

# 第 3 章 TDOA 方法

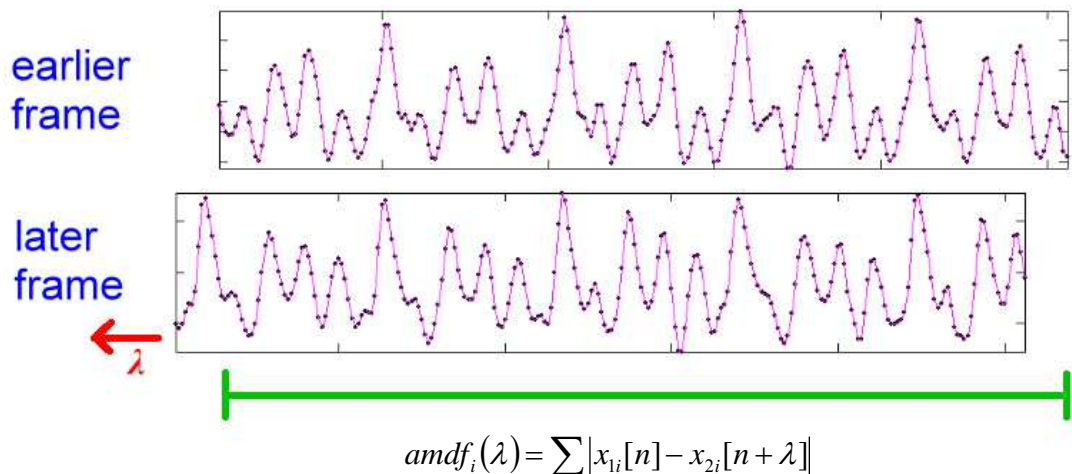
## 3.1 AMDF

### 3.1.1 AMDF 方法介紹

AMDF 的全名為 Average Magnitude Difference Function。x<sub>1</sub> 及 x<sub>2</sub> 分別為兩個麥克風所取樣到的兩段訊號，則

$$amdf_i(\lambda) = \sum |x_{1i}[n] - x_{2i}[n + \lambda]| \quad (9)$$

其中 i 表示音框編號，λ 為位移，上式即可求得 AMDF 值；也就是說，移動兩段訊號中的其中一段，位移 λ 個取樣點，將訊號兩兩相減後求總和，其值就是 AMDF 值。而兩段訊號最相近的時間點其差異必最小，所以我們在合理的平移範圍內(後面的章節會有詳盡的說明)，找出 AMDF 的最小值，其最小值所在的位移值即可推算出兩段訊號的時間差 TDOA。如圖：



圖表 4 - AMDF 示意圖

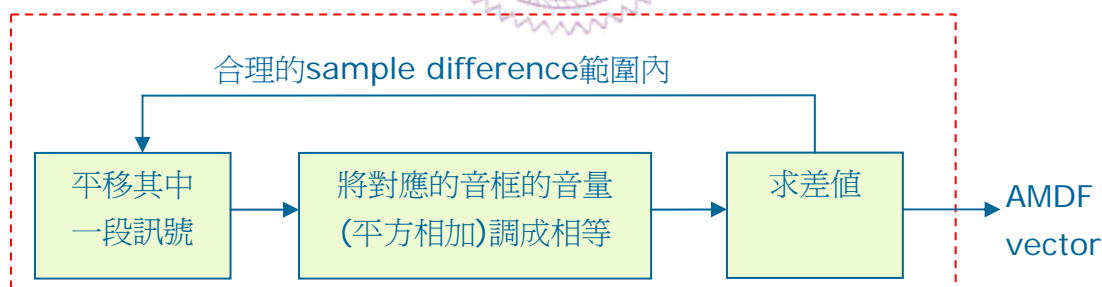
### 3.1.2 AMDF 的實作流程

AMDF 實作流程詳細說明如下（流程圖請參考圖 5）：

1. 前處理：將對應的音框的音量（平方相加）調成相等。求出對應的兩個音框的音量比值，再將音框中的點都乘上比值的開根號，使得其中一個音框的音量相對放大或縮小，如下式：

