國　立　交　通　大　學

電信工程學系碩士班

碩士論文

自發性對話語音音節合併現象之分析及辨識改進

An Analysis and Modeling of Syllable Contraction in Spontaneous Mandarin Speech Recognition

研究生：孫立諺

指導教授：王逸如 博士

中華民國九十三年七月

自發性對話語音音節合併現象之分析及辨識改進

An Analysis and Modeling of Syllable Contraction in Spontaneous Mandarin Recognition

研 究 生：孫立諺　　Student： Li-Yen Sun

指導教授：王逸如 博士　　Advisor： Dr. Yih-Ru Wang

A Thesis
Submitted to Department of Communication Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Master of Science
In
Electrical Engineering
June, 2004
Hsinchu, Taiwan, Republic of China

中華民國九十三年七月

自發性對話語音音節合併現象之分析及辨識改進

研究生：孫立諺　　指導教授：王逸如 博士

國立交通大學電信工程學系碩士班

中文摘要

自發性對話語音是最接近人們在自然情況下的對話語音，結合語音辨識的技術可以應用在許多領域上。在本論文中，我們專注於改善自發性語音中的重要現象－音節合併現象（Syllable Contraction）的辨識，我們統計分析語料中有關音節合併現象的特性，據以對音節合併現象建立特別的聲學模型。實驗結果顯示所建立的多音節聲學模型對於發生合併現象音節之辨識效果有明顯提昇，但當音節合併現象嚴重到合併為單一音節時，仍難以正確辨識。

關鍵詞：自發性對話語音、音節合併現象、聲學模型

An Analysis and Modeling of Syllable Contraction in Spontaneous Mandarin Speech Recognition

Student：Li-Yen Sun　　Adviser：Dr. Yih-Ru Wang

Department of Communication Engineering
National Chiao Tung University

ABSTRACT

In this thesis, the effect of syllable contraction in spontaneous Mandarin speech recognition is exploited. Syllable contraction is a phenomenon of serious coarticulation between two consecutive syllables and is a major factor to degrade the performance of a spontaneous-speech recognizer. In the study we propose to construct separate HMM models for syllables with and without contraction. Performance of the proposed approach was examined by simulations using a Mandarin dialogue speech database called MCDC (Mandarin Conversational Dialogue Corpus). Experimental results showed that the recognition performance for contracted syllables could be greatly improved via building HMM models for contracted syllable pairs with high frequency. But the recognition is still difficult for the case of very serious contraction.

Keywords: syllable contraction, spontaneous Mandarin speech recognition, MCDC

致謝

其實在研究所的大部份時間，我一直都沒有學好做一件事該有的態度跟方法，所以論文能夠在最後關頭順利完成，真的打從心裡感謝王逸如老師的督導、訓練我做事情的方法及對於任何問題都能夠竭盡所能的回答我們，也感謝陳信宏老師在繁忙之時還能抽空給予我們在語音方面的指導。

實驗室裡面大家相處的氣氛十分融洽，所以能夠來到這個實驗室，結識這些學業上的夥伴，我實在感到相當慶幸。感謝郭威志學長和輝哥學長兩年的照顧，室友阿德、生命共同體阿樹、人生由黑白變彩色的俊良、電話來了的小Z、講冷笑話的性獸、嘉俊別在叫我苦瓜了、動作最快的祺翰和出家人智合，還有小傅學長、實驗室的學弟妹及我在研究所兩年中所認識的每一個人，謝謝你們陪我度過這兩年的時光。

最後，我要感謝我的父母及家人，在這兩年中對我的關心、支持及鼓勵，讓我能夠在學業上能夠心無旁騖，順利畢業。所以僅將這本論文獻給我的父母、關心我及我所關心的每一個人。

目錄

表目錄

圖目錄

第一章 緒論

1.1 研究動機

語音辨識技術近年來已經有很長足的進步，成熟的語音辨識技術應用在商品上，可以大大提升了商品本身的價值或者是附加價值，也便利了人們的生活。但就目前為止，對於更接近於人們日常生活對話的自發性對話語音（Spontaneous speech），卻還沒有非常多的研究；所以如果我們想要將語音辨識技術完全落實應用至日常生活中，勢必要對自發性對話語音辨識有更深一步的瞭解。而自發性對話語音因為它本身的特點：說話速度快，使得音節與音節間會發生更嚴重的相互影響；所以自發性對話語音在辨識上將會比一般的中文辨識還要困難許多。這時如果我們能夠有效解決自發性對話語音在辨識上的問題，相信對於便利人們生活這方面將有非常大的提昇。

1.2 研究方向

在中研院所提供的語料－現代漢語口語對話語料庫（MCDC），由於 MCDC 語料庫中已由語言學家標記音節間相互影響效應，所以我們可以做有關音節合併現象之研究。故我們藉著這份語料去研究音節間相互影響的情況，並且期望能對於連續語音的辨識率有所提升。

1.3 章節概要

本篇論文章節綱要如下，第一章說明研究動機與論文研究領域；第二章為音節合併現象說明與特性統計；第三章為針對音節合併現象所建立的聲學模型及實驗結果與數據分析，並嘗試對合併現象做自動標示；第四章結論與未來展望。

第二章 自發性對話語音音節合併現象之特性統計

在自發性對話語音語料中，由於說話速度比較快或者是某些音節經常使用，會使得某些音節被合併（Contraction）或省略，造成音節間的相互影響，使得音節本身的結構產生變化；最常見的例如：「這樣」常常會發出近似單一音節的「降」的音；「覺得」的「得」這個音節的聲母甚至整個是整個音節常常會被忽略。

本論文主要的研究方向為自發性對話語音音節合併現象的分析，因此必須找一套具有自發性語音特性的語料庫，目前國內符合此種特性且較為完整有系統的即為現代漢語口語對話語料庫（ Mandarin Conversational Dialogue Corpus），以下我們簡稱 MCDC，此語料庫不同於一般朗讀式（ Read Speech ）資料庫，如 MAT（Mandarin Speech Data Across Taiwan）。MCDC 語料的語音是較自發性、較口語化的，同時對於音節合併現象有標記，所以非常的適合用於做音節合併現象的初步研究。

2.1 現代漢語口語對話語料庫（MCDC）

現代漢語口語對話語料庫（Mandarin Conversation Dialogue Corpus, MCDC），是由中央研究院調查研究工作室蒐集完成，語料發音人的選取是依據 16-25 歲、26-35 歲以及 36-45 歲三大年齡層由台北市市民中隨機抽樣選出，最後選取出 16 位參與錄音，共 9 位女性，7 位男性，發音者雙方在錄音前未曾謀面，在一開始自我介紹後選定主題聊天，因此語料並非朗讀式的，相當適合用於自發性對話語音的研究。

語音口語標注是使用工研院自行開發的轉寫工具 TranList，其中標注了發音人、標記員、檔案起訖時間、語料中文內容、語料漢語拼音內容及各種語言學

上的現象。

2.1.1 自發性語料庫的特性

由上節對 MCDC 語料庫的簡介，我們已知其所蒐集的語料是非常貼近日常對話的，也就是語音是自發性的，因此若要對此語料庫作研究，必須先了解自發性語料與朗讀式語料間特性的差異，我們知道自發性語料是非常口語化的，也就是因為口語化而產生了許多朗讀式語料中不會出現的特殊現象，在 MCDC 語料庫中標註了許多這方面特性的資訊，如：拖長音、音節合併、發音偏差、不確定音...等，而標註種類詳細的介紹，請參考中文詞知識庫小組為此語料庫所出的一本技術報告--現代漢語口語對話語料庫標註系統說明【1】，下面僅對於本論文所使用到的特性做簡單的介紹。

1、 **非語音現象（ Paralinguistic Phenomena ）**

分成兩類，一、凡非語音但確定由人所發出的聲音，包括笑聲、咳嗽聲、呼吸聲、吐氣聲.....等，和其他口腔發出無法辨識的聲音等等。二、非語音且確定非人所發。

2、 **感嘆詞 （ Particle ）**

不具標準語意的感嘆詞，其與用成分居多如回應或同意。

3、 **不確定字/音（ Uncertain）**

一、標記員根據前後語意，可以猜測出大概的語意內容，但無法百分之百確定。二、標記員無法根據語意猜測出對應字詞，但可清楚紀錄出其發音。下面列出一些的例句。

接著我們要詳細介紹本論文中欲研究的主題－音節合併現象做詳細的介紹與分析。

2.1.2 音節合併現象

由於 MCDC 語料庫中已由語言學家標記音節間相互影響效應－音節合併（Syllable Contraction）之資料，所以我們可以做音節合併現象之研究。在國語自發性口語語音中，當說話者話說得太快或不清楚時出現的音節合併現象，合併現象有三【2】:

(1) 清楚可辨的音節短少,像是從原本正常的三個字三個音節變成三個字兩個音節，或者是兩個字兩個音節變成兩個字一個音節。

(2) 音節雖無短少，但卻都連在一起，難以切割。

(3) 音節無短少且音節可切割，只是音節結構有變。

在 MCDC 語料庫中，音節合併現象在標記時包括所有音節合併的字。但拼音部份仍以標準發音的漢語拼音轉寫，而非以實際發音轉寫。經由初步統計得知其占整份語料約 17%，可想像其對辨識將產生相當大的影響。

2.2 音節合併現象初步分析

我們目前為止已經對於音節合併現象有初步的瞭解,所以我們針對音節合併現象可能會發生的一些現象做分析討論。如前述，當音節合併現象發生時，清楚可辨的音節短少，原本正常的三個字三個音節變成三個字兩個音節，或者是兩個字兩個音節變成兩個字一個音節。所以在嚴重的音節合併現象發生下，不但音節間互相影響造成辨識率的下降,還可能因為兩個音節合併為一個音節而造成刪除型的錯誤發生,接下來我們將針對音節合併現象在辨識上所可能發生的狀況作分析。在此我們所使用的自發性語音基本辨認器將在下一章中詳述，在此我們只是想由辨認器結果探討音節合併現象對於辨認結果之影響。

