第四章　加入韻律模型的語音合成實作

4.1 實驗設定

4.1.1 實驗流程

本論文的實驗方式，是對各種韻律參數的產生模型，做好壞的評估比較，其流程如圖 12。

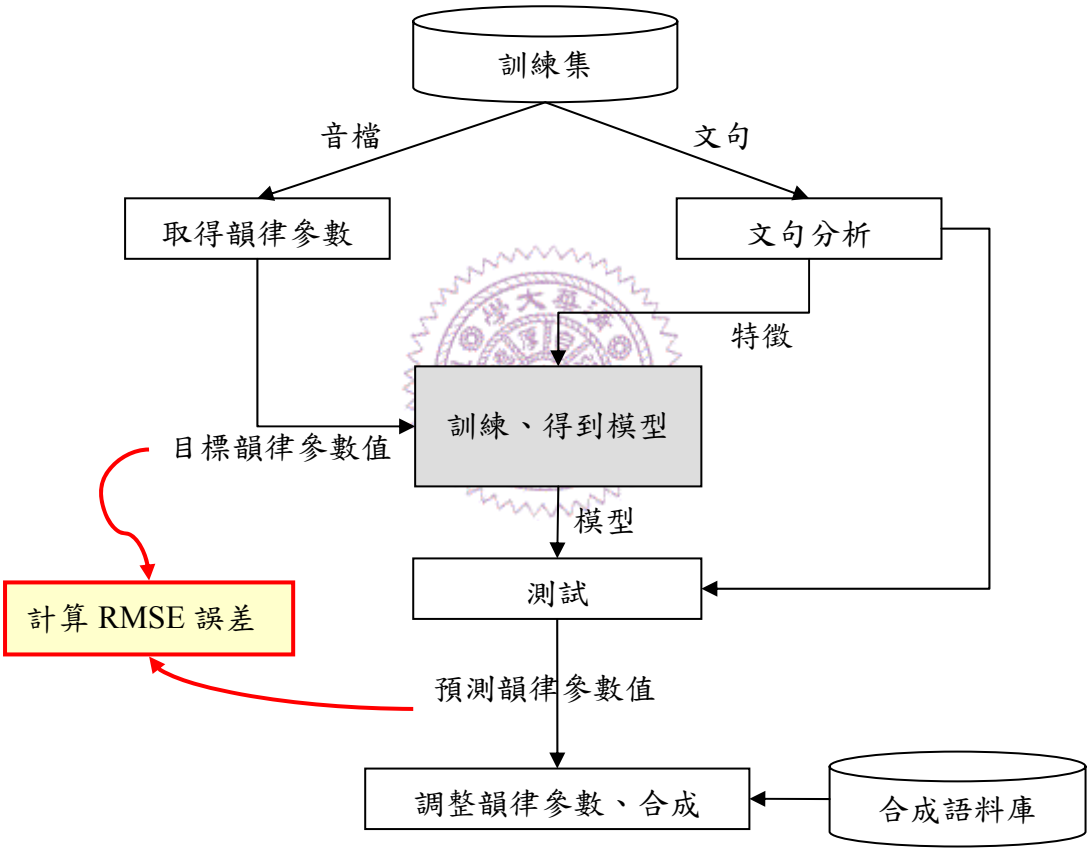

圖 15　本論文的實驗流程圖

4.1.2 韻律模型評估方式

實驗結果的好壞評估，是以訓練語料庫擷取出的韻律參數，當作標準答案，對模型實作結果輸出，計算均方根誤差值（Root-Mean-Square Error, RMSE），同

時進行聽測實驗之結果討論。RMSE 公式如下式（18）。其中 N 為總資料筆數，X 為預估結果，T 為標準答案。其誤差值越小越好。

$$Error = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - T_i)^2} \qquad (18)$$

4.1.3 迴歸模型的輸入

一般而言，迴歸模型的輸入參數越多，其預測的結果應該會比較準確。台科大古鴻炎教授的 ANN 模型的輸入，只有音節層的特徵參數。而與交通大學陳信宏教授的 RNN 模型相比較，我們又取了更多的特徵。其比較表列如下表 9。

