

國立交通大學
電機與控制工程研究所

碩士論文

以 OMAP5912 實現
語者方位判定於輪式機器人平台



Implementation of DOA for Speech
Using OMAP5912 on a Wheeled Robot

研究生：孫藍蕙

指導教授：胡竹生 博士

中華民國九十五年七月

以 OMAP5912 實現語者方位判定於輪式機器人平台

研究生：孫 藍 蕙

指導教授：胡 竹 生 博士

國立交通大學電機與控制工程研究所碩士班



本論文以聲源方位判定演算法(DOA, Direction of arrival)判斷語者方位，其中結合了聲源啟動偵測系統(VAD, Voice Activity Detection) 以及聲源方位判定系統(GCC, Generalized Cross-Correlation)應用於機器人聽覺系統，由這兩個聲音處理系統整合成為 DOA 演算法。本理論預期要以 OMAP5912 作為機器人的核心系統，在其上架設雙聲道麥克風，偵測環境中的聲源以及噪音源。VAD 系統主要針對環境偵測是否存在語音，並且在偵測語音後，紀錄語音頻帶，提出新的係數調整演算法，傳送到下級 GCC 系統計算出聲源角度，可適應即時環境噪音之改變。因為 OMAP5912 平台包含有 DSP5510 系統核心以及 ARM 系統核心，所以論文初期以 PC 驗證演算法，進而以 DSP 平台 DSK5510 做驗證，後期預計整合進 OMAP5912 中並實際與輪式機器人整合，為系統增加智慧型功能，可應用於多樣環境追蹤人聲以及環境監測等功能。

Implementation of DOA for Speech Using OMAP5912 on a wheeled Robot

Student : Lai-Hui Sun

Advisor : Prof. Jwu-Sheng Hu

Institute of Electrical and Control Engineering
National Chiao-Tung University



ABSTRACT

This thesis proposes a DOA algorithm (Direction of arrival) used in source direction detection system-GCC (Generalized Cross-Correlation) and voice activity detection (VAD) system for applications in robot where speech reception is contaminated by various noise and sound sources. Two microphones connected with TI OMAP5912 were made for the implementation platform. To cope with the changes of noise and environmental characteristics, the adaptive VAD system with on-line noise channel calibration is designed and implemented. The VAD result is used to ensure the correct timing for on-line calibration and choose the voice's frequency without consulting the noise's. The algorithms are implemented on OMAP platform with DSP and ARM dual core. To verify the effectiveness of the proposed system, the speech signal and angel of voice are combined with robot system that demo the movement of robot. The experimental results show that the algorithms are able to detect the direction of source and reduce the error of angel.

誌謝

對於本論文的完成，首先要感謝我的指導教授胡竹生博士。猶記得大四下開始跟著老師做專題，接著協助老師當課程助教，如此繁瑣卻又蘊藏著寶貴經驗的實作課程中，老師總是親身指導著我，一起研究課程內容，一起討論課程程式，教導我作學問需有的態度和方法，亦耳濡目染的從老師身上學習到研究應有的積極精神，特別是論文研究中，遭遇瓶頸時，那不分晝夜的研究與討論，種種感恩之情，難以用文字描述詳盡。在此，對於老師多年來的指導，誠摯的致上最真摯的謝意！

另外，也感謝在這許多實驗室日子中，在我身邊共同努力奮鬥的夥伴：首先要感謝佳興學長還有 alphas 學弟(崇維)，他們對於我研究生的日子有著舉足輕重的地位，不論是聲音理論、TI 競賽、系統架設，或是 OMAP 的問題，如果沒有他們，也不會有今天種種成果！感謝价呈學長的時時鼓勵以及加油打氣、陪我打桌球；感謝維瀚學長在協助解決聲音理論問題之外，還與我分享許多旅遊經驗，讓我大開眼界；感謝立偉學長帶著我勇闖 Hands on 比賽，讓我更加鍛鍊自己；感謝宗敏學長平時的關心以及數不清的協助，比賽忙碌的時候還義不容辭的幫大家到天亮；感謝鏗元學姊打開實驗室料理的大門；還有，講義氣卻悶騷的螞蟻，什麼都會的工具箱永融，常常在我後面搞笑的恆嘉，一起有氧健身的佩靜，帥氣卻偷吃步的耀賢，成熟的真男人朱木，認真為生活打拼的楷祥，可愛的弘齡；大家都是我實驗室生活中不可或缺的一員，少了誰，『實驗室就不實驗室了』！

最後我要感謝我的父母家人，他們供給我無憂無慮的生活，讓我能安心研究，也傾聽我所有的心聲，尤其是我的男友柄逢，在我忙碌的日子中，更能體諒我、協助我，甚至犧牲自己寶貴的時間，因為有你們，我才能順利完成學業，最後，再此感謝每一位曾經協助過、關心過我的人，謝謝你們。

目錄

摘要	ii
ABSTRACT	iii
誌謝	iv
目錄	v
表列	vii
圖列	viii
第一章 緒論	1
1.1 研究動機	1
1.2 研究內容與目標	1
1.3 文獻回顧	3
1.4 論文內容概述	6
第二章 語者方位判別系統	7
2.1 語音活動偵測系統	7
2.1.1 系統架構	9
2.1.2 電腦 MATLAB 浮點模擬結果	10
2.2 聲源定位系統	13
2.2.1 系統架構	13
2.2.2 電腦 MATLAB 浮點模擬結果	15
2.3 結合語音活動偵測之適應性頻帶選擇架構與流程	17
2.3.1 傳統固定頻帶聲源定位問題	17
2.3.2 適應性頻帶調整方法	17
2.3.3 系統架構	18
第三章 基於 DSK5510 實現定點語者方位判別系統	20
3.1 簡介	20
3.2 原理	20
3.2.1 定點化規格考量	20
3.2.2 記憶體配置	22
3.2.3 方位判別系統運算修正	23
3.3 定點實驗結果	26

第四章	實驗平台	28
4.1	平台架構說明	28
4.2	硬體環境	30
4.2.1	OMAP5912 處理器	30
4.2.2	麥克風放大電路	33
4.2.3	海星號輪式機器人平台運動控制	34
4.3	軟體環境	37
4.3.1	Linux 作業系統	37
4.3.2	ARM 處理器以及轉換編譯器	38
4.3.3	數位訊號處理器	39
4.3.4	程式組合開發軟體 (Code Composer Studio)	40
4.4	實驗環境與流程	41
第五章	實驗結果與分析	43
5.1	機器人平台對聲源偵測影響實驗	43
5.2	電腦浮點運算實驗結果	45
5.3	DSK5510 定點運算實驗結果	47
第六章	結論	49
6.1	研究成果	49
6.2	未來展望	49
參考文獻	50



表列

表 1 DSP 定點化規格表	21
表 2 數位類比輸入輸出取樣頻率設定對照表	22
表 3 TMS320C5510 特性表	39
表 4 浮點電腦端辨識率比較表	46
表 5 定點 DSP 端辨識率比較表	47



圖列

圖 1	海星號系統架構圖	2
圖 2	機器人 SIG 及 ROBITA	4
圖 3	機器人 KISMET	4
圖 4	VAD 演算法流程圖	9
圖 5	有經過長時音框處理 $N=6$ ，上：LTSD 與 Γ 關係圖，下：VAD 模擬結果	11
圖 6	無經過長時音框處理 $N=0$ ，上：LTSD 與 Γ 關係圖，下：VAD 模擬結果	11
圖 7	不同環境參數對於 VAD 結果的影響圖—雜訊能量調整較小	12
圖 8	不同環境參數對於 VAD 結果的影響圖—雜訊能量調整較大	12
圖 9	聲音訊號在麥克風 I 和 J 之間的時間延遲	14
圖 10	無雜訊判斷聲源方位	15
圖 11	聲源加入雜訊判斷方位	16
圖 12	適應性聲源定位系統架構圖	18
圖 13	取樣頻率參數設定	21
圖 14	乒乓暫存器機制圖	23
圖 15	Γ 公式內插圖解	24
圖 16	浮點 MATLAB(左)與定點 DSP(右)的原始語音輸入能量	26
圖 17	浮點 MATLAB(左)與定點 DSP(右)的 LTSD 與 GAMMA 關係結果比較	26
圖 18	浮點 MATLAB 的 VAD 結果	27
圖 19	定點 DSP 的 VAD 結果	27
圖 20	海星號系統架構圖	28
圖 21	發展平台架構圖	29
圖 22	OMAP5912 內部架構	30
圖 23	OMAP5912 發展平台	31
圖 24	麥克風放大電路	33
圖 25	麥克風系統架構圖	33
圖 26	海星號機器人平台實體圖	34
圖 27	輪式機器人運動模式示意圖	35
圖 28	馬達控制流程圖	36
圖 29	CCS 整合發展環境介面	40
圖 30	海星號主體實體圖	41
圖 31	實驗環境	41
圖 32	實驗架構圖	42
圖 33	機器人旋轉時，麥克風錄製人聲的語音分析	43
圖 34	圖 33 經過 VAD 系統後的聲音訊號	43
圖 35	針對語音部份分析結果	44

圖 36	針對恆定出現馬達噪音做分析	44
圖 37	實驗環境實際照片 1	45
圖 38	實驗環境平面關係圖 1	45
圖 39	實驗環境實際照片 2	47
圖 40	實驗環境平面關係圖 2	47



第一章 緒論

1.1 研究動機

機器人於家庭生活中逐漸扮演重要的角色，如何配置合適且靈敏的聽覺系統於機器人上更是一個重要的研究方向。因此，設計出配合家庭需求，或是應答型機器人，不可或缺的重要元素便是智慧型聽覺系統。

首先，應答型機器人需要了解聲源的來源，針對這項需求，必須克服環境其他噪音影響的因素，所以必須構思具有偵測聲源起動裝置的智慧型聽覺系統。另外，也需要判斷出聲源方位，加強機器人跟隨使用者的基本功能。

OMAP5912 處理器包含 RISC(ARM)與 DSP 雙核心處理器，透過這兩種類型的處理器的合作，讓一些演算法得以在機器人系統上快速執行。在 OMAP 的 ARM 上所執行的作業系統讓更多類型的程式得以執行，讓機器人系統整合更多周邊；DSP 則可以輔助 ARM 讓耗時間的演算法加速運算，減少運算負擔。

1.2 研究內容與目標

在架設實驗平台方面，首先將針對 OMAP 處理器修後的 Linux 核心安裝到平台上，並建立檔案系統。其次，透過 DSP/BIOS Link 建立 ARM 與 DSP 間的通訊。在演算法方面，針對海星號機器人環境設計聲源定位系統，利用時間延遲偵測 (Time Delay Estimation, 簡稱 TDE)，對於常使用的量測標準 (Generalized Cross-Correlation, 簡稱 GCC)，提出新的頻帶調整演算法，針對有興趣的頻域訊號做處理，減輕運算量以及運算時間，並結合語音活動偵測演算法 (Voice Activity Detection, 簡稱 VAD) 方法，藉此提升聲源定位準確度。

本論文目標可分為：

1. 選定真人語音活動偵測及適應性的聲源定位系統，用電腦做初步驗證。
2. 將演算法實現於 DSP5510 核心平台，作即時性且定點化的處理。
3. 以德州儀器 OMAP5912 為平台，運用 OMAP 的 DSP 晶片對聲音作前置處理，搭配 ARM 核心嵌入式系統，結合 PC 端運用語音偵測及適應性聲源定位演算法進行即時偵測，驗證 OMAP 系統以及其與週邊傳輸合作功能。其系統架構圖如圖 1 所示。

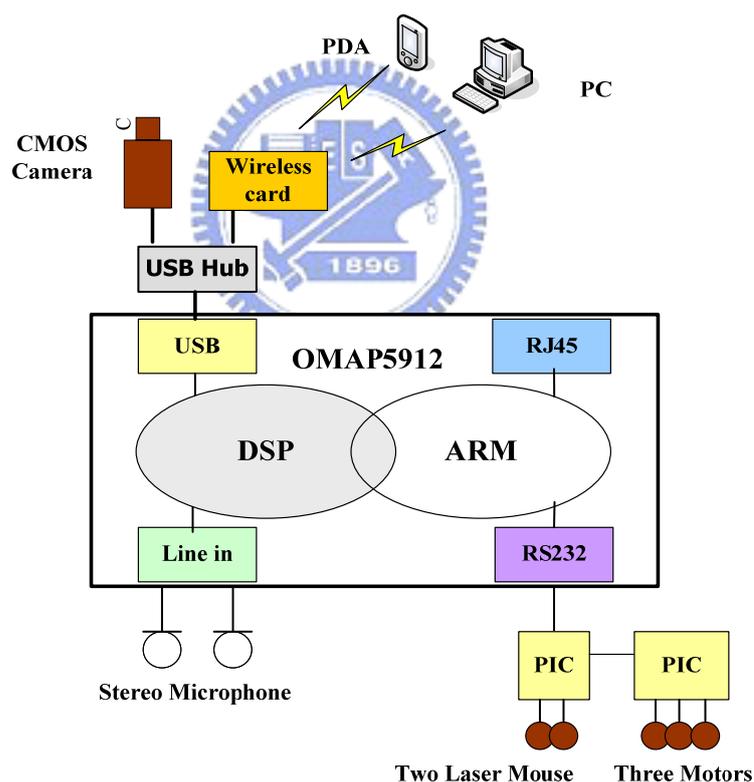


圖 1 海星號系統架構圖

1.3 文獻回顧

即使如麥克風系統已有相當多的研究，在機器人聽覺系統上的應用或研究仍然十分有限。機器人聽覺一直到近幾年才逐漸受到重視，其中有大部份是引用人類聽覺方面的研究成果，如 Chun and Caudell [1]嘗試使用人類中腦 (mid brain) 之 Inferior Colliculus 結構，以及 HRTF (Head Related Transfer Function) 的資料，運用影像處理之技巧來尋找聽覺的規則。Schauer and Gross[2] 基於腦神經學上的發現，利用 ITD (Interaural Time Difference) 與 IID (Interaural Intensity Difference) 的訊號來作 360°聲源方向的判斷。

除了以人為主的聽覺研究外，亦有從其他生物的角度來研究，例如 Webb 及 Harrison [3]研究以蟋蟀的角度來設計一個聲音及運動結合 ITD 與 IID 及 HRTF 資料來分辨 360°的聲源方向。

機器人聽覺及其整合在全世界較有名的研究團隊有：

(a) **SIG (日本 Kyoto University, Japan Science and Tech. Corp., 及 Sony Computer Science Laboratories, Inc.)**

以 Kazuhiro Nakadai 為首，此團隊設計仿效真人的頭、耳及視訊，使用簡單的雙麥克風 Beamforming 方法來分離不同角度的聲源[4-9]。Nakadai 等人所使用的方法是基於 HRTF 及 ITD 與 IID 的資料，以及一個基於假設的 reasoning engine 來分離部圖角度的語音源。為求與人類近似，Nakadai 等人並發展了一個 Robot Head 稱作 SIG [10][11] (見圖 2)。

(b) **Project ROBITA (Real-world Oriented BI-modal Talking Agent)**

ROBITA 是日本 Waseda University 發展之機器人，由 Matsusaka [12-14] 等人共同研發。其重點在實現一個與人類自然溝通的介面，包含視覺、聽覺、及對談。ROBITA 在 sound source separation 並無深入研究，在聲源方向判斷上亦採用雙麥克風及 HRTF 的觀念如 Nakadai 等人的做法，整個計畫整合的範圍亦相當大，但在 spatial acoustics 方面著墨甚少，圖 2 為 ROBITA 在多人

對談時的照片。由其發表的結果來看(在 <http://www.tk.elec.waseda.ac.jp/robota/> 的影片)，背景噪音很低且交談訊號不重複。

(c) **Kismet at MIT**

Kismet[15]是一個機器人系統著重於感知與社交性方面的研究(見圖 3)，其中語音交談自然是研究的重點，但 Kismet 在聽覺方面幾乎沒有放置重點，僅用了單一麥克風。然而其在感知語言所代表的喜怒哀樂上有相當豐富的成果。

