

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

客語文句轉語音及語音辨認之研究(1/3)

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 96-2221-E-009-030-MY3

執行期間 96 年 8 月 1 日至 97 年 7 月 31 日

計畫主持人：陳信宏

共同主持人：余秀敏、羅烈師

計畫參與人員：蕭希群、楊智合、黃信德、江振宇

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

處理方式：可公開查詢

執行單位：國立交通大學

中 華 民 國 97 年 5 月 30 日

摘要

本計畫使用中文語音合成及辨認技術，將現有的客語語音合成系統與辨認模組做進一步的改進，改進項目包括：詞典資訊的擴充、語料庫的收錄與整理、客語 Text Analysis 斷詞正確性的改進、以及建立以 word n-gram 為基礎的客語辨認用 Language Model，以強化客語語音合成與辨認的效能。

關鍵詞：客語語音合成、客語語音辨認

一、前言

現有的客語語音合成與辨認系統所共同面臨的狀況，就是詞彙資訊量及資料量的不足所造成的各種問題。在語音合成方面，詞典資訊量不足所造成的不良結果除了在針對文章作斷詞分析(Text Analysis)時無法得到正確的分詞結果外，對於合成聲音所需的韻律(Prosody)也會造成不順暢的預估。另外，訓練韻律產生器(Prosody Generator)以及實現 Corpus-based TTS 系統所需的語料若不足，亦會造成合成效果不如預期。在語音辨識方面，現有的客語語音辨認系統受限於語料庫的不足，無法建立良好的聲學模型(Acoustic Model)及語言模型(Language Model)，導致辨認效能無法提升。

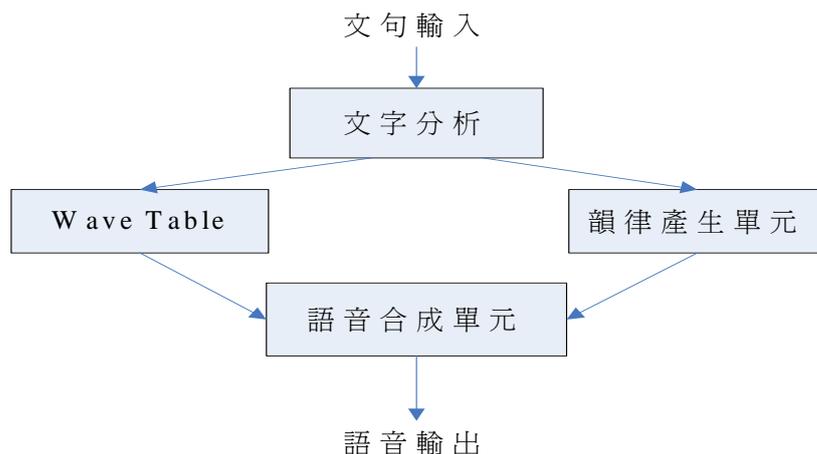
二、研究目的

基於上述論點，本計畫除致力於建立客語語音合成及辨認所需的語料庫外，亦希望取用目前中文語音合成與辨認技術與資源[3, 4, 5, 6]，輔助客語系統的建立。技術部分包含引用中文定量複合詞規則、詞綴構詞，以及專有名詞(地名、機構等)的結構化資訊，來建立相對的客語構詞規則，以改進客語 TTS 文句分析的效能。另外，我們也進行以大量的中文語言模型，利用語言模型調適方式(Language Model Adaptation)，調整成客語能夠使用的語言模型，以提升客語語音辨認的效能。

三、研究方法

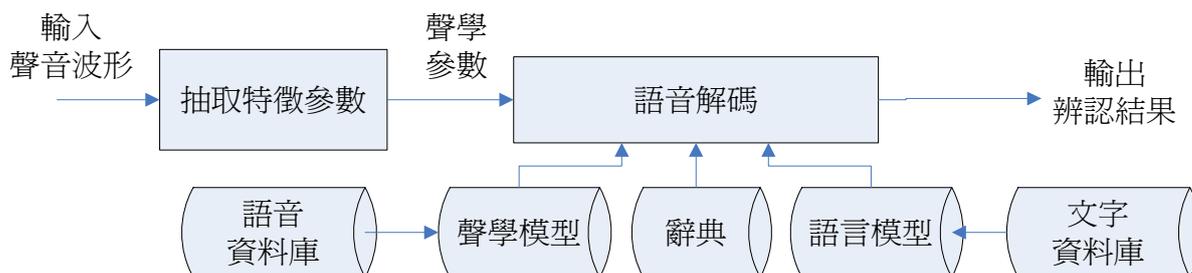
1. 客語語音合成與辨認模組簡介

我們已建立一個基本的四縣客語文句轉語音(Hakka Text-to-Speech, HTTS)系統[1, 2]，其基本模組如圖一所示，它係採用我們過去發展國語文句轉語音子系統相同的架構，系統包含四個模組：文句分析(Text analysis)、韻律產生單元(Prosody generation)、基本波形單元(Wave table)、語音合成單元(Speech synthesis)等。文句分析單元將輸入文句做斷詞分析，以產生詞串、詞類串、音節串；韻律信息產生單元則是由文句分析的結果抽取一些語言參數，來產生適當的韻律信息，用以合成自然流利的語音；基本波形單元儲存 671 個基本音節的波形訊號；語音合成單元則是使用韻律信息來及調整基本音節波形，以串接的方式產生自然流利的語音。



