(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 201432668 A
(43)公開日： 中華民國 103 (2014) 年 08 月 16 日

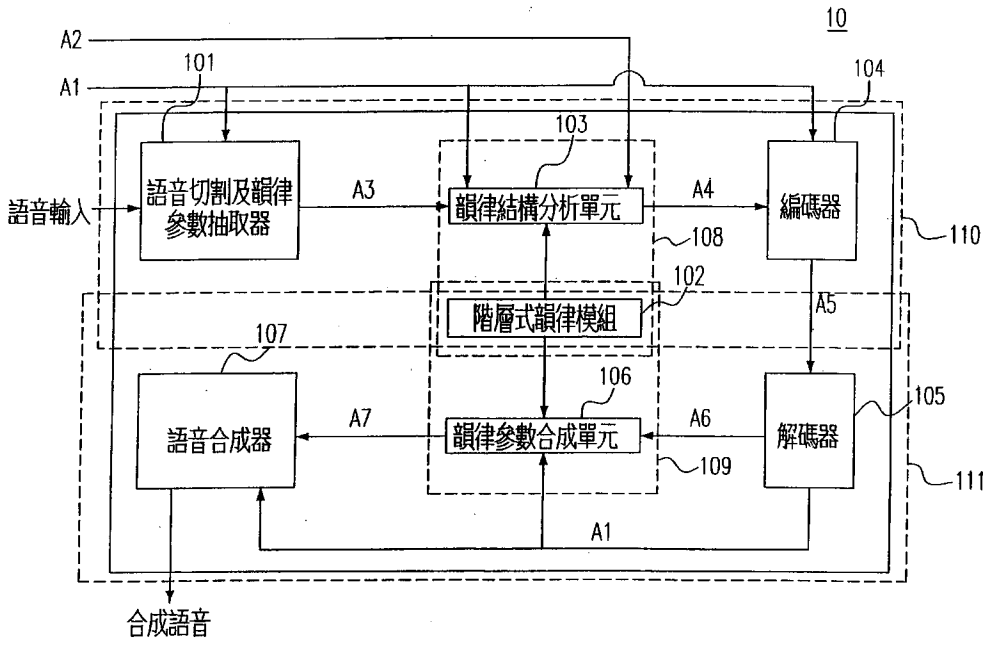
(21)申請案號：102104478 (22)申請日： 中華民國 102 (2013) 年 02 月 05 日
(51)Int. Cl.： *G10L13/08 (2013.01)*
(71)申請人：國立交通大學 (中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)
新竹市大學路 1001 號
(72)發明人：陳信宏 CHEN, SIN HORNG (TW) ；王逸如 WANG, YIH RU (TW) ；江振宇 CHIANG, CHEN YU (TW) ；謝喬華 HSIEH, CHIAO HUA (TW)
(74)代理人：蔡清福
申請實體審查：有 申請專利範圍項數：14 項 圖式數：5 共 36 頁

(54)名稱

編碼串流產生裝置、韻律訊息編碼裝置、韻律結構分析裝置與語音合成之裝置及方法
STREAMING ENCODER, PROSODY INFORMATION ENCODING DEVICE, PROSODY-ANALYZING DEVICE, AND DEVICE AND METHOD FOR SPEECH-SYNTHESIZING

(57)摘要

本案係提供一種語音合成之裝置，其包括一階層式韻律模組，提供一階層式韻律模型；一韻律結構分析單元，接收一低階語言參數、一高階語言參數及一第一韻律參數，且根據該高階語言參數、該低階語言參數、該第一韻律參數及該階層式韻律模組，產生至少一韻律標記；以及一韻律參數合成單元，根據該階層式韻律模組、該低階語言參數及該韻律標記來合成一第二韻律參數。



第一圖

10：語音合成裝置
101：語音切割及韻律參數抽取器
102：階層式韻律模組
103：韻律結構分析單元
104：編碼器
105：解碼器
106：韻律參數合成單元
107：語音合成器
108：韻律結構分析裝置
109：韻律參數合成裝置
110：韻律訊息編碼裝置

111：韻律訊息解碼裝置

A1：低階語言參數

A2：高階語言參數

A3：第一韻律參數

A4：第一韻律標記

A5：編碼串流

A6：第二韻律標記

A7：第二韻律參數

201432668

發明摘要

※ 申請案號：102104478

※ 申請日：102.2.05 ※IPC 分類：G10L 13/08 (2013.01)

【發明名稱】(中文/英文)

編碼串流產生裝置、韻律訊息編碼裝置、韻律結構分析裝置與語音合成之裝置及方法/ Streaming Encoder, Prosody Information Encoding Device, Prosody-Analyzing Device, and Device and Method for Speech-Synthesizing

【中文】

本案係提供一種語音合成之裝置，其包括一階層式韻律模組，提供一階層式韻律模型；一韻律結構分析單元，接收一低階語言參數、一高階語言參數及一第一韻律參數，且根據該高階語言參數、該低階語言參數、該第一韻律參數及該階層式韻律模組，產生至少一韻律標記；以及一韻律參數合成單元，根據該階層式韻律模組、該低階語言參數及該韻律標記來合成一第二韻律參數。

【英文】

The present invention provides a speech-synthesizing device, including a hierarchical prosodic module for providing a hierarchical prosodic model; a prosody-analyzing device for receiving a low-level linguistic feature, a high-level linguistic feature and a first prosodic feature, and producing at least one prosodic mark based on the low-level linguistic feature, the high-level linguistic feature and the first prosodic feature; and a prosody-synthesizing unit for synthesizing a second prosodic feature based on the low-level linguistic feature, the high-level linguistic feature and the first prosodic feature.

【代表圖】

【本案指定代表圖】：第（一）圖。

【本代表圖之符號簡單說明】：

10：語音合成裝置

101：語音切割及韻律參數抽取器

102：階層式韻律模組

103：韻律結構分析單元

104：編碼器

105：解碼器

106：韻律參數合成單元

107：語音合成器

108：韻律結構分析裝置

109：韻律參數合成裝置

110：韻律訊息編碼裝置

111：韻律訊息解碼裝置

A1：低階語言參數

A2：高階語言參數

A3：第一韻律參數

A4：第一韻律標記

A5：編碼串流

A6：第二韻律標記

A7：第二韻律參數

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

發明專利說明書

（本說明書格式、順序，請勿任意更動）

【發明名稱】（中文/英文）

編碼串流產生裝置、韻律訊息編碼裝置、韻律結構分析裝置與語音合成之裝置及方法/ Streaming Encoder, Prosody Information Encoding Device, Prosody-Analyzing Device, and Device and Method for Speech-Synthesizing

【技術領域】

【0001】 本發明係關於一種語音裝置，尤指一種語音合成裝置。

【先前技術】

【0002】 在傳統以音段爲基礎之語音編碼中，音段對應之韻律訊息通常使用量化直接對韻律參數進行編碼，而沒有考慮到使用具有語言意義之韻律模型來進行參數化韻律編碼。其中有以將音節內音素對應之長度及音高軌跡進行編碼，編碼方式是以預儲存之具有代表性的音節內音素長度及音高軌跡群組樣版，來表示音節內音素的音長及音高軌跡資訊，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；以對於音高軌跡進行編碼，將音高軌跡以片段之直線表示其值，音高軌跡之訊息以對這些片段直線的斜率及端點值表示，於碼書(codebook)中儲存具有代表性的片段直線樣板，音高軌跡便以此碼書進行編碼，此方法簡單，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；還有以對於詞的音長進行純量量化，對於詞的音高軌跡以詞平均音高及詞音高斜率表示之，並對平均值及斜率進行純量量化，並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；以對於音素的音長、音高位階先進行正規化，其正規化方法爲是將音素音長及音高位階的觀察值，分別扣掉該音素類別之平均音長及平均音高位階，最後將正規化之音素音長及音高位階進行量化編碼，此方法可降低傳輸位元率，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；還有以將語音切成不等音框數的語音音段，每個音段的音高軌跡以此音段的平均音高表示之，而能量軌跡是以向量量化表示之，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；

以將語音切成音段，對於音段音高軌跡、音段長度及音段能量軌跡進行編碼，將音高軌跡以片段之直線表示其值，音高軌跡之訊息以對這些片段直線的端點值及時間值表示編碼，而音段長度以正規化地音段長度用純量量化表示，其正規化方法爲是將音段長度的觀察值扣掉該音段類別之平均長度，音段能量軌跡是以 DTW 的方式對於預儲存之樣版進行比對，以誤差值最小之樣版編號爲編碼所需資訊，另外也對 DTW 之路徑、音段起頭及結尾以樣板表示之能量誤差進行編碼，此方法並未考慮韻律產生模型，對於編碼後之語音亦不易進行韻律轉換；目前已有文獻關於將音段的音高軌跡以平均值表示之，並將此平均值以純量量化，此方法簡單，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；還有將音高軌跡以片段之直線表示其值，音高軌跡之訊息以對這些片段直線的端點的音高值及時間資訊表示之，並將這些端點值以純量量化表示之，此方法簡單，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；還有以分段線性近似法(piecewise linear approximation，PLA)表示音對的音高，PLA 裡面包含音段端點的音高及時間資訊、以及折點 (critical point)的音高及時間資訊，其中有文獻係以純量量化表示這些資訊，及以向量量化表示這些 PLA 資訊；還有文獻以傳統 frame-based speech coder 的方法將每個 frame 的音高資訊進行量化，雖然可將音高資訊正確地表示，但相對 data rate 較高；還有將音段的音高軌跡以儲存於 codebook 中的音高軌跡樣板量化並編碼，此方法可以用極低的 data rate 將音高資訊編碼，但 distortion 較大；還有文獻是將音段的時長直接進行純量量化，方法簡單，可完全保留原本音段的長度，但並未考慮韻律產生模型，對於編碼後之語音不易進行韻律轉換；還有將連續三個音段的長度以向量量化編碼，方法簡單，但並未考慮韻律產生模型，對於編碼後之語音亦不易進行韻律轉換；還有文獻提出一個以語音辨認爲基礎的韻律編碼，它會有辨認錯誤引起的合成錯誤聲音的缺點，並且沒有後處理做聲音速度轉換的功能。