2.2.1 音節合併現象對於辨識產生的影響

當發生輕微的音節合併效應時，雖然音節間會產生相互影響，但音節本身主要的結構還未嚴重改變，所以我們也許可以藉著 411 音節辨識器將其正確的辨識出來。下例中雖然標記發生音節合併現象但卻被單一音節模型正確辨識出來的例子，可能該處發生較輕微的音節合併現象。這裡我們使用漢語拼音；”A”為感嘆詞（Particle）；”~”為不確定字音（Uncertain）起始標記。其中”_”連接處為發生音節合併現象位置。

標示內容：**dui_A** jiu shi tao kong ~Na

（**對_A** 就 是 掏 空 ~Na）

辨識結果：**dui A** zhi ~ri hao kong A

當音節合併現象發生加重時，音節結構受到相互影響而改變，此時辨識可能會受到影響而辨識為相近的音節。甚至如果音節合併現象太過嚴重，造成整個音節結構的改變，例如：”zhe_yang（這樣）”因為音節合併效應發音近似”jiang（降）”，則容易被辨識成單一音節（此時會造成一個刪除型錯誤 或者是 一個替代型錯誤加上一個刪除型錯誤）。我們接下來觀察幾個例子，下例中為音節合併現象標記處因為受到前後音的相互影響而發生辨識錯誤的情形。

標示內容： **dui_A** nei …

（ **對_A** 那… ）

辨識結果： **dui la** nei …

下例中有兩處標記發生音節合併現象卻被辨識為單一音節，可能該處發生嚴重的音節合併，其中清楚可看出標示內容中 ”yin_wei（因為）”發生音節合併，

但是辨識結果只辨識出”yin（因）”，而 ”wei （為）”這個音節發生了刪除型的錯誤；第二處音節合併標記則是兩個音節完全的合併為第三個音節，不但發生了刪除型錯誤同時也發生了替代型的錯誤。

標示內容：**yin_wei** wu lai ne bian de wen quan **bi_jiao** mei you…

（因為烏來那邊的溫泉比較沒有…）

辨識結果：**yin** wu lai nen bie wen xuan **biao** mei ying…

分析了幾個音節合併在辨識上可能發生的現象後，接下來我們利用觀察幾個發生音節合併現象的音節（Contraction Syllable）和未發生音節合併現象的音節（Non-Contraction Syllable）的辨識率比較，如表 2.1 所示，看看在辨識上發生合併現象的音節是否會如我們所預期較未發生合併現象的音節來得低。我們嘗試以下方法來觀察音節合併效應確實會影響辨識率。首先定義單一音節辨識率如下式，我們可以觀察單一音節替代型錯誤的發生狀況；同時我們也觀察刪除型錯誤的情況。

音節辨識率 = 正確數（Hits）/（ 正確數 + 替代型錯誤數（Substitutions））

表 2.1：使用基本 411 音節模型觀察發生與未發生合併現象音節辨識狀況

統計 MCDC 語料			音節辨識率與刪除型錯誤（測試語料）				
音節	出現	發生合併現	未發生合併現象音節		整體音節	發生合併現象音節	
漢語拼音	次數	象音節總數	辨識率	刪除型錯誤	辨識率	辨識率	刪除型錯誤
shi ㄕ	6329	2005	70.9%	41	64.4%	45.9%	52
wo ㄨㄛ	3954	1366	66.4%	23	60.9%	44.2%	29
yi ㄧ	3923	1656	60.8%	44	47.8%	29.1%	46
de ㄉㄜ	3701	1567	56%	34	43.4%	24.3%	62
jiu ㄐㄧㄡ	2727	755	71.5%	10	67.1%	56.3%	6
you ㄧㄡ	2839	967	68.4%	19	55.6%	25.3%	14
bu ㄅㄨ	2443	713	80.3%	9	70.5%	33.9%	11
dui ㄉㄨㄟ	2833	1174	76.7%	6	71.9%	60.9%	28

ge　ㄍㄜ	1472	407	53.3%	9	45.2%	27.1%	9
na　ㄋㄚ	974	337	22.2%	9	17.8%	11.1%	10
ni　ㄋㄧ	1935	477	64.5%	16	60.8%	28.9%	8
ta　ㄊㄚ	2049	654	73%	14	57.8%	26.2%	23
hen　ㄏㄣ	1181	218	62.9%	3	54.8%	21.1%	6

我們可以很明顯的看到未發生合併現象的音節和發生合併現象的音節在辨識率上都有一定的差距（發生合併現象的音節部分明顯偏低），所以說音節合併現象確實會影響辨識。同時在刪除型錯誤的發生狀況比較上，發生合併現象的音節明顯的比率高於未發生合併現象的音節甚多。

2.2.2 初步分析歸納

經過以上對於音節合併現象的初步分析可知，其不但造成音節辨識上困難度的上升，同時也容易造成刪除型錯誤的發生。所以接下來我們將針對音節合併現象作更詳細的分析統計，進而希望能夠找出適當的方法來解決因為音節合併現象所造成辨識率低落的問題。

2.3 音節合併現象統計與分析

經過統計 MCDC 整份語料可分析資料全部共約 14,5000 字，含 411 音節、感嘆詞（Particle）、不確定字/音（Uncertain）。其中標記有音節合併現象者共有 19,766 處。例如：”就-是”之間發生音節合併現象，則算一處；”這-樣-子”連續發生音節合併現象，則算兩處。所有的前後音節組合（包含發生與未發生音節合併現象）共有 118,143 組（跨句分開不算），所以以音節組合來看，音節合併現象發生率約 17%。

若以音節受合併現象影響來看，所有標記受合併現象影響音節為 35,309 個，總音節數約 145,000，所以音節受合併現象影響率約 24%。

2.3.1 音節合併現象在音節層次上的統計

因為音節合併現象是音節間相互影響的效應，所以我們試著觀察音節合併現象的發生是否與音節的各種性質有關；例如：空聲母音節是否容易與前一音節合併，爆破音、摩擦音等…非韻律聲母起始的音節是否就較不易與前一音節合併等等…一些有關於音節耦合效應（Coarticulation）的現象。接下來的統計分析將先以音節間的組合開始。

（1） 音節組合（依照前後音節 411 音分類組合）

我們發現某些常見音節組合音節合併現象發生率遠高於 MCDC 語料中音節合併現象的平均發生率。下表 2.2 中列出語料中幾個常見音節組合及其合併現象發生率。

表 2.2：常出現音節組合之合併現象發生率

音節組合	合併現象發生次數	合併現象發生率
zhe_yang （ㄓㄜ_ㄧㄤ）	452	76.7%
wo_men （ㄨㄛ_ㄇㄣ）	447	58.2%
ran_hou （ㄖㄢ_ㄏㄡ）	447	66.4%

上表中列出了最常發生音節合併現象的幾組音節組合。可觀察到其中第一組與第七組的第二音節之聲母為摩擦音，還有第六組的第二音節之聲母為爆破音，這些都是我們所知道較不易發生音節耦合現象的情況，所以似乎與我們所猜想的結果不符。但我們仍需仔細分析音節合併現象在音節間之關係。

（2） 音節組合（含聲調）

若進一步統計含聲調之音節組合發生音節合併現象之情況，我們可以得到結果如表 2.3。

表 2.3：常出現音節組合（含聲調） 之合併現象發生率

音節組合	合併現象發生次數	合併現象發生率
zhe4_yang4 （ㄓㄜˋ_ㄧㄤˋ）	452	76.8%
ran2_hou4 （ㄖㄢˊ_ㄏㄡˋ）	447	67.7%
wo3_men5 （ㄨㄛˇ_ㄇㄣˊ）	415	58.7%

經由觀察含聲調的音節組合，可以發現這些音節組合似乎是我們在講話時所會常常使用到的，而且其發生率與不含聲調的音節組合幾乎相同。

（3） 字元組合（Character pair）

下表 2.4 中列出常見字元組合發生合併現象狀況。比較表 2.3 和表 2.4，除了一小部分受發音錯誤和破音字（們：• 或ˊ）的影響，音節組合（含聲調）幾乎可以確定其字元組合為何，其發生率也幾乎是相同的，而且發現這些字元組合皆為我們所常用的詞。所謂常用詞是有些詞是一般人常用,但往往其所含的資訊量卻很少的，例如:”所以”，”這樣”…，因為是較無關資訊的傳達，往往在發音時易產生前後音混淆的情形。反之如：人名、地名…等，則較不易發生音節合併現象。

表 2.4：常出現字元組合之合併現象發生率

字元組合	合併現象發生次數	合併現象發生率
這樣	452	76.9%
然後	447	67.7%
我們	446	58.1%

下表 2.5 為我們統計 MCDC 語料中最常發生音節合併現象的二字詞（依出現次數做排序）。我們可以發現它們都是國語口語中的常用詞。

表 2.5：常見二字詞之合併現象統計

二字詞	詞性	音節合併現象發生次數	出現次數	音節合併現象發生率
就是	關聯連接詞	292	787	37.00%
我們	代名詞	442	776	56.90%
然後	副詞	444	657	67.58%
因為	關聯連接詞	404	598	67.56%
覺得	狀態句賓動詞	364	565	64.42%
沒有	狀態及物動詞	237	444	53.38%
所以	關聯連接詞	362	414	87.44%
可是	關聯連接詞	167	371	45.00%
什麼	指代定詞	69	370	18.65%
這樣	狀態不及物動詞	267	351	76.07%
現在	時間詞	225	329	68.40%
對對	動作及物動詞	171	327	52.30%
他們	代名詞	154	323	47.68%
比較	動詞前程度副詞	108	309	35.00%
其實	副詞	143	278	51.44%
那邊	位置詞	68	248	27.42%
那個	定量複合詞	79	239	33.05%

（ 註：文字資料之斷詞結果是由交大語音處理實驗室之斷詞器產生 ）

由於斷詞會受到前後文的影響，所以出現次數會與之前所觀察的字元組合有部分的出入，而且這邊只是統計二字詞，未考慮三字詞涵蓋的影響；但我們仍可看出其發生音節合併現象的機率是差不多的，不但出現次數多，發生率也高。而且在詞性觀察上這些也是較無關資訊傳達的詞類。