$$estimatedFrame2 = \sqrt{\frac{energy1}{energy2}} \times frame2 \quad (10)$$

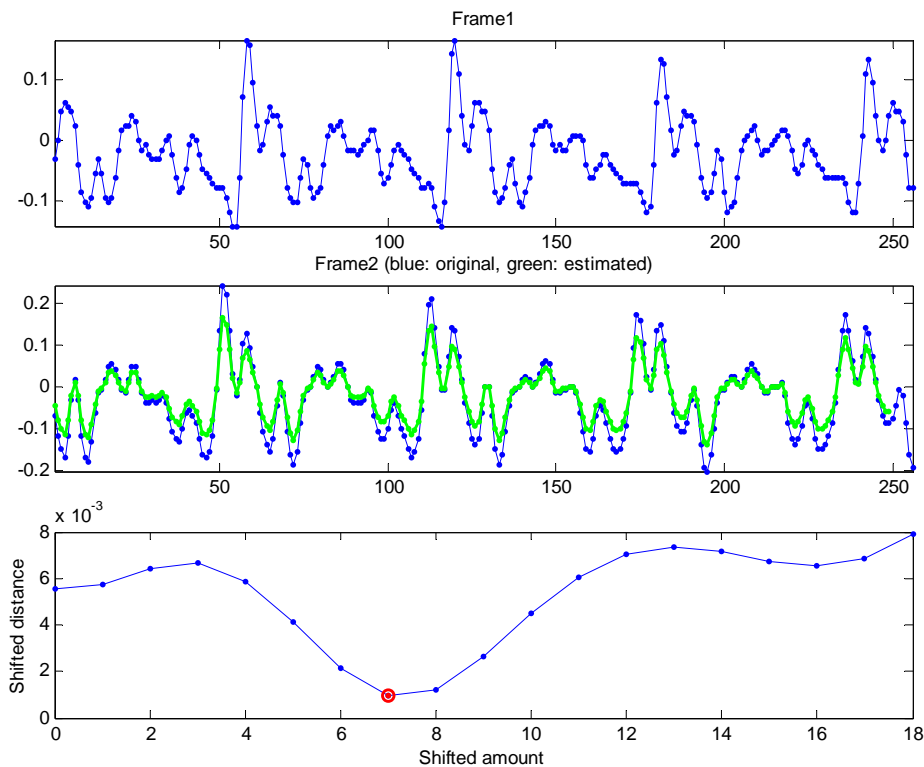
2. 藉由音量我們可判斷出哪個麥克風距離音源較遠，在合理的平移值範圍內（0~maxShift），平移此麥克風所取樣到的訊號，並作 AMDF，則每個音框都會有一個相對應的 AMDF 向量。
3. 找出 AMDF 向量的最小值，藉著其最小值所在的位移值，再除以取樣率即可推算出 TDOA 值。



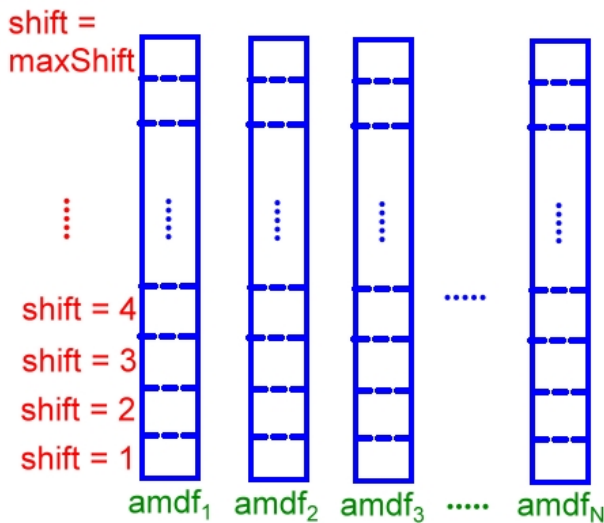
圖表 5 - AMDF 流程示意圖

以一個音框的角度來看，我們舉圖 6 為例來說明，圖 6 的上圖和圖 6 的中圖分別為兩個相對應的音框，而圖 6 的中圖的綠色部分則代表音量調整過後的訊號，圖 6 的下圖即為 AMDF 向量，而其最小值所在的平移值（在此例中為平移值等於 7 的地方）即為兩段訊號相差的取樣點數，藉著這個平移值我們可以進一步推算出 TDOA。