【0003】 由習知技術可歸納出其編碼過程如下：(1)語音切割成音段;(2)對音段的頻譜及韻律訊息進行編碼，通常一個音段是對應到音素(phoneme)、音節(syllable)或該系統定義之聲學單元，語音的切割可以採用

語音辨認系統(automatic speech recognition)或用給定已知文本進行強迫對齊(forced alignment)而得到切割好的音段。接下來每個音段要對其頻譜資訊及韻律訊息進行編碼。另一方面，以音段爲基礎之語音編碼系統的語音還原包含了：(1)頻譜及韻律訊息解碼與還原;(2)語音合成。習知技術大多偏重於頻譜資訊的編碼，而於韻律訊息編碼方面較少著墨，通常以量化的方式對於韻律訊息進行編碼，並無考慮韻律訊息其背後的產生模型，因此不易得到較低的編碼位元率，並且較不易以系統化之方法對編碼後的語音進行語音轉換。

【0004】 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「編碼串流產生裝置、韻律訊息編碼裝置、韻律結構分析裝置與語音合成之裝置及方法」，用以改善上述習知技術之缺失。

【發明內容】

【0005】 本案之一面向係提供一種語音合成之裝置，其包括一階層式韻律模組，提供一階層式韻律模型；一韻律結構分析單元，接收一低階語言參數、一高階語言參數及一第一韻律參數，且根據該高階語言參數、該低階語言參數、該第一韻律參數及該階層式韻律模組，產生至少一韻律標記；以及一韻律參數合成單元，根據該階層式韻律模組、該低階語言參數及該韻律標記來合成一第二韻律參數。

【0006】 本案之另一面向係提供一種韻律訊息編碼裝置，包含一語音切割及韻律參數抽取器，接收一語音輸入及一低階語言參數，用以產生一第一韻律參數；一韻律結構分析單元，接收該第一韻律參數、該低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數及該高階語言參數，產生一韻律標記；以及一編碼器，接收該韻律標記及該低階語言參數，用以產生一編碼串流。

【0007】 本案之又一面向係提供一種編碼串流產生裝置，包含一韻律參數抽取器，產生一第一韻律參數；一階層式韻律模組，賦予該第一韻律參數一語言結構意義;一編碼器，根據該語言結構意義之該第一韻律參數來產生一編碼串流，其中該階層式韻律模組包含至少二參數，其中各該參數

係選自一音長、一音高軌跡、一停頓時機、一停頓出現頻率、一停頓時長或其組合。

【0008】 本案之再一面向係提供一種語音合成之方法，包含下列步驟：提供一第一韻律參數、一低階語言參數、一高階語言參數及一階層式韻律模組；根據該第一韻律參數、該低階語言參數、該高階語言參數、及該階層式韻律模組來對該第一韻律參數進行韻律結構分析，以產生一韻律標記；以及根據該韻律標記來輸出一語音合成。

【0009】 本案之再一面向係提供一種韻律結構分析單元，包含一第一輸入端，接收一第一韻律參數；一第二輸入端，接收一低階語言參數；一第三輸入端，接收一高階語言參數；以及一輸出端，其中該韻律結構分析單元根據該第一韻律參數、該低階語言參數及該高階語言參數，而於該輸出端產生一韻律標記。

【0010】 本案之再一面向係提供一種語音合成裝置，包含一解碼器，接收一編碼串流，並還原該編碼串流以產生一低階語言參數及一韻律標記；一階層式韻律模組，接收該低階語言參數及該韻律標記，以產生一韻律參數；以及一語音合成器，根據該低階語言參數及該韻律參數來產生一語音合成。

【0011】 本案之再一面向係提供一種韻律結構分析裝置，包含一階層式韻律模組，提供一階層式韻律模型；以及一韻律結構分析單元，接收一第一韻律參數、一低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數、該高階語言參數及該階層式韻律模組，產生一韻律標記。

【圖式簡單說明】

【0012】

第一圖：本案一較佳實施例之語音合成裝置之示意圖。

第二圖：本案一較佳實施例之漢語語音階層式韻律結構示意圖。

第三圖：本案一較佳實施例之使用 HMM-based speech synthesizer 產生語音合成的流程圖。

第四圖：顯示本案一較佳實施例之語者相關和語者獨立原始(original)及編碼/解碼後重建(reconstruction)之韻律參數韻律範例。

第五圖：顯示本案一較佳實施例之原始語音、韻律訊息編碼後語音合成及轉換爲不同語速之語音之波形、音高軌跡的差異。

【實施方式】

【0013】 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。

【0014】 爲達上述之發明目的，使用階層式韻律模組於語音韻律編碼中，其方塊圖如第一圖所示，包含語音切割及韻律參數抽取器 101、階層式韻律模組 102、韻律結構分析單元 103、編碼器 104、解碼器 105、韻律參數合成單元 106、語音合成器 107、韻律結構分析裝置 108、韻律參數合成裝置 109、韻律訊息編碼裝置 110 及韻律訊息解碼裝置 111。

【0015】 以下介紹本發明的概念：首先將一語音訊號及其對應之低階層語言參數輸入至語音切割及韻律參數抽取器 101，其功能在於使用聲學模型(acoustic model)將輸入語音做音節邊界切割、以及求取音節韻律參數，提供下一級韻律結構分析單元 102 使用；

【0016】 階層式韻律模組 102 之主要用途是用來描述中文語音之韻律階層結構，它包含了韻律狀態模型、韻律停頓模型、音節韻律模型及音節間韻律模型等多種韻律模型。

【0017】 韻律結構分析單元 103 之用途爲利用階層式韻律模組 102，解析輸入語音之韻律參數 A3(由方塊 101 語音切割及韻律參數抽取器產生)，將語音韻律解析爲韻律結構以韻律標記表示之。

【0018】 編碼器 104 之主要功能爲將重建語音韻律所需要的訊息進行編碼(encoding)並進行編碼串流(bit streaming)，這些訊息包含韻律結構分析單元 103 所產生的韻律標記 A4、以及輸入之低階語言參數 A1。

【0019】 解碼器 105 之主要功能是將編碼串流 A5 解碼，將韻律參數合成單元 106 所需要的韻律標記 A6 以及低階語言參數 A1 解碼出來。

【0020】 韻律參數合成單元 106 之主要功能爲利用解碼出的韻律標

5

記 A6 以及低階語言參數訊息 A1，使用階層式韻律模組 102 爲旁資訊(side information)將語音韻律參數合成還原。

【0021】 語音合成器 107 之主要功能爲利用還原之韻律參數 A7、低階語言參數 A1，將語音合成，其係以馬可夫模型爲基礎。

【0022】 韻律結構分析裝置 108 包含階層式韻律模組 102 及韻律結構分析單元 103，其利用階層式韻律模組，以韻律結構分析單元解析輸入語音之韻律參數 A3(由語音切割及韻律參數抽取器 101 產生)，將語音韻律解析爲韻律結構以韻律標記 A4 表示之。

【0023】 韻律參數合成裝置 109 包含階層式韻律模組 102 及韻律參數合成單元 106，其利用解碼器 105 還原出的一第二韻律標記 A6 及低階語言參數 A1，根據該第二韻律標記 A6 及低階語言參數 A1，使用階層式韻律模組 102 作爲旁資訊(side information)以韻律參數合成單元 106 合成出第二韻律參數 A7。

【0024】 韻律訊息編碼裝置 110 包含語音切割及韻律參數抽取器 101、階層式韻律模組 102、韻律結構分析單元 103、韻律結構分析裝置 108 及編碼器 104，其先以語音切割及韻律參數抽取器 101 對一輸入語音及一低階語言參數 A1 作解析以得出一第一韻律參數 A3，然後該韻律結構分析裝置 108 根據該第一韻律參數 A3、該低階語言參數 A1 及一高階語言參數 A2 來形成一第一韻律標記 A4，接著該編碼器 104 根據該第一韻律標記 A4 及該低階語言參數 A1 來形成一編碼串流 A5。

【0025】 韻律訊息解碼裝置 111 包含解碼器 105、階層式韻律模組 102、韻律參數合成單元 106、韻律參數合成裝置 109 及語音合成器 107，其係以解碼器 105 將韻律訊息編碼裝置 111 所輸出之編碼串流 A5 還原爲一第二韻律標記 A6 及一低階語言參數 A1，並透過韻律參數合成裝置 109 來合成一第二韻律參數 A7，該第二韻律參數 A7 經由語音合成器 107 合成出一語音合成。

【0026】 爲了介紹本發明之最佳實施例，以下列式子來表示，這個式子是用於韻律結構分析單元 103，將語音韻律解析爲韻律結構以韻律標記表示之，方法是將韻律聲學特徵參數序列(**A**)以及語言參數序列(**L**)輸入韻律

