

## 智慧手機客語拼音輸入法之研發-以臺灣海陸腔為例

### Research and Implementation of Hakka Pinyin Input Method for Mobile Cell - An Example of Taiwan HioLiuk Accent

黃豐隆, 劉名峻

Feng-Long Huang, Ming-Chan Liu

國立聯合大學資訊工程學系

Department of Computer Science and Information Engineering, National United University

ncat70, gn01932987@gmail.com

#### 摘要

根據客委會 105 年度所進行的全國客家人口暨語言基礎資料調查資料，全國最新客家人口比例為 19.3%，總數為 453 萬 7 千多人，成為台灣第二大族群，因研究者本身為新竹客家人，新竹客語以海陸腔為主，故以海陸腔為基礎。

本文開發智慧手機平台上客語拼音輸入法之海陸腔系統，在客語海陸腔輸入法中，除了基礎功能使用者偏好輸入、客語拼音字首快速輸入、前後詞預測、並提出了解決客語特殊字在智慧手機上無法顯示的問題之方法，讓客語特殊字以其拼音方式顯示。為讓完全無客語基礎之使用者也可以打出客語文章，我們增加中轉客輸入法，中文轉客語輸入法設計給毫無客語語言基礎或是尚在學習客語的使用者，讓使用者使用注音符號無聲調方式來輸入，輸入中文後、系統會自動比對其中文的客語詞彙，再進一步將輸入的中文轉成客語詞彙顯示。

在系統客語語料庫中，包含客語單字拼音字庫 5366 筆、客語詞彙拼音字庫 19351 筆、客語前後詞字庫 8505 筆與中文客語對照字庫 1679 筆；語音合成系統縮需之語音檔有單詞 2949 個、詞彙 1566 個，這些資料為客委會和教育部所提供。

在使用者評估調查中，在客語海陸腔輸入法的總平均 4 分，中文轉客語輸入法總平均分數為 4.46 分，客語語音總平均分數為 4.3 分，在台灣大多數使用者使用注音做拼音輸入，所以使用者的輸入習慣大大的影響其評估分數，在語音調查方面，使用者認為此系統有助於學習客語海陸腔的發音，也有助於數位學習的發展。

**關鍵字:** 智慧型手機，客語海陸腔輸入法、中文轉客語輸入法、語言模型。

## Abstract

According to a survey data about the Hakka dialect made by Hakka Affairs Council in 2016, Hakka people in Taiwan is a population ratio of 19.3%, with 4.537 million peoples, making it the second largest cultural group in Taiwan. Coincidentally, the researcher of this thesis is a Taiwanese from Hsinchu city with a Hakka origin. Also, the Hakka accent in this area in Hsinchu is called HioLiuk accent. Thus, this thesis is based on HioLiuk accent.

The Hakka input method not only has the basic functions such as Preference, Quick-type, words- prediction, but also displays a pinyin selection on the smart phone monitor for the user to key-in the desired word and instantly provides the special Hakka characters. This input method is designed for people who do not understand Hakka language well, and allows the users to use phonetic notation (without tones) as input. After keying in, the system will check its meaning and output the Hakka characters to finally choose from.

There are 5,366 records in Hakka Single Word Pinyin database (table), 19,351 records in Hakka Word Pinyin database (table), 8,505 records in Hakka Previous and Successive Word database (table) and 1,679 records in Chinese To Hakka database (table). Speech synthesis system requires many voice files: 2,949 word files and 1,566 phrases provided by Hakka Affairs Council and Ministry of Education.

In the survey questionnaire filled-out by users, HioLiuk input method got 4.0 points on average. Chinese To Hakka input method got 4.5 points on average, Hakka Language pronunciation got 4.3 points on average. Most users in Taiwan use pinyin input method, which has the big influence on the points. In the questionnaire about voice input, users believe that this system is conducive to those people who are learning Hakka and effective for the development of e-learning.

