國立中興大學資訊網路與多媒體研究所

碩士學位論文

以資料探勘之技術解決線上客語語音合成系統中多音字發音歧義之研究

A Research on Resolving Pronunciation Ambiguity of Polyphonic Characters by using Data Mining Techniques in an On-Line Hakka Text-to-Speech System

指導教授：余明興博士Ming-Shing Yu

 黃豐隆博士 Feng-Long Huang

研 究 生：羅丞邑 Cheng-Yi Luo

中華民國一百年七月



# 誌謝

 首先我要感謝我的指導教授余明興老師與黃豐隆老師，謝謝老師們在這兩年的指導，以及當我在學習或是實驗上有困難的時候，適時的給我指引方向，多虧兩位老師的幫忙，才能順利完成這篇論文。

 感謝實驗室的旺松、炯榮、書豪、昭甫、俊賢、振淵、俊毅、昶毅學長們，在咪聽的時候總是能夠不畏艱難的挑起大樑，把先發和後援的局數都吃光，很懷念大家一起在實驗室玩的殭屍和籃球團康活動。感謝跟我同屆的任余、家銘、又箴、榮現、亦傑，在課業上可以互相討論幫忙，在球場上則是一起揮灑汗水，在平日生活互相的吐槽，在團康活動中則是運用心機互相利用，有你們當同學是我的榮幸。感謝學弟妹尚毅、珮均、健祐、尉綸，有你們的加入讓實驗室更加熱鬧，而且也幫忙做了很多苦工，希望你們在未來的一年能夠順利如期畢業。

 最後感謝我的父母、家人以及女友，沒有你們默默的支持與鼓勵，就沒有今天的我。也謝謝我許多好友的陪伴，感謝大家！

# 摘要

近年來由於政府的政策，以及少數族群意識的抬頭，鄉土文化與方言教學逐漸受到重視，由於現代網際網路非常發達，因此靜態的學習資訊幾乎是隨處可得，但動態輔助學習的工具則略顯不足。因此我們著手建立客語系統，採用以詞為合成單元的架構，並在各詞間加入適當的停頓，以及對多音字的判斷，使發音結果清晰，讓人可以輕鬆聽懂。

我們的重點放在多音字的處理，從一字兩音、一字三音等，各挑選出幾個來做模型訓練。我們用決策樹分類器訓練出多音字模型，並以此預估當文句中含有不在字典中的多音字時，該多音字應發哪種音。最後把預估的結果合成語音，並進行線上的聽測實驗，結果顯示，本系統所產生的客語合成語音具有真人說話之自然與清楚的特性。

關鍵字：客語；多音字；語音合成系統；詞義辨識

# Abstract

This thesis aims at the implementation for Hakka Text-to-Speech (TTS) System on Internet. Our system is composed of four components as follows: Text analysis, Mandarin to Hakka, Prosody prediction, and Speech generation module. More than 5400 monosyllabic speech units and 4063 word speech units of Hakka and several silences with various durations have been recorded as basic unit for speech synthesis. By adding breaks to Hakka sentences and finding out the pronunciation of polyphonic characters appropriately, we can provide real synthesis speech with frequent, prosodic and natural quality on Internet .

We focus on solving pronunciation ambiguity of polyphonic characters, i.e., to determine which pronunciation should be chosen. We predict pronunciation by using Bayesian network classifier、 C4.5 decision tree classifier、 CART classifier, and SVM classifier. The result of our experiments show that we can handle the prediction of some words very well in our Hakka Text-to-Speech System.

**目錄**

[誌謝. ii](#_Toc300756231)

[摘要. iii](#_Toc300756232)

[Abstract iv](#_Toc300756233)

[第一章 緒論 1](#_Toc300756234)

[1.1 研究動機 1](#_Toc300756235)

[1.2 數位學習簡介 2](#_Toc300756236)

[1.3 研究方向 3](#_Toc300756237)

[1.4文獻探討 6](#_Toc300756238)

[1.5論文架構 7](#_Toc300756239)

[第二章 客語四縣腔語海陸腔介紹 9](#_Toc300756240)

[2.1客語在臺灣的分佈 9](#_Toc300756241)

[2.2 四縣及海陸腔調簡介 11](#_Toc300756242)

[2.3 聲調表示 13](#_Toc300756243)

[2.4 四縣腔及海陸腔連音變調 15](#_Toc300756244)

[第三章 詞典、語音庫與語音合成系統模組介紹 19](#_Toc300756245)

[3.1 詞典與語音庫 19](#_Toc300756246)

[3.1.1 中文詞典 20](#_Toc300756247)

[3.1.2 國客語對照詞典 21](#_Toc300756248)

[3.1.3 語音庫 23](#_Toc300756249)

[3.2文句分析模組 24](#_Toc300756250)

[3.3韻律訊息模組 27](#_Toc300756251)

[3.3.1能量前處理-正規化 27](#_Toc300756252)

[3.3.2停頓的類型 29](#_Toc300756253)

[3.4語音合成模組 31](#_Toc300756254)

[第四章 多音字介紹及研究方法 37](#_Toc300756255)

[4.1 多音字介紹 37](#_Toc300756256)

[4.2 監督式與非監督式方法 37](#_Toc300756257)

[4.3 貝氏網路分類器 39](#_Toc300756258)

[4.4 C4.5決策樹與CART決策樹 42](#_Toc300756259)

[4.5 SVM支援向量機 45](#_Toc300756260)

[第五章 語音合成系統實作 48](#_Toc300756261)

[5.1 系統架構 48](#_Toc300756262)

[5.1.1 運作流程 48](#_Toc300756263)

[5.1.2 資料庫 49](#_Toc300756264)

[5.2 文句分析 51](#_Toc300756265)

[5.2.1實驗的方法 54](#_Toc300756266)

[5.2.2實驗的結果 58](#_Toc300756267)

[5.3 韻律訊息 73](#_Toc300756268)

[5.3.1實驗的方法 73](#_Toc300756269)

[5.4語音合成與聽覺測試 75](#_Toc300756270)

[5.4.1 語音合成 75](#_Toc300756271)

[5.4.2 聽覺測試與多音字聽覺測試 77](#_Toc300756272)

[第六章 結論 79](#_Toc300756273)

[參考文獻 81](#_Toc300756274)

[附錄. 86](#_Toc300756275)

**表目錄(從頁首開始)**

[表一：客語聲母符號表 12](#_Toc299632279)

[表二：客語韻母符號表 12](#_Toc299632280)

[表三：客語四縣腔調值、調號與調形 14](#_Toc299632281)

[表四：客語海陸腔調值、調號與調形 14](#_Toc299632282)

[表五：客語四縣腔連音變調規則 16](#_Toc299632283)

[表六：客語海陸腔連音變調規則 17](#_Toc299632284)

[表七：中研院46種簡化詞性標籤列表 21](#_Toc299632285)

[表八：國客語對照詞典的客語詞長(詞中的字數)統計 22](#_Toc299632286)

[表九：詞典內容範例 23](#_Toc299632287)

[表十 ：停頓類型的情況與說明 30](#_Toc299632288)

[表十一：系統開發環境 48](#_Toc299632289)

[表十二：多音字的類型、發音及語料數目 56](#_Toc299632290)

[表十三：「行」的各種特徵組合的正確率(門檻值74%) 60](#_Toc299632291)

[表十四：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「行」的confusion matrix 61](#_Toc299632292)

[表十五：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「行」的各發音正確率 61](#_Toc299632293)

[表十六：CART決策樹於前二後二詞性、目標詞性，多音字「行」的confusion matrix 62](#_Toc299632294)

[表十七：CART決策樹於前二後二詞性、目標詞性，多音字「行」的各發音正確率 62](#_Toc299632295)

[表十八：「調」的各種特徵組合的正確率(門檻值60%) 63](#_Toc299632296)

[表十九：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「調」的confusion matrix 64](#_Toc299632297)

[表二十：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「調」的各發音正確率 64](#_Toc299632298)

[表二十一：「重」的各種特徵組合的正確率(門檻值48%) 65](#_Toc299632299)

[表二十二：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「重」的confusion matrix 66](#_Toc299632300)

[表二十三：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「重」的各發音正確率 66](#_Toc299632301)

[表二十四：「差」的各種特徵組合的正確率(門檻值75%) 67](#_Toc299632302)

[表二十五：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「差」的confusion matrix 68](#_Toc299632303)

[表二十六：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「差」的各發音正確率 68](#_Toc299632304)

[表二十七：「易」的各種特徵組合的正確率(門檻值58%) 69](#_Toc299632305)

[表二十八：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「易」的confusion matrix 70](#_Toc299632306)

[表二十九：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「易」的各發音正確率 70](#_Toc299632307)

[表三十：「口」的各種特徵組合的正確率(門檻值80%) 71](#_Toc299632308)

[表三十一：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「口」的confusion matrix 72](#_Toc299632309)

[表三十二：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「口」的各發音正確率 72](#_Toc299632310)

[表三十三：各種標記的停頓長度 76](#_Toc299632311)

**圖目錄(從頁首開始)**

[圖一：線上客語語音合成系統架構 5](#_Toc299632312)

[圖二 ：台灣客家語言主要分佈 10](#_Toc299632313)

[圖三：四聲調之音高軌跡變化 15](#_Toc299632314)

[圖四：句子中可能詞之間的前後關係圖，圖中數字為起點至該節點的最少詞數 26](#_Toc299632315)

[圖五：音檔「實實在在」的波型 35](#_Toc299632316)

[圖六：語音合成流程圖 36](file:///L%3A%5C%E6%88%91%E7%9A%84%E8%AB%96%E6%96%87%5C%E6%88%91%E7%9A%84%E8%AB%96%E6%96%87%5C%E8%AB%96%E6%96%87%E5%88%9D%E7%A8%BF0728.docx#_Toc299632317)

[圖七：貝氏網路與條件機率表 41](file:///L%3A%5C%E6%88%91%E7%9A%84%E8%AB%96%E6%96%87%5C%E6%88%91%E7%9A%84%E8%AB%96%E6%96%87%5C%E8%AB%96%E6%96%87%E5%88%9D%E7%A8%BF0728.docx#_Toc299632318)

[圖八：二元決策樹樹狀圖 44](#_Toc299632319)

[圖九：SVM示意圖 46](#_Toc299632320)

[圖十：線上客語語音合成系統操作介面 49](#_Toc299632321)

[圖十一：Database schema 50](#_Toc299632322)

[圖十二：從中文詞查找出客語讀音流程 53](#_Toc299632323)

[圖十三：TreeBank範例示意圖 55](#_Toc299632324)

[圖十四：自然度測試結果 77](#_Toc299632325)

[圖十五：多音字測試結果 78](#_Toc299632326)

# 第一章 緒論

語言是人類用來溝通表達的工具，目的是交流思想和觀念等。科技日新月異，現代科技產品功能越來越先進，其中讓開發人員注重的一個重要項目就是如何讓電腦能夠像人一樣聽說讀寫，因此才會有這許多人致力於語音辨識和語音合成的領域。文字轉語音合成系統(Text-to-Speech, TTS)是語音技術發展中非常重要的一個環節，主要的功能在於經由使用者輸入文字，而讓電腦發出近似於人類的流暢自然且易懂的語音。

## 1.1 研究動機

電子書、語言翻譯機和語言學習系統等完善的語音合成系統不僅可以帶給我們生活極大的便利，對於弱勢文化的保存也有幫助，因為我們可以利用這些工具去學習不熟悉的語言。目前台灣客家族群約占總人口的七分之一，但是最大的通病是有的客家人本身並不太會講客家話，甚至聽不懂。相對於國語和閩南語，客語和原住民語比較起來相對就顯得弱勢許多。弱勢族群語言往往因為使用頻率低，傳承不容易，又因近代社會變遷、地理環境等因素影響，族群中會使用方言年輕人逐漸減少，最後不再有人使用此方言。文化是人類寶貝的資產，語言的傳承，代表著族群文化可以一直生生不息，方言消失即代表文化的傳承有可能會有斷層，甚至是有消失的危機，因此客語的永續傳承是相當重要的。

