聲源三維方位之偵測系統
A Three-dimensional Sound Source Localization System

古鴻炎(*Hung-Yan Gu*)　　楊善翔(*Shan-Hsiang Yang*)
國立台灣科技大學 資訊工程系
E-mail: {guhy, M9515043}@mail.ntust.edu.tw

摘要 — 本論文研究、製作了一個聲源三維方位的偵測系統，僅使用三個麥克風組成的麥克風陣列來輸入聲音訊號。我們提出以頻譜亂度加 SNR 驗證之方法，來分辨出語音音框；然後使用廣義交互相關函數的逼近算法來估計聲音信號到達麥克風的延遲時間(TDOA)，並且提出以同步式相位複製來改進相位不穩定的問題，再者提出以拋物線內插法來提高 TDOA 的準確度；接著以計算 TDOA 估計值與理論值之間的距離，來尋找出聲源方位，進一步以所提出的不等值投票機制，來對數個音框作綜合的方位角度計算。經由實際偵測之實驗可知，我們系統花費的計算量少，可輕易達成即時處理的目標，並且平均的角度偵測誤差很小，水平角與仰角分別的平均誤差是 3.43 度與 2.08 度。[1]

關鍵詞:語音處理、聲源方位、麥克風陣列、VAD、TDOA。

一、 前言

聲源方位之偵測，可應用於電傳會議(telconference)、互動玩具等。此外，隨著智慧型機器人的蓬勃發展，應用於機器人聽覺之聲源方位偵測的研究也逐漸增加。目前聲源方位偵測的研究方向主要可分爲二類，一類是將麥克風陣列所接收到的資料，形成相關矩陣，再使用beamforming或子空間理論作處理，以求得聲源的角度，如Multiple Signal Classification (MUSIC) [1]；另一類則是估計聲音到達兩兩麥克風之間的時間延遲Time Delay of Arrival (TDOA)，再利用聲源與麥克風陣列的幾何關係，以估計出聲源的角度[2, 3]。由於第一類研究方向的計算量大，而且必需事先量測陣列中每支麥克風在各角度與距離所分別對應的脈衝頻率響應，因此我們選擇第二類的研究方向，除了改進偵測的方法之外，也把聲源方位偵測系統的實作設爲目標。

我們所製作的聲源方位偵測系統，它的系統架構如圖 1 所示，包含如下列出的組件:

(a) **麥克風陣列**，我們使用三個 Primo 公司所生產的 EM-147 麥克風[10]，來形成正三角形的平面陣列，三角形的邊長設爲 20cm，如圖 2 所示。

(b) **信號放大與濾波電路**，用以將麥克風感應出的電壓信號加以放大(使用 LM386 IC)，再作低通濾波(使用 LM324 IC)，截止頻率則設爲 7,000Hz。

(c) **資料擷取電路**，經由 ADC 得到數位資料，再經 USB 介面將資料傳輸至電腦(或處理器)，在此我們使用美商國家儀器的 NI-9215 DAQ 產品[11]，並且把取樣率設爲 16,000Hz。

(d) Voice Activity Detection (**VAD**)**計算**，對各個語音音框(frame)計算亂度(entropy)及 SNR 值，再據以判斷此音框是語音音框或雜訊音框。