**Keyword: Mobile Cell, Hakka Input Methods, Chinese-Hakka Input Methods, Language Model.**

## 一、概論

台灣客家人是指具有漢族客家人血統的臺灣人，為臺灣第二大族群。其母語為台灣客家話，歷經明鄭時期、清治時期，數百年來通行於臺灣的客家地區，日治、光復時期至今，仍是多數客家族群的母語。「逢山必有客、凡客必住山」西部的丘陵與山區漸形成客家人為主的村莊，保有濃厚的客家文化，亦有客家人居住在平原或沿海鄉鎮等地。在客家委員會根據「客家基本法定義」調查顯示，臺灣客家人口比例較高的客家地區即桃竹苗、高屏、花東縱谷地區。

### 臺灣客語與輸入法現況

根據客委會 105 年度所進行的全國客家人口暨語言基礎資料調查資料[25]，全國最新客家人口比例為 19.3%，總數為 453 萬 7 千多人，而有子女的客家人中，子女和父母交談使用的語言以國語占大多數，造就現在客語有逐漸失傳的現象。

目前台灣的 3 C 產品普及率高，幾乎人手一機，如果能在智慧型手機上推廣客語，應能提升客家語言，並傳承發揚客語文化。基於四縣腔客語輸入法[2]，本文進一步開發出海陸腔版本之輸入法。再以二十六個羅馬拼音方式[4]作為客語輸入，並提出一種替代字的拼音，解決臺灣客語中十分特別之「特殊字」，原本無法顯示於手機平台的問題，將文句中的特殊字以其拼音方式顯示。

語言與文字是人們溝通方式中最常使用的方法，如果可以更有效率使語言的學習與流傳，勢將有助於語言與文字的推展有益，進而促使文化的推展更加順暢。如前所述，臺灣的客語使用率不高，導致客語與客家文化面臨嚴重流失，未來甚至可能消失。因此，在手機環境中研發客語的拼音輸入法，提供一種具效率方式，使用者可以直接輸入客語文字，並且具有語音輸出，提供便利學習客語的方式，這是研發本輸入法的目的。

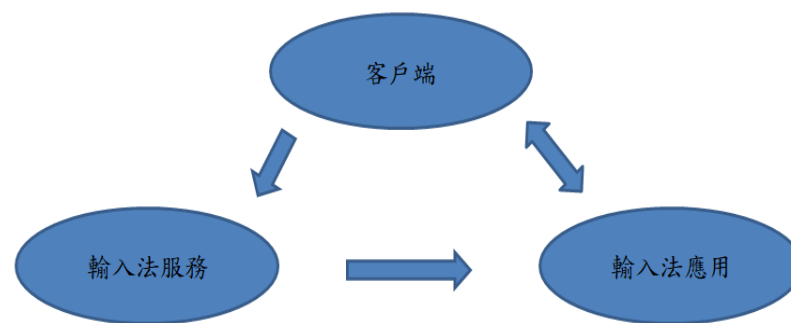
在「中轉客輸入法」中，以教育部的注音符號標示[5]為基準作為其輸入，以輸入中文詞的無聲調注音符號後，系統會以對應之客語詞做轉換，將其客語做輸出，也就是輸入中文詞輸出為客語詞；為讓使用者能快速的選出想要的字，

提出與實作注音字首縮寫輸入，以減少使用者的操作動作。這兩套輸入法的研開目的都是為了使客語能夠永續與易於推廣，經由手機的便利性與流通性以提升客語的使用與傳承。

目前在電腦平台上已有的輸入法: 教育部台灣客語拼音輸入法[5]，客語無聲調拼音輸入法[6]，信望愛客語輸入法[7]。相關拼音輸入法相關的研究與成果有:閩南語拼音輸入之台語兒[20]，手機實體按鍵與觸控介面之新式輸入法[10]，開發新式輸入法改善手機現有輸入問題;智慧手機平台四縣客語拼音輸入法，客語四縣腔輸入法;客語語音合成之初步研究[12];客語文章翻語音系統之實作[13];行動電話注音輸入法鍵盤配置之研究[21]中文文轉音系統中韻律階層的求取[29];以及客語《陸豐方言》百年語言演變析探[30]等。

## 二、智慧手機平台好客輸入法

自 Android 平台 1.5 版本以後，Google 已開放 Android 平台輸入法框架 (Input Method Framework, IMF)，IMF 是 Android 平台的特色設計，大大推動不具實體鍵盤之設備的產生。IMF 結構中包含三個主要的部分，如圖一所示：



