

# 行政院客家委員會獎助客家研究優良博碩士論文精要

## 一、論文名稱

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

## 二、作者

羅丞邑

## 三、獎助年度

100 年度

## 四、獎助金額

五萬元整

## 五、研究過程(含研究方法、研究對象、研究工具等)

研究對象：客語語音合成系統中多音字發音的判斷。

本文中提到的語音合成系統就是把使用者輸入的文句發音，就像讓電腦說話，又稱文轉音系統。良好的文轉音系統有許多應用，如電子書、語言翻譯機與語言學習等，帶給社會許多貢獻，除了讓生活更便利，還兼具保存語言文化的功能。現在的文轉音系統大都是國語發音，我們實作了客語發音的文轉音系統，希望為幫助客語教學，保存客家文化做出貢獻。

本研究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。

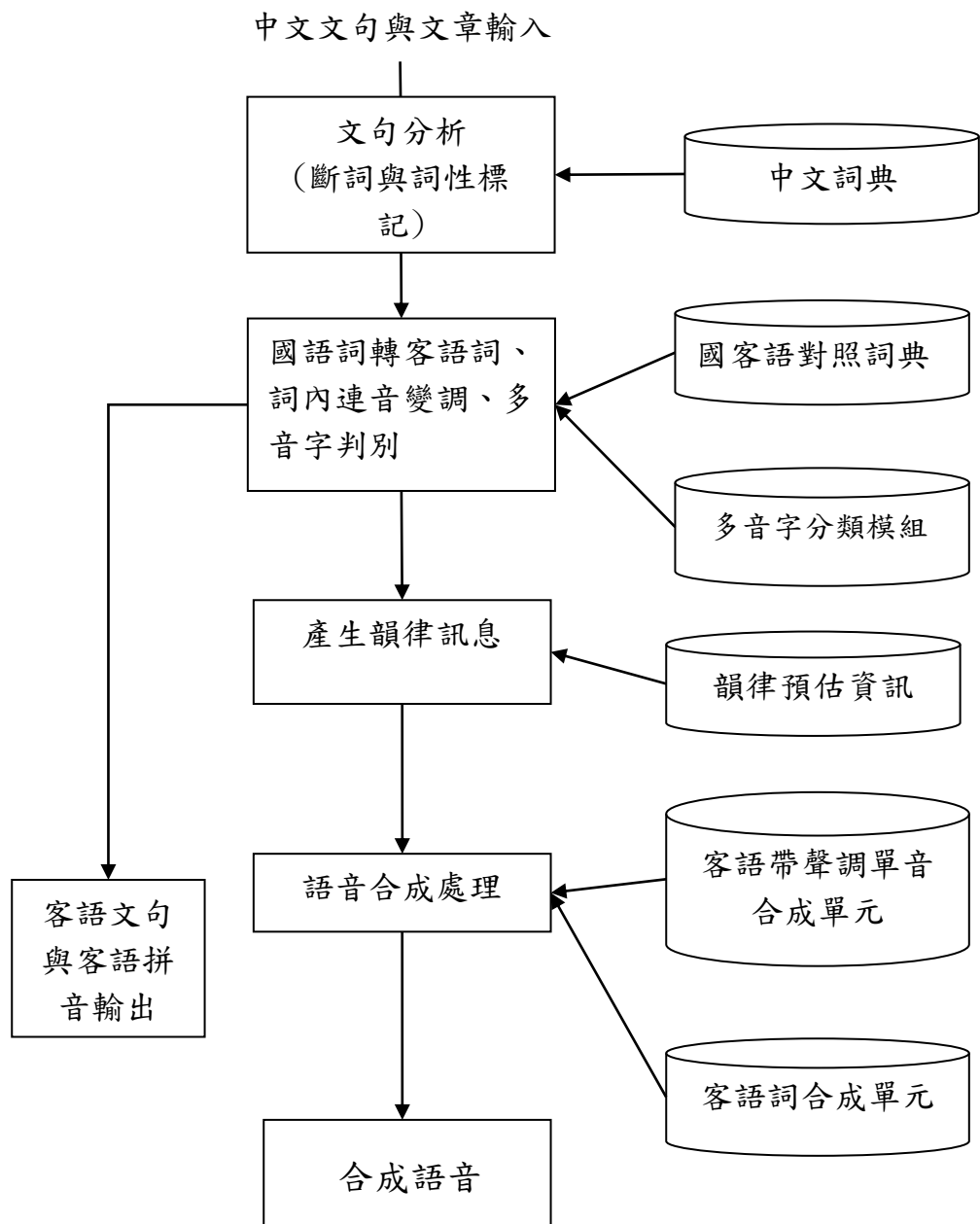
本系統使用的語音合成法即為串接合成法，預錄的合成單元有單音與多音節音檔。因客語是一種具有聲調之語言，每一單音有不同聲調，需事先錄製，以便作為後續之合成使用。

