國 立 交 通 大 學

電信工程學系

碩士論文

台語語音辨識及智慧型口語對話汽車導航系統

Taiwanese Speech Recognition System and ITS Application

研 究 生：梁振豐

指導教授：王逸如　博士

中華民國九十五年八月

台語語音辨識及智慧型口語對話汽車導航系統

Taiwanese Speech Recognition System and ITS Application

研 究 生：梁振豐　　Student：Chen-Feng Liang

指導教授：王逸如 博士　　Advisor：Dr. Yih-Ru Wang

國立交通大學

電信工程學系碩士班

碩士論文

A Thesis

Submitted to Institute of Commmunication Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in Electrical Engineering

June 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年八月

台語語音辨識及智慧型口語對話汽車導航系統

研究生：梁振豐　　指導教授：王逸如 博士

國立交通大學電信工程學系碩士班

中文摘要

在本論文中建立了一個台語音節辨認器並且將之應用於一個實際應用系統-「慧型口語對話汽車導航系統」中。在論文中首先針對台語語音特性入聲調的部份深入研究，我們由 confusion matrix 去分析入聲調對辨識系統的影響性，且根據語言學上的知識修改文字拼音資料庫，使得訓練出來的聲學模型更為可靠，進而提升辨識率；並利用 Kullback Leibler(KL) distance 去觀察聲學模型在修改前與修改後之間的差異。同時論文中也將 sub-syllable bigram 語言模型(Language Model)加入辨識系統中，使的辨識率得以提升。在論文最後把台語語音辨識系統應用在-「慧型口語對話汽車導航系統」。

Taiwanese Speech Recognition System and ITS Application

Student：Chen-Feng Liang Advisor：Dr.Yih-Ru Wang

Institute of Communication Engineering

National Chiao Tung University

Abstract

In the thesis, a syllable recognition system for Taiwanese was established and applied to a real application - Intelligent Transportation System(ITS).

First , a syllable-based Taiwanese speech recognition system was implemented. And the effect on the recognition performance of the syllable with entering tone was carefully exiamined. Based on the linguistics knowledge, the syllable with entering tone in database was re-labeled according to the tone sandhi of Taiwanese in order to improve the recognition rate. In addition, Kullback Leibler distance between the acoustic model before and after re-labeling was examined to verify the linguistics knowledge. And the syllable bigram language model to the recognition system will obtain the rate improvement. Finally, the Taiwanese Speech Recognition System was applied to an Intelligent Transportation System(ITS)

致謝

由決定辭去工作準備研究所考試至今也經歷了將近 3 年半的時間，這段時間的經歷足夠我一生細細品味了。

感謝陳信宏老師與王逸如老師細心的指導，除了在研究方面的引導，讓我們能順利畢業；在工作態度與做事方法上，更是不停的耳提面命，讓我們了解到，方法可以加快事情的完成，態度卻是成就一切事物最為關鍵的因素。

感謝博班學長：志合、阿德、性獸，不斷的在研究上幫我們解惑；謝謝上屆學長姐：順哥、佩穎、Lubo、希群、榮勳、金翰，在剛進研究所時的教導；也感謝學弟們在最後一年衝刺中的陪伴;最重要的是與我這 2 年來一起經歷研究所這段時光的好同學：東毅、世帆、阿勇、國興、鴻彥、見偟以及世哲，沒有你們這 2 年就少了很多歡笑，少了撐下去的動力，很高興能認識你們，這段一起渡過的日子我想我是無法忘記了。

這段時間裡認識多年的朋友：陳欣德、謝岱清、胡耿豪、梁家玲、劉源宏、梁福聯，更是時刻陪伴我經歷風風雨雨，尤其欣德在我當初困頓時無條件提供我住宿，這份情我永遠不會忘的；二技同學：老詹、宜雯、JUJU、家寧、嘉琪、阿杰、給我很多幫助的巧雯、還有在天堂裡面與我一起同在的叔叔（孟儒），感謝陪伴我一起哭泣、一起歡笑、一起吶喊的朋友，我的人生因你們而豐富。

最感激的是我的家人的陪伴，這麼多年來你們對我的包容與愛護，當我在外面的挫折折磨的精疲力盡時，提供了一個地方幫我療傷，讓我得以喘息，補充繼續在外面闖蕩的能量。奶奶、爸、姐、弟，我愛你們勝過世上的一切。

目錄

中文摘要.....I
英文摘要.....II
致謝.....III
目錄.....IV
表目錄.....VI
圖目錄.....VIII
第一章 緒論.....1
1.1 研究動機.....1
1.2 研究方向.....1
1.3 章節概要.....2
第二章 台語語音特性.....3
2.1 台語結構與分類.....3
2.2 入聲韻母變調規則.....5
2.3 鼻化韻母特性與鼻音特徵的本質.....7
2.4 輕聲調的特性.....8
第三章 台語語音辨識.....10
3.1 簡介.....10
3.2 資料庫與現有資源介紹.....10
3.3 台語音節辨識系統基本架構.....11
3.4 特徵參數擷取.....12
3.5 聲學模型.....13
3.6 辨識網路 14
3.7 基本台語辨認系統之效能評計.....14

第四章 台語入聲調特性分析與改進方法.....18
4.1 由 confusion matrix 分析台語入聲調.....18
4.2 利用 Kullback Leibler(KL)2 distance 觀察入聲調.....25
4.3 添加語言模型至台語辨識器.....29
4.3.1 語言模型簡介.....29
4.4 辨識結果比較.....30
第五章 智慧型口語對話汽車導航系統.....32
5.1 系統簡介.....32
5.2 架設工具與系統內部功能簡介.....32
5.3 辨認模型的架設與訓練.....35
5.3.1 Subsyllable HMMs.....36
5.3.2 Background Model.....36
5.3.3 語法架構.....36
第六章 結論與未來展望.....41
參考文獻.....42

表目錄

表格 2.1 漢語音節內部結構.....3
表格 2.2 台語八聲例表.....4
表格 2.3 口元音與鼻化元音.....7
表格 2.4 口元音與鼻化元音的例句.....7
表格 3.1 語料的統計資料.....11
表格 3.2 特徵參數設定.....13
表格 3.3 Outside test 辨識率.....14
表格 3.4 Dictionary 校正後辨認率.....15
表格 3.5 音檔校正後辨認率.....16
表格 3.6 聲母辨認率.....16
表格 3.7 韻母辨認率.....16
表格 3.8 入聲韻母辨認率.....16
表格 4.1 入聲韻母與基本韻母之相互辨識關係一.....19
表格 4.2 入聲韻母與基本韻母之相互辨識關係二.....20
表格 4.3 入聲韻母與基本韻母之相互辨識關係三.....20
表格 4.4 入聲韻母與基本韻母之相互辨識關係四.....21
表格 4.5 入聲韻母與基本韻母之相互辨識關係五.....21
表格 4.6 入聲韻母與基本韻母之相互辨識關係六.....22
表格 4.7 入聲韻母與基本韻母之相互辨識關係七.....23
表格 4.8 入聲韻母與基本韻母之相互辨識關係八.....23
表格 4.9 修改促聲韻母後的辨識率.....25
表格 4.10 修改辨識網路後的辨識率.....25
表格 4.11 syllable database.....31

表格 4.12 Syllable unigram 辨識率.....31
表格 4.13 Syllable Bigram 辨識率.....31
表格 5.1 特徵參數設定.....36
表格 5.2 grammar 架構.....37
表格 5.3 19 speech acts in ITS Dialogue System.....38
表格 5.4 The 23 categories in ITS Dialogue System.....38

圖目錄

圖 2.1 台語八聲調之基頻軌跡.....4
圖 3.1 語音音節辨識基本架構.....11
圖 3.2 辨識網路.....14
圖 4.1 train data 修改情形.....24
圖 4.2 Outside test data 修改情形.....24
圖 4.3 未根據變調規則修改之前 HMM 之間 confusion 的情況.....28
圖 4.4 根據變調規則修改之後 HMM 之間 confusion 的情況.....29
圖 4.5 Back-off bigram Word-Loop Network.....30
圖 5.1 口語汽車導航系統方塊圖.....32
圖 5.2 由 ATK 架設的基本即時辨認器.....33
圖 5.3 Galaxy Communicator software 內部結構圖.....34
圖 5.4 系統運作實例.....39

第一章 緒論

1.1.研究動機

現代科技產品中主要的操作介面，大部分建立在需要人體接觸才可操作，而語音辨識技術建立人與機器之間無接觸的操作方式。

在台灣這片土地上，除了國語之外，普遍被使用的語言就是台語。用台語念古詩相對國語而言，來的有意境。在文字方面，雖然台語並無一套真正屬於自己的文字，但卻可藉由漢字與拼音的組合來表達台語真正的意思，例如台語的 “sian3”和 “sam3”在國語方面卻只能用 “打”來替代，但是兩者意義卻不同， “sian3”是屬於力道較小的打而 “sam3”是屬於力道較大的打[1]，這也是台語獨特之處，希望電腦也能夠聽的懂這種獨特的語言。

1.2.研究方向

本論文包含了台語語音辨識以及將台語語音辨識器應用於智慧型口語對話汽車導航系統(Intelligent Transportation System，ITS)兩個部分，在語音辨識方面，之前本實驗室畢業的王文德學長，有做過台語語音辨識初步的研究，本篇論文承接之前的成果，我們建立屬於台語本身的聲學模型(acoustic model)並利用此聲學模型來進行單音節的辨識，深入探討台語語音特性，加入新的辨識網路，以期能提升台語語音辨識率。

