

手機平台 APP 之四縣客語輸入法的研發

Research and Implementation of Sixian Hakka Pinyin Input Method for Mobile Cell APP

黃豐隆, 劉桂森, 曾勝億

Feng-Long Huang, Kuei-Sen Liu, Sheng-Yi Tseng

國立聯合大學資訊工程學系

Department of Computer Science and Information Engineering, National United University

ncat70, c519546, jeff810123@gmail.com

摘要

本論文研究智慧型的客語拼音輸入法為一基於 Android 輸入法框架 (Input Method Framework, IMF), 使用者能在任何文書的 APP (Application, APP) 輸入客語文字。

使用者輸入客語單字或客語詞彙拼音縮寫時, 輸入法會依照使用者的輸入搜尋儲存於 Android 上的 SQLite 資料庫的單字拼音對照字庫、縮寫詞對照詞庫及前後詞對照詞庫, 根據搜尋的結果產生出候選字或候選詞彙, 提供使用者選擇輸出。

單字拼音字庫和縮寫詞對照詞庫分別包含 9361 個字數及 32453 個詞彙數; 客語音檔資料庫包含單音節檔 2427、詞彙檔 3392 及靜音檔 27, 總計 5846。除了基本的客語拼音輸入外, 輸入法本身提供了幾種功能:

- (1) 使用者偏好輸入: 記錄使用者平常輸入的字詞, 目的讓使用者能依照自己偏好更快速將常用的字或詞彙做輸出。
- (2) 客語詞字首快速輸入: 使用者可透過縮寫詞對照詞庫內搜尋字母, 快速得到該客語詞, 節省打字次數。
- (3) 前後詞預測輸出: 此功能具備了讓使用者快速輸出客語句, 讓客語詞彙或句達到更快輸出的效率, 並訓練出客語的前後詞 bi-gram。
- (4) 客語詞人聲發音: 此目的是讓客語初學者能聽到正確的客語拼音唸出, 達到學習的目的。

關鍵詞: 客語無聲調拼音輸入法、好客輸入法、中文轉客文語音合成系統、拼音輸入法。

Abstract

The proposal scheme called Hakka pinyin input method is based on Android (IMF) Input Method Framework. Users can input Hakka texts in any APP of mobile cell. When user inputs a Hakka character or Hakka vocabulary phonetic abbreviation, the input method will refer to the input of user and search for a single character phonetic transcription font stored in the SQLite database. The data will send to database Single Word Pinyin the Based Word Library, Acronym, and Previous and Successive the Based Word Library. According to the results of the search produce a candidate word or vocabulary, and provide the user to select the output.

The Single Word Pinyin and Abbreviation in our systems contain 9361 words and 32453 vocabularies. In addition to the basic message Pinyin input, the input method self provides several functions: (1) User preference input: Record the frequency of the words normally entered by the user in pref. the purpose to allow users follow their own preferences more quickly, that will commonly used words or vocabulary for output. (2) Hakka fast input word: User can search for letters by acronyms, or APP quickly guest the Hakka word, and save the number of typing. (3) Previous and Successive Word Prediction Output: This feature has the ability to let the user quickly generate the Hakka sentence, let the Hakka words or sentences to achieve faster output efficiency, and training the Bi-gram probability. (4) Hakka Language pronunciation: The purpose is to let the beginner to hear the correct Hakka pronunciation, to achieve the purpose of language learning.

Keywords: Pinyin Input Method for Hakka; Hao Ke Pinyin Input Method; Text To Speech; Pinyin Input Method.

一、緒論

由於科技蓬勃發展，個人電腦及智慧型手機越來越普及化，對於使用漢字當母語的我們，在系統內，一定都會使用輸入法來當中文的輸入工具，目前手機內已建置各式各樣的輸入法文體，例如：中文、台語、日文及英文等。

客家族群為台灣第二大的族群，約有四百二十萬人。由於社會環境的變遷因素，導致客家文化不斷且快速流失之中，客委會藉由推廣與傳承文化，希望客家民眾望子女從日常生活及學校學習客家語言，達到容易理解與學習此種語言的地步，讓客語不會瀕臨「滅音」，讓語言得以保存，在每一個社會的角落都均有保存的必要性。隨著資訊的日新月異，在電腦平台上，已經有教育部及客委會提供的客語拼音輸入法，例如：信望愛客語輸入法[13]、教育部台灣客家語拼音輸入法[4]及客語無聲調拼音輸入法等[10]，然而在手機平台不論是 Android 或 IOS 環境，目前均無任何客語的拼音輸入法，因此本論文研究與實作一種使用者能方便且快速輸入四縣腔客語文句的輸入法 APP，希望有助於客語文化的推廣與傳承。