(e) **TDOA 估計**，計算兩兩麥克風之間的廣義交互相關函數，再依據廣義交互相關函數去找出兩兩麥克風之間的 TDOA 估計值。

(f) **方位計算**，依據各音框的 TDOA 估計值與理論值之間的距離，來尋找聲源的方位角度；然後以不等值投票機制，計算多個音框的綜合方位角度。

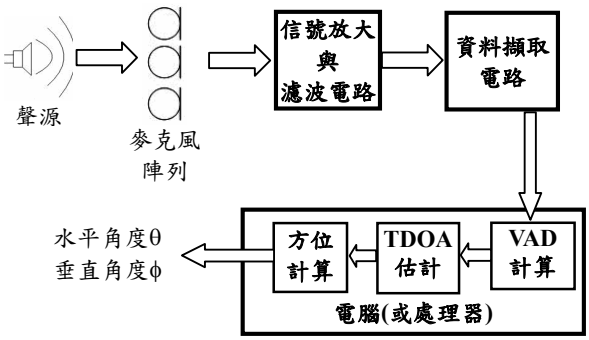

圖1　聲源方位偵測系統之系統架構

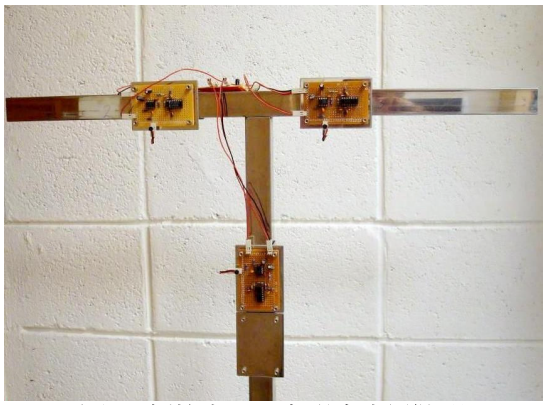

圖2　倒放之正三角形麥克風陣列

關於VAD之計算，我們提出一種以頻譜亂度加SNR驗證之方法，來對持續輸入進來的信號音框作分辨，以區分出含有人類語音之音框，及只含有背景雜訊之音框，以避免使用雜訊音框來作聲源偵測。關於TDOA的估計，我們採取了基於廣義交互相關generalized cross correlation (GCC)的方法[4]，然後對GCC在實務上面臨的缺點，做了幾點改進，以得到更精確的TDOA估計值，例如提出以同步式相位複製來改進相位不穩定的問題，再進一步提出以拋物線內插法來提高TDOA的準確度。關於聲源方位的計算，我們參考了Rabikin等人的研究[5]，但是把他們的方法作修改，以便能夠使用於三個麥克風組

[1] 本研究由國科會贊助，計畫編號 NSC 96-2218-E-011-002。

成的平面正三角形陣列的情況，此外，我們提出一種不等值投票的機制，以便對多個音框作綜合的角度計算，而找出聲源的方位角及仰角。

二、 VAD:語音活動偵測

對於持續接收到的聲音信號音框，我們必須先作語音活動偵測(VAD)，以判斷各個輸入音框是否爲語音音框，若是語音音框才拿去估計 TDOA 及計算聲源方位。我們提出以頻譜亂度加 SNR 驗證來作爲 VAD 的判斷方法。

2.1 頻譜亂度

人類語音的主要能量，約集中於 3kHz 以下，並且頻譜強度隨著頻率軸顯現出明顯的高低起伏，相對地，噪音信號的頻譜強度則較爲平坦。若對一個信號音框的頻譜去計算頻譜亂度，則語音音框的頻譜會有較低的亂度值，因此可訂一亂度值門檻來區分語音和噪音。

在 Renevey 等人的研究中[6]，對於一個信號音框的頻譜，定義以一個頻帶的功率和全頻帶的功率比值，作爲該頻帶的機率，也就是令 w 頻帶的機率 $P(w)$爲

$$P(w,t)=\frac{|Y(w,t)|^2}{\sum_{w=6}^{256\text{-}1}|Y(w,t)|^2},\quad w=6,\ 7,\ \cdots,\ 256\text{-}1 \tag{1}$$

其中 t 代表第 t 個音框，w 表示第 t 個音框頻譜的第 w 個頻帶，$|Y(w,t)|$表示頻帶強度，256 表示 FFT 點數 512 的一半，實驗顯示 w 從 6 開始比從 0 (DC)好。然後，就可依據 $P(w)$ 去計算第 t 個音框的頻譜亂度 H_t，即

$$H_t=-\sum_{w=6}^{255}P(w,t)\cdot\log(P(w,t)) \tag{2}$$

實務上，雜訊的頻譜並不是十分的平坦，並且雜訊信號有隨機變動的現象，因此使得雜訊音框所計算出的亂度值會有明顯的跳動情形，有時會很接近、甚至於低於語音音框的亂度值。因此，Renevey 等人的研究中，提出的一種解決方法是，在計算亂度值之前，先將信號加入少量的白雜訊。

