(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 201227714 A1
(43)公開日： 中華民國 101 (2012) 年 07 月 01 日

(21)申請案號：099145318 (22)申請日： 中華民國 99 (2010) 年 12 月 22 日
(51)Int. Cl.： *G10L13/00 (2006.01)*
(71)申請人：財團法人工業技術研究院 (中華民國) INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE (TW)
新竹縣竹東鎮中興路 4 段 195 號
(72)發明人：林政源 LIN, CHENG YUAN (TW) ；黃健紘 HUANG, CHIEN HUNG (TW) ；郭志忠 KUO, CHIH CHUNG (TW)
(74)代理人：洪堯順
申請實體審查：有 申請專利範圍項數：30 項 圖式數：15 共 47 頁

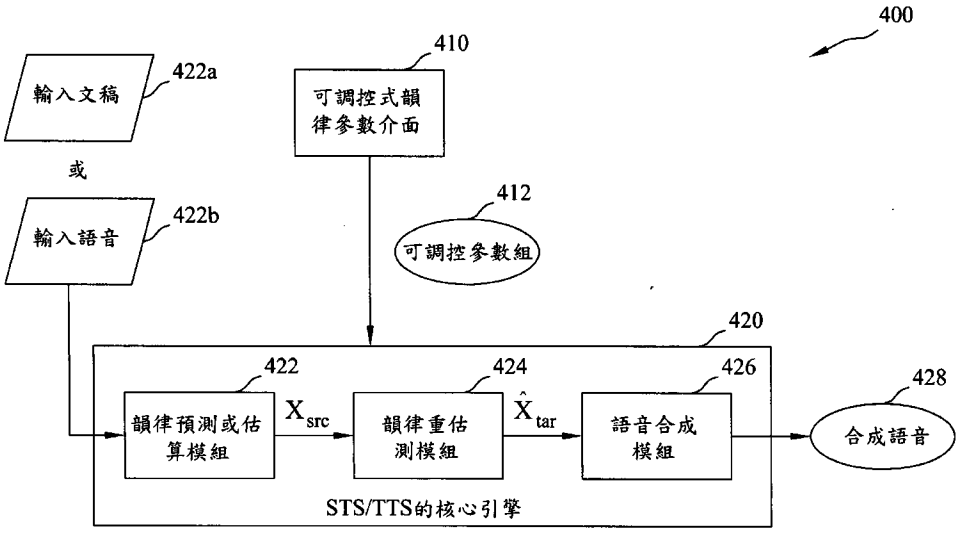
(54)名稱

可調控式韻律重估測系統與方法及電腦程式產品

CONTROLLABLE PROSODY RE-ESTIMATION SYSTEM AND METHOD AND COMPUTER PROGRAM PRODUCT THEREOF

(57)摘要

在一種可調控式韻律重估測系統的一實施範例中，一個語音或文字轉語音的核心引擎係由一韻律預測或估算模組、一韻律重估測模組、以及一語音合成模組所組成。此韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至此韻律重估測模組。此韻律重估測模組根據由一個可調控式韻律參數介面所提供的一可調控參數組及收到的韻律資訊，將此韻律資訊重估測後，產生新的韻律資訊，再提供給此語音合成模組以產生合成語音。



400：韻律重估測系統
410：可調控式韻律參數介面
412：可調控參數組
420：STS/TTS 的核心引擎
422：韻律預測或估算模組
422a：輸入文稿
422b：輸入語音
424：韻律重估測模組
426：語音合成模組
428：合成語音
X_{src}：韻律資訊
$\hat{X}_{tar}$ ：調整後的韻律資訊

六、發明說明：

【發明所屬之技術領域】

本揭露係關於一種可調控式韻律重估測(controllable prosody re-estimation)系統與方法及電腦程式產品。

【先前技術】

韻律預測在文字轉語音(Text-To-Speech，TTS)系統上，對語音合成的自然性有很大的影響。文字轉語音合成系統主要有基於大語料庫(Corpus-based)之最佳單元選取合成方法以及隱藏式馬可夫(HMM-based)統計模型方法。隱藏式馬可夫模型統計方法的合成效果比較有一致性，不會因為輸入的句子不同而有明顯差異性。而訓練出的語音模型檔案通常都很小(例如 3MB) ，這些特點都優於大語料庫的方法，所以此 HMM-based 的語音合成最近變得很普及。然而，利用此方法在產生韻律時，似乎存在著過度平滑化(over-smoothing)的問題。雖然有文獻提出全域變異數的方法(global variance method)來改善(ameliorate)此問題，使用此方法去調整頻譜有明顯正向效果，但用於調整基頻(F0)則無聽覺上的偏好效果，有時候似乎會因為伴隨產生的副效應(side effect)而降低語音品質。

最近一些關於TTS的文獻也提出加強TTS之豐富表現的技術，這些技術通常需要大量收集多樣式的語料庫(corpora)，因此往往需要很多的後製處理。然而，建構

[5]

一個韻律豐富性的TTS系統是十分耗時的，因此有部分的文獻提出採用外部工具的方式提供 TTS 產生更多樣化的韻律資訊。例如，基於工具(tool-based)的系統提供使用者多種更新韻律的可行方案，像是提供使用者一個圖形使用者介面(GUI)工具，來調整音高曲線(pitch contour)以改變韻律，並且根據新韻律重新合成語音;或是使用標記語言(markup language)來調整韻律等。然而，多數使用者無法正確地透過圖形使用者介面來修改音高曲線，同樣地，一般人並不熟悉如何撰寫標記語言，所以，基於工具的系統在實際使用上也是不方便的。

關於TTS的專利文獻有很多，例如可控制TTS輸出品質、控制TTS不同速度輸出的、用於電腦合成語音的中文語音音韻轉換、使用韻律控制的中文文本至語音拼接合成、TTS韻預測方法、以及語音合成系統及其韻律控制方法等。

舉例來說，如第一圖所揭露的中文語音音韻轉換系統 100，是利用一個音韻分析單元 130，接收一來源語音及相對應的文字，透過此分析單元裡面的階層拆解模組 131、音韻轉換函式選擇模組 132、音韻轉換模組 133 擷取音韻資訊，最後套用到語音合成單元 150 以產生合成語音(synthesized speech)。

如第二圖所揭露的語音合成系統與方法是一種針對

外來語的 TTS 技術，以語言分析模組(language analysis module)204 分析文字資料(text data)200 而得之語言資訊(language information)204a，透過韻律預測模組(prosody prediction module)209 產生韻律資訊(prosody information)209a，接著由語音單元挑選模組(speech-unit selection module)208 至特徵參數資料庫(characteristic parameter database)206 中，挑選一序列較符合文字內容與預測韻律資訊的語音資料，最後由語音語合成模組(speech synthesis module)210 合成出語音 211。

【發明內容】

本揭露實施的範例可提供一種可調控式韻律重估測系統與方法及電腦程式產品。

在一實施範例中，所揭露者是關於一種可調控式韻律重估測系統。此系統包含一個可調控式韻律參數介面以及一個語音或文字轉語音(Speech-To-Speech or Text-To-Speech，STS/TTS)的核心引擎。此可調控式韻律參數介面用來輸入一可調控參數組。此核心引擎由一韻律預測或估算模組(prosody predict/estimation module)、一韻律重估測模組(prosody re-estimation module)、以及一語音合成模組(speech synthesis module)所組成。此韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至此韻律重估測模組。此韻律重估測模組根據輸入的可調控參數組及收

到的韻律資訊，將此韻律資訊重估測後，產生新的韻律資訊，再提供給此語音合成模組以產生合成語音。

在另一實施範例中，所揭露者是關於一種可調控式韻律重估測系統，此韻律重估測系統係執行於一電腦系統中。此電腦系統備有一記憶體裝置，用來儲存一原始錄音語料庫與一合成之語料庫。此韻律重估測系統可包含一可調控式韻律參數介面及一處理器。此處理器備有一韻律預測或估算模組、一韻律重估測模組、以及一語音合成模組。此韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至此韻律重估測模組，此韻律重估測模組根據輸入的可調控參數組及收到的韻律資訊，將此韻律資訊重估測後，產生新的韻律資訊，再套用至此語音合成模組以產生合成語音。其中，此處理器統計此兩語料庫之韻律差異來建構一韻律重估測模型，以提供給此韻律重估測模組使用。

