I441163

(19)中華民國智慧財產局

(12)發明說明書公告本 **(11)證書號數：TW I441163 B**

(45)公告日： 中華民國 103 (2014) 年 06 月 11 日

(21)申請案號：100142341 (22)申請日： 中華民國 100 (2011) 年 11 月 18 日

(51)Int. Cl.： ***G10L15/00 (2013.01)***

(30)優先權：2011/05/10 中華民國 100116350

(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：楊智合 YANG, JYH HER (TW)；江振宇 CHANG, CHEN YU (TW)；劉銘傑 LIU, MING CHIEH (TW)；王逸如 WANG, YIH RU (TW)；廖元甫 LIAO, YUAN FU (TW)；陳信宏 CHEN, SIN HORNG (TW)

(74)代理人：林火泉

(56)參考文獻：

TW 201021024A US 2011/0035216A1

Roger Peng Yu, Kit Thambiratnam, and Frank Seide, "Word-Lattice Based Spoken-Document Indexing with Standard Text Indexers," SIGIR'08, July 20-24, 2008, Singapore, ACM 978-1-60558-164-4/08/07.

審查人員：涂淑惠

申請專利範圍項數：24 項 圖式數：6 共 0 頁

(54)名稱

中文語音辨識裝置及其辨識方法

CHINESE SPEECH RECOGNITION DEVICE AND SPEECH RECOGNITION METHOD THEREOF

(57)摘要

本發明係揭露一種中文語音辨識裝置及其辨識方法，首先，接收一語音訊號，以進行辨識後，輸出一格狀詞組(word lattice)。接著，接收格狀詞組，並根據一韻律停頓模型、一韻律狀態模型、一音節韻律聲學模型、一音節間韻律聲學模型與一因子化語言模型，重新計算格狀詞組中詞弧上的分數，將其重新排名，以輸出語音訊號對應之一語言標籤、一韻律標籤與一音段標記。本發明利用兩階段方式重新計分，不僅可提升基本語音資訊的辨識率，更可標記出語言、韻律、音段等標籤，以供後級語音技術所需的韻律結構與語言資訊。

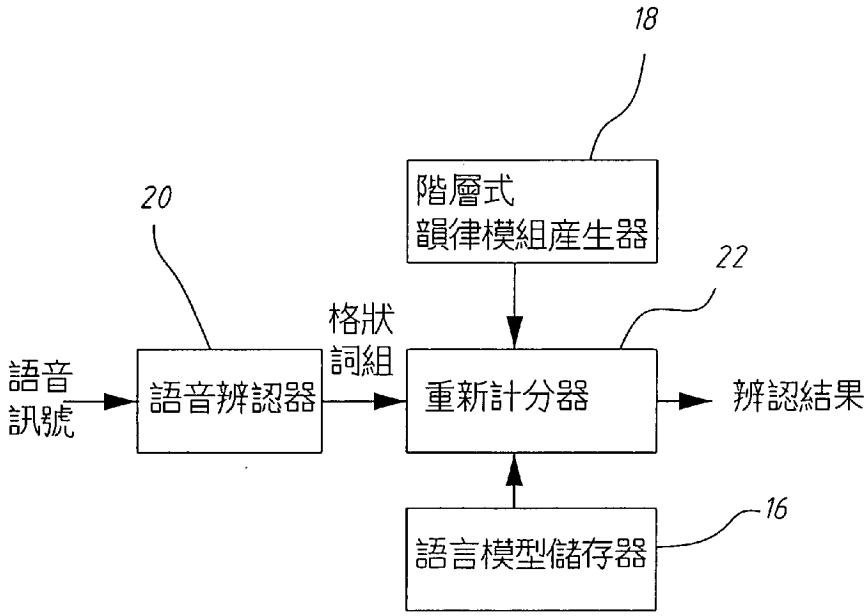
A Chinese speech recognition device and speech recognition method thereof is disclosed. Firstly, a speech signal is received and recognized to output a word lattice. Next, the word lattice is received, and a grade at each word arc of the word lattice is recalculated and ranked by a break syntax model, a prosodic state model, a syllable prosodic acoustic model, a syllable-juncture prosodic-acoustic model and a factored language model, so as to output a language tag, a prosodic tag and a phonetic tag, which correspond to the speech signal. The present invention estimates the grades by a two stage way to enhance the recognition rate of the basic speech information and labels the language tag, prosody tag and phonetic tag, which are used as the prosody structure and the language information that the speech rear stage requires.

16・・・語言模型儲存器

18・・・階層式韻律模型產生器

20・・・語音辨識器

22・・・重新計分器



第 2 圖

公告本

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：100142341

※申請日：100.11.18　※IPC分類：G10L 15/00 (2013.01)

一、發明名稱：（中文/英文）

中文語音辨識裝置及其辨識方法 / Chinese speech recognition device and speech recognition method thereof

二、中文發明摘要：

本發明係揭露一種中文語音辨識裝置及其辨識方法，首先，接收一語音訊號，以進行辨識後，輸出一格狀詞組（word lattice）。接著，接收格狀詞組，並根據一韻律停頓模型、一韻律狀態模型、一音節韻律聲學模型、一音節間韻律聲學模型與一因子化語言模型，重新計算格狀詞組中詞弧上的分數，將其重新排名，以輸出語音訊號對應之一語言標籤、一韻律標籤與一音段標記。本發明利用兩階段方式重新計分，不僅可提升基本語音資訊的辨識率，更可標記出語言、韻律、音段等標籤，以供後級語音技術所需的韻律結構與語言資訊。

三、英文發明摘要：

A Chinese speech recognition device and speech recognition method thereof is disclosed. Firstly, a speech signal is received and recognized to output a word lattice. Next, the word lattice is received, and a grade at each word arc of the word lattice is recalculated and ranked by a break syntax model, a prosodic state model, a syllable prosodic acoustic model, a syllable-juncture prosodic-acoustic model and a factored language model, so as to output a language tag, a prosodic tag and a phonetic tag, which correspond to the speech signal. The present invention estimates the grades by a two stage way to enhance the recognition rate of the basic speech information and labels the language tag, prosody tag and phonetic tag, which are used as the prosody structure and the language information that the speech rear stage requires.

四、指定代表圖：

(一)本案指定代表圖為：第（ 2 ）圖。

(二)本代表圖之元件符號簡單說明：

16 語言模型儲存器　　18 階層式韻律模型產生器

20 語音辨識器　　22 重新計分器

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】

本發明係有關一種辨識技術，特別是關於一種中文語音辨識裝置及其辨識方法。

【先前技術】

韻律輔助語音辨認是近幾年來重要的研究議題。韻律是指在連續語音中的超音段（suprasegmental）的特徵現象，如重音、聲調、停頓、語調及節奏等；如果將韻律現象以物理特性表現出，通常會出現在語音中音高軌跡、能量強度、語音長度及停頓的變化之中。而且韻律與各種層次的語言特徵參數有高度的關聯性，從音素（phone）、音節（syllable）、詞（word）、片語（phrase）到句子（sentence）甚至是更高層次的語言參數，由於它們之間的關係，所以韻律資訊對於提升語音辨認的準確度是會有幫助的。

從過去韻律輔助語音辨認的文獻中，歸納出如第 1 圖的語音模型產生裝置方塊圖，其包含一韻律模式訓練器 10、一特徵參數抽取器 12 與一人工標記韻律語料庫 14。運作上，人工標記韻律語料庫 14 是輸入語音資料，然後請專家標記韻律標籤；特徵參數抽取器 12 是根據人工標記韻律語料庫 14 抽取出頻譜特徵參數、各種層次之語言特徵參數及韻律聲學特徵參數；韻律模式訓練器 10 是根據特徵參數抽取器 12 的各種輸出特徵參數，與人工標記韻律語料庫 14 中找出的韻律線索或事件，如音高重音（pitch accent）及語調短語（intonational phrase）邊界等韻律線索，建立韻律相依聲學模型、韻律相依語言模型及韻律模型，以描述不同層次之語言特徵參數上的韻律線索與及其韻律聲學特徵參數的關係。

