

# 第一章 緒論

## 1.1 研究背景與動機

國內外發展汽車導航從最初期的筆記型電腦配置電子地圖光碟與導航軟體，發展至今已有成熟的車用導航產品，並且逐漸成為車輛標準配備。而目前國內之車用導航系統除了全球定位訊號接收單元（GPS Receiver）之外，多屬於封閉之系統，所謂的封閉，意指此導航方式不易結合動態資訊如：道路施工、交通事故，或車輛擁塞等資訊，達到符合實際狀況的最佳路線導航。而其電子地圖儲存於光碟中，此種儲存方式最大的優點是容量大，但未來電子地圖的更新將成為問題。另外，車用導航系統中的微電腦運算能力有限，無法如單機地理資訊系統般處理龐大的資料運算。

近年來地理資訊系統之發展趨勢為結合網際網路之 Web-GIS。在這種趨勢中，最大的特色就是以 Server-client 之層級觀念為系統架構。Web-GIS 結合了網路伺服器與 GIS 工作站，使用者可經由網際網路讀取 GIS 工作站中的圖形資料或者是屬性資料，這種趨勢也提高了一般人取得地理資訊的便利性，並降低了使用成本。

本所張晉維[1]先生在此發展趨勢下，基於使用者端顯示版面之限制及資料傳送量之考量，將 GIS 伺服器所產生的導航資訊加以簡化，僅傳送最短路徑之座標資料，在使用者端以簡圖的方式顯示。林詠傑[2]先生在此成果上更加入了圖隨路轉功能、簡圖放大、縮小、平移等，使得使用者能更清楚看到簡圖資訊。張晉維與林詠傑之研究成果不但解決了上述車用導航系統所遭遇的限制，並且將使用者從商用車輛與個人小客車擴大為行人，此與行動通訊市場結合行動電話與個人數位助理（Personal Digital Assistant, PDA）之發展完全吻合。

雖然個人行動導航未來發展具有潛力，但是使用者端的設備為了達到便於攜帶的特點，體積通常較小，以觸控式螢幕來操作軟體。如此也造成使用者在駕駛車輛的同時，或者是步行的同時，因為安全的考量而無法分散注意力來觀看螢幕並操作軟體。因此，本研究即在此需求下，以及張晉維[1]、林詠傑[2]之研究成果上，加入語音辨識及語音提示之功能，將語音辨識及語音提示應用於導航軟體中，提高使用者操作軟體的便利性與行車安全，使導航系統更加完備。

## 1.2 研究目的與課題

本研究的研究目的是在 Server-client 架構下的導航系統中，在使用者端的設備上加入語音辨識以及語音提示功能。語音辨識功能使駕駛者得以簡單、扼要的語音指令控制 PDA 螢幕上地圖顯示的倍率，中心位置，由駕駛者依其需求自行調整至最佳顯示狀態。語音提示功能則盡可能地減少駕駛者注視 PDA 螢幕的需求，並使駕駛者獲得清楚的轉向提示。當駕駛者想獲得前方設施的資訊時，可以以語音的方式聽取設施資訊。在導航系統中加入此二項功能前，需要探討的相關課題主要有以下五點：

- 探討車輛導航系統中，人機雙向之語音需求。
- 建立車輛導航系統中，語音提示之功能。
- 建立車輛導航系統中，駕駛人之指令及其系統操作之關聯。
- 瞭解語音辨識相關技術之發展，及其建立過程。
- 探討提升語音辨識效率之應有作法。

### 1.3 研究架構

由本研究在 1.2 節列出之相關課題所形成的研究架構如圖 1-1 所示

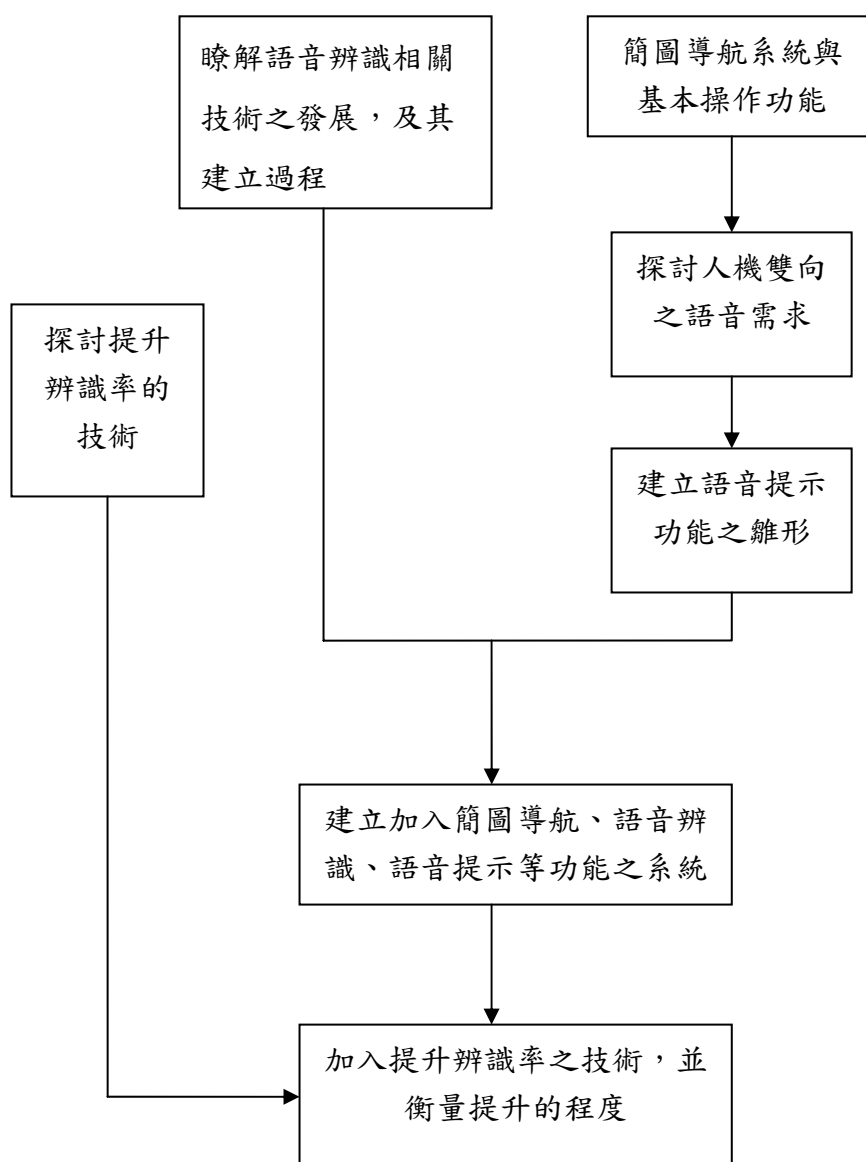


圖 1-1 研究架構圖

#### 1.4 研究範圍

本研究建立於張晉維與林詠傑之導航系統，其系統架構如圖 1-2 所示。本研究著重的語音辨識技術與語音提示功能則應用於使用者端之微電腦系統，因此在開發個人數位助理上的導航軟體時，由全球定位系統(Global Positioning System, GPS) 以及通訊設備所傳入 PDA 的資料，將視為已知的資料。這些資料中，由 GPS 傳至 PDA 的資料為經度、緯度、衛星數、速度；由通訊設備傳至 PDA 的資料為文字格式的導航路徑；由 PDA 傳至通訊設備的資料包含了 GPS 資料與文字格式之起迄點。

語音辨識技術牽涉的問題層面相當廣，尤其是當句子長度增加時，要從句子中拒絕無關的字詞並萃取關鍵詞將會使問題顯得複雜，而且一方面在駕駛車輛時，最重要的是以簡單扼要的命令取代觸控式螢幕，達到操作軟體的目的；另一方面 PDA 的運算能力有限，不容易達到萃取關鍵詞的功能。

因此，在語音辨識方面的研究範圍將集中於特定詞彙的辨識，這項限制也影響到本研究建構使用者輸入起迄點的功能，因為若以語音輸入起迄點，以特定詞辨識的技術無法辨識出大量的地名，必須以大字彙辨識技術才可能達到輸入起迄點的功能，關於大字彙辨識技術，在第三章將有詳細的說明，此辨識技術對於 PDA 主要的限制是在運算能力。但本研究使用較少資料量的路名供使用者輸入迄點，以特定詞辨識技術建構起迄點輸入功能的雛形，示範性地實現語音輸入迄點的功能。未來隨著 PDA 處理器速度達到較高的水準後，將可望使用大字彙辨識技術辨識完整的路名資料庫。

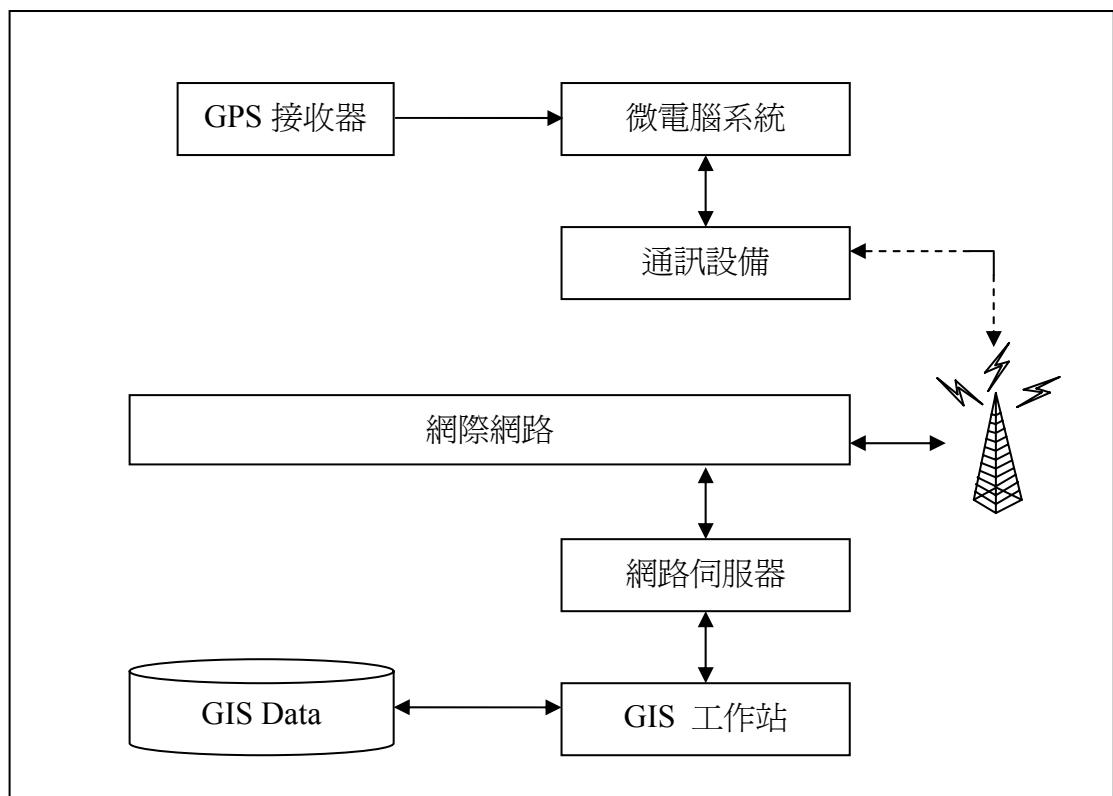


圖 1-2 本研究之導航系統架構圖

## 1.5 研究內容與方法

本論文共分為七章，其內容與方法分述如下：

### 第一章 緒論

說明本論文之研究動機與背景、目的與課題、範圍、內容、方法與流程。

### 第二章 文獻回顧

本章由說明 Server-client 架構之導航系統開始，回顧張晉維與林詠傑之研究成果。而各電機資訊學院有許多關於視覺辨識，語音辨識的相關論著，因此第二章將著重於收集與本研究有關之語音辨識文獻，並加以整理，作為建立導航與語音操作系統之前置作業，以及語音辨識方法的基礎。

### 第三章 人機雙向之語音需求

本章前半部份探討原先導航系統軟體的操作流程，思考以語音作為使用者介面的情況下，操作流程將如何重新設計。並探討使用者在何種情況下需要獲得語音提示。而由於語音提示在技術上是以文字轉換聲音的方式呈現，難度較語音辨識來得低，因此在本章後半部份將以 Visual Basic 6.0 之程式語言建立語音提示的雛形，在第五章與簡圖導航、語音辨識等功能結合。

### 第四章 語音辨識的原理與技術

語音辨識的主要概念是透過資料處理的方法求得語音的「特徵值」，在這過程中所要探討的方法如下：

- 以能量法去除靜音部分
- 預強調與漢明窗處理
- 由自我相關係數求線性預估係數—杜賓演算法 (Durbin algorithm)
- 動態校準法 (Dynamic Time Warping)

### 第五章 建立加入簡圖導航、語音辨識、語音提示之系統

本章在 Microsoft eMbedded Visual Basic 3.0 上之 Pocket PC Emulator 的環境下開發結合了簡圖導航、語音辨識、語音提示等功能之導航軟體，並依第三章所探討之程式流程讓使用者以語音為介面順利操作軟體，最後記錄其辨識率。

### 第六章 提升辨識率的技術

透過第二章之文獻回顧以及第四章之語音辨識基本技術，整理出文獻中提高辨識率的技術，並將其應用於第五章所開發之導航軟體，比較此技術提高或降低語音辨識率之程度。

### 第七章 結論與建議

本章將根據本研究之最後結果提供建議，供日後相關研究參考用。

## 1.6 研究流程

本研究之研究流程圖如圖 1-3 所示。

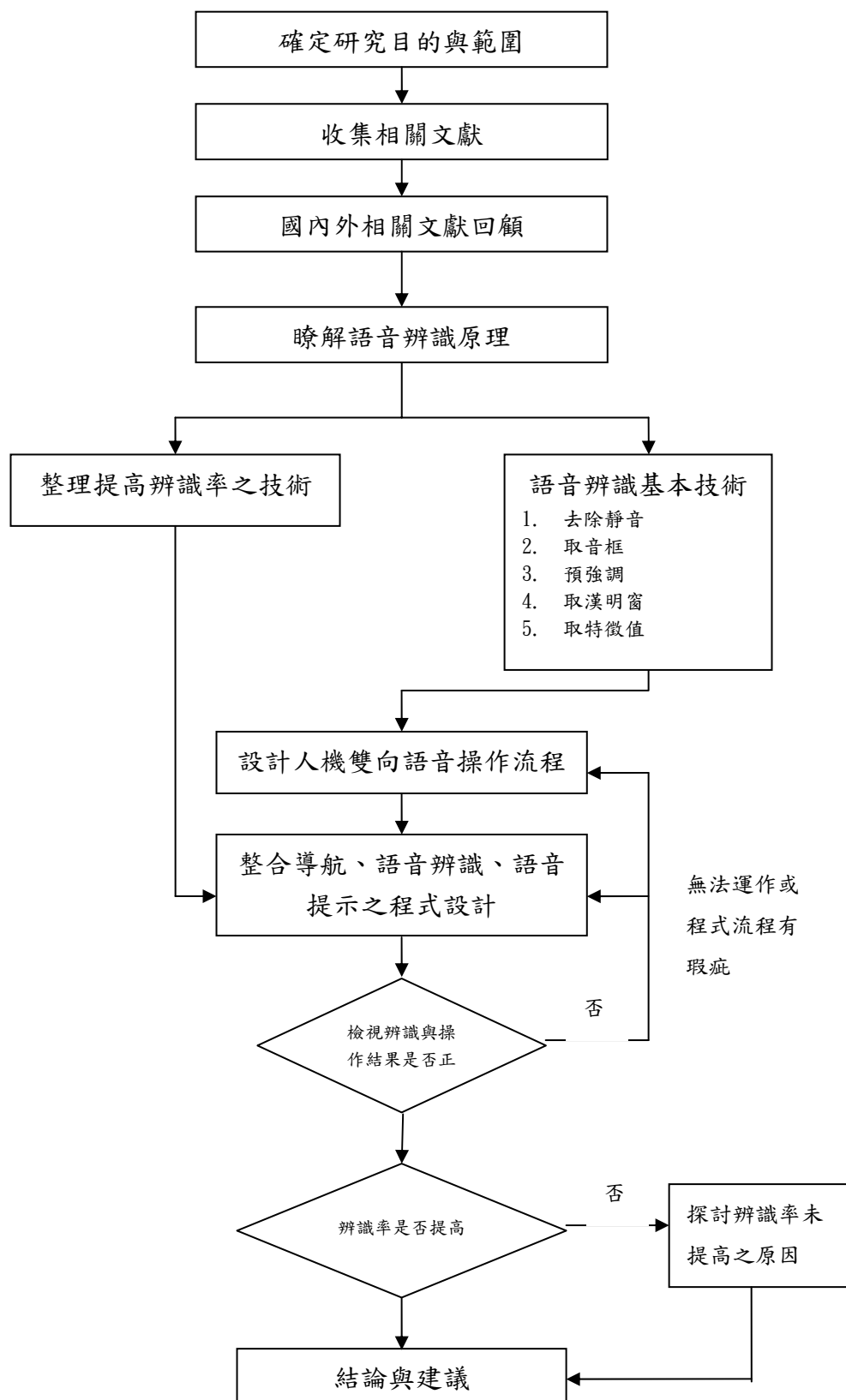


圖 1-3 研究流程圖

## 第二章 文獻回顧

本章將回顧與本研究有關之文獻，首先為兩篇簡圖顯示之路徑導航系統架構下的相關文獻，以此回顧為出發點可瞭解相關研究在此課題上所獲得之成果以及建議。第三篇與第四篇文獻為語音辨識之相關文獻，其語音辨識方法為本研究建立語音辨識功能之基礎。第五篇文獻主要重點為 GPS、GIS 之技術原理敘述，以及駕駛感知時間之相關研究，供本研究於導航系統中設計語音辨識與語音提示之參考。

### 2.1 簡圖顯示之路徑導航方面文獻

張晉維[1]指出，以往在GIS導航路徑系統中大多呈現單機作業的方式，利用單機之GPS接收器與GIS電子地圖系統，透過單機之微電腦系統處理資訊，產生如最短路徑指派等資訊，由於單機之運算能力有限，使得運算速度受到影響；此外最短路徑不應只有考慮到距離的限制，其他諸如道路現況、壅塞情形等均是影響最短路徑產生的關鍵因素，因此作者主要以Server-Client的觀念將最短路徑處理之功能放置於資訊中心的主機上，利用主機強大的運算功能並結合最新之交通道路狀況進而產生最短路徑資訊，透過無線傳輸至汽車駕駛人的接收裝置上，並將結果展示出來。

利用此一觀念作者針對GIS導航路線簡化圖示方法作一探討，主要考量為GIS在無線通訊傳輸上的限制，並進一步將導航路徑資訊簡化成關鍵資訊（點位+道路名稱）加以傳送，縮小檔案容量以增進無線傳輸上的可能性，並將此架構利用跨平台之J2ME程式構建一初步展圖模型於PDA（PALM作業系統）上，證實其可行性。在此篇文章中作者主要是建立GIS導航路徑簡化圖示的模型，藉此能讓汽車駕駛人透過簡圖或是文字敘述的方式快速了解行進路線，方便駕駛人了解系統所產生之導航資訊。

該研究成果包含了：

#### 1. GIS系統回顧

整理GIS系統於台灣現況應用之限制與問題，作一深入探討，並將所有地圖檔在MapInfo中轉為二度分帶座標系統，再轉為台灣經緯度座標系統，方便使用。

GIS系統應用層面廣泛，作者認為與交通運輸相結合，將有效地處理交通問題與決策，以現階段而言將可直接建立運輸地理資訊系統(GIS-T)，從整體的觀點出發。常見的應用有電子地圖與汽車導航，以現代人對時間的重視與要求，透過先進的資訊系統來獲得所需要的交通資訊則可增進現代人在交通上的效益。作者比較傳統地理資訊系統與網際網路地理資訊系統(Web-GIS)的特性，以Web-GIS而言，是以Server-Client的層級觀念，在用戶端只需要一個終端機就可存取大型的地理資訊系統，但功能亦有所限制，大多用於查詢地圖、地圖縮放移動等。

無線網際地理資訊系統則是將網際地資訊系統擴展至無線通訊領域，除具備網際地理資訊系統的架構外，再加上無線通訊的應用，將較

重的計算工作建構於Server端，而Client端只負責查詢與顯示結果，作者即以此架構構建車用導航系統之雛形。

## 2. 無線通訊系統探討

探討台灣無線通訊系統的現況，分析目前台灣普遍通行的GSM系統在傳輸資料上之限制，因此在GSM系統頻寬目前無法立即有效擴大的情形下，作者將GIS導航路徑資訊簡化視為車用導航系統必經之路。

## 3. GIS簡化圖示資料決定

在傳送導航路徑點位資訊於用戶端設備展圖時，需先在GIS Server端產生導航路徑點位數值資訊，並在行動設備端開發展圖程式，如此才能在用戶端如PDA上顯示簡化後之導航路徑。首先作者分析原本GIS所包含的資料，如座標系統格式、圖層格式等。由於GIS系統軟體中所儲存之座標格式與台灣2度分帶座標系統不相同，且軟體無內建轉換參數，因此作者將檔案格式轉換成台灣2度分帶座標系統。在圖層資料方面，GIS資料庫中包含有地形地物、道路、縣市界等多種圖層，但並非每一種圖層都對導航資訊有用，因此作者只以台灣省縣市區界圖層和道路圖層為主，其餘刪除。

由於作者並非以最短路徑為研究主題，因此最短路徑之產生係由GIS Server所決定，作者將結果取出後，呈現在簡化後的道路圖層上，並同時取出關鍵點位資訊和道路名作為在PDA上展圖的基礎。故最後輸出資導航路徑資料有點座標、道路名稱。

## 4. 圖示之簡化

作者認為GIS導航路徑圖示簡化之主要目的是因無線通訊之頻寬限制，行動設備端也無法處理龐大的資料量。在路徑訊息認知方面，作者認為GIS所提供之資訊過於龐大，且有些甚至很難在短時間以簡單的方式去理解，可由圖2-1所看出作者的想法，在無線網際GIS中，駕駛者是需要簡單、明瞭之路徑導引訊息。

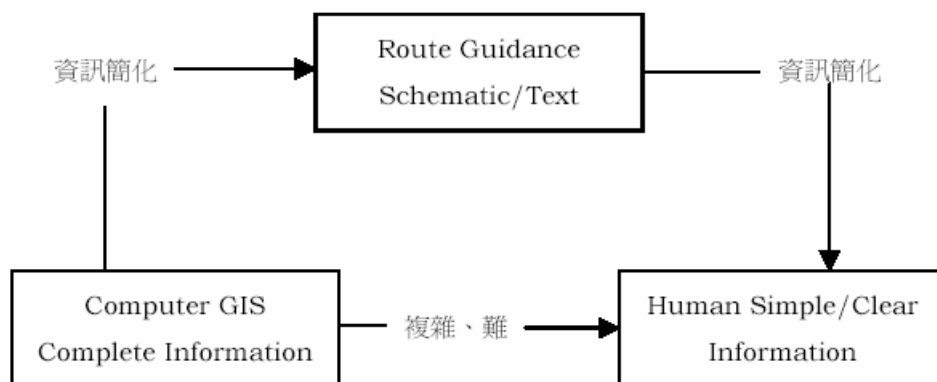


圖2-1 路徑訊息認知圖

作者以JAVA之J2ME語言開發程式，以PALM為作業平台。由於作者考量到未來行動通訊產品，皆有導向嵌入式程式發展的趨勢，故選用此語言加以開發一展圖程式雛形。開發流程如圖2-2所示。展圖程式將所接收到的導航路徑的點位資料（經緯度資料）和道路名稱，經過計算產生簡圖顯示於Palm上如圖2-3所示，展圖程式雛形檔案量大小為2.02KB，檔行路徑數值檔檔案大小為.32KB（以作者案例而言），因此可應用於無線傳輸上，證實其可行性。

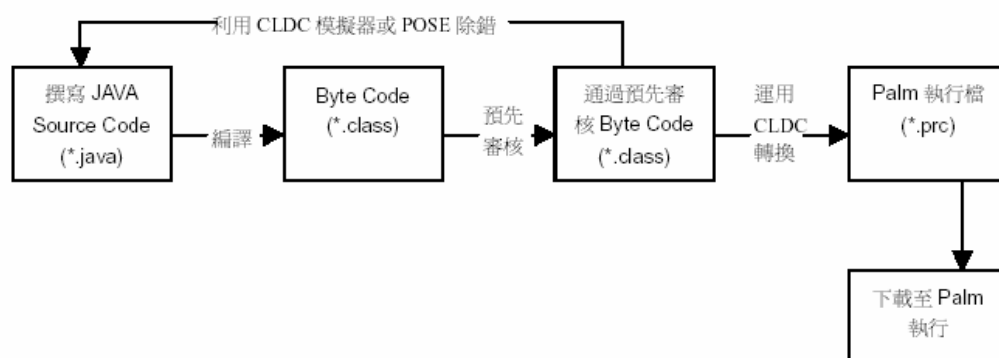


圖2-2 J2ME程式編輯流程



圖 2-3 Palm 導航資訊簡圖

## 5. 導航路徑文字導引

導航文字導引應用相同點位資訊，透過空間幾何做為圖形轉換文字之依據，描述路徑包含：起始點位、行進方向、行進距離、轉折方向、轉折後位置等。在轉向語意方面，由於台灣可能有五叉路甚至六叉路口的情形，因此在轉向文字導引上就顯得特別重要，作者以轉向角度大小依序給道路編號，然後依編號轉為文字導引，其判斷方式為：