2.4 結論

由前面統計分析我們可以發現：

1. 在自發性對話語音中音節合併現象發生率是相當高的。像在 MCDC 語料中就發生了約 17%的音節合併現象，可見其影響不小。
2. 音節合併現象確實會影響辨識率。經過初步的辨識結果分析，發生合併現象的音節辨識率比未發生合併現象的音節辨識率要低的多，同時發生合併現象的音節刪除型錯誤發生次數也來的較高。
3. 發生合併現象多為常用詞，前後文相關模型可能對辨識沒多大幫助。經過統計，常用詞不但出現次數多，且其發生音節合併現象的機率也高，所以合併現象應該主要是跟常用詞有關，而跟前後音節的發音特性沒有多大的關係。

第三章 考慮音節合併現象之自發性語音辨認器

在第二章表 2.1 中我們經過統計得知，音節合併現象確實會影響辨識，在音節辨識率上普遍發生合併現象音節明顯較未發生合併現象音節為低，且在自發性語料中音節合併現象發生比例高達 17%。所以我們將針對音節合併現象另外建立更精確之聲學模型，以解決因為合併現象所造成辨識率降低的問題。首先我們將先建立一個基本 411 音節辨識系統做為基準系統；然後考慮合併現象建立獨立音節模型，也就是針對受合併現象之音節各自建立其 411 音節模型；更進一步我們將針對發生合併現象之音節建立相依音節模型，也就是建立多音節聲學模型與右相關音節模型來描述合併現象。之後將比較新增模型對於辨識上的影響與音節合併現象在辨識上所造成的錯誤分析。

3.1 基本 411 音節辨識系統

在辨識系統的建立方面，我們使用英國劍橋大學所開發的工具 HTK (HMM Toolkit)【3】。我們首先建立 411 音節的基本辨識系統，做為辨認器效能比對之用。十分之九的 MCDC 語料做為訓練語料，模型建立過程中當訓練語料不足時，用相近音模型替代之。訓練語料統計如表 3.1 所示。

表 3.1：MCDC 訓練語料文字統計表

	411 音節	Particles	Paralinguistic phenomena	Filler	Uncertain
字數	101464	9235	7744	4665	3342
百分比	80.24%	7.3%	6.12%	3.69%	2.64%
總字數	126450				
Sub-turn 數	6011				

總共建立了 100 個 Final-Dependent Initial 和 40 個 Final 以組成 411 音節模型。所有 Initial 及 Final Model 的狀態轉換都只允許 1-state forward only。加上其他現象模型如:40 個 Particle 模型、76 個 Uncertain 模型、15 個 others(包含了呼吸聲、咳嗽聲等…一些非語音的模型),所以基本辨識系統共建立了 271 個模型。其中模型狀態混合數(Model State Mixture Number)調整是依據訓練語料中每個模型狀態(Model State)訓練語料音框(frame)數量多寡,設定其狀態混合數(state mix no.)。如下式,每個狀態(state)累積訓練語料超過 50 個音框(frame)時增加一個混合數(mix no.)。所有模型詳細設定如表 3.2。

$$N_S = \min\left(\lfloor n / 50 \rfloor , 32 \right) \qquad (1)$$

其中 N_S : State Mix no.

n : State Frame no.

表 3.2:基本辨識系統音節模型種類數量

模型種類	411 音節模型		Particle	Uncertain	other	ALL
數量	100 RCD Initial	40 Final	40	76	15	271
狀態數	3	5	3	3	3	--
模型混合數	依資料量最大 32 個					

在建立基本 411 音節模型後我們可以對測試語料得到初步辨識率如表 3.3。首先定義辨識率如下式所示:

辨識率=(正確音節數-(替代型錯誤+插入型錯誤+刪除型錯誤))/正確音節數

表 3.3:MCDC 基本 411 音節模型辨識結果

	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
基本辨識系統	39.81%	45.8%	12.9%	41.3%	6%

3.2 考慮合併現象建立獨立音節模型

在這裡我們先不考慮音節間相互影響的效應，我們將 MCDC 語料只依據 411 音節是否受合併現象影響做分類，分別建立兩個組的 411 模型，也就是將發生合併現象之音節與未發生合併現象之音節分別建立 411 模型，這裡我們對「感嘆詞」和「不確定字/音」還是使用基本系統之模型，獨立音節模型建立步驟如下：

1. 如圖 3.1 所示，首先利用 MCDC 語料訓練出基本 411 音節辨識系統模型。
2. 將 MCDC 語料中針對發生合併現象和未發生合併現象的音節分開建立各自的 411 音節模型。
3. 當發生合併現象 411 音節模型訓練語料出現次數足夠（大於三次），我們建立其模型；如果出現次數不足三次，我們用基本辨識系統 411 音節模型替代之。

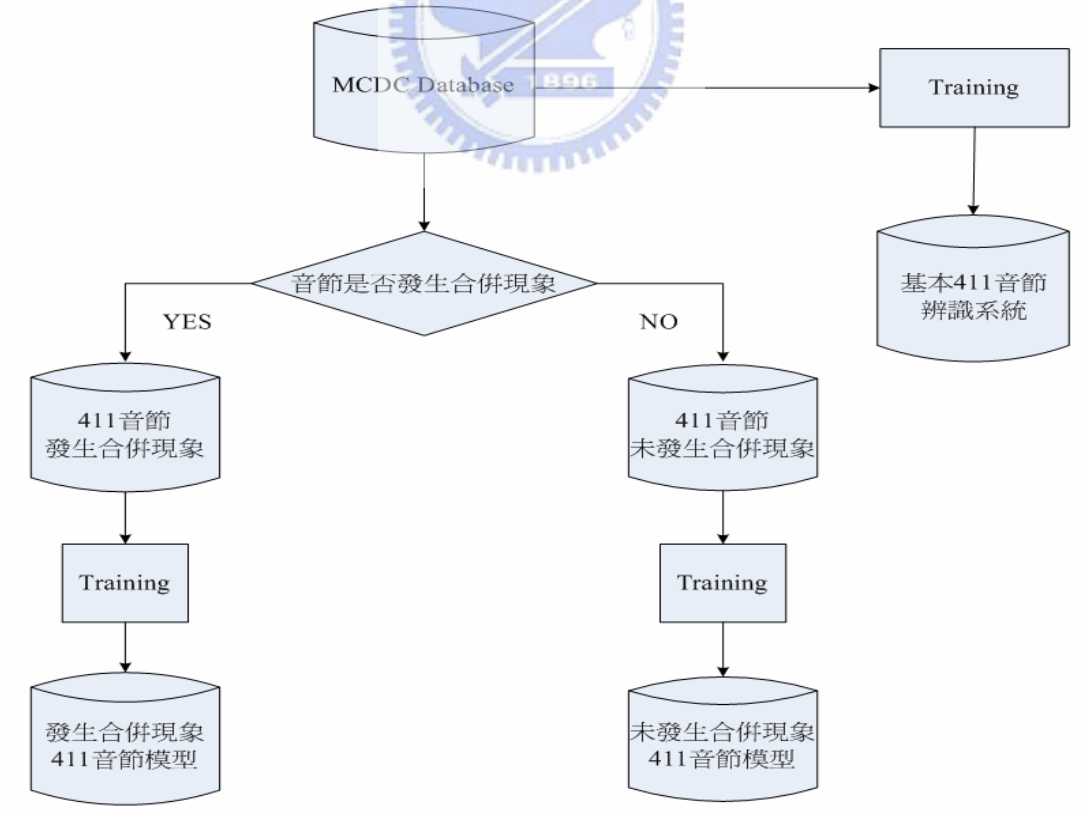

圖 3.1：獨立音節模型建立流程圖

3.2.1 發生及未發生合併現象之音節模型

針對未發生合併現象的音節，我們建立其 411 音節模型。當訓練語料不足時，我們用基本系統之 411 音節模型替代之。所以可以建立出未受音節合併現象影響的 411 音節模型，簡稱非合併型 411 音節模型，共訓練出不受合併現象影響的 100 個右相關 Initial 和 40 個 Final 模型。針對發生合併現象的音節，我們分別建立合併型之 411 音節模型。當訓練語料不足時，我們用已經建立好的基本 411 音節模型替代之。所以可以建立出受音節合併現象影響的 411 音節模型。

3.2.1.1 模型建立過程

在訓練發生合併現象的音節模型前，我們必需先定義發生合併現象的 Initial 及 Final，以便訓練其模型。首先我們來觀察一下音節間發生合併現象的狀況。如表 3.4 所示，觀察音節 Y，音節合併現象可依合併位置分為三類，底線表示音節合併狀況，例如：”XY”表示音節 Y 與前一音節 X 發生合併現象。

表 3.4：音節(Y)發生合併狀況

音節(Y)	音節未發生合併現象	音節發生合併現象		
種類	1	2	3	4
音節合併情況	**X Y Z**	**X Y Z**	**X Y Z**	**X Y Z**

當前後音節未發生音節合併現象時，我們拿來建立未發生音節合併現象的模型；當前後音節發生音節合併現象時，我們認定前一音節的 Final 部分和後一音節的 Initial 部分受到音節合併效應的影響，例如：上表中種類二為音節(X)的 Final 與音節(Y)的 Initial 受到音節合併效應影響，所以拿來建立發生合併現象的 Initial 及 Final 模型；雖然音節合併現象為前後音節互相影響的效應，但在這裡我們先不考慮前後文的相關性。同樣的當訓練語料不足以建立模型時，我

們拿基本系統的 411 音節模型替代之。所以可以建立出受音節合併現象影響的 411 音節模型，簡稱合併型 411 音節模型，共訓練出受合併現象影響的 100 個右相關 Initial 和 40 個 Final 模型。