近年來大家越來越重視方言的問題，因此方言教學與認證的資訊越來越多，又因為網路的盛行，因此許多單位紛紛把教學資源建立成網站，但是以閩南語居多，例如互動功能的有聲詞典網站或線上語音合成系統[13][17]，而客語教學以靜態的文章居多，較具有公信力的則以[14][16]為主。我們的語音合成系統目的是讓使用者自行輸入中文文句，進而轉換出對應的客語詞及發音，合成出流利且自然的語音，可以讓原本不會客語的使用者做客語的語言學習。

## 1.2 數位學習簡介

數位學習(E-Learning)是指經由數位媒介，例如網路或多媒體來進行學習，可以是很正式的學習，例如學業上的課程，也可以是很輕鬆的學習，例如專長興趣的課程。透過網際網路，即時傳遞各種資訊和知識，E-Learning的E除了代表「Electronic電子的」，同時也包含了下列的其它意涵[26]：

(1)Exploration：學習者可自由探索學習。

(2)Experience：學習者可經由與電腦互動、模擬、及學習社群的支援等，改變學習經驗，進化和增進學習水平。

(3)Excitement：參與數位學習能帶給學習者刺激和興奮感。

(4)Extend：能提供更多的學習選擇，將學習變成持續進行的流程，延伸學習的時間。

(5)Expand：把教學管道的提供擴大到課堂外、從紙上到線上。

(6)Easy：數位教材介面，容易學習者使用與操作，且使用者可隨時上線，容易學習。

(7)Effective：學習者能獲得有效的學習成效。

 而數位學習的優點主要是降低學習成本、內容及時可靠且易於更新、不受時間空間的限制、富彈性的學習內容及全球通用的數位媒介。缺點則是學習者易感到孤獨、較適合主動的學習者、無法因材施教、數位媒介設備的問題等。而我們的語音合成系統就是希望應用在數位學習上，幫助不會客語的使用者能夠輕鬆的學習客語不同腔調的正確讀音，以及國語詞和客語詞之間的對應。

## 1.3 研究方向

近來各研究單位紛紛建置文字轉語音合成系統，但供線上使用與支援客語發音的系統卻少之又少。我們實作了線上客語語音合成系統，其運作流程為使用者輸入國語文句，經國語斷詞處理後，將斷出的各國語詞分別轉成對應的客語詞，並處理有多音字的客語詞，接著找出客語詞對應的音檔，在音檔間加入適當的停頓，合成連續語音之後輸出。

本研究的重點放在客語多音字的判斷上。從國語文句中萃取多音字特徵，使用資料探勘(Data Mining)的方法，分別以C4.5決策樹(Decision tree)、貝氏網路分類器、CART決策樹以及SVM分類器預估多音字的類型。最後依判斷出的多音字和加入停頓的結果實作線上客語語音合成系統，系統的整體架構如圖一所示。

中文文句與文章輸入

文句分析

(斷詞與詞性標記)

中文詞典

國語詞轉客語詞、

詞內連音變調、多音字判別

國客語對照詞典

語音合成處理

客語帶聲調單音合成單元

客語文句與客語拼音輸出

產生韻律訊息

韻律預估資訊

客語詞合成單元

 合成語音

 多音字分類模組

圖一：線上客語語音合成系統架構

## 1.4文獻探討

在此之前，我們研讀了幾篇有關的論文，大部分都是語音合成系統的研究，包括客語與國語：

客語語音合成之初步研究[24]，論文焦點放在韻律參數產生及音節信號波型上，採用的是基週波形時間比例內插法(TIPW)，此技術合成的語音信號，其音質和其他時域方法合成的一樣清晰，並且可以接受基頻軌跡和音長的調整。在文句分析的國語對應到客語的轉換，只以詞語對應、轉換的方式處理，不過客語的多音字以及讀音語音有分別的字，並沒有做特別的處理。最後驗證出借用國語基週軌跡來合成客語是可行的。

客語文句翻語音系統之實作[6]，此篇論文中，收集36259個客語詞建成詞庫，錄製不帶聲調的671個基本音節做為基本合成單元。錄製客語短文當訓練語料，使用遞迴式類神經網路取得韻律訊息，再採用PSOLA方法做語音合成。發現平滑的波形起始與結尾，能合出較悅耳的語音，且韻律訊息產生器所需的訓練語料庫經由人工調整會比從HTK求得的位置更準確，加上校正基頻軌跡不連續的情況，可以得到較好的均方根值(RMSE)。但未對入聲韻尾做特殊處理。

基於隱藏式馬可夫模型之中文語音合成系統[25]，。此論文在訓練部分使用特徵向量和標記(label)來建立內文相關隱藏式馬可夫模型(Context-Dependent HMM)，包括相關內文的語音、語言、韻律等資訊。合成部份則是根據狀態持續時間機率密度函式來決定每一個模型中的狀態持續時間，挑選出能使得輸出機率最大化的參數序列，通過梅爾對數頻譜近似濾波器之後輸出合成語音。

## 1.5論文架構

本論文共分六章，章節概述如下：

1. 緒論：介紹研究動機，研究方向與過去的文獻。
2. 客語四縣腔與海陸腔介紹：簡介客語四縣腔和海陸腔，包括在臺灣的客語族群分佈、四縣腔調拼音表、海陸腔調拼音表與連音變調規則，
3. 詞典、語音庫與語音合成系統模組介紹：介紹詞典及語音庫的來源與內容與前處理工作，以及斷詞器的功能、韻律訊息模組和語音合成模組。
4. 多音字介紹及研究方法：先簡單介紹多音字的問題，然後描述資料探勘的概念，接著列出本論文所處理的六個多音字的類型、發音及語料個數，最後一一簡述本論文所採用的不同分類演算法的概念。
5. 語音合成系統實作：本章把語音合成系統完整敘述一遍，包括環境、系統架構、運作流程和預估多音字的實驗數據。最後再附上聽測的結果。
6. 結論與未來展望：對本論文的實作結果分析後下結論，並說明未來改進的方向。

# 第二章 客語四縣腔語海陸腔介紹

漢語有許多方言，大體而言有七大方言：官話、吳語、湘語、贛語、閩語、客語、粵語。他們之間有很大的差別，基本上沒有學習是不能互通的，而這些方言中很多還有各自的次方言。客家話原本屬於北方方言，在客家先民南遷的過程中和南方的其他方言互相融合，而形成了一種和北方方言不同，卻又異於南方方言的一種新語言，就是客家話。客語在臺灣有五個主要腔調，分別是四縣、海陸、大埔、饒平、和詔安。其中最多人使用的是四縣腔，其次是海陸，故本系統目前先以四縣腔及海陸腔客語為研究主體。

## 2.1客語在臺灣的分佈

臺灣地區的客語分為許多腔調，最多人使用的是「四縣腔」，四縣指的是原本屬於廣東省梅州附近的四個縣：興寧、五華、平遠、蕉嶺。說四縣腔的地方主要是桃園縣的一部份、新竹縣的一部份、苗栗縣以及南部六堆地區。四縣腔可以說是臺灣所用的客語次方言中人數最多的一支，約占四分之三。在臺灣第二多人使用的是「海陸腔」，原本屬於廣東省的海豐和陸豐縣，說海陸腔的地方主要是桃園和新竹的一部份地區。

再次是原屬廣東省潮州大埔和豐順的「大埔腔」，現在分佈在臺中東勢、石岡、新社。還有原屬廣東省潮州饒平的「饒平腔」，現散居在桃竹苗間、彰化員林、永靖、田尾。至於原屬福建省漳州秀篆、官陂、霞葛、太平的「詔安腔」，現分佈於雲林二崙、崙背、西螺及桃園八德、大溪等地區。如圖二顯示臺灣五個客語主要腔調的分布地區。



圖二 ：台灣客家語言主要分佈

另外，在南部六堆使用的四縣腔，事實上與北部使用的四縣腔有很大的差異，加上在桃竹苗區域，四縣與海陸的交界附近，兩腔調混合衍生出「四海話」。因此，若把臺灣的各客語腔調詳細列出，共有：四縣、海陸、大埔、饒平、詔安、六堆、四海，以及零星的永定、豐順、武平、五華、揭西、揭陽等。

## 2.2 四縣及海陸腔調簡介

在客語中四縣腔的聲母最少，參考臺灣客家語拼音方案2009年2月的版本[15]，如表一是客語聲母符號表，表二則是客語韻母符號表。四縣腔變調的規則只有陰平加去聲變陽平加去聲、陰平加陽入變陽平加陽入及陰平加陰平變成陽平加陰平三種，是最容易學習的客語腔調。海陸腔的變調規則是低聲調上聲後面不論接什麼調時，皆要變為中平調陽去；以及高入聲調陰入聲後面不論接什麼調時，皆要變為低入調陽入聲兩種。

臺灣北部的四縣腔和南部六堆的四縣腔略有差異[1]。例如南部的回去說「**歸去**」，而不說「**轉去**」，雞蛋說「**雞春**」而不說「**雞卵**」等。音韻系統上也略有差別，部份是受到閩南語的影響。

表一：客語聲母符號表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **客家語拼音** | b | p | m | f | v | \*bb |
| **國際音標** | [p] | [ph] | [m] | [f] | [v] | [b] |
| **注音符號** | ㄅ | ㄆ | ㄇ | ㄈ | 万 | ㆠ |
|  |  |  |  |  |  |  |
| **客家語拼音** | d | t | n | l | g | k |
| **國際音標** | [t] | [th] | [n] | [l] | [k] | [kh] |
| **注音符號** | ㄉ | ㄊ | ㄋ | ㄌ | ㄍ | ㄎ |
|  |  |  |  |  |  |  |
| **客家語拼音** | ng | h | \*j | \*q | \*x | z |
| **國際音標** | [ŋ] | [h] | [tɕ] | [tɕh] | [ɕ] | [ts] |
| **注音符號** | ㄫ | ㄏ | ㄐ | ㄑ | ㄒ | ㄗ |
|  |  |  |  |  |  |  |
| **客家語拼音** | c | s | \*zh | \*ch | \*sh | \*rh |
| **國際音標** | [tsh] | [s] | [t∫] | [t∫h] | [∫] | [Z] |
| **注音符號** | ㄘ | ㄙ | \*ㄓ | \*ㄔ | \*ㄕ | \*ㄖ |

\*bb可用於詔安腔、南投國姓鄉及部分南部客家地區。

\*j、\*q、\*x僅用於四縣腔； \*zh、\*ch、\*sh、\*rh用於大埔、海陸、饒平、詔安等四腔；\*ㄓ、\*ㄔ、\*ㄕ、\*ㄖ之發音時舌頭放平，與翹舌音有別。

表二：客語韻母符號表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **客家語拼音** | a | \*i | \*u | e | o |
| **國際音標** | [a] | [i] | [u] | [e] | [o] |
| **注音符號** | ㄚ | │ | ㄨ | ㄝ | ㄛ |
|  |  |  |  |  |  |
| **客家語拼音** | ii | \*er | \*m | \*n | \*ng |
| **國際音標** | [i`] | [γ] | [m] | [n] | [ŋ] |
| **注音符號** | 帀 | ㄜ | ㄇ | ㄋ | ㄫ |
|  |  |  |  |  |  |
| **客家語拼音** | \*b | \*d | \*g |
| **國際音標** | [p] | [t] | [k] |
| **注音符號** | ㄅ | ㄉ | ㄍ |

\*-i、-u可用在陰聲韻尾；韻尾-m、-n、-ng只用在陽聲韻尾（鼻音韻尾）。-b、-d、-g只用在入聲韻尾（塞音韻尾）。
\*er用於海陸、饒平二腔。

海陸腔在聲調上因為去聲分陰陽，所以比四縣腔多了一個聲調，變成七個聲調。此外它有舌尖面聲母存在，所以有的音聽起來略有一些捲舌的感覺，聲母的個數也比四縣多。它的調值和廣東話有一些很相像的地方。海陸腔和四縣腔間用的詞語略有不同。例如花生海陸腔講「地豆」，蚯蚓海陸腔講「蟲憲」等。此外海陸腔的變調也比四縣來的複雜。

由於海陸腔各聲調的調值和四縣腔差不多相反，也就是說四縣腔唸為高音的，海陸腔唸為低音；四縣腔唸為低音的，海陸腔則唸為高音。所以在轉換上有規則可循。

## 2.3 聲調表示

 四縣客語的聲調有六種，即高、低、升、降，及入聲的高、低。客語保留古代漢語的入聲，這部分國語並沒有保留。入聲的發音通常急而短促，只有高低兩種。四縣腔及海陸腔調號與調形依據臺灣客家語拼音方案[15]，分別如及表四，聲調音高軌跡變化如[18]。

