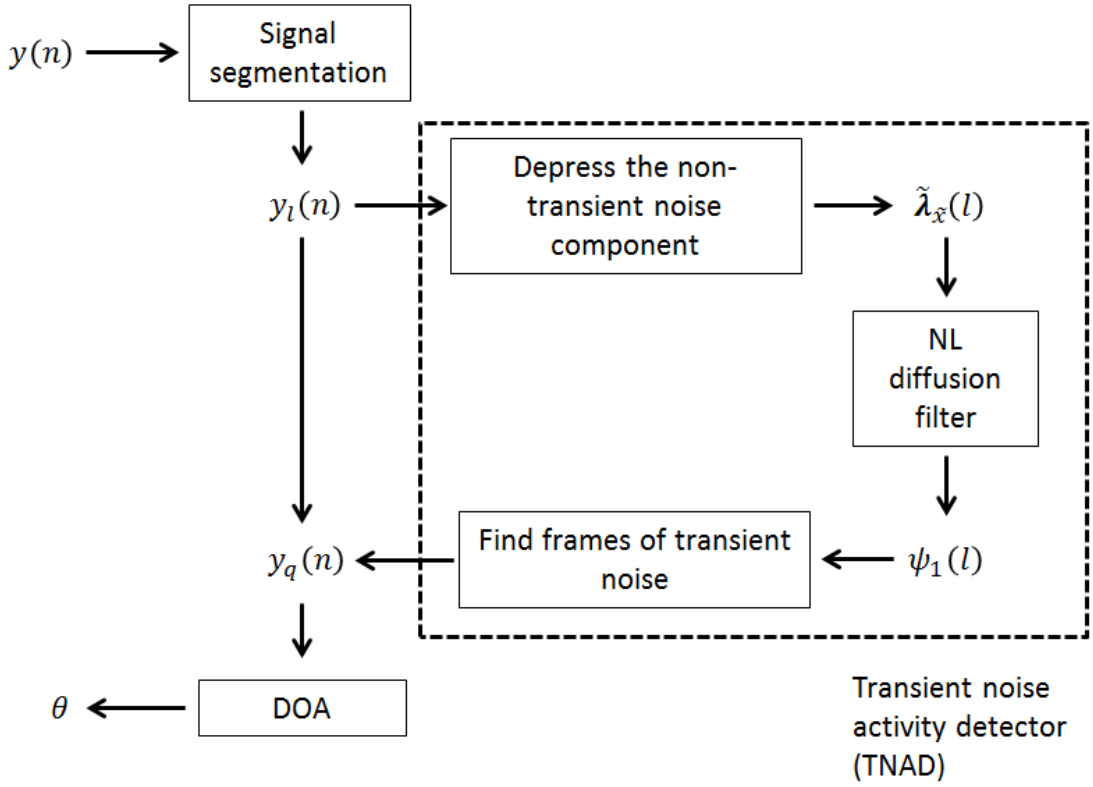

圖 3.1 演算法架構圖

暫態噪音聲源方位估測演算法說明：

1. 麥克風接收聲源訊號為 $y(n)$，將分成若干個音框 $y_l(n)$ 分別處理，並抑制非暫態噪音的成分，再計算音框中各個頻帶的能量 $\tilde{\lambda}_{\tilde{x}}(l)$。

2. $\tilde{\lambda}_{\tilde{x}}(l)$ 經過 NL diffusion filter 處理後可以得到描述各音框頻帶能量關係的馬可夫鏈轉移機率矩陣，將特定階數的馬可夫鏈轉移機率矩陣乘上音框頻帶能量矩陣，再特徵值分解後得到的特徵向量 $\psi_1(l)$ 可作為估測音框 l 內存在暫態噪音的指標。

3. 選取存在暫態噪音的音框 q 內的訊號 $y_q(n)$，經過 DOA 演算法估算暫態噪音聲源訊號的位置角度 θ 。

3.2 暫態噪音活動偵測

暫態噪音(Transient noise)又叫做脈衝噪音(Impulse noise)，基本結構由脈衝訊號組成。初始有能量很強的峰值然後迅速在短時間內衰減，通常持續時間在 10ms 到 50ms 之間。可以用(3.2.1)的方程式定義它：

$$x(n) = h(n) * (b(n)v(n)) \tag{3.2.1}$$

$h(n)$ 是通過特定的濾波器的脈衝響應，以描述每一個暫態事件時域訊號形狀。$b(n)$ 是二元隨機序列 $\{0, 1\}$，描述暫態噪音在時域上發生與否。$v(n)$ 是一串經過連續隨機過程的值，描述暫態噪音的振幅大小。

利用暫態噪音與其它聲音頻帶能量分佈特性的不同，暫態噪音活動偵測(Transient Noise Activity Detection, TNAD)可以將暫態噪音從一般的非暫態噪音以及語音中分離出來。$y(n)$ 是麥克風接收到的時域訊號，由暫態噪音 $x(n)$ 和非暫態的一般聲音 $u(n)$ 所組成：

$$y(n) = x(n) + u(n) \tag{3.2.2}$$

利用短時距傅利葉轉換(STFT)將接收到的時域訊號轉換至頻域：

$$Y(l,k) = X(l,k) + U(l,k) \tag{3.2.3}$$

使用週期圖法(periodogram)計算頻譜能量：

$$\lambda_y(l,k) \triangleq \frac{1}{N}|\mathrm{Y(l,k)}|^2 \tag{3.2.4}$$

$$\boldsymbol{\lambda}_y(l) = \left[\lambda_y(l,1), \dots, \lambda_y(l,N)\right]^T \tag{3.2.5}$$

$$Y(l,k) = X(l,k) + U(l,k) \xrightarrow[periodogram]{} \boldsymbol{\lambda}_y(l) = \boldsymbol{\lambda}_x(l) + \boldsymbol{\lambda}_u(l) \tag{3.2.6}$$

圖 3.2 可以看到未經處理的訊號頻帶能量集中在低頻，從各個頻帶能量差異下不易區分暫態噪音與一般聲音，因此需要進一步的處理讓它們之間的差異更大，使得 TNAD 的結果更為準確。

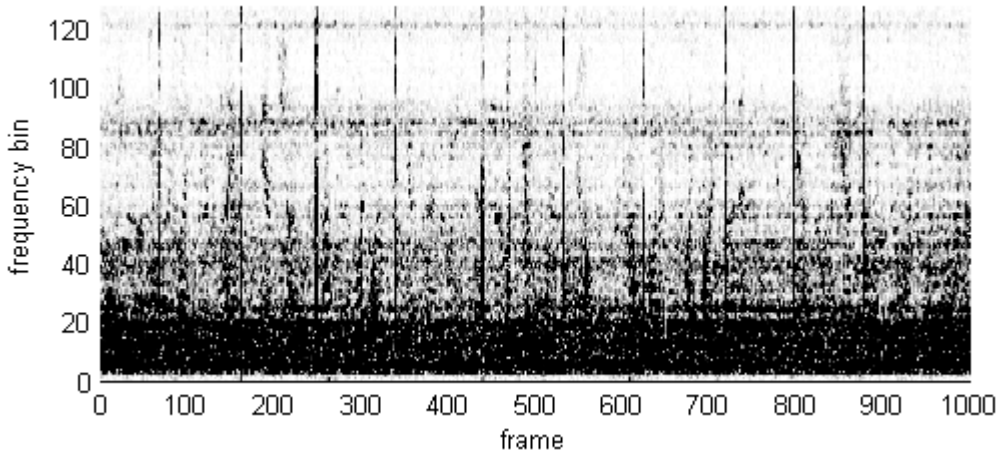

圖 3.2 未經任何處理的原始訊號頻譜

暫態噪音在極短的時間內訊號振幅變化很大，與鄰近取樣點的震幅相關性較低。一般情況下非暫態噪音或語音在極短的時間內訊號振幅變化較小，與鄰近取樣點的震幅相關性較高。利用這樣性質差異，可以計算線性預估係數(linear prediction coefficients)並建立自回歸模型(autoregressive model)估算原本的訊號：

$$\hat{y}_l(\mathrm{n}) = \sum_{r=1}^{N} a_r^l \, y_l(n-r) \qquad (3.2.7)$$

非暫態的聲音得到的線性預估係數可以在誤差較小的情況下估算原本的訊號，而暫態噪音得到的線性預估係數無法準確的估算原本的訊號，預估誤差會很大。因此以估算的誤差作為處理後的訊號，這樣的處理稱它為 ”Whiten” 或是 ”decorrelate” ，如圖 3.3，可以壓抑訊號中一般聲音的部分，並將暫態噪音的性質保留下來：

$$\tilde{y}_l(n) = y_l(n) - \hat{y}_l(\mathrm{n}) \qquad (3.2.8)$$

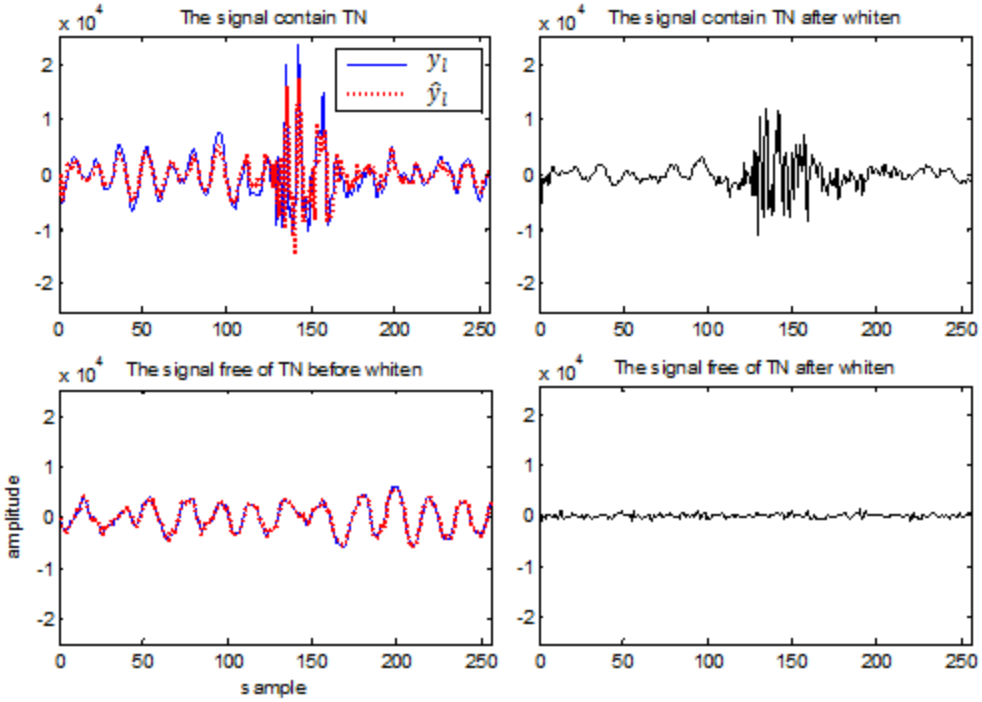

圖 3.3 暫態噪音和一般聲音在 whiten 後的結果

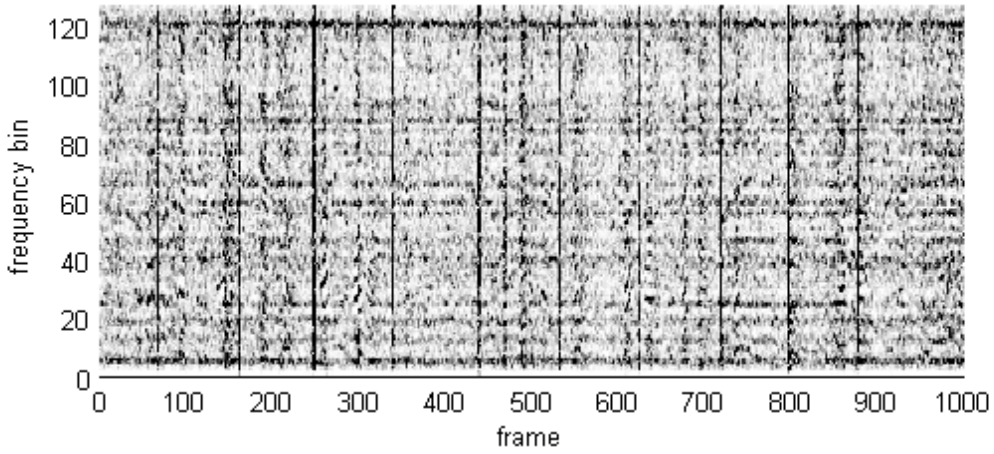

圖 3.4 以預估誤差作為處理後的訊號頻譜

利用經過 whiten 後存在暫態噪音的訊號與一般聲音的訊號在頻域和時域上的差異，可以再做進一步的處理抑制訊號中非暫態噪音的成分。

在頻域的部分，如圖 3.4，因為暫態噪音是脈衝訊號，因此在各頻帶上的能量都很強，而且分配較為平均。而一般聲音相較於暫態噪音在各頻帶上能量較弱，偏重於某些頻帶。

利用頻譜刪減(Spectral subtraction)的方法，估計一般聲音的頻譜，再從接收到的訊號頻譜上將一般聲音的頻譜刪減掉，以抑制一般聲音在訊號頻譜上的成分。