使用 Renevey 等人的解決方法之後，我們發現雜訊音框的亂度值仍然會有不小的跳動情形，因此我們再研究了一種改進方法，就是將雜訊音框的亂度值作如下的濾波處理:

$$\bar{H}_t=max\{H_t,H_{t-1}\} \tag{3}$$

也就是把每個音框算出的亂度值與前一個音框的作比較，再取較大者作爲此音框的亂度值。

2.2 SNR 驗證

藉由 2.1 節之頻譜亂度量測，大體上可區分出語音音框及非語音音框。但是在觀察這些被判斷爲語音的音框之後，可發現其中有部分的音框其實不是語音音框，這是因爲我們把門檻值訂得較寬鬆一些，以便讓語音音框儘量不被誤判，如此就必須再作進一步的驗證，來把那些非語音的音框過濾出來，以提高 VAD 判斷的準確性。

我們經由觀察發現，這些非語音音框與真正的語音音框比較起來，其能量較小。所以我們想到的一個驗證方法是，計算每個音框的能量，並且對於被判斷爲語音的音框，將它的能量與被判斷爲雜訊音框的能量作比較，比較的方法是計算訊號雜訊比，其公式如下：

$$SNR=10\log_{10}\frac{S}{N} \tag{4}$$

其中 S 表示目前被判斷爲語音音框的能量值，而 N 表示先前被判斷爲雜訊音框的平均能量，就是把最近幾個被頻譜亂度判斷爲雜訊的音框，拿去計算能量平均值。根據實驗的觀察，我們訂出一個適當的門檻值(即 12dB)，來把 SNR 低於此門檻的音框改判爲雜訊音框。

三、 時間延遲估計

關於時間延遲(TDOA)的估計，我們採取 Knapp 所提出的基於廣義交互相關(GCC)的方法[4]，並且使用 Phase Transform (PHAT)作爲濾波器的頻率響應函數，如此廣義交互相關的計算方式就如公式(5)：

$$R_{y_1y_2}(\tau)=\int_{-\infty}^{\infty}\frac{G_{x_1x_2}(f)}{|G_{x_1x_2}(f)|}\cdot e^{j2\pi f\tau}df \tag{5}$$

其中 $G_{x_1x_2}(f)$ 表示輸入信號 x_1 及 x_2 的交互功率頻譜 (cross power spectrum)，其定義爲

$$G_{x_1x_2}(f)=\int_{-\infty}^{\infty}R_{x_1x_2}(\tau)\cdot e^{-j2\pi f\tau}d\tau \tag{6}$$

$$R_{x_1x_2}(\tau)=\mathrm{E}\left[x_1(t)\cdot x_2(t-\tau)\right] \tag{7}$$

公式(7)裡，E[·]表示取期望值。理論上當我們找到一個 τ 值使得公式(5)的 $R_{y_1y_2}(\tau)$ 爲最大時，則此 τ 值即爲我們所要的時間延遲估計值。

考慮計算量及方便實作，我們並不是直接依據公式(6)和(7)來計算 $G_{x_1x_2}(f)$，而是參考 Omologo 等人的研究[7]，先分別計算 x_1 及 x_2 的信號音框的 DFT 頻譜 $X_1(f)$和 $X_2(f)$；接著，再使用頻譜 $X_1(f)$和 $X_2(f)$來逼近公式(5)中的 $G_{x_1x_2}(f)$ 以及 $|G_{x_1x_2}(f)|$，也就是以 $X_1(f)\cdot X_2^*(f)$ 逼近算式 (5) 中的 $G_{x_1x_2}(f)$，並以 $|X_1(f)\cdot X_2^*(f)|$ 來逼近 $|G_{x_1x_2}(f)|$。

其實公式(5)相當於對 $G_{x_1x_2}(f)$ 的相位作反向傅立葉轉換，可用以求得一個脈衝函數，該脈衝函數的最大值的水平位置即爲延遲的時間。但是由於實作上是以 DFT 頻譜來作逼近，所以並不會得到理想的脈衝函數。再者，實務上若某些頻帶的 $G_{x_1x_2}(f)$ 很小或爲 0，其相位會變得不穩定。對於相位不穩定的頻帶，通常是頻率大於 4kHz 且頻譜強度很小的頻帶，我們研究了一個改進方法，就是令

$$\theta_i(f)=2\theta_i(\tfrac{f}{2}),\ f=\frac{16000}{512}\times k,\quad k=128,\cdots,255 \tag{8}$$

