使用韻律階層及大量詞彙的中文文轉音系統

A Mandarin Text-to-Speech System Using Prosodic Hierarchy and a Large Number of Words

余明興、張唐瑜、許燦煌、蔡育和
國立中興大學資訊科學所
msyu@dragon.nchu.edu.tw, s9256047@cs.nchu.edu.tw, s9256040@cs.nchu.edu.tw, s9256013@cs.nchu.edu.tw

摘要

本論文實作了一個中文的文轉音系統(Test-to-Speech)系統，它使用大量的詞彙來做爲合成單元(Synthesis units)，並且配上適當的韻律階層。韻律階層可以使語意更加清晰，也可以幫助選取適當的合成單元。因此本篇論文主要包含兩個重點：韻律階層的求取和以大量詞彙作爲合成單元的架構，在韻律階層的求取上，我們實驗了利用剖析器爲基礎的方法以及著名的統計式方法-CART(Classification And Regression Trees)來進行求取。我們使用大量詞彙來當成我們的合成單元，可以免去許多語音處理不易做好的連音處理。我們也利用韻律預估模組所得到的參數，進行音量和音長的調整。最後我們完成一套包含 12224 個二字詞以及 2690 個三字詞的中文文轉音系統，並開放於線上試用。

關鍵字：Text-to-Speech, Parser, Prosodic Hierarchy.

1.緒論

1.1.中文文轉音系統

近年來文轉音系統在實作上，最常見到的爲波形拼接法（waveform- concatenation）。這種作法主要是利用預先錄製好的聲音，稱之爲合成單元（synthesis units），存放在語音資料庫中，要用時再將其取出拼接，來合成所要讀出的語句。這些合成單元要能包含所有可能的發音，這些預錄的單元可能是音素(phoneme)、雙音素(di-phone)、音節(syllable) …等。

傳統的 TTS 包含三個模組：（一）文句分析(Text analysis)：這部份包含斷詞以及一些語言知識上的標記，例如詞類（Part-of-Speech, POS）等，另外也會處理字轉音。（二）韻律預估(Prosody prediction)：預估合成音的音長（Duration）、音量（Energy）、音高(Pitch)等聲學參數。（三）語音合成(Speech generation)：利用已經預估好的韻律參數來進行韻律的調整。常見的方法，有 PSOLA（Pitch-Synchronous Overlap and Add）[9]…等。傳統的架構發音不自然，主要的問題[8]爲（一）連音無法獲得充分解決，（二）韻律調整範圍過大。因爲早期記憶體的限制，錄製大量語音來做爲合成單元非常困難，所以合成語音品質遲遲無法突破。

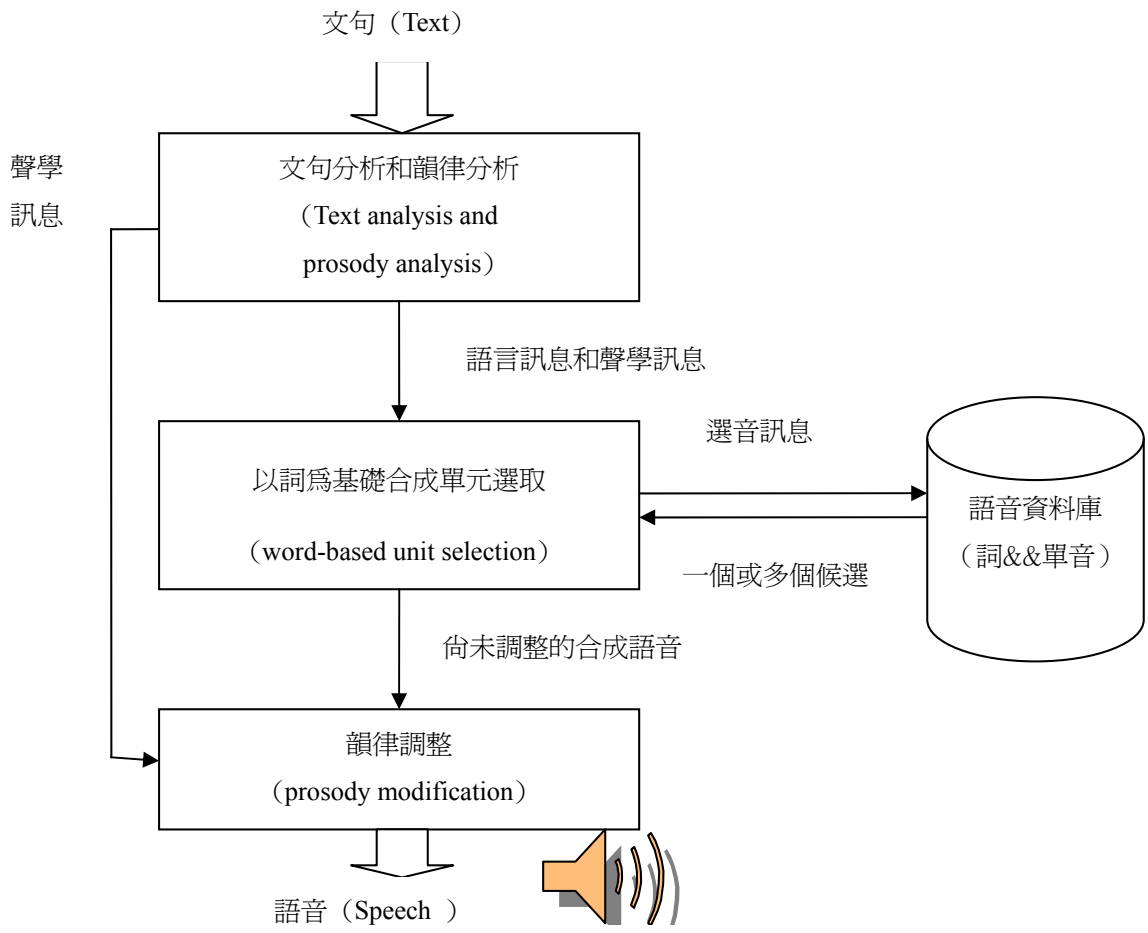
近年來，文轉音(Text-to-Speech)系統漸漸以 Corpus-based 的架構來實作[6][9]，並以波形拼接來做爲合成方法。該方法的優點在於可以得到不錯的語音品質，相較於傳統以音節爲合成單元的架構，更容易被聽者接受。這種架構的特徵在於：（一）會錄製大量語料，（二）盡可能不做訊號處理，（三）選擇適當的合成單元。軟體界的巨人－微軟，所推出的「木蘭」雙語(中文以及英語)TTS

系統[6]，便是使用 Corpus-based 的方式所組成的。錄音的語料爲錄製句子(sentence)。在連續音語料上採用一個階層式的韻律模型。最後再利用 Decision Tree[6][9]的方法來找尋適當的非固定長度合成單元來進行拼接，過程中，完全不做訊號處理。

本論文所提出的系統，使用大量的詞彙來做爲合成單元，並配合適當的語音處理。我們認爲從詞(主要是二字詞和三字詞)中抽取出來的合成單元，在音程上較爲完整，聽的較清楚。而且當我們用許多的詞來做爲合成單元時，大部分的連音現象都已包含在合成單元中，而連音是語音合成的各種處理中非常難以做好的部份。語音合成中的音量調整只要不超出位元數的最大音量限制，並不會影響聲音的品質，所以可以做較大程度的調整。語音合成中的音長調整，只要範圍不大，可以用切音加上淡出的方式來處理，對語音品質的影響也很小。