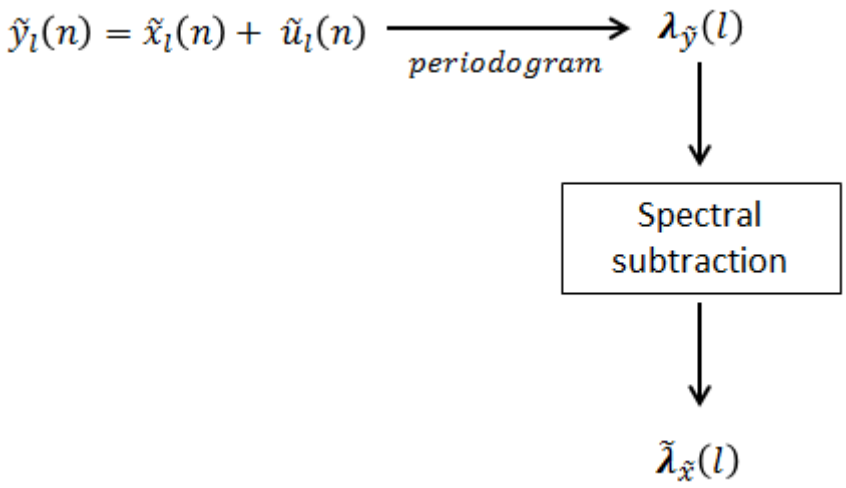

圖 3.5 頻譜刪減法流程圖

圖 3.4 為頻譜刪減法流程圖，$\tilde{y}_l(n)$是麥克風所接收到經過 whiten 之後第 l 個音框內的訊號，由暫態噪音 $\tilde{x}_l(n)$ 和一般聲音 $\tilde{u}_l(n)$ 所組成，透過短時距傅利葉轉換後用週期圖法計算頻帶能量 $\boldsymbol{\lambda}_{\tilde{y}}(l)$，再利用頻譜刪減法抑制一般聲音的在頻譜上的能量，估算只包含暫態噪音的訊號頻譜能量 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$ 。

頻譜刪減法的公式如下：

$$\bar{\boldsymbol{\lambda}}_{\tilde{u}} = E\{\boldsymbol{\lambda}_{\tilde{u}}(h)\} \tag{3.2.9}$$

$$\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l) = \begin{cases} \boldsymbol{\lambda}_{\tilde{y}}(l) - \beta\,\bar{\boldsymbol{\lambda}}_{u}, & \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l) > \gamma\bar{\boldsymbol{\lambda}}_{\tilde{u}} \\ \gamma\bar{\boldsymbol{\lambda}}_{\tilde{u}}, & \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l) \le \gamma\bar{\boldsymbol{\lambda}}_{\tilde{u}} \end{cases}, \quad 1 < \beta, \; 0 < \gamma < 1 \tag{3.2.10}$$

在(3.2.9)式中，h 是不包含暫態噪音的一般聲音的音框索引參數，計算這些音框的頻帶能量平均值以估算一般聲音的頻帶能量 $\bar{\boldsymbol{\lambda}}_{\tilde{u}}$。再代入(3.2.10)式，將從訊號頻譜上減去 $\beta\bar{\boldsymbol{\lambda}}_{u}$ 得到估算的暫態噪音頻譜 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$，其中 β 是一個過估計系數(overestimation factor)，它是依賴暫態噪音與一般聲音的能量大小進行定義， γ 是一個頻譜下限系數。

圖 3.6 是經過頻譜刪減法後的結果，對於穩態(stationary)或是非穩態(non-stationary)噪音都有一定的抑制效果，但是對於頻帶變化較大的語音則無法估計適當的頻帶，因此幾乎無法將語音的部分消除。

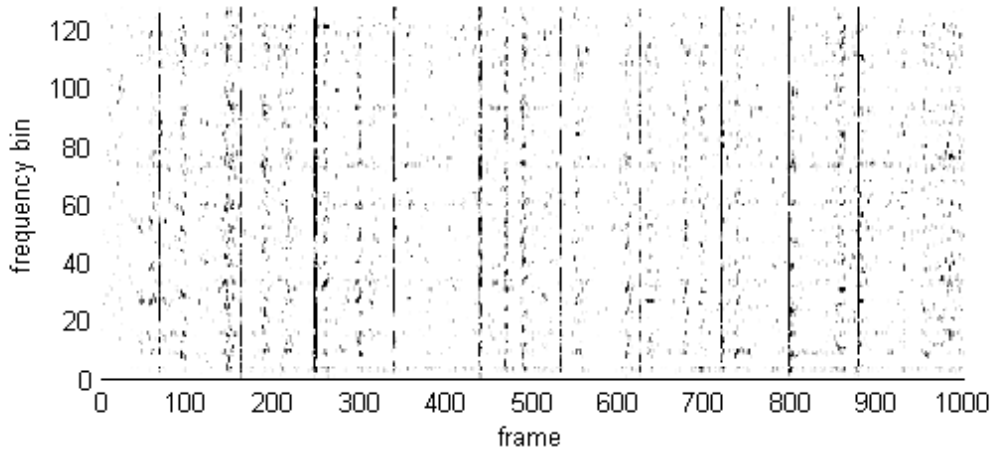

圖 3.6 whiten 後的訊號經過頻譜刪減法處理後的訊號頻譜

觀察 whiten 後的時域訊號，因為保留暫態噪音的特性，因此在存在暫態噪音音框內的訊號，短時間內振幅變化很大。而一般聲音的振幅較小變化也較緩和。這邊嘗試從每一個音框中計算一般聲音的平均振幅，再從接收到的時域訊號中將一般聲音的訊號刪減掉。

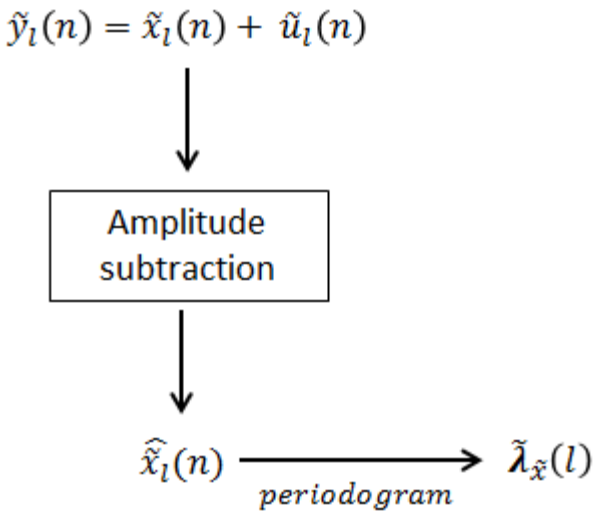

圖 3.7 時域振幅刪減法流程圖

圖 3.7 為時域振幅刪減法流程圖。$\tilde{y}_l(n)$ 是麥克風所接收到第 l 個音框經過 whiten 之後的訊號，由暫態噪音 $\tilde{x}_l(n)$ 和一般聲音 $\tilde{u}_l(n)$ 所組成，經過時域振幅刪減法消除一般聲音後估算的暫態噪音 $\hat{\tilde{x}}_l(n)$ ，再計算估算的暫態噪音頻譜能量 $\tilde{\lambda}_{\tilde{x}}(l)$ 。

時域振幅刪減法的公式如下：

$$\mu_l = \frac{1}{N}\sum_{n=(l-1)N+1}^{lN}|\tilde{y}_l(n)| \tag{3.2.11}$$

$$\tilde{x}_l(n) = sign\big(\tilde{y}_l(n)\big)\left\{\max(\,|\tilde{y}_l(n)| - \alpha\,\mu_l,\,0\,)\right\}, \quad \alpha > 1 \tag{3.2.12}$$

頻譜刪減法和時域振幅刪減法不同的是，前者必須選取沒有暫態噪音的音框計算一般聲音的頻帶能量以消除一般聲音，而後者不需要對於音框的選擇作事先的篩選。在(3.2.11)式中， μ_l 是對於每一個音框去計算音框內訊號的平均振幅，再代入(3.2.12)，不論訊號的正負，振幅均減去 $\alpha\,\mu_l$ ，最多到 0，可以得到估算的暫態噪音時域訊號 $\tilde{x}_l(n)$ ，其中 α 是一個過估計係數，第四章的實驗會討論它在不同的環境干擾聲源下最適合的值。

圖 3.8 是經過域振幅刪減法後的結果。在存在暫態噪音的音框中，它可以在保留暫態噪音的特性下，對 whiten 後的訊號作進一步的處理。而不存在暫態噪音的音框的訊號幾乎可以被消除掉。

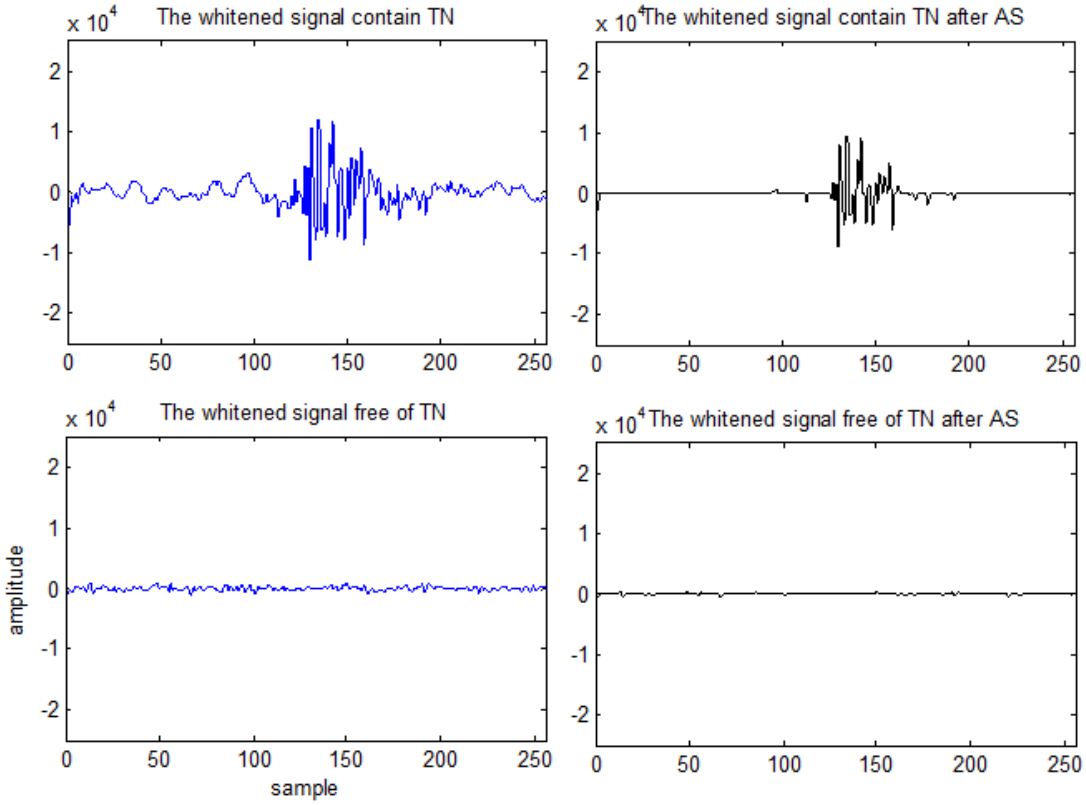

圖 3.8 暫態噪音和一般聲音經過時域振幅刪減法後的結果

最後再計算 $\tilde{x}_l(n)$ 的頻譜能量 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$，如圖 3.8。從頻域上也可以看到，只有存在暫態噪音音框的頻帶有能量，其它不存在暫態噪音的音框的頻帶能量會被抑制，如圖 3.9。

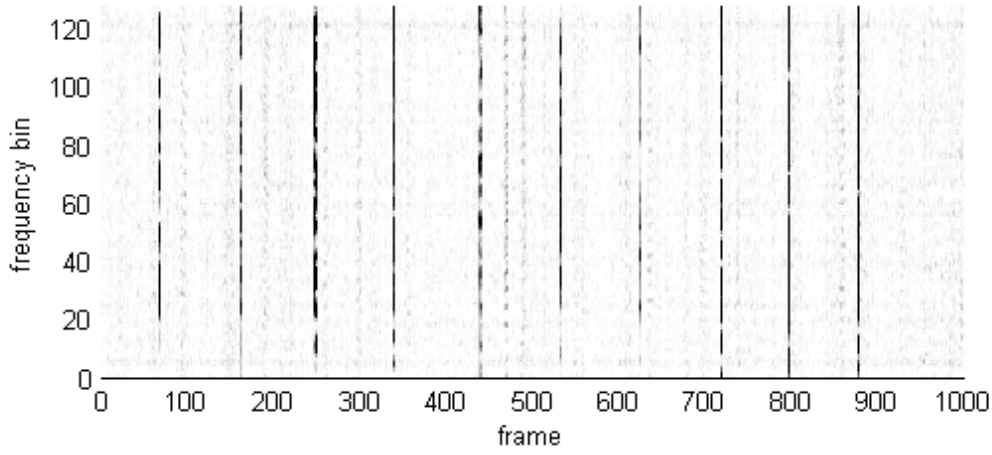

圖 3.9 whiten 後的訊號經過時域振幅刪減法後的訊號頻譜

接下來根據暫態噪音和一般聲音的頻帶能量分佈差異，利用 NL diffusion filter 將它們所屬的音框分離出來。

根據 diffusion framework，定義一個描述 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$ 和 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)$ 之間關係的核心方程式：

$$k(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)) = \exp\left\{-\frac{\|\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)-\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\|^2}{2\sigma^2}\right\} \quad (3.2.13)$$

在(3.2.13)式中，$\left\|\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)-\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right\|^2$ 代表 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$ 和 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)$ 頻帶間的距離，σ 是控制核心方程式的大小的參數，依照頻帶點間的情況決定。