表三：客語四縣腔調值、調號與調形

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **調類** | **陰平** | **陽平** | **上聲** | **陰去** | **陰入**(b,d,g) | **陽入**(b,d,g) |
| **四縣調值** | 24 | 11 | 31 | 55 | 21(ˋ) | 5 |
| **四縣調號** | 2 | 3 | 4 | 1 | 2 | 5 |
| **四縣調形** | fuˊ | fuˇ | fuˋ | fu | fugˋ | fug |
| **例字** | 夫 | 扶 | 虎 | 富 | 福 | 服 |
| **國語(近似)** | 2聲ˊ | 3聲ˇ | 4聲ˋ | 1聲 |  |  |
| 表四：客語海陸腔調值、調號與調形

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **調類** | **陰平** | **陽平** | **上聲** | **陰去** | **陰入**(b,d,g) | **陽入**(b,d,g) | **陽去** |
| **海陸調值** | 31 | 55 | 24 | 11 | 5 | 21(ˋ) | 33 |
| **四縣調號** | 4 | 1 | 2 | 3 | 5 | 2 | 5 |
| **海陸調形** | fuˋ | fu | fuˊ | fuˇ | fug | fugˋ | fu+ |
| **例字** | 夫 | 扶 | 虎 | 富 | 福 | 服 | 護 |
| **海陸(近似)** | 4聲ˋ | 1聲 | 2聲ˊ | 3聲ˇ |  |  |  |

 |

****

圖：四聲調之音高軌跡變化

## 2.4 四縣腔及海陸腔連音變調

客語四縣腔的連音變調規則有三種，如表五所示，表六則是海陸腔的連音變調規則。

表五：客語四縣腔連音變調規則

|  |
| --- |
| **規則1**：由兩個陰平（調號ˊ）字構成的字彙，讀時前字變調讀**陽平（**調號**ˇ）**               陰平 + 陰平 → **陽平** + 陰平 |
| **範****例** | **詞彙** | **變調前之拼音** | **變調後之拼音** |
| 新衫 | sin**ˊ** sam**ˊ** | sin**ˇ** sam**ˊ** |
| 買新衫 | mai**ˊ** sin**ˊ** sam**ˊ** | mai**ˇ** sin**ˇ** sam**ˊ** |
| **規則2**：陰平（調號ˊ）與陰去構成的詞彙，讀時前字變調讀**陽平（ˇ）**               陰平 + 陰去 → **陽平** + 陰去 |
| **範****例** | **詞彙** | **變調前之拼音** | **變調後之拼音** |
| 針線 | ziim**ˊ** sien | ziim**ˇ** sien |
| 拿針線 | na**ˊ** ziim**ˊ** sien | na**ˇ** ziim**ˇ** sien |
| **規則3**：陰平（調號ˊ）與陽入字構成的詞彙，讀時前字變調讀**陽平（ˇ）**               陰平 + 陽入 → **陽平** + 陽入 |
| **範****例** | **詞彙** | **變調前之拼音** | **變調後之拼音** |
| 音樂 | im**ˊ** ngog | im**ˇ** ngog |
| 聽音樂 | tang**ˊ** im**ˊ** ngog | tang**ˇ** im**ˇ** ngog |

表六：客語海陸腔連音變調規則

|  |
| --- |
| **規則1**：上聲變調 即低聲調上聲（調號ˊ）後面不論接什麼調時，皆要變為中平調**陽去****（**調號**＋）** |
| **範****例** | **詞彙** | **變調前之拼音** | **變調後之拼音** |
| 打球 | daˊ kiu | da**＋** kiu |
| 古板 | guˊ banˊ | gu**＋** banˊ |
| 好漢 | hoˊ honˇ | ho**＋** honˇ |
| 啟發 | kiˊ fad | ki**＋** fad |
| 老牛 | loˊ ngiuˊ | lo**＋** ngiuˊ |
| 等路 | denˊ lu＋ | den**＋** lu＋ |
| 解決 | gaiˊ giedˋ | gai**＋** giedˋ |
|  |
| **規則2**：陰入聲變調 即高入聲調陰入聲後面不論接什麼調時，皆要變為低入調**陽入聲** **（**ˋ**）** |
| **範****例** | **詞彙** | **變調前之拼音** | **變調後之拼音** |
| 壁頂 | biagdang**ˊ** | biagˋdang**ˊ** |
| 決賽 | gied soi**ˇ** | giedˋ soi**ˇ** |
| 目珠 | mugzhuˋ | mugˋzhuˋ |
| 發覺 | fadgog | fadˋgog |
| 八字 | badsii**＋** | badˋsii**＋** |
| 法律 | fadludˋ | fadˋludˋ |

# 第三章 詞典、語音庫與語音合成系統模組介紹

本研究的主要目的是，在網際網路(Internet)上建置一個客語語音合成系統。使用者輸入正體中文文句後，經由文句分析模組處理可產生對應的客語文句，然後再經過韻律訊息模組與語音合成之後，輸出具有自然且流暢之客語語音，並標注客語拼音。此系統可提供國人學習客語之用，作為線上學習( Online e-learning)的平台。

在進行建置語音合成系統之前，要先進行許多前置準備工作，我們先建立詞典與錄製語音檔，這兩項工作同時分頭並行。系統使用的詞典有兩種，國語詞典與國客語對照詞典，分別用在系統中的不同模組。錄製語音檔需要準確無誤的發音，所以委託陳婷芳老師錄製四縣腔各種音檔，海陸腔則是委託詹益雲老師錄製。此章節介紹本系統中幾個重要模組。各模組的背景知識也會簡略描述，並舉例說明。

## 3.1 詞典與語音庫

 本系統使用的詞典有二，分別是供中文斷詞使用的中文詞典，與把國語詞轉成對應的客語詞所使用的國客語對照詞典。

### 3.1.1 中文詞典

 在許多的斷詞方法中，常需要查詞典，才能初步辨識出句子中存在哪些可能的詞。其中若又是以詞庫式斷詞為主的斷詞方法，對於斷詞結果的效能而言，詞典內容具有非常決定性的影響力。

 本研究目前使用的中文詞典，共有236509個詞。其來源是有二：(1)13萬目詞詞典，其詞典是以中研院的八萬目詞(ASCED)為基礎，另外再從中研院平衡語料庫(ASBC, 3.0)抽取未在八萬目詞內的詞，一共約13萬詞的詞典，新詞典也以此為基礎。

(2)Chinese GigaWord語料庫，把這個語料庫中的文章斷詞後，挑出不在13萬詞典中的詞，並且計算詞頻，若高於一門檻值，才將其加至新詞典中[19]。最後得到一部超過23萬詞的詞典。詞典內每一詞目有中文詞、讀音、詞性、和詞頻。

本研究所採用之詞性標記集合有43種基本標記，再加上3種特殊標記，共計46種詞類標記，如表七。

表七：中研院46種簡化詞性標籤列表

|  |  |  |  |
| --- | --- | --- | --- |
| **詞性標籤** | **詞性意義** | **詞性標籤** | **詞性意義** |
| A | 非謂形容詞 | Neu | 數量定詞 |
| Caa | 對等連接詞 | Nf | 量詞 |
| Cab | 連接詞 | Ng | 後置詞 |
| Cba | 連接詞 | Nh | 代名詞 |
| Cbb | 關聯連接詞 | P | 介詞 |
| D | 副詞 | SHI | 是 |
| Da | 數量副詞 | T | 語助詞 |
| DE | 的、之、得、地 | VA | 動作不及物動詞 |
| Dfa | 動詞前程度副詞 | VAC | 動作使動動詞 |
| Dfb | 動詞後程度副詞 | VB | 動作類及物動詞 |
| Di | 時態標記 | VC | 動作及物動詞 |
| Dk | 句副詞 | VCL | 動作接地方賓語動詞 |
| FW | 外文標記 | VD | 雙賓動詞 |
| I | 感嘆詞 | VE | 動作句賓動詞 |
| Na | 普通名詞 | VF | 動作謂賓動詞 |
| Nb | 專有名詞 | VG | 分類動詞 |
| Nc | 地方詞 | VH | 狀態不及物動詞 |
| Ncd | 位置詞 | VHC | 狀態使動動詞 |
| Nd | 時間詞 | VI | 狀態類及物動詞 |
| Nep | 指代定詞 | VJ | 狀態及物動詞 |
| Neqa | 數量定詞 | VK | 狀態句賓動詞 |
| Neqb | 後置數量定詞 | VL | 狀態謂賓動詞 |
| Nes | 特指定詞 | V\_2 | 有 |

### 3.1.2 國客語對照詞典

 國客語對照詞典目前約有3萬詞目，主要來源含行政院客委會、台北市客委會等單位，並有一部份自建。包括客委會所編輯發行的「客語能力認證 基本詞彙」初級與中高級的詞彙[3][4][5]，與幾位學者專家所提供的資料。每一詞目含有四縣腔與海陸腔資訊，及其拼音與例句，我們使用這些資訊來實作四縣腔與海陸腔語音合成系統。本系統所使用之客語拼音與聲調表示，是依據教育部09年修正之版本。若同一客語詞有多種拼音，則分別存成多筆資料。目前共29596個詞彙，其客語詞的詞長(詞中的字數)統計如八，詞典內容範例則如九。

表八：國客語對照詞典的客語詞長(詞中的字數)統計

|  |  |
| --- | --- |
| **字詞** | **個數** |
| 一字詞 | 926 |
| 二字詞 | 18060 |
| 三字詞 | 6154 |
| 四字詞 | 3949 |
| 五字詞 | 277 |
| 六字詞 | 80 |
| 七字詞 | 66 |
| 八字詞及以上 | 84 |
| 總計 | 29596 |

表九：詞典內容範例

|  |  |
| --- | --- |
| **欄位** | **內容** |
| 國語詞 | 年輕人 |
| 四縣客語詞 | 後生人 |
| 四縣客語拼音 | heu sangˊ nginˇ |
| 海陸客語詞 | 後生人 |
| 海陸客語拼音 | heu├ sangˋ ngin  |
| 英語詞 | youth |
| 國語例句 | 年輕人要多打拼。 |
| 四縣客語例句 | 後生人愛煞猛打拼做頭路。 |
| 海陸客語例句 | 後生人愛煞猛打拼做頭路。 |
| 資料來源 | 客委會 |

### 3.1.3 語音庫

 我們委託熟悉客語四縣腔的陳婷芳老師(台中市北屯區陳平國小老師)錄製客語四縣腔基本之合成單元，以客語單音節之語音為單位，含四縣腔之六種聲調，即高、低、升、降四種聲調，加上高、低二種入聲音，共有2427個基本合成單元，每個音檔以客語拼音開頭，接上阿拉伯數字聲調做為檔名。此外，也錄製客委會發行的「客語能力認證基本詞彙」的詞語音檔共2234個，以及不同時長之靜音檔。這些合成單元音檔錄製格式為：44.1kHz、16 bits，儲存成Windows PCM格式(wav檔)。

 海陸部分則是委託從新竹縣芎林鄉五龍國小退休的詹益雲老師錄製，以客語單音節之語音為單位，含海陸腔之七種聲調變化，即高、低、升、降、中五種聲調，加上高、低二種入聲音，共有3005個基本合成單元。同樣也錄製客委會發行的「客語能力認證基本詞彙」的詞語音檔共1829個。這些合成單元音檔錄製格式與上述四縣腔皆一致

。 本研究目前只錄製四縣腔與海陸腔的語音檔，以後的相關研究有可能加入其他次方言腔調。

## 3.2文句分析模組

 本模組接受中文文句輸入，先經過中文斷詞，斷出數個中文詞，再至國客語對照詞典中，找出各中文詞對應的客語詞，並且輸出客語詞與客語讀音。本節的重點在介紹此模組使用的中文斷詞器。

 常見的基本斷詞法有連續字串比對法、語法分析法……等[35][36]。連續字串比對法是將文句拆成小段至詞典中查找，列出所有可能的詞，並將當下找到最長的詞輸出，又稱「長詞組成優先」。語法分析法是利用語言學為基礎，加上經驗法則，分析句子的型態做斷詞處理。