上述這些方法主要的限制在於缺乏大量具可靠及多元韻律標記的大型語料庫，因此只能使用一些少量且十分明顯的韻律線索，因而導致對語音辨認效能的改善十分有限。

因此，本發明係在針對上述之困擾，提出一種中文語音辨識裝置及其辨識方法，以解決習知所產生的問題。

【發明內容】

本發明之主要目的，在於提供一種中文語音辨識裝置及其辨識方法，其係利用韻律狀態模型、韻律停頓模型、音節韻律模型及音節間韻律模型，來改善中文搶詞及聲調的問題，並同時提升中文字、詞、基本音節的辨識率，亦標記出詞類、標點符號、韻律停頓及韻律狀態等標籤，日後可供後級語音轉換及語音合成所需的韻律結構與語言資訊。

為達上述目的，本發明提供一種中文語音辨識裝置，包含存有一因子化語言模型（Factored Language Model）之一語言模型儲存器、產生一韻律停頓模型、一韻律狀態模型、一音節韻律聲學模型與一音節間韻律聲學模型之一階層式韻律模型產生器與一語音辨識器，語音辨識器係接收一語音訊號，以對其進行辨識後，輸出一格狀詞組（word lattice）。上述元件皆連接一重新計分器，其係接收格狀詞組，且重新計分器根據韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型與因子化語言模型，重新計算格狀詞組中詞弧上的分數，將其重新排名，以輸出語音訊號對應之一語言標籤、一韻律標籤與一音段標記。

本發明亦提供一種中文語音辨識方法，首先，接收一語音訊號，以對其進行辨識後，輸出一格狀詞組。接著，接收格狀詞組，且根據一韻律停

頓模型、一韻律狀態模型、一音節韻律聲學模型、一音節間韻律聲學模型與一因子化語言模型，重新計算格狀詞組中詞弧上的分數，將其重新排名，以輸出語音訊號對應之一語言標籤、一韻律標籤與一音段標記，便完成辨識方法。

茲為使　貴審查委員對本發明之結構特徵及所達成之功效更有進一步之瞭解與認識，謹佐以較佳之實施例圖及配合詳細之說明，說明如後：

【實施方式】

為了介紹本發明之最佳實施例，以式（1）來表示，此式（1）是用來解碼出最佳的語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$、韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，與音段標記 γ_S：

$$\Lambda_l^*,\Lambda_p^*,\gamma_S^*=\arg\max_{\Lambda_l,\Lambda_p,\gamma_S} P(\mathrm{W,POS,PM,B,P},\gamma_S \mid \mathrm{Xa,Xp}) \approx$$

$$\arg\max_{\Lambda_l,\Lambda_p,\gamma_S} \{P(\mathrm{Xa},\gamma_S \mid \mathrm{W})\,P(\mathrm{W,POS,PM})$$

$$\cdot P(\mathrm{B} \mid \Lambda_l)\,P(\mathrm{P} \mid \mathrm{B})\,P(\mathrm{X} \mid \gamma_S,\Lambda_p,\Lambda_l)\,P(\mathrm{Y,Z} \mid \gamma_S,\Lambda_p,\Lambda_l)\} \quad (1)$$

式（1）中的 $P(\mathrm{B}|\Lambda_l)$、$P(\mathrm{P|B})$、$P(\mathrm{X} \mid \gamma_S,\Lambda_p,\Lambda_l)$、$P(\mathrm{Y,Z} \mid \gamma_S,\Lambda_p,\Lambda_l)$ 分別表示韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型。其中，$\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為韻律停頓序列，P={p,q,r}為韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為語音訊號之音節之總數量，韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲

學特徵參數、Z 為一音節間之差分特徵參數。

以下請參閱第 1 圖，本發明包含一語言模型儲存器 16，其係存有複數型態的語言模型，包含一因子化語言模型（Factored Language Model），其係模擬詞、詞類及標點符號，以提供不同層次的語言參數來幫助預估韻律模型。另有一階層式韻律模型產生器 18，其係產生複數型態的韻律模型，包含上述之韻律停頓模型、韻律狀態模型、音節韻律聲學模型與音節間韻律聲學模型，以改善中文搶詞及聲調之問題。此外，更利用一語音辨識器 20 接收一語音訊號。由於語音辨識器 20 存有一基礎聲學模型（acoustic model）與一雙連文語言模型（bigram language model），因此可藉此對語音訊號進行辨識，以輸出一格狀詞組（word lattice）。語言模型儲存器 16、階層式韻律模型產生器 18 與語音辨識器 20 皆連接一重新計分器 22，其係接收格狀詞組，且重新計分器 22 根據基礎聲學模型、韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型、因子化語言模型與式（2），重新計算格狀詞組中詞弧上的分數，將其重新排名，以輸出語音訊號對應之上述之語言標籤 Λ_l、韻律標籤 Λ_p 與音段標記 γ_S，此不但可提升中文字、詞、基本音節的辨識率，亦標記出詞類、標點符號、韻律停頓及韻律狀態等標籤，日後可供後級語音轉換及語音合成所需的韻律結構與語言資訊。

$$L(S,\Lambda_a)=\sum_{j=1}^{16}\alpha_j \log p_j \qquad (2)$$

其中 $S=[p_1,\ldots,p_{16}]$ 是一個向量，$p_1 \sim p_{16}$ 為依據基礎聲學模型、韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型與

因子化語言模型所構成之 16 個機率，$\Lambda_a = [\alpha_1,...,\alpha_{16}]$ 為利用鑑別式模型組合（discriminative model combination）演算法決定之權重向量。

請同時參閱第 3 圖與第 4 圖。階層式韻律模型產生器更包含一原始語料庫 24，其係存有複數聲音檔及其文字內容。原始語料庫 24 連接一特徵參數抽取器 26，其係依據聲音檔與文字內容，抽取複數種低層次語言參數、複數種高層次語言參數、一音高（pitch）、一音節長度（syllable duration）與一韻律能量之相關複數韻律聲學參數輸出之，其中低層次語言參數包含聲調 t、基本音節 s 與韻母 f；高層次語言參數則包含詞序列 W、詞類序列 POS 與標點符號序列 PM。另有一中文韻律階層結構儲存器 28，其係存有複數種韻律成分與複數種韻律停頓標籤，每一韻律停頓標籤係區分每一韻律成分。在此實施例中，韻律停頓標籤以四種為例，如第 4 圖所示，即第一類韻律停頓 B0/B1、第二類韻律停頓 B2、第三類韻律停頓 B3、第四類韻律停頓 B4，又韻律成分包含音節 SYL、韻律詞 PW、韻律片語 PPh 與呼吸群組 BG 或韻律片語群組 PG 兩者之其一者。特徵參數抽取器 26 與中文韻律階層結構儲存器 28 連接一韻律模式訓練器 32，其係擷取韻律停頓標籤、低層次語言參數、高層次語言參數、音高、音節長度與韻律能量之相關韻律聲學參數，以藉此預估韻律聲學特徵參數序列 Xp、韻律狀態序列 P 與韻律停頓序列 B，並使韻律狀態序列 P 與韻律停頓序列 B 藉由相關之韻律聲學特徵參數序列 Xp 強化之。韻律模式訓練器 32 以最大似然性原則（maximum likelihood criterion）調整韻律狀態序列 P 與韻律停頓序列 B，以藉此與韻律聲學特徵參數序列 Xp、以依次序最佳化演算法（sequential optimal algorithm）訓練出韻律停頓模型、韻律狀態模型、音節韻律聲學模

