

Taiwanese Hakka Across Taiwan Corpus and Formosa Speech Recognition Challenge 2025 – Dapu & Zhao'an Accents

廖元甫 Yuan-Fu Liao
陽明交大 National Yang
Ming Chiao Tung University
yfliao@nycu.edu.tw

郭志忠 Chih-Chung Kuo
陽明交大 National Yang
Ming Chiao Tung University
cck@nycu.edu.tw

黃昭世 Chao-Shih Huang
陽明交大 National Yang
Ming Chiao Tung University
hsf@nycu.edu.tw

藍瑀翔 Yu-Siang Lan
陽明交大 National Yang
Ming Chiao Tung University
yslan168@nycu.edu.tw

賴翰俊 Han-Chun Lai
華碩雲端 ASUS Cloud
patrick.lai@asuscloud.com

許文翰 Wen-Han Hsu
長問科技 BRONCI
Technology Inc
jeff.hsu@bronci.com.tw

摘要

為了重振瀕危的臺灣客家話，臺灣首個大規模客語語音語料庫 (HAT) 應運而生，該語料庫預計蒐集涵蓋臺灣各地腔調客家話的錄音。本文介紹 HAT 語料庫近兩年擴充的第二部分：大埔腔和詔安腔。此外，為了推廣此新建語料並評估目前最先進客語 ASR 系統的效能，特舉辦 2025 年福爾摩沙語音辨識挑戰賽—第二屆客語 ASR 競賽 (FSR-2025-Hakka ASR II)。共有 16 隊參加兩個賽道—客語語音辨識轉漢字、客語語音辨識轉拼音。最佳結果分別為：漢字一字元錯誤率 7.50%；拼音—音節錯誤率 14.81%。

Abstract

To revive the endangered Hakka language in Taiwan, the first large-scale Hakka speech corpus covering all aspects of Taiwanese Hakka across Taiwan (HAT) was created. This paper introduces the second part of the HAT corpus: the Dapu and Zhao'an accents. Furthermore, to promote this newly constructed corpus and evaluate the performance of the most advanced Hakka ASR system, the 2025 Formosa Speech Recognition Challenge, FSR-2025-Hakka ASR II, was held. Sixteen teams participated on two tracks: speech-to-Hakka-Hanzi and speech-to-Hakka-Pinyin. The best results were: Hanzi character error rate (CER) 7.50%; Pinyin syllable error rate (SER) 14.81%.

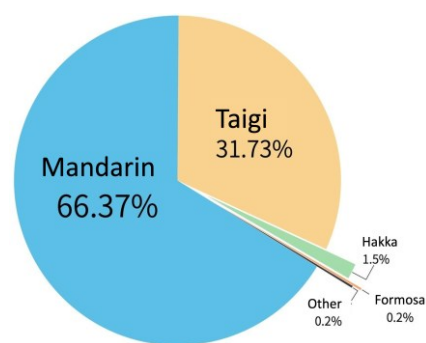


圖 1. 臺灣 2020 年人口及住宅普查報告中，本國籍常住人口主要使用語言統計。

關鍵字：臺灣客語、大埔腔、詔安腔、語音語料庫、自動語音辨識

Keywords: Taiwanese Hakka, Dapu accent, Zhao'an accent, speech corpus, automatic speech recognition (ASR)

1 Introduction

臺灣客語，簡稱為客語，曾是臺灣第二大語言，約有 15-20% 的居民使用。然而，由於華語的主導地位，客語如今已瀕臨滅絕。尤其是在 1949 年後，政府積極推廣國語（華語）作為臺灣的官方語言和主要語言。這種推廣導致其他語言的使用率和地位下降，包括臺灣客語。根據臺灣 2020 年人口及住宅普查報告 (主計總處 2020)，目前只有 1.5% 的人口使用客語作為主要語言 (見圖 1)；以客語為次要使用語言者也僅有 4.0%。換句話說，若不積極復振，客語在臺灣可能很快就會消亡。