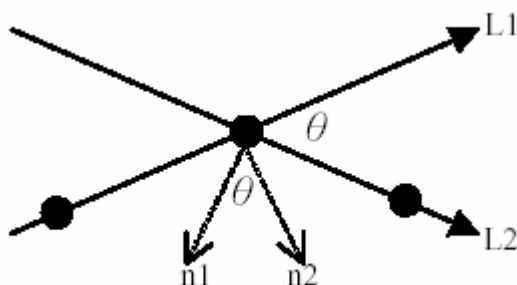


圖2-4：路徑角度關係

以圖2-4來說明如何求得兩直線之交角  $\theta$ ：

$$\begin{aligned} \text{兩直線方程式為：} \quad L1: a_1x + b_1y + c_1 &= 0 \\ L2: a_2x + b_2y + c_2 &= 0 \end{aligned} \quad (2-1)$$

$$\text{設 } n1、n2 \text{ 為 } L1、L2 \text{ 的法向量，} \quad \begin{aligned} \vec{n}_1 &= (-b_1, a_1) \\ \vec{n}_2 &= (-b_2, a_2) \end{aligned} \quad (2-2)$$

則兩直線之交角  $\theta$  為  $\vec{n}_1, \vec{n}_2$  之夾角  $\theta$ ，另一夾角為  $\pi - \theta$ 。

$$\cos \theta = \frac{\vec{n}_1 \cdot \vec{n}_2}{|\vec{n}_1| \cdot |\vec{n}_2|} = \frac{a_1a_2 + b_1b_2}{\sqrt{a_1^2 + b_1^2} \cdot \sqrt{a_2^2 + b_2^2}} \quad (2-3)$$

亦即： $a_1a_2 + b_1b_2 = 0$ ， $\theta$  為直角，

$a_1a_2 + b_1b_2 > 0$ ， $\theta$  為銳角，

$a_1a_2 + b_1b_2 < 0$ ， $\theta$  為鈍角。

在最後作者並構建敘述性文字語意詞彙以供導航程式轉化之用，如表2-1所示。

作者完成了上述之成果，本研究將以此成果為基礎，更進一步選擇適當之作業平台發展展圖程式，並且將此展圖程式之操作介面以語音取代。

表2-1 敘述性文字語意詞彙轉換值表

項目	位置詞彙	方位詞彙	動作詞彙
起始點位	道路交差口名稱或地點名	東西南北	出發
行進方向			直行
行進距離	公尺/公里		
轉折提示距離	公尺/公里	前方	
轉折方向	道路名稱		左轉、右轉、左右迴轉
轉折角度判斷		(左或右方) 第N條道路	
轉折後位置	道路名稱		

林詠傑[2]對無線通訊機制、行動設備、定位系統與 GIS 系統加以探討，選擇以 CDMA 作為通訊技術、選用 GPS 系統定位，並利用 WinCE 作為行動設備之作業系統，以 Microsoft eMbedded Visual Basic 3.0 開發客戶端簡圖展圖程式，達成「圖隨路轉」、「轉向提示」、「地圖縮放」等功能。研究成果經實際路徑驗證在資訊接收時，每個點資訊量平均為 58.2 Bytes，以一次展現五個路口之路段而言，客戶端仍有相當充裕的容量提供其他相關資訊展現使用。

## 2.2 語音辨識方面文獻

李上銘[3]在語音辨識的流程中選擇了特徵參數擷取及隱藏式馬可夫模型的參數這兩個地方來研究。最主要觀念是主成份分析。在特徵參數的抽取中，對語料的頻率領域做「主成份分析」以取代傳統的三角濾波器。從實驗結果發現在沒有雜訊時，主成分分析的方法和傳統梅爾倒頻譜係數的效能差不多。但因為在白色雜訊的假設下這方法會有最大信噪比，實驗證明在加了白色雜訊 30db, 20db, 10db 的雜訊環境下其效能會比梅爾倒頻譜係數好。這個方法與頻譜消去法以及倒頻譜平均消去法的結合時也證實有不錯的加成性。

在隱藏式馬可夫模型的參數分析方面，對最大相似性線性回歸 (MLLR) 參數做主成份分析，並比較其與特性語音 (EigenVoice)、最大相似性線性回歸的效能。在這個部份中，一個大字彙的辨認實驗被用來測試。實驗證明在調適語料少的時候這個方法會比最大相似性線性回歸好。

李健平[4]研究認為語音輸入操作電腦是語音訊號處理之重要應用。該研究以 C 語言撰寫語音辨識軟體，因為 C 語言可以得到最低階的硬體控制。主要研究目的

是透過聲控指令，來操作一些基本的執行動作，因此該系統使用特定的詞彙辨識技術。

首先以 API Function 來達到錄音、播放，以及存檔的功能，整理如表 2-2。

表 2-2 聲音輸出輸入介面 API 功能敘述

API 功能敘述	API 名稱
查詢與打開的音頻波形的輸出入設備	waveInGetDevCaps 、 waveInGetNumDevs 、 waveInOpen 、 waveOutGetDevCaps 、 waveOutGetNumDevs 、 waveOutOpen
配置音頻波形的資料緩衝區塊	waveInPrepareHeader 、 waveOutPrepareHeader
播放與錄製音頻波形檔案	waveInStart、waveOutWrite
使用視窗訊息到管理處理系統音頻波形的播放與錄製	waveInAddBuffer 、 waveOutWrite
結束音頻波形的播放與錄製	waveInStop、waveOutReset
釋放音頻波形的資料記憶體區塊	waveInUnprepareHeader 、 waveOutUnprepareHeader
關閉音頻波形的輸出設備	waveInClose、waveOutClose

達到錄音的功能之後，接下來是對語音訊號作前置處理以及特徵參數抽取。前置處理包含了以下三點：

1. 端點偵測：目的是將語音訊號中靜音或背景雜音的部分去除，也就是要確定語音訊號中，什麼時候是使用者所輸入的語音，什麼時候是背景雜音或靜音。使用的方法除了包含端點量測配合越零率量測，更加入了去除直流偏移的處理，以準確地判定語音訊號起點與終點。
2. 預強調處理：根據人耳的研究指出，人耳在 1KHz 到 4KHz 的頻率帶內具有+6dB/octave 的高頻提升特性，通常使用一階的有限脈衝響應濾波器（FIR Filter）來模擬人耳的特性。其濾波器時域公式如下：

$$y(n) = x(n) - 0.95 \cdot x(n-1) \dots\dots\dots (2-4)$$

其中  $x(n)$  為時刻  $n$  的輸入訊號， $y(n)$  為時刻  $n$  的輸出訊號

3. 取信號窗：在處理數位訊號時，常考慮到訊號的時變特性，因此一次只取一小段處理，為了使相鄰兩音框的頻譜變化不至於太大，通常會讓相鄰兩

音框有部分的重疊。且為了減少音框邊界不連續的影響，在處理每個音框前會先乘上一個視窗函數 (Window Function)，使兩邊界壓低保持音框之間適當的聯繫性，一般採用漢明窗 (Hamming Window)

$$W(n) = \begin{cases} 0.54 - 0.46 \cdot \cos\left(\frac{2\pi \cdot n}{N-1}\right) & , 0 \leq n < N \\ 0 & , otherwise \end{cases} \dots\dots\dots (2-5)$$

在特徵參數的抽取部分，該研究以離散傅利葉轉換 (Discrete Fourier Transform, DFT) 將訊號由時域轉為頻域：

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j \frac{2\pi kn}{N}} \quad , k = 0, 1, \dots, N-1 \dots\dots\dots (2-6)$$

其中  $x(n)$  為時域上的訊號波形，但真正在撰寫程式時，通常是採用快速傅利葉轉換 (Fast Fourier Transform, FFT)，以大幅降低計算量。在將訊號轉換為頻域後，經過梅爾三角濾波器組，便可以得到語音的特徵參數。

至於最重要的辨識核心，目前常見的語音辨識方法主要有隱藏式馬可夫模型 (Hidden Markov Model, HMM) 及動態時間校準 (Dynamic Time Warping, DTW) 兩種，該研究基於演算的速度考量，選擇運算量較低的動態校準法。將測試語音的參數與標準語音資料庫中的參數一一比對，做失真量的計算，進而從中找出一組失真量最小的，即為辨識的結果。

最後的結論指出，由於在語音特徵參數抽取是採用 MFCC，雖然在混淆音的辨識上比 LPC 更能分辨出不同之處，但會使得受到環境雜訊以及使用者音量影響的相同語音指令，卻辨識出不同的結果。所以未來可以再加上一些雜訊消除的方法，把環境雜訊對語音辨識的影響降低。

而在使用的方便性上，使用者要使用某語音指令前，必須先經過訓練的過程，才能使用此語音指令。若能將此系統推廣至非特定語者系統，利用事先訓練好的語料，免除使用者在使用前必須先訓練的麻煩，便能提升使用的便利性。

## 2.3 汽車導航系統之原理與發展

吳玉珍、王穆衡、何毓芬[5]認為全球衛星定位系統(Global Position System; GPS)技術已成熟，再結合詳細精確的數值地圖以及完整豐富的地理資訊系統 (Geographic Information System; GIS)，使得國內外從初期的以筆記型電腦配置導航軟體，漸漸發展出成熟的車用導航系統產品，甚至已有部份車商將車用導航軟體視為車內配備的一部份，除得以幫助使用者拋棄過去傳統邊開車邊找地圖的不便，更可結合即時路況資訊，避開壅塞路段以迅速抵達目的地，讓使用者真正感受到科技帶給人們生活品質的提昇與便利。

該研究針對車用導航系統的架構做一介紹，進而瞭解國外 (包括歐美及日本) 車用導航系統之演進發展，並將國內目前於市場上已發展的車用導航系統之裝置方式、資訊顯示方式及資訊顯示內容做一整理分析，最後提出我國要發展本土車用導航系統之重要關鍵課題 (包括系統未來發展之重要趨勢、系統發展可能會帶

來之行車安全影響)。本研究針對有關的部分，例如全球定位系統原理與誤差、電子地圖的種類，以及車內設備輸出輸入介面對行車安全的影響做一整理，以充分瞭解汽車導航系統之子系統的構成與原理。

目前市面上常見的車用導航系統，大多是在個人之小客車上裝置 GPS 衛星接收器、GPS 天線、配合所設計之電子地圖及展圖之軟體、LCD (或 TFT) 螢幕…等，其主要原理是利用各種感測器結合信號處理、人工智慧及資訊科技，求得車輛正確位置，將個人汽車即時位置及相關道路顯示出來，以提供駕駛者進行查詢、最佳路徑指引、導航等功能。為增加其車輛定位之準確性，也有車輛另外配置了指南針、輪胎行走距離感應器、揚聲器、壓電陀螺儀等儀器輔助，如圖 2-5 所示。

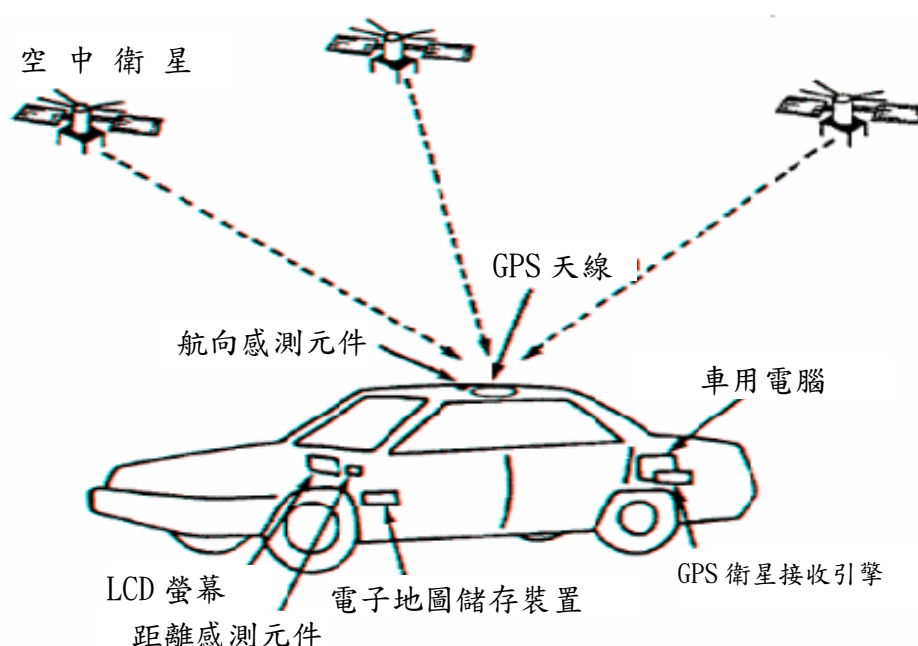


圖 2-5 車用導航系統架構圖

本研究之簡圖路徑導引系統中，全球定位系統與電子地圖可謂是汽車導航系統中最基本也是最重要的兩部分，建立語音輸入與語音提示功能於其上之前，實有必要對兩者有清楚的瞭解，如此可掌握全球定位系統與電子地圖之工作原理、合作關係，並由使用者需求的角度來妥適地加入語音功能。

## 一、全球定位系統

簡稱 GPS 的全球衛星定位系統，為美國國防部開發，利用規模遍及全球的人造衛星之航法系統，是一種結合衛星及通訊發展的技術，原本是美國國防部因應軍事定時、定位及導航系統之目的而發展，現多作為商業及測量用途。

目前 GPS 衛星已發展至 Block II 型式的定位衛星，由 Rockwell International 製造，在軌道上重量約 1,900 磅，太陽能接收板長度約 17 呎，預期壽命為 7.5 年，於 1994 年完成第 24 顆衛星的發射。因此

目前太空中有 24 顆 GPS 衛星可供定位運用，其中包括三個預備衛星，它們平均分佈於 6 個軌道面，每個軌道面上各有 4 顆，距離地面高度約 10,900 海哩（大約 20,000 公里），呈  $55^\circ$  角傾斜繞行地球運轉，繞行地球一周需 12 恆星時，每日可繞行地球 2 周，這也就是說，不論任何時間、任何地點，至少有 4 顆以上的衛星出現在我們的上空。

目前全球有五個地面衛星監控站，分佈於夏威夷、亞森欣島、迪亞哥加西亞、瓜加林島、科羅拉多州，這些衛星地面控制站，同時監控 GPS 衛星的運作狀態及它們在太空中的精確位置，主地面控制站更負責傳送衛星瞬時常數(Ephemeris's Constant)及時脈偏差(Clock Offsets)的修正量，再由衛星將這些修正量提供給 GPS 接收器做為定位運用。

GPS 的定位是利用衛星基本三角定位原理，GPS 接收裝置以測量無線電信號的傳輸時間來量測距離，以距離來判定衛星在太空中的位置，這是一種高軌道與精密定位的觀測方式。假設衛星在 11,000 英哩高處，測量我們的距離，首先以 11,000 英哩為半徑，以此衛星為圓心畫一圓，而我們位置正處於球面上。

再假設第二顆衛星距離我們 12,000 英哩，而我們正處於這二顆球所交集的圓周上。現在我們再以第三顆衛星做精密定位，假設高度 13,000 英哩，我們即可進一步縮小範圍到二點位置上，但其中一點為非我們所在的位置極有可能在太空中的某一點，因此，我們捨棄這一點參考點，選擇另一點為位置參考點。

如果要獲得更精確的定位，則必定要再測量第四個顆衛星，從基本物理的觀念上來說，以訊號傳輸的時間乘以速度即是我們與衛星的距離，我們將此測得的距離稱為虛擬距離，在 GPS 的測量上，我們測的是無線信號，速度幾乎達 18 萬 6 千英哩/Sec 的光速，而時間卻短的驚人，甚至只要 0.06 秒，時間的測量需要二個不同的時錶，一個時錶裝置於衛星上以記錄無線電信號傳送的時間，另一個時錶則裝置在接收器上，用以記錄無線電信號接收的時間，雖然衛星傳送信號至接收器的時間極短，但時間上並不同步，假設衛星與接收器同時發出聲音給我們，我們會聽到二種不同的聲音，這是因為衛星從 11,000 英哩遠的地方傳來，所以會有延遲的時間，因此，我們可以延遲接收器的時間，從此延遲的時間乘與速度，就是接收器到衛星的距離，此即為 GPS 的基本定位原理。

而 GPS 誤差的來源主要有：

#### 1. 氣層上的延遲誤差

當 GPS 訊號經過電離層 (Ionosphere) 上一些帶電性的粒子及對流層 (Troposphere) 上的水汽時，訊號便會產生角度的偏離而產生遲滯的現象。

#### 2. 多路徑訊號傳送誤差 (Multipath Error)

GPS 訊號傳送並非從衛星上直接到達地面接收器，在到達接收

器時，通常會經過各種地面物質的反射，因此會有一種訊號多重定位的疊合誤差，這種情形就類似您有時在電視上會看到影像重疊不合的鬼影現象，即是此種誤差效應的影響。

### 3. 衛星時錶誤差 (Ephemeris Error)

即使衛星是非常的精密複雜，它可以計算出一些極微小的訊息資訊，如原子鐘 (Cesium) 即是如此一個精準的裝置，但是精準並不代表完美，因此仍會有一些微小的誤差產生，即使衛星的定位會持續的被監控著，但並不是每一秒都處於被監視的狀態之中，這期間一旦有微小的定位誤差或衛星星曆的誤差產生，便會影響到接受器在定位計算時的準確性。

### 4. 幾何精度稀釋 (Geometric Dilution of Precision, 簡稱 GDOP)

幾何精度稀釋聽起來相當的複雜抽象，其實它的原理非常的簡單，在上述我們所提到的，一個接收器可以在同一時間得到許多顆衛星定位資訊，但在精密定位上，只要四顆衛星訊號即已足夠了，一個好的接收器便可判斷如何在這些衛星靠近的訊號來計算，如果接收器所選取的訊號當中，有二顆衛星距離甚近，則這二顆衛星訊號在角度較小的地方會有一個重疊的區域產生，隨著距離愈近，此區域便愈大，影響精度的誤差亦愈大。如果選取的衛星彼此相距有一段距離，則訊號相交之處便較為明確，誤差當然就縮減了不少。

### 5. 接收器內部誤差 (Receiver Error)

市面上 GPS 接收器琳瑯滿目，目前世界上 GPS 接收機之機型已超過一百種，有五十家以上的製造廠商。好的接收器可精準計算出一個位置方位量，相對的，較差的接收器亦可能存在著自己本身的誤差，而嚴重的影響了精度，接收器除了訊號的接收及計算能力之外，誤差的修正能力也是一個好的接收器所必須考慮的重點，當然，接收器穩定性及壽命亦是影響精度的其中因素。

### 6. 選效誤差 (Selective Availability, 簡稱 SA)

這是美國國防部花了一千二百萬美元所發展的干擾衛星定位精度的技術，其目的在於防止不法之徒或恐怖組織，利用 GPS 來發展高精密的武器，危害到國家安全，因此基於國家安全的考量，美國國防部會不定期的傳送一些不正確的軌道訊號及干擾訊號至衛星，而使其精度降低，只有國防軍事用的接收器具有解碼裝置，可排除 SA 的誤差。但有鑑於 GPS 之廣泛應用，多年來業者一直促請政府放寬對 GPS 精確通訊的管制，因此美國已於 2000 年 5 月宣佈解除此項管制。

有鑑於全球衛星定位系統 (GPS) 在商業及救難任務之運用日益普遍，美國總統柯林頓下令軍方停止多年來為保護國防機密而故意降低民用全球定位系統精確度的做法，讓配有 GPS 接收器的汽車、飛機、漁船、登山者、滑雪者今後可以接收準確度極高的衛星訊號。這道行政令自台

灣時間 2000/5/2 上午八時起正式生效。

在美國決定解除 GPS 精確通訊的管制前，白宮、五角大廈、中情局、其他相關單位針對此問題討論了四年，直到開發出可以封鎖某個地區 GPS 通訊的新技術才同意放行。根據這道行政令，美國當局已不再擾亂民間 GPS 用戶發送的信號編碼，不過藉由新的技術，仍可隨時封鎖某個地區的 GPS 通訊，避免敵軍或恐怖組織使用此一準確度極高的衛星通訊。而軍方也會繼續使用編碼、精確度高的 GPS 通訊，導引諸如飛彈等精確命中目標的武器。

報導指出，停止降低 GPS 訊號精確度之後，民間現在所用的各式衛星導航設備，判讀的精確度將大幅提升到之前的十多倍，所以無須汰舊換新，購買新的接收器。除此之外，警方、消防人員、救難小組今後可以憑著 GPS 接收器找到確實的失事地點，將原本相當於足球場的搜尋範圍縮小到和網球場一般大小。以重大交通事故為例，目前醫護人員靠著 GPS 設備只能鎖定方圓一百公尺的範圍，若剛好在多條公路交錯的交通要道，找尋目標更是困難重重，進而延誤救援時間，若能將精度提高，就可避免彎繞多餘道路而延誤急救時機。

未來可見，在管制解除後，GPS 未來的商業用途將更加廣泛，包括陸海空交通導航、農作物栽種、採礦、探勘石油、通信、電子數據傳輸、營建、娛樂、緊急救援等。少了軍方干擾，預料 GPS 應用市場的營業額未來三年將倍增，由每年的八十億美元增至一百六十億美元。

另外一方面，民眾使用 PPS (Precis Position System) 接收器的障礙可能已比以前更較容易掃除，由於共產鐵幕已大部份被破除，美國已不再有蘇聯這個頭號敵人，因此精密定位訊號被敵人盜用的風險已大為減小，因此美國政府已同意非軍事用途的 PPS 系統開放申請使用，但是仍須依個案審核。

不管怎麼說，目前市面上已經有差分型 GPS (Differential GPS-DGPS) 的系統被廣泛應用，此系統利用地面的 DGPS 無線電基地台對特定的區域播送誤差修正訊號，這些基地台均經過細密的測量，在美國其訊號提供給海岸防衛隊及一些商業團體使用。這些基地台不斷傳送該區域的衛星誤差值給具備差分解調能力的 GPS 接收器，以得到精確的定位訊號。

另外有一種非常精確的 DGPS 地理測量系統，基本上這種系統是由兩個或三個測量參考站及幾個移動式的接收器所組成，它們要花很長的時間搜集定位數據以得到很小範圍的相關資料，參考站的數據可以用來修正移動接收器的定位數據，這種設備有些非常的昂貴，高級的測量設備其精密度可以達到 1 公分。

根據實際的測試結果，基本的 DGPS 系統即使是 SPS-GPS 而且是在 SA 干擾的情況下，其精確度也在三公尺以下。即使是在高速公路行駛，系統也會顯示你是行駛在內車道或外車道上(假設此高速公路有 DGPS 訊號的廣播服務)。美國計劃從 1999 年開始利用通訊衛星播送 DGPS 訊

號，但其服務範圍只限於北美地區，也就是說即使一般大眾僅能使用SPS標準定位系統，屆時也能擁有相當精密的定位服務。

## 二、電子地圖

早期人們僅使用一般的紙圖來記錄所要表示的相對關係，而這類的資料型態稱之為類比式（Analog）資料。這種記錄方式容易因年代久遠而有圖紙毀損、圖紙伸縮等保存問題的產生，加上無法以類比式資料進行套疊分析（Overlay Analysis）及路網分析（Network Analysis）等空間分析，資料無法與圖形產生關連，這些問題與障礙使一般以圖紙來表示空間地理關係的方式已無法滿足社會大眾。

自電腦科技迅速發展後，使地圖的記錄方式從傳統的類比式資料變成數值化（Digital）資料，將地圖數位化（Digitize）後以電腦來記錄圖形的資料，解決了上述問題。

而後為使數值化地圖能發揮更大的功效，而不僅侷限於展示的功能，便將地圖的製作移植到GIS的軟體，建置電子地圖GIS資料庫；藉由GIS能將圖形和屬性資料相連結之特性，可作更廣泛的應用、查詢與空間分析，使電子地圖所能達到的功能更為龐大。

電子地圖（Electronic Map, E-map）是數值地圖（Digital Map）的通俗名稱，係將普通地圖或地形地物現況以數值方式紀錄，所建置而成的數值化地圖資料，稱為數值地圖。廣義的電子地圖泛指各領域的數值地圖；狹義的電子地圖僅指運輸領域所使用的數值地圖。在ITS的應用方面，電子地圖是指車輛定位系統、行車導引系統、用路人及乘客資訊系統等在車上或資訊中心或管理中心所使用的數值道路地圖（Digital Road Map）。因此就智慧型運輸系統（Intelligent Transport System, ITS）術語而言，電子地圖專指數值道路地圖。