整個系統的流程圖如圖一所示。文句分析和韻律分析模組負責提供語言訊息和聲學訊息，聲學訊息包含音長與音量，語言訊息主要爲韻律階層，這是決定停頓長度和選取合成單元的根據。接下來經由單元選取模組來選音拼接，最後的韻律調整模組則是利用韻律分析預估出的聲學參數來進行最後的微調，例如做淡入（fading-in）和淡出（fading-out）等。

在文句分析的工作方面，我們使用了 rule-based 的斷詞方法[10]，用 bigram 機率模型來標記詞類，還有做字轉音…等。韻律階層的求取我們實驗了使用 CART 以及使用剖析器兩種方法。在本節的其它篇幅，我們會介紹中文剖析器、韻律階層和我們所使用的大量詞彙。

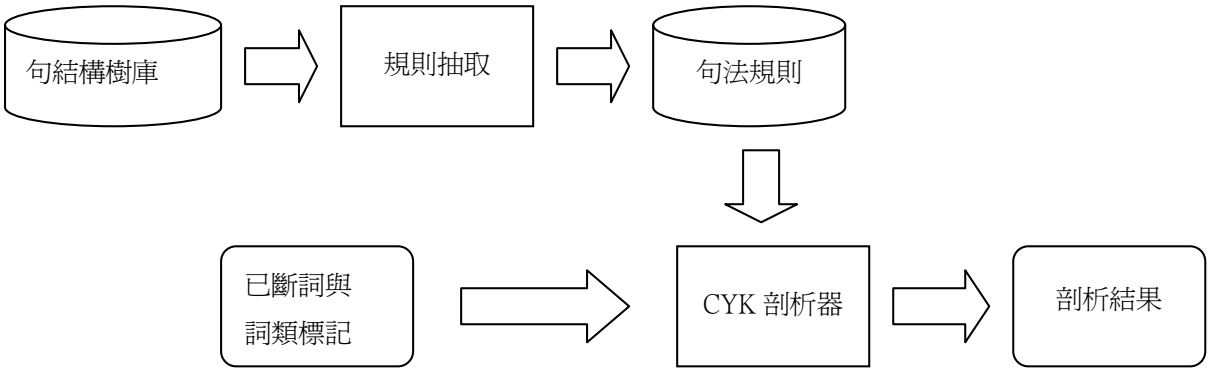


圖一 － 系統流程圖

1.2. 中文剖析器

句法剖析[1]在自然語言處理的過程中有許多用途，如：問答系統、機器翻譯、關鍵字擷取等等。在本文中我們完成一個剖析器，其用途是用來求取韻律階層，利用此韻律階層讓TTS(Text-to-Speech)系統的語音可以聽起來更加好聽。

在過去研究中，利用樹庫(treebank)訓練出來的 probabilistic context-free grammar(以下簡稱PCFG)，拿來對句子做剖析是很常用的技術。在英文部分，因爲有大量的英文樹庫資料，所以利用此英文樹庫所訓練出來的 PCFG 來剖析英文句子，依目前的資料顯示正確率約可至九成[5]。中文剖析器方便目前由中研院所完成的中文剖析器其正確率約在六成[15]。本文中的剖析器也是利用 PCFG 來剖析句子，所訓練出來的 PCFG 是由中研院的中文句結構樹庫中所擷取出來的。，我們是使用 Bottom-up 的 Cocke-Younger-Kasami（CYK）演算法[1][7]來實做我們的中文剖析器。此剖析器的系統架構如圖二所示。



圖二 - 剖析器系統架構圖

1.3.韻律階層

我們的 TTS 系統使用到韻律階層架構，它主要的用途是用來預測停頓的位置和類型。韻律階層由上到下爲句子(Sentence，簡寫爲 S)、韻律片語(Prosodic phrase，簡稱 PP)、韻律詞(Prosodic word，簡稱 PW)、和詞(word)。韻律階層可以描繪成一個樹狀的結構，稱爲韻律階層樹。一顆典型的韻律階層樹如圖三所示。這個架構下的階層共有 4 個，下面爲各階層的定義：

1． 詞(word)

定義：詞爲有意義的基本單位。

例：台北。

2． 韻律詞（Prosodic Word，簡稱爲 PW）

定義：組成單元爲一個或多個詞。

例：台北去玩。

3． 韻律片語（Prosodic Phrase，簡稱爲 PP）

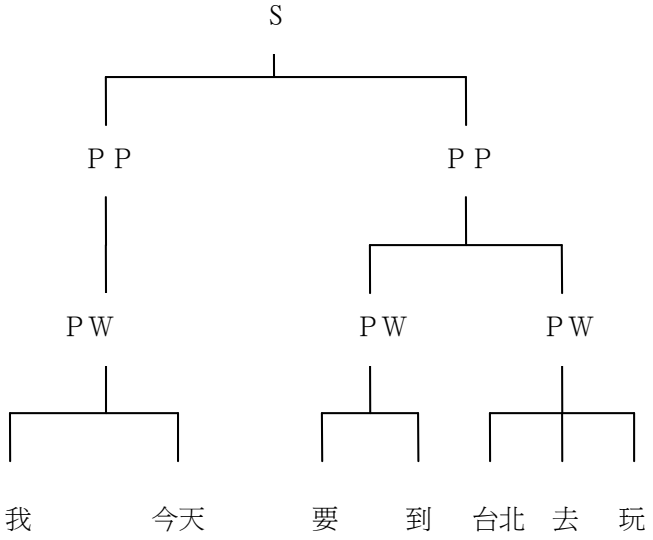
定義：組成單元爲一個或多個韻律詞。

例：要到台北去玩。

4． 句子（Sentence，簡稱爲 S ）

定義：以五大標點符號（包含：“，” “。” “；” “！” “？”）爲區隔，組成單元爲一個或多個韻律片語。

例：我今天要到台北去玩。



圖三 - 韻律階層樹

韻律詞裡面不停頓，爲一個連續發音的單位，韻律詞跟韻律詞之間爲小停頓（minor break），韻律片語跟韻律片語之間爲中停頓（major break)，另外還有一種大停頓是發生在標點符號出現的時候。停頓對於語音的可辨度和自然度有相當程度的影響，良好的停頓可以讓人易於理解句中涵義。

1.4.大量詞彙

我們有鑒於錄製句子，可能無法錄到一些罕見詞，以及考慮到詞跟詞之間的連音現象較微弱[6] ，所以嘗試使用大量詞彙(word-based)當成我們的合成單元。換句話說，如果可以錄製大量詞彙來做爲合成單元，這些合成單元會包含相當豐富的連音訊息。表一所示就是 word-based 對於 corpus-based 主要的優缺點比較。主要的缺處在於句子韻律上的音高問題。人在講話時，詞的發音會因爲在句中位置的不同，而有不同的音高變化。而錄詞的方式在選音的時候，合成單元聲調高低的變化範圍較小。除此之外，以音節平均音長來說，當我們錄製詞或單音的時候，合成單元會比錄製句子時來的長，錄音者必須注意這個問題。