以整句語料的角度來看，所有的音框的 AMDF 向量將構成一個 AMDF 矩陣，如圖 7，而在實驗結果中，往水平的方向將可看出有最小值集中在某個平移值的趨勢。



圖表 6 - 一個音框的 AMDF 結果

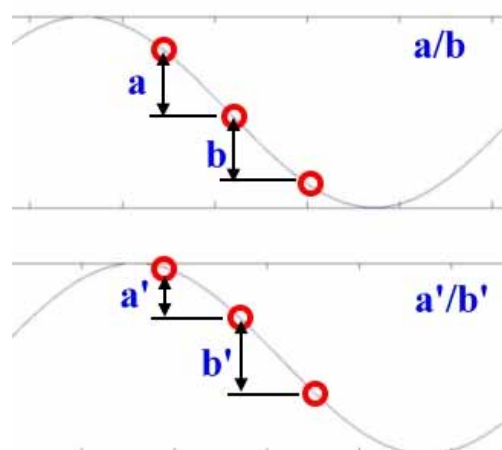


圖表 7 - AMDF 矩陣示意圖

## 3.2 ratioAMDF

### 3.2.1 ratioAMDF 方法說明

在前一節 AMDF 方法中，我們在將訊號作 AMDF 運算之前，必須先將兩段訊號的音量調成相等，使兩段訊號的特性更相近，而比值的 AMDF 方法的想法來源，即是針對這一點。我們直接以取樣點與取樣點之間距離的比值來做運算，而不用去在意音量的衰減對訊號產生的差異。取樣點與取樣點之間距離的比值求法如下圖：



圖表 8 - 比值的 AMDF 示意圖

圖中  $a$ 、 $b$  及  $a'$ 、 $b'$  分別為兩段訊號的取樣點與取樣點之間的距離（即兩個取樣點相減），我們則直接以比值  $a/b$  及  $a'/b'$  來做 AMDF 運算，在合理的平移範圍內每個對應的音框都可算出一個 ratioAMDF 向量，而 ratioAMDF 向量的最小值所對應的平移值，即可推算成 TDOA。

### 3.3.2 ratioAMDF 實作流程

ratioAMDF 實作流程的詳細說明如下：

1. 對每個音框計算出取樣點與取樣點之間距離的比值：

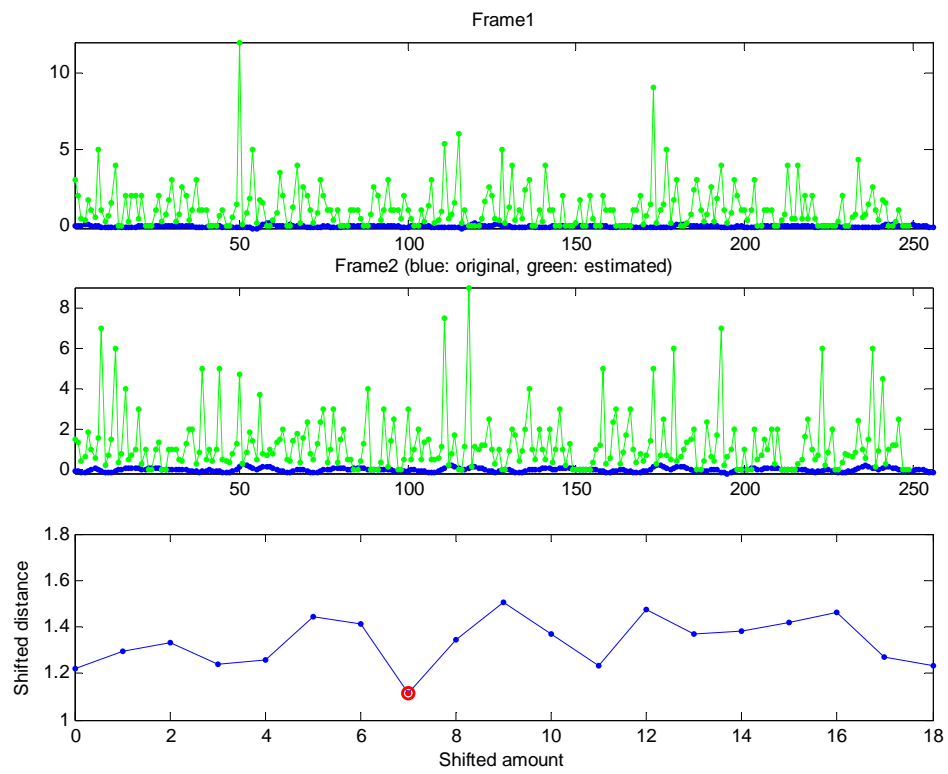
$$ratio_i(n) = \frac{|x_i[n] - x_i[n+1]|}{|x_i[n+1] - x_i[n+2]|} \quad (11)$$