圖 2. 臺灣客語五種腔調及其使用地區。

調號	1	5	2	3	7	4	8	
調名	陰平	陽平	上聲	陰去	陽去	陰入	陽入	
四縣	調值	24	11	31	55	2	5	
	調型	v'	v'	v'	v	vd'	vd	
南四縣	調值	24	33	11	31	55	2	5
	調型	v'	v'	v'	v'	v	vd'	vd
海陸	調值	53	55	24	11	33	5	2
	調型	v'	v	v'	v'	v+	vd	vd'
大埔	調值	33	35	113	31	53	21	54
	調型	v'	v'	v'	v^	v'	vd'	vd
饒平	調值	11	55	53		24	2	5
	調型	v'	v	v'		v'	vd'	vd
詔安	調值	11	53	31		55	24	43
	調型	v'	v'	v^		v	vd'	vd'

表 1. 六種腔調之臺灣客語拼音聲調表。

要振興客家話，不僅要妥善保存，更要引進一些現代語音人工智慧 (AI) 技術，例如自動語音辨識 (ASR) 和語音合成 (TTS)，以支援人們的日常生活。這些技術將幫助並鼓勵人們使用客家話。因此，一個大型臺灣客語語料庫 HAT (Hakka Across Taiwan)，即「HAT 臺灣客語語音資料庫」應運而生。並於 2023 年推出第一版 HAT-Vol-1 (客委會 2024)，包含臺灣客語最主要的兩大腔調：四縣與海陸。

HAT 語料庫旨在建立自動語音辨識 (ASR) 和語音合成 (TTS) 系統。為此當年也舉辦了 FSR-2023 - Hakka ASR (Formosa Speech Recognition Challenge 2023 - Hakka ASR)，即「福爾摩莎語音辨識挑戰賽 2023—客語自動語音辨識」(Liao et al. 2023；陽明交大 2023)，以推廣 HAT 語料庫，並評估正在開發的先進客家語自動語音辨識系統的性能。FSR-2023 只針對 HAT-Vol-1 所包含的四縣與海陸腔客語。今年隨著其他腔調語料逐漸蒐集完成，特別針對大埔與詔安兩種腔調再次舉辦 FSR-2025 - ASR II 客語語音辨識競賽(陽明交大 2025)。

本論文將首先簡要說明臺灣客語的腔調種類。其次介紹 HAT 語料庫在 2023 年之後增加的大埔腔和詔安腔，其中包含 251 位說話人，約 400 小時的 ASR 語音數據；以及 4 位說話人，共 125 小時的 TTS 數據。最後，主要說明 FSR-2025 挑戰賽，包括熱身賽與決賽的賽制、使用的資料集、參加的隊伍、比賽成績，以及成績較好的隊伍所使用的技術概述。

2 臺灣客語

臺灣客語是臺灣地區使用的客家話方言，主要由客家人使用。它是臺灣官方認可的國家語言之一，傳統上主要共有五種次方言腔調：

四縣、海陸、大埔、饒平、和詔安腔。圖 2 展示了臺灣客家話的五種腔調及其在臺灣的使用區域(賴維凱 2008)。另外，2011 年教育部國語推行委員會在審查《臺灣客家語常用詞辭典》時，正式決定將南四縣腔從四縣腔分開來；也就是將以屏東內埔為代表的南部六堆客家聚落，與北部以苗栗為代表的四縣腔，其詞彙與腔調變異在辭典中分別出來，稱為南四縣腔。

臺灣五種客家次方言中，使用最廣泛的是四縣和海陸。四縣有六個聲調，起源於廣東梅州；海陸有七個聲調，起源於廣東海豐和陸豐。四縣腔為目前使用人口最多的客家話，幾乎涵蓋北部、中部、南部與東部；海陸腔居次，使用人口主要以新竹、南桃園、苗栗北境及花東一帶。FSR-2023 挑戰賽，就是以四縣與海陸兩種腔調為語音辨識標的。

剩下較少人口的客語腔調中，大埔腔分布在臺中東勢、石岡、新社、和平、豐原及苗栗卓蘭；饒平腔主要分布在桃園、新竹、苗栗，彰化永靖次之；詔安腔是使用人口最少的，以雲林境內的二崙、崙背為主，桃園大溪次之(賴維凱 2008)。HAT 語料庫目前已經完成大埔、詔安兩種腔調的蒐集，因此今年的 FSR-2025 客語語音辨識挑戰賽就以此兩種語料資源極度稀少的腔調作為競賽標的。