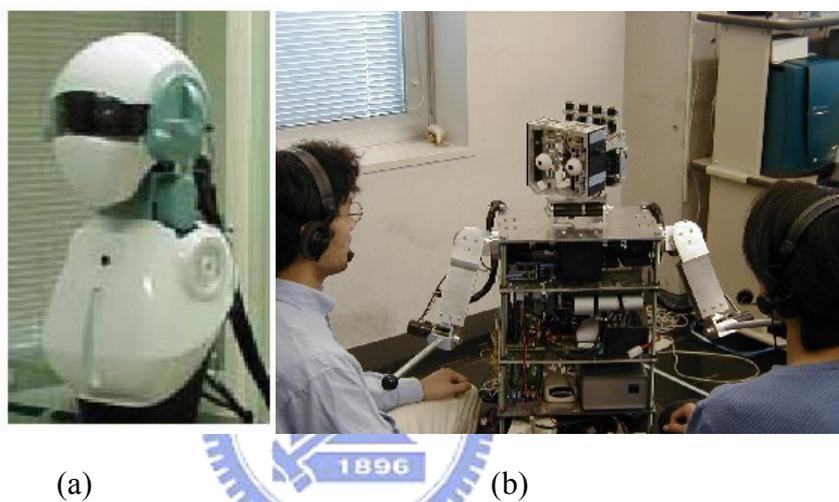


圖 2 機器人 SIG 及 ROBITA

(a) Nakadai 等人所研發之 Acoustic Head SIG [16][17] (b) ROBITA

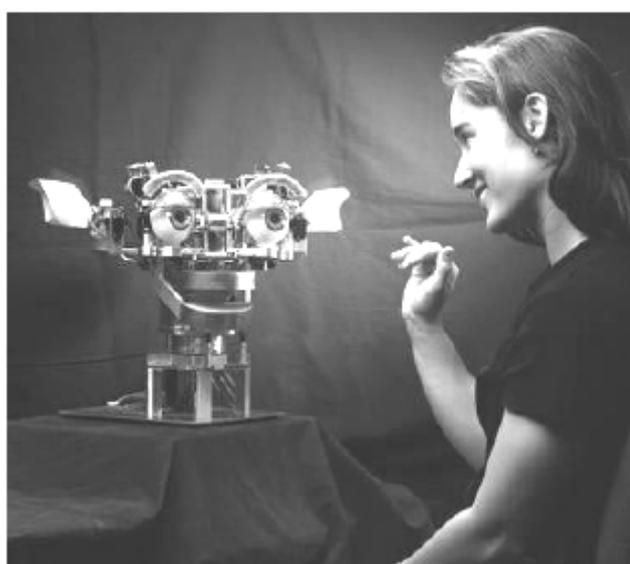


圖 3 機器人 Kismet

(d) **Cog at MIT**

Cog project 由國際知名的 Rodney Brooks 教授領導的團隊所組成，的主要研究方向為 Humanoid Intelligence 及 Humanoid Interaction，曾經展示機器人由聽取鼓聲的節奏來控制擊鼓的動作，以達成預先教授的節奏。此外大部分的研究均著重於 vision information。

(e) **SmartHead**

日本東京大學的 Ando 教授所帶領的團隊研發具備四個麥克風的機器人稱 SmartHead[18][19]。Ando 教授的重點在研究 sensor fusion 及 sensory-motor 整合。其聽覺方面使用類似於影像處理的光流法，計算四個麥克風的 spatiotemporal gradient，並利用此資訊來產生聲源方向及距離的資訊。此方法相當的直覺(heuristic)，其理論依據較弱，全仰賴後續整合及訊號處理的解釋，來達成目標。

從上述的文獻整理來看，機器人聽覺方面的研究很少利用到 Array Signal Processing 方面的技術，而以擬人化的研究方向居多。所以本論文也採用模擬雙耳，架設雙聲道麥克風系統。

1.4 論文內容概述

本篇論文包含了三個主要的部分，分別是即時性演算法的設計、實驗平臺架構與實現及即時性演算法的驗證。結果將以下面章節描述三個主要部分：

第二章：介紹語音活動偵測演算法、聲源定位系統以及整合兩個系統的設計

第三章：介紹 DSK5510 定點化實現的原理及結果

第四章：介紹實驗平台架構

第五章：演算法在 PC 與 DSK5510 與實驗平台的驗證結果與分析

第六章：結論



第二章 語者方位判別系統

2.1 語音活動偵測系統

語音活動偵測 VAD(Voice Activity Detection)是用來判定是否有真人語音，近年來已廣泛用於通訊上達到節省能量耗損的目的。若用於語音辨識方面是屬於語音辨識的前處理，對辨識結果的影響很大，精確的語音活動偵測可降低噪音影響並提高辨識率。傳統的語音活動偵測大多使用語音能量或過零率（zero-crossing rate）等資訊來判別，本節將介紹的語音活動偵測演算法是使用長時間語音資訊（long-term speech information）來判別是否有真人語音[20]。

最常見的判定真人語音資訊為語音能量和過零率，雜訊及氣音的過零率都很高，語音能量都較低。例如，由歐洲電信標準協會(ETSI)所制定用於 GSM(Global System for Mobile Communications) 系統中的 AMR (Adaptive Multi Rate) VAD 判定方法就採用了能量、週期、頻譜失真等三種參數來判定[21][22]。另外由國際電信聯盟 (ITU) 所制定的 G.729-VAD 採用了全頻帶能量差、低頻帶能量差、頻譜失真和過零率四種參數來判定[23][24]。論文中使用的 VAD 演算法是使用長時間語音的資訊而非傳統瞬間音框 (instantaneous frame) 資訊，針對長時間語音資訊，定義出下列定義。

■ 長時頻譜能量封包(Long-Term Spectrum Envelope，簡稱 LTSE)

若 $x(n)$ 為一段包含有雜訊的語音訊號，而 $X(k,l)$ 代表著 $x(n)$ 中第 l 個音框第 k 個頻率的值，那麼 N 階的 LTSE 定義為：

$$LTSE_N(k,l) = \max_{j=-N}^{j=+N} \{X(k,l+j)\} \quad (2-1)$$

其 $LTSE_N(k,l)$ 代表的意義為，從第 $l-N$ 個音框(frame)到第 $l+N$ 個音框，這 $2N+1$ 個音框分別對其取頻譜絕對值 (Amplitude Spectrum) 後，在第 k 個頻率下，這

2N+1 個頻域絕對值內的最大值。而 LTSE 則代表了長時間語音資訊的意義，因為 LTSE 不只是對單一音框取值，而是針對 2N+1 個音框取最大值，這樣的好處是不容易忽略某些字頭的子音或是摩擦音。除了 LTSE 外，為了判定是否為真人語音，必須定義另一項定義 LTSD。

■ 長時頻譜分布(Long-Term Spectral Divergence，簡稱 LTSD)

LTSD 的定義如 2-2 式：

$$LTSD_N(l) = 10 \log_{10} \left(\frac{1}{NFFT} \sum_{k=0}^{NFFT-1} \frac{LTSE^2(k, l)}{N^2(k)} \right) \quad (2-2)$$

其中 NFFT 代表了作 FFT (Fast Fourier Transform) 的點數，而 $N(k)$ 代表了雜訊的頻譜絕對值平均，定義如 2-3 式：

$$N_K(k) = \frac{1}{2K+1} \sum_{j=-K}^{j=K} X(k, l+j) \quad (2-3)$$

從 2-3 式可看出， $N_K(k)$ 代表在第 k 個頻率下，第 l 個音框及前後 K 個音框的頻譜絕對值平均， $X(k, l)$ 和先前定義一樣，代表現階段語音的頻譜絕對值。因此 LTSD 的意義為：現階段長時間語音的頻譜能量佔了雜訊頻譜能量的比例，換句話說判定是否為真人語音是用了現階段語音能量的大小來判定，而此能量大小包含了長時間語音資訊，並非只有單一音框資訊。當 LTSD 大於某個臨界值則判定為真人語音，反之則非真人語音，而此臨界值 γ 定義如下：

$$\gamma = \begin{cases} \gamma_0 & E \leq E_0 \\ \gamma_0 + \frac{\gamma_1 - \gamma_0}{E_1 - E_0} (E - E_0) & E_0 < E < E_1 \\ \gamma_1 & E \geq E_1 \end{cases} \quad (2-4)$$

其中 E_0 和 E_1 代表了在最乾淨和最吵雜的情況下之雜訊的能量，而 E 是指現階段雜訊的能量。 γ_0 和 γ_1 代表在最乾淨和最吵雜的情況下與 LTSD 比較的臨界值，因此 E_0, E_1, γ_0 和 γ_1 是先設定好的初始值。從 (2-4) 式可觀察出當現階段雜訊能量介於 E_0 和 E_1 時，則 γ 會依 $E - E_0$ 在 $E_1 - E_0$ 所佔的比例，作出 γ_0 的線性調整。

2.1.1 系統架構

而 VAD 演算法的流程如圖 4 所示：

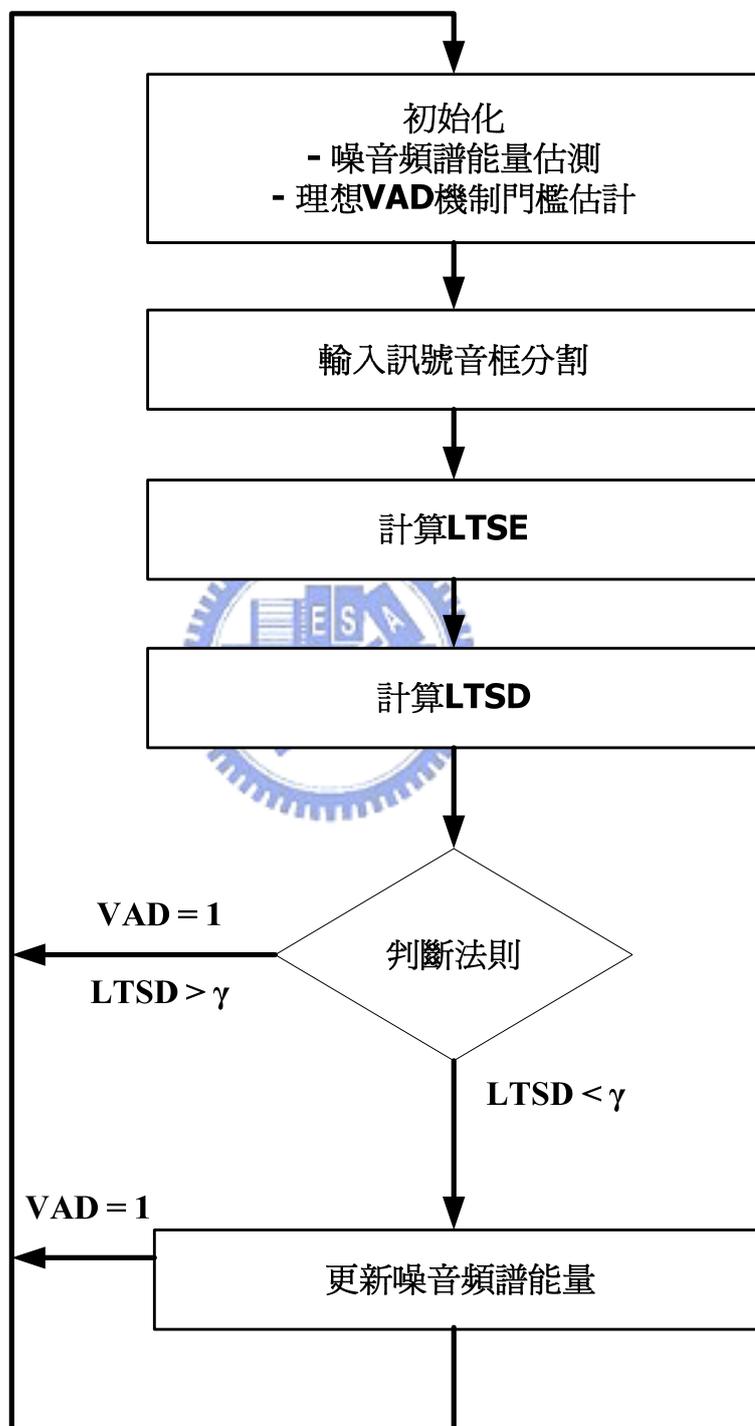


圖 4 VAD 演算法流程圖

1. 設定初始值 E_0, E_1, γ_0 和 γ_1 。
2. 將語音作切割，一個音框(frame)為 30ms，而音框和音框的交疊為 20ms。
3. 計算 LTSE 和 LTSD。
4. 將 LTSD 與 γ 作比較，若 $LTSD > \gamma$ 則判定為真人語音，若 $LTSD < \gamma$ ，則非真人語音，則更新雜訊頻譜絕對值平均 $N(k)$ 。

另外在更新雜訊頻譜絕對值平均 $N(k)$ 方面，並非完全的更新，而是利用了適應性訊號處理的觀念，定義如下：

$$N(k,l) = \alpha N(k,l-1) + (1-\alpha)N(k) \quad (2-5)$$

其中， k 代表頻率， l 代表音框，從 2-5 式可看出， $N(k)$ 的更新，除了有現階段 $N(k)$ 的資訊外，也包含了上一個音框的 $N(k)$ 資訊，而此權重 α 可依照環境自行調整。

2.1.2 電腦 MATLAB 浮點模擬結果

本章節將上述 VAD 演算法，用 Matlab 模擬，先以取樣頻率為 16k Hz，用單一麥克風錄製 12 秒的語音，並對其作 VAD 的判定。圖 5 展示出有做長時間語音資訊 VAD 的結果，也就是 2-1 式取 $N=6$ 。而圖 6 展示出沒做長時間語音資訊 VAD 的結果，也就是 2-1 式取 $N=0$ 。

圖 5 和圖 6 的上半部分展示出 LTSD 與 γ 的關係，變動較大的為 LTSD，從圖中可觀察出，沒做長時間語音資訊 ($N=0$) 的 LTSD 變動較大，這是因為每前進一個音框，LTSD 都會有新的值。比較兩張圖可發現，沒做長時間語音資訊 ($N=0$) 的 VAD 結果很容易將字頭字尾部分，並判定為非真人語音，而有做長時間語音資訊 ($N=6$) 的 VAD 就不容易忽略字頭字尾部分。

圖 7 和圖 8 的差異主要是環境參數中，對於雜訊能量的設定值，可以看出，雜訊能量的設定，可以影響雜訊門檻值的及時改變，也就是圖 7 的 VAD 效果較好，能適應環境雜訊變化。