在又一實施範例中，所揭露者是關於一種可調控式韻律重估測方法。此方法包含:準備一個可調控式韻律參數介面，以供輸入一可調控參數組;根據輸入文稿或輸入語音來預測出或估算出韻律資訊;建構一韻律重估測模型，並根據此可調控參數組及預測出或估算出的韻律資訊，藉由此韻律重估測模型來調整出新的韻律資訊;以及將此新的韻律資訊提供給一語音合成模組以產生合成語音。

[S]

在又一實施範例中，所揭露者是關於一種可調控式韻律重估測的電腦程式產品。此電腦程式產品包含一記憶體以及儲存於此記憶體的一可執行的電腦程式。此電腦程式藉由一處理器來執行:準備一個可調控式韻律參數介面，以供輸入一可調控參數組;根據輸入文稿或輸入語音來預測出或估算出韻律資訊;建構一韻律重估測模型，並根據此可調控參數組及預測出或估算出的韻律資訊，藉由此韻律重估測模型來調整出新的韻律資訊;以及將此新的韻律資訊提供給一語音合成模組以產生合成語音。

茲配合下列圖示、實施範例之詳細說明及申請專利範圍，將上述及本發明之其他目的與優點詳述於後。

【實施方式】

本揭露實施的範例是要提供一個基於韻律重估測之可調控式的系統與方法及電腦程式產品，來提升韻律豐富性以更貼近原始錄音的韻律表現，以及提供可控制的多樣式韻律調整功能來區別單一種韻律的 TTS 系統。因此，本揭露中，利用系統先前所估測的韻律資訊當作初始值，經過一個韻律重估測模組後求得新的韻律資訊，並且提供一個可調控韻律參數的介面，使其調整後韻律具有豐富性。而此核心的韻律重估測模組是統計兩份語料庫的韻律資訊差異而求得，此兩份語料庫分別是原始

[S]

錄音的訓練語句以及文字轉語音系統的合成語句的語料庫。

在說明如何利用可調控韻律參數來產生具有豐富性的韻律之前，先說明韻律重估測的建構。第三圖是一範例示意圖，說明多樣式韻律分佈的表示法，與所揭露的某些實施範例一致。第三圖的範例中，X_{tts}代表 TTS 系統所產生的韻律資訊，並且 X_{tts}的分佈是由它的平均數 μ_{tts} 以及標準差 σ_{tts} 來規範，表示為(μ_{tts}，σ_{tts})。X_{tar} 代表目標韻律(target pitch)，並且 X_{tar}的分佈是由(μ_{tar}，σ_{tar})來規範。如果(μ_{tts}，σ_{tts})與(μ_{tar}，σ_{tar})都為已知的話，則根據兩分佈，(μ_{tts}，σ_{tts})與(μ_{tar}，σ_{tar})，之間的統計差異(statistical difference)，X_{tar}可以被重估測而得出。正規化後之統計上的均等(normalized statistical equivalent)公式如下:

$$(X_{tar}-\mu_{tar})/\sigma_{tar} = (X_{tts}-\mu_{tts})/\sigma_{tts} \qquad (1)$$

將韻律重估測的觀念延伸，則如第三圖所示，可以在(μ_{tts}，σ_{tts})與(μ_{tar}，σ_{tar})之間使用內插法(interpolation)，計算出多樣式之調整後的韻律分佈($\hat{\mu}_{tar}$，$\hat{\sigma}_{tar}$)。依此，就容易產生出豐富的(rich)調整後的韻律$\hat{X}_{tar}$以提供給 TTS 系統。

無論使用何種訓練方法，來自 TTS 系統的合成語音

[S]

與來自它的訓練語料庫(training corpus)的錄音(recorded speech)之間始終存在著韻律差異(prosody difference)。換句話說，如果有一個 TTS 系統的韻律補償機制可以減少韻律差異的話，就可以產生出更自然的合成語音。所以，本揭露實施的範例所要提供的一種有效的系統，係以基於一種重估測的模式，來改善韻律預測(pitch prediction)。

第四圖是一種可調控式韻律重估測系統的一個範例示意圖，與所揭露的某些實施範例一致。第四圖的範例中，韻律重估測系統 400 可包含一個可調控式韻律參數介面 410 以及一個語音或文字轉語音(Speech-To-Speech or Text-To-Speech，STS/TTS)的核心引擎 420。可調控式韻律參數介面 410 用來輸入一可調控參數組 412。核心引擎 420 可由一韻律預測或估算模組 422、一韻律重估測模組 424、以及一語音合成模組 426 所組成。韻律預測或估算模組 422 根據輸入文稿 422a 或輸入語音 422b 來預測出或估算出韻律資訊 X_{src}，並傳送至韻律重估測模組 424。韻律重估測模組 424 根據輸入的可調控參數組 412 以及收到的韻律資訊 X_{src}，將韻律資訊 X_{src} 重估測後，產生新的韻律資訊，也就是調整後的韻律資訊 $\hat{X}_{tar}$，再套用至語音合成模組 426 以產生合成語音 428。

在本揭露實施的範例中，韻律資訊 X_{src} 的求取方式是根據輸入資料的型態來決定，假如是一段語音，則採

用韻律估算模組進行韻律萃取，假如是一段文字，則是採用韻律預測模組。可調控參數組 412 至少包括有三個參數，彼此是獨立的。此三個參數可由外部輸入 0 個或 1 個或 2 個，其餘未輸入者可採用系統預設值。韻律重估測模組 424 可根據如公式(1)的韻律調整公式來重估測韻律資訊 X_{src}。可調控參數組 412 裡的參數可採用兩個平行語料庫的方式統計而得。兩個平行語料庫分別是前述提及的原始錄音的訓練語句以及文字轉語音系統的合成語句的語料庫。而統計方式則分為靜態分佈法(static distribution method)及動態分佈法(dynamic distribution method)。

第五圖與第六圖是韻律重估測系統 400 分別應用在 TTS 與 STS 上的範例示意圖，與所揭露的某些實施範例一致。第五圖的範例中，當韻律重估測系統 400 應用在 TTS 上時，第四圖中的 STS/TTS 的核心引擎 420 扮演 TTS 核心引擎 520 的角色，而第四圖中的韻律預測或估算模組 422 扮演韻律預測模組 522 的角色，根據輸入文稿 422a 來預測出韻律資訊。而第六圖的範例中，當韻律重估測系統 400 應用在 STS 上時，第四圖中的 STS/TTS 的核心引擎 420 扮演 STS 核心引擎 620 的角色，而第四圖中的韻律預測或估算模組 422 扮演韻律估算模組 622 的角色，根據輸入語音 422b 來估算出韻律資訊。

[S]

承上述，第七圖與第八圖是當韻律重估測系統 400 分別應用在 TTS 與 STS 上時，韻律重估測模組與其他模組的關聯示意圖，與所揭露的某些實施範例一致。第七圖的範例中，當韻律重估測系統 400 應用在 TTS 上時，韻律重估測模組 424 接收韻律預測模組 522 預測出的韻律資訊 X_{src}，及參考可調控參數組 412 中的三個可調控參數，記為(μ_{shift}，μ_{center}，γ_{σ})，然後採用一韻律重估測模型，來調整韻律資訊 X_{src}，產生新的韻律資訊，即調整後的韻律資訊 $\hat{X}_{tar}$，並傳送至語音合成模組 426。

第八圖的範例中，當韻律重估測系統 400 應用在 STS 上時，與第七圖不同的是，韻律重估測模組 424 所接收的韻律資訊 X_{src} 是韻律估算模組 622 根據輸入語音 422b 估算出的韻律資訊。而韻律重估測模組 424 後續的運作與第七圖中所載相同，不再重述。關於三個可調控參數(μ_{shift}，μ_{center}，γ_{σ})與韻律重估測模型將再詳細說明。