圖一：輸入法整體架構

輸入法是用戶、應用程式之間的互動的渠道，所有的輸入法應用都需要繼承特定的 Android 平台提供的服務，Android 平台的輸入法框架為輸入法應用定義一個父類 `InputMethodService`，`InputMethodService` 提供一個輸入法的標準實現流程，定義輸入法生命周期內的重要函數，以方便開發人員對 Android 輸入法進行擴展。

## 手機鍵盤介面

在 Android 系統平台裡，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。系統模組包含好客輸入法 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，0 一為我們設計智慧型手機的鍵盤介面，本 APP 開發完成後即可加入手機，新增為使用者的客語拼音輸入法。



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。

### 一)客語詞典及功能

目前針對客語斷詞做標記的論文不多，交大電信工程蔡依玲的碩士論文[9]運用隱藏式馬可夫模型，我們採用客家委員會[11][12][13]和教育部[14][15]提供之客語例句與詞彙，並參考林昕緯斷詞標記模組工具[10]，如圖七所示，以結合人工標記與中文斷詞功能半自動式之客語斷詞及詞性等資訊而取得語料，並用該語料來建置客語辭典。

本文客語輸入法中，海陸腔客語語料詞典包含：

- 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。在單字拼音對照字庫資料表，含有 5366 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)，如表八所示。

輸入法會依照使用者輸入查詢【單字拼音】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。

【前後詞對照詞庫】資料表，有 8505 個詞彙數，則包含了 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)，如表九所示。前後詞對照詞庫資料表資料來源為客語能力認證基本詞彙-初級、中級、高級之例句做斷詞，再加以合成前後詞後，統計其在例句中出現之次數(Cnt)

臺灣客家語言中，因應客語文化的獨特性，另定一些自創之特殊字型，稱為「客語特殊字」。為了可以正確顯示出字型，客委會已電腦平台上提供造字編碼表及造字檔**錯誤! 找不到參照來源。**，供使用者下載安裝，約有 300 個特殊字，例如：「房間□」的「□」字，或者中文「我」的客語「**𠵼**」等。現在手機平台並無客語輸入法，導致手機上不易普遍使用客語文章，客語文句中的特殊字勢必增加客語使用與推廣的難度。手機作業系統中，客委會或相關軟體目前尚無提供相關的造字檔，因此無法於手機平台上顯示任何特殊字字彙。

### 三、輸入法之功能設計與結果分析

海陸腔輸入法的三個資料庫是藉由蒐集來的客語字及客語詞彙，在大量的客語文章內進行斷字及斷詞所整理來，因此提供四種功能：

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。
2. **前後詞預測模式：**鍵入客語客客、詞時，快速正確顯示下一個客語詞彙。
3. **使用者偏好輸入：**目的讓使用者能依照自己偏好快速將常用的字輸出。
4. **客語特殊字的處理：**解決手機端客語特殊字無法顯示之問題。
5. **中轉客輸入法：**中文輸入轉客語的功能可提供一般人不會客語、想學習客語的使用者，一個友善學習環境。

#### 一) 無聲調輸入模式

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內，並做輸出單字與詞彙的功能，表十四為輸入「向」客語拼音時，候選詞窗會顯示相關聯的拼音，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。

#### 二) 前後詞預測模式

在前後詞模式中，提供使用者能以最少操作動作來達成想要輸出的句子。在大量的客語文章及客語的句子做斷詞的動作後，在資料庫內設（**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，將斷詞完成的結果歸類在資料庫內。

*n*-gram model 是一種機率統計式的語言模型，依據樣本語料庫之文句序列（**sequence**）文字(character)或詞彙（**wrod**）單元加以訓練，獲得文句上下文文字與詞彙出現的機率值。*n*-gram 語言模型是自然語言處理技術的基礎，已廣泛應用於機器翻譯、語音識別與合成、手寫體識別、拼寫糾錯、漢字輸入，以及文獻分類與查詢等研究。

設  $S$  序列包含  $m$  個詞彙  $w_1, w_2, \dots, w_m$ ,  $S$  出現的機率  $P(w_1, w_2, \dots, w_m)$ , 為每一個詞出現的條件機率相乘：

$$P(S) = P(w_1, w_2, \dots, w_m) = P(w_1) * P(w_2|w_1) * \dots * P(w_m|w_1, w_2, \dots, w_{m-1}) \quad (1)$$