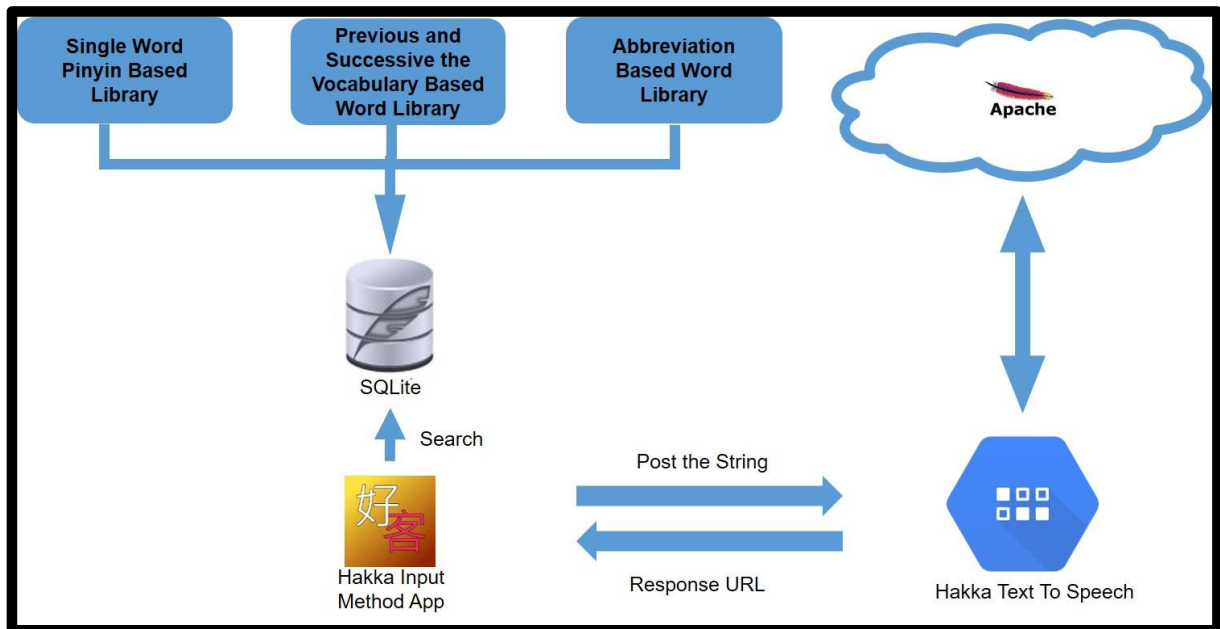
二、手機平台之好客輸入法研發

在輸入法的系統架構內，本論文基於以 Windows 平台的客語無聲調拼音輸入法[10]以及中文轉客文語音合成系統 (Text To Speech, TTS) 的文句分析模組[11]，研發一套 Android 平台的客語輸入法，使用的架構以 Android 輸入法框架為基礎，此為一套輸入法及文字輸出的處理架構，且為開放原始碼的 Android 應用程式套件 (Android Application Package, APK)，因為智慧型手機 (Smart Mobile Phone) 普及化，在手機平台尚無客語輸入法代表客語在手機上作為溝通的語言，因此在 Android 平台上開發好客輸入法 APP，並且穩定支援 Android2.3 版以上。

(一) 系統架構

在 Android 系統裡，輸入法編輯器 (Input Method Editor, IME) 是一個 Android app，也是輸入法框架 (Input Method Framework, IMF) 的組件之一，允許應用程式向用戶提供其他輸入的方法，例如螢幕上的鍵盤，甚至語音的輸入。

如圖一：好客輸入法系統架構所示，系統包含好客輸入法 APP (Application, APP)、語音合成雲端伺服器兩部分，在 APP 部分，為一基於 Android IMF 設計的智慧客語拼音輸入法，使用者能在任何 APP 需要輸入客語文字、英文或標點符號時，鍵盤進行切換並輸入客語文字，如圖二：鍵盤介面所示。



圖一、好客輸入法系統架構圖



圖二、輸入法鍵盤介面

使用者鍵入欲輸出的客語單字拼音或客語詞彙拼音縮寫，輸入法便會以使用者輸入進行搜尋，搜尋儲存於 Android 上的 SQLite 資料庫的單字拼音對照字庫和詞彙拼音縮寫對照詞庫，根據搜尋的結果產生候選字或候選詞彙，供使用者選擇並輸出。

(二) 客語詞典及功能介紹

1. 客語詞典介紹

不像中文使用之普遍，客語文章並非目前台灣一般人常使用的文體，大量採集客語語料需花費許多人力物力，十分不易。本論文包含三個客語詞彙資料庫，是由大量的客語文章內進行斷字及客語斷詞整理所得，經過語料蒐集、前處理而來的客語字 (Characters) 及客語詞彙 (Words)，以及其客語拼音。

本論文的客語詞典分別有 (1) 單字拼音對照字庫 (2) 前後詞對照詞庫 (3) 縮寫詞拼音對照字庫，在手機 SQLite 內的結構以資料表呈現。

如表一：單字拼音對照字庫所示，表內含有 9361 個字數，包含了 Val、Key、Cnt、Pref 四個欄位，分別代表了客語單字 (Val)、拼音 (Key)、字頻 (Cnt)、偏好 (Pref)。

表一、單字拼音對照字庫

	Val	Key	Cnt	Pref
1	ㄚ	a2	4	0
2	鴉	a2	6	0
3	啞	a4	7	0
4	隘	ai1	8	0
5	挨	ai2	11	0
6	唉	ai3	6	0
7	矮	ai4	18	0
8	暗	am1	86	0
9	諳	am1	0	0
10	闇	am1	2	0
11	黯	am1	0	0
12	黥	am1	0	0
13	庵	am2	2	0
14	俺	am3	0	0
15	掩	am4	5	0

本輸入法會依照使用者輸入查詢此資料表，藉由輸入的客語拼音，查詢資料表拼音 (Key) 的欄位，列出所有可能的字詞，並依照字頻和使用者的偏好輸入計算優先權重，然後按照優先權重進行排序；在字頻 (Cnt) 是將所有蒐集的文章以字為單位，進行斷詞所統計出的數量；在偏好 (Pref) 則是記錄使用者習慣輸出的字詞，此部分在第三章的 (三) 小節詳細說明。

如表二：前後詞對照詞庫所說明，有 32453 個詞彙數，包含 Uniq、Prev、Next、Cnt

及 Pref 五個欄位，分別代表了客語詞彙 (Uniq)、上一個詞 (Prev)、下一個詞 (Next)、次數總和 (Cnt) 及偏好 (Pref)。

表二、前後詞對照字庫

	Uniq	Prev	Next	Cnt	Pref
1	這隻	這	隻	38	0
2	隻房間	隻	房間	1	0
3	房間个	房間	个	1	0
4	个光線	个	光線	1	0
5	光線毋	光線	毋	2	0
6	毋好	毋	好	27	0
7	好，	好	，	52	0
8	，毋	，	毋	62	0
9	毋適合	毋	適合	1	0
10	適合做	適合	做	1	0
11	做書房	做	書房	1	0
12	書房。	書房	。	1	0
13	今晡日个	今晡日	个	5	0
14	个天時	个	天時	6	0
15	天時當	天時	當	4	0

使用者在輸入字詞時，會依照上一個詞 (prev) 的欄位在資料庫內做搜尋，並預測出下一個詞 (next) 出現的最高機率做排序提供使用者輸出，並訓練出 bi-gram；然而次數總和 (cnt) 是藉由客家委員會蒐集來的客語詞，以詞的為單位在大量文章以進行斷詞，所統計出在文章內該詞會出現的次數；在偏好 (pref) 是紀錄使用者所使用該詞彙的頻率，並將常使用的詞依照權重優先進行排序，並提供使用者更快的輸出，此說明如上個資料表所言，會在第三節 (三) 小節詳細說明。

客家語言有自創之特殊字型，為了可以正確顯示出字型，在電腦平台客委會有提供造字編碼表及造字檔[14]，供使用者安裝，例如：表二「房間个」的「个」，但礙於手機市面上並無客語輸入法，導致客語在手機上不盛行，因此客委會目前尚未提供手機上相關的造字檔呈現字型。

2. 客語拼音輸入功能介紹

本輸入法提供四種功能：

- (1) 客語無聲調輸出模式：讓使用者更快輸出客語詞及客語文句，此功能在第三節(一)小節會詳細說明。

- (2) 使用者偏好輸入：目的讓使用者能依照自己偏好更快速將常用的字做輸出，此功能在第三節（三）小節會詳細說明。
- (3) 客語詞字首快速輸入：藉由輸入縮寫詞的拼音，在表三：縮寫詞拼音對照字庫搜尋字母，快速得到該縮寫的客語詞，節省打字次數。此功能引用縮寫詞對照字庫，在資料表包含了詞彙（val）、字首縮寫拼音（key）、拼音（pinyin）、詞類（part）及偏好（pref）等五個欄位，輸入法將輸入的縮寫字母在詞庫內以 key 欄位作為搜尋條件，並將搜尋到的結果會產生候選字及候選詞彙，提供使用者選擇並輸出。

