

第三章 韻律產生器的設計

本論文所提的韻律參數預測模型之設計流程如圖 9。本章將針對各部份流程做詳細分析與介紹。

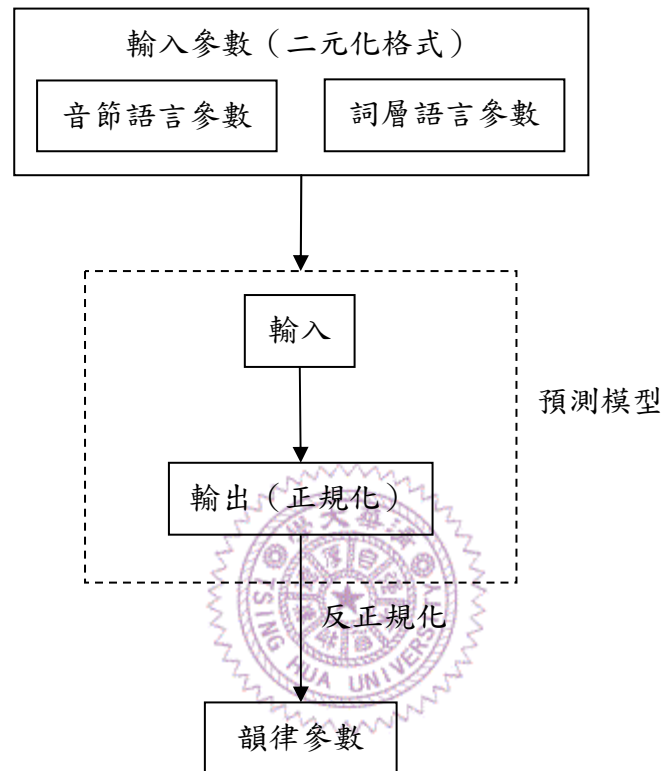


圖 13 韻律預測模型流程圖

3.1 韻律模型的輸入與輸出

要估算韻律模型，首先要取得會影響韻律參數的特徵參數群作為輸入，並且以訓練語音資料庫中已經標好的韻律參數，作為目標輸出，以期調整此模型至輸出為最接近訓練語料庫的韻律參數。

3.1.1 文句分析得到之特徵

要對韻律參數進行訓練，首先要得到影響韻律參數的語言學特徵群，我們進行文句分析處理來得到欲合成之語句的語言學特徵。

文句分析的處理是採用長詞優先（Longest Matching）法；由前往後斷詞，即由句子的第一個音節開始往後查找是否有符合的詞，直到查到最後一個音節為止，查找時以可斷出之最長詞為優先考慮。接著，經過後續的構詞處理與變調處理後，得到十六個特徵參數[14]，詳列於表 1。

其中屬於音節層的特徵參數包含：聲母（Consonant）、韻母（Vowel），與音調（Tone）。另外考慮到前後音節也會對韻律參數造成影響，故加入了前一音節的韻母與音調，和次一音節的聲母與音調。其分類如表 2-表 4 所列。

屬於詞層的特徵參數則有詞性（POS）、詞長（Word Length）、位於詞中的位置，和該詞前後的標點符號（PM）。一樣考慮到前後詞對韻律參數造成的影響，加入了前後詞的詞性和詞長。其分類如表 5-表 8 所列。

韻律模型輸入的方式，是以輸入參數採用一對一的映對。若有 5 種可能的輸入值，而輸入值為 3 時，以 5 個位元來表示，第三個輸入位元為 1，其他 4 個輸入位元皆為 0。映對結果每一音節的文句特徵總共包含 95 位元。

3.1.2 欲估算之韻律參數

本論文研究的韻律參數預測模型，以音長（Duration）、基週頻率（Pitch Contour）、與音量（Amplitude）三種。音長以秒為單位，音量以 dB 值為單位，而基週軌跡表示則是參考交通大學陳信宏教授所提出的四項正交基週軌跡函數