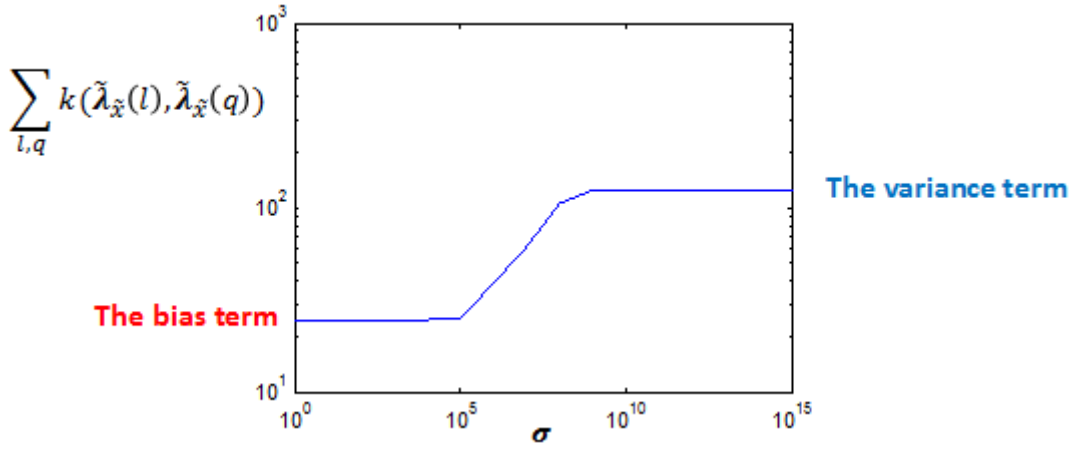

圖 3.10 不同核心參數下的核心方程式總和

當預估核心參數遠小於適當的值時，會趨近於偏差項(bias term)：

$\sigma \to 0$, we have $k\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right) \to \delta_{l,q}$ (3.2.14)

$\sum_{l,q} k(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)) \to M$ (3.2.15)

當預估核心參數遠大於適當的值時，會趨近於變異項(variance term)：

$\sigma \to \infty$, we have $\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right) \to 1$ (3.2.16)

$\sum_{l,q} k(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)) \to M$ (3.2.17)

方程式處於這兩種情況下時都會得到不正確的結果。當核心參數在中間的線性區，偏差項和變異項的影響都會最小，因此最適當的核心參數要選取在線性區內的數值。

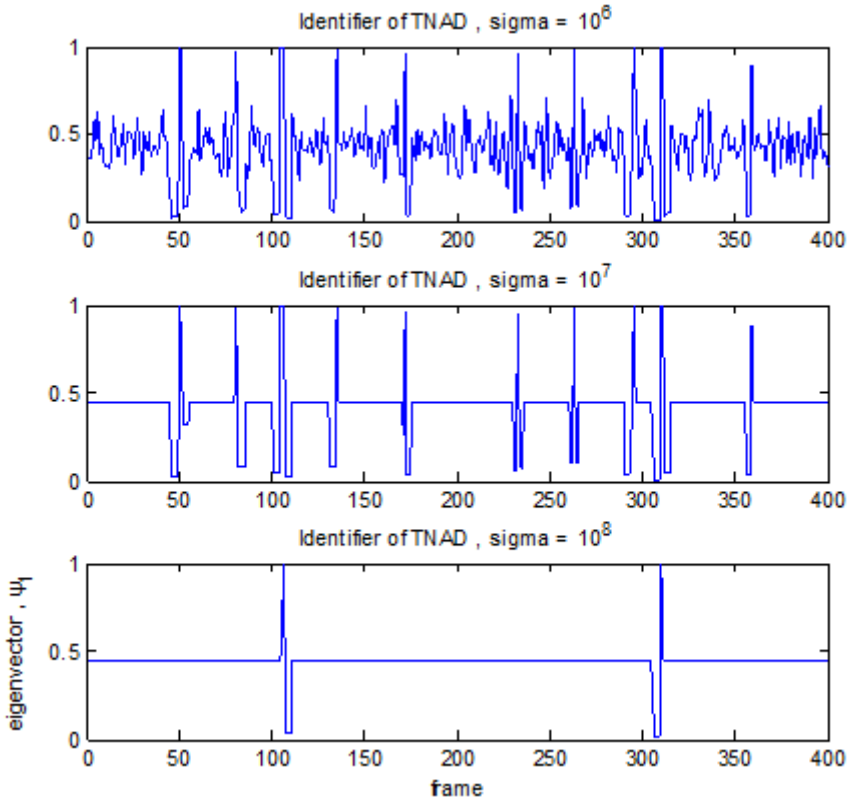

圖 3.11 不同核心參數下 TNAD 的結果

圖 3.11 是表示資料內存在 10 組暫態噪音時，在不同核心參數下，TNAD 所得到的結果。可以得知最適當的核心參數是 10^7 ，選取太小的核心參數會使得 non-local diffusion filter 的效果很差，導致 TNAD 的結果容易受到干擾，選取過大的核心參數則會讓能量較小的暫態噪音被 TNAD 歸類為是一般聲音而在最後的結果被壓抑掉。

將描述 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$ 和 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)$ 之間關係的核心方程式正規化後，建立馬可夫鏈轉移機率方程式：

$$\begin{cases} d(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)) = \sum_{j=1}^{M} k\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(j)\right) \\ p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right) = \frac{k\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right)}{d(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l))} \end{cases} \tag{3.2.18}$$

$$\sum_{j=1}^{M} p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(j)\right) = 1 \tag{3.2.19}$$

(3.2.18)式中 $p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right)$ 表示一階隨機行走從 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l)$ 到 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)$ 的轉移機率方程式。將 $p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right)$ 寫成矩陣的型式 **P** ：

$$\mathbf{P} = \begin{bmatrix} p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1)\right) & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2)\right) & \cdots & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M)\right) \\ p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1)\right) & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2)\right) & \cdots & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M)\right) \\ \vdots & \vdots & \ddots & \vdots \\ p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1)\right) & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2)\right) & \cdots & p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M)\right) \end{bmatrix} \tag{3.2.20}$$

將轉移機率矩陣 **P** 進行特徵值分解：

$$\mathbf{P} = \boldsymbol{\Psi}\boldsymbol{D}\boldsymbol{\Psi}^T = \begin{bmatrix} \boldsymbol{\psi}_1 & \cdots & \boldsymbol{\psi}_M \end{bmatrix} \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_1{}^T \\ \vdots \\ \boldsymbol{\psi}_M{}^T \end{bmatrix}$$

$$= \boldsymbol{\psi}_1 \rho_1 \boldsymbol{\psi}_1{}^T + \cdots + \boldsymbol{\psi}_M \rho_M \boldsymbol{\psi}_M{}^T \tag{3.2.21}$$

可得到特徵值 ρ_j 以及特徵向量 $\boldsymbol{\psi}_j$ ：

$$1 = \rho_1 \geq \rho_2 \geq \cdots \geq \rho_M > 0 \tag{3.2.22}$$

$$\boldsymbol{\psi}_j = \left[\psi_j(1)\ \psi_j(2)\ \cdots \psi_j(M)\right]^T \tag{3.2.23}$$

轉移機率方程式 $p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(j)\right)$ 可寫成以下形式：

$$\begin{aligned} p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(q)\right) &= \psi_1(l)\rho_1\psi_1(q) + \cdots + \psi_M(l)\rho_M\psi_M(q) \\ &= \sum_{j=1}^{M} \rho_j \psi_j(l)\psi_j(q) \end{aligned} \tag{3.2.24}$$

$\tilde{\boldsymbol{\Lambda}}$ 是 $\tilde{\boldsymbol{\lambda}}_{\tilde{x}}$ 組成的音框頻帶能量矩陣：

$$\begin{aligned} \tilde{\boldsymbol{\Lambda}} &= \left[\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2), \ldots, \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M)\right]^{\boldsymbol{T}} \\ &= \begin{bmatrix} \tilde{\lambda}_{\tilde{x}}(1,1) & \tilde{\lambda}_{\tilde{x}}(1,2) & \cdots & \tilde{\lambda}_{\tilde{x}}(1,N) \\ \tilde{\lambda}_{\tilde{x}}(2,1) & \tilde{\lambda}_{\tilde{x}}(2,2) & \cdots & \tilde{\lambda}_{\tilde{x}}(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\lambda}_{\tilde{x}}(M,1) & \tilde{\lambda}_{\tilde{x}}(M,2) & \cdots & \tilde{\lambda}_{\tilde{x}}(M,N) \end{bmatrix} \end{aligned} \tag{3.2.25}$$

其中 M 是音框個數，N 是頻帶個數， $\tilde{\boldsymbol{\Lambda}}$ 的維度為 M × N，音框個數 M 遠小於頻帶個數 N。為了保留原始資料的資訊並縮減音框頻帶能量矩陣不必要的維度，利用轉置的方法將 $\tilde{\boldsymbol{\Lambda}}$ 映射到較低維度的空間，如式(3.2.26)

$$\mathbf{A} = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^{\boldsymbol{T}}\ ,\quad \mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,M} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,M} \end{bmatrix} \tag{3.2.26}$$

將轉移機率矩陣 **P** 乘上縮減維度的音框頻帶能量矩陣 **A** 將不同音框的頻帶差異分群，**PA** 的結果可以解釋為估測的暫態噪音頻帶以及估測的一般聲音頻帶， $[\mathbf{PA}]_{l,q}$ 是 **PA** 第 l 列第 q 行的元素：

$$[\mathbf{PA}]_{l,q} = p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(1)\right)\alpha_{1,\,q} + p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(2)\right)\alpha_{2,\,q} + \cdots +$$

$$p\left(\tilde{\boldsymbol{\lambda}}_{\tilde{x}}(l), \tilde{\boldsymbol{\lambda}}_{\tilde{x}}(M)\right)\alpha_{M,\,q} = \sum_{j=1}^{M}\rho_j b_{j,q}\psi_j(l) \quad (3.2.27)$$

$$b_{j,q} = \psi_j(1)\alpha_{1,\,q} + \psi_j(2)\alpha_{2,\,q} + \cdots + \psi_j(M)\alpha_{M,\,q} = \langle \alpha_{(\cdot,\,q)}, \boldsymbol{\psi}_j \rangle \quad (3.2.28)$$

從 (3.2.27) 可以得知，第一特徵向量中第 l 個元素的 $\psi_1(l)$ 可以作為暫態噪音出現與否的指標。

為了讓分群的效果更顯著，可以增加隨機行走的次數：

$$[\mathbf{P}^t\mathbf{A}]_{l,q} = \sum_{j=1}^{M}\rho_j{}^t b_{j,q}\psi_j(l) \quad (3.2.29)$$

但是次數趨近於無限大時，會讓原本的資料趨近到一個定值：

$$\rho_j{}^t \xrightarrow{t\to\infty} 0\ ,\ \forall\rho_j < 1 \quad (3.2.30)$$

$$[\mathbf{P}^t\mathbf{A}]_{l,q} \xrightarrow{t\to\infty} b_{j,q} \quad (3.2.31)$$

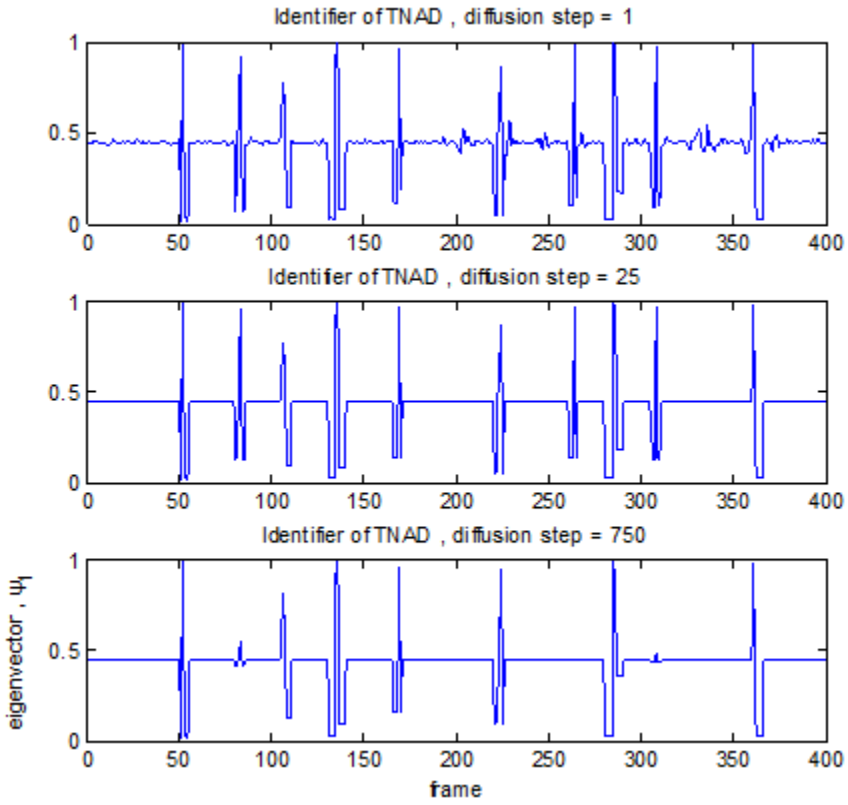

圖 3.12 不同的隨機行走的次數下 TNAD 的結果

3.3 暫態噪音聲源方位估測

麥克風接收到的訊號經過 TNAD 後，可以得到估測存在暫態噪音的音框索引參數。但是不同麥克風陣列的形狀配置方式，會影響接收到的訊號大小，進而影響 TNAD 的結果。以環形麥克風 8 顆麥克風為例，圖 3.13 為第 1、3、5、7 顆麥克風接收到的訊號有 5 組暫態噪音在- 90 度的方向經過 TNAD 的結果，面對聲源的麥克風可以收到振幅完整的聲源訊號，因此可以偵測到所有的暫態噪音。某些能量較小的暫態噪音衰減較為嚴重，所以與聲源反方向的麥克風接收到的暫態噪音較微弱，造成 TNAD 不準確。