表三、縮寫詞拼音對照字庫

	Val	Key	Pinyin	Part	Pref
1	值勤	ck	ciid1 kiun3	VA	0
2	植入	cn	ciid1 ngib1	VC	0
3	植皮	cp	ciid1 pi3	Na	0
4	植物	cv	ciid1 vud1	Na	0
5	植被	cp	ciid1 pi2	Na	0
6	植髮	cf	ciid1 fad4	Na	0
7	植樹	cs	ciid1 su1	VA	0
8	殖民地	cmt	ciid1 min3 ti1	Nc	0
9	叱喝	ch	ciid4 hod4	VC	0
10	螫伏	cf	ciid4 fug1	VA	0
11	螫居	cg	ciid4 gi2	VCL	0
12	沉思	cs	ciim3 sii2	VE	0
13	沉重	cc	ciim3 cung1	VH	0
14	沉淪	cl	ciim3 lun3	VH	0
15	沉痛	ct	ciim3 tung1	VH	0

- (4) 客語自然語音輸出：此功能是 APP 提供使用者使用學習的功能，目的是讓客語初學者能聽到正確的客語拼音唸出，達到學習的目的，此功能將在第三章（四）小節做說明。

三、輸入法之功能設計與結果分析

在此節中，我們會詳細介紹好客拼音輸入法 APP 在輸入的模式中，從無聲調輸出模式、前後詞預測輸出、使用者偏好輸入功能及客語自然語音輸出，進行驗證說明與討論分析。

(一) 無聲調輸出模式

客語單字與詞彙是蒐集大量的客語文章，進行斷字與斷詞的動作，在提供在輸入法框架（Input Method Framework，IMF）內，並做輸出單字與詞彙的功能，在這個模式上我們運用了資料探勘（Data mining，DM）裡面的聚類分析（Cluster Analysis，亦稱集群分析）[17]的觀點，將比較相似的樣本聚集在一起，形成集群(cluster)，把「距離」作為分類的依據，「相對距離」愈近的，相似度就愈高，以達到分群的目的，如表四：聚類分析表所示，目的輸入客語拼音時，將相對應的客語字詞更快顯示在候選詞窗中，並提供使用者輸出。

表四、聚類分析表

Input	b	bi	bia	bian	biang	biang1
Output	ba1	bi1	biag1	bian1	biang1	biang1
	ba2	bi2	biag2	bian2	biang2	.
	biang3	.
	bai1	bien1	biang4	.
	bai2	bien2	biang1	.	.
	bai3	bian1	biang2	.	.

正如上述所言，在此要輸出 **biang1** 也就是「拚」的客語單字，當在鍵盤上輸入 **b** 時，將會搜尋客語單字拼音對照字庫、客語縮寫詞對照詞庫內所有為 **b** 的客語字或詞彙，在候選詞內提供選取，當組字窗內多加一個 **i** 時，此時在單字與縮寫對照字庫內的搜尋範圍會縮小，並列出 **bi** 相關的單字或辭彙，在候選詞內，也就是每當多輸入一個相關的字母時，所依照輸入的字母搜尋的相關範圍會逐漸縮小，此時相關的客語字就會逐漸出現，並提供使用者輸出，如圖三：聚類分析所示。



圖三、聚類分析說明圖

(二) 前後詞預測輸出

此功能目的是為了減少打字的次數，提供使用者能更快輸出，在此使用大量的客語文章及客語的句子做斷詞的動作，正如上述 3.2 的客語詞典介紹所言，在 DataBase 內設 (uniq) 客語詞彙、前一個詞 (prev)、下一個詞 (next)、個數 (cnt) 及偏好 (pref) 等五個欄位；利用斷詞完成的結果歸類在 DataBase 內。例如：要輸出一句客語句子「有九降風來吹燥。」，當輸出「有」，在候選詞窗內會出現所有「有」後可能出現的客語字或詞彙「禮貌」、「十」或「九降風」.....等，如圖四：前後詞預測輸出圖所說明，供使用者選取，選取「九降風」時，後面會接著「九降風」後面可能出現的詞彙或字，幫助使用者減少打字次數；並且在輸入字詞時，會依照 prev 的欄位做搜尋，並預測出下一個詞出現的最高機率做排序提供使用者輸出，並訓練出 bi-gram。