表一 – word-based 相對於 corpus-based 的優缺表

優點	缺點
1． 包含較多連音訊息。 2． 合成單元音程較完整。	1． 較無法錄製到帶有句子韻律的音。 2． 錄音時需注意音長。

現階段在系統的實作上，總共錄製約 12224 個二字詞和 2690 個三字詞，這些詞是出現頻率較高的詞。另外，還有一個單音庫，這個單音庫包含所有中文可能出現的音。在不管聲調的情況下，中文約有 409 個音，而中文有五個聲調，所以這個單音庫整整錄製了 409*5 種音，雖然說實際上中文的發音只會有約 1300 種。

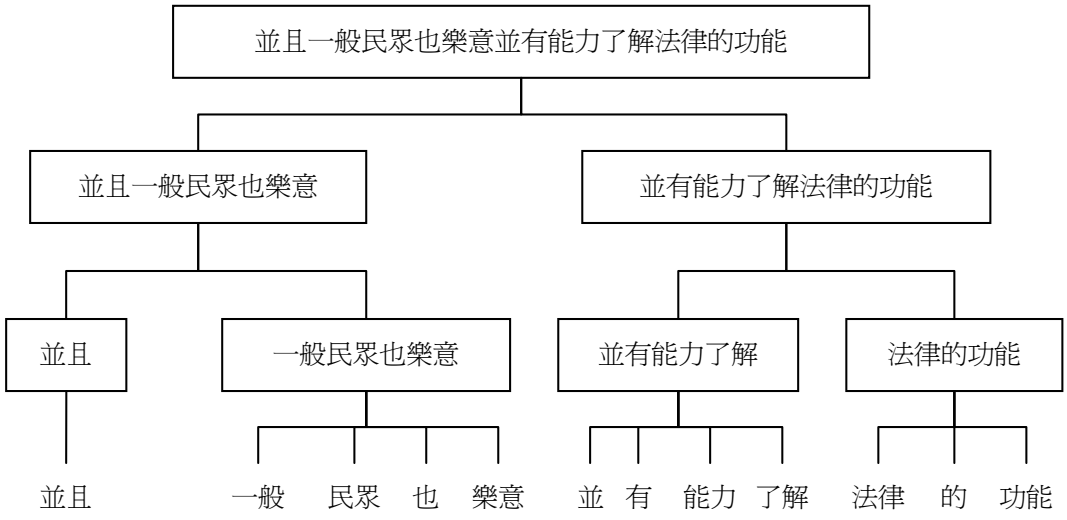
2.韻律階層的求取

在本節中，我們先描述如何得到帶有停頓標記的實驗語料。然後說明求取停頓標記的兩種方式：使用 CART 和使用剖析器的方式。

2.1.標記實驗語料

我們求取韻律階層的實驗語料，是從中研院的中文句結構樹庫 2.1 裡的句結構樹，拿掉文法剖析的結果，還原成原本的句子。再由實驗室的成員，以自己的感覺下去標記的。標記的方法是實際唸過一次，並從自己唸的方式來對句子標記韻律片語和韻律詞的中斷。總共有九份語料，除了第九份只有 1,923 句外，其於八份每份有 6,250 句。由實驗室的成員一人標記一份語料。在檢查標記完的結果時，若某句的標記有將中斷標在原本結構樹的詞內的話，則視爲無效的標記，並刪除該句。經過這項檢查後，全部有效的標記句子有 51,525 句。以下列出一個帶有標記的句子和它所對應的韻律階層樹（圖四），其中＊表示韻律片語的邊界，空白表示韻律詞的邊界。

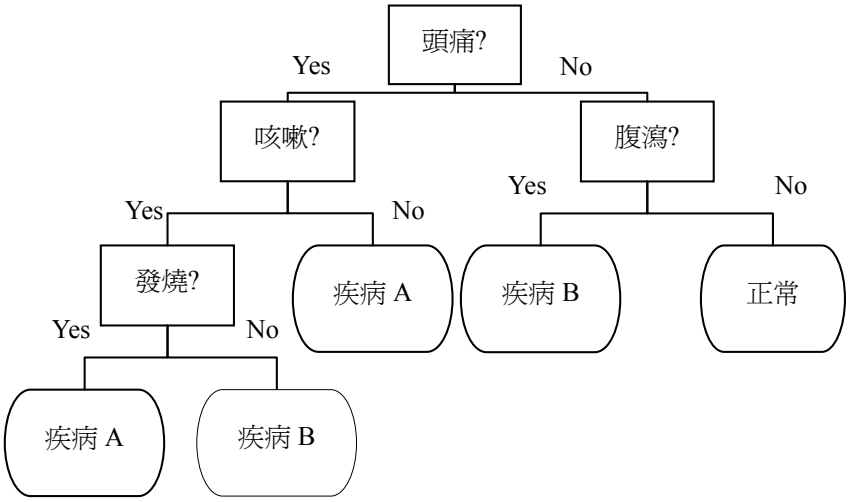
EX:並且　一般民眾也樂意＊並有能力了解　法律的功能



圖四 - 韻律階層樹

2.2. CART(Classification And Regression Trees)

CART 是一種二元(binary)分割的方法。分割條件的選擇是根據資料的分類數及其屬性來決定，並依據 Gini 規則[2]來決定分割的條件。每經過一次分割後，資料會被分成兩群，如圖五所示。經由不斷的切割資料，達到分類的目的。

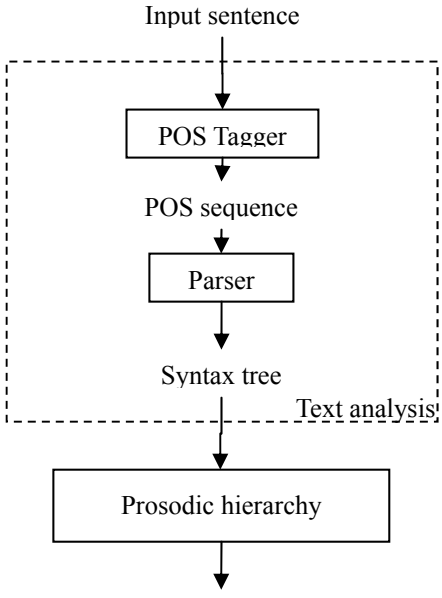


圖五 CART 分類樹

以圖五爲例子，假設我們要分類的是得到 A 疾病和 B 疾病以及正常的人。首先我們先確定是否有頭痛的現象，若是沒有，則走到腹瀉（即檢查腹瀉的節點），若是有，則到咳嗽（即檢查咳嗽的節點）。接下來再繼續問問題，此時腹瀉這個節點的集合已經可以分類出來了，一個是正常，而另一個是有 B 疾病。其它的類推，便是以此種方法來分類。

我們將詞和詞之間的停頓分成下列三類：無停頓(no break)、小停頓(minor break)、以及中停頓(major break)，所以我們可以使用 CART 這樣的方法來分類。而 CART 的特點是，它是自動來產生這樣的決策樹，也就是我們只需準備語料和所用到的特性，便可以使用 CART 來對資料作分類的工作。