以下以應用在 TTS 為例，先以第九圖的範例示意圖來說明如何建構韻律重估測模型，與所揭露的某些實施範例一致。在韻律重估測模型建構的階段，需要有兩份平行的語料庫，也就是句子內容相同的兩份語料庫，一個定義為來源語料庫(source corpus)，另一個定義為目標語料庫(target corpus)。在第九圖的範例中，目標語料庫是根據一個給定的(given)文字語料庫(text corpus)910 而

[S]

錄製(record)的原始錄音語料庫(recorded speech corpus)920，是作TTS訓練之用。然後，可利用一種訓練方法，例如 HMM-based，來建構TTS系統930。一旦TTS系統930建立後，根據相同的文字語料庫910輸入的文稿內容，可使用此訓練出的TTS系統930來產生一個合成之語料庫(synthesized speech corpus)940，此即來源語料庫。

因為原始錄音語料庫920與合成之語料庫940是兩份平行的語料庫，可直接經由簡單的統計來估測此兩平行語料庫之韻律差異950。在本揭露實施的範例中，利用韻律差異950，提供兩種統計法來獲得一韻律重估測模型960，其中一種是全域統計法，另一種是單句統計法。全域統計法是一靜態分佈法(static distribution method)，而單句統計法是動態分佈法(dynamic distribution method)。此兩種統計法說明如下。

全域統計法是以全體語料為統計單位，統計原始錄音語料庫與合成語音語料庫的方式，並以整體語料庫的韻律來衡量彼此之間的差異，而希望文字轉語音系統所產生之合成語音韻律可以盡量近似於原始錄音的自然韻律，因此對於原始錄音語料庫整體之平均數μ_{rec}和標準差σ_{rec}，以及合成語音語料庫整體之平均數μ_{tts}和標準差σ_{tts}而言，這兩者之間存在一個正規化統計均等(Normalized Statistical Equivalent)關係，如下式。 [S]

$$\frac{X_{rec}-\mu_{rec}}{\sigma_{rec}}=\frac{X_{tts}-\mu_{tts}}{\sigma_{tts}} \qquad (2)$$

其中，X_{tts}表示由 TTS 系統所預測的韻律，而X_{rec}表示原始錄音的韻律。換句話說，假設給予一個X_{tts}，則應該依下式來修正:

$$\hat{X}_{tts}=\mu_{rec}+(X_{tts}-\mu_{tts})\frac{\sigma_{rec}}{\sigma_{tts}}\ ,$$

才能使得修正後的韻律有機會近似於原先錄音的韻律表現。

單句統計法是以一個句子當作基本的統計單位。並以原始錄音語料庫及合成語料庫的每一句子為基本單位，比較該兩語料庫的每一句的韻律差異性來觀察與統計彼此的差異，做法說明如下:(1)對於每一平行序列對，亦即每一合成語句及每一原始錄音語句，計算其韻律分佈(μ_{tts}，σ_{tts})及(μ_{rec}，σ_{rec})。(2)假設共計算出 K 對韻律分佈，標記為(μ_{tts}，σ_{tts})$_1$ 及(μ_{rec}，σ_{rec})$_1$ 至(μ_{tts}，σ_{tts})$_K$ 及(μ_{rec}，σ_{rec})$_K$，則可利用一回歸法(regression method)，例如最小平方誤差法、高斯混合模型法、支持向量機方法、類神經方法等，建立一回歸模型(regression model)RM。(3)在合成階段(synthesis stage)時，由 TTS 系統先預測出輸入語句的初始韻律統計(μ_s，σ_s)，爾後套用回歸模型 RM 就可得出新的韻律統

[S]

計($\hat{\mu}_s$，$\hat{\sigma}_s$)，即輸入語句的目標韻律分佈。第十圖是產生回歸模型 RM 的一個範例示意圖，與所揭露的某些實施範例一致。其中，回歸模型 RM 採用最小平方誤差法而建立，所以套用時只需將初始韻律資訊乘上 RM 即可，此回歸模型 RM 是用來預測任一輸入語句的目標韻律分佈。

當韻律重估測模型建構完成後(不論是採用全域統計法或是單句統計法)，本揭露實施的範例還提供一個可由參數調控(parameter controllable)的方式，來讓 TTS 或 STS 系統能夠產生更豐富的韻律。其原理先說明如下。

將方程式(1)中的 *tts* 替換成 *src*，並且引入參數 α 及 β，在(μ_{src}，σ_{src})與(μ_{tar}，σ_{tar})之間使用插入法，如下列方程式。

$$\begin{cases}\hat{\mu}_{tar} = \alpha \cdot \mu_{tar} + (1-\alpha)\cdot \mu_{src} \\ \hat{\sigma}_{tar} = \beta \cdot \sigma_{tar} + (1-\beta)\cdot \sigma_{src}\end{cases}, \quad 0 \le \alpha, \beta \le 1$$

其中，μ_{src} 與 σ_{src} 分別是來源語料庫的韻律平均值 μ_{src} 以及韻律標準差 σ_{src}。所以，欲計算出多樣式之調整後的韻律分佈，韻律重估測模型可用下列的形式來表達，X_{src} 是來源語音。

$$\hat{X}_{tar} = \hat{\mu}_{tar} + (X_{src} - \mu_{src})\frac{\hat{\sigma}_{tar}}{\sigma_{src}}$$

韻律重估測模型也可用下列的另一形式來表達。 [S]

$$\hat{X}_{tar} = \mu_{shift} + (X_{src} - \mu_{center}) \cdot \gamma_{\sigma}$$

其中，μ_{center} 就是上一形式中的 μ_{src}，也就是所有 X_{src} 的平均值，μ_{shift} 就是上一形式中的 $\hat{\mu}_{tar}$，γ_{σ} 就是上一形式中的 $\hat{\sigma}_{tar} / \sigma_{src}$。當韻律重估測模型採用此種表達形式時，共有三種參數(μ_{shift}，μ_{center}，γ_{σ})可調整。透過此三種參數(μ_{shift}，μ_{center}，γ_{σ})的調整，可使調整後的韻律更具有豐富性。以 γ_{σ} 值的變化說明如下。

當 $\gamma_{\sigma}=0$ 時，調整後的韻律 $\hat{X}_{tar}$ 等於參數 μ_{shift} 的值，表示調整後的韻律 $\hat{X}_{tar}$ 等於一個輸入的常數值，例如合成之機器人的聲音(synthetic robotic voice)。當 $\gamma_{\sigma}<0$ 時，即 $\hat{\sigma}_{tar} / \sigma_{src} < 0$，表示調整後的韻律 $\hat{X}_{tar}$ 是特殊韻律的調整，例如外國腔調的語音(foreign accented speech)。當 $\gamma_{\sigma}>0$ 時，表示調整後的韻律 $\hat{X}_{tar}$ 是正規韻律的調整，其中，當 $\gamma_{\sigma}=1$ 時，$\hat{\sigma}_{tar} = \sigma_{src}$;當 $\gamma_{\sigma}>1$ 時，$1<\gamma_{\sigma}<\sigma_{tar}/\sigma_{src}$;當 $\gamma_{\sigma}<1$ 時，$\sigma_{tar}/\sigma_{src}<\gamma_{\sigma}<1$。

因此，透過適當參數的調控，可適合某些情境或語氣或不同語言的表達，可視終端需求而定。而本揭露實施的範例中，韻律重估測系統 400 只需開放一個可調控式韻律參數介面 410 供終端輸入此三個參數即可。當此三個參數有未輸入者時，也可採用系統預設值。此三個參數的系統預設值可設定如下。

$\mu_{center} = \mu_{src}$； $\mu_{shift} = \mu_{tar}$； $\gamma_{\sigma} = \sigma_{tar}/\sigma_{src}$。 [S]

