

中英文文句翻語音系統中連音處理之研究

吳宗憲，陳昭宏，林超群

國立成功大學 資訊工程研究所

e-mail address: chwu@server2.iie.ncku.edu.tw

摘要

在本論文中，我們建構了一套高音質的中英文文句翻語音系統，本系統可以將中英文參雜的輸入文句經由系統處理後，轉為合成的語音訊號輸出，同時我們亦提出了一些在語音合成處理上的實用技術，這些技術能同時滿足在中文及英文合成處理的需求。為了能使系統合成出的語音更自然流利，除了在英文語音合成中採用包含連音資訊的雙音(Diphone)合成單元外，我們並且嘗試在中文連音(Coarticulation)處理方面能有些突破，我們可以將中文連音型態分成三大類來解決連音處理的問題，分別針對中文單音節(Monosyllable)間及英文雙音間的連接，就能量及基週軌跡進行平滑化，以期達到連音處理的目的。根據實驗評估結果顯示，在自然度方面，以 MOS 法評估出來的結果為 3.6 分，若以等級區分則約在良好及尚可之間。而在可辨度評估方面，正確率達 84.2%。

一、緒論

中文文句翻語音系統的研發，主要分佈在華人地區：中國大陸、香港、台灣。至今仍未見一套成熟、可商品化的系統，主要是自然度、流利度及破音字的處理上遇到瓶頸，此外，由於中文文章中也常含有英文文句，因此吾人希望藉由本論文的研究，開發一套發音清晰且具有高度自然語調的中英文語音合成系統。

就中文語音合成模組而言，首先我們錄製了 1432 個由女性發聲的國語單音(含四聲及輕聲)，同時對每一單音除了在錄音過程中要求音長標準化(Normalization)之外，事後又對音高、音量依聲調做標準化的處理。發音模組中我們採用由 CNET 實驗室提出的時域基週同步疊加法(Time Domain Pitch Synchronous Overlap and Add, TD-PSOLA)，在時域上對單音做音高、音長、音量的調整。音韻處理方面，我們均知中

文法鬆散任意組合之語句甚多，而具有意義的最小代表單位為『詞』，因此我們將中研院八萬詞庫擴充，將日常口語化、習慣化、通俗化之詞加入，成為系統適用的詞庫。並且將詞庫中之詞以四聲相接的情形及字數做分類，統計各個分類之音高、音長、音量的變化情形，並對需要例外處理之詞額外記錄。因此雖然我們採用規則條列法則(Rule Base)來決定如何調整音韻的變化，但我們細分音韻調整為『詞內音韻處理』及『整句音韻處理』兩個模組，先後分別調整音韻，當語音合成時根據輸入的文字，做整合性的音韻調整。在破音字處理上採用破音字分類法，並逐類用有限的規則解決。

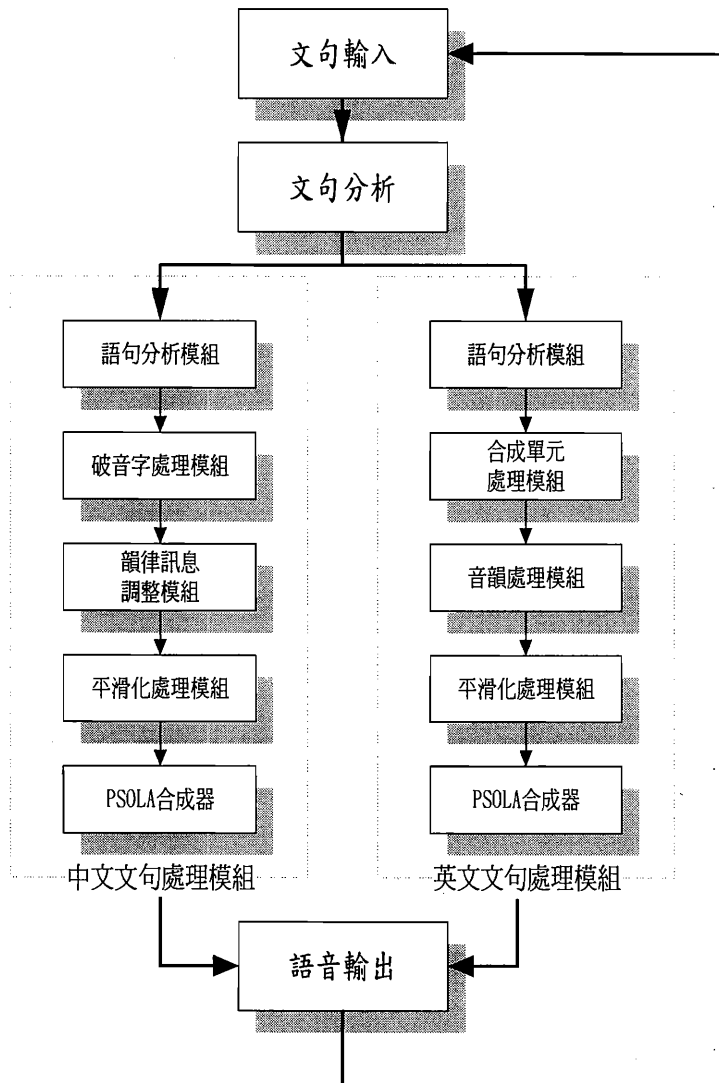
就英文語音合成模組而言，本系統採雙音(Diphone)為基本合成單元。首先我們製作出包含約十三萬英文字之電子字典，再根據電子字典找出 691 個包含所有雙音組合之平衡字(Balanced Word)，我們根據此預先錄製所有平衡字的音檔，再由音檔中分離出所要的雙音使成為單一合成單元。錄音至儲存的過程不做任何失真的壓縮處理；同時對每一平衡字要求在錄音過程中，音長、音高、音量均限制在一範圍內。發音模組我們仍採用時域基週同步疊加法，在時域上對各個組合之雙音作音高、音長、音量的調整。另外我們亦製作出自然拼音法的『字轉音規則庫』，處理創造字或不存在電子字典中之單字。詞綴庫的建立也是必要的，協助字轉音規則庫處理，如此就能完善的建立起英文文句翻語音系統。