結構分析單元 103，韻律結構分析單元 103 輸出最佳的韻律標記序列($\mathbf{T}^*$)，這個最佳的韻律標記便可以用來表示語句的韻律參數，進而用於韻律參數編碼，其對應的數學式爲：

$$\mathbf{T}^* = \{\mathbf{B}^*, \mathbf{P}^*\} = \arg\max_{\mathbf{T}} P(\mathbf{T} \mid \mathbf{A}, \mathbf{L}) = \arg\max_{\mathbf{T}} P(\mathbf{T}, \mathbf{A} \mid \mathbf{L})$$
$$= \arg\max_{\mathbf{T}} P(\mathbf{A} \mid \mathbf{T}, \mathbf{L}) P(\mathbf{T} \mid \mathbf{L}) = \arg\max_{\mathbf{B},\mathbf{P}} P(\mathbf{X}, \mathbf{Y}, \mathbf{Z} \mid \mathbf{B}, \mathbf{P}, \mathbf{L}) P(\mathbf{B}, \mathbf{P} \mid \mathbf{L})$$
$$\approx \arg\max_{\mathbf{B},\mathbf{P}} \underbrace{P(\mathbf{X} \mid \mathbf{B}, \mathbf{P}, \mathbf{L}) P(\mathbf{Y}, \mathbf{Z} \mid \mathbf{B}, \mathbf{L}) P(\mathbf{P} \mid \mathbf{B}) P(\mathbf{B} \mid \mathbf{L})}_{\text{階層式韻律模型}}$$

其中 $\mathbf{A} = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} = \{A_1^N\} = \{X_1^N, Y_1^N, Z_1^N\}$ 爲韻律聲學特徵參數序列，N 爲語句音節數，X、Y 和 Z 分別表示音節爲基礎的韻律特徵參數、音節間及差分韻律聲學特徵參數；

$\mathbf{L} = \{\mathbf{POS}, \mathbf{PM}, \mathbf{WL}, \mathbf{t}, \mathbf{s}, \mathbf{f}\} = \{L_1^N\} = \{POS_1^N, PM_1^N, WL_1^N, t_1^N, s_1^N, f_1^N\}$ 爲語言參數序列，其中 $\{\mathbf{POS}, \mathbf{PM}, \mathbf{WL}\}$ 爲高階語言參數序列，**POS**、**PM** 及 **WL** 分別爲詞類序列、標點符號序列及詞常序列，而 $\{\mathbf{t}, \mathbf{s}, \mathbf{f}\}$ 爲低階語言參數序列，**t**、**s** 級 **f** 分別爲聲調、基本音節類別及韻母類別序列；$\mathbf{T} = \{\mathbf{B}, \mathbf{P}\}$ 爲韻律標記序列，其中 $\mathbf{B} = \{B_1^N\}$ 爲韻律停頓序列，$\mathbf{P} = \{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$ 爲韻律狀態序列，其中 p 表示音節音高韻律狀態，q 表示音節長度韻律狀態，r 表示音節能量韻律狀態。韻律標記序列是用來描述階層式韻律模組 102 所考量的中文韻律階層結構，如第二圖所示。此結構包含四種韻律成分：音節、韻律詞、韻律片語及呼吸群組或韻律片語群組。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四種韻律成分的邊界；另一個韻律標記 P 爲韻律狀態可表示爲 $\mathbf{P} = \{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$，用來表示上層韻律成分，也就是韻律詞、韻律片語及呼吸群組或韻律片語群組這三層綜合的音節韻律聲學特徵。

【0027】 <階層式韻律模組> $P(\mathbf{X} \mid \mathbf{B}, \mathbf{P}, \mathbf{L}) P(\mathbf{Y}, \mathbf{Z} \mid \mathbf{B}, \mathbf{L}) P(\mathbf{P} \mid \mathbf{B}) P(\mathbf{B} \mid \mathbf{L})$

【0028】 爲了實現階層式韻律模組，我們在此更詳細地描述此模型。此模型包含了四個子模型：音節韻律聲學模型 $P(\mathbf{X} \mid \mathbf{B}, \mathbf{P}, \mathbf{L})$、音節間韻律聲學模型 $P(\mathbf{Y}, \mathbf{Z} \mid \mathbf{B}, \mathbf{L})$、韻律狀態模型 $P(\mathbf{P} \mid \mathbf{B})$ 以及韻律停頓模型 $P(\mathbf{B} \mid \mathbf{L})$：

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

如下式所示再以以三個子模型來近似：

$$P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L}) \approx P(\mathbf{sp}|\mathbf{B},\mathbf{p},\mathbf{t})P(\mathbf{sd}|\mathbf{B},\mathbf{q},\mathbf{t},\mathbf{s})P(\mathbf{se}|\mathbf{B},\mathbf{r},\mathbf{t},\mathbf{f})$$

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

其中子模型$P(sp_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1})$、$P(sd_n \mid q_n, s_n, t_n)$以及$P(se_n \mid r_n, f_n, t_n)$分別代表第$n$個音節的音高輪廓模型、音節長度模型、能量層次模型，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$t_{n-1}^{n+1} = (t_{n-1}, t_n, t_{n+1})$分別表示韻律停頓序列及聲調序列，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個音節的音高輪廓爲例，我們可得：

$$sp_n = sp_n^r + \beta_{t_n} + \beta_{p_n} + \beta_{B_{n-1},tp_{n-1}}^{f} + \beta_{B_n,tp_n}^{b} + \mu_{sp}$$

其中$sp_n = [\alpha_{0,n}, \alpha_{1,n}, \alpha_{2,n}, \alpha_{3,n}]$爲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，其係數由下述數學求得：

$$\alpha_{j,n} = \frac{1}{M_n + 1}\sum_{i=0}^{M_n} F_n(i) \cdot \phi_j(\frac{i}{M_n}) \qquad j = 0 \sim 3$$

其中$F_n(i)$代表第n個音節第i個音框音高值(frame pitch)，$M_n + 1$代表第n個音節具有音高(pitch)地音框數，$\phi_j(\frac{i}{M_n})$代表第j個正交化基底，其數學式如下：

$$\phi_0(\frac{i}{M}) = 1$$

$$\phi_1(\frac{i}{M}) = [\frac{12 \cdot M}{M+2}]^{1/2} \cdot [\frac{i}{M} - \frac{1}{2}]$$

$$\phi_2(\frac{i}{M}) = [\frac{180 \cdot M^3}{(M-1)(M+2)(M+3)}]^{1/2} \cdot [(\frac{i}{M})^2 - \frac{i}{M} + \frac{M-1}{6 \cdot M}]$$

$$\phi_3(\frac{i}{M}) = [\frac{2800 \cdot M^5}{(M-1)(M-2)(M+2)(M+3)(M+4)}]^{1/2}$$

$$\cdot [(\frac{i}{M})^3 - \frac{3}{2}(\frac{i}{M})^2 + \frac{6M^2 - 3M + 2}{10 \cdot M^2}(\frac{i}{M}) - \frac{(M-1)(M-2)}{20 \cdot M^2}]$$

sp_n^r爲正規化的sp_n，β_{t_n}和β_{p_n}分別爲聲調和韻律狀態的影響參數，$\beta_{B_{n-1},tp_{n-1}}^{f}$和$\beta_{B_n,tp_n}^{b}$爲向前及向後連音影響參數；$tp_{n-1} = t_{n-1}^{n}$以方便表示；$\mu_{sp}$爲音高的全域平均值。基於假設$sp_n^r$爲零平均值和正規分佈，所以我們以常態分佈來表示，可得

$$P(sp_n \mid B_{n-1}^{n}, p_n, t_{n-1}^{n+1}) = N(sp_n; \beta_{t_n} + \beta_{p_n} + \beta_{B_{n-1},tp_{n-1}}^{f} + \beta_{B_n,tp_n}^{b} + \mu_{sp}, R_{sp})$$

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。

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

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 的影響參數。

(2) 音節間韻律聲學模型$P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$：

音節間韻律聲學模型則以五個子模型近似之，如下式所示：

$$P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L}) \approx P(\mathbf{pd},\mathbf{ed},\mathbf{pj},\mathbf{dl},\mathbf{df}|\mathbf{B},\mathbf{L}) \approx \prod_{n=1}^{N-1} P(pd_n, ed_n, pj_n, dl_n, df_n|\mathbf{B},\mathbf{L})$$
$$\approx \prod_{n=1}^{N-1}\Big\{g(pd_n;\alpha_{B_n,L_n},\eta_{B_n,L_n})N(ed_n;\mu_{ed,B_n,L_n},\sigma^2_{ed,B_n,L_n})\cdot N(pj_n;\mu_{pj,B_n,L_n},\sigma^2_{pj,B_n,L_n})N(dl_n;\mu_{dl,B_n,L_n},\sigma^2_{dl,B_n,L_n})$$
$$\cdot N(df_n;\mu_{df,B_n,L_n},\sigma^2_{df,B_n,L_n})\Big\}$$

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

$$pj_n = (sp_{n+1}(1) - \chi_{t_{n+1}}) - (sp_n(1) - \chi_{t_n})$$

其中$sp_n(1)$爲sp_n的第一維度(即音節音高平均值)；χ_t爲聲調 t 平均音高位階。

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

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

爲跨越第 n 個接合點的兩個正規化的音節拉長因子，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；因爲對韻律停頓而言L_n的空間仍是太大，所以將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$P(\mathbf{P}|\mathbf{B})$

韻律狀態模型$P(\mathbf{P} \mid \mathbf{B})$以三個子模型近似之，如下式所示：

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

(4) 韻律停頓模型 $P(\mathbf{B}|\mathbf{L})$

韻律停頓模型 $P(\mathbf{B}\,|\,\Lambda_l)$ 如下式所示

$$P(\mathbf{B}\,|\,\mathbf{L}) \approx \prod_{n=1}^{N-1} P(B_n\,|\,L_n)$$

其中 L_n 爲第 n 個音節的文本相關的語言特徵參數，此機率可用任何方法預估，本實施例中使用決策樹演算法去預估此機率。

此階層式韻律模式之訓練，在適當的韻律斷點和韻律狀態初始化後，是以依次序最佳化演算法(sequential optimal algorithm)來訓練韻律模型，同時對於訓練語料以最大似然性原則(maximum likelihood criterion)作韻律標記且得到此階層式韻律模式之參數。