臺灣客語的官方書寫系統是臺灣客語漢字。教育部分別於 2009 年和 2010 年提出了兩批「臺灣客語書寫推薦用字」(教育部 2025/7/10 修正；教育部 2024/8/26 修正)。漢字承載著客語的歷史文化內涵，並將其與更廣泛的漢語語言傳統聯繫起來。

大埔(Dapu) 語者人數						
年齡	~17	18~30	31~50	51~65	66~	合計
男性	9	5	12	20	20	66
女性	7	9	21	32	10	79
合計	16	14	33	52	30	145
詔安(Zhao'an) 語者人數						
年齡	~17	18~30	31~50	51~65	66~	合計
男性	2	3	7	23	16	51
女性	7	9	9	18	12	55
合計	9	12	16	41	28	106

表 2. HAT 語料庫中大埔腔和詔安腔 ASR 子集錄音語者的年齡和性別分佈。

除了正規文字以外，教育部也制定了客語音標，採用拉丁字母，稱為「臺灣客語拼音方案」(教育部 2024/8)。此拼音方案由教育部於 2012 年推出，旨在規範客家語發音的拼音，使不熟悉漢字客語發音的人更容易使用。

表 1 展示了六種腔調之臺灣客語拼音聲調。臺灣客語拼音方案中的聲調符號有兩種：調型與調值。調型類似注音符號中的聲調符號，有「/」、「∨」、「\」、「^」、「+」等；調值則是以阿拉伯數字，如 24、53，表示聲調的高低變化。聲調無論調型或調值，統一採音節右上標。例：客家 hag` ga` / hag² ga²⁴。上表中的 v 表舒聲韻音節；vd 表示入聲韻音節。大埔、詔安腔入聲調下加底線，表調值之短促，如：21、54。

教育部臺灣客語辭典的聲調標記，除了調型與調值外還有調號。客語調號和台語調號一樣，都是對應傳統聲調調名一平、上、去、入的序號，且分成陰陽兩組：調號 1~4 對應陰平~陰入；調號 5~8 代表陽平~陽入。不同腔調的客語，部分聲調種類會合併，因此實際上的聲調種類只有 6 或 7 種。其中，大埔腔另有超陰平調（調值 33），調號為 9，是陰平調的一種特殊變調。另外，南四縣腔部分地區（如美濃）之陰平調調值 33，與標準調值 24 不同，調型以「+」表示，但調號仍是 1。

在 HAT 語料庫與 FSR 競賽資料集中，聲調拼音一律採用調值表示，且不遵循客語拼音方案的上標表示規定，以方便電腦處理。但應當注意的是，不同腔調的相同調值，不一定代表相同的聲調種類。例如：調值 53，在海陸腔是陰平、在詔安腔是陽平、在饒平

音檔	大埔	詔安	合計
音檔筆數	75,239	98,119	173,358
語音句數	151,719	179,164	330,883
時分秒數	204:10:55	213:34:03	417:44:58

表 3. HAT 語料庫中大埔腔和詔安腔 ASR 子集語音資料的數量和長度統計。

音檔	大埔男	大埔女	詔安男	詔安女	合計
音檔數	17,592	16,012	17,616	16,661	67,881
語句數	37,701	34,473	35,773	33,852	141,799
時分數	31:30	30:21	32:25	31:27	125:44

表 4. HAT 語料庫中大埔腔和詔安腔 TTS 子集語音資料的數量和長度統計。

腔是上聲、在大埔腔則是陰去聲。所以，這會使得混合多種腔調的語料進行語音辨識模型訓練時，在拼音辨識上會變得更複雜。

3 HAT 臺灣客語語音資料庫

HAT 語料庫包含兩類錄音語料子集：自動語音辨識 (ASR)、文字轉語音合成 (TTS)。關於 ASR 與 TTS 子集的錄音設計、錄製方式，以及四縣腔與海陸腔的語料統計請參考 (Liao, 2023)。基本上大埔腔與詔安腔語料的錄製方式與四縣與海陸一致，主要差別在於合適的錄音人比較難找，所以蒐集到的語料之語者人數較少。另外語者的年紀也偏大，因此將年齡統計 50 歲以上再細分為 65 歲上下兩群。