二、系統架構

此套中英文文句翻語音系統主要可分為文句分析模組、中文文句處理模組、英文文句處理模組等幾個部份，其中，中文處理模組又可細分為語句分析模組、破音字處理模組、韻律訊息調整模組、平滑化處理模組及 PSOLA 合成器等五個基本模組。而英文處理模組又可細分為語句分析模組、合成單元處理模組、音韻處理模組、平滑化處理模組及 PSOLA 合成器等五個基本模組。透過這些模組的運作，可將任意輸入的中英文夾雜之文句，轉成語音輸出。基本架構圖如圖(一)所示，各模組功能簡介如下：

◆ 文句分析模組：

此模組之功能乃是分辨輸入之文句，何處是中文句(含標點符號及阿拉伯數字)的起始點及結束點；何處是英文文句及空白的起始點及結束點；如此分離出不同文句之屬性，即可將文句送入適當的處理模組加以處理。



圖(一) 系統基本架構圖

◆ 中文文句處理模組：

中文處理模組專責處理中文語句，包含中文字串、標點符號及阿拉伯數字的語音合成，其中又細分為：

1. 語句分析模組：

利用系統中文詞庫對輸入的中文語句進行構詞、斷詞的工作，找出詞邊界、詞屬性及斷句邊界，並配合發音字典取得各個字相對應的注音符號

與音檔編號。

2. 破音字處理模組：

使用破音字分類法及配合系統中文詞庫、讀音字典等來處理破音字，且修正相對應的注音符號與音檔編號。

3. 韻律訊息調整模組：

主要細分為詞內音韻處理及整句的音韻處理，因為使用我們特別處理過的詞庫包含了口語化之詞，況且詞庫是已知且有限的，所以詞內音韻處理我們事先可以做好，執行時只要針對句型判斷，做整句的音韻處理即可。

4. 平滑化處理模組：

針對可連音之中文字做連音處理，並儘量使之前後基週頻譜平滑化。

5. PSOLA 合成器：

採用時域基週同步疊加法，參考單音的基週資料，在時域上作音長、音高、音量的調整。

◆ 英文文句處理模組：

英文處理模組專責處理英文語句，包含英文字串、空白(Space)的語音合成，其中又細分為

1. 語句分析模組：

此模組主要是處理輸入之英文字串，配合電子字典、詞綴庫及字轉音規則庫，將英文字串轉換成相對應的音標及重音組合。

2. 合成單元處理模組：

此模組的功能乃是接收由『字轉音標及重音處理模組』產生之音標及重音字串，由此字串至音檔庫中找出相對應之音檔。

3. 音韻處理模組：

根據音標標定之重音及次重音調整整個字的音韻，在調整此音韻變化之前必須對各個雙音(Diphone)的基週軌跡(Pitch Contour)做標準化，再由音

韻規則做調整。

4. 平滑化處理模組：

由於雙音是由其他平衡詞之音檔中切出來的，所以前後雙音的基週可能並不一樣(或相差極大)，即使經過音韻處理模組處理之後仍有基週不平滑之情形，此模組將前後雙音正規化(Normalization)，如此便能輸出自然、平順的合成語音。

5. PSOLA 合成器：

跟中文文句處理模組一樣，我們同樣也採用時域基週同步疊加法，參考合成單元的基週資料，在時域上作英文字內各個雙音的音長、音高、音量的調整。

三、系統資料庫建立

◆ 中文資料庫建立

1. 語料庫：

在考慮實用性及高品質的原則下，本系統係採用成大資訊所中文實驗室錄製之音檔，此音檔採用 16 bits 音效卡及單向收音之單聲道麥克風錄製合成單元，且使錄音過程標準化(音長大約 0.27ms、音高為錄音者之正常發音音高、音量差異在 20%以下)。