 本系統使用的斷詞法為「少詞組成優先搭配詞的Unigram機率斷詞」[19]，「少詞組成優先」指的是句子斷詞後，詞數越少越好，與常見的「長詞組成優先」有類似的效果。例如：「臺灣人口眾多」，經過少詞組成優先斷詞後的結果可以為：「臺灣/人口/眾多」與「臺灣人/口/眾多」兩種，它們都斷成三個詞。但是我們必需選出比較好的一種，這時就用到Unigram機率斷詞，我們分別算出這些詞的機率，以機率大小來決定輸出的詞串，使得其機率值****最大。

 再以另一個實際的例子來說明這個方法。現在有個短句「一般電信業者」，先查詞典找出所有可能的詞(candidate)，然後列出句子中可能的詞與彼此間的前後關係，如圖四。首先我們先由前往後掃描一次，並將從起點至目前節點的最少詞數給記錄下來。接著從結尾節點T由後往前掃描，每次往回找尋累積詞數最少的節點，如此一來本應該可以找到一條擁有最少詞數的斷詞結果。但是以這個例子來說，卻有可能同時出現多種同樣詞數的結果，像「一般/電信業/者」及「一般/電信/業者」同樣詞數都為3。因此我們藉由比較這兩種情況unigram機率的大小，以此決定最終的斷詞結果。

1

2

2

1

2

2

3

3

3

3

圖：句子中可能詞之間的前後關係圖，圖中數字為起點至該節點的最少詞數

決定斷詞結果之後，斷詞器會對斷出的詞標註詞性。假設斷詞後得到一個詞串序列 ，其詞串序列對應的詞性序列為 ，而我們使用trigram 語言模型(language models)來找出最佳的詞性序列，也就是找一條擁有最大trigram機率的詞性序列，即

 (3-1)

(3-1)中，<S>表示為一個起始符號。表示已知詞的詞性為的機率。的機率值代表在前兩個連續詞性分別為的情況下，出現的機率是多少。加入此機率值的計算較能幫助我們找到與詞相關的詞性序列。

## 3.3韻律訊息模組

 此模組會根據文字分析模組的結果，為每個發音的單位估計出合適的韻律訊息，一般來說，包括音高(pitch)、音長(duration)、音量(volume)和停頓(break)。本系統錄製了許多客語詞的音檔，在這些音檔中已經包含了詞內的音長與音高變化，故只要把詞間的連接處做接合處理，並且調整音量，最後加上停頓，所合成的語音就比較清晰好聽。故合成時毋需調整音高，也幾乎不必調整音長。本節先敘述調整音檔能量此一重要的前處理，及找出文句中適當的停頓，決定停頓的長短類型。

### 3.3.1能量前處理-正規化

 語音庫的音檔由專業人士錄製，惟人工錄音時很難保證每個音檔的音量一致，導致合成出的連續語音音量忽大忽小。在這部分我們使用[9]提到的方法，針對語音庫的音檔做前處理，把音檔的能量正規化(normalization)。

(1) 單字音部分：

首先，對每一個單音語音求出其音量之平均值如下:

 (3-2)

(3-2)中，*i*為客語之聲調代號(*i* = 1表示第1聲，*i* =2表示第2聲，依此類推)，表示客語第*i*聲調中第*j*個語音檔，為語音之第*k*點取樣值，為此音檔取樣量化之點數，表示的語音能量之平均值。

 (3-3)

(3-3)中，為第*i*聲調的能量平均值，*Ni*為客語語音中聲調*i*之語音個數，本系統中*Ni* = 408。音量正規化則用公式(3-4)調整。

 (3-4)為原始語音信號，為標準化後之語音信號。

 (2) 多字音部分

一樣的方法，求出每一個多字音語音的平均音量，如下：

 (3-5)

(3-5)中，為多字音語音第k點的取樣值，此音檔共有個點，表示平均音量。

在做正規化前，先算出單字音的平均音量，

 (3-6)

為各單字音檔的平均音量，t為單字音音檔個數。求出所有單音檔音量平均值。

然後把單字音的平均音量除以此多音字平均音量，得一比值，再把此音檔乘上這個比值，即完成正規化。

 (3-7)

為原始語音信號，為標準化後之語音信號。

### 3.3.2停頓的類型

在語音合成系統中，停頓(Break)是一項重要的特徵。人類在閱讀時通常會把一個長句分割成幾個部分來讀，或是因為語氣或強調重點的不同，而在不同的時機產生大小不一的停頓。因此在語音合成系統中加入停頓預估，合成出的語音會更容易讓人理解。

在本系統中，把詞和詞之間分為三種不同的邊界，分別是：韻律片語(prosodic phrase)間、韻律詞(prosodic word)間和詞間。

若要更詳細的區分，在任兩個中文字間，所有可能的停頓可以分為五大類。以下照停頓的長到短，依序列出，見表：

表十 ：停頓類型的情況與說明

|  |  |
| --- | --- |
| **情況與說明** | **停頓類型** |
| 在五大標點之間(五大標點“。！？；，＂之間) | 大停頓 |
| 韻律片語間(五大標點之內) | 中停頓 |
| 韻律詞間(韻律片語之內) | 小停頓 |
| 詞之間(韻律詞之內) | 無停頓 |
| 字元之間(詞之內) | 連音 |

在此舉個實際的例子，來說明為何我們要把句子斷詞後，再做加入停頓的處理，藉此得到更自然的合成語音。假設我們現在輸入的句子是：

瀕臨絕種動物增加的原因在於棲息環境的破壞惡化及人類的撲殺。

經過斷詞器的處理後，可以得到結果如(E1)：

瀕臨 絕種 動物 增加 的 原因 在於 棲息 環境 的 破壞 惡化 及 人類 的 撲殺。 (E1)

 如果直接用斷完詞的結果，各詞之間不加入停頓，會因句子太長而讓人有不符合自然口語的感覺。若在各詞之間都加入停頓，則合成出斷斷續續的語音，會讓人聽起來覺得過於冗長。如果可以加入特定的少量停頓，類似我們講話的流暢情形，讓整個長句斷成幾個較短的部分，如下：

瀕臨絕種動物 增加的原因 在於 ＊ 棲息環境的破壞惡化 及 人類的撲殺

(E2)

 例中空白的部分，是韻律詞的邊界，即前所述的小停頓(minor break)；而句子下方長線代表韻律片語的範圍，「＊」符號為韻律片語的邊界，就是中停頓(medium break)；原先被斷詞器斷開但最後又被合併的部分，稱為無停頓(no break)。在合成長句的時候，把小停頓和中停頓插入文句中對應的位置，可讓合成語音的語意更加清楚。另外，五大標點符號對應到大停頓以及詞內直接連音，這些由斷詞完的結果就可以處理。本模組的重點放在詞與詞之間，預估(predict)停頓類型是中停頓、小停頓或無停頓。我們求取停頓類型的方法則是採用資料探勘的方式[11]。

## 3.4語音合成模組

 語音合成的主要功能是根據韻律訊息模組的結果，取出相對應的語音合成單元，利用特定的語音合成技術對語音合成單元進行韻律特性的調整和修正，而目前較常見語音之產生技術有二類：

(1)波形拼接法(Formant Synthesis)：

同步疊加法(Pitch Synchronous Overlap and Add，PSOLA)在時域(Time Domain) 上改變語音波形的基頻和時長，以產生合成語音[28]，此法可改變語音之音長(Duration)與音高(Pitch)韻律訊息，改良頻域(Frequency Domain)處理耗時的缺點，如此就可產生出較自然且具有韻律的合成語音。由於PSOLA是利用音高週期標記法將訊號取出其音高標記，經過計算得到新的音高標記位置，將分析訊號疊加在新的標記位置上，因此可調整語音之音高，音高之變化正好可以呈現語音之聲調，因此只需錄較少的合成單元，減少錄製成本與後製處理時間。但是由於PSOLA是一種基頻同步的語音合成技術，因此基頻週期或其起始點的判定誤差將會影響合成的效果，所以有時合成之語音有部份之雜訊(Burst Noise)與較重的鼻音，使合成語音比較不自然且機器合成的味道較重，且其清晰度亦較低，這些是PSOLA主要缺點。

(2)串接合成法(Waveform Concatenation)：

 主要是利用預先錄製好之合成單元(Synthesis Units)，存放在語音資料庫中，使用時再將欲發音的字句斷成字詞，而選用之語音單元將其串接起來，合成出所要的語音。這做法簡化了計算拼接與口音等複雜的計算，特別適合在少量字彙的輸出時使用。通常合成單元要包含所有可能的發音種類，這些預錄的單元可以是音素(Phoneme)、雙音素與音節(Syllable)等。

本系統使用的語音合成法即為串接合成法，預錄的合成單元有單音與多音節音檔。因客語是一種具有聲調之語言，每一單音有不同聲調，需事先錄製，以便作為後續之合成使用。

相對於像過去電腦系統產生較單調與不自然的語音，這種方法直接選用人們所發出的語音作為合成單元，語音訊號之處理比較單純，聽起來較具有一般講話的特性，語音輸出具有自然性，且處理時間較短。但是採用串接合成法時，需要錄製全部所需之語音單元，在建構語音庫時就要花較多的心力。

例如，客語四縣腔670種單音節，其中非入聲單音的四種聲調均需錄製，入聲單音也要錄兩種聲調，約兩千四百多個語音；而海陸腔740種單音節，約三千個語音。此外，常用的客語詞也為數不少，可以個別錄製，上述語音經切音後存入語音庫中，經由語音合成模組選取合成單元，合成時不作音高與時長處理。系統合成之語音不會出現鼻音太重的現象，具有真人發聲之自然與親切效果，因此目前許多語音合成系統都採用串接合成法來合成語音。

本模組所接收的輸入是音檔檔名串列。使用者輸入的國語文句經中文斷詞，再將斷詞好的各個中文詞轉成對應的客語詞，並且得到客語發音，最後找出該拼音對應的合成單元，回傳音檔檔名串列。之後合成模組再把對應的音檔找出，並且依照韻律訊息模組預測的停頓，加入適當的靜音檔。

但是我們發現欲合成的文句中含有入聲字的時候，會因為入聲字的單音檔長度太短，造成合出的語音聽起來急且短促，幾乎無法聽懂，這是必需解決的問題。例：實實在在 (siid**ˊ** siid**ˊ** cai cai)

我們查看了語音庫中的單音音檔，入聲字的音長大約在0.15秒到0.19秒之間，而非入聲字都在0.3秒以上。我們希望讓入聲字發音時不要太急促，又能保留它本身音長較短的特性，於是決定幫入聲字的音檔補足0.29秒，讓它的音長不超過非入聲字，聽起來又不會太急。做法為在選取音檔的時候，判斷該單音是否為入聲字，若是就在該音檔後加上一個短的靜音檔，讓該入聲音檔和靜音檔的長度總合為0.29秒，再輸出音檔內容與其他音檔做串接處理，結果如圖五。如此一來，串接處理完成後的連續語音聽起來會更自然。此時，本系統的語音合成流程如六。

在

在

實

實

圖五：音檔「實實在在」的波型

圖中靜音部分即為插入的靜音檔。此例中「實」的音長為0.19秒，插入的靜音檔長為0.1秒

音檔檔名串列(經過多音字處理)

結束

接合

把入聲字音檔的長度補至特定長度

判斷該音檔是否為入聲字音檔

開始

是

否

圖六：語音合成流程圖

# 第四章 多音字介紹及研究方法

 對於求取多音字該發哪個音的方法，文獻上有許多相關的研究結果，其中包含語言模型[20][21][31]、決策樹[12]、還有組合式策略的方法 [12]，或是粗糙集合理論[22]。我們所使用的方法是資料探勘的方法。

## 4.1 多音字介紹

 在國語中，常常會遇到一些多音字的例子，例如「隊長」的長這時發音做ㄓ ㄤˇ，而「長度」的長則發音為ㄔㄤˊ。而客語中同樣也會發生類似的情況，有的是國語發音和客語發音都含有多音字的現象，例如「感覺」和「睡覺」的覺這個字，國語發音分別為「ㄐㄩㄝˊ」和「ㄐㄧㄠˋ」，客語發音則為「gog˙」和「gau」。有的則是國語發音皆一致，但是客語發音有多音字的情形，例如「屋頂」和「頂天立地」的頂這個字，國語發音皆為「ㄉㄧㄥˇ」，客語發音則分別為「dangˋ」和「dinˋ」。