在 ASR 子集的部分，經過兩年的努力，我們招募了 251 名錄音語者，共錄製了 330,883 句語音，相當於 417 小時的客家話語音，涵蓋了大埔和詔安兩個腔調。表 2 詳細列出了最終完成的大埔和詔安語音辨識 ASR 子集中說話者的年齡和性別分佈。從中可看出超過三分之一的語者年齡在 51~65 歲。表 3 則分別列出了大埔和詔安 ASR 子集中的音檔筆數、語音句數、和音檔時間總長的時分秒數。每個音檔可能包含多句語音，而其分句是根據錄音提示稿的標點符號。

在 TTS 子集的部分，每個腔調由一名男性和一名女性母語者錄製。基於 TTS 模型訓練的需求，每位錄音語者各自錄製了約 30 小時的語音。表 4 顯示了四名 TTS 錄音語者的音檔統計數據，包含音檔的檔案數、語句數、和長度時分數。

上述 HAT 語料庫中的每個語音檔都帶有相應的 JSON 格式元數據 (metadata)，其中包括錄音提示卡編號、音檔編號、音檔長度、提示句客語漢字與拼音及其華語翻譯，以及說話人特徵和錄音環境等資訊。例如：

```
DF1010001J2003_1.json
{
  "提示卡編號": "J2003"
  "音檔編號": 1
  "發音員編號": "DF101"
  "音檔長度": "00:04"
  "客語漢字": "五月節愛食粽。"
  "客語拼音": "ng31 ngied54 zied21 oi33
    shid54 zung53 。"
  "華語字": "端午節要吃粽子。"
  "性別": "女性"
  "年齡": 72
  "身分別": "薪傳師 (教師)"
  "現居地": "臺中市東勢區"
  "18 歲前居住地": "臺中市東勢區"
  "教育程度": "大學"
  "錄音腔調": "大埔"
  "錄音環境": "一般辦公室"
  "流暢度": "普通"
}
```

4 FSR-2025 - ASR II 客語語音辨識競賽

為了推廣 HAT 語料並評估先進客語自動語音辨識系統的性能，FSR-2025 挑戰賽於 2025 年 6 月 2 日至 10 月 6 日舉行，歡迎學術界和產業界組隊參加。最後共有 20 隊報名，並有 16 隊完賽，包含學生組 13 隊、社會組 3 隊。

FSR-2025 挑戰賽的主要任務是建造一個能辨識臺灣大埔腔與詔安腔客語之語音辨識器，該任務根據語音辨識輸出分為兩個賽道 (Track)：

(1) Track1: 臺灣客語漢字—語音辨識輸出漢字
例如：今晡日係拜二。

辨識率計算漢字字元錯誤率 (CER)。

(2) Track2: 臺灣客語拼音—語音辨識輸出拼音
例如：gim24 bu24 ngid2 he55 bai55 ngi55。

辨識率計算拼音音節錯誤率 (SER)。

挑戰賽共進行兩次評測：熱身賽、決賽。熱身賽的目的是讓參賽者先進行一次模擬測

	語者數	語句數	字元數	時數	
total	> 51	36,316	542,617	82.23	
讀稿	17	12,197	180,055	31.43	大
讀稿	16	15,152	199,870	30.59	詔
Train	33	27,349	379,925	62.02	
讀稿	3	1,304	18,639	4.00	大
即席	3	445	11,877	1.08	埔
小計	> 3	1,749	30,516	5.08	腔
讀稿	6	2,154	26,105	4.00	詔
即席	6	501	15,006	1.13	安
小計	> 6	2,655	41,111	5.13	腔
Eval	> 9	4,404	71,627	10.21	
讀稿	5	1,122	15,390	2.97	大
即席	5	904	27,496	2.03	埔
小計	> 5	2,026	42,886	5.00	腔
讀稿	4	1,322	15,119	2.57	詔
即席	4	1,215	33,060	2.43	安
小計	> 4	2,537	48,179	5.00	腔
Test	> 9	4,563	91,065	10.00	