型、音節間韻律聲學模型輸出之，且自動標記韻律狀態序列 P 與韻律停頓序列 B 於語音訊號上。本發明利用大型未標記韻律的原始語料庫 24，進行韻律標記及建立韻律模式，不但省時間又省成本。

以下介紹各模型，首先介紹因子化語言模型，其係以式（3）表示：

$$P(\mathrm{W,PM,POS}) = \prod_{i=1}^{M} \{P(w_i \mid w_{i-2}^{i-1}) \cdot P(pos_i \mid pos_{i-1}, w_i) \cdot P(pm_{i-1} \mid pos_{i-1}^{i}, w_{i-1})\} \quad (3)$$

其中 w_i 為第 i 個詞，pos_i 為第 i 個詞類標籤，pm_i 為第 i 個標點符號標籤。

韻律停頓模型以式（4）表示：

$$P(\mathrm{B} \mid \Lambda_l) = \prod_{n=1}^{N-1} P(B_n \mid L_n) \quad (4)$$

其中 L_n 為第 n 個音節的文本相關的語言特徵參數。

韻律狀態模型以式（5）表示：

$$P(\mathrm{P|B}) = P(\mathrm{p|B})\ P(\mathrm{q|B})P(\mathrm{r|B}) =$$

$$P(p_1)P(q_1)P(r_1)[\prod_{n=2}^{N} P(p_n \mid p_{n-1}, B_{n-1})P(q_n \mid q_{n-1}, B_{n-1})P(r_n \mid r_{n-1}, B_{n-1})] \quad (5)$$

其中 p_n、q_n、r_n 分別為第 n 個音節的音節音高層次、音節長度層次與音節能量層次。

音節韻律聲學模型以式（6-1）表示：

$$P(\mathrm{X} \mid \gamma_S, \Lambda_p, \Lambda_l) =$$

$$P(\mathrm{sp} \mid \gamma_S, \mathrm{B,p,t})\ P(\mathrm{sd} \mid \gamma_S, \mathrm{B,q,t,s})\ P(\mathrm{se} \mid \gamma_S, \mathrm{B,r,t,f})$$

$$= \prod_{n=1}^{N} P(sp_n \mid p_n, B_{n-1}^{n}, t_{n-1}^{n+1}) P(sd_n \mid q_n, s_n, t_n) P(se_n \mid r_n, f_n, t_n) \quad (6\text{-}1)$$

其中 sp 為音高輪廓，sd 為音節長度，se 為音節能量，sp_n、sd_n、se_n、t_n、s_n、f_n 分別為第 n 個音節的音高輪廓、音節長度、音節能量、聲調、基本音節與韻母。

$P(sp_n \mid p_n, B_{n-1}^n, t_{n-1}^{n+1})$、$P(sd_n \mid q_n, s_n, t_n)$、$P(se_n \mid r_n, f_n, t_n)$ 分別為第 n 個音節的音高輪廓、音節長度、音節能量之子模型。$B_{n-1}^n=(B_{n-1},B_n)$；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，這些影響因子並以加成方式去結合一塊，以第 n 個音節的音高輪廓為例，可得式（6-2）：

$$sp_n = sp_n^r + \beta_{t_n} + \beta_{p_n} + \beta^f_{B_{n-1},t_{n-1}^n} + \beta^b_{B_n,t_n^{n+1}} + \mu_{sp} \qquad (6\text{-}2)$$

其中 sp_n 為一四維正交化係數用以表達第 n 個音節觀察到的音高輪廓，sp_n^r 為正規化的 sp_n，β_{t_n} 和 β_{p_n} 分別為聲調和韻律狀態的影響因子，$\beta^f_{B_{n-1},t_{n-1}^n}$ 和 $\beta^b_{B_n,t_n^{n+1}}$ 為向前及向後連音影響因子，μ_{sp} 為音高的全域平均值。基於假設 sp_n^r 為零平均值和正規分佈，所以以常態分佈來表示，可得式（6-3）：

$$P(sp_n \mid p_n, B_{n-1}^n, t_{n-1}^{n+1})$$
$$= N(sp_n; \beta_{t_n} + \beta_{p_n} + \beta^f_{B_{n-1},t_{n-1}^n} + \beta^b_{B_n,t_n^{n+1}} + \mu_{sp}, R_{sp}) \qquad (6\text{-}3)$$

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

音節間韻律聲學模型以式（7-1）表示：

$$P(\mathrm{Y,Z} \mid \gamma_S, \Lambda_p, \Lambda_l) =$$

$$P(\mathrm{pd,ed,pj,dl,df} \mid \gamma_S, \Lambda_p, \Lambda_l$$

$$= \prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n, df_n \mid \gamma_S, B_n, \Lambda_{l,n}) \qquad (7\text{-}1)$$

其中 pd、ed、pj 分別為短停頓長度、能量下降程度、正規化音高差序，dl、df 皆為正規化音高拉長因子，且 pd_n 、ed_n 、pj_n 分別為在第 n 個音節所跟隨的接合點（juncture）的短停頓長度、能量下降程度、正規化音高差序，dl_n 、df_n 皆為在第 n 個音節所跟隨的接合點的正規化音高拉長因子。pj_n 、dl_n 、df_n 分別以式（7-2）、(7-3)、(7-4）定義之：

$$pj_n = (sp_{n+1}(1) - \beta_{t_{n+1}}(1)) - (sp_n(1) - \beta_{t_n}(1)) \quad (7\text{-}2)$$

其中 $sp_n(1)$ 為 sp_n 的第一維度（即音節音高平均值）；$\beta_{t_n}(1)$ 為聲調影響因子的第一維度。

$$dl_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n-1} - \gamma_{t_{n-1}} - \gamma_{s_{n-1}}) \quad (7\text{-}3)$$

$$df_n = (sd_n - \gamma_{t_n} - \gamma_{s_n}) - (sd_{n+1} - \gamma_{t_{n+1}} - \gamma_{s_{n+1}}) \quad (7\text{-}4)$$

上述 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，$\Lambda_{l,n}$ 的空間仍是太大，所以將 $\Lambda_{l,n}$ 分成幾類，然後同時估計 Gamma 及常態分佈的參數。

上述四種韻律模型所使用的分佈及方法可視實際情況調整，而非用來限制本發明之範圍。

以下介紹本發明之兩階段式運作過程，請參閱第 2 圖。首先，語音辨認器 20 接收語音訊號，以利用基礎聲學模型與雙連文語言模型對其進行辨識後，輸出格狀詞組。接著，重新計分器 22 接收格狀詞組，且根據基礎聲學模型、韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型、因子化語言模型與式(2)，重新計算格狀詞組中詞弧上的分數，

將其重新排名，以輸出語音訊號對應之語言標籤Λ_l、韻律標籤Λ_p與音段標記γ_S。

以下介紹階層式韻律模組產生器 18 產生韻律停頓模型、韻律狀態模型、音節韻律聲學模型與音節間韻律聲學模型之過程，請繼續參閱第 3 圖。首先，特徵參數抽取器 26 依據原始語料庫 24 中的複數聲音檔及其文字內容，抽取低層次語言參數、高層次語言參數、音高、音節長度與韻律能量輸出之。接著，韻律模式訓練器 32 從中文韻律階層結構儲存器 28 與特徵參數抽取器 26 擷取韻律停頓標籤、低層次語言參數、高層次語言參數、音高、音節長度與韻律能量，以藉此預估韻律聲學特徵參數序列 Xp、韻律狀態序列 P 與韻律停頓序列 B，並使韻律狀態序列 P 與韻律停頓序列 B 藉由相關之韻律聲學特徵參數序列 Xp 強化之。最後，韻律模式訓練器 32 以最大似然性原則調整韻律狀態序列 P 與韻律停頓序列 B，以藉此與韻律聲學特徵參數序列 Xp、以依次序最佳化演算法訓練出韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型輸出之，且自動標記韻律狀態序列 P 與韻律停頓序列 B 於語音訊號上。

當韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型皆被訓練出來後，其係與低層次語言參數、高層次語言參數、韻律狀態序列 P、韻律停頓序列 B 、音節韻律聲學特徵參數 X、音節間之韻律聲學特徵參數 Y、音節間之差分特徵參數 Z 所建立的關係如第 5 圖所示。由圖可知，高層次語言參數係藉式（4）之韻律停頓模型得到韻律停頓序列 B；韻律停頓序列 B 與高層次語言參數藉由式（7-1）之音節間韻律聲學模型係得到音節間之韻律聲學特徵參數 Y、音節間之差分特徵參數 Z；韻律停

頓序列 B 藉由式（5）之韻律狀態模型係得到韻律狀態序列 P；以及韻律狀態序列 P、韻律停頓序列 B 與低層次語言參數藉由式（6）之音節韻律聲學模型得到音節韻律聲學特徵參數 X。

下表一為語音辨認的實驗結果，它是在多語者中文連續語音資料庫中，實地測試第 2 圖實施例之語者不相關辨認結果。此資料庫包含 303 個語者，隨機從中挑選約 90%其包含 274 個語者約 23 小時的語料來訓練系統，剩餘約 10%的部分其包含 29 個語者約 2.43 小時當作測試語料，但是為了觀察豐富標記輸出的結果，本發明挑選出長文部分其包含 19 個語者約 2 小時來做系統測試。由表一看出本發明比只使用因子化語言模型的基礎系統有更好的效能，本發明在詞、字和基本音節的錯誤率分別是 20.7%、14.4%和 9.6%，當此結果與基礎系統作比較時，其絕對的錯誤下降率分別為 3.7%、3.7%和 2.4% （或相對錯誤下降為 15.2%、20.4%和 20%）。

表一

	詞錯誤率	字錯誤率	基本音節錯誤率
基礎系統	24.4	18.1	12.0
本發明	20.7	14.4	9.6

表二為詞類解碼的實驗結果，其基礎系統的精確度、召回率及 F 量測分別為 93.4%、76.4%及 84.0%；而本發明分別為 93.4%、80.0%及 86.2%。表三為標點符號解碼的實驗結果，其基礎系統的精確度、召回率及 F 量測分別為 55.2%、37.8%及 44.8%；而本發明分別為 61.2%、53.0%及 56.8%。表四為聲調解碼的實驗結果，其基礎系統的精確度、召回率及 F 量測分別為 87.9%、87.5%及 87.7%；而本案發明分別為 91.9%、91.6%及 91.7%。

表二

	精確度	召回率	F 量測
基礎系統	93.4	76.4	84.0
本發明	93.4	80.0	86.2

表三

	精確度	召回率	F 量測
基礎系統	55.2	37.8	44.8
本發明	61.2	53.0	56.8

表四

	精確度	召回率	F 量測
基礎系統	87.9	87.5	87.7
本發明	91.9	91.6	91.7

本發明之聲音波形及其對應之各種語音標記結果範例如第 6 圖所示。在第 6 圖中，由上依序而下分別為範例音檔之聲音波形、音高層次之韻律狀態、音節長度層次之韻律狀態、音節能量層次之韻律狀態、韻律停頓的標記（不含 B0 與 B1，為簡潔表示）、範例音檔之正確內容文字、根據韻律停頓的標記所建構出來的範例音檔之語法片語結構、解碼出的詞彙、解碼出的詞類及標點符號及符號意義表示。

此聲音波形的時間單位為秒，其中表示三角形的符號為短停頓（short pause, sp），由波形可以觀察出有四個韻律片語（PPh），而本實施例也確實解碼出四個 PPh 由 B3 所分開出來，每一個 PPh 甚至解碼出韻律詞的結果（PW）是由 B2 所區分出來，如語法片語結構所示；從音高層次之韻律狀態中可以觀察出，在 B3 位置時出現重大的音高重置現象；在音節長度層次之韻律狀態中，B2-3 的位置顯示出前一個音節長度有拉長現象，由這些標

記結果顯示韻律停頓與韻律狀態呈現出階層式韻律結構。

綜上所述，本發明利用兩階段方式重新計分，不但能提升基本語音辨識率，更標記出語言、韻律、音段等標籤，以供後續使用。

以上所述者，僅為本發明一較佳實施例而已，並非用來限定本發明實施之範圍，故舉凡依本發明申請專利範圍所述之形狀、構造、特徵及精神所為之均等變化與修飾，均應包括於本發明之申請專利範圍內。

【圖式簡單說明】

第 1 圖為先前技術之語音模型產生裝置方塊圖

第 2 圖為本發明之裝置方塊圖。

第 3 圖為本發明之階層式韻律模型產生器方塊圖。

第 4 圖為本發明之韻律成分與韻律停頓標籤之示意圖。

第 5 圖為本發明之韻律停頓模型、韻律狀態模型、音節韻律聲學模型、音節間韻律聲學模型與各種語音參數關係示意圖。

第 6 圖為本發明之聲音波形及其對應之各種語音標記示意圖。

【主要元件符號說明】

10　韻律模式訓練器

12　特徵參數抽取器

14　人工標記韻律語料庫

16　語言模型儲存器

18　階層式韻律模型產生器

20　語音辨識器

22　重新計分器

24　原始語料庫

26　特徵參數抽取器

28　中文韻律階層結構儲存器

32　韻律模式訓練器

七、申請專利範圍：

1. 一種中文語音辨識裝置，包含：

一語言模型儲存器，可存放複數型態的語言模型，包含一因子化語言模型（Factored Language Model）；

一階層式韻律模型產生器，可產生複數型態的韻律模型，包含一韻律停頓模型、一韻律狀態模型、一音節韻律聲學模型與一音節間韻律聲學模型；

一語音辨識器，接收一語音訊號，以對其進行辨識後，輸出一格狀詞組（word lattice）；以及

一重新計分器，連接該語言模型儲存器、該階層式韻律模型產生器與該語音辨識器，以接收該格狀詞組，且該重新計分器根據該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型與該因子化語言模型，重新計算該格狀詞組中詞弧上的分數，將其重新排名，以輸出該語音訊號對應之一語言標籤、一韻律標籤與一音段標記。

2. 如請求項 1 所述之中文語音辨識裝置，其中該階層式韻律模型產生器更包含：

一原始語料庫，存有複數聲音檔及其文字內容；

一特徵參數抽取器，連接該原始語料庫，並依據該些聲音檔與該文字內容，抽取複數種低層次語言參數、複數種高層次語言參數、一音高（pitch）、一音節長度（syllable duration）與一韻律能量之相關複數韻律聲學參數輸出之；

一中文韻律階層結構儲存器，其係存有複數種韻律成分與複數種韻律停頓標籤，該些韻律停頓標籤係區分每一該韻律成分；以及

一韻律模式訓練器，連接該特徵參數抽取器與該中文韻律階層結構儲存器，並擷取該些韻律停頓標籤、該些低層次語言參數、該些高層次語言參數、該音高、該音節長度與該韻律能量之相關該些韻律聲學參數，以藉此預估一韻律聲學特徵參數序列 Xp、一韻律狀態序列 P 與一韻律停頓序列 B，該韻律模式訓練器更調整該韻律狀態序列 P 與該韻律停頓序列 B，以藉此與該韻律聲學特徵參數序列 Xp 訓練出該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型輸出之，且自動標記該韻律狀態序列 P 與該韻律停頓序列 B 於該語音訊號上。