### 1. 電子地圖的技術種類

數值地圖依其資料紀錄儲存的方式區分為以下三種：

#### （1）向量式（Vector）數值地圖

以一連串的座標紀錄相關空間資料，在運輸領域的應用上，一般可再分為具有網路特性與不具網路特性兩種。

##### ①具網路特性向量式數值地圖

具備節點與節線位相關係之數值地圖，在運輸領域之應用包括運輸規劃之交通量指派、商用車動態車輛指派及行前旅次規劃等課題均需採用此類型的數值地圖。

##### ②不具網路特性向量式數值地圖

不具備節點與節線位相關係之數值地圖，此類數值地圖由於無法從事相關網路分析作業，一般適用於不需涉及網路特性之分析應

用，包括道路設施管理、肇事分析、即時路況資訊展示等課題。

## (2) 網格式 (Raster) 數值地圖

以均一的網格為基本單位來紀錄空間資料，適合以網格方式建置的數值地圖主要為地質、地形資料，在運輸領域之應用包括新闢運輸系統的選線、工程廢棄土的選址等課題。

## (3) 影像式 (Image) 數值地圖

以均一的像元 (Pixel) 為基本單位來紀錄空間資料，此類數值地圖一般以掃描器 (Scanner) 直接紀錄儲存地圖資料。在運輸領域上，除可作為向量式數值地圖與網格式數值地圖之背景資料方便空間資料查閱之外，亦可作為車輛導航、即時路況資訊展示及觀光資訊查詢顯示等課題之應用。

## 2. 電子地圖的功能與效益

數值化道路地圖是許多智慧型運輸系統發展必備的基本圖，是顯示車輛位置、道路交通狀況、及導引車輛與管理交通不可缺少的基本工具，也是系統間相互溝通最基本的資料庫，對行車導引系統的使用者及道路管理者均有很大的助益。此外，電子地圖也可用於道路維修與道路規劃，如果缺乏標準化的數值化道路地圖，則交通資訊的傳播及溝通會遭遇困難與阻礙，因此，無論政府機關或民間業者都亟需標準化的數值化道路地圖，未來道路管理者與用路人如有必要進行雙向通訊時，使用相同的數值化道路地圖才能有效的通訊，並得知用路人所在的位置。

## 3. 電子地圖的應用

電子地圖可應用的範圍相當廣泛，在運輸方面可應用的領域如下：

- (1) 交通量指派
- (2) 客運業車隊管理 (排班與調度)
- (3) 商用車動態車輛派遣
- (4) 交通監控系統
- (5) 災害搶救及疏導指揮系統
- (6) 車輛導航系統
- (7) 行前旅次規劃
- (8) 旅行者資訊系統
- (9) 即時路況資訊展示
- (10) 先進交通管理系統

## (11) 先進大眾運輸系統

## (12) 網路地圖服務

目前車輛導航系統已越來越普遍，許多家車廠甚至將此配備視為基本配備，顯示其普及性。車輛導航系統除了導航功能外，同時提供定位功能，駕駛者於意外發生後或是車輛遭竊遺失時，行控中心可透過車上安裝的 GPS 發射器，確定目前車子的位置以協助車主處理。

由於網際網路的廣泛使用，使得網路地圖的服務越來越普遍，其熱潮已從國外延燒到國內，將地圖放置網站上，可供使用者透過網路進行地圖查詢；查詢基本方式包括路名、交叉路口點、重要地標點及地名等等，亦提供使用者自製地圖的功能，並可將成果列印或儲存，有些網站甚至還可提供最短路徑選取的功能，提供更為便利的查詢與服務。

除了發展出一套完整的電子地圖之外，電子地圖的資料更新及維護是很重要而無止境的工作，地圖更新的來源方式包括軍事調查、定期野外調查、地圖使用後的報告、最新航空照片或衛星影像的取得、公路新建或養護單位之道路異動資料、民間或政府單位出版品及網路資料的蒐集，另外也可發展一些系統處理這方面的工作，如資料錯誤回饋系統 (Data Error Feedback System)、輸入資料登錄系統 (Input Data Registration System)。當資料庫的使用者及其顧客反映資料有錯誤之處，這些資訊可用來改進資料庫的正確性，可利用資料錯誤回饋系統來解決這些問題。資料庫的規模若很龐大，則有必要使用輸入資料登錄系統，以確保對資料更改紀錄之掌握。目前有許多電子地圖廠商是透過使用者主動地提出地圖中錯誤的資訊，經由廠商確認後即可取得優惠，如現金回饋或是折價券等等，如此不但可減少廠商資料更新的費用，更可降低地圖與現實狀況之間的落差。

而本研究所謂「簡圖顯示」基本精神乃將電子地圖龐大的資料庫置於 Server 端，形成 Web-GIS，也就是上述的網路地圖，一方面可讓使用者經由網路查詢地圖，另一方面也使得地圖資料的更新與維護問題變得簡單。當使用者輸入起迄點後可接收導航簡圖，該簡圖即為系統所計算出的最短路徑，且在使用者端顯示的地圖中並不包含所有的街廓；也就是說，本研究因為考慮到 PDA 本身螢幕較小，需顯示較容易一目了然的簡圖，於是將電子地圖的主體置於 Server 端，使用者端獲得的簡圖雖然仍具有節點、節線等位相關係可稱做向量式電子地圖，但由於此簡圖只需顯示最短路徑以及本研究增加之最短路徑旁的設施資訊，因此在電子地圖的數量規模上比起 Server 端有相當大的差距，路徑簡化顯示的結果，讓使用者理解地圖資訊的時間縮短。

而 GPS 接收器最主要的功能在經由衛星訊號往返時間差以計算出使用者目前所在的經度、緯度。配合簡圖顯示，可以在簡圖上標示出使用者與道路節點、節線的相對位置。而前述一般 GPS 的誤差平均為十公尺，經過差分之後可達三公尺，這對於未來構建語音提示有相當程度的影響。因為當 GPS 誤差在三公尺以下時，甚至可以在電子地圖上顯示使用者在內車道或外車道上，進而以語音提示使用者是否需要變換車道。

但目前提供 DGPS 廣播服務的地區僅限於北美，因此本研究在建立語音提示前，仍假設 GPS 未經差分，提示語音中將不包含車道資訊。

### 三、車內設備輸出輸入介面對於行車安全的影響

駕駛者的注意力分散是交通安全上最主要的關鍵，當然使駕駛員分散注意力的東西包括有收音機、速食餐、上網和車上導航系統…等，最近三十幾年來（自 1966 年），美國 NHTSA（National Highway Transportation Safety Administration）致力於解決各種可能危害公路安全的威脅，所面臨的課題包括酒精的危害及安全帶的防護測試，事實上，駕駛者的注意力分散並不是一個新的問題，只是與以往有所不同的是，有許多創新的科技迅速的加入了汽車製造的市場，如無線電話、網路服務、導航系統、娛樂設備…等。

駕駛者的反應操縱了車輛的安全性，而注意力分散降低了駕駛者的反應動作，若是有更多更複雜的事物吸引駕駛者，會使駕駛者的反應性能更低，現今我們可以從許多項資訊管道得知這些危險性正在成長中。

近幾年來科技的進步，使得汽車上的視覺顯示結合了更複雜的資訊顯示及控制系統，在這些車上裝置能提供汽車使用者安全的駕車環境前，針對安全性和人為因素（human factors）上的考量進行詳細的調查是必要的，以決定系統的設計準則、政策指導方針、標準和法規。

在資訊顯示、控制指令與安全性之間做權衡取捨（trade-offs）的考量，是汽車製造業者的責任，一個新的顯示系統裝置至少要和"舊車"一樣安全，任何額外的視覺顯示或控制都不應該減少駕駛環境的安全性。以下將介紹人為處理能力和駕駛能力以及人機介面的考量。

#### 1. 人為處理能力

在過去的研究中，Card et al[6]提出了「人類處理裝置」的概念，認為該裝置係由以下三個子處理裝置所組成的（如圖 2-6）：

- (1) 知覺的處理裝置（perceptual processor）
- (2) 行動的處理裝置（motor processor）
- (3) 認識的處理裝置（cognitive processor）

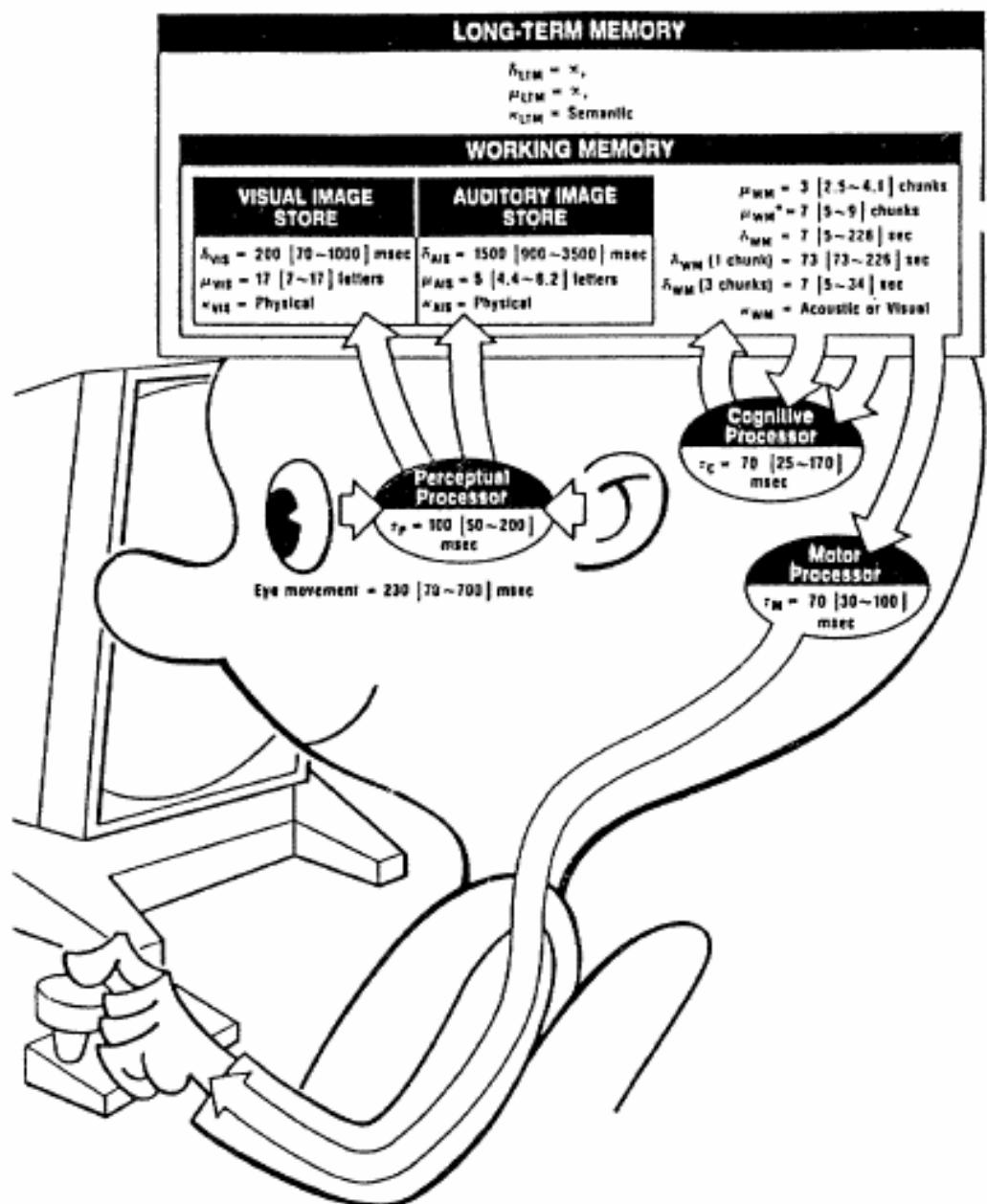


圖 2-6 人類資訊處理裝置 (The human information processor)

上述每一個子處理裝置都僅能使用有限的時間處理資訊，從圖 3-2 可得知每一個處理裝置的平均處理時間，如知覺的平均處理時間為 100 毫秒、認識的平均處理時間為 70 毫秒、行動的平均處理時間為 70 毫秒，其實這三個子處理裝置與交通工程學中用路者的反應時間 (reaction time) 類似，反應時間係指由視聽外界狀況之感覺、辨識刺激、以致決策並採取反應行動所經歷的時間，簡稱 PIJR Time，由以下四個時間所組成：

- (1) 感識時間 (perception time): 由許多目標中察覺某特殊狀況而加入注意。
- (2) 辨明時間 (identification time): 運用智慧，對該特殊狀況加以

充分瞭解與比較。

- (3) 判斷時間 (judgement time): 激發情緒, 以決定該特殊狀況之處理。
- (4) 行動反應時間 (reaction time): 執行判斷所做之決定。

圖 2-7 表示了資訊呈現的數量與資訊處理的數量間的關係, 由圖可知, 理想的狀況下兩者間應為線性成長 (亦即資訊呈現的數量越多, 資訊處理的時間越長), 然而實際上, 在資訊呈現數量少時, 資訊處理時間會較理想情形更快些, 當資訊呈現數量達到某個"極值"時, 資訊處理的時間就會迅速的衰退, 且錯誤和疏忽就會發生了, 因此如能找出這個"極值", 對車上裝置所造成的安全影響有極大的幫助。

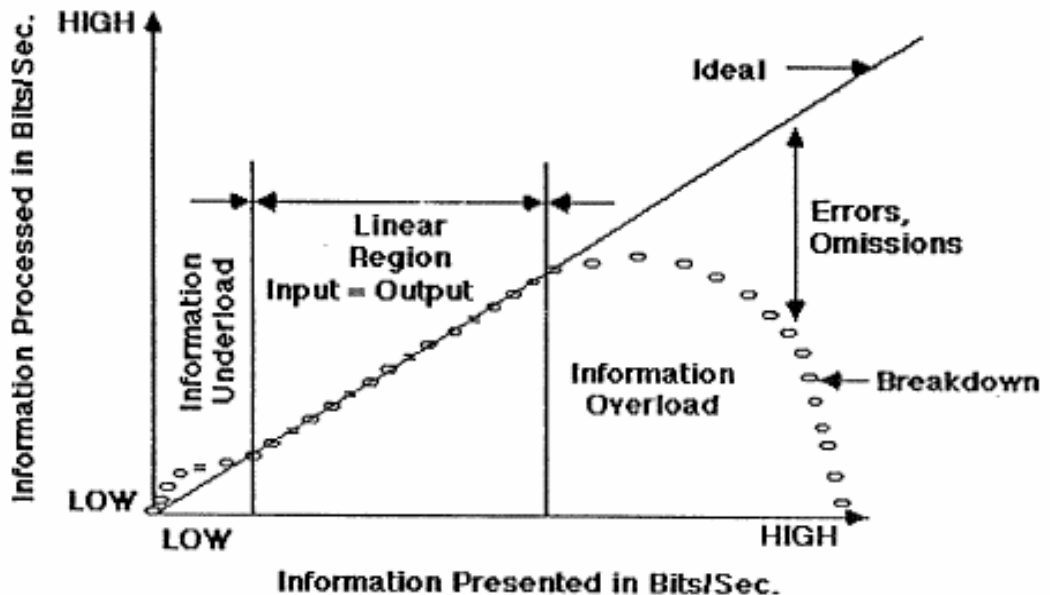


圖 2-7 資訊呈現數量與資訊處理數量關係圖

許多新的汽車顯示裝置 (如汽車音響、VCD、DVD、導航系統...等) 都可視為一種控制裝置, 和其他車輛控制裝置 (如車速表、排檔裝置、煞車...等) 一樣重要, Fitt[7]所提出的定律計算出駕駛者移動手到距離  $D$ 、目標物大小  $S$ , 所需花費的時間  $T_{pos}$ , 公式如下所示。舉例來說, 目標物大小為  $1.3\text{cm}$ 、移動距離為  $25\text{cm}$ , 則所需花費的時間約為  $0.44$  秒。

$$T_{pos} = I_m \log_2(D/S + 0.5); \text{ where } I_m = 100 [70-120] \text{ msec/bit}$$

年齡、酒精和毒品對人類的處理能力具有不利的影響。老化所引起的緩慢行為, 大部份是因為較微弱的訊號已無法接收至腦中, 而酒精和毒品也有類似的影響, Zwahlen[8]曾提出一項調查, 測量 12 位沒有酒精影響的實驗體和 12 位受酒精影響的實驗體 (BAC 值[Blood Alcohol Content]約  $0.1\%$ ) 的選擇反應, 結果顯示沒有酒精影響的人, 平均反應時間為  $0.83$  秒; 而受酒精影響的人, 平均反應時間為  $0.88$  秒, 且可預

期的，BAC 值越高，反應時間越緩慢。

## 2. 駕駛能力

McCormick 和 Sanders[9]指出，55%的交通事故是由單獨的人為錯誤所造成，30%可歸因於人為疏失和環境因素的組合，因此在汽車裝置的設計上就必須同時考量人類身體和心裡的能力。

人類的駕駛能力受限於視覺資訊的影響，視覺資訊是駕駛中最重要的一部份，駕駛者必須負責處理車子內外所發生的視覺資訊，如果視覺資訊中有大量的不確定性存在，則會產生很大的錯誤影響。已經有許多研究描述，駕駛人將注意力維持於道路上是非常重要的，若須轉移駕駛者的注意力時，應確保適當且安全的控制存在，而在汽車外部視覺範圍內所偵測到的物件，對駕駛者也有影響，過去研究指出：晚上開車的安全限度較白天開車來的低，因為晚上開車時預視的距離減少了、周邊視野有限，另外當一個目標原本固定在視覺範圍內時，當目標移動到後方時，對視覺也會有些微的影響。

至於在汽車內部的視覺資訊對維持車輛安全也很重要，Gramberg-Danielsen[10]曾提出，將視線從車外 20 公尺處移回車內里程表位置所需的時間，再加上眼睛適應新的注視環境所需的時間，在 40 歲以下約須花費 0.4-0.8 秒；在 60 歲以上則須花費 2.0-2.7 秒。由此可知，駕駛者大部份時間都應注視於前方道路上，因此在使用車內其他裝置時，對於車內裝置所發送出來的訊息或畫面，引起駕駛人將視線從車外移至車內所需花費的時間和危險性，也必須加以考量。

茲將人類於駕駛時的主要性能和限制彙整如下：

- (1) 有限的資訊獲得
- (2) 有限的周邊視覺範圍
- (3) 受限於眼球的環境適應速度
- (4) 受年齡、酒精、毒品的影響
- (5) 受限於控制時的移動速度
- (6) 受強光（或刺眼的光）、黑暗適應能力的影響
- (7) 受情緒、疲倦…等心理上的影響

## 3. 人機介面

人機之間的交互作用是車內裝置設計時重要的考量因素之一，每一個設計在車內使用的裝置，其系統的使用必須被設計為可讓大多數人毫無困難的使用，人機間的介面主要可分類成兩個子系統：輸出介面與輸入介面。

輸出介面

在討論人機間可能的資訊輸出介面時，有兩種表示方式：一為視覺呈現方式；另一為聽覺呈現方式，也有部份研究人士提出第三種為觸覺呈現方式，在此係將觸覺呈現方式納入資訊輸入介面的部份來探討。

一般提到資訊輸出予駕駛者時，通常會先想到用影像、畫面或文字的方式呈現，但是對於某些資訊而言，並不一定是最好的呈現方式，以下歸納適用以視覺呈現輸出資訊的條件環境：

- (1) 訊息是複雜的
- (2) 訊息是長的
- (3) 訊息不具有即時性
- (4) 使用者的聽覺系統超過負荷
- (5) 接收訊息的環境太過吵雜

另一種資訊輸出方式是以聽覺（即聲音）呈現，例如當鑰匙遺留在車上時，車子會發出嗶嗶的叫聲，以下亦歸納適用以聽覺呈現輸出資訊的條件環境：

- (1) 訊息是短的
- (2) 訊息是簡單的
- (3) 訊息具有即時性
- (4) 使用者的視覺系統超過負荷
- (5) 接收訊息的環境太亮或無法適應黑暗環境

#### 輸入介面

在汽車裝置上有許多系統是必需要由駕駛者來操控的，不論是排檔的動作、打方向燈，乃至於許多新的車內裝置（如汽車音響、VCD/DVD、導航系統…等），其操控的動作都會影響行車上的安全，過去研究顯示太過繁雜的操作對行車也造成相對的危險，因此後續的車內裝置發展中，多以單鍵或觸控面板來執行系統功能選擇，可簡化操控時所需花費的時間及注意力，本研究屬於雙向語音導引控制功能，目的是降低駕駛者操作的複雜度，以語音建立單鍵、觸控面板之外的操作方式，進而提高行車上的安全。

### 第三章 路徑導引之語音功能需求

本章主要內容為敘述為何駕駛者需要語音功能，以及語音功能出現的時機，並擬定語音應用於行動路徑導航之應有作法與流程。其中較為特別的功能是將駕駛者轉向角度以十二時方位表示，由於程式語言以正切反函數推定道路節線間夾角之弧度，因此在該主題中亦解釋本研究如何以程式之條件式判斷而將轉向弧度歸納為十二時方位。

#### 3.1 路徑導引之功能與架構

本研究路徑導引之功能主要是由使用者需求的角度產生。簡單的說，使用者的需求不外乎是清楚的地圖，且在車輛該轉彎時有清楚的提示。

也就是說，使用者輸入起迄點後，得到一最佳路徑，並且在駕駛的同時可藉由說出操作地圖的指令，調整地圖顯示，清楚地瞭解目前車輛所在的路名或者是車輛在地圖上的位置，在車輛需要轉向時能告知使用者，附帶的需求之一是告知使用者附近設施的相關資訊，路徑導引系統的功能架構圖如圖 3-1 所示。基本上，這就像是一位對該路段熟悉的朋友坐在副座指引駕駛。

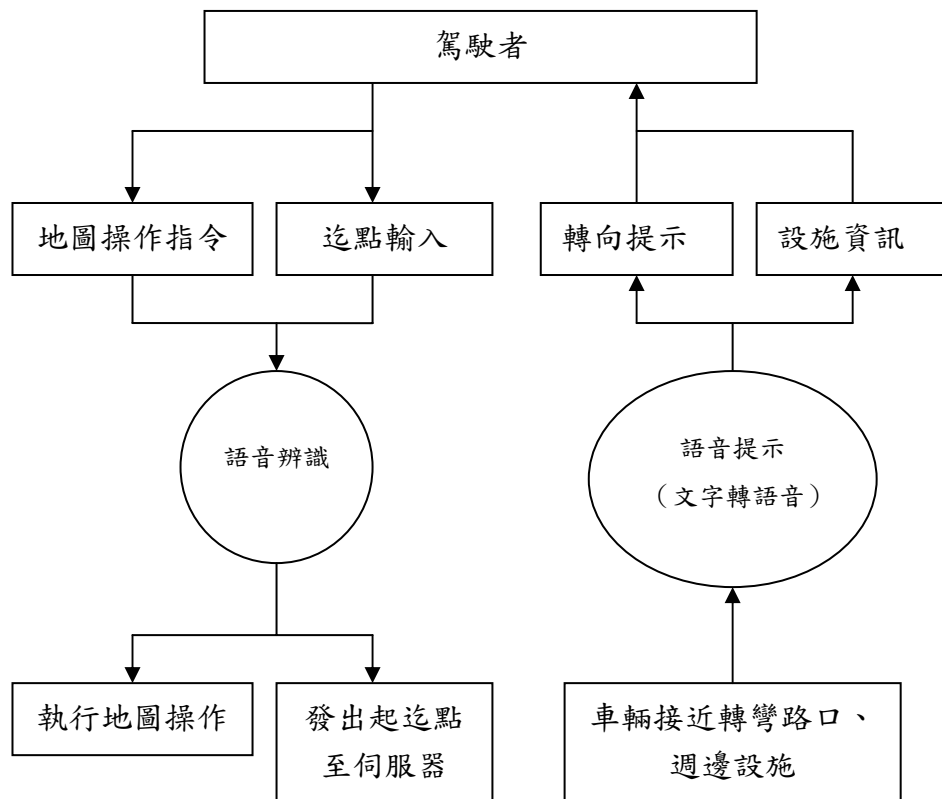


圖 3-1 路徑導引系統功能架構圖

在全球定位系統與電子地圖尚未普及時，駕駛人若要駕車前往不熟悉的地點，多半要在車內準備一本地圖，事前規劃好車輛路線。但在真正上路時，常常事與願違，因為不是每一條道路的標示都非常清楚，或者是道路名稱等標示被廣告招牌遮蔽，常使得駕駛人不知自己身在何處，這時若遇上該轉彎的路口，自然容易錯過，一旦錯過，駕駛人還是得再將地圖拿出來重新規劃一番。