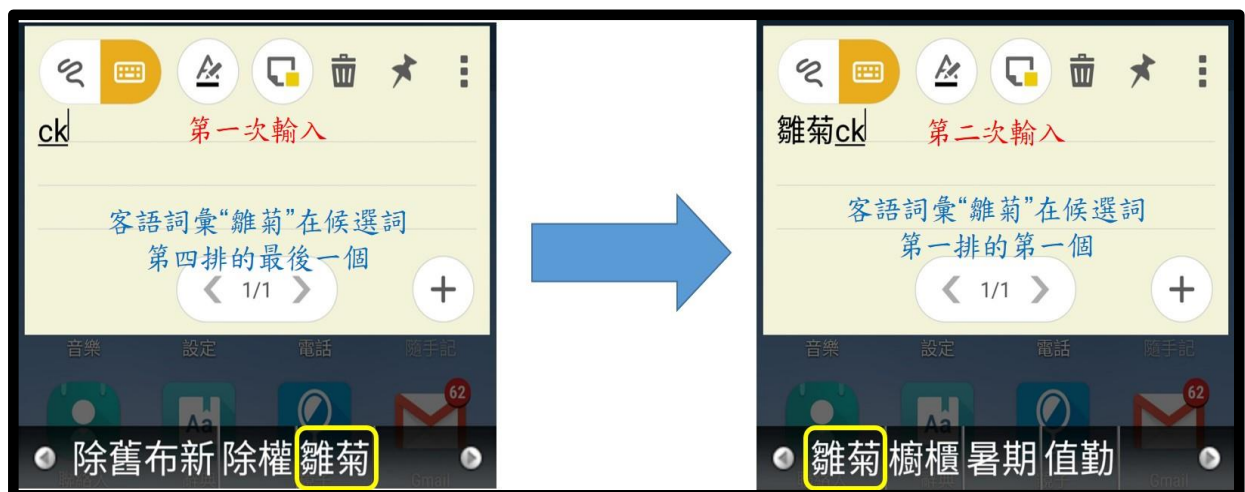
圖四、前後詞預測輸出圖

(三) 使用者偏好輸入功能

依照使用者所輸入的習慣不同，因而影響打字輸入的效率，此功能的目的是讓使用者能依照自己偏好，更快將常用的字作輸出，以提高輸入法的效率。

1. 偏好輸入功能

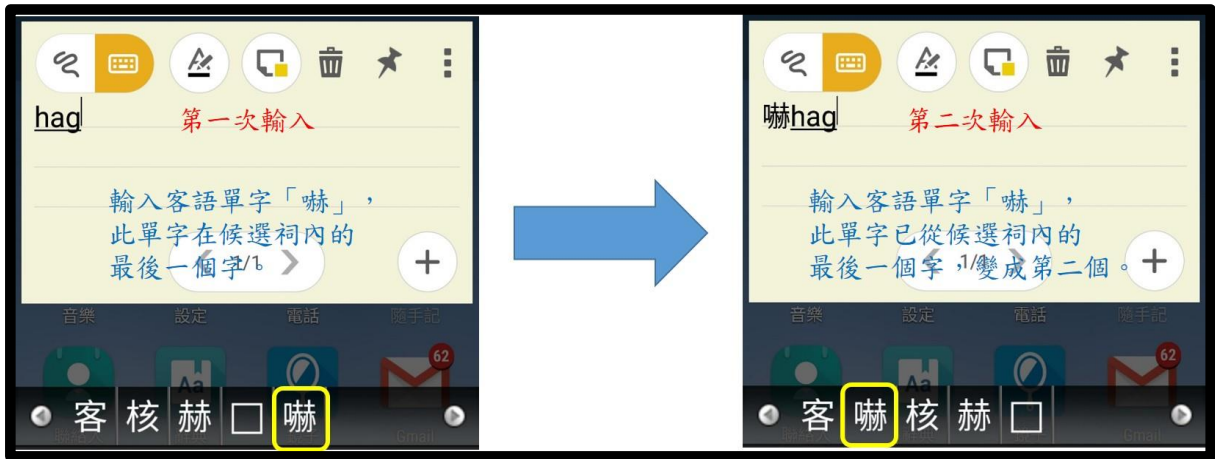
在縮寫詞拼音對照字庫上，由於沒有 **cnt** 欄位，為了將使用者常用的詞，更快輸出，因此在資料庫欄位上設置 **pref** (偏好)，紀錄使用者平常輸入的字與詞彙的頻率；在系統內的 **pref** 一開始的初始值為 0，每當輸入一個字或者詞彙時，將紀錄在 **DataBase** 上的 **pref** 欄位做加 1 的動作，以此類推，並依照 **pref** 排列在作搜尋，排列出下一個可能習慣輸出的客語單字或詞彙，並提供在候選詞窗中的前面供使用者選取做輸出，如下圖五：Pref 排序所示。



圖五、偏好輸入功能圖_Pref 排序

2. 權重比偏好輸入

在單字拼音對照字庫及前後詞對照字庫內，由於有 **cnt** 跟 **pref** 的欄為，因此我們將這兩欄位依照自定權重比相加後，再作排序，首先我們設一個可放隨機變數的字串，將 **cnt** 與 **pref** 這兩個欄位依照自訂的權重比相加 (比例為 1:10)，並將結果回傳至字串內，在將字串依照權重比相加所計算出的結果做排序，下圖六：範例為程式內輸入「hag」為「嚇」計算權重比相加後所產生排序。



圖六、偏好輸入功能圖_權重比排序

3. 比較與討論分析

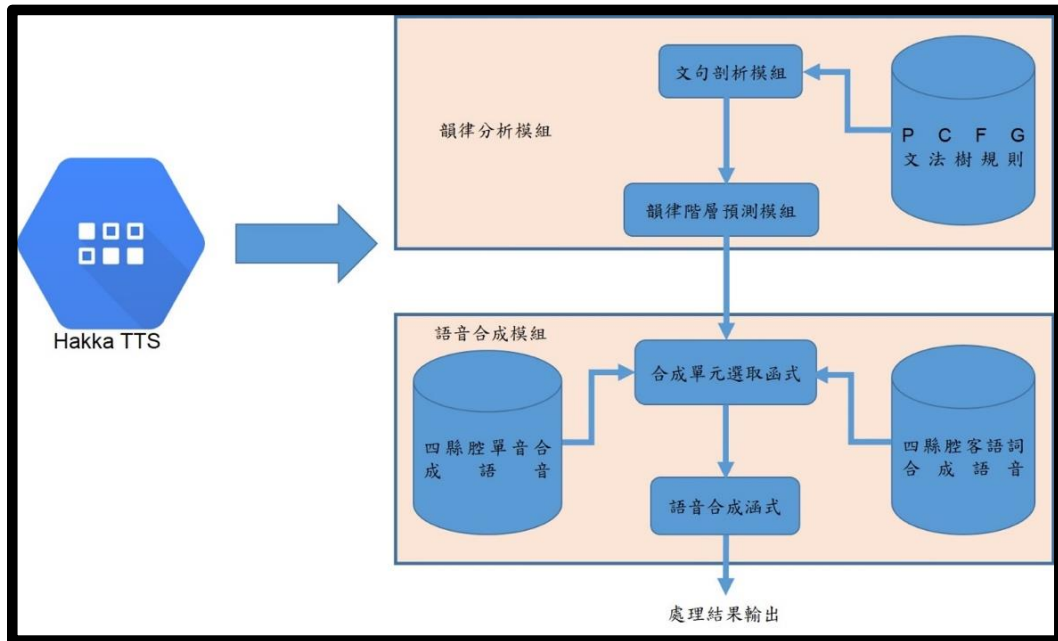
在本章節將提出的偏好輸入與權重比偏好輸入做詳細比較與討論，在偏好輸入部分採用每輸入一次，在 **pref** 初始值為零狀況直接加一，意指原在候選詞最後一頁的最後一個詞，輸出後則直接跳到候選詞第一頁第一個；相較於權重比偏好輸入，依照 $\text{cnt} * 1: \text{pref} * 10$ 比例相加後做排序，會依照 **cnt** 是否為零，候選詞的排序則不會從最後一頁的最後一個，直接跳到第一頁第一個，而是可能先在第五頁的第四個再來第三頁的第二個，慢慢往前排序，因此得出的結論是權重比偏好輸入有自動學習的功能輸出的效果為最好，反之偏好輸入如果要輸入曾未輸入過的詞要慢慢找，甚至可能找到最後一個。

(四) 客語自然語音輸出

輸入法提供客語自然語音輸出功能，目的讓客語初學者及非客家人在輸入法輸出客語字詞，能聽到正確的客語拼音念初，達到學習目的。此章節說明藉由模組之間的關係而形成語音合成模組，以及語音合成後輸出的整體流程。