其中 16,000 是取樣率，512 是 FFT 點數及音框長度。公式(8)對於相位不穩定的頻帶 f，取其半頻 0.5*f 之相位的二倍來作爲頻帶 f 的相位值。

由於以 DFT 作逼近的緣故，實際上只能求得離散時間之廣義交互相關函數，因此 TDOA 只能精確到樣本點之單位，而不是連續的秒數。對此問題，前人採取的一

種補救方法是提高取樣率，但相對的計算量也會增加。因此，我們研究以拋物線內插法來作補救，當以反向 DFT 計算出公式(5)之廣義交互相關函數後，先找出整數個樣本點之 TDOA 估計值τ，再取τ與τ的前後各一點來建立一個拋物線，然後依此拋物線的頂點算出它所對應的橫軸值來作爲較精確的 TDOA 估計值。

四、 方位偵測

應用 TDOA 來估計聲源方位的方法，目前已經有一些論文被提出。在此我們參考了 Rabikin 等人提出的方法[5]，但是做了一些修改，以便應用於三個麥克風所形成的平面陣列的情況。

使用此方法時，必需事先計算出各種角度上的 TDOA 理論值。假設任意兩個麥克風的座標分別爲<xm_1, ym_1, zm_1>和<xm_2, ym_2, zm_2>，並且聲源的座標爲<xs, ys, zs>，則聲波由聲源位置到達這二支麥克風的時間爲

$$t_i = \frac{\sqrt{(xm_i - xs)^2 + (ym_i - ys)^2 + (zm_i - zs)^2}}{c}, \quad i = 1, 2 \qquad (9)$$

其中 c 代表聲音在空氣中傳播的速度，如此就可算出這二支麥克風之間的 TDOA 理論值爲 $D=t_1-t_2$。

接著，我們把陣列中的麥克風兩兩作組合，並且爲每一組合去計算出一個 TDOA 的理論值，設可組合出 M 個麥克風組，再將這 M 個 TDOA 理論值並列成爲一個 M 維向量 $DV=<D_1, D_2, \cdots, D_M>$。依此作法，對於各個可能的聲源的位置，都可以得到一個對應的 DV 向量，我們把這些不同的聲源位置所計算出的向量收集起來，形成一個集合，就稱此集合爲 acoustic map (ACMP) [8]。至於聲源位置的佈放，如果在水平及垂直方向都每間隔 d 度去設定一個聲源位置，再者考慮聲源可能出現的方位角及仰角之範圍都是在±90 度之內，則 ACMP 中總共需要去計算 $(180/d)^2+1$ 個 DV 向量，我們將這些向量依照(方位角、仰角)之大小次序作排列。在系統製作時，我們設 d=5。

實際作聲源方位偵測時，在求得各組麥克風的 TDOA 估計值之後，亦可將它們並列成一個待測向量 $EV=<E_1, E_2, \cdots, E_M>$，然後去計算此待測向量與 ACMP 中所有向量的幾何距離，並找出距離最小的向量，公式如下：

$$k = \arg\min_i \ dist(DV_i, EV), \ \ i = 0, \ \cdots, \ (180/d)^2 \qquad (10)$$

其中 $dist(DV_i,EV)$代表第 i 個 DV 向量與待測向量之間的幾何距離。

根據 acoustic map 中向量的排列方式，我們可將 k 代入公式(11)與(12)中，

$$\theta = 90 - d \times k \,/(180/d - 1) \qquad (11)$$

$$\phi = \begin{cases} 90, & k=0 \\ 90 - [(k\text{-}1) \bmod (180/d - 1) + 1] \times d, & 0 < k < (180/d)^2 \\ -90, & k=(180/d)^2 \end{cases} \qquad (12)$$