2. 音韻規則庫：

詞內音韻規則庫：我們先將詞庫中的詞以四聲相接的情形及字數做分類，統計各個分類之音高、音長、音量變化情形並對需要例外處理的詞額外記錄。

整句音韻規則庫：整句音韻調整我們採用規則條列式(Rule Base)的作法，在執行時，根據輸入的文句做整體的調整。其規則條列如下：

- 換氣循環：模擬人的換氣動作，約每六字為一個循環，遇到詞則提前或延後換氣。換氣前音量、音高逐漸降低，換氣時略為停頓，換氣後音量、音高較為提高。

- 句首處理：每個句子的句首音高略為提高。
- 句尾處理：句尾要拉長音長、降低音高、能量遞減。
- 二聲接二聲：提高第一字的音高。
- 介詞處理：遇到常用的字，如：的、著、是、到、和、得等，縮短音長、降低音量，並拉長前一字的音長。
- 標點符號：在文章中遇有各類標點符號時加入長短不等的停頓，如表(一)。
- 停頓：詞之前、句尾亦加入適當的停頓。
- 句型：將文句分為直述句、疑問句、驚嘆句、命令句，分別調整其音高、音長、音量的整體走勢。

表(一) 標點符號與停頓時間關係表

標點符號	，	；	。	？	！
停頓時間	480ms	500ms	520ms	540ms	560ms

3. 基週資料庫：

基週標記位置的取得首先使用程式整批(Batch)作業，再由人工逐一檢查、校正，並連同每一個合成單元的音高、音量平均值(mean)、及音長、穩定區位置等特徵參數一起儲存，以增加系統效率。

◆ 英文資料庫建立

- 電子字典

一套包含大約十三萬字、約佔 2.8 Mega Bytes 儲存空間的英文電子字典。

- 雙音語料庫

在評估過各種合成單元的合成效果之後，我們採用了合成效果既不錯，所佔儲存空間又適當的『雙音』(Diphone)為本系統的基本合成單元。首先由電子字典中只考慮常用字，再根據音素(Phoneme)為組合雙音的基礎，找出

且錄製平衡字共 691 個，再由我們發展的切音工具程式在這些平衡字中標定出 1187 個雙音邊界。同樣的，我們仍然採用語錄製中文合成單元相同的語者及錄音規格，採用相同規格及語者是為了不使合成中英文句時，會有兩個不同人之聲音，而且如此音高、音量及唸法習慣皆相同，將有助於合成出高品質、高自然度之語音。

- 詞綴庫

我們另外從電子字典中分離出 383 個詞綴，其中包含前置詞綴 205 個後置詞綴 178 個，為處理英文時態字及複數字、組合字之用。

- 字轉音規則庫

當我們遇到創造字及組合字或不包含於此電子字典之中的單字時，若系統跳過不唸則會使使用者對本系統的效能大打折扣，[15] 對此種問題提出解決的方法。

- 基週資料庫

與中文模組相同採用時域基週同步疊加法(PSOLA)，不同的是用來調整英文字之字調，因此我們需要知道每個被合成單元聲音波型的基週標記位置(Pitch Mark Position)。所以我們利用程式標定再經過人工檢查，預先儲存每個被合成單元聲音波型的基週標記位置及音高、音量、音長及穩定區位置等特徵參數。

四、中文文句處理模組

- ◆ 中文文句處理流程

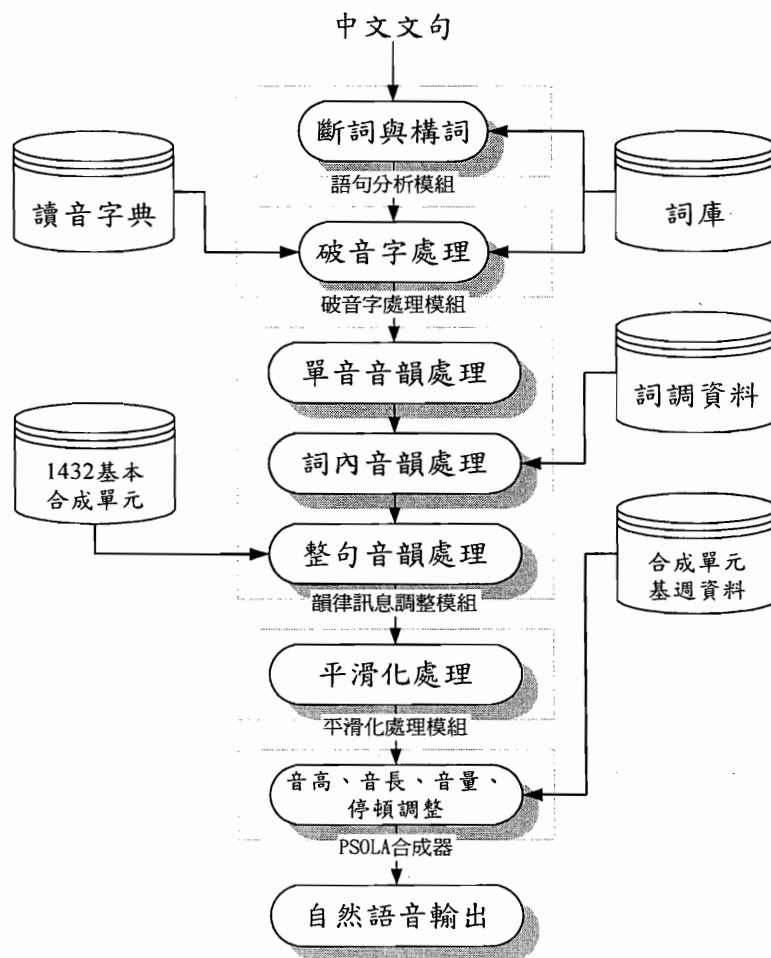
中文文句處理模組(如圖二)又可細分為語句分析模組、破音字處理模組、韻律訊息調整模組、平滑化處理模組及 PSOLA 合成器等五個基本模組。其中除平滑化處理模組將留在連音處理討論外，其餘模組將列在以下各節逐一說明。