1. 語音處理系統架構圖

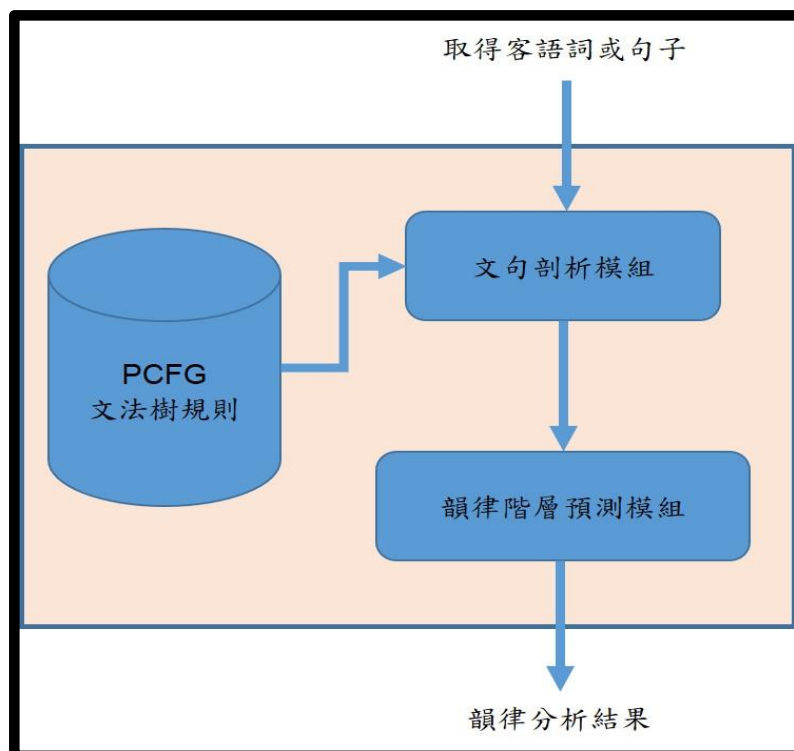
在系統內的客語語音合成功能模組 (Hakka Text To Speech, HTTS)，如圖七：HTTS 系統架構圖所示，是由韻律分析模組及語音合成模組所組成 (此模組引用興大資訊科學與工程學系林昕緯)。



圖七、HTTS 系統架構圖

2. 韻律分析模組

在韻律分析模組中分為文句剖析器與韻律階層預測模組，文句剖析負責將客語斷詞及標記詞性後的客語句子做文法的剖析，韻律階層模組再根據文法的剖析結果，找出句子中的韻律階層，在根據韻律階層給予適當的停頓訊息，因此韻律分析模組目的是客語字詞或句子，藉由語音合成模組合成後聽起來很流暢，如圖八：韻律分析模組架構圖。

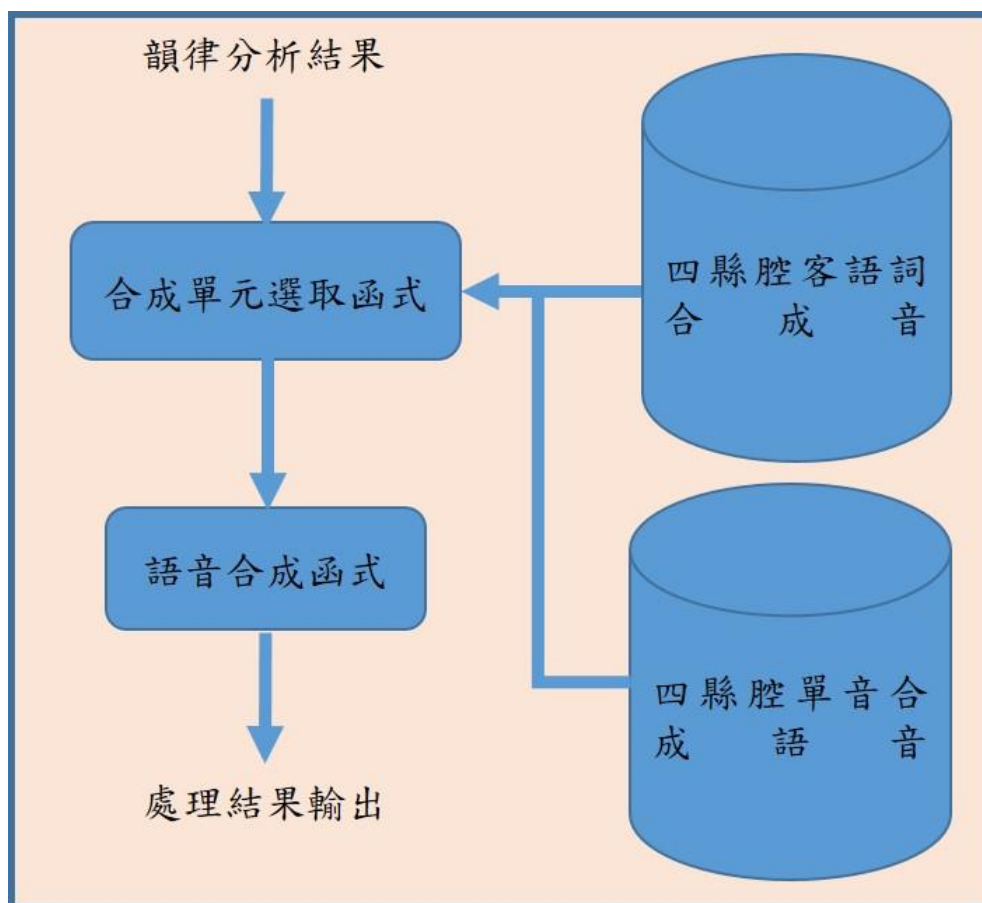


圖八、韻律分析模組架構圖

3. 語音合成模組


此模組是由(1)單元選取模組(2)語音合成器兩個子模組所合成，在單元選取模組內有兩個客語語音資料庫(1)客語單音節合成音(2)客語詞彙合成音，此語音資料庫則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)，錄製此合成音檔，在音檔的單音節共有2454個音檔，而客語詞彙音檔則是3392個音檔，詞彙音檔之錄製範本採用「客語能力認證基本詞彙」[15]，音檔格式皆屬：44.1kHz、16bits，儲存為Windows PCM格式(wav檔)。

