(19)中華民國智慧財產局
(12)發明說明書公開本 (11)公開編號：TW 201411602 A
(43)公開日： 中華民國 103 (2014) 年 03 月 16 日

(21)申請案號：101133059 (22)申請日： 中華民國 101 (2012) 年 09 月 10 日
(51)Int. Cl.： *G10L13/08 (2013.01)* *G10L15/08 (2006.01)*
G10L15/02 (2006.01)
(71)申請人：國立交通大學 (中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)
新竹市大學路 1001 號
(72)發明人：陳信宏 CHEN, SIN HORNG (TW) ；王逸如 WANG, YIH RU (TW) ；江振宇 CHIANG, CHEN YU (TW) ；謝喬華 HSIEH, CHIAO HUA (TW)
(74)代理人：蔡清福
申請實體審查：有 申請專利範圍項數：18 項 圖式數：7 共 35 頁

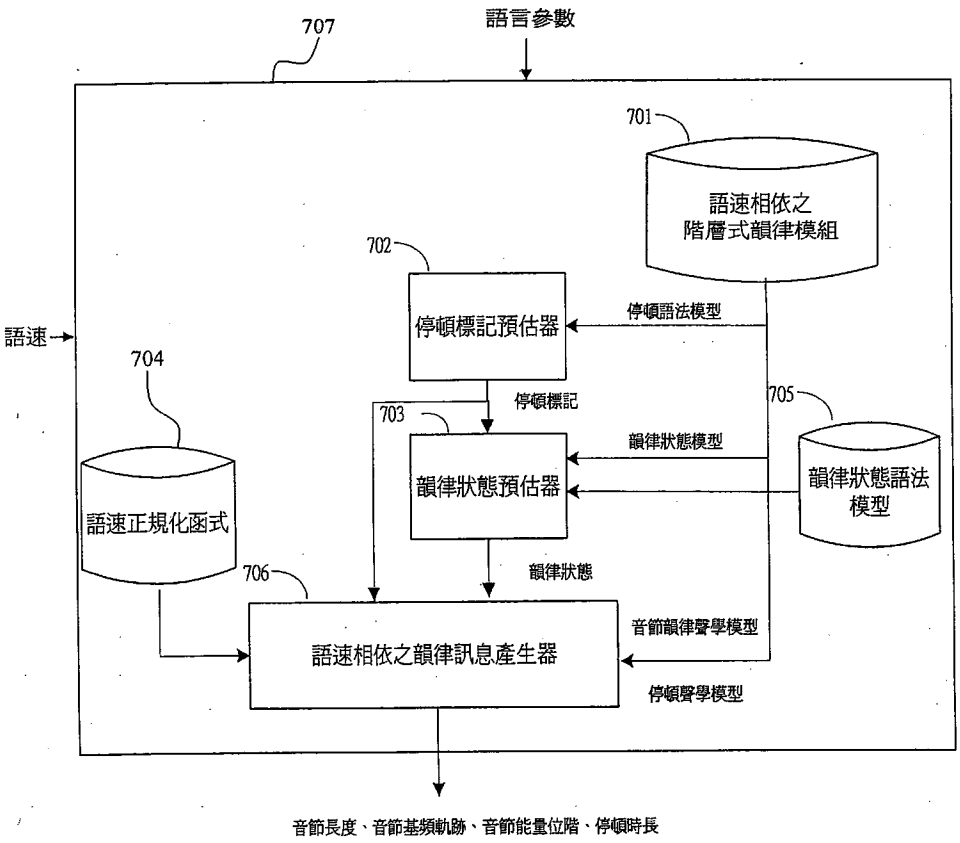
(54)名稱

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組

SPEAKING-RATE CONTROLLED PROSODIC-INFORMATION GENERATING DEVICE AND SPEAKING-RATE DEPENDENT HIERARCHICAL PROSODIC MODULE

(57)摘要

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一文字；一文字分析器，用以接收該文字，以產生一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。



第七圖

701：語速相依之階層式韻律模組

702：停頓標記預估器

703：韻律狀態預估器

704：語速正規化函式

705：韻律狀態語法模型

706：語速相依之韻律訊息產生器

707：語速相依之韻律生成模組

201411602

發明專利說明書

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101137059

※ 申請日：101.9.10

※IPC 分類：G10L 13/08 (2013.01)
G10L 15/08 (2006.01)
G10L 15/02 (2006.01)

一、發明名稱：（中文/英文）

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組/Speaking-Rate Controlled Prosodic-Information Generating Device and Speaking-Rate Dependent Hierarchical Prosodic Module

二、中文發明摘要：

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一文字；一文字分析器，用以接收該文字，以產生一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。

三、英文發明摘要：

The present invention provides a speaking-rate controlled prosodic-information generating device, including a first input for receiving a speaking rate; a second input for receiving a text; a text analyzer for receiving a text to produce a linguistic feature ; a speaking-rate dependent hierarchical prosodic module uses a linguistic feature incorporating with a speaking rate to produce a SR-dependent

prosodic-acoustic feature; and an output for outputting a SR-dependent prosodic-acoustic feature.

四、指定代表圖：

(一)本案指定代表圖為：第（七）圖。

(二)本代表圖之元件符號簡單說明：

701：語速相依之階層式韻律模組

702：停頓標記預估器

703：韻律狀態預估器

704：語速正規化函式

705：韻律狀態語法模型

706：語速相依之韻律訊息產生器

707：語速相依之韻律生成模組

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

prosodic-acoustic feature; and an output for outputting a SR-dependent prosodic-acoustic feature.

四、指定代表圖：

(一)本案指定代表圖為：第（七）圖。

(二)本代表圖之元件符號簡單說明：

701：語速相依之階層式韻律模組

702：停頓標記預估器

703：韻律狀態預估器

704：語速正規化函式

705：韻律狀態語法模型

706：語速相依之韻律訊息產生器

707：語速相依之韻律生成模組

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

六、發明說明：

【發明所屬之技術領域】

本發明係關於一種韻律訊息產生裝置，尤指一種以語速相依之階層式韻律模組為基礎之韻律訊息產生裝置。

【先前技術】

目前對於相異語速語音合成之討論不少，但始終不能達成一流暢的自然合成語音。其中有以每個語音音框的時間軸進行伸張及壓縮，藉以調整說話速度慢及快之研究，但並未考慮到語速對於韻律結構的影響；以線性預估的方式對輸入語音進行語速修正，對輸入的語音信號以線性插入或移除信號本身之研究，該方法雖簡易有效率，但對於語速的考量過於粗糙；以清化元音(devoiced vowel)的決定中考慮了語速影響，有效地改進清化元音在慢語速的退化程度之研究，但其韻律的產生方法並未考量語速的影響；以對不同語速語料庫建立韻律結構的轉換關係，藉以達到語速轉換的目的之研究，但該方法並不能掌握到連續語速的轉換變化；雖有文獻實現了可控制語速之 TTS，首先對三種速度(快、正常、慢)各自建立音長模型，對三個音長模型以內插方式來產生目標語速所需之音長，最後結合於 HMM 為基礎之語音合成器，此方法僅考慮韻律之中的音長部份，並未對其他韻律參數進行語速影響調整，且由於不同語速需各自建立自己的音長模型，會使得模型參數量大增；再則它使用內插法去產生音長，無法獲得準確的語速控制；另有文獻對正常及快速語料分別建立 HSMM 模型，再以 CMLLR 對音長模型進行音長平均值的語速調適，該方法僅考慮韻律之中的音長部份，且由於不同語速需各自建立自己的音長模型，會使得模型參數量大增；及有進行大規模主觀測試三種語速控制的方法研究，分別為：(1)針對目標語速選取相近語速之語料來訓練 HMM 模型，(2)依比例去伸縮合成語句的發音長度，及(3)基於 ML 準則去決定狀態長度(state duration)，這些方法都是建立於 HMM-based 的語音合成系統，實驗結果發現方法(2)

最適合用於快語速合成語音，而方法(1)較適合慢速語音，不同的語速控制方法都只適於某種語速，並沒有一種方法能掌握所有語速的控制。

因此，可知習知技術大多以等比例拉長或縮短各個合成單元(如音節、詞)之長度來達到語速控制，而於韻律結構、音高軌跡、停頓時間長度及停頓出現頻率方面較少著墨，並無考慮聲學韻律訊息其背後的產生模型，因此並不能以系統化的方式掌握語速對於韻律多層面的影響，進而用以產生韻律訊息；這些韻律訊息可充分應用於語音合成之語速控制，使各種語速之合成語音應用在語音合成之領域聽起來都很流利自然。

爰是之故，申請人有鑑於習知技術之缺失，乃經悉心試驗與研究，並一本鍥而不捨的精神，終發明出本案「韻律訊息產生器及語速相依之階層式韻律模組」，用以改善上述習用手段之缺失。

【發明內容】

本案之一面向係提供一韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一語言參數；一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之韻律聲學特徵參數。

本案之另一面向係提供一種語速相依之階層式韻律模組，包含至少二模型，其中各該模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

本案之又一面向係提供一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

【實施方式】

本發明將可由以下的實施例說明而得到充分瞭解，使得熟習本技藝之人士可以據以完成之，然本案之實施並非可由下列實施案例而被限制其實施型態。

本發明建立一個可應用於電子有聲書、手機、PDA 及電腦等裝置上之考慮語速對於音長、音高軌跡、停頓時機、停頓出現頻率、停頓時長所造成的影響之語速相依之階層式韻律模組；以及基於語速相依之階層式韻律模組，可以產生出各種語速的韻律聲學特徵參數，幫助語音合成達到良好的語速控制效果。其步驟包含兩個階段：模組建立及韻律合成。模組建立階段請參閱第一圖，其中是以階層式韻律模型為基礎建立語速相依之階層式韻律模組。請參閱第二圖，韻律合成階段是語速相依之韻律聲學特徵參數產生方法階段，其中是以語速相依之階層式韻律模組為基礎，產生語音合成所需要的各種韻律聲學特徵參數，且滿足不同語速之要求。

如前所述之模組建立階段之步驟包含對語音語料庫 101 中的每一句話，先做音節切割，再由每一音節抽取韻律聲學特徵參數；接著由語速估計 102 求取音節平均長度做為語速 SR；然後由語速正規化函式之建構 103 依據整個語音資料庫各句話的韻律聲學特徵參數對語速的統計分布來求取正規化函式；接著由韻律聲學特徵參數之語速正規化 104 來將韻律聲學特徵參數做正規化而獲得正規化韻律聲學特徵參數，再由修正型階層式韻律模型訓練演算法 105 使用整個語音語料庫每一句話的語速、語言參數、及正規化韻律聲學特徵參數來訓練獲得語速相依之階層式韻律模組 106，其中修正型階層式韻律模型訓練演算法，考慮語速之進一步影響，修正原本的階層式韻律模型訓練演算法，將其中兩個子模型：停頓語法模型及韻律狀態模型，加入語速考量，藉此補償語速對停頓時機(或出現頻率)、以及韻律狀態轉移所造成之影響。

如前所述之韻律合成階段之步驟包含:首先由文字分析器 201 將輸入文字做斷詞及詞類標記分析，獲得語言參數，再由語速相

依之韻律聲學特徵參數產生方法 202 使用語言參數、語速、語速相依之階層式韻律模組 204、以及語速正規化函式 203 來產生四種韻律聲學特徵參數。語速相依之階層式韻律模組 204 主要是決定整個語句的韻律架構(依據語速)及基本韻律參數合成，而語速正規化函式 203 是將基本韻律參數的統計特性調到指定語速的統計特性。

請參閱表一及第三圖，其分別為本發明中使用語料庫大小之統計資訊及語料庫語速之統計分佈圖。該語料庫是以一女性專業播音員依四種語速所錄製之平行語音資料庫當作實施目標，由該圖中可知四種語速所錄製之平行語音資料語速分佈在 0.15-0.3second/syllable 之間。

表一

	語句數	音節數	小時數
快語速	368	50691	3.4
一般語速	376	51868	3.9
中等語速	362	49956	4.8
慢語速	372	51231	6.0

對於韻律聲學特徵參數的正規化函式建構方法，其中一般正規化方法是對每個語句各自的資料統計參數做正規化，該方法簡易且具有效率，但可能造成過度正規化，導致除了語速之外的其它影響因素亦被調整而扭曲，進而使模組建造錯誤。本發明採用一較合理之正規化方法，即使用平滑曲線去模擬每個語句的正規化參數與語速的關係，藉由這些平滑曲線來形成語速正規化函式。

對於韻律聲學特徵參數中的音節長度，採取高斯正規化的方法，並使用二階多項式曲線來模擬音節長度的標準差，如下列式子所示：

$$sd_n' = \left(sd_n - \mu_k^{sd}\right) \Big/ \tilde{\sigma}^{sd}(SR(k)) \times \sigma_g^{sd} + \mu_g^{sd}$$

其 中

$$\tilde{\sigma}^{sd}(SR) = a_1 (SR)^2 + b_1 \cdot SR + c_1$$

為平滑化後的標準差，$\mu_k^{sd}=SR(k)$為語句k之音節平均長度(也就是語速)，sd_n和sd_n'分別代表原始音節長度和語速正規化之音節長度；μ_g^{sd}和σ_g^{sd}為語料庫整體的音節長度平均值與標準差。

對於停頓長度，使用Gamma分佈來表示其分佈，同樣使用二階多項式曲線來模擬語句之停頓長度平均值與標準差對語速SR的關係，其數學式子如下：

$$\tilde{\mu}^{pd}(SR)=a_2(SR)^2+b_2\cdot SR+c_2$$

$$\tilde{\sigma}^{pd}(SR)=a_3(SR)^2+b_3\cdot SR+c_3$$

接著利用平滑化的平均值$\tilde{\mu}^{pd}(SR(k))$和標準差$\tilde{\sigma}^{pd}(SR(k))$去對停頓長度$pd_n$做分佈正規化，其使用之公式為：

$$pd_n'=G^{-1}(G(pd_n,\tilde{\alpha}^{pd}(SR(k)),\tilde{\beta}^{pd}(SR(k))),\alpha_g^{pd},\beta_g^{pd})$$

其中$G(pd,\alpha,\beta)$為Gamma分佈的累積分佈函數(cumulative distribution function)，G^{-1}為G的反函數；

$$\tilde{\alpha}^{pd}(SR(k))=(\tilde{\mu}^{pd}(SR(k)))^2/(\tilde{\sigma}^{pd}(SR(k)))^2$$

和

$$\tilde{\beta}^{pd}(SR(k))=(\tilde{\sigma}^{pd}(SR(k)))^2/\tilde{\mu}^{pd}(SR(k))$$

為Gamma函數的兩個參數的平滑值，α_g^{pd}和β_g^{pd}為語料庫整體的停頓長度平均值和標準差。

對於音節音高軌跡，先進行正交展開(orthogonal expansion)，使用四個Legendre多項式為基底，用所得到的四維正交參數來表示基頻軌跡，即$\mathbf{sp}_n=\left[a_n^0\ a_n^1\ a_n^2\ a_n^3\right]^T$，接著依每一詞彙聲調(lexicon tone)之每一維度來正規化SR

對 $\mathbf{sp}_n$ 的影響，公式如下：

$$\mathbf{sp}_n'(i)=\frac{\mathbf{sp}_n(i)-\tilde{\mu}^{sp}(SR(k),t_n,i)}{\tilde{\sigma}^{sp}(SR(k),t_n,i)}\times\sigma_g^{sp}(t_n,i)+\mu_g^{sp}(t_n,i)$$

其中

$$\tilde{\mu}^{sp}(SR,t,i)=b_4(t,i)\cdot SR+c_4(t,i)$$

$$\tilde{\sigma}^{sp}(SR,t,i)=b_5(t,i)\cdot SR+c_5(t,i)$$

分別為 sp 第 i 維、第 t 聲調的平滑化平均值與標準差，它們都以一階函數來表示；$\mu_g^{sp}(t,i)$ 和 $\sigma_g^{sp}(t,i)$ 為整個語料庫的 sp 第 i 維、第 t 聲調的平均值與標準差。

對於音節能量位階，由於該與錄音條件有很大的相關性，包含麥克風與語者距離、麥克風本身的錄音品質、錄音的環境等等因素之影響遠遠大於語速所造成的，因此本實施案例採取非語速相依的高斯正規化。

在完成參數正規化後，再對所有訓練語句以實施方塊 105 修正型階層式韻律模型訓練演算法來自動產生一個語速相依之階層式韻律模組，該模組包括四個子模型，用來描述觀察到的韻律聲學特徵參數、語言參數及韻律階層架構標記之間的關係。雖然我們在之前參數正規化時已把語速對韻律聲學特徵參數之影響做適當補償消除，但停頓出現的頻率及韻律狀態的轉移仍與語速有很大的相關性，因此我們以決策樹描述七種停頓類型(請參閱第四圖)出現頻率與語言參數之間的關係來修正停頓語法子模型；以及使用一階馬可夫模型來描述前一個韻律狀態和目前韻律狀態之間的轉移關係來修正韻律狀態子模

型，使所述之二個子模型與語速相依。修正型韻律模型訓練演算法為一參數最佳化問題求解的方法，在已知正規化韻律聲學特徵參數{X, Y, Z}、語言參數{L}及語速SR之情況下找到最佳的韻律標記序列T={B,PS}，即下列數學式子：

$$\mathbf{B}^*,\mathbf{PS}^* = \arg\max_{\mathbf{B},\mathbf{P}} P(\mathbf{B},\mathbf{PS} \mid \mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{L},SR)$$
$$\approx \arg\max_{\mathbf{B},\mathbf{P}} \underbrace{P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{PS}|\mathbf{B},SR)P(\mathbf{B}|\mathbf{L},SR)}_{\text{語速相依之階層式韻律模型}}$$