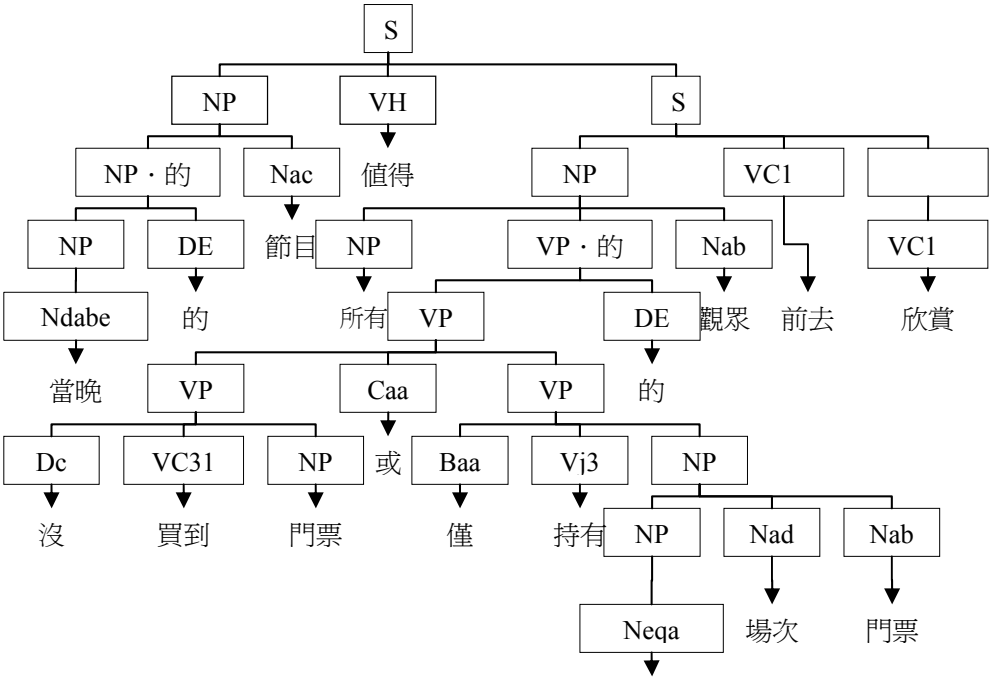
2.3.剖析器爲基礎的方法



圖六 - 模組關係圖

我們的另一個求取停頓型態的實驗，是以文法剖析爲基底的方式[13]。先求得整句的文法樹之後，再從該文法樹來得到整段句子的韻律階層(prosodic hierarchy)。這個模組在整個系統中是包含在文句分析(Text analysis)這個大模組下。大略的架構如圖六，句子進來先經過斷詞器斷詞並標記詞性，將輸出的結果輸入給文法剖析器，接著得到該句子的文法樹，再從文法樹中得到韻律階層。

接下來我們用實例來說明這種方法。假設輸入的句子爲：當晚的節目值得所有沒買到門票或僅持有其它場次門票的觀眾前去欣賞，此句的文法樹結構如圖七所示。我們會依照由底層往上層合併的方式來求取韻律詞和韻律片語，在合併時我們會給韻律詞（片語）字數的限制。我們依序說明字數的限制和合併的方式。



圖七 - 文法樹圖 1

2.3.1.韻律詞及韻律片語的字數

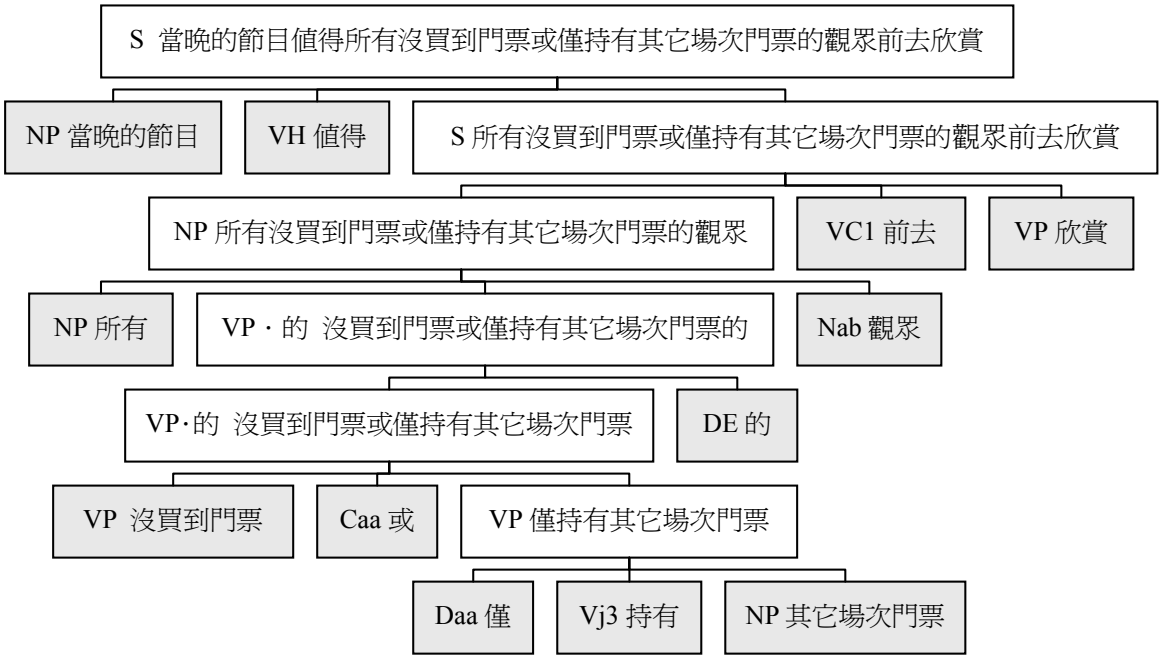
圖七表現出這個句子的文法結構。接著我們便從這樣的結構中找出可能的韻律詞(prosodic word)和韻律片語(prosodic phrase)。我們認爲韻律詞和韻律片語所含的字數會隨著句子的長度(句中字數)而改變。也就是說，在較長的句子中，韻律詞(片語)的字數也會比較多。經過觀察和實驗，我們給予韻律詞(片語)一個最大字數的限制，這個限制會隨著句長而變動。表二顯示我們使用的句長與最大韻律詞(片語)的字數關係。

表二　句長與最大韻律詞(片語)字數關係表

句長 / 最大字數	1-3	4-8	9-12	13-18	19-22	23-30	31-39	40-41	>42
PW 最大字數	句長	6	6	6	6	6	7	7	8
PP 最大字數	句長	句長	9	12	13	14	15	16	16