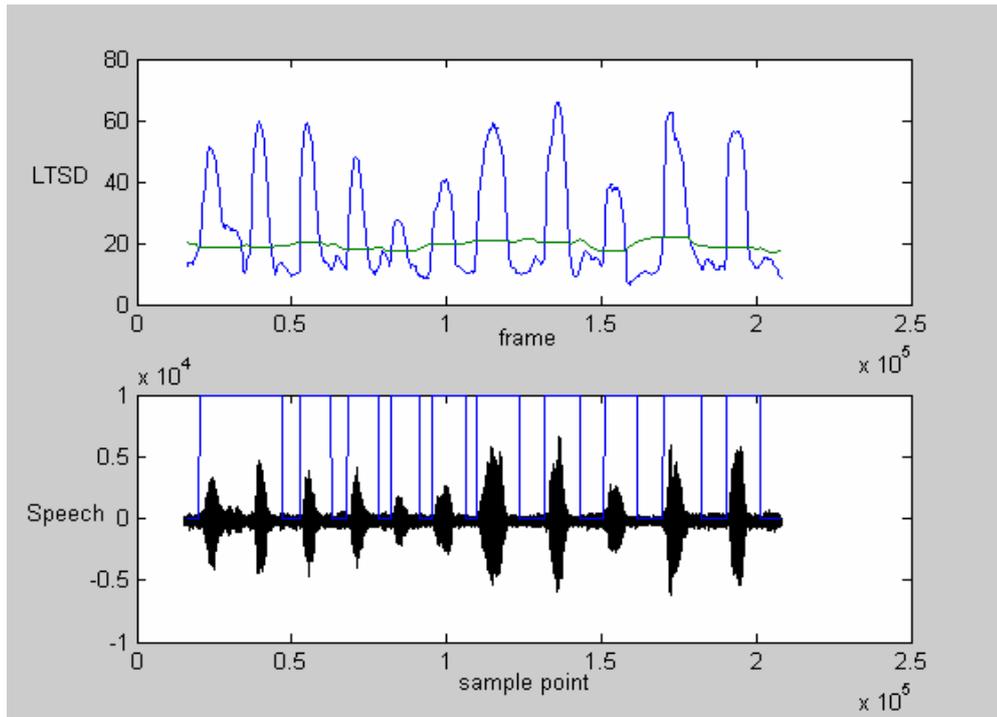


圖 5 有經過長時音框處理 $N=6$ ，上：LTSD 與 γ 關係圖，下：VAD 模擬結果

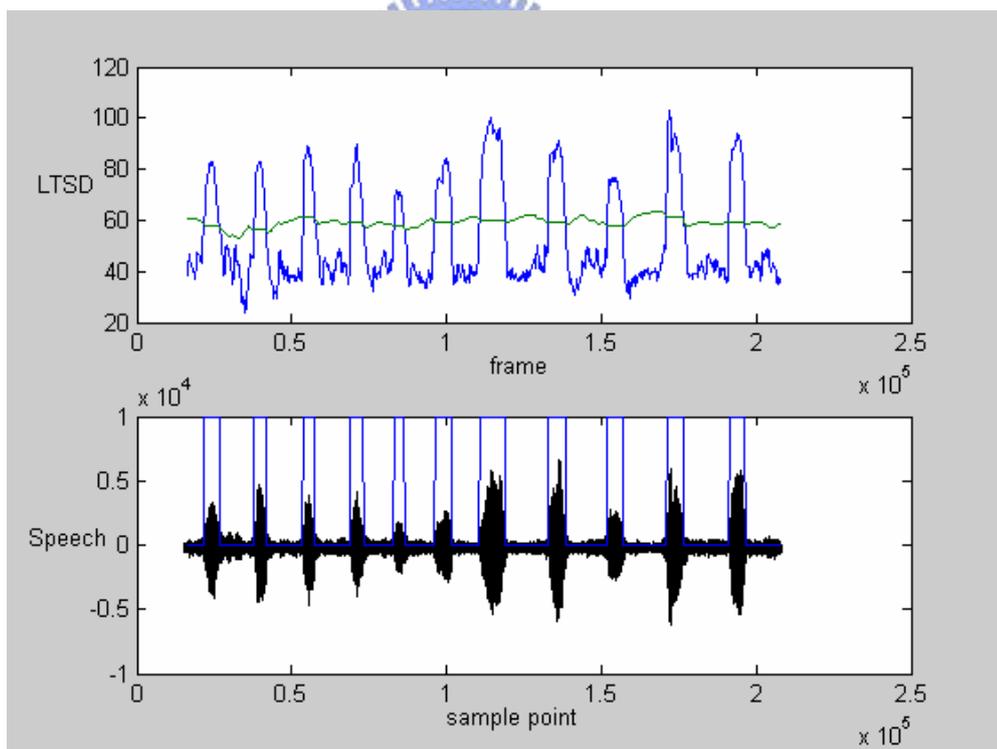


圖 6 無經過長時音框處理 $N=0$ ，上：LTSD 與 γ 關係圖，下：VAD 模擬結果

圖 5 以及圖 6 兩個差異可以很明顯的發現對於語音處理的效果上，圖 6 比圖 5 沕出的語音較窄，也就是容易忽略氣音的影響。

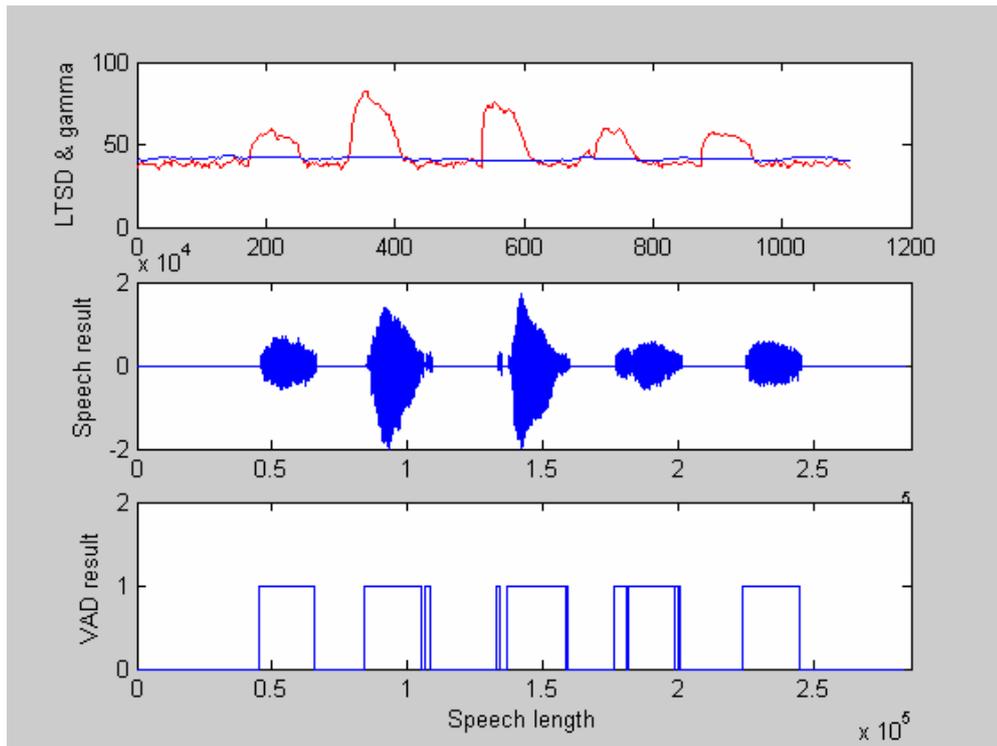


圖 7 不同環境參數對於 VAD 結果的影響圖—雜訊能量調整較小
 $\gamma_0=45, \gamma_1=5, \text{noise_e0}=50, \text{noise_e1}=75$

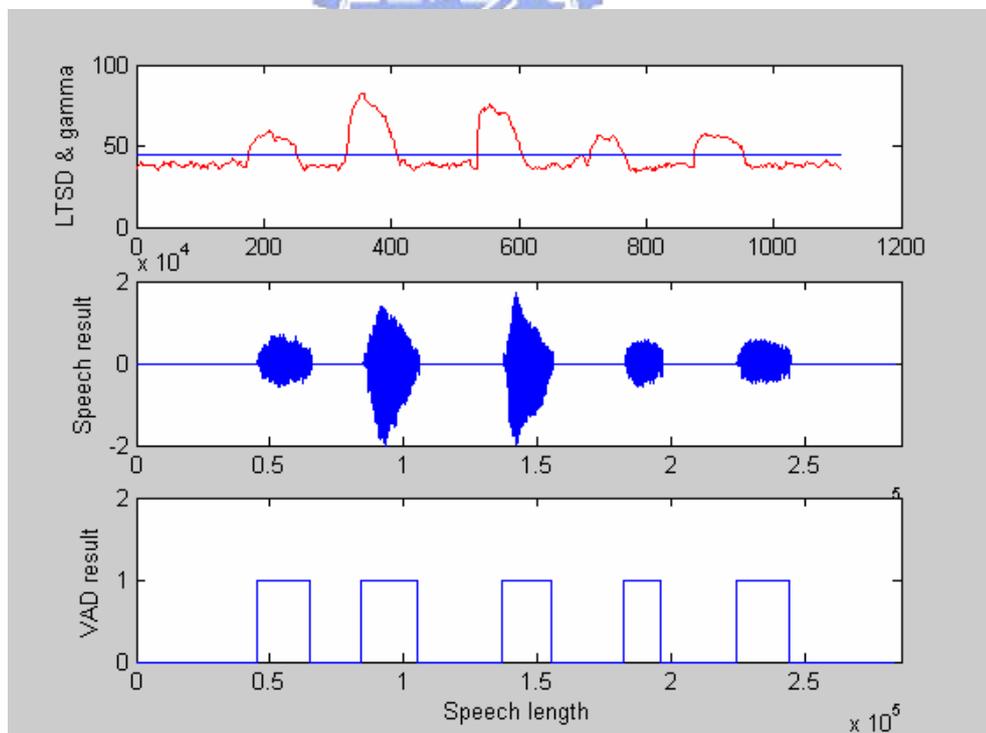


圖 8 不同環境參數對於 VAD 結果的影響圖—雜訊能量調整較大
 $\gamma_0=45, \gamma_1=5, \text{noise_e0}=100, \text{noise_e1}=120$

2.2 聲源定位系統

麥克風作定位的研究已有眾多文獻技術[25][26]，DOA(Direction of Arrival)的技術層面巨觀上可以分成兩大類，第一類是 TDE(Time Delay Estimation)，第二類是 Eigenspace Method。TDE 的方法都只有針對一個聲音來源來作方位的辨別，沒有辦法在同一時間內辨別出數個聲源的位置，如果要同時辨別數個聲源的方位就必須要使用 Eigenspace Method，但是 TDE 的好處是只需要使用兩顆麥克風，所需的硬體架構較為簡單，並且其運算量相較之下也比較符合即時應用的要求。我們利用 TDE 方法，只需要使用兩顆麥克風便能針對一個聲音來源做方位辨別，所需的硬體架構較簡單，其運算量也較符合即時應用的要求。

2.2.1 系統架構

聲源定位系統利用 TDE 方法，需要使用兩顆麥克風針對一個聲音來源做方位辨別，因為聲源從不同的方位傳遞到這兩顆麥克風都會造成不同的響應，此法便是利用聲音從空間中傳遞的特性所造成的時間差來作方位的估計。也因為只有使用兩顆麥克風，所以假設從聲源傳遞過來的聲音訊號是平面波是相當合理的。此兩顆麥克風所接收到的訊號可以 model 成為：

$$y_i(kT_s) = s_0(kT_s - \tau_{0i}) + \sum_{m=1}^M \alpha_m s_0(kT_s - \tau_{mi}) + n_i(kT_s) \quad i = 1, \dots, Q \quad (2-6)$$

其中 Q 代表麥克風數量， $s_0(kT_s)$ 代表語音訊號， $n_i(kT_s)$ 代表雜訊訊號， α_m 代表改變大小的 Scale value， τ_{0i} 和 τ_{mi} 則代表聲源到達第 m 個麥克風所花的時間。如下頁圖 9 所示， $\tau_{0i} = d_i / c$ 。此 model 包含了訊號在空間中回音反彈的成分，不過對於單一麥克風的接收訊號，可以簡以 $y_i(kT_s)$ 表示。

傅立葉轉換單一麥克風接收訊號為：

$$\hat{G}_i(\kappa, \omega) = \sum_m y_i[\kappa + m] \omega[m] e^{-j\omega m} \quad (2-7)$$

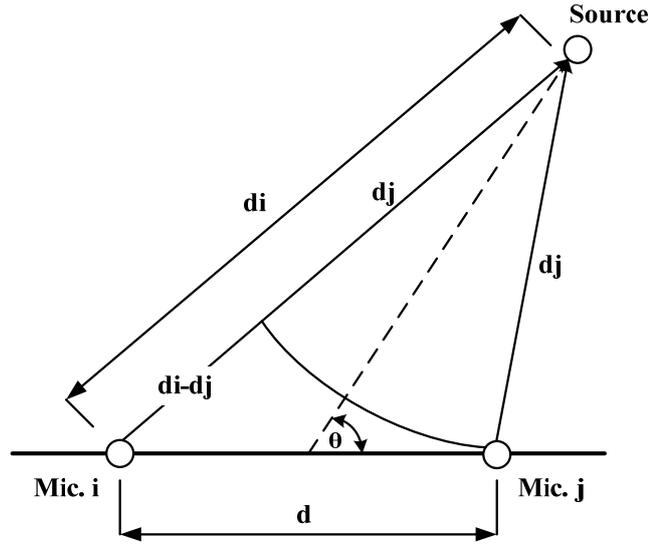


圖 9 聲音訊號在麥克風 i 和 j 之間的時間延遲

其中， κ 表示音框(frame)數目， m 代表一個 frame 內含的離散聲音點數， $\omega[m]$ 則是 rectangular window，負責切取聲音音框。接著計算兩顆麥克風的聲音音框訊號關係，關係如下：

$$\hat{G}_{ij}(\kappa, \omega) = \hat{G}_i(\kappa, \omega) \hat{G}_j^*(\kappa, \omega) \quad (2-8)$$

G_i 和 G_j 分別是兩顆麥克風的 FFT 訊號，此關係式可表示兩顆麥克風的相關關係。以 G_j 為主，計算 G_i 訊號超前或是延遲的 sample 點數。接著定義 Generalized cross correlation function 為：

$$\hat{R}_{ij}[\kappa, l] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_{ij}(\kappa, \omega) \hat{G}_{ij}(\kappa, \omega) e^{j\omega l} d\omega \quad (2-9)$$

$\psi_{ij}(\kappa, \omega)$ 是權重函數，GCC 可針對特定頻率、頻帶運算。計算的方法是針對特定頻率 ω 計算 GCC，對應 κ frame 以及延遲的 sample 數目 l ，每位移一音框點數計算一次，最大值的發生處即等效於聲音傳遞到這兩顆麥克風的時間差，

$$\hat{l}_{ij}[\kappa] = \arg \max_l \{ \hat{R}_{ij}[\kappa, l] \} \quad (2-10)$$

於此藉由時間差的資訊搭配餘弦定理便可以推測出聲源相對於這兩顆麥克風的方位了。

■ GCC 演算法流程解說

1. 先取得兩顆麥克風訊號的離散傅立葉轉換，並將訊號拆解成以 frame 為單位。
2. 計算兩顆麥克風之間的相關關係-cross correlation。
3. 針對有興趣的頻帶計算 GCC function。
4. 得到延遲或超前點數，進而計算花費時間、距離，找出聲源方向。

2.2.2 電腦 MATLAB 浮點模擬結果

本章節將上述 GCC 演算法，用 Matlab 模擬，先以取樣頻率為 16k Hz，用單一麥克風錄製 12 秒的語音，對其作 GCC 的判定。針對 GCC 演算法，主要能更將兩個聲道的聲音訊號加以比對，計算出延遲或是超前的點數，所以將單一麥克風錄製的聲音訊號複製成兩個聲道，刻意延遲四個音訊點數，然後將兩個聲道聲音送入 GCC 系統中，觀察最後輸出延遲點數是否貼進四個音訊點數。另外也針對訊號做了增加雜訊的處理，包括有雜訊、反射等等，分別對於角度判定有些影響，以下將逐步分析。

(a) 無雜訊語音訊號

將訊號送入系統後，計算出每個音訊音框(frame)的角度資訊，累加計算判斷結果，是 histogram 統計的結果。

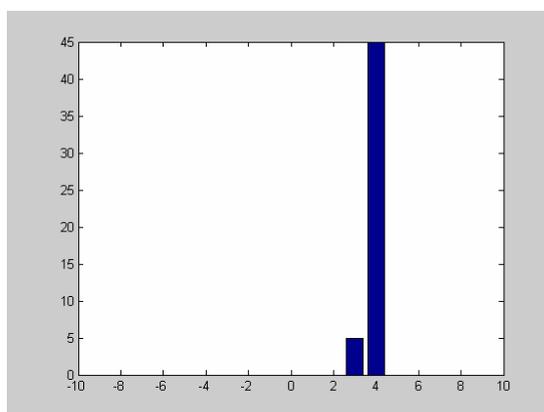


圖 10 無雜訊判斷聲源方位

所以由此得知，多數運算結果都指向 histogram，精確度有九成。這也是很合理的運算結果，因為 GCC 本來就會有所誤差，並非完全精確。

(b) 加入雜訊語音訊號，包含雜訊以及反射

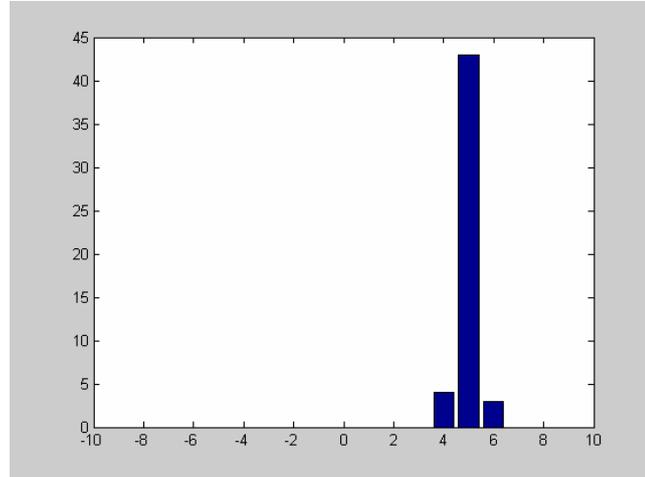


圖 11 聲源加入雜訊判斷方位

所以從圖 11 中可以看到，的確在加入雜訊後，GCC 的判斷就有比較多誤差。但是仍以延遲四個取樣點為主要運算結果，精確度掉到 86%。

雖然 GCC 的演算法只能針對簡易環境或是需求來達成，但是實際上在需求不高的況下，GCC 仍然也會有較好的判斷結果。

2.3 結合語音活動偵測之適應性頻帶選擇架構與流程

本方法針對機器人系統設計語音起動裝置以及聲源方向定位，利用時間延遲偵測(Time Delay Estimation, TDE)，以 GCC(Generalized Cross-Correlation)做方位判定，但結合了語音活動偵測 VAD(Voice Activity Detection)方法，提出新的頻帶調整演算法，藉此提升聲源定位準確度。

2.3.1 傳統固定頻帶聲源定位問題

在 2.2 節中提到 Generalized Cross-Correlation 定義為：

$$\hat{R}_{ij}[\kappa, l] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_{ij}(\kappa, \omega) \hat{G}_{ij}(\kappa, \omega) e^{j\omega l} d\omega \quad (2-9)$$

$\psi_{ij}(\kappa, \omega)$ 是權重函數，GCC 可針對特定頻率、頻帶運算。計算的方法是針對特定頻率 ω 計算 GCC，對應 κ frame 以及延遲的 sample 數目 l ，每位移音框點數計算一次，最大值的發生處即等效於聲音傳遞到這兩顆麥克風的時間差點數，

$$\hat{l}_{ij}[\kappa] = \arg \max_l \{ \hat{R}_{ij}[\kappa, l] \} \quad (2-10)$$

通常而言，GCC 的 $\psi_{ij}(\kappa, \omega)$ weighting function model 頻帶皆為固定，並不會自動的變動。所以在固定 $\psi_{ij}(\kappa, \omega)$ model 的情形下，選擇針對特定頻率 ω 計算 GCC，以人為因素考量聲源頻帶分布，比如說去除較高頻或是較低頻的聲音資訊，以中間頻帶分步作為計算，藉以減輕所有頻帶計算的計算量。