其中B代表停頓標記序列，PS = {p, q, r}分別為音節基頻、長度及能量位階的韻律狀態標記序列，此兩類韻律標記是用來描述第四圖所考量的中文韻律階層結構，此結構包含四種韻律成分：音節、韻律詞、韻律片語、及呼吸或韻律片語群組；韻律停頓B_n是用來描述音節n和音節n+1之間的停頓狀態，共使用七種韻律停頓狀態來描述此四種韻律成分的邊界；A={X,Y,Z}為韻律聲學特徵參數序列，其中X={sp, sd, se}、Y={pd, ed}和Z={pj, dl, df}分別代表與音節相關的韻律聲學特徵參數、音節間及差分之韻律聲學特徵參數序列；L={POS,PM,WL,t,s,f}為語言參數序列，其中{POS,PM,WL}為高階語言參數序列，POS、PM及WL分別為詞類序列、標點符號序列及詞長序列，而{t,s,f}為低階語言參數序列，t、s及f分別為聲調、基本音節類別及韻母類別序列；SR為語句之語速。詳細符號定義請參閱表二。

表 二

T: 韻律標籤	**B**: 停頓類型	
	P: 韻律狀態	**p**: 基頻韻律狀態 **q**: 時長韻律狀態 **r**: 能量位階韻律狀態
A: 韻律聲學特徵參數	**X**: 音節韻律聲學特徵參數	**sp**: 音節基頻軌跡 **sd**: 音節時長 **se**: 音節能量位階
	Y: 音節間韻律聲學特徵參數	**pd**: 停頓時長 **ed**: 能量低點位階
	Z: 音節差分韻律聲學特徵參數	**pj**: 正規化基頻跳躍 **dl**: 正規化時長拉長因子 1 **df**: 正規化時長拉長因子 2
SR: 語速		
L: 語言參數	**POS**: 詞類 **PM**: 標點符號 **WL**: 詞長	
	t: 聲調 **s**: 基本音節類型 **f**: 韻母類型	

語速相依之階層式韻律模組可以下列方程式表示 $P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{PS}|\mathbf{B},SR)P(\mathbf{B}|\mathbf{L},SR)$。該模組包含四個子模型：音節韻律聲學模型 $P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L})$、停頓聲學模型 $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$、韻律狀態模型 $P(\mathbf{PS}|\mathbf{B},SR)$ 以及停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)$：

(1) 音節韻律聲學模型 $P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L})$：

如下式所示，它再以三個子模型來近似：

$$P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L}) \approx P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{t})P(\mathbf{sd}|\mathbf{B},\mathbf{q},\mathbf{t},\mathbf{s})P(\mathbf{se}|\mathbf{B},\mathbf{r},\mathbf{t},\mathbf{f})$$
$$\approx \prod_{n=1}^{N} P(\mathbf{sp}_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})P(sd_n \mid q_n, s_n, t_n)P(se_n \mid r_n, f_n, t_n)$$

其中子模型 $P(\mathbf{sp}_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})$、$P(sd_n \mid q_n, s_n, t_n)$ 以及 $P(se_n \mid r_n, f_n, t_n)$ 分別代表第 n 個音節的音高軌跡、音節長度、能量位階之模型，t_n、s_n 及 f_n 分別表示第 n 個音節的聲調、基本音節、及韻母類型；$B_{n-1}^{n} = (B_{n-1}, B_{n+1})$；和 $t_{n-1}^{n+1} = (t_{n-1}, t_n, t_{n+1})$。

在本實施例中，這三個子模型各考慮了多個影響因子(Affecting Factors, AFs)，這些影響因子以加成方式結合，以第 n 個音節的音高軌跡為例，我們可得：

$$\mathbf{sp}_n = \mathbf{sp}_n^r + \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu}_{sp}$$

其中 $\mathbf{sp}_n = [\alpha_{0,n}, \alpha_{1,n}, \alpha_{2,n}, \alpha_{3,n}]$ 為一四維正交化係數向量，用以表示第 n 個音節觀察到的音高軌跡，$\mathbf{sp}_n^r$ 為正規化後的殘餘值，$\boldsymbol{\beta}_{t_n}$ 和 $\boldsymbol{\beta}_{p_n}$ 分別為聲調和韻律狀態兩影響因子(AF)的影響數值(Affecting Factor, AF)，$\boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}}$ 和 $\boldsymbol{\beta}^b_{B_n,tp_n}$ 為向前及向後連音兩 AF 的影響數值；$tp_{n-1} = t^n_{n-1}$；μ_{sp} 為音高的全域平均值。基於假設 $\mathbf{sp}_n^r$ 為零平均值之高斯常態分佈，我們可以高斯常態分佈來表示 $\mathbf{sp}_n$ 如下所示

$$P(\mathbf{sp}_n \mid B^n_{n-1}, p_n, t^{n+1}_{n-1}) = N(\mathbf{sp}_n; \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu}_{sp}, R_{sp})$$

其中 $N(\mathbf{x};\boldsymbol{\mu},R)$ 表示向量 x 為 mean vector $\boldsymbol{\mu}$ 及 covariance matrix R 之常態分佈。

音節長度 $P(sd_n \mid q_n, s_n, t_n)$ 及能量位階 $P(se_n \mid r_n, f_n, t_n)$ 亦是以此方式去實現：

$$P(sd_n \mid q_n, s_n, t_n) = N(sd_n; \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n} + \mu_{sd}, R_{sd})$$

$$P(se_n \mid r_n, f_n, t_n) = N(se_n; \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se}, R_{se})$$

其中 γ_x 及 ω_x 分別代表音節長度以及音節能量位階受影響因素 x 的影響數值(AF)。

(2) 停頓聲學模型 $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$：

音節間韻律聲學模型則以五個子模型近似之，如下式所示：

$$P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L}) \approx P(\mathbf{pd},\mathbf{ed},\mathbf{pj},\mathbf{dl},\mathbf{df}|\mathbf{B},\mathbf{L}) \approx \prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n, df_n | B_n, L_n)$$

$$\approx \prod_{n=1}^{N-1} \{ g(pd_n; \alpha_{B_n,L_n}, \eta_{B_n,L_n}) N(ed_n; \mu_{ed,B_n,L_n}, \sigma^2_{ed,B_n,L_n}) \cdot N(pj_n; \mu_{pj,B_n,L_n}, \sigma^2_{pj,B_n,L_n})$$

$$\cdot N(dl_n; \mu_{dl,B_n,L_n}, \sigma^2_{dl,B_n,L_n}) N(df_n; \mu_{df,B_n,L_n}, \sigma^2_{df,B_n,L_n}) \}$$

其中在第 n 個音節所跟隨的接合點(juncture n，之後以第 n 個接合點表示)的停頓長度 pd_n 以 Gamma 分佈模擬，ed_n 為第 n 個接合點的能量低點位階；pj_n 為跨越第 n 個接合點的正規化音高差，其定義如下：

$$pj_n = (\mathbf{sp}_{n+1}(1) - \chi_{t_{n+1}}) - (\mathbf{sp}_n(1) - \chi_{t_n})$$

其中 $\mathbf{sp}_n(1)$ 為 $\mathbf{sp}_n$ 的第一維度(即音節音高平均值)，χ_t 為聲調 t 平均音高位階；dl_n 及 df_n 分別為跨越第 n-1 及第 n 個接合點的兩個正規化的音節拉長因子，其定義如下：

$$dl_n = (sd_n - \pi_{t_n} - \pi_{s_n}) - (sd_{n-1} - \pi_{t_{n-1}} - \pi_{s_{n-1}})$$

$$df_n = (sd_n - \pi_{t_n} - \pi_{s_n}) - (sd_{n+1} - \pi_{t_{n+1}} - \pi_{s_{n+1}})$$

其中 π_x 代表影響因素 x 的平均音長。除了 pd_n 以 Gamma 分佈模擬外，其他四種模型皆以常態分佈模擬；因為對韻律停頓而言 L_n 的參數空間仍是太大，可以使用 CART (Classification And Regression Trees) 決策樹分類法將 L_n 分成幾類，然後同時估計 Gamma 及常態分佈的參數。

(3) 韻律狀態模型 $P(\mathbf{PS}|\mathbf{B}, SR)$

韻律狀態模型 $P(\mathbf{PS}|\mathbf{B}, SR)$ 以三個子模型近似之，分別用來模擬音節音高、長度及能量三種韻律狀態，並以語速等分成小段 bin 來區

分不同語速所造成的影響，如下式所示：

$$P(\mathbf{P}\mid\mathbf{B},SR)=P(\mathbf{p}|\mathbf{B},SR)P(\mathbf{q}|\mathbf{B},SR)P(\mathbf{r}|\mathbf{B},SR)$$
$$\approx P(p_1|\text{bin}(SR(k)))P(q_1|\text{bin}(SR(k)))P(r_1|\text{bin}(SR(k)))$$
$$\cdot\left[\prod_{n=2}^{N}P(p_n\mid p_{n-1},B_{n-1},\text{bin}(SR(k)))P(q_n\mid q_{n-1},B_{n-1},\text{bin}(SR(k)))P(r_n\mid r_{n-1},B_{n-1},\text{bin}(SR(k)))\right]$$

其中，p_n, q_n, r_n表示音節 n 的音高、長度及能量韻律狀態；$\text{bin}(SR(k))$為語句 k 的語速 $SR(k)$所屬的小段(bin)。

(4) 停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)$

停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)\cong\prod_n P(B_n\mid L_n,SR(k))$由兩個步驟建構成，第一步先由 CART 決策樹分析演算法來估計 $P(B_n\mid L_n)$，第二步再使用多項式曲線來模擬 7 種停頓類型在每個決策樹子結點的出現頻率和語速 SR 的關係，最後估計出 $P(B_n\mid L_n,SR)$，其公式如下所示：

$$P(B_n=m\mid L_n,SR(k))=\frac{P(B_n=m\mid L_n,SR(k))}{\sum_{x\in\text{all break types}}P(B_n=x\mid L_n,SR(k))}\approx\frac{c_{m,j}SR(k)+d_{m,j}}{\sum_{x\in\text{all break type}}c_{x,j}SR(k)+d_{x,j}}$$

其中 Bn 為第 k 個語句第 n 個音節後的停頓類型，j 為決策樹子結點的索引值，Ln 為對應的語言參數向量，cm, j 和 dm, j 為停頓類型 m、子結點 j 的線性迴歸係數。

此修正型階層式韻律模式訓練演算法，在適當的韻律斷點和韻律狀態初始化後，是以依序最佳化程序(sequential optimization procedure)來訓練韻律模型，同時對於訓練語料以最大似然性法則(maximum likelihood criterion)來產生韻律標記及獲得語速相依之

階層式韻律模式之參數。

下列為該模組訓練之實驗結果。請參閱表三，其列出在使用不同影響因子組合下，各韻律聲學參數重建之總殘餘誤差值(Total Residual Error, TRE)，即扣除各種影響因子之AF組合後，韻律聲學特徵參數殘餘值變異數與原始韻律聲學特徵參數變異數之比值，其中，加入韻律狀態之AF後，各韻律聲學特徵參數之TRE都變得非常小。

表三

音節基頻軌跡		音節時長		音節能量位階	
影響因子	TRE	影響因子	TRE	影響因子	TRE
+ 語調	67.3%	+語調	70.6%	+ 語調	61.4%
+ 前後連音	63.2%	+ 基本音節類型	50.1%	+ 聲母類型	48.0%
+ 基頻韻律狀態	0.8%	+ 時長韻律狀態	1.4%	+ 基頻韻律狀態	1.9%

停頓時長為音節間韻律聲學子模型最重要的參數，請參閱第五圖，其顯示出七種停頓類別的平均值對語速的關係，其中在B0、B1、B2-1及B2-3四種不明顯停頓時長的類別，它們與語速相關性甚小，其餘停頓類別之停頓時長皆隨著SR呈非線性增加。而表四為對每種停頓類別計算重建停頓時長的均方根誤差，發現只有B2-2、B3及B4之誤差會比較大，這是因為這些停頓類別通常發生在MINOR BREAK 或MAJOR BREAK位置，因其變異較大所以重建誤差也自然較大，此結果是在合理的範圍。

表四

停頓類型	B0	B1	B21	B22	B23	B3	B4
均方根誤差	3 毫秒	19 毫秒	25 毫秒	90 毫秒	30 毫秒	104 毫秒	149 毫秒

請參閱第六圖，其是用聲調 AF 來產生快、慢兩種語速的音高軌跡，可觀察到每一聲調的基頻軌跡受語速的影響程度皆不盡相同。

請參閱表五，其顯示一個停頓類別的標記例子，此例子對四個不同語速的平行語料標記，在此只標示出 B4(@)、B3(/) 及 B2-2(*)三種具明顯停頓時長之類別，其顯示出語速越慢時越容易出現明顯類別的停頓，符合預期之結果。

表五

依據 行政院 主計處 的 統計 @，十月份 * 一 到 二十日 / ，我國 出口 及進口 金額 / 比起 去年 同期 * 均有 增加 @，
依據 行政院 主計處 的 統計 @，十月份 * 一 到 二十日 /，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加@，
依據 * 行政院 主計處 的 統計 @，十月份 / 一 到 * 二十日 /，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加 @，
依據 / 行政院 * 主計處 的 統計 @，十月份 / 一 * 到 * 二十日 @，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加 @，

上述各項實驗數據顯示該模組可有效地描述漢語語音韻律參數之各種變化。

對於可控制語速之韻律聲學特徵參數產生方法可經由參閱第七圖得到進一步瞭解，

其為第二圖的較詳細圖示，其是基於訓練出來的語速相依之階層式韻律模組 701 之可控制語速之漢語韻律聲學特徵參數產生法流程圖。方塊 702 為停頓標記預估器，其使用該韻律模型中的停頓語法模型來做停頓標記預估的方法：

$$B_n^* = \arg\max_{B_n} P(B_n \mid L_n, SR)$$

其中 L_n 為輸入的語言參數，SR 為指定的語速。

方塊 703 為韻律狀態標記預估器，其使用此韻律模型中的韻律狀態模型搭配一組額外的韻律狀態語法模型 705，以維特比演算法(Viterbi algorithm)來預估之，如以下數學式所示：

$$\mathbf{p}^*, \mathbf{q}^*, \mathbf{r}^* = \arg\max_{\mathbf{p},\mathbf{q},\mathbf{r}} \left(\begin{array}{l} P(p_1|\text{bin}(SR))P(q_1|\text{bin}(SR))P(r_1|\text{bin}(SR)) \\ \cdot \prod_{n=2}^{N} P(p_n \mid p_{n-1}, B_{n-1}^*, \text{bin}(SR))P(q_n \mid q_{n-1}, B_{n-1}^*, \text{bin}(SR))P(r_n \mid r_{n-1}, B_{n-1}^*, \text{bin}(SR)) \end{array} \right) \cdot \left(\prod_{n=1}^{N} P(p_n \mid L_n)P(q_n \mid L_n)P(r_n \mid L_n) \right)$$

其中 $p(p_n \mid L_n)$、$p(q_n \mid L_n)$、$p(r_n \mid L_n)$ 為韻律狀態語法模型，它們係使用做完韻律標記之訓練語料以 CART 演算法實現之，B_{n-1}^* 為停頓標記預估結果。

有了韻律標記預估結果後，可利用韻律模型中的的音節韻律聲學模型 $P(\mathbf{PS}|\mathbf{B},\mathbf{L})$ 和停頓聲學模型 $P(\mathbf{X},\mathbf{Y}|\mathbf{B},\mathbf{L})$ 來產生語速正規化之韻律聲學特徵參數，再藉由語速正規化函式 704 之反函式來還原產生指定語速之韻律聲學特徵參數，各韻律聲學特徵參數之產生說明如下：

語速控制的停頓時長產生方法為

$$pd_n' = \mathrm{G}^{-1}(\mathrm{G}(pd_n^*, \alpha_g^{pd}, \beta_g^{pd}), \tilde{\alpha}^{pd}(SR), \tilde{\beta}^{pd}(SR))$$

其中

$$pd_n^* \equiv \mu_n^* = \alpha_n^* \beta_n^*$$

為語速正規化之停頓時長，它使用停頓聲學模型中由 B_n^* 和前後文參數 L_n 所找到的節點的 Gamma 分布的參數 α_n^* 及 β_n^* 去計算的平均值 μ_n^* 來估計；語速控制的音節音高軌跡產生方法為

$$\mathbf{sp}_n'(i) = \frac{\mathbf{sp}_n^*(i) - \mu_g^{sp}(t_n, i)}{\sigma_g^{sp}(t_n, i)} \times \tilde{\sigma}^{sp}(SR, t_n, i) + \tilde{\mu}^{sp}(SR, t_n, i)$$

其中語速正規化之基頻軌跡 $\mathbf{sp}_n^*$ 的預估如下面數學式所示，它是以預估之韻律標記和聲調語言參數來挑選對應的 AFs 所疊加產生：

$$\mathbf{sp}_n^* = \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n^*} + \boldsymbol{\beta}^f_{B_{n-1}^*, t_{n-1}^n} + \boldsymbol{\beta}^b_{B_n^*, t_n^{n+1}} + \boldsymbol{\mu}_{sp}$$

語速控制的音節長度產生方法如下：

$$sd_n' = \left(sd_n^* - \mu_g^{sd}\right) / \sigma_g^{sd} \times \tilde{\sigma}^{sd}(SR) + \mu_k^{sd}$$

其中語速正規化之音節長度 sd_n^* 是以對應的 AFs 所疊加產生：

$$sd_n^* = \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n^*} + \mu_{sd}$$

最後音節能量位階的產生方法為

$$se_n^* = \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se}$$

以下為語音合成範例。本發明所產生的韻律聲學特徵參數能結合於任何語音合成

器，以達到語速控制之語音合成。在此以一隱藏式馬可夫為基礎之語音合成技術(HMM-based speech synthesis)為例將語音合成出來，此技術為習知技術，在此簡短說明其參數設定：中文的21個聲母及39個韻母都各以一個HMM表示，每個HMM包含5個HMM狀態，每一個狀態內的觀察向量包含兩個類別串：一個為維度75的頻譜參數，另一個為離散的事件來表示清音(unvoiced)或濁音(voiced)的狀態，每一個狀態皆以多變量單一高斯函數(multi-variate single Gaussian)表示其觀察機率。訓練HMM模型的方法是以習知方法(embedded-trained 及決策樹方法對HMM狀態分群)訓練其參數，上述之參數設定及訓練方法可視實際情況而調整，其並非用以限制本發明之範圍。