的係數(以下我們簡稱為 OGN)[19],把基週頻率軌跡轉成四項係數,以供訓練。

表 1 韻律模型的輸入

輸入位置	輸入形式	長度	說明	備註
1-6	Binary	6	前一音節韻母類別	音節層 語言參數
7-11	Binary	5	前一音節聲調	
12-17	Binary	6	本音節韻母類別	
18-26	Binary	9	本音節聲母類別	
27-31	Binary	5	本音節聲調	
32-40	Binary	9	次一音節聲母類別	
41-45	Binary	5	次一音節聲調	
46-54	Binary	9	前一詞的詞性	詞層 語言參數
55-63	Binary	9	本詞的詞性	
64-72	Binary	9	次一詞的詞性	
73-75	Binary	3	前一詞的詞長	
76-78	Binary	3	本詞的詞長	
79-81	Binary	3	次一詞的詞長	
82-85	Binary	4	位於詞中位置	
86-90	Binary	5	本詞前的標點符號	
91-95	Binary	5	本詞後的標點符號	

表 2 韻母分類表

類別	韻母
1	空韻母
2	ㄚ、一ㄚ、ㄨㄚ
3	ㄛ、一ㄛ、ㄨㄛ
4	ㄜ、儿
5	ㄝ、一ㄝ、ㄛㄝ
6	ㄝ、一ㄝ、ㄨㄝ、一、ㄟ、ㄨㄟ
7	ㄞ、一ㄞ、ㄛ、ㄨ、ㄨ、一ㄨ
8	ㄟ、一ㄟ、ㄨㄟ、ㄛㄟ、ㄟ、一ㄟ、ㄨㄟ、ㄛㄟ
9	ㄠ、一ㄠ、ㄨㄠ、ㄠ、一ㄠ、ㄨㄠ、ㄠㄠ

表 3 聲母分類表

類別	聲母
1	ㄇ、ㄋ、ㄌ、ㄍ、空聲母
2	ㄈ、ㄊ、ㄑ
3	ㄏ、ㄒ、ㄍ
4	ㄣ、ㄗ、ㄛ
5	ㄨ、ㄜ、ㄛ
6	ㄍ、ㄑ、ㄒ、ㄗ、ㄛ

表 4 聲調分類圖

類別	聲調
1	一聲
2	二聲
3	三聲
4	四聲
5	五聲

表 5 詞性分類表

類別	詞性
1	形容詞
2	名詞
3	動詞
4	介詞
5	連接詞
6	副詞
7	感歎詞
8	冠詞
9	特殊動詞（是）

表 6 詞長分類表

類別	詞長
1	單字詞
2	二字詞或三字詞
3	多字詞

表 7 詞中位置分類表

類別	詞中位置
1	詞首
2	詞末
3	詞中
4	單字詞

表 8 標點符號分類表

類別	標點符號
1	、，：
2	？！
3	。；
4	其他符號
5	無

3.2 訓練語料準備

3.2.1 訓練語音資料庫設計

我們的訓練資料庫共有兩種。其一為我們實驗室原有的語音資料庫 (NeutralSentence)，此語料庫與承載句語料庫為同一個專業女錄音員所錄製。另一個則為台北科大黃紹華老師所提供的語料庫 (HSF) [4]，此為另一個專業的男廣播員所錄製。

實驗室的語音資料庫共 1,000 句，其中取 500 句為訓練集，共 6,787 個單音節；剩下 500 句為測試集，共 7,155 個單音節。分別進行模型預測與目標韻律參數值的誤差評估，並對語音合成結果作聽測實驗。

另外，黃紹華老師的語料庫前 455 句短句作為訓練集，包含 112 句 13 個音節以內、315 句 40 個音節以內的語音平衡短句，和 28 句新聞短句，總共有 6,038 個音節；測試集為 200 句新聞長句語料庫，有 28,378 個音節。其中新聞語句的選取來源，新聞內容包含商務(12.5%)、醫學(12%)、社會事件(12%)、運動(10.5%)、文學(9%)、電腦(8%)、食品和營養(8%)、電影(6.5%)、家庭生活(6.5%)、旅行(6%)、政治(2.5%)、交通和運輸(2.5%)等等。聲音來源為同一個男生。