直到現今，車內導航系統免除了駕駛人翻閱地圖的不便，但似乎是將翻閱地圖的右手轉變成為操作導航系統的右手。也有許多車內導航產品具有轉向提示的功能，但是提示的內容往往不盡詳細，在複雜的路口甚至有誤導駕駛人的可能。以語音作為介面，改善以上兩點不盡理想之處便是本研究導航系統的兩大功能，第三大功能是建立設施資訊的語音提示雛形，使駕駛人在開車的同時能獲得周邊設施的資訊。以下對本研究中的三大功能作說明：

### 3.1.1 語音辨識

駕駛人對於操作車用導航系統的需求可分為駕駛中的需求，與非駕駛中的需求，駕駛中的需求是調整地圖顯示與開啟或關閉語音提示，非駕駛中的需求是設定起迄點，關於駕駛中與非駕駛中的需求分類，在下一節會有較詳細的說明。在駕駛中的情況，本研究從無至有，撰寫語音辨識程式，經由 PDA 對駕駛人發出的命令錄音，以該段錄音進行語音辨識。駕駛者將以簡單，扼要的口語指令直接控制 PDA 螢幕上簡圖顯示的倍率，中心位置。以達到最佳的顯示效果，亦降低駕駛者理解地圖資訊的時間。

非駕駛中的需求主要是設定起迄點。在第一章研究範圍中，說明了本研究假設起點為由 GPS 所獲得的經、緯度。且考量 PDA 運算能力的限制，應用大字彙辨識技術辨識 GIS 伺服器中資料庫的路名、景點名稱將對 PDA 造成過多的負擔，本研究在發展輸入迄點功能時，以特定詞辨識技術達到示範性的目標。所謂示範性是指本研究將設定十個地點名稱作為語音模型，當使用者說出其中之一的地點時，特定詞辨識技術可以從眾語音模型中辨識出該地點，完成迄點輸入的功能。語音辨識功能的操作流程如圖 3-2 所示。

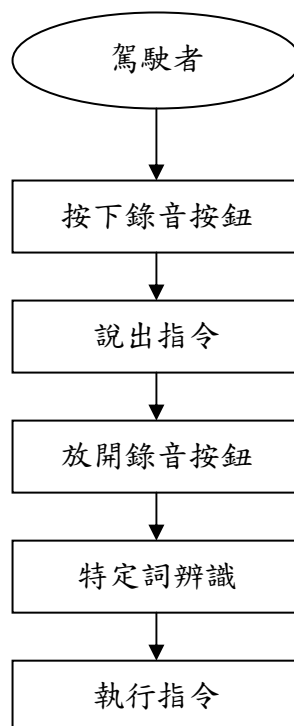


圖 3-2 語音辨識功能的操作流程

### 3.1.2 轉向提示

在駕駛中的情況，駕駛者無法持續注視 PDA 螢幕查看前方路口是否該轉彎；若駕駛者只是瞄一下 PDA，在第二章裡亦指出駕駛者眼睛適應新的注視環境需要 0.4 秒至 2.7 秒的適應時間，隨著年齡而不同。因此，雖然第一項功能中的語音操作地圖使駕駛者得以順利調整地圖顯示，但以語音轉向提示的方式告知駕駛者前方路口轉彎的資訊，能減少駕駛者注視 PDA 螢幕的需求。所以本研究除了滿足使用者調整地圖的需求外，另採語音轉向提示功能併行的方式，期能增進行車安全性。

轉向提示功能主要是藉由車輛目前座標、車輛行進方向，以及車輛前方與車輛行進方向具有夾角的相鄰節線等三項資訊配合 PDA 記憶體中簡圖各道路的名稱，可以得到車輛前方轉彎路口的距離，轉彎的角度，轉彎後的道路名稱，這些文字將經由文字轉語音的技術成為轉向提示的內容。其中，為了在複雜的路口前，使駕駛者獲得清楚的轉彎提示，本研究將轉彎角度的文字轉換為十二點鐘方位，取代以往只告知左轉或右轉的提示。關於轉向提示的詳細內容，將在 3.2 節說明。圖 3-3 說明轉向提示的輸入輸出資訊。

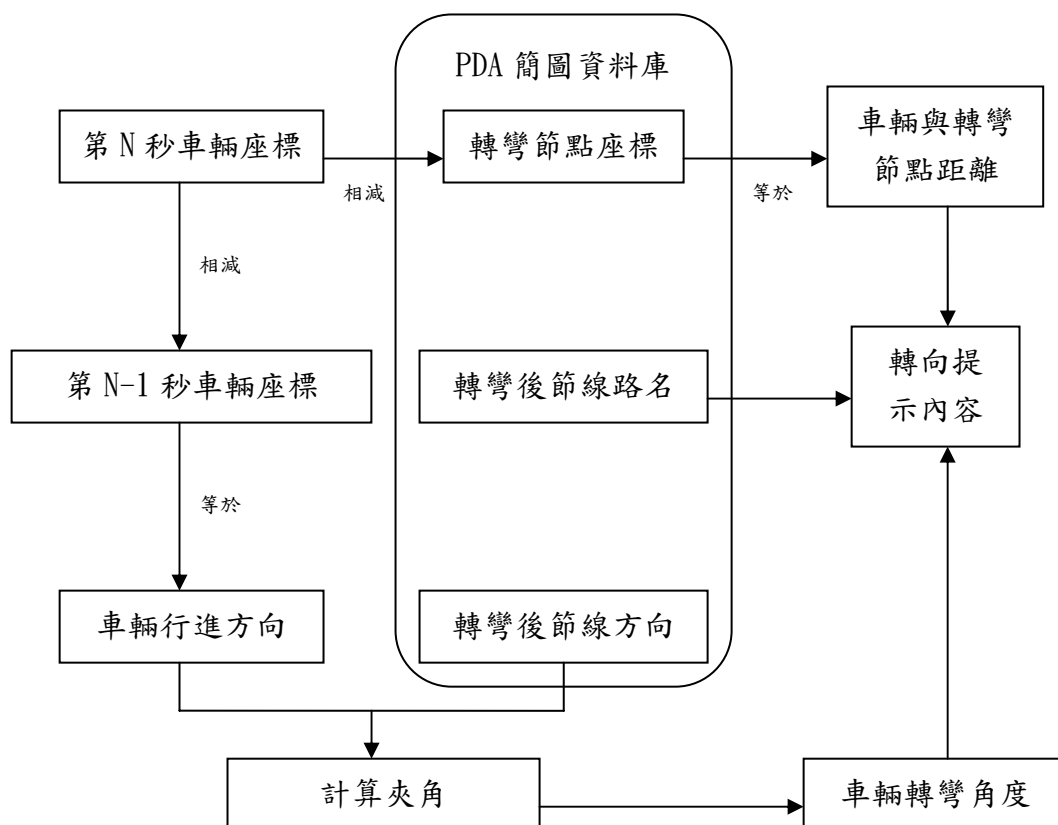


圖 3-3 轉向提示的輸入輸出資訊

圖 3-3 中計算夾角一項在程式運算中佔了較高的運算量，因為程式由節線兩端節點間的座標相減，計算其正切反函數 ( $\tan^{-1} x$ ) 而得該道路於座標軸之角度，與下一節線之角度相減後得到轉彎角度。但正切函數的特性，如  $\tan\left(\frac{\pi}{4}\right)$  與  $\tan\left(\frac{\pi}{4} + \pi\right)$  的結果同樣是 1，使得  $\tan^{-1}(1)$  的角度

出現兩種可能，程式所判定的角度結果總在  $-\frac{\pi}{2}$  至  $\frac{\pi}{2}$  之間。因此，本研究在進行節線角度的計算時，加入數個條件判斷式，以正確地計算節線在 0 至  $2\pi$  間的角度。地圖上可能出現的節線角度可分為四個象限，如圖 3-4 之節線 a, b, c, d。

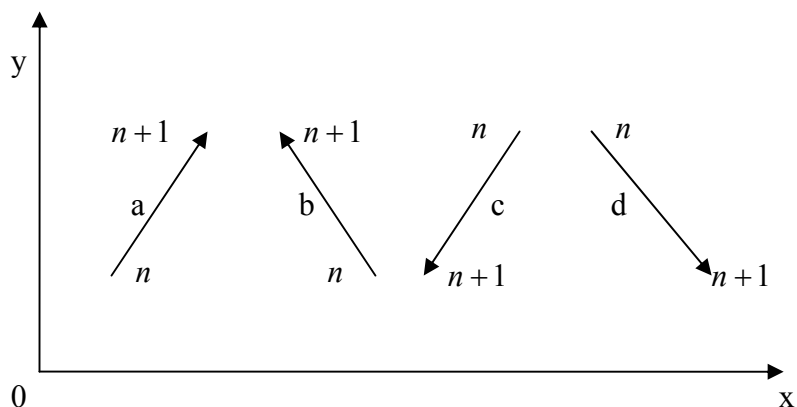


圖 3-4 分別代表四個象限的節線 a, b, c, d

而在角度未知的情況下，a, b, c, d 之間的差異便是  $X_{n+1} - X_n$  與  $Y_{n+1} - Y_n$  在正負號上的差異，由此差異便可以在程式中加入條件判斷式：

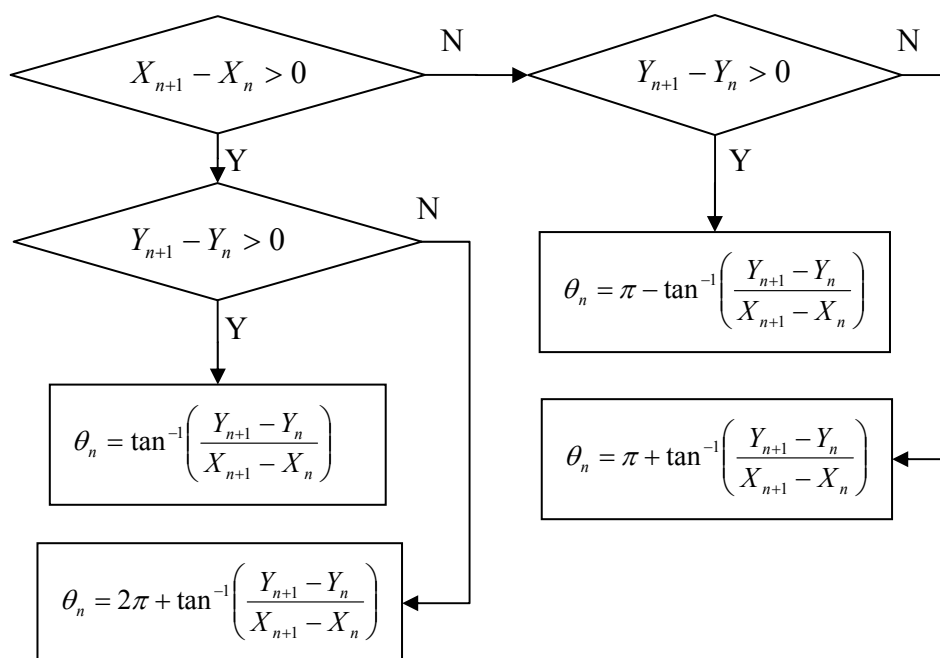


圖 3-5 計算節線角度之條件判斷式

$\theta_{n+1} - \theta_n$  兩節線角度相減後，所得結果即為由駕駛人觀點之兩節線夾角，正數的角度為順時鐘方向，負數的角度為反時鐘方向。如圖 3-6 所示，若計算結果為 30 度，將落入一點鐘方向的範圍中，若為 -50 度，則為十點鐘方向。

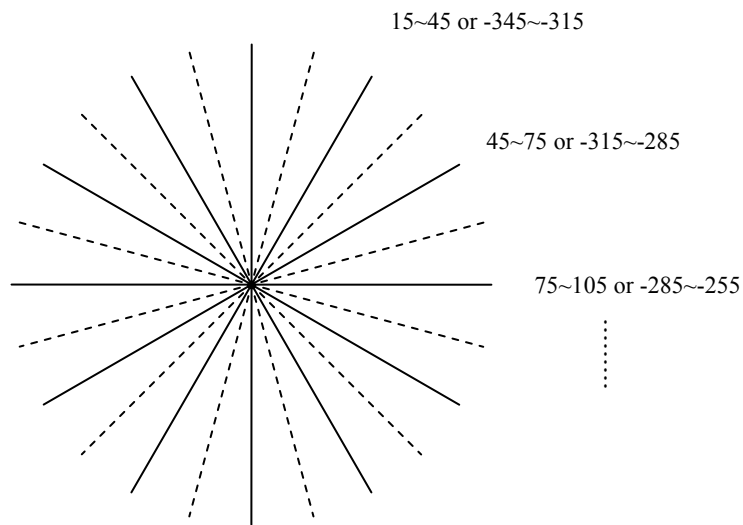


圖 3-6 節線夾角之 12 時方位

### 3.1.3 設施資訊

設施資訊的提供是本研究在張晉維與林詠傑所建立之簡圖路徑導引系統之後，一項延伸的功能。主要是考量駕駛者有獲得車輛附近設施資訊的需求，例如駕駛車輛的衍生需求：停車，雖然目前臺北市內許多停車場入口都能具體顯示該停車場目前剩餘車位數等資訊，但如果遇上停車場無剩餘車位的情況，駕駛者仍然必須要駕車到停車場入口前閱讀無剩餘車位的資訊，再駕車找尋下一停車場。

如果以語音提示的方式告知駕駛人停車設施資訊，當駕駛人距離停車場尚有一段距離時，便發出設施資訊供駕駛者瞭解該停車場目前使用狀況。當然，本研究假設該停車場資訊可由伺服器端獲得，並且以文字的方式儲存於 PDA 中，由文字轉語音技術發出設施資訊的提示。此功能中所謂設施不僅指停車場，設施還可以包括飯店、加油站、醫院、銀行、郵局、警察局、消防隊等，提示的資訊內容並不一定是價格，可以只提示設施名稱，達到告知駕駛人前方有郵局、銀行、警察局等設施的目的。本研究擬選取停車場、飯店、加油站共三類設施建構設施提示功能，關於此三類設施資訊的詳細內容，於 3.2 節說明。圖 3-7 說明提供停車場設施資訊流程：

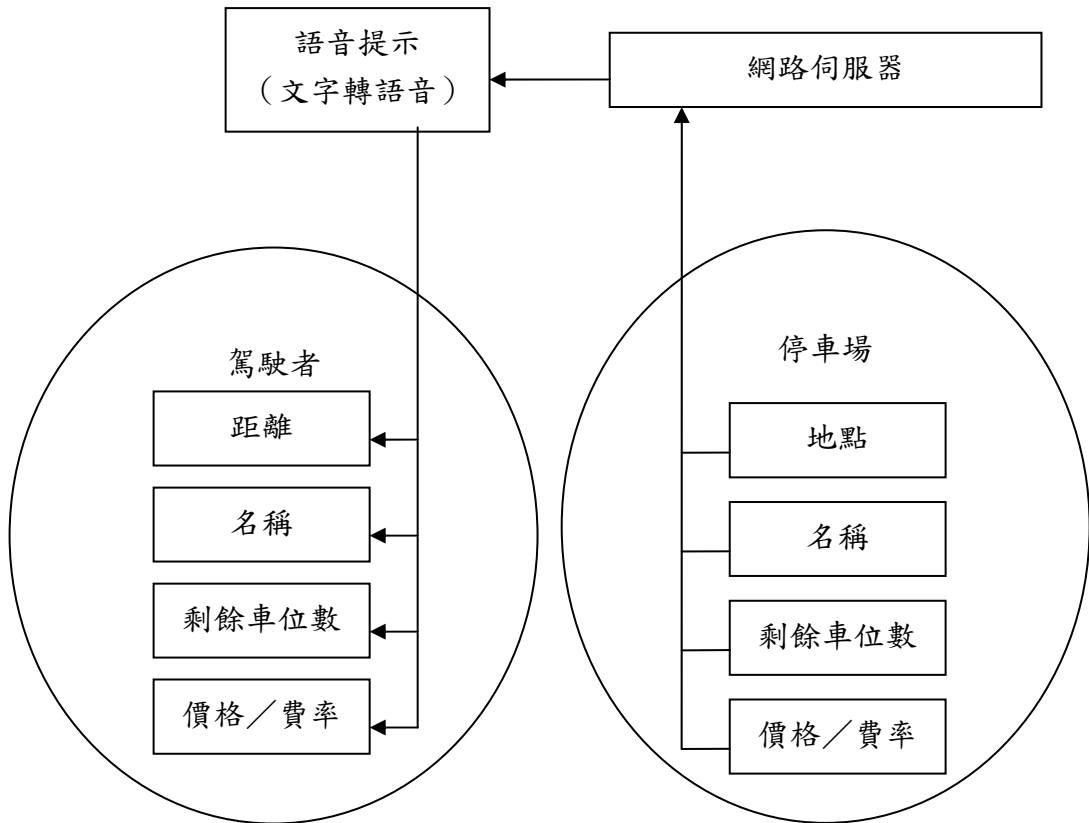


圖 3-7 語音提示之設施資訊流程

### 3.2 駕駛人下達之關鍵詞指令

在語音辨識的領域中，依可辨識的數量來分，大致分為特定詞辨識（或稱少量字彙辨識）的技術，以及大字彙辨識技術兩種類別。特定詞辨識技術意指將各特定詞之語音數位化並儲存成各語音模型，有了語音模型之後，就可以將待辨識的語音訊號與各語音模型比較，從各語音模型中逐一找出與待辨識語音最近似的語音模型，進而執行其所對應的相關動作。例如：將「0」、「1」、「2」…「9」等數字的語音訊號分析之後成為 10 種語音模型，當使用者對麥克風說出「3」時，「3」的語音訊號便與「0」、「1」、「2」…「9」等語音模型比較，找出其中最相似的模型，這樣子的技術就可以稱之為特定詞辨識。

而中文字詞可能產生的組合數量非常龐大，若以「字」或「詞」作為語音模型，在逐一搜尋近似的語音模型時，將具有高度複雜性，並且耗費大量的處理時間，如此使得語音辨識的實用性大大地降低。因此，當語音辨識系統欲辨識大量字詞（也就是不限定於幾種特定詞）時，便採用大字彙的辨識技術。在大字彙辨識技術中，語音模型的形式不再是特定的字詞，而是單獨建立ㄅ、ㄆ、ㄇ、ㄋ等聲母與韻母模型，也可能是建立ㄅㄩ、ㄅㄨ等注音符號與子母音的模型，在中文裡聲母、韻母列表 3-1。

表 3-1 中文聲母與韻母內容

子音（聲母）	母音（韻母）	
	單母音	雙母音
ㄅ、ㄆ、ㄇ、ㄌ、ㄋ、 ㄍ、ㄑ、ㄒ、ㄔ、ㄕ、 ㄆ、ㄇ、ㄏ、ㄉ、ㄊ、 ㄒ、ㄙ、ㄗ、ㄘ、ㄙ、 ㄌ	一、ㄨ、ㄛ、ㄜ、ㄝ、 ㄖ、ㄗ、ㄘ、ㄙ、ㄞ、 ㄟ、ㄠ、ㄡ、ㄢ、 ㄣ	一ㄅ、一ㄆ、一ㄇ、一ㄌ、 一ㄋ、一ㄍ、一ㄑ、一ㄒ、 一ㄔ、一ㄕ、一ㄖ、一ㄗ、 一ㄘ、一ㄙ、一ㄞ、一ㄟ、 一ㄠ、一ㄡ、一ㄢ、一ㄣ、 一ㄤ、一ㄥ、一ㄨㄛ、 一ㄨㄜ、一ㄨㄝ、一ㄨㄞ、 一ㄨㄟ、一ㄨㄠ、一ㄨㄡ、 一ㄨㄢ、一ㄨㄣ、一ㄨㄤ、 一ㄨㄥ

大字彙辨識技術可從使用者說出的語音訊號中，辨識出聲母與韻母的組成。但此辨識技術需要較複雜的辨識演算法，例如隱藏式馬可夫模型、或者是類神經網路等演算法，且此技術需要充足的訓練語料來訓練語音模型，以獲得較精準的語音模型。所謂的訓練語料，一般國內所採用的訓練語料是由「中華民國計算語言學會」(ROC Computational Linguistic Society)所提供的「台灣之國語語音資料庫」(Mandarin speech database Across Taiwan, 簡稱 MAT)，它是由清華大學電機系、台灣大學電機系、交通大學電信系、成功大學資訊所、中央大學電機系、淡江大學電機系、中興大學、中研院史語所、交通部電信研究所和工研院電通所的專家學者共同合作，分別在各單位建立同一規格的錄音系統軟硬體，制訂標準的檔案格式，透過電話網路經由個人電腦錄製，其取樣頻率為 8kHz，取樣位元數為 16 位元來進行語音收集及建檔工作，以供國內發展語音技術及開發語音系統之用。MAT 語料庫統計表如表 3-2 所示。

表 3-2 MAT 語料庫統計表

類別	性別	人數（人）		句數（句）		音節數（個）		時間（小時）	
MAT2000	男	1002	2229	22591	50472	115940	260133	13.62	31
	女	1227		27881		144193		17.38	
MAT2500	男	1265	2570	25096	49865	163364	324712	17.86	36.2
	女	1305		24769		161348		18.33	

另外，大字彙辨識技術除了需要大量語料訓練外，尚需對語音訊號作聲學解碼、語言學解碼。這是因為同一個中文字常因為前後文的不同，而有不一樣的發音，因此需要對語音訊號解碼，從前後文的關係獲得更正確的辨識

結果，大字彙辨識技術的種種過程對於 PDA 裝置恐將造成較大的負擔，且本研究主要是要以語音取代使用者以 PDA 觸控筆操作導航軟體的動作，以林詠傑所建立的簡圖導航系統軟體為本研究簡圖導航系統的基本雛形（如圖 3-8 所示），可以清楚瞭解到除了導航模組中的轉向提示功能之外，幾乎所有的觸控筆操作動作都能以語音辨識的方式取代：

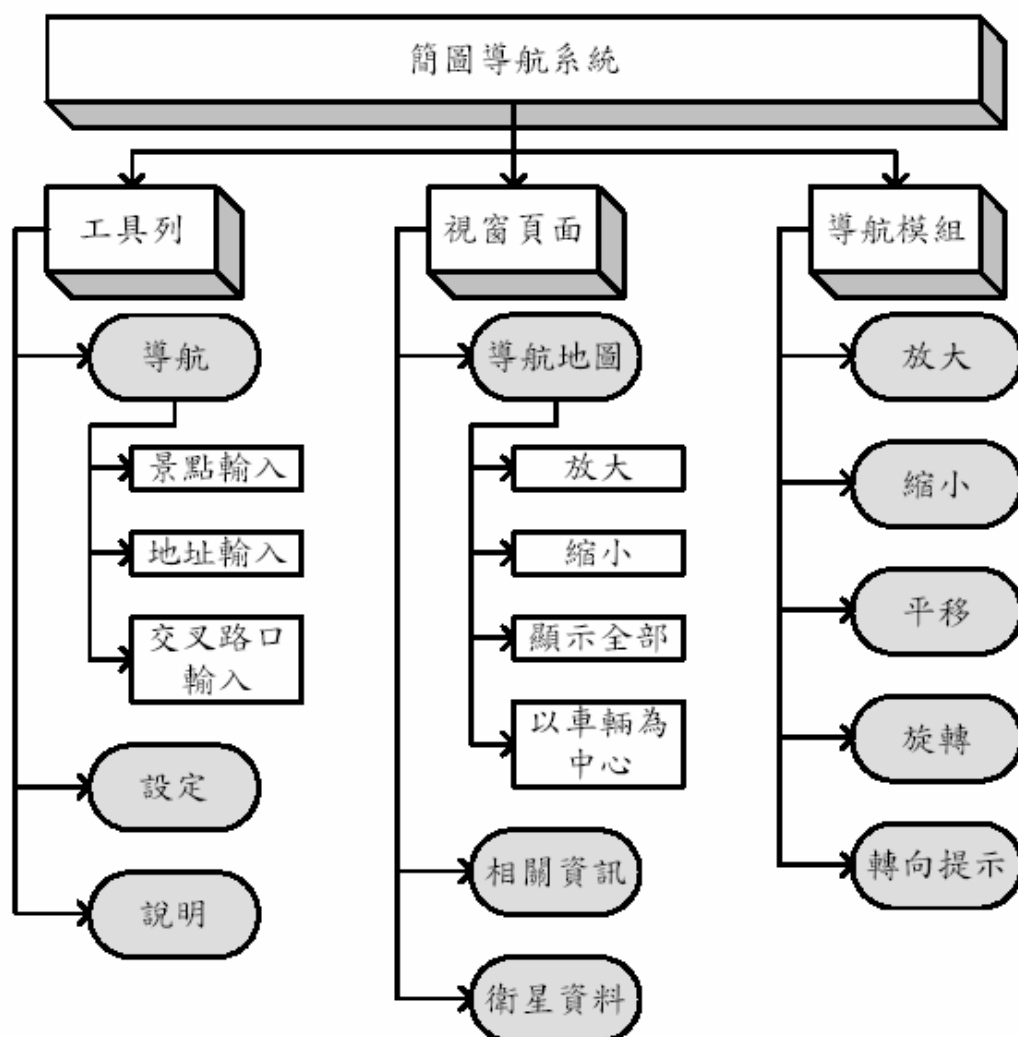


圖 3-8 林詠傑建立之簡圖導航軟體系統架構圖

車用導航系統未來的發展方向之一是提供更友善的使用者介面，避免繁瑣的操作，導航軟體的操作應力求簡單而有效。因此，以特定字詞為「指令」，使得導航系統產生應對該指令的「動作」，這種稱為特定詞辨識的技術是非常適合車內環境的語音辨識技術。

