基於 HNM 及半音節接合之英語歌聲合成系統

An English Singing Voice Synthesis System Based on HNM and Concatenation of Demisyllables

古鴻炎
Hung-Yan Gu
國立台灣科技大學
資訊工程系
guhy@mail.ntust.edu.tw

梁弘學
Hung-Hsueh Liang
國立台灣科技大學
資訊工程所
M9515052@mail.ntust.edu.tw

摘要

本論文研究、製作了一個英語歌聲合成的系統。由於語料中只有 1,389 個不同的音節，因此我們提出一個音節單元的建造流程，用以解決音節單元不足的問題。信號模型採用的是諧波加噪音模型(harmonic-plus-noise model, HNM)，亦即音節單元的形成是在 HNM 參數的層次。此外，我們應用國語的 ANN 模型來產生英語音節的抖音參數，製作動態滿度設定，以合成出較為自然的歌聲信號。進行主觀的聽測實驗的結果是，以半音節作接合和優先選擇音節單元，兩種方式的合成歌聲幾乎無差異；另外，與一套市面販售的軟體作比較，兩者的評分亦很接近。

關鍵詞：歌聲合成、半音節、諧波加雜音模型。

1. 前言

在國語歌聲合成的研究方面，過去我們已有了不錯的成果[1, 2, 3]。最近我們參與了本校智慧型機器人中心的計畫『機器人劇場』，為了因應未來有國際性表演的需求，因此我們著手以先前國語歌聲合成的成果為基礎，來研究英語歌聲合成的方法及建立系統。關於英語歌聲的合成，我們曾找到幾篇文獻[4, 5]，但是考慮到實作上的問題，我們仍決定採取自己的方向(approach)。

英語與國語最大的一個不同點是，英語的音節結構為:子音(Consonant) + 母音(Vowel) + 子音(Consonant)，比國語的一般音節結構: 聲母(C)+韻母(V)，多了母音後的子音。另外，相對於國語只有 412 個不同的音節，英語的不同音節的數量則顯得很龐大，依據 CMU 詞典[6]來統計，至少有 14,906 個不同音節，並且隨著新 word 的出現，音節數量也可能會再增加。所以要錄製英語的所有音節在實作上是不容易的，因此我們將研究如何以有限的語料來作英語歌聲的合成。

研究歌聲合成的一個準備工作是，收集一個可供擷取出合成單元的語料庫。我們收集英語語料來源是，從『空中英語教室』的 MP3 錄音檔中摘錄出同一位女性所發音的 35 篇文章，然後使用 WaveSurfer 軟體[7] 來對每一篇裡的音節作逐一的標記，標記出音節的邊界以及音節的拼音。之後我們將所有的音節切割出來存成獨立的音檔，總共約有 10,000 個音節，但是去除重複的之後，只剩下 1,389 種不同發音的音節。爲了程式處理上的方便，我們統一採取以 CMU 音標代碼來命名音檔的檔名，以方便程式作音標特性(如有聲、無聲)的剖析判斷，表 1 裡列出 CMU 大學制訂的音標代碼。

表 1 CMU 音標代碼

母音

音標	代碼	音標	代碼	音標	代碼	音標	代碼
i	iy	ɑ	aa	u	uw	ɑ ɪ	ay
ɪ	ih	ʌ	ah	ʊ	uh	ɑ ʊ	aw
e	ey	ə	ax	ə˞	er	ɔ ɪ	oy
ɛ	eh	ɔ	ao	ɜ˞	er	l.	ah l
a	ae	o	ow			n.	ah n

子音

音標	代碼	音標	代碼	音標	代碼	音標	代碼
p	p	f	f	ʃ	sh	ŋ	ng
b	b	v	v	ʒ	zh	l	l
t	t	s	s	ʧ	ch	r	r
d	d	z	z	ʤ	jh	j	y
k	k	Θ	th	m	m	h	hh
g	g	d	dh	n	n	w	w