3.2.1.2 HMM Model 狀態轉換之設定

由於受到音節合併現象的影響，音節長度(duration)將會比一般的音節來的短，所以我們針對模型狀態的轉移作新的設定，如表 3.5 所示。當發生音節合併現象時，該音節 Initial or Final 的音節長度（Syllable Duration）可能會相當短，故我們設定模型中狀態（State）轉移規則為：除了往前一個狀態外，允許跳過一個狀態(1-state skip)。我們設定兩種狀態轉換方式，之後將觀察其對辨識率的增進情況。

表 3.5：狀態轉移方法及模型種類

發生音節合併現象的模型		跳躍的模型	狀態跳躍方式
Method 1	Initial	ALL	如圖 3.2
	Final	--	--
Method 2	Initial	ALL	如圖 3.3
	Final	ALL	如圖 3.4

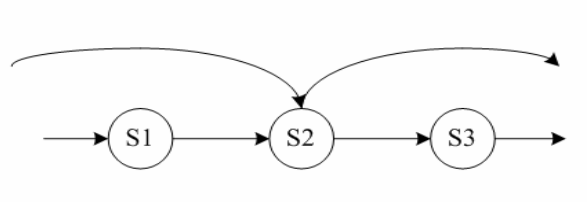

圖 3.2：Initial 的 State Skip（State 2 不可跳過）

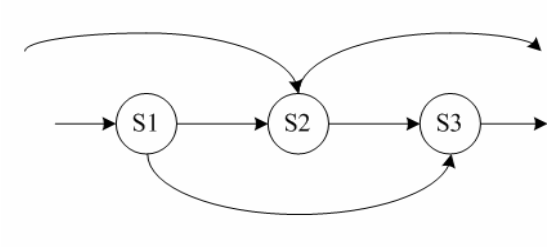

圖 3.3：Initial 的 State Skip（可以 Skip 下個 State）

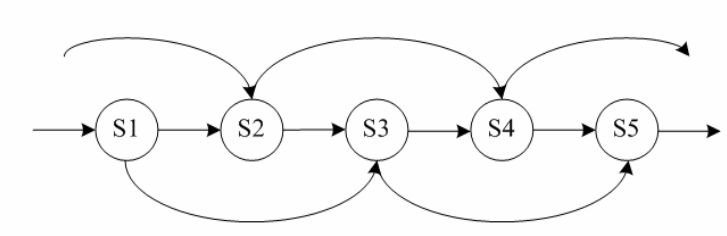

圖 3.4：Final 的 State Skip（可以 Skip 下個 State）

3.2.2 使用合併現象之音模型在辨識器之模型連接

當音節合併現象發生時至少為兩個音節以上的互相影響，所以我們對於音節合併的連接狀況要有所限制。例如：當音節的 Final 發生合併現象時，其後必定連接 Initial 發生合併現象的音節。另外當音節合併現象發生時，其音節間合併的情況相當的嚴重，所以我們在發生合併現象的音節間不留短暫停的模型(Short Pause Model)。

所以在辨識上我們將音節（Y）依據發生音節合併的情況分為四種情況(如表 3.6 所示)，底線部份表示音節合併現象的發生位置。

表 3.6：音節合併現象發生狀況

種類	1	2	3	4
音節合併情況	X Y Z	X Y Z	X Y Z	X Y Z
Initial	I_{NC}	I_{C}	I_{NC}	I_{C}
Final	F_{NC}	F_{NC}	F_{C}	F_{C}

註： I_{NC} & F_{NC}：未發生音節合併現象的 Initial & Final

I_{C} & F_{C} ： 發生音節合併現象的 Initial & Final

圖 3.3 為我們針對音節辨識所加的限制，可行路徑上我們不加其它分數，也就是類似無文法規則(Free Grammar)的作法。注意紅框範圍為我們針對音節合併現象所增加的限制，也就是有受到合併現象影響的韻母聲母必須要相連接。

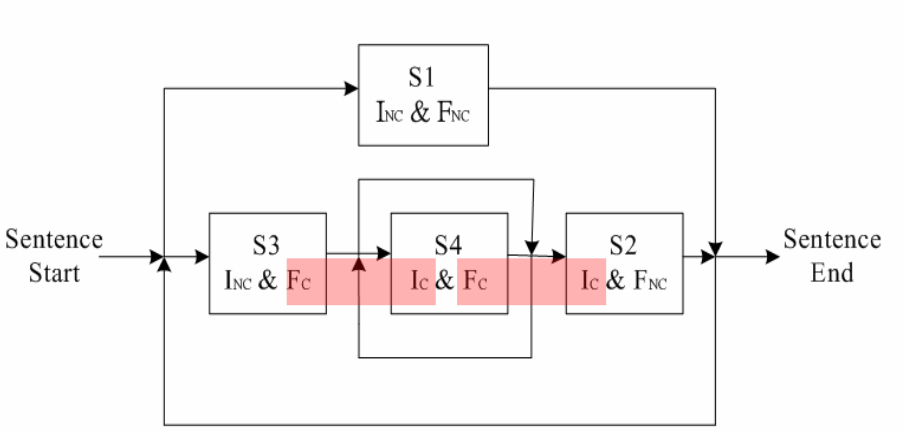

圖 3.5：合併現象音節辨識器模型連接

3.2.3 實驗及結果分析

之前我們設定兩種狀態轉移規則，表 3.7 為比較兩種方法在辨識方面的表現。由表可看出，方法一的狀態跳躍方式略優於方法二，因為每一個 Initial 或 Final 其狀態應該有其穩定部份，我們不應該允許它被省略，故我們採用方法一。

表 3.7：比較兩種不同狀態轉移方式對於音節辨識率的影響

	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
方法一	39.20%	46.6%	10.6%	42.8%	7.4%
方法二	38.84%	47.1%	9.8%	43.1%	8.2%

分別建立非合併及合併型 411 音節模型之後，我們比較它們與基本辨識系統在整體辨識率上的表現，表 3.8 為兩組模型在辨識率上的比較，整體來說辨識率些微下降，正確數與刪除型錯誤雖有改進，但替代型錯誤與插入型錯誤相對增加。

表 3.8：新增音節合併現象模型後與基本音節模型在辨識上的比較

	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
基本辨識系統	39.81%	45.8%	12.9%	41.3%	6%
非合併及合併型 411 音節模型辨識系統	39.20%	46.6%	10.6%	42.8%	7.4%

我們再針對常發生音節合併現象的音節，觀察音節辨識率在新增模型後的改善狀況，同時觀察音節合併現象所容易造成的刪除型錯誤數量的狀況。如表 3.9 所示，在辨識上普遍音節的辨識率沒有提升(只有刪除型錯誤改善)，這是可以理解的；因為僅將音節合併現象分開建立一組共同模型而未考慮它們是何類音節相連接(未考慮左右文)本來就是一種十分粗糙的做法。

單一音節辨識率 = 正確數 / (正確數 + 替代型錯誤)

表 3.9：新增音節合併現象模型之後對於常見音節的辨識情況

模型	基本 411 音節模型辨識系統		非合併及合併型 411 音節模型辨識系統	
常見音節	音節辨認率	刪除型錯誤	音節辨認率	刪除型錯誤
shi 是	64.4%	120	64.8%	83
wo 我	60.9%	61	63.3%	43
yi 一	47.8%	104	51.4%	80
de 的	43.4%	98	43.8%	71
jiu 就	67.1%	15	67.4%	12
you 有	55.6%	37	55.9%	28
bu 不	70.5%	30	69.8%	25
dui 對	71.9%	46	73.7%	29
ge 個	45.2%	18	44.4%	15
na 那	17.8%	18	17.9%	13
ni 你	60.8%	31	60.8%	21
ta 他	57.8%	47	57.3%	31
hen 很	54.8%	9	53.7%	8

3.3 針對合併現象建立相依音節模型

因為音節合併現象為音節間相互影響的效應，所以在辨識上可能無法單純的只以建立合併音節之 411 音節模型來解決，接下來我們嘗試建立音節間相依的（Dependent）模型來嘗試解決因音節合併現象所造成辨識率降低的問題。

3.3.1 多音節聲學模型(Syllable Pair Acoustic Model)

依據第二章的統計，我們針對常發生音節合併現象的音節組合，單獨建立它們的聲學模型結。所以依據發生音節合併現象的音節組合(或 詞)數量的多寡，建立音節組合(或 詞)的聲學模型。這裡列出幾個常發生的音節組合，如表 3.10 所示：

表 3.10：數個常發生音節合併現象的音節組合

雙音節組合	三音節組合
dui_A（對 A）	dui_dui_dui（對對對）
zhe_yang（這樣）	suo_yi_wo（所以我）
wo_men（我們）	dui_bu_dui（對不對）
ran_hou（然後）	