【0029】 <韻律結構分析單元>

【0030】 韻律結構分析單元工作的目的在解析輸入語句的韻律階層性結構，也就是由韻律聲學特徵參數序列(**A**)以及語言參數序列(**L**)去找到最佳的韻律標記 $\mathbf{T}=\{\mathbf{B},\mathbf{P}\}$ ，數學式表示如下：

$$\mathbf{T}^* = \{\mathbf{B}^*,\mathbf{P}^*\} = \arg\max_{\mathbf{B},\mathbf{P}} Q$$

其中

$$Q = P(\mathbf{B}|\mathbf{L})P(\mathbf{P}|\mathbf{B})P(\mathbf{X}|\mathbf{B},\mathbf{P},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})$$
$$=\left(\prod_{n=1}^{N-1} P(B_n\,|\,L_n)\right)\cdot\left(P(p_1)P(q_1)P(r_1)\left[\prod_{n=2}^{N}\frac{P(p_n\,|\,p_{n-1},B_{n-1})P(q_n\,|\,q_{n-1},B_{n-1})}{P(r_n\,|\,r_{n-1},B_{n-1})}\right]\right)$$
$$\cdot\left(\prod_{n=1}^{N} P(sp_n|B_{n-1}^{n},p_n,t_{n-1}^{n+1})P(sd_n\,|\,q_n,s_n,t_n)P(se_n\,|\,r_n,f_n,t_n)\right)$$
$$\cdot\left(\begin{matrix} g(pd_n;\alpha_{B_n,L_n},\eta_{B_n,L_n})N(ed_n;\mu_{ed,B_n,L_n},\sigma^2_{ed,B_n,L_n}) \\ \prod_{n=1}^{N-1} N(pj_n;\mu_{pj,B_n,L_n},\sigma^2_{pj,B_n,L_n})N(dl_n;\mu_{dl,B_n,L_n},\sigma^2_{dl,B_n,L_n}) \\ N(df_n;\mu_{df,B_n,L_n},\sigma^2_{df,B_n,L_n}) \end{matrix}\right)$$

韻律結構分析單元的工作方法可以用以下的疊代法求最佳解實現：

(1)初始化：使 $i=0$，由下式找到最佳韻律斷點序列：

$$\mathbf{B}^i = \arg\max_{\mathbf{B}} P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{B}|\mathbf{L})$$

(2)重複疊代：以下列三步驟重複疊代得到韻律斷點序列及韻律狀態序列：

步驟一：給定$\mathbf{B}^{i-1}$，使用維特比(Viterbi)演算法標記韻律狀態序列，使得 Q 值增加：，

$$\mathbf{P}^i = \arg\max_{\mathbf{P}} P(\mathbf{X}|\mathbf{B}^{i-1},\mathbf{P},\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B}^{i-1},\mathbf{L})P(\mathbf{P}|\mathbf{B}^{i-1})P(\mathbf{B}^{i-1}|\mathbf{L})$$

步驟二：給定$\mathbf{P}^i$,使用維特比(Viterbi)演算法標記韻律斷點序列，使得 Q 值增加：

$$\mathbf{B}^i = \arg\max_{\mathbf{B}} P(\mathbf{X}|\mathbf{B},\mathbf{P}^i,\mathbf{L})P(\mathbf{Y},\mathbf{Z}|\mathbf{B},\mathbf{L})P(\mathbf{P}^i|\mathbf{B})P(\mathbf{B}|\mathbf{L})$$

步驟三：若 Q 值達到收斂(convergence)，跳出此(2)重複疊代，否則將$i=i+1$且跳回步驟一。

(3) 結束：得到最佳韻律標記$\mathbf{B}^*=\mathbf{B}^i$ 及$\mathbf{P}^*=\mathbf{P}^i$

【0031】 **<韻律訊息的編碼>**

【0032】 由階層式韻律模組 102 可知，音節音高輪廓sp_n、音節長度sd_n以及音節能量位階se_n皆爲考慮多個影響因子之線性組合，這些因子包含低階語言參數：聲調t_n、基本音節型態s_n、韻母型態f_n，另外就是用來表示階層式韻律結構的韻律標記 (由方塊 103 爲韻律結構分析單元得到)：韻律斷點B_n以及韻律狀態p_n、q_n以及r_n。因此，音節音高輪廓sp_n、音節長度sd_n以及音節能量位階se_n只需要將以上的這些因子編碼傳送即可，其中使用下式於韻律參數合成單元 106 以還原其參數：

$$sp'_n = \beta_{t_n} + \beta_{p_n} + \beta^f_{B_{n-1},tp_{n-1}} + \beta^b_{B_n,tp_n} + \mu_{sp}$$
$$sd'_n = \gamma_{t_n} + \gamma_{s_n} + \gamma_{q_n} + \mu_{sd}$$
$$se'_n = \omega_{t_n} + \omega_{f_n} + \omega_{r_n} + \mu_{se}$$

值得注意的是sp^r_n、sd^r_n以及se^r_n可以被忽略不須被傳送，因爲它們的變量十分小可以被忽略。

【0033】 另外音節間的停頓長度pd_n是由 Gamma 分佈模擬，也就是$g(pd_n;\alpha_{B_n,L_n},\eta_{B_n,L_n})$，這個 Gamma 分佈模型描述停頓長度$pd_n$如何受到前後文語言參數及韻律停頓的影響，由於前後文語言參數的組合很多，因此利

用七個決策數(decision tree)分別代表七種韻律斷點下，不同前後文語言參數對音節間停頓的影響 pd_n，稱此七個決策樹爲韻律斷點相關決策樹(break type-dependent decision trees, BDTs)，每一個 BDT 下的葉節點(leaf node) T_n 可以代表某一種韻律斷點下、某一種前後文語言參數的音節間停頓長度分佈，這些分佈即當作傳送音節間停頓長度資訊時使用的旁資訊(side information)，因此只要以葉節點的編號(leaf-node index)以及韻律斷點 B_n 就可以表示音節間停頓長度。值得注意的是，每個音節對應的葉節點編號可由韻律結構分析單元 103 得到，而音節間停頓長度，根據韻律參數合成單元 106 中葉節點的編號(leaf-node index)以及韻律斷點資訊，查詢 BDT 上對應 $\mu_{T_n}^{pd}$ 值來還原音節間停頓長度。

【0034】 總結以上的說明，編碼器 104 需要編碼的符號(Symbol)包含：聲調 t_n、基本音節型態 s_n、韻母型態 f_n、韻律斷點 B_n、三種韻律狀態(p_n、q_n、r_n)以及葉節點(leaf node) T_n。編碼器 104 依據以上 symbol 的種類數以不同的位元長度(bit length)編碼，最後串接爲位元串(bit stream)送至解碼端經由解碼器 105 解碼，然後送至韻律參數合成單元 106 還原韻律訊息，並經由語音合成器 107 語音合成。除了位元串，部分階層式韻律模組 102 的參數爲旁資訊(side information)，用於還原韻律參數使用的參數，其包含音節音高輪廓影響參數：$\{\beta_t, \beta_p, \beta_{B,tp}^f, \beta_{B,tp}^b, \mu_{sp}\}$、音節音長影響參數：$\{\gamma_t, \gamma_s, \gamma_q, \mu_{sd}\}$、音節能量位階影響參數：$\{\omega_t, \omega_f, \omega_r, \mu_{se}\}$、BDT 音節間停頓長參數 $\mu_{T_n}^{pd}$。

【0035】 <語音合成>

【0036】 語音合成器 107 的工作目的是經由給定的基本音節型態、音節音高輪廓、音節長度、音節能量準位、音節間停頓長度，利用隱藏式馬可夫爲基礎之語音合成技術(HMM-based speech synthesis)將語音合成出來。HMM-based speech synthesis 技術爲習知技術，在此僅簡短說明其參數設定：中文的 21 個聲母及 39 個韻律都各以一個 HMM 表示，每個 HMM 包含 5 個 HMM 狀態，每一個狀態內的觀察相量包含兩個類別串：一個爲維度 75 的頻譜參數，另一爲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，以維度爲 5 的 multi-variate single Gaussian 向量表示每個聲母或韻律 HMM 裡面 5 個狀態的長度機率分布。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。

【0037】 圖三爲使用 HMM-based speech synthesizer 產生語音合成的流程圖。於 HMM 狀態及清濁音產生器 303 我們首先用以下的習知方法的 HMM 狀態時長模型 301 產生每一個 HMM 狀態的時長：

$$d_{n,c} = \mu_{n,c} + \rho \cdot \sigma^2_{n,c} \quad \text{for } c = 1 \sim C$$

其中 $\mu_{n,c}$ 及 $\sigma^2_{n,c}$ 分別代表的 n 個音節的第 c 個 HMM 狀態，對應高斯函數模型的平均值參數及變異量參數，ρ 爲伸縮係數，由以下式子得到：

$$\rho = \left(sd'_n - \sum_{c=1}^{C} \mu_{n,c} \right) / \left(\sum_{c=1}^{C} \sigma^2_{n,c} \right)$$

值得注意的是上式中 sd'_n 即是韻律參數合成單元 106 還原的音節音長。由於每一個 HMM 狀態皆有標示其清音及濁音的狀態，因此在產生 HMM 狀態長度後，便可利用 HMM 狀態清濁音模型 302 得到音節內濁音的時長或音框數 $M'_n + 1$，進而音節音高輪廓於對數音高軌跡及激發信號產生器 306 可以以下式還原：

$$F'_n(i) = \sum_{j=0}^{3} \alpha'_{j,n} \cdot \phi_j\left(\frac{i}{M'_n}\right) \quad \text{for } i = 0 \sim M'_n$$