台語語音辨識器應用於 ITS 部分，建立所需的聲學模型，以及根據台語的語法建構出台語適用的語法結構。

1.3. 章節概要

第一章　介紹研究動機，方向以及章節概要。

第二章　介紹台語語音特性

第三章　介紹台語語音辨識的基本架構原理，與初步辨識率

第四章　辨識率提升方法與實驗

第五章　台語智慧型口語對話汽車導航系統相關介紹

第六章　結論與未來展望

第二章台語語音特性

2.1 台語結構與分類

台語和國語一樣都是聲調語言，音節由聲母、韻母和聲調(tone)所組成。所謂聲母與韻母，都是傳統中國聲韻學上的名詞，前者指音節內的第一個輔音，而韻母則包括介音及韻步，韻步又包括主要元音及韻尾[2]。易言之，漢語音節的內部結構如表格 2.1 所示。

表格 2.1 漢語音節內部結構

<table>
<tr><td colspan="4">聲調</td></tr>
<tr><td rowspan="3">聲
母</td><td colspan="3">韻母</td></tr>
<tr><td rowspan="2">介
音</td><td colspan="2">韻步</td></tr>
<tr><td>主
要
元
音</td><td>韻尾</td></tr>
<tr><td></td><td></td><td>元音/輔音</td></tr>
</table>

台語因無標準文字，音節數目也多有爭議，但本論文中所使用的是漢羅拼音，此外台語的基本音節為 877 個，較國語多，其列表詳見附件一；另外，台語的聲調共為八種，其中二、六聲已合併而只剩七種，也較國語的聲調多，各聲調之特徵及例字如表格 2.2 所示，其典型基頻軌跡(pitch contour)如圖 2.1 所示。

表格 2.2 台語八聲例表

聲調	台文字	羅馬拼音
一聲(陰平)	衫	saN
二聲(陰上)	短	te2
三聲(陰去)	褲	kho3
四聲(陰入)	闊	khoah
五聲(陽平)	人	lang5
六聲(陽上)	矮	e2
七聲(陽去)	鼻	phiN7
八聲(陽入)	直	tit8

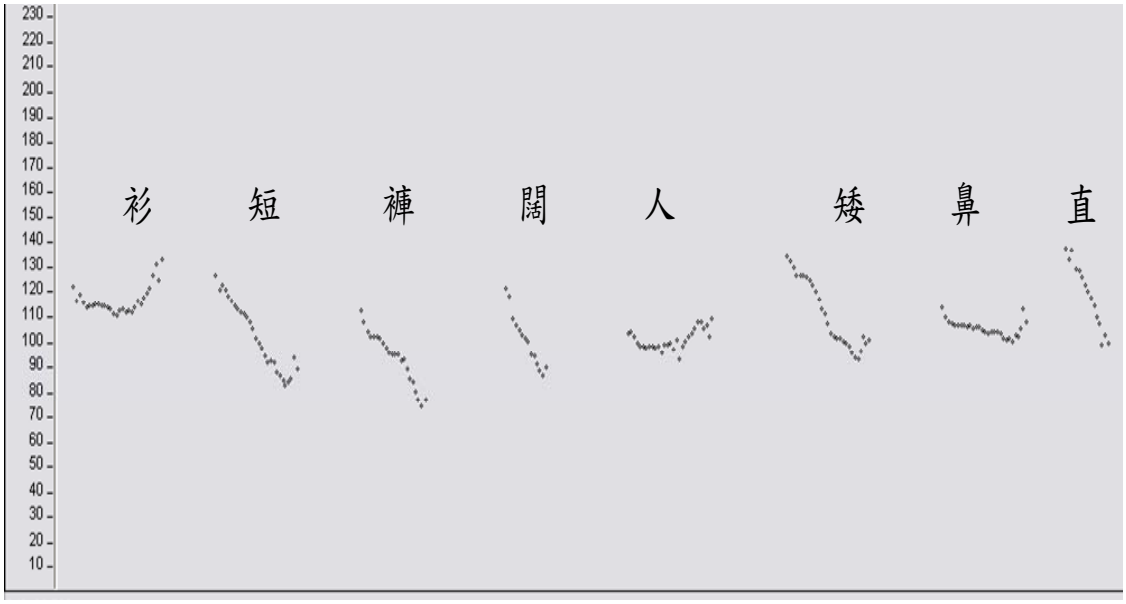


圖 2.1 台語八聲調之基頻軌跡

台語之聲母包含空聲母共 18 類，依其發音特點可分五類：(1) 脣音：p, ph, b, m，(2) 舌尖非齒音：t, th, l, n，(3) 舌根音：k, kh, g, ng，(4)舌尖齒音：ch, chh, s, j，(5)喉音：0(空聲母), h, 。

台語之韻母包含空韻母共 84 類如付件二所示，其中又可分為基本韻母、入聲韻母、鼻化韻母與入聲鼻化韻母。基本韻母泛指不屬於其他三類之韻母，入聲韻母其符號特性是以 p，t，k，h 結尾，鼻化韻母其符號特性是以 N 結尾，而入聲鼻化韻母其符號特性為 hN 結尾之韻母。而入聲韻母及鼻化入聲韻母也是國語中所沒有的。

2.2 入聲韻母變調規則

在本論文中所製作的台語語音辨認器雖不做音調辨認，但台語的入聲韻母之變調規則中，不但聲調會受前文之影響而改變，有時其基本音節也會受到改變。所以我們在此將討論那些會改變基本音節之變調規則。

一般我們將促聲韻母(h 收尾)及入聲韻母(p，t，k 收尾)統稱為入聲韻母，雖然都是喉化的短音，不過在變調規則中有所不同，例如：

A. 單獨： chioh8 ‘借’ chio3 ‘照’

B. 在其他音節前：chio3 kiaN3 ‘借鏡子’ chio3 kiaN3 ‘照鏡子’

上述例句中如果獨立念時，由於聲調不同而音質有異。其中，chio3 ‘照’ 由於是舒聲，它的調可以念的很長，而且讓人完全聽的到它的音值。但是 chioh ‘借’ 由於是入聲其調值會變的急促而短，這是因為緊喉的關係。但是，這兩個字如在其他音節之前，就變的完全一樣，這正好表示 chioh ‘借’的入聲已經變成舒聲 chio，因此會與 chio ‘照’念的完全相同，也就是一般而言，促聲韻母在其他音節之前會掉落，掉落的促聲韻母自然使聲調變回舒聲。但是入聲韻母連音的時候沒有這種變化，p，t，k 韻尾仍然存在。

2.2.1 入聲韻母變調特例

(1)入聲韻母下一音節接後綴詞

一般而言，形容詞後綴-e 沒有它自己的聲調，它表音上的聲調其實來自於它前面的音節，如下例子:

A. tua7 e5　“大的”

se3 e5　“小的”

B. thih8 e5　“鐵的”

eh8 e5　“窄的”

值得注意的是B中的例子，其第一音節都沒有刪除喉塞音h。這跟喉塞音在其他音節之前會掉落的規則相牴觸。喉塞音是否在其他後綴之前也會保存呢?如下例子:

A. kim1 a2　“金子”

kau5 a2　“猴子”

B. thih8 a2 → 變調成 thi3 a2　“鐵”

ioh8 a2 → 變調成 io3 a2　“藥”

B中的例子顯示名詞後綴-a有其本身的聲調，而使其前的音節產生變調，最後導致喉塞音的脫落。

在拼音表示法中，我們發現只要是後接的音節沒有聲調(即表音上的聲調其實是來自它前面的音節)，則不論有沒有聲母，其前的喉塞音都不會脫落。

(2)入聲韻母後一音節接輕聲調

輕聲調的特性在2.5章節會闡述，這邊我們討論入聲調後一音節接輕聲調時，對入聲調的影響。輕聲以”。”表示，如下例句：

A. kah8 khi。　“跟某人或物一起送去”

B. ah8 lai。　“押來”

台語的lai “來”和khi “去”，如接在其他動詞之後，往往變成輕聲。這種情形下，之前的音節不會變調。如果之前正好是入聲調，其喉塞音不會脫落。

(3)位於聲調群最後一個音節

入聲變調與其他變調一樣，因句法結構而有不同的聲調群，如下例句：

A. pe8 ah4 ia7 ho2(白鴨也好)

B. pe8 ah4 pi2 ou2 ah4 a2 ho2(白鴨比黑鴨還好)

我們可以看到所有的喉塞音 ah4 的喉塞音都完好如初，儘管之後都有其他的音節，卻不見有任何的喉塞音遭刪除。這是因為白鴨（名詞詞組 NP）構成一個聲調群，ah4 不變調，因為它是聲調群的最後一個音節。由於 ah4 保持了入聲，因而有喉塞音。

總結這部分的討論，我們得到幾個通則：(a)導致變調的是後一音節的聲調，而非音節本身；(b)並非由於喉塞音的刪除而引起聲調的改變，而是由於聲調的改變導致喉塞音無法出現。

2.3 鼻化韻母特性與鼻音特徵的本質

台語有六個口元音和四個鼻化元音如表格 2.3 所示，而且這兩種原因有辨義的功能，由表格 2.4 中的例子可知同樣的元音會因鼻音之有無而有不同的意思：

表格 2.3 口元音與鼻化元音

口元音	i	e	a	ou	o	u
鼻化元音	iN	eN	aN	ouN		

表格 2.4 口元音與鼻化元音的例句

口元音	詞義	鼻化元音	詞義
i7	玩	iN7	院
pi2	比	piN2	扁
sa1	拿	saN1	嬰兒
pe7	爸爸	peN7	病
ta2	焦	taN2	擔

在台語的鼻化有四個通則；(1) 在開音節及閉音節中，鼻化的展延方向有矛盾。在開音節裡，左向展延是必要的，但在閉音節裡，左向展延卻絕對不可以；

(2) 元音之後的元音韻尾與輔音韻尾表現大不相同:輔音韻尾不能與聲母同為鼻音,但是元音韻尾卻必須與聲母一致;(3) 鼻音韻尾之前的元音不鼻化;(4) 鼻音的展延方向可以是向左也可以是向右。