以求出θ代表的方位角，及ϕ代表的仰角，而得知聲源目前所在的三維空間方位。

一個語音命令的發音，我們希望只計算出一組聲源空間方位的θ與ϕ值，因此我們研究了一種投票機制，來把一次發音的所有語音音框所計算出的聲源方位向量<θ, ϕ>作投票，以決定此次發音的聲源方位。設一個語音命令中共有 F 個語音音框，則此次發音的聲源方位向量<θ, ϕ>就由下列的不等票值投票公式去計算得到：

$$<\theta, \ \phi> \ = \left(\sum_{i=0}^{F\text{-}1} w_i \right)^{-1} \times \sum_{i=0}^{F\text{-}1} w_i <\theta_i, \phi_i> \qquad (13)$$

其中 w_i 表示第 i 個語音音框的票值權重。由實驗的結果顯示，使用音框能量作爲權重，會比使用音框 rms 振幅好。

五、 系統製作 與 偵測實驗

5.1 系統製作

麥克風輸入的聲訊經由我們製作的放大電路將信號放大，並濾除高頻雜訊，然後透過 DAQ 擷取出數位訊號，轉換出的數位資料再經由 USB 介面傳輸至電腦中，之後再進行 VAD、TDOA、方位估計等運算。

在三個麥克風的聲音信號傳輸至電腦後，我們對各麥克風的信號樣本依序取 512 點作爲一個音框(32ms)，每個音框會先乘上漢寧(Hanning)窗[9]，再經過 512 點之 FFT 運算轉換至頻域，接著，在頻域將信號與白雜訊的頻譜相加。接著使用(1)式及(2)式來計算頻譜亂度，若是被判斷爲語音音框，接著再使用 SNR 進行驗證。我們取最近 100 個非語音音框的能量平均值作爲 SNR 計算中所需的雜訊能量值。三個麥克風訊號的音框經 SNR 驗證後，若皆爲語音音框才進行 TDOA 的估計。

關於 TDOA 的估計，先分別取得各麥克風信號音框的 FFT 頻譜，再計算各組麥克風之間的交互功率頻譜 $X_1(f) \cdot X_2^*(f)$，然後作反向 FFT 運算而估計出公式(5)裡的 $R_{y_1y_2}(\tau)$，之後從τ值的合理範圍中，找出具有最大 $R_{y_1y_2}(\tau)$ 值的τ值，來作爲 TDOA 的估計值。

關於τ 的合理範圍，當聲源的位置與二麥克風形成一直線時，可得到最大的 TDOA 估計值，再者考慮溫度爲攝氏 25 度的室內環境時，聲音的傳播速度約爲 346m / sec，因此可計算出 TDOA 的最大數值爲 20cm / 346m = 5.7803×10^{-4} 秒。由於取樣頻率設爲 16kHz，因此 TDOA 的最大數值約爲 9.2485 個樣本點，但是估計出的 TDOA 值如果只能精確到樣本點之整數值，則會引發不小且不均勻的角度誤差。因此，對於公式(5)作反向 FFT 所求取出的 TDOA 整數樣本點值τ，我們再取左右$\tau+1$和$\tau-1$兩點上的值來計算出拋物線，然後以此拋物線的頂點所對應的橫座標位置作爲新的 TDOA 估計值。

5.2 角度偵測之模擬實驗

我們使用事先錄好的語音信號來進行線外(off-line)模擬實驗，語音的內容包含了十個控制機器人的命令，如表 1 所列。語音信號是由兩男兩女在隔音錄音室中錄音，然後在我們的實驗室裡，使用具有功率放大的喇叭來播放預錄的聲音，我們將喇叭分別放置在距離麥克風陣列中心 100cm 及 150cm 處，測試的方位角包括 ±90°、±60°、±30°、0°，而仰角則包括 ±30°、0° 等，如此共有 42 個空間位置來作語音信號播放及麥克風陣列的收音與

偵測。

表 1 機器人控制之語音命令

前進	開始	左轉	準備	向後轉
後退	停止	右轉	煞車	機器人

在前述條件下，我們系統作角度偵測得到的各錄音者的平均角度誤差及標準差如表 2 所列，由此表可以觀察出，男女生之間的角度誤差以及標準差並沒有明顯的差異，方位角的平均誤差約在 4 度左右，標準差約在 2 度左右，而仰角的平均誤差約在 2 度左右，標準差也約 2 度左右。