請參閱表六，其為MOS主觀聽覺評估結果，其係經由十五位測試者聆聽三種語速各十句所做主觀音質評定的MOS分數平均，由該表中可看出合成語音在不同語速皆有不錯的聲音品質。

表六

語速	快(*SR*=0.17)	中(*SR*=0.20)	慢(*SR*=0.25)
MOS	3.35	3.44	3.28

雖然本發明已以較佳實施例揭露如上，然其並非用以限定本發明之範圍，任何熟習此技藝者，在不脫離本發明之精神和範圍內，當可作各種更動與潤飾，因此本發明之

保護範圍當視後附之申請專利範圍所界定者為準。

實施例:

1. 一種可控制語速的韻律訊息產生裝置，包含：

一第一輸入端，用以接收一語速；

一第二輸入端，用以接收一語言參數；

一文字分析器，用以接收一文字，以產生一語言參數;

一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及

一輸出端，用以輸出與該語速相依之該韻律聲學特徵參數。

2. 如實施例 1 所述的裝置，其中該語速相依之韻律生成模組包含一語速相依之階層式韻律模組、一語速相依之韻律訊息產生器、以及至少一個預估器，其中各該預估器係選自由包含一停頓標記預估器及一韻律狀態預估器；

3. 如實施例 1-2 所述的裝置，其中該語速相依之階層式韻律模組包含一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型，俾與一語速相依。

4. 如實施例 1-3 所述的裝置，其中該停頓標記預估器，根據該語速、該語言參數和該語速相依之階層式韻律模組之停頓語法模型而執行一停頓標記預估操作，以產生一停頓標記預估結果。

5. 如實施例 1-4 所述的裝置，其中該韻律狀態預估器，根據該語

速、該語速相依之階層式韻律模組之韻律狀態模型、一韻律狀態語法模型和該停頓標記預估結果而執行一韻律狀態預估操作，以產生一韻律狀態預估結果。

6. 如實施例 1-5 所述的裝置，其中該語速相依之韻律訊息產生器，根據一語速正規化函式、該語速相依之階層式韻律模組之音節韻律聲學模型及停頓聲學模型、該韻律狀態預估結果、該停頓標記預估結果、該輸入語速及語言參數，以產生對應語速之韻律聲學特徵參數。

7. 如實施例 1-6 所述的裝置，其中該語速正規化函式用以調整韻律聲學特徵參數的統計特性成任一語速的統計特性; 其所使用的正規化參數係採用整體語料的統計分佈經平滑化而得到。

8. 如實施例 1-7 所述的裝置，其中該語言參數至少包含兩參數，其中各該參數係選自由包含詞類、標點符號、詞長、聲調、基本音節類型及韻母類型及其組合之一。

9. 一種語速相依之階層式韻律模組，包含至少二子模型，其中各該子模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

10. 如實施例 9 所述的模組，其中該語速相依之階層式韻律模組根據一漢語語音資料庫之語言參數、正規化韻律聲學參數及各語句的語速，再以一修正型階層式韻律模型訓練演算法來建構。

11. 如實施例 9-10 所述的模組，其中該正規化韻律聲學參數根據各

語句之語速，使用語速正規化函式對韻律聲學參數做一正規化操作所產生。

12. 如實施例 9-11 所述的模組，其中該音節韻律聲學模型、該停頓聲學模型、該韻律狀態模型及該停頓語法模型各包含至少兩種的子模型來建構。

13. 如實施例 9-12 所述的模組，其中該修正型階層式韻律模型訓練演算法亦施用於至少一停頓語法子模型與一韻律狀態子模型。

14. 如實施例 9-13 所述的模組，該語速相依之階層式韻律模組根據一輸入語速、一輸入語言參數於該模組中，以產生相對應之一停頓類型機率用以協助停頓標記之預估、一韻律狀態機率用以協助韻律狀態之預估、一音節韻律聲學特徵參數機率及一音節間停頓時長之機率用以協助產生一語速相依之韻律聲學特徵參數。

15. 一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

16. 如實施例 15 所述的模組，其中該韻律狀態包含基頻韻律狀態、時長韻律狀態及能量位階韻律狀態。

17. 如實施例 15-16 所述的模組，其中該音節韻律聲學特徵參數包含音節基頻軌跡、音節時長及音節能量位階；

該音節間韻律聲學特徵參數包含停頓時長及能量低點位階；及

該音節差分韻律聲學特徵參數包含基頻跳躍、時長拉長因子 1 及時長拉長因子 2。

18. 如實施例 15-17 所述的模組，其中根據所產生的語速相依之韻律聲學特徵參數，可使用習知之語音合成器來合成出相對應之任一指定語速之合成語音。

【圖式簡單說明】

第一圖：本案一較佳實施例之架構語速相依之階層式韻律模組流程圖。

第二圖：本案一較佳實施例之產生語速相依之韻律聲學特徵參數簡易流程圖。

第三圖：本案一較佳實施例之語料庫語速統計圖。

第四圖：本案一較佳實施例之漢語語音階層式韻律結構示意圖。

第五圖：本案一較佳實施例之七種停頓類別的停頓時長平均值對語速之關係圖。

第六圖(a)~(b)：本案一較佳實施例之不同聲調之基頻軌跡於不同語速之差異圖。

第七圖：本案一較佳實施例之產生語速相依之韻律聲學特徵參數流程圖。

【主要元件符號說明】

101：語音資料庫

102：語速估計

103：語速正規化函式之建構

104：韻律聲學特徵參數之語速正規化

105：修正型階層式韻律模型訓練演算法

106：語速相依之階層式韻律模組

201：文字分析器

202：語速相依之韻律參數產生方法
203：語速正規化函式
204：語速相依之階層式韻律模組
701：語速相依之階層式韻律模組
702：停頓標記預估器
703：韻律狀態預估器
704：語速正規化函式
705：韻律狀態語法模型
706：語速相依之韻律訊息產生器
707：語速相依之韻律生成模組

七、申請專利範圍：

1. 一種可控制語速的韻律訊息產生裝置，包含：

一第一輸入端，用以接收一語速；

一第二輸入端，用以接收一語言參數；

一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及

一輸出端，用以輸出與該語速相依之該韻律聲學特徵參數。

2. 如申請專利範圍第 1 項所述的裝置，其中該語速相依之韻律生成模組包含一語速相依之階層式韻律模組、一語速相依之韻律訊息產生器、以及至少一個預估器，其中各該預估器係選自由包含一停頓標記預估器及一韻律狀態預估器；

3. 如申請專利範圍第 1 項所述的裝置，其中該語速相依之階層式韻律模組包含一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型，俾與一語速相依。

4. 如申請專利範圍第 3 項所述的裝置，其中該停頓標記預估器，根據該語速、該語言參數和該語速相依之階層式韻律模組之停頓語法模型而執行一停頓標記預估操作，以產生一停頓標記預估結果。

5. 如申請專利範圍第 3 項所述的裝置，其中該韻律狀態預估器，根據該語速、該語速相依之階層式韻律模組之韻律狀態模型、一韻律狀態語法模型和該停頓標記預估結果而執行一韻律狀態預估操作，以產生一韻律狀態預估結果。

6. 如申請專利範圍第 2 項所述的裝置，其中該語速相依之韻律訊息產生器，根據一語速正規化函式、該語速相依之階層式韻律模組之音節韻律聲學模型及停頓聲學模型、該韻律狀態預估結果、該停頓標記預估結果、該輸入語速及該語言參數，以產生一對應語速之韻律聲學特徵參數。

7. 如申請專利範圍第 6 項所述的裝置，其中該語速正規化函式用以調整韻律聲學特徵參數的統計特性成任一語速的統計特性；其所使用的正規化參數係採用整體語料的統計分佈經平滑化而得到。

8. 如申請專利範圍第 6 項所述的裝置，其中該語言參數至少包含兩參數，其中各該參數係選自由包含詞類、標點符號、詞長、聲調、基本音節類型及韻母類型及其組合之一。

9. 一種語速相依之階層式韻律模組，包含至少二子模型，其中各該子模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

10. 如申請專利範圍第 9 項所述的模組，其中該語速相依之階層式韻律模組根據一漢語語音資料庫之語言參數、一正規化韻律聲學參數及各語句的語速，再以一修正型階層式韻律模型訓練演算法來建構。

11. 如申請專利範圍第 9 項所述的模組，其中該音節韻律聲學模型、該停頓聲學模型、該韻律狀態模型及該停頓語法模型各包

含至少兩種的子模型來建構。

12. 如申請專利範圍第 10 項所述的模組，其中該正規化韻律聲學參數根據各語句之語速，使用語速正規化函式對韻律聲學參數做一正規化操作所產生。

13. 如申請專利範圍第 10 項所述的模組，其中該修正型階層式韻律模型訓練演算法亦施用於至少一停頓語法子模型與一韻律狀態子模型。

14. 如申請專利範圍第 10 項所述的模組，該語速相依之階層式韻律模組根據一輸入語速、一輸入語言參數於該模組中，以產生相對應之一停頓類型機率用以協助停頓標記之預估、一韻律狀態機率用以協助韻律狀態之預估、一音節韻律聲學特徵參數機率及一音節間停頓時長之機率用以協助產生一語速相依之韻律聲學特徵參數。

15. 一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

16. 如申請專利範圍第 15 項所述的模組，其中該韻律狀態包含基頻韻律狀態、時長韻律狀態及能量位階韻律狀態。

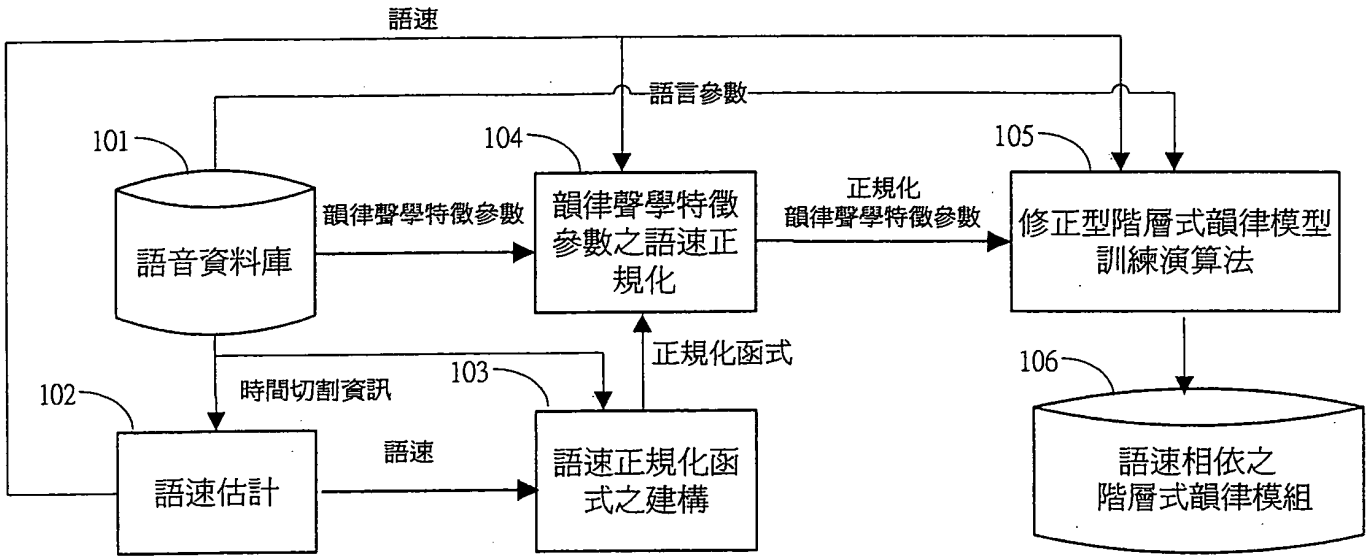
17. 如申請專利範圍第 15 項所述的模組，其中該音節韻律聲學特徵參數包含音節基頻軌跡、音節時長及音節能量位階；

該音節間韻律聲學特徵參數包含停頓時長及能量低點位階；及

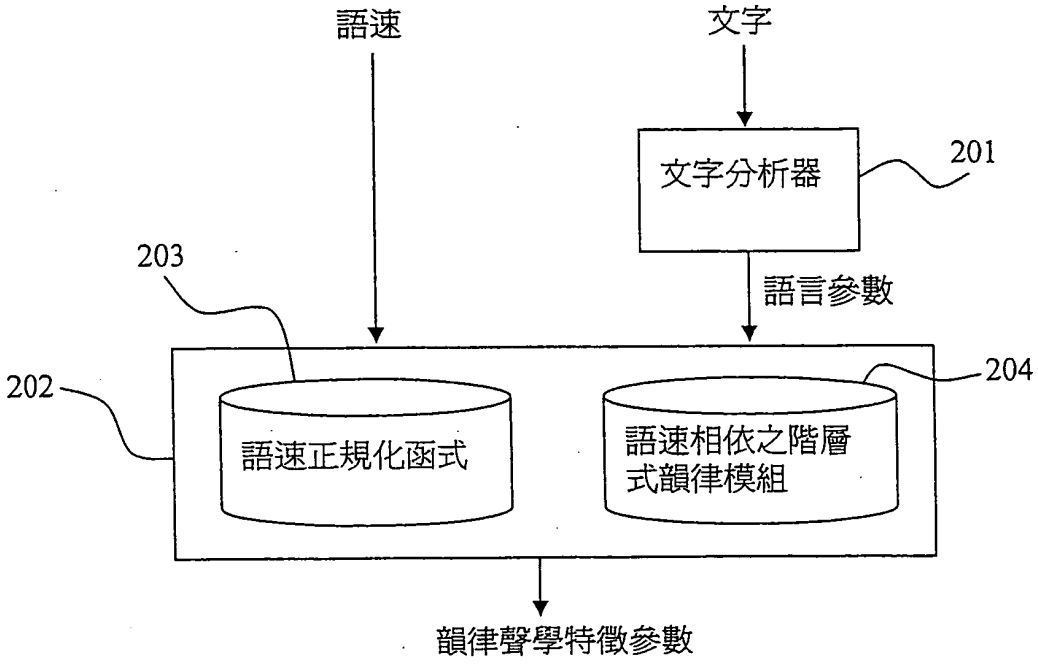
該音節差分韻律聲學特徵參數包含基頻跳躍、時長拉長因子 1 及時長拉長因子 2。

18. 如申請專利範圍第 15 項所述的模組，其中根據所產生的語速相依之韻律聲學特徵參數，可使用習知之語音合成器來合成出相對應之任一指定語速之合成語音。

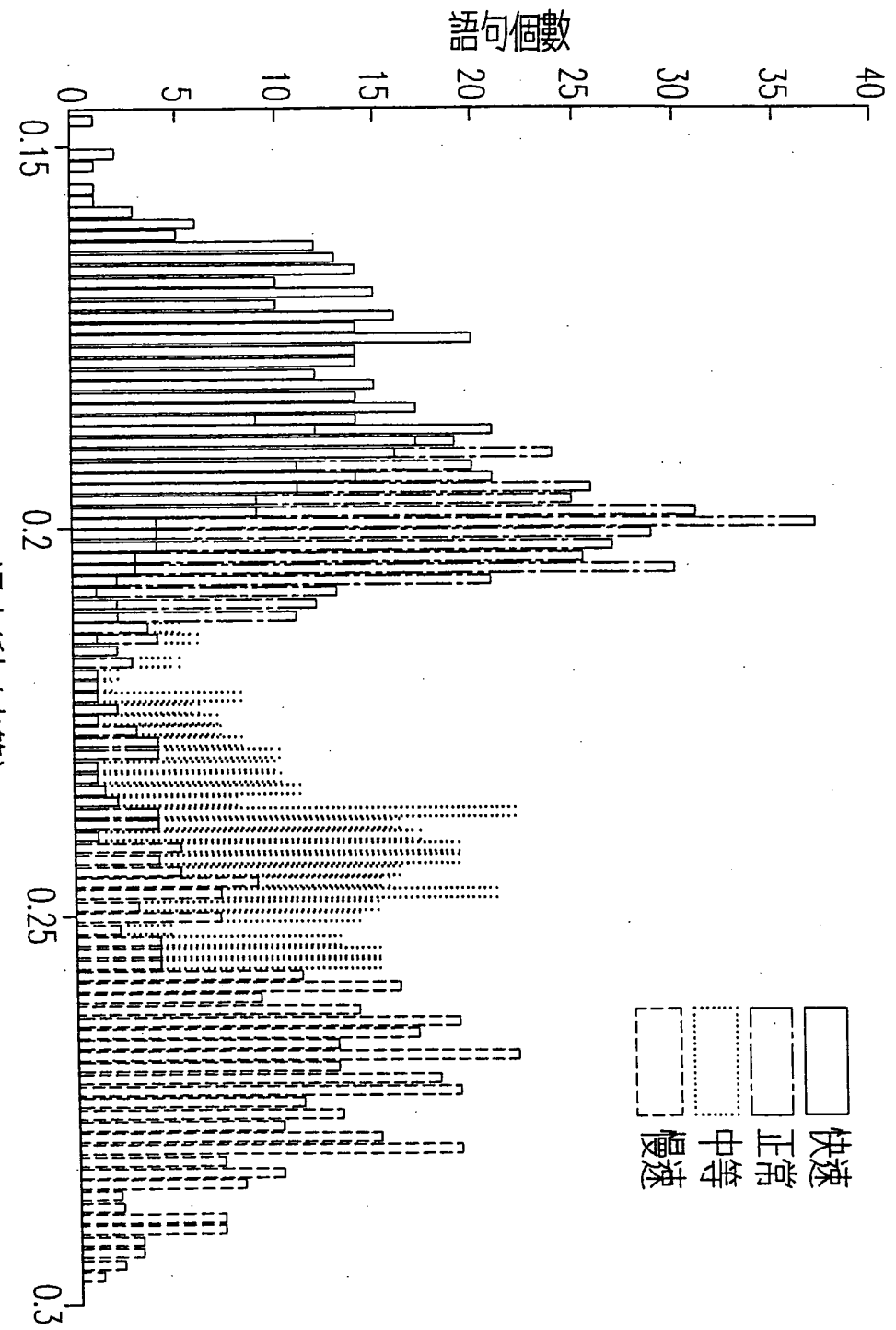
八、圖式：


語速
語言參數
101
104
105
語音資料庫
韻律聲學特徵參數
韻律聲學特徵參數之語速正規化
正規化
韻律聲學特徵參數
修正型階層式韻律模型訓練演算法
正規化函式
102
103
106
時間切割資訊
語速
語速估計
語速正規化函式之建構
語速相依之階層式韻律模組