而鼻化元音與鼻輔音的[鼻音]應是相同的,均屬於詞素,而且是漂浮性的特徵。[鼻音]在特徵樹上附著在自然發聲(SV)點上。

2.4 輕聲調的特性

在台語中聲調扮演著相當重要的辨詞功能,而輕聲調在台語裡是較特殊的聲調,底下我們將介紹輕聲的特性[3]。輕聲的發音與第三聲很類似,主要是靠前後音節間的長短來分辨。而輕聲主要在多語詞之間才有辨語的功能,因此不將之歸於第三聲。輕聲化規律條件有二:(1) 能輕聲化的音節只出現在主要語法範疇動詞詞組(VP)、名詞詞組(NP)、語氣句尾詞(S)等的末尾;(2) 關連到代名詞或數量語時,只有在無語意重點時才可輕聲,因此延伸了以下的功能:

1. 標誌出主要句法單位 NP,VP,S 的界線

例如:

a. 走= 出來　　VP　　跑出來

b. 走出來外口　　VP　　跑到外頭來

其中 a 句的出來為輕聲,因是在 VP 末尾。

2. 標誌語意重點

例如:

a. 讀= 兩本　(重點在動詞)

b. 讀兩本　　(重點在數量語)

3. 分辨語詞

例如:

a. 後= 日　　後天

b. 後日　　改天

4. 標誌虛詞所在

台語雖然不是所有的虛詞都輕聲，但是所有輕聲出現的情形都是NP，VP，S末尾的虛詞。這些虛詞可能是詞尾、詞組尾、或是句尾。

第三章台語語音辨識

3.1 簡介

目前國內台語語音辨認系統已有長庚大學呂仁園教授的不特定語者大辭彙華台雙語辨認系統，在一萬詞時其辨識率為 73.50%而七萬詞時其辨識率為 58%[4]。這章節將介紹台語語音基本特性、基本語音辨識流程與原理，以及將原有台語辨識器加以擴充[5]。

3.2 資料庫與現有資源介紹

目前我們擁有的台語語料庫有：

3.2.1 Database1：含男女兩人以自然流麗朗讀之大型 database，所錄製之內容是為故事類，最長段落其字數約為 250 個字，原來做為訓練 TTS 之用，其錄音之 sampling rate 為 20kHz，我們將其 down sampling 為 16kHz，以配合其他語料。

3.2.2 Database2：由 99 人所錄製之語料，其 sampling rate 為 16kHz，在 99 人的語料中有些人讀相同的文章，相異文章的數目只有 57 篇。

3.2.3 Database3：由男十人及女八人錄音，取樣頻率為 16kHz，文字取自”台語通用會話：一套讓你能說出台語美辭雅語的教材，第二級[6]”及”一分鐘台語單字速成[7]”。

3.2.4 Database4：重新錄製 database2 的文章，男三十四人及女四十五人。

語料整理後用於訓練，以及測試的字數結果列於表格 3.1。

表格 3.1 語料的統計資料

	音檔格式	錄製人數	訓練字數	測試字數
男	無檔頭音檔 (pcm),取樣率為 16kHz	91	44889(6.13hr)	4921(0.69hr)
女		106	60798(8.42hr)	7290(1.007hr)
總數		197	105687 (14.55hr)	12211(1.697hr)

台語的基本音節有 877 個，而在 877 個基本音節裡，台語語料中只涵蓋了 685 個基本音節，其中有 101 個聲母，79 個韻母(其中有 29 個入聲韻母)。

3.3 台語音節辨識系統基本架構

先前的台語音節辨識系統的基本架構[8]，它的方塊圖見圖 3.1，主要包含三個部份：(1) 特徵參數擷取；(2) 聲學模型訓練；(3) 辨識比對。

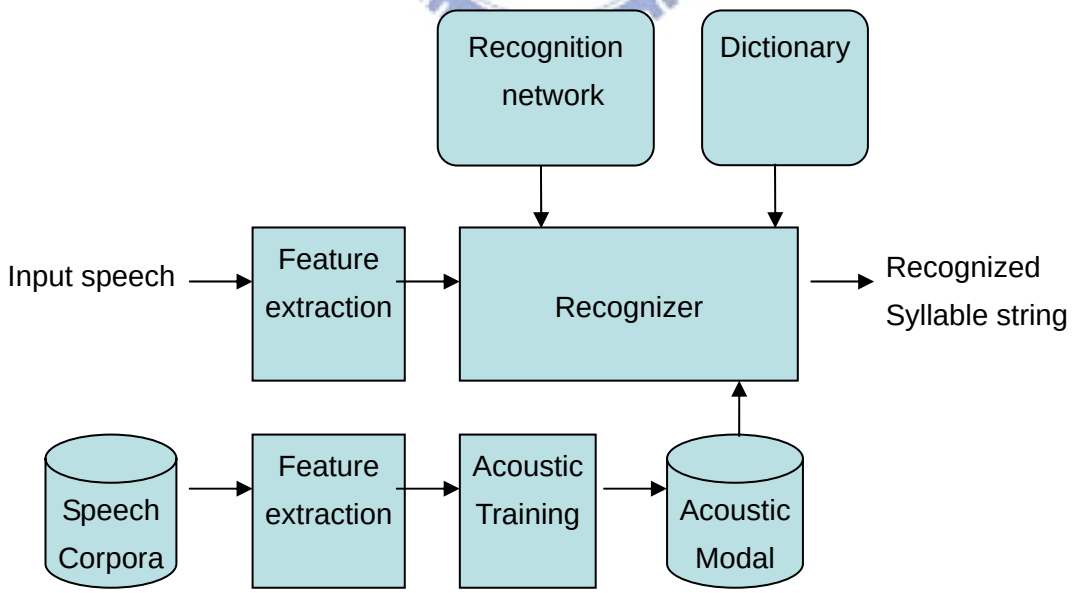


圖 3.1 語音音節辨識基本架構

各個方塊之功能說明如下：

(1) Feature extraction ：對音訊做處哩，求出能表示此語音的特徵訊息，以作為語音訓練及辨識的參數。

(2) Training ：利用特徵參數與隱藏式馬可夫模型(Hidden Markov Model，HMM)，求出一組聲學模型。

(3) Dictionary ：此檔案記錄了聲學模型與基本音節的對應關係。在辨識時可經由查詢此檔案的動作，來將辨識後的聲學模型符號，轉成基本音節的拼音符號。

(4) Recognition network ：作為辨識時所依據的搜尋網路，先前的辨識網路並無加任何限制。

接著，將簡述各方塊之系統參數。

3.4 特徵參數擷取

將原始的語音訊號經數位化後，雖然可直接拿來做辨識之用，不過由於資料量過於龐大，因而造成處理速度的緩慢，而特徵參數擷取是對音訊做處理，求出能表示此語音的特徵訊息，再利用所擷取出的特徵向量序列(feature vector sequence)，配合 HMM，來訓練聲學模型，作為語音辨識的參考模型。在此我們採取語音界廣泛使用的，梅爾刻度之倒頻譜系數(Mel-Frequency Cepstral Coefficients，MFCC)，經統計錄音者的說話速度為 3~~4 字／秒因而將 Delta window size 與 Delta delta windows size 設為 7，在這邊我們將取 38 維 MFCC 參數當特徵參數，相關之設定見表格 3.2。

表格 3.2 特徵參數設定

取樣頻率	16 kHz
預強濾波器	$1-0.97Z^{-1}$
視窗形式	Hamming window
音框長度(Frame size)	32ms
音框平移(Frame shift)	10ms
Filter bank	22
Feature vector	MFCC_E_D_A_N_Z
Delta window size	7
Delta delta window size	7

其中符號 MFCC_E_D_A_N_Z 之意義為 12 維 MFCC,13 維 Delta MFCC 和 13 維 Delta delta MFCC，並且做 Cepstral Mean Normalization。

3.5 聲學模型

將台語基本音節分解成聲母右相關韻母(101 類)和韻母(84 類)(如附件二表示)，並利用 HMM 來建立台語的聲學模型，因此我們所需要訓練的模型個數共為 185 個。聲母右相關韻母和 silence model 我們使用 3 個 states 的 HMM model 來建立，韻母則用 5 個 states 的 HMM model 來建立，並且每個 state 都升到 64 個 mixtures。Short pause(sp)是 syllable 與 syllable 之間的 silence,其 model 是一個 state 的 HMM model，採用與 silence model 中間的 state 共用。

在訓練的過程中,在本論文裡我們採用 HTK[9]裡所提及的方式 *flat start*，其方法是認為開始時所有的 HMM model 參數皆令為一樣，再利用 Baum-Welch 的方式更新各個 HMM model 的參數，雖然此法會花費很多的時間，但我們所擁有的語料並不多因此不用花費太多的時間即可得到不錯的模型,其 *initial model* 建立過程如下：

以類似 uniform segmentation 的方式對訓練語進行切割並訓練出一個 global HMM model 其 mean 和 covariance 等於 global speech mean 和 covariance，並指定每一個 HMM model 皆為此 global HMM model。

3.6 辨識網路

一開始我們採用的辨識網路並沒有加任何的限制，即為任何音節皆可接任何音節，圖 3.2 為示意圖。

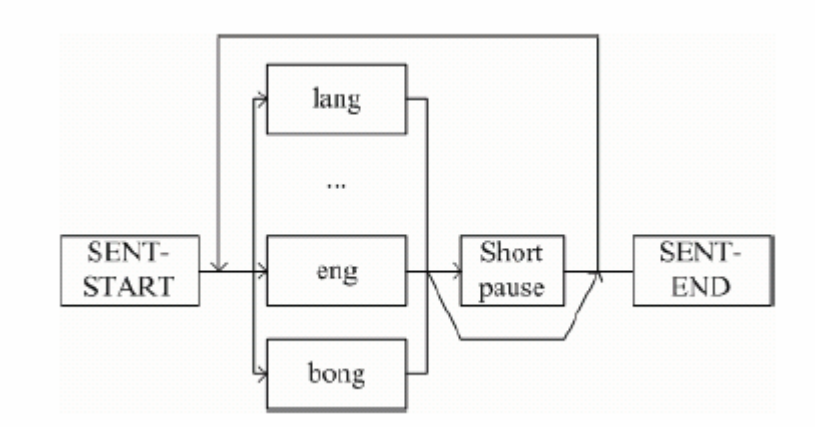


圖 3.2 辨識網路

3.7 基本台語辨認系統之效能評計

由之前王文德學長[5]所留下的結果，訓練語料是 Database1、Database2、Database3 的語料庫，外部測試(Outside test)辨識率如表格 3.3。