表 5. FSR-2025-Hakka II 語音資料集。

試，以驗證其系統，故成績僅供參考。最後的決賽成績才做為競賽的結果與排名。

配合競賽的時程，依序釋出下列競賽用語音資料集 (相關統計參見表 5)：

- 訓練集 (FSR-2025-Hakka-Train)：來自 HAT 語料庫約 60 小時、33 位語者的語料，包含大埔腔和詔安腔約各半的語料和語者。參賽者報名後即可獲得此訓練集。
- 驗證集 (FSR-2025-Hakka-Eval)：這是在熱身賽階段釋出作為驗證測試之用。如表 5 中所示，除了來自 HAT 語料庫的讀稿 (read) 錄製 (Record) 語料外；為了提高語音辨識難度，我們另外增加了從媒體 (Media) 節目中擷取的即席 (spontaneous) 語音。驗證集語音總時數約為 10 小時，兩腔調各 5 小時，讀稿和即席語音的時數大約為 4 比 1。
- 測試集 (FSR-2025-Hakka-Test)：這是在決賽中的測試語音。與熱身賽驗證集大小一樣都是 10 小時，且也都有來自 HAT 錄製的讀稿語音和來自媒體的即席語音，但是兩者的比例約為 1 比 1。因為較難辨識的即席語音比例提高，因此決賽的挑戰更大。

隊伍編號	漢字 CER	拼音 SER	漢字名次	拼音名次	幾何平均
U	6.84%	19.57%	1	3	2
M	8.84%	13.44%	3	1	2
D	10.51%	14.72%	8	2	4
P	7.64%	89.49%	2	13	5
T	10.01%	47.14%	4	10	6
H	15.92%	20.49%	10	4	6
B	10.42%	23.40%	7	6	6
F	17.92%	21.91%	12	5	8
R	10.33%		5	17	9
E	16.06%	29.09%	11	8	9
J	10.37%		6	17	10
K	13.36%	70.68%	9	12	10
O	101.63%	25.52%	17	7	11
C	37.64%	35.75%	15	9	12
L	31.93%	51.47%	13	11	12
G	75.58%	100.97%	16	14	15
N	36.68%		14	17	15

表 6. 熱身賽辨識結果與排名。

	讀稿	即席	綜合	大埔	詔安
漢字 CER	1.82%	15.03%	6.84%	3.96%	8.99%
拼音 SER	6.51%	21.66%	13.44%	9.02%	16.75%

表 7. 熱身賽各種語料類別之最佳辨識結果。

4.1 熱身賽結果

表 6 列出熱身賽參賽 16 隊與主辦單位建置之基準 (Baseline) 系統 (隊伍編號 B)，共 17 組針對驗證集 (FSR-2025-Hakka-Eval) 之辨識結果與排名。總排序是依據漢字排名和拼音排名的幾何平均，若平均相同則依據漢字排名。隊伍編號除 B 以外，A~P 為學生組，Q~U 為社會組。最後 A, I, Q, S 四隊未提交結果參與評測，故不予排名。剩下參賽 16 隊中有 3 隊未提出拼音辨識結果，故假定其拼音排名並列 17。Track1 漢字賽道參賽 16 隊中有 6 隊成績優於基準；Track2 拼音賽道參賽 13 隊中則有 5 隊成績優於基準。總排序則共有 6 隊在基準系統 B 前面。

表 7 顯示最佳結果是 Track1 漢字 CER = 6.84%；Track2 拼音 SER = 13.44%。顯然漢字的辨識率優於拼音，錯誤率相差一倍。若依據語音性質分類，可以看到即席語音辨識率遠低於讀稿語音，這也是因為訓練語料中只有讀稿語音。另外若依據腔調分類，則可以