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。

文轉音系統在實作上會遇到許多難題，從最基本的國語斷詞，國語轉客語的詞與讀音是否正確，到客語合成的韻律調整，都要花許多心力去處理。像人工錄音時音檔的音量差異太大，使合成出的連續語音聽起來音量忽大忽小，此時我們要對合成單元進行能量正規化處理，調整音量。

研究方法：使用資料探勘中的「分類」方法，判斷多音字的發音。

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，系統的整體架構如下圖所示。



首先介紹文句分析模組，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。

常見的基本斷詞法有連續字串比對法、語法分析法……等。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」，「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。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  $\hat{W} = W_1, W_2, \dots, W_n$ ，其詞串序列對應的詞性序列為  $\hat{pos} = pos_1, pos_2, \dots, pos_n$ ，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。

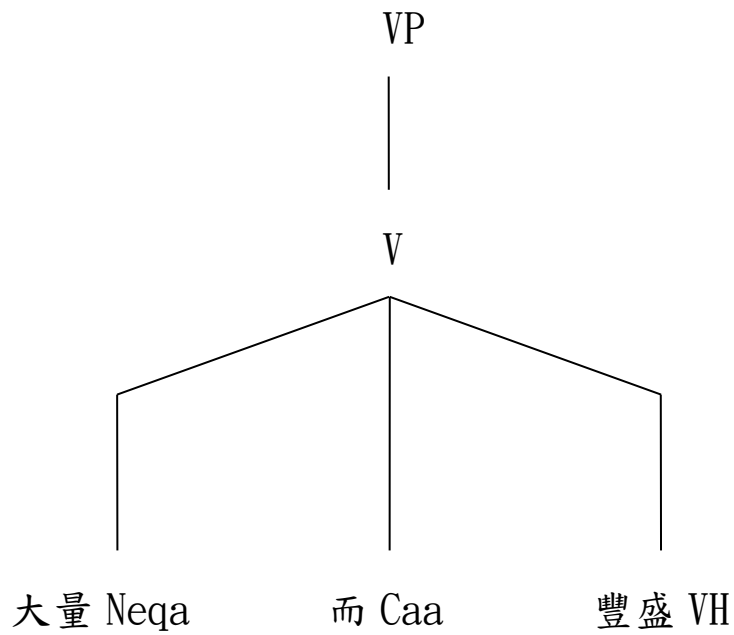
接著是本研究最重要的韻律訊息模組。一般來說韻律訊息可粗分為：音高、音長、音量、停頓。因本系統錄製單音的所有聲調，故不故需調整音高。在錄製「單音」與「詞」的合成單元時，力求發音速度一致，且發音速度太慢的合成單元會被捨棄，以正常速度重錄，故合成時不調整音長。因人工錄製合成單元時，音量大小難以保持一致，將導致合成後語音音量忽大忽小。故要先進行音檔能量前處理：能量正規化。

為使錄音後合成單元之能量調整至較理想之情況，本系統以相同聲調之所有語音進行正規化處理，計算出各個聲調之平均值，依此作為調整能量之依據。意即比能量平均值大之單音檔將調降，反之將調