表格 3.3 Outside test 辨識率

	音節總數(N)	Sub(%)	Del(%)	Ins(%)	Recognition rate(%)
男	900	52.0	3	2.3	42.67
女	900	49.4	1.8	1.4	47.33

Sub：number of substitutions;

Del：number of deletions;

Ins： number of insertions;

N：total number of labels in the defining transcription file

其中　$\text{Recong rate}=\frac{\text{N-(Sub+Del+Ins)}}{\text{N}}\times 100\%$

接著檢查所有的 Database 後發現，原本的 dictionary 結構，有些分類不理想將其加以修正，修正如下：

1　將原本編為第 6 類的韻母 ou、ouN、ouh、ouhN，併入第 5 類

2　將原本屬於第 8 類的韻母 ng、nhg，以及第 9 類韻母 m，mh，併入第 1 類

3　原本的韻母分為 9 類，改變後變 6 類

4　Sub-syllables 的數目，由 206 個(含 SP)縮減為 187 個

並擴大訓練及辨認語料後，即訓練語料總共有 105687 個 syllable，且總共有 12211 個 syllable 來當作 outside test 的辨識語料，表格 3.4 為 Dictionary 校正後之辨認率。

表格 3.4 Dictionary 校正後辨認率

	音節總數 (N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	48.2	2.9	2.8	46.1

在資料庫介紹中，Database2 的語料所錄的音檔中，每一個音檔一般都超過 50 個 syllables，有些甚至於超過 100 個以上的 syllable。由於我們訓練聲學模型是採用的方式是 flat start，一個音檔若是過長，可能會影響到聲學模型的正確性。在此我們將 database2 的音檔用人工的方式下去檢查，並將音檔分段成好幾個小段音檔，且每個音檔中包含的 syllable 數目控制在 20 以下。在將分段後的音檔重新下去訓練聲學模型，在利用出來的聲學模型，以同樣的測試語料跑辨認率，得到的辨認結果列在表格 3.5。

表格 3.5 音檔校正後辨認率

	音節總數(N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	48.2	2.7	2.7	46.4

我們進一步分析聲母與韻母的辨識率，聲母的辨識率如表格 3.6，韻母的辨識率如表 3.7。

表格 3.6 聲母辨認率

	音節總數(N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	31.9	2.7	2.7	62.7

表格 3.7 韻母辨認率

	音節總數(N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	37.2	2.7	2.7	57.4

我們可以發現，在台語中聲母的辨認率高於韻母的辨認率，而在國語系統中卻是韻母的辨認率高於聲母辨認率，所以我們進一步的統計非入聲韻母與入聲韻母辨認率如表 3.8。

表格 3.8 非入聲韻母與入聲韻母辨認率

	音節總數(N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
非入聲韻母辨認率	10202	34.3	2.7	2.4	60.6
入聲韻母辨認率	2009	51.5	3.2	3.8	41.5

發現入聲韻母的辨認率比非入聲韻母的辨認率相差 19.1%，入聲韻母的辨認率影響到整個系統的辨認率，下一章節將更深入分析台語入聲韻母語音特性，更在辨識網路上作些調整與改變。

第四章台語入聲調特性分析與改進方法

在台語辨識系統跟國語辨識系統最大的不同在於，台語辨識系統中辨認單元中含有聲調部分，也就是台語中的入聲調有加入辨認單元中，使的變調的問題會影響到辨識率。接下來我們分析入聲調對於辨識系統的影響。

4.1 由 confusion matrix 分析台語入聲調(Entering tone)

在之前 2.2 章節入聲韻母特性中提到，促聲韻母(h)在其他音節之前會掉落，掉落的促聲韻母自然使聲調變回舒聲。以下我們根據 confusion matrix 分析，觀察是否有此現象發生。

由觀察 confusion matrix 中發現 ，{ai,aih}，{au,auh}，{ui,uih}，{oe,oeh}，{iu,iuh}，{iau,iauh}無論 inside test 或 outside test 完全都沒有 confusion 的情況發生，也就是說 ai 不會變是成 aih，aih 也不會辨識成 ai。但這六組由於促聲韻母出現次數不多(aih 出現 18 次, auh 出現 4 次, uih 出現 18 次, oeh 出現 53 次, iuh 出現 18 次, iauh 出現 18 次)，使的訓練出來的聲學模型不是很可靠，而無法確定是否與語言學上說的有所矛盾。

接著我們將 entering tone 分為 8 組，分別為{a,ah,ak,ap,at }，{e,eh,ek}，{i,ih,ip,it}，{ia,iah,iak,iap,iat}，{io,ioh,iok}，{o,oh,ok,op,ou}，{oa,oah,oat}，{u,uh,ut}並觀察其相互辨識關係。

我們在表格中 O.D:表示分析 Outside data 得到的結果， I.D:表示分析 Inside data 得到的結果，左邊表示我們欲辨認的單元，上方表示辨識為何種單元。

在表格 4.1 中，我們可以看到比較容易 confusion 的情況是{a,ah}，只討論 substitution 的情況下，我們可以發現 subsyllable {a} 在 inside test 中，發生 substitution 的總數為 1656 次，而被 subsyllable{ah}取代的次數為 550 次，如此一來，我們可以發現 subsyllable{a}有被取代的情形發生時，有

33.3%是被 subsyllable{ah}所取代。 Outside test 情況中有 33.1%是被 subsyllable{ah}所取代。

而 subsyllable {ah}在有被取代的情形發生時，在 inside test 中，有 53.8%被 subsyllable {a}所取代，outside test 中，有 42.4%被 subsyllable {a}所取代。

表格 4.1 入聲韻母與基本韻母之相互辨識關係一

Ref	a		ah		ak		ap		at		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
a	479	4828	166	550	46	111	0	4	8	20	282	971
ah	78	422	284	2045	12	70	0	5	2	13	92	316
ak	16	13	16	6	22	676	0	0	3	2	13	38
ap	1	0	1	0	1	0	1	259	0	0	11	3
at	8	8	1	1	6	0	0	0	1	345	9	9

表格 4.2 中，可以看到比較容易 confusion 的情況是{e,eh}，subsyllable {e}在有被取代的情形發生時，在 inside test 中，有 41%被 subsyllable {eh}所取代，outside test 中，有 32.1%被 subsyllable {eh}所取代。

而 subsyllable {eh}在有被取代的情形發生時，在 inside test 中，有 82.1%被 subsyllable {e}所取代，outside test 中，有 75.7%被 subsyllable {e}所取代。

表格 4.2 入聲韻母與基本韻母之相互辨識關係二

Ref	e		eh		ek		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
e	894	8767	78	448	1	7	164	638
eh	88	266	45	1111	0	0	22	58
ek	8	5	0	0	1	301	29	18

表格 4.3 中，可以看到比較容易 confusion 的情況是{i, it}，subsyllable {i}在有被取代的情形發生時，在 inside test 中，有 66.8%被 subsyllable {it}所取代，outside test 中，有 57%被 subsyllable {it}所取代。

而 subsyllable {it}在有被取代的情形發生時，在 inside test 中，有 67.1%被 subsyllable {i}所取代，outside test 中，有 51%被 subsyllable {i}所取代。

表格 4.3 入聲韻母與基本韻母之相互辨識關係三

Ref	i		ih		ip		it		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
i	832	8598	2	10	0	2	183	1248	136	608
ih	5	5	3	302	0	1	5	3	9	4
ip	10	1	0	0	0	201	13	1	2	0
it	86	438	0	0	0	2	86	3697	83	213

表格 4.4 中，可以看到比較容易 confusion 的情況是{ia, iah}，subsyllable {ia}在有被取代的情形發生時，在 inside test 中，有 22.9%被 subsyllable {iah}所取代，outside test 中，有 21.3%被 subsyllable {iah}所取代。

而 subsyllable {iah}在有被取代的情形發生時，在 inside test 中，有 30.1%被 subsyllable {ia}所取代，outside test 中，有 35.6%被 subsyllable {ia}所取代。

表格 4.4 入聲韻母與基本韻母之相互辨識關係四

Ref	ia		iah		iak		iap		iat		Other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
ia	70	1444	22	74	0	0	0	0	0	0	81	249
iah	32	34	86	1077	0	0	0	0	1	2	57	77
iak	1	0	2	0	0	60	0	0	0	0	0	0
iap	1	0	5	1	0	0	0	176	0	0	4	0
iat	0	0	0	0	0	0	0	0	1	248	18	13

表格 4.5 中，可以看到比較容易 confusion 的情況是{io, ioh}，subsyllable {io}在有被取代的情形發生時，在 inside test 中，有 17.9%被 subsyllable {ioh}所取代，outside test 中，有 39%被 subsyllable {ioh}所取代。

而 subsyllable {ioh}在有被取代的情形發生時，在 inside test 中，有 36.5%被 subsyllable {io}所取代，outside test 中，有 48.6%被 subsyllable {io}所取代。

表格 4.5 入聲韻母與基本韻母之相互辨識關係五

Ref	io		ioh		iok		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
io	86	1265	44	28	0	6	69	122
ioh	34	57	52	1105	0	2	36	97
iok	3	2	0	1	8	368	12	9

這組跟前面幾組比較不一樣，比較特別的地方在於 subsyllable{o,ou}之間 confusion 的情形跟 subsyllable{o,oh,ok,op}基本韻母與入聲韻母之間 confusion 的情形更為嚴重，如表格 4.6 所示。

Subsyllable{o}在有被取代的情形發生時，在 inside test 中，有 6.21% 被 subsyllable {oh}所取代，卻有 48.7%被 subsyllable{ou}所取代，outside test 中，只有 4.5%被 subsyllable {oh}所取代，卻有 41.5%被 subsyllable{ou}所取代。