## 4.2 監督式與非監督式方法

監督式學習[34]是一個機器學習中的技巧，可以由訓練資料中學到或建立一個模型，並依此模型推測新的資料。訓練資料是由輸入物件和預期輸出所組成。函數的輸出可以是一個連續的值（稱為迴歸分析），或是預測一個「分類」。

非監督式學習則不需要預先定義好的分類以及標記好的資料，不針對任何屬性學習及設計模型。「分群」就是屬於非監督式的學習，目的在於具及高相似性的資料，使得群體與群體之間產生較低的相似度。有別於監督式學習網路，無監督式學習網路在學習時並不知道其分群結果是否正確，亦即沒有告訴它何種學習是正確的，當學習完畢並經測試後，也可以將之應用到新的資料上。

資料探勘可從形式不拘的大量資料中，發掘隱藏有用的資訊，經過分析之後得到有用的知識(Knowledge)。一般而言，Data Mining包含下列五項功能：

(1)分類(classification)

(2)推估(estimation)

(3)預測(prediction)

(4)關聯分組(affinity grouping)

(5)同質分組(clustering)

本研究用「分類」功能解決多音字問題，是屬於監督式學習的方法。「分類」是根據已知資料及其分類屬性，建立資料的分類模型，再以此分類模型對欲分類的新對象預估出類別(class)。現在有著各式各樣的分類器，各自都有強項或弱項，分類器的表現很大程度上跟要被分類的資料特性有關， 並沒有某一單一分類器可以在所有給定的問題上都表現最好。

首先準備一份訓練語料，經過程式處理後，當遇到需要處理的多音字，就在每個句子都用人工標上多音字的類型，接著取出這些語料的特徵，送去分類器訓練，得到一個分類模組，然後用此分類模組對測試資料做分類，分類的結果就是我們所要的多音字類型。因此分類流程大概可概括成：尋找影響發音的某一重要變項的因素、了解語料的特徵、最後則是建立分類規則。

 分類器的種類有許多種，我們選擇了常見的貝氏網路(Bayesian network)分類器[27]、C4.5決策樹[30]、CART決策樹[29]以及SVM[32]分類器，並將分類結果做比較。以下章節先簡單介紹各個分類器的原理。

## 4.3 貝氏網路分類器

貝氏網路是以貝氏定理為基礎所發展出來的理論，是一種機率圖型模型，能將特定領域中的不確定性組合成模型，以供推論與診斷之用，利用事件發生的機率來推測未知資料的類別。由於機率的計算是可隨著已知樣本的增加而逐次調整的，在新樣本加入時只需局部調整某些機率值，即可得到新的分類模型，其分類模型的機動性相當高，在資料不斷增加的情況下，可以得到較好的分類效能。

藉由有向非循環圖型中得知一組隨機變數及其n組條件機率分配(Conditional Probability Distributions)的性質。圖形中每個節點會包含相關連的變數的機率值，並且由父點來決定下一個子點的相關機率值，如此可以很簡單的去合併一個新的資料到貝氏網路中，並且維持貝氏網路的資料始終是最新的。

貝氏網路是由有向非循環圖 (Directed Acyclic Graph, DAG, G=(V,E)) 組成，節點表示隨機變數，可以是觀察到的變量、潛在變量、未知參數等。連接兩個節點的箭頭代表此兩個隨機變數是具有因果關係或是非條件獨立的；而節點中變數間若沒有箭頭相互連接一起的情況就稱其隨機變數彼此間為條件獨立。若兩個節點間以一個單箭頭連接在一起，表示其中一個節點是「因(parents)」，另一個是「果(descendants or children)」，兩節點就會產生一個條件機率值。比方說,我們以*Xi*表示第i個節點，而*Xi*的「因」以*Pi*表示，*Xi*的「果」以*Ci*表示。圖七就是一種典型的貝氏網路結構圖，依照先前的定義，我們就可以輕易的從圖七可以得知：

*P1={ψ},C1=*{*X2,X3*}*, P2=*{*X1,X3*} *, C2={ψ}, P3=*{*X1*}*,C3=*{*X2*} (4-1)

X11

0

1

X3

0 1

0.4 0.6

0.7 0.3

X1

0 1

0.4 0.6

X3 X1

0 0

0 1

1 0

1 1

X2

0 1

 0.4 0.6

 0.7 0.3

 0.2 0.8

 0.8 0.2

圖七：貝氏網路與條件機率表

貝氏網路適用在離散型的資料集，它依照P(*Xi* | *Pi*)此條件機率寫出條件機率表，此條件機率表的每一列(row)列出所有可能發生的*Pi*，每一行(column)列出所有可能發生的*Xi*，且任一列的機率總和必為1。寫出條件機率表後就很容易將事情給條理化，且輕易地得知此貝氏網路結構圖中各節點間之因果關係。

基於此理論，貝氏網路分類器就變成一個分類的方法，用來判斷未知類別的資料應該最接近哪一個類別。其目標是希望能透過機率統計的分析，達到最小誤差的一種分類方式。貝氏網路分類器回傳的分類結果為測試資料是各個類型的機率，除了預測最可能的類型，也能分別得知其它類型的機率高低。

## 4.4 C4.5決策樹與CART決策樹

決策樹是一個預測模型，用來處理類別型或連續型變數的分類預測問題，屬於監督式的學習方法，採用自頂端而下的貪婪演算法(Greedy Algorithm)。它代表的是對象屬性與對象值之間的一種映射關係，主幹是類似流程圖的樹架構，每個內部節點代表一個測試屬性，分支代表一個可能的測試輸出結果，最底層的樹葉節點則代表不同分類的類別標記。所以決策樹的主要功能是藉由分類已知的資料來建立一個樹狀結構，並從中歸納出一些規則。回到本研究，我們要分類的多音字可以套用在決策樹模型上，來幫助我們區分出該發音的類型，故可用決策樹方法分類。

CART(Classification and Regression Trees) 演算法是建構決策樹時最常用的演算法之一，1984年布里曼 （L. Brieman）與其同僚發表這種方法，是一種二元(binary)分割的方法。CART以每個節點的動態臨界值做為判斷式，藉由單一輸入的變數函數，在每個節點分隔資料，並建立一個二元決策樹，根據資料的分類數及其屬性並依據Gini規則來決定分割條件。使用 Gini Ratio來衡量指標，如果分散的指標程度很高，表示資料中分佈許多類別，相反的，如果指標程度越低，則代表單一類別的成員居多。一般來說，決策樹分類的正確性有賴於資料來源的多寡，若是透過龐大資料建構的決策樹，其預測和分類結果往往是符合期望的。

 以圖八為例子，假設要分類的是運動的類型。首先觀察有無球類，若沒有則走到水上，若是有則到工具。接下來繼續問題，此時水上這個節點的集合已經可以分類出結果，若是水上就是游泳，否則是跑步。其它節點以此類推，二元決策樹便是以此方法來分類。

工具?

水上?

圓型?

桌上?

球類?

 棒球

 撞球

橄欖球

 籃球

游泳

跑步

Yes

Yes

Yes

Yes

Yes

No

No

No

No

No

圖八：二元決策樹樹狀圖

C4.5則是改良自ID3 (Iterative Dichotomizer 3, J. Ross Quinlan於1979年所提出)演算法，主要是先建構一顆完整的決策樹，再針對每一個內部節點，依使用者定義的預估錯誤率(Predicted Error Rate)來做決策樹修剪的動作。而不同的節點之特徵值離散化結果是不相同的。C4.5與CART之間的第一個差異是CART在每一個節點都呈現二分法，因此產生二元決策樹，而C4.5則在每一個節點產生不同數目的分支。這是因為C4.5對持續性變項的處理方式和CART相當類似，但對類別變項的處理就相當不同，根據類別值來選擇最好的屬性進行區別，屬性選擇指標使用資訊獲利 (Information Gain)。下列是資訊獲利的算法。

期望訊息用於判別中值組：

 (4-2)

其中為任一中的值組被歸類為的機率，可由得之，在分割後，為了達到一致的判別我們需要下列的訊息：

 (4-3)

分割屬性的資訊獲利：

 (4-4)

另外一個差異則是決策樹的修剪，CART使用決策樹的分散度為度量，來標記不同的分支樹，然後以沒有見過的預先分類好的資料（測試組）來測試這些分支樹。相反的，C4.5並不參考其他資料，嘗試以只用訓練資料的情況下來修剪決策樹。

## 4.5 SVM支援向量機

SVM(Support Vector Machine)也是一種可用來做分類或迴歸分析的方法，是根據 Vapnik的 max margin strategy 發展出來的分類器，具有容易使用、高正確率、相對快速的特性，因此常被用於自然語言處理中或需要自動辨別分類研究上。核心函數能將資料映射到更高維的空間而不會增加計算複雜度，而且就算在高維特徵向量空間下還是能產生不錯的效能。

假定我們現在要區別一個平面上兩種不同類別的點，要如何用一條直線把它們分開呢？在平面上可能無法做到，但如果通過某種映射，將這些點映射到其他空間(例如球面)，就有可能在這個空間找到一條直線將其分開。SVM基本上就是這樣的原理，但是更為複雜，因為SVM一般的做法是將所有待分類的點映射到高維空間，然後在高維空間中找到一個能將這些點分開的超平面(Hyperplane)。如圖九，虛線就是可以把資料分為兩類的平面，兩條實線中間的距離叫做邊界(margin)，也就是SVM演算法試著最大化的目標，在虛線兩邊的點稱為支持向量(support vectors)。



圖九：SVM示意圖

SVM對於線性不可分割的情況，通過使用非線性映射演算法，將輸入空間線性不可分割的低維樣本轉化為高維特徵空間，使其線性可分割，高維特徵空間再採用線性演算法，而對樣本的非線性特徵進行線性分析成為可能。此外它植基於結構風險最小化理論之上，在特徵空間中建構最優分割超平面，使得學習器得到全域最優化，並且在整個樣本空間的期望風險以某個概率滿足一定上界。

# 第五章 語音合成系統實作

 我們建置了一個線上客語語音合成系統，開發環境與使用之軟體如表所示。

表十一：系統開發環境

|  |  |
| --- | --- |
| **Software** | **Version** |
| Operating System: | Windows XP |
| Web server | Apache Server v2.5.10 |
| Database | MySQL 5.0.51b |
| DBMS platform | phpMyAdmin 2.10.3 |
| Web Programming: | PHP 5,Java |
| classifier | Weka 3.7  |

## 5.1 系統架構

### 5.1.1 **運作流程**

 本系統是使用者以國語文句輸入，經由中文詞典斷詞後，預估出詞間的停頓類型，插入停頓之後，搜尋國客語對照詞典取得每個詞的客語讀音，然後去找是否有對應的客語詞音檔並使用之，若無客語詞音檔，再從客語的單音語料庫中選取適當的合成單元，此外再判斷句子之內是否含有需處理的多音字，若有包含的話則交由多音字分類器處理，判斷此時最適合的發音，最後產生使用者所需的客語合成語音，介面如圖十。



圖十：線上客語語音合成系統操作介面

### 5.1.2 資料庫

Database Schema如圖十一，haka2資料表是國客語對照詞典，allspell2資料表是客語四縣腔單字讀音表，allsealand2資料表是客語海陸腔單字讀音表，multiwavname資料表是四縣腔客語詞與其音檔檔名的對照表，sl\_multiwavname資料表是海陸腔客語詞與其音檔檔名的對照表。

haka2

multiwavname

fourhaka

後生人

fourspell

sealandspell

chinese

english

etc

source

hakaexample

chiexample

sealand

seaexample

後生人愛煞猛打拼做頭路。

年輕人要多打拼。

heu¯ sangˊ nginˇ

heu├ sangˋ ngin¯

年輕人

youth

客委會

後生人

後生人愛煞猛打拼做頭路。

hakkaword

Fname

後生人

後生人

sl\_multiwavname

hakkaword

Fname

後生人

後生人

allspell2

spell

haka

source

客委會

年

ngian3

allsealand2

spell

haka

source

ngien

年

客委會

圖十一：Database schema

## 5.2 文句分析

本系統支援的輸入為國語中文文句。假設使用者輸入一句中文文句「中興大學建國百年畢業典禮即將於六月舉行」，這個文句先送入中文斷詞器做斷詞與詞性標記，斷完詞的結果為：「中興大學 建國 百年 畢業典禮 即將 於 六月 舉行」，共有8個中文詞，詞性串列為：「Nc,Nc,Nd,Na,D,P,Nd,Vc」。詞與詞性的對照如(E3)：