而這些μ_{src}、μ_{tar}、σ_{tar}、σ_{src}的值可透過前述所提的兩個平行語料庫的方式統計而得。也就是說，本揭露中的系統也提供參數未輸入者的預設值。因此，在本揭露實施的範例中，此可調控參數組412，例如μ_{shift}，μ_{center}，γ_{σ})，是可彈性調控的(flexible control)。

承上述，第十一圖是一範例流程圖，說明一種可調控式韻律重估測方法的運作，與所揭露的某些實施範例一致。第十一圖的範例中，首先，準備一個可調控式韻律參數介面，以供輸入一可調控參數組，如步驟 1110 所示。然後，根據輸入文稿或輸入語音來預測出或估算出韻律資訊，如步驟 1120 所示。建構一韻律重估測模型，並根據此可調控參數組及預測出或估算出的韻律資訊，藉由此韻律重估測模型來調整出新的韻律資訊，如步驟 1130 所示。最後，將此新的韻律資訊提供給一語音合成模組以產生合成語音，如步驟1140所示。

在第十一圖的範例中，各步驟之實施細節，例如步驟1110之可調控參數組的輸入與調控、步驟1120之韻律重估測模型的建構與表達形式、步驟 1130 之韻律重估測等，如同上述所載，不再重述。

本揭露實施之韻律重估測系統也可執行於一電腦系統上。此電腦系統(未示於圖示)備有一記憶體裝置，用來儲存原始錄音語料庫920與合成之語料庫940。如第

十二圖的範例所示，韻律重估測系統 1200 包含可調控式韻律參數介面 410 及一處理器 1210。處理器 1210 裡可備有韻律預測或估算模組 422、韻律重估測模組 424、以及語音合成模組 426，來執行韻律預測或估算模組 422、韻律重估測模組 424、以及語音合成模組 426 之上述功能。處理器 1210 可經由統計記憶體裝置 1290 中此兩語料庫之韻律差異，來建構上述之韻律重估測模型，以提供給韻律重估測模組 424 使用。處理器 1210 可以是電腦系統中的處理器。

本揭露之實施範例也可以用一電腦程式產品(computer program product)來實現。此電腦程式產品至少包含一記憶體以及儲存於此記憶體的一可執行的電腦程式(executable computer program)。此電腦程式可藉由一處理器或電腦系統來執行第十一圖之可調控式韻律重估測方法的步驟 1110 至步驟 1140。此處理器還可韻律預測或估算模組 422、韻律重估測模組 424、以及語音合成模組 426、及透過可調控式韻律參數介面 410 輸入可調控式韻律參數，來執行韻律預測或估算模組 422、韻律重估測模組 424、以及語音合成模組 426 之上述功能。藉由這些模組來執行步驟 1110 至步驟 1140。當前述三個參數(μ_{shift}，μ_{center}，γ_{o})有未輸入者時，也可採用前述之預設值。各實施細節如同上述所載，不再重述。

[S]

在本揭露中，進行一系列的實驗來證明其實施範例的可行性。首先，以全域統計法以及單句統計法來進行音高準位(pitch level)的驗證實驗，例如可採用音素、韻母(final)、或音節(syllable)等當作基本單位來求取音高曲線(pitch contour)後再求其平均數。這裡採用音高作為實驗的依據是因為韻律的變化與音高變化是十分密切相關，所以可以透過觀察音高的預測結果來驗證所提的方法可行性。另外，以微觀的方式進一步作比較，來觀察比較音高曲線的預測差異程度。例如，以韻母當作基本單位為例，先以 2605 個中文句子(Chinese Mandarin sentence)的語料庫並採用基於HMM之TTS方法來建構一 TTS 系統。然後，建立韻律重估測模型。再給予前述可調控參數組，並觀察有使用與無使用其韻律重估測模型之 TTS 系統之間的效能差異(performance difference)。

第十三圖是對一句子之四種音高曲線的範例示意圖，包括原始錄音語料、採用 HTS 方法的 TTS、採用靜態分佈法的 TTS、及採用動態分佈法的 TTS，其中橫軸代表句子的時間長度(單位為秒)，縱軸代表韻母的音高曲線(Final's pitch contour)，其單位為 log Hz。從第十三圖的範例可以看出，在基於 HTS 方法(基於 HMM 的其中一種方法)的 TTS 之音高曲線 1310 中，有明顯之過度平滑化的現象。第十四圖是 8 個相異句子在第十三圖所示四種情況下之音高平均值及標準差的範例示意圖，其中橫軸代表句子的號碼(sentence number)，縱軸

[S]

代表平均值±標準差，其單位為 log Hz。從第十三圖及第十四圖的範例可以看出，相較於採用傳統 HTS 方法的 TTS，本揭露實施範例之 TTS(無論是採用動態或靜態分佈法)可以產生與原始錄音語料更具相似韻律的結果。

在本揭露中，分別進行兩項聽覺測試(listening test)，包括偏好度測試(preference test)及相似度測試(similarity test)。相較於傳統基於 HMM 之 TTS 方法，其測試結果顯示本揭露之經重估測後的合成語音有非常好的效果，特別是偏好度測試的結果。主要是因為本揭露之重估測後的合成語音已經妥善補償原始之 TTS 系統所產生之過度平滑的韻律，而產生更逼真的韻律。

在本揭露中，也進行另一實驗來觀察給予前述可調控參數組後，其實施範例中的 TTS 的韻律是否變得更豐富。第十五圖是給予不同的三組可調控參數所產生之三種音高曲線的範例示意圖，這三種音高曲線分別由三種合成聲音所估算而得，包括原始 HTS 方法的合成聲音、合成之機器人的聲音、及外國腔調的語音，其中橫軸代表句子的時間長度(單位為秒)，縱軸代表韻母的音高曲線，其單位為 log Hz。從第十五圖的範例可以看出，對於合成之機器人的聲音，經重估測後的音高曲線是幾乎接近於平坦(flat);至於外國腔調的語音，經重估測之音高曲線的形狀(pitch shape)與 HTS 方法所產生之音高曲線

[S]

相較，是呈現相反方向(opposite direction)。經過非正式的語音聽測實驗，多數聽者認為，提供這些特殊的合成語音對目前 TTS 系統韻律表現上有加分的效果。

所以，從實驗與量測顯示本揭露實施的範例都有優異的實現結果。本揭露實施的範例在 TTS 或 STS 的應用上，可提供豐富的韻律及更貼近原始錄音的韻律表現，也可提供可控制的多樣式韻律調整功能。從本揭露實施的範例中，也觀察到當給予某些值的可調控參數後，經重估測後的合成語音，例如機器人的聲音或外國腔調的語音，會有特殊的效果。

綜上所述，本揭露實施的範例可提供一種有效率的可調控式韻律重估測系統與方法，可應用於語音合成。本揭露之實施範例利用先前所估測的韻律資訊當作初始值，經過一個重估測模型後求得新的韻律資訊，並且提供一個可調控式韻律參數介面，使其調整後韻律具有豐富性。重估測模型可藉由統計兩平行語料庫的韻律資訊差異而求得，此兩平行語料庫分別是原始錄音的訓練語句以及文字轉語音系統的合成語句。

以上所述者僅為本揭露實施的範例，當不能依此限定本揭露實施之範圍。即大凡本發明申請專利範圍所作之均等變化與修飾，皆應仍屬本發明專利涵蓋之範圍。

[S]