註: A 為 Particle，其發音以漢語拼音表示為 ”a” 。

所以針對音節合併現象發生次數達一定數量的音節組合，我們另外建立其聲學模型，聯合參與辨識。

3.3.1.1 模型建立過程

如圖 3.4 所示，在發生音節合併現象的語料中，對於出現次數足夠的音節組合我們建立其聲學模型；出現次數不足的我們暫時不特別處理，仍然訓練發生音節合併現象的模型。

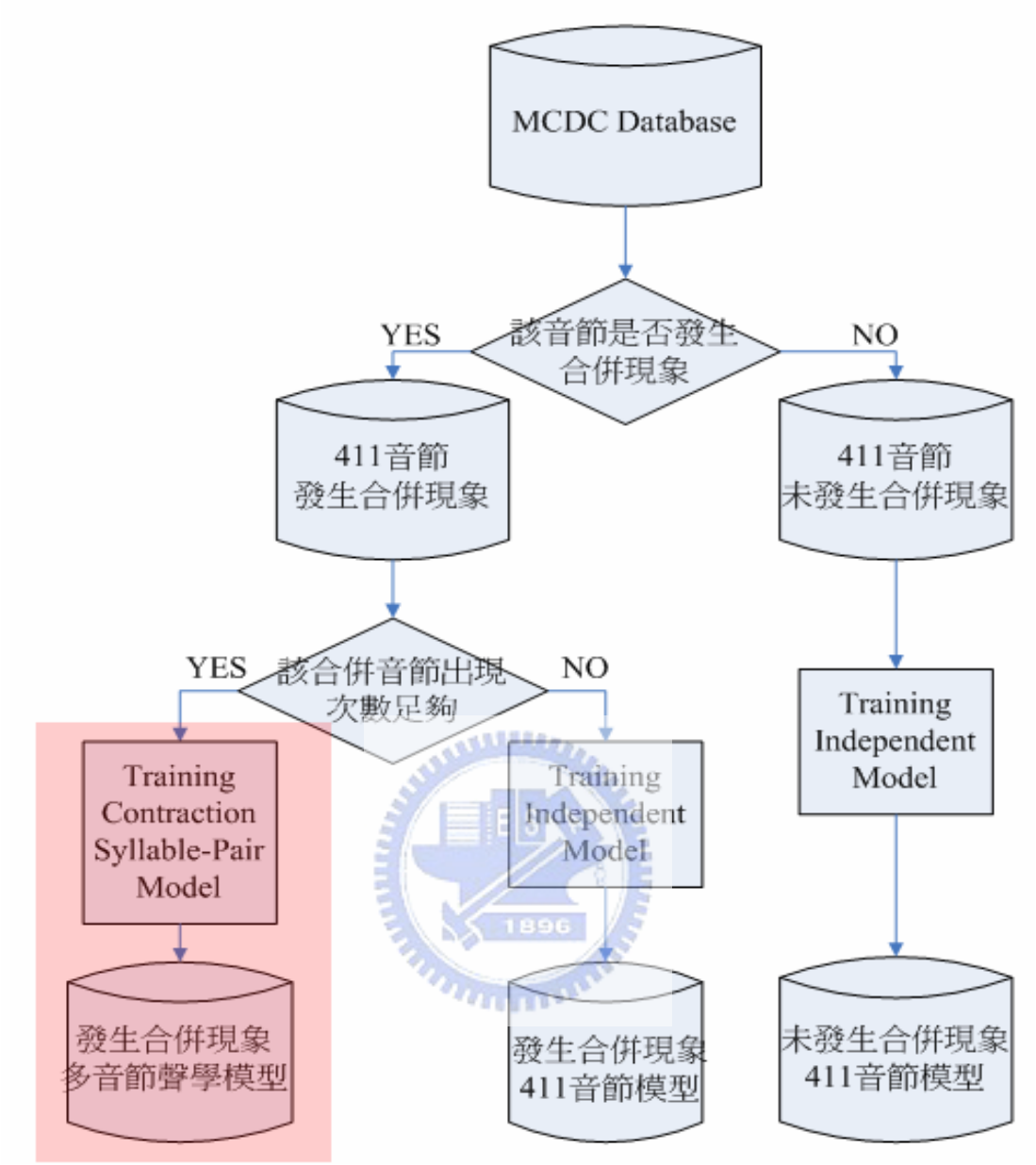

圖 3.6：多音節聲學模型建立流程圖

接著我們詳細描述各類模型之設定。對於合併音節組合之模型，因為音節合併現象的關係，所以我們在設定 Model 的狀態數(State)時將會所修正。原本正常的 411 音節模型為 8 個狀態數，但是雙音節的音節組合模型狀態數因為發生合併現象我們設定小於 16 個狀態，為 12 個狀態數；三音節的 Syllable Pair Model 狀態數設定也一同縮減。其模型主要的設定如表 3.11 所示：

表 3.11：多音節聲學模型詳細設定

	雙音節組合	三音節組合
Model No.	59	4
State No.	12	16
State Skip	No	No
State Mix No.	依資料量最多 32 個	

3.3.1.2 新增多音節聲學模型後之辨認器架構

因為多音節聲學模型已經 Model 了一些主要會發生合併現象的常見詞，所以為了簡化 Grammar，我們不考慮發生音節合併現象的多音節聲學模型再與其他發生合併現象音節連接的情況，辨認器架構如圖 3.5 所示。

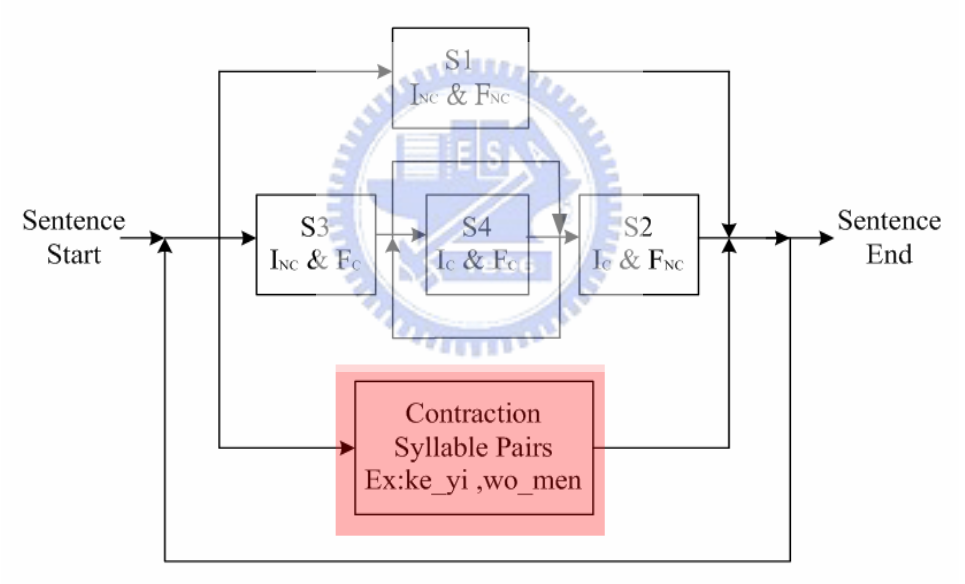

圖 3.7：多音節聲學模型辨識器模型連接

3.3.1.3 新增多音節聲學模型之辨識結果分析

接下來我們將觀察新增音節模型後對於辨識的改善狀況，這裡我們聯合兩組的模型參與辨識：「非合併及合併型 411 音節模型」和「多音節聲學模型」；在多音節聲學模型的數目上，我們依照足夠訓練語料，先訓練出 63 組音節組合模型參與辨識。

針對使用圖 3.5 的辨認系統之實驗其辨認結果如表 所表 3.12 示。正確數、替代型錯誤和刪除型錯誤都有改善。

表 3.12：比較新增多音節聲學模型後的辨識狀況

參與辨識的模型	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
非合併及合併型 411 音節模型	39.20%	46.6%	10.6%	42.8%	7.4%
非合併及合併型 411 音節模型 & 多音節聲學模型	39.89%	47.5%	10.3%	42.1%	7.7%

我們同樣再針對常發生音節合併現象的音節，觀察音節辨識率在新增多音節聲學模型後的改善狀況，同時觀察音節合併現象所容易造成的刪除型錯誤數量的狀況。如表 3.13 所示，明顯的發生音節合併現象且出現次數多的音節，在新增合併音之多音節聲學模型後辨識率提升相當明顯；刪除型錯誤則是第三欄比第二欄增加了許多，但比起第一欄仍有下降，更可以看出發生合併現象 411 音節模型對於辨識並無幫助，只是造成替代型錯誤 和 插入型錯誤的增加，這樣刪除型錯誤自然下降。

單一音節辨識率 = 正確數 / （ 正確數 + 替代型錯誤 ）

表 3.13：比較常發生音節合併現象之音節在不同模型下辨識率提升狀況

	基本 411 音節模型		非合併及合併型 411 音節模型&多音節聲學模型		非合併型 411 音節模型 & 多音節聲學模型	
常見音節	音節辨識率	刪除型錯誤數量	音節辨識率	刪除型錯誤數量	音節辨識率	刪除型錯誤數量
shi 是	64.4%	120	68.6%	83	67.4%	95
wo 我	60.9%	61	64.6%	47	65.4%	50
yi 一	47.8%	104	52.6%	80	52.4%	100
de 的	43.4%	98	47.9%	74	47.1%	78
jiu 就	67.1%	15	69.7%	14	70.3%	21
you 有	55.6%	37	56.8%	33	62.2%	37
bu 不	70.5%	30	68.3%	27	68.9%	32

dui 對	71.9%	46	69.2%	36	72.9%	44
ge 個	45.2%	18	41.0%	13	44.4%	26
na 那	17.8%	18	20.0%	11	19.7%	20
ni 你	60.8%	31	64.4%	25	63.6%	29
ta 他	57.8%	47	62.6%	30	62.0%	44
hen 很	54.8%	9	55.6%	9	54.8%	9

接下來我們想觀察目前為止所建立的模型對於整體辨識的影響。如預期的，由表 3.14 中我們可以看出第四列的模型組合對於辨識有最好的效果，也就是「沒有發生合併現象的 411 音節模型」和「針對合併現象所建立的多音節聲學模型」來辨識的效果最佳。

表 3.14：不同模型對於辨識的狀況

Model	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
基本 411 音節辨識系統	39.81%	45.8%	12.9%	41.3%	6.0%
非合併及合併型 411 音節模型	39.20%	46.6%	10.6%	42.8%	7.4%
非合併及合併型 411 音節模型 & 多音節聲學模型	39.89%	47.5%	10.3%	42.1%	7.7%
未發生合併現象 411 音節模型 & 多音節聲學模型	41.54%	47.6%	12.3%	40.1%	6.0%
基本 411 音節辨識系統 & 多音節聲學模型	41.28%	47.5%	12.0%	40.5%	6.2%

比較表 3.14 中第一欄和第四欄的結果，正確數、刪除型錯誤、替代型錯誤 都有明顯的改善（插入型錯誤部份幾乎不變），這也說明了音節合併現象的確會造成刪除型錯誤的發生，所以新增音節組合模型之後也就造成了辨識方面的改善。

由前面幾種不同模型聯合起來的辨識比較，我們可以發現發生合併現象 411 音節模型對於辨識其實是幾乎完全沒有幫助，而且造成替代型錯誤、插入型錯誤的上升，所以我們之後的辨識只以「不受音節合併現象影響的 411 音節模型」與「多音節聲學模型」聯合辨識。