參考林詠傑所建立的簡圖導航軟體系統架構圖（圖 3-8）後，本研究依此架構規劃出駕駛者下達之關鍵詞指令，其指令樹狀圖如圖 3-9 所示。

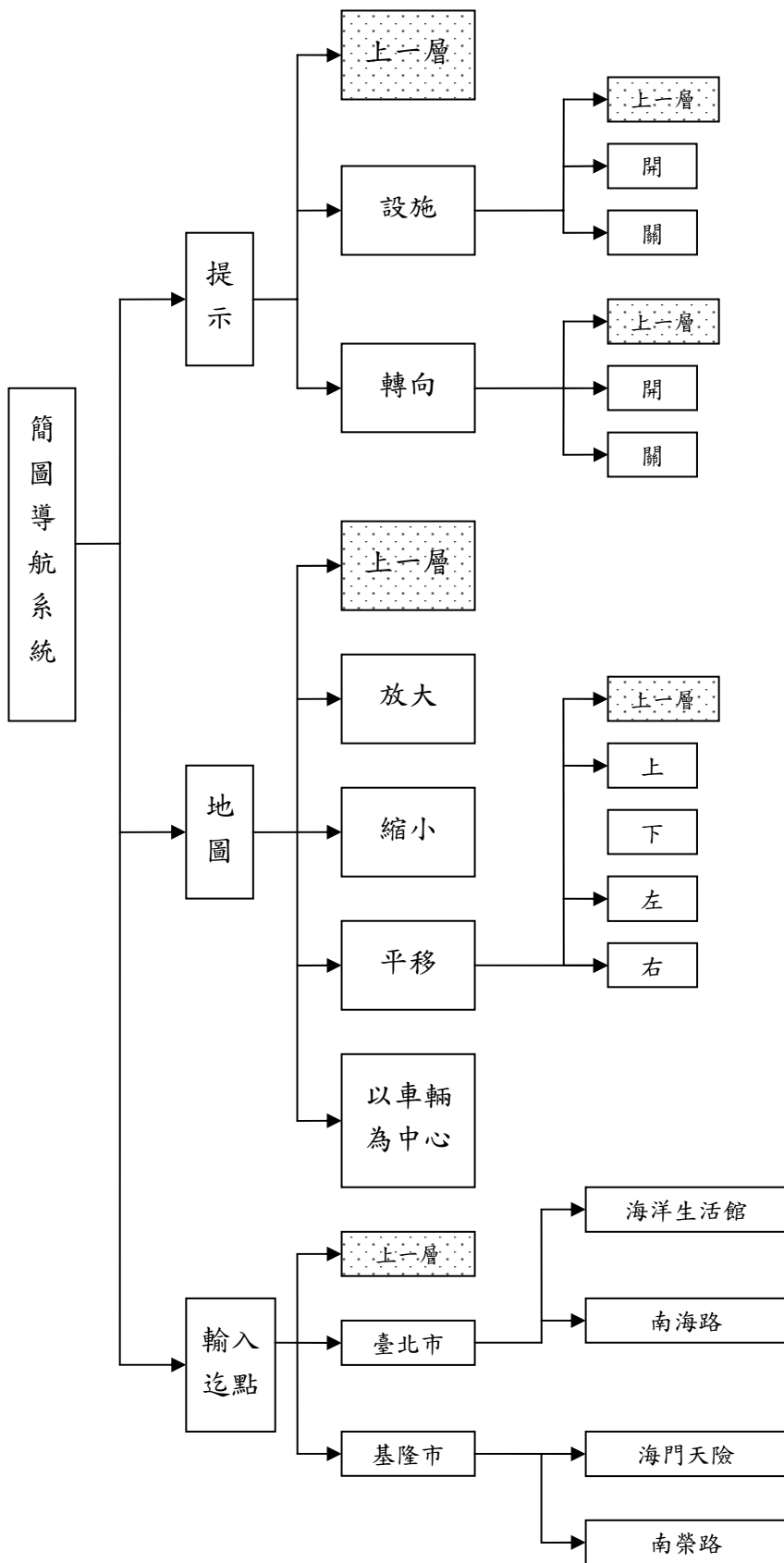


圖 3-9 駕駛者下達之關鍵詞指令樹狀圖

在建構本研究中駕駛者下達之關鍵詞指令時，是著重於車輛行進中的情況，駕駛者對於導航系統較為必要的操作，在這部分省略了對於導航較不重要的功能。圖 3-8 中，說明與衛星資料等功能，在駕駛時，駕駛者對其需求較低，因為說明是關於此軟體的使用說明，在使用導航軟體前應已有所認識，在駕駛的同時閱讀說明也有安全之虞；衛星資料的功能對於駕駛者也不大具有需求上的迫切性。相對的，調整導航簡圖的顯示使得駕駛者能以適當的顯示比例理解地圖資訊，對駕駛者而言，具有較高的使用需求。

圖 3-9 中，簡圖導航軟體的使用分為地圖、提示與輸入迄點三大部分，這是考慮到三點駕駛者需求：

1. 在駕駛的同時，駕駛者常依其需要而調整地圖顯示比例（放大、縮小），位置（平移、以車輛為中心）。
2. 在駕駛的同時，車輛於轉彎處前方會發出轉向提示，若有設施位於道路旁，則會發出設施資訊，若駕駛者已對該路段熟悉，或者是希望安靜地駕駛，便有關閉提示的需求。
3. 在駕駛的同時，駕駛者有可能臨時改變其原本的行程，在這種情況下就會產生重新指定新迄點的需求。

### 3.3 語音提示內容

在簡圖路徑導引系統中，語音提示功能分為轉向提示與設施資訊兩類，本節延續 3.1 節功能說明，定義其提示內容。

#### 3.3.1 轉向提示

由圖 3-3，轉向提示是由車輛與轉彎節點距離、轉彎後節線路名，與車輛轉彎角度三者所構成。轉向提示功能啟動的時機是當車輛距離轉彎節點一百公尺時，且提示完第一次後，將重新計算與轉彎節點的距離，在第一次結束後三秒，開始第二次提示。

而上述的啟動規則若無更詳細的機制設計，將在駕駛者尚未轉第一個彎時，第二個彎也進入車輛前方一百公尺處的情況下產生問題，如示意圖 3-10。

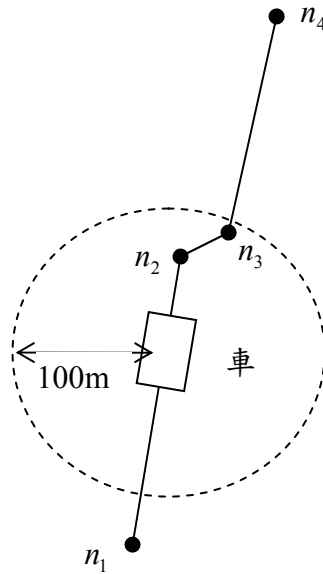


圖 3-10：轉向提示可能產生問題的情況

在這種情況下， $n_2$ 、 $n_3$ 兩轉彎節點在導引系統內會發生衝突。為了避免這種情況發生，在撰寫轉向提示程式時，應加入一判斷機制，若系統已偵測到 $n_2$ ，則接下來偵測到的 $n_3$ 儲存於記憶體中，待GPS指出車輛已駛離 $n_2$ ，再進行 $n_3$ 的轉向提示。

考慮上述連續兩轉彎節點的問題後，本研究可針對單一轉彎節點進行轉向提示，其內容中需要使用文字轉換語音技術的部分將包括隨著節線不同而變動的轉彎後節線路名，隨著車輛移動而變動的車輛與轉彎節點距離，以及轉彎後節線方向與車輛行進方向夾角。內容不變的語音則包括「前方」、「公尺」、「點鐘方向轉入」，其句型結構如圖 3-11。

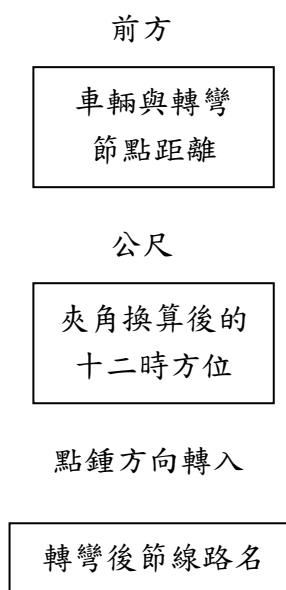


圖 3-11 轉向提示句型結構

### 3.3.2 設施資訊

在 3.1 節中曾指出本研究所謂設施，泛指停車場、飯店、加油站、醫院、銀行、郵局、警察局、消防隊等。每種設施性質不盡然相同，因此，設施資訊的提示不若轉向提示具有固定的句型結構。由停車場設施資訊流程（圖 3-7），停車場可經由網際網路將其相關資訊提供予網路伺服器；同樣的，飯店與加油站亦經由網際網路提供資訊。在這種情況下，設施單位提供的資訊，再加上設施與車輛的距離，就是駕駛者所聽取的資訊，以停車場、飯店，與加油站三種設施為例，本研究制定的設施資訊內容如表 3-3 所示。

表 3-3 設施資訊內容

停車場	距離
	剩餘車位數
	名稱
	價格／費率
飯店	距離
	空房數
	名稱
	價格／費率
加油站	距離
	名稱

## 第四章 建立語音辨識與提示功能

本章由介紹開發工具開始，進入語音辨識與語音提示程式實作的部分。以 PDA 實機進行錄音時遇到直流偏移的問題，在本章亦提出討論與解決辦法。語音辨識需要多個音訊處理步驟，在 4.2 節將詳述每一步驟之原理，並擬程式流程。

### 4.1 開發工具

本研究開發語音辨識，語音提示程式所使用的工具是 Microsoft eMbedded Visual Tools 3.0 版，該工具所開發的程式可應用於 Handheld PC Pro 3.0 版、Palm-size PC 1.2 版、Pocket PC 2.0 版、Pocket PC 2002 版等作業系統，惟開發前必須選定作業系統。而該工具中包括了 eMbedded Visual Basic 與 eMbedded Visual C++兩種程式語言供程式開發者選擇，因為 eMbedded Visual Basic 語言在介面與語法上與 Visual Basic 語言相當類似，對於程式開發具有相當的便利性。因此，本研究選擇 eMbedded Visual Basic 作為開發工具所使用的語言。其開發環境如圖 4-1 所示。

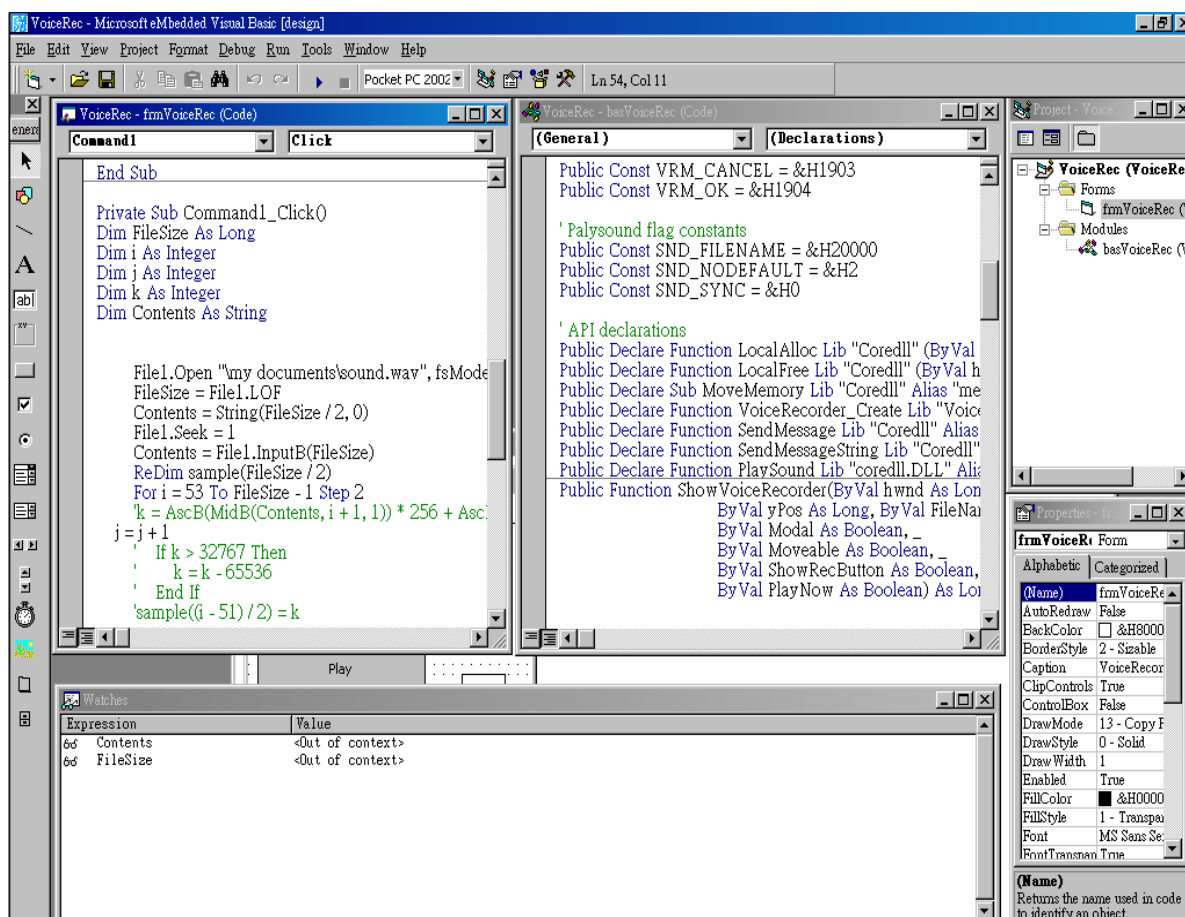


圖 4-1 eMbedded Visual Basic 開發環境

eMbedded Visual Tools 除了能透過 USB 傳輸介面將程式下載至實體 PDA 執行外，亦提供了 PDA 模擬器，方便程式開發者進行程式除錯。但模擬器尚未完全模擬實體 PDA，尤其是本研究語音辨識中的錄音功能，模擬器無法支援。因此，本研究在撰寫聲音輸出輸入之錄音程式時，是以實機來做測試的。圖 4-2 是 Pocket PC 2002 模擬器執行畫面。



圖 4-2 Pocket PC 2002 模擬器執行畫面

## 4.2 辨識原理

簡而言之，語音辨識的原理是將語音數位化後，進行訊號處理與分析的工作，分析目的是取得語音訊號的特徵值，並將特徵值儲存成為語音模型。當使用者再次說出同樣的字或詞時，可由語音的特徵值找出最為相似的語音模型。因此，語音辨識包含的技術層面相當廣泛，在接下來的章節將論述本研究撰寫語音辨識程式所根據的理論基礎。

### 4.2.1 類比／數位轉換

語音在空氣中是以波的形式來傳遞，這種形式的訊號稱為類比訊號 (Analog signal)。電腦無法直接處理類比訊號，因為在電腦中，所有的資料都是以 0 或 1 表示。訊號以數值大小表示時稱為數位訊號 (Digital Signal)。因此，若要以電腦來處理類比訊號，必須先將類比訊號轉為數位

訊號。

原來連續性的訊號，經過數位化的處理後，變成一種不連續的訊號，這個訊號只在某些固定的時間刻度上有值，這些刻度稱為取樣點，取樣點上的振幅大小稱之為取樣值，兩個取樣點之間的時間間隔稱之為取樣週期，取樣週期的倒數稱為取樣頻率。取樣頻率的另一種說法是，在一秒中內要對原始訊號做幾次取樣，取樣頻率的大小可以由使用者自定，但是一般來說要符合取樣定理（Sampling Theorem）。就是取樣頻率需大於兩倍取樣頻率，否則會造成取樣失真（Aliasing），一旦取樣失真發生，所得到的聲音取樣值便不能代表原來的訊號。因此在對語音訊號作取樣時，由於語音的最高頻率不會超過 4000Hz，所以取樣頻率定在 8000Hz 以上，就保證取樣失真不會發生。

在每個取樣點上，訊號的振幅大小也被數位化，在麥克風上的訊號是一種細微交流電壓的型態，例如：原始訊號的振幅在-50mv ~50mv 之間變動，數位化的過程就是把電壓轉換成數字大小，如以-128 代表-50mv，127 代表 50mv。

而一般語音辨識所使用的 Waveform 聲音檔案分為 8-bit PCM 與 16-bit PCM。在 8-bit PCM 聲音檔案中，每個取樣點是由單一無正負號的 Byte 所組成，而 16-bit PCM 聲音檔案中，每個取樣點是由一個有正負號的 16-bit 數值所組成，其中低位元組的資料放置於前面，高位元組的資料放置於後面，即所謂的 Little-Endian 記錄方式。而單音（Mono）以及立體聲（Stereo）的資料編排方式中，立體聲的左聲道（Left Channel）放置於右聲道（Right Channel）之前。表 4-1 說明 8-bit PCM 與 16-bit PCM 取樣值範圍。

表 4-1 取樣資料範圍表

資料格式	最大值	最小值	中間值
8-bit PCM	255 (0xFF)	0	128 (0x80)
16-bit PCM	32767 (0x7FFF)	-32768 (0x8000)	0

本研究所使用的錄音規格為 8000Hz，16-bit PCM，而在建立 PDA 錄音程式時，使用實機測試該程式的錄音效果，實機的規格為 iPAQ 3950，400 MHz Intel PX A250 Applications 處理器，64MB RAM。在普遍觀察錄音結果後，發現聲音訊號的起點至 0.2 秒左右的取樣值為零，且 0.2 秒至 0.35 秒之間的取樣值均不大於零，如圖 4-3 與圖 4-4。

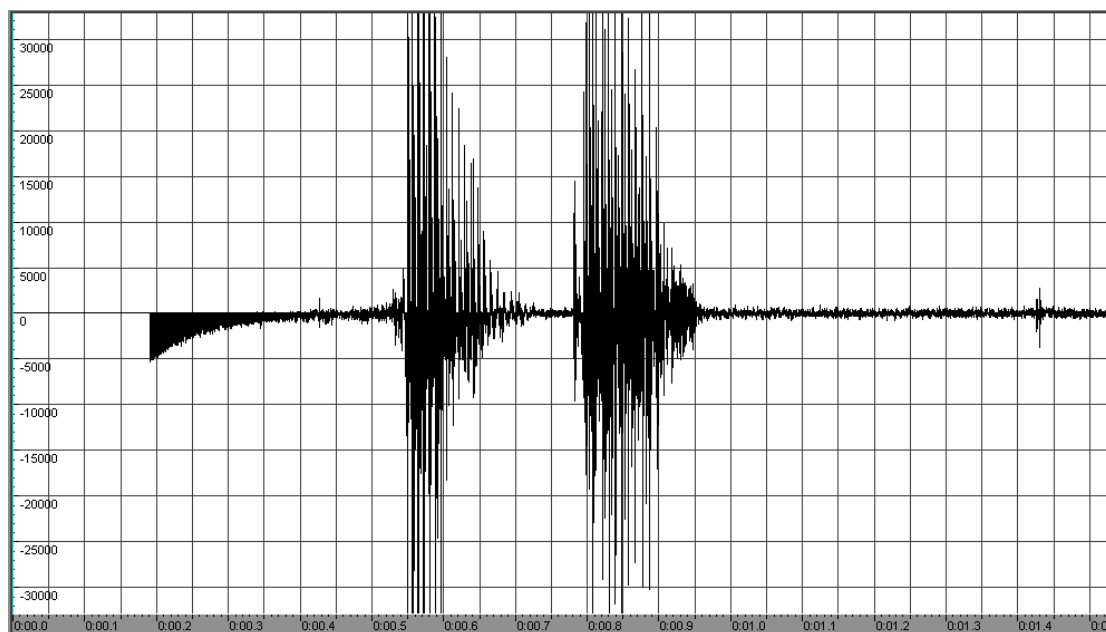


圖 4-3 實機錄製「放大」一詞之取樣值

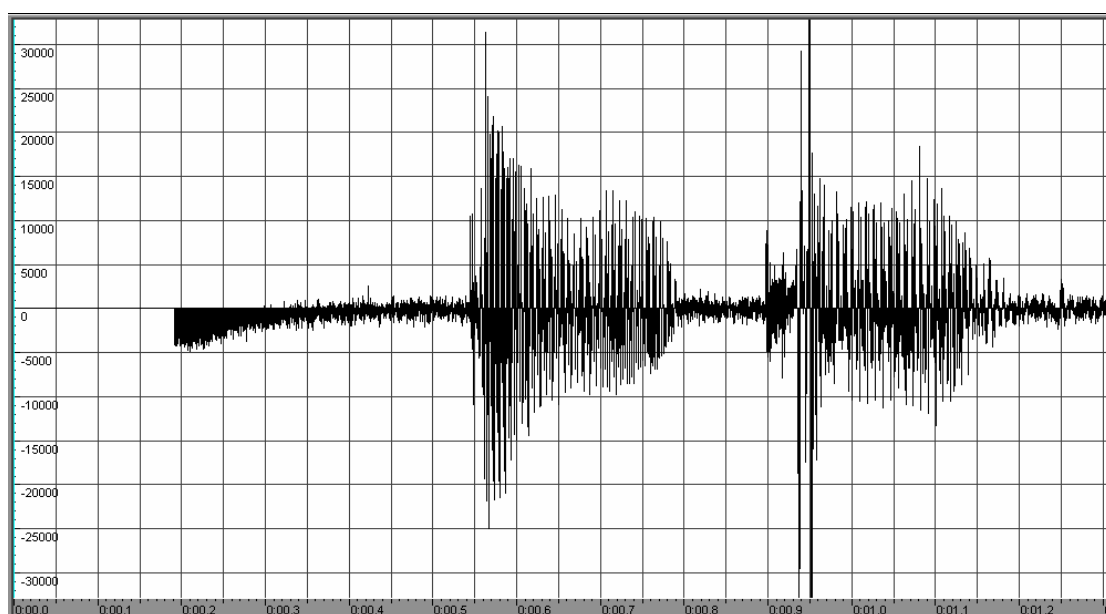


圖 4-4 實機錄製「地圖」一詞之取樣值

上述的錄音取樣問題並不會出現在一般桌上型電腦，這是由 Compaq iPAQ 系列 PDA 聲音輸入硬體設計所導致。探究其原因，可以確定原因的部分是 0.2 秒至 0.35 秒左右。該系列 PDA 採用電容式麥克風，當電容器尚未充滿電荷時就開始錄音，會導致聲音訊號在 0.2 秒至 0.35 秒左右產生直流偏壓 (DC Offset) 的現象。

至於訊號起點至 0.2 秒間的靜音現象，可能的原因是當時程式正在配置聲音訊號緩衝區，或者是電容器尚未達到工作電壓。由於缺乏相關技術文件，本研究無法確定造成靜音現象的原因。本研究將此二現象視為硬體上的限制。

#### 4.2.2 語音訊號前置處理

在語音訊號錄製完成後，接下來進行的工作是對語音訊號作前置處理，這包括了端點偵測（去除靜音部分）、訊號預強調、切割音框、漢明窗處理三個步驟，主要目的是加強語音訊號的特徵，使得語音訊號更具代表性。三個步驟分述如下：

##### 1. 端點偵測：

端點偵測的主要目的是將語音訊號中靜音或背景雜音的部分去除，一方面可以找出使用者輸入語音的起點與終點；另一方面去除不需要處理的部分，自然使得比對的範圍縮小，減少比對所耗費的時間。

本研究以能量量測法判定語音訊號的起點與終點，考慮靜音現象與直流偏移現象，從第 0.35 秒後，開始讀入取樣值至記憶體中的二維陣列（名為 Sample），如表 4-2 所示。

表 4-2 每八十筆取樣資料存入陣列

語音訊號				
第 0.35 秒	第 0.36 秒	第 0.37 秒	第 0.38 秒	.....

Sample(1, 1)	Sample(2, 1)	Sample(3, 1)	Sample(4, 1)	.....
Sample(1, 2)	Sample(2, 2)	Sample(3, 2)	Sample(4, 2)	.....
Sample(1, 3)	Sample(2, 3)	Sample(3, 3)	Sample(4, 3)	.....
Sample(1, 4)	Sample(2, 4)	Sample(3, 4)	Sample(4, 4)	.....
Sample(1, 5)	Sample(2, 5)	Sample(3, 5)	Sample(4, 5)	.....
⋮	⋮	⋮	⋮	
Sample(1, 79)	Sample(2, 79)	Sample(3, 79)	Sample(4, 79)	.....
Sample(1, 80)	Sample(2, 80)	Sample(3, 80)	Sample(4, 80)	.....