如此在選擇頻帶上容易因人為直觀判斷頻率而誤判聲源或雜訊，且在選擇頻率上也並無量化依據，這也導致聲源定位處理過程中的雜訊誤判結果或是遺失重要聲音資訊的事情發生。

2.3.2 適應性頻帶調整方法

GCC 的定義延伸出相當多的變化，也是推動此類方法持續演進的關鍵技術，

這類方法都是轉換到頻率域，並且定義出 $\psi_{ij}(\kappa, \omega)$ 為 Weighting function，進而增加時間差估計的正確性。比如說調整 $\psi_{ij}(\kappa, \omega)$ weighting function 來配適不同的環境，這對於 GCC 的改善也有幫助。

在此所定義之適應性頻帶調整指的則是能根據輸入信號，用訊號處理的技巧來適應性地調整 GCC 的頻帶選擇，讓 GCC 效果更能適應現在環境，如不同場景或是適應吵雜的環境等等。

因此，針對 GCC 頻帶選擇的依據是前端 VAD 的結果而來，根據前端 VAD 篩選語音資料後，將結果記錄下來，分析辨別出語音或是雜訊，將雜訊頻域能量記錄下來，擇取語音能量超過雜訊能量的數個重要的頻帶，扣除語音頻帶中非人聲的雜訊頻帶，轉給 GCC，用以調整 $\psi_{ij}(\kappa, \omega)$ ，使其在頻帶的選擇上能更有依據以及去除雜訊的影響。

2.3.3 系統架構



系統流程圖如下：

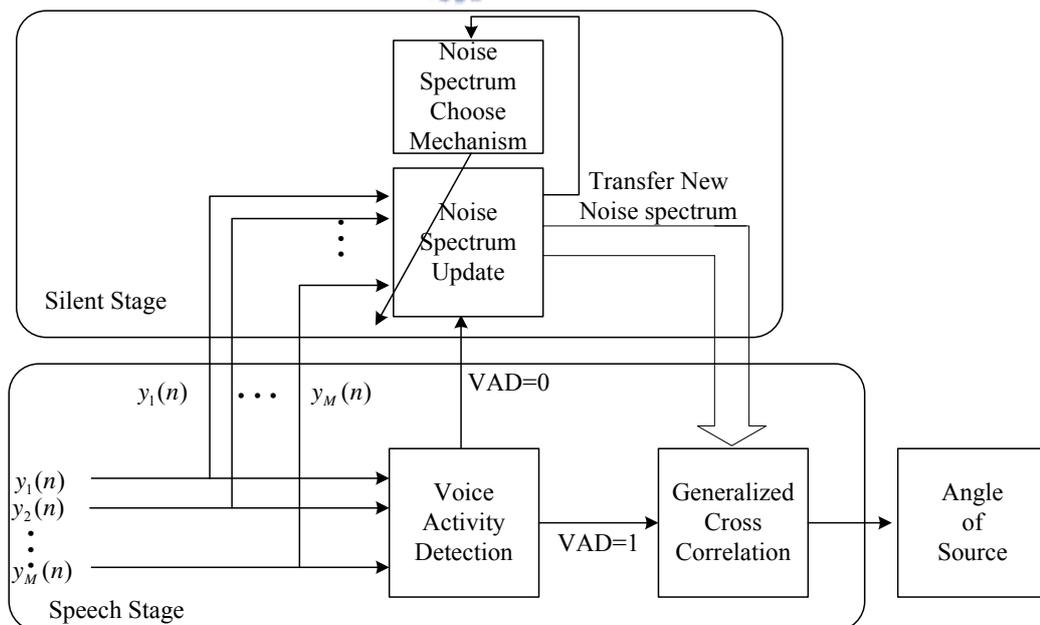


圖 12 適應性聲源定位系統架構圖

流程步驟如下：

- (a) 透過雙聲道麥克風，錄取輸入訊號 $y(n)$ 。
- (b) 經過 VAD 系統，判斷環境中是否有聲源。
- (c) 若 $VAD=0$ ，表示判斷沒有聲源發生，則進入 silent 階段，針對噪音頻帶紀錄；若 $VAD=1$ ，表示判斷環境中有聲源，則進入 GCC 系統，作聲源定位判斷。
- (d) 經過 GCC 系統運算後，得到聲源角度資訊，往後傳遞以控制機器人運轉。

在 2.1 節中有介紹過論文中使用的 VAD 演算法是使用長時間語音的資訊而非傳統瞬間音框 (instantaneous frame) 資訊，針對長時間語音資訊，定義出 LTSE--代表了長時間語音資訊的意義，和 LTSD--現階段長時間語音的頻譜能量佔了雜訊頻譜能量的比例，利用此兩個定義，判定是否為真人語音是用了現階語音能量的大小來判定，而此能量大小包含了長時間語音資訊，並非只有單一音框資訊。

在搭配臨界值 γ 對應調整環境中最吵雜以及最安靜情形的環境參數後，經過 Hang-over 機制便能得到 VAD 的語音篩檢結果，因此，針對 GCC 頻帶選擇的依據是前端 VAD 的結果而來，根據前端 VAD 篩選語音資料後，將結果記錄下來，分析辨別出語音或是雜訊，將雜訊頻帶記錄下來，轉給 GCC，用以調整 $\psi_{ij}(k, \omega)$ ，使其在頻帶的選擇上能更有依據以及去除雜訊的影響。

綜合以上兩節，對於適應性頻帶選擇的方法，即為對於聲源偵測 GCC 演算法的 ω 作智慧型的選擇，以前段 VAD 聲源啟動裝置中所得到的聲音資訊，分辨出聲源的頻帶或是雜訊的頻帶，而紀錄沿用至後段 GCC 的頻帶選擇。此法相較於以往使用直覺選定頻帶則顯得更有依據。相關實驗結果，請參考第 5.2 小節。

第三章 基於 DSK5510 實現定點語者方位判別系統

3.1 簡介

經過前面第二章所提到的相關理論及電腦端浮點 MATLAB 的驗證後，我們要把此演算法在 OMAP 上實現。由於 OMAP 上包含有兩個核心系統，一個為 ARM，另一個則是 DSP 5510，希望讓聲音演算法交由 DSP5510 來處理，針對 FFT 等數位訊號所常用的運算發揮其 DSP 的處理效能，而讓 ARM 接管系統管理的工作，所以首先針對 DSP 環境做實現。為了節省除錯時間，首先選擇在 DSK5510 這個平台上先將演算法實現。

DSK5510 相較於 OMAP 雙核心系統擁有較為單純的除錯環境，在程式開發上比較容易著手。從電腦端轉換平台到 DSP5510 上，最大的差異在於浮點運算和定點運算的不同，在定點化的過程首先我們有種種考量，包含有硬體限制、記憶體配置、數學運算式子修正以及硬體架構規劃，將在以下各小節中詳細說明。

3.2 原理

3.2.1 定點化規格考量

在定點化的過程首先我們有以下考量

1. 硬體支援的取樣頻率
2. 暫存器的精度
3. 記憶體空間
4. 計算過程造成量化誤差的影響

根據以上幾點的考量搭配 55 系列的 DSP，所以我們選定實現上的規格如下：

ADC 取樣頻率	48K
ADC 取樣精度	16-bits
內部暫存器精度	16-bits
定點乘法器	17-bits X 17-bits (signed)
FFT 點數	512 點
VAD 計算頻帶	100Hz~4KHz
LTSE 取樣音框	-3~3(7 個音框)
噪音音框	-2~1(4 個音框)
GCC 計算頻帶	依據當時計算狀況決定
GCC 延遲取樣數目	19
正弦餘弦精度	16-bits
GCC 統計量	19 次

表 1 DSP 定點化規格表

ADC 取樣頻率是由 DSK5510 硬體規格所決定，聲音取樣頻率由以下圖 13 所決定。

所以若要設定取樣頻率，就必須要在程式中針對 0001000 這個位址設定。搭配 USB 模式，我們能夠使用到的取樣頻率有 ADC 以及 DAC 的兩種配合的組合情形，組合情形如下頁表 2 所示。在這裡我們的輸出入的設定是 48K 的取樣頻率。

Sample Rate Control (Address: 0001000)									
BIT	D8	D7	D6	D5	D4	D3	D2	D1	D0
Function	X	CLKOUT	CLKIN	SR3	SR2	SR1	SR0	BOSR	USB/Nor- mal
Default	0	0	0	0	0	0	0	0	0
CLKOUT	Clock output divider			0 = MCLK	1 = MCLK/2				
CLKIN	Clock input divider			0 = MCLK	1 = MCLK/2				
SR[3:0]	Sampling rate control (see Sections 3.3.2.1 and 3.3.2.2)								
BOSR	Base oversampling rate								
	USB mode:			0 = 250 f_s	1 = 272 f_s				
	Normal mode:			0 = 256 f_s	1 = 384 f_s				
USB/Normal	Clock mode select:			0 = Normal	1 = USB				
X	Reserved								

圖 13 取樣頻率參數設定

SAMPLING RATE†		FILTER TYPE	SAMPLING-RATE CONTROL SETTINGS				
ADC (kHz)	DAC (kHz)		SR3	SR2	SR1	SR0	BOSR
96	96	3	0	1	1	1	0
88.2	88.2	2	1	1	1	1	1
48	48	0	0	0	0	0	0
44.1	44.1	1	1	0	0	0	1
32	32	0	0	1	1	0	0
8.021	8.021	1	1	0	1	1	1
8	8	0	0	0	1	1	0
48	8	0	0	0	0	1	0
44.1	8.021	1	1	0	0	1	1
8	48	0	0	0	1	0	0
8.021	44.1	1	1	0	1	0	1

† The sampling rates are derived from the 12-MHz master clock. The available oversampling rates do not produce exactly 8-kHz, 44.1-kHz, and 88.2-kHz sampling rates, but 8.021 kHz, 44.117 kHz, and 88.235 kHz, respectively. See Figures 3–17 through 3–34 for filter responses

表 2 數位類比輸入輸出取樣頻率設定對照表

在 DSP 內部，所有資料都被設定為 Int16 的格式，這是有號數，所以暫存器的精確度是 16 bit。而在 DSP 程式中所有的乘法運算都利用組合語言來處理，所以精確度可以利用 DSP 內部 40bit 累加器來保留下來。

另外，根據圖 9 所示，如果我們根據表 1 所設定，將 GCC 延遲取樣數設為 19，這個意思是指，圖 9 中的兩顆麥克風之間的距離，最多延遲 9 個取樣點。而根據取樣頻率 f_s 、聲速 c ，可以求得麥克風之間的距離 x 。如下所示：

$$a \cos\left(9 \times \frac{1}{f_s} \times c\right) / x = 0$$

$$f_s = 48000 \text{ sample/sec}, c = 34000 \text{ cm/sec}$$

$$\therefore x = 6.375 \text{ cm}$$

所以當配合硬體規格定出相關參數時，其實就已經算好麥克風應該擺放的距離，這也影響了 GCC 的結果角度準確度。

3.2.2 記憶體配置

我們採用 ping-pong buffer 的機制來處理前段 ADC 取樣後的資料，當 ping buffer 在擷取 ADC 取樣資料的同時，pong buffer 則進行演算法的運算，在下一個取樣時間之前必須完成所有運算並且將對應計算的結果存回指定的位置，等到下一個取樣時間到達，ping buffer 和 pong buffer 互相交換功能，依此類推去處理輸

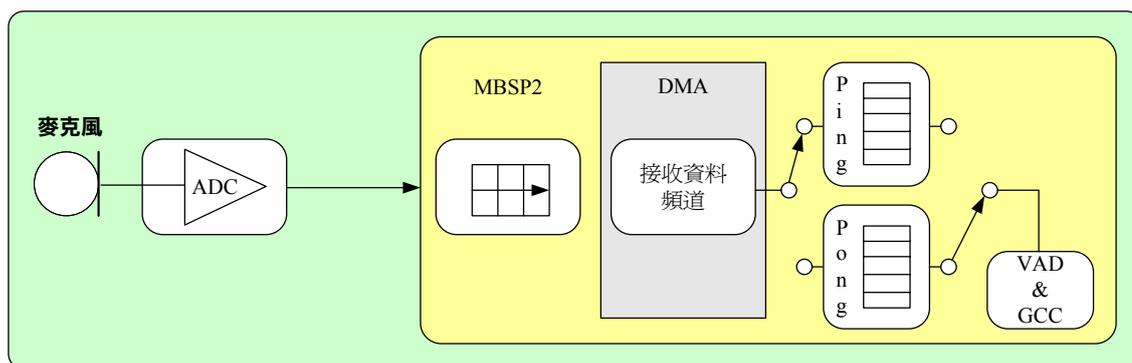


圖 14 乒乓暫存器機制圖

入的訊號，如此可以避免資料遭到覆蓋或是處理不完全的情形發生，如圖 14。

而表 1 提到內部主要會利用 7 個 512 點儲存緩衝區來記憶運算所需的資料，最新取樣的資料會以環形指標的方式找到對應的緩衝區而覆蓋 7 個取樣時間之前的資料，如此我們接下來的運算都只要到固定的位置去存取即可。

3.2.3 方位判別系統運算修正

■ VAD

根據第二章的理論介紹，VAD 演算法的主要精神在於找到適應性的背景噪音去比對出能量而辨認真正聲源是否發生，原本的計算 LTSD 和噪音的公式如式 (2-2).(2-5)，搭配適應性的臨界值(式 2-4)來決定是否有聲源存在。

在實現上我們必須對這些公式進行化簡，以減少 DSP 的運算時間和運算誤差，首先針對 LTSD 進行化簡，我們拿掉了對數以及平方的運算，因為 LTSD 的目的只是要與臨界值做比較，所以這也表示我們同步的改變了原來臨界值公式的表示單位，不過這是相對的大小作幅度調整。

在計算噪音的公式中，我們帶有一個 α 的權重，MATLAB 模擬的採用的數值為 0.95，這意味著最新取樣的噪音只佔最後更新後噪音的 5%，假若噪音訊號不夠大的時候，容易因為有限精度運算而消失，所以我們適度的放大噪音以確保 α 權重能夠忠實呈現，當然相對應的 LTSD 比值則會因為噪音放大而變小，導致在與 γ 做比較的精度相對變低。

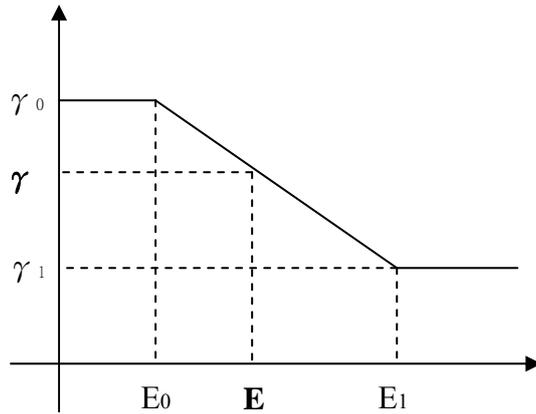


圖 15 γ 公式內插圖解

最後對於 γ 公式的實現，我們可以拆解成同一斜率上找出內差的觀點來思考，如圖 15 所示，因此不同的斜率會影響到 γ 的變化敏感度，所以在選擇將噪音變大的同時， γ 的敏感度會相對的下降。

總結以上的考量，我們可以得知定點化所帶來的誤差和必須在每個變數上進行取捨，以減少運算過程精度的損失。



■ GCC

同樣的考量模式應用在式(2-9)，我們計算的目的在於找到最大值而推估出聲音與兩個麥克風的時間差，因此簡化分母項的開根號運算，此外在式(2-9)中也必須將指數運算拆解成正弦餘弦表示法，因為有對稱的關係所以我們只需要預先計算一半的角度存入記憶體，最後得到新的運算式如下：

$$\hat{R}_{ij}[\kappa, l] = \sum \frac{G_i \times G_j}{G_i^2 + G_j^2} (\cos(\omega l) + \sin(\omega l))$$

根據新公式，分母項次因為沒有取根號，所以會遠大於分子項，必須作適當擷取，才能換取整個運算式的解析度。

另外，由式(2-9)來看，實際上是將頻域訊號再經過 inverse FFT，轉回時域，但是如今在 GCC 運算時，並不是針對所有的點，所以個數上並不對稱，也不應該採計虛部的部份，所以運算過後，當要計算出最大值時，如式(2-10)，則只需要針對實部作判斷就好。

最後我們經過 19 次運算得到統計後要回推角度，必須經過反餘弦計算，而計算這個部份則在下個階段交由 OMAP 中的 ARM 端來計算，由 C 語言來處理較為複雜的反餘弦函數，如此分工較能達到效益。

■ GCC 頻率選擇

在之前 2.3 小節中有提到，我們將 VAD 運算完畢的結果，記錄其噪音能量，傳送到後端 GCC 來做頻帶篩選，所以在作 GCC 之前，對於 FFT 完的結果還要再經過一次能量的選擇，也就是針對我們認知上的語音訊號主頻來做 GCC，來比較出延遲點數，降低雜訊的影響。