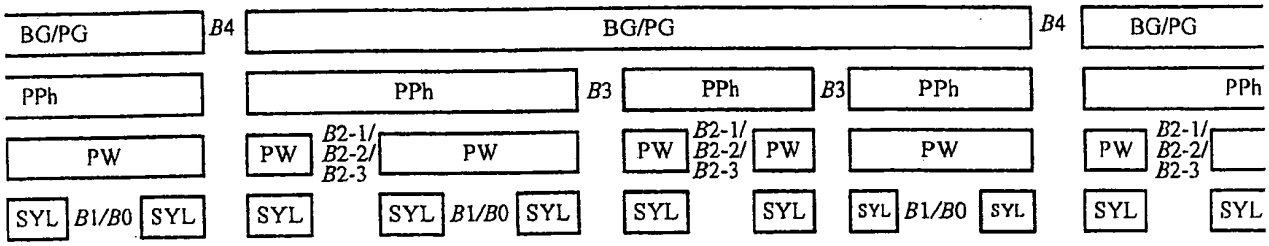
第一圖


語速
文字
文字分析器
201
203
語言參數
202
204
語速正規化函式
語速相依之階層式韻律模組
韻律聲學特徵參數

第二圖



第二圖



SYL: 音節

PW: 韻律詞

PPh: 韻律片語

BG/PG: 呼吸或韻律片語群組

B4: 呼吸或韻律片語群組邊界韻律斷點

B3: 韻律片語邊界韻律斷點

B2-1: 第一類韻律詞韻律斷點，表示音高重置

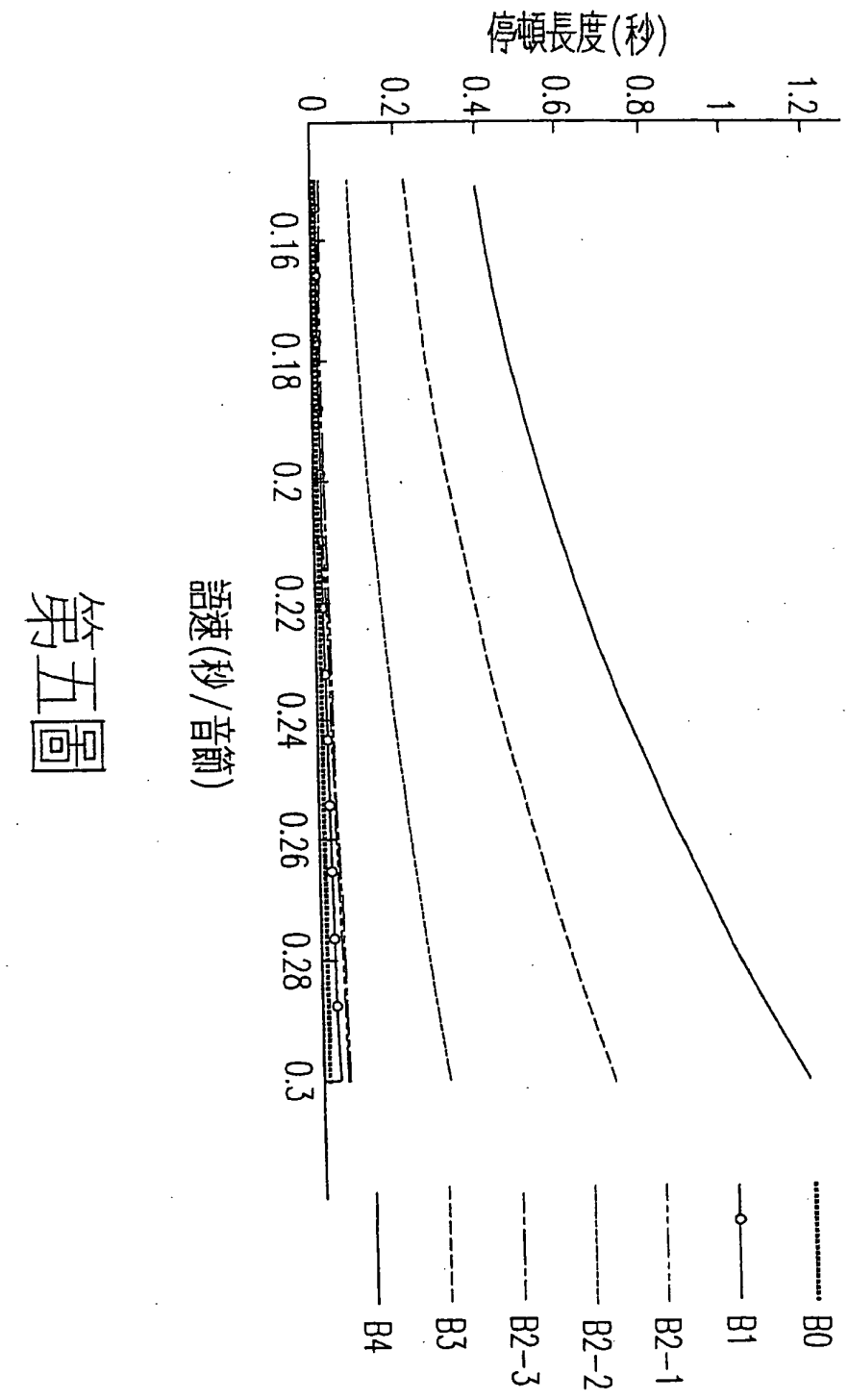
B2-2: 第二類韻律詞韻律斷點，表示短靜音停頓

B2-3: 第三類韻律詞韻律斷點，表示音節拉長停頓

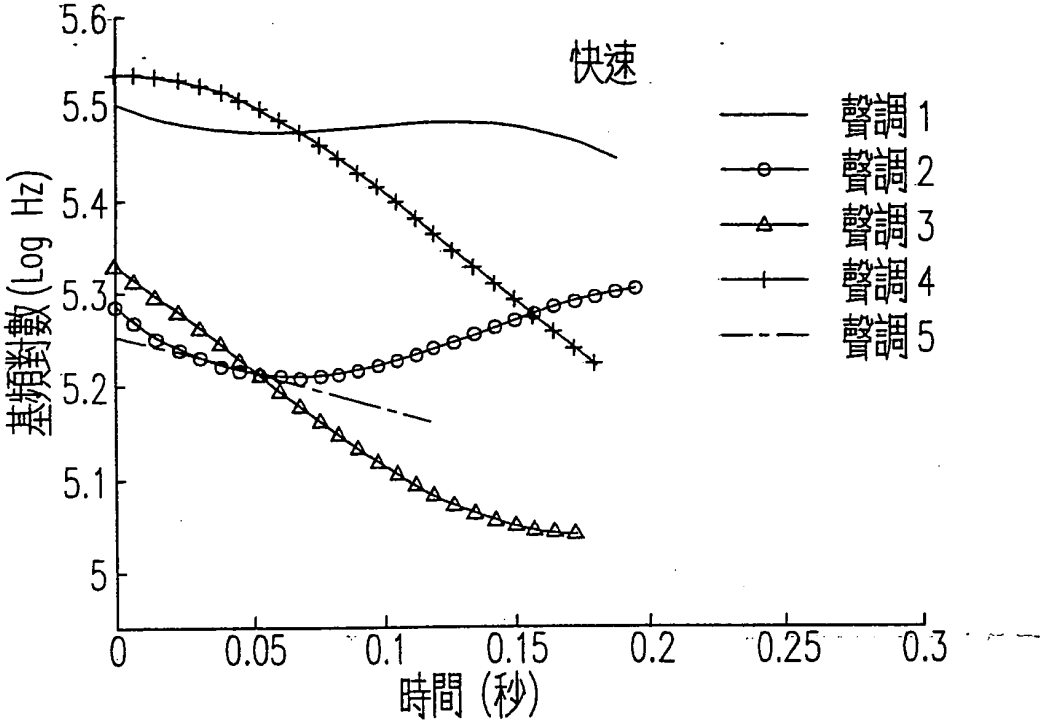
B1: 韻律詞內正常韻律斷點

B0: 韻律詞內強連音韻律斷點

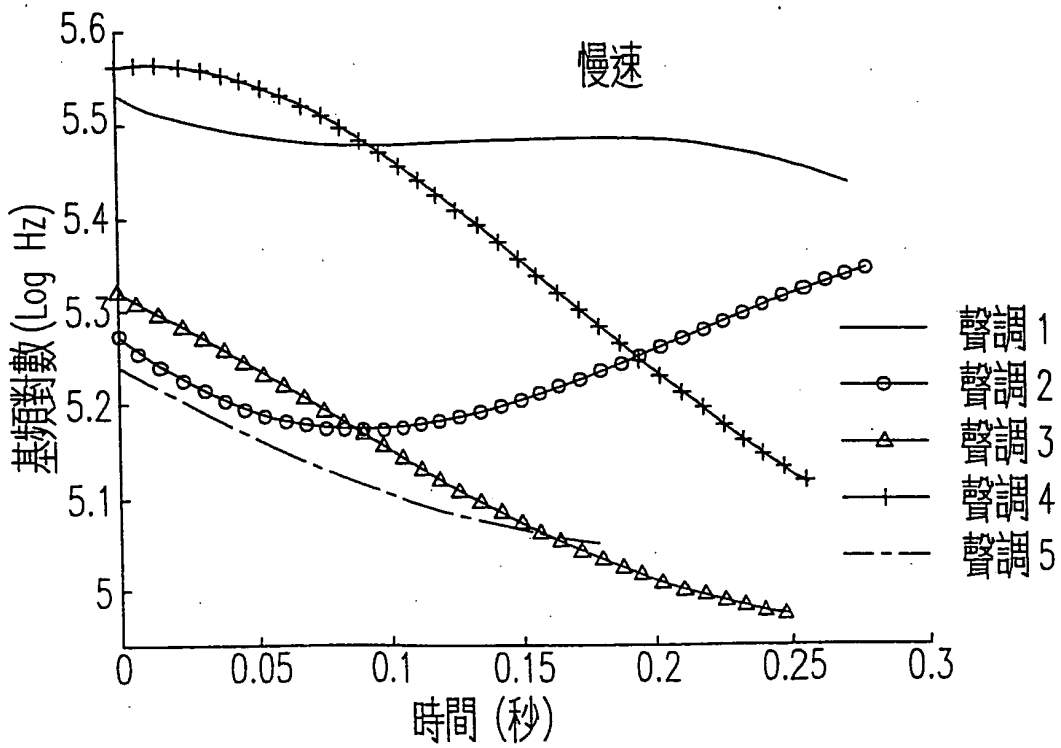
第四圖


停頓長度(秒)
0
0.2
0.4
0.6
0.8
1
1.2
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.3
語速(秒/音節)
B0
B1
B2-1
B2-2
B2-3
B3
B4

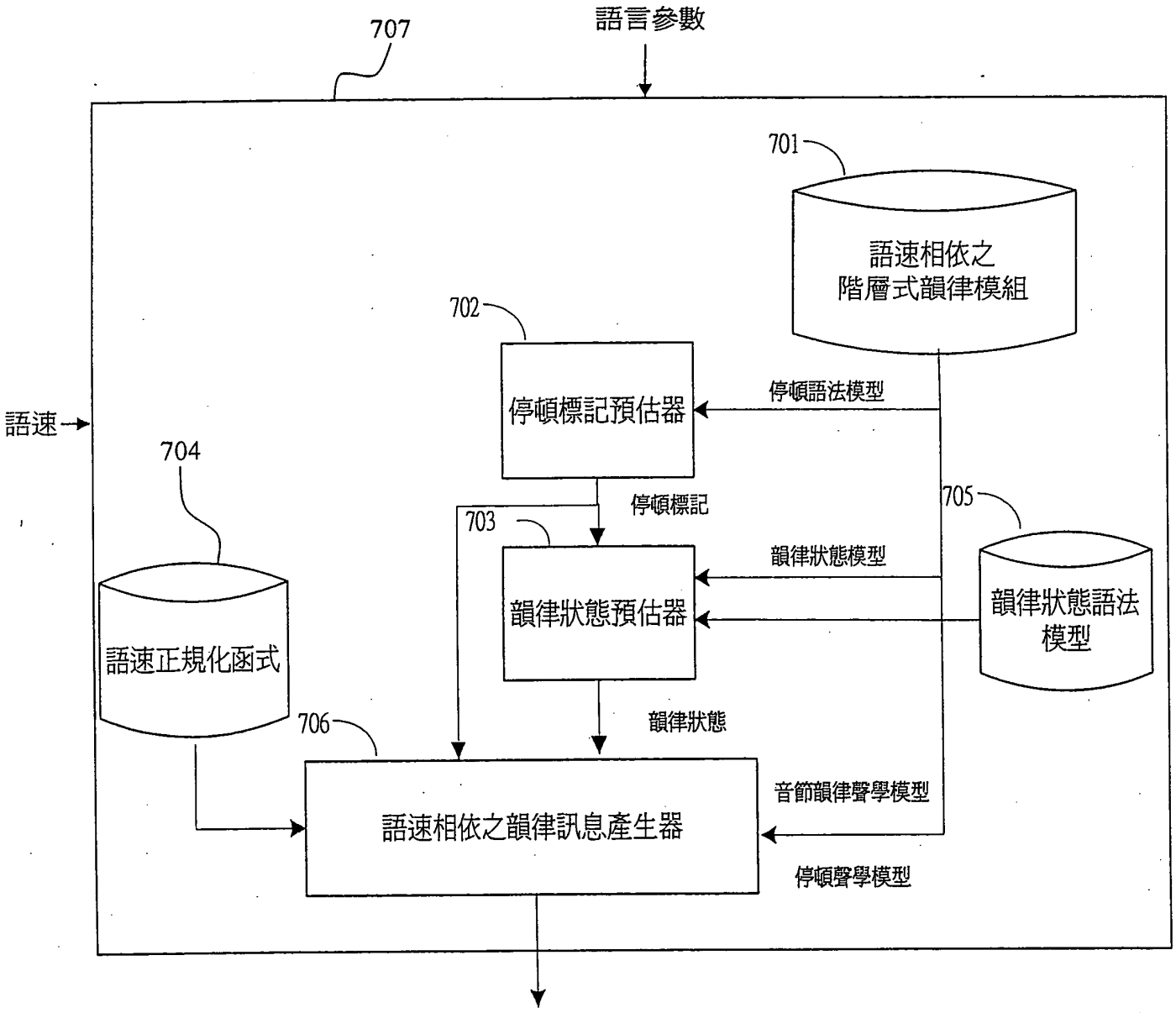
第五圖


快速
聲調 1
聲調 2
聲調 3
聲調 4
聲調 5
基頻對數 (Log Hz)
5.6
5.5
5.4
5.3
5.2
5.1
5
0
0.05
0.1
0.15
0.2
0.25
0.3
時間 (秒)

第六圖(a)


慢速
聲調 1
聲調 2
聲調 3
聲調 4
聲調 5
基頻對數 (Log Hz)
5.6
5.5
5.4
5.3
5.2
5.1
5
0
0.05
0.1
0.15
0.2
0.25
0.3
時間 (秒)

第六圖(b)


語言參數
707
701
語速相依之
階層式韻律模組
702
停頓標記預估器
停頓語法模型
語速
704
停頓標記
703
705
韻律狀態模型
韻律狀態預估器
韻律狀態語法
模型
語速正規化函式
706
韻律狀態
音節韻律聲學模型
語速相依之韻律訊息產生器
停頓聲學模型
音節長度、音節基頻軌跡、音節能量位階、停頓時長

第七圖

201411602

申請補充、修正之日期：102 年 01 月 17 日

發明專利說明書

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101133059

※ 申請日：　　※IPC 分類：

一、發明名稱：（中文/英文）

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組/Speaking-Rate Controlled Prosodic-Information Generating Device and Speaking-Rate Dependent Hierarchical Prosodic Module

二、中文發明摘要：

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一文字；一文字分析器，用以接收該文字，以產生一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。

三、英文發明摘要：

The present invention provides a speaking-rate controlled prosodic-information generating device, including a first input for receiving a speaking rate; a second input for receiving a text; a text analyzer for receiving a text to produce a linguistic feature ; a speaking-rate dependent prosody generation module uses a linguistic feature incorporating with a speaking rate to produce a SR-dependent

201411602

申請補充、修正之日期:102 年 03 月 7 日修正
年 月 日

發明專利說明書

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101133059

※ 申請日： 101.9.10 ※IPC 分類：

一、發明名稱：（中文/英文）

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組/Speaking-Rate Controlled Prosodic-Information Generating Device and Speaking-Rate Dependent Hierarchical Prosodic Module

二、中文發明摘要：

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。

三、英文發明摘要：

The present invention provides a speaking-rate controlled prosodic-information generating device, including a first input for receiving a speaking rate (SR); a second input for receiving a linguistic feature; a speaking-rate dependent prosody generation module which uses the linguistic feature incorporating with the speaking rate to produce a SR-dependent prosodic-acoustic feature; and an output for outputting the SR-dependent prosodic-acoustic feature.