- ◆ 語句分析模組

- 斷詞

基本上語句分析模組所處理的第一個步驟便是斷詞(word identification)，斷詞的目的在標示出輸入的中文文句中所包含之所有的詞，由於中文的文法極為鬆散且字與字之間的組合極多，所以同一中文字串可能會有許多不同的斷詞方式，也就是語意上及斷詞上的混淆(ambiguity)，因此中文的斷詞事很困難的事。

本論文採用的斷詞方式是利用統計的方式，將我們常用的詞事先建立一套詞庫，經過我們建立好我們適用的資料結構之後，如此就可正確且快速地找出所有包含於輸入中文句中的詞組合，再輸入構詞程序找出最佳之詞組合。



圖(二) 中文文句處理流程

● 構詞

雖然我們已經將原來引用的中研院詞庫修正注音成為口語化詞並且加入新詞，但想要包含所有的詞是不可能的，所以我們只能將非規律性組合的詞與出現頻率較高的詞蒐入詞庫中，具有規律性的部份則由構詞規則來處理。

本系統採用的規則如下：

- 長詞優先：以最少詞數涵蓋最多字數者優先。
- 左詞優先：兩個相鄰詞發生搶中間詞之時，以左邊的詞為優先。
- 詞長平均：例如『台南市政』可以斷成(台南、市政)或(台南市、政)兩種組合，我們採用兩詞長較相近的組合。
- 前綴詞處理：包括小、可、老、好等處理。
- 數量詞處理：利用詞性將數詞與量詞合併輸出。
- 重疊詞處理：如『快快樂樂』在斷詞後會斷成(快、快樂、樂)，處理之後將(快快樂樂)當成一個詞輸出。

◆ 破音字處理模組

為研究破音字的問題，我們總共統計出中文字共有 960 個字擁有兩個以上的讀音，而處理方面，我們採用本實驗室研發出之『破音字分類處理法』[13]來處理合成破音字語音時所遇到的問題。

◆ 音韻訊息調整模組

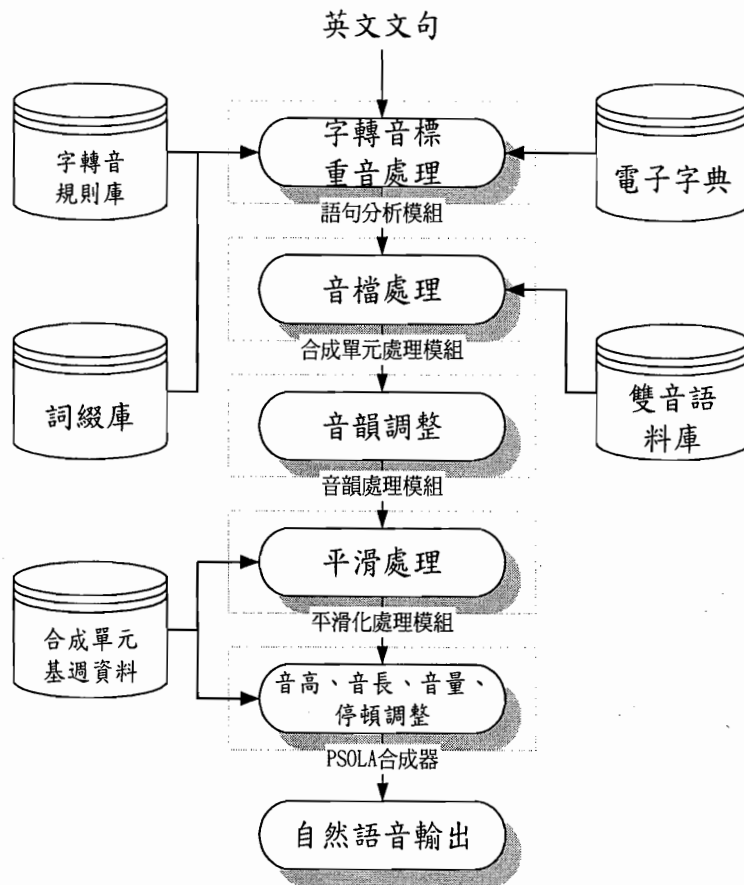
韻律訊息的調整本論文主要是採用音韻規則庫來作為調整韻律訊息的依據，並且使用 PSOLA 合成器調整音高、音長、音量，詞內音韻利用統計法建立規則庫，整句音韻規則庫採用條列法