3.3 定點實驗結果

下面是 VAD 的定點化實驗結果。比較圖 16 的左右兩張圖，可以發現右邊的圖是定點化的結果，數值分佈變動較少，不自然的起伏。實測的語音內容是『一二三四五』。

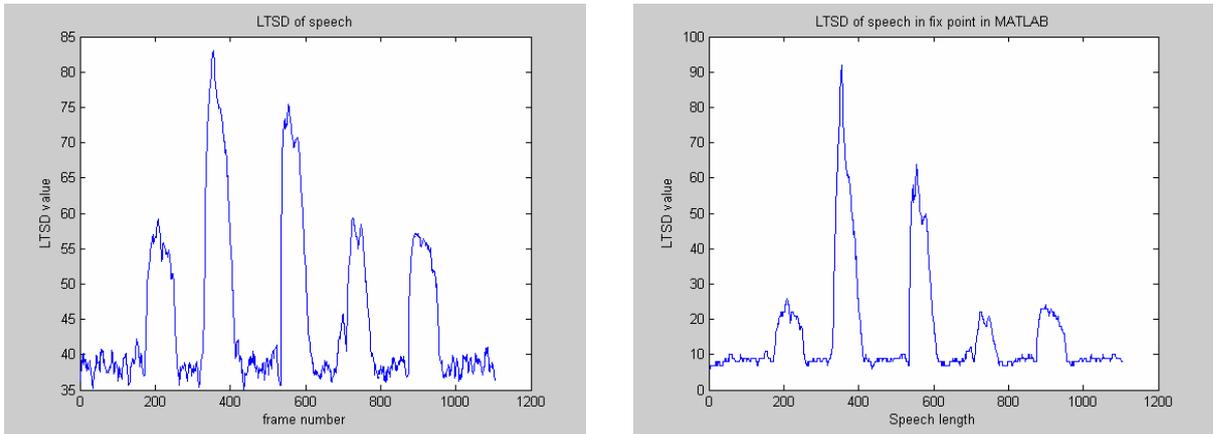


圖 16 浮點 MATLAB(左)與定點 DSP(右)的原始語音輸入能量

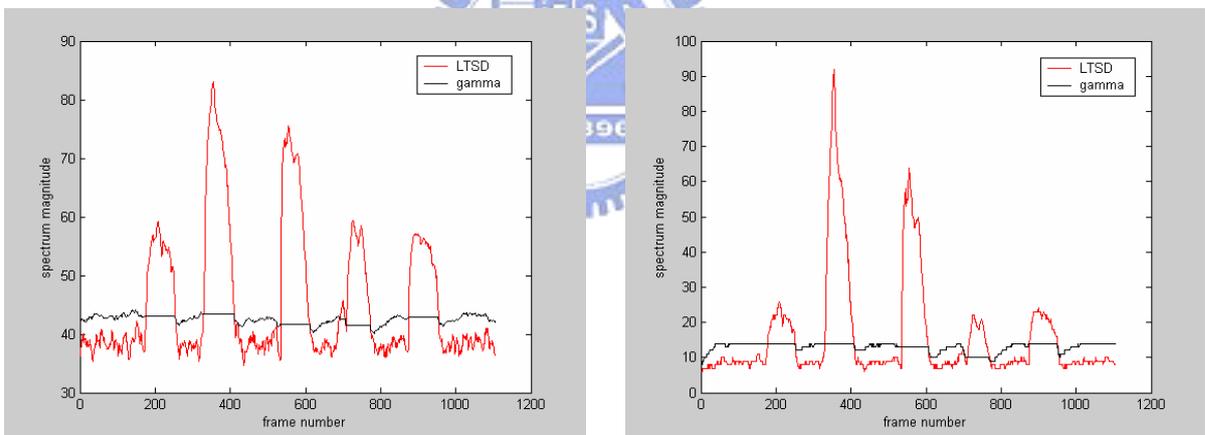


圖 17 浮點 MATLAB(左)與定點 DSP(右)的 LTSD 與 gamma 關係結果比較

比較最後浮點的 VAD 結果(下頁圖 18)與定點的 VAD 結果(下頁圖 19)，可以發現結果相差不大，缺乏一些氣音的部分。不過因為定點的數值範圍較小，相對於浮點所能表現得雜訊適應能力較低，不過因為 VAD 結果是 LTSD 以及 gamma 兩者之間的比值，所以只要關係存在，都能表現出 VAD 的結果。

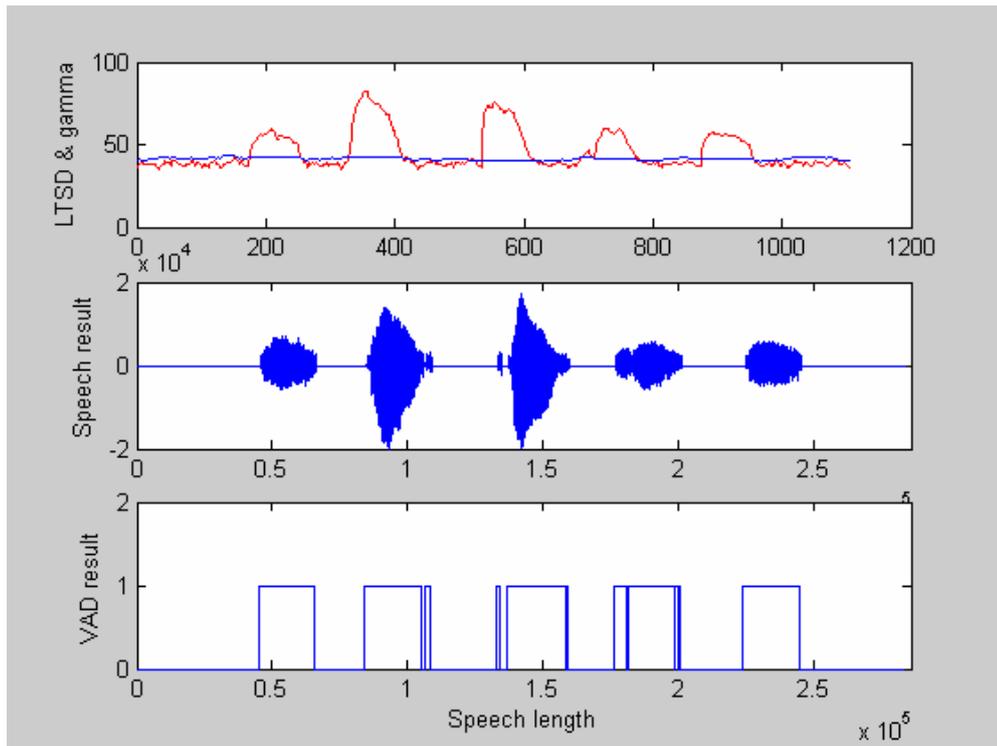


圖 18 浮點 MATLAB 的 VAD 結果

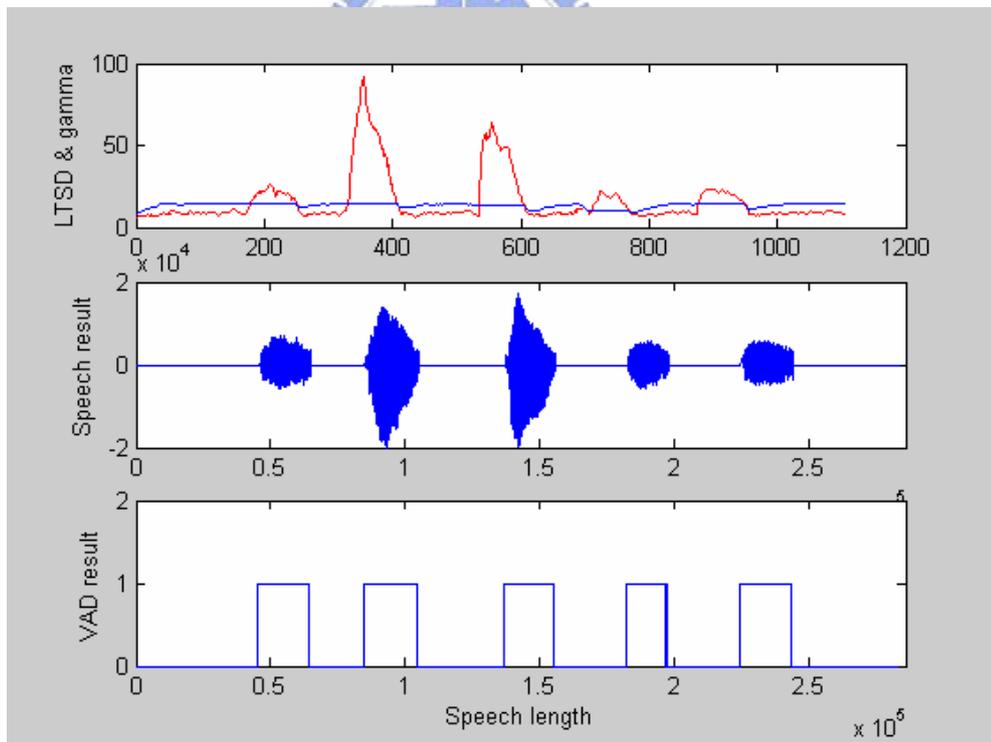


圖 19 定點 DSP 的 VAD 結果

環境參數設為： $\gamma_0=45$, $\gamma_1=5$, $\text{noise_e0}=50$, $\text{noise_e1}=75$ 。

第四章 實驗平台

4.1 平台架構說明

圖 20 說明整個系統架構：

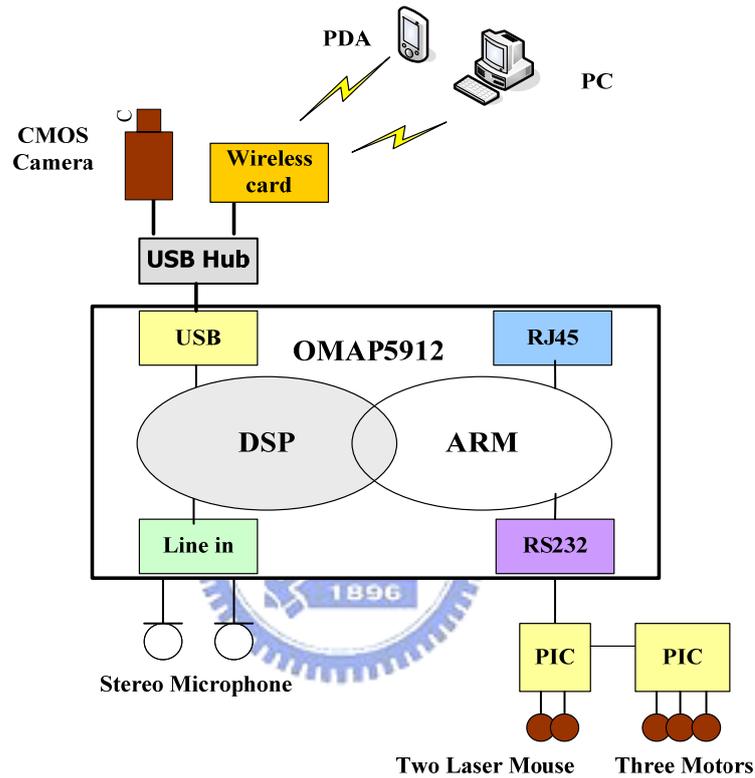


圖 20 海星號系統架構圖

整體運作以 OMAP5912 為平台核心，負責執行各項周邊子程序及聲音前處理，建構於海星號機器人上，硬體周邊有：

- (1) CMOS 攝影機
- (2) 雙聲道麥克風
- (3) 無線網路
- (4) 馬達控制模組
- (5) 光流影像移動感測器定位模組

雙聲道麥克風負責接收環境中的音源訊號，也同時接收了環境噪音；無線網

路負責傳輸 telnet 主控模式以及架設 web server，讓使用者可以遠端透過網路操控系統，讀取機器人接收資訊；同時也將聲音資料透過 socket 傳遞到 PC 端處理。此外，透過 RS232 介面與 PIC 模組相連，可以控制馬達以及接收光流影像移動感測器回傳位置資訊。

下圖 21 就是整個開發平台的架構，當在個人電腦上的 ARM 程式與 DSP 程式開發完成後，不需要下載到硬體平台上，只要經過 NFS 網路檔案系統，使用 RS-232 介面來控制平台的運作，即可直接在平台上使用寫好的程式。此雙核心系統分別是 ARM 和 DSP，由 ARM 負責主控系統運作，DSP 負責運算聲音前處理，各司其職又能相輔相成，互相溝通，充分將雙核心功能發揮到極致，本系統便是利用此特質，著眼開發海星號機器人系統。

發展平台架構圖如下：

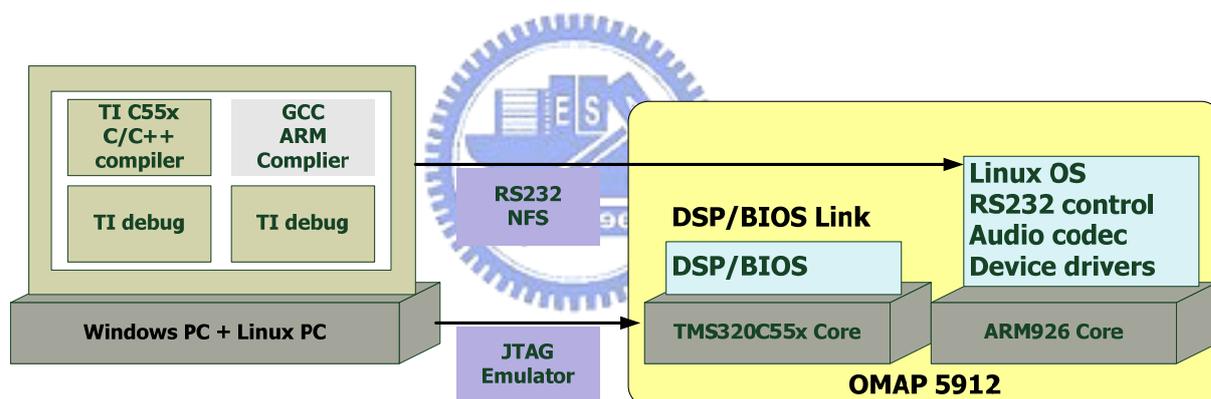


圖 21 發展平台架構圖



圖 23 OMAP5912 發展平台

雙核心架構提供了整合 DSP 和 RISC 這兩項不同的技術 - TI TMS320C55x DSP 核心與高效能的 TI926 ARM 核心。

➤ Inter-processor communication

在 OMAP 處理器中的 ARM 與 DSP 資料交互傳遞的方法有下列三種

◇ Mailbox

ARM 與 DSP 處理器可以透過 Mailbox 中斷的機制做資料交換，這樣的機制可以提供軟體自定通訊協定讓處理器之間的通訊更有彈性。當 ARM 把資料寫進 DSP 時，DSP 會收到一個 Mailbox 中斷，藉由中斷 DSP 可以在 ARM 寫進資料後，就可以取出使用。DSP 也可以用類似的方法與 ARM 通訊。

Mailbox 不適合拿來作大量資訊的傳輸，比較適合傳送命令、參數或是指標。

◇ MPU Interface(MPUI)

MPUI 可以讓 ARM 和系統的 DMA 控制器與 DSP 和 DSP 的週邊互相通訊。MPUI 可以存取整個 DSP 的記憶體空間與 DSP 的週邊匯流排。因此 ARM 與系統的 DMA 可以有 DSP 輸入輸出空間的完整存取權限。

MPUI 適合用來做下列工作

- ◆ 由 ARM 將 DSP 的程式載入到 DSP 的程式記憶體中。
- ◆ ARM 與 DSP 之間的資料分享
- ◆ 透過 Share Memory 達成行程間通訊的通訊協定
- ◆ 讓 ARM 可以使用與控制 DSP 的週邊

☆ Share Memory

這樣的架構是依靠 OMAP 上的 Traffic Controller 達成，藉由 ARM 與 DSP 存取相同的記憶體空間(含 SRAM 與其它記憶媒介)。ARM 可以控制 DSP 的 MMU 讓 DSP 可以存取這些記憶媒介。使用 Share Memory 的機制需要 Mailbox 與 MPUI 的幫助。



選用 OMAP5912 這塊平台來開發，最主要著重於其 ARM based 系統能讓更多類型的程式得以執行，讓機器人系統整合更多周邊，這對於開發嵌入式系統來說非常方便，可以針對不同的需求而開發不同的硬體設施；DSP 則可以輔助 ARM 讓耗時間的演算法得以實現。

4.2.2 麥克風放大電路

C55x DSP 是一顆高效能、低耗電量的處理器。處理器中含有兩個乘累加器 (MAC)，每一個 MAC 可以在一個週期內將 17-bit 的乘法完成。算術邏輯單元 (ALU) 有兩組，一為 40-bit 另一為 16-bit。由上述說明可知，DSK5510 擁有高速的 DSP，可實現複雜的演算法，而 AIC23 Codec 在 DSP 端，提供 AD/DA 轉換，利用 Flash 可燒入已編輯的程式碼，使得 DSK 可脫離 PC 獨自運作。