其中， $1 \leq i \leq N$ ， $N$  為音框總數， $1 \leq n \leq M-2$ ， $M$  為音框大小。

2. 經由音量大小，我們可判斷出哪個麥克風距離音源較遠，在合理的平移值範圍內，平移此麥克風所取樣到的訊號，並做 AMDF，則每個音框都會有一個相對應的 ratioAMDF 向量。

3. 找出 ratioAMDF 向量的最小值，其最小值所對應的平移值，再除以取樣頻率即可推算出 TDOA 值。

以一個音框的角度來看，我們舉下圖為例來說明。圖表 9 的上圖與中圖藍色部份分別為兩段訊號的波形，由於 wave 檔的範圍為-1~1，所以以圖的比例來說，真實訊號的部份看起來會較不明顯，而綠色部分則是取樣點與取樣點之間的距離比值結果；圖表 9 下圖即是 ratioAMDF 向量，其最小值所在的平移值（在此例中為平移值等於 7 的地方）即為兩段訊號相差的取樣點數，藉著這個平移值我們可以進一步推算出 TDOA。



圖表 9— 一個音框的 ratioAMDF 結果



## 3.3 最小平方方法

### 3.3.1 最小平方方法方法說明

$X$  及  $Y$  分別為兩個麥克風所取樣到的兩段訊號， $X = [x_1, x_2, \dots, x_N]$ ， $Y = [y_1, y_2, \dots, y_N]$ ， $N$  為音框數。由於兩段訊號是源自同一個音源，所以兩段訊號之間是振幅大小的不同及延遲的關係，則我們可以下式來表示  $X$ 、 $Y$  之間的關係：

$$Y = \alpha \cdot X + \beta \quad (12)$$

其中  $\alpha$  及  $\beta$  為常數。我們改以矩陣的方式來表示，則  $X$  及  $Y$  的關係可寫成  $Y = A * b$ ，展開為：

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} \quad (13)$$

$Y = A * b$

為了逼近  $X$  及  $Y$  兩段訊號，我們要找出一組  $b$  使得  $(Ab - Y)^2$  為最小，透過線性代數的運算，可得：

$$b = (A^T A)^{-1} A^T Y \quad (14)$$

則  $X$  逼近  $Y$  後的結果為  $\hat{Y} = A * b$ 。最後我們以  $\hat{Y}$  及  $Y$  來作比較，在合理的平移範圍內，求  $\hat{Y}$  和  $Y$  的差值，則最小的差值所對應的平移值即可推算出 TDOA。