【圖式簡單說明】

第一圖是一種中文語音音韻轉換系統的一個範例示意圖。

第二圖是語音合成系統與方法的一個範例示意圖。

第三圖是一範例示意圖，說明多樣式韻律分佈的表示法，與所揭露的某些實施範例一致。

第四圖是一種可調控式韻律重估測系統的一個範例示意圖，與所揭露的某些實施範例一致。

第五圖是第四圖之韻律重估測系統應用在 TTS 上的一個範例示意圖，與所揭露的某些實施範例一致。

第六圖是第四圖之韻律重估測系統應用在 STS 上的一個範例示意圖，與所揭露的某些實施範例一致。

第七圖是當韻律重估測系統應用在 TTS 上時，韻律重估測模組與其他模組的一個關聯示意圖，與所揭露的某些實施範例一致。

第八圖是當韻律重估測系統應用在 STS 上時，韻律重估測模組與其他模組的一個關聯示意圖，與所揭露的某些實施範例一致。

第九圖是一範例示意圖，以應用在 TTS 上為例，說明如何建構一韻律重估測模型，與所揭露的某些實施範例一致。

第十圖是產生回歸模型的一個範例示意圖，與所揭露的某些實施範例一致。

第十一圖是一範例流程圖，說明一種可調控式韻律重估測方法的運作，與所揭露的某些實施範例一致。

[S]

第十二圖是韻律重估測系統執行於一電腦系統中的一範例流程圖，與所揭露的某些實施範例一致。

第十三圖是對一句子之四種音高曲線的範例示意圖，與所揭露的某些實施範例一致。

第十四圖是 8 個相異句子在第十三圖所示四種情況下之音高平均值及標準差的範例示意圖，與所揭露的某些實施範例一致。

第十五圖是給予不同的三組可調控參數所產生之三種音高曲線的範例示意圖，與所揭露的某些實施範例一致。

【主要元件符號說明】

100 中文語音音韻轉換系統　130 音韻分析單元

131 階層拆解模組　132 音韻轉換函式選擇模組

133 音韻轉換模組　150 語音合成單元

200 文字資料　204 語言分析模組

204a 語言資訊　206 特徵參數資料庫

208 語音單元挑選模組　209 韻律預測模組

209a 韻律資訊　210 語音合成模組

211 合成語音

X_{tts} TTS 系統所產生的韻律資訊

X_{tar} 目標韻律　$\hat{X}_{tar}$ 調整後的韻律

$(\mu_{tts}, \sigma_{tts})$ X_{tts}的分佈　$(\mu_{tar}, \sigma_{tar})$ X_{tar}的分佈

$(\hat{\mu}_{tar}, \hat{\sigma}_{tar})$ 調整後的韻律分佈

[S]

400 韻律重估測系統　　410 可調控式韻律參數介面

412 可調控參數組　　420 STS/TTS 的核心引擎

422 韻律預測或估算模組　　422a 輸入文稿

422b 輸入語音　　424 韻律重估測模組

426 語音合成模組　　428 合成語音

X_{src} 韻律資訊　　$\hat{X}_{tar}$ 調整後的韻律資訊

520 TTS 核心引擎　　522 韻律預測模組

620 STS 核心引擎　　622 韻律估算模組

(μ_{shift}，σ_{center}，γ_{σ}) 三個可調控參數

910 文字語料庫　　920 原始錄音語料庫

930 TTS 系統　　940 合成之語料庫

950 韻律差異　　960 韻律重估測模型

1110 準備一個可調控式韻律參數介面，以供輸入一可調控參數組

1120 根據輸入文稿或輸入語音來預測出或估算出韻律資訊

1130 建構一韻律重估測模型，並根據此可調控參數組及預測出或估算出的韻律資訊，藉由此韻律重估測模型來調整出新的韻律資訊

1140 將此新的韻律資訊提供給一語音合成模組以產生合成語音

1200 韻律重估測系統　　1210 處理器

1290 記憶體裝置

1310 基於 HMM 之 TTS 方法的 TTS 的音高曲線

[S]

201227714

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：99145318

※ 申請日： 99.12.22 ※IPC 分類： G10L 13/00 (2006.01)

一、發明名稱：（中文/英文）

可調控式韻律重估測系統與方法及電腦程式產品/
CONTROLLABLE PROSODY RE-ESTIMATION SYSTEM AND METHOD AND COMPUTER PROGRAM PRODUCT THEREOF

二、中文發明摘要：

在一種可調控式韻律重估測系統的一實施範例中，一個語音或文字轉語音的核心引擎係由一韻律預測或估算模組、一韻律重估測模組、以及一語音合成模組所組成。此韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至此韻律重估測模組。此韻律重估測模組根據由一個可調控式韻律參數介面所提供的一可調控參數組及收到的韻律資訊，將此韻律資訊重估測後，產生新的韻律資訊，再提供給此語音合成模組以產生合成語音。

三、英文發明摘要：

In one embodiment of a controllable prosody re-estimation system, a TTS engine consists of a prosody prediction/estimation module, a prosody re-estimation module and a speech synthesis module. The prosody prediction/estimation module generates predicted or estimated prosody information according to input text or speech, and transmits the generated prosody information to the prosody re-estimation module. The prosody re-estimation module re-estimates the generated prosody information and produces new prosody information, according to a set of controllable parameters provided by a controllable prosody parameter interface. The new prosody information is provided to the speech synthesis module to produce a synthesized speech.

七、申請專利範圍：

1. 一種可調控式韻律重估測系統，該系統包含:

一個可調控式韻律參數介面，用來輸入一可調控參數組;

以及

一個語音或文字轉語音的核心引擎，該核心引擎至少由一韻律預測或估算模組、一韻律重估測模組、及一語音合成模組所組成，其中該韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至該韻律重估測模組，該韻律重估測模組根據輸入的該可調控參數組及收到的韻律資訊，將該韻律資訊重估測後，產生新的韻律資訊，再提供給該語音合成模組以產生合成語音。

2. 如申請專利範圍第 1 項所述之系統，其中該可調控參數組中的參數彼此是獨立的。

3. 如申請專利範圍第 1 項所述之系統，該韻律重估測系統應用在文字轉語音上時，該韻律預測或估算模組扮演一韻律預測模組的角色，根據該輸入文稿來預測出該韻律資訊。

4. 如申請專利範圍第 1 項所述之系統，該韻律重估測系統應用在語音轉語音上時，該韻律預測或估算模組扮演一韻律估算模組的角色，根據該輸入語音來估算出該韻律資訊。

5. 如申請專利範圍第 1 項所述之系統，該系統還建構一韻律重估測模型，並且該韻律重估測模組採用該韻律資訊重估測模型來將該韻律資訊重估測，以產生該新的

[S]

韻律資訊。

6. 如申請專利範圍第 5 項所述之系統，該系統係透過一原始錄音語料庫以及一合成之語料庫來建構該韻律重估測模型。

7. 如申請專利範圍第 1 項所述之系統，其中該可調控參數組包括多個可調控參數，並且當其中至少一參數未輸入時，該系統提供該未輸入之至少一參數的預設值。

8. 如申請專利範圍第 5 項所述之系統，其中該韻律重估測模型以下列的形式來表達:

$$\hat{X}_{tar} = \mu_{shift} + (X_{src} - \mu_{center}) \cdot \gamma_{\sigma}$$

其中，X_{src}代表由一來源語音所產生的韻律資訊，$\hat{X}_{tar}$代表該新的韻律資訊，μ_{center}、μ_{shift}、及γ_{σ}是三個可調控參數。

9. 如申請專利範圍第 8 項所述之系統，其中當μ_{center}未輸入時，該系統設定μ_{center}的預設值為一來源語料庫的韻律平均值，當μ_{shift}未輸入時，該系統設定μ_{shift}的預設值為一目標語料庫的韻律平均值，當γ_{σ}未輸入時，該系統設定γ_{σ}的預設值為$\sigma_{tar}/\sigma_{src}$，$\sigma_{tar}$為一目標語料庫的韻律標準差，$\sigma_{src}$為一來源語料庫的韻律標準差。

10. 一種可調控式韻律重估測系統，係執行於一電腦系統中，該電腦系統備有一記憶體裝置，用來儲存一原始錄音語料庫與一合成之語料庫，該韻律重估測系統包含:

一可調控式韻律參數介面，用來輸入一可調控參數組;