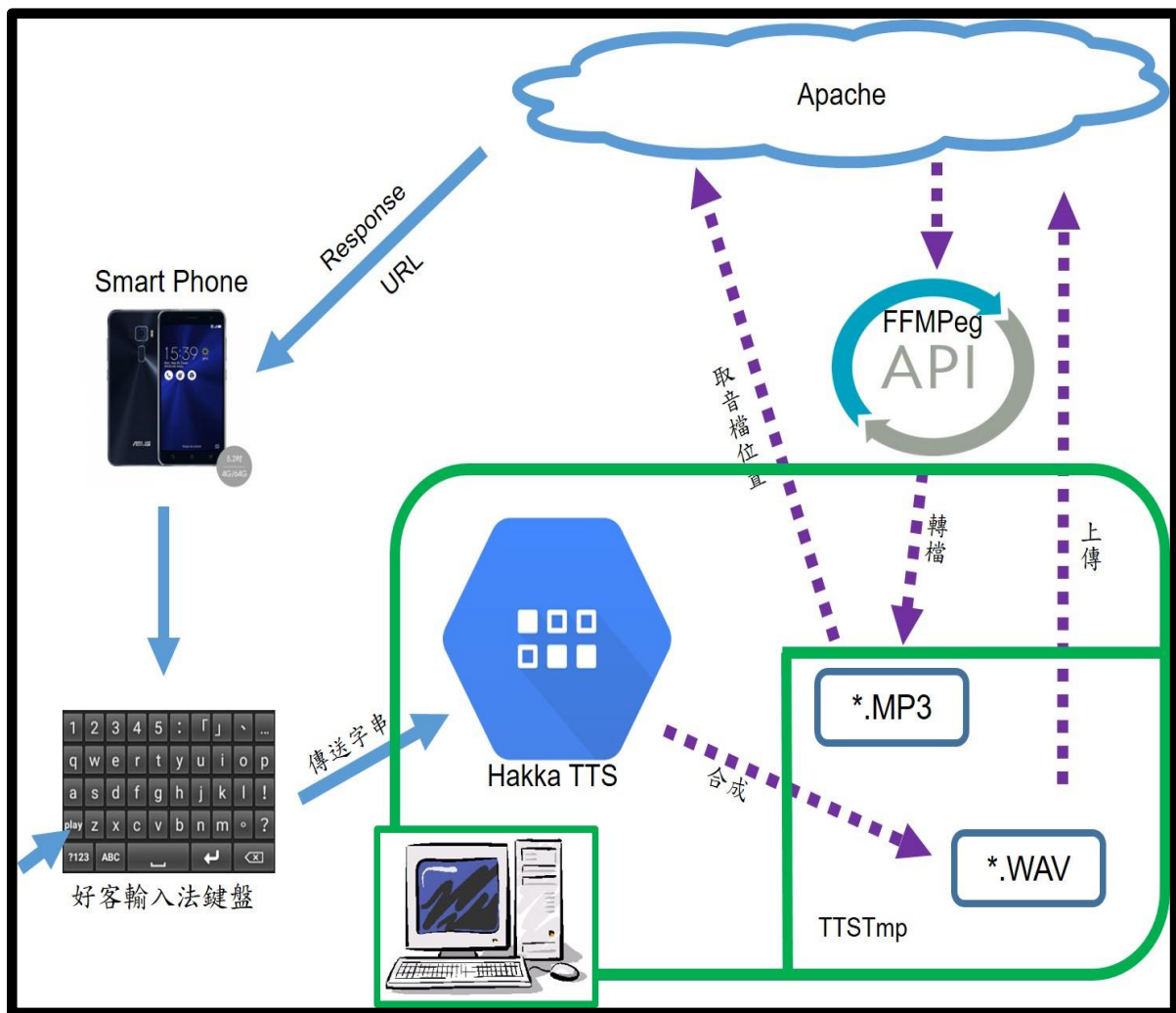
首先單元選取模組接收韻律分析的結果，從客語詞彙合成音及單音節合成音的音檔資料庫中，找尋適合的詞彙合成音及單音；在藉由語音合成器將選出來的語音檔作序列串接合成，如圖九：語音合成模組架構圖所示。



圖九、語音合成模組架構圖

4. 語音輸出操作流程

在好客輸入法 APP 鍵盤上設置一個 play 鍵  (如圖十：客語語音系統流程圖所示)，將暫存輸出或輸出後的客語字或客語詞彙存放在系統 buffer 內，當使用者觸發 play 鍵後，將 buffer 內的暫存客語字詞，藉由 Socket 回傳至本端伺服器的 HTTS 資料庫內合成，此時合成出的 wav 格式，再由 Apache 取至 FFmpeg 的 API 內轉檔成 mp3 格式 [16]，再由 Apache 來 Response URL 至系統的 buffer 內做播放，當想聽第二次播放時，此時無需再合成一次，直接取 buffer 內 URL 做播放，客語自然語音輸出系統流程圖如圖十所示。



圖十、客語自然語音輸出系統流程圖

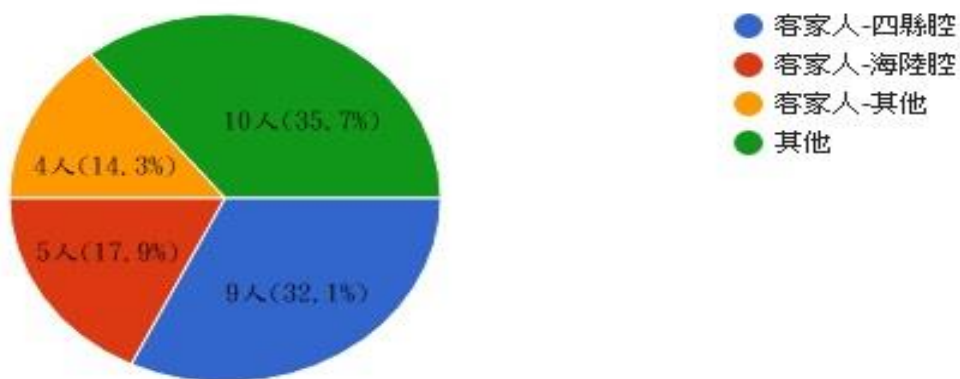
四、結果分析

在資訊的時代裡，台語兒[18]藉由每個人每天都會上網，將台語的字、詞及句子發音正確的唸出來，來傳承與推廣，並幫助社會上學習台語，由此功能，讓我們領悟藉由

客語詞彙或客語句的正確發音，可以提供客語初學者得知客語的念法是否正確，幫助使用者學習目的，對於客語熟練者，可以給不同腔調的客家人得知四縣腔的發音，達到客家人之間互相溝通的環境。

(一) 客語語音比較與分析討論

使用者依照好客輸入法所輸入的客語字詞或文句，傳送至 HTTS 內合成，藉由韻律分析模組將客語字詞或文句做文句剖析，並將剖析結果送至韻律階層內輸出結果，其目的讓語音合成模組取得韻律階層的結果，所合成音檔輸出聲音更有流暢度 (Fluency)；因此將客語合成音檔提供不同腔調的客家人 18 人與非客家人 10 人，總共 28 人，如圖十一客語聽測者背景比例所示，實際做客語音檔的聽測，過程中我們針對客家人與非客家人人口做分析，並取得聽測清晰度 (Intelligibility) 與流暢度結果，藉此兩者可得知是否有效幫助學習客語發音。