圖一：客語文字轉語音雛型系統之基本架構圖

一個完整的客語辨認系統基本模組則如圖二所示，我們已建立其中的四縣腔客語的聲學模型，四縣客語語音中共有 671 個基本音節，其中聲母有 17 個(包含空聲母)而韻母有 71 個。我們使用和國語語音辨認器相同的次音節 HMM 模型，包括 87 個右相關聲母模型和 71 個韻母模型；其中右相關聲母模型狀態數目為 3 而韻母模型狀態數目為 5。建立此聲學模型使用的語料庫是由 35 個語者(含 18 男性及 17 女性)所錄製，總共有 34,818 個音節。



圖二：客語語音轉文字雛型系統之基本架構圖

2. 客語語音合成與辨認模組的改進

在已建立的客語 TTS 系統及聲學模型上，我們持續進行以下的技術改進，希望能大幅提升客語 TTS 及語音辨認的效能：

(1) 文字分析正確性的改進：文字分析模組效能的好壞最主要依靠兩個主要部分，一個是詞典的收錄，另一個則是構詞規則的套用。目前詞典收錄情形如表一，共含 36,258 個詞條，主要來源為台北市客委會以及台大客家社所提供的客語電子詞典，此外，也從《客語實用手冊》(龔萬灶著)、《中級客語認證詞庫》(行政院客委會)、以及陳碧娥老師撰寫之客語文稿等資料中，補充詞典所缺列的詞彙。表一為依詞長度統計的詞數。

表一：客語詞典的詞長統計

一字詞	二字詞	三字詞	四字詞	五字詞	六字詞	七字詞	八字詞	總詞數
7325	18549	5927	3776	413	139	99	30	36258

在詞典詞條收錄方面，今年我們引進了由行政院客委會所編定的《中級客語認證詞庫》，此詞庫以四縣腔調之客語為主，但尚包含海陸、饒平、大埔、及福建的詔安等腔調，收錄共約五千詞條，並將各詞條依據語意及用法上區分為三十類，每類詞彙數從數十到數百不等。除了詞彙上的分類，此詞典並含有各詞條對應的中文詞，這對於往後希望將中文文字處理方式引入客語來說是一個很豐富的資源。

然而由文字分析的角度而言，光是詞典的收錄並不足以將文字做正確斷詞，從中文文字分析模組發展的過程中我們發現，在針對文句中定量複合詞、詞綴以及專有名詞定義上的精確度，影響著整體的文字分析效能，若無法針對上述情況建立完整的分析方法，則不只是影響上述類別詞條本身的正確性，也會造成詞條周遭的詞受到影響，而造成搶詞或者過份構詞的情況。

目前在中文定量複合詞的分析上，針對各類的定詞、量詞、時間詞、地方詞等基本構詞單元有著詳細的定義，並且針對各類集合建立各種構詞規則及構詞限制，並且以實際自動處理大量語料後的結果分析各類規則所產生的例外情形。另外對於詞綴詞以及專有名詞的分析上，我們引用部分中央研究院對於詞綴及名詞的分類方式，建立了階層性的分析方法，以更精確的方式定義詞條與詞條間的關連性。舉例來說，我們可以針對一串地名詞條建立專有名詞前後間的從屬階層關係，將此階層架構所組成的詞組(word chunk)從文句中定義出與其他詞條的邊界，以降低其他詞條與所構成詞搶詞的機率。

我們目前正進行將此分析方法套用在客語文字分析模組裡，以定義符合客語語言特性的構詞單元集合以及構詞規則；對於詞綴及專有名詞部分，我們比較中文與客語之間結構上的異同，建立屬於客語的詞綴及專有名詞處理模組。期望在加入部分語法結構資訊後，達到更好的斷詞結果。