其中 $\alpha'_{j,n}$ 代表由韻律參數合成單元 106 還原的音節音高輪廓向量的第 j 維，也就是 $sp'_n = [\alpha'_{0,n}, \alpha'_{1,n}, \alpha'_{2,n}, \alpha'_{3,n}]$。接著，MLSA 合成濾波器(synthesis filter)所需要的激發信號(excitation signal)便可由還原的對數音高軌跡產生。另一方面，除了激發信號以外，每個音框頻譜資訊是以習知技術在給定 HMM 狀態長度和 HMM 的狀態觀察向量參數後，於音框 MGC 產生器 305 利用 HMM 聲學模型 304 以習知技術之參數產生法產生出適當的每個音框之 MGC 參數，並將每個音節之能量位階調整至韻律參數合成單元 106 還原的音節能量位階。最後，將激發信號及每個音框之 MGC 參數輸入至 MLSA 濾波器 307，便可合成出語音。

【0038】 **<實驗結果>**

【0039】 表一顯示實驗語料的重要統計資訊，實驗語料分爲兩大部分：(1)單一語者語料庫 Treebank speech corpus、以及(2)多語者中文連續語音資料庫 TCC300，這兩分語料分別用於實地測試的第一圖實施例之語者相關(speaker dependent, SD)及語者獨立(speaker independent)之韻律訊力編碼效

能。

表一

語料名稱	子集合	用途	語者數目	語句數目	音節數目	語料小時數
Treebank	TrainTB	*訓練階層式韻律模組 *訓練語音切割及語音合成之聲學模型	1	376	51,868	3.9
	TestTB	評估語者相關韻律訊息編碼	1	44	3,898	0.3
TCC300	TrainTC1	訓練語音切割之聲學模型	274	8,036	300,728	23.9
	TrainTC2	訓練階層式韻律模組	164	962	106,955	8.3
	TestTC	評估語者獨立韻律訊息編碼	19	226	26,357	2.4

【0040】 表二爲各編碼符號(symbol)所需要的編碼位元長度(codeword length)，表三爲旁資訊的參數量說明。

表二

Symbol	Symbol 數目	位元數
聲調 t_n	5	3
基本音節類別 s_n	411	9
音節音高韻律狀態 p_n	16	4
音節長度韻律狀態 q_n	16	4
音節能量韻律狀態 r_n	16	4
韻律停頓 B_n	7	3
BDT 葉節點	5/7/3/2/4/3/1(SI)	3/3/2/1/2/2/0(SI)
*B*0/1/2-1/2-2/2-3/3/4	3/9/3/9/5/11/9(SD)	2/4/2/4/3/4/4(SD)
每個音節總位元數（最大值）	30 (SI) 31(SD)	

表三

參數類別	參數數目
聲調影響參數: $\beta_t / \gamma_t / \omega_t$	20/5/5
向前及向後連音影響參數: $\beta^f_{B,tp} / \beta^b_{B,tp}$	720/720

韻律狀態影響參數: $\beta_p/\gamma_q/\omega_r$	16/16/16
全域平均值: $\mu_{sp}/\mu_{sd}/\mu_{se}$	1/1/1
基本音節型態及韻母型態影響參數: γ_s/ω_{f_n}	411/40
BDT 葉節點靜音長平均值: $\mu_{T_n}^{pd}$	25 (SI)/49 (SD)
總和	1997 (SI)/2021 (SD)

【0041】 表四爲韻律參數合成單元 106 還原的各韻律參數的方均根誤差 (root-mean-square errors, RMSE)，由表四中可看出誤差皆十分小。

表四

		音高軌跡 (Hz/semitone)	音節時長 (ms)	音節能量位階 (dB)	音節靜音時長(ms)
Treebank	訓練 (TrainTB)	16.2/1.42	4.81	0.68	38.7
	測試 (TestTB)	15.7/1.22	4.74	0.70	30.9
TCC300	訓練 (TrainTC2)	12.1/1.26	8.54	1.05	46.9
	測試 (TestTC)	11.7/1.13	12.49	1.86	63.0

表五

		平均±標準差	最大值	最小值
Treebank	訓練 (TrainTB)	116±5.25	131.5	91.5
	測試 (TestTB)	114.9±4.78	124.1	99.1
TCC300	訓練 (TrainTC2)	113.3±9.2	138.0	66.1
	測試 (TestTC)	114.9±14.9	158.8	84.7

表六

語料	語句數	音節數	小時數	發音速度 =(語句音節數) / (語句音節時長總合秒數)	語音速度 =(語句音節數) / (語句總合秒數)
FastTB	368	50,691	3.4	5.52	4.40
TrainTB	376	51,868	3.9	5.05	3.82
TestTB	44	3,895	0.3	4.89	3.78
SlowTB	372	51231	6.0	3.78	2.46

【0042】 表五爲本案之位元率表現。在語者相關和語者獨立平均的傳輸位

元率分別爲 114.9±4.78 位元每秒及 114.9±14.9 位元每秒，此位元率十分低。第四圖(a)及第四圖(b)顯示語者相關(401、402、403、404)和語者獨立(405、406、407、408)原始(original)及編碼/解碼後重建(reconstruction)之韻律參數韻律範例，包含語者相關的音高層次 401、音節長度 402、音節能量位階 403、音節間靜音時長及韻律斷點標記 404 (不含 B0 與 B1，爲簡潔表示)，以及語者獨立的音高層次 405、音節長度 406、音節能量位階 407、音節間靜音時長及韻律斷點標記 408。由第四圖(a)及第四圖(b)可明顯發現還原韻律及原始韻律十分接近。

【0043】 **<語速轉換範例>**

【0044】 本案之韻律編碼方法亦提供系統化的語速轉換平台，方法爲於韻律參數合成單元 106 將原本語速之階層式韻律模組 102 抽換爲目標語速之階層式韻律模組 102。實地測試所採用的訓練語料相關統計資訊如表六所示，原本於實驗結果中使用的語者相關語料是正常速度語料，以此語料爲標準，另外兩個不同語速語料分別爲快速語料及慢速語料，它們對應之階層式韻律模組皆可以相同於正常速度之訓練方法完成。第五圖(a)顯示原始語音之波形 501、音高軌跡 502；第五圖(b)顯示韻律訊息編碼後語音合成之波形 505、音高軌跡 506；第五圖(c)顯示轉換爲語速較快之語音的波形 509、音高軌跡 510；第五圖(d)顯示轉換爲語速較慢之語音的波形 513、音高軌跡 514，其中第五圖(a)~第五圖(d)直線的部分表示音節切割位置（可以漢語拼音 503、507、511 及 515 表示）及實驗所使用的時間爲 504、508、512 及 516。由第五圖(a)~第五圖(d)可以明顯的看到原始語速、快速、慢速語音上音節長度及音節間停頓時長的差異。由非正式的聽覺實驗聆聽不同語速的語音合成，其韻律相當流暢且自然。

【0045】 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爲準。

【0046】 <u>實施例:</u>

1. 一種語音合成之裝置，其包括：

一階層式韻律模組，提供一階層式韻律模型；

一韻律結構分析單元，接收一低階語言參數、一高階語言參數及一第一韻律參數，且根據該高階語言參數、該低階語言參數、該第一韻律參數及該階層式韻律模組，產生至少一韻律標記；以及一韻律參數合成單元，根據該階層式韻律模組、該低階語言參數及該韻律標記來合成一第二韻律參數。

2. 如實施例 1 所述之裝置，更包括：

一韻律參數抽取器，接收一語音輸入及一低階語言參數，切割該語音輸入來形成一切割的語音，根據該低階語言參數及該切割的語音產生該第一韻律參數；以及

一韻律參數合成裝置，其中：

該第一階層式韻律模組係根據一第一語速而被產生；

當該韻律參數合成裝置欲產生與該第一不同的一第二語速時，該第一階層式韻律模組被抽換爲具該第二語速的一第二階層式韻律模組且該韻律參數合成單元將該第二韻律參數改變爲一第三韻律參數；以及該語音合成器根據該第三韻律參數及該低階語言參數產生具有該第二語速之語音合成。

3. 如實施例 1-2 所述之裝置，更包括：

一編碼器，接收該韻律標記及該低階語言參數，且根據該韻律標記及該低階語言參數而產生一編碼串流；以及

一解碼器，接收該編碼串流，並還原該韻律標記及該低階語言參數，其中該編碼器包含一碼書，提供一相對應於該韻律標記所需的編碼位元以產生該編碼串流，且該解碼器亦包含一碼書，提供該編碼位元對該編碼串流進行該韻律標記之還原。

4. 如實施例 1-3 所述之裝置，更包括：

一韻律參數合成裝置，接收經解碼器還原之該韻律標記及該低階語言參數來產生該第二韻律參數，該第二韻律參數包含一音節基頻軌跡、一音節時長、一音節能量位階、及一音節間靜音時長。

5. 如實施例 1-4 所述之裝置，其中:

該第二韻律參數係以一加法模組還原;以及

該音節間靜音時長係以一碼書查表還原。

6. 一種韻律訊息編碼裝置，包含:

一韻律參數抽取器，接收一語音輸入及一低階語言參數，用以產生一第一韻律參數；

一韻律結構分析單元，接收該第一韻律參數、該低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數及該高階語言參數，產生一韻律標記；以及

一編碼器，接收該韻律標記及該低階語言參數，用以產生一編碼串流。

7. 一種編碼串流產生裝置，包含:

一韻律參數抽取器，產生一第一韻律參數；

一階層式韻律模組，賦予該第一韻律參數一語言結構意義;