五、英文文句處理模組

◆ 英文文句處理流程

英文處理模組(如圖三)可細分為語句分析模組、合成單元處理模組、音韻處理模組、平滑化處理模組及 PSOLA 合成器等五個基本模組。其中平滑化處理模

組將在下一節報告。

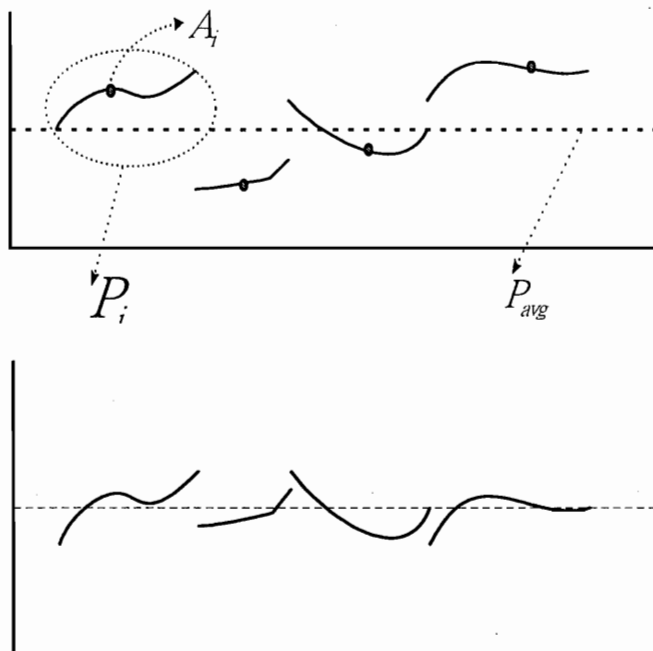


圖(三) 英文文句處理流程

◆ 語句分析模組

英文語句分析模組的處理較中文語句分析為簡單，因為英文的每一單字之間均有空白間隔存在。此模組的功能在處理輸入之英文字串，並且由字串中分離出英文單字逐一參考電子字典，找出此單字的音標、音節及重音標示，另外在此步驟中亦找出對應雙音(Diphone)檔案的索引(Index)及標示出那一雙音需做重音或輕音處理。

此外，為處理不存在於電子字典中之英文單字(包括組合字、創造字及時態字等)，我們解決的方法是參考系統中建立的字轉音規則庫及詞綴庫，若經由詞綴刪除之後可以在電子字典中找到單字，則以字典中之音標為主，若此詞綴處理



圖(四) 基週正規化

之後仍不在電子字典之中，則採用參考字轉音規則庫的自然拼音法將單字的音標、音節數、重音及輕音標示等創造出來。並仍需做雙音及音標的對應工作。

◆ 合成單元處理模組

此模組接收由語句分析模組處理過且輸出的音標及重音、輕音節標記，配合雙音索引檔由語料庫中找到雙音音檔並載入記憶體中。

◆ 音韻處理模組

● 正規化(Normalization)

由於各個雙音是由不同的平衡字語料中切割下來的，所以組合為另一單字發音時，會有前後音大小聲不同，基週變化太快的現象，雖然在錄音的過程中我們已對音量及基週大小做了一些限制，但仍有必要使被合成單元的能量、基週正規化(Normalization)，所以我們在音韻處理模組中首先處理的動作就是正規化，如圖(四)，過程簡述如下：

一、求各個 Diphone 之 Pitch Contour P_i 及平均 Pitch A_i

$$\text{二、求 Pitch 之總平均 } P_{avg} = \frac{\sum_{i=1}^n A_i}{n} \quad (5.1)$$

$$\text{三、Normalize } \forall P_i, P_i = P_i \times \frac{P_{avg}}{A_i} \quad (5.2)$$

● 音韻處理

本論文解決韻律訊息的問題是採用音韻規則的建立，音韻規則簡介如下：

一、音量規則：

1. 母音音量較子音為大。
2. 重音節的音量比非重音節大。
3. 有聲音節的音量通常隨著音高上升而增加。
4. 句子末了音節的音量通常較小，尤其是非重音節。

二、音長規則：

1. 重音節的音長比非重音節長。
2. 在停頓之前的音節中，增加其母音或子音的音長。
3. 在片語末端的音節中，增加其母音或子音的音長。
4. 在單字音節的音節段中，稍微縮短其中非最後音節段的音長。
5. 略為縮短多音節單字中的音節段的音長。
6. 縮短單字所有的子音中非字首的子音的音長。
7. 明顯增加被強調的單字中的母音的音長。

三、音高週期規則

1. 句首較句末有較高的音高。
2. 在無聲子音附近的音高比有聲子音附近的要高一點。
3. 重音節的地方有較高的音高。

六、連音處理

在中文的連音處理方面，由於本系統的中文音韻訊息調整模組已經將詞內音韻及整句音韻先後處理，且已有相當不錯的效果，所以我們只要針對單音(syllable)

與單音連接時做必要的處理即可。在比較過一些方法及試聽實作後的音質之後，我們決定應用中興大學應數所發展的連音處理模型[14]，加入我們發展的基週平滑法，改良成為新的連音處理模組來處理本系統合成語音的連音問題，連音型態分類及處理表如表(二)。