201411602

申請補充、修正之日期：102 年 01 月 17 日

發明專利說明書

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101133059

※ 申請日：　　※IPC 分類：

一、發明名稱：（中文/英文）

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組/Speaking-Rate Controlled Prosodic-Information Generating Device and Speaking-Rate Dependent Hierarchical Prosodic Module

二、中文發明摘要：

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一文字；一文字分析器，用以接收該文字，以產生一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。

三、英文發明摘要：

The present invention provides a speaking-rate controlled prosodic-information generating device, including a first input for receiving a speaking rate; a second input for receiving a text; a text analyzer for receiving a text to produce a linguistic feature ; a speaking-rate dependent prosody generation module uses a linguistic feature incorporating with a speaking rate to produce a SR-dependent

201411602

申請補充、修正之日期:102 年 03 月 7 日修正
年 月 日

發明專利說明書

（本說明書格式、順序，請勿任意更動，※記號部分請勿填寫）

※ 申請案號：101133059

※ 申請日： 101.9.10 ※IPC 分類：

一、發明名稱：（中文/英文）

可控制語速的韻律訊息產生裝置及語速相依之階層式韻律模組/Speaking-Rate Controlled Prosodic-Information Generating Device and Speaking-Rate Dependent Hierarchical Prosodic Module

二、中文發明摘要：

本案係提供一種可控制語速的韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一語言參數；一語速相依之韻律生成模組，用以接收該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之一韻律聲學特徵參數。

三、英文發明摘要：

The present invention provides a speaking-rate controlled prosodic-information generating device, including a first input for receiving a speaking rate (SR); a second input for receiving a linguistic feature; a speaking-rate dependent prosody generation module which uses the linguistic feature incorporating with the speaking rate to produce a SR-dependent prosodic-acoustic feature; and an output for outputting the SR-dependent prosodic-acoustic feature.

四、指定代表圖：

(一)本案指定代表圖為：第（七）圖。

(二)本代表圖之元件符號簡單說明：

701：語速相依之階層式韻律模組

702：停頓標記預估器

703：韻律狀態預估器

704：語速正規化函式

705：韻律狀態語法模型

706：語速相依之韻律訊息產生器

707：語速相依之韻律生成模組

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

六、發明說明：

【發明所屬之技術領域】

本發明係關於一種韻律訊息產生裝置，尤指一種以語速相依之階層式韻律模組為基礎之韻律訊息產生裝置。

【先前技術】

目前對於相異語速語音合成之討論不少，但始終不能達成一流暢的自然合成語音。其中有以每個語音音框的時間軸進行伸張及壓縮，藉以調整說話速度慢及快之研究，但並未考慮到語速對於韻律結構的影響；以線性預估的方式對輸入語音進行語速修正，對輸入的語音信號以線性插入或移除信號本身之研究，該方法雖簡易有效率，但對於語速的考量過於粗糙；以清化元音(devoiced vowel)的決定中考慮了語速影響，有效地改進清化元音在慢語速的退化程度之研究，但其韻律的產生方法並未考量語速的影響；以對不同語速語料庫建立韻律結構的轉換關係，藉以達到語速轉換的目的之研究，但該方法並不能掌握到連續語速的轉換變化；雖有文獻實現了可控制語速之 TTS，首先對三種速度(快、正常、慢)各自建立音長模型，對三個音長模型以內插方式來產生目標語速所需之音長，最後結合於 HMM 為基礎之語音合成器，此方法僅考慮韻律之中的音長部份，並未對其他韻律參數進行語速影響調整，且由於不同語速需各自建立自己的音長模型，會使得模型參數量大增；再則它使用內插法去產生音長，無法獲得準確的語速控制；另有文獻對正常及快速語料分別建立 HSMM 模型，再以 CMLLR 對音長模型進行音長平均值的語速調適，該方法僅考慮韻律之中的音長部份，且由於不同語速需各自建立自己的音長模型，會使得模型參數量大增；及有進行大規模主觀測試三種語速控制的方法研究，分別為：(1)針對目標語速選取相近語速之語料來訓練 HMM 模型，(2)依比例去伸縮合成語句的發音長度，及(3)基於 ML 準則去決定狀態長度(state duration)，這些方法都是建立於 HMM-based 的語音合成系統，實驗結果發現方法(2)

最適合用於快語速合成語音，而方法(1)較適合慢速語音，不同的語速控制方法都只適於某種語速，並沒有一種方法能掌握所有語速的控制。

因此，可知習知技術大多以等比例拉長或縮短各個合成單元(如音節、詞)之長度來達到語速控制，而於韻律結構、音高軌跡、停頓時間長度及停頓出現頻率方面較少著墨，並無考慮聲學韻律訊息其背後的產生模型，因此並不能以系統化的方式掌握語速對於韻律多層面的影響，進而用以產生韻律訊息；這些韻律訊息可充分應用於語音合成之語速控制，使各種語速之合成語音應用在語音合成之領域聽起來都很流利自然。

爰是之故，申請人有鑑於習知技術之缺失，乃經悉心試驗與研究，並一本鍥而不捨的精神，終發明出本案「語速相依之韻律訊息產生器及語速相依之階層式韻律模組」，用以改善上述習用手段之缺失。

【發明內容】

本案之一面向係提供一韻律訊息產生裝置，包含一第一輸入端，用以接收一語速；一第二輸入端，用以接收一語言參數；一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及一輸出端，用以輸出與該語速相依之韻律聲學特徵參數。

本案之另一面向係提供一種語速相依之階層式韻律模組，包含至少二模型，其中各該模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

本案之又一面向係提供一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

【實施方式】

本發明將可由以下的實施例說明而得到充分瞭解，使得熟習本技藝之人士可以據以完成之，然本案之實施並非可由下列實施案例而被限制其實施型態。

本發明建立一個可應用於電子有聲書、手機、PDA 及電腦等裝置上之考慮語速對於音長、音高軌跡、停頓時機、停頓出現頻率、停頓時長所造成的影響之語速相依之階層式韻律模組；以及基於語速相依之階層式韻律模組，可以產生出各種語速的韻律聲學特徵參數，幫助語音合成達到良好的語速控制效果。其步驟包含兩個階段：模組建立及韻律合成。模組建立階段請參閱第一圖，其中是以階層式韻律模型為基礎建立語速相依之階層式韻律模組。請參閱第二圖，韻律合成階段是語速相依之韻律聲學特徵參數產生方法階段，其中是以語速相依之階層式韻律模組為基礎，產生語音合成所需要的各種韻律聲學特徵參數，且滿足不同語速之要求。

如前所述之模組建立階段之步驟包含對語音語料庫 101 中的每一句話，先做音節切割，再由每一音節抽取韻律聲學特徵參數；接著由語速估計 102 求取音節平均長度做為語速 *SR*；然後由語速正規化函式之建構 103 依據整個語音資料庫各句話的韻律聲學特徵參數對語速的統計分布來求取正規化函式；接著由韻律聲學特徵參數之語速正規化 104 來將韻律聲學特徵參數做正規化而獲得正規化韻律聲學特徵參數，再由修正型階層式韻律模型訓練演算法 105 使用整個語音語料庫每一句話的語速、語言參數、及正規化韻律聲學特徵參數來訓練獲得語速相依之階層式韻律模組 106，其中修正型階層式韻律模型訓練演算法，考慮語速之進一步影響，修正原本的階層式韻律模型訓練演算法，將其中兩個子模型：停頓語法模型及韻律狀態模型，加入語速考量，藉此補償語速對停頓時機(或出現頻率)、以及韻律狀態轉移所造成之影響。

如前所述之韻律合成階段之步驟包含:首先由文字分析器 201 將輸入文字做斷詞及詞類標記分析，獲得語言參數，再由語速相

依之韻律聲學特徵參數產生方法 202 使用語言參數、語速、語速相依之階層式韻律模組 204、以及語速正規化函式 203 來產生四種韻律聲學特徵參數。語速相依之階層式韻律模組 204 主要是決定整個語句的韻律架構(依據語速)及基本韻律參數合成，而語速正規化函式 203 是將基本韻律參數的統計特性調到指定語速的統計特性。

請參閱表一及第三圖，其分別為本發明中使用語料庫大小之統計資訊及語料庫語速之統計分佈圖。該語料庫是以一女性專業播音員依四種語速所錄製之平行語音資料庫當作實施目標，由該圖中可知四種語速所錄製之平行語音資料語速分佈在 0.15-0.3second/syllable 之間。

表一

	語句數	音節數	小時數
快語速	368	50691	3.4
一般語速	376	51868	3.9
中等語速	362	49956	4.8
慢語速	372	51231	6.0

對於韻律聲學特徵參數的正規化函式建構方法，其中一般正規化方法是對每個語句各自的資料統計參數做正規化，該方法簡易且具有效率，但可能造成過度正規化，導致除了語速之外的其它影響因素亦被調整而扭曲，進而使模組建造錯誤。本發明採用一較合理之正規化方法，即使用平滑曲線去模擬每個語句的正規化參數與語速的關係，藉由這些平滑曲線來形成語速正規化函式。

對於韻律聲學特徵參數中的音節長度，採取高斯正規化的方法，並使用二階多項式曲線來模擬音節長度的標準差，如下列式子所示：

$$sd_n' = \left(sd_n - \mu_k^{sd}\right) \Big/ \tilde{\sigma}^{sd}(SR(k)) \times \sigma_g^{sd} + \mu_g^{sd}$$

其 中

$$\tilde{\sigma}^{sd}(SR) = a_1 \left(SR\right)^2 + b_1 \cdot SR + c_1$$

為平滑化後的標準差，$\mu_k^{sd} = SR(k)$為語句 k 之音節平均長度(也就是語速)，sd_n和 sd_n'分別代表原始音節長度和語速正規化之音節長度；μ_g^{sd}和 σ_g^{sd}為語料庫整體的音節長度平均值與標準差。

對於停頓長度，使用 Gamma 分佈來表示其分佈，同樣使用二階多項式曲線來模擬語句之停頓長度平均值與標準差對語速 SR的關係，其數學式子如下：

$$\tilde{\mu}^{pd}(SR) = a_2(SR)^2 + b_2 \cdot SR + c_2$$

$$\tilde{\sigma}^{pd}(SR) = a_3(SR)^2 + b_3 \cdot SR + c_3$$

接著利用平滑化的平均值 $\tilde{\mu}^{pd}(SR(k))$和標準差 $\tilde{\sigma}^{pd}(SR(k))$去對停頓長度 pd_n做分佈正規化，其使用之公式為：

$$pd_n' = G^{-1}(G(pd_n, \tilde{\alpha}^{pd}(SR(k)), \tilde{\beta}^{pd}(SR(k))), \alpha_g^{pd}, \beta_g^{pd})$$

其中 $G(pd, \alpha, \beta)$為 Gamma 分佈的累積分佈函數(cumulative distribution function)，G^{-1}為 G 的反函數；

$$\tilde{\alpha}^{pd}(SR(k)) = (\tilde{\mu}^{pd}(SR(k)))^2 / (\tilde{\sigma}^{pd}(SR(k)))^2$$

和

$$\tilde{\beta}^{pd}(SR(k)) = (\tilde{\sigma}^{pd}(SR(k)))^2 / \tilde{\mu}^{pd}(SR(k))$$

為 Gamma 函數的兩個參數的平滑值，α_g^{pd}和 β_g^{pd}為由語料庫整體的停頓長度平均值和標準差所計算的 Gamma 函數參數。

對於音節音高軌跡，先進行正交展開(orthogonal expansion)，使用四個 Legendre 多項式為基底，用所得到的四維正交參數來表示基頻軌跡，即 $\mathbf{sp}_n = \left[a_n^0\ a_n^1\ a_n^2\ a_n^3\right]^T$，接著依每一音節聲調(lexical tone)之每一維度來正規化 SR

對 $\mathbf{sp}_n$ 的影響，公式如下：

$$\mathbf{sp}_n'(i)=\frac{\mathbf{sp}_n(i)-\tilde{\mu}^{sp}(SR(k),t_n,i)}{\tilde{\sigma}^{sp}(SR(k),t_n,i)}\times\sigma_g^{sp}(t_n,i)+\mu_g^{sp}(t_n,i)$$

其中

$$\tilde{\mu}^{sp}(SR,t,i)=b_4(t,i)\cdot SR+c_4(t,i)$$

$$\tilde{\sigma}^{sp}(SR,t,i)=b_5(t,i)\cdot SR+c_5(t,i)$$

分別為 **sp** 第 i 維、第 t 聲調的平滑化平均值與標準差，它們都以一階函數來表示；$\mu_g^{sp}(t,i)$ 和 $\sigma_g^{sp}(t,i)$ 為整個語料庫的 **sp** 第 i 維、第 t 聲調的平均值與標準差。

對於音節能量位階，由於它與錄音條件有很大的相關性，包含麥克風與語者距離、麥克風本身的錄音品質、錄音的環境等等因素之影響遠遠大於語速所造成的，因此本實施案例採取非語速相依的高斯正規化。

在完成參數正規化後，再對所有訓練語句以實施方塊 105 修正型階層式韻律模型訓練演算法來自動產生一個語速相依之階層式韻律模組，該模組包括四個子模型，用來描述觀察到的韻律聲學特徵參數、語言參數及韻律階層架構標記之間的關係。雖然我們在之前參數正規化時已把語速對韻律聲學特徵參數之影響做適當補償消除，但停頓出現的頻率及韻律狀態的轉移仍與語速有很大的相關性，因此我們以決策樹描述七種停頓類型(請參閱第四圖)出現頻率與語言參數之間的關係來修正停頓語法子模型；以及使用一階馬可夫模型來描述前一個韻律狀態和目前韻律狀態之間的轉移關係來修正韻律狀態子模

型，使所述之二個子模型與語速相依。修正型韻律模型訓練演算法為一參數最佳化問題求解的方法，在已知正規化韻律聲學特徵參數{X, Y, Z}、語言參數{L}及語速 *SR* 之情況下找到最佳的韻律標記序列 T={B,PS}，即下列數學式子：

$$\mathbf{B}^*,\mathbf{PS}^* = \arg\max_{\mathbf{B},\mathbf{PS}} P(\mathbf{B},\mathbf{PS} \mid \mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{L},SR)$$
$$\approx \arg\max_{\mathbf{B},\mathbf{PS}} \underbrace{P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{PS}|\mathbf{B},SR)P(\mathbf{B}|\mathbf{L},SR)}_{\text{語速相依之階層式韻律模型}}$$

其中 **B** 代表停頓標記序列，**PS** = {**p**, **q**, **r**}分別為音節基頻、長度及能量位階的韻律狀態標記序列，此兩類韻律標記是用來描述第四圖所考量的中文韻律階層結構，此結構包含四種韻律成分：音節、韻律詞、韻律片語、及呼吸或韻律片語群組；韻律停頓 B_n 是用來描述音節 n 和音節 $n+1$ 之間的停頓狀態，共使用七種韻律停頓狀態來描述此四種韻律成分的邊界；**A**={**X**,**Y**,**Z**}為韻律聲學特徵參數序列，其中 **X**={**sp**, **sd**, **se**}、**Y**={**pd**, **ed**} 和 **Z**={**pj**, **dl**, **df**}分別代表與音節相關的韻律聲學特徵參數、音節間及差分之韻律聲學特徵參數序列；**L**={**POS**,**PM**,**WL**,**t**,**s**,**f**}為語言參數序列，其中{**POS**,**PM**,**WL**}為高階語言參數序列，**POS**、**PM** 及 **WL** 分別為詞類序列、標點符號序列及詞長序列，而{**t**,**s**,**f**}為低階語言參數序列，**t**、**s** 及 **f** 分別為聲調、基本音節類別及韻母類別序列；*SR* 為語句之語速。詳細符號定義請參閱表二。

表 二

T: 韻律標籤	**B**: 停頓類型	
	PS: 韻律狀態	**p**: 基頻韻律狀態 **q**: 時長韻律狀態 **r**: 能量位階韻律狀態
A: 韻律聲學特徵參數	**X**: 音節韻律聲學特徵參數	**sp**: 音節基頻軌跡 **sd**: 音節時長 **se**: 音節能量位階
	Y: 音節間韻律聲學特徵參數	**pd**: 停頓時長 **ed**: 能量低點位階
	Z: 音節差分韻律聲學特徵參數	**pj**: 正規化基頻跳躍 **dl**: 正規化時長拉長因子 1 **df**: 正規化時長拉長因子 2
SR: 語速		
L: 語言參數	**POS**: 詞類 **PM**: 標點符號 **WL**: 詞長	
	t: 聲調 **s**: 基本音節類型 **f**: 韻母類型	

語速相依之階層式韻律模組可以下列方程式表示 $P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{PS}|\mathbf{B},SR)P(\mathbf{B}|\mathbf{L},SR)$。該模組包含四個子模型：音節韻律聲學模型 $P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})$、停頓聲學模型 $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$、韻律狀態模型 $P(\mathbf{PS}|\mathbf{B},SR)$ 以及停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)$：

(1) 音節韻律聲學模型 $P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L})$：

如下式所示，它再以三個子模型來近似：

$$P(\mathbf{X}|\mathbf{B},\mathbf{PS},\mathbf{L}) \approx P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{t})P(\mathbf{sd}|\mathbf{B},\mathbf{q},\mathbf{t},\mathbf{s})P(\mathbf{se}|\mathbf{B},\mathbf{r},\mathbf{t},\mathbf{f})$$
$$\approx \prod_{n=1}^{N} P(\mathbf{sp}_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})P(sd_n \mid q_n, s_n, t_n)P(se_n \mid r_n, f_n, t_n)$$

其中子模型 $P(\mathbf{sp}_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})$、$P(sd_n \mid q_n, s_n, t_n)$ 以及 $P(se_n \mid r_n, f_n, t_n)$ 分別代表第 n 個音節的音高軌跡、音節長度、能量位階之模型，t_n、s_n 及 f_n 分別表示第 n 個音節的聲調、基本音節、及韻母類型；$B_{n-1}^{n} = (B_{n-1}, B_n)$；和 $t_{n-1}^{n+1} = (t_{n-1}, t_n, t_{n+1})$。