Subsyllable{oh}在有被取代的情形發生時，在 inside test 中，有 16.1% 被 subsyllable {o}所取代，有 27.9%被 subsyllable{ou}所取代，outside test 中，只有 34.9%被 subsyllable {o}所取代，有 18.6%被 subsyllable{ou}所取代。

Subsyllable{ou}在有被取代的情形發生時，在 inside test 中，有 45.0% 被 subsyllable {o}所取代，outside test 中，有 33.9%被 subsyllable {o}所取代。

表格 4.6 入聲韻母與基本韻母之相互辨識關係六

Ref	o		oh		ok		op		ou		Other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
o	90	2335	7	46	1	14	0	0	64	361	82	320
oh	20	19	7	730	1	3	0	0	11	33	27	63
ok	3	8	0	2	3	468	0	0	8	6	16	19
op	0	0	0	0	0	0	0	18	0	0	0	0
ou	136	482	23	66	9	38	0	0	332	3997	145	484

表格 4.7 中，基本韻母與入聲韻母之間 confusion 的情形沒那麼明顯。

表格 4.7 入聲韻母與基本韻母之相互辨識關係七

Ref	oa		oah		oat		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
oa	191	2137	3	2	2	6	61	181
oah	27	2	2	385	0	2	11	10
oat	11	2	0	0	6	337	4	11

表格 4.8 中，基本韻母與入聲韻母之間 confusion 的情形也沒有很明顯。

表格 4.8 入聲韻母與基本韻母之相互辨識關係八

Ref	u		uh		ut		other	
	O.D	I.D	O.D	I.D	O.D	I.D	O.D	I.D
u	295	2997	0	0	5	34	87	176
uh	0	1	0	5	0	0	0	0
ut	11	11	0	0	28	630	21	12

綜合以上結果，我們可以清楚的知道容易產生 confusion 的組別有{a,ah}、{e,eh}、{i,it}、{ia,iah}、{io,ioh}、{o,ou}。與語言學中提到的促聲韻母(h)在其他音節之前會掉落，掉落的促聲韻母使聲調變回舒聲互相驗證。

接著我們把訓練語料與測試語料文字檔作修改，根據語言上的變調規則，將需要變調的促聲韻母都改成基本韻母的音節，train data 總共修改了 6147 個 subsyllable，各 subsyllable 修改的比例如圖 4.1，Outside test data 總共修改了 702 個 subsyllable，各 subsyllable 修改的比例如圖 4.2。

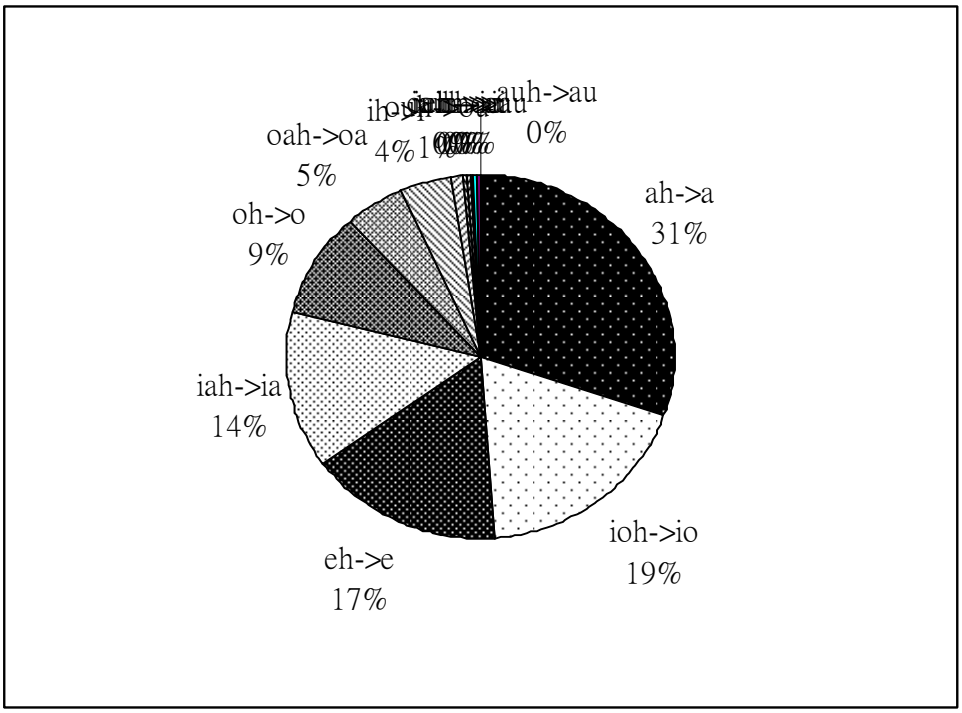


圖 4.1 train data 修改情形

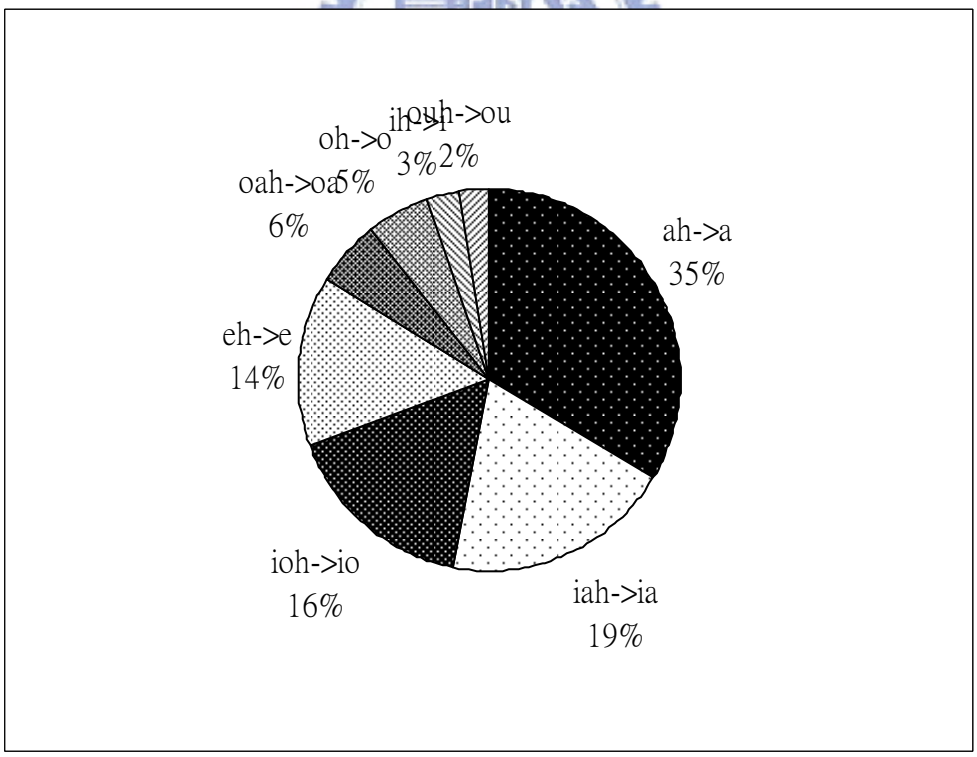


圖 4.2 Outside test data 修改情形

再次訓練聲學模型，得到辨識率列在表格 4.9 中。

表格 4.9 修改促聲韻母後的辨識率

	音節總數 (N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	45.5	3.0	2.2	49.3

接著我們根據變調規則，修改台語辨識網路，也就是讓有促聲韻母(h)的音節只允許出現在句尾。修改辨識網路之後辨識率如表 4.10。

表格 4.10 修改辨識網路後的辨識率

	音節總數 (N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	44.3	2.9	2.6	50.2

辨識率提升了 3%，我們可以發現台語入聲調變調在辨識系統中，對於辨識率的影響，是一個很大的問題點。

4.2 利用Kullback Leibler(ＫＬ)2 distance觀察入聲調 HMM confusion情形

KL distance[10]主要用來觀察兩個相異的機率分佈，你估計機率分佈P1(x)與另一實際機率分佈P2(x)之間的相似度。P1(x),P2(x)為 Gaussian distribution，且P1(x)的mean與variance分別為 μ_1 、 σ_1^2 ，P2(x)的mean與variance分別為 μ_2 、 σ_2^2 。

從P1(x)到P2(x)的KL distance KL(P1,P2)定義為

$$KL(P1,P2)=\int P1(x)\cdot\log\frac{P1(x)}{P2(x)}dx \tag{4.1}$$

$$=\frac{1}{\sqrt{2\pi\sigma_1^2}}\int e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\log(\frac{P1(x)}{P2(x)})\cdot dx$$

$$\log\frac{P1(x)}{P2(x)}=\log(\frac{\sigma_2}{\sigma_1})+[-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}] \tag{4.2}$$

所以D(P1,P2)可寫成

$$KL(P1,P2)=\frac{1}{\sqrt{2\pi\sigma_1^2}}\int e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot[\log\frac{\sigma_2}{\sigma_1}-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}]\cdot dx \tag{4.3}$$

我們將D(P1,P2)分成3項

令 $I_1=\frac{1}{\sqrt{2\pi\sigma_1^2}}\int e^{-\frac{(x-\mu1)^2}{2\sigma_1^2}}\cdot\log(\frac{\sigma_2}{\sigma_1})\cdot dx=\log(\frac{\sigma_2}{\sigma_1})$ (4.4)

$$I_2=-\frac{1}{\sqrt{2\pi\sigma_1^2}}\int\frac{(x-\mu1)^2}{2\sigma_1^2}\cdot e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot dx \tag{4.5}$$

$$=-E[\frac{(x-\mu_1)^2}{2\sigma_1^2}]=-\frac{1}{2\sigma_1^2}\cdot E[x^2-2\mu_1x+\mu_1^2]$$

$$=-\frac{1}{2\sigma_1^2}\cdot[E[x^2]-2\mu_1^2+\mu_1^2]$$

$$=-\frac{1}{2\sigma_1^2}\cdot[\sigma_1^2+\mu_1^2-\mu_1^2]=-\frac{1}{2} \tag{4.6}$$

$$I_3=\frac{1}{\sqrt{2\pi\sigma_1^2}}\int\frac{(x-\mu_2)^2}{2\sigma_2^2}\cdot e^{(-\frac{(x-\mu_1)^2}{2\sigma_1^2})}\cdot dx \tag{4.7}$$