因為 AIC23 Codec 的麥克風輸入放大電路只支援單聲道及為了量測受控體的頻率響應，所以另外做了雙聲道的麥克風放大電路，再與 AIC23 Codec 的 Line In 連接。所以整個系統架構圖與實體圖如下所示：

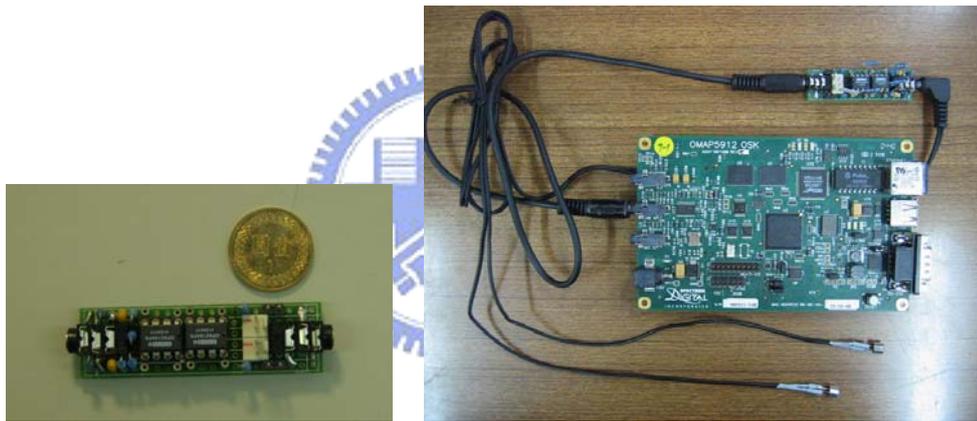


圖 24 麥克風放大電路

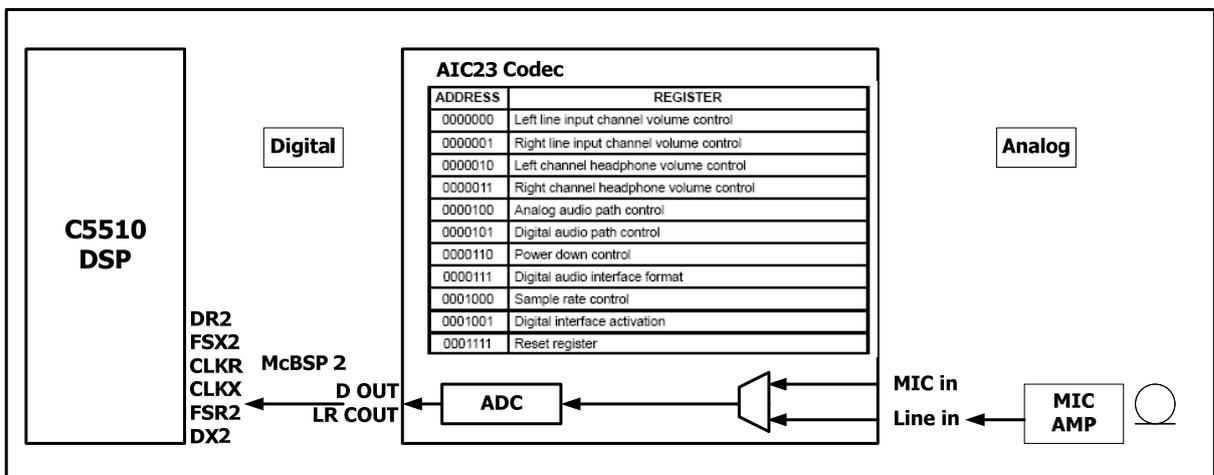


圖 25 麥克風系統架構圖

4.2.3 海星號輪式機器人平台運動控制

海星號輪式機器人是使用全向輪來建構出一個擁有全方位運動能力又具備有對地面之光流影像偵測法定位功能的嵌入式網路整合操控的運動平台。

此平台底盤的下方包含 2 個光流影像移動感測器(兩個托盤下方)、3 個馬達以及 3 個全向輪，如圖 26 所示；光流影像移動感測器是用來感測地面影像，並使用地面影像的光流變化來量測該平台的運動狀態；馬達是用來提供該平台運動動力的，可以透過微控制器使用 PWM 訊號來控制該馬達的正反轉向和轉速快慢；全向輪則是提供該平台全向運動的能力。

海星號的控制程序以 OMAP 為主，當聲音演算法計算完角度後，便從 RS232 送出控制訊號，連接到馬達控制板，控制馬達轉動速度，讓平台旋轉。

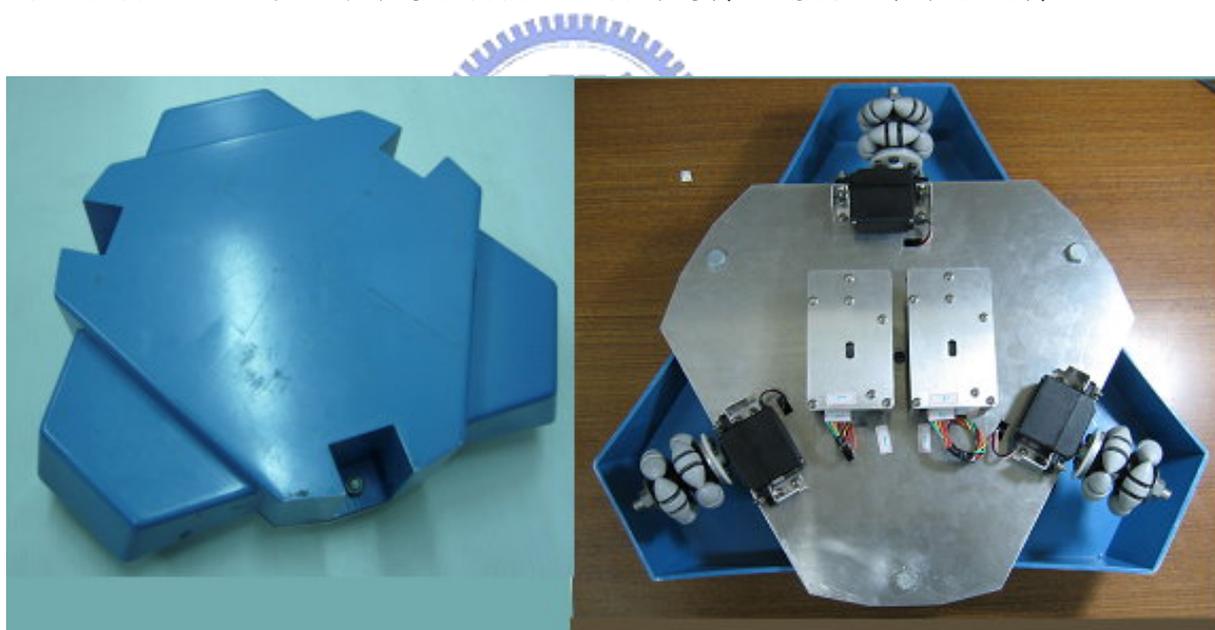


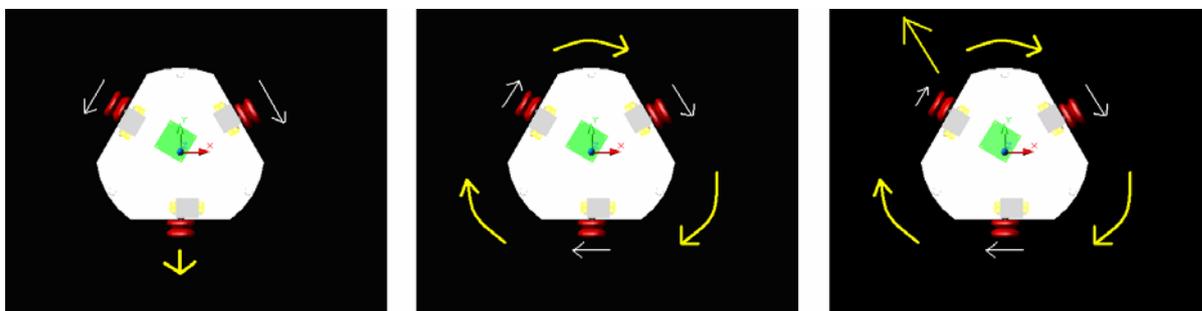
圖 26 海星號機器人平台實體圖

■ 運動模式

海星號主要的組成機構為三個特殊設計之全向輪。全向輪的結構特徵，為在一個圓形輪軸的圓周上環繞著數個橢圓形狀的滾筒，這些滾筒的軸心和輪軸的平面之間的角度是可以調整的；這些滾筒的作用便是用來將輪子轉動時產生的垂直輪軸的力量轉換至平行於輪軸的方向上面。藉由此三個全向輪的運動控制可以達到：(1)原地旋轉，(2)對頭直行，(3)對頭差速轉彎，(4)平移，(5)平移且自轉。如此高的自由度動作是一般傳統輪式機器人無法同時具備的。由簡至繁解說如下：

- (1) 原地旋轉：三個輪子的角速度保持一定且相等、轉向相等，則本平台可作原地等角速度旋轉。
- (2) 對頭直行：當三輪中其中一輪不運轉，其他兩輪以相同的角速度旋轉但轉向相反時，則本平台會朝不運轉的輪子方向(一個「頭」)前進。
- (3) 對頭差速轉彎：以對頭直行模式為基礎，轉動的兩輪若轉速不同，則本平台會進行轉彎的動作，此時「頭」的指向會改變。這種運動模式與一般兩輪平台差速轉彎的道理類似。
- (4) 平移：所謂平移即是本體指向永遠不改變的移動，而且移動方向可以任意選擇。這種運動方式正是二輪平台無法呈現的運動模式。
- (5) 平移且自轉：這是本平台所能呈現最複雜的一種運動模式，一面對固定方向作平移，一方面又以固定角速度作自轉。

以下列出其中幾種運動模式的示意圖：



(a)沿某一個方向直線前進

(b)原地正反向旋轉

(c)沿一方向前進且同時旋轉

圖 27 輪式機器人運動模式示意圖

■ 聲源追蹤系統整合

海星號機器人系統與聲源追蹤合作，功能是在聲源方向判定後，自動轉向聲源位置，並利用攝影機拍攝聲源方向有無異常環境變化，所以搭配此系統，機器人運動平台必須轉換聲源方向角度資訊，將角度資訊轉換成馬達轉動資訊，並傳送給平台上的全向輪馬達，使其運動朝向正確的角度。

海星號機器人的馬達控制流程圖如下圖 x 所示：

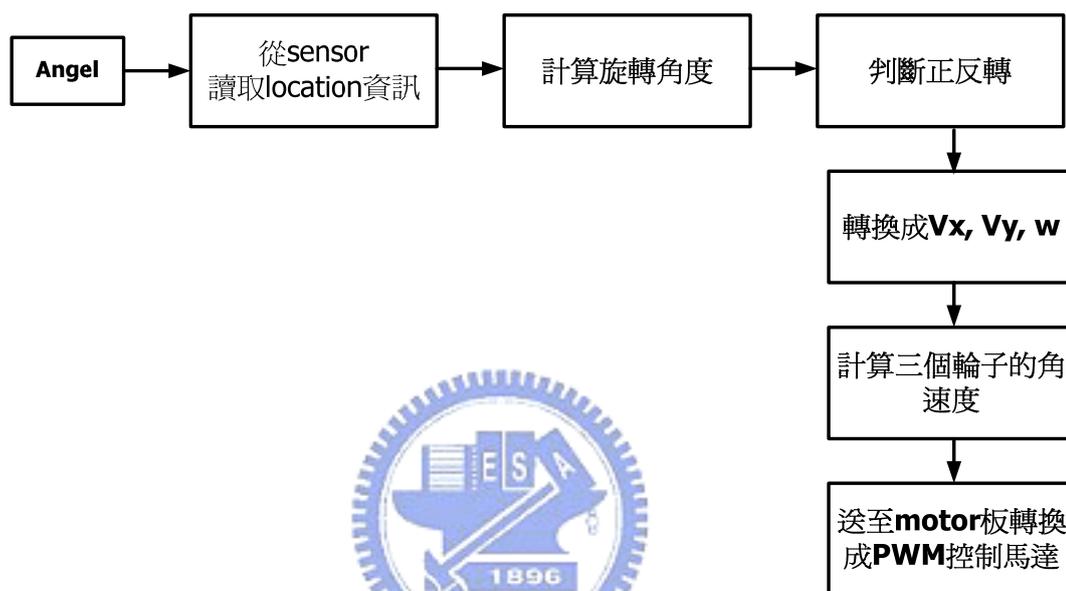


圖 28 馬達控制流程圖

4.3 軟體環境

4.3.1 Linux 作業系統

自從 1991 年 Linux 由 Linux Torvalds 將第一版 Linux 公開在網路上後，這套系統已經受到世界各地普遍的歡迎。現在的 Linux 已經支援相當多的處理器核心，所以現在 Linux 的應用相當的廣泛，由高階的伺服器、工作站，到常見的個電腦中都有它的蹤跡，甚至是一些手持式裝置例如手機、PDA 等一些現代人常用的嵌入式系統上都看得到 Linux 的存在。

所謂的嵌入式系統指的就是為了提高產品的穩定性、可攜性，或者是降低產品成本、增加特定功能等不同的目的，對一般的標準系統進行修改以符合特殊設計上的需求。在硬體方面靠的主要就是被稱為系統單晶片（SOC - System On a Chip）的半導體整合技術，而嵌入式作業系統則是軟體方面的技術核心。Linux 的可靠、高效能、伸縮性、低成本，以及最重要的握有控制權等幾項特點，使得 Linux 在嵌入式作業系統上站在非常有利的地位。

為了讓 Linux 可以在 OMAP 5912 上執行，必須使用以下修正檔，讓 OMAP5912 的處理器與週邊可以順利使用。

- ARM linux patch
- Montavista OMAP Processor and kernel
- DSP/BIOS Link

另外，針對本系統要求，也著手開發相關週邊，如 USB wireless 以及 USB Camera，如此更可彰顯出 Linux 系統在開放源程式碼之下，提高了系統整合的優勢，對於系統開發者更為方便。

4.3.2 ARM 處理器以及轉換編譯器

(a) gcc

gcc[28] 可以說是 Richard Stallman 所創立的 GNU 計畫中最重要的作品之一，它提供了自由軟體世界高品質的編譯器 (compiler)，實現了我們在自由軟體平台上開發程式的夢想。如果沒有 GCC，恐怕今日我們也不會有這麼多形形色色的自由軟體可用。

gcc 一個很大的特色是高度可移植性，目前已知有超過三十種硬體平台與作業系統可以執行 gcc，其中硬體平台包括了: x86, arm, ia64, alpha, hppa, m68k, Power PC, mips, IBM rs6000, sparc/sparc64, 等，而作業系統則從 Microsoft 平台 (DOS/Win32) 到 IBM OS/2 到各家的 UNIX 都有。此高度可移植性正是 gcc 廣為流傳散佈的主要原因。在許多商業版的 UNIX 系統中，如果沒有特別買其專屬的編譯器的話 (其價格往往不便宜)，人們通常就選擇安裝 gcc 來使用。而且，儘管 gcc 是自由軟體計畫開發出來的，但其所編譯出來的程式品質並不輸給商業版的編譯器，甚至在某些平台上所編譯出來程式有更好的執行效能。

(b) Cross-compiler

通常在嵌入式平台上不會有多餘的記憶體與儲存空間來直接編譯程式，所以需要由其他機器產生嵌入式系統的執行檔，然後放到嵌入式系統上執行。Cross-compiler 是要在發展平台上產生目的平台的程式碼，在這裡發展平台是 x86 架構的個人電腦，目的平台則是 OMAP 5912 上的 ARM 處理器。

在開發平台上使用的是 GCC 3.3.2 Cross compiler 工具組，這些軟體工具將放在 Linux 上執行。

4.3.3 數位訊號處理器

TMS320C5510 DSP 晶片為德州儀器公司的 DSP 處理器，為目前工業界最省電的 DSP 之一，相當適合應用在可攜式產品，晶片內部擁有一個程式匯流排(program bus)，三個資料讀取匯流排(data read bus)和兩個資料寫入匯流排(data write bus)增加運算的平行度，藉此達到高效能且省電的目的，並且可自動關閉未使用的周邊，記憶體或核心工作單元來增加電池的壽命。

此 DSP 擁有完整的發展平台及工具，例如 C-Compiler，使用者可用 C 語言發展軟體，容易發展複雜的演算法，提供許多常用的函式庫，例：DSPLIB（包含了 FIR/IIR filters，FFT 等等），加速常用的運算，還有 DSP/BIOS 工具，讓產品能容易的擁有即時多工的功能，上述的這些工具使得此 DSP 能快速的實現複雜的演算法，也兼顧了程式碼執行的效率。

在週邊部分，已內建了多樣化的輸出入通道，包括主機埠界面(Host Port Interface，HPI)、外部記憶體界面(External Memory Interface，EMIF)、高速串聯埠界面(Multichannel Buffered Serial Ports，McBSP)、即時資料交換(Real-Time Data Exchange，RTDX)，充分提供與外部輸出入訊號的傳輸界面，表 3 為 TMS320C5510 一些基本的特性。