茲就重疊相加法分為能量平滑及基週平滑簡述如下：

表(二) 連音型態分類及處理表

前音節之最末音素	後音節之最前音素	連音型態	處理方式
出 彳 尸 日 卩 ㄗ ム ヲ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ム ㄗ ㄗ ㄗ ㄗ	清子音： ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ	停頓處理	加入 10ms 的靜音停頓
	濁子音： ㄗ ㄗ ㄗ ㄗ ㄗ	緊密連接	無停頓
	母音： ム ヲ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ ㄗ	重疊連接	重疊相加法

● 能量平滑：

$$s[i] = s_1[i] \times w_1[i] + s_2[i] \times w_2[i] \quad \text{for } i=0 \text{ to } M \quad (6.1)$$

$$\text{其中 } w_1[i] = \begin{cases} 1 & 0 \leq i \leq \alpha L \\ \frac{1}{(\alpha-1)L} i + \frac{1}{1-\alpha} & \alpha L \leq i \leq L \\ 0 & \alpha L \leq i \leq M \end{cases} \quad 0 < \alpha < 1 \quad (6.2)$$

$$w_2[i] = 1 - w_1[i] \quad (6.3)$$

S_1 為前一 syllable 之數位訊號， S_2 為後一 syllable 之數位訊號，且根據實驗， α 取 10/12 至 11/12 為佳。

● 基週平滑：

我們觀察到如果屬於重疊連接連音型態，則前後語音訊號的基週軌跡 (pitch contour) 大部分都是平滑的，而我們錄製的語料是以單音錄製，所以當前後音連接時基週軌跡絕大多數並不平滑，導致音調上的不自然，為此我們也提出一基週軌跡平滑化的方法，簡述如下：

為了要保持原基週軌跡的走勢，我們採用旋轉的觀念

$$\begin{bmatrix} \sin \phi & \cos \phi \\ \cos \phi & -\sin \phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix} \quad (6.4)$$

其中

$$\phi = \arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right) \quad (6.5)$$

$$a = \sqrt{((1-\alpha)L)^2 + p(L)^2} \quad (6.6)$$

$$b = \sqrt{((1-\alpha)L)^2 + A} \quad (6.7)$$

$$c = p(L) - A \quad (6.8)$$

L 為基週軌跡長度，p(L)代表軌跡上第 L 點之值，A 為前一音之基週軌跡的最後一點值，與後一音之基週軌跡的第一點值之平均， αL 為旋轉起始點。另外由於旋轉過後之基週軌跡，長度上可能會比我們需求的要長或短，因此我們再使用 up-sampling 或 down-sampling 的方法來達到我們的要求。

以上處理是以中文為例，在英文平滑化處理上亦可採用相同的方法，唯因為英文的合成單元-雙音比中文的合成單元要短的多，而且相連接的雙音之間的自相關性更高，若經由一般的方法處理很難會有效果，經由以上的平滑化處理流程之後，我們可以相當輕易地得到自然度較高的合成語音。

七、實驗結果與系統效能評估

一般評估文句翻語音系統的效能大多採用『自然度』(naturalness)與『可辨度』(intelligibility)對系統進行測試，我們相同也朝這兩方面對系統效能進行評估，但此種評估方式較為主觀，因此我們也以圖表示出語音波型、基週軌跡等數據作為客觀評估的依據。實驗對象除實驗室同仁(專業領域)八人之外，另也對非實驗室(一般人士)五人進行測試，評估項目說明如下：

◆ 主觀評估：

- 自然度評估(MOS 法)：

在自然度評估的方法中，測試樣本分別包括中文(詞、句子、短文)及英文單字的內部測試(inside test)及外部測試(outside test)，其評估方法採用平均鑑定分數(Mean Opinion Scores, MOS)法，將結果分為優良(excellent)、良好(good)、尚可(fair)、差(poor)及極差(unsatisfactory)五個等級，分別給予 5 至 1 不等的分數，測試結果如表(三)所示。

表(三) 自然度評估表

測試種類		數量	自然度(MOS)
中文	二字詞	200	4.0
	三字詞	200	4.0
	四字詞	200	3.8
	句子	100	3.6
	短文	10	3.6
英文	inside test(word)	100	3.1
	outside test(word)	100	2.9
平均			3.6