表 9　與 ANN、RNN 的輸入特徵比較表

輸入特徵	MIR TTS	ANN	RNN
前一音節韻母類別	○	○	
前一音節聲調	○	○	
本音節韻母類別	○	○	○
本音節聲母類別	○	○	○
本音節聲調	○	○	○
次一音節聲母類別	○	○	○
次一音節聲調	○	○	○
前一詞的詞性	○		
本詞的詞性	○		○
次一詞的詞性	○		○
前一詞的詞長	○		
本詞的詞長	○		○
次一詞的詞長	○		○
位於詞中位置	○		○
位於句中位置		○	
本詞前的標點符號	○		○
本詞後的標點符號	○		○

其中，由於我們已經使用承載式語料庫設計我們的語音合成器，在單元選取時，取出對應句中位置（句首、句中、句末）的合成單元，所以我們才會沒有把「位於句中位置」這項特徵放入模型的輸入。

4.2 實驗結果

4.2.1 類神經網路單輸出與多輸出的比較

關於第三章提出的韻律獨立訓練法，我們做了實驗來證實這個方法確實合理且有效。首先，我們假設，獨立訓練應該更專注於模型的建立，理論上應該會表現得較佳。我們先針對傳統作法與獨立訓練這兩種不同的作法，使用基本的類神經網路做訓練實驗與比較。結果如表 10 所列。

經由實驗證實了韻律獨立訓練法的表現較佳，並且花費較少時間，也讓我們方便分別討論各個模型，故以下實驗皆採用獨立訓練的方式進行。

表 10　類神經網路傳統多輸出與獨立訓練法的 RMSE 值比較表

語料庫	方法	測試	聲母長 (sec)	韻母長 (sec)	OGN1 (Hz)	OGN2	OGN3	OGN4	音量 (dB)
Neutral Sentence	傳統	內部	0.0186	0.0360	35.008	23.709	18.537	15.885	2.2308
		外部	0.0212	0.0341	34.921	28.236	21.418	19.267	2.5090
	獨立	內部	0.0177	0.0379	33.684	18.717	14.542	9.7477	2.1702
		外部	0.0233	0.0335	35.110	29.734	22.514	20.090	2.5933
HSF	傳統	內部	0.0237	0.0405	23.389	7.7729	4.6862	3.0278	3.8904
		外部	0.0253	0.0425	27.104	9.7455	5.8121	4.0312	6.3035
	獨立	內部	0.0235	0.0387	22.590	7.8548	4.7588	2.9033	3.6222
		外部	0.0254	0.0432	27.769	9.6875	5.7612	4.1435	6.4286

4.2.2 三種迴歸模型 RMSE 比較

依照第三章的設計方法，我們以兩筆資料庫分別作訓練迴歸模型的實驗。其中類神經網路與支撐向量機的使用，參照一般使用的設計做測試。

倒傳遞類神經網路的訓練環境設計，採用兩層隱藏層，每層中有 20 個神經元，作 30 次的學習循環。

而會影響支撐向量機的表現的有式（13）中的 C 值與核心函數 Radial Basis Function 中的 σ 值。我們使用交叉比對法（Cross Validation），大範圍嘗試各種 C 與 σ 值的組合。C 的實驗範圍為 $2^{-5},2^{-3},...,2^{15}$，σ 的實驗範圍為 $2^{-15},2^{-13},...,2^{3}$，針對每一個韻律參數皆挑選最佳參數的組合，在此最佳組合的附近，再進一步做搜尋。其最佳參數如表 11 所列，並且以這些挑選出來的參數組合，訓練我們的支撐向量機韻律模型。

表 11　SVM 各個韻律參數的最佳 C 與 σ 值

語料庫	方法	聲母長	韻母長	OGN1	OGN2	OGN3	OGN4	音量
Neutral Sentence	C	612	412	28	58	58	58	52
	σ	0.0625	0.0625	0.0625	0.03125	0.03125	0.00781	0.25
HSF	C	2148	612	132	82	52	50.125	228
	σ	0.01563	0.0625	0.01563	0.01563	0.03125	0.00781	0.00391

依線性迴歸法、倒傳遞類神經網路、與支撐向量機三種迴歸模型，得到的實驗結果，其 RMSE 誤差值如表 12 所列。

觀察 RMSE 誤差值的表現，我們發現線性迴歸法已有一定的表現，但是整體誤差量較其餘二者略高。SVM 訓練出的模型表現並不會比類神經網路差，有時候甚至還有更好的表現。現比較這三者的優缺點如下表 13。