表 2 各錄音者的平均角度誤差及標準差

	男 1		男 2		女 1		女 2	
	方位角	仰角	方位角	仰角	方位角	仰角	方位角	仰角
平均誤差	3.46	2.04	4.89	2.71	3.65	1.89	4.09	2.10
標準差	1.22	0.86	3.36	2.42	2.07	1.42	2.82	1.84

如果以各種距離、方位角、仰角的組合，分別統計四位錄音者所發的語音命令的角度偵測誤差，則結果如表 3 所列，從表 3 中對兩種距離都作觀察可以發現，當方位角爲正負 90 度時，偵測出角度的誤差及標準差都比其它角度的大；此外，在距離爲 150cm 且方位角爲正 90 度的情形下，角度誤差及標準差變得最大，我們檢討其原因是，此位置剛好是最靠近噪音源(風扇)的位置，因此在估計 TDOA 時容易受到干擾而產生較大的誤差。

一開始爲了取得較多的語音音框來計算聲源的方位，所以在 VAD 處理時盡量將 SNR 門檻值設得較低(即 12dB)，即使如此，由表 2 與 3 的實驗結果可以看出，我們的系統已可獲得不錯的準確度。接著，我們考慮提高 SNR 門檻值，來將能量較弱的語音音框剔除，以排除雜訊的干擾，所以在後續的實驗，我們將 SNR 門檻值提高到 18dB。

另外，使用拋物線內插法得到的 TDOA 估計值，若是超出 TDOA 值的合理範圍，原先的作法是，採用廣義交互相關所求取出的 τ 值作爲估計值，這樣作可能會減弱拋物線內插法的效果。所以，當超出合理範圍時，這裡改成以合理範圍的邊界(即±9.2485)來作爲估計值。

我們依上述的二種改進方法來進行實驗，實驗結果如表 4 所列。由表 4 可知，當只使用方法 1 時(提高 SNR 門檻值)，方位角的平均角度誤差可減少 0.45 度；另外，當只使用方法二時(修正 TDOA 超出合理範圍的處理方式)，方位角的平均角度誤差可減少 0.46 度；如果同時使用上述二種方法，則方位角的平均偵測誤差，更可以減少 0.59 度，而得到 3.43 度之平均角度誤差，這說明了前述二種方法能夠用以提升聲源方位估計的準確度。

5.3 角度偵測之實際實驗

線上實際測試是由二個男生分別坐在麥克風陣列周圍，仰角設定爲 –30°，而方位角分別設定爲 ±90°、±60°、±30°、0° 等方位，並且嘴唇與麥克風陣列中心的距離調整爲約 100 公分，然後依序唸出表 1 中的命令語音，以

表 3 不同位置下角度偵測之平均誤差與標準差

(a) 方位角測試角度 0,30,60,90

距離	仰角	平均角度誤差與標準差	方位角							
			90		60		30		0	
			方位角	仰角	方位角	仰角	方位角	仰角	方位角	仰角
100 Cm	0	AVG	**8.00**	**2.75**	**3.21**	**2.09**	**3.77**	**1.93**	**1.96**	**2.92**
		STD	3.18	0.97	1.40	0.57	0.85	0.76	0.71	1.07
	30	AVG	**4.25**	**2.00**	**2.50**	**0.80**	**3.52**	**0.38**	**2.77**	**1.66**
		STD	2.10	2.04	1.11	1.03	0.66	0.79	1.36	1.09
	-30	AVG	**4.92**	**0.97**	**1.96**	**0.69**	**2.84**	**0.59**	**1.60**	**2.44**
		STD	2.49	1.14	1.76	0.65	1.21	0.62	1.11	1.00
150 Cm	0	AVG	**11.39**	**1.78**	**2.82**	**1.07**	**4.03**	**1.99**	**2.02**	**2.85**
		STD	4.12	1.45	1.55	0.64	1.59	0.98	0.90	1.32
	30	AVG	**13.72**	**3.66**	**2.90**	**2.87**	**2.93**	**1.47**	**3.38**	**1.89**
		STD	11.10	5.30	3.72	3.96	1.52	0.88	1.63	2.20
	-30	AVG	**7.90**	**2.58**	**3.26**	**1.02**	**1.67**	**0.95**	**1.60**	**2.44**
		STD	4.71	3.45	3.90	1.36	1.62	1.31	1.11	1.00