以及

一處理器，該處理器備有一韻律預測或估算模組、一韻律重估測模組、及一語音合成模組，該韻律預測或估算模組根據輸入文稿或輸入語音來預測出或估算出韻律資訊，並傳送至該韻律重估測模組，該韻律重估測模組根據輸入的該可調控參數組及收到的韻律資訊，將該韻律資訊重估測後，產生新的韻律資訊，再提供給該語音合成模組以產生合成語音；

其中，該處理器統計該兩語料庫之韻律差異來建構一韻律重估測模型，以提供給該韻律資訊重估測模組使用。

11. 如申請專利範圍第 10 項所述之系統，該電腦系統包括該處理器。

12. 如申請專利範圍第 10 項所述之系統，其中該韻律重估測模型以下列的形式來表達：

$$\hat{X}_{tar} = \mu_{shift} + (X_{src} - \mu_{center}) \cdot \gamma_{\sigma}$$

其中，X_{src} 代表由一來源語音所產生的韻律資訊，$\hat{X}_{tar}$ 代表該新的韻律資訊，μ_{center}、μ_{shift}、及 γ_{σ} 是三個可調控參數。

13. 如申請專利範圍第 12 項所述之系統，其中當 μ_{center} 未輸入時，該系統設定 μ_{center} 的預設值為一來源語料庫的韻律平均值，當 μ_{shift} 未輸入時，該系統設定 μ_{shift} 的預設值為一目標語料庫的韻律平均值，當 γ_{σ} 未輸入時，該系統設定 γ_{σ} 的預設值為 $\sigma_{tar}/\sigma_{src}$，$\sigma_{tar}$ 為一目標語料庫的韻律標準差，σ_{src} 為一來源語料庫的韻律標準差。

14. 如申請專利範圍第 10 項所述之系統，該系統利用一單

[S]

句統計法來獲得該韻律重估測模型。

15. 一種可調控式韻律重估測方法，係執行於一可調控式韻律重估測系統或一電腦系統中，該方法包含:

準備一個可調控式韻律參數介面，以供輸入一可調控參數組;

根據輸入文稿或輸入語音來預測出或估算出韻律資訊;

建構一韻律重估測模型，並根據該可調控參數組及該預測出或估算出的韻律資訊，藉由該韻律重估測模型來調整出新的韻律資訊;以及

將該新的韻律資訊套用至一語音合成模組以產生合成語音。

16. 如申請專利範圍第 15 項所述之方法，其中該可調控參數組包括多個可調控參數，並且當其中至少一參數未輸入時，該方法還包括設定該未輸入之至少一參數的預設值，並且該至少一參數的預設值係統計兩平行語料庫的韻律分佈而得出。

17. 如申請專利範圍第 15 項所述之方法，其中該韻律重估測模型係經由統計兩平行語料庫的韻律差異而建構，該兩平行語料庫為一原始錄音語料庫以及一合成之語料庫。

18. 如申請專利範圍第 17 項所述之方法，其中該原始錄音語料庫是根據一個給定的文字語料庫而錄製的原始錄音語料庫，而該合成之語料庫是經由該原始錄音語料庫訓練出的一文字轉語音系統所合成語句的語料庫。

19. 如申請專利範圍第 15 項所述之方法，該方法係利用一 [S]

靜態分佈法來獲得該韻律重估測模型。

20. 如申請專利範圍第 17 項所述之方法，該方法係利用一單句統計法來獲得該韻律重估測模型。

21. 如申請專利範圍第 15 項所述之方法，其中該韻律重估測模型以下列的形式來表達:

$$\hat{X}_{tar} = \mu_{shift} + (X_{src} - \mu_{center}) \cdot \gamma_{\sigma}$$

其中，X_{src}代表由一來源語音所產生的韻律資訊，$\hat{X}_{tar}$代表該新的韻律資訊，μ_{center}、μ_{shift}、及γ_{σ}是三個可調控參數。

22. 如申請專利範圍第 20 項所述之方法，其中該單句統計法還包括:

以該原始錄音語料庫及該合成語料庫的每一句子為基本單位，比較該兩語料庫的每一句子間的韻律差異性並統計彼此的差異;

根據該統計出的差異，利用一回歸法，建立一回歸模型;以及

在合成語音時，以該回歸模型來預測一輸入語句的目標韻律分佈。

23. 如申請專利範圍第 21 項所述之方法，其中當μ_{center}未輸入時，該方法設定μ_{center}的預設值為一來源語料庫的韻律平均值，當μ_{shift}未輸入時，該方法設定μ_{shift}的預設值為一目標語料庫的韻律平均值，當γ_{σ}未輸入時，該方法設定γ_{σ}的預設值為$\sigma_{tar}/\sigma_{src}$，$\sigma_{tar}$為一目標語料庫的韻律標準差，$\sigma_{src}$為一來源語料庫的韻律標準差。

[S]

24. 一種可調控式韻律重估測的電腦程式產品，該電腦程式產品包含一記憶體以及儲存於該記憶體的一可執行的電腦程式，該電腦程式藉由一處理器來執行:

準備一個可調控式韻律參數介面，以供輸入一可調控參數組;

根據輸入文稿或輸入語音來預測出或估算出韻律資訊;

建構一韻律重估測模型，並根據該可調控參數組及預測出或估算出的韻律資訊，藉由一韻律重估測模型來調整出新的韻律資訊;以及

將該新的韻律資訊提供給一語音合成模組以產生合成語音。

25. 如申請專利範圍第 24 項所述之電腦程式產品，其中該韻律重估測模型係經由統計兩平行語料庫的韻律差異而建構，該兩平行語料庫為一原始錄音語料庫以及一合成之語料庫。

26. 如申請專利範圍第 25 項所述之電腦程式產品，其中該韻律重估測模型係利用一單句統計法來獲得。

27. 如申請專利範圍第 24 項所述之電腦程式產品，其中該韻律重估測模型以下列的形式來表達:

$$\hat{X}_{tar} = \mu_{shift} + (X_{src} - \mu_{center}) \cdot \gamma_{\sigma}$$

其中，X_{src}代表由一來源語音所產生的韻律資訊，$\hat{X}_{tar}$代表該新的韻律資訊，μ_{center}、μ_{shift}、及γ_{σ}是三個可調控參數。

28. 如申請專利範圍第 26 項所述之電腦程式產品，其中該

[S]

單句統計法還包括:

以該原始錄音語料庫及該合成語料庫的每一句子為基本單位，比較該兩語料庫的每一句子間的韻律差異性並統計彼此的差異;