3. 如請求項 2 所述之中文語音辨識裝置，其中該些韻律成分包含音節、韻律詞、韻律片語與呼吸群組或韻律片語群組兩者之其一者。

4. 如請求項 2 所述之中文語音辨識裝置，其中該韻律模式訓練器以最大似然性原則（maximum likelihood criterion）調整該韻律狀態序列 P 與該韻律停頓序列 B。

5. 如請求項 2 所述之中文語音辨識裝置，其中該韻律模式訓練器以依次序最佳化演算法（sequential optimal algorithm），並藉該韻律狀態序列 P、該韻律停頓序列 B 與該韻律聲學特徵參數序列 Xp 訓練出該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型。

6. 如請求項 2 所述之中文語音辨識裝置，其中該因子化語言模型係以下列公式表示：

$$P(\mathrm{W,PM,POS})=\prod_{i=1}^{M}\{P(w_i \mid w_{i-2}^{i-1})\cdot P(pos_i \mid pos_{i-1}, w_i)\cdot P(pm_{i-1} \mid pos_{i-1}^{i}, w_{i-1})\}$$

，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，w_i 為第 i 個該詞，pos_i 為第 i 個詞類標籤，pm_i 為第 i 個標點符號標籤。

7. 如請求項 2 所述之中文語音辨識裝置，其中該韻律停頓模型 $P(\mathrm{B}\mid\Lambda_l)=\prod_{n=1}^{N-1}P(B_n \mid L_n)$，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，L_n 為第 n 個該音節的文本相關的語言特徵參數。

8. 如請求項 2 所述之中文語音辨識裝置，其中該韻律狀態模型 $P(\mathrm{P}\mid\mathrm{B})=$

$P(\mathrm{p}|\mathrm{B})\ P(\mathrm{q}|\mathrm{B})P(\mathrm{r}|\mathrm{B})=$

$P(p_1)P(q_1)P(r_1)[\prod_{n=2}^{N}P(p_n\mid p_{n-1},B_{n-1})P(q_n\mid q_{n-1},B_{n-1})P(r_n\mid r_{n-1},B_{n-1})]$，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$為標點符號序列，$M$ 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，p_n、q_n、r_n 分別為第 n 個該音節的該音節音高層次、該音節長度層次與該音節能量層次。

9. 如請求項 8 所述之中文語音辨識裝置，其中該音節韻律聲學模型 $P(\mathrm{X}|\gamma_S,\Lambda_p,\Lambda_l)=$

$P(\mathrm{sp}|\gamma_S,\mathrm{B,p,t})\ P(\mathrm{sd}|\gamma_S,\mathrm{B,q,t,s})\ P(\mathrm{se}|\gamma_S,\mathrm{B,r,t,f})$

$=\prod_{n=1}^{N}P(sp_n\mid p_n,B_{n-1}^n,t_{n-1}^{n+1})P(sd_n\mid q_n,s_n,t_n)P(se_n\mid r_n,f_n,t_n)$，其中該些低層次語言參數包含聲調 t、基本音節 s 與韻母 f，該些高層次語言參數包含該詞序列 W、該詞類序列 POS 與該標點符號序列 PM，sp 為音高輪廓，sd 為音節長度，se 為音節能量，sp_n、sd_n、se_n、t_n、s_n、f_n 分別為第 n 個該音節的該音高輪廓、該音節長度、該音節能量、該聲調、該基本音節與該韻母。

10. 如請求項 2 所述之中文語音辨識裝置，其中該音節間韻律聲學模型 P

$(\mathrm{Y,Z}|\gamma_S,\Lambda_p,\Lambda_l)=$

$P(\mathrm{pd,ed,pj,dl,df}|\gamma_S,\Lambda_p,\Lambda_l)=$

$\prod_{n=1}^{N-1} P(pd_n,ed_n,pj_n,dl_n,df_n|\gamma_S,B_n,\Lambda_{l.n})$，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$為標點符號序列，$M$ 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$為該韻律停頓序列，$\mathrm{P}=\{\mathrm{p,q,r}\}$為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 $\mathrm{Xp}=\{\mathrm{X,Y,Z}\}$，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，pd、ed、pj 分別為短停頓長度、能量下降程度、正規化音高差序，dl、df 皆為正規化音高拉長因子，且 pd_n、ed_n、pj_n 分別為在第 n 個該音節所跟隨的接合點的該短停頓長度、該能量下降程度、該正規化音高差序，dl_n、df_n 皆為在第 n 個該音節所跟隨的接合點的該正規化音高拉長因子。

11. 如請求項 1 所述之中文語音辨識裝置，其中該語音辨識器存有一基礎聲學模型（acoustic model）與一雙連文語言模型（bigram language model），並藉此對該語音訊號進行辨識，以輸出該格狀詞組。

12. 如請求項 11 所述之中文語音辨識裝置，其中該重新計分器係利用下列公式重新計算該分數：

$L(S,\Lambda_a)=\sum_{j=1}^{16}\alpha_j\log p_j$，其中 $S=[p_1,...,p_{16}]$ 是一個向量，$p_1 \sim p_{16}$ 為

依據該基礎聲學模型、該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型與該因子化語言模型所構成 16 個機率，$\Lambda_a = [\alpha_1, ..., \alpha_{16}]$ 為利用鑑別式模型組合（discriminative model combination）演算法決定之權重向量。

13.一種中文語音辨識方法，包含下列步驟：

接收一語音訊號，以對其進行辨識後，輸出一格狀詞組（word lattice）；以及

接收該格狀詞組，且根據一韻律停頓模型、一韻律狀態模型、一音節韻律聲學模型、一音節間韻律聲學模型與一因子化語言模型，重新計算該格狀詞組中詞弧上的分數，將其重新排名，以輸出該語音訊號對應之一語言標籤、一韻律標籤與一音段標記。

14.如請求項 13 所述之中文語音辨識方法，其中該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型與該音節間韻律聲學模型之產生方法，包含下列步驟：

依據複數聲音檔及其文字內容，抽取複數種低層次語言參數、複數種高層次語言參數、一音高（pitch）、一音節長度（syllable duration）與一韻律能量輸出之；

擷取區隔複數種韻律成分之複數種韻律停頓標籤、該些低層次語言參數、該些高層次語言參數、該音高、該音節長度與該韻律能量，以藉此預估一韻律聲學特徵參數序列 Xp、一韻律狀態序列 P 與一韻律停頓序列 B；以及

調整該韻律狀態序列 P 與該韻律停頓序列 B，以藉此與該韻律聲學特徵

參數序列 Xp 訓練出該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型輸出之，且自動標記該韻律狀態序列 P 與該韻律停頓序列 B 於該語音訊號上。

15.如請求項 14 所述之中文語音辨識方法，其中該些韻律成分包含音節、韻律詞、韻律片語與呼吸群組或韻律片語群組兩者之其一者。

16.如請求項 14 所述之中文語音辨識方法，其中在調整該韻律狀態序列 P 與該韻律停頓序列 B 之步驟中，係以最大似然性原則（maximum likelihood criterion）實行之。

17.如請求項 14 所述之中文語音辨識方法，其中在藉該韻律狀態序列 P、該韻律停頓序列 B 與該韻律聲學特徵參數序列 Xp 訓練出該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型之步驟中，係以依次序最佳化演算法（sequential optimal algorithm）訓練之。

18.如請求項 14 所述之中文語音辨識方法，其中該因子化語言模型係以下列公式表示：

$$P(\mathrm{W,PM,POS})=\prod_{i=1}^{M}\{P(w_i \mid w_{i-2}^{i-1})\cdot P(pos_i \mid pos_{i-1}, w_i)\cdot P(pm_{i-1} \mid pos_{i-1}^{i}, w_{i-1})\}$$