在信號波形產生方面，我們採用了諧波加雜音模型(harmonic-plus-noise model, HNM) [8]，這是因為我們從過去的經驗得知，以 HNM 合成出的聲音信號相當清晰。在分析階段，要對收集的英語音節信號作頻譜分析，並且記錄各音框的諧波參數和噪音參數，如此在作歌聲合成時，就可以依據音高、諧波和雜音參數，來產生出對應的信號波形。

我們系統的製作，分成分析和合成兩個階段，分析階段的處理流程如圖 1 所示，首先將收集的 1,389 個英語音節作 ASR (attack, sustain, release)的標記[9]，根據 ASR 的資訊，就可以計算出音節核心母音的中心點，也就是半音節的切點。接著我們將這些音節作 HNM 分析，以求得各音框的諧波及噪音參數，然後存入各自的參數檔案。

合成階段的處理流程如圖 2 所示，首先是讀入、分析歌譜檔案；分析後讀取對應音節的 HNM 參數；由於可能找不到所需的音節單元，此時就以音節前或後串接子音，或以半音節接合的方式來解決；然後依據歌譜規定的音高，去調整各音節的 HNM 參數；接著使用抖音 ANN 模型產生出

抖音參數；再對一些音樂性參數作調整；最後合成出歌聲信號。

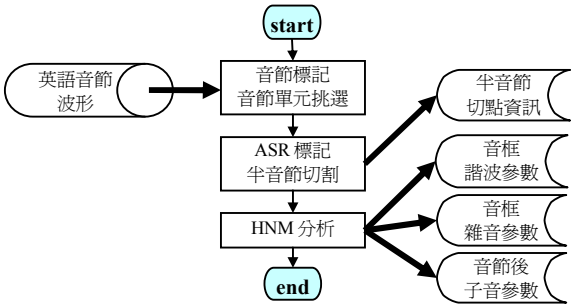

圖 1 分析階段之處理流程

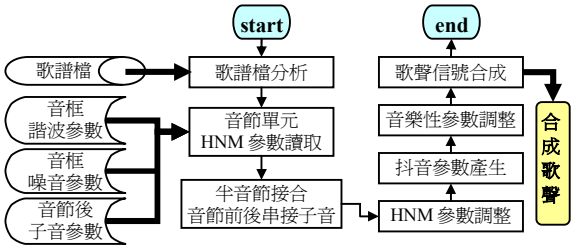

圖 2 合成階段之處理流程

2. HNM 及音節單元 HNM 參數形成

我們以先前研究的國語歌聲合成系統[1, 3]為基礎，對它作修改，以用來合成英語歌聲。由於先前的國語歌聲合成是以 HNM 作為信號模型，所以我們將沿用 HNM 模型，不過對於音節單元不足的問題，我們必需在 HNM 之上，考慮不存在的音節單元，要如何將它的 HNM 參數組合出來，然後才能把整個音節的 HNM 參數帶入信號合成程式，以合成出各個英語音節的歌聲信號。HNM 參數指的是一序列音框的諧波和雜音參數，一個音框的諧波參數包括各個諧波的頻率、振幅、和相位數值，而雜音參數是 30 個倒頻譜(cepstrum)係數[10, 11]。

2.1. HNM 簡介

HNM 翻譯成諧波加雜音模型，從名稱來看就是要把聲音信號 $s(t)$ 分解成諧波 $h(t)$ 及噪音 $n(t)$ 兩部分，如公式(1)所示:

$$s(t)=h(t)+n(t) \tag{1}$$

HNM 提供了一個最大有聲頻率(maximum voiced frequency, MVF)的訂定方法，且以 MVF 作為分界點，對於頻率低於 MVF 的頻帶，就產生出諧波信號，而對於頻率高於 MVF 的頻帶，就產生出雜音音信號，然後再將這兩種信號加起來作為 HNM 的合成信號。