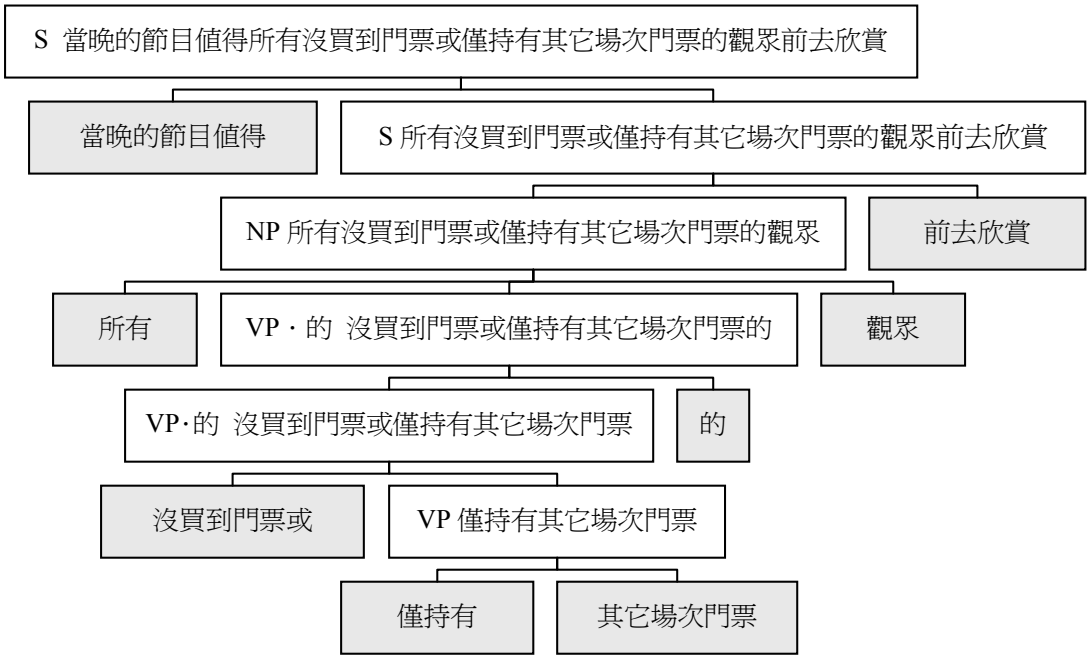
2.3.2. 由底層往上層合併

定好韻律詞和韻律片語的最大字數後，接下來是用由最底層往上層(bottom-up)合併的方式來找出韻律詞和韻律片語。我們以三個步驟來實作這個方法，首先是將小樹縮成韻律詞，接著再同層節點合併成韻律詞，最後由底向上合併。這個方法直接依照長度合併韻律詞和韻律片語。第一個步驟會掃描文法樹，將文法樹中較小的樹縮減成一個節點。經過第一個步驟，我們將小樹的部分縮成韻律詞，縮減的結果如圖八所示。



圖八 - 小樹縮減結果

接下來是第二步驟，同層葉節點的合併。由圖八中可以很清楚的看出同層的單元。在此步驟中，合併的限制是不能超過韻律詞的最大字數。在此步驟我們會將 NP(當晚的節目)和 VH(值得)合併成一個單位，VC1(前去)和 VP(欣賞)合併成一個單位。以此類推，同層合併的結果，如圖九所示。最底層有 Daa(僅)、Vj3(持有)、NP(其它場次門票)，我們優先合併字數較短的節點，所以會合併 Daa(僅)和 Vj3(持有)。

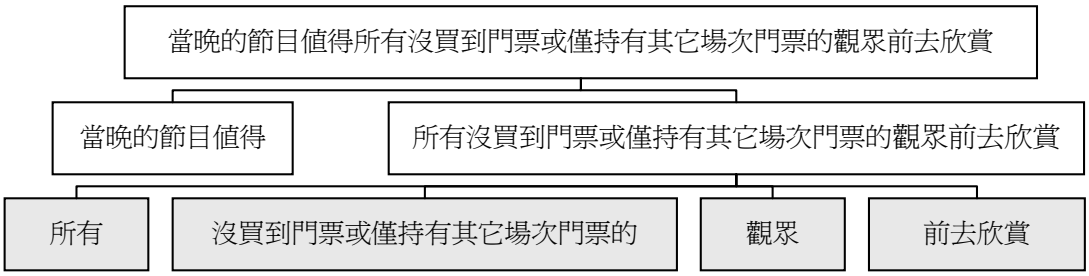


圖九 - 同層合併結果

最後我們由底向上合併韻律詞(片語)，從最底層開始(第六層)，首先處理的就是“僅持有”和“其它場次門票”，兩個節點的字數和爲 9，雖然超過最大韻律詞的字數，但未超過最大韻律片語的字數，所以合併成一新的韻律片語。最底層的部分合併完畢。

接著往上昇一層處理，要處理的單元爲：“沒買到門票或”、“僅持有其它場次門票”。由於兩個單元的字數合爲 15，剛好到最大的韻律片語字數，所以可以再合併成韻律片語。接著往上昇一層，處理“沒買到門票或僅持有其它場次門票”、“的”。雖然兩個單元的字數和會 16，但是由於“的”是詞綴[16]，而在我們的方法中，詞綴合併不受字數限制，因此合併成“沒買到門票或僅持有其它場次門票的”。

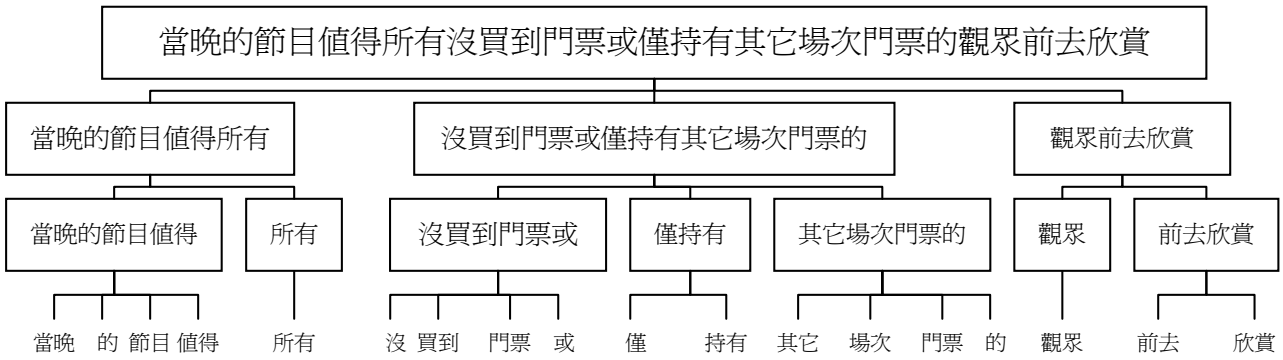
再往上昇一層，處理“所有”、“沒買到門票或僅持有其它場次門票的”、“觀眾”。在這次的處理中，由於字數的限制，因此不會產生任何合併的動作。所以將這三個單元往上昇一層處理，如圖十所示。



圖十 - 上昇處理圖示

接下來處理的單元變成了：“所有”、“沒買到門票或僅持有其它場次門票的”、“觀眾”、“前去欣賞”。而在這個部分，我們可以合併的單元爲“觀眾”、“前去欣賞”，合併的形態爲韻律詞。合併後再上昇，得到四個單元：“當晚的節目值得”、“所有”、“沒買到門票或僅持