一編碼器，根據具有該語言結構意義之該第一韻律參數來產生一編碼串流，其中:

該階層式韻律模組包含至少二參數，其中各該參數係選自一音長、一音高軌跡、一停頓時機、一停頓出現頻率、一停頓時長或其組合。

8. 一種語音合成之方法，包含下列步驟：

提供一第一韻律參數、一低階語言參數、一高階語言參數及一階層式韻律模組；

根據該第一韻律參數、該低階語言參數、該高階語言參數、及該階層式韻律模組來對該第一韻律參數進行韻律結構分析，以產生一韻律標記；以及

根據該韻律標記來輸出一語音合成。

9. 如實施例 8 所述之方法，更包含下列步驟：

對一輸入語音及該低階語言參數執行語音切割及韻律參數抽取，以產生該第一韻律參數；

分析該第一韻律參數以產生該韻律標記；

編碼該韻律標記以形成該編碼串流；

解碼該編碼串流;

根據該低階語言參數及該韻律標記來合成一第二韻律參數；以及

根據該第二韻律參數及該低階語言參數來輸出該語音合成。

10.一種韻律結構分析單元，包含：

一第一輸入端，接收一第一韻律參數；

一第二輸入端，接收一低階語言參數；

一第三輸入端，接收一高階語言參數；以及

一輸出端，其中該韻律結構分析單元根據該第一韻律參數、該低階語言參數及該高階語言參數，而於該輸出端產生一韻律標記。

11.一種語音合成裝置，包含:

一解碼器，接收一編碼串流，並還原該編碼串流以產生一低階語言參數及一韻律標記；

一階層式韻律模組，接收該低階語言參數及該韻律標記，以產生一韻律參數；以及

一語音合成器，根據該低階語言參數及該韻律參數來產生一語音合成。

12.一種韻律結構分析裝置，包含:

一階層式韻律模組，提供一階層式韻律模組；以及

一韻律結構分析單元，接收一第一韻律參數、一低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數、該高階語言參數及該階層式韻律模組，產生一韻律標記。

13.如實施例 12 所述之韻律結構分析裝置，其中:

該低階語言參數包含一中文基礎音節類別及聲調；

該高階語言參數包含一詞、一詞類、及一標點符號；以及

該韻律參數包含一音節基頻軌跡、一音節時長、一音節能量位階及一音節間靜音時長。

14.如實施例 12-13 所述之韻律結構分析裝置，係使用一階層式韻律模組，並以一最佳化演算法輔以該低階語言參數及該高階語言參數對該第一韻律參數進行韻律結構分析，以輸出該韻律標記。

【符號說明】

【0047】

10：語音合成裝置

101：語音切割及韻律參數抽取器

102：階層式韻律模組

103：韻律結構分析單元

104：編碼器

105：解碼器

106：韻律參數合成單元

107：語音合成器

108：韻律結構分析裝置

109：韻律參數合成裝置

110：韻律訊息編碼裝置

111：韻律訊息解碼裝置

301：HMM 狀態時長模型

302：HMM 狀態清濁音模型

303：HMM 狀態時長及清濁音產生器

304：HMM 聲學模型

305：音框 MGC 產生器

306：對數音高軌跡及激發信號產生器

307：MLSA 濾波器

401：語者相關的音高層次

402: 語者相關的音節長度

403: 語者相關的音節能量位階

404: 語者相關的音節間靜音時長及韻律斷點標記

405: 語者獨立的音高層次

406: 語者獨立的音節長度

407: 語者獨立的音節能量位階

408: 語者獨立的音節間靜音時長及韻律斷點標記

501、505、509 及 513：語音之波形

502、506、510 及 514：語音之音高軌跡

503、507、511 及 515：漢語拼音（音節切割位置）

504、508、512 及 516：實驗所使用的時間

A1：低階語言參數

A2：高階語言參數

A3：第一韻律參數

A4：第一韻律標記

A5：編碼串流

A6：第二韻律標記

A7：第二韻律參數

【生物材料寄存】

國內寄存資訊【請依寄存機構、日期、號碼順序註記】

國外寄存資訊【請依寄存國家、機構、日期、號碼順序註記】

【序列表】(請換頁單獨記載)

申請專利範圍

1. 一種語音合成之裝置，其包括：

一階層式韻律模組，提供一階層式韻律模型；

一韻律結構分析單元，接收一低階語言參數、一高階語言參數及一第一韻律參數，且根據該高階語言參數、該低階語言參數、該第一韻律參數及該階層式韻律模組，產生至少一韻律標記；以及

一韻律參數合成單元，根據該階層式韻律模組、該低階語言參數及該韻律標記來合成一第二韻律參數。

2. 如申請專利範圍第 1 項所述之語音合成之裝置，更包括：

一韻律參數抽取器，接收一語音輸入及一低階語言參數，切割該語音輸入來形成一切割的語音，根據該低階語言參數及該切割的語音產生該第一韻律參數；以及

一韻律參數合成裝置，其中：

該第一階層式韻律模組係根據一第一語速而被產生；

當該韻律參數合成裝置欲產生與該第一不同的一第二語速時，該第一階層式韻律模組被抽換爲具該第二語速的一第二階層式韻律模組且該韻律參數合成單元將該第二韻律參數改變爲一第三韻律參數；以及

該語音合成器根據該第三韻律參數及該低階語言參數產生具有該第二語速之語音合成。

3. 如申請專利範圍第 1 項所述之裝置，更包括：

一編碼器，接收該韻律標記及該低階語言參數，且根據該韻律標記及該低階語言參數而產生一編碼串流；以及

一解碼器，接收該編碼串流，並還原該韻律標記及該低階語言參數，其中該編碼器包含一碼書，提供一相對應於該韻律標記所需的編碼位元以

產生該編碼串流，且該解碼器亦包含一碼書，提供該編碼位元對該編碼串流進行該韻律標記之還原。

4. 如申請專利範圍第 1 項所述之裝置，更包括：

一韻律參數合成裝置，接收經解碼器還原之該韻律標記及該低階語言參數來產生該第二韻律參數，該第二韻律參數包含一音節基頻軌跡、一音節時長、一音節能量位階、及一音節間靜音時長。

5. 如申請專利範圍第 4 項所述之裝置，其中:

該第二韻律參數係以一加法模組還原;以及

該音節間靜音時長係以一碼書查表還原。

6. 一種韻律訊息編碼裝置，包含:

一語音切割及韻律參數抽取器，接收一語音輸入及一低階語言參數，用以產生一第一韻律參數；

一韻律結構分析單元，接收該第一韻律參數、該低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數及該高階語言參數，產生一韻律標記；以及

一編碼器，接收該韻律標記及該低階語言參數，用以產生一編碼串流。

7. 一種編碼串流產生裝置，包含:

一韻律參數抽取器，產生一第一韻律參數；

一階層式韻律模組，賦予該第一韻律參數一語言結構意義;

一編碼器，根據具有該語言結構意義之該第一韻律參數產生一編碼串流，其中:

該階層式韻律模組包含至少二參數，其中各該參數係爲一音長、一音高軌跡、一停頓時機、一停頓出現頻率、一停頓時長或其組合。

8. 一種語音合成之方法，包含下列步驟：

提供一第一韻律參數、一低階語言參數、一高階語言參數及一階層式韻律模組；

根據該第一韻律參數、該低階語言參數、該高階語言參數、及該階層式韻律模組來對該第一韻律參數進行韻律結構分析，以產生一韻律標記；以及

根據該韻律標記來輸出一語音合成。

9. 如申請專利範圍第 8 項所述之方法，更包含下列步驟：

對一輸入語音及該低階語言參數執行語音切割及韻律參數抽取，以產生該第一韻律參數；

分析該第一韻律參數以產生該韻律標記；

編碼該韻律標記以形成該編碼串流；

解碼該編碼串流;

根據該低階語言參數及該韻律標記來合成一第二韻律參數；以及

根據該第二韻律參數及該低階語言參數來輸出該語音合成。

10.一種韻律結構分析單元，包含：

一第一輸入端，接收一第一韻律參數；

一第二輸入端，接收一低階語言參數；

一第三輸入端，接收一高階語言參數；以及

一輸出端，其中該韻律結構分析單元根據該第一韻律參數、該低階語言參數及該高階語言參數，而於該輸出端產生一韻律標記。

11.一種語音合成裝置，包含:

一解碼器，接收一編碼串流，並還原該編碼串流以產生一低階語言參數及一韻律標記；

一階層式韻律模組，接收該低階語言參數及該韻律標記，以產生一韻

律參數；以及

一語音合成器，根據該低階語言參數及該韻律參數來產生一語音合成。

12.一種韻律結構分析裝置，包含:

一階層式韻律模組，提供一階層式韻律模組；以及

一韻律結構分析單元，接收一第一韻律參數、一低階語言參數及一高階語言參數，且根據該第一韻律參數、該低階語言參數、該高階語言參數及該階層式韻律模組，產生一韻律標記。

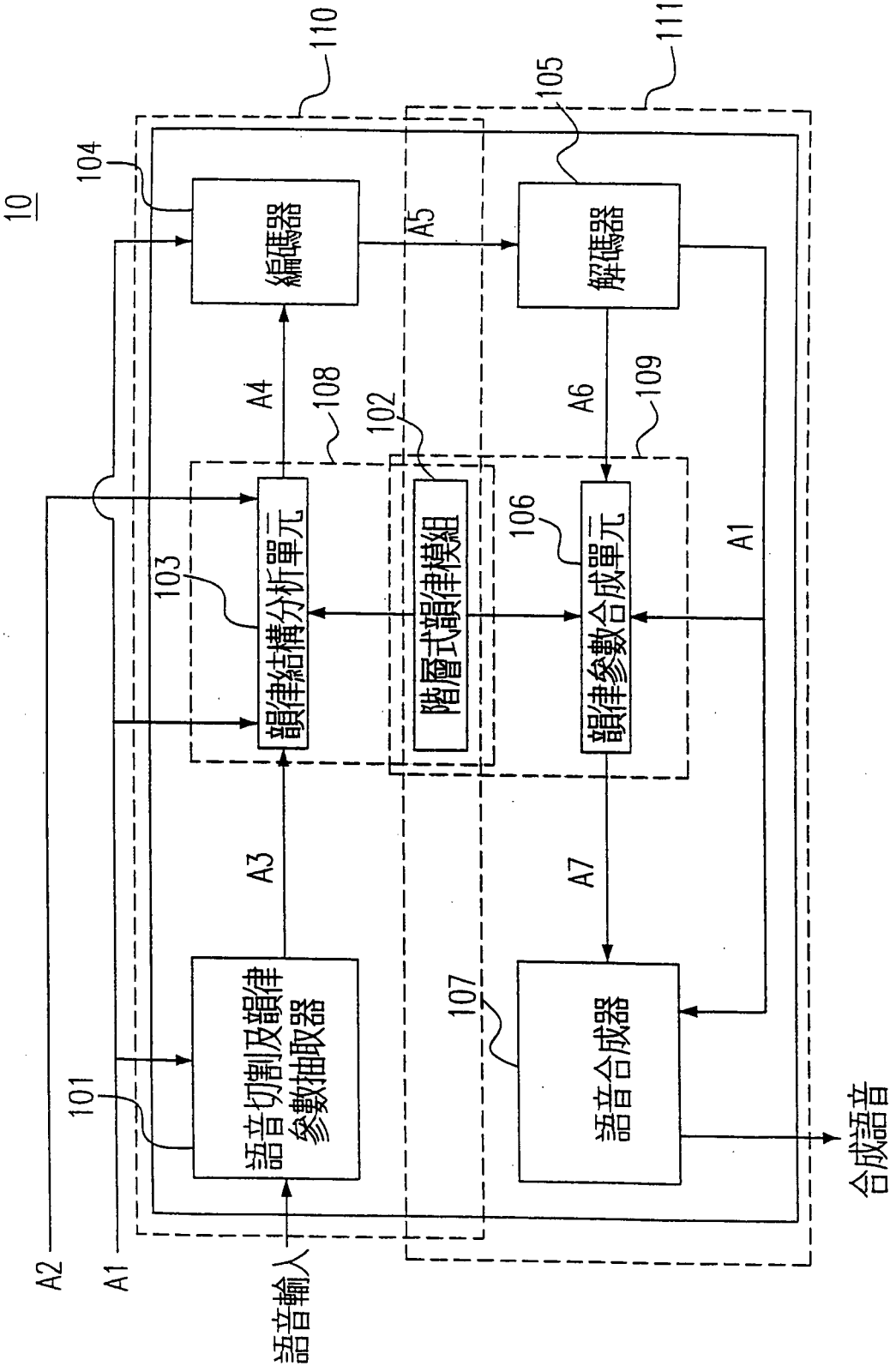
13.如申請專利範圍第 12 項所述之韻律結構分析裝置，其中:

該低階語言參數包含一中文基礎音節類別及聲調；

該高階語言參數包含一詞、一詞類、及一標點符號；以及

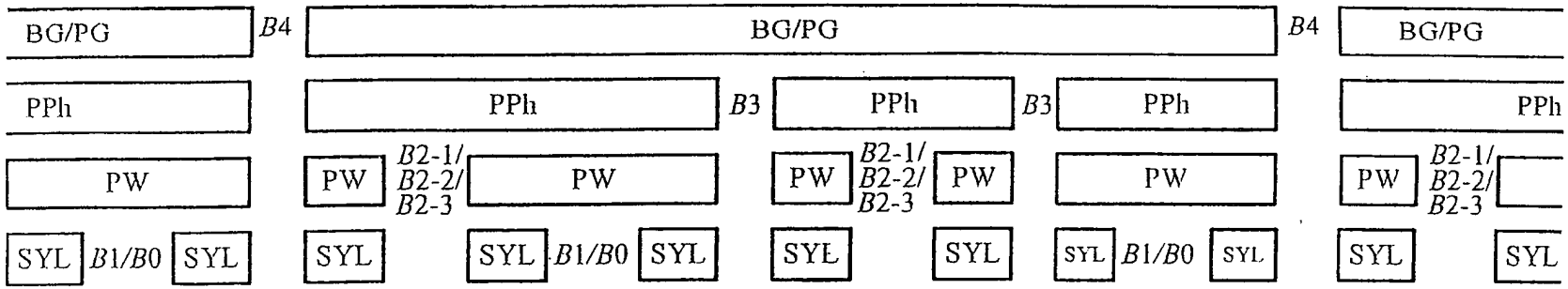
該韻律參數包含一音節基頻軌跡、一音節時長、一音節能量位階及一音節間靜音時長。

14.如申請專利範圍第 12 項所述之韻律結構分析裝置，係使用一階層式韻律模組，並以一最佳化演算法輔以該低階語言參數及該高階語言參數對該第一韻律參數進行韻律結構分析，以輸出該韻律標記。


10
110
111
104
105
編碼器
解碼器
A5
A4
A6
108
102
109
韻律結構分析單元
階層式韻律模組
韻律參數合成單元
103
106
A1
A3
A7
語音切割及韻律參數抽取器
語音合成器
107
101
A2
A1
語音輸入
合成語音