產生諧波信號時，以頻率值有倍數關係的多個弦波來作合成，如公式(2)所示：

$$h(t)=\sum_{k=1}^{K(t)} a_k(t)\cos\left(\phi_k(t)\right) \tag{2}$$

其中 $a_k(t)$ 及 $\phi_k(t)$ 表示時間 t 時，第 k 個弦波的振幅及相位，$K(t)$ 則表示時間 t 時諧波的數目。當產生雜音信號時，則是以間隔固定為 100Hz 的弦波信號來作為信號成分，至於弦波的振幅，則依倒頻譜(cepstrum)係數轉換出的頻譜包絡來決定。對於無聲音素(如/p/)的信號合成，可把 MVF 設為 0，即整個頻譜都視為雜音部分。較詳細的作法可參考原始文獻[8]或我們先前的論文[1,2]。

2.2. 音節 HNM 參數形成

對於不同種類的子音信號，必需採取不同的 HNM 參數分析方式，所以我們先對英語音節結尾的子音作分類，在此依**有聲**、**無聲**，和是否為**爆破音**而分成四類，如表 1 所示。對於以**有聲**非**爆破音**類子音結尾的音節，就把整個音節一起作分析並儲存整個音節的 HNM 參數檔，這是因為這類子音與母音之間並沒有明顯的區隔。而對於另三類的子音，則是分析後，將結尾子音部分的 HNM 參數另外存檔，以便往後用來處理音節單元不足之問題。

表 2 音節尾部子音之分類

有聲爆破子音	b、d、g
有聲非爆破子音	m、n、l、r、v
無聲爆破子音	p、t、k
無聲非爆破子音	s、z、f 等摩擦音

2.2.1 音節後串接子音

在 HNM 參數分析時，某些音節後子音的 HNM 參數會另外存檔。以音節/g aa d/為例，經過分析後會得到兩個參數檔，一個是/g aa/部分的參數檔，另一個則是/d/部分的參數檔。如此在合成階段，假設找不到音節/n aa d/的 HNM 參數檔，我們就可尋找一個以/n aa/開頭的音節，取出它的/n aa/部分所分析出的 HNM 參數，來和/d/的 HNM 參數作串接，而形成所需的/n aa d/音節的 HNM 參數。其實所收集的語料中，音節後子音被省略的情形非常多，所以也可以此方式來解決。

2.2.2 前半、後半音節接合

上述音節後串接子音的作法，主要是針對音節結尾的子音有另外存 HNM 參數檔的情況。如果音節尾部是有聲非爆破音的子音(如/m/)時，則不能以此作法來解決，此時我們就改採半音節(demisyllable)接合的作法。為了實施半音節接合，在分析階段裡必需標記各音節的 ASR 位置，以便計算各音節的半音節切點位置(即 sustain 部分的中間點)，一個 ASR 標記的例子如圖 3 所示。此外切點前後各三個音框的 MFCC (mel frequency cepstrum coefficients)係數[10, 11]也必需儲存起來。

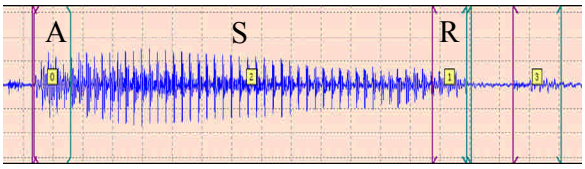

圖 3 音節/g aa d/之 ASR 標記例子

當準備進行半音節接合時，首先是讀取所有的前半和後半音節的候選音節的 MFCC 係數檔，接著依據公式(3)來計算每一組前、後半音節在接合點的頻譜差異[12]，

$$C\left(s_{pre}, s_{post}\right)=\sum_{n=1}^{3} d\left({s_{pre}}^{n}, {s_{post}}^{n+3}\right) \qquad (3)$$

其中 ${s_{pre}}^{n}$ 表示前半音節之候選音節 s_{pre} 的切點附近第 n 個音框的 MFCC 係數向量，而 ${s_{post}}^{n+3}$ 表示後半音節之候選音節 s_{post} 的切點附近第 n+3 個音框的 MFCC 係數向量，而 $d(\bullet,\bullet)$ 表示兩個音框的 MFCC 係數的歐基里德距離量測。圖 4 說明一個候選音節在切點前後共有 6 個音框的 MFCC 係數向量。