，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，

該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，w_i 為第 i 個該詞，pos_i 為第 i 個詞類標籤，pm_i 為第 i 個標點符號標籤。

19.如請求項 14 所述之中文語音辨識方法，其中該韻律停頓模型 $P(\mathrm{B}|\Lambda_l)=\prod_{n=1}^{N-1}P(B_n|L_n)$，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，L_n 為第 n 個該音節的文本相關的語言特徵參數。

20.如請求項 14 所述之中文語音辨識方法，其中該韻律狀態模型 $P(\mathrm{P|B})=P(\mathrm{p|B})\,P(\mathrm{q|B})P(\mathrm{r|B})=$

$P(p_1)P(q_1)P(r_1)[\prod_{n=2}^{N}P(p_n|p_{n-1},B_{n-1})P(q_n|q_{n-1},B_{n-1})P(r_n|r_{n-1},B_{n-1})]$，其中該語言標籤 $\Lambda_l=\{\mathrm{W,POS,PM}\}$，該韻律標籤 $\Lambda_p=\{\mathrm{B,P}\}$，該音段標記 γ_S，且 $\mathrm{W}=\{w_1^M\}=\{w_1,...,w_M\}$ 為詞序列，$\mathrm{POS}=\{pos_1^M\}=\{pos_1,...,pos_M\}$ 為詞類序列，$\mathrm{PM}=\{pm_1^M\}=\{pm_1,...,pm_M\}$ 為標點符號序列，M 為該語音訊號之詞之總數量，$\mathrm{B}=\{B_1^N\}=\{B_1,...,B_N\}$ 為該韻律停頓序列，P={p,q,r}為該韻律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層

次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，p_n 、q_n 、r_n 分別為第 n 個該音節的該音節音高層次、該音節長度層次與該音節能量層次。

21.如請求項 20 所述之中文語音辨識方法，其中該音節韻律聲學模型 P

$(\mathrm{X}|\gamma_S,\Lambda_p,\Lambda_l)=$

P (sp| γ_S ,B,p,t) P (sd| γ_S ,B,q,t,s) P (se| γ_S ,B,r,t,f)

$=\prod_{n=1}^{N} P(sp_n \mid p_n, B_{n-1}^{n}, t_{n-1}^{n+1}) P(sd_n \mid q_n, s_n, t_n) P(se_n \mid r_n, f_n, t_n)$，其中該些低層次語言參數包含聲調 t、基本音節 s 與韻母 f，該些高層次語言參數包含該詞序列 W、該詞類序列 POS 與該標點符號序列 PM，sp 為音高輪廓，sd 為音節長度，se 為音節能量，sp_n 、sd_n 、se_n 、t_n 、s_n 、f_n 分別為第 n 個該音節的該音高輪廓、該音節長度、該音節能量、該聲調、該基本音節與該韻母。

22.如請求項 14 所述之中文語音辨識方法，其中該音節間韻律聲學模型 P

$(\mathrm{Y,Z}|\gamma_S,\Lambda_p,\Lambda_l)=$

P (pd,ed,pj,dl,df|$\gamma_S,\Lambda_p,\Lambda_l$) =

$\prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n, df_n \mid \gamma_S, B_n, \Lambda_{l,n})$ ，其中該語言標籤 Λ_l = {W,POS,PM} ，該韻律標籤 Λ_p = {B,P} ，該音段標記 γ_S ，且 W={ w_1^M }={ $w_1,...,w_M$ }為詞序列，POS={ pos_1^M }={ $pos_1,...,pos_M$ }為詞類序列，PM={ pm_1^M }={ $pm_1,...,pm_M$ }為標點符號序列，M 為該語音訊號之詞之總數量，B={ B_1^N }={ $B_1,...,B_N$ }為該韻律停頓序列，P={p,q,r}為該韻

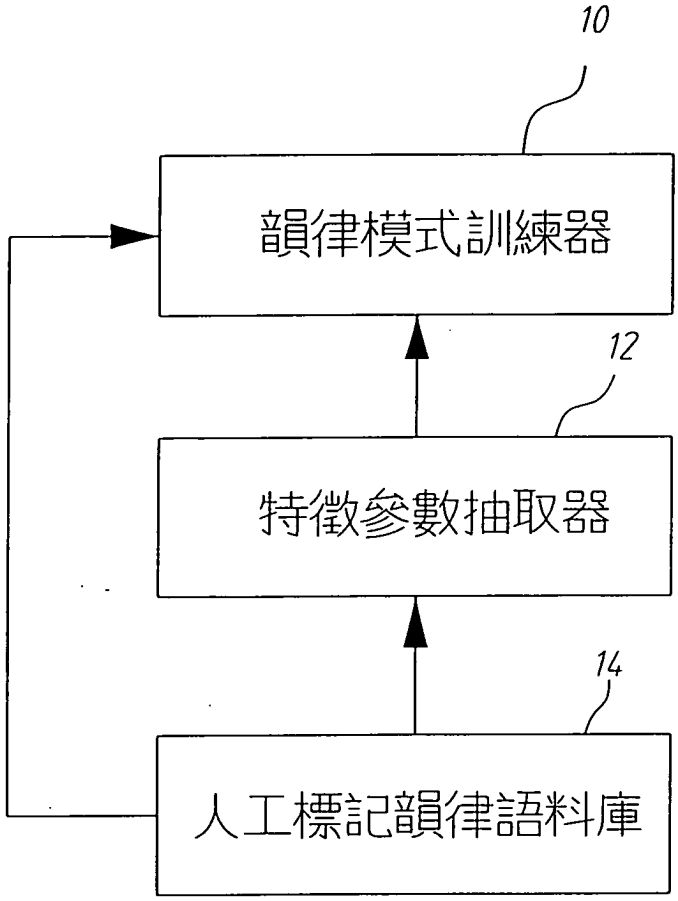
律狀態序列，p 為音節音高層次，q 為音節長度層次，r 為音節能量層次，N 為該語音訊號之音節之總數量，該韻律聲學特徵參數序列 Xp={X,Y,Z}，X 為一音節韻律聲學特徵參數、Y 為一音節間之韻律聲學特徵參數、Z 為一音節間之差分特徵參數，pd、ed、pj 分別為短停頓長度、能量下降程度、正規化音高差序，dl、df 皆為正規化音高拉長因子，且 pd_n、ed_n、pj_n 分別為在第 n 個該音節所跟隨的接合點的該短停頓長度、該能量下降程度、該正規化音高差序，dl_n、df_n 皆為在第 n 個該音節所跟隨的接合點的該正規化音高拉長因子。

23. 如請求項 13 所述之中文語音辨識方法，其中在對該語音訊號進行辨識之步驟中，係藉一基礎聲學模型（acoustic model）與一雙連文語言模型（bigram language model）辨識之。