中興大學 建國 百年 畢業典禮 即將 於 六月 舉行 (E3)

 Nc Nc Nd Na D P Nd Vc

接下來把斷出的中文詞各自到國客語對照詞典查出對應的客語詞並做連音變調，像這句中的「即將」所對應的客語詞「臨將」，此時系統會列出所有找到的客語詞。若在國客語對照詞典中找不到詞，可能的原因有：

(1) 國語斷詞的輸出詞太長，客語詞典中只有此詞的部分資訊。舉個簡單的例子，本系統的中文斷詞器會把「中興大學」當成一個詞輸出，但是在客語詞典中，「中興」與「大學」卻是兩個詞，造成找不到正確的對應客語詞，且無法使用客語詞的音檔。當這種情況發生時，我們把長度三個字以上的詞取出，使用國客語對照詞典中的國語詞欄位，用長詞優先斷詞，把此詞拆成較短的詞，即可把中文詞「中興大學 」轉成對應的客語詞「中興」「大學」。

(2) 客語詞典中沒有此國語詞的資訊。此時只能把此國語詞拆成單字，去查找單字拼音表，找出單字的讀音。

(3)

上述過程可以簡化為十二的流程圖。找出客語詞之後，把客語詞列出，如：「中興大學 建國 百年 畢業典禮 臨將 於 六月 舉行」，並且同時列出這些客語詞的客語讀音，其中「中興大學」的「中」已經變調過了，如(E4)：

中 興 大 學 建 國 百 年 畢 業 典 禮 臨 將 於

zungˇ hinˊ tai hog gien gued bagˋ ngienˇ bidˋ ngiab dienˋ liˊ limˇ ziongˊ iˇ

 六 月 舉 行

liugˋ ngied giˋ hi (E4)

No

中文詞

是否有對應客語詞

當成單字查客語讀音

Yes

輸出讀音

拆成短詞

No

Yes

連音變調

是否有對應客語詞

是否含有須處理的多音字

Yes

No

圖十二：從中文詞查找出客語讀音流程

### 5.2.1實驗的方法

 進行多音字實驗前需要標記訓練語料，我們所使用的語料是Chinese GigaWord語料庫[33]，來源是由許多個單位收集新聞語料而來，並且由Linguistic Data Consortium發行，授權讓人取得使用。目前我們所使用的版本為2007年所發行的第三版，其新聞語料主要有以下三個來源：Central News Agency of Taiwan、 Xinhua News Agency of Beijing和Zaobao Newspaper of Singapore。

 我們所使用的語料要以繁體中文為主，所以選擇從台灣中央通訊社(Central News Agency)所收集而來的語料。在此份語料中，含有約五千多萬行句子，包含詞性標記的檔案大小約6GB。語料庫中的文章都是斷好詞，並且標上詞性標記的語料，其詞性標記是由台灣中央研究院所標記。

 而我們也有採用中文結構樹資料庫(TreeBank)的資訊[23]，其是由86年起由中央研究院詞庫小組(CKIP)從中央研究院現代漢語平衡語料庫(Sinica Corpus)中抽取句子，經由電腦自動剖析成結構樹，再加以人工修正、檢驗後所得的成果。目前版本是3.0版，包含了6個檔案，61,087個中文樹圖，361,834個詞。中文句子的語法結構表達採取中心語主導原則(Head-Driven Principle)，剖析中文句子時，詞組類型由中心語決定，並且參照中心語和其他成分所記載的語法和語意訊息，表達出句子中詞和詞之間的語法結構和語意角色關係。下面圖十三是TreeBank的舉例[33]：

 V

大量Neqa

 而Caa

豐盛VH

 VP

圖十三：TreeBank範例示意圖

　　圖中「大量」詞性為Neqa，是表示物品數量的定詞；「豐盛」詞性則為一般狀態不及物動詞VH。此時「大量而豐盛」的上層結構為V，而V的最上層結構為VP。

然後我們取出部分的GigaWord來標記訓練語料與測試語料，我們所處理的多音字總共有：行、調、重、差、易、口這六個，其中行、調、重各有三種發音，差、易、口則各有兩種發音，發音類型和語料數目如表十二所示：

表十二：多音字的類型、發音及語料數目

|  |  |  |
| --- | --- | --- |
| **多音字** | **四縣腔發音類型** | **海陸腔發音類型** |
| **行** | hangˇ、hongˇ、hen | hang、hong、hen＋ |
| **調** | tiauˇ、tiau、diau | tiau、tiau＋、diauˇ |
| **重** | cung、ciungˇ、cungˊ | chung＋、chung、chungˋ |
| **差** | caˊ、caiˊ | caˋ、caiˋ |
| **易** | id˚、i | rhid˚、rhiˇ |
| **口** | kieuˋ、heuˋ | kieuˊ、heuˊ |
|  |
|  | **訓練語料** | **測試語料** | **總數** |
| **行** | 4961筆 | 895筆 | 5856筆 |
| **調** | 2426筆 | 389筆 | 2815筆 |
| **重** | 4336筆 | 383筆 | 4719筆 |
| **差** | 730筆 | 186筆 | 916筆 |
| **易** | 546筆 | 237筆 | 783筆 |
| **口** | 3201筆 | 387筆 | 3588筆 |

從多音字的目標詞的詞性組合開始實驗，且慢慢增加特徵數，從前一詞、後一詞與目標詞的詞性組合，前二詞、後二詞與目標詞的詞性組合，前三詞、後三詞與目標詞的詞長詞性組合…等。總共實驗了以下特徵：目標詞與後一詞可以生成的詞性、目標詞與前一詞可以生成的詞性、目標詞與前一詞跟後一詞可以生成的詞性、目標詞前第一詞詞性、目標詞前第二詞詞性、目標詞前第三詞詞性、目標詞後第一詞詞性、目標詞後第二詞詞性、目標詞後第三詞詞性、目標詞詞性。以例子「我 今天 早上 去 學校 差點 遲到」解說，在「差點」這詞裡面含有多音字「差」：

我 今天 早上 去 學校 差點 遲到

Nh Nd Nd VCL Nc D VH

以目標詞「差點」為例，記錄下需要的特徵，需注意的是當遇到句首沒有詞的時候，則詞性標記為「Begin」；當遇到句尾沒有詞的時候，則詞性標記為「End」：

該目標詞前第一詞詞性： Nc

該目標詞前第二詞詞性： VCL

該目標詞前第三詞詞性：Nd

該目標詞後第一詞詞性：VH

該目標詞後第二詞詞性：END

該目標詞後第三詞詞性：END

目標詞與前一詞可以生成的語法類別(取TreeBank頻率最高者)：NULL

目標詞與後一詞可以生成的語法類別(取TreeBank頻率最高者)：VP

目標詞與前一詞跟後一詞可以生成的語法類別(取TreeBank頻率最高者)：NULL

目標詞詞性：D

### 5.2.2實驗的結果

我們依序將特徵從目標詞性開始，慢慢增加特徵的數目，內部測試、外部測試、訓練語料與測試語料一起測試(全部當訓練)、訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料)這四種測試是最終的實驗結果，結果發現，不同多音字分類器正確率最高的特徵組合幾乎皆是採用C4.5演算法，且選取特徵為「前三後三詞性、目標詞與後一詞可以生成的語法類別、目標詞與前一詞可以生成的語法類別、目標詞與前一詞跟後一詞可以生成的語法類別、目標詞性」的時候；只有在「行」這個字時，是CART演算法的正確率較佳，不過跟C4.5演算法的正確率也相去不遠。

 在修改特徵組合的過程中，可以看出目標詞性，對於正確率有相當大的影響，而再加上目標詞前後的詞性以及和前後詞可以生成的詞性，能使預測的正確率上升。整體而言，預測多音字的正確率都有80%以上。但是要注意的是，如果當某多音字的某個發音，可以選出的訓練語料太少時，將可能因為資料稀疏，使得訓練出來的分類器無法有效分類未知的語料，導致分類器的正確率降低。

下列表十三至表三十二為這六個多音字實驗的數據，其中表格上層的I.內部測試、II.外部測試、III.訓練語料與測試語料一起測試(全部當訓練) 、IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料)這四個實驗使用特徵為「目標詞前第一詞詞性、目標詞前第二詞詞性、目標詞前第三詞詞性、目標詞後第一詞詞性、目標詞後第二詞詞性、目標詞後第三詞詞性、目標詞與後一詞可以生成的語法類別、目標詞與前一詞可以生成的語法類別、目標詞與前一詞跟後一詞可以生成的語法類別、目標詞性」。