升。接著用同樣的方法，對多字音檔做調整。

處理完音量後，要預估停頓。在文轉音系統中，停頓是一項重要的特徵。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。因此在文轉音系統中加入停頓預估，合成出的語音會更容易讓人理解。

我們也有採用中文結構樹資料庫(TreeBank)的資訊，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。下圖是 TreeBank 的舉例：



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。

研究工具：貝氏網路(Bayesian network)分類器、C4.5 決策樹、CART 決策樹以及 SVM 分類器，呼叫 weka 軟體中的函式做分類。

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、C4.5 決策樹、CART 決策樹以及 SVM 分類器，並將分類結果做比較。

## 六、主要研究發現

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

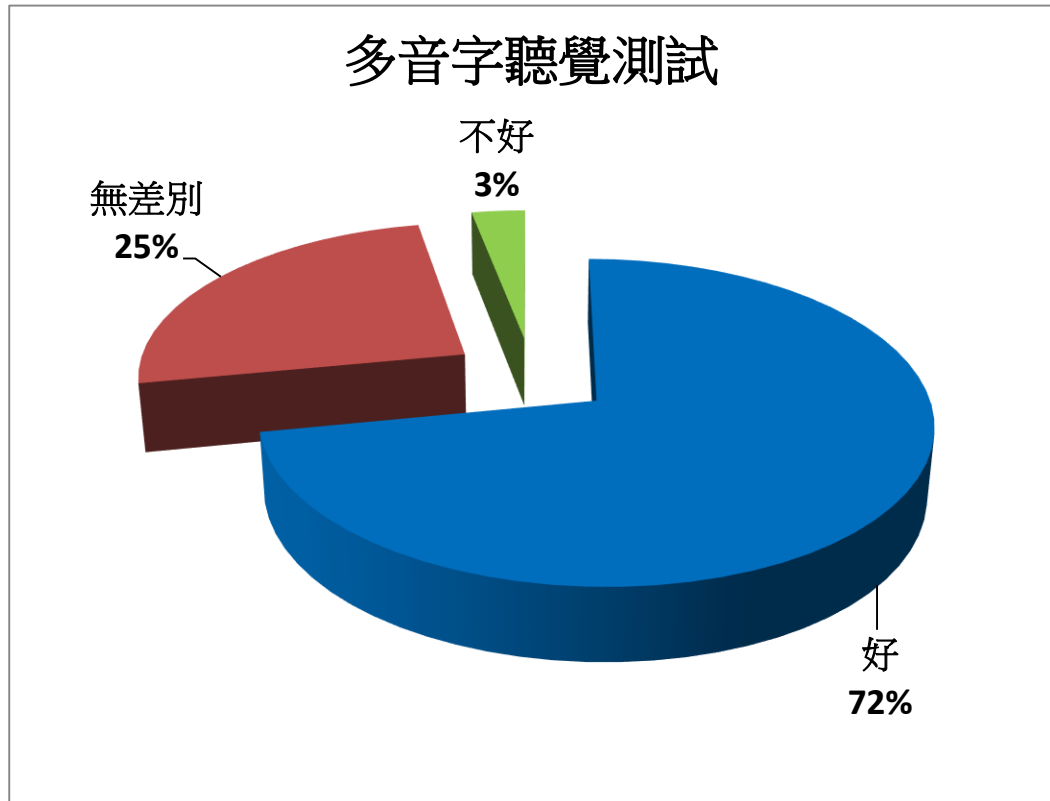


測試是最終的實驗結果，結果發現，不同多音字分類器正確率最高的特徵組合幾乎皆是採用 C4.5 演算法，且選取特徵為「前三後三詞性、目標詞與後一詞可以生成的語法類別、目標詞與前一詞可以生成的語法類別、目標詞與前一詞跟後一詞可以生成的語法類別、目標詞性」的時候；只有在「行」這個字時，是 CART 演算法的正確率較佳，不過跟 C4.5 演算法的正確率也相去不遠。

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。

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，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。



## 七、結論及建議事項

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。

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。