圖式

第一圖



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

B4: 呼吸群組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

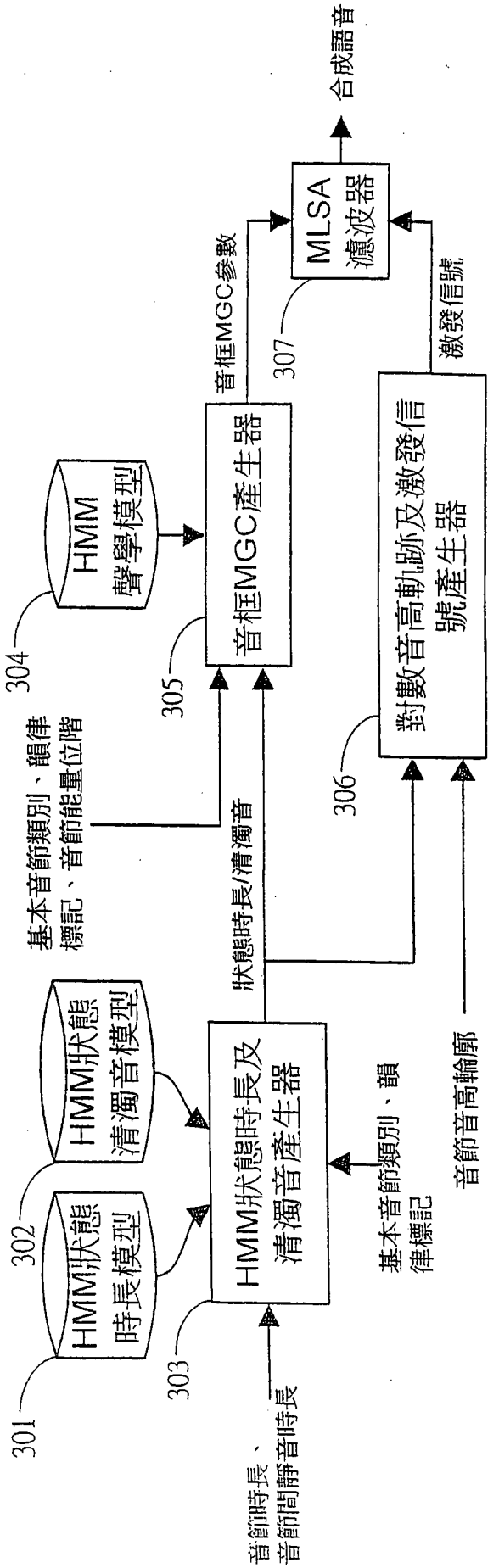
B2-2: 第二類韻律詞韻律斷點，表示短靜音停頓

B2-3: 第三類韻律詞韻律斷點，表示音節拉長停頓

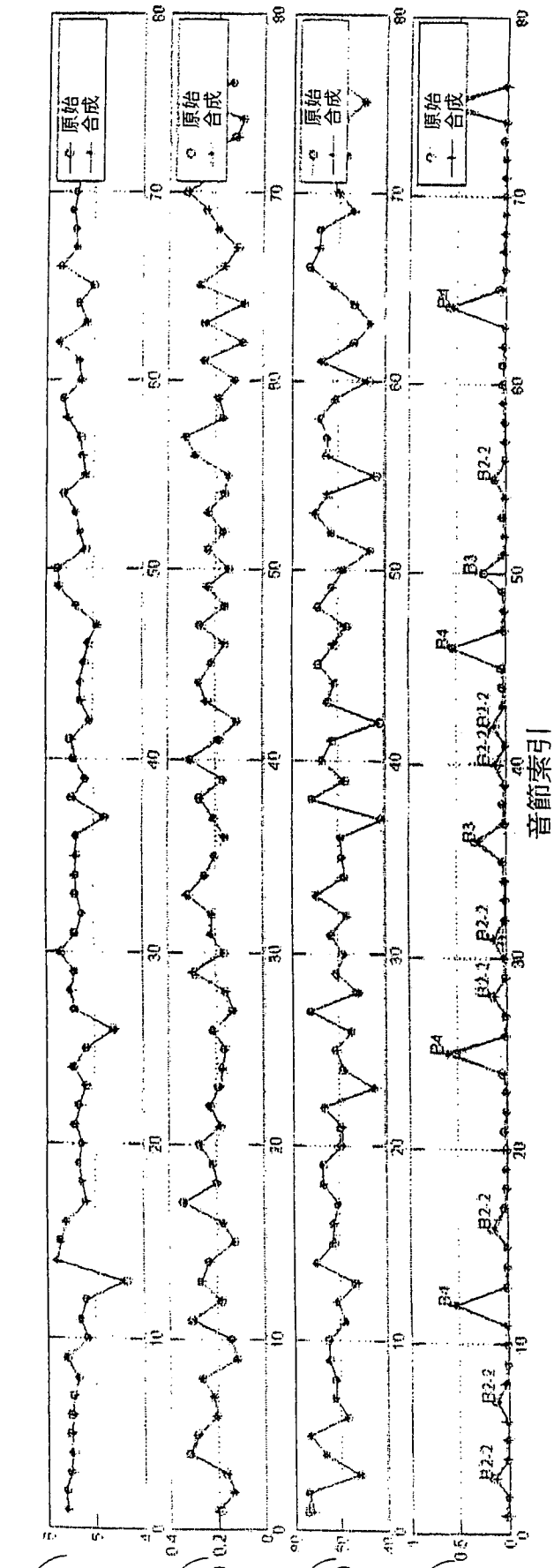
B1: 韻律詞內正常韻律斷點

B0: 韻律詞內強連音韻律斷點

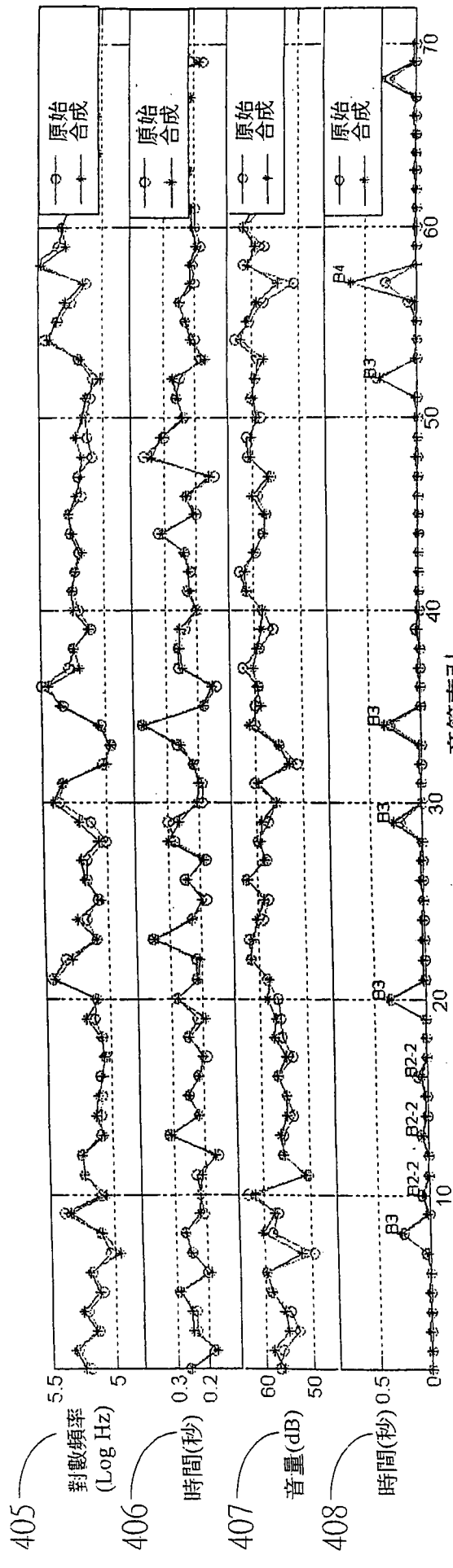
第二圖


HMM狀態時長模型
HMM狀態清濁音模型
HMM聲學模型
HMM狀態時長及清濁音產生器
音框MGC產生器
對數音高軌跡及激發信號產生器
MLSA濾波器
合成語音
音節時長、音節間靜音時長
基本音節類別、韻律標記
音節音高輪廓
基本音節類別、韻律標記、音節能量位階
狀態時長/清濁音
音框MGC參數
激發信號
301
302
303
304
305
306
307

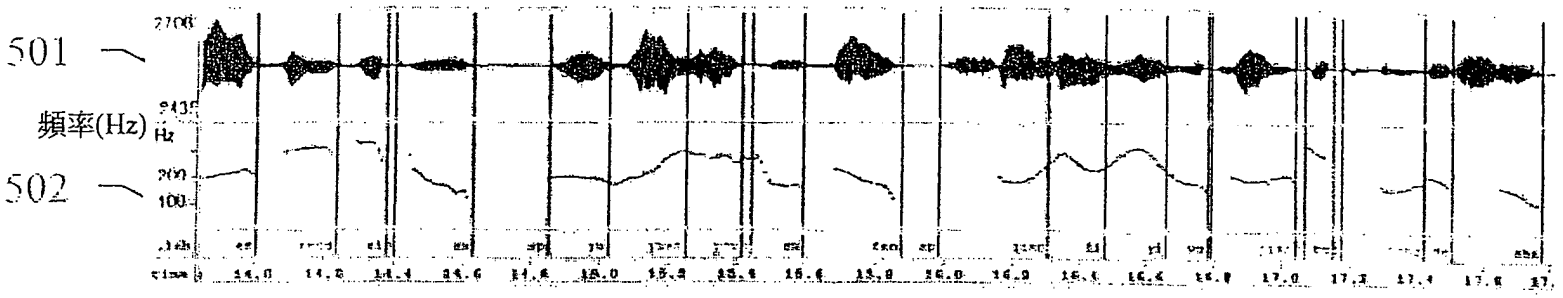
第三圖



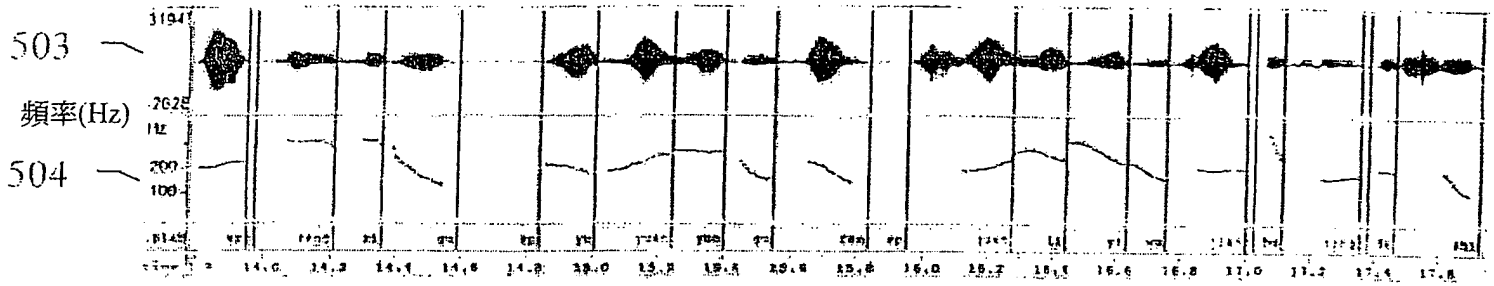
第四圖(a)



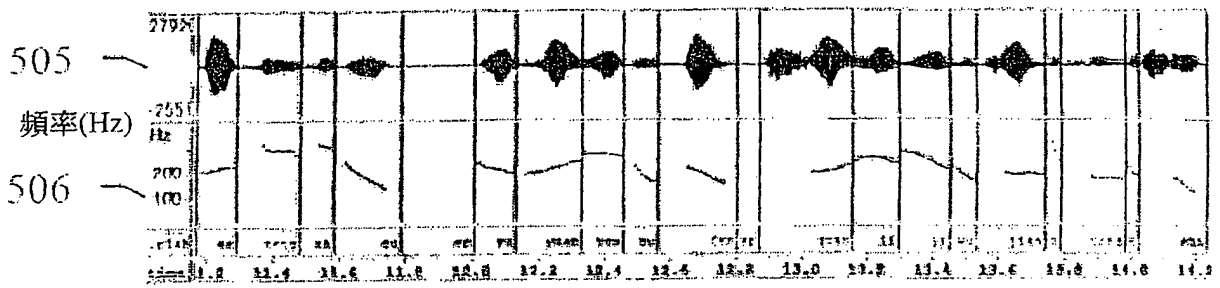
第四圖(b)



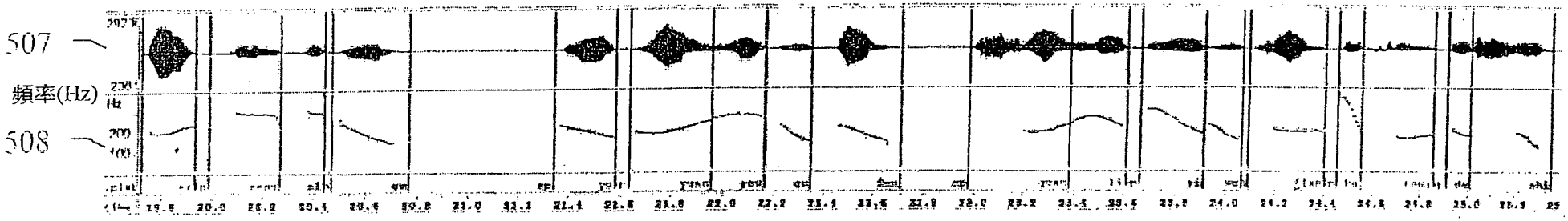
第五圖(a)



第五圖(b)



第五圖(c)



第五圖(d)