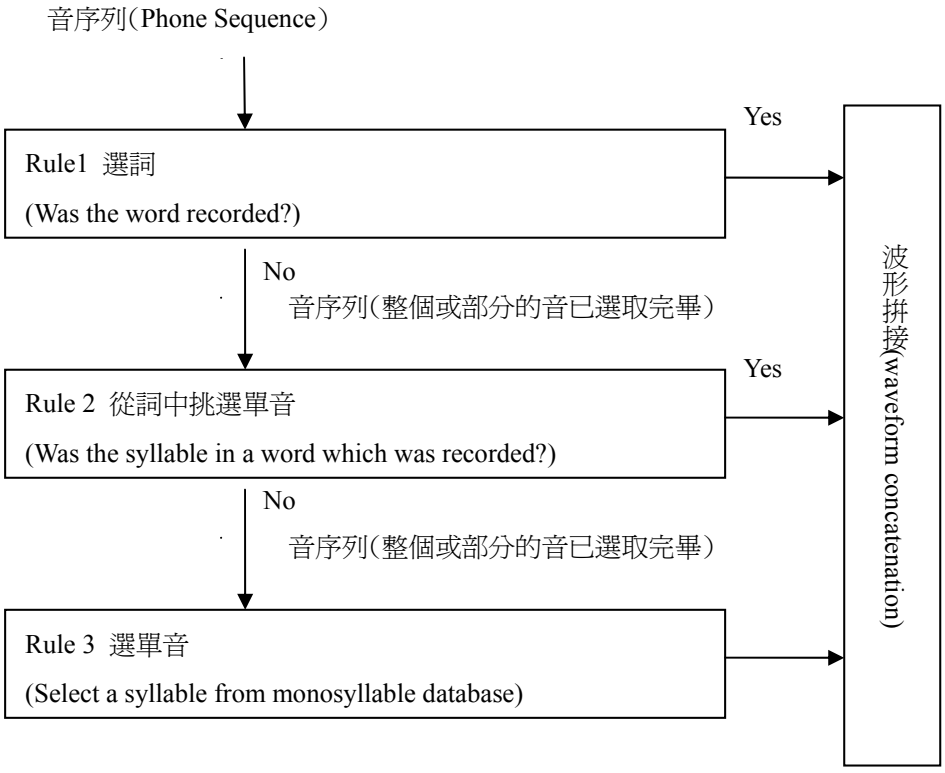
有其它場次門票的”、“觀眾前去欣賞”。而這四個單元可以合併的單元爲“當晚的節目值得”、“所有”，合併的形態爲韻律片語。最後可以得到如圖十一的韻律階層。



圖十一 - 韻律階層圖(Bottom-up)

3.合成單元選取

在選擇適當的合成單元方面，主要是使用 rule-based 的方法來進行挑音，規則爲「有同音詞（二字以上）找同音詞，沒有則從詞（二字以上）中找尋單音，沒有再從單音庫裡面尋找單音拼接」，選音的流程如圖十二所示。



圖十二 - 選音流程圖

在流程圖中，通常 rule1 和 rule3 都只有一個候選，在 rule2 上，可能有許多的候選音，我們使用了一個代價公式來幫助我們選音。這個代價公式的設計，是依據（1）詞中位置（Position in word）和（2）前音與後音的聲調（Tone）和（3）子音/母音結構（C/V 結構）來做音節單元的

挑選。這個代價公式所呈現的實際上的意義爲「先選詞中位置相同的，再考慮前後音的子音、母音和聲調的代價」。先介紹詞中位置的定義，詞中位置包含：詞首、詞中、詞尾，例如：後悔莫及（“後”爲詞首、“悔”和“莫”爲詞中、“及”爲詞尾）。我們以一個範例解釋從詞中選單音的概念。表三爲選音時的ㄉㄤ的候選。

表三 – 選音範例

候選音	候選所在詞	前音是否在詞內	前音	後音是否在詞內	後音
ㄉㄤ	**當**天	否	ㄑㄩㄢˊ	是	ㄊㄧㄢ
ㄉㄤ	便**當**	是	ㄅㄧㄢˋ	否	ㄧㄡˇ
ㄉㄤ	便**當**店	是	ㄅㄧㄢˋ	是	ㄉㄧㄢˋ
ㄉㄤ	**當**選	否	ㄕˋ	是	ㄒㄩㄢˇ

假設我們合成的句子的斷詞結果爲「當晚　我們　要　到　台北」，而(ㄉㄤ　ㄨㄢˇ)這個同音詞並沒有錄過，接下來便要從詞中去選擇ㄉㄤ。因爲目標音爲詞首所以只要考慮詞中位置和後音的代價。“當天”和“當選”的詞中位置和目標位置都是在詞首，所以優先選之。在這兩個ㄉㄤ的候選中，“當選”的ㄉㄤ因爲“選”的聲調和“晚”的聲調是相同的，因此計算出來的代價爲最小。所以最後會選擇“當選”的ㄉㄤ來做拼接。（詳細請參考[11]）

4.韻律調整

通常中文 TTS 要調整的韻律包含音調、停頓、音長以及音量。我們以合成單元選取的方式來得到適當音調的合成單元，而不做音調的調整。在音長和音量的調整上，我們用[13]的統計方式韻律預估模型來預估各音節的音長和音量，然後盡量調整成所要求的音長和音量。

在韻律調整模組上，主要處理工作包含停頓值的給予、適當的重疊（overlapping）、音長調整和音量調整。停頓的部份，我們採取給予固定停頓值的作法，我們根據聽測結果來決定適當的停頓值，如表四所示。

表四 – 停頓時間表

停頓類型	停頓長度(ms)
Minor break	50
Major break	250
逗號（，）	400
句號（。）	625
問號（？）	625
驚嘆號（！）	625
分號（；）	500
頓號（、）	250
冒號（：）	300

在我們的韻律階層架構中，Prosodic Word 內是不會發生停頓的。但是在實際利用合成單元來拼接的時候，某些合成單元間直接拼接會造成連接處停頓的感覺，這跟韻律階層的概念有所出入，所以我們在合成單元間給予短暫的重疊（Overlapping）。重疊的長度爲我們參考[8]所制定出來的，如表五所示（ㄐ代表爲無聲母）。

表五 - 重疊長度比例表

後音節子音的類型	重疊的部分佔目前音長的比值
ㄅㄆㄉㄊㄍㄎ	0
ㄐㄓㄕ	0.05
ㄑㄔㄘ	0.1
ㄈㄏㄒㄕㄙ	0.15
ㄇㄋㄌㄖㄐ	0.2

在音長的調整上，我們並不使用 PSOLA 來進行調整，而是使用 cut off（切音，切掉單音後段）的作法，這個方法簡單而且能確保音的清晰。主要的原則爲「如果目標音比合成單元短則進行切音，否則不進行調整」。在切音之後，我們會做一個淡出（fading-out）的處理，使其聽起來較爲自然。

在音量上，只要不要調整到爆掉，皆不會發生如調整音長時的失真情形。淡入（fading-in）和淡出（fading-out）主要用來降低音量上的突兀，假設我們要調整的部分總共有 n 個點，每點振幅值爲＜$x_1, x_2, x_3, ..., x_n$＞，淡出的公式如(式 1)所示。同樣的假設下，淡入的公式如(式 2)所示。我們認爲有三個地方是需要 fading-in 和 fading-out，分別是（1）合成單元與後音有較強的連音現象（fading-out），（2）合成單元與前音有較強的連音現象（fading-in），以及（3）用切音縮短過音長（fading-out）。

$$x_{i(fading\_out)} = x_i * \frac{n-i+1}{n+1} \quad , \quad i = 1 \sim n \quad , \quad \cdots\cdots\cdots\text{(式 1)}$$

$$x_{i(fading\_in)} = x_i * \frac{i}{n+1} \quad , \quad i = 1 \sim n \quad 。 \quad \cdots\cdots\cdots\text{(式 2)}$$

5.實驗結果

5.1. 剖析器實驗數據

測試與訓練語料我們用中研院句結構樹庫 Version 2.1 來作測試與句法規則抽取，在中研院句結構樹庫中總共有 54,902 個中文句結構樹。但是在資料庫中有一些標示“%”符號的句子，它代表意義是句子不完整導致剖析器無法剖析，或是語意錯誤不合文法，我們必須事先將它剔除。所以剩下 51939 句來作爲訓練語料，從中抽取句法規則和統計機率；測試語料也是利用訓練語料來做內部測試（inside test），另外對於詞類標記我們也有做縮減，原因在於可以提升覆蓋率[15]，讓無法剖析出的句子減少。所以我們利用縮減詞類的語料來製作剖析器，同時也拿來作測試，詞