我們分解$(x-\mu_2)^2=(x-\mu_1+\mu_1+\mu_2)^2$

$$=[(x-\mu_1)+(\mu_1-\mu_2)]^2$$

$$=(x-\mu_1)^2+2(x-\mu_1)(\mu_1-\mu_2)+(\mu_1-\mu_2)^2 \tag{4.8}$$

在將I_3分成I_{31}、I_{32}、I_{33}3個部份分別做運算

$$I_{31} = \frac{1}{\sqrt{2\pi\sigma_1^2}} \int \frac{(x-\mu_1)^2}{2\sigma_2^2} \cdot e^{(-\frac{(x-\mu_1)^2}{2\sigma_1^2})} \cdot dx$$

$$= \frac{1}{2\sigma_2^2} \cdot E[(x-\mu_1)^2] = \frac{\sigma_1^2}{2\sigma_2^2} \tag{4.9}$$

$$I_{32} = \frac{1}{\sqrt{2\pi\sigma_1^2}} \int \frac{2(x-\mu_1)(\mu_1-\mu_2)}{2\sigma_2^2} e^{(-\frac{(x-\mu_1)^2}{2\sigma_1^2})} dx$$

$$= \frac{2(\mu_1-\mu_2)}{2\sigma_2^2} E[(x-\mu_1)] = 0 \tag{4.10}$$

$$I_{33} = \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot \int \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2} e^{(-\frac{(x-\mu_1)^2}{2\sigma_1^2})} dx$$

$$= \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2} \cdot E[1]$$

$$= \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2} \tag{4.11}$$

$$\therefore I_3 = I_{31} + I_{32} + I_{33}$$

$$= \frac{\sigma_1^2}{2\sigma_2^2} + \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2} \tag{4.12}$$

所以非對稱性KL distance的公式如下：

$$KL(P1,P2) = I_1 + I_2 + I_3$$

$$= \log(\frac{\sigma_2}{\sigma_1}) + \frac{[(\mu_1-\mu_2)^2 + \sigma_1^2 - \sigma_2^2]}{2\sigma_2^2} \tag{4.13}$$

由於非對稱性KL(P1,P2)≠KL(P2,P1)，因此定義KL2(P1,P2)如下：

KL2(P1,P2)=KL(P1,P2)+KL(P2,P2) (4.14)

當P1，P2有相同的PDF時KL2 distance等於零。

我們將HMM的每個state每個維度以一個mean與variance來近似，利用原本訓練出來的HMM來求出要用來進似的mean與variance。新的mean與variance由以下公式求出：

$$\hat{\mu} = \sum_i^n C_i \cdot \mu_i \tag{4.15}$$

$\hat{\mu}$：表示新的*mean*

C_i：表示第*i*個*mixture*的*weigh*

$$E[x_i^2] = \sigma_i^2 + \mu_i^2$$

$$\hat{E}[x^2] = \sum_{i=1}^{n} C_i \cdot E[x_i^2]$$

$$\hat{\sigma}^2 = \hat{E}[x^2] - \hat{\mu} \qquad (4.16)$$

$\hat{\sigma}^2$：表示新的 var*iance*

如此一來，我們可以比較2個HMM每個state之間相似程度，甚至2個sub-syllable HMMs之間相似度。這裡我們只比較基本韻母與入聲韻母HMM之間confusion的情況，在未根據變調規則修改之前confusion的情況如圖4.3，以及根據變調規則作修改之後confusion的情況如圖4.4，圖中顏色越黑代表confusion的情形越嚴重。

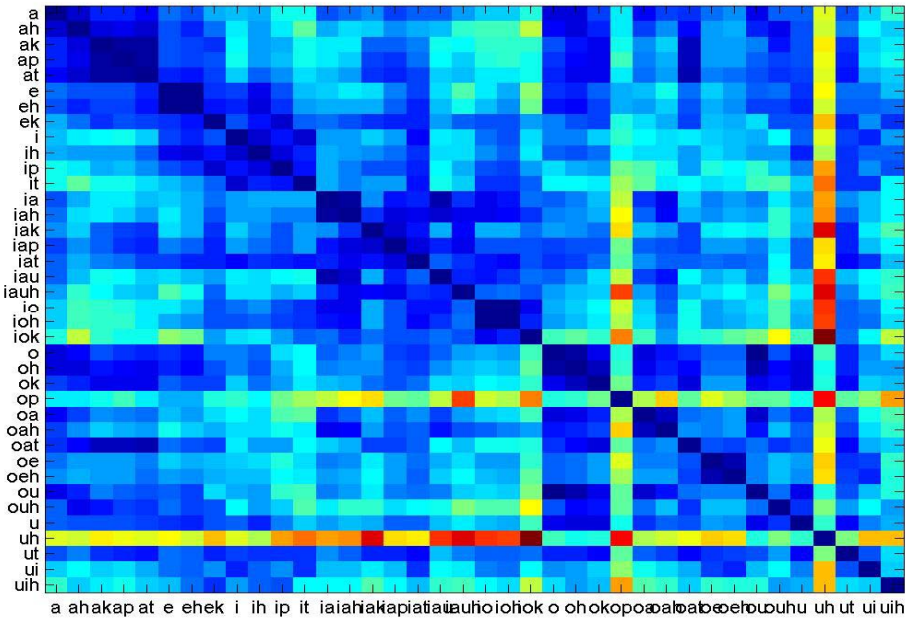


圖4.3未根據變調規則修改之前HMM之間confusion的情況

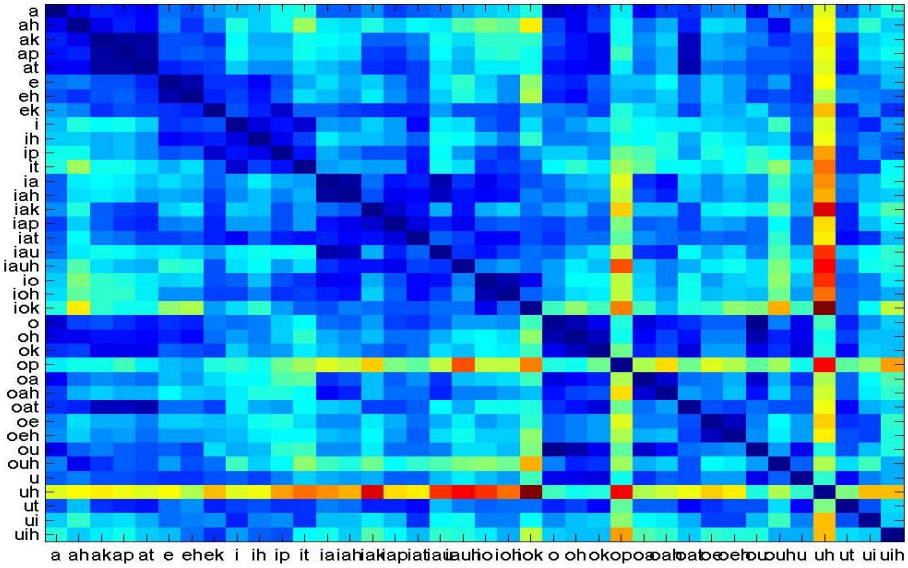


圖4.4根據變調規則修改之後HMM之間confusion的情況

每個state之間confusion的情況列在附件三，我們可以發現在根據變調規則修改之後，基本韻母跟入聲韻母之間比還未修改之前confusion的情形有所改善。

4.3 將語言模型加入台語辨識器

語言模型(Language Model, LM)[11]，可分為兩種，一種是依據語言的文法、字詞的詞性，訂定一些規則使得文章出現一定要符合規則之語言模型(Rule-Based LM)；另一種則是藉由處理大量文字資料，統計出詞與詞之間的聯接規則而建立的語言模型(Statistic-Based LM)。在本篇論文中我們採用統計式之語言模型。

4.3.1 語言模型簡介

所有的語言都有其文法規則，利用這類文法規則可求得一個機率模型，則我們稱此為語言模型。在進行語音辨識時，將 LM 所求的資訊結合聲音模型（Acoustic Model）資訊，通常可以大幅提高辨識系統的效能。

建立語言模型通常是以 Word 為基本單位，而在台語中，其對應的是「詞」為單位。原因是以「詞」為基本單位來建構會比較符合語言規則，也比較能看出所代表的意義。例如有一個詞（Word）為「球賽」，若將這個詞拆成兩個字「球」和「賽」（Character），在意義的表達上遠遠不及原本清楚。因此，我們會先建立一個詞典（lexicon），詞典裡面定義了所有我們要使用的詞，以便建立 LM。

由於缺乏大量文字語料，我們無法有效統計出可靠的 word-based LM，因此目前我們採用的 Back-off syllable bigram word-loop network，圖 4.5 為示意圖。

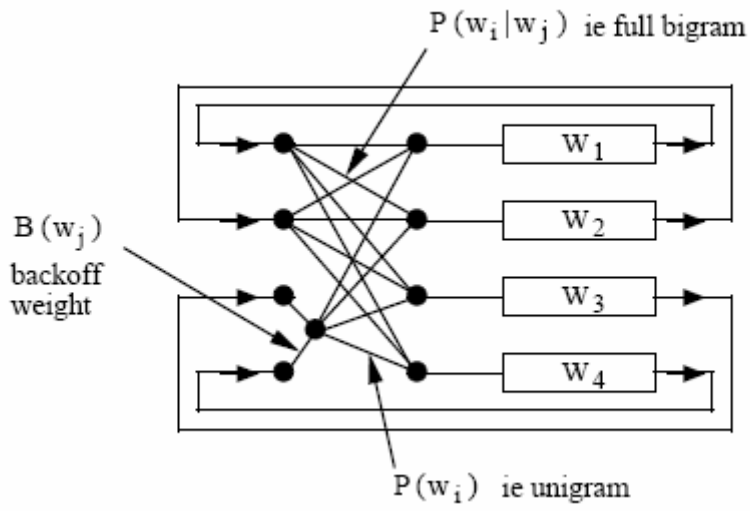


圖 4.5 Back-off bigram Word-Loop Network(form HTK book[9])