(b) 方位角測試角度-30,-60,-90

距離	仰角		方位角					
			-30		-60		-90	
			方位角	仰角	方位角	仰角	方位角	仰角
100 cm	0	AVG	**2.04**	**2.13**	**2.14**	**2.30**	**7.51**	**4.20**
		STD	0.79	0.99	1.58	1.25	2.02	1.31
	30	AVG	**2.21**	**1.31**	**4.65**	**0.98**	**3.29**	**2.22**
		STD	2.12	1.96	3.17	1.71	1.71	1.47
	-30	AVG	**2.17**	**2.43**	**1.64**	**3.77**	**5.42**	**3.27**
		STD	1.57	1.04	1.93	0.86	1.92	1.24
150 cm	0	AVG	**2.36**	**2.40**	**2.86**	**3.21**	**7.56**	**3.18**
		STD	1.32	1.49	1.36	1.22	2.17	1.23
	30	AVG	**3.47**	**2.41**	**4.77**	**3.59**	**6.03**	**1.72**
		STD	1.63	2.03	5.38	4.40	4.38	2.32
	-30	AVG	**2.18**	**2.48**	**2.06**	**2.49**	**7.72**	**3.86**
		STD	2.70	2.46	2.43	2.85	5.82	3.14

表 4 VAD 門檻及 TDOA 邊界值處理之平均偵測誤差

	方位角		仰角	
	平均誤差	標準差	平均誤差	標準差
基準	**4.02**	2.88	**2.18**	1.93
方法 1	**3.57**	2.62	**2.07**	1.95
方法 2	**3.56**	2.87	**2.11**	1.91
同時使用	**3.43**	2.74	**2.08**	1.96

進行線上即時偵測，結果得到的平均角度誤差，如表 5 中所列。

將表 5 與表 3 中的數據作比較，以 ”兩人平均” 和 ”方位平均” 來看，線上實際測試與線外模擬測試的結果並沒有太大的差異。以 ”方位平均” 來看，實際測試與模擬測試的結果差距約 1 到 2 度，我們認爲這是由於實際測試只在距離 100cm、仰角 –30° 的各個方位角所偵測，所以方位角 90° 所偵測到的角度誤差較小，因而使得實際測試的角度誤差平均值看起來較好。若只將實際測試與模擬測試距離 100cm 且仰角 –30° 的數據相互比較，則可發現差異其實不大。

本系統進行線上實際測試時所使用的是筆記型電腦，其 CPU 爲 Intel Core 2 Duo T8300，記憶體爲 2GB，

程式執行時 CPU 使用率之最大時只達 13%，所以符合本論文目標，就是製作具有不錯的效能，同時計算量較少的聲源方位偵測系統。

表 5　線上實際偵測之平均角度誤差與標準差

		方位角							
距離	仰角	90		60		30		0	
100cm	-30	方位角	仰角	方位角	仰角	方位角	仰角	方位角	仰角
男 1	誤差	**4.40**	**1.61**	**2.50**	**4.56**	**2.52**	**2.43**	**1.24**	**2.56**
	標準差	4.73	1.44	2.57	0.51	1.62	1.02	1.14	1.52
男 2	誤差	**5.14**	**3.72**	**2.08**	**0.57**	**0.95**	**0.71**	**2.80**	**2.44**
	標準差	3.34	3.26	2.60	1.12	1.01	1.19	1.35	1.41
兩人平均	誤差	**4.77**	**2.66**	**2.29**	**2.56**	**1.74**	**1.57**	**2.02**	**2.50**
	標準差	4.03	2.35	2.59	0.81	1.32	1.10	1.25	1.46
		方位角						方位平均	
距離	仰角	-30		-60		-90			
100cm	-30	方位角	仰角	方位角	仰角	方位角	仰角	方位角	仰角
男 1	誤差	**2.50**	**0.94**	**3.43**	**0.89**	**2.88**	**0.23**	**2.78**	**1.89**
	標準差	0.81	0.71	1.80	1.20	0.33	0.43	1.86	0.98
男 2	誤差	**3.83**	**1.46**	**3.04**	**1.91**	**3.06**	**2.34**	**2.99**	**1.88**
	標準差	2.16	1.22	2.45	1.13	0.79	2.36	1.96	1.67
兩人平均	誤差	**3.17**	**1.20**	**3.24**	**1.40**	**2.97**	**1.28**		
	標準差	1.49	0.96	2.13	1.17	0.56	1.39		