表 12　三種迴歸模型的 RMSE 值比較表

（LR:線性迴歸／NN:倒傳遞類神經網路／SVM:支撐向量機）

語料庫	方式	測試	聲母長 (sec)	韻母長 (sec)	OGN1 (Hz)	OGN2	OGN3	OGN4	音量 (dB)
Neutral Sentence	LR	內部	0.0192	0.0382	33.750	18.793	15.716	11.680	2.4806
		外部	0.0206	0.0333	34.765	27.657	20.200	18.303	2.4403
	NN	內部	0.0177	0.0379	33.684	18.717	14.542	9.7477	2.1702
		外部	0.0233	0.0335	35.446	27.707	20.242	19.337	2.5933
	SVM	內部	0.0049	0.0125	11.126	9.9982	8.9142	10.518	0.7461
		外部	0.0210	0.0329	33.758	27.438	20.088	18.331	2.3506
語料庫	方式	測試	聲母長 (sec)	韻母長 (sec)	OGN1 (Hz)	OGN2	OGN3	OGN4	音量 (dB)
HSF	LR	內部	0.0251	0.0431	24.064	8.1444	4.8131	3.1096	4.0612
		外部	0.0242	0.0416	26.639	9.5350	5.7097	3.9565	6.2233
	NN	內部	0.0235	0.0387	22.590	7.8548	4.7588	2.9033	3.6222
		外部	0.0254	0.0432	27.769	9.6875	5.7612	4.1435	6.4286
	SVM	內部	0.0040	0.0038	18.621	6.6954	3.9031	3.0121	3.4483
		外部	0.0238	0.0401	26.683	9.4555	5.7002	3.9433	6.2723

表 13　類神經網路與支撐向量機的優缺點比較

	優點	缺點
線性迴歸法	• 簡單快速 • 適合嵌入式系統、網頁程式	• 沒有參數可進一步調整 • 預測能力稍差
類神經網路	• 學習能力強 • 具有容錯能力	• 訓練結果不穩定，一樣的輸入與輸出，每次訓練結果卻不一定相同，造成研究者的麻煩 • 網路過大，十分耗費記憶體
支撐向量機	• 多維度的輸出則不容易求解 • 速度較慢	• 數學方法求解，訓練結果穩定 • 預測能力佳

4.2.3 以韻律模型預測結果加入語音合成的改良

一旦完成韻律參數值的預測，我們便可以採用語音合成器來合成出所需的語句。預測誤差值越低，不必然表示合成結果聽起來會比較流暢，所以我們也進行主觀的聽測評估，以確認我們的韻律模型是否對整體的合成自然度有幫助。

我們對兩筆訓練語料庫，分別以其內部測試與外部測試的文句，進行我們韻律產生器的預測且合成語音輸出，經由實驗發現，當使用我們合成系統時確實比原本單純的承載式合成系統較接近真人說話的腔調。圖 16-圖 19 是幾個合成結果例子。

4.2.4 聽測實驗

我們設計一個聽測實驗，包含內部與外部測試的句子共 10 句，句子內容請參考附錄。每句皆含有韻律未經調整前、與經過韻律模型改良後的合成音檔各一句。由 12 個使用者分別對其自然度作評分，分數由 1 分至 5 分，以 0.25 分為間距。

其結果，韻律未經調整前的平均分數為 3.87，經過韻律模型改良後的平均分數為 4.13。經過統計，其中認為未經調整前的自然度較佳的佔了 31.7%，而認為改良後的自然度較佳的佔了 62.5%，認為兩者不相上下的佔了 5.8%。由此可以證明，我們的實驗不僅降低 RMSE 誤差量，在主觀的聽覺感受上，其合成結果自然度亦有顯著的進步。