### 3.3.2 最小平方方法實作流程

以下最小平方方法實作流程的說明：

1.將矩陣 A 代入公式(13)求出 b。

2.計算  $\hat{Y}=A * b$ 。

3.在一個合理的平移範圍內，將  $\hat{Y}$  及 Y 相減後，求其總和，得出一個差值向量。如下式：

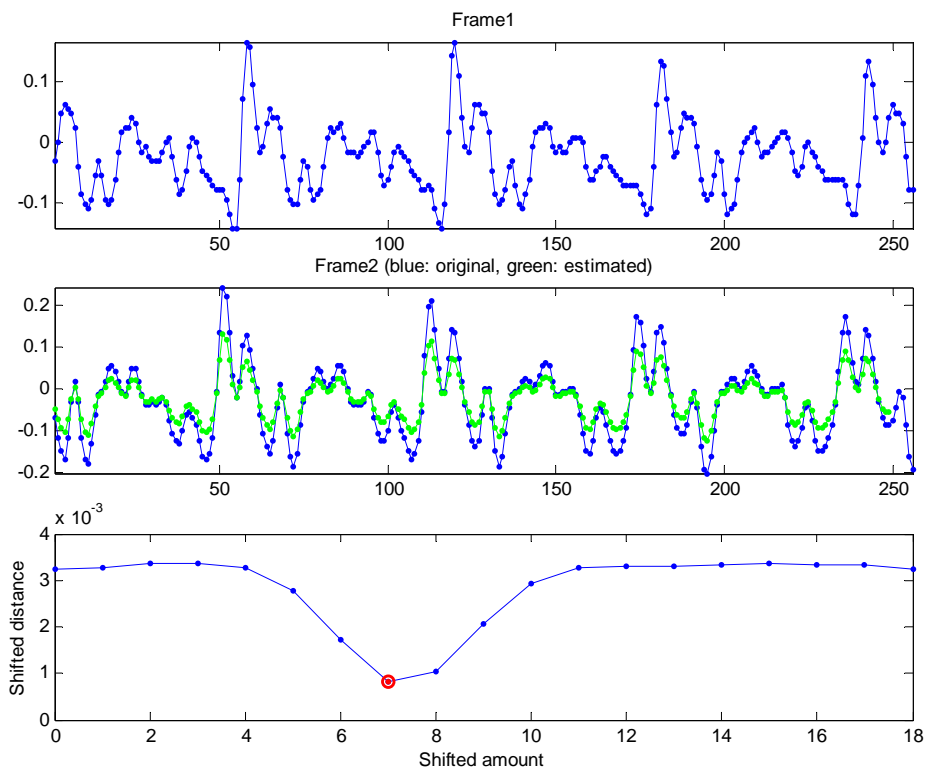
$$difference_i = \sum_{k=1}^M |\hat{y}_k - y_k| \quad (15)$$

其中，M 為音框的大小，i 為平移值。

4.找出差值向量的最小值，其最小值所在的平移值，再除以取樣頻率，即可推算出 TDOA。

以一個音框的角度來看，我們舉下圖為例。圖表 10 的上圖與中圖藍色部份分別為兩段訊號的波形，而中圖的綠色部分則是以最小平方方法來逼近其中一段訊號的結果，圖表 10 的下圖即是差值向量，其最小值所在的平移值（在此例中為平移值等於 7 的地方）即為兩段訊號相差的取樣點數，藉著這個平移值我們可以進一步推算出 TDOA。





圖表 10—一個音框的最小平方法結果



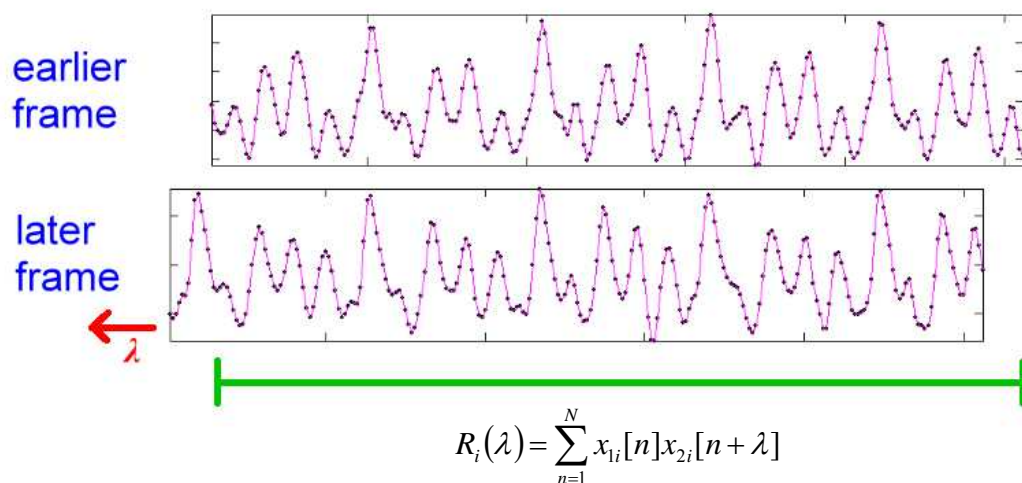
## 3.4 Cross Correlation

### 3.4.1 Cross Correlation 方法說明

$x_1$  及  $x_2$  分別為兩個麥克風所取樣到的兩段訊號，則

$$R_i(\lambda) = \sum_{n=1}^N x_{1i}[n]x_{2i}[n+\lambda] \quad (16)$$

其中  $i$  表示音框編號， $\lambda$  為平移。經由音量大小，我們可判斷出哪個麥克風距離音源較遠，平移此麥克風所取樣到的訊號，經由上式我們就可以計算出 cross correlation 值。而兩段訊號最相近時的時間點，其相關性最大，也就是說 cross correlation 值必最大，所以我們在合理的平移範圍內，找出 cross correlation 的最大值，其對應的位移值即可推算出兩段訊號的時間差 TDOA。如下圖所示：