(2) 以客語 corpus-based TTS 為基礎的語料庫的建立：目前客語 TTS 系統仍舊為 PSOLA-based 架構，主要原因是以目前語料庫收錄的狀況仍不足以建立需要大量語料的 corpus-based TTS 系統。我們在針對語音合成所收錄的語料部分，針對四縣腔，我們請一位發音純正的龔萬灶老師錄製，使用的文本為龔老師所出刊的客語散文集「阿鶯箭介故鄉」。錄音文章數為 42 篇，共 72064 音節，目前已錄製完成，正著手將所錄製的音檔先經由自動切割後，以人工方式修正為正確切割位置。由於 corpus-based TTS 系統對於語音切割位置精確度的要求極高，這項工作必須以極細緻且正確的方式進行。另外在文字檔部分，我們也請龔老師以人工標示上正確發音，再以自動斷詞加上人工修正等方式標示出正確的斷詞及 pos 標示。另外，海陸腔部分，由羅烈師老師自己錄製，使用文本為范文芳老師所著「頭前溪的故事」，全書約兩萬五千字，已依文意切割成 150-250 字數之短文，目前錄音程序與環境皆已穩定，約錄完 2000 音節，將可快速增加錄音進度。文字檔方面已逐篇進行通用拼音系統之註記，待錄製到一個階段後，將開始著手斷詞、pos 標示以及語音波形的切割。

(3) 客語語音辨識語言模型(Language Model)的改進：我們由新增語音錄製及語言模型建立兩方面著手。在新增語音錄製方面，我們對四縣腔客語及海陸腔客語分別進行，在四縣

腔客語的錄製上分為兩種類型，分別為單一語者及多語者兩類。單一語者語音資料方面，我們請一位多年來從事客語教學的陳姓老師幫忙撰寫客語文章，並為這些文章標上四縣音標後錄製成語音，總共錄製 30918 個音節。而在多語者語音資料方面，所使用的文字資料則是從陳老師所寫文章中抽取適合的短句錄製。這個錄音計畫規畫錄製男女各 50 人共 100 人，目前男女錄製人數已經各達 22 人，總共的音節數量為 18620 個音節。海陸腔客語的錄製方面，目前已完成錄音卡的製作，正進行語音錄製中。

目前客語語音辨識系統並沒有加入 word n-gram 統計資訊，原因是客語的文字語料不足，訓練語言模型所需要的斷詞資訊也還不完整。目前我們正著手以語言模型調適方式，將中文語音辨認所建立起來的語言模型，以少量客語資料作為調適用的訓練語料，以建立能提供客語辨認使用的語言模型。具體而言，我們是以目前中文辨認系統在建立的階層式語言模型(Structured Language Model)技術，將其運用在客語語言模型上。在中文語言模型的建立上，我們將定量複合詞、詞綴、專有名詞、人名等有特殊結構的詞，抽取其結構，並以 Finite State Machine 等方式建立統計式的機率模型，並在語音解碼步驟採取兩階段的方式，將詞彙的結構資訊加入語音辨認的流程。以此語言模型為基礎，再由少量客語文字結構的分析得到可用於 Language Adaptation 上的統計資訊後，將中文的語言模型調整成客語的語言模型。我們相信以語言調適的方式，將可解決客語語音辨認中對於 Language Model 訓練語料不足的困境。

四、結果與討論

本計畫正進行客語 TTS 及辨認系統的改進，除由基本的詞典擴增、新語料錄製上著手外，亦使用中文語音合成及辨認技術及資源，對 TTS 的文句分析及辨認的語言模型進行改進，在本年度結束時將可完成初步的系統改進，開發可用的四縣客語 TTS 系統及四縣客語辨認系統。

成果發表之論文：

- [1] Hsiu-Min Yu, Lie-Shih Lo, Hsin-Te Hwang, Hsi-Chun Hsiao, and Sin-Horng Chen, "On Constructing Speech Corpus for Implementing Hakka Text-To-Speech Synthesis," in Proceeding of OCOCOSDA, Dec. 2007.

文獻探討

- [1] Hsi-Chun Hsiao, Hsiu-Min Yu, Yih-Ru Wang and Sin-Horng Chen, "Multilingual Speech Corpora for TTS System Development", Int. Symp. on Chinese Spoken Language Processing, Dec. 2006, Singapore; and Lecture Note in Computer Science, Vol. 4274/2006, Chinese Spoken Language Processing, Springer, pp.748-759 (SCI)
- [2] Hsiu-Min Yu, Hsin-Te Hwang, Dong-Yi Lin and Sin-Horng Chen, "A Hakka Text-to-Speech System", Int. Symp. on Chinese Spoken Language Processing, Dec. 2006,

- Singapore; and Lecture Note in Computer Science, Vol. 4274/2006, Chinese Spoken Language Processing, Springer, pp.241-247 (SCI)
- [3] S. H. Chen, S. H. Hwang, and Y. R. Wang, “An RNN-based Prosodic Information Synthesizer for Mandarin Text-to-Speech”, IEEE Trans. Speech and Audio Processing, Vol.6, No.3, pp.226-239, May 1998.
 - [4] Min Chu, Hu Peng, Yong Zhao, Zhengyu Niu and Eric Wang, “Microsoft Mulan – A Bilingual TTS System,” ICASSP 2003,
 - [5] C. H. Lee, H. Li, L. S. Lee, R. H. Wang and Q. Huo, *Advances in Chinese Spoken Language Processing*, World Scientific Publishing Co., 2006
 - [6] R. Sproat, *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, Kluwer Academic Publishers, 1998.