隊伍編號	漢字 CER	無調拼音 SER	拼音 SER	漢字名次	拼音名次	幾何平均
E	7.50%	17.45%	25.04%	1	6	2
P	8.99%	12.36%	19.22%	2	3	2
D	11.21%	11.32%	15.08%	4	2	3
M	22.50%	10.49%	14.81%	13	1	4
U	9.46%	20.97%	30.44%	3	8	5
H	15.73%	13.82%	20.68%	7	4	5
B	17.13%	14.43%	23.50%	9	5	7
L	16.05%	18.97%	27.57%	8	7	7
O	15.61%	26.60%	35.20%	6	11	8
J	13.35%			5	17	9
S	18.78%	21.30%	33.38%	10	10	10
N	28.70%	20.70%	30.45%	15	9	12
K	21.21%	37.80%	47.95%	12	12	12
R	19.08%			11	17	14
C	30.40%	56.67%	59.35%	16	13	14
F	26.43%			14	17	15
G	503.0%	503.4%	503.4%	17	14	15

表 8. 決賽辨識結果與排名。

看到大埔腔的結果優於詔安腔，其中原因尚待探究。

4.2 決賽結果

表 8 列出決賽參賽 16 隊與基準系統 B，共 17 組針對測試集 (FSR-2025-Hakka-Test) 之辨識結果與排名 (注意：T 隊只參加熱身賽；S 隊只參加決賽)。有 3 隊沒有提出拼音辨識結果，故假設其拼音排名並列 17。共有 6 隊平均排名領先基準系統 B，其中編號 E 的隊伍成績進步最大，從熱身賽平均第 9 名進步到決賽時並列最優的第 2 名；其它 5 名在熱身賽時也排在基準系統 B 之前。當中兩次拼音排名第一的 M 隊，其漢字成績異常，名次從第 3 掉到 13，可能是決賽的系統有問題。

決賽額外計算了「無調拼音 SER」，就是拼音結果不管聲調對錯，只計算非聲調部分的音節錯誤率。從結果來看，拼音不管聲調後 SER 明顯進步很多，除了漢字成績異常的 M 以外，還有 H, B, N 這三組的無調拼音 SER 低於漢字 CER。這可能表示目前混合大埔與詔安腔的辨識模型，在聲調方面採取調值表示會產生混淆問題；而且目前都是標示本調，

編號	CER SER	名次	賽道	主要特色	Whisper	variant model	語料 1	語料 2	TTS	DA	DA Robust strategy
E	7.5%	1	漢	兩階段 FT 全參數 FT	Large-v2		HAT-Vol-1	FSR-train		SpecAugment	
	25.0%	6	拼	後處理：①二腔漢字 G2P 比較拼音，取 WER 小者②字典比對修正	同上		同上	FSR-train → 漢字 Encoder + 拼音 Decoder		同上	
P	9.0%	2	漢	dialect-aware special tokens	Large-v3	FormoSpeech Hakka	FSR-train FSR-eval		VoxHakka VC speaker DA	SpecAugment -freq. masking -time masking	Progress. Augm. : Speed+Spec+Noise
	19.2%	3	拼	同上	同上		同上		同上	同上	同上
D	11.2%	4	漢	降噪前處理	Medium		FSR-train FSR-eval	LoRA 效果不佳		SpecAugment Audio Concat.	
	15.1%	2	拼	同上	同上		同上	同上		同上	
M	22.5%	13	漢	N-best 文本候選 mBART 文本修正 x5 RNNLM 重評分	Large-v3 +LoRA	3 種模型 (大埔/詔安/混合) 輸出 5-best 候選句	FSR-train FSR-eval 處理合音字+刪錯讀			①MUSAN Noise SNR ∈{5,10,15}dB ②Speed perturb.	
	14.8%	1	拼	腔調辨識+腔調 ASR 模型 RNNLM 重評分	Medium	2 種模型 (大埔/詔安) 2 選 1	同上	同上		同上	
U	9.5%	3	漢		Large-v3 +AdaLoRA	FormoSpeech Hakka	FSR-train FSR-eval	詔安 OOV (TTS)	formospeech/yourtts-htia-240704		速度、音高變動與空氣吸收
	30.4%	8	拼	Kaldi K2 模型	n/a	Wenet-Zipformer	HAT-Vol-1 = 四 396h + 海 300h	FSR-2023 FSR-train FSR-eval	同上		
H	15.7%	7	漢		Large-v3		FSR-train FSR-eval	Radio 11hr E-learning 16hr	VITS: 辭典例句+Media 語料	Static: MUSAN, MetricAug	Dynamic: Audiomentations
	20.7%	4	拼		同上		同上			同上	同上
B	17.1%	9	漢		Large-v3 Turbo		FSR-train	串接音檔 159.5 小時		Audio Concat. SpecAugment	Audiomentations
	23.5%	5	拼		同上		同上	同上		同上	同上