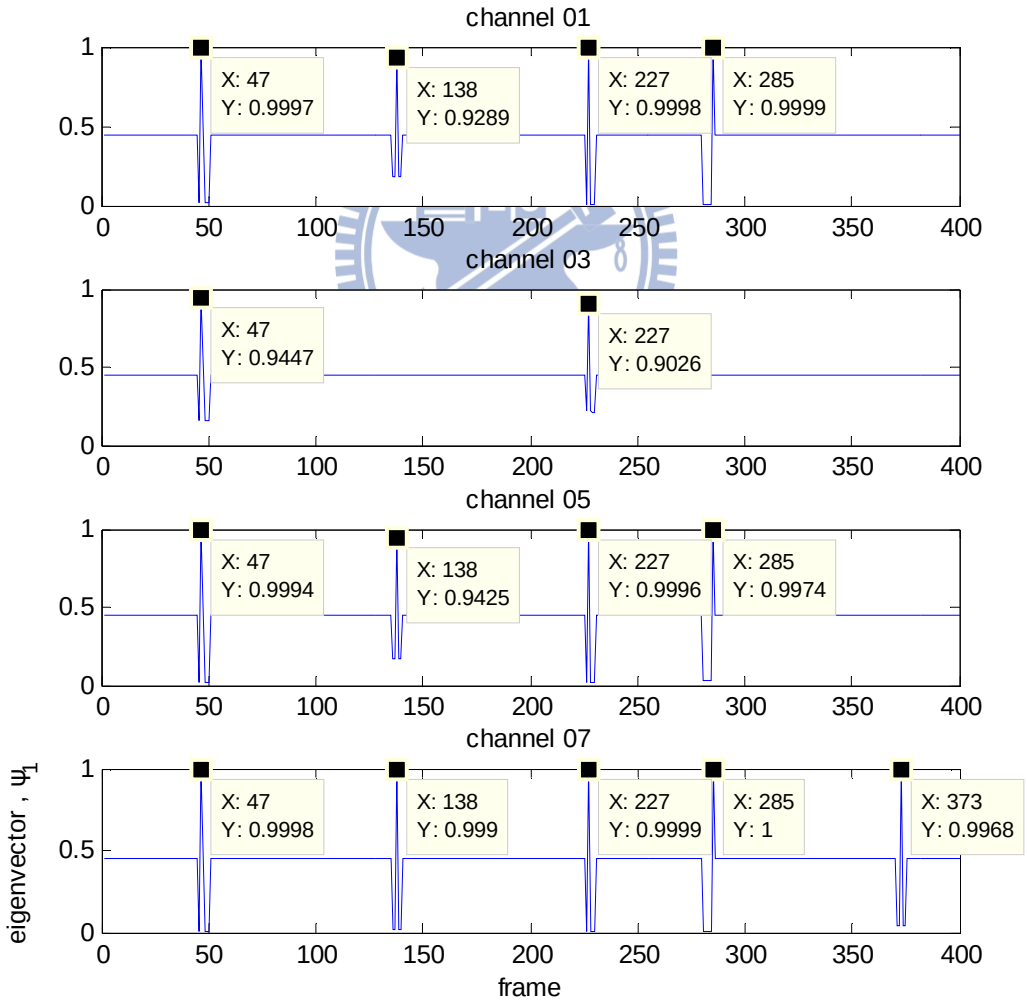

圖 3.13 陣列中不同麥克風 TNAD 的結果

線性麥克風陣列中所有麥克風都是面對聲源，因此不同麥克風接收的振幅差異不同影響較輕微。但是不論麥克風的結構為何，我們都無法預知哪顆麥克風可以接收到最完整的暫態噪音，所以必須統計所有麥克風經過 TNAD 得到的結果。在圖 3.13 中，暫態噪音活動偵測的結果是在第 47、138、227、285、373 音框存在暫態噪音，因此我可以只針對這 5 個音框做聲源方位估測。

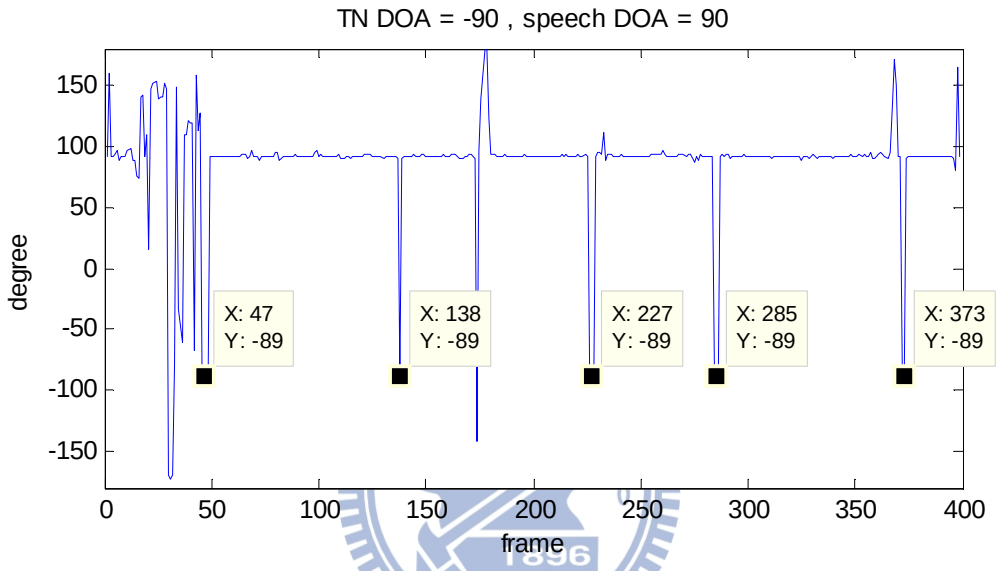

圖 3.14 MUSIC DOA spectrum

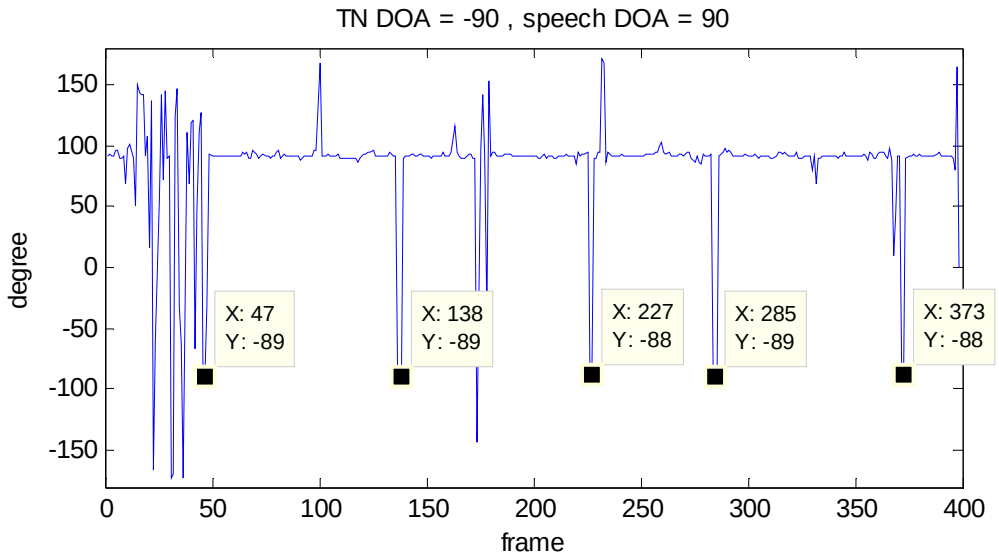

圖 3.15 SBF DOA spectrum

圖 3.13 和圖 3.14 是兩種不同聲源方位估測演算法對於一段包含 5 組聲源方向為 -90 度的暫態噪音，以及聲源方向為 90 度的語音的資料估測的結果。我們可以看到存在暫態噪音的音框內所估測到的聲源方位就是暫態噪音的聲源方位。因為暫態噪音在極短的時間內的能量相較於其它種類聲源要大，在一般的情況下，存在暫態噪音的音框內的主要訊號成分就是暫態噪音，所以對於選定的音框做聲源估測就可以得到我們要的結果。

第四章 實驗結果與分析

本章節將介紹本論文的方法對於 TNAD 效能提升程度的實驗結果，以及利用 TNAD 所得到的結果，進行暫態噪音聲源方位估測。

本論文利用具有八顆麥克風的環型數位麥克風陣列作為訊號接收平台，圖 4.1 為錄音環境的實際照片，暫態噪音是由受試者在麥克風陣列平台的不同方位擊掌產生，並於麥克風陣列平台的不同方位(0 度、90 度、180 度、270 度)距離 1M 的位置錄音，如圖 4.2。

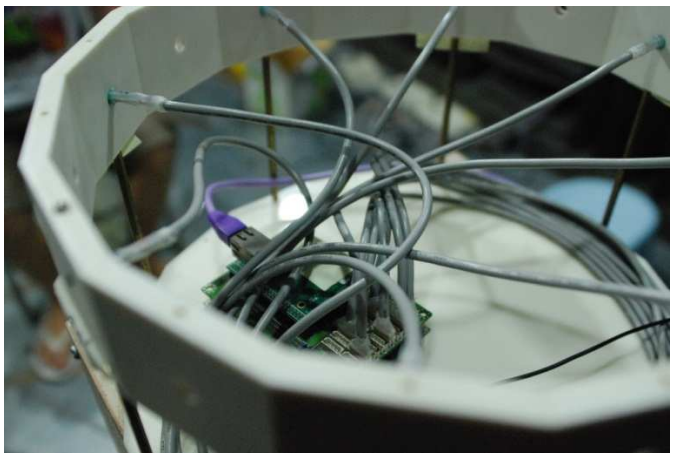

圖 4.1 環形數位麥克風陣列平台

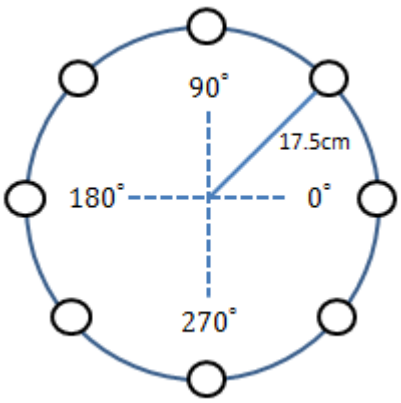

圖 4.2 環形麥克風陣列平台的平面圖

實驗錄音與訊號處理的規格如表 4.1。

陣列架構	環形麥克風陣列
陣列半徑	17.5 cm
麥克風個數	8
取樣頻率	8 kHz
FFT size	256 sample
shift size	128 sample
Block size	5 frame
Block overlap size	0

表 4.1 平台錄音與訊號處理的詳細數據

實驗使用兩種不同性質無方向性的非暫態噪音以及特定方向的語音做為干擾聲源，測試 TNAD 以及暫態噪音聲源方位估測的穩健度：第一種非暫態噪音是穩態(stationary)的 F16 noise；第二種非暫態噪音是非穩態(non-stationary)的，這裡使用 Babble noise。這兩種非噪音的錄製方法：利用喇叭播放並將喇叭放置於錄音環境中的對稱四個角落，用以製造 Diffusion Noise。語音的錄製是使用人工嘴在陣列不同方位(0 度、90 度、180 度、270 度)距離 1M 進行語音的播放。

4.1 暫態噪音活動偵測實驗結果與分析

在本實驗中，利用環境存在不同種類干擾聲源測試暫態噪音活動偵測(TNAD)演算法效能，並比較使用 whiten 後的訊號經過頻譜刪減法(SS)或時域振幅刪減法(AS)處理過後對於提升準確度的表現。

暫態噪音與干擾聲源的比值為 TSR(Transient-to-signal ratio)：

$$TSR = 10log_{10}\frac{\mathbb{E}\{x_q^{\,2}(n)\}}{\mathbb{E}\{y_l^{\,2}(n)\}} \tag{4.1.1}$$

在實驗中使用 TSR 表示暫態噪音訊號相對於干擾聲源大小的標準，在(4.1.1)中，$x_q(n)$為包含乾淨暫態噪音第 q 個音框的訊號，$y_l(n)$為乾淨干擾聲源第 l 個音框的訊號。

實驗測試音檔包含 100 次不同振幅大小的暫態噪音，音檔長度為 1000 個音框，音框長度為 256，每次 TNAD 處理一個 block，block 長度為 5 個音框，資料中總共有 200 個 block，存在 100 個包含暫態噪音的 block。

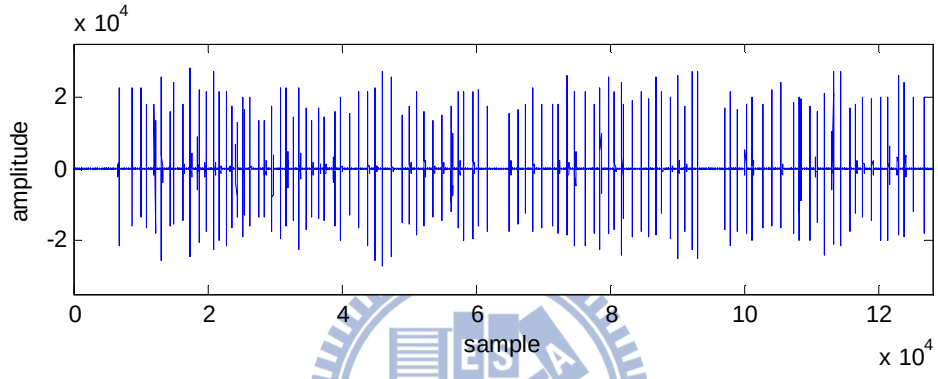

圖 4.3 包含 100 次不同震幅大小暫態噪音的實驗音檔

有三種 TSR 平均為 0 的不同干擾聲源，分別是 F16 noise、Babble noise 以及語音。語音在不同時間 TSR 的變化程度較大，如圖 4.4。