根據該統計出的差異，利用一回歸法，建立一回歸模型;以及

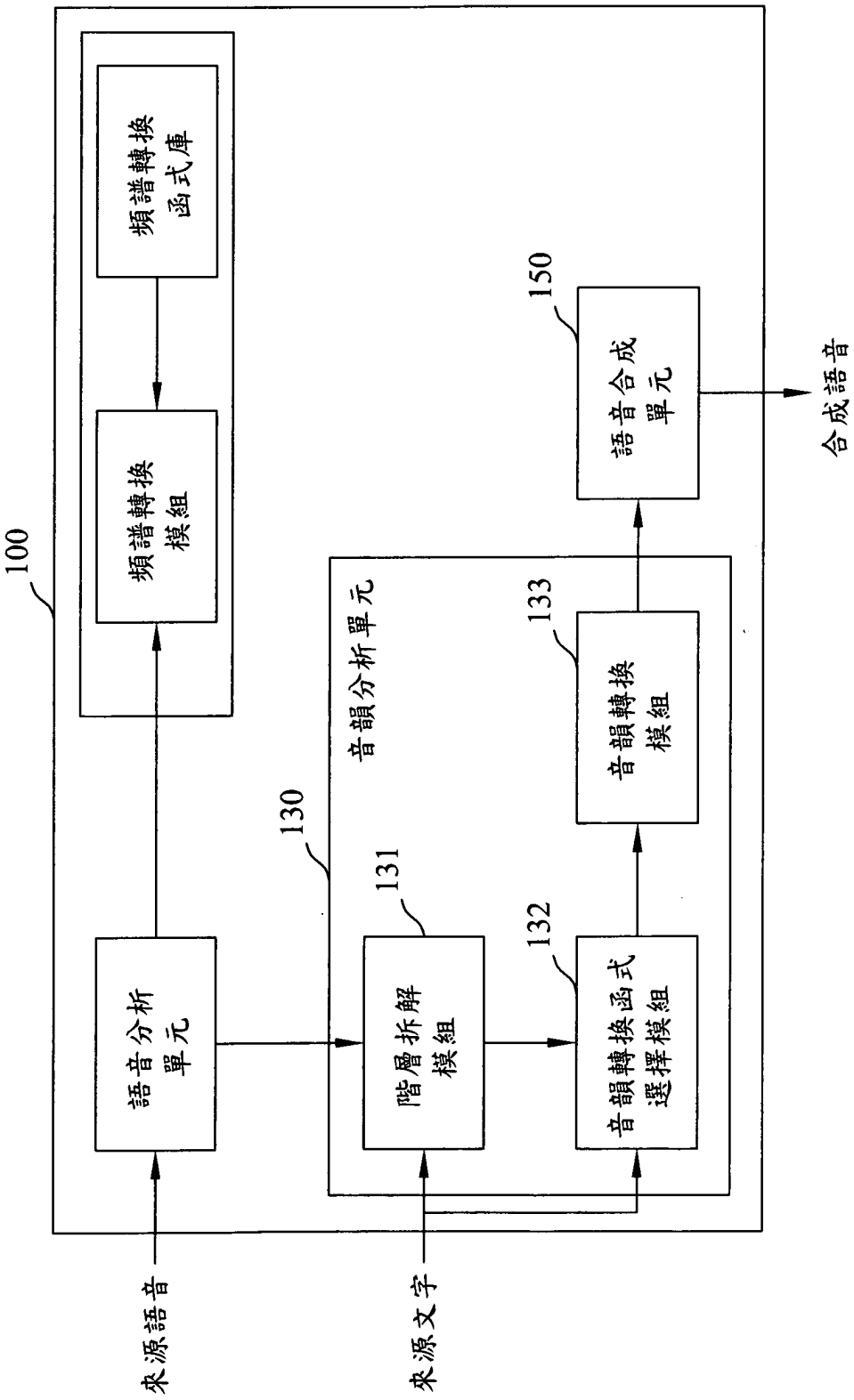
在合成語音時，以該回歸模型來預測一輸入語句的目標韻律分佈。

29. 如申請專利範圍第 28 項所述之電腦程式產品，其中當 μ_{center} 未輸入時，該方法設定 μ_{center} 的預設值為一來源語料庫的韻律平均值，當 μ_{shift} 未輸入時，該方法設定 μ_{shift} 的預設值為一目標語料庫的韻律平均值，當 γ_{σ} 未輸入時，該方法設定 γ_{σ} 的預設值為 $\sigma_{tar}/\sigma_{src}$，$\sigma_{tar}$ 為一目標語料庫的韻律標準差，σ_{src} 為一來源語料庫的韻律標準差。

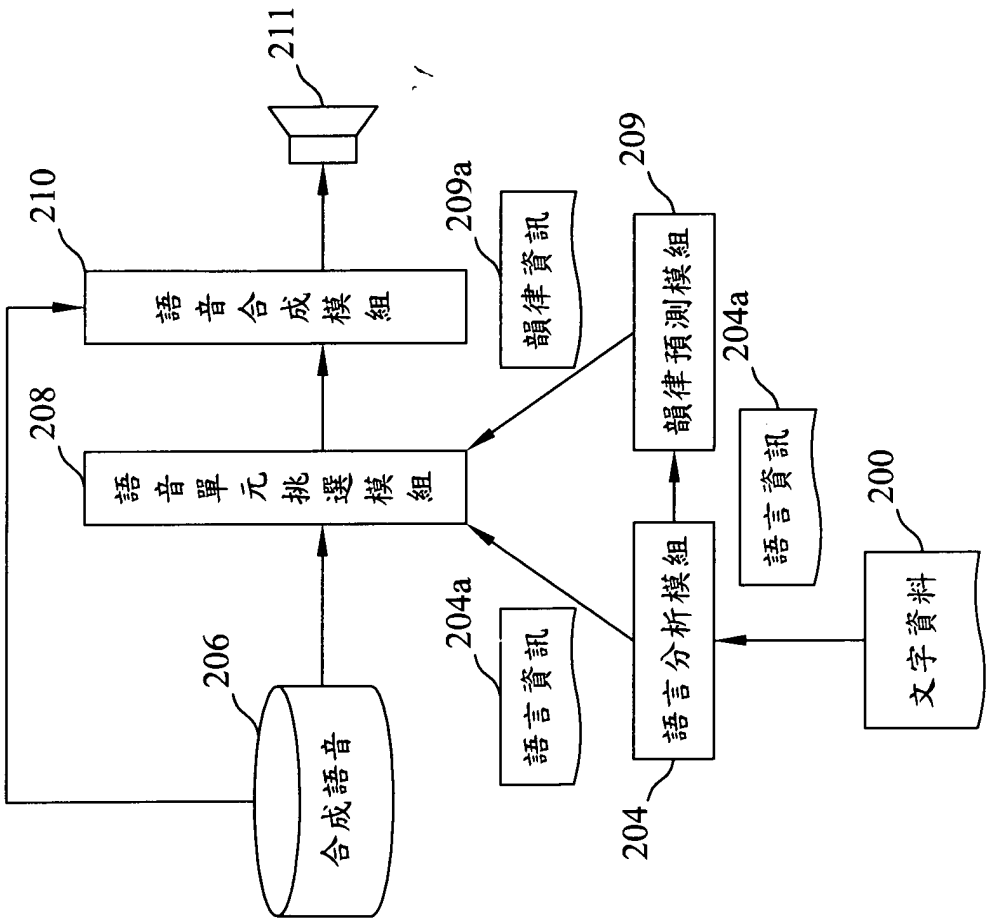
30. 如申請專利範圍第 25 項所述之電腦程式產品，其中該韻律重估測模型係利用一靜態分佈法來獲得。

[S]

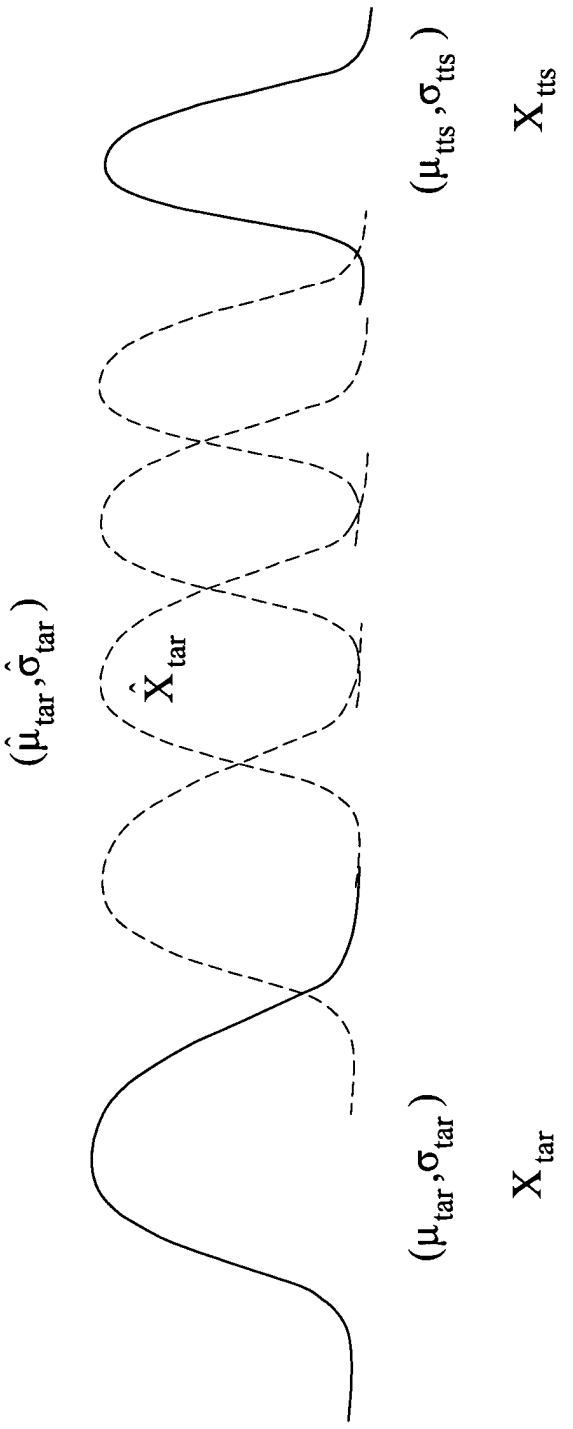
八、圖式


100
頻譜轉換函式庫
頻譜轉換模組
語音分析單元
130
音韻分析單元
131
階層拆解模組
132
音韻轉換選擇模組
133
音韻轉換模組
150
語音合成單元
合成語音
來源語音
來源文字