其中  $P(w_1)$  表示第一個詞  $w_1$  出現的機率； $P(w_2|w_1)$  則是已知第一個詞  $w_1$ 、第二個詞  $w_2$  出現的條件機率；同理，第  $m$  詞  $w_m$  出現的機率決定  $S$  序列開頭起至  $w_m$  之前面所有  $(m-1)$  的詞彙  $w_1, w_2, \dots, w_{m-1}$ 。

依拜氏定理(Bayes theorem)，條件機率  $P(B|A)$  可以計算求得機率值，惟面臨參數空間過太模型不易訓練以及資料嚴重稀疏(Data sparsity)，雖然可以運用平滑化(data Smoothing)方法克服部份問題，惟這些問題導致模型效益大大降低。實際上，任意一個詞的出現的機率僅僅與它前面出現的有限的一個或者幾個詞有關。

如果一個詞的機率僅限於它前面出現的  $n$  個詞有關，公式(2)即可化簡如下：

$$P(S) = P(w_1, w_2, \dots, w_m) = P(w_1) * P(w_2|w_1) * \dots * P(w_m|w_{m-n+1}, w_{m-n+2}, \dots, w_{m-1}) \quad (2)$$

在  $n$ -gram, 一般常用的模型多見  $n=1, 2, 3$ ,

如果一個詞的出現的機率僅於它前面出現的一個詞有關，那麼我們就稱之為 bigram model，公式(2)可表示如下：

$$P(S) = P(w_1, w_2, \dots, w_m) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) * \dots * P(w_m|w_{m-1}) \quad (3)$$

本文在  $n$ -gram 語言模型，目前訓練客語單字與客語詞之 1-gram 與 2-gram 語言模型，以計算出前後字或前後詞出現之條件機率。

### 三) 偏好輸入

每個人輸入文句時，會有個人常用的字或詞，這些常用字或詞會因為每個人的使用習慣而不同，因此提供有為個人使用習慣而建立的偏好功能，減少使用者在輸入時尋找的時間，此即個人化的使用詞彙環境。

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或詞的頻率；在系統內的 **pref** 一開始的初始值



為 0，每當輸入一個字或詞時，將紀錄在資料庫上的 **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，

#### 四) 客語特殊字顯示處理

客家委員會之客家語言有自創之特殊字型，為了可以正確顯示出字型，在電腦平台客委會提供造字編碼表及造字檔[17]，供使用者安裝，但礙於智慧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智慧手機上相關的造字檔呈現字型。

在徵詢多位客委會「客語薪傳師」後，我們採用此方式解決手機平台顯示客語特殊字的問題。

智慧手機中因為有特殊字無法顯示的問題，在這裡我們將特殊字以其「無聲調拼音」方式顯示，智慧手機能顯示出來，讓使用者能夠辨別，而不再是無法顯示的亂碼，圖三為客語特殊字(1)「𠵼(ngai)」與(2)「𠵼」拼音顯示。



圖三：特殊字顯示示意圖(1) (2)

#### 五) 中轉客輸入法

在目前臺灣的社會，年輕客家一輩多不會說自己母語，導致客語能力只剩聽的能力，甚至連聽的能力也已流失；此外，對於非客家的朋友，常因身旁沒會講客語者可以請教，因此學習客語成為一件困難的事。為此，我們提出中文輸入轉成客語的功能，可提供一般不會客語、想學習客語的使用者，一個友

善學習環境，並且在數位學習上新增客語學習方式。這項功能整合於我們的客語拼音的系統，以鍵盤的功能鍵切換「客語／中文」輸入法。

## 中文注音無聲調輸入

本輸入法設定使用對象為一般客語初學者，由於初學者大多數不會客語拼音，為此，我們特別設計使用「中文注音方式」輸入，並自動轉譯成客語輸出顯示於手機畫面上，如此可以讓不知客語詞彙的情形下，以替代方式輸入客語詞彙。