如表 3.15 所示，雖然增加多音節聲學模型的數量對於辨識率提升有正比的關係，但是隨著我們 Model 到的發生合併現象的音節組合數量增加量越來越少，辨識率的提升也就跟著越來越有限。在刪除型錯誤方面，隨著所建立多音節聲學模型數量的增加，刪除型錯誤也有明顯的下降。在此實驗中我們將測試語料中語句中不包含 411 音節之測試語料移除，所以總測試語料句數由 716 句降為 444 句，字數由 17733 字降為 17182 字，基本 411 音節辨識系統辨識率由 39.81%昇為 41.05%。

表 3.15：新增多音節聲學模型數量對於整體辨識率的提升情況

	基本 411 音節辨識系統	未發生合併現象 411 音節模型 & 多音節聲學模型				
多音節聲學模型	--	60	70	80	90	100
在測試語料中所 Model 到的音節合併現象比例（All 2285）	--	約 45%	約 47%	約 48%	約 50%	約 51%
整體辨識率 （Acc）	41.05%	42.70%	42.75%	42.82%	42.86%	42.88%
刪除型錯誤數量	2172	2072	2067	2064	2061	2060

上表 3.15 中可看出新增多音節聲學模型數量約 60 個時整體辨識率提升約 1.7%，但是再新增 40 個模型後整體辨識率卻只提升 0.18%，所以多音節聲學模型的效果都集中在前 60 個模型，符合常用詞的趨勢。整體的辨識率提升了 1.7%，接下來觀察發生合併現象音節的改善狀況又是如何。之前第二章我們曾經觀察過發生與未發生合併現象音節的辨識狀況，那時明顯的可看出發生合併現象的音節辨識率偏低，接下來我們針對合併現象新增多音節聲學模型後對發生合併現象的音節辨識方面的狀況作觀察。由表 3.16 可看出新增多音節聲學模型之後，未發生合併現象音節辨識率相差不多，但是在發生合併現象音節辨識率方面則有相當明顯的提昇，同時刪除型錯誤也明顯的下降了，證明了我們所建立的多音節聲學模型在辨識上確實對發生合併現象之音節有明顯的提昇。

單一音節辨識率 = 正確數 / （ 正確數 + 替代型錯誤 ）

表 3.16：新增多音節聲學模型對於發生和未發生合併現象之音節辨識比較

		基本 411 音節模型		未發生合併現象 411 音節模型 & 多音節聲學模型	
有建立之多音節聲學模型音節		未發生合併現象之音節（刪除型錯誤）	發生合併現象之音節（刪除型錯誤）	未發生合併現象之音節（刪除型錯誤）	發生合併現象之音節（刪除型錯誤）
對 A	dui	76.7% （6）	60.9% （28）	78.8% （10）	63.2% （26）
	A	35.4% （12）	46.1% （26）	39.7% （13）	63.7% （24）
這樣	zhe	33.3% （4）	8.7% （11）	37.3% （2）	28.2% （9）
	yang	49.0% （1）	10.2% （5）	63.3% （3）	33.3% （4）
我們	wo	66.4% （23）	44.2% （29）	68.8% （27）	58.8% （23）
	men	61.5% （3）	39.3% （19）	63.1% （7）	62.0% （11）
然後	ran	67.6% （6）	29.8% （14）	62.9% （8）	51.7% （3）
	hou	63.0% （7）	37.8% （17）	61.9% （4）	51.9% （9）
因為	yin	45.5% （3）	57.9% （7）	54.5% （3）	62.2% （8）
	wei	57.1% （6）	14.7% （18）	66.7% （4）	45.8% （4）
所以	suo	22.2% （2）	39.5% （14）	33.3% （2）	71.1% （7）
	yi	60.8% （44）	29.1% （46）	61.6% （41）	38.9% （44）

其中音節組合”zhe_yang(這樣)”的個別音節辨識率比其它音節都要低的多，可能為其容易合併為音節”jiang(降)”所導致。至於在發生合併現象的音節組合其中之一音節辨識率偏低，則可能為其合併為單一音節組合所導致。

3.3.2 新增多音節聲學模型後在辨識上的錯誤分析

接下來我們觀察嚴重的音節合併現象在辨識上的難題。如果音節合併現象太過嚴重，造成整個音節結構的改變，例如：”zhe_yang（這樣）”因為合併現象發音近似”jiang（降）”，則容易被辨識成單一音節 （此時會造成一個刪除型錯誤或者是一個替代型錯誤加上一個插入型錯誤）。所以即使我們針對合併現象建立多音節聲學模型，仍然無法正確的將這些發生合併現象的音節正確辨識出來，之後我們將分析幾個常見的例子。

下例中有兩處發生音節合併現象卻被辨識為單一音節，可能該處發生嚴重的合併現象，所以基本 411 音節模型較多音節聲學模型佔優勢；有一處雖然沒標記發生合併現象，但仍被多音節聲學模型辨識出來，可能該處有發生不嚴重的音節合併現象。

標準答案：**yin_wei** wu lai ne bian de wen quan **bi_jiao** **mei you** MHM zhong

（因為 烏來 那邊 的溫泉 比較 沒有 MHM 重…）

「基本 411 音節模型」辨認結果

辨識結果：**yin** wu lai nen bie wen xuan **biao** **mei ying** zhong

「非合併型 411 音節模型」和「多音節聲學模型」辨認結果

辨識結果：**yan** wu lai ne bie wen xuan **biao** **mei_you** MHM zhong

由表 3.17 可看出其實蠻多情況是因為發生嚴重的音節合併現象而被辨識為單一音節的結果，可見我們所建立的多音節聲學模型並不能解決大部份的音節合併現象所造成辨識率低落的問題，因為有一部份在嚴重的合併現象下發音已近似單一音節，而這些是我們目前所無法解決的。

表 3.17：音節合併現象嚴重程度觀察

標準答案 發生合併現象	ALL	辨識結果 為單一音節（分數較高）		
dui_A 對 A	68	dui 對（6）		
wo_men 我們	52	weng 溫（5）	men 們（5）	dong 東（3）
ran_hou 然後	48	hao 豪（2）	nao 腦（2）	
zhe_yang 這樣	53	jiang 降(11)	yang 樣(2)	shi 是（2）
yin_wei 因為	43	yin 因（5）		
suo_yi 所以	47	sui 雖（4）		
jue_de 覺得	51	jue 覺（14）	que 缺（2）	jiu 就（2）
jiu_shi 就是	41	jiu 就（7）	shi 是（1）	

由上表可看出，某些音節組合發生合併現象時特別容易合併為另一個單一音節，例如：”zhe_yang（這樣）”特別容易合併為”jiang（降）”的音；有的則是容易忽略掉其中一個音節，例如：”jue_de（覺得)”特別容易忽略掉 ”de(得)”的音等等…。

3.3.3 使用多音節聲學模型對 MCDC 語料中音節合併現象做自動標示

為檢查 MCDC 語料庫中音節合併現象標示是否一致，我們對音節合併現象的標記做自動標示。當音節合併現象不嚴重時，在辨識時我們所訓練的多音節聲學模型相當有可能會跟（兩個）單一音節模型發生互相競爭的情況；我們藉著已經建立好的多音節聲學模型，重新對訓練語料做辨識，也就是對音節合併自動標示的工作，並與人工標示結果做比較。其一可確認音節合併模型之精確性；其二檢查人工標示之一致性。

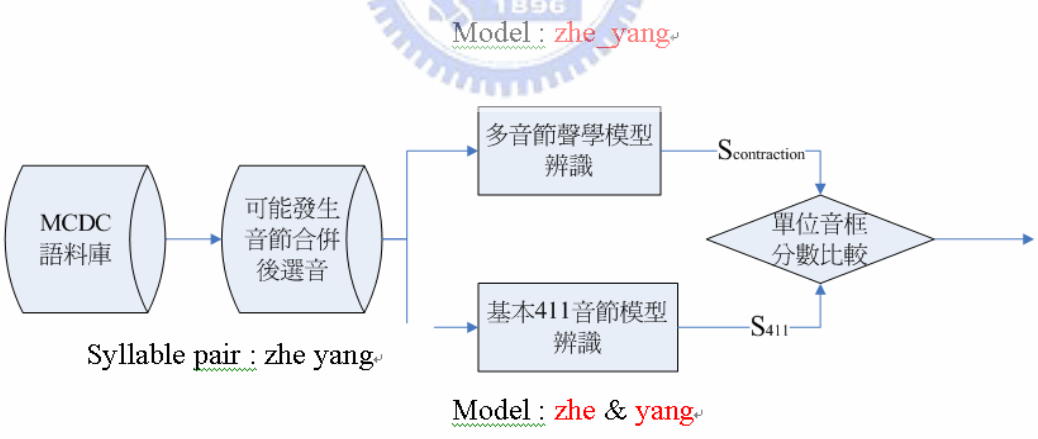

圖 3.8：對音節合併現象做自動標示流程圖

如圖 3.6 所示，我們藉著專門為音節合併現象所建立的多音節聲學模型，對訓練語料中有標記發生音節合併現象的音節組合重新作分數（ log probability ）上的評估；同時我們也拿基本 411 音節辨識系統對同個地方作分

數上的評估；如此一來，便可觀察出是否音節合併現象會具有相同特性及音節合併現象的標記狀況等等…。為了觀察不同模型在同一處的辨識狀況，我們定義分數的求取方法，其步驟如下列：

(1)我們利用專門為音節合併現象所建立的多音節聲學模型，對訓練語料中有標記發生音節合併現象的音節組合觀察其音框的平均辨識分數 （ log p / per frame ），是為 $S_{con.}$。

(2)再利用基本 411 音節模型對訓練語料中有標記發生合併現象的音節組合觀察其音框的平均辨識分數 （ log p / per frame ），為 S_{411}。