PARAMETER NAME	TMS320C5510
Clock frequency (MHz)	200
Data/program Memory (words)	160k
DMA	6
External Memory Interface	32
MCBSP	3
IO Supply (Volts)	3.3
Core Supply (Volts)	1.6

表 3 TMS320C5510 特性表

4.3.4 程式組合開發軟體 (Code Composer Studio)

CCS 是一套整合式開發環境的發展軟體，它提供 C 語言的編譯器、組合語言的組譯器與其他工具，使用者也可以快速簡單的在軟體開發時除錯。在一般情況下它可以直接將 DSP 程式下載並在 DSP 上執行程式，但是在搭配 DSP/BIOS Link 使用時無法直接將程式下載到 DSP 中，但是在開發 DSP 程式時還是要搭配 CCS 來編譯出 DSP 的執行檔，並藉由 DSP/BIOS Link 讓 Linux 應用程式可以控制 DSP 的工作。

在 DSP 執行時可以藉由 Load symbol 的方式來對 DSP 程式除錯，但是必須 CCS 必須先啟動，讓 DSP 與 ARM 處理器由 Halt 的狀態改變成 Running 狀態，讓作業系統先執行，然後將 DSP 的程式藉由 DSP/BIOS Link 讀進 DSP 後，才去 Load symbol。下圖 29 是 CCS 的整合開發介面(IDE)。

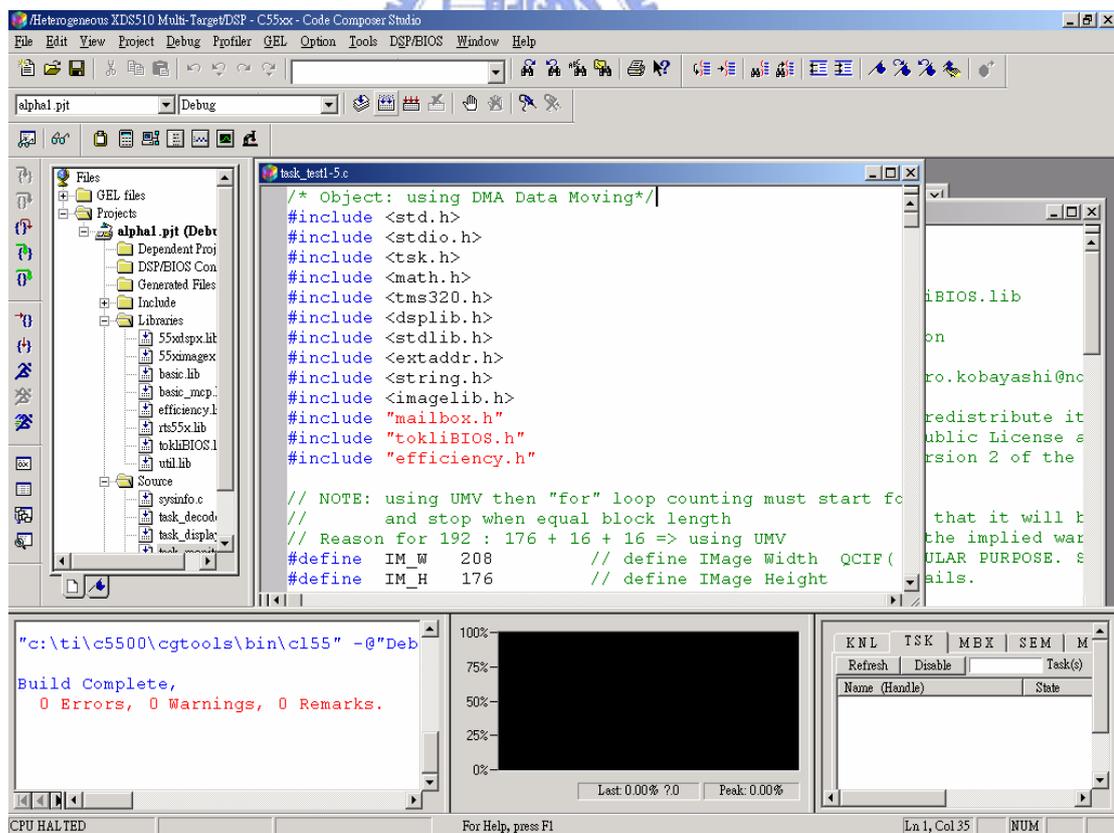


圖 29 CCS 整合發展環境介面

4.4 實驗環境與流程

平台之系統電路架構組織如圖 30 所示。海星號平台底盤的上方放置平台的系統電路架構；此系統電路架構包含 1 組 OMAP5912 發展平台、1 組 USB wireless card、1 組 USB Camera、1 組 USB Hub、1 組馬達控制板、1 組感測器控制板、1 組雙聲道麥克風電路、1 組麥克風電路獨立直流電源、4 組充電電池組以及 4 組電力供應系統板。

海星號主體以 OMAP5912 為核心系統，負責偵測環境中的聲源，其上配置有雙聲道麥克風，可以錄取環境聲音；系統並包含有 USB wireless card 以及 USB Camera，初期 USB wireless card 負責傳輸聲音資料至遠端 PC 驗證運算、接收運算完畢的聲源角度資訊，而 USB Camera 則是對環境捕捉影像，提供遠端使用者觀看。



圖 30 海星號主體實體圖

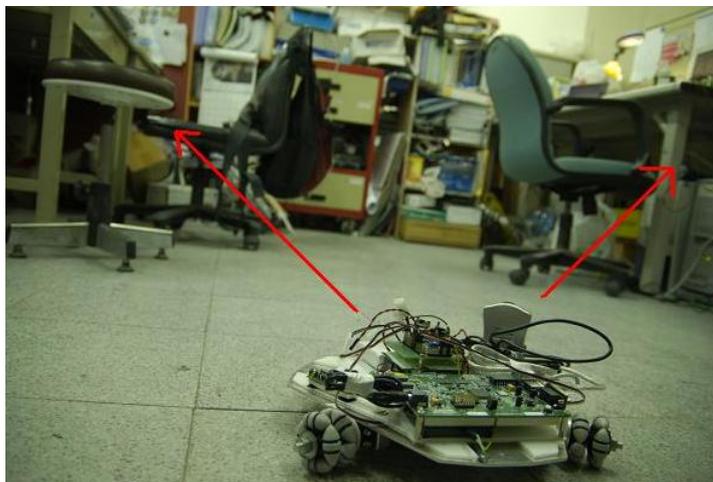


圖 31 實驗環境

實驗環境如圖 31。選擇在一開放環境中來進行語者方位判定，四周環境如同一般生活環境，包含有冷氣、電腦、細微人聲等等環境雜訊。藉由量角器事先定出特定角度進行實驗，觀察實驗結果輸出角度資訊是否符合。

整體實驗架構圖如圖 32，流程步驟如下：

- (a) 透過雙聲道麥克風，錄取輸入訊號 $y(n)$ ，進入 Speech 階段。
- (b) 經過 VAD 系統，判斷環境中是否有聲源。
- (c) 若 $VAD=0$ ，表示判斷沒有聲源發生，則進入 silent 階段，針對噪音頻帶紀錄；若 $VAD=1$ ，表示判斷環境中有聲源，則進入 GCC 系統，作聲源定位判斷。
- (d) 經過 GCC 系統運算後至 angle 階段，得到聲源角度資訊往後傳遞，以控制機器人運轉。

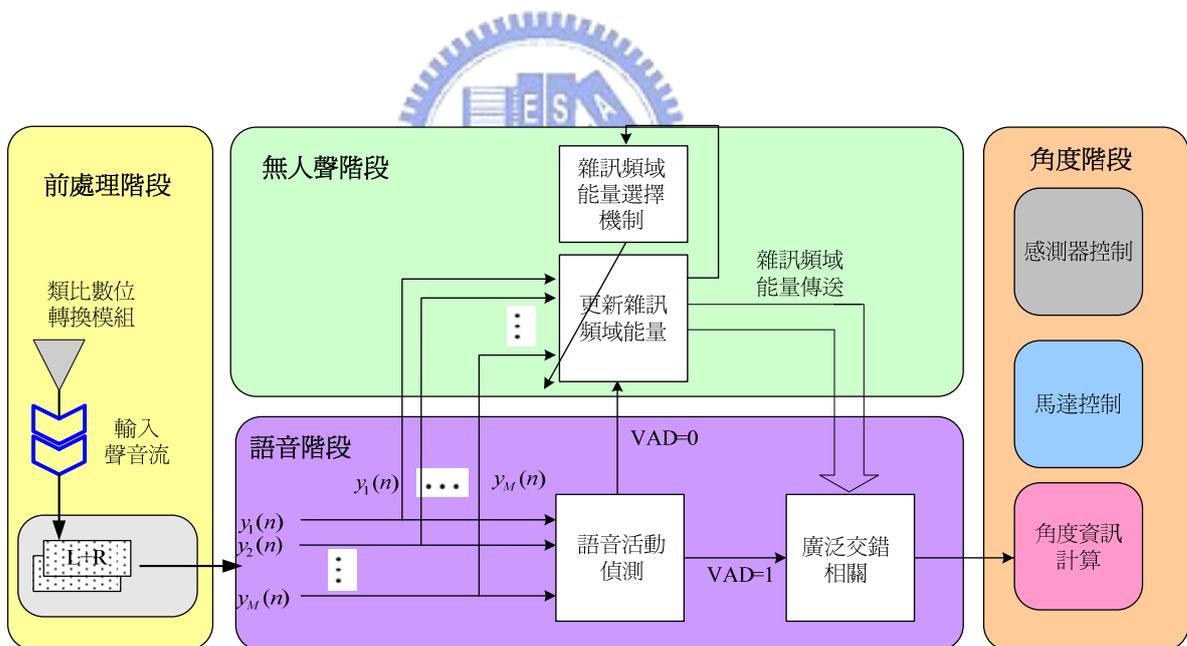


圖 32 實驗架構圖

第五章 實驗結果與分析

5.1 機器人平台對聲源偵測影響實驗

圖 33 是在機器人左右旋轉時，對雙聲道麥克風錄製一段人聲，可以觀察出其中比較明顯的六個聲音強度，為語音『喂喂喂一二三』，至於從語音訊號一開始到結束為止都存在的低強度小幅聲源，即為馬達發出的聲音。

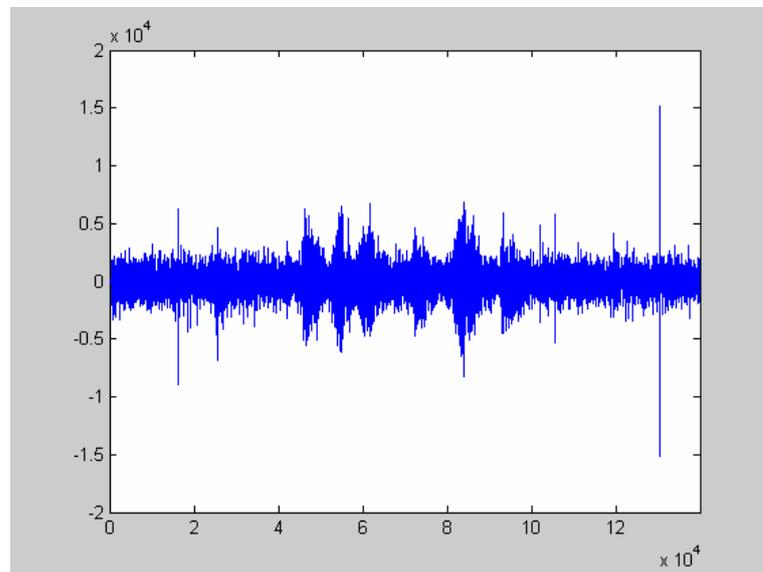


圖 33 機器人旋轉時，麥克風錄製人聲的語音分析

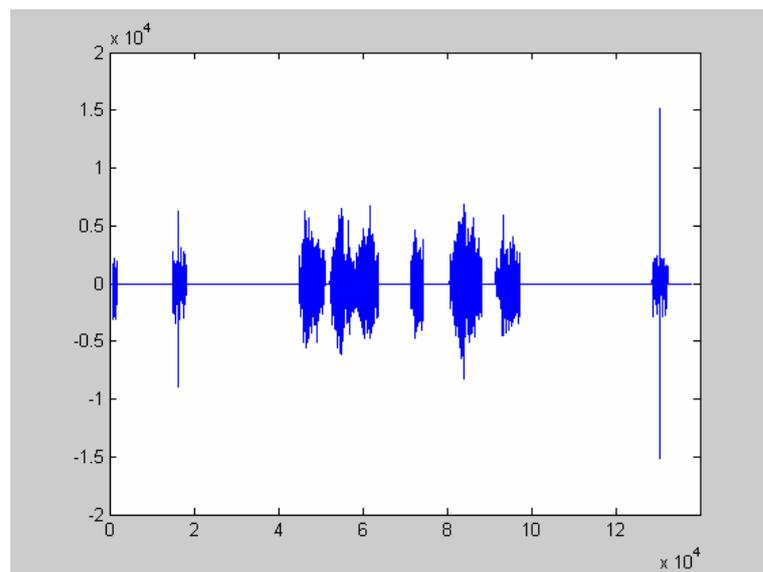


圖 34 圖 33 經過 VAD 系統後的聲音訊號

上頁圖 34 表示圖 33 經過 VAD 系統後的結果。可以發現，經過 VAD 系統後，馬達規律的聲音已經被排除為背景噪音，餘下的僅剩人聲的部份；至於圖中前後端的異常突起部分，則是環境中因為馬達在地面運轉所發生的地面障礙影響。

針對圖 33 中的語音分析部份，可以得到下圖 35 結果；單純的取出語音部分來做分析，可以發現語音部分主要頻帶是在 580Hz，其餘為共振效應的響應。於是針對人聲考量，可以預估頻帶是分布在 580Hz 附近。

至於圖 36，則是單純對馬達運轉聲音作分析，可以發現馬達頻帶主要在 48Hz 附近，針對此分析結果，我們可以把此頻帶在語音頻譜分析中予以剔除，以去除馬達對於分析錄音中的影響。

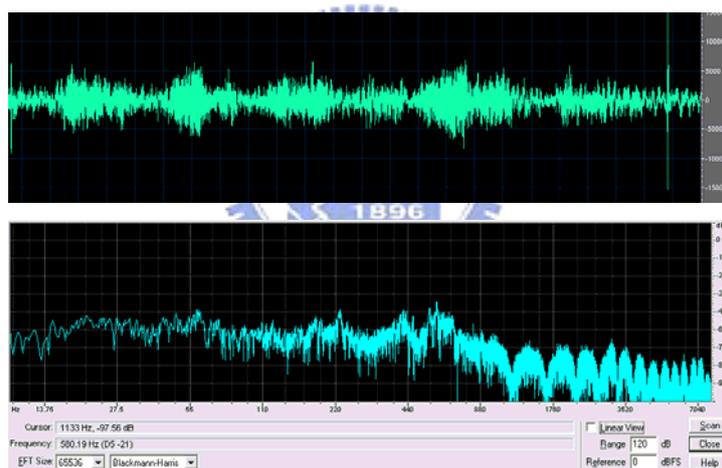


圖 35 針對語音部份分析結果

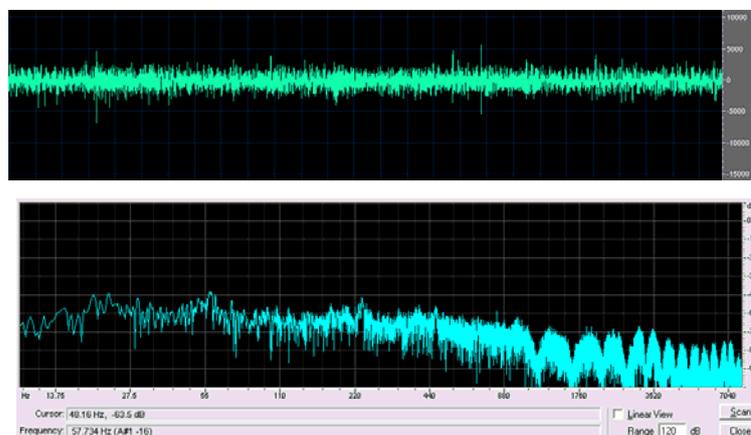


圖 36 針對假定出現馬達噪音做分析

5.2 電腦浮點運算實驗結果

實驗環境的平面關係圖如下：

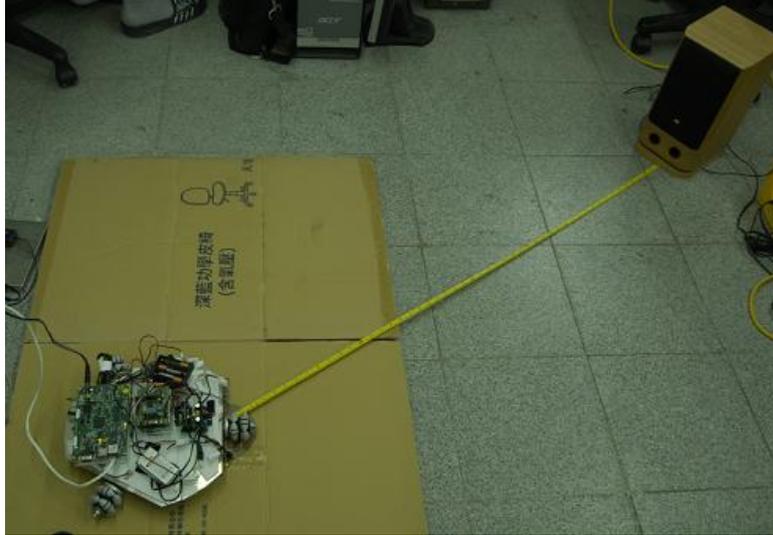


圖 37 實驗環境實際照片 1

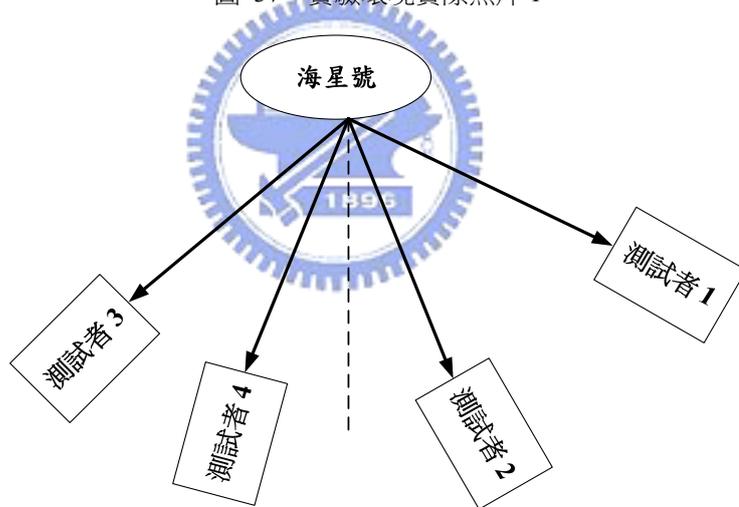


圖 38 實驗環境平面關係圖 1

本章節將探討語音訊號通過聲源啟動系統以及角度判別系統的實驗結果。圖 37 為實驗環境的實際照片，實驗環境中有一個喇叭，用來播放固定人聲，而透過電腦用來將真人語音錄製後的語音訊號經過處理計算出角度。