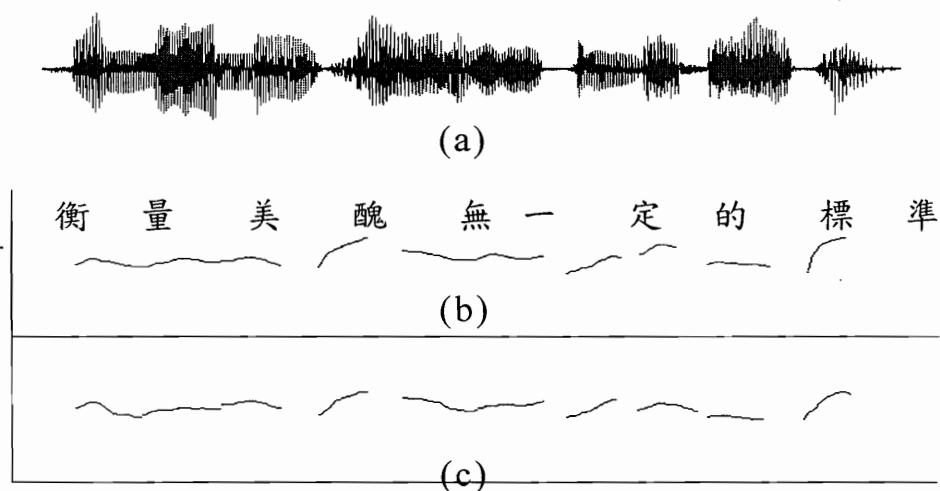
表(四) 可辨度評估表

測試種類		數量	可辨度
中文	單字詞	1432	88.9%
	二字詞	200	92.7%
	三字詞	200	90.4%
	四字詞	200	93.0%
	句子	100	93.5%
英文	inside test(word)	100	72.8%
	outside test(word)	100	57.9%
平均			84.2%

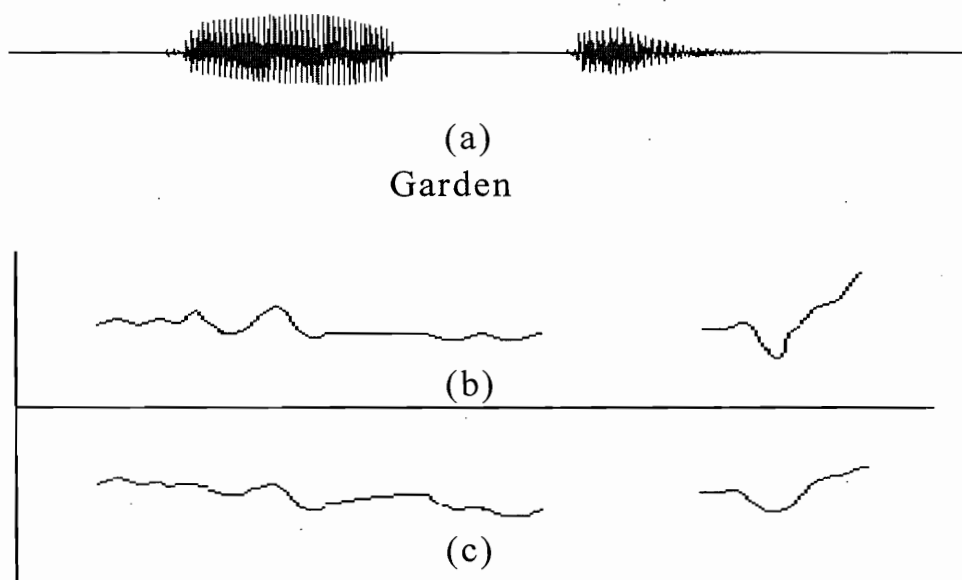
● 可辨度評估：

在可辨度的評估方面，測試樣本以系統合成的語音為主，分別包括中文(詞、句子)及英文單字的內部測試(inside test)及外部測試(outside test)，評估方法是我們由系統輸出合成的語音由受測人員將聽到的中文語音以聽寫的方式寫出標示的注音符號，聽到的英文語音默寫出單字拼法，若拼出的單字其唸法與原語音類似則以正確計算，最後統計正確的文句與受測者默寫結果之間的差異，可辨度評估結果如表(四)所示。

根據實驗評估結果顯示，在自然度方面，以 MOS 法評估出來的結果為 3.6 分，若以等級區分則約在良好及尚可之間。而在可辨度評估方面，平均有 84.2% 的正確率。綜合以上兩項的評估，雖然未能達到極高分數的評估結果，但已經能達到可以接受的程度了。



圖(五) (a)原始的語音波型(b)原始基週軌跡(c)合成語音之基週軌跡圖



圖(六) (a)原始的語音波型(b)原始基週軌跡(c)合成語音之基週軌跡圖

◆ 客觀評估：

我們以一句中文句及一個英文單字的語音波型圖為例，分別展示出原始的語音波形、基週軌跡及合成之後的基週軌跡，如圖(五)、(六)所示。

由基週軌跡比較圖中我們可一發現平滑化的處理已發生效果，他將原本基週軌跡相差過多的雙音合成單元都做了平滑化處理，使合成出的語音能有較自然的輸出。

八、結論與討論

在本論文中，我們建構了一套中英文文句翻語音系統，本系統可以將中英文參雜的文句(包括檔案)經由演算法的運算，配合語料庫及其他資料庫，轉為語音合成的聲音輸出，一般而言合成的語音還算自然流利，在時效上也可以達到即時處理的要求，值得一提的是，本系統提供了一些在語音訊號處理上的實用技術，這些技術均能在中文及英文處理上共同使用。經由實驗，本系統各方面的綜合表現均能令人滿意。