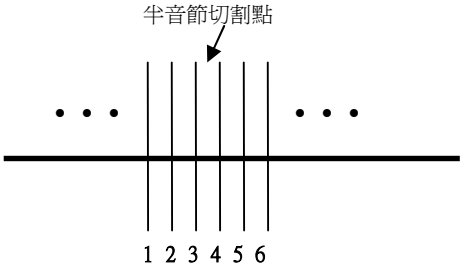

圖 4 候選音節在切點附近之 6 個音框

依據公式(3) 的計算結果，我們可選出頻譜差異最小的一組前、後半音節的候選音節，然後去讀取這兩個候選音節的 HNM 參數檔案。假設要接合出音節/s ey m/的 HNM 參數，而頻譜差異最小的一組前、後半音節的候選音節，分別是/s ey/和/g ey m/。因此，我們就從前半音節的候選音節/s ey/取出切點前的 HNM 參數，而從後半音節的候選音節/g ey m/取出切點後的 HNM 參數，再把這兩部分的參數串接成所需音節/s ey m/的 HNM 參數。

接合時，因為前半後、半音節來源不同，音量也不一致，可能造成在接合點附近的振幅會有很大的落差，使得合成出的信號變得不平順。所以我們要對前半及後半音節做音量的匹配，其作法是取接合點前三和後三個音框來作短時能量的比較和調整。

2.2.3 音節前串接子音

另一種情況是缺少音節前子音的情形，一個例子如找不到音節/sh ay/的 HNM 參數檔，這時就可從另外儲存的子音參數檔中搜尋/sh/的 HNM 參數檔，若有找到，就可以/sh/的參數串接上/ay/的參數來形成音節/sh ay/的 HNM 參數；若沒有找到，此時就只好以類似發音的子音來作替換。前人的論文曾提到[13]，歌聲中較重要的是母音的表現，子音相對地較不重要， 所以替換子音應是可以忍收的。

3. 歌譜與歌詞處理

合成階段的第一個步驟是輸入歌譜，然後依據歌譜中的資訊，讀取、或組合出對應音節的 HNM 參數，接著配合歌譜中音高和音長的資訊，來合成出英語歌聲的信號。

3.1. 歌譜檔輸入

先前曾提到研究動機是要因應國際性表演的需求，其實表演的名稱是**認譜唱歌**，就是要透過影像處理來辨識樂譜，然後由我們的軟體去合成出歌聲信號及播放。並且樂譜辨識程式和歌聲合成程式之間，需要經由歌譜檔來作溝通，所以歌譜檔的格式必須先作定義。

我們定義了一個可以由樂譜辨識程式產生或由使用者鍵入的文字式歌譜格式(一)。此格式的第一行記錄《歌曲名稱, 速度, 滿度設定》，可以從圖 5(a)看到，歌曲名稱為"eyes on me"，速度為一分鐘 96 拍，%0.85 %0.8 %0.8 則表示合成信號時的滿度設定。第二行開始記錄歌詞和音符的訊息，採取的格式為《編號, 英文 word, 音符, 音長》。

但是一個英文 word 可能會包含多個音節，例如 ever 就包含了/eh/和/v er/兩個音節，所以爲了方便後續的處理，我們對歌譜新增一些表示的格式，以『^』代表前面處理的 word 仍有音節尚未處理，這可以從圖 5(b)中編號 44 和 45 的列裡看到。另外關於轉音的表示，我們在音符和音長的欄位以減號『-』來連接轉音的音符，圖 5(c)中編號 52 的列即是一個轉音的例子。休止符的表示，採取的格式是《0, 0, 音長》，圖 5(c)中編號 47 的列表示要休息半拍。

```
eyes_on_me 96 %0.85 %0.8 %0.8
01 my A2 0.5
02 last D3 1
03 night E3 1
04 here F3# 1
05 for A3 0.5
06 you F3# 4.5
```

(a)第一行之格式

```
41 oh A2 0.5
42 did D3 1
43 you E3 1.5
44 ever F3# 0.5
45 ^ G3 0.5
46 know A3 2.5
```