在本實施例中，這三個子模型各考慮了多個影響因子(Affecting Factors, AFs)，這些影響因子以加成方式結合，以第 n 個音節的音高軌跡為例，我們可得：

$$\mathbf{sp}_n = \mathbf{sp}_n^r + \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu}_{sp}$$

其中 $\mathbf{sp}_n = [\alpha_{0,n}, \alpha_{1,n}, \alpha_{2,n}, \alpha_{3,n}]^T$ 為一四維正交化係數向量，用以表示第 n 個音節觀察到的音高軌跡，$\mathbf{sp}_n^r$ 為正規化後的殘餘值，$\boldsymbol{\beta}_{t_n}$ 和 $\boldsymbol{\beta}_{p_n}$ 分別為聲調和韻律狀態兩影響因子(AF)的影響數值(Affecting Pattern, AP)，$\boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}}$ 和 $\boldsymbol{\beta}^b_{B_n,tp_n}$ 為向前及向後連音兩 AF 的影響數值；$tp_{n-1} = t^n_{n-1}$；μ_{sp} 為音高的全域平均值。基於假設 $\mathbf{sp}_n^r$ 為零平均值之高斯常態分佈，我們可以高斯常態分佈來表示 $\mathbf{sp}_n$ 如下所示

$$P(\mathbf{sp}_n \mid B^n_{n-1}, p_n, t^{n+1}_{n-1}) = N(\mathbf{sp}_n; \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n} + \boldsymbol{\beta}^f_{B_{n-1},tp_{n-1}} + \boldsymbol{\beta}^b_{B_n,tp_n} + \boldsymbol{\mu}_{sp}, R_{sp})$$

其中 $N(\mathbf{x};\boldsymbol{\mu},R)$ 表示向量 $\mathbf{x}$ 為 mean vector $\boldsymbol{\mu}$ 及 covariance matrix R 之常態分佈。

音節長度 $P(sd_n \mid q_n, s_n, t_n)$ 及能量位階 $P(se_n \mid r_n, f_n, t_n)$ 亦是以此方式去實現：

$$P(sd_n \mid q_n, s_n, t_n) = N(sd_n; \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n} + \mu_{sd}, R_{sd})$$

$$P(se_n \mid r_n, f_n, t_n) = N(se_n; \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se}, R_{se})$$

其中 γ_x 及 ω_x 分別代表音節長度以及音節能量位階受影響因素 x 的影響數值(AP)。

(2) 停頓聲學模型 $P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$：

音節間韻律聲學模型則以五個子模型近似之，如下式所示：

$$P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L}) \approx P(\mathbf{pd},\mathbf{ed},\mathbf{pj},\mathbf{dl},\mathbf{df}|\mathbf{B},\mathbf{L}) \approx \prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n, df_n | B_n, L_n)$$

$$\approx \prod_{n=1}^{N-1} \{ g(pd_n; \alpha_{B_n,L_n}, \beta_{B_n,L_n}) N(ed_n; \mu_{ed,B_n,L_n}, \sigma^2_{ed,B_n,L_n}) \cdot N(pj_n; \mu_{pj,B_n,L_n}, \sigma^2_{pj,B_n,L_n})$$

$$\cdot N(dl_n; \mu_{dl,B_n,L_n}, \sigma^2_{dl,B_n,L_n}) N(df_n; \mu_{df,B_n,L_n}, \sigma^2_{df,B_n,L_n}) \}$$

其中在第 n 個音節所跟隨的接合點(juncture n，之後以第 n 個接合點表示)的停頓長度 pd_n 以 Gamma 分佈模擬，ed_n 為第 n 個接合點的能量低點位階；pj_n 為跨越第 n 個接合點的正規化音高差，其定義如下：

$$pj_n = (\mathbf{sp}_{n+1}(1) - \chi_{t_{n+1}}) - (\mathbf{sp}_n(1) - \chi_{t_n})$$

其中 $\mathbf{sp}_n(1)$ 為 $\mathbf{sp}_n$ 的第一維度(即音節音高平均值)，χ_t 為聲調 t 平均音高位階；dl_n 及 df_n 分別為跨越第 n-1 及第 n 個接合點的兩個正規化的音節拉長因子，其定義如下：

$$dl_n = (sd_n - \pi_{t_n} - \pi_{s_n}) - (sd_{n-1} - \pi_{t_{n-1}} - \pi_{s_{n-1}})$$

$$df_n = (sd_n - \pi_{t_n} - \pi_{s_n}) - (sd_{n+1} - \pi_{t_{n+1}} - \pi_{s_{n+1}})$$

其中 π_x 代表影響因素 x 的平均音長。除了 pd_n 以 Gamma 分佈模擬外，其他四種模型皆以常態分佈模擬；因為對韻律停頓而言 L_n 的參數空間仍是太大，可以使用 CART (Classification And Regression Trees) 決策樹分類法將 L_n 分成幾類，然後同時估計 Gamma 及常態分佈的參數。

(3) 韻律狀態模型 $P(\mathbf{PS}|\mathbf{B},SR)$

韻律狀態模型 $P(\mathbf{PS}|\mathbf{B},SR)$ 以三個子模型近似之，分別用來模擬音節音高、長度及能量三種韻律狀態，並以語速等分成小段 bin 來區

分不同語速所造成的影響，如下式所示：

$$P(\mathbf{PS}\,|\,\mathbf{B}, SR) = P(\mathbf{p}|\mathbf{B}, SR)P(\mathbf{q}|\mathbf{B}, SR)P(\mathbf{r}|\mathbf{B}, SR)$$
$$\approx P(p_1|\text{bin}(SR(k)))P(q_1|\text{bin}(SR(k)))P(r_1|\text{bin}(SR(k)))$$
$$\cdot\left[\prod_{n=2}^{N} P(p_n\,|\,p_{n-1}, B_{n-1}, \text{bin}(SR(k)))P(q_n\,|\,q_{n-1}, B_{n-1}, \text{bin}(SR(k)))P(r_n\,|\,r_{n-1}, B_{n-1}, \text{bin}(SR(k)))\right]$$

其中，p_n, q_n, r_n表示音節 n 的音高、長度及能量韻律狀態；$\text{bin}(SR(k))$為語句 k 的語速 $SR(k)$所屬的小段(bin)。

(4) 停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)$

停頓語法模型 $P(\mathbf{B}|\mathbf{L},SR)\cong\prod_n P(B_n\,|\,L_n,SR(k))$由兩個步驟建構成，第一步先由 CART 決策樹分析演算法來估計 $P(B_n\,|\,L_n)$，第二步再使用多項式曲線來模擬 7 種停頓類型在每個決策樹子結點的出現頻率和語速 SR的關係，最後估計出 $P(B_n\,|\,L_n,SR)$，其公式如下所示：

$$P(B_n = m\,|\,L_n, SR(k)) = \frac{P(B_n = m\,|\,L_n, SR(k))}{\sum_{x\in\text{all break types}} P(B_n = x\,|\,L_n, SR(k))} \approx \frac{c_{m,j}SR(k) + d_{m,j}}{\sum_{x\in\text{all break type}} c_{x,j}SR(k) + d_{x,j}}$$

其中 B_n為第 k 個語句第 n 個音節後的停頓類型，j 為決策樹子結點的索引值，L_n為對應的語言參數向量，$c_{m,j}$和 $d_{m,j}$為停頓類型 m、子結點 j 的線性迴歸係數。

此修正型階層式韻律模式訓練演算法，在適當的韻律斷點和韻律狀態初始化後，是以依序最佳化程序(sequential optimization procedure)來訓練韻律模型，同時對於訓練語料以最大似然性法則(maximum likelihood criterion)來產生韻律標記及獲得語速相依之

階層式韻律模式之參數。

下列為該模組訓練之實驗結果。請參閱表三，其列出在使用不同影響因子組合下，各韻律聲學參數重建之總殘餘誤差值(Total Residual Error, TRE)，即扣除各種影響因子之 AP 組合後，韻律聲學特徵參數殘餘值變異數與原始韻律聲學特徵參數變異數之比值，其中，加入韻律狀態之 AP 後，各韻律聲學特徵參數之 TRE 都變得非常小。

表三

音節基頻軌跡		音節時長		音節能量位階	
影響因子	TRE	影響因子	TRE	影響因子	TRE
+ 語調	67.3%	+語調	70.6%	+ 語調	61.4%
+ 前後連音	63.2%	+ 基本音節類型	50.1%	+ 聲母類型	48.0%
+ 基頻韻律狀態	0.8%	+ 時長韻律狀態	1.4%	+ 基頻韻律狀態	1.9%

停頓時長為音節間韻律聲學子模型最重要的參數，請參閱第五圖，其顯示出七種停頓類別的平均值對語速的關係，其中在*B*0、*B*1、*B*2-1及*B*2-3四種不明顯停頓時長的類別，它們與語速相關性甚小，其餘停頓類別之停頓時長皆隨著*SR*呈非線性增加。而表四為對每種停頓類別計算重建停頓時長的均方根誤差，發現只有*B*2-2、*B*3及*B*4之誤差會比較大，這是因為這些停頓類別通常發生在MINOR BREAK 或MAJOR BREAK位置，因其變異較大所以重建誤差也自然較大，此結果是在合理的範圍。

表四

停頓類型	*B*0	*B*1	*B*2-1	*B*2-2	*B*2-3	*B*3	*B*4
均方根誤差	3 毫秒	19 毫秒	25 毫秒	90 毫秒	30 毫秒	104 毫秒	149 毫秒

請參閱第六圖，其是用聲調 AP 來產生快、慢兩種語速的音高軌跡，可觀察到每一聲調的基頻軌跡受語速的影響程度皆不盡相同。

請參閱表五，其顯示一個停頓類別的標記例子，此例子對四個不同語速(由上往下語速漸慢)的平行語料標記，在此只標示出 *B*4 (@)、*B*3 (/) 及 *B*2-2 (*)三種具明顯停頓時長之類別，其顯示出語速越慢時越容易出現明顯類別的停頓，符合預期之結果。

表五

依據 行政院 主計處 的 統計 @，十月份 * 一 到 二十日 / ，我國 出口 及進口 金額 / 比起 去年 同期 * 均有 增加 @，
依據 行政院 主計處 的 統計 @，十月份 * 一 到 二十日 /，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加@，
依據 * 行政院 主計處 的 統計 @，十月份 / 一 到 * 二十日 /，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加 @，
依據 / 行政院 * 主計處 的 統計 @，十月份 / 一 * 到 * 二十日 @，我國 出口 * 及 進口 金額 / 比起 去年 同期 * 均有 增加 @，

上述各項實驗數據顯示該模組可有效地描述漢語語音韻律參數之各種變化。

對於可控制語速之韻律聲學特徵參數產生方法可經由參閱第七圖得到進一步瞭解，

其為第二圖的較詳細圖示，其是基於訓練出來的語速相依之階層式韻律模組 701 之可控制語速之漢語韻律聲學特徵參數產生法流程圖。方塊 702 為停頓標記預估器，其使用該韻律模型中的停頓語法模型來做停頓標記預估的方法：

$$B_n^* = \arg\max_{B_n} P(B_n \mid L_n, SR)$$

其中 L_n 為輸入的語言參數，SR 為指定的語速。

方塊 703 為韻律狀態標記預估器，其使用此韻律模型中的韻律狀態模型搭配一組額外的韻律狀態語法模型 705，以維特比演算法(Viterbi algorithm)來預估之，如以下數學式所示：

$$\mathbf{p}^*, \mathbf{q}^*, \mathbf{r}^* = \arg\max_{\mathbf{p},\mathbf{q},\mathbf{r}} \left(\begin{array}{l} P(p_1|\text{bin}(SR))P(q_1|\text{bin}(SR))P(r_1|\text{bin}(SR)) \\ \cdot\prod_{n=2}^{N} P(p_n \mid p_{n-1}, B_{n-1}^*, \text{bin}(SR))P(q_n \mid q_{n-1}, B_{n-1}^*, \text{bin}(SR))P(r_n \mid r_{n-1}, B_{n-1}^*, \text{bin}(SR)) \end{array} \right) \cdot \left(\prod_{n=1}^{N} P(p_n \mid L_n)P(q_n \mid L_n)P(r_n \mid L_n) \right)$$

其中 $p(p_n \mid L_n)$、$p(q_n \mid L_n)$、$p(r_n \mid L_n)$ 為韻律狀態語法模型，它們係使用做完韻律標記之訓練語料以 CART 演算法實現之，B_{n-1}^* 為停頓標記預估結果。

有了韻律標記預估結果後，可利用韻律模型中的的音節韻律聲學模型 $P(\mathbf{PS} \mid \mathbf{B}, \mathbf{L})$ 和停頓聲學模型 $P(\mathbf{X}, \mathbf{Y} \mid \mathbf{B}, \mathbf{L})$ 來產生語速正規化之韻律聲學特徵參數，再藉由語速正規化函式 704 之反函式來還原產生指定語速之韻律聲學特徵參數，各韻律聲學特徵參數之產生說明如下：

語速控制的停頓時長產生方法為

$$pd_n' = \mathrm{G}^{-1}(\mathrm{G}(pd_n^*, \alpha_g^{pd}, \beta_g^{pd}), \tilde{\alpha}^{pd}(SR), \tilde{\beta}^{pd}(SR))$$

其中

$$pd_n^* \equiv \mu_n^* = \alpha_n^* \beta_n^*$$

為語速正規化之停頓時長，它使用停頓聲學模型中由 B_n^* 和前後文參數 L_n 所找到的節點的 Gamma 分布的參數 α_n^* 及 β_n^* 去計算的平均值 μ_n^* 來估計；語速控制的音節音高軌跡產生方法為

$$\mathbf{sp}_n'(i) = \frac{\mathbf{sp}_n^*(i) - \mu_g^{sp}(t_n, i)}{\sigma_g^{sp}(t_n, i)} \times \tilde{\sigma}^{sp}(SR, t_n, i) + \tilde{\mu}^{sp}(SR, t_n, i)$$

其中語速正規化之基頻軌跡 $\mathbf{sp}_n^*$ 的預估如下面數學式所示，它是以預估之韻律標記和聲調語言參數來挑選對應的 AP 所疊加產生：

$$\mathbf{sp}_n^* = \boldsymbol{\beta}_{t_n} + \boldsymbol{\beta}_{p_n^*} + \boldsymbol{\beta}^f_{B^*_{n-1}, t^n_{n-1}} + \boldsymbol{\beta}^b_{B^*_n, t^{n+1}_n} + \boldsymbol{\mu}_{sp}$$

語速控制的音節長度產生方法如下：

$$sd_n' = \left(sd_n^* - \mu_g^{sd}\right) / \sigma_g^{sd} \times \tilde{\sigma}^{sd}(SR) + \mu_k^{sd}$$

其中語速正規化之音節長度 sd_n^* 是以對應的 AP 所疊加產生：

$$sd_n^* = \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n^*} + \mu_{sd}$$

最後音節能量位階的產生方法為

$$se_n^* = \omega_{t_n} + \omega_{f_n} + \omega_{r_n^*} + \mu_{se}$$

以下為語音合成範例。本發明所產生的韻律聲學特徵參數能結合於任何語音合成器，以達到語速控制之語音合成。在此以一隱藏式馬可夫為基礎之語音合成技術

(HMM-based speech synthesis)為例將語音合成出來，此技術為習知技術，在此簡短說明其參數設定：中文的 21 個聲母及 39 個韻母都各以一個 HMM 表示，每個 HMM 包含 5 個 HMM 狀態，每一個狀態內的觀察向量包含兩個類別串：一個為維度 75 的頻譜參數，另一個為離散的事件來表示清音(unvoiced)或濁音(voiced)的狀態，每一個狀態皆以多變量單一高斯函數(multi-variate single Gaussian)表示其觀察機率。訓練 HMM 模型的方法是以習知方法(embedded-trained 及決策樹方法對 HMM 狀態分群)訓練其參數，上述之參數設定及訓練方法可視實際情況而調整，其並非用以限制本發明之範圍。

請參閱表六，其為 MOS 主觀聽覺評估結果，其係經由十五位測試者聆聽三種語速各十句所做主觀音質評定的 MOS 分數平均，由該表中可看出合成語音在不同語速皆有不錯的聲音品質。

表六

語速	快(*SR*=0.17)	中(*SR*=0.20)	慢(*SR*=0.25)
MOS	3.35	3.44	3.28

雖然本發明已以較佳實施例揭露如上，然其並非用以限定本發明之範圍，任何熟習此技藝者，在不脫離本發明之精神和範圍內，當可作各種更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者

為 準 。

實施例:

1. 一種可控制語速的韻律訊息產生裝置，包含：

一第一輸入端，用以接收一語速；

一第二輸入端，用以接收一語言參數；

一文字分析器，用以接收一文字，以產生一語言參數;

一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及

一輸出端，用以輸出與該語速相依之該韻律聲學特徵參數。

2. 如實施例 1 所述的裝置，其中根據所產生的語速相依之韻律聲學特徵參數，可使用習知之語音合成器來合成出相對應之任一指定語速之合成語音。

3. 如實施例 1-2 所述的裝置，其中該語言參數至少包含兩參數，其中各該參數係選自由包含詞類、標點符號、詞長、聲調、基本音節類型及韻母類型及其組合之一。

4. 如實施例 1-3 所述的裝置生，其中該語速相依之韻律生成模組包含一語速相依之階層式韻律模組、一語速相依之韻律訊息產生器、以及至少一個預估器，其中各該預估器係選自由包含一停頓標記預估器及一韻律狀態預估器。

5. 如實施例 1-4 所述的裝置，其中該語速相依之韻律訊息產生器，根據一語速正規化函式、該語速相依之階層式韻律模組之音節

韻律聲學模型及停頓聲學模型、該韻律狀態預估結果、該停頓標記預估結果、該輸入語速及語言參數，以產生對應語速之韻律聲學特徵參數。

6. 如實施例 1-5 所述的裝置，其中該語速正規化函式用以調整韻律聲學特徵參數的統計特性成任一語速的統計特性; 其所使用的正規化參數係採用整體語料的統計分佈經平滑化而得到。