經由數次實驗之後，我們發現中文文句翻語音模組的整體表現較英文文句翻語音模組為佳，原因之一是因為本實驗室針對中文文句翻語音系統發展已行有多年，而英文的語音合成技術除了從一些國外的論文、期刊中可以得到一些有價值的資料外，有關雙音語料庫建立、英文語法規則等基本知識我們均可以說是從頭開始建立，另外，非母語的限制更是從錄製語料庫開始便困擾著我們，以下列出三大方向提供未來參考改進之：

◆ 雙音語料庫：

就中文語音合成而言，我們的單音語料是相當標準的，而且以此語料合成的效果也有口皆碑，但相較中文合成而言我們的英文合成就成熟許多，我們認為最主要的原因是出在英文非我們的母語，首先在製作平衡字時，雖然我們已經儘量從常用字著手，但仍有許多字的讀音對語料錄製者而言是非常難錄製的，再者，我們在實驗中發現，有些字無論如何處理，合成的效果仍有瑕疵，我們判斷可能是有一些平衡字含有『破音』成份(如：sk、kr、scr等)，而這些『破音』成份若被切成雙音且被合成於另一不含『破音』成

份的單字時，合成語音的品質就會大打折扣。最簡單的解決方法是不採用這些含有『破音』成份的雙音，如果儲存空間允許，甚至可以將固定成形字首及字尾連同『破音』成份的雙音均由三音(Triphone)或多音(Polyphone)取代，如此將可大大提昇合成語音的品質。

◆ 音韻資料庫：

雖然我們已根據我們提出的音韻處理規則分別對重音及次重音加以處理，但效果仍有待改善，原因乃是除了具有重音、或次重音標記的音標符號我們要對其特別處理之外，對於重音或次重音標記附近的雙音合成單元要如何處理，我們至今仍無法提出一種有效的規則對其處理，對此我們構想未來應朝向類似連續語音資料庫中詞音韻歸納的方法，對應出合成單元在平衡字中的相對位置，並擷取其音長、音高、音量、停頓等韻律訊息特徵，加以參考之，另外，整句的韻律訊息調整亦可由此方法實現。

◆ 整句連音處理：

本論文的研究之中，為降低系統複雜度，因此我們並未對整句的連音做適當的處理，簡單的說，我們合成出的英文語音中單字與單字間應有的連音資訊都被我們忽略了(例如：make up on....、would you mind...等)，未來應統計出那些音標的組合會產生連音，會產生『變調連音』或『不變調連音』還是連音之後省略後面母音或變化為輕音的各個類型，分別加以處理之。

九、誌謝

感謝國科會經費補助(計畫編號 NSC85-2622-E-006-003)，使得本論文能順利完成。

參考文獻

- [1] F.J. Charpentier and M.G. Stella, "Diphone Synthesis Using an Overlap-Add Technique For Speech Waveforms Conca-tenation", Intern. conf. on ASSP, ICASSP-86, pp. 2015-2019, 1986.
- [2] J. Allen, "Synthesis of speech from unrestriced text", Proc, IEEE, vol. 64, pp. 422-

433, 1976.

- [3] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signals", p398-p403, 1987.
- [4] L. S. Lee, C. Y. Tseng and M. Ouh-Young, "The Synthesis Rules in a Chinese Text-to-Speech System", IEEE Trans. On Acoust Speech, and Signal Processing, vol. 37, No. 9, pp. 1309-1319, September 1989.
- [5] E.Moulines, F. Charpentier, "Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis using Diphones", Speech Comm, Vol. 9, pp.453-467, 1990.
- [6] 熊明德, "中文語音規則合成之研究", 淡江資工所碩士論文, 民國 80 年
- [7] 陳克建、陳正佳、林隆基, "中文語句分析的研究—斷詞與構詞", 技術報告, TR-86-004, 中央研究院, 1986
- [8] H. Klatt, "Review of Text-to-Speech Conversion for English", J. Acoust. Soc. Amer., vol. 82, no. 3, pp. 737-793, Sep. 1987.
- [9] Lin-Shan. Lee, Chiu-Yu. Tseng, and Ching-jiang. Hsieh, "Improved Tone Concatenation Rules in a Formant-Based Chinese Text-to-Speech System" IEEE Trans., Speech, Audio Processing, vol. 1, no. 3, pp. 287-294, Jul. 1993.
- [10] 盧承周、余秀敏、劉繼謚, "中文發音辭典的建立和破音字的發音辨別系統", 電信研究雙季刊, pp. 287-295, 19 卷 13 期, 1989.
- [11] Bigorgne, et al, "Multilingual PSOLA Text-to-Speech System", in Proc. ICASSP93, pp. II-187-II-190.
- [12] Ngor-Chi Chan, Chorkin Chan, "Prosodic Rules for Connected Mandarin Synthesis", Journal of Information Science and Engineering, pp. 261-281, 1992.
- [13] 蔡祈岩, "文句翻語音系統中破音字及音韻處理之研究", 成大資工所碩士論文, 民國 84 年
- [14] 陳志祥, "國語連續語音連音型態之初步研究", 中興大學應數所碩士論文, 民國

83 年

- [15] 林超群,“中英文文句翻語音系統中連音處理之研究”,成大資工所碩士論文,民國 85 年