(b)一 word 有多個音節

```
47 0 0 0.5
48 that A3 0.5
49 i G3 0.5
50 had D3 0.5
51 mine F3# 3
52 on E3-D3 0.5-0.5
53 you D3 4
```

(c) 轉音的表示

圖 5 文字式歌譜格式(一)

3.2. 歌詞轉音節單元

爲了方便輸入歌譜檔，使用者只需輸入英文 word 作為歌詞，而不需另外去查詢、輸入歌詞的音標，而由程式透過查詢 CMU 電子詞典，來將歌詞 word 自動轉成音標代碼序列。

可是我們的合成單元為音節，所以得到歌詞 word 的音標代碼序列仍然是不夠的，因為會面臨音標代碼序列切割成音節的問題: 當一個 word 包含多個音節時，如何將這些音節的邊界界定出來? 例如含有雙音節的音標代碼序列/ae f t er/，要把它分割成/ae f/和/t er/兩音節，而不是/ae/和/f t er/。

我們的作法是，先尋找母音符號，來得知音節的數量及相鄰母音之間的子音個數，然後對於相鄰母音之間的子音再作考慮。將所有位於相鄰母音之間的各種子音排列的情形列出並作觀察，發現當子音的數量為 1 個時，幾乎都是歸屬於後面的母音；當數量為 2 個時，幾乎都是平分給前後兩個母音；至於例外情況及多於兩個子音的組合，經觀察而歸納出一些子音群分割的規則，例如/ah b r ah p t l iy/要分割成/ah/、/b r ah p t/及/l iy/，而/s ao f t b ao l/要分割成/s ao f t/及/b ao l/。爲了寫作程式來處理，我們建造了一個分割規則表，來讓歌詞轉音節單元的程式據以作音節的分割。

根據此規則表，我們的程式就可將歌譜格式(一)轉換成歌聲合成程式較方便處理的格式(二)，一個歌譜轉換前後的對照例子如圖 6 所示，由圖 6(b)可以看到，歌詞 did 的音標/d ih d/會被表示成 IH-D-D，也就是把母音放到前面以方便後續的處理；此外，我們用『^』替換編號 4 來表示此音節為某個歌詞 word 的一部份，一直到下個數字編號出現，此例為編號 5，編號 5 和之前的『^』代表同一個歌詞 word 的組成音節。另外，圖 6(a)的音長單位為拍數，而圖 6(b)的音長單位為秒，這可由圖 6 (a)的拍數和歌譜速度換算得到。

```
01 oh A2 0.5
02 did D3 1
03 you E3 1.5
04 ever F3# 0.5
05 ^ G3 0.5
06 know A3 2.5
07 0 0 0.5
08 that A3 0.5
09 i G3 0.5
10 had D3 0.5
11 mine F3# 3
12 on E3-D3 0.5-0.5
13 you D3 4
```

(a) 歌譜格式(一)

```
1 OW-O-O A2 0.31
2 IH-D-D D3 0.62
3 UW-Y-O E3 0.93
^ EH-O-O F3# 0.31
5 ER-V-O G3 0.31
6 OW-N-O A3 1.56
7 0 0 0.31
8 AE-DH-T A3 0.31
9 AY-O-O G3 0.31
10 AE-HH-D D3 0.31
11 AY-M-N F3# 1.87
12 AA-O-N E3-D3 0.31-0.31
13 UW-Y-O D3 2.5
```

(b) 歌譜格式(二)

圖 6 歌譜格式轉換

3.3. 音節單元不足之處理

我們在 2.2 節裡提到，解決音節單元不足的方法是，以音節前、後串接子音和半音節接合的方式來組合出所需音節的 HNM 參數，但是如果仍然無法組合出所需的音節時，我們就會退一步以相近發音的音節來取代，這樣的音節組合處理之流程就如圖 7 所示。