如此一來在某個有標記發生音節合併現象的地方，我們可以得到兩個不同的分數，一個是用多音節聲學模型去觀察所得到的分數，一個是用基本 411 音節模型去觀察所得到的分數，我們將兩個分數相減得到

$$S_d = S_{con.} - S_{411}$$

為兩個分數的差距，若 S_d 為正則代表多音節聲學模型所觀察到的分數較高（確實發生音節合併現象）；反之若為負，則代表基本 411 音節模型所觀察到的分數較高（可能音節合併現象較輕微，未標記），我們對所有音節合併現象標記的作觀察。

(1) 多音節聲學模型對標記處的合併現象作觀察(訓練語料)

因為是對訓練語料作分數觀察，所以多音節聲學模型分數高於基本 411 音節模型是正常的，如圖 3.7，觀察重點為那些基本 411 音節模型分數高於多音節聲學模型的資料點應該回歸為未發生合併現象的標記。

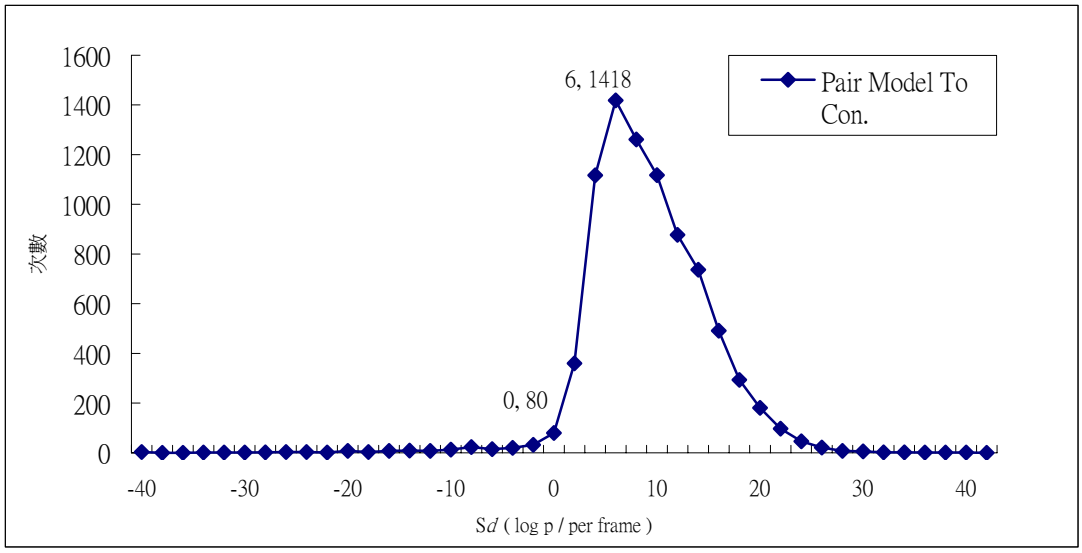

圖 3.9：多音節聲學模型對標記處的合併現象觀察(訓練語料)

(2) 多音節聲學模型對未標記合併現象處音節組合作觀察(訓練語料)

接下來我們反過來觀察是否有發生音節合併的音節組合卻未被標記員標記出來 的狀況，做法同上。如圖 3.8 所示：在 7446 筆資料中僅約 420 筆資料點利用多音節聲學模型來辨識的分數是較基本 411 音節模型高的，這些地方應該是歸為發生音節合併現象的標記 （下圖中紅色曲線中偏右的資料點），但是整體看來所標記的音節合併現象是蠻一致的。

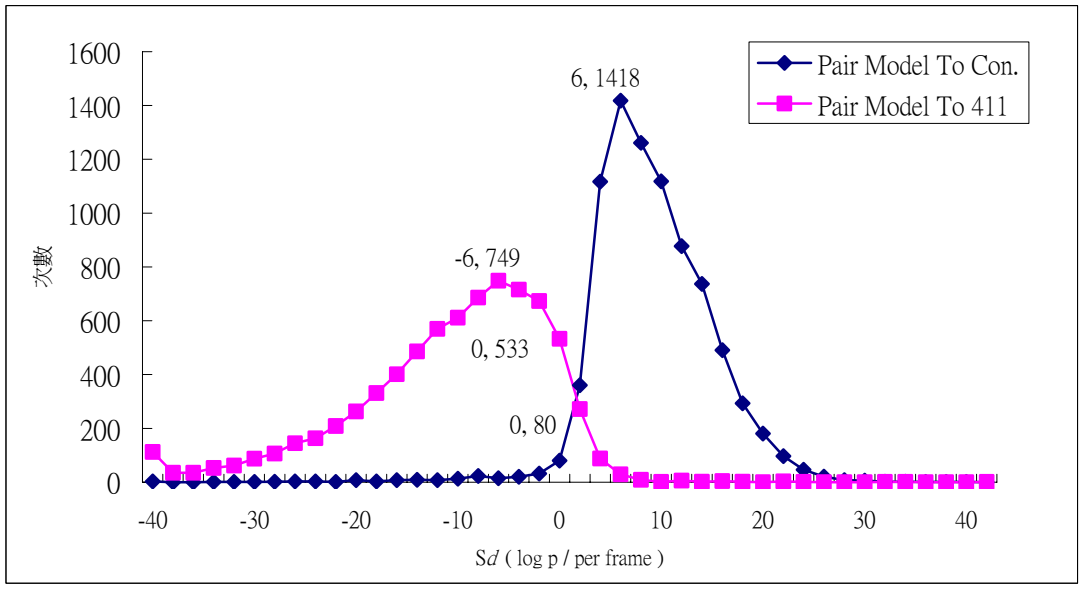

圖 3.10：多音節聲學模型對標記與未標記合併處音節組合觀察(訓練語料)

(3) 多音節聲學模型對標記處的合併現象作觀察(測試語料)

接下來我們觀察測試語料的狀況。如圖 3.9，看到多音節聲學模型對於標記音節合併處的辨識，稍較基本 411 音節模型為佳；雖然可看出區隔，但是偵測效果差。

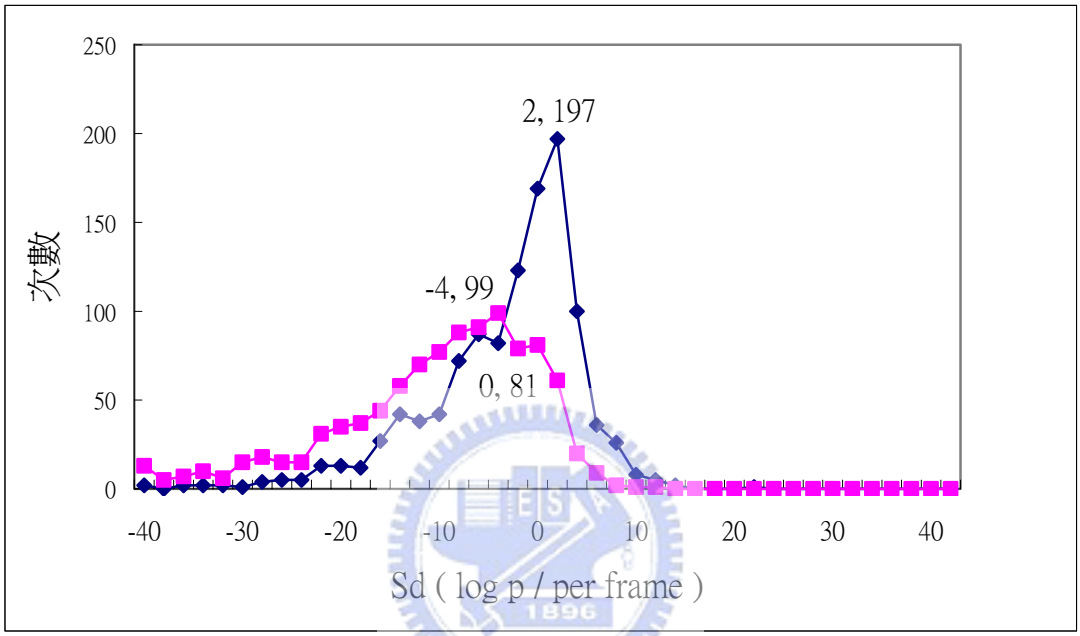

圖 3.11：多音節聲學模型對標記與未標記合併處音節組合觀察(測試語料)

(4) 針對 Top7 Model 的音節組合觀察

如果我們只針對 Model 最好的前 7 個模型來看其音節組合標示的正確性，我們可以觀察到對於標示發生合併現象處約有 64%是利用多音節聲學模型來辨識的分數是較高的(橘色曲線)；至於未標示發生合併現象處約有 77%是利用 411 音節模型來觀察分數是較高的(綠色曲線)。整體來說，如果我們能夠擁有足夠多發生合併現象的人為標記，則我們將可以將多音節聲學模型 Model 的更好，這樣一來就可以在未標記合併現象的語料中，有效的偵測合併現象的發生，進而訓練較不受合併現象影響的精確 411 音節模型及受合併現象影響的多音節聲學模型，提升整體的辨識率。