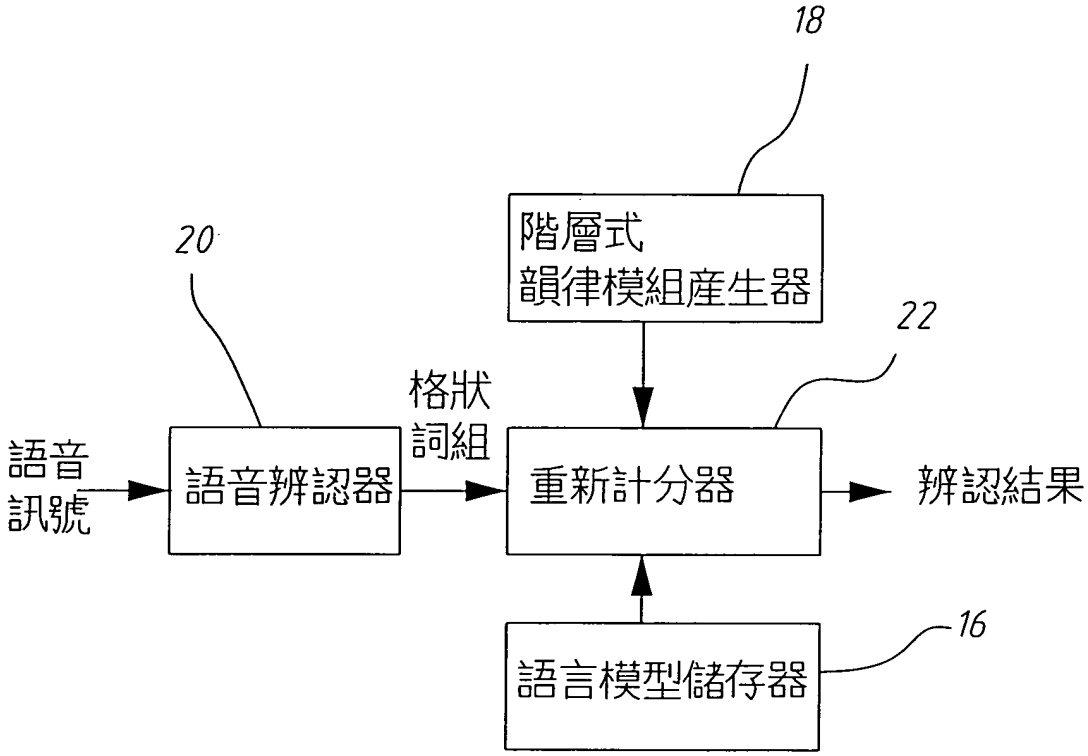
24. 如請求項 23 所述之中文語音辨識方法，其中在重新計算該分數之步驟中，係利用下列公式：

$L(S,\Lambda_a)=\sum_{j=1}^{16}\alpha_j \log p_j$，其中 $S=[p_1,\ldots,p_{16}]$ 是一個向量，$p_1 \sim p_{16}$ 為依據該基礎聲學模型、該韻律停頓模型、該韻律狀態模型、該音節韻律聲學模型、該音節間韻律聲學模型與該因子化語言模型所構成 16 個機率，$\Lambda_a=[\alpha_1,\ldots,\alpha_{16}]$ 為利用鑑別式模型組合（discriminative model combination）演算法決定之權重向量。