在此我們先定義"基本音節"之觀念，它代表音節的主體，例如音節 CVC 結構裡的 CV 部分就是基本音節，而基本音節的 HNM 參數所存的檔

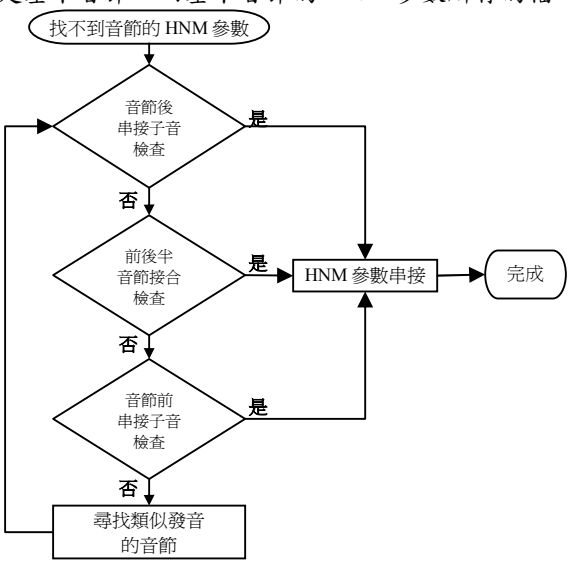

圖 7 音節單元組合處理之流程

案，稱為基本檔；然而有聲非爆破音類的子音因為是與母音一起作 HNM 分析和存檔，所以此類子音的基本音節仍然維持為 CVC_n，這裡的 C_n 表示有聲非爆破音類的子音(如/m/)。

圖 7 的流程裡，首先是檢查基本檔是否存在，如果存在，就可取出尾部子音的 HNM 參數來進行串接，如果不存在，就往下作半音節接合之檢查。半音節接合的檢查是，先嘗試找出所有可以成為基本音節的前半和後半音節的候選音節，如果存在候選音節，接著依據公式(3)找出頻譜差異最小的一組候選音節作為前半和後半音節的來源音節，再對前半、後半來源音節取出前半、後半的 HNM 參數，串接成所需音節的 HNM 參數。當上述兩個步驟仍無法組合出所需音節的 HNM 參數時，此時就檢查是否可用音節前串接子音的方式來解決。如果仍然無法解決，我們就以刪減子音的方式來作音節代替，目前試驗合成了 7 首歌，共 357 個音節，刪減子音的情況共發生了 10 次。

4. 系統製作與評估

根據前人研究國語歌聲合成的經驗[1, 2, 3]，可知要合成出自然的歌聲，必需考慮到下拍點、滿度、音量、和抖音等參數的設定，這幾個參數的設定將在以下的子節裡說明。整個系統建造好之後，我們在筆記本電腦(CPU: Intel T5600 1.83GHz)上測試它的速度，結果是平均每合成 1 秒的歌聲信號，需花費 0.51 秒的 CPU 時間，所以它可作到即時的歌聲合成處理。

4.1. 音樂性參數調整

當把合成的音節信號作串接時，必須考慮音節前端子音的影響，因為不同子音的長度長短不一，若只是將每個音節信號的起始位置對齊音符的時間起點，則子音較長的音節會造成歌聲的主音，也就是音節核心母音的部份延後出現，這樣會產生拖拍(拍子不準)的現象，使得歌聲的節奏變得忽快忽慢。因此我們依據前人的作法[1,2]，將音節的母音起始位置對齊音符的起點，而讓子音的部份提前唱。

音符的滿度是控制歌聲表情的一個因素，滿度指的是一個音符實際所唱的長度與定義上長度的時間比例，如果每個音符都唱滿了，會使得歌聲聽起來很黏膩；但是若滿度不夠高，則會使每個音變得過於獨立，而不能展現出歌聲表情。基本上在此也採用前人的研究成果[1,2]，但是使用固定的滿度設定，會使得同一個歌詞 word 裡的音節之間變得過於獨立(顯得無關係)，所以我們以動態方式增加同一個 word 裡音節的滿度，以提升 word 裡音節的相關性。