3.2.2 正規化

由於人說話的時候並不能一直維持一樣的語氣和語調，語句長短也都不一，所以要正規化，使之變異量變小，才不會被這些差異影響了訓練的結果。

其音長的正規化是以聲母為分類，音量的正規化則是以韻母為分類，而代表基週軌跡的四項 OGN 係數的正規化則以音調為分類。正規化方式為減去該分類

的平均值，再除以該分類的標準差。假設該韻律參數屬於分類中的第 k 類， N 為該分類中的資料筆數，則第 i 筆韻律參數的正規化公式如下式：

$$\begin{aligned} Nor_i &= \frac{X_i - mean_k}{std_k}, \text{where} \\ mean_k &= \frac{1}{N} \sum_{j=1}^N X_j \\ std_k &= \sqrt{\frac{1}{N} \sum_{j=1}^N (X_j - mean_k)^2} \end{aligned} \quad (14)$$

因為韻律參數需經過正規化才去訓練其對應的韻律模型。所以經由模型產生的預測結果必須反正規化後，才能得到正確的預測實際值。反正規化公式如下式：

$$deNor_i = X_i \times std_k + mean_k \quad (15)$$

另外，因為我們合成單元語料庫為女生的聲音，而黃紹華老師的語料庫為男生的聲音，所以在韻律參數上勢必要做調整。本論文中，我們依照聲母、韻母、音調做為分類依據，對音長、音量、與平均基週軌跡各求兩者的平均量的差，以黃紹華老師的語料庫訓練出來的預測值，作一個平移。

3.3 韻律模型的訓練

3.3.1 線性迴歸

在實作上，我們首先使用線性迴歸法來預測韻律參數，線性迴歸的原理如第二章所描述。我們把特徵群當作輸入值，韻律參數值經過正規化後當作模型的輸出值，對每一種特徵單元的權重值作預測。其測試結果與合成分析請參閱第四章所述。

$$\underbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,95} \\ x_{2,1} & x_{2,2} & & x_{2,95} \\ \dots & & & \\ x_{N,1} & x_{N,2} & \dots & x_{N,95} \end{bmatrix}}_{\text{features}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_{95} \end{bmatrix}}_{\text{weight}} \cong \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}}_{\text{prosody}} \quad (16)$$



3.3.2 倒傳遞類神經網路

一般中文的語音合成研究中，韻律參數的模型訓練上，大多使用的是類神經網路，例如本文第一章所提到的交通大學陳信宏教授，與台科大古鴻炎教授的研究。因為類神經網路具有適應性學習能力，並且可以使用在多輸入且多輸出的情形我們試著模擬其他研究的類神經方法，並且提出我們的想法。

類神經網路與倒傳遞類神經網路的介紹請參考第二章。我們以倒傳遞類神經網路為基礎，設計我們的韻律模型。

在神經元的設計上，我們所使用的集成函數為加權乘積和，作用函數是直接使用集成函數輸出，而轉換函數為非線性的雙曲線正切函數，如圖 4 圖 5。

層的設計是正規化輸出，經過兩層隱藏層，每一層包含的神經元的個數則經

過測試，決定為 20 個。

網路結構是屬於監督式學習的前向式網路架構，每個神經元皆與相鄰層內的所有神經元之間有網路連結，每次訓練作 20 次迭代求最佳網路連結權重值。我們的類神經韻律模型如圖 14。