在中轉客辭典只有中轉客注音對照料表，有 5853 個詞彙數，包含了 Val、Chinese、Key、Pinyin、afterabber 及 Pref 六個欄位，分別代表了客語詞彙 (Val)、中文意義(Chinese)、注音縮寫(Key)、完整注音(Pinyin)、第一個字完整後面都縮寫(afterabber)及偏好偏好 Pref。

中轉客注音對照料表之建立，資料來源為客語能力認證基本詞彙-初級、中級、高級，將其正規劃後(如:E3)，並利用重編國語辭典修訂本[15]建立客語的中文意義注音，在建立中文注音的字首縮寫與第一個字拼音完整其他都字首縮寫(如:E4、E5)，並轉成對映按鍵之注音編碼其注音編碼方式，如表二所示。

在中轉客輸入法中，也運用在海陸腔無聲調輸入模式使用的資料探勘中的聚類分析的觀點。其目的為輸入中文注音拼音時，將中文對應的客語詞彙能更快顯示在候選詞窗中，並提供使用者輸出，圖七為輸入中文無聲調「土地公生日」之注音拼音轉成客語詞彙之過程。中文:土地公生日=>客語:伯公生。

首先輸入中文「土」字之注音拼音「ㄊㄨˇ」，其後選欄位將列出所有有關「ㄊㄨˇ」注音拼音之轉換客語後相關詞彙，當使用者輸入的關鍵拼音越多時，所顯示在候選欄的相關詞彙，將越接近使用者想輸入的詞彙，提供使用者做輸入動作。當輸入動作到第四步時，已經顯示出我們預先想要輸入的中文轉成客語詞彙「伯公生」。惟中文輸入詞如找不到合適客語詞，則無法轉換客語詞。

## 四、輸入法評量分析

本文在 Andorid 手機平台上研發海陸腔客語拼音輸入法，希望提供 3C 使用者具效率之客語輸入法，可以快速便利輸入客語文句，並且具有單字與

詞彙之客語語音輸出，提供一種手機平台上經由聽語音學習客語的方式。簡而言之，本輸入法的特色可以歸納如下表所示。

表一：本輸入法的特性說明

本輸入法功能	說明	備註
1 手機平台	可在 Andorid 手機或 PAD 平台環境輸入客語文句。	以海陸腔為主
2 輸入單字與詞彙	可以輸入單字、詞彙	
3 邊輸入邊搜尋	使用者一面鍵入拼音符號，即開始比對可能的單字或詞彙。	邊鍵入邊搜尋客語詞典
4 前後字預測	確定一字之後，可以預測該字之後可能的下一個字。	
5 前後詞預測	確定一個詞彙之後，可以預測該字之後可能的下一個詞。	運用統計式語言模型 bigram
6 客語特殊字呈現	目前客語單字中有數百個特殊字，以拼音顯示出來。	特殊字無法正常顯示在手上
7 語音學習	可對一個單字或一個詞彙作語路輸出。	
8 記憶個人化輸入語彙環境	可以記憶個人化輸入語彙環境	個人常用單字與詞彙，其出現順序往前遞增。
9 中轉客輸入法	本輸入法對於客語初學者或某些忘記的客語拼音，提供輸入中文轉換為客語的功能，可提供所需之客語拼音與詞彙。	

本輸入法具有上述之功能，為提供正確、效率化的輸入法，需經使用者的測試，作為修正與改進的依據。本文有關海陸腔客語輸入法的評量測試，採用平均主觀分數(Mean Opinion Score, MOS)的方式做為操作滿意度之分析，操作調查的受測者為懂客語拼音並且擁有客語初級、中級或高級其中之一認證者:客家委員會人員含客家薪傳師[18]羅文生、張旭英、何惠美等老師，

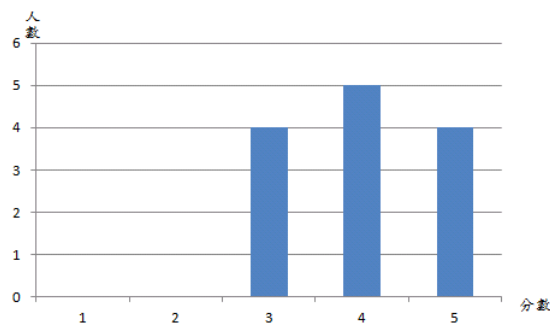
與懂客語但不會客語拼音的新竹在地海陸腔之使用者 10 人，總共合計 13 人，這些受測者使用客語海陸腔最少都有 20 年以上。