唱歌時由於各音節母音發聲嘴型大小的不同，信號振幅大小也會有所不同，因此，合成出一個音節的信號後，必需再調整該音節的振幅值。在此我們採用規則式的處理方式，規則參考自前人的論文[14]，但是配合英語的母音作了一些修改。首先將音節的振幅調整到一個定值，再依音節的母音(也就是主要嘴型)來調整音量，規則如下：(1)母音為/aa/、/ah/時，該音節音量不變。(2)母音為/ae/、/eh/、/ey/時，音節音量下降 1dB。(3)母音為/ow/、/oy/、/aw/、/ao/時，音節音量下降 2dB。(4)母音為/ih/、/iy/時，音節音量下降 4dB。(5)母音為/er/、/ax/時，音節音量下降 5dB。(6)其餘情況音量降低 3dB。

4.2. 抖音產生

抖音參數包括音位軌跡(intonation)、抖音範圍(extent)和抖音頻率(rate) 等三項，我們可依據這三個參數來產生出瞬間頻率，其公式為：

$$f(t) = VD(t) + VE(t) \cdot \cos\left(\phi(t)\right) \qquad (4)$$

其中 $f(t)$表示瞬間頻率，$VD(t)$表示音位軌跡，$VE(t)$表示抖音範圍曲線，而 $VR(t)$表示抖音頻率曲線，$\phi(t)$可以由 $VR(t)$ 積分得到。

關於抖音參數的產生，我們也是沿用前人的成果[3]，但是由於此 ANN 抖音模型是針對國語歌聲合成而建立的，所以我們若要直接使用此抖音模型，就必需將英語音節分類並對應到國語音節，以便套入該模型去產生出抖音參數。

我們訂定的分類規則如下：(1)先將母音分類，如/aa/、/ah/對應至/a/。(2)再檢查音節後子音是否為/m/、/n/、/l/、/r/、/v/，如/eh n/對應/en/。(3)音節前子音照發音方式分類，如摩擦音、爆破音等。(4)音節後的摩擦音及爆破音子音，因為國語音節不包含此成分，所以不加以考慮。幾個對應的例子如下，英語音節/g aa d/對應至國語音節 ga，英語音節/s ey m/對應至國語音節 sen。

4.3. 聽測實驗

我們合成了兩首歌曲來作聽測實驗，分別是較輕快的"Jingle Bells"和抒情的"Eyes On Me"，可從網頁 http://guhy.csie.ntust.edu.tw/EngSong/ 去下載試聽這兩首合成的歌曲。參與實驗的聽測者為大學生及碩士生共 15 名，其中 6 人為本實驗室同學，其他 9 人則為校外人士。評分的範圍為-2~2 分，0 分代表分不出差異，1 分是好一些，2 分是好很多，反之亦然。

聽測實驗分成兩部份，內部聽測比較的是自然度，比較對象為**優先選擇音節**的方式和**完全使用半音節接合**的方式，在此以完全使用半音節的方式為基準；外部聽測是與市面販售的一套女聲歌聲合成軟體"VOCALOID LOLA"[15]作比較，檢測我們系統與商業軟體的差距。

表 2 為聽測實驗之結果，在內部聽測方面，其實大多數人給的都是 0 分，雖然兩種合成方式的差別在於合成單元的選擇方式不一樣，一個是先選音節為主，另一個是全部以半音節接合方式，但是合成出的歌聲差異不大。在外部聽測方面，分數的分布範圍很廣，從-2 到 2 分都有人給分，因此詢問聽測者的意見，評正分的人表示，對於音節信號而言，我們程式的合成信號聽起來比較自然；但是評負分的人表示，在音節之間的連接性方面，我們程式是分別合成出各個音節的信號再作串接，所以整體聽起來會有不夠平順、不夠柔和的感覺。

表 3 聽測實驗之結果

	Jingle Bells	Eyes On Me
內部聽測-平均評分	0.25	0.20
外部聽試-平均評分	0.08	-0.03
內部聽測-標準差	0.42	0.40
外部聽試-標準差	1.32	0.93