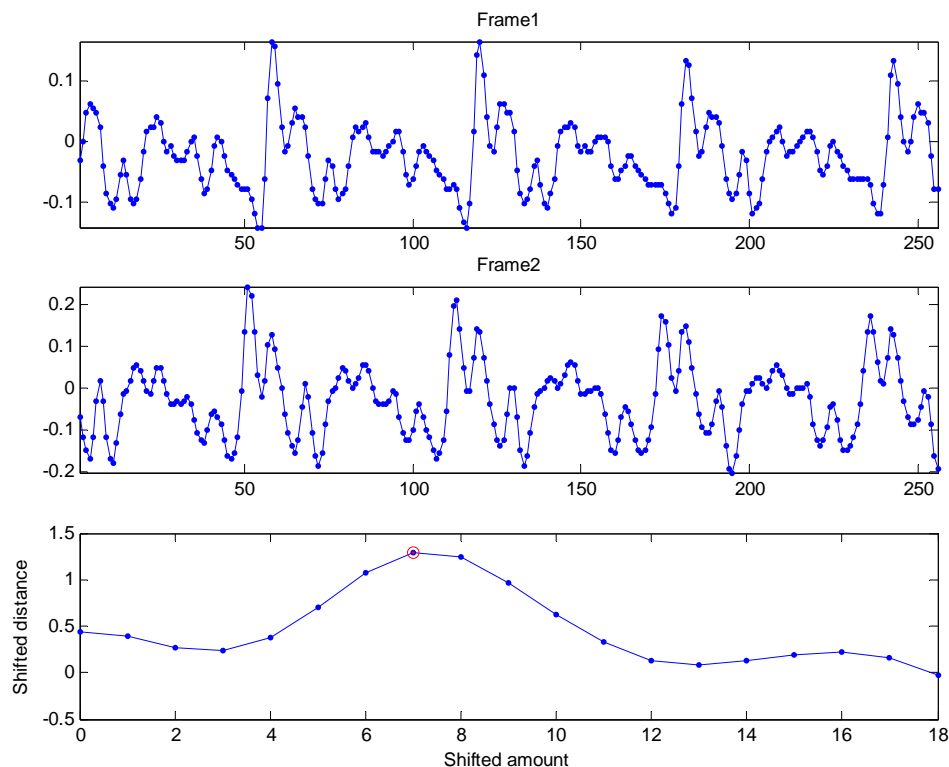
圖表 11 - Cross Correlation 示意圖

### 3.4.2 Cross Correlation 實作流程

以下為 Cross Correlation 實作流程的說明：

1. 在合理的平移範圍內，重複平移距離音源較遠的麥克風所取樣到的訊號，並，並將每個音框的取樣點與取樣點相乘後求總和，則每個音框都會有一個相對應的 Cross Correlation 向量。
2. 找出 Cross Correlation 向量的最大值，其最大值所對應的平移值，再除以取樣頻率，即可推算出 TDOA 值。

以一個音框的角度來看，我們舉下圖為例。圖表 12 的上圖與中圖藍色部份分別為兩段訊號的波形，圖表 12 的下圖即是 Cross Correlation 向量，其最大值所在的平移值（在此例中為平移值等於 7 的地方）即為兩段訊號相差的取樣點數，藉著這個平移值我們可以進一步推算出 TDOA 值。



圖表 12 – 一個音框的 Cross Correlation 結果

## 3.5 整句語料的 TDOA

在前幾節所介紹的 TDOA 方法中，都是以音框的角度來說明 TDOA 的方法內容，而實際上我們在處理整句語料時，則稍有變化，本節將說明處理整句語料和處理音框的不同，以及辨識率的計算方法。（在往後的章節，TDOA 的值我們以取樣點來表示）