類簡化對應如附錄一。

針對剖析器的評估，我們有對樹結構(tree structure)、標記(Label)、括號(Bracket)作正確率的評估，其中對於評估標記與括號好壞的評估模型是 PARSEVAL[4]。我們採用如下的評估項目與公式[15]：

- 結構樹正確率 SP(Structure Precision)

$$SP=\frac{\#\text{correct parsing tree of testing data}}{\#\text{treebank parsing tree of testing data}}$$

- 詞組標記正確率 LP（Labeled Precision）

$$LP=\frac{\#\text{label correct constituents in parser's parse of testing data}}{\#\text{label constituents in parser's parse of testing data}}$$

- 詞組標記召回率 LR（Labeled Recall）

$$LR=\frac{\#\text{label correct constituents in parser's parse of testing data}}{\#\text{label constituents in treebank's parse of testing data}}$$

- 詞組標記效能評估 LF（Labeled F-measure）

$$LF=\frac{LP*LR*2}{LP+LR}$$

- 括號精確率 BP（Bracketed Precision）

$$BP=\frac{\#\text{bracket correct constituents in parser's parse of testing data}}{\#\text{bracket constituents in parser's parse of testing data}}$$

- 括號召回率 BR（Bracketed Recall）

$$BR=\frac{\#\text{bracket correct constituents in parser's parse of testing data}}{\#\text{bracket constituents in treebank's parse of testing data}}$$

- 括號效能評估 BF（Bracketed F-measure）

$$BF=\frac{BP*BR*2}{BP+BR}$$

實驗數據如表六所示。

表六 剖析器實驗結果(單位：%)

SP	LP	LR	LF	BP	BR	BF
38.78	61.96	64.31	63.11	70.04	72.80	71.39

雖然樹結構正確率不高，不過實驗數據中我們最主要的括號正確率與召回率分別爲 70.04%與 72.80%。這兩個數字很重要，因爲在我們求取韻律階層時用到的是語法樹(parse tree)的括號結構，詞性並未使用到。

5.2. 韻律階層求取實驗數據

接下來的部分是韻律階層的求取。我們會將分類的結果和原始人工標記的結果作比對，產生一混淆矩陣，如表七所示。接著用此混淆矩陣來計算三種停頓型態預估的正確率和召回率。

表七 – confusion matrix

True labels	Predicted labels		
	B_0	B_1	B_2
B_0	C_{00}	C_{01}	C_{02}
B_1	C_{10}	C_{11}	C_{12}
B_2	C_{20}	C_{21}	C_{22}

在表七中，$B_i\,(i = 0,1,2)$ 表示詞邊界(詞間)的停頓型態，B_0 表示該邊界在韻律詞內，不加停頓(no break)；而 B_1 表示該邊界爲韻律詞間，應加入小停頓(minor break)；B_2 表示該邊界爲韻律片語間，加入中停頓(major break)。其中對角線 C_{ii} (i =1 ,2 ,3)表示所預測的結果和標準答案一致的次數，而 C_{ij} (i ,j= 1 ,2 ,3 ; $i \neq j$)表示真實答案的標記爲 B_i 而程式標記成 B_j 的次數。

某一類別的召回率(Recall)的計算方式爲：

$$\mathrm{Rec_i} = C_{ii} \Big/ \sum_{j=0}^{2} C_{ij} \quad (i = 0, 1, 2)$$

以表七的混淆矩陣而言，我們可得到 B_0 (no break)的召回率

$$\mathrm{Rec_0} = C_{00} \Big/ (C_{00} + C_{01} + C_{02})$$

某一類別的精確率(Precision)的計算方式爲：

$$\mathrm{Pre_i} = C_{ii} \Big/ \sum_{j=0}^{2} C_{ji} \quad (i = 0, 1, 2)$$

以表七的混淆矩陣而言，我們所得到的 B_0 (no break)的正確率

$$\mathrm{Pre_0} = C_{00} \Big/ (C_{00} + C_{10} + C_{20}) \text{ 。}$$

全部類別的正確率(Accuracy)的計算方式爲：

$$\mathrm{Acc} = \sum_{i=0}^{2} C_{ii} \Big/ \sum_{i=0}^{2} \sum_{j=0}^{2} C_{ij}$$

以表七的混淆矩陣而言，我們所得到的正確率 Acc 爲

$$(C_{00} + C_{11} + C_{22}) / (C_{00} + C_{01} + C_{02} + C_{10} + C_{11} + C_{12} + C_{20} + C_{21} + C_{22})$$

接著是實驗的數據，表八是將所有的人工標記的語料合成一份，接下來將語料分成五等分，

四等分當成訓練語料來訓練 CART 的分類樹，而一等分爲測試語料，用來計算正確率。表八爲 CART 得到的結果。表九是以剖析樹的方式，對語料標記韻律階層。由於這個方法是規則式的，所以不用分訓練語料和測試語料。表九是九份語料合在一起標記所得到的結果。而表八和表九中的 Acc1 表示總正確率，Acc2 表示將 B1 和 B2 視爲同一類標記所得到的總正確率。

表八 - CART 結果

True labels	Predicted labels		
	B0	B1	B2
B0	30,434	3,198	126
B1	5,758	6,810	372
B2	635	1,381	514
Acc1：0.767	Pre0：0.826	Pre1：0.598	Pre2：0.508
Acc2：0.791	Rec0：0.902	Rec1：0.526	Rec2：0.203

表九 - Bottom-Up 結果

True labels	Predicted labels		
	B0	B1	B2
B0	156090	7384	4502
B1	54390	5471	5062
B2	7854	1828	2911
Acc1：0.669	Pre0：0.715	Pre1：0.372	Pre2：0.233
Acc2：0.698	Rec0：0.929	Rec1：0.084	Rec2：0.231

5.3. 文轉音系統實驗數據

在我們的實驗中，我們會分別進行自然度(naturalness)測試、偏好測試（preference testing）和可辨度(intelligibility)測試。我們利用 MOS（Mean Opinion Score）來評量我們合成語音的自然度。實驗中，測試方法爲播放語音，請一些人來當測試者，爲這些語音來評分數。分數分成五個等級：5 分：非常好（excellent）、4 分：好（good）、3 分：普通（fair）、2 分：差（poor）、1 分：極差（unsatisfactory）。偏好測試測試方法爲連續播放兩個合成語音，請聽者選取較佳的一個。偏好測試與自然度測試使用相同的文句，可辨度測試以句子爲單位，測試方法爲請聽者寫下合成語音的音或文字。

第一次的測試者爲 8 名本實驗室的成員，計有研究生 6 人和教師 2 人。在 MOS 測試和偏好測試中，我們是以段落（paragraphs）爲單位，測試總共有 20 段，每段長度介於 15 至 25 個字之間。而測試段落的來源爲新聞語料，每個段落選自不同主題，有政治、體育、影劇…等，偏好測試語自然度測試使用相同的段落。實驗目的方面，我們較感興趣於 Prosodic Word 間的小停頓是否要停頓？所以進行了自然度以及偏好測試，實驗結果如表十和表十一。兩者的結果相反，我們認爲聽者不易區別 Prosodic Word 間的小停頓。在可辨度測試中，得到 97.2%的正確率。