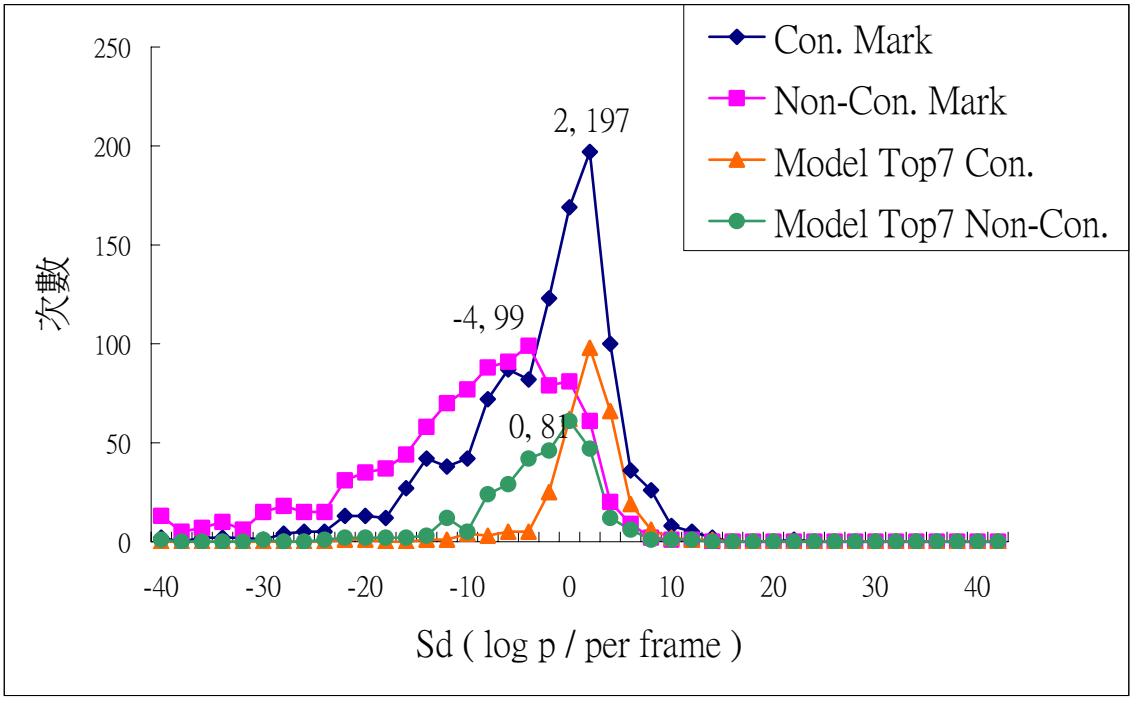

圖 3.12：多音節聲學模型對標記與未標記合併處音節組合觀察(top7 Model)

在曾淑娟老師的研究報告中曾指出不同人在自發性語音發生音節合併現象時會有不同的狀況，也就是合併後的音節結構可能會有所不同，也許這是多音節聲學模型對於測試語料自動標示是否能夠再提升的關鍵之一。

3.3.4 右相關音節模型描述合併現象(RCD Final Model)

之前我們已經針對經常出現合併現象的音節組合建立多音節聲學模型，但對於出現次數較少之發生合併現象音節組合則只能建立右相關韻母模型。雖然右相關音節模型不足以完整 Model 音節合併現象，但是比起之前對於發生合併的音節所建立的 411 音節模型仍來得好許多。

3.3.4.1 模型建立過程

我們依合併現象後一音節聲母發音特性作分類，分類如表 3.18 所示。我們

將獨立的聲母依 爆破音/鼻音…等分為五類；空聲母音節則依韻母發音特性分為六大類。依照下表的分類，對合併現象的前一音節建立右相關韻母模型。

表 3.18：合併現象後一音節聲母分類

類別	後一音節	
1	爆破音	ㄅ、ㄆ、ㄉ、ㄊ、ㄍ、ㄎ
2	鼻音	ㄇ、ㄋ
3	摩擦音	ㄈ、ㄏ、ㄒ、ㄕ、ㄖ、ㄙ
4	邊音	ㄌ
5	塞擦音	ㄐ、ㄑ、ㄓ、ㄔ、ㄗ、ㄘ
6	a-	ㄚ、ㄢ、ㄠ、ㄢ、ㄤ
7	o-	ㄛ、ㄡ
8	e-	ㄜ、ㄝ、ㄟ、ㄥ、ㄣ、ㄦ
9	yi-	ㄧ、ㄧㄚ、ㄧㄝ、ㄧㄢ、ㄧㄠ、ㄧㄡ、ㄧㄢ、ㄧㄣ、ㄧㄤ、ㄧㄥ、ㄧㄛ
10	wu-	ㄨ、ㄨㄚ、ㄨㄛ、ㄨㄢ、ㄨㄟ、ㄨㄢ、ㄨㄣ、ㄨㄤ、ㄨㄥ
11	yu-	ㄩ、ㄩㄝ、ㄩㄢ、ㄩㄣ、ㄩㄥ

在模型建立數量方面我們仍然依據足夠的訓練語料，總共訓練出發生合併現象的 114 個右相關韻母模型。

3.3.4.2 新增右相關韻母模型後之辨認器架構

在辨識上我們聯合「非合併型 411 音節模型」、「多音節聲學模型」和「右相關韻母模型」參與辨識。辨識路徑如圖 3.10 所示，與之前不同的為我們針對發生合併現象音節建立右相關韻母模型，所以在音節的連接上必定有路徑上的限制；例如：假使該音節的韻母為發生合併現象的右相關爆破音韻母模型，則所連接的下個音節的聲母必定為發生合併現象的爆破音聲母模型等等...。同樣的這裡我們在辨識路徑上不留其它分數。

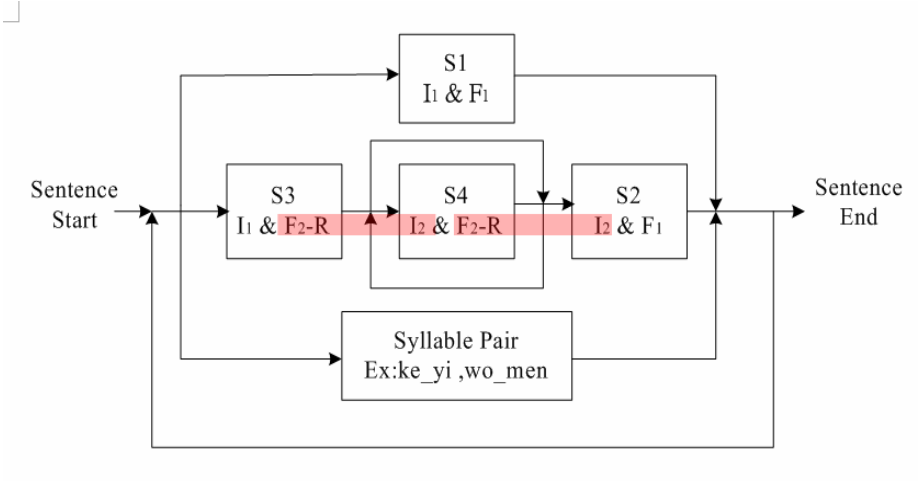

圖 3.13：RCD 模型辨識器連接

3.3.4.3 新增右相關韻母模型辨識結果與分析

下表 3.19 為比較新增右相關韻母模型後在辨識方面的改進狀況。整體來說音節辨識率些微的下降，刪除型錯誤雖有改善，但正確音節數、替代型錯誤數、插入型錯誤數都變差了，有點類似之前新增「合併型 411 音節模型」參與辨識後的結果。其可能原因其一可能是基本上發生合併現象的訓練語料已經不是很多，再加上我們在訓練時是拿不足以建立多音節聲學模型的語料來訓練模型，所以可能因訓練語料數量不足導致模型不精確；其二可能為合併現象造成音節結構嚴重改變，音節的聲母或韻母被改變或已經忽略掉了，此時我們建立其右相關韻母模型在辨識上便無法得到好的結果。

表 3.19：比較新增多音節聲學模型後的辨識狀況

參與辨識的模型	辨認率	正確音節數	刪除型錯誤	替代型錯誤	插入型錯誤
非合併型 411 音節模型 & 多音節聲學模型	42.88%	48.2%	12.0%	39.8%	5.4%
非合併型 411 音節模型 & 多音節聲學模型 & 右相關韻母模型	42.12%	47.5%	11.1%	40.5%	6.3%

3.4 實驗結論

由前面的實驗結果我們可以發現：

1. 建立常用詞的多音節聲學模型為一有效方法解決因合併現象所造成音節辨識率低落的問題。經由之前實驗結果可知新增多音節聲學模型後，原本標記未發生合併現象的音節辨識率並無明顯提升，但是標記發生合併現象的音節辨識率確實有相當大的提升，同時在刪除型錯誤方面也有相當的改善，達到我們改善發生合併現象音節辨識率的目標。

2. 嚴重的合併現象為無法利用聲學模型來辨識。如果音節合併現象太過嚴重，造成整個音節結構的改變，例如：”zhe_yang（這樣）”因為合併現象發音近似”jiang（降）”時即使我們針對合併現象建立多音節聲學模型，仍然無法將這些發生合併現象的音節正確辨識出來。

第四章 結論與未來展望

在本論文中，我們先針對自然語音對話語料中的音節合併現象做分析與統計，得知音節合併現象的發生多為我們日常對話常用詞的特性，而跟音節在聲學上的特性並無多大相關，所以對於在 Read Speech 語料中能夠有效提升辨識率的前後文相關模型在音節合併現象上並無效果。

由實驗結果得知，建立常用詞的多音節聲學模型的確能有效的解決因發生合併現象所造成音節辨識率低落和刪除型錯誤發生的問題，但是對於嚴重的音節合併現象，例如：”zhe_yang（這樣）”因為合併現象發音近似”jiang（降）”則即使是一般正常人來聽(只針對發生合併現象處)可能也難以正確的辨識出來，這時就必須參考前後文的語意才能夠正確辨識出來。

未來研究方向可朝下列幾點進行：

1. 對於發生合併現象後的聲學特性是否會跟音節本身有關，例如：是否是以其中一個音節為主體或是合併成另外一個音節在本論文中並無詳細探討，在第三章中只有列出幾個常見例子稍加觀察，所以我們或許可以對這方面加以研究，以建立更精確的聲學模型，以提升辨識率。
2. 對於發生嚴重合併現象的音節，在辨識上我們無法單純的以多音節聲學模型來解決，這時必須參考上層的語意內容來輔助辨識，例如加入 Language Model 參與辨識，才有可能正確辨識出來。

參考文獻

[1] 曾淑娟、劉怡芬, ”現代漢語口語對話語料庫標註系統說明”, 中*文詞知識庫小組*，民國九十一年一月

[2] Shu-Chuan Tseng, “Feature of Contraction Syllable of Spontaneous Mandarin” EUROSPEECH 2003 – GENEVA, PP. 77-80.

[3] S. Young, G.. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, “The HTK Book（for HTK Version 3.2.1）”