其中Wi 代表音節表中第i 個syllable，Pwj|wi=P(Wj|Wi)代表syllable Wi之後會接Wj 之機率。

4.4 辨識結果比較

在這邊我們建立 syllable unigram 和 syllable bigram 的 language model(LM)，加入辨識網路中，檢視辨識率提升的情形。

我們採用的訓練文字檔，列在下面表格 4.11：

表格 4.11 syllable database

	syllable 數	資料來源	標音方式
故事類	156084	鄭良偉教授 所提供	人工標音
社會新聞	54604	網路文章整理而成	人工標音
一般文章類	8892	台語通用會話：一套讓你能說出台語美辭雅語的教材	原來就有標音

採用 syllable unigram LM 辨識網路後，調整 penalty 與 LM weight 後得到最好的辨識結果，列在表格 4.12：

表格 4.12 Syllable unigram 辨識率

	音節總數 (N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	39.8	3.6	1.5	55.1

其中 $\text{Recong rate}=\frac{\text{N-(Sub+Del+Ins)}}{\text{N}}\times 100\%$

我們可以發現，只加入 syllable unigram 之後，辨識率可達 55.1%，提升了 6% 之多，接著我們改用 syllable bigram LM 辨識網路，調整 penalty 與 LM weight 得到最好的辨識結果，列在表格 4.13：

表格 4.13 Syllable Bigram 辨識率

	音節總數 (N)	Sub(%)	Del(%)	Ins(%s)	Recognition rate(%)
Outside test	12211	30.4	3.1	1.4	65.1

第五章 智慧型口語對話汽車導航系統（Intelligent Transportation System，ITS）

接著，我們將台語語音辨識系統使用於一個實際的應用系統上，並實際操作系統檢視成效如何。

5.1 系統簡介

系統架構圖如圖5.1所示[12]。包括即時語音辨認器，以辨認使用者的語音；對話管理模組，做對話流程控制，以及控制大哥大與汽車導航系統和擷取資訊；文字轉語音模組則將對話管理做出的回應轉換成語音輸出。

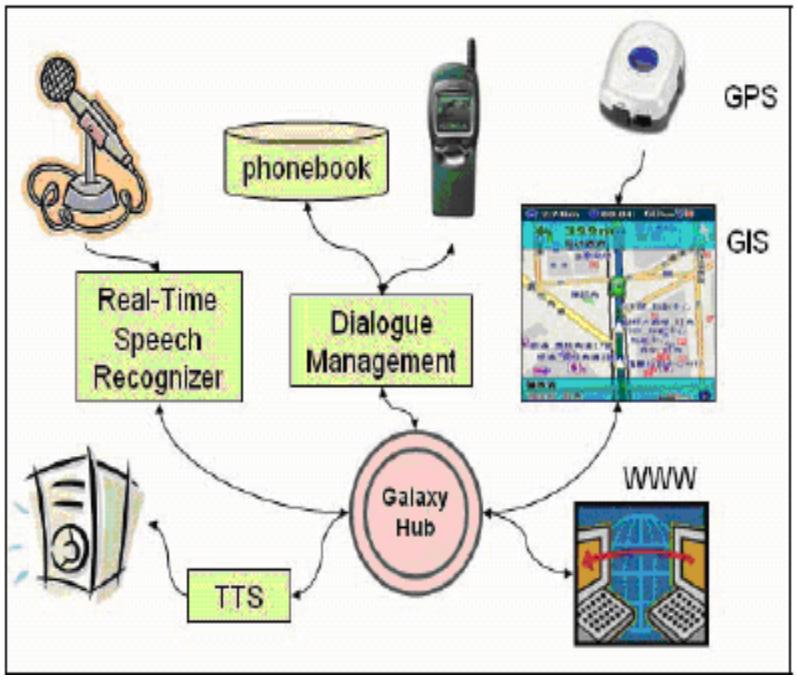


圖5.1　口語汽車導航系統方塊圖

5.2 架設工具與系統內部功能簡介

在圖5.1中即時語音辨認器裡使用的語音聲學模型與語言模型，我們使用英國劍橋大學釋放出來的Hidden Markov Model Toolkit（HTK）訓練產生。HTK可以作大字彙語音辨認，包括使用複雜的tri-phone模型，訓練tri-gram模型，也可以作cluster-based的語言模型。架設即時語音

辨認器的部份，我們是採用同是英國劍橋大學釋放出來的Application Toolkit for HTK（ATK），如圖2.2，除了方便與HTK相容外，ATK還可以在Windows作業環境下執行，支援多工，所以可以很方便架設一個Windows下的即時辨認器。此外ATK還提供了辨認結果可信度（Confidence）的功能，視每次辨認結果的分數高低附上相對應的可信度，可用來提升對話流程控制能力。

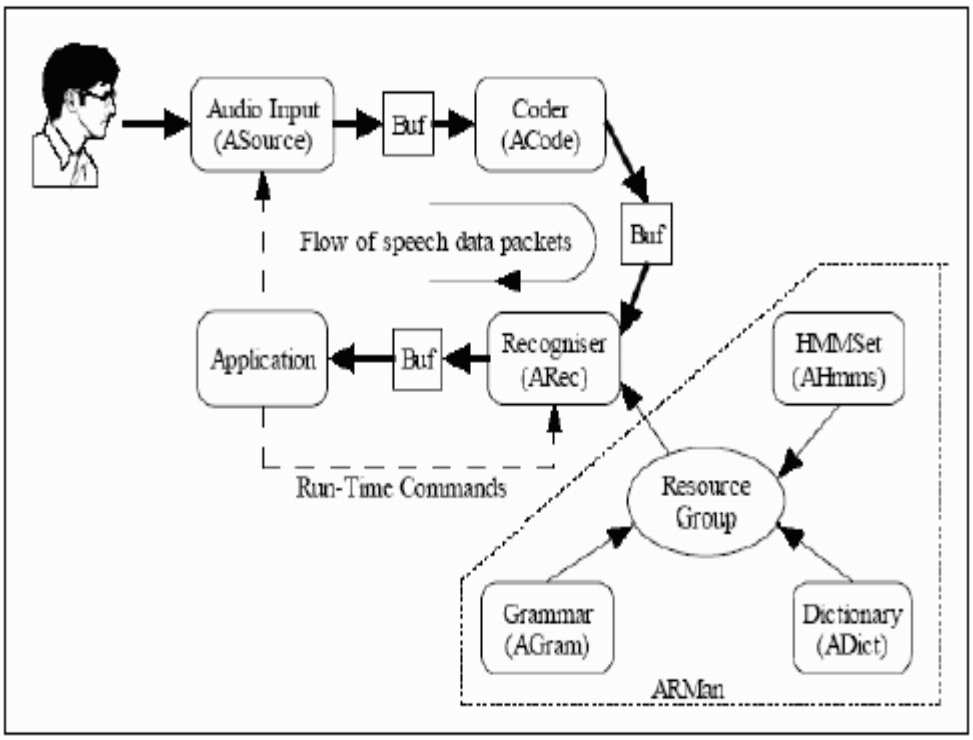


圖5.2 由ATK架設的基本即時辨認器(from Application Toolkit for HTK)

最後使用MIT釋放出來的Galaxy communicator將各模組的輸出輸入連接起來。如圖5.3所示Galaxy communicator為一Hub-Server架構，各伺服器可獨立執行，再透過網路連接Hub來傳遞訊息。好處是其為分散式系統，各伺服器端可在網路上的不同機器執行，所以各伺服器端可獨立開發執行，方便多人開發系統各個部份，以減少後續的程式碼維護負荷；而且若單台機器計算能力不足，亦可以用多台機器來分散計算負荷。

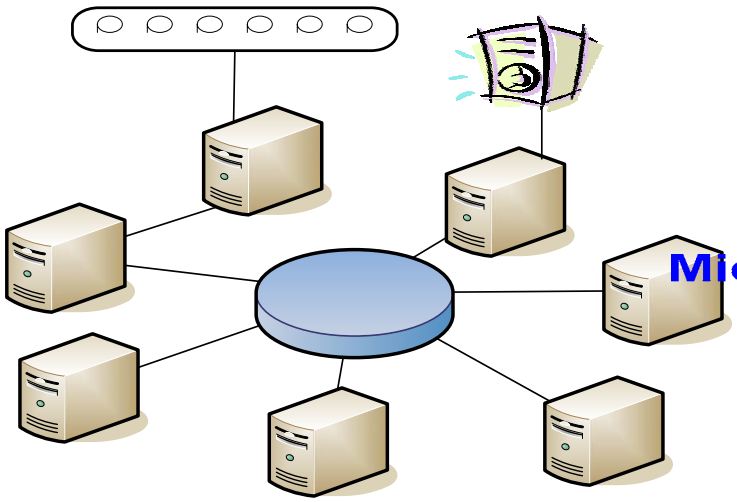


圖5.3 Galaxy Communicator software內部結構圖

ITS 對話系統是由 Microphone Array Server, ASR (Automatic Speech Recognition) Server, GIS (Geographic Information Systems) Server, Parser Server, Dialog Management Server, Corpus-based TTS(Text-To-Speech) Server, Natural Language Generation Server 和 Galaxy Hub. Servers 所組成，並透過 Galaxy Hub. Servers 溝通。 每一個 Server 負責的功能如下說明:

Microphone Array Server:

The Microphone Array Server 主要減低聲音訊號受車內環境所受的影響，能補償聲音訊號被影響部分。 然後將處理的聲音信號傳送給 ASR Server.