7. 如實施例 1-6 所述的裝置，其中該語速相依之階層式韻律模組包含一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型，俾與一語速相依。

8. 如實施例 1-7 所述的裝置，其中該停頓標記預估器，根據該語速、該語言參數和該語速相依之階層式韻律模組之停頓語法模型而執行一停頓標記預估操作，以產生一停頓標記預估結果。

9. 如實施例 1-8 所述的裝置，其中該韻律狀態預估器，根據該語速、該語速相依之階層式韻律模組之韻律狀態模型、一韻律狀態語法模型和該停頓標記預估結果而執行一韻律狀態預估操作，以產生一韻律狀態預估結果。

10. 一種語速相依之階層式韻律模組，包含至少二子模型，其中各該子模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

11. 如實施例 10 所述的模組，其中該語速相依之階層式韻律模組根據一漢語語音資料庫之語言參數、正規化韻律聲學參數及各

語句的語速，再以一修正型階層式韻律模型訓練演算法來建構。

12. 如實施例 10-11 所述的模組，其中該音節韻律聲學模型、該停頓聲學模型、該韻律狀態模型及該停頓語法模型各包含至少兩種的子模型來建構。

13. 如實施例 10-12 所述的模組，其中該正規化韻律聲學參數根據各語句之語速，使用語速正規化函式對韻律聲學參數做一正規化操作所產生。

14. 如實施例 10-13 所述的模組，其中該修正型階層式韻律模型訓練演算法亦施用於至少一停頓語法子模型與一韻律狀態子模型。

15. 如實施例 10-14 所述的模組，該語速相依之階層式韻律模組根據一輸入語速、一輸入語言參數於該模組中，以產生相對應之一停頓類型機率用以協助停頓標記之預估、一韻律狀態機率用以協助韻律狀態之預估、一音節韻律聲學特徵參數機率及一音節間停頓時長之機率用以協助產生一語速相依之韻律聲學特徵參數。

16. 一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

17. 如實施例 16 所述的模組，其中該韻律狀態包含基頻韻律狀態、時長韻律狀態及能量位階韻律狀態。

18. 如實施例 16-17 所述的模組,其中該音節韻律聲學特徵參數包含音節基頻軌跡、音節時長及音節能量位階;

該音節間韻律聲學特徵參數包含停頓時長及能量低點位階;及該音節差分韻律聲學特徵參數包含基頻跳躍、時長拉長因子 1 及時長拉長因子 2。

【圖式簡單說明】

第一圖:本案一較佳實施例之架構語速相依之階層式韻律模組流程圖。

第二圖:本案一較佳實施例之產生語速相依之韻律聲學特徵參數簡易流程圖。

第三圖:本案一較佳實施例之語料庫語速統計圖。

第四圖:本案一較佳實施例之漢語語音階層式韻律結構示意圖。

第五圖:本案一較佳實施例之七種停頓類別的停頓時長平均值對語速之關係圖。

第六圖(a)~(b):本案一較佳實施例之不同聲調之基頻軌跡於不同語速之差異圖。

第七圖:本案一較佳實施例之產生語速相依之韻律聲學特徵參數流程圖。

【主要元件符號說明】

101:語音資料庫
102:語速估計
103:語速正規化函式之建構
104:韻律聲學特徵參數之語速正規化
105:修正型階層式韻律模型訓練演算法
106:語速相依之階層式韻律模組
201:文字分析器

202：語速相依之韻律參數產生方法

203：語速正規化函式

204：語速相依之階層式韻律模組

701：語速相依之階層式韻律模組

702：停頓標記預估器

703：韻律狀態預估器

704：語速正規化函式

705：韻律狀態語法模型

706：語速相依之韻律訊息產生器

707：語速相依之韻律生成模組

七、申請專利範圍：

1. 一種可控制語速的韻律訊息產生裝置，包含：

一第一輸入端，用以接收一語速；

一第二輸入端，用以接收一語言參數；

一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及

一輸出端，用以輸出與該語速相依之該韻律聲學特徵參數。

2. 如申請專利範圍第 1 項所述的裝置，其中根據所產生的語速相依之韻律聲學特徵參數，可使用習知之語音合成器來合成出相對應之任一指定語速之合成語音。

3. 如申請專利範圍第 1 項所述的裝置，其中該語言參數至少包含兩參數，其中各該參數係選自由包含詞類、標點符號、詞長、聲調、基本音節類型及韻母類型及其組合之一。

4. 如申請專利範圍第 1 項所述的裝置，其中該語速相依之韻律生成模組包含一語速相依之階層式韻律模組、一語速相依之韻律訊息產生器、以及至少一個預估器，其中各該預估器係選自由包含一停頓標記預估器及一韻律狀態預估器。

5. 如申請專利範圍第 4 項所述的裝置，其中該語速相依之韻律訊息產生器，根據一語速正規化函式、該語速相依之階層式韻律模組之音節韻律聲學模型及停頓聲學模型、該韻律狀態預估結果、該停頓標記預估結果、該輸入語速及該語言參數，以產生一對應語速之韻律聲學特徵參數。

七、申請專利範圍：

1. 一種可控制語速的韻律訊息產生裝置，包含：

一第一輸入端，用以接收一語速；

一第二輸入端，用以接收一語言參數；

一語速相依之韻律生成模組，用以配合該語言參數及該語速，以產生該語速相依之一韻律聲學特徵參數；及

一輸出端，用以輸出與該語速相依之該韻律聲學特徵參數。

2. 如申請專利範圍第 1 項所述的裝置，其中根據所產生的語速相依之韻律聲學特徵參數，可使用習知之語音合成器來合成出相對應之任一指定語速之合成語音。

3. 如申請專利範圍第 1 項所述的裝置，其中該語言參數至少包含兩參數，其中各該參數係選自由包含詞類、標點符號、詞長、聲調、基本音節類型及韻母類型及其組合之一。

4. 如申請專利範圍第 1 項所述的裝置，其中該語速相依之韻律生成模組包含一語速相依之階層式韻律模組、一語速相依之韻律訊息產生器、以及至少一個預估器，其中各該預估器係選自由包含一停頓標記預估器及一韻律狀態預估器。

5. 如申請專利範圍第 4 項所述的裝置，其中該語速相依之韻律訊息產生器，根據一語速正規化函式、該語速相依之階層式韻律模組之音節韻律聲學模型及停頓聲學模型、該韻律狀態預估結果、該停頓標記預估結果、該輸入語速及該語言參數，以產生一對應語速之韻律聲學特徵參數。

6. 如申請專利範圍第 5 項所述的裝置，其中該語速正規化函式用以調整韻律聲學特徵參數的統計特性成任一語速的統計特性；其所使用的正規化參數係採用整體語料的統計分佈經平滑化而得到。

7. 如申請專利範圍第 4 項所述的裝置，其中該語速相依之階層式韻律模組包含一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型，俾與一語速相依。

8. 如申請專利範圍第 4 項所述的裝置，其中該停頓標記預估器，根據該語速、該語言參數和該語速相依之階層式韻律模組之停頓語法模型而執行一停頓標記預估操作，以產生一停頓標記預估結果。

9. 如申請專利範圍第 4 項所述的裝置，其中該韻律狀態預估器，根據該語速、該語速相依之階層式韻律模組之韻律狀態模型、一韻律狀態語法模型和該停頓標記預估結果而執行一韻律狀態預估操作，以產生一韻律狀態預估結果。

10. 一種語速相依之階層式韻律模組，包含至少二子模型，其中各該子模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

11. 如申請專利範圍第 10 項所述的模組，其中該語速相依之階層式韻律模組根據一漢語語音資料庫之語言參數、一正規化韻律聲學參數及各語句的語速，再以一修正型階層式韻律模型訓練

6. 如申請專利範圍第 5 項所述的裝置，其中該語速正規化函式用以調整韻律聲學特徵參數的統計特性成任一語速的統計特性;其所使用的正規化參數係採用整體語料的統計分佈經平滑化而得到。

7. 如申請專利範圍第 4 項所述的裝置，其中該語速相依之階層式韻律模組包含一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型，俾與一語速相依。

8. 如申請專利範圍第 4 項所述的裝置，其中該停頓標記預估器，根據該語速、該語言參數和該語速相依之階層式韻律模組之停頓語法模型而執行一停頓標記預估操作，以產生一停頓標記預估結果。

9. 如申請專利範圍第 4 項所述的裝置，其中該韻律狀態預估器，根據該語速、該語速相依之階層式韻律模組之韻律狀態模型、一韻律狀態語法模型和該停頓標記預估結果而執行一韻律狀態預估操作，以產生一韻律狀態預估結果。

10. 一種語速相依之階層式韻律模組，包含至少二子模型，其中各該子模型係選自由一音節韻律聲學模型、一停頓聲學模型、一韻律狀態模型、一停頓語法模型及其組合之一，俾與一語速相依。

11. 如申請專利範圍第 10 項所述的模組，其中該語速相依之階層式韻律模組根據一漢語語音資料庫之語言參數、一正規化韻律聲學參數及各語句的語速，再以一修正型階層式韻律模型訓練

演算法來建構。

12. 如申請專利範圍第 10 項所述的模組，其中該音節韻律聲學模型、該停頓聲學模型、該韻律狀態模型及該停頓語法模型各包含至少兩種的子模型來建構。

13. 如申請專利範圍第 11 項所述的模組，其中該正規化韻律聲學參數根據各語句之語速，使用語速正規化函式對韻律聲學參數做一正規化操作所產生。

14. 如申請專利範圍第 11 項所述的模組，其中該修正型階層式韻律模型訓練演算法亦施用於至少一停頓語法子模型與一韻律狀態子模型。

15. 如申請專利範圍第 11 項所述的模組，該語速相依之階層式韻律模組根據一輸入語速、一輸入語言參數於該模組中，以產生相對應之一停頓類型機率用以協助停頓標記之預估、一韻律狀態機率用以協助韻律狀態之預估、一音節韻律聲學特徵參數機率及一音節間停頓時長之機率用以協助產生一語速相依之韻律聲學特徵參數。

16. 一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

17. 如申請專利範圍第 16 項所述的模組，其中該韻律狀態包含基頻韻律狀態、時長韻律狀態及能量位階韻律狀態。

演算法來建構。

12. 如申請專利範圍第 10 項所述的模組，其中該音節韻律聲學模型、該停頓聲學模型、該韻律狀態模型及該停頓語法模型各包含至少兩種的子模型來建構。

13. 如申請專利範圍第 11 項所述的模組，其中該正規化韻律聲學參數根據各語句之語速，使用語速正規化函式對韻律聲學參數做一正規化操作所產生。

14. 如申請專利範圍第 11 項所述的模組，其中該修正型階層式韻律模型訓練演算法亦施用於至少一停頓語法子模型與一韻律狀態子模型。

15. 如申請專利範圍第 11 項所述的模組，該語速相依之階層式韻律模組根據一輸入語速、一輸入語言參數於該模組中，以產生相對應之一停頓類型機率用以協助停頓標記之預估、一韻律狀態機率用以協助韻律狀態之預估、一音節韻律聲學特徵參數機率及一音節間停頓時長之機率用以協助產生一語速相依之韻律聲學特徵參數。

16. 一種語速相依之階層式韻律模組，包含至少二參數，其中各該參數係選自由一停頓類型、一韻律狀態、一音節韻律聲學特徵參數、一音節間韻律聲學特徵參數及一音節差分韻律聲學特徵參數及其組合之一，俾該模組與一語速相依。

17. 如申請專利範圍第 16 項所述的模組，其中該韻律狀態包含基頻韻律狀態、時長韻律狀態及能量位階韻律狀態。

18. 如申請專利範圍第 16 項所述的模組，其中該音節韻律聲學特徵參數包含音節基頻軌跡、音節時長及音節能量位階；

該音節間韻律聲學特徵參數包含停頓時長及能量低點位階；及

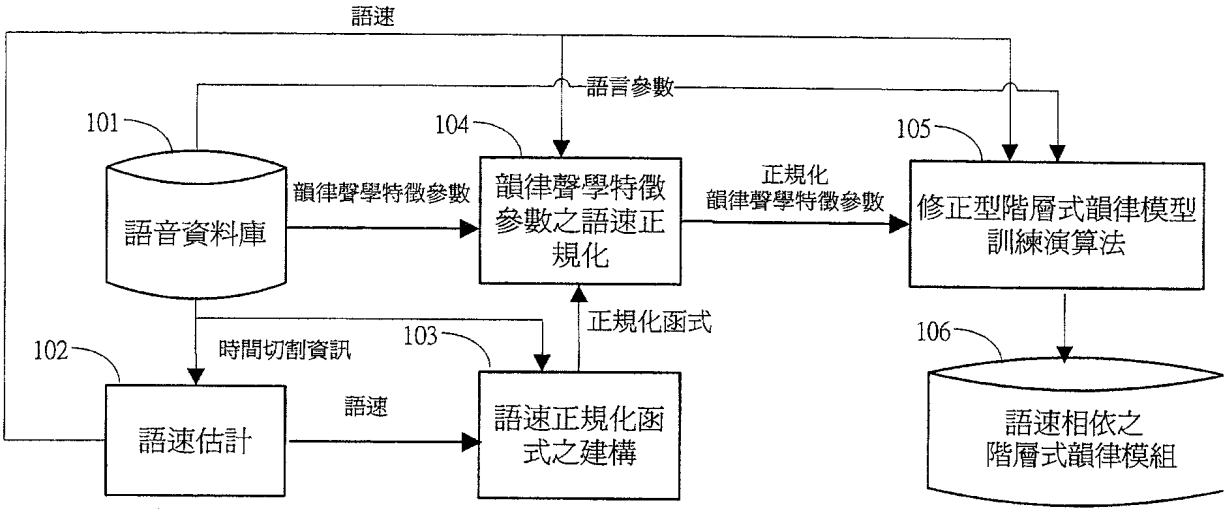
該音節差分韻律聲學特徵參數包含基頻跳躍、時長拉長因子 1 及時長拉長因子 2。

18. 如申請專利範圍第 16 項所述的模組，其中該音節韻律聲學特徵參數包含音節基頻軌跡、音節時長及音節能量位階；

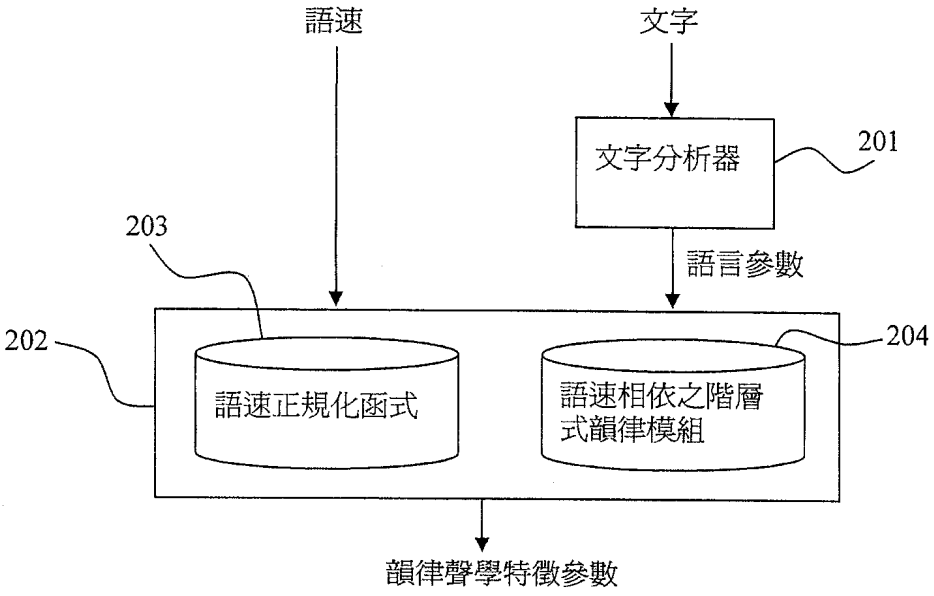
該音節間韻律聲學特徵參數包含停頓時長及能量低點位階；及

該音節差分韻律聲學特徵參數包含基頻跳躍、時長拉長因子 1 及時長拉長因子 2。

八、圖式：

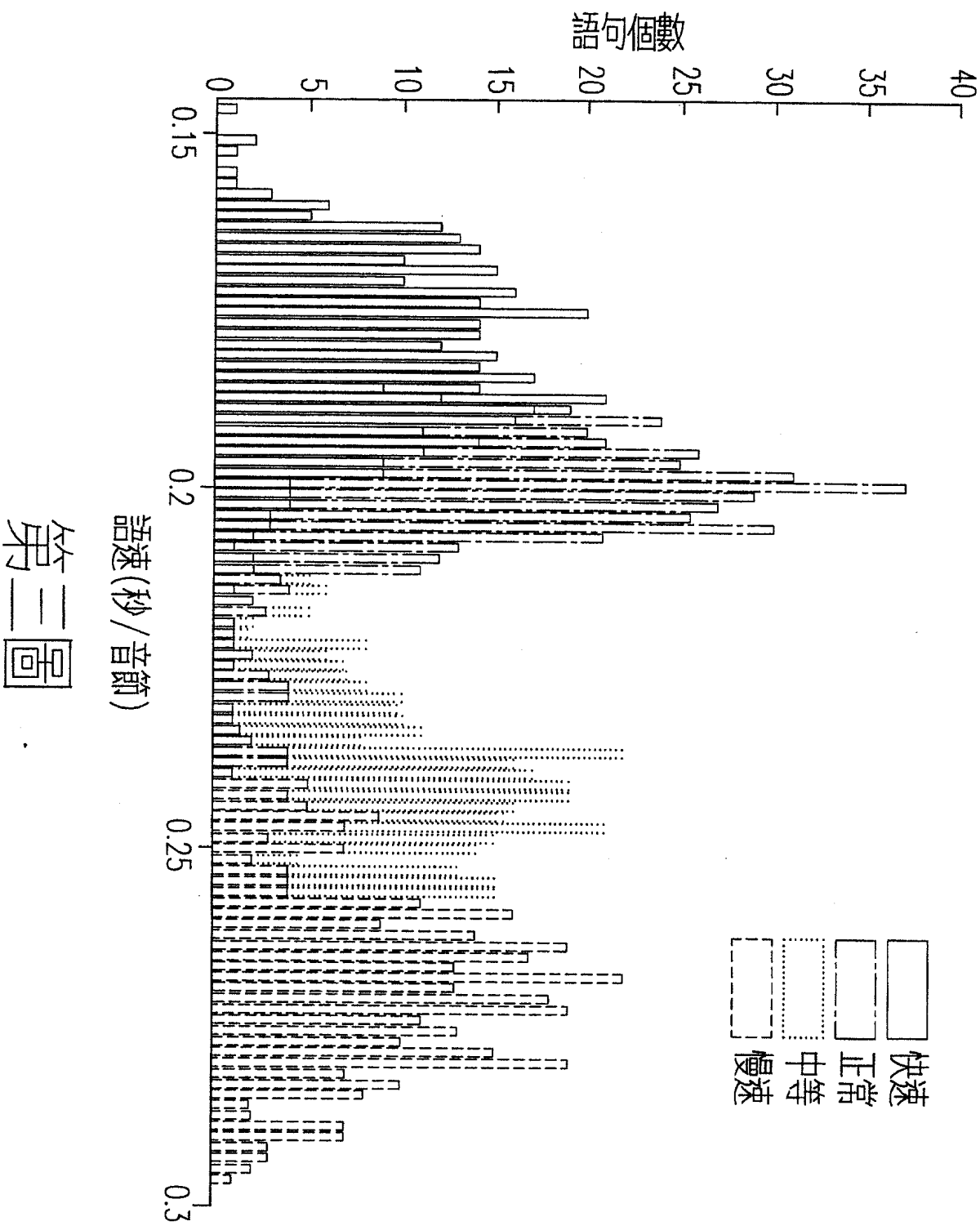

語速
語言參數
101
104
105
語音資料庫
韻律聲學特徵參數
韻律聲學特徵參數之語速正規化
正規化
韻律聲學特徵參數
修正型階層式韻律模型訓練演算法
正規化函式
102
時間切割資訊
103
106
語速估計
語速
語速正規化函式之建構
語速相依之階層式韻律模組