這些使用者提供五句客語句子跟五句中文句子輸入，操作完成後，依照受測者自己對者兩套輸入法做滿意度評分，滿意度評分分數為 1~5 分。

### 一)客語海陸腔輸入法之滿意度評量

由圖八可知，受測者在使用客語海陸腔輸入法後的滿意程度，大部分受測者覺得此輸入法入門門檻有點高，因為使用客語海陸腔輸入法所採用的拼音，需要經過相關的訓練才能使用此輸入法，因此覺得對不懂客語拼音的使用者來說很難操作，其總平均分數為 4.0 分，達良好的程度。

受測者之意見，主要有：本輸入法操作上滿順手，對於平常使用倉頡或注音輸入，反而比較慢，建議考慮具有操作者習慣的輸入方式；使用者必須具備客語能力能操作使用；很直覺化的拼音系統，感覺非常實用。

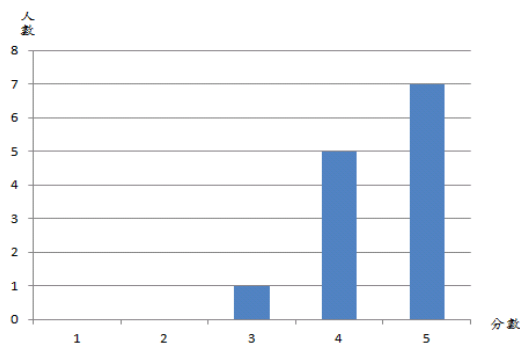


圖四: 海陸腔輸入法使用者評分

### 二) 中轉客輸入法之滿意度評量

由圖九所示，因臺灣人的輸入習慣大部分都是使用注音符號，因此中轉客輸入法對使用者來說，比較親切與順手，故大部分使用者的評分分數都較使用拼音方式的海陸腔輸入法高，其總平均分數為 4.46 分，接近很好的程度，這亦顯示提供「中文學習客語」之數位學習功能受到測試者肯定。

受測者意見主要有，本系統構想、做法很好，可讓想學習客語的人容易上手；希望後端詞彙庫能多點，讓系統能夠更佳完整發揮功能；希望能將海陸腔輸入法與中轉客輸入法兩套系統整合在一起，才不用做麻煩的切換動作。



圖五: 中轉客輸入法使用者評分

## 五、結論與未來方向

本文提出的智慧型手機客語海陸腔輸入法及中轉客輸入法，希望熟悉客語和不熟悉客語的使用者能夠使用，對於現在眾多輸入法當中，缺少的就是客家語言的輸入法，本文以研發海陸腔客語輸入法與相關應用為主，客語使用族群能夠運用母語在手機上交談，進一步達到客語的傳承與發揚。

在語料庫的收集過程中，我們使用客委會提供的初級、中級、高級基本詞彙及教育部台灣客家語常用詞辭典，總共 19315 個客語詞彙、5366 個客語單字、8050 筆客語前後詞詞彙，並使用中興大學資訊科學與工程學系林昕緯的斷詞標記模組工具[10]，做大量的客語文章斷詞作業來建立前後詞對照詞庫。

海陸腔輸入法及中轉客輸入法都以聯合大學資訊工程學系劉桂森的論文為基底，並加以提出使用客語特殊字的拼音解決客語特殊字無法顯示之問題，中轉客輸入法，以輸入客語詞之對應中文注音來達成輸入中文輸出客語之功能，並且加入散打功能，讓使用者能夠快速的找出想輸入的字詞。

在使用者評估調查中，調查了十三位客語海陸腔的使用者，其中有 3 位為現任客家薪傳師；在海陸腔輸入法的總平均 4 分，中轉客輸入法總平均分數為 4.46 分。由評估調查中可以發現，在中轉客輸入法分數較高，很可能因國人的