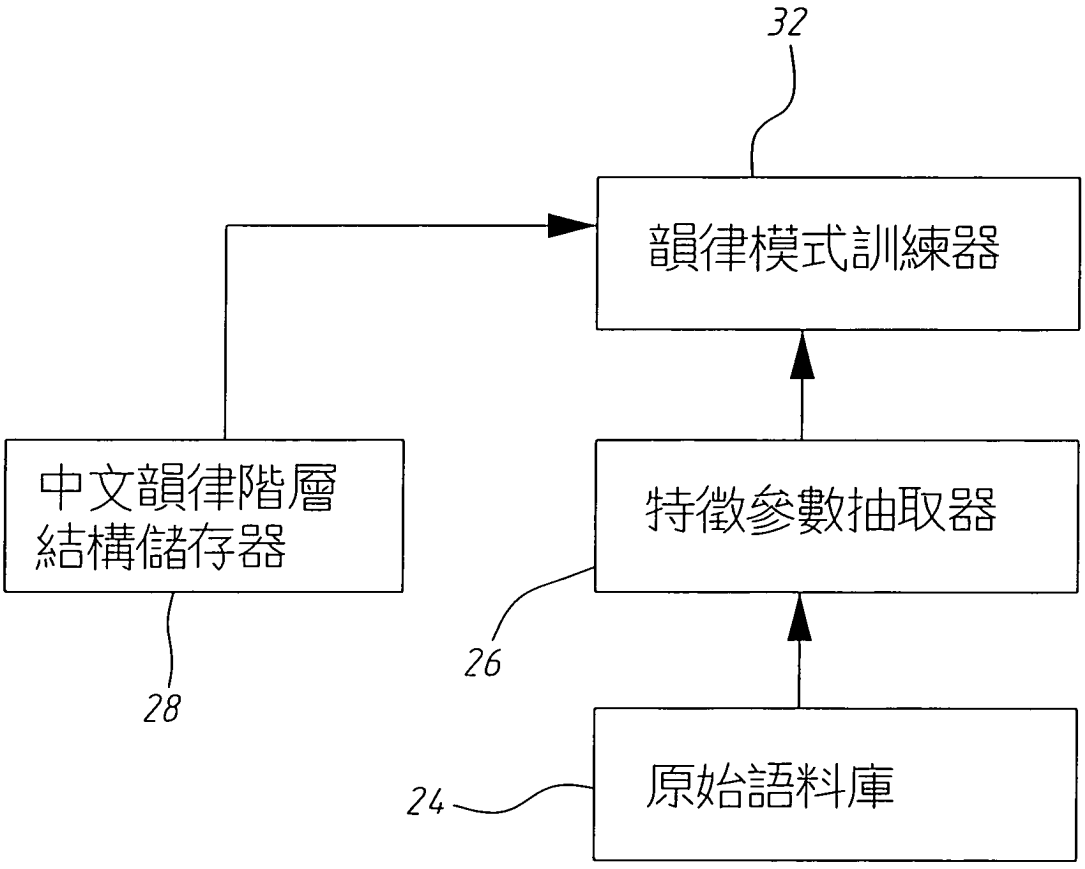
八、圖式：



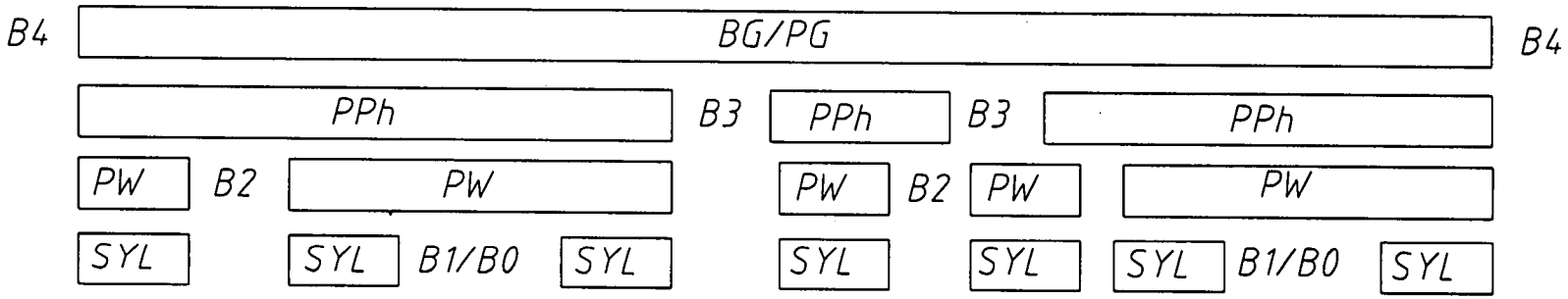
第 1 圖



第 2 圖



第 3 圖



SYL:音節

PW::韻律詞

PPh:韻律片語

BG/PG:呼吸群組或韻律片語群組

B1/B0::第一類韻律停頓

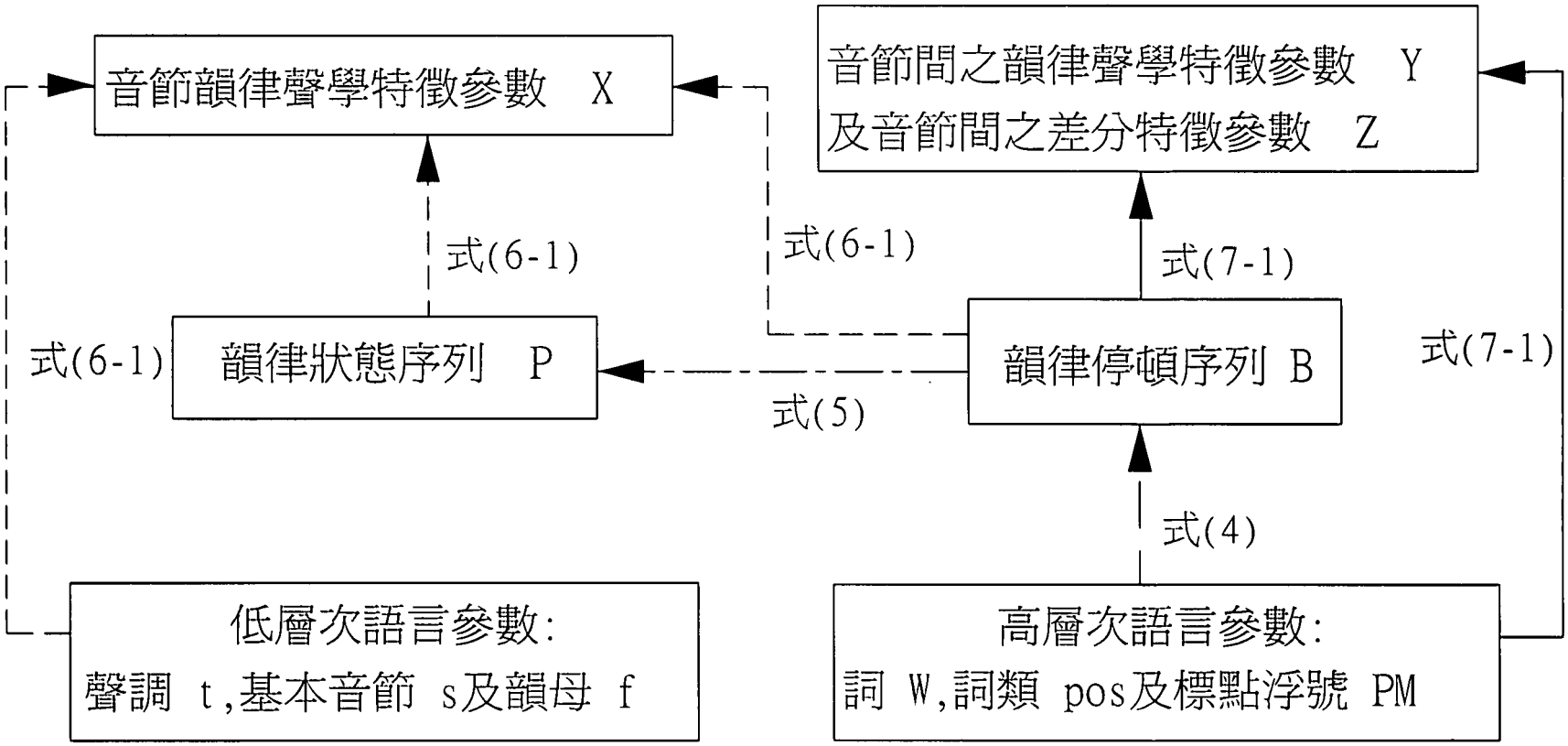
B2:第二類韻律停頓

B3:第三類韻律停頓

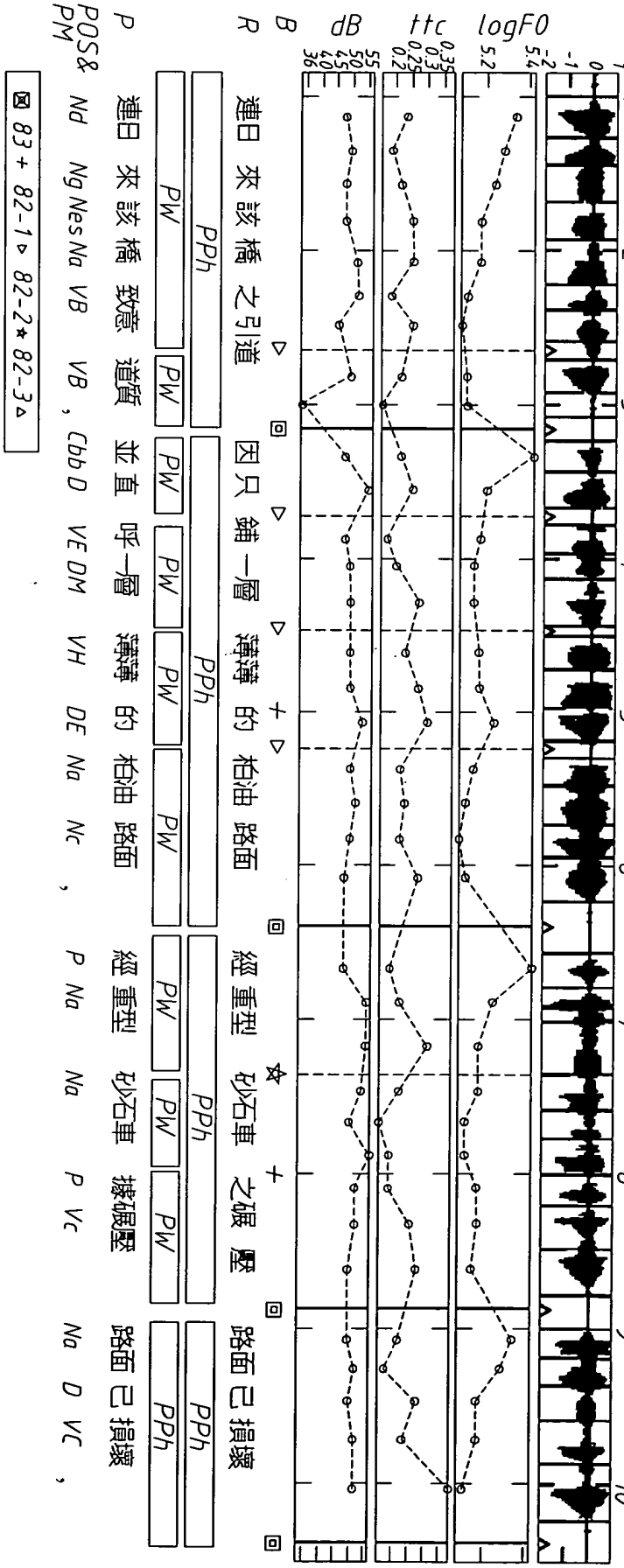
B4:第四類韻律停頓

第 4 圖

I441163


音節韻律聲學特徵參數 X
音節間之韻律聲學特徵參數 Y
及音節間之差分特徵參數 Z
式(6-1)
式(6-1)
式(7-1)
式(6-1)
韻律狀態序列 P
韻律停頓序列 B
式(7-1)
式(5)
式(4)
低層次語言參數:
聲調 t,基本音節 s及韻母 f
高層次語言參數:
詞 W,詞類 pos及標點浮號 PM

第 5 圖


時間單位:秒
logF0
ttc
dB
B
R
連日 來 該 橋 之引道 因只 鋪一層 薄薄 的 柏油 路面 經 重型 砂石車 之碾 壓 路面 已 損壞
PPh
PW
P
POS&
PM
連日 來 該 橋 致意 道質 並 直 呼一層 薄薄 的 柏油 路面 經 重型 砂石車 擄碾壓 路面 已 損壞
Nd Ng Nes Na VB VB , CbbD VE DM VH DE Na Nc , P Na Na P Vc Na D VC ,
83 + 82-1 82-2 82-3

第 6 圖