每取完八十筆資料，便計算該 0.01 秒中，樣本取樣值絕對值的平均數，也就是平均能量（Energy）。

$$Energy(i) = \frac{\left[ \sum_{j=1}^{80} |Sample(i, j)| \right]}{80} \dots\dots\dots (4-1)$$

一般判斷語音起點的方法是當平均能量值大於 240 的情況持續 0.02 秒時，便判定為語音起點。但本研究考慮車內環境雜音的音量較不固定，若設定一固定的門檻值將使得語音起點被環境雜音所影響。因此，本研究使用能量比例的方法，將環境雜音的影響降低，但使用者需要靠近麥克風才能有較準確的判定。若公式 4-2 成立，則語音訊號起點由  $i$  開始。

$$\frac{[Energy(i) + Energy(i-1)]}{[Energy(i-2) + Energy(i-3)]} > 2 \quad \dots\dots\dots (4-2)$$

至於偵測語音的終點則是採用能量門檻與能量比例並行的方式，兩者同時成立後，才判斷終點。

## 2. 預強調處理

聲學的研究指出，語音訊號經由嘴唇發出後，會有高頻的損失。但人耳的研究亦指出，人耳在 1KHz 到 4KHz 的頻率帶內具有 +6Db/octave 的高頻提升特性，相當於自動增益 (Auto Gain Control, AGC) 的功能。為了補償高頻的損失，將語音訊號通過一高頻濾波器，也就是所謂一階有限脈衝響應濾波器 (FIR Filter)：

$$\begin{cases} y(n) = x(n) - 0.95 \cdot x(n-1) \\ y(0) = x(0) \end{cases} \quad \dots\dots\dots (4-3)$$

## 3. 切割音框

由於聲帶無法在高低頻率之間快速的變化，因此語音具有短時距穩定的特性。切割音框的主要目的是在短時距穩定的情況下分析語音訊號，也就是以音框為單位，分析語音訊號，計算音框的特徵值。本研究以 240 個取樣值為一音框，各音框間重疊 160 個取樣點，避免語音訊號因為切割音框而產生不連續的情形。而本研究在端點偵測的步驟中，已將取樣值置入 Sample 二維陣列，因此，在切割音框的步驟中，只需重新定義二維陣列 (Frame)，並將 Sample 陣列依表 4-3 置入其中。

## 4. 漢明窗處理

在處理語音訊號時，因為考慮短時距穩定的特性，將語音訊號切割成為音框分段處理，雖然音框之間有相互重疊以避免發生訊號不連續的情況。但再將音框乘上漢明窗有兩項好處：第一，因為漢明窗為一鐘形函數 (圖 4-5)，音框乘上漢明窗後可將音框兩邊界壓低，使各音框之間更為連續。第二，音框乘上漢明窗後，在頻譜上的能量將更為集中，加強音框的語音特徵。漢明窗公式如 4-4，漢明窗處理是指公式 4-4 中預強調後的訊號  $y(n)$  乘上漢明窗  $h(n)$ ，如公式 4-5。其中  $N$  為一音框內取樣值數目，在本研究中  $N=240$ ，而  $n$  則由 0 至 239 之間， $n=0$  為音框第一筆資料， $n=239$  為音框最後一筆資料。

$$h(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad \dots\dots\dots (4-4)$$

$$s(n) = y(n) \times h(n) \quad \dots\dots\dots (4-5)$$

表 4-3 Sample 陣列與 Frame 陣列之關係

Sample(1, 1)~Sample(1, 80)					
Sample(2, 1)~Sample(2, 80)	Frame(1, 1) ~ Frame(1, 240)				
Sample(3, 1)~Sample(3, 80)			Frame(2, 1) ~ Frame(2, 240)		
Sample(4, 1)~Sample(4, 80)	Frame(3, 1) ~ Frame(3, 240)	Frame(4, 1) ~ Frame(4, 240)			
Sample(5, 1)~Sample(5, 80)					
Sample(6, 1)~Sample(6, 80)					
Sample(7, 1)~Sample(7, 80)					
⋮				⋮	

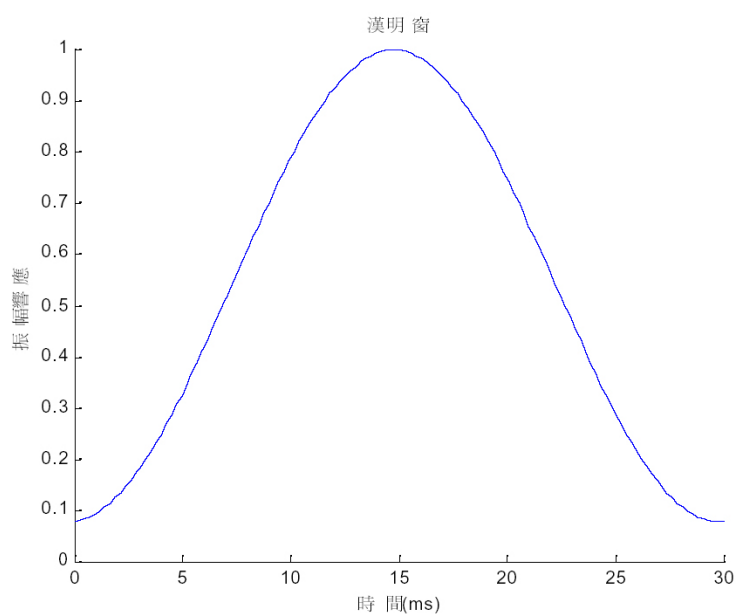


圖 4-5 漢明窗函數時域圖

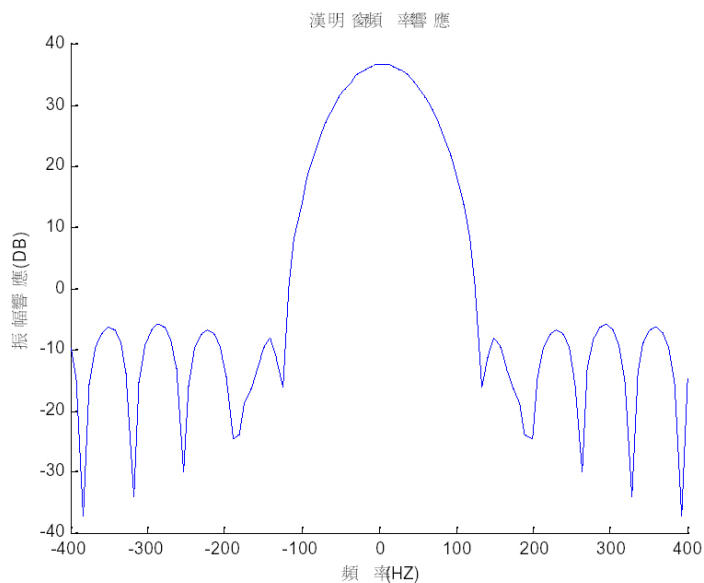


圖 4-6 漢明窗頻域圖

#### 4.2.3 計算音框特徵參數

本研究所使用之特徵參數為倒頻譜參數以及轉移倒頻譜參數，每音框共 28 組，其中倒頻譜參數 14 組，轉移倒頻譜參數 14 組。

經過漢明窗處理的音框訊號，至求得音框特徵參數，需以下幾個步驟：

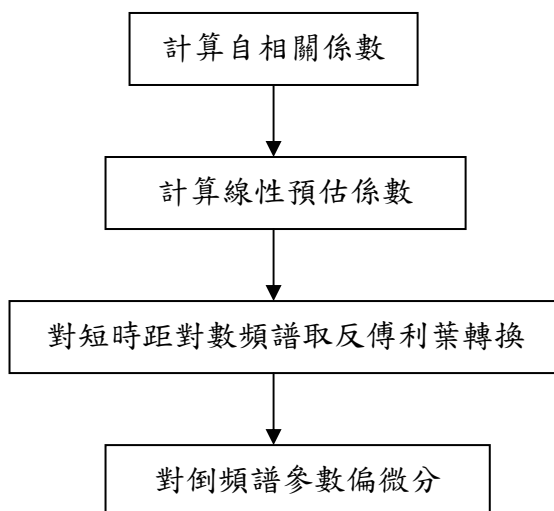


圖 4-7 特徵參數計算步驟

由於語音訊號的短時距穩定特性，因此，音框中某一訊號值可由鄰近訊號值的線性組合來預估。這也是語音辨識技術中已使用很普遍的語音參數之一，所謂線性組合的預估方式，如以下式子的說明。

$$\hat{S}_n = a_1 \times S_{n-1} + a_2 \times S_{n-2} + \cdots + a_p \times S_{n-p} + e_n \cdots \cdots (4-6)$$

$$\hat{S}_n = \sum_{k=1}^p a_k \cdot S_{n-k} + e_n \cdots \cdots (4-7)$$

其中  $\hat{S}_n$  是預估的語音訊號， $S_{n-1} \cdots S_{n-p}$  是  $S_n$  之前  $p$  個語音訊號， $a_1 \cdots a_p$  是線性預估係數 (Linear Predictive Coefficient)，也稱為線性預估編碼 (Linear Predict Coding)，使得音框中所有的  $\hat{S}_n$  與實際值差異最小，亦即求下列式子的最小值。

$$E = \sum e_n^2 \cdots \cdots (4-8)$$

將 (4-8) 代入 (4-9) 後，

$$E = \sum \left( \hat{S}_n - \sum_{k=1}^p a_k \cdot S_{n-k} \right)^2 \cdots \cdots (4-9)$$

解最小值的問題，在計算線性預估係數前，必須先求得自相關係數 (Autocorrelation Coefficient)，一般採用「自相關係數法」以及「變異法」，其中自相關係數法運算速度較快，因此本研究採用該演算法。該演算法利用矩陣以及偏微分等運算求解最小值問題，將運算過程簡化為一般式後，得自相關係數  $r_k$  求法如下所示：

$$r_k = \sum_{n=0}^{N-k-1} S_n \cdot S_{n+k}, \quad k = 0, 1, \cdots, p \cdots \cdots (4-10)$$

其中  $N$  是取樣點數目， $p$  代表線性預估係數的階數，本研究中  $p$  值為 12，使用 12 個線性預估係數。以一個音框為例，程式流程如圖 4-8。

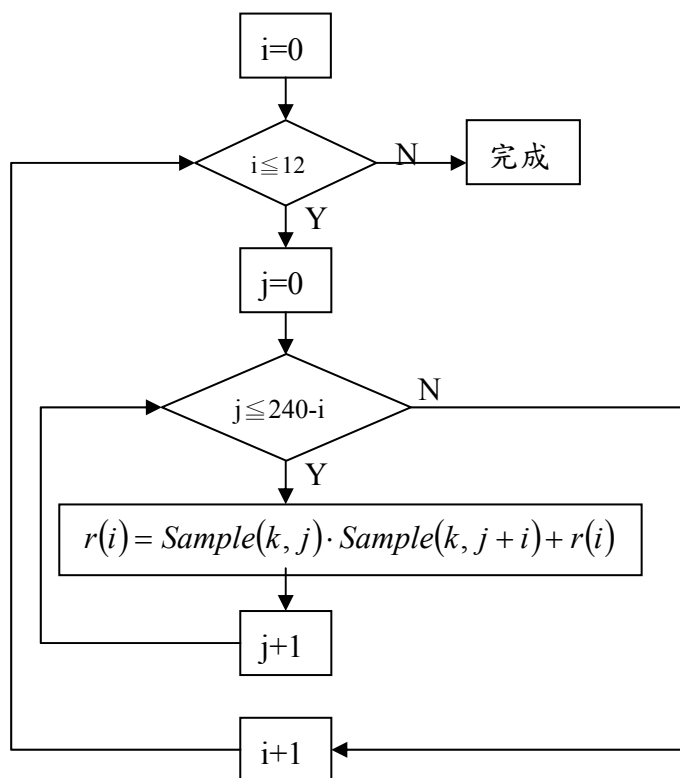


圖 4-8 求取自相關係數程式流程圖

自相關係數計算完畢後，接下來便將自相關係數套入杜賓演算法 (Durbin algorithm) 中以計算線性預估係數，此演算法步驟如下。

$$E_0 = r_0 \dots\dots\dots (4-11)$$

$$k_i = \left[ r_i - \sum_{j=1}^{i-1} a_j^{(i-1)} \cdot r_{i-j} \right] / E_{i-1}, \quad 1 \leq i \leq p \dots\dots\dots (4-12)$$

$$a_i^{(i)} = k_i \dots\dots\dots (4-13)$$

$$a_j^{(i)} = a_j^{(i-1)} - k_i \cdot a_{i-j}^{(i-1)}, \quad 1 \leq j \leq i-1 \dots\dots\dots (4-14)$$

$$E_i = (1 - k_i^2) \cdot E_{i-1} \dots\dots\dots (4-15)$$

$$a_j = a_j^{(p)}, \quad 1 \leq j \leq p \dots\dots\dots (4-16)$$

(4-12) 至 (4-15) 間的運算是以遞迴的方式計算各  $k_i$ ，直到所有  $k_i$ ， $1 \leq i \leq p$  都求出來為止。(4-16) 式的意義是各線性預估係數  $a_j$  的值是由  $a_j^{(p)}$  指定，例如： $a_1 \dots\dots a_{12}$  的值分別是  $a(1,12)$ 、 $a(2,12)$ 、 $\dots\dots a(12,12)$ 。以一個音框為例，程式流程如圖 4-9。

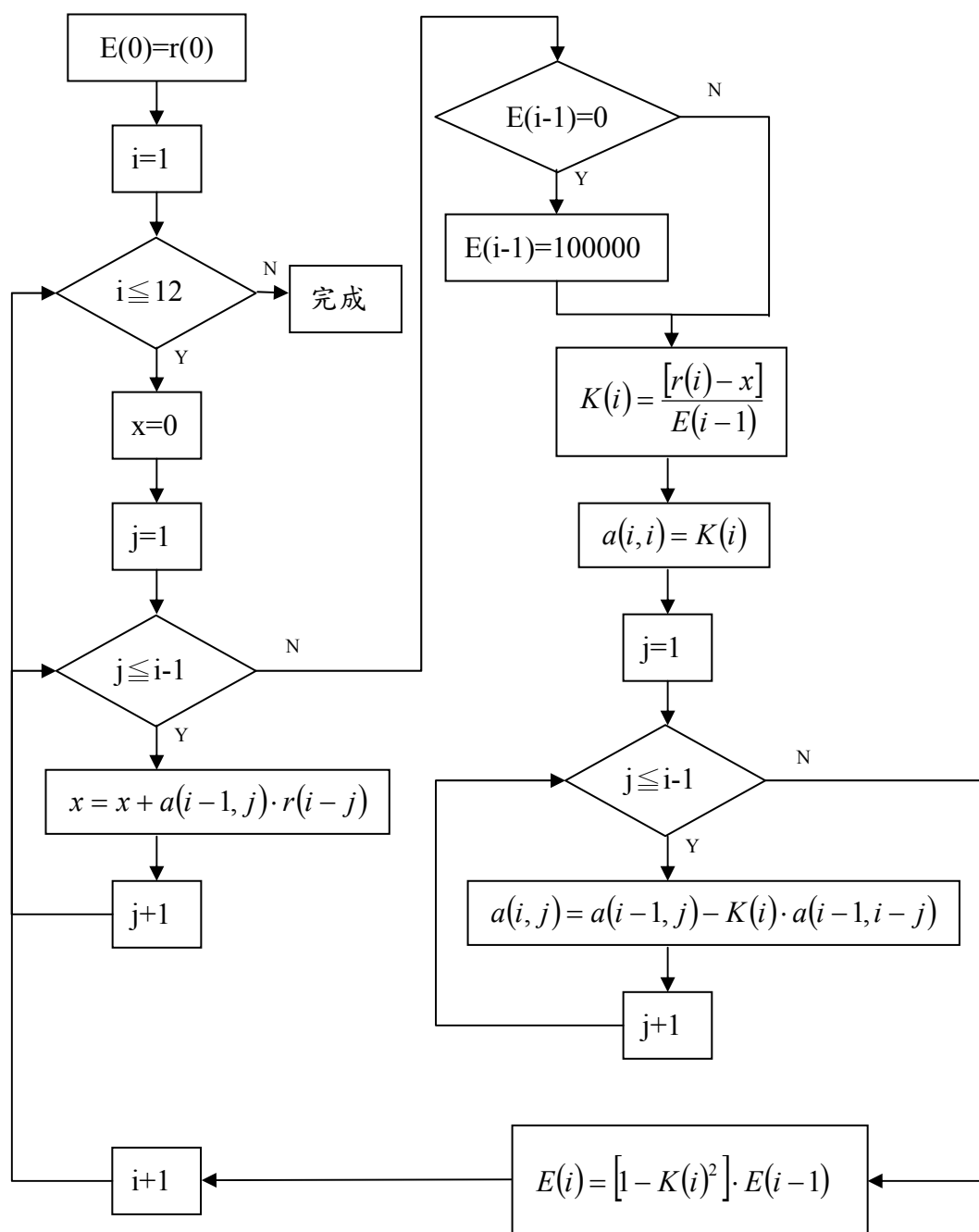


圖 4-9 求取線性預估係數之程式流程圖

在經由自相關係數計算線性預估係數後，便可以進行倒頻譜(Cepstrum)參數的計算。由線性預估係數進入倒頻譜領域的方式是對短時距對數頻譜取反傅利葉轉換，聲音訊號在頻域的特徵往往比時域訊號來得更能顯示出訊號的特徵，12 階線性預估係數計算出 14 組倒頻譜參數，為本研究使用之 28 組語音特徵值中的前 14 組，後 14 組為轉移倒頻譜參數。倒頻譜參數一般式可表示如下。

$$C_1 = a_1 \dots\dots\dots (4-17)$$

$$C_n = a_n + \sum_{m=1}^{n-1} \left(1 - \frac{m}{n}\right) \cdot a_m \cdot c_{n-m} \quad , \quad 1 < n \leq p \dots\dots\dots (4-18)$$

$$C_n = \sum_{m=1}^p \left(1 - \frac{m}{n}\right) \cdot a_m \cdot c_{n-m} \quad , \quad n \geq p \dots\dots\dots (4-19)$$

其中代數  $a$  為上述已求得之線性預估係數， $C$  為倒頻譜參數， $p$  為線性預估係數階數。以一個音框為例，程式流程如圖 4-10。

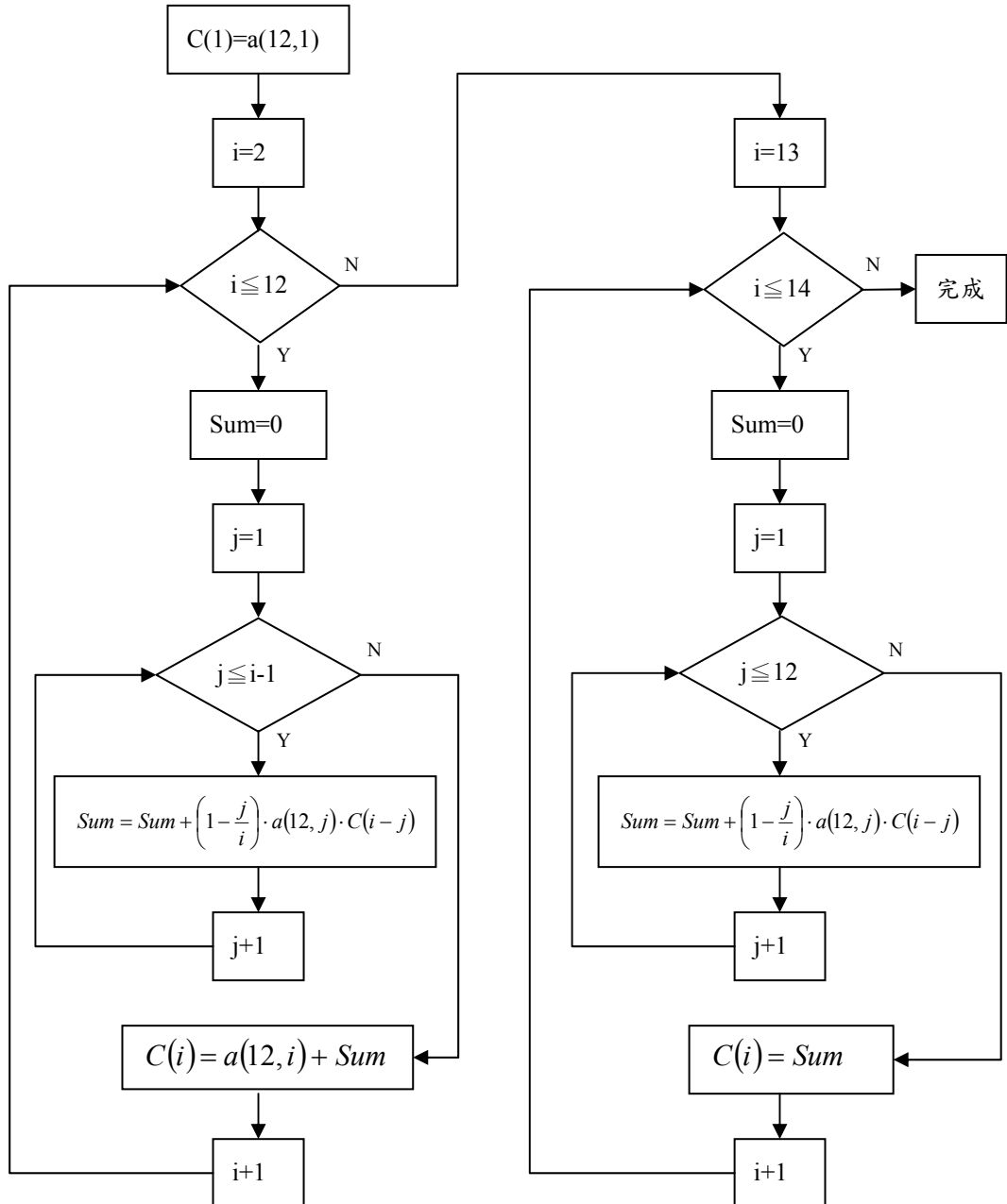


圖 4-10 計算倒頻譜參數之程式流程圖

圖 4-11 為將「放大」一詞之語音訊號經過上述處理後之倒頻譜參數，

可看出該詞各頻率能量在頻譜上的變化，共 41 個音框，以 S1 至 S41 表示，每音框有 14 個倒頻譜參數。

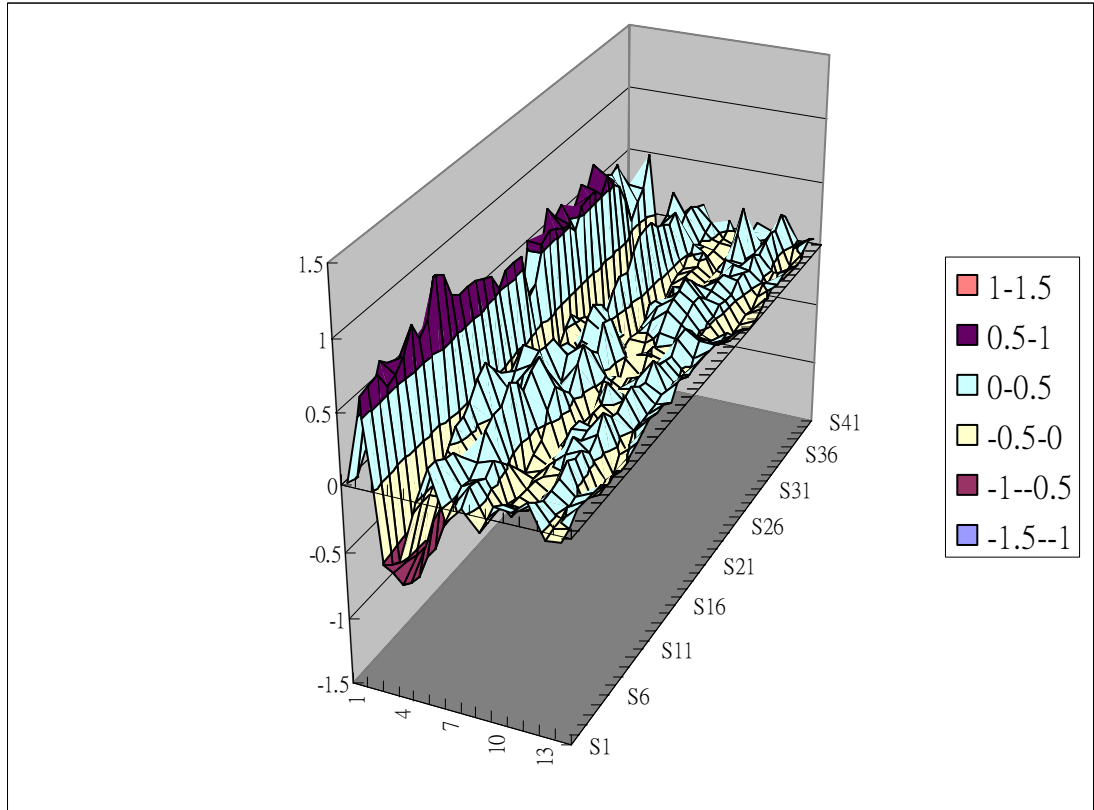


圖 4-11 「放大」一詞之倒頻譜參數

轉移倒頻譜參數是將倒頻譜參數對時間做偏微分後所得到的特徵值。因為語音大多會受到外界雜訊的影響，使得頻譜受到干擾而有所扭曲，但是雜訊對轉移倒頻譜參數影響並不大，因為倒頻譜參數對時間偏微分後，所得之轉移倒頻譜參數是一種動態參數。計算方式如下。

$$\Delta C_n(t) = \frac{\partial C_n(t)}{\partial t} = \frac{\sum_{k=-K}^K k \cdot C_n \cdot (t+k)}{\sum_{k=-K}^K k^2} \dots\dots\dots (4-20)$$