第一圖

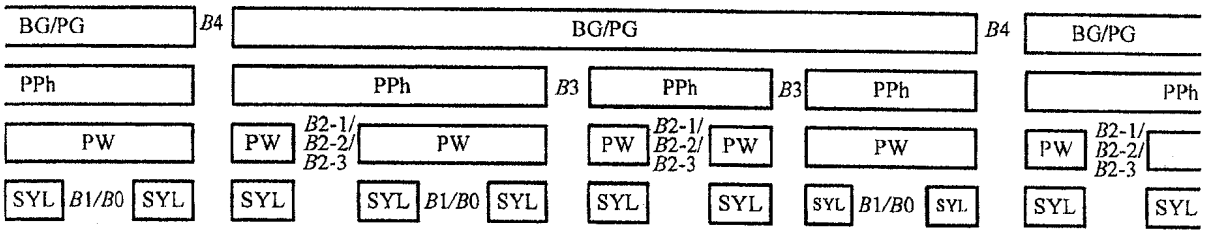

語速
文字
文字分析器
201
203
語言參數
202
204
語速正規化函式
語速相依之階層式韻律模組
韻律聲學特徵參數

第二圖

申請補充、修正之日期:102 年 03 月 6 日



第二圖



SYL: 音節

PW: 韻律詞

PPh: 韻律片語

BG/PG: 呼吸或韻律片語群組

B4: 呼吸或韻律片語群組邊界韻律斷點

B3: 韻律片語邊界韻律斷點

B2-1: 第一類韻律詞韻律斷點，表示音高重置

B2-2: 第二類韻律詞韻律斷點，表示短靜音停頓

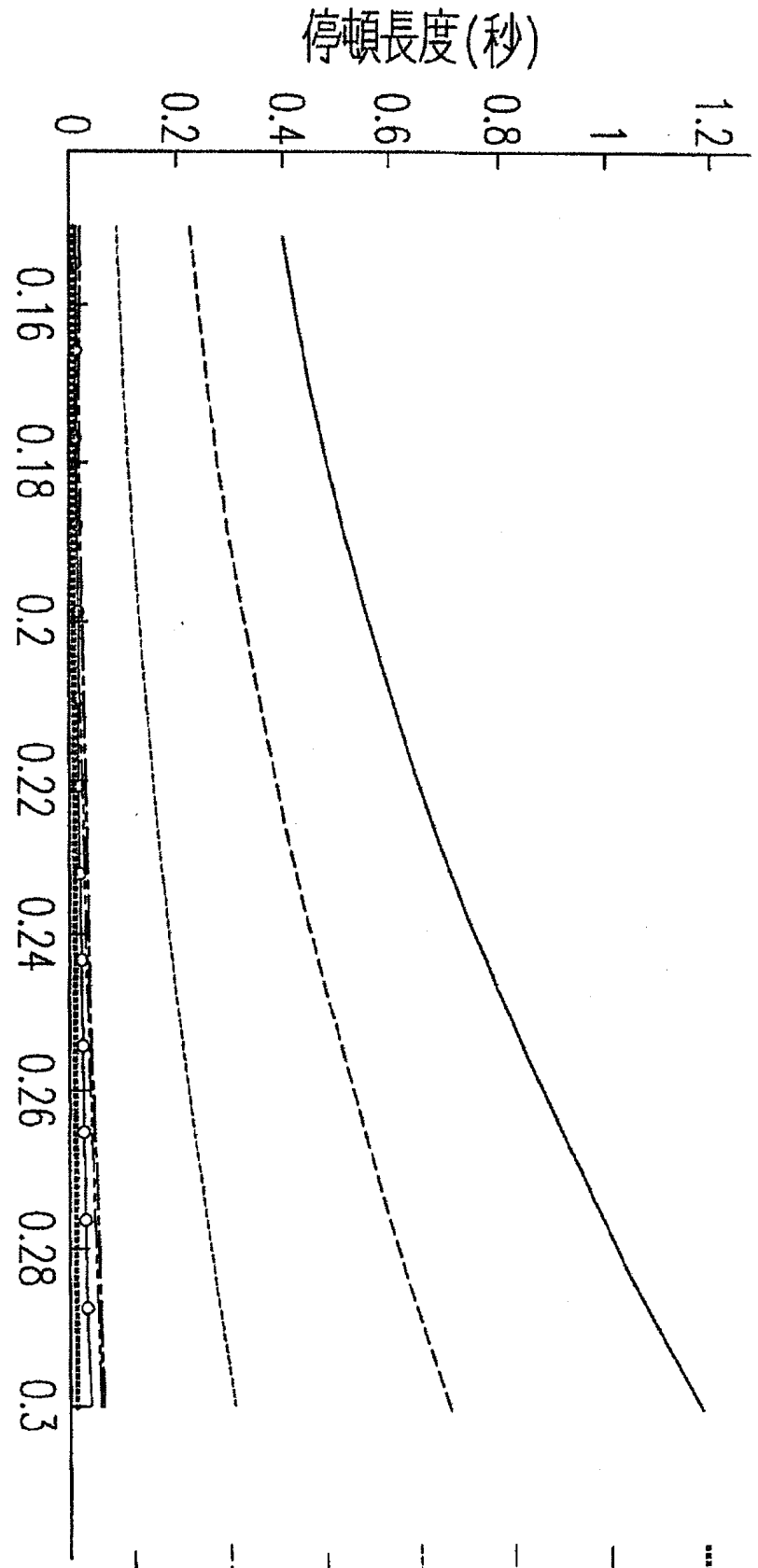
B2-3: 第三類韻律詞韻律斷點，表示音節拉長停頓

B1: 韻律詞內正常韻律斷點

B0: 韻律詞內強連音韻律斷點

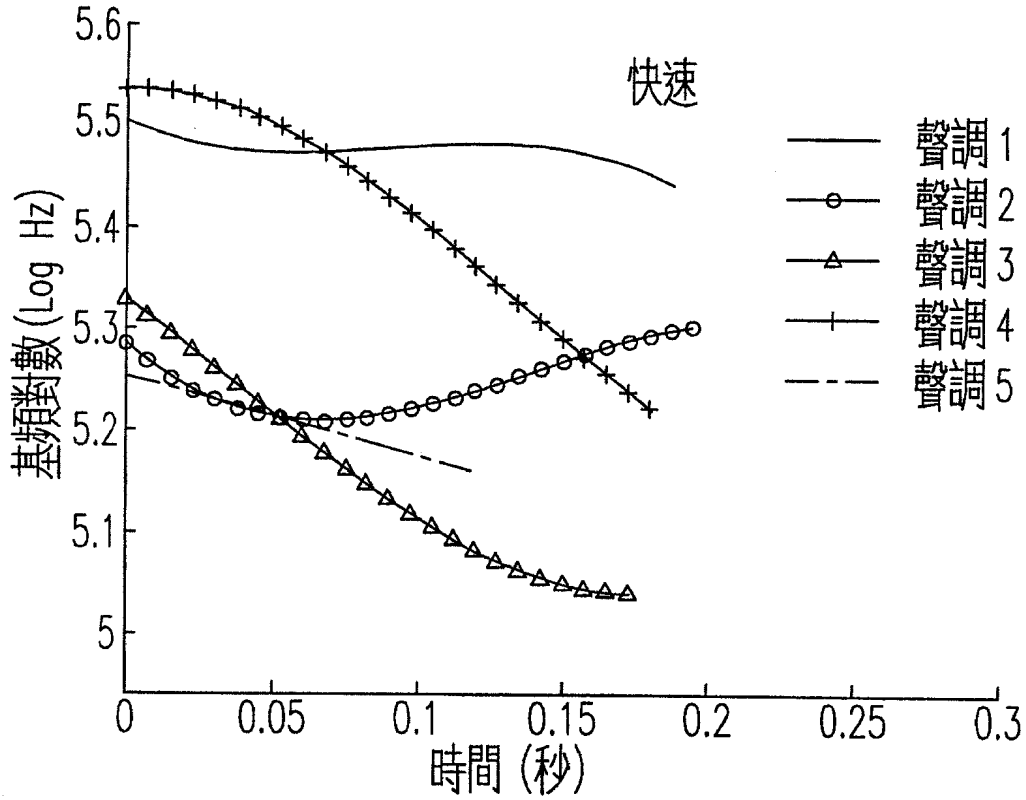
第四圖

申請補充、修正之日期:102 年 03 月 6 日

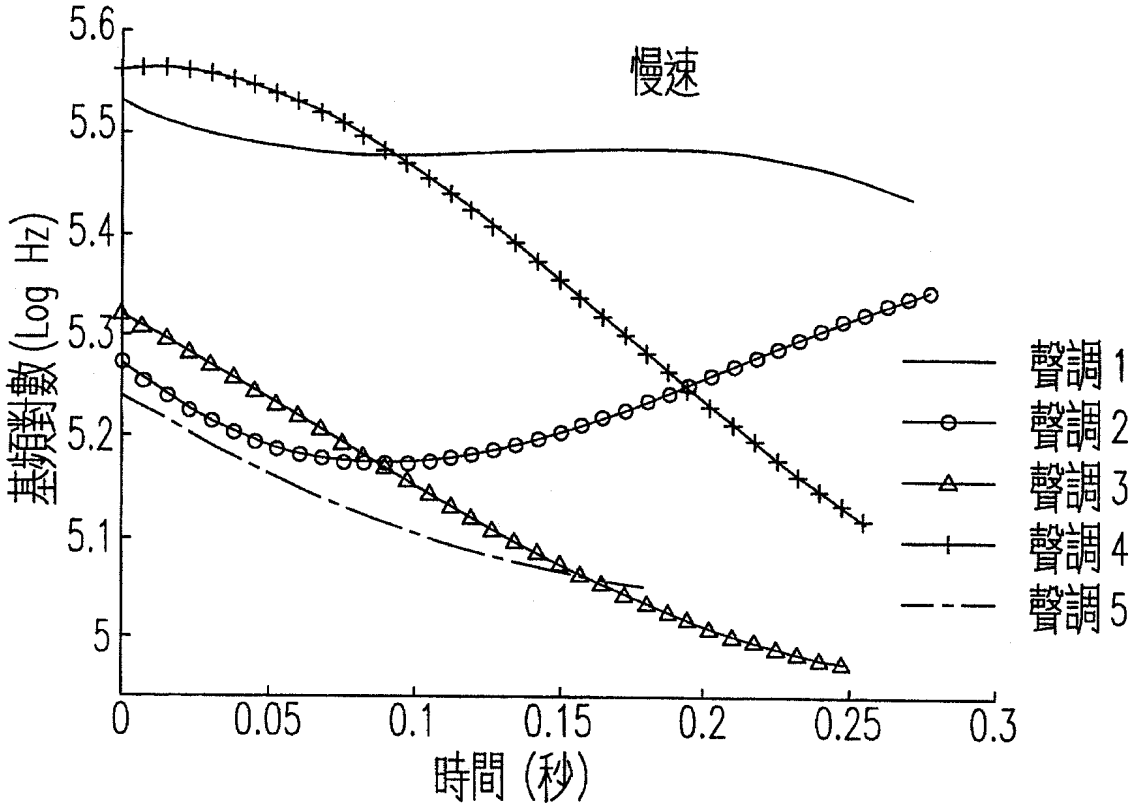


第五圖

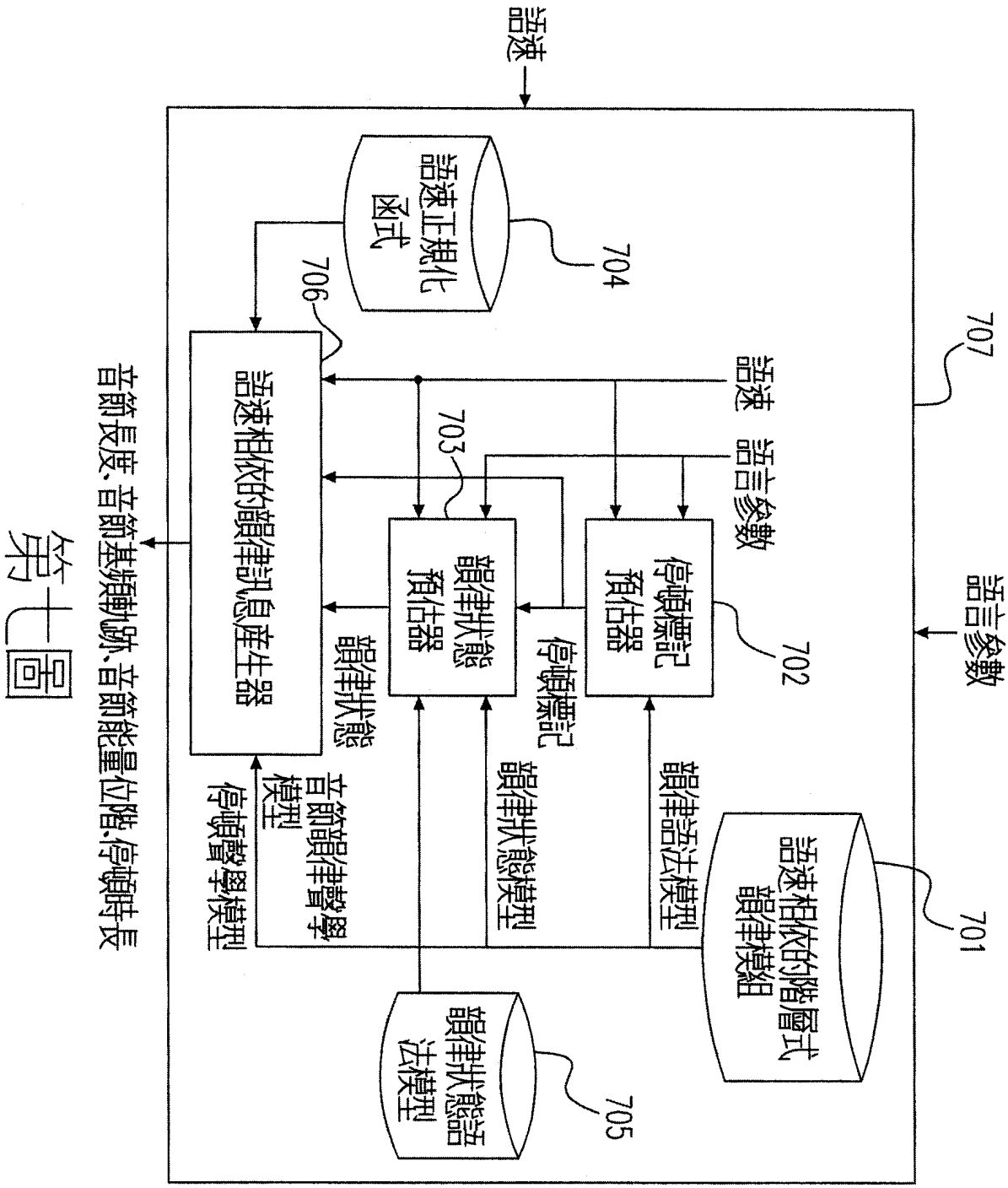
申請補充、修正之日期:102 年 03 月 6 日



第六圖(a)



第六圖(b)


語速
語速正規化函式
704
706
707
語速 語言參數
703
語速相依的韻律訊息產生器
韻律狀態預估器
停頓標記預估器
702
語言參數
韻律狀態
停頓標記
音節韻律聲學模型
停頓聲學模型
韻律狀態模型
韻律語法模型
語速相依的階層式韻律模組
701
韻律狀態語法模型
705
音節長度、音節基頻軌跡、音節能量位階、停頓時長

第七圖