圖十一、客語聽測者背景比例

本論文將依平均主觀分數 (Mean Opinion Score, MOS) 方法取操作滿意度調查，在聽測過程中，每一聽測客語音檔由一至七個客語字詞所合成的文句，由聽測者聽完一音檔，給予評分範圍為五分至一分，內容分別：1 分_非常不清晰、2 分_不清晰、3 分_普通、4 分_清晰及 5 分_非常清晰，在計算平均；礙於每人給評分標準不一樣，在分析平均之間設落點區別，內容分別如下 0：聽測度量表。

表五、聽測平均主觀分數的度量表

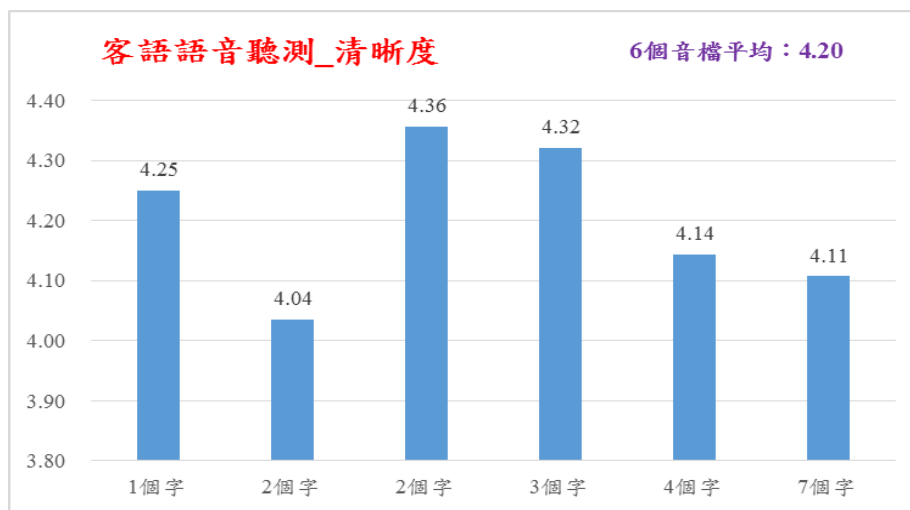
分數範圍	品質	失真情形
1~1.5	非常不清晰	句子有很多字念錯，聽起來很煩躁。
	非常不流暢	聽起來非常不舒服，流暢不自然。
1.6~2.5	不清晰	句子有很多字念錯，聽起來很煩躁。
	不流暢	流暢不自然，但還不至於不舒服。
2.6~3.5	普通	句子有些念錯，流暢有點不自然，語音感覺失真，而且有點煩躁。
3.6~4.5	清晰	句子念對，語音有點失真。
	流暢	流暢自然，但不覺得不順。
4.5~5	非常清晰	句子念對，語音無失真。
	非常流暢	流暢自然。

如下表六：平聽測分數平均顯示，針對客家人、非客家人及全部聽測者分開調查，並給予合成音檔分數加總後的平均。

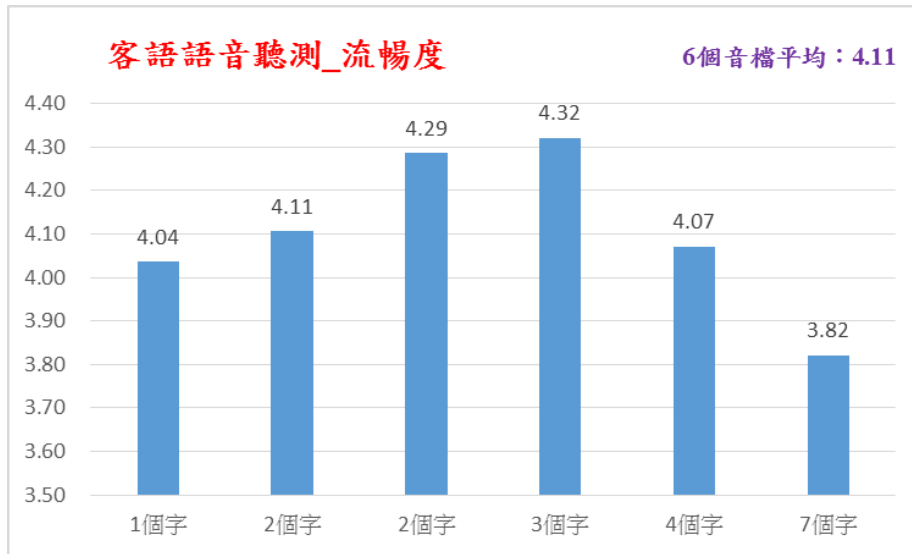
表六、聽測分數平均表

	客家人	非客家人	全部聽測者
清晰度	4.38	3.88	4.20
流暢度	4.23	3.88	4.11

當輸入越多客語字送至客語語音合成處理時，所合出的音質不會因為字越多，而咬字不清晰，缺點在於流暢度靠近流暢（4分）的階段，句子念對，流暢自然，語音有點失真，但不覺得不順，由此階段測驗可證明本論文提出的客語語音輸出是可用於幫助客語發音學習，由圖十二客語語音清晰度與圖十三客語語音流暢度說明與顯示。



圖十二、客語語音清晰度結果



圖十三、客語語音流暢度結果

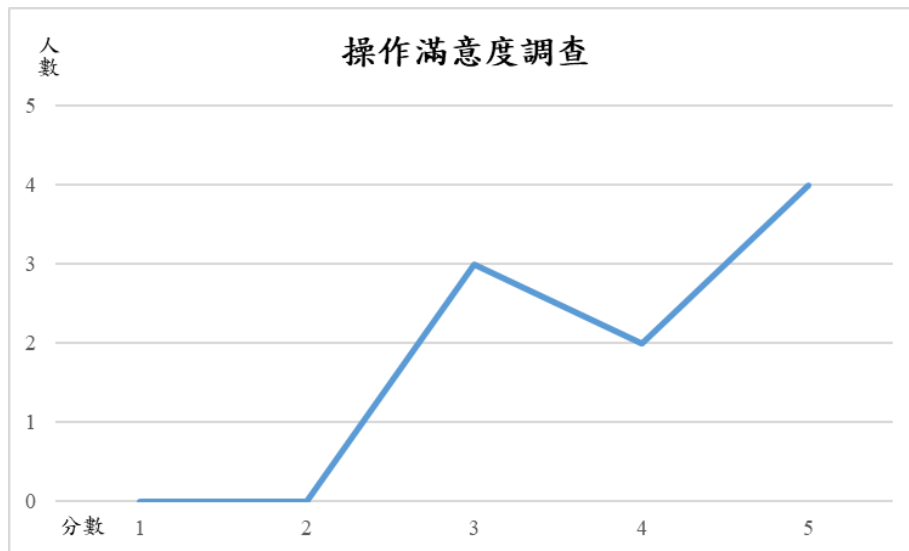
(二) 輸入法操作滿意度分析討論

受測者主要以懂得客語拼音為主，並提供於使用客語年限已超過五年，並擁有客語初、中或中高級其之一認證者：客家委員會人員含客語薪傳師：鍾惠容、張旭英、林錫霞等 6 人、國立聯合大學客家研究學院的文觀系學生陳毓泰等 3 人，合計 9 人，使用腔調以四縣腔為使用年限較長（20 年以上）。