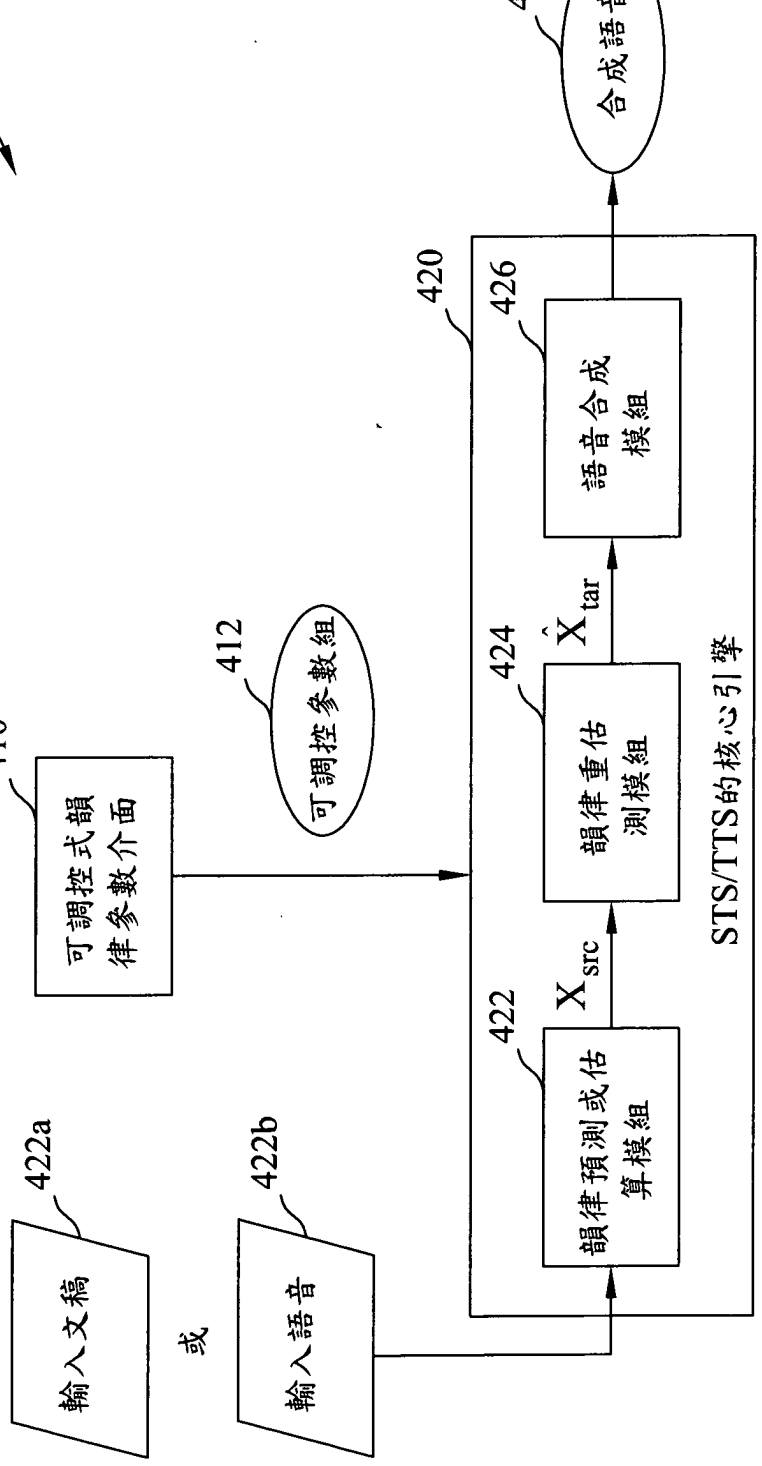
第一圖


211
210
語音合成模組
209a
韻律資訊
209
韻律預測模組
208
語音單元挑選模組
204a
語言資訊
206
合成語音
204a
語言資訊
語言分析模組
200
文字資料
204

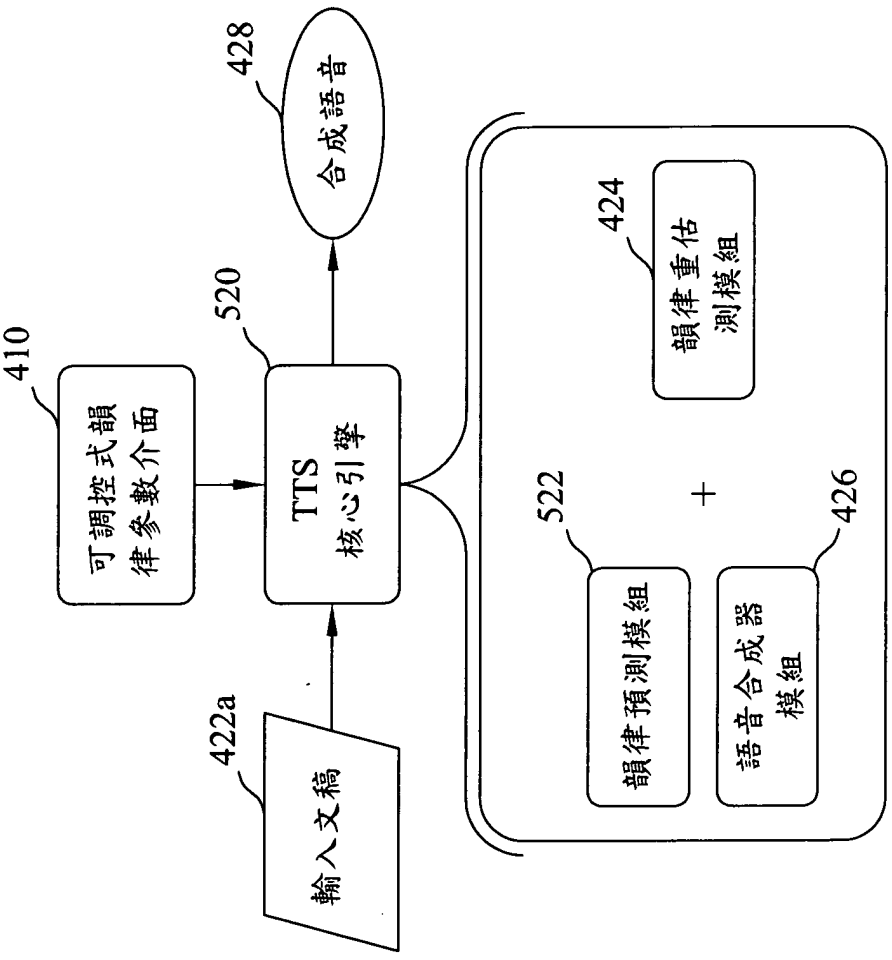
第二圖


$(\hat{\mu}_{tar},\hat{\sigma}_{tar})$
$\hat{X}_{tar}$
(μ_{tts},σ_{tts})
X_{tts}
(μ_{