表十三：「行」的各種特徵組合的正確率(門檻值74%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 91.3122 | 88.9538 | 90.7277 | 99.3751 |
| II.外部測試 | 89.5833 | 87.9952 | 90.4713 | 80.1742 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 90.9665 | 88.9686 | 92.9303 | 99.2828 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 90.4571 | 88.9503 | 90.6077 | 78.7042 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 82.7357 | 78.4665 | 87.0048 | 97.763 |
| A+B+C、前三後三詞性(套用IV) | 80.0603 | 76.9965 | 79.8091 | 77.7499 |
| C、前三後三詞性、目標詞性(套用III) | 90.9665 | 90.0273 | 92.7937 | 99.1974 |
| C、前三後三詞性、目標詞性(套用IV) | 90.4571 | 89.3521 | 90.6077 | 78.8046 |
| 前三後三詞性、目標詞性(套用III) | 89.515 | 90.181 | 92.7083 | 99.1633 |
| 前三後三詞性、目標詞性(套用IV) | 90.4571 | 89.6032 | 91.0095 | 78.8046 |
| 前二後二詞性、目標詞性(套用III) | 89.515 | 90.3176 | 92.0253 | 97.319 |
| 前二後二詞性、目標詞性(套用IV) | 90.4571 | 90.1055 | 91.1602 | 82.2702 |
| 前一後一詞性、目標詞性(套用III) | 89.515 | 90.1639 | 91.8374 | 92.4522 |
| 前一後一詞性、目標詞性(套用IV) | 90.4571 | 90.9091 | 90.005 | 87.7951 |
| 目標詞性(套用III) | 89.515 | 89.515 | 89.515 | 89.515 |
| 目標詞性(套用IV) | 90.4571 | 90.4571 | 90.4571 | 90.4571 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表十四：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「行」的confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
| **實際 預測** | **hangˇ** | **hen** | **hongˇ** |
| **hangˇ** | 1412 | 31 | 31 |
| **hen** | 59 | 28 | 0 |
| **hongˇ** | 63 | 3 | 364 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表十五：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「行」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **hangˇ** | 0.92 | 0.958 | 0.939 |
| **hen** | 0.452 | 0.322 | 0.376 |
| **hongˇ** | 0.922 | 0.847 | 0.882 |
| **加權平均** | 0.9 | 0.906 | 0.902 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表十六：CART決策樹於前二後二詞性、目標詞性，多音字「行」的confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
| **實際 預測** | **hangˇ** | **Hen** | **hongˇ** |
| **hangˇ** | 1407 | 23 | 44 |
| **Hen** | 55 | 32 | 0 |
| **hongˇ** | 50 | 4 | 376 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表十七：CART決策樹於前二後二詞性、目標詞性，多音字「行」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **hangˇ** | 0.931 | 0.955 | 0.942 |
| **Hen** | 0.542 | 0.368 | 0.438 |
| **hongˇ** | 0.895 | 0.874 | 0.885 |
| **加權平均** | 0.906 | 0.912 | 0.908 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表十八：「調」的各種特徵組合的正確率(門檻值60%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 86.8096 | 83.4295 | 88.8293 | 99.3405 |
| II.外部測試 | 86.6075 | 86.6945 | 86.3233 | 73.1439 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 88.881 | 85.7889 | 90.9414 | 99.2895 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 86.2069 | 85.5062 | 85.8934 | 70.8464 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 79.5737 | 74.1030 | 85.8615 | 98.0462 |
| A+B+C、前三後三詞性(套用IV) | 72.5183 | 72.9363 | 75.4441 | 69.5925 |
| C、前三後三詞性、目標詞性(套用III) | 88.8810 | 86.2877 | 91.0835 | 99.2895 |
| C、前三後三詞性、目標詞性(套用IV) | 86.2069 | 84.117 | 84.6395 | 70.4284 |
| 前三後三詞性、目標詞性(套用III) | 88.3481 | 85.6838 | 91.4032 | 99.2540 |
| 前三後三詞性、目標詞性(套用IV) | 86.1024 | 83.6991 | 85.0575 | 70.3239 |
| 前二後二詞性、目標詞性(套用III) | 87.389 | 84.7247 | 91.4032 | 96.9499 |
| 前二後二詞性、目標詞性(套用IV) | 86.1024 | 83.908 | 86.2069 | 73.9812 |
| 前一後一詞性、目標詞性(套用III) | 87.9218 | 84.6181 | 88.8009 | 90.4085 |
| 前一後一詞性、目標詞性(套用IV) | 86.1024 | 84.117 | 85.7889 | 82.1317 |
| 目標詞性(套用III) | 77.6199 | 77.6199 | 77.6199 | 77.6199 |
| 目標詞性(套用IV) | 78.0564 | 78.2654 | 78.2654 | 78.0564 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表十九：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「調」的confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
| **實際 預測** | **tiauˇ** | **tiau** | **diau** |
| **tiauˇ** | 580 | 21 | 2 |
| **Tiau** | 32 | 96 | 0 |
| **Diau** | 77 | 0 | 149 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「調」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **tiauˇ** | 0.842 | 0.962 | 0.898 |
| **tiau** | 0.821 | 0.75 | 0.784 |
| **diau** | 0.987 | 0.659 | 0.79 |
| **加權平均** | 0.873 | 0.862 | 0.857 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十一：「重」的各種特徵組合的正確率(門檻值48%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 82.0341 | 73.3164 | 85.7934 | 96.8865 |
| II.外部測試 | 77.6483 | 70.2542 | 77.3305 | 61.0169 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 82.7119 | 73.178 | 81.9915 | 96.9703 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 77.4455 | 70.0935 | 74.8287 | 59.7508 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 75.2331 | 63.072 | 82.0763 | 95.4661 |
| A+B+C、前三後三詞性(套用IV) | 66.5421 | 60.2492 | 66.3551 | 59.6262 |
| C、前三後三詞性、目標詞性(套用III) | 82.4364 | 75.0847 | 83.2415 | 96.8008 |
| C、前三後三詞性、目標詞性(套用IV) | 77.5701 | 71.5265 | 75.8255 | 60.1246 |
| 前三後三詞性、目標詞性(套用III) | 82.3729 | 75.0424 | 86.3136 | 97.0339 |
| 前三後三詞性、目標詞性(套用IV) | 77.2586 | 71.7134 | 74.6417 | 59.9377 |
| 前二後二詞性、目標詞性(套用III) | 81.589 | 73.7076 | 84.3008 | 92.4788 |
| 前二後二詞性、目標詞性(套用IV) | 77.134 | 71.9003 | 74.8287 | 64.6106 |
| 前一後一詞性、目標詞性(套用III) | 78.8136 | 72.7754 | 79.6398 | 82.1610 |
| 前一後一詞性、目標詞性(套用IV) | 75.7009 | 70.8411 | 75.1402 | 73.1464 |
| 目標詞性(套用III) | 65.9534 | 65.9534 | 65.9534 | 65.9534 |
| 目標詞性(套用IV) | 66.3551 | 66.3551 | 66.4174 | 66.3551 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表二十二：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「重」的confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
| **實際 預測** | **cung** | **ciungˇ** | **cungˊ** |
| **cung** | 705 | 34 | 61 |
| **ciungˇ** | 57 | 336 | 18 |
| **cungˊ** | 164 | 28 | 202 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十三：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「重」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **cung** | 0.761 | 0.881 | 0.817 |
| **ciungˇ** | 0.844 | 0.818 | 0.831 |
| **cungˊ** | 0.719 | 0.513. | 0.599 |
| **加權平均** | 0.772 | 0.774 | 0.767 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十四：「差」的各種特徵組合的正確率(門檻值75%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 94.2466 | 96.3014 | 97.6712 | 99.863 |
| II.外部測試 | 94.214 | 93.4498 | 95.3057 | 78.7118 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 94.214 | 95.6332 | 97.7074 | 99.8908 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 94.2122 | 91.3183 | 93.2476 | 76.8489 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 91.9214 | 88.1004 | 96.3974 | 99.5633 |
| A+B+C、前三後三詞性(套用IV) | 86.8167 | 82.6367 | 80.3859 | 76.8489 |
| C、前三後三詞性、目標詞性(套用III) | 94.214 | 95.3057 | 97.7074 | 99.8908 |
| C、前三後三詞性、目標詞性(套用IV) | 94.2122 | 90.9968 | 93.2476 | 77.1704 |
| 前三後三詞性、目標詞性(套用III) | 94.214 | 94.7598 | 97.7074 | 99.8908 |
| 前三後三詞性、目標詞性(套用IV) | 94.2122 | 90.6752 | 93.2476 | 77.1704 |
| 前二後二詞性、目標詞性(套用III) | 94.214 | 94.7598 | 98.4716 | 99.0175 |
| 前二後二詞性、目標詞性(套用IV) | 94.2122 | 92.6045 | 93.2476 | 79.2412 |
| 前一後一詞性、目標詞性(套用III) | 94.214 | 94.3231 | 94.214 | 95.8515 |
| 前一後一詞性、目標詞性(套用IV) | 94.2122 | 93.5691 | 94.2122 | 90.9968 |
| 目標詞性(套用III) | 94.214 | 94.214 | 94.214 | 94.214 |
| 目標詞性(套用IV) | 94.2122 | 94.2122 | 94.2122 | 94.2122 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表二十五：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「差」的confusion matrix

|  |  |  |
| --- | --- | --- |
| **實際 預測** | **caˊ** | **caiˊ** |
| **caˊ** | 233 | 0 |
| **caiˊ** | 18 | 60 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十六：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「差」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **caˊ** | 0.928 | 1 | 0.963 |
| **caiˊ** | 1 | 0.769 | 0.87 |
| **加權平均** | 0.946 | 0.942 | 0.939 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十七：「易」的各種特徵組合的正確率(門檻值58%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 100 | 98.9011 | 100 | 100 |
| II.外部測試 | 100 | 97.4457 | 100 | 65.1341 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 100 | 98.4962 | 100 | 100 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 100 | 98.7229 | 100 | 60.301 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 88.1226 | 86.4623 | 97.5734 | 99.7446 |
| A+B+C、前三後三詞性(套用IV) | 84.9624 | 85.3383 | 88.7218 | 61.6541 |
| C、前三後三詞性、目標詞性(套用III) | 100 | 99.4891 | 100 | 100 |
| C、前三後三詞性、目標詞性(套用IV) | 100 | 98.8722 | 100 | 61.6541 |
| 前三後三詞性、目標詞性(套用III) | 100 | 99.7446 | 100 | 100 |
| 前三後三詞性、目標詞性(套用IV) | 100 | 99.2481 | 100 | 61.2782 |
| 前二後二詞性、目標詞性(套用III) | 100 | 99.8723 | 100 | 100 |
| 前二後二詞性、目標詞性(套用IV) | 100 | 99.6241 | 100 | 70.6767 |
| 前一後一詞性、目標詞性(套用III) | 100 | 100 | 100 | 100 |
| 前一後一詞性、目標詞性(套用IV) | 100 | 100 | 100 | 93.985 |
| 目標詞性(套用III) | 100 | 100 | 100 | 100 |
| 目標詞性(套用IV) | 100 | 100 | 100 | 100 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表二十八：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「易」的confusion matrix

|  |  |  |
| --- | --- | --- |
| **實際 預測** | **id**˚ | **i** |
| **id**˚ | 116 | 0 |
| **I** | 0 | 150 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表二十九：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「易」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **id**˚ | 1 | 1 | 1 |
| **I** | 1 | 1 | 1 |
| **加權平均** | 1 | 1 | 1 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表三十：「口」的各種特徵組合的正確率(門檻值80%)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **C4.5** | **貝氏網路** | **CART** | **SVM** |
| I.內部測試 | 88.1912 | 84.911 | 89.0034 | 98.5629 |
| II.外部測試 | 86.4548 | 82.5251 | 85.563 | 83.4448 |
| III.訓練語料與測試語料一起測試(全部當訓練) | 88.5452 | 84.9777 | 91.1929 | 98.6343 |
| IV.訓練語料與測試語料一起測試(取66%當訓練語料，其餘當測試語料) | 85.8197 | 83.1967 | 84.1803 | 83.5246 |
| **特徵類型改變比較** |
| A+B+C、前三後三詞性(套用III) | 83.6678 | 80.7971 | 83.9744 | 97.9933 |
| A+B+C、前三後三詞性(套用IV) | 80.9016 | 79.3443 | 79.918 | 83.1967 |
| C、前三後三詞性、目標詞性(套用III) | 88.573 | 86.0089 | 91.1929 | 98.5229 |
| C、前三後三詞性、目標詞性(套用IV) | 85.8197 | 83.8525 | 84.1803 | 83.2787 |
| 前三後三詞性、目標詞性(套用III) | 88.6009 | 86.1761 | 91.1929 | 98.495 |
| 前三後三詞性、目標詞性(套用IV) | 85.8197 | 84.6721 | 84.1803 | 83.2787 |
| 前二後二詞性、目標詞性(套用III) | 87.1795 | 86.0089 | 90.8863 | 96.1817 |
| 前二後二詞性、目標詞性(套用IV) | 85.7377 | 85.1639 | 84.1803 | 84.0984 |
| 前一後一詞性、目標詞性(套用III) | 86.9844 | 84.922 | 89.8272 | 89.7157 |
| 前一後一詞性、目標詞性(套用IV) | 85.9016 | 83.2787 | 85.9016 | 85.3279 |
| 目標詞性(套用III) | 80.7971 | 80.7971 | 80.7971 | 80.7971 |
| 目標詞性(套用IV) | 80.9016 | 82.0492 | 80.9016 | 82.0492 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別 門檻值:在取出的語料中發音最多個數佔總發音個數之百分比)(單位：%)

表三十一：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「口」的confusion matrix

|  |  |  |
| --- | --- | --- |
| **實際 預測** | **kieuˋ** | **heuˋ** |
| **kieuˋ** | 915 | 72 |
| **heuˋ** | 101 | 132 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

表三十二：C4.5決策樹於前三後三詞性、A、B、C、目標詞性，多音字「口」的各發音正確率

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F-Measure** |
| **kieuˋ** | 0.901 | 0.927 | 0.914 |
| **heuˋ** | 0.647 | 0.567 | 0.604 |
| **加權平均** | 0.852 | 0.858 | 0.855 |

(A:目標詞與後一詞可以生成的語法類別 B:目標詞與前一詞可以生成的語法類別 C:目標詞與前一詞跟後一詞可以生成的語法類別)

## 5.3 韻律訊息

 有了句子的讀音，即可找出對應音檔，合成出連續語音。經過韻律訊息模組調整韻律，使合成的連續語音更自然好聽。因此把輸入文句做完斷詞與國語詞轉客語詞之後，要對這個階段性的結果抽取特徵，用貝氏網路分類器預估停頓[7]。

 我們認為停頓的長短類型，受到該停頓前後的詞影響很大。故利用文句分析模組處理後的結果，得到詞長與詞性的組合，希望能從中取得有用的資訊。把該停頓的前後詞長與前後詞性都當成特徵，並設法找出正確率最高的特徵組合。

### 5.3.1實驗的方法

 進行實驗前需要標記訓練語料，這些語料來自我們詞典內部的國語例句欄位，依來源不同，包括客委會發行的「客語能力認證基本詞彙」中的部分例句等。這些國語句子先經過文句分析模組斷詞與標註詞性，並且經過國語詞轉客語的程序將它們轉成客語文句。實驗時給定已斷詞的客語文句，用人工在斷詞的位置標記停頓的類型，一共標了1670句，把其中一些斷詞與詞性錯誤和句子內容不符需求的部分去掉，剩下1200句，共有10196個停頓，其中80%當訓練資料，20%當測試資料，利用分類軟體weka中的貝氏網路分類器與CART分類器功能來分類。