首先受測者利用本論文研發客語拼音輸入法，輸入五句客語句子或受測者想表達客語句詞，並依操作滿意度給予分數 1~5 分，滿意度標準如表七：滿意度量表。

表七、滿意度平均主觀分數的度量表

滿意度分數	操作滿意度	操作情況
1 分	操作非常不易	輸入時間花費 2.5 分鐘以上，輸入非常慢。
2 分	操作不易	輸入時間花費 1.5~2.5 分鐘之間，輸入慢。
3 分	操作普通	輸入時間 1~1.5 分鐘之間，輸入普通。
4 分	操作容易	輸入時間 30 秒~1 分鐘之間，輸入快。
5 分	操作非常容易	輸入時間 30 秒以內，輸入極快。



圖十四、輸入法操作滿意度統計圖

由上圖十四：操作滿意度可得知，受測者在科技蓬勃發展時代裡，對於本論文研發客語拼音輸入法的操作滿意度：5分（操作非常容易）人數4人、4分（操作容易）則有2人及3分（操作普通）則有3人。並表示受測者給語輸入法評價與建議：

1. 5分：在於可以使用母語在手機平台上作為溝通橋樑，且系統操作容易上手。
2. 4分：客語特殊字體無法正確呈現為整體的美中不足。
3. 3分：受測者以傳承與教學立場表示，對於不懂拼音的客家人及客語初學者，建議輸入法使用輸入漢字拼音並輸出客語字表示，例如：輸入「今天」，系統提示輸出相對應客語字詞「今晡日」。

因此為了解決特殊字呈現及不懂拼音的客家人及客語初學者，未來研究工作實有賴進一步研發客語特殊字體編碼及漢字輔助客語輸出。

五、結論

本論文提出智慧行動裝置中的四縣客語拼音輸入-好客輸入法，目的是讓熟悉客語使用者及客語初學者使用，對於客家族群來說，客語輸入法是他們在眾多輸入法中，可以用自己的母語輸出表達知其意義。

以往聚類分析的結果是用英文拼音、部首拼音及注音拼音等所組成的漢字，使用此觀點並提供無聲調輸出模式，在拼音組字時越接近，相對應的客語字詞就越快出現，並在候選詞窗提供客語字或詞彙輸出；藉由此模式，在前後詞預測輸出訓練詞與詞之間 bi-

gram，及使用者偏好輸入依使用者習慣，並記錄平常輸入的字詞，並幫助使用者使用更快輸出客語字詞及文句。

在客語自然語音輸出功能，在有效幫助學習客語四縣腔的正確發音時，並針對客家人及非客家人進行客語聽測實驗，測量合成音檔的清晰度與流暢度，平均分數皆在 4.20 及 4.11，落點在於清晰與流暢，結果分析出合成 7 個字詞甚至越多，音檔本質不會不清晰，反而咬字很清楚，缺點在於流暢度靠近普通至流暢的階段，藉此結果驗證語音合成功能，可以幫助初學者學習客語四縣腔正確發音，有助於其他客語腔調學習。

對於本論文所探討的主題而言，仍有不少改善的空間，例如：手機平台上無法回朔「所選取的選字」導致使用者所輸入的文具分析無法邁行文具處理，客語特殊字無法有效在手機內呈現字型，客語初學者不懂拼音等。因此，在未來可進行的工作及研究方向可包含下列項目：

- (1) 錯誤提示功能：藉由系統建議選字，讓使用者無須再打一次，可自動更改成想要輸出客語字詞。
- (2) 中文輔助客語輸出：輸入法藉由輸入漢字來輔助客語輸出，並藉此系統幫助不會拼音客家人及客語初學者使用。
- (3) 客語特殊字型呈現：為了有效呈現特殊字作為溝通橋樑，因此向客委會索取客語造字編碼研究造字檔，並與各大智慧型手機廠牌結合，希望有朝一日能夠在手機上呈現客語的特殊字型。

藉由系統目前所提供的功能，並結合未來研究的方向，讓客語輸入法充滿實用性，並有效幫助在手機平台做為溝通橋樑，以及提供客語初學者在客語級別能力的檢測學習，希望有助於客語文化的推廣與傳承。

六、參考文獻

- [1] 客家委會出版，臺閩地區客家人口推估及客家認同委託研究成果，2014。
- [2] 客家委會出版，台灣客家民眾客語使用狀況，2012-2013。
- [3] 教育部客委會出版，客家語言拼音方案，2012。
- [4] 教育部客委會出版，台灣客家語言拼音輸入法，2013。
- [5] 客家委會出版，客語能力認證基本辭彙-中級、中高級暨語料選粹，2016。
- [6] 客家委會出版，客語能力認證基本詞彙-初級四縣版。

- [7] 教育部客委會出版，教育部台灣客家語常用詞辭典。
- [8] 客家委會出版，客語能力認證基本詞彙-中級、中高級暨語料選粹四縣版上冊。
- [9] 客家委會出版，客語能力認證基本詞彙-中級、中高級暨語料選粹四縣版下冊。
- [10] 魏俊瑋，客語無聲調拼音輸入法之研究與實做，國立中興大學資訊工程系碩士論文，2013。
- [11] 林昕緯，中文轉客文語音合成系統中的文句分析模組之研究，國立中興大學資訊科學與工程學系碩士論文，2014。
- [12] KerKerInput 輸入法，2010，<https://github.com/itszero/KerKerInput>，[存取日期：28 6 2016]。
- [13] 信望愛台語客語輸入法，2005，<http://taigi.fhl.net/TaigiIME/>，[存取日期：15 7 2016]。
- [14] 哈客網路學院出版，客語造字檔，2009，<http://elearning.hakka.gov.tw/ver2015/kaga/dontseehakkafont.aspx>，[存取日期：15 6 2016]。
- [15] 哈客網路學院出版，客語能力認證基本詞彙，https://elearning.hakka.gov.tw/ver2015/kaga/Kaga_QD.aspx?type=primary，[存取日期：15 6 2017]。
- [16] FFmpeg，2011，<https://ffmpeg.org/about.html>，[存取日期：15 6 2017]。
- [17] DATA MINING Concepts and Techniques(3th ed.). (2011). Morgan Kaufmann, Jiawei Han, Micheline Kamber, Jian Pei.
- [18] BaconBao，台語兒，google chrome store，2015，[存取日期：28 12 2016]。