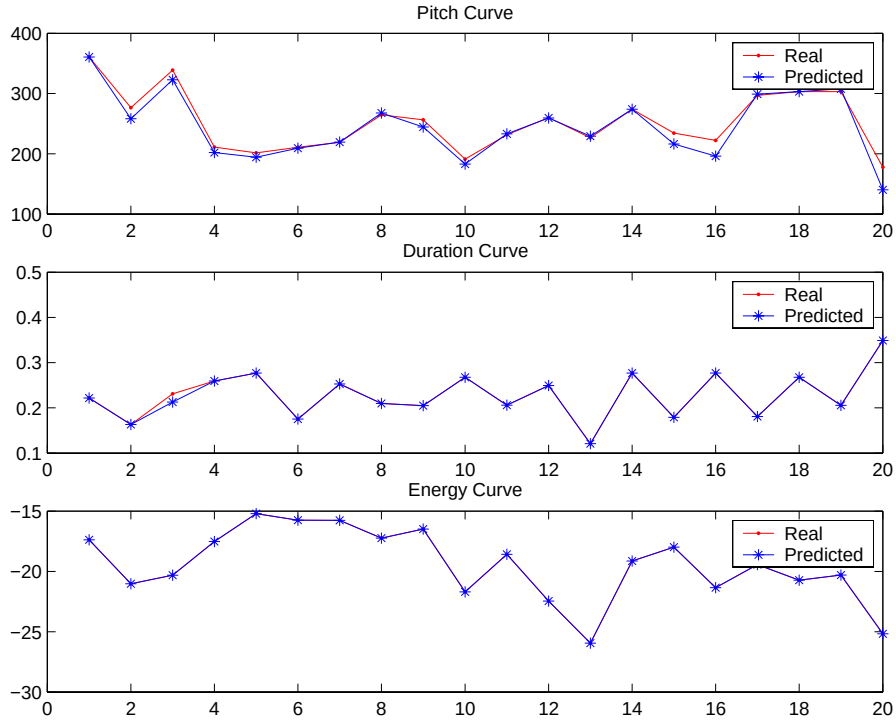

圖 16　NeutralSentence 訓練語料庫的內部測試範例

「太不公平了！明明大家條件差不多，他卻高分錄取。」

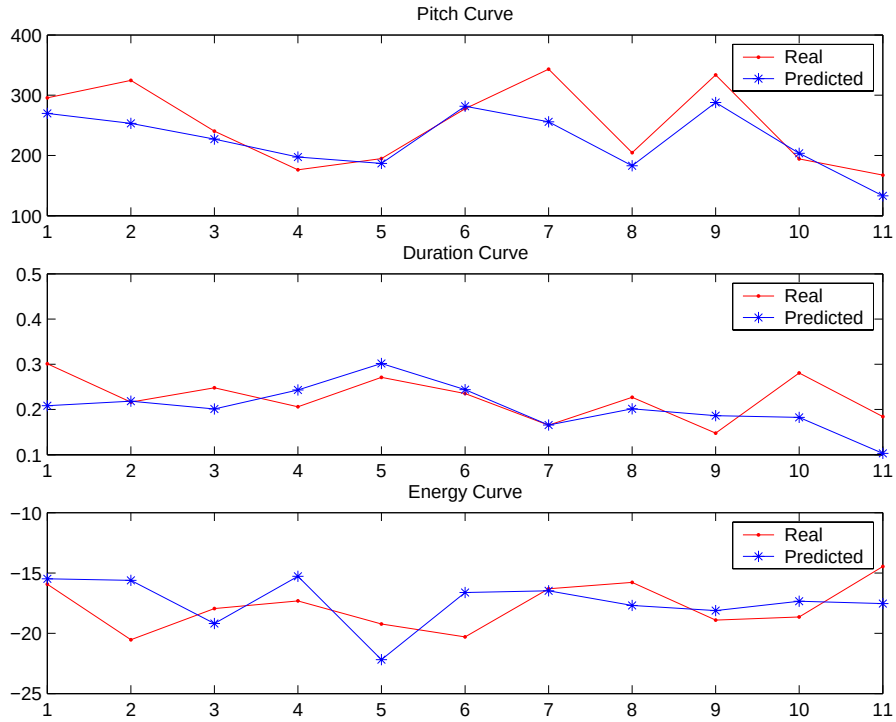

圖 17　NeutralSentence 訓練語料庫的外部測試範例

「誰允許你隨便動我電腦的。」