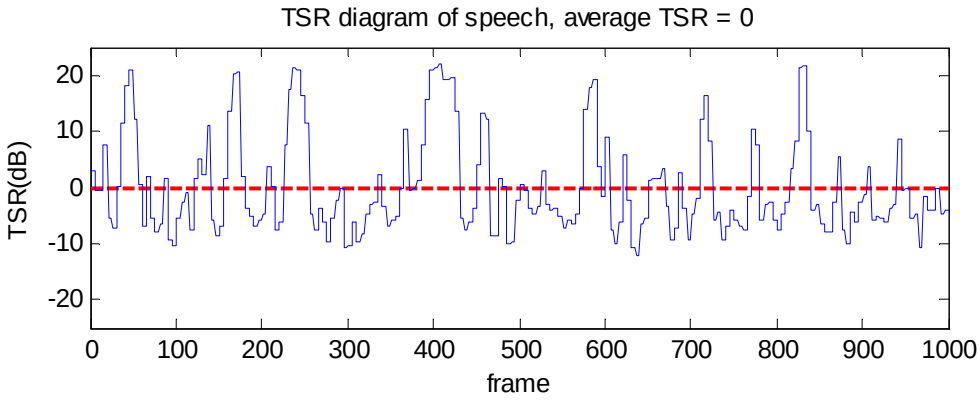

圖 4.4 干擾聲源為語音時 TSR 的變化情形

將暫態噪音音檔和相同長度的干擾聲源音檔混音成以下三種音檔：

測試音檔 1：　暫態噪音(擊掌) ＋ 非暫態穩態噪音(F16)

測試音檔 2：　暫態噪音(擊掌) ＋ 非暫態非穩態噪音(Babble)

測試音檔 3：　暫態噪音(擊掌) ＋ 語音

在所有的實驗中 TNAD 的門檻值固定在 0.45，核心參數 σ 固定為 10^7，隨機行走階數 t 固定為 25。選擇 block 長度為 5 以減少運算量。

對於 TNAD 偵測暫態噪音的準確率，使用以下實驗指標參數：

	Transient noise	Other case
Decided as Transient noise	True Positive	False Positive
Decided as other case	False Negative	True Negative

表 4.2 暫態噪音活動偵測實驗指標參數定義表

Detection rate：True Positive / (True Positive + False Negative)

在給定存在暫態噪音的 block 下，被判斷為存在暫態噪音的機率。

False alarm rate：False Positive / (False Positive + True Negative)

在不存在暫態噪音的 block 下，誤判為存在暫態噪音的機率。

圖 4.5 為只使用 whiten 處理後的訊號，經過 TNAD 後，在 detection rate 為 100%的情況下，不同的 TSR 中存在 false alarm 的機率。當 TSR 很低的時候，穩態噪音和非穩態噪音會幾乎完全的蓋過暫態噪音，存在暫態噪音的音框和不存在暫態噪音的音框經過 whiten 之後，差異不大，因此會有非常高的 false alarm rate。而語音由於有能量較小的部分，因此在 TSR 很低的時候不會完全蓋過暫態噪音，所以 false alarm rate 會緩慢上升。

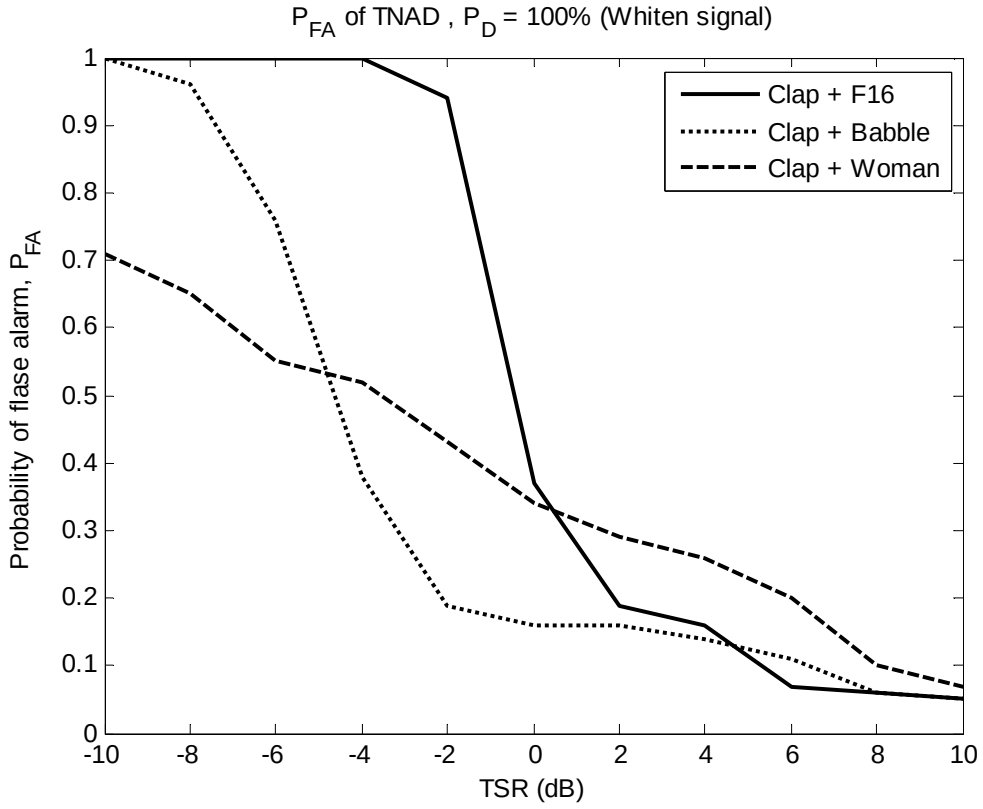

圖 4.5 不同 TSR 下三種干擾聲源經過 whiten 後 TNAD 的結果

以下三小節分別對於不同種類的干擾音源進行個別討論，並比較經過 SS 或 AS 處理過後，對於 TNAD 的提升效果，並找出適當的參數。

4.1.1 干擾聲源為穩態噪音

在這小節分析中，我們針對干擾聲源為非暫態穩態噪音(F16)的情況來進行對於 SS 和 AS 處理過後的訊號對於提升 TNAD 效能的比較，以及找出 detection rate 為 100%時，false alarm rate 最低時的參數。

圖 4.6 和圖 4.7 為在不同的參數 β 下，SS 處理後的訊號經過 TNAD 的結果。為了維持 detection rate 為 100%，在高 TSR 時，調大 β 值可以將 false alarm rate 降低為 0，在低 TSR 時，SS 降低 false alarm rate 的效果有限， β 值增加可以使得 false alarm rate 降低，但是 detection rate 也會同時跟著快速的下降。

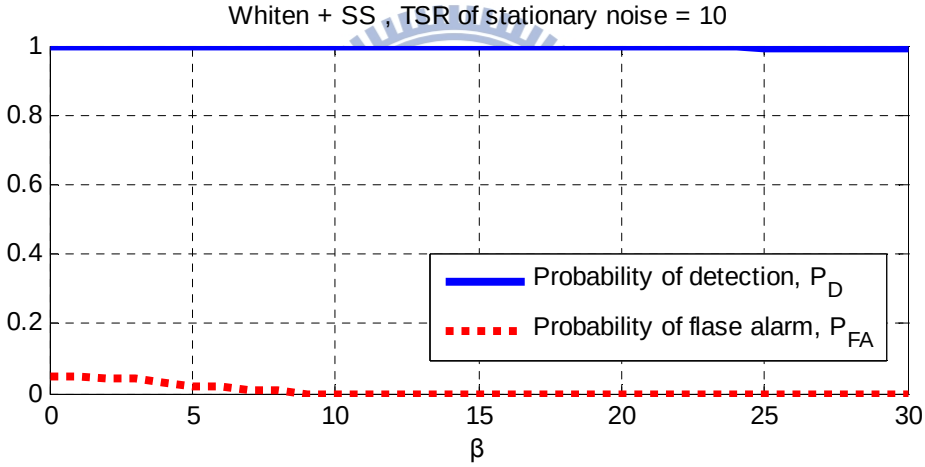

圖 4.6 干擾為 F16 在 TSR=10 的情況下經過 SS 後 TNAD 的結果

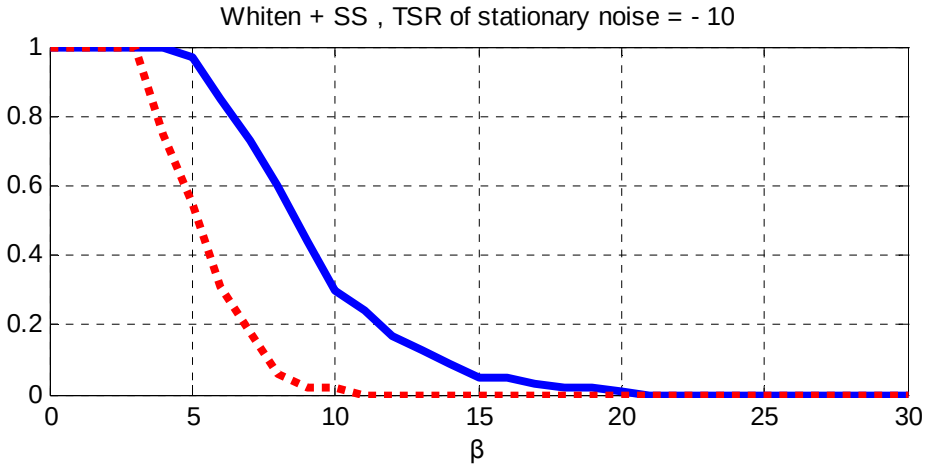

圖 4.7 干擾為 F16 在 TSR=-10 的情況下經過 SS 後 TNAD 的結果

圖 4.8 和圖 4.9 為在不同的參數 α 下，AS 處理後的訊號經過 TNAD 的結果。在維持 detection rate 為 100% 的情況下，高 TSR 時，α 值增加可以將 false alarm rate 降低為 0，低 TSR 時，也可以將 false alarm rate 降低至 0.1 以下，可以選定 α 值讓 false alarm rate 降低為 0，但是 detection rate 也會些許下降。

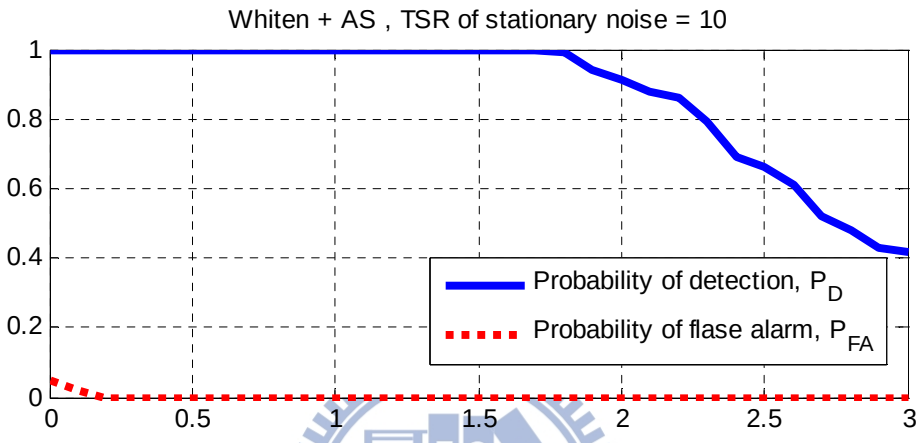

圖 4.8 干擾為 F16 在 TSR = 10 的情況下經過 AS 後 TNAD 的結果

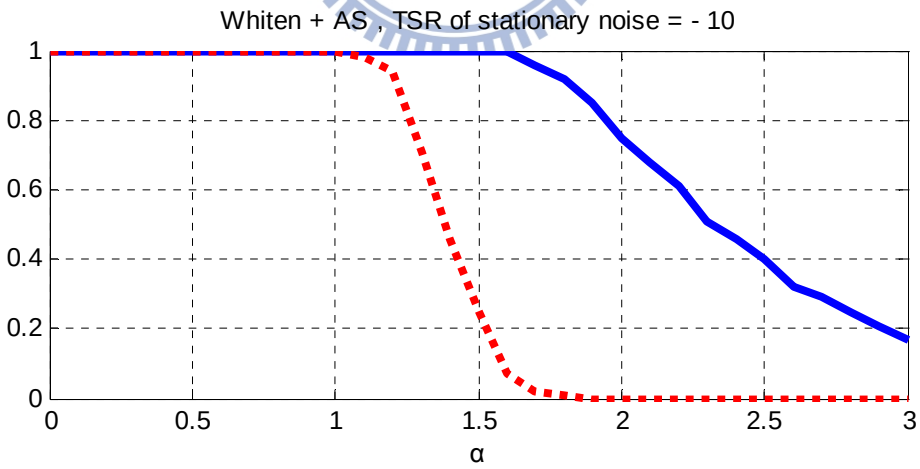

圖 4.9 干擾為 F16 在 TSR = -10 的情況下經過 AS 後 TNAD 的結果

由以上實驗結果得知，隨著 TSR 下降要維持 detection rate 為 100%，適合的 α 值會從 1.7 下降到 1.6，變化不大，而適合的 β 會大幅降低。為了在不同的 TSR 下都維持 100% 的 detection rate，參數 α 必須在 1.6 以下，β 必須在 4 以下。圖 4.10 是選定適合的參數在不同的 TSR 下干擾聲源為 F16 時 whiten 後的訊號經過 SS 或 AS 處理後，TNAD 所得到 detection rate 為 100%時 false alarm rate 的結果。經過 SS 處理，在 TSR 為 0 以上時，可以將 false alarm rate 降低至 0.1 以下。而經過 AS 處理後，在 TSR 為 -6 以上時，可以將 false alarm rate 降低為 0。