在程式的實作上，可採用轉移倒頻譜之速算公式，假設  $C(m,n)$  表示第  $m$  個音框第  $n$  個倒頻譜參數， $\Delta C(m,n)$  表示第  $m$  個音框第  $n$  個轉移倒頻譜參數，共  $L$  個音框。對應五種音框位置，有五種速算公式。

$$\Delta C(1,n) = \frac{[2 \cdot C(3,n) + C(2,n)]}{5}, \text{ 第一個音框} \dots\dots\dots (4-21)$$

$$\Delta C(2,n) = \frac{[2 \cdot C(4,n) + C(3,n) - C(1,n)]}{6}, \text{ 第二個音框} \dots\dots\dots (4-22)$$

$$\Delta C(L-1,n) = \frac{[2 \cdot C(L,n) - C(L-2,n) - 2 \cdot C(L-3,n)]}{6}, \text{ 第 } L-1 \text{ 個音框} \dots\dots\dots (4-23)$$

$$\Delta C(L, n) = \frac{[-C(L-1, n) - 2 \cdot C(L-2, n)]}{5}, \text{ 第 } L \text{ 個音框} \dots\dots\dots (4-24)$$

$$\Delta C(i, n) = \frac{[2 \cdot C(i+3, n) + C(i+2, n) - C(i, n) - 2 \cdot C(i-1, n)]}{10}, \text{ 第 } i \text{ 個音框} \dots\dots\dots (4-25)$$

計算轉移倒頻譜參數之程式流程如下。

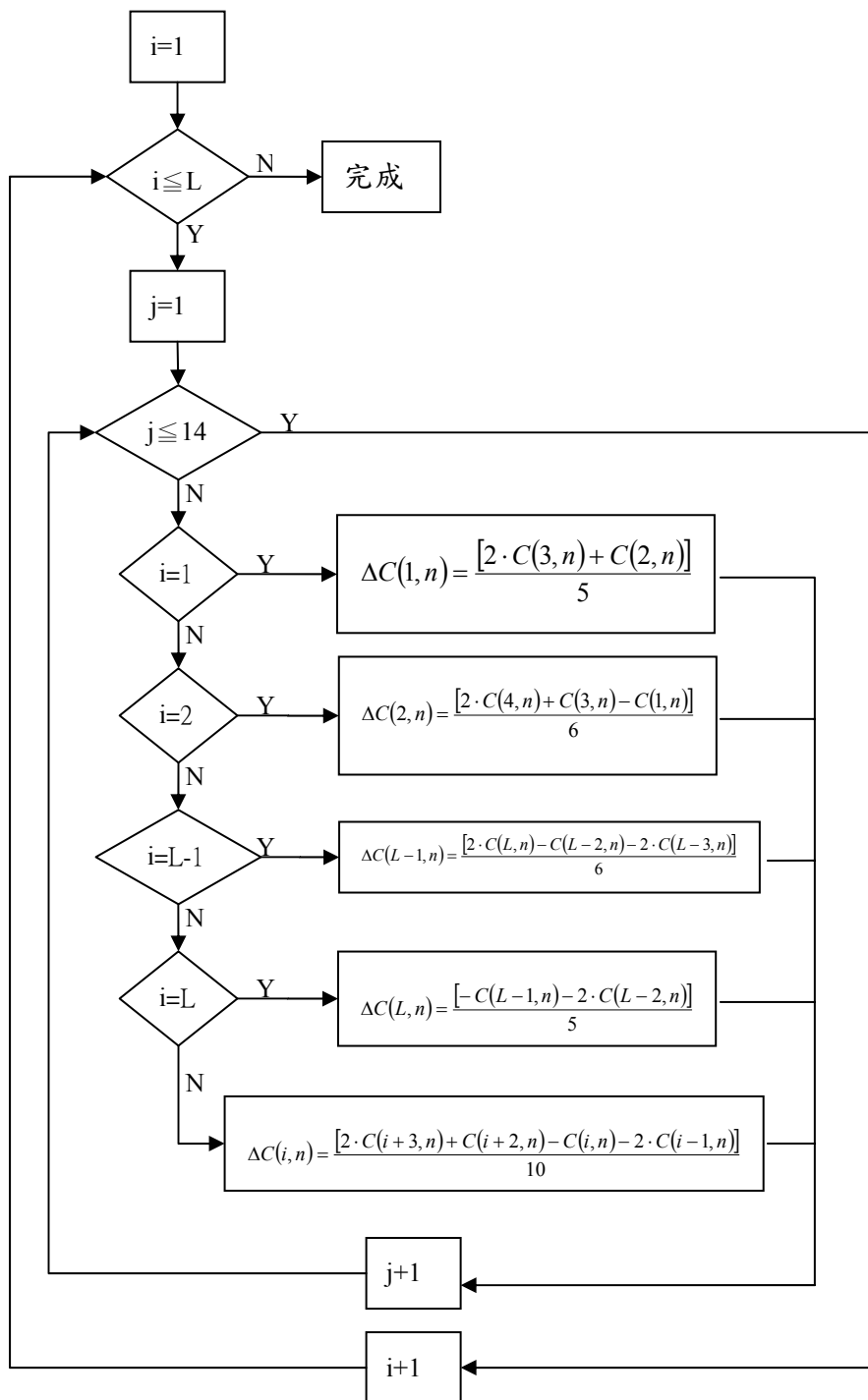


圖 4-12 計算轉移倒頻譜參數之程式流程圖

#### 4.2.4 動態時間校準法

由於每次錄製的語音訊號時間長短會隨著說話的快慢而不盡相同，因此若直接比較參考樣本（語音模型）與測試樣本（待測試之訊號）之音框特徵值差異，勢必會因為兩樣本之音框數不同而在比對過程中產生誤差。

較好的比對方法是考慮到參考樣本與測試樣本的訊號長短，當測試樣本音框數較少時就代表說話速度較快，等於是將聲音訊號壓縮；而說話速度慢時，像是將聲音訊號擴張。動態校準法能隨兩訊號時間長短的差異而改變其比對路徑，且在運算量方面比其他辨識方法如隱藏式馬可夫鏈、類神經網路等要來得少。動態校準法是由動態規劃（Dynamic Programming, DP）而來，概念是將參考樣本與測試樣本中的參數比對，計算單一節點失真量，再與上一節點之最小累積失真量相加。如此遞迴最終累積失真量即為總體失真量。

在理想的狀況下，可以假設參考樣本與測試樣本之音框數相等，即  $M = N$ ，則整體失真量  $D$  可表示為：

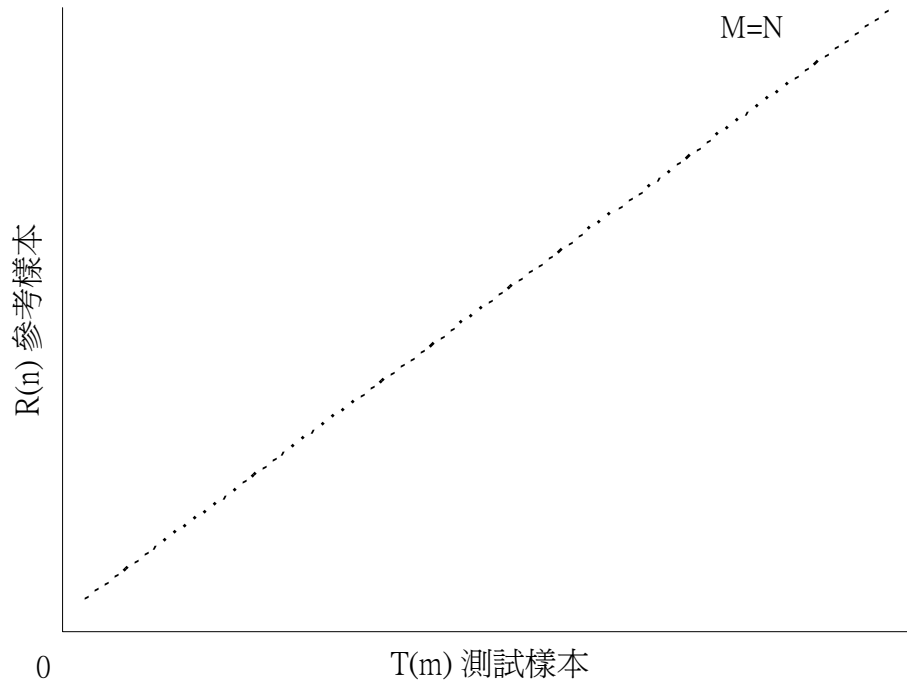


圖 4-13 理想狀況之動態校準法路徑

$$D = \sum_{m=1}^M d(T(m), R(n)) \dots\dots\dots (4-26)$$

其中， $T$  為測試樣本，共有  $M$  個音框， $T = \{T(m) | m = 1, 2, \dots, M\}$

$R$  表示參考樣本，共有  $N$  個音框， $R = \{R(n) | n = 1, 2, \dots, N\}$

$T(m)$  表示測試樣本中，第  $m$  個音框的特徵參數向量

$R(n)$ 表示參考樣本中，第 $n$ 個音框的特徵參數向量

$d$ 為參考樣本與測試樣本在某一音框的局部失真量

然而上述假設 $M = N$ 之筆直的比對路徑在實際的情況下，出現機率不大。因為隨著語音訊號長度的變異，通常是 $M > N$ 或 $M < N$ 的情形。動態校準法配合時間軸扭曲，在 $d(T(1), R(1))$ 至 $d(T(m), R(n))$ 間搜尋一最佳路徑，使得整體的失真量達到最小。

但是若直接在 $M \times N$ 個 $d$ 值中搜尋最佳路徑，將是耗費龐大計算時間的作法。對於語音辨識來說，不論測試樣本與參考樣本之間相似性，比對路徑總是在斜率 $N/M$ 附近，便可以決定辨識結果。距離該斜率線太遠的比對，因為已經超出合理的時間擴張、壓縮情況，所以可將搜尋的範圍作一限制，以降低運算量。動態校準法之限制式可分為兩種：整體搜尋路徑限制（Global Path Constrains）與局部搜尋路徑限制（Local Path Constrains）。

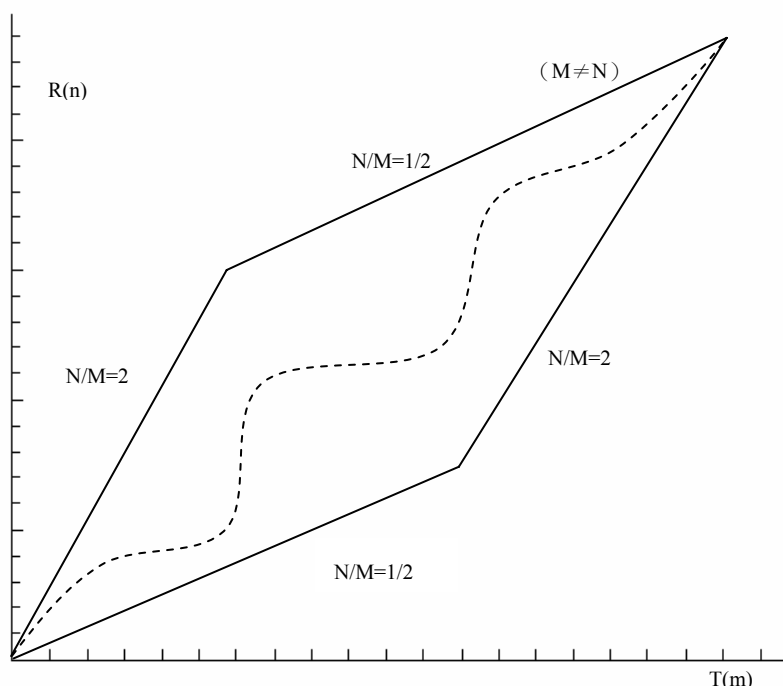


圖 4-14 整體搜尋路徑限制之平行四邊形

在整體搜尋路徑限制中，搜尋路徑限制於斜率 2 與 1/2 的平行四邊行之間，如圖 4-14 所示，而其限制式可表示如下。

$$1 + \frac{[\Phi_x(k(m, n)) - 1]}{2} \leq \Phi_y(k(m, n)) \leq 1 + 2 \cdot [\Phi_x(k(m, n)) - 1] \dots\dots\dots (4-27)$$

$$N + 2 \cdot [\Phi_x(k(m, n)) - M] \leq \Phi_y(k(m, n)) \leq N + \frac{[\Phi_x(k(m, n)) - M]}{2} \dots\dots\dots (4-28)$$

其中 $k(m, n)$ 為比對空間中的一點， $\Phi_x$ 函數能取出該點 $x$ 座標值， $\Phi_y$ 函

數則取出  $y$  值。即  $\Phi_x(k(m,n))=m$  ,  $\Phi_y(k(m,n))=n$  。

而除了整體搜尋路徑限制之外，尚有局部搜尋路徑限制使得演算法能在平行四邊形內找尋失真量最小的路徑。以  $D_{\min}(m,n)$  表示累積的最小失真量，其搜尋方法例如：目前演算法進行至節點  $(m,n)$  時，在鄰近三組資料中找最小值，也就是找  $D_{\min}(m-1,n)$  ,  $D_{\min}(m-1,n-1)$  ,  $D_{\min}(m-1,n-2)$  三者中的最小值，找到後再將該  $D_{\min}$  加上目前所在節點之  $d$  值，成為  $D_{\min}(m,n)$ 。其公式可以表示如下。

$$D_{\min}(m,n) = d(T(m), R(n)) + \min \begin{cases} D_{\min}(m-1,n) \\ D_{\min}(m-1,n-1) \dots\dots\dots (4-29) \\ D_{\min}(m-1,n-2) \end{cases}$$

如果目前搜尋的進度位於平行四邊形內的節點  $(m,n)$ ，則被比較三個點為  $(m-1,n)$  ,  $(m-1,n-1)$  ,  $(m-1,n-2)$ ，只往左方、左下方搜尋是因為以時間順序來說，音框  $n+1$  在音框  $n$  之後，比對路徑若由  $(m-1,n+1)$  回到  $(m,n)$ ，將會造成時間上的矛盾。因此，搜尋路徑在時間上只向下一個音框或同一個音框移動，圖形有由左下往右上的趨勢。搜尋路徑示意圖如圖 4-15。

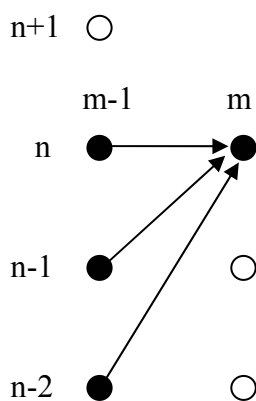


圖 4-15 決定  $D_{\min}(m,n)$  時搜尋三節點示意圖

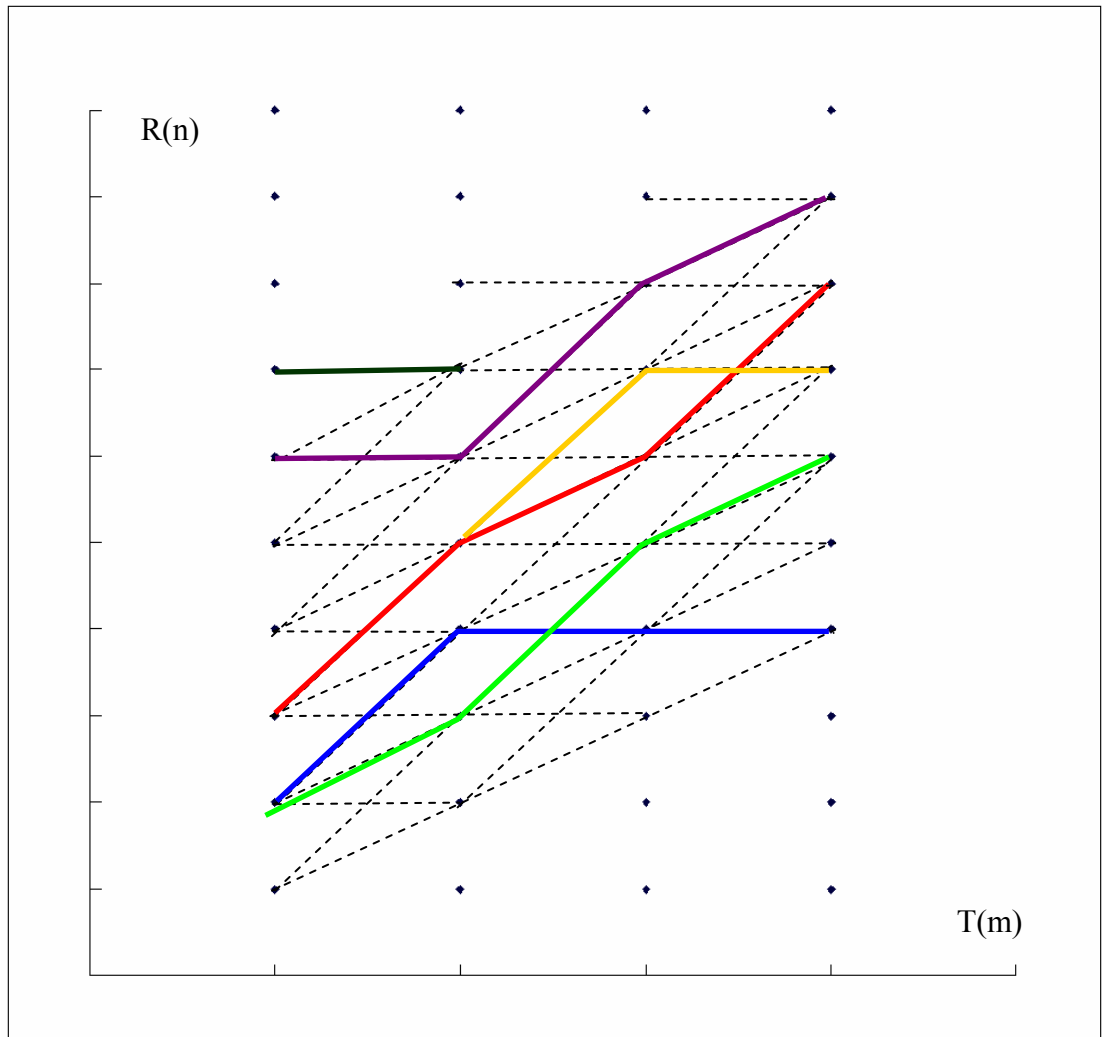


圖 4-16 在解空間中可能出現的搜尋路徑示意圖

圖 4-17 為本研究中動態校準法之程式流程，測試樣本音框數為  $M$ ，參考樣本音框數為  $N$ 。

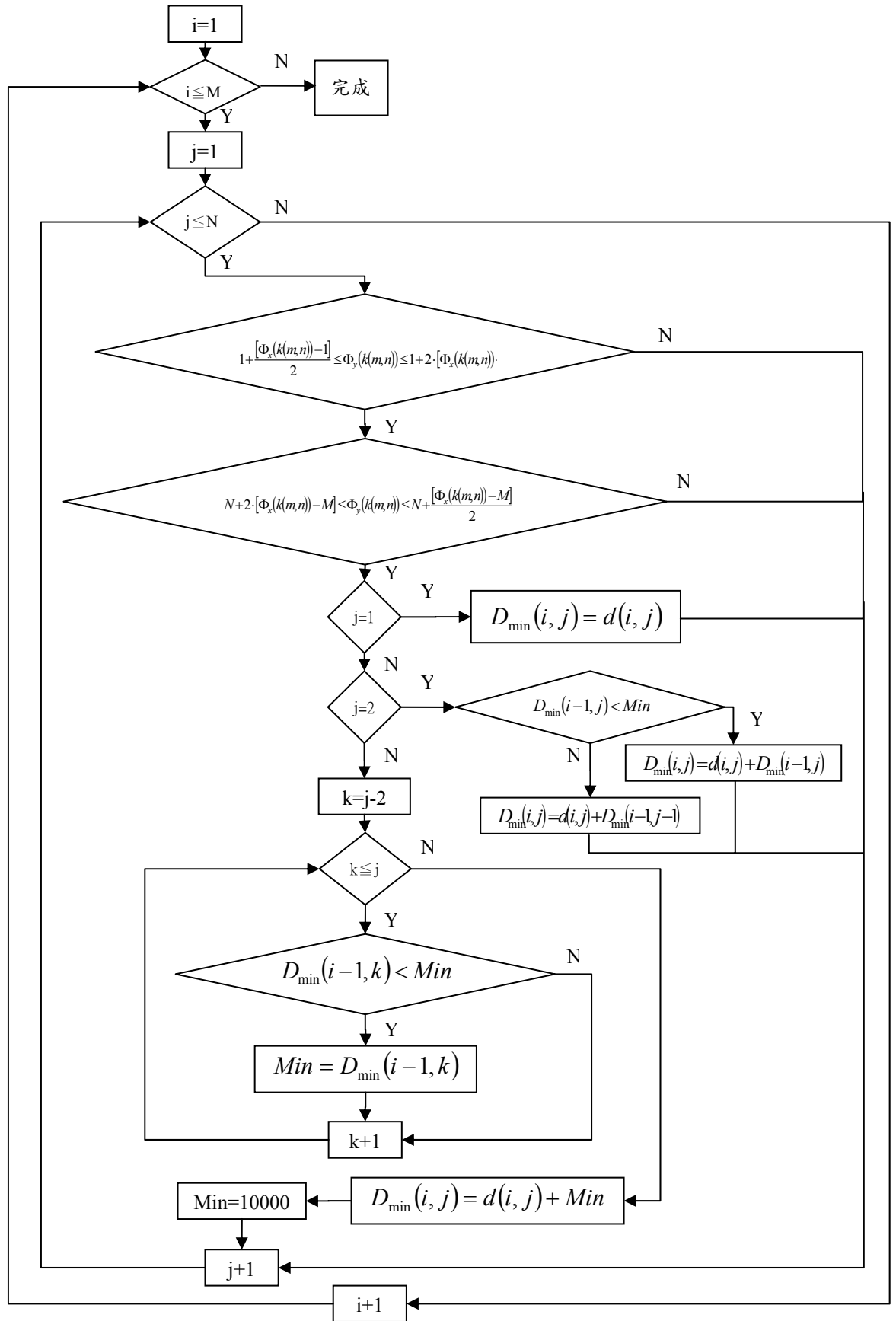


圖 4-17 動態校準法程式流程圖

而本研究第一次完成「放大」、「縮小」二詞的倒頻譜參數，轉移倒頻譜參數的計算後，將二詞之特徵參數紀錄於程式中成為參考樣本（或稱為語音模型），接著重新錄製二詞，同樣進行特徵參數計算，並將記憶體內之計算結果寫入文字檔案儲存，最後將此內含測試樣本特徵參數之文字檔案與參考樣本特徵參數兩者經由其他軟體繪製立體曲面圖，如此能更具體的看出參考樣本與測試樣本之特徵值差異，其中圖 4-18 橫軸的數列為「放大」測試樣本，縱軸數列 S 為「放大」參考樣本；圖 4-19 橫軸的數列為「縮小」之參考樣本，縱軸數列 S 為「放大」之參考樣本。