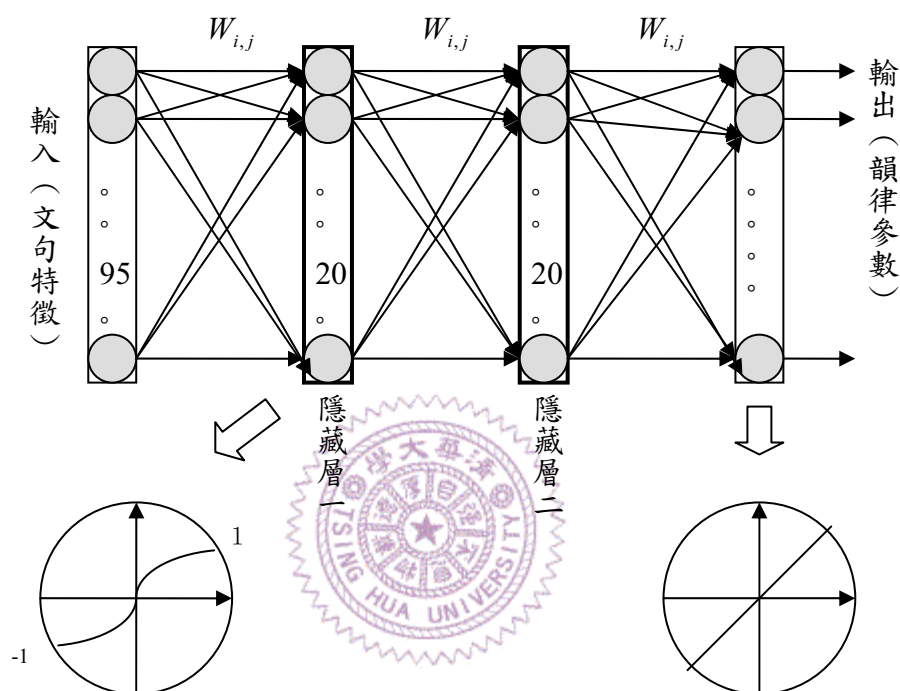


圖 14 本論文以倒傳遞類神經網路為基礎的韻律模型架構示意圖

一般的研究都是利用類神經網路多輸入多輸出的特性，方便而有效率的使用這個方式。但我們在這邊提出一個假設，對每種韻律參數做分開預測，換言之，假如有 4 種韻律參數，則我們需要訓練出 4 個不同的迴歸模型，這是與前人研究較為不同的地方，稱之為韻律獨立訓練法。

韻律獨立訓練法，有其利弊，優點是準確，但相反的，我們卻必須犧牲更多空間，來存放多個網路模型，同時，也必須花費更多時間，來對韻律參數一一作訓練。我們對單輸出與多輸出的類神經網路模型做實驗，實驗結果請參考第四章。

3.3.3 支撐向量機

除了一般研究使用的類神經網路來作迴歸分析，我們嘗試使用了以數學求解的支撐向量機迴歸。我們的支撐向量機的核心函數是使用 Radial Basis Function，計算迴歸誤差的 Loss Function 則是 ε -Insensitive，如第二章所介紹。

我們使用台灣大學林智仁教授所發展的支撐向量機的工具箱（LIBSVM）[20]。除了傳統的支撐向量機的迴歸，請見第二章的說明；另外並提供多加入一個參數 ν 的支撐向量機迴歸（ ν -Support Vector Regression），詳細理論請見[21]，即式（13）改寫為下式（17）。

$$R_{w,b,\xi,\xi^*,\varepsilon} = \frac{1}{2} \|w\|^2 + C(\nu\varepsilon + \frac{1}{N}(\sum_{i=1}^N \xi + \sum_{i=1}^N \xi^*)) \quad (17)$$

經過實驗證實，確實比傳統的支撐向量機更有效降低誤差值，所以我們決定採用包含參數 ν 的支撐向量機模型。同樣地，把特徵群當作輸入值，韻律參數值當作模型的輸出值，求得預測超平面的同時，即可得到式（17）中的最佳 w 、 b 、 ξ 、 ξ^* 與 ν 值，然後我們可以依此定義出我們的預測模型。另外，我們也採用韻律獨立訓練法來求分別求出對應的支撐向量機迴歸模型。