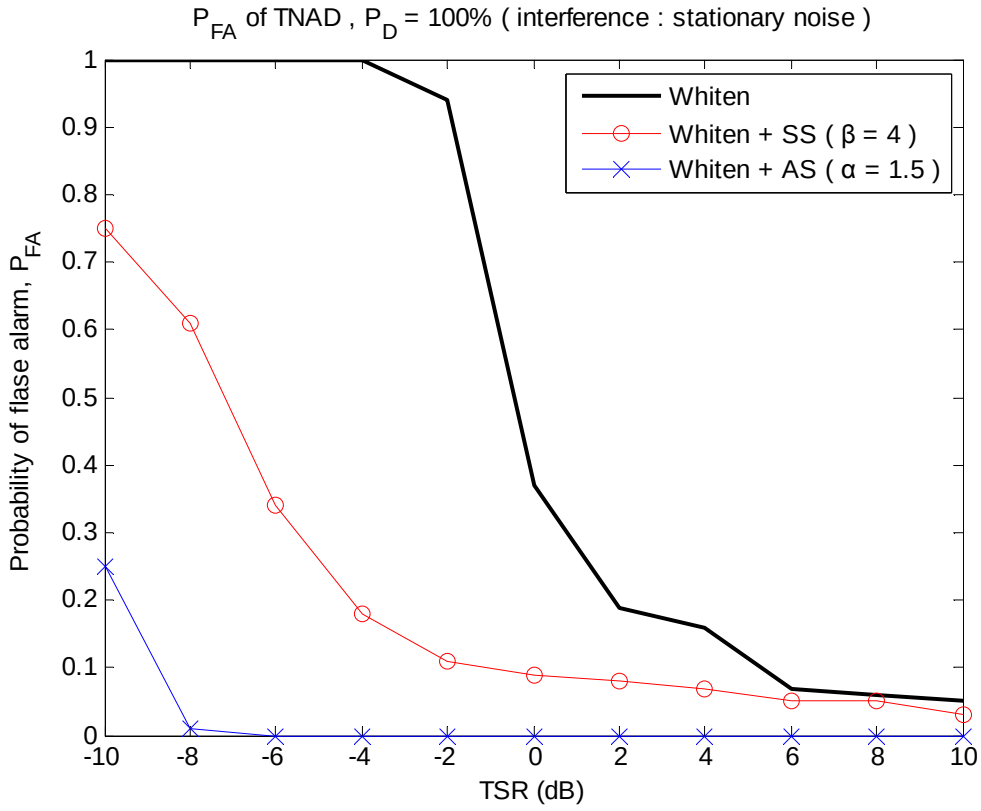

圖 4.10 不同 TSR 下干擾聲源為 F16 經過三種方法處理後 TNAD 的結果

4.1.2 干擾聲源為非穩態噪音

在這小節分析中，我們針對干擾聲源為非暫態非穩態噪音(Babble)的情況來進行對於 SS 和 AS 處理過後的訊號對於提升 TNAD 效能的比較，以及找出 detection rate 為 100%時，false alarm rate 最低時的參數。

圖 4.11 和圖 4.12 為在不同的參數 β 下，SS 處理後的訊號經過 TNAD 的結果。在維持 detection rate 為 100%的情況下，SS 對於非穩態噪音干擾壓抑的效果和對於穩態噪音的結果差不多。在高 TSR 時，調大 β 值可以稍微降低 false alarm rate，在低 TSR 時， β 值增加會使得 false alarm rate 和 detection rate 同時快速的下降。

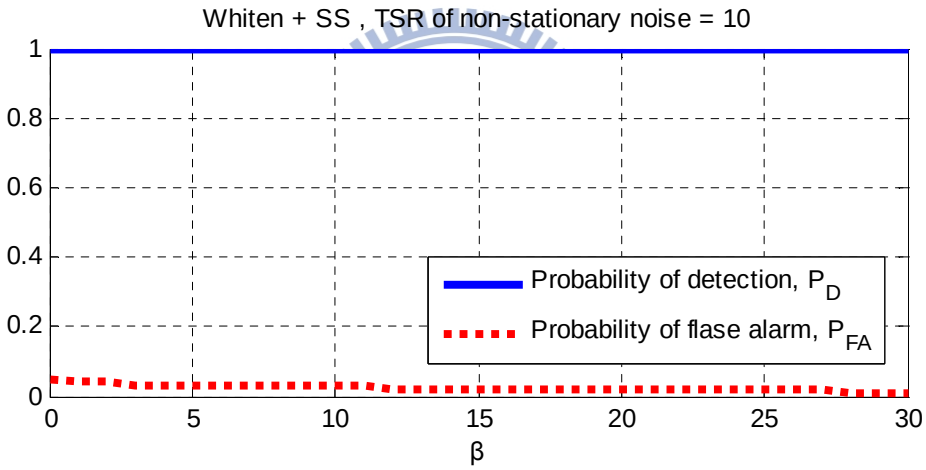

圖 4.11 干擾為 Babble 在 TSR = 10 的情況下經過 SS 後 TNAD 的結果

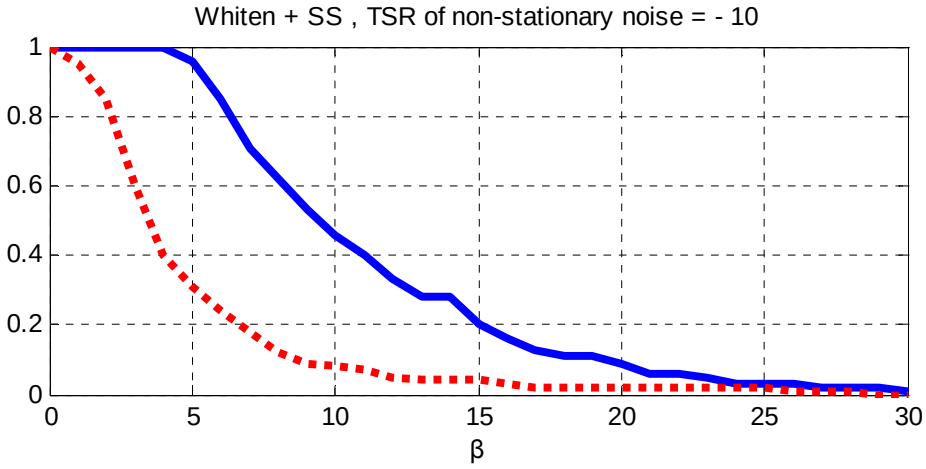

圖 4.12 干擾為 Babble 在 TSR = -10 的情況下經過 SS 後 TNAD 的結果

圖 4.13 和圖 4.14 為在不同的參數 α 下，AS 處理後的訊號經過 TNAD 的結果。在維持 detection rate 為 100% 的情況下，AS 對於非穩態噪音干擾壓抑的效果和對於穩態噪音的結果差不多。高 TSR 時，α 值增加可以將 false alarm rate 降低為 0，低 TSR 時，也可以將 false alarm rate 降低接近 0，可以選定 α 值讓 false alarm rate 為 0，但是 detection rate 也會稍微降低。

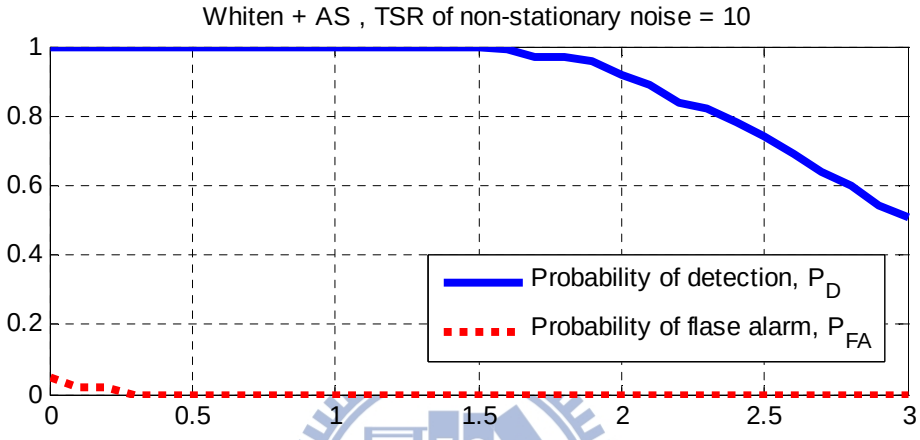

圖 4.13 干擾為 Babble 在 TSR = 10 的情況下經過 AS 後 TNAD 的結果

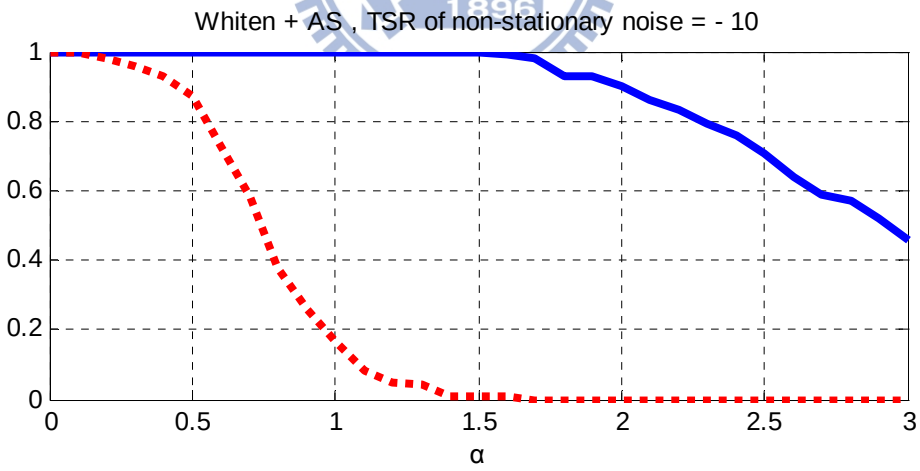

圖 4.14 干擾為 Babble 在 TSR = 10 的情況下經過 AS 後 TNAD 的結果

由以上實驗結果得知，隨著 TSR 下降要維持 detection rate 為 100%，適合的 α 值固定不變，而適合的 β 會大幅降低。為了在不同的 TSR 下都維持 100% 的 detection rate，參數 α 必須在 1.5 以下，β 必須在 4 以下。圖 4.15 是選定適合的參數在不同的 TSR 下干擾聲源為 Babble 時 whiten 後的訊號經過 SS 或 AS 處理後，TNAD 所得到 detection rate 為 100%時 false alarm rate 的結果。經過 SS 處理，在 TSR 為 -2 以上時，可以將 false alarm rate 降低至 0.1 以下。而經過 AS 處理後，在 TSR 為 -8 以上時，可以將 false alarm rate 降低為 0。

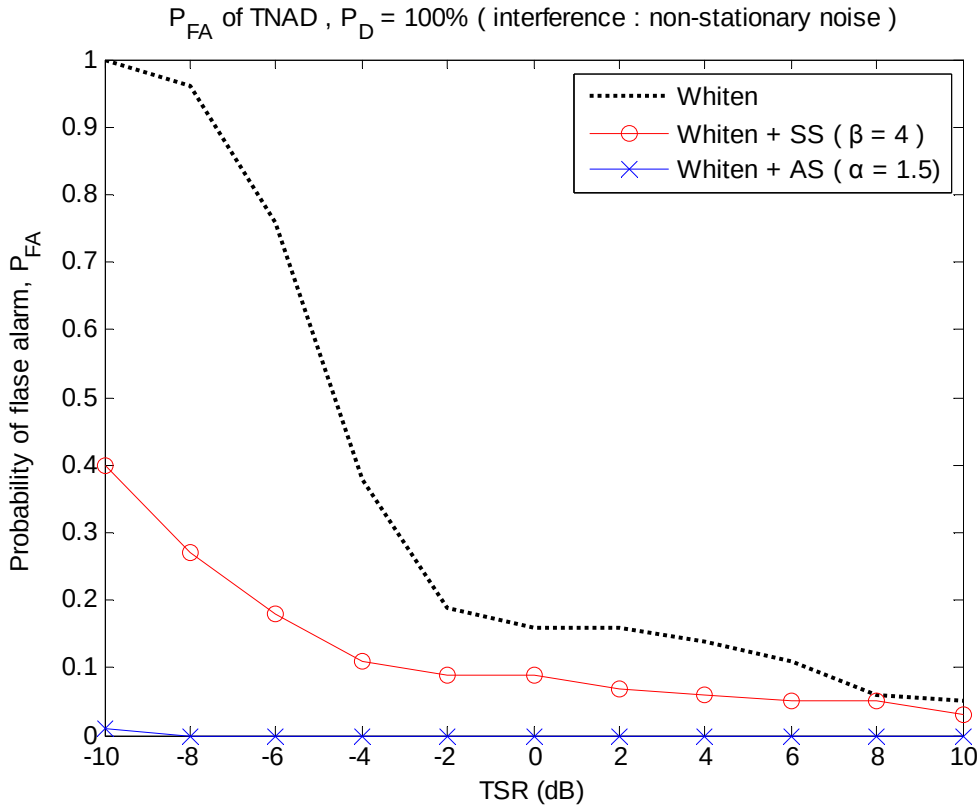

圖 4.15 不同 TSR 下干擾聲源為 Babble 經過三種方法處理後 TNAD 的結果

4.1.3 干擾聲源為語音

在這小節分析中，我們針對干擾聲源為語音的情況來進行對於 SS 和 AS 處理過後的訊號對於提升 TNAD 效能的比較，以及找出 detection rate 為 100%時，false alarm rate 最低時的參數。

圖 4.16 和圖 4.17 為在不同的參數 β 下，SS 處理後的訊號經過 TNAD 的結果。因為不可能估測準確的語音頻帶，因此 SS 無法消除語音，在不同 TSR 下，SS 對於提升 TNAD 準確度效果都不好。如果估測的語音頻帶能量較弱，SS 降低 false alarm rate 的效果有限，如果估測的語音頻帶能量較強，則是 detection rate 和 false alarm rate 都會隨著 β 的增加快速的降低。