5. 結論

我們初步建置了一個英語歌聲的合成系統。由於英語多了音節尾部的子音，因此我們將尾部子音分成四類，以分別考慮 HNM 參數的分析方式。此外，從所收集的英語朗讀語句，只能找出 1,389 個不同的音節，也就是面臨了音節單元不足的問題，因此我們提出一個音節建構的流程，當遇到語料裡不存在的音節時，先考慮以音節後串接子音的方式，若不行，再考慮前、後半音節接合的方式，若不行，再考慮音節前串接子音的方式，如果仍不行，則考慮以類似音節來替換。

關於歌譜格式，我們設計了一種使用者方便輸入的歌譜格式(一)，接著藉由程式查詢 CMU 英語詞典以及我們整理的音節分割規則表，可把該歌譜格式轉換成合成程式方便處理的另一種歌譜格式(二)。

在音樂性參數的設定方面，除了採用前人的研究成果(如節拍對齊)，我們也針對英語歌聲加入了動態滿度設定和子音音量調整，另外在半音節的接合點前後，也作了音量的修正。關於抖音參數的數值，我們透過所設計的子、母音對應表的轉換，來使用前人的 ANN 抖音參數模型作產生，以便合成出較為自然的歌聲。

對於所建造的英語歌聲合成系統，我們經由聽測實驗來評估，發現優先選擇音節會比完全使用半音節的方式自然，但是差異不大。此外，與市面上販售的合成軟體作比較，我們系統所合成出的歌聲也有不錯的表現，但是在相鄰音節作連接的處理方面，則仍需要改進；並且，我們使用的是英語朗讀的語料，合成出的歌聲欠缺歌者共振峰(singer's formant)的特性，這點未來可再作改進。

參考文獻

[1] 古鴻炎、廖皇量，「用於國語歌聲合成之諧波加噪音模型的改進研究」，*WOCMAT 2006 國際電腦音樂與音訊技術研討會*，台北，*session 2* (音訊處理 I)， 2006。

[2] 古鴻炎、王如江，「基於歌聲表情分析之國語歌聲合成」，*第十二屆人工智慧與應用研討會* TAAI 2007，雲林，Application-of-Intelligent-System Session，2007。

[3] H. Y. Gu and Z. F. Lin, "Mandarin Singing Voice Synthesis Using ANN Vibrato Parameter Models", *Int. Conf. Machine Learning and Cybernetics* (ICMLC2008), Kunming, China, pp. 3288-3293, 2008.

[4] J. Bonada and X. Serra, "Synthesis of the Singing Voice by Performance Sampling and Spectral Models", *IEEE Signal Processing Magazine*, pp. 67-79, 2007.

[5] Y. E. Kim, "A Framework for Parametric Singing Voice Analysis/Synthesis", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, NY, USA, pp. 123-126, 2003.

[6] Carnegie Mellon University, *The CMU Pronouncing Dictionary*, http://www.speech.cs.cmu.edu/speech/

[7] K. Sjölander and J. Beskow, *Centre of Speech Technolodge at KTH*, http://www.speech.kth.se/wavesurfer/

[8] Y. Stylianou, *Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification*, Ph.D. thesis, Ecole Nationale Superieure des Telecommunications, Paris, France, 1996.

[9] C. Dodge and T. A. Jerse, *Computer Music: Synthesis, Composition, and Performance,* 2nd ed., Schirmer Books, 1997.

[10] 王小川，*語音信號處理*，全華科技圖書公司，台北，2004。

[11] D. O'Shaughnessy, *Speech Communication: Human and Machine*, 2nd ed., IEEE Press, 2000.

[12] 王國憲，*基於聲學特性之國語語音合成流暢度改進之研究*，國立台灣科技大學電機工程研究所碩士論文，2003。

[13] V. Kyritsi, A. Georgaki and G. Kouroupe-troglou, "A score-to-singing voice synthesis system for the greek language", *International Computer Music Conference* (ICMC07), Copenhagen, 2007.

[14] 周彥佐，*基於 HNM 之國語、閩南語的語音合成研究*，國立台灣科技大學資訊工程研究所碩士論文，*2007*。

[15] Zero-g, *Virtual female vocalist from zero-g*, powered by Yamaha VOCALOID singing synthesis technology, http://www.zero-g.co.uk/