從停頓的前後詞的詞長與詞性組合開始實驗，且慢慢增加特徵數，從前二詞與後一詞的詞長詞性組合、前一詞後兩詞的詞長詞性組合、前兩詞後兩詞的詞長詞性組合…，直到正確率沒有增加為止。

我們總共實驗了以下特徵：該停頓前後詞的詞長詞性、至該停頓為止的韻律詞長度、前一個停頓的類型、整句的字數、整句的詞數。以例子「中興大學 建國 百年 畢業典禮 即將 於 六月 舉行」解說，把此句加上停頓，以M代表中停頓(medium break)、m代表小停頓(minor break)，n代表無停頓：

中興大學m建國n百年n畢業典禮M臨將n於m六月n 舉行。

以「於」與「六月」中間的停頓m為例：

該停頓前一詞的詞長：1

該停頓後一詞的詞長：2

該停頓前一詞的詞性：P

該停頓後一詞的詞性：Nd

至該停頓為止的韻律詞長度：3，因「臨將於」為同一韻律詞。

前一個停頓的類型：n無停頓

整句的字數：19

整句的詞數：8

觀察所有實驗結果，發現貝氏網路分類器在特徵為「前三後三詞長詞性、韻律詞長、前停頓類型、句字數、句詞數」時，預估中停頓有最佳的分類表現，總正確率為77.538%，雖然不是所有特徵組合中最高的，但與已相去不遠，而且還有最好的中停頓預測正確率。在修改特徵組合的過程中，可以看出詞長、詞類以及韻律詞的長度，對於正確率的提升有相當大的影響，而再加上句子包含的詞數與字數，能使預測中停頓的正確率上升。整體而言，預測中停頓的正確率仍不佳。我們認為主因是人工標記時，標出的中停頓太少，導致資料稀疏，難以用分類獲得正確的結果。

## 5.4語音合成與聽覺測試

 經過文句分析與韻律訊息的處理之後，我們就可以進行語音合成，語音結果為使用者輸入國語文句所對應的客語文句語音。

### 5.4.1 語音合成

語音合成系統有了讀音串列與停頓訊息之後，便開始進行語音合成。在挑選音檔時，發現某個客語詞有專屬的詞音檔，便會使用此詞音檔，若沒有詞音檔時，先經過多音字分類器處理，看是否還有須處理的多音字，若沒有最後才用單音合成。至於句中的停頓，我們採用固定停頓的長度，為音檔加入適當的停頓。句子中有兩種停頓，分別是小停頓和中停頓，句子間則有標點符號的停頓，如

表三十三所示。

表三十三：各種標記的停頓長度

|  |  |
| --- | --- |
| **停頓類型** | **停頓長度(ms)** |
| 小停頓(minor break) | 200 |
| 中停頓(medium break) | 350 |
| 逗號(，) | 550 |
| 句號(○) | 650 |
| 問號(？) | 650 |
| 驚嘆號(！) | 650 |
| 分號(；) | 600 |
| 頓號(、) | 400 |
| 冒號(：) | 450 |

 上述的固定停頓值，是參考[10]所錄製訓練的語料庫統計中得到。又因為本系統的語速較慢，故把停頓的長度也調整，比國語更長些，合成出的語音可聽出有加入停頓處理。

### 5.4.2 聽覺測試與多音字聽覺測試

本節我們對最後TTS合成語音的效果來測試，目前並沒有一個客觀且有效的方法來評斷不同的合成結果優劣。我們採用測量MOS(Mean Opinion Score)來評量合成自然度。總共有11人參與測試。

測試方法是播放語音檔，請人當受測者，聽完音檔後打上評分，分數分五個等級：1分：極差(unsatisfactory)，4分：差(poor)，6分：普通(fair)，8分：好(good)，10分：非常好(excellent)。受測者根據自己聽到的感覺，打上分數。測試內容共有10句客語文句，分別播放四縣腔與海陸腔的合成結果音檔。結果如圖十四：

圖十四：自然度測試結果

圖十四中粗糙面灰色長條圖為四縣腔的合成語音測試平均分數，淺粉色長條圖為海陸腔的合成語音測試平均分數。X軸為受測者編號，Y軸為分數。

四縣腔的測試平均分數7.97分，海陸腔的測試平均分數則為7.81分。從此結果可看出，用本論文的方法合成語音具有不錯的自然度。

我們再對這六個不同的多音字，從GigaWord中各取出2個句子，共有12句，並使用本系統合成客語語音，給11人測試。 然後測試重點主要是多音字處理前後的差異性，無差別的話是代表說剛好在處理前選到的多音字就是正確的，跟處理後的是一樣的結果。

圖十五：多音字測試結果

# 第六章 結論

本論文主要探討網際網路上建置客語語音合成系統，本系統包含有三個模組：文句分析、律韻訊息分析與語音合成，並建立相關之語料庫與語音庫。我們錄製客語單字之基本語音合成單元與不同時長之靜音檔，配合韻律參數作為合成之用。為改善所錄製合成單元之能量差異，對所有單元進行能量正規化處理；並處理有多音字的客語詞，接著找出客語詞對應的音檔，並且加入適當的停頓，以提升輸出的語音品質。

語音合成中的串接法合出的客語連續語音，會在遇到入聲字時產生發音急促的現象，本論文將入聲字音檔的長度補足至特定長度，以解決此問題。此外當遇到含有「行」、「調」、「重」、「差」、「易」、「口」這幾個多音字的詞且詞典無此詞的音檔時，則經由多音字分類器模組，根據輸入文句的特徵，選出一個此時發音機率最高的發音，而回傳給系統的語音合成模組。

 而系統由11位熟悉客語之人士進行10分等級之聽測實驗，進行合成語音品質之自然度分析。聽測實驗之統計結果顯示，合成語音的自然度分數已達7.9分，證明合成語音具有自然度。

目前實作之客語語音合成系統已可供線上使用，透過使用者自行輸入中文文句，合成出對應客語語音。而且只要連上網路，就可線上立即聽取並學習，每一詞彙均標示出拼音，以便於學習。我們所建置的系統符合數位學習的精神，達到幫助使用者學習客語的目標，期望能對推廣客語貢獻一份心力。

關於客語的TTS系統，我們認為未來可以朝下列主題繼續研究：

 (1) 擴充詞典：目前使用的國客語詞典只有約三萬詞，若能增加詞彙數，便可提升國語詞轉客語詞的正確率。未來若建置客語有聲詞典，詞彙內容也會更豐富。

(2) 擴充多音字個數：目前處理的多音字有六個，可視需要繼續新增多音字，讓語音合成系統遇到多音字的時候能夠適當地處理。

(3) 其他客語次方言：本論文已完成客語四縣腔與海陸腔的語音合成系統，之後將可以視需要增加其它客語腔調的研究工作，例如大埔腔或饒平腔等，希望建置出多腔調的客語語音合成系統。

# 參考文獻

1. 鍾榮富，“台灣客家語音導論”，五南圖書出版股份有限公司，2004。
2. 龔萬灶，“客話實用手冊”，龔萬灶，2007。
3. “客語能力認證基本詞彙–初級(四縣腔)”，行政院客家委員會，2009。
4. “客語能力認證基本詞彙–中高級(四縣腔)”，行政院客家委員會，2009。
5. 客語能力認證基本詞彙–初級(海陸腔)”，行政院客家委員會，2009。
6. 林東逸，“客語文句翻語音系統之實作”，交通大學電信工程學系碩士論文，2006。
7. 吳俊毅，“線上客語語音合成系統中產生韻律訊息之研究”，中興大學資訊科學與工程學系研究所碩士論文，2010。
8. 黃豐隆，“線上國客雙語有聲詞典建置之研究”，全國計算機會議(NCS-2009)，台北，台灣，2009。
9. 黃豐隆，“線上語音合成系統之研製與應用,”全國電信研討會 (NST-2008),雲林,台灣, 2008。
10. 潘能煌，“中文文轉音系統的韻律預估及其改進”，國立中興大學應用數學系博士論文， 2004。
11. 蔡育和，“中文文轉音系統中韻律階層的求取”，中興大學資訊科學學系碩士論文，2005。
12. 林金玉， “中文轉台語文轉音系統中一詞多音之預測”， 國立中興大學資訊科學與工程學系研究所碩士論文， 2008。
13. 台文/華文線頂辭典<http://203.64.42.21/iug/Ungian/soannteng/chil/Taihoa.asp>
14. 行政院客委會 客語能力認證網站<http://api1.game.hakka.gov.tw/hkaword.php>
15. 臺灣客家語拼音方案<http://www.edu.tw/mandr/bulletin.aspx?bulletin_sn=3824>
16. 臺灣客家語常用詞詞典網路版 <http://hakka.dict.edu.tw/>
17. 臺灣話語音漢字辭典及外語辭典<http://www.edutech.org.tw/DICT/All-Dictionaries.htm>
18. 聯合大學線上國客語有聲詞典<http://203.64.183.226/public2/hakka-edu/hakka-index.htm>
19. 賴亦傑，“應用多詞及多詞性語言模型的中文斷詞及詞性標記方法”，中興大學資訊科學與工程學系碩士論文，2011。
20. 張子榮、 初敏， “解決多音詞字-音轉換的一種統計學習方法”， 中文信息學報， 2002， 第3期， pp.39-45.
21. 蔡文鴻， “語言模型訓練與調適技術於中文大詞彙連續語音辨識之初步研究”，國立臺灣師範大學資訊工程研究所碩士論文，2005。
22. 廖振淵， “利用粗糙集理論解決中文轉台語文轉音系統中一詞多音問題”，中興大學資訊科學與工程學系研究所碩士論文，2010。
23. 中研院中文詞知識庫小組<http://godel.iis.sinica.edu.tw/CKIP/treebank.htm>
24. 李雪貞，“客語語音合成之初步研究”，台灣科技大學資訊工程系研究所碩士論文，2002。
25. 蔡昀庭，“基於隱藏式馬可夫模型之中文語音合成系統”，國立清華大學電機工程研究所碩士論文，2007。
26. 高瑜璟，“數位學習-學習的新趨勢”，高師大資訊教育研究所。<http://www.nhu.edu.tw/~society/e-j/57/57-22.htm>
27. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann publishers, 1999.
28. Min Chu, Hu Peng, Hong-yun Yang and E.Chang, " Selecting Non-Uniform Units from A Very Large Corpus for Concatenative Speech Synthesizer ", Proceedings of ICASSP 2001, IEEE, Volume 2, pp.785 - 788, Salt Lake City.
29. Classification and Regression Trees (C&RT)，<http://www.statsoft.com/textbook/classification-and-regression-trees/>
30. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
31. Frederick Jelinek, "Statistical Methods for Speech Recognition " ,The MIT Press, Cambridge Massachusetts, 1997.
32. Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995. <http://www.springerlink.com/content/k238jx04hm87j80g/>
33. David Graff and Ke Chen “Chinese Gigaword”, Linguistic Data Consortium (LDC) catalog number LDC2007T03 and isbn 1-58563-409-3. <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2007T38>
34. S. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica Journal 31, 2007, 249-268.
35. Keh-Jiann Chen and Shing-Huan Liu, "Word Identification For Mandarin Chinese Sentences", Proceeding COLING '92 Proceedings of the 14th conference on Computational linguistics 1992, Vol. 1, pp.101-107.
36. Andi Wu, Zixin Jiang, "Word Segmentation In Sentence Analysis", International Conference on Chinese Information Processing in Beijing China 1998, pp.169-180.

# 附錄

**附錄一：線上客語語音合成系統**



 使用者先在文字框輸入欲查詢的國語文句，然後選擇想要合成是客語的四縣腔或是海陸腔，接著按「送出」鍵。經過程式處理之後，會採用AJAX的技術顯示結果，第一個項目「輸入中文字串」是使用者輸入的國語文句；第二個項目「中文斷詞結果」是國語文句經過斷詞器所斷詞完的結果；第三個項目「中文轉客語詞」是根據斷詞完的結果，每個詞依序去國客語對照詞典查詢所得；第四個項目「客語拼音輸出」也是從國客語對照詞典查詢所得出的；第五個項目「客語拼音輸出(客語詞+單字語音+停頓)」則是所合成的客語語音，使用者按下播放鍵就可聽取合成的結果，並且可以重複播放；第六個項目「處理到的多音字」是顯示所處理到的多音字。