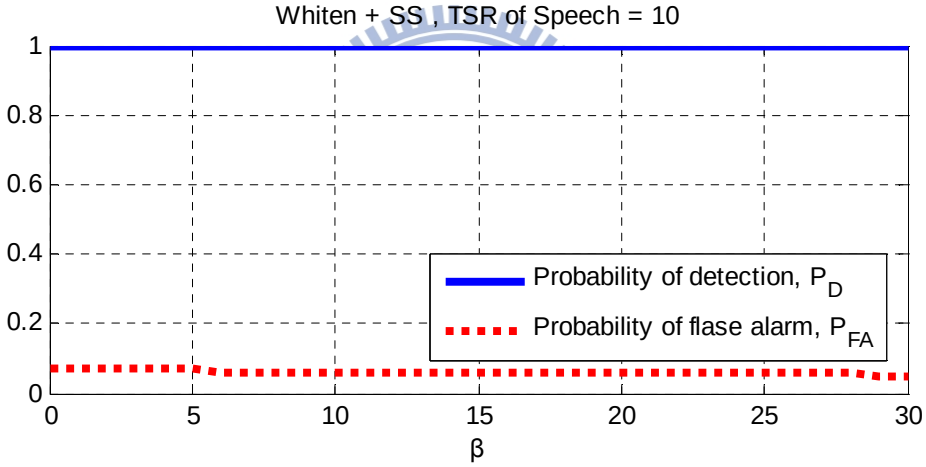

圖 4.16 干擾為 Speech 在 TSR = 10 的情況下經過 SS 後 TNAD 的結果

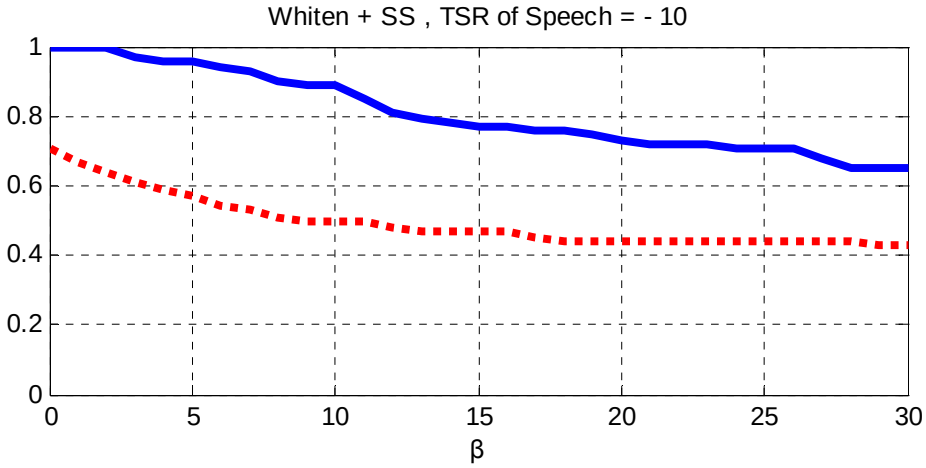

圖 4.17 干擾為 Speech 在 TSR = -10 的情況下經過 SS 後 TNAD 的結果

圖 4.18 和圖 4.19 為在不同的參數 α 下，AS 處理後的訊號經過 TNAD 的結果。因為是從時域上縮減振幅，可以在保留暫態噪音的成分下，將 whiten 後的語音訊號振幅較小的部分完全消除，所以對於降低 false alarm rate 有一定的效果。在維持 detection rate 為 100% 的情況下，高 TSR 時，調整 α 值可以將 false alarm rate 降低為 0，低 TSR 時， α 值增加也可以有效的降低 false alarm rate。

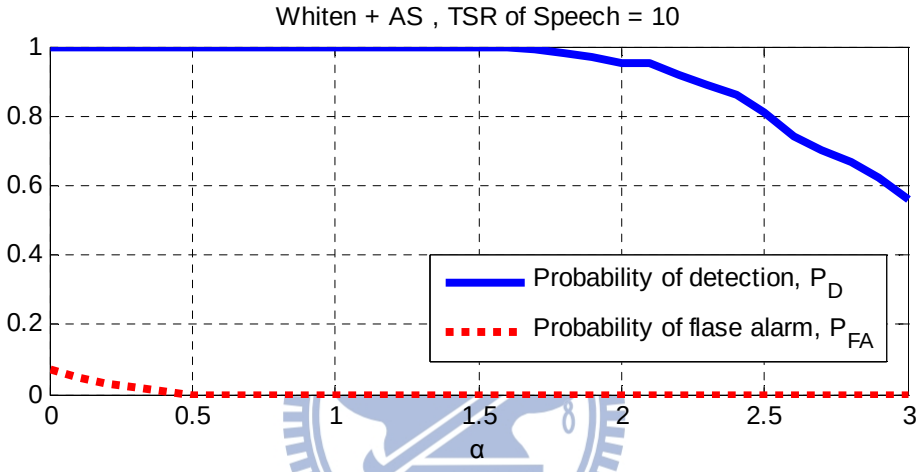

圖 4.18 干擾為 Speech 在 TSR = 10 的情況下經過 SS 後 TNAD 的結果

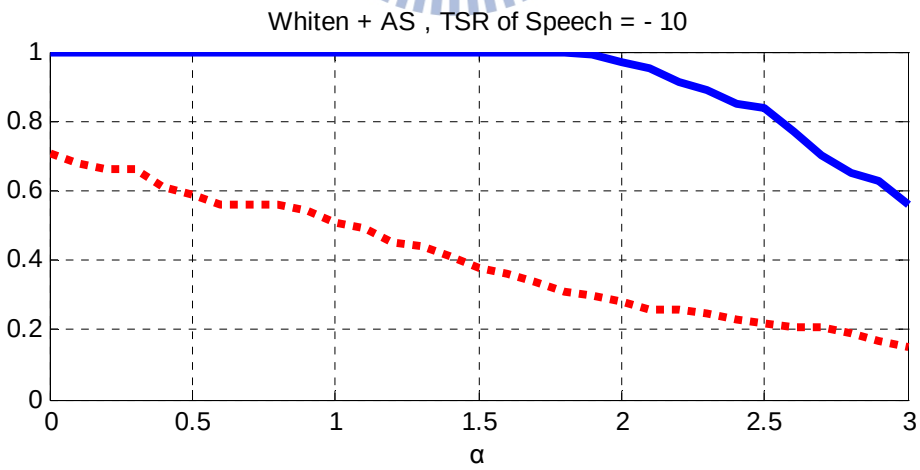

圖 4.19 干擾為 Speech 在 TSR = -10 的情況下經過 SS 後 TNAD 的結果

由以上實驗結果得知，隨著 TSR 下降要維持 detection rate 為 100%，適合的 α 值會從 1.6 上升到 1.8，變化不大，而適合的 β 會大幅降低。為了在不同的 TSR 下維持 100% 的 detection rate，參數 α 必須在 1.6 以下，β 必須在 2 以下，因此 SS 在有語音的情況下對於提升 TNAD 的準確度效果不大。圖 4.20 是選定適合的參數在不同的 TSR 下干擾聲源為語音時 whiten 後的訊號經過 SS 或 AS 處理後，TNAD 所得到 detection rate 為 100% 時 false alarm rate 的結果。經過 SS 處理， false alarm rate 降低的幅度有限。而經過 AS 處理後，在 TSR 為 4 以上時，可以將 false alarm rate 降低為 0。

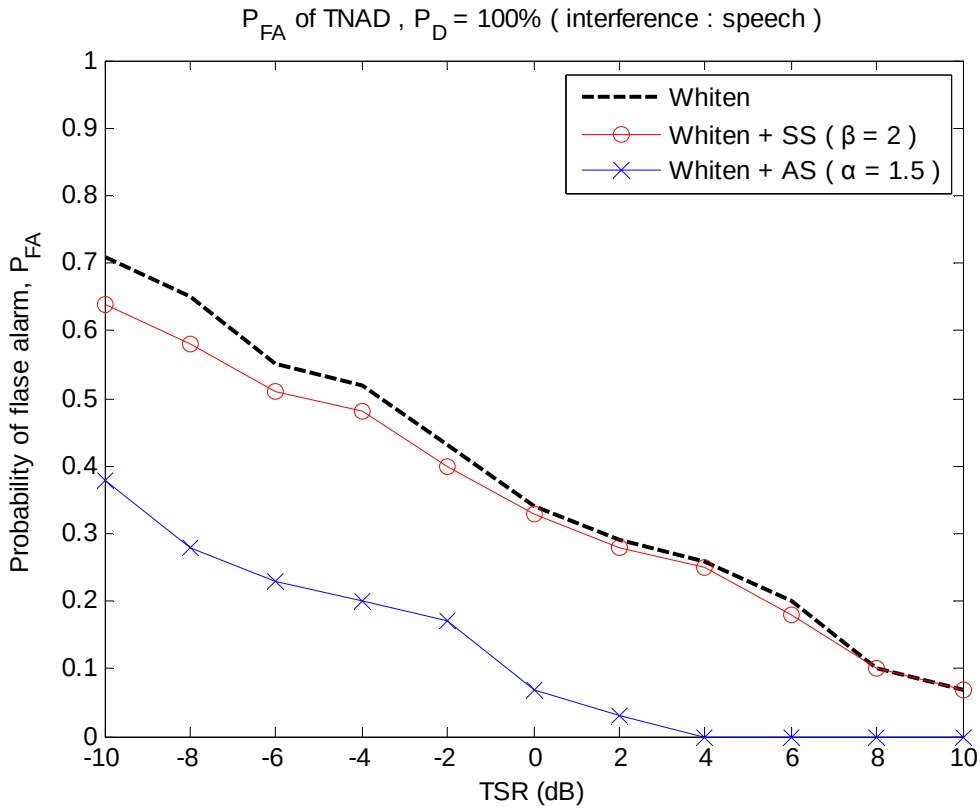

圖 4.20 不同 TSR 下干擾聲源為 Babble 經過三種方法處理後 TNAD 的結果

4.2 暫態噪音聲源方位估測實驗結果與分析

暫態噪音聲源方位估測可以分為兩部分：第一部分是 TNAD 得到的暫態噪音活動偵測結果，第二部分是利用偵測結果針對存在暫態噪音的音框做聲源方位估測。

在本實驗中，為了找出適合偵測暫態噪音的 DOA 演算法，先單獨測試第二部分的結果，利用環境存在不同種類干擾聲源測試 SBF DOA 與 MUSIC DOA 對於暫態噪音聲源方位估測的效果，討論優缺點找出適合的 DOA 演算法。最後測試本論文提出的方法對於不同干擾聲源的穩健程度。

實驗測試音檔共有 8 組，為陣列中 8 顆麥克風接收到的訊號，每組包含 50 次聲源方向為 90 度不同振幅大小的暫態噪音，音檔長度為 2750 個音框，每次 DOA 處理一個音框，音框長度為 256，每次 TNAD 處理一個 block，block 長度為 5 個音框，資料中總共有 550 個 block，存在 50 個包含暫態噪音的 block。

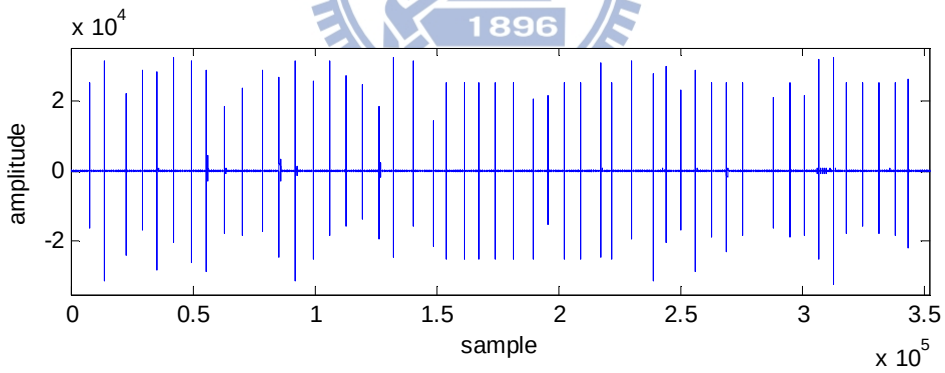

圖 4.21 包含 50 次不同震幅大小暫態噪音的實驗音檔

有三種 TSR 平均為 0 的不同干擾聲源，分別是無方向性的 F16 noise、無方向性的 Babble noise 以及聲源角度為 180° 的語音。語音在不同時間 TSR 的變化程度較大，如圖 4.22。

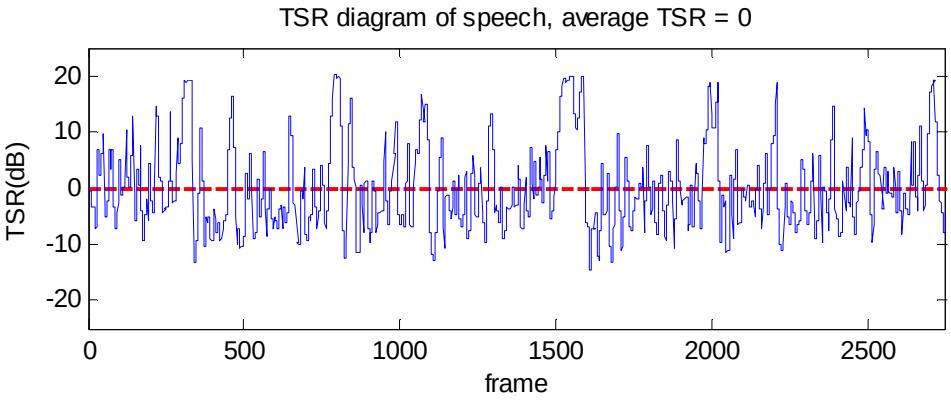

圖 4.22 干擾聲源為聲源角度 180° 的語音時 TSR 的變化情形

將暫態噪音音檔和相同長度的干擾聲源音檔混音成以下三種音檔：

測試音檔 1： 聲源角度為90°的暫態噪音(擊掌) ＋
無方向性的非暫態穩態噪音(F16)