ASR Server:

ASR Server 主要是 Microphone Array Server 輸入的聲音訊號做語音辨識，並將辨識出來的字串傳送給 Parser Server。這部分主要使用的工具軟體為 ATK or HTK.

GIS Server:

GIS server 主要提供道路以及使用者感興趣的地點相關資訊(Point of Interests, POI)。 這部分主要使用的工具軟體為 PaPaGo SDK。

Parser Server:

Parser server 主要是解析語者所要傳達的意思。將 ASR server 輸入的辨識字串做解析以便瞭解語者需要何種服務或功能，並將訊息傳送給 Dialog Management Server。這部分主要使用的軟體工具為 “Phoenix: Semantic frame parser” of University of Colorado。

Dialog Management Server:

Dialog Management Server 主要控制對話流程。 將 Parser Server 傳送來的訊號做排序並驅動系統功能執行。

Natural Language Generation Server:

Natural Language Generation Server 主要根據語者輸入各種不同情況，產生自然流利的應答句子，並將這些句子輸出給 Corpus-base TTS Server。

Corpus-based TTS Server:

Corpus-based TTS server 將 Natural Language Generation Server 產生的句子轉換成聲音波形。

Galaxy Hub:

Galaxy hub 負責各個 server 之間訊號傳輸流程，確保每個 server 都能夠確實執行各自的功能。.

5.3 辨認模型的架設與訓練

台語ITS是將國語ITS做一些修改後所得，主要修改ASR Server 的部分，修改的項目包含：sub-syllable HMMs、Background Model、以及語法結構

5.3.1 Sub-syllable HMMs：

採用台語辨認系統的聲學模型。在特徵參數擷取方面， 取39維MFCC參數當特徵參數，相關之設定見表格5.1：

表格5.1特徵參數設定

取樣頻率	16kHz
視窗形式	Hamming window
音框長度(Frame size)	25ms
音框平移(Frame shift)	10ms
Feature vector	MFCC_D_A_Z_0

其中符號MFCC_D_A_Z_0 之意義為13維MFCC，13維Delta MFCC和 13維 Delta delta MFCC，並且做Cepstral Mean Normalization。

5.3.2 Background Model：

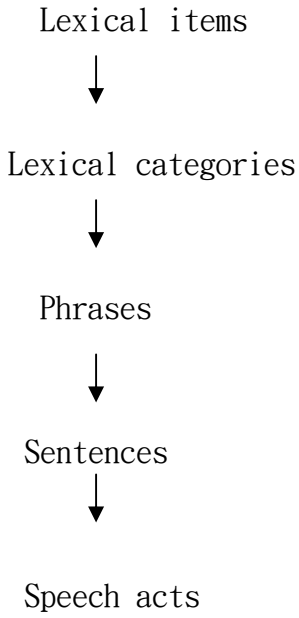
給 ATK 計算 confidence score 使用，一個 bghmm 模型 8 個 states 128 mixtures。

5.3.3 語法結構：

由清大資工所同學從收集的語料庫整理出來，再根據台語的語法作修改，並將一些習慣用國語來表達的道路名稱、人名、地點等等，採用國台語並用的機制來實行，這些詞類不管用國語或台語都可以順利的辨認出來。

台語 grammar 架構，基本上跟國語 grammar 架構相同，而國語的 grammar

架構如下說明：



例句如表格 5.2

表格 5.2 grammar 架構

Lexical items	我、要、去、交大、該、怎麼、走
Lexical categories	我[ID_user]、要[AUX]、去[V_Motion]、交大[PN_destination]、該[AUX]、怎麼[Q_WH]、走[V_motion]
Phrases	[ID_user] →[ID_user_pp]、 [AUX] →[AUX_pp]、 [V_Motion]→ [V_Motion_pp]、 [PN_destination] →[PN_destination_pp]、 [AUX] [Q_WH] [V_motion]→[Q_WH_pp]
Sentences	[ID_user_pp][AUX_pp][V_Motion_pp][PN_destination_pp] [Q_WH_pp]
Speech acts	user_inquire_destination

在整個國語 ITS 對話系統中，有 19 個 speech acts 和 23 個 syntactic-semantic categories，如表格 5.3，表格 5.4，還有實際系統運作時的例子:圖 5.4。

表格 5.3 19 speech acts in ITS Dialogue System

1	system_opening	12	user_opening
2	system_prompt	13	user_inquire_destination
3	system_hold_I	14	user_comfirm_arrive
4	system_hold_F	15	user_inquire_route
5	system_navigate	16	user_comfirm_request_a
6	system_correct	17	user_comfirm_request_b
7	system_confirm_a	18	user_pre-closing
8	system_confirm_b	19	user_arrive_destination
9	system_answer		
10	system_arrive_destination		
11	system_closing		

表格 5.4 The 23 categories in ITS Dialogue System

[ID_system] 系統	[ID_user] 你	[V_opening] 你好	[V_prompt] 請問	[V] 使用
[ADV] 繼續	[AUX] 想要	[V_motion] 去	[V_position] 在	[V_direction] 左轉
[CN_route] 路徑	[PN_route] 園區一路	[CN_destination] 交叉路口	[PN_destination] 清華大學	[N_anaphor] 那個
[Q_WH] 哪個	[Q_particle] 嗎	[Q_alternative] 還是	[NEG] 不是	[affirmative] 是的
[SPEC_position] 前方	[ASP] 了	[CONJ] 和		

圖 5.4 系統運作實例

台語 grammar 架構完全跟國語 grammar 相同，而有差異的地方，在於一些國語句子，當用台語來說時需要做一些修改，可以看出國語 ITS grammar 架構，轉移到台語時，可以完全的沿用，但首先要注意到的是，國台語中一些用詞的不同，以下有整理一些國語句子轉成台語句子的例子。

例如：

user_opening:

國:系統你好

台:系統你好

user_inquire_destination:

國:現在 我 要 去 交大 應該 怎麼 走

台:即馬 我 欲 拄 交大 應該 安怎 走

user_comfirm_arrive:

國:我 現在 在 交大

台:我 即馬 在 交大

user_inquire_route:

國:再來 要 怎麼 走

台:再來 欲 安怎 走

user_comfirm_request_a:

國:工業東二路 左轉 還 右轉

台:工業東二路 幹倒ㄐ 抑是 幹正ㄐ

user_comfirm_request_b:

國:我 順著 光復路一段 走 嗎

台:我 順 光復路一段 走 嗎

user_pre-closing:

國: 謝謝系統

台: 多謝系統

user_arrive_destination:

國:我 已 到達 六合夜市

台:我 已經 到 六合夜市

第六章 結論與未來展望

6.1 結論

本論文包含台語語音辨識與將台語辨識系統應用在 ITS 上，在台語語音辨識部分，我們針對原先資料的瑕疵做修正，並擴充語料庫，將原本的辨識率提升到 46.4%，且進一步的探討台語入聲調變調規則，根據變調規則下去修改訓練語料與測試語料，並改變辨識網路，將辨識率提升到 50.2%，我們可以發現入聲調變調在台語辨識系統中是一個很大的問題。除此之外，我們還加入 syllable bigram language model 將辨識率改善至 65.1%，相信只要擴充台語文字庫到足夠建語言模型，可以再進一步的提升辨識率。

ITS 系統部分，將台語應用加入之後，提升的系統使用的方便性，讓使用者有多一種語言選擇，增加了系統的彈性。

6.2 未來展望

在台語語音辨識方面，有很多問題需要去解決，語料庫的不足，以致於有些聲學模型無法得到一個可靠的效果；各地口音的差異，也是問題的關鍵，這些可以靠大量收集語料庫來讓問題減少。另一部份，就台語鼻音與鼻話韻母以及輕聲調方面的台語語音特性，都是未來可以努力的方向，相信一定可以將台語辨識率在向上提升。

在實際的對話系統『智慧型口語對話汽車導航系統』上，可以將語言模型（Language model）添加到及時語音辨認器上，並設計新的文字翻譯與會話分析模組，使系統更能在實際口語對話中的情境下使用，並讓系統可以國台語並用。

參考文獻

【1】 陳珮玥，”台灣閩南語中首部動作特指「打」的語意探析”，第五屆漢語詞彙語意學術研討會論文集，新加坡，2004 年 6 月。

【2】 王閔鴻，”不特定語者大辭彙華台雙語辨識引擎之研製及其應用”，私立長庚大學碩士論文，民國九十二年六月。

【3】 鄭良偉，”台語的語音與詞法”，遠流出版社，1997 年

【4】 鍾榮富，”台語的語音基礎”，文鶴出版有限公司，2002 年 11 月。

【5】 王文德，”台語語音辨識與文字處理之研究”，國立交通大學碩士論文，民國九十三年七月。

【6】 方南強，”台語通用會話:一套讓你能說出台語美辭雅語的教材，第二集”，開拓出版商，1997 年。

【7】 鄭如玲，”一分鐘台語單字速成”，三思堂文化事業有限公司，2002 年

【8】 S. Young，”A Review of Large-vocabulary Countinuous-speech Recognition”，IEEE Signal Processing Magazine，1996.

【9】 S. Young，G. Evermann，T. Hain，D. Kershaw，G. Moore，J. Odell，D. Ollason，D. Povey，V. Valtchev，P. Woodland，” The HTK book(for HTK version 3.3”

【10】A. Seigler，Uday Jain，Bhiksha Raj，Richard . Stern，” Automatic Segmentation, Classification and Clustering of Broadcast News Audio”， ECE Department – Speech Group Carnegie Mellon University Pittsburgh, PA 15213

【11】張榮勳，”國語廣播新聞語音基本辨認系統之建立”，國立交通大學碩士論文，民國九十四年七月。

【12】蔡金翰，”語音對話系統和對話策略之研究”，國立交通大學碩士論文，民國九十四年七月。