表 9. 前六名領先隊伍與基準系統(編號 B)的主要作法摘要。

但兩種腔調有各自不同的變調規則，混為一體恐難以訓練建模。

最後，比較表 8 與表 6 的結果，可以看出相對於熱身賽階段，決賽時領先群隊伍的系統都有進步。因為雖然決賽測試集的難度比熱身賽驗證集的難度要高不少（即席語音比例從 20% 增加到 50%），但最佳成績差異不大。決賽最佳結果分別是 Track1 漢字 CER = 7.50%；Track2 拼音 SER = 14.81%。這當中的改進至少應該包含模型訓練增加了熱身賽驗證集 (FSR-2025-Hakka-Eval) 的語料，尤其當中所含的媒體語料是原來訓練集中完全欠缺的即席語音類型。

4.3 領先隊伍作法摘要

表 9 是從參賽隊伍投稿的論文中，簡單摘要出領先六隊跟基準系統的主要作法。除了 U 隊在拼音賽道採用 Kaldi K2 模型外，全部參賽系統都是基於 Whisper 核心模型的各種版本。P、U 兩隊的漢字賽道使用 FormoSpeech 公開的模型 whisper-large-v3-taiwanese-hakka (FormoSpeech, 2025) 做為基底，此模型是將 Whisper Large-v3 經過六種客語腔調的微調訓練而得。其餘大部分系統都直接使用 Whisper Large-v3，只有兩隊使用 Medium、一隊使用 Large-v2，唯有基準系統採用 Large-v3 Turbo。

若觀察漢字賽道第一名的 E 隊，其主要特色是以兩階段微調訓練 Large-v2 模型：第一階段訓練語料是 2023 年釋出的 HAT-Vol-1 語料庫，包含四縣腔 396 小時、海陸腔 300 小時；第二階段訓練語料是大會提供的大埔腔與詔安腔語料—FSR-train 62 小時、FSR-eval 10 小時。另外也簡單提到使用 SpecAugment (Park et al., 2019) 進行資料增強。他們的模型則是使用 ESPNet toolkit (Watanabe et al., 2018) 進行實現。

漢字第二名 P 隊與第三名 U 隊辨識率接近，兩隊的作法也有很多相似之處，都是使用 FormoSpeech 的客語基礎模型，也都使用 TTS 來擴充訓練語音。但 P 隊的特色是針對 Whisper 設計特殊 token 來辨識腔調，包括大埔腔、詔安腔、與未知腔，藉此讓模型對腔調有意識的進行區分來影響語音辨識結果。另外與第一名相同，P 隊也採用 SpecAugment 進行資料擴增，U 隊則無。

在拼音賽道方面，第一名 M 隊與第二名 D 隊的辨識率相近；兩隊的共通點就是都採用較小的 Whisper Medium 模型。M 隊的特色就是額外訓練一個 RNNLM 語言模型（漢字採用 LSTM、拼音採用 GRU），對於 Medium 模型輸出的 10-best 候選句來進行重評分；另外就是利用漢字辨識結果進行斷詞與查辭典，藉以辨識腔調種類，再以腔調專屬拼音辨識模型來進行辨識。此外，M 隊對於大會提供的語料也進行過濾，刪除備註語者錯讀的音檔，也處理合音字的特殊表示。

拼音第二名 D 隊的特色則是對語音進行降噪前處理，包括訓練語音和測試語音；另外也採用 SpecAugment 和串接短音段來進行訓練語料擴充。拼音第三名是 P 隊，其辨識率跟前兩名差距較大。P 隊的拼音作法幾乎與漢字作法完全一樣，只差在拼音是使用 Large-v3 原生模型，而不是 FormoSpeech 的客語模型。