輸入使用習慣所造成的，在台灣大多數使用者使用注音輸入文句，所以使用者的輸入習慣大大的影響其評估分數。

對於本論文之研發與相關研究尚有需多待改進之部分，因此，在未來可進行的研究方向：

1. 將客語拼音輸入法與中轉客輸入法做整合，可省去切換輸入法之間的時間與操作。
2. 海陸腔相關資料較四縣腔來的少，因此還需要收集並整理出更多海陸腔相關資料，並建立出更加完整的資料庫系統。
3. 新增錯誤提示功能，藉由建議選字，使用者無須再打一次，可自動修正成正確的客語字詞。
4. 持續擴充客語語料庫內容，並訓練語言模型。

## 參考文獻

- [1] 黃河、陳信木，全國客家認同與客家人口之抽樣調查研究，台北：行政院客家委員會，2002。
- [2] 劉桂森，智慧手機平台中四縣客語拼音輸入法之研究與應用，聯合大學資訊工程學系碩士論文，2017。
- [3] 教育部客委會出版，客家語言拼音方案，2012。
- [4] 教育部，重修國語詞典修訂本，2015。
- [5] 教育部客委會出版，台灣客家語言拼音輸入法，2013。
- [6] 魏俊璋，客家無聲調拼音輸入法之研究與實作，國立中興大學資訊工程系碩士論文，2013。
- [7] 信望愛台語客語輸入法，2005，<http://taigi.fhl.net/TaigiIME/>，[存取日期：10.1.2018]。
- [8] 維基百科，客家裔臺灣人，<https://zh.wikipedia.org/wiki/客家裔臺灣人>。
- [9] 蔡依玲，基於隱藏式馬可夫模型之客語文具轉語音系統，國立交通大學電信成所碩士論文，2009。
- [10] 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。
- [11] 客家委員會出版，105年客語能力認證基本詞彙-中級暨中高級[海陸腔]-上冊，2016。
- [12] 客家委員會出版，105年客語能力認證基本詞彙-中級暨中高級[海陸腔]-下

- 冊，2016。
- [13] 客家委員會，105 年度客語能力認證基本詞彙-初級【海陸腔】，2016
- [14] 教育部客委會出版，教育部台灣客家語常用詞辭典。
- [15] 教育部客委會出版，重編國語辭典修訂本。
- [16] DATA MINING Concepts and Techniques(3th ed.). (2011). Morgan Kaufmann, Jiawei Han, MichelineKamber, Jian Pei.
- [17] 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：15/1/2018]。
- [18] 客家委員會-客語薪傳師，<https://master.hakka.gov.tw/>，[存取日期：15/1/2018]。
- [19] Ffmpeg，2011，<https://ffmpeg.org/about.html>，[存取日期：15/1/2018]
- [20] BaconBao，台語兒，google chrome store，2015，[存取日期：15 /1/2018]
- [21] 吳孟龍，手機實體按鍵與觸控介面之新式輸入法，國立成功大學工業設計學系碩士論文，2011。
- [22] 周宇軒，行動電話注音輸入法鍵盤配置之研究，華梵大學資訊管理學系碩士論文，2010。
- [23] Yabin Zheng, Chen Li & Maosong Sun,2011 “CHIME:An Efficient Error-Tolerant Chinese Pinyin Method”, IJCAI’11 Proceedings of the Twenty-Second international joint conference on Artificial Intelligence – Volume Three pp. 2551-2556.
- [24] Yabin Zheng, Lixing Xie, Zhiyuan Liu, Maosong Sun, Yang Zhang & Liyun Ru, 2011"Why Press Backspace? Understanding User Input Behaviors in Chinese Pinyin Input Method", HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics Volume 2 pp. 485-490.
- [25] 客家委員會，2016，<http://www.ihakka.net/epaper/1060203/epaper.htm>，[存取日期：22 /1/2018]
- [26] 五南圖書出版，”客語發音學”，2006。
- [27] 李雪貞，客語語音合成之初步研究”，國立臺灣科技大學資訊工程所，碩士論文，2002。
- [28] 林東毅，客語文句翻語音系統之實作，國立交通大學電信工程所，碩士論文，2007。
- [29] 蔡育和，中文文轉音系統中韻律階層的求取，國立中興大學資訊科學與工程所，碩士論文，2005。
- [30] 呂嵩雁，客語《陸豐方言》的百年語言演變析探，國立東華大學台灣語文學系，研究計畫報告書，2007。