測試音檔 2： 聲源角度為90°的暫態噪音(擊掌) ＋
無方向性的非暫態非穩態噪音(Babble)

測試音檔 3： 聲源角度為90°的暫態噪音(擊掌) ＋
聲源角度為180°的語音

在這節所有的實驗中，TNAD 的門檻值固定在 0.45，核心參數 σ 固定為 10^7，隨機行走階數 t 固定為 25。為了讓演算法在 real-time 上實現，我們選擇 block 長度為 5 以減少運算量。

首先測試 SBF DOA 與 MUSIC DOA 對於暫態噪音聲源方位估測的效果：從測試音檔中選出 50 個包含暫態噪音的音框，只固定對這 50 個音框進行聲源估測。

當干擾聲源為 F16 noise 和 Babble noise 這兩種無方向性的非暫態噪音時，在不同的 TSR 下 SBF DOA 和 MUSIC DOA 都可以在範圍為 ±5 度的標準內估測出正確的位置。這樣的情形下，我們選擇估測結果的 RMSE 做為表示聲源方位估測的效能的標準。

圖 4.23 和圖 4.24 是 SBF DOA 和 MUSIC DOA 在干擾聲源為非暫態噪音時，對選定的 50 個音框進行聲源估測的 RMSE。我們可以看到利用特徵結構估測訊號到達角度的 MUSIC DOA 可以比偵測各方向能量決定聲源方位的 SBF DOA 得到更精確的估算結果。

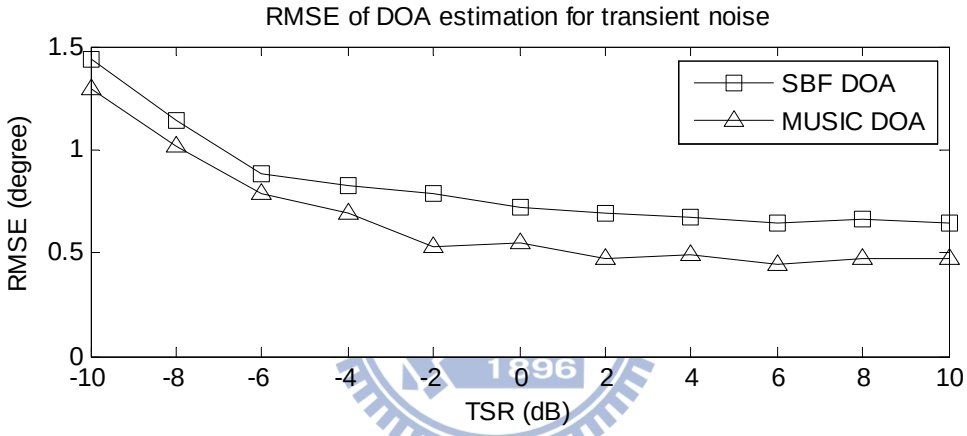

圖 4.23 干擾聲源為非暫態穩態噪音時估測聲源方位的 RMSE

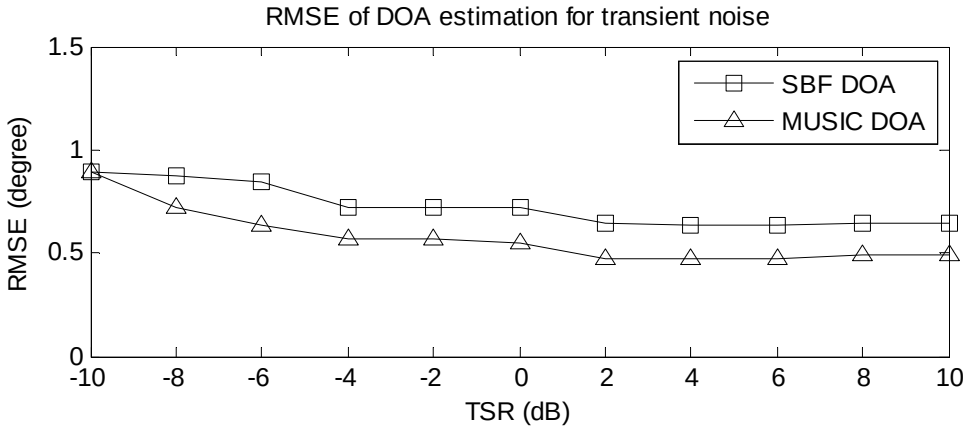

圖 4.24 干擾聲源為非暫態非穩態噪音時估測聲源方位的 RMSE

在干擾聲源為聲源角度 180° 的語音時，SBF DOA 和 MUSIC DOA 已經無法在不同的 TSR 下，於範圍為 ±5 度的標準內估測出正確的位置。因此在干擾聲源為有方向性語音的情況下，我們用偵測到正確位置的機率來當估測角度方法效能的標準。當估測角度為正確角度±5 度的範圍內時，判定為正確位置。

表 4.3 是 SBF DOA 和 MUSIC DOA 在干擾聲源為角度 180° 的語音時，對選定的 50 個音框進行聲源估測的 RMSE。我們可以看到在干擾聲源是有方向性語音的情況下， SBF DOA 可以比 MUSIC DOA 估測暫態噪音聲源方位更為準確，並且在 TSR 為 2 dB 以上可以有 100% 的 detection rate。因為 MUSIC DOA 會估算音框內相關程度較高訊號的聲源位置，而 SBF DOA 是估算音框內能量最強的訊號的位置。因此音框內同時存在語音以及暫態噪音的時候，MUSIC DOA 會傾向於估測語音聲源的位置。

TSR(dB)	SBF DOA	MUSIC DOA
10	100 %	100 %
8	100 %	100 %
6	100 %	100 %
4	100 %	98 %
2	100 %	96 %
0	98 %	88 %
-2	94 %	84 %
-4	90 %	78 %
-6	82 %	74 %
-8	78 %	68 %
-10	64 %	60 %

表 4.3 干擾聲源為有方向性的語音時估測聲源方位的 Detection rate

從 4.1 暫態噪音活動偵測的實驗結果與分析以及以上實驗中，我們可以找到一套適合偵測暫態噪音以及估測暫態噪音聲源位置的演算法：

1. 在各種不同大小的干擾聲源下，不論干擾聲源的種類為何，時域振幅刪減法不需要訓練資料(training data)就能夠直接壓抑非暫態噪音的成分，可以有效的降低 false alarm rate，以提升暫態噪音活動偵測的準確率。
2. 對於估測暫態噪音聲源位置，在干擾聲源為非暫態噪音時，MUSIC DOA 對於聲源估測的精準度比起 SBF DOA 要來的正確，但是這兩套演算法都可以在容許的範圍內估測正確的位置。而干擾聲源為語音時，SBF DOA 的穩定性較高。因此在目標聲源為暫態噪音時，SBF DOA 是最適合作為聲源估測的演算法，計算量低就可以達到偵測正確位置的效果，很適合在 Real-time 的情況下應用。

最後我們測試本論文的演算法對於不同干擾聲源時，追蹤暫態聲源方位的效能，在這部分實驗指標參數定義如下：

	TN source location	Other case
Decided as TN source location	True Positive	False Positive
Decided as other case	False Negative	True Negative

表 4.4 暫態聲源方位追蹤實驗指標參數定義表

Detection rate：True Posirtive / (True Positive + False Negative)

在暫態噪音聲源位置下，被判斷為暫態噪音位置的機率。

False alarm rate：False Positive / (False Positive + True Negative)

在不是暫態噪音聲源位置下，誤判為暫態噪音位置的機率。

表 4.5 和表 4.6 為暫態噪音聲源方位估測的結果。在干擾聲源為非暫態噪音時，因為 SBF DOA 在估測暫態噪音聲源位置幾乎不受影響，因此效能主要是由 TNAD 的結果決定。在干擾聲源為語音時，當 TSR 下降時，TNAD 和 SBF DOA 的 false alarm rate 都會上升。

TSR(dB)	Stationary noise	Non-stationary noise	Speech
10	100 %	100 %	100 %
8	100 %	100 %	100 %
6	100 %	100 %	100 %
4	100 %	100 %	100 %
2	100 %	100 %	100 %
0	100 %	100 %	98 %
-2	100 %	100 %	94 %
-4	100 %	100 %	92 %
-6	100 %	100 %	80 %
-8	100 %	100 %	76 %
-10	100 %	100 %	60 %

表 4.5 各種情況干擾聲源下暫態噪音聲源方位估測的 Detection rate

TSR(dB)	Stationary noise	Non-stationary noise	Speech
10	0 %	0 %	0 %
8	0 %	0 %	0 %
6	0 %	0 %	0 %
4	0 %	0 %	0.2 %
2	0 %	0 %	0.8 %
0	0 %	0 %	1.7 %
-2	0 %	0 %	4.9 %
-4	0 %	0 %	7.9 %
-6	0.6 %	0 %	12.4 %
-8	14.4 %	0.5 %	17.7 %
-10	43.7 %	2.9 %	22.9 %

表 4.6 各種情況干擾聲源下暫態噪音聲源方位估測的 False alarm rate

第五章 結論

5.1 研究成果

本論文提出了一套偵測暫態噪音並追蹤暫態噪音聲源方位的演算法並由大量的樣本實驗證明它的效能。麥克風陣列接收的訊號經過 whiten 以及時域振幅刪減法處理之後，可以非常有效的抑制穩態與非穩態的非暫態噪音的訊號，而對於語音也有一定的抑制效果，使得此方法在環境不理想時也有相當良好的辨識率。暫態噪音活動偵測的結果在經過 SBF DOA 估測聲源方位後，對於干擾聲源為非暫態噪音的 TSR 為 -4 dB 以上，以及對於語音的 TSR 為 6 dB 以上時，暫態噪音聲源方位估測可以達到 100% detection rate 以及為 0 的 false positive rate 的表現。因此在環境存在干擾聲源時，本方法仍然可以準確的偵測暫態噪音，並估測正確的聲源方向，在裝置有遠距喚醒的需求時，可以取代關鍵字做為另一種喚醒機制。

5.2 未來展望

對於加強暫態噪音偵測準確度或許可以嘗試以下的方法：

1. 利用陣列信號處理的方法再加大暫態噪音與一般聲音的差異
2. 隨著接收訊號大小適應性的選取適合的核心參數
3. 在語音開始或是音位(phoneme)轉換的時候，在暫態噪音活動偵測中容易被辨識為暫態噪音，開發能夠辨別這兩者之間差異的演算法。

加入時域振幅刪減法改良後的暫態噪音活動偵測演算法對抗語音干擾有一定的效果，因此也可以準確的偵測暫態噪音在一串語音中的位置及成分，可以應用於一般常用的語音強化演算法上。

Reference

[1] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol.27, pp. 113-120, Apr. 1979.

[2] R. Talmon, I. Cohen, and S. Gannot, “Speech enhancement in transient noise environment using diffusion filtering,” *Proc. 35th IEEE Internat. Conf. Acoust. Speech and Signal Process. (ICASSP-2010), Dallas, Texas*, pp. 4782–4785, Mar. 2010.

[3] Wen-Jun Zeng and Xi-Lin Li, “High-Resolution Multiple Wideband and Nonstationary Source Localization With Unknown Number of Sources,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3125–3136, 2010.

[4] Eric A. Lehmann, ”Particle Filtering Methods for Acoustic Source Localisation and Tracking”, *Ph.D. thesis, Australian National University (ANU), Canberra, Australia*, July 2004.

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

[6] J.-S. Hu, M.-T. Lee, and T.-C. Wang, “Wake-Up-Word Detection for Robots Using Spatial Eigenspace Consistency and Resonant Curve Similarity,” *Robotics and Automation, 2011. ICRA ’11. IEEE International Conference on*, pp. 3901–3906, 2011.

[7] B. Scholkopf, A. Smola, and K. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Comput.*, vol. 10, pp. 1299–1319, 1996.

[8] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Comput.*, vol. 15, pp. 1373–1396, 2003.

[9] D. L. Donoho and C. Grimes, “Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data,” *PNAS*, vol. 100, pp. 5591–5596, 2003.

[10] J.B. Tenenbaum, V. de Silva and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, volume 290, pages2319-2323, 2000

[11] S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear embedding. Science,volume 290, pages 2323–2326, 2000

[12] H.-T. Chen, H.-W. Chang, and T.-L. Liu. Local Discriminant Embedding and Its Variants. In Proc. Int’l Conf. on Computer Vision and Pattern Recognition, volume 2, pages 846-853, 2005.

[13] R. Coifman and S. Lafon, “Diffusion maps,” *Appl. Comput. Harmon.Anal.*, vol. 21, pp. 5–30, Jul. 2006}

[14] B. Nadler, S. Lafon, R. Coifman, and I. G. Kevrekidis, “Diffusion maps,spectral clustering and reaction coordinates of dynamical systems,” *Appl.Comput. Harmon. Anal.*, pp. 113–127, 2006.

[15] A. Singer, Y. Shkolnisky, and B. Nadler, “Diffusion interpretation of nonlocal neighborhood filters for signal denoising,” *SIAM Journal ImagingSciences*, vol. 2, no. 1, pp. 118–139, 2009.

[16] R. Talmon, I. Cohen, and S. Gannot, “Transient noise reduction using nonlocal diffusion filters,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1584–1599, Aug. 2011.

[17] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, “Supervised Graph-Based Processing for SequentialTransient Interference Suppression,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol.20, no. 9, pp. 2528–2538, Aug. 2011.

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

[19] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, ”Computer-steered microphone arrays for sound transduction in large rooms,” *J. Acoust. Soc. Am.*, vol. 78 Issue 5 pp. 1508-1518, July 1985.