5 結論與展望

本文不僅回顧了 HAT 語料庫第二階段中的大埔腔與詔安腔子集，也介紹了針對這兩種腔調的 FSR-2025 客語語音辨識挑戰賽。HAT 計畫也將持續錄製剩餘的客家方言。由於目前的 HAT 語料庫專注於讀稿語音，因此也開始著手轉錄來自電視、廣播等傳統媒體以及 YouTube、播客等線上平台的自發性即席語音，

尤其是訪談形式的語音。最終目標是收集足夠的台灣自發性客家話語音數據，藉以建立更先強韌可靠的客家語自動語音辨識系統。

Acknowledgments

This work was supported in part by the Co-creation Platform of the *Speech-AI Research Center (SARC)*, *Industry-Academia Innovation School (IAIS)*, NYCU, under the framework of the National Key Fields Industry-University Cooperation and Skilled Personnel Training Act, from the Ministry of Education (MOE), the National Development Fund (NDF), industry partners in Taiwan, and in part by the *Hakka Affairs Council (客委會)*.

References

- FormoSpeech. 2025. whisper-large-v3-taiwanese-hakka. <https://huggingface.co/formospeech/whisper-large-v3-taiwanese-hakka>. Accessed: 2025-09-10.
- Y.-F. Liao et al. 2023. *Taiwanese Hakka Across Taiwan Corpus and Formosa Speech Recognition Challenge 2023 - Hakka ASR*. *26th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, Delhi, India, 2023, pp. 1-6, doi: 10.1109/O-COCOSDA60357.2023.10482979.
- Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, and Quoc V. Le. 2019. SpecAugment: A simple data augmentation method for automatic speech recognition. In *Interspeech 2019*, pages 2613 – 2617.
- Shinji Watanabe, Takaaki Hori, Shigeki Karita, Tomoki Hayashi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalta Soplín, Jahn Heymann, Matthew Wiesner, Nanxin Chen, Adithya Renduchintala, and Tsubasa Ochiai. 2018. Espnet: End-to-end speech processing toolkit. In *Interspeech 2018*, pages 2207 – 2211.
- 主計總處 2020：表 6、6 歲以上本國籍常住人口使用語言情形。109 年人口及住宅普查>綜合報告統計表。Accessed: Nov. 2, 2025. [Online]. https://www.stat.gov.tw/News_Content.aspx?Create=1&n=2755&state=1327FD6AD8DCDA52&s=230300&ccms_cs=1&sms=11065
- 客委會 2024：臺灣客語語音資料庫(HAT)。中華民國計算語言學學會/語料庫。https://www.aclclp.org.tw/use_mat_c.php#hat。
- 教育部 2024/8：臺灣客語拼音方案使用手冊。教育部語文成果網/字音類/臺灣客語拼音。

Accessed: Nov. 2, 2025. [Online]. Available:
<https://language.moe.gov.tw/index.aspx>

教育部 2024/8/26 修正：臺灣客語書寫推薦用字
(第 2 批)。教育部語文成果網/字形類/臺灣客
語用字。Accessed: Nov. 2, 2025. [Online].
Available: <https://language.moe.gov.tw/index.aspx>

教育部 2025/7/10 修正：臺灣客語書寫推薦用字
(第 1 批)。教育部語文成果網/字形類/臺灣客
語用字。Accessed: Nov. 2, 2025. [Online].
Available: <https://language.moe.gov.tw/index.aspx>

陽明交大 2023：“Formosa Speech Recognition
Challenge 2023 - Hakka ASR.” Accessed: Aug. 23,
2023. [Online]. Available:
[https://sites.google.com/nycu.edu.tw/fsw/home/cha
llenge-2023](https://sites.google.com/nycu.edu.tw/fsw/home/challenge-2023)

陽明交大 2025：“Formosa Speech Recognition
Challenge 2025 - Hakka ASR II.” Accessed: Aug.
23, 2023. [Online]. Available:
[https://sites.google.com/nycu.edu.tw/fsw/home/cha
llenge-2025](https://sites.google.com/nycu.edu.tw/fsw/home/challenge-2025)

賴維凱 2008：臺灣客家話的分布及使用概況。教
育部臺灣客語辭典/客語資源/客語知識庫/綜
論。Accessed: Nov. 2, 2025. [Online]. Available:
<https://hakkadict.moe.edu.tw/resource/>