### 3.5.1 最短路徑法

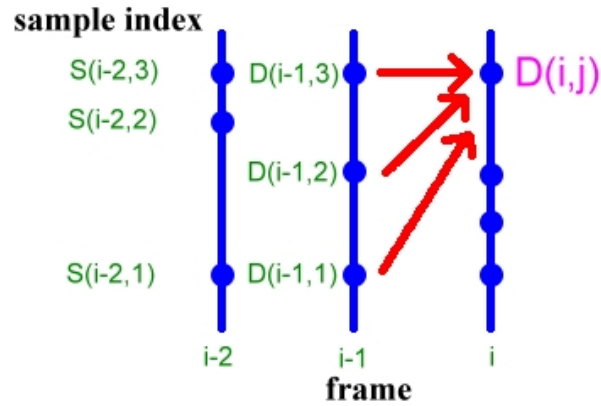
理論上，由同一句語料，所估算出來的所有音框的 TDOA 都應相同，但是在實際的環境中，可能因為環境雜訊 (background noise)、聲音的反射 (reflection) 以及混響 (reverberation) 等，造成實驗上的誤差，所以我們以 Dynamic Programming 的方式來找出一段誤差最小的路徑，來表現 TDOA 及方向角的變化趨勢。

在此，我們以 AMDF 的方法為例來說明如何以最短路徑法來找出整句語料的 TDOA 變化趨勢。找出所有對應的音框所估算出的 AMDF 向量的局部最小值 (local minima)，然後以下面的遞迴關係式算出所有 local minima 所對應的 cost：

$$D(i, j) = amdf(i, j) + \min \begin{cases} D(i-1, 1) + weight * (S(i, j) - S(i-1, 1)) \\ D(i-1, 2) + weight * (S(i, j) - S(i-1, 2)) \\ \vdots \\ D(i-1, L(i-1)) + weight * (S(i, j) - S(i-1, L(i-1))) \end{cases} \quad (17)$$

$D(i, j)$  為第  $i$  個 AMDF 向量中的第  $j$  個 local minimum 所對應的 cost，其中， $1 \leq i \leq N$ ， $N$  表音框數， $1 \leq j \leq L(i)$ ， $L(i)$  表第  $i$  個 AMDF 向量的 local minimum 個數， $S(i, j)$  表示第  $i$  個 AMDF 向量中第  $j$  個 local minimum 所對應的平移值。

如下圖所示：



圖表 13－最短路徑法示意圖

橫軸代表音框，縱軸代表 local minimum 所對應的取樣點差，也就是平移值，而每個音框上的點即為 AMDF 向量的 local minimum。  $D(i,j)$  的求法，即是前一個音框所有的 AMDF 向量之 local minima 所對應的 cost，分別加上其 local minima 對應的平移值與目前的 AMDF 向量中第  $j$  個 local minimum 對應的平移值的差值再乘上一個權重，其中值最小者，即為  $D(i,j)$ 。

在計算遞迴關係式時，要記錄每一個點所對應的路徑，直到我們求出所有點的 cost 後，找出最後一個音框所對應的所有 cost 的最小值，再以 back tracking 的方式，反覆回推前一個最佳路徑的位置，如此一再反覆，才能算出整個最佳路徑。

不同 TDOA 方法的 weight 值都不一樣，觀察不同 TDOA 方法下的 local minima 的分佈範圍，經由實驗我們得到不同 TDOA 方法的 weight 值，在下一章的實驗結果中將有較詳細的說明。

## 3.5.2 辨識率的計算

前一小節中所提到的處理整句語料的方法，能讓我們看出一句語料中 TDOA 與方向角的走勢，但對於系統而言，尚不能以此方法來評量整個系統。為了方便計算本系統的效能，此小節將說明兩種計算本系統辨識率的方法。

### 1. 整句語料的辨識角度

以最短路徑法找出一條整句語料的 TDOA 變化路徑，然後以投票的方式，在此路徑中統計所有的音框所對應的 TDOA 是落在哪個平移值上，而佔有最多 TDOA 的平移值，我們即當作此句語料所代表的 TDOA，並以此 TDOA 套入方向角公式，求出此句語料的音源入射方向角。

### 2. 辨識結果表達方式

共有以下兩種計算辨識率的方式：

#### a. 以曲線圖來表達辨識結果：

分別在同一個距離、不同角度下的音源位置，將所有的方向角估算結果取平均，再以曲線呈現，所以每張辨識結果的曲線圖中，將會有四條曲線分別代表不同距離下的方向角辨識結果。

#### b. 以語料筆數來表達辨識結果：

將所有語料估算出來的音源入射方向角和實際的音源入射方向角作比較，統計方向角相差值在某些誤差範圍下的語料筆數，其所佔百分比即為辨識率。