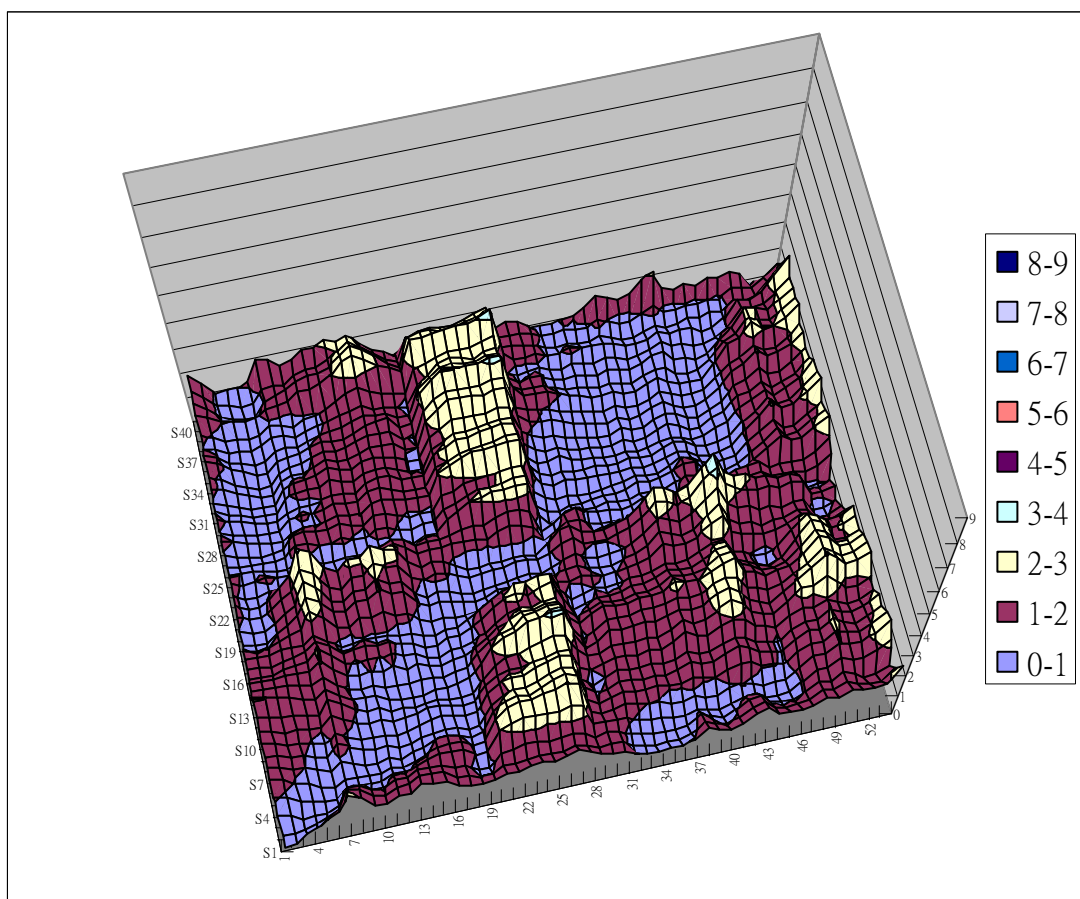


圖 4-18 「放大」一詞之測試與參考樣本差異之立體曲面圖

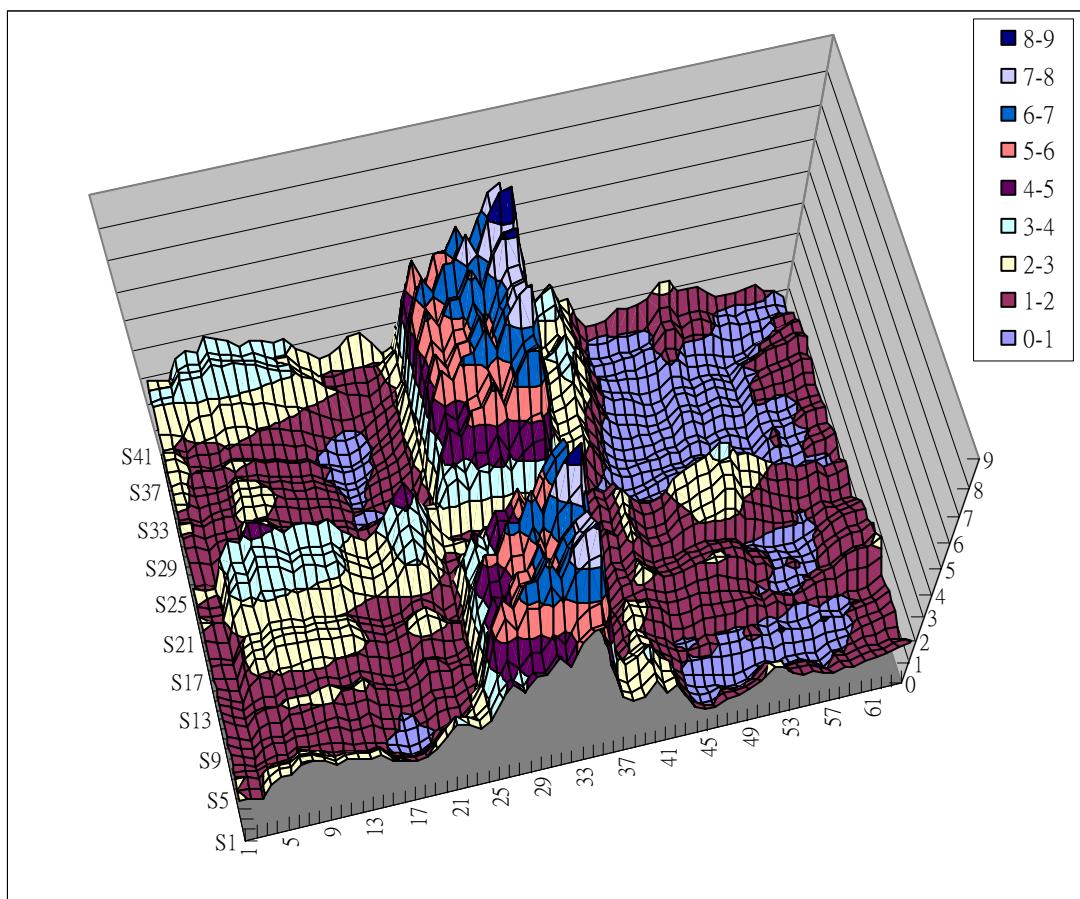


圖 4-19 「放大」、「縮小」二詞參考樣本差異之立體曲面圖

觀察圖 4-18 與圖 4-19，圖中曲面的高度代表兩軸特徵值的差異大小，也就是前述之  $d(T(m), R(n))$ ，高度越高表示差異越大。在圖 4-18 中清楚地表現出由左下角之  $(1,1)$  至右上角  $(m,n)$  間的確有持續低曲面高度的路線，這是「放大」一詞之測試樣本與參考樣本差異的情況。

而圖 4-19 為「放大」與「縮小」之參考樣本差異的情況，圖 4-18 中持續低曲面高度的路線不復存在，「縮小」約在第 25 音框至第 45 音框間的特徵值與「放大」的所有特徵值均有較大的差異，在經過動態校準法的計算後，將呈現較高的總體失真量。語音辨識的原理就是將測試樣本特徵值與各關鍵詞做比較，找出總體失真量最小的關鍵詞，並執行相關動作。

### 4.3 文字轉語音技術

在電腦中出現的文字起初是以 1byte，也就是 8-bit 表示 256 個字母。但 256 種字母對於中文來說是不敷使用的，因此出現 Big-5 碼，以 2byte，16-bit 達到可表示 65536 種字體的水準。而 PDA 上使用的文字均屬於 Unicode 編碼方式，與 Big-5 相同，使用 2byte 紀錄一個中文字。因此，在 PDA 上若比較兩中文字是否為同一字，是由 Unicode 號碼來判定。

國字有許多同音字，例如「八」、「巴」、「叭」、「扒」、「吧」都是發ㄅㄚ

的拼音，因此可建立國字與發音的對照關係，輸入文字可輸出該文字拼音方式。國字共 1256 種拼音，以下舉例經本研究整理後的第一種至第十四種發音：

八 0001 巴 0001 叭 0001 扒 0001 吧 0001 芭 0001 疤 0001 捌 0001 笆 0001 伯 0002  
 把 0002 爸 0002 耙 0002 罷 0002 霸 0002 壩 0002 灞 0002 拔 0003 跋 0003 鉞 0003  
 把 0004 靶 0004 吧 0005 罷 0005 玻 0006 剝 0006 般 0006 鉢 0006 發 0006 菠 0006  
 撥 0006 北 0007 亳 0007 播 0007 擘 0007 薄 0007 薛 0007 簸 0007 藥 0007 白 0008  
 百 0008 伯 0008 孛 0008 佰 0008 帛 0008 泊 0008 勃 0008 柏 0008 悖 0008 脖 0008  
 舶 0008 萆 0008 博 0008 渤 0008 蔽 0008 搏 0008 鉈 0008 箔 0008 膊 0008 駁 0008  
 蔔 0008 魄 0008 磔 0008 薄 0008 爆 0008 簿 0008 跛 0009 簸 0009 拜 0010 敗 0010  
 白 0011 百 0012 伯 0012 佰 0012 擺 0012 襪 0012 卑 0013 杯 0013 盃 0013 背 0013  
 悲 0013 搯 0013 痺 0013 碑 0013 裨 0013 北 0014 孛 0014 貝 0014 沸 0014 邶 0014

在 PDA 中，上述資料的儲存方式是以字串 (String) 的方式儲存於記憶體中，字串長度為 900byte，包含了 450 個 Unicode 字體。當「北」需要被轉換為語音時，程式於字串中尋找「北」出現的位置，即為上述資料中倒數第五個國字，並讀出該位置後四位 Unicode，即為編號第 14 的發音，播放該發音，完成文字轉語音動作，如圖 4-20 所示。

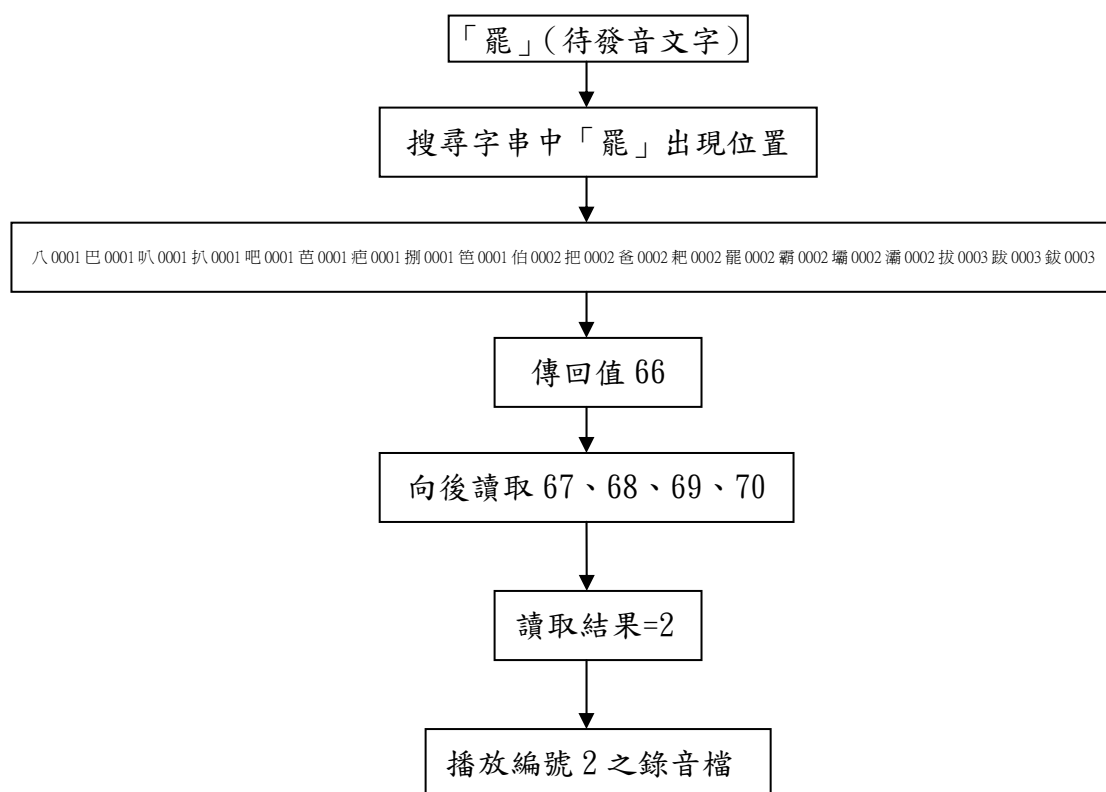


圖 4-20 文字轉語音程式流程

但是文字轉語音最大的困難是在破音字的處理，因為上述的方法若輸入字為破音字，則容易有發音錯誤的情況發生。一般在處理類似問題的方法通常是建立破音字詞庫，由上下文的關係減少發音錯誤的情況，並利用字詞庫的功能將「1234 元」正確唸出「一千兩百三十四元」，這些功能都將是本系統未來需要改進的地方。

## 第五章 導航系統展現

語音辨識與提示的原理，方法，程式流程已於第四章完成。本章將導航系統所能達到的功能作一展示。範例路徑為臺北郵局至基隆海門天險，路徑旁的設施包括飯店、加油站，與停車場共三種，分別以紅色、綠色，與藍色表示該設施。設置的按鈕有「放大」、「縮小」、向上、下、左、右平移，讀入是指將最短路徑座標檔案讀入記憶體中，「Run」按鈕是執行語音辨識，「錄音」按鈕按下後將顯示錄音控制列供使用者開始錄音，停止錄音，「Go」按鈕則是開始讀入車輛移動路徑，為了方便展示，車輛設定以每秒 44 公尺移動，「停」按鈕則是停止車輛移動。

### 5.1 地圖操作

#### 5.1.1 讀入路徑與設施座標檔案

啟動軟體後按下「讀入」按鈕，此時系統會讀入儲存於文字檔中的路徑與設施座標檔案，並進行座標系統轉換，搜尋其中最大與最小的座標值，使得讀入後的地圖顯示自動調整為顯示全部地圖。另外，各關鍵詞的語音模型特徵參數此時也讀入記憶體中。



圖 5-1 讀入導航路徑簡圖

### 5.1.2 放大與縮小功能

由於按下讀入按鈕後，顯示的地圖為全部路徑，由臺北郵局至海門天險約三十公里，此時若顯示路名、設施名稱等資訊，將造成文字重疊不清的狀況。因此，經過試驗後，程式設定在放大 32 倍或更高時才顯示路名、設施名稱，會有較佳的效果。每次放大、縮小的倍率為 2 倍，因此相當於按了五次「放大」按鈕。在放大按鈕累積次數與縮小按鈕累積次數相差 5 次以上，都會顯示路名、設施資訊，避免縮小至容易使文字重疊不清的倍率時，文字仍然持續顯示。



圖 5-2 放大 32 倍之後顯示文字

### 5.1.3 地圖平移功能

考慮到使用語音辨識時做地圖平移功能時，無法如觸控筆一般拖曳畫面達到平移效果，程式設計成具有上、下、左、右按鈕，以達到平移的功能，每次移動 500Twips，Twips 為 eVB 中座標單位，1Twip 約合 1/567 公分。圖 5-3 顯示圖 5-2 向上平移後的結果。



圖 5-3 向上平移後的結果

## 5.2 錄音操作

按下「錄音」按鈕後，將出現錄音控制列，供使用者錄音以及停止錄音，並播放錄音內容。當使用者按下 OK 按鈕後，錄音檔案將被儲存，待使用者按下「Run」按鈕後，語音辨識所分析的聲音訊號即為此檔案。圖 5-4 為使用者按下「錄音」按鈕後的畫面。



圖 5-4 錄音控制列

## 5.3 轉向與設施提示功能

### 5.3.1 轉向提示

按下「Go」按鈕後，車輛便開始沿著導航路徑前進，車輛周邊並顯示一圓形範圍，表示當前方轉向路口接近並進入此圓形範圍時，便發出轉向提示。為了使語音提示簡潔，也為了避免在同一條道路上重複提示，當轉向角度小於 15 度時，提示內容便不包括轉向角度，僅提示距離以及路名。除了在語音上提示，為了方便在書面上呈現效果，提示內容也同時顯示於畫面上方的文字方塊，如圖 5-5。



圖 5-5 轉向提示

### 5.3.2 設施提示

為了使駕駛人聽到設施提示時能分辨該設施位於道路左邊或右邊，在設施提示方面於文字方塊中加入該設施目前與車輛所在位置之十二時方位角，文字顏色亦與設施顏色相同，如果駕駛人有察看的需求，可以由顏色更容易地瞭解是何種設施，視線可以很快的抓住該顏色的文字內容。而語音提示內容則如同第三章所述。



圖 5-6 設施提示

## 5.4 語音辨識結果顯示

按下「Run」按鈕後，將把錄音檔案讀入記憶體中進行語音辨識的處理，並且與按「讀入」時所讀入的各語音模型特徵值比較。比較結果顯示於文字方塊中，如圖 5-7 所示，結果顯示「放大」失真量最小，因此執行地圖放大的動作。



圖 5-7 語音辨識後的結果

## 第六章 結論與建議

本研究之導航系統主要建構於張晉維與林詠傑所提出的主從架構，使得 PDA 端經無線通訊由伺服器端下載簡圖導航路徑，且簡圖導航路徑所佔的檔案非常小，適合無線傳輸以及 PDA 儲存容量較小的特性。而本研究一方面思考如何提高駕駛者之行車安全，另一方面也因為 PDA 技術的進步，目前市面上已有中央處理器達到 400Mhz 之高速 PDA 機種，能快速處理數值計算。因此，本研究以加入語音辨識與提示功能於簡圖導航系統為研究重點，讓原本的簡圖導航系統功能更加完備。

### 6.1 結論

#### a. 於導航系統中實現語音辨識

將語音辨識實現於 PDA 之簡圖導航系統中，辨識的範圍屬於特定詞辨識。在本研究中所使用的訊號處理技術，均是語音辨識領域中所慣用的，或者是計算量較少的。從讀入錄音檔案開始，本研究將理論公式化為程式實作，包括語音訊號的端點偵測、預強調處理、切割音框、漢明窗處理、計算自相關係數、線性預估係數、倒頻譜參數、轉移倒頻譜參數等計算，最後是動態校準法，包含了計算以及搜尋最佳路徑之動作。

而本研究語音辨識功能之辨識成功率在八成左右，但隨著關鍵詞數目減少，成功率隨著提高。環境的雜音在關鍵詞數目較多的情況下，仍然影響辨識成功率。但在 PDA 上進行上述複雜的計算，較大的問題是在辨識速度。在 Celeron 1.0 Ghz, 256 MB SDRAM 的環境下經由 Pocket PC 2002 模擬器執行本研究之語音辨識部分，平均需耗時 5 至 7 秒，這是經由 eMbedded Visual Basic 內之計時器所計得的時間。在程式中設定中斷點以觀察各部分運算所耗費的時間，發現運算量最高的部分在將 Wave Form 檔案讀入記憶體中陣列的過程，佔了 3 至 4 秒。由於取樣頻率為 8000Hz，兩秒的錄音檔案便有 16000 筆資料，每筆資料佔兩位元組，因此程式在此進行一萬次以上十六進位轉換十進位的動作，是較耗時的地方。

計算自相關係數花了 2 至 3 秒，一個音框約有三千次的計算，語音經過端點偵測後，音框數目多在 40 至 60 個之間，也就是十萬次以上的計算，因此計算自相關係數也相當耗時。

#### b. 感知轉彎路口與設施

一般汽車導航在遇到轉彎路口時的處理多半只告訴駕駛人左轉或右轉，但在遇到較複雜的路口時，這樣子的資訊便不夠充分。本研究經由正切反函數計算出節線角度後，再計算與下一節線之角度差，並歸納為十二時方位，在轉向提示內容中告知駕駛人正確的轉彎角度。而除了

轉彎角度，亦可計算出車輛與轉彎路口的距離，以及下一節線之路名，使駕駛人對於轉向能掌握充分的資訊，降低駕駛人缺乏資訊所引發的緊張，也使駕駛人更專心地駕駛。

在路徑旁的設施包括飯店、加油站、停車場均能順利地顯示，並且以語音的方式播放。未來可以加入更多種類的設施提示，經由無線傳輸的方式，亦可提供多元的即時資訊。

#### c. 由文字檔模擬簡圖路徑檔案

本研究以文字檔案模擬由伺服器所下載的路徑座標檔案，以及 GPS 所傳送之車輛座標。並且可經由座標轉換將卡式直角座標系統之路徑顯示於 PDA 座標系統。地圖放大時重新計算所有座標點，包括停車場、加油站，與飯店，清除原畫面後再將重新計算結果作一次顯示。地圖平移時原理相同。

### 6.2 建議

#### a. 地圖顯示可建立索引

雖然簡圖顯示講求的是快速且直接地顯示導航路徑，但是在程式實作時可以感受到隨著節點數的增加，平移、放大、縮小等地圖操作速度將變得緩慢。但若能在讀入路徑座標的同時，將地圖劃分為如同九宮格般，並建立索引。當地圖放大時，在螢幕畫面外的節點便不需計算，待車輛移動至畫面邊界或者是使用者將地圖平移至邊界時再參考索引以顯示邊界外的圖。如此將使得計算量大幅減少，不需要每次進行地圖操作之動作時，都將每個點再算一次。

#### b. 加快語音辨識速度

本研究雖然達到語音功能，但其辨識速度仍然有待改進，若以解析度較低的 8-bits 檔案格式，在讀入語音資料時可以降低一半的運算量，雖然有可能使得辨識率降低，但若能有效提升辨識速度，亦是可以考慮的辦法。

而運算量也很高的部分是在自相關係數的計算，而在語音辨識領域中，特徵參數的種類不只是倒頻譜參數，另有經由傅利葉轉換與梅爾濾波器之梅爾頻譜參數。在 1965 年由 Cooley 及 Tukey 提出之快速傅利葉轉換法，原本 256 點取樣值的音框傅利葉轉換需要  $256 \times 256$  次計算，快速傅利葉轉換合併了許多重複的計算，使得計算次數降為  $256 \times \log_2 256$ ，但前提是取樣點必須為二的冪次方。與自相關係數相較，雖然快速傅利葉轉換在演算法上需要採用蝴蝶式（Butterfly）的交錯算法，在撰寫程式時顯得較為複雜，但運算量較少將使得辨識速度有些許的提升。

#### c. 語音提示的時機

在進行轉向提示與設施提示時，仍然有配合不夠周全的地方。例如有可能會發生設施正好位於轉向路口時，當系統提示完設施資訊後，其實車輛已經駛離轉向路口的情形。

這時候轉向與設施提示兩者間只能擇一提示，因此，可以設計供使用者選擇提示項目的功能，依使用者意願自行調整。

#### d. 提示中含破音字的處理

由於語音提示的發音是由比對字碼決定，而一個中文字可能有許多種發音。雖然在本研究的程式中，尚未遇到此破音字問題，但可以預見的是隨著接收的資料量提高，勢必會遭遇該問題。由於破音字與前後字詞有密切關係，可能的解決方法是建立破音字的前後字詞表，由此表決定正確的發音。或者是伺服器傳送路名屬性資料時，已將各種破音字加以註明，並標示正確發音，語音提示功能便可針對這些特例的發音作調整。

### 參考文獻：

1. 張晉維，「GIS 導航路徑簡化圖示在無線通訊端應用之研究」，國立交通大學，碩士論文，民國 90 年。
2. 林詠傑，「行動導航系統路徑簡圖顯示之研究」，國立交通大學，碩士論文，民國 91 年。
3. 李上銘，「語音辨認中基於主成份分析之進一步技術」，國立台灣大學，碩士論文，民國 90 年。
4. 李健平，「語音辨認應用於 PDA 之作業控制研究」，中原大學，碩士論文，民國 89 年。
5. 吳玉珍，王穆衡，何毓芬，「車用導航系統之初探」，交通部運輸研究所，民國 90 年七月。
6. <http://cyberg.wits.ac.za/cyberg/sessiondocs/symposia/age/age1/age1.htm>
7. <http://www.evl.uic.edu/aej/578/week3.html>
8. <http://www.umtri.umich.edu/library/acq/acq0142.html>
9. <http://thehfe.com/design.htm>
10. [http://www.unfallrekonstruktion.de/pdf/vuf\\_1996\\_reaction\\_english.pdf](http://www.unfallrekonstruktion.de/pdf/vuf_1996_reaction_english.pdf)
11. Jelinek. F, “Continuous Speech Recognition by Statistical Methods,” Proc. IEEE, Vol. 64, pp. 532-536, Apr. 1976.
12. 王漢蘭，隱藏式馬可夫模型應用於語音訊號辨識之研究，中原大學資訊工程所，碩士論文，民國 88 年。
13. 吳昭興，「應用 GPS/GIS 於汽車導航系統設計之研究」，國立台灣大學造船及海洋學工程研究所，碩士論文，民國 89 年。
14. Huang. X. D. and Lee. K. F., “On Speaker-Independent, Speaker-Dependent, and Speaker-Adaptive Speech Recognition,” IEEE Transactions on Speech and Audio Processing, Vol. 1, No. 2, pp. 150-157, Apr. 1993.
15. T. Mandel, “The Elements of User Interface Design” , 1997.
16. Shaughnessy. D. O’ , “Speech Communication: Human and Machine” , Addison Wesley, 1987.
17. Deller. J. R., “Discrete-Time Processing of Speech Signals” , Macmillan, 1993.
18. Oppenheim. A.V. and Schafer. R. W., “Discrete-Time Signal Processing” , Prentice Hall, 1989.