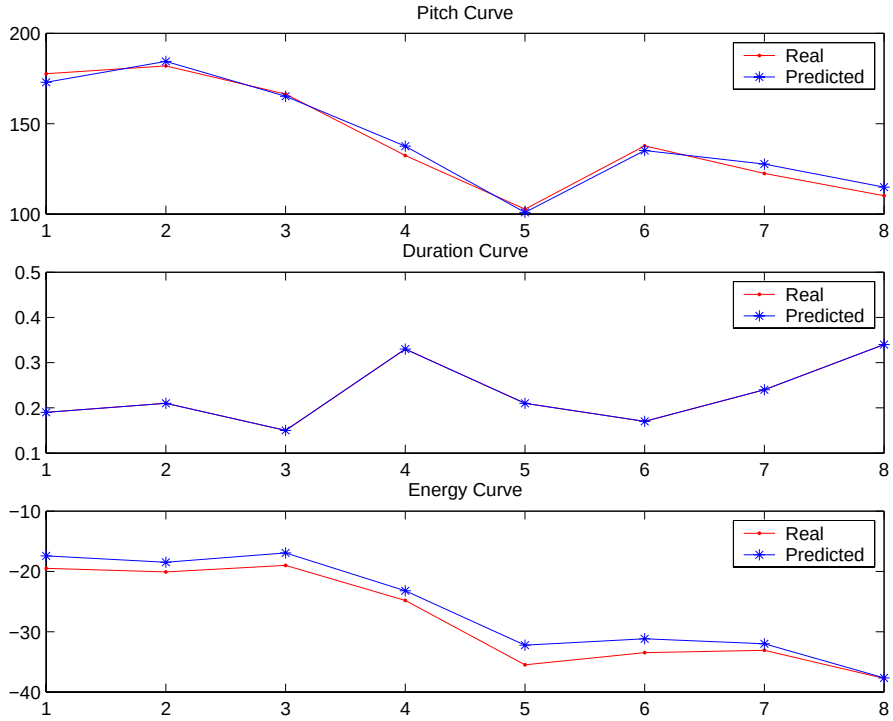

圖 18　HSF 訓練語料庫的內部測試範例

「哈哈大笑有助健康」

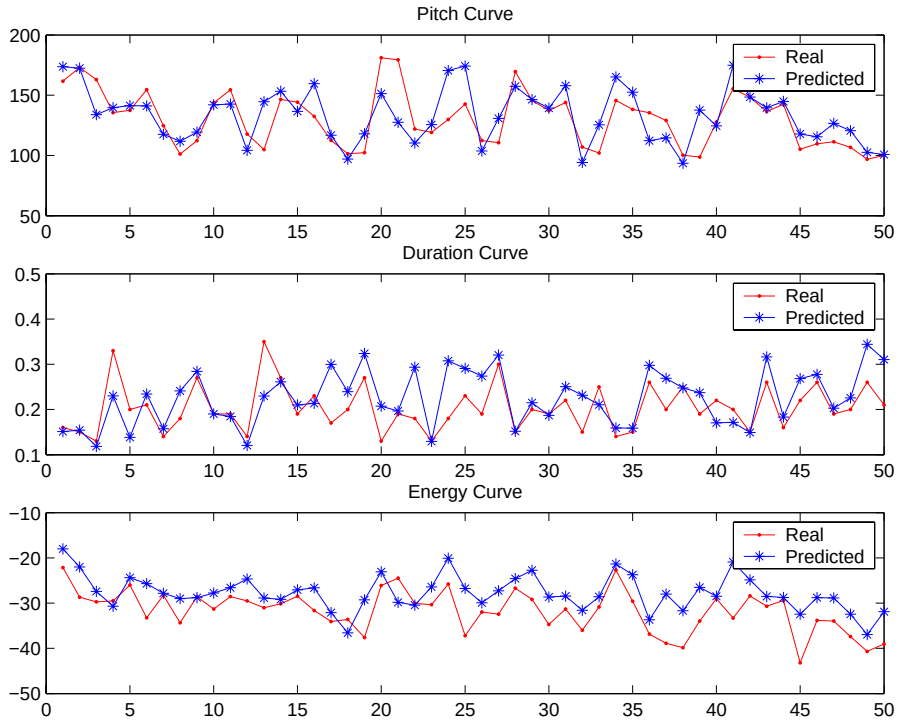

圖 19　HSF 訓練語料庫的外部測試範例

「嘉義地區各中等學校暨救國團鄉鎮市團委會，為配合「國家清潔週」，特訂今天舉辦「大地清淨」活動，協助各鄉鎮民眾整理環境。」