結論

我們研究製作了一個三維方位的聲源偵測系統，相對於許多系統使用四個以上之麥克風，我們系統僅使用三個麥克風組成的正三角形之平面陣列。

關於 VAD 的處理，我們對三個麥克風的信號音框計算出頻譜亂度後，經公式(3)的濾波處理、再作 SNR 驗證，以判斷各音框是語音音框或非語音音框。這裡的語音活動偵測方法，加入了新的想法，實驗後也驗證可以把大部分的語音音框分辨出來。關於 TDOA 的估計，我們基於廣義交互相關的方法(即公式(5)、(6) 與(7))，在實作上以 FFT 計算交互功率頻譜來作逼近；此外，爲了排除某些頻帶的交互功率頻譜值太小而發生相位值的不穩定，我們研究以同步式相位複製來改善這個問題；再者，由於以反向 FFT 計算廣義交互相關函數(即公式(5))，所估計出的 TDOA 是整數個樣本點，因此我們研究以拋物線內插法來提高 TDOA 估計值的精確度。在聲源方位的計算上，我們將 TDOA 估測值向量與事先建立好的 ACMP 中的向量作比較，以找出其中距離最小的 ACMP 向量所對應的空間方位角度。此外，我們提出了一種不等值投票機制及其權重設定方式，以便對於一個命令語音的數個音框，作綜合式的聲源方位偵測。

關於系統效能的測試，先進行了線外模擬測試，在 42 個不同的空間位置播放事先預錄的音檔，從角度偵測的結果可以看出，水平方位角的平均偵測誤差可減少至 3.43 度，而垂直仰角的平均偵測誤差可減少至 2.08 度。另外，線上實際偵測的結果與線外模擬偵測的一致。並且由線上實際偵測可以看出，我們系統執行時所佔用的 CPU 資源不多，速度比即時處理的要求快許多。

參考文獻

[1] R. O. Schmidt, “Multiple emitter location and signal parameter estimation”, IEEE trans. Antennas and Propagation, Vol. AP-34, No.3, pp. 276-280, 1986.

[2] B. Kwon, G. Kim and Y. Park, “Sound source localization methods with considering of microphone placement in robot platform”, The 16th IEEE Int. Symposium on Robot and Human Interactive Communication, pp. 127-130, 2007.

[3] X. Lv and M. Zhang, “Sound source localization based on robot hearing and vision”, Int. Conf. Computer Science and Information Technology, pp. 942-946, 2008.

[4] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay”, IEEE trans. Acoustics, Speech and Signal Processing, Vol. 24, No. 4, pp. 320-327, 1976.

[5] D. V. Rabikin, R. J. Renomeron, A. Dahl, J. C. French, and J. Flanagan, "A DSP implementation of source location using microphone arrays", in Proc. 131st Meeting of the Acoustical Society of America, pp. 88-99, 1996.

[6] P. Renevey and A. Drygajlo, “Entropy based voice activity detection in very noisy conditions”, European Conference on Speech Communication and Technology (EuroSpeech), 2001.

[7] M. Omologo and P. Svaizer, “Use of the crosspower spectrum phase in acoustic event location”, IEEE trans. Speech and Audio Processing, Vol. 5, No. 3, pp. 288-292, 1997.

[8] A. Brutti, M. Omologo, and P. Svaizer, “Comparison between different sound source localization techniques based on a real data collection”, Hands-Free Speech Communication and Microphone Arrays, pp. 69-72, 2008.

[9] D. O’Shaughnessy, Speech communications: human and machine, IEEE Press, Piscataway, NJ, 2000.

[10] Primo, EM-147, http://www.primo.com.sg/ourproducts-jap-microphone.html

[11] National Instruments, NI-9215 DAQ, http://sine.ni.com/nips/cds/view/p/lang/zht/nid/13881