實際在環境中實驗時，有對應於該環境的相關參數必須先設定，主要是針對 VAD 系統做調校，如 2.1.2 節所描述，

$$\gamma = \begin{cases} \gamma_0 & E \leq E_0 \\ \gamma_0 + \frac{\gamma_1 - \gamma_0}{E_1 - E_0} (E - E_0) & E_0 < E < E_1 \\ \gamma_1 & E \geq E_1 \end{cases} \quad (2-4)$$

其中 E_0 和 E_1 代表了在最乾淨和最吵雜的情況下，雜訊的能量，而 E 是指現階段雜訊的能量。 γ_0 和 γ_1 代表在最乾淨和最吵雜的情況下與 LTSD 比較的臨界值，因此 E_0, E_1, γ_0 和 γ_1 是先設定好的初始值。所以每當進入一新環境時，隨即可以針對環境先錄製聲音觀察。以目前在實驗室模擬一般家庭環境的情形，環境參數的設定如下：

$E_0 : 45, E_1 : 220, \gamma_0 : 20, \gamma_1 : 180。$

從 2-4 式可觀察出當現階段雜訊能量介於 E_0 和 E_1 時，則 γ 會依 $E - E_0$ 在 $E_1 - E_0$ 所佔的比例，作出 γ 的線性調整。

測試方式為：分別在四個不同的方向上透過喇叭播放真人語音訊號，持續運算 100 次，紀錄每次運算完畢的角度，將結果予以平均和計算標準差。四個角度距離機器人平台各是以下距離：測試者一：115 cm；測試者二：127 cm；測試者三：102 cm；測試者四：102 cm。角度的定義為機器人平台前方半圓，以海星號為主，由右方向左方遞增 180° ，因此，測試者三是相對於海星號 45° ，測試者四是相對於海星號 75° ，測試者二是相對於海星號 120° ，測試者一是相對於海星號 150° 。

Source	SNR	Real angle	Mean	Standard Deviation
Test 1	16.12 dB	150°	128.74°	5.55
Test 2	15.03 dB	120°	111.62°	5.704
Test 3	21.4 dB	45°	56°	5.347
Test 4	22.13 dB	75°	79.01°	4.275

表 4 浮點電腦端辨識率比較表

5.3 DSK5510 定點運算實驗結果

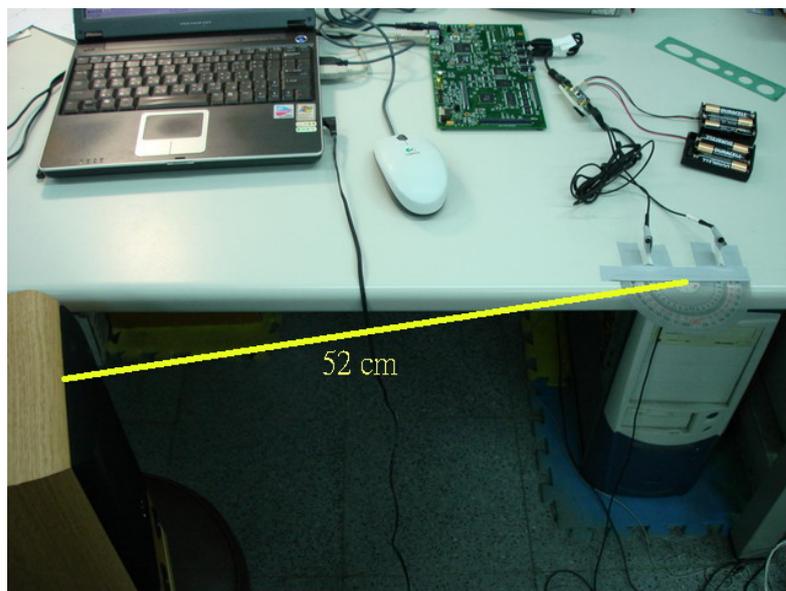


圖 39 實驗環境實際照片 2

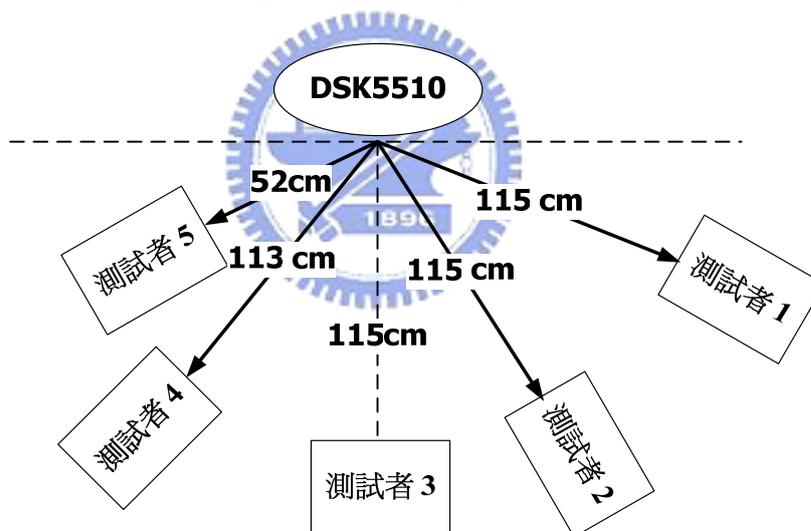


圖 40 實驗環境平面關係圖 2

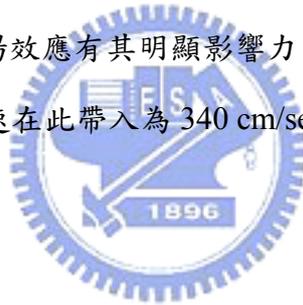
聲源資訊	真實角度	平均延遲點數	平均度數	誤差百分比
Test 1	150°	3	131.81°	12.1%
Test 2	120°	5	116.39	3%
Test 3	90°	9	90°	0
Test 4	50°	11	77.16°	54.3%
Test 5	15°	16	38.94°	159%

表 5 定點 DSP 端辨識率比較表

測試方式為：分別在五個不同的方向上透過喇叭播放真人語音訊號，持續運算 100 次，紀錄每次運算完畢的角度，將結果予以平均和。五個角度距離機器人平台各是以下距離：測試者一：115 cm；測試者二：115cm；測試者三：115 cm；測試者四：113 cm，測試者五：52 cm。角度的定義為機器人平台前方半圓，以海星號為主，由右方向左方遞增 180°，因此，測試者三是相對於海星號 90°，測試者四是相對於海星號 50°，測試者二是相對於海星號 120°，測試者一是相對於海星號 150°。

上頁實驗結果為定點 DSP 的語者方位判定結果，利用圖 39 中的喇叭穩定傳送單一人聲，測試不同的擺放角度以及擺放距離所造成的影響，因為是定點運算，所以最後角度資訊為環境預設可辨別的 19 個角度值其中之一(參考表 1)，接著要計算出實驗結果對應到麥克風的角度資訊，在此帶入聲速為 340 cm/sec。

分析誤差來源，遠近聲場效應有其明顯影響力，另外，聲源是否為點聲源也足以影響平面波的行進，再者，聲速在此帶入為 340 cm/sec，這也會有少許誤差在其中。



第六章 結論

6.1 研究成果

本論文已實作完成 OMAP5912 為平台，結合 PC 端計算真人語音偵測(VAD)與聲源角度判定(GCC)的聲音處理系統，並安裝於機器人平台上，可以使用於室內環境。另外也針對 DSP5510 核心作理論驗證與實作，完成聲音演算法於定點處理器平台之上。

測試結論為：

1. VAD 結合 GCC 適用於噪音源變動的環境。
2. 在機器人平台上，馬達一啟動就有一低頻雜訊，可透過 VAD 將此低頻雜訊濾除乾淨，還可透過 VAD 消去路面顛簸的影響。
3. 透過 PC 驗證以及 DSK5510 驗證，可以逐步解決 DSP 端資料處理的問題，也在整合進 OMAP 時較順利。
4. DSK5510 的定點化過程，容易遭到環境音源能量的影響，所以對於環境變數調整必須事前先測量、定義好環境參數。

6.2 未來展望

在 OMAP 平台上整合系統是非常艱辛且耗時的，所以必須透過 PC 驗證以及 DSK5510 來確保每一步驟的正確性，未來在開發此種雙核心平台時，除了對應的 debug 環境上要充分，加強雙核心系統的開發效能，同時也可以繼續開發相關周邊硬體，比如說利用影像演算法加強辨識方位準確度，或是利用影像計算出使用者與機器人平台的相對位置等等，皆可以將機器人智慧型功能再提升一個層次。

参考文献

- [1] Chun, G.D. and Caudell, T.P., “A model for auditory localization in robotic systems based on the neurobiology of the inferior colliculus and analysis of HRTF data,” Proceedings IJCNN '01. International Joint Conference on Neural Networks, 2001, Volume: 2, 15-19 July 2001, Page(s): 1107 -1111 vol.2
- [2] C. Schauer, H.-M. Gross: “Model and application of a binaural 360 degree sound localization system,” International Joint INNS-IEEE Conference on Neural Networks, Washington DC, July 14-19, 2001
- [3] Webb, B., and R. R. Harrison, “Eyes and ears: combining sensory motor systems modelled on insect physiology,” pp. 3913–3918 in IEEE International Conference on Robotics and Automation. San Francisco, 24–28 April 2000.
- [4] Nakadai, K.; Hidai, K.; Okuno, H.G.; Kitano, H., “Real-time speaker localization and speech separation by audio-visual integration,” Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on, Volume: 1 , 11-15 May 2002, Page(s): 1043 -1049 vol.1.
- [5] Nakadai, K.; Okuno, H.G.; Kitano, H.; “Epipolar geometry based sound localization and extraction for humanoid audition,” Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on , Volume: 3 , 29 Oct.-3 Nov. 2001, Page(s): 1395 -1401 vol.3
- [6] Kitano, H.; Okuno, H.G.; Nakadai, K.; Sabisch, T.; Matsui, T., “Design and architecture of SIG the humanoid: an experimental platform for integrated perception in RoboCup humanoid challenge” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 1 , 31 Oct.-5 Nov. 2000, Page(s): 181 -190 vol.1
- [7] Nakadai, K.; Matsui, T.; Okuno, H.G.; Kitano, H., “Active audition system and humanoid exterior design,” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 2 , 31 Oct.-5 Nov. 2000, Page(s): 1453 -1461 vol.2
- [8] Nakadai, K.; Okuno, H.G.; Kitano, H., “Robot recognizes three simultaneous speech by active audition,” Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on, Volume: 1 , Sept. 14-19, 2003, Page(s): 398 -405
- [9] Okuno, H.G.; Nakadai, K.; Hidai, K.I.; Mizoguchi, H.; Kitano, H., “Human-robot interaction through real-time auditory and visual multiple-talker tracking” Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on , Volume: 3 , 29 Oct.-3 Nov. 2001, Page(s): 1402 -1409 vol.3
- [10] Kitano, H.; Okuno, H.G.; Nakadai, K.; Sabisch, T.; Matsui, T., “Design and architecture

- of SIG the humanoid: an experimental platform for integrated perception in RoboCup humanoid challenge” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 1 , 31 Oct.-5 Nov. 2000, Page(s): 181 -190 vol.1
- [11] Nakadai, K.; Matsui, T.; Okuno, H.G.; Kitano, H., “Active audition system and humanoid exterior design,” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 2 , 31 Oct.-5 Nov. 2000, Page(s): 1453 -1461 vol.2
- [12] Y.Matsusaka, T.Tojo, S.Kubota, K.Furukawa, D.Tamiya, K.Hayata, Y.Nakano, T.Kobayashi, "Multi-person Conversation via Multi-modal Interface - A Robot who Communicate with Multi-user -", Proceedings of 6th European Conference on Speech Communication and Technology, vol.4, pp.1723-1726, 1999.
- [13] Y.Matsusaka, T.Kobayashi, "Human Interface of Humanoid Robot Realizing Group Communication in Real Space", Proceedings of the Second International Symposium on Humanoid Robots, pp.188-193, 1999.
- [14] Y.Matsusaka, S.Fujie, T.Kobayashi, "Modeling of Conversational Strategy for the Robot Participating in the Group Conversation," Proc. ISCA-EUROSPEECH2001, pp.2173-2176, 2001.
- [15] C. Breazeal and B. Scassellati, “A context-dependent attention system for a social robot,” in Proc. of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99), 1999, pp. 1146–1151.
- [16] Kitano, H.; Okuno, H.G.; Nakadai, K.; Sabisch, T.; Matsui, T., “Design and architecture of SIG the humanoid: an experimental platform for integrated perception in RoboCup humanoid challenge” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 1 , 31 Oct.-5 Nov. 2000, Page(s): 181 -190 vol.1
- [17] Nakadai, K.; Matsui, T.; Okuno, H.G.; Kitano, H., “Active audition system and humanoid exterior design,” Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, Volume: 2 , 31 Oct.-5 Nov. 2000, Page(s): 1453 -1461 vol.2
- [18] S. Ando, “An autonomous three–dimensional vision sensor with ears,” IEICE Transactions on Information and Systems, vol. E78–D, no. 12, pp. 1621–1629, 1995.
- [19] Ando, S.; Intelligent sensor systems: integrating advanced automatism and optimality into sensors, Solid State Sensors and Actuators, 1997. TRANSDUCERS '97 Chicago., 1997 International Conference on , Volume: 1 , 16-19 June 1997 Page(s): 291 -294 vol.1
- [20] Javier Ramírez , José C. Segura , Carmen Benítez , Ángel de la Torre and Antonio Rubio ,”Efficient voice activity detection algorithms using long-term speech information,” Speech Communication, Volume 42, Issues 3-4, April 2004, Pages271-287
- [21] European Digital Cellular Telecommuni- cations System; Half rate speech; Voice Activity Detection

- (VAD), ETSI GSM 06.42 (ETS 300-581-6), 1995.
- [22] European Digital Cellular Telecommunications System; Half rate speech; Half rate speech transcoding, ETSI GSM 06.20 (ETS 300-581-2), 1995.
- [23] ITU-T G.729, Coding of Speech at 8kbit/s Using CS-ACELP, March, 1996.
- [24] A. Benyassine, E. Shlomot, H. Y. Su, D. Massaloux, C. Lamblin, and J. P. Petit, "ITU recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," IEEE Commun. Mag., vol. 35, pp. 64–73, Sept. 1997.
- [25] RALPH O. SCHMIDT, MEMBER, "Multiple emitter location and signal parameter estimation," IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. AP-34, NO. 3, MARCH 1986
- [26] Parham Aarabi *, Sam Mavandadi, "Robust sound localization using conditional time–frequency histograms." 2003 Elsevier Science
- [27] OMAP5912 Dual-Core Processor DSP Subsystems Reference Guide. Texas Instruments. Dallas, Texas. [Online]. Available: <http://www.ti.com>.
- [28] GNU Compiler Collection, <http://gcc.gnu.org>.