表十　自然度測試數據

測試者編號	有給停頓值(平均 MOS)	標準差	不給停頓值(平均 MOS)	標準差
M01	4.05	0.497	4	0.474
M02	3.15	0.963	3.15	0.792
M03	3.3	0.714	3.3	0.640
M04	4.385	0.504	4.67	0.181
M05	3.55	0.668	3.65	0.852
M06	3.9	0.538	4	0.632
M07	4.2	0.748	4	0.707
M08	2.95	0.804	2.95	0.804
平均	3.68		3.715	

表十一　偏好測試數據

測試者編號	有給停頓值（%）	不給停頓值（%）
M01	40	60
M02	50	50
M03	55	45
M04	50	50
M05	70	30
M06	70	30
M07	50	50
M08	55	45
平均	55	45

另一個測試是請本校的八位研究生作 MOS 測試，測試的句子共有四十句，測試的語音以 CART 和文法樹方式兩種方法所得到的韻律階層實際合成的語音各二十句。除了句子外並合成六篇文章，用 CART 和文法樹的韻律階層各合成三篇。測試前先定義 MOS 給分的準測[3]。我們定義：

5 分：難以分辨是合成語音還是自然語音。

4.5 分：清楚可辨度佳，在半小時內聽不累。

4 分：可以很清楚的了解在語音的意思，且沒有特別的斷詞錯誤。會有一或二個音節發音不清。

3 分：大部分能聽懂合成語音的意思，有明顯的錯誤。聽者無法連續聽十分鐘。

2 分：聽者無法聽出一些關鍵字，且此種的合成語音聽起來像是直接由音節連在一起。

1 分：聽起來像是機器人講話的聲音，並且聽不出語音所表達的意思。

測試結果如表十二所示。表十二表示在 MOS 測試中，八位聽者所給的平均分數。上表的數字表示在 MOS 的測試中，語者給分的平均值。用文法樹所產生的韻律階層分數比用 CART 所產生的韻律階層稍高，但相差不多。雖然文法樹在正確率的數據輸給 CART，但是實際合成語音的表現比 CART 稍好。

表十二　MOS 測試結果

		聽者 1	聽者 2	聽者 3	聽者 4	聽者 5	聽者 6	聽者 7	聽者 8	平均
句子	CART	3.45	3.65	3.55	3.35	3.5	3.9	4.2	4.1	3.71
	文法樹	3.48	3.9	3.45	3.95	3.45	4.18	3.88	4.25	3.82
文章	CART	4	4	4.66	4.66	4.66	4.66	3.66	5	4.42
	文法樹	4	4	4.33	5	5	4.83	4	5	4.52

6.結論與未來研究

在文轉音系統方面，目前已經得到一套發音清晰（可辨度測試 97.2%）的中文文轉音系統（線上系統網址：http://140.120.15.239/onlineTTS/cgitest.html）。未來還有需要處理的工作有（一）構詞部分的加強。例如：等看看。（二）連音變調需要的語意分析。例如：老李買好酒。（三）破音字的判別。例如：得 （ㄉㄜˊ）or（ㄉㄜ•）。（四）停頓型態的預測的正確率。（五）自動切音的工作。（六）錄製大量語料與罕見詞。在韻律階層預測上，實驗上韻律片語比韻律詞更爲重要，所以未來的工作朝向更準確更合理的韻律片語預測。

參考文獻

[1]Aho, A. V. and Ullman, J. D., "The Theory of Parsing, Translation, and Compiling ",1972, Vol. 1, Prentice-Hall, Englewood Cliffs, NJ.

[2]Breiman L, Friedman J. H., Olshen R. A., et al, "Classification and Regression Trees", Wadsworth, Inc, 1984.

[3]Bao H., Wang A., Lu S., "A Study of Evaluation Method for Synthetic Mandarin Speech", Proceedings of ISCSLP 2002, PP:383-386, Taipei, Taiwan.

[4]Charniak, E., "Treebank Grammars", In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 1031-1036. AAAI Press/MIT Press, 1996.

[5]Collins, M. J., "Head-Driven Statistical Models for Natural Language Parsing. ", Ph.D. Thesis, University of Pennsylvania, Philadelphia, 1999.

[6] Chu M., Peng H., Yang H. Y. and Chang E., " Selecting Non-Uniform Units from A Very Large Corpus for Concatenative Speech Synthesizer ", Proceedings of ICASSP 2001, IEEE, Volume 2, pp.785 - 788, Salt Lake City.

[7]Ney, H. "Dynamic Programming Parsing for Context-Free Recognition", IEEE Transactions on Signal Processing 1991, 39(2), 336-340.

[8] Hwang S. H. and Yei C. Y., "The Synthesis Unit Generation Algorithm for Mandarin TTS", Proceedings of ICASSP 2002, IEEE, Volume 1, pp. 457 - 460, Orlando, Florida.

[9]周福強, "以語料庫爲基礎之新一代中文文句翻語音合成技術", 國立臺灣大學電機工程學研究所博士論文, 1998 年。

[10]唐大任, "中文斷詞器之研究", 國立交通大學電信工程學所碩士論文, 2001 年。

[11]張唐瑜, "以大量詞彙作爲合成單元的中文文轉音系統", 國立中興大學資訊科學所碩士論文, 2005 年。

[12]許燦煌, "機率式中文剖析器之設計與實作", 國立中興大學資訊科學所碩士論文, 2005 年。

[13]潘能煌, "中文文轉音系統的韻律預估及其改進", 國立中興大學應用數學所博士論文, 2004 年。

[14]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所碩士論文, 2005 年。

[15]謝佑明, 楊敦淇, 陳克健, "語法規律的抽取及普遍化與精確化的研究", Proceedings of ROCLING XVI, 2004, pp.141-150。

[16]中央標準局委辦「中文資料分類處理分詞規範」計畫公聽會, 1998。

附錄一

附錄 詞類縮減對應表

中研院詞類標記	本剖析器所用詞類	說明
A	A	非謂形容詞
Caa	C	對等連接詞
Cab	C	連接詞
Cba	C	連接詞
Cbb	C	關聯連接詞
D	D	副詞
Da	D	數量副詞

DE	DE	的, 之, 得, 地
Dfa	D	動詞前程度副詞
Dfb	D	動詞後程度副詞
Di	D	時態標記
Dk	D	句副詞
FW	N	外文標記
I	I	感嘆詞
Na	N	普通名詞
Nb	N	專有名稱
Nc	N	地方詞
Ncd	Ncd	位置詞
Nd	N	時間詞
Nep	Ne	指代定詞
Neqa	Ne	數量定詞
Neqb	Ne	後置數量定詞
Nes	Ne	特指定詞
Neu	Ne	數詞定詞
Nf	N	量詞
Ng	Ng	後置詞
Nh	N	代名詞
P	P	介詞
SHI	V	是
T	T	語助詞
VA	V	動作不及物動詞
VAC	V	動作使動動詞
VB	V	動作類及物動詞
VC	V	動作及物動詞
VCL	V	動作接地方賓語動詞
VD	V	雙賓動詞
VE	V	動作句賓動詞
VF	V	動作謂賓動詞
VG	V	分類動詞
VH	V	狀態不及物動詞
VHC	V	狀態使動動詞
VI	V	狀態類及物動詞
VJ	V	狀態及物動詞
VK	V	狀態句賓動詞

VL	V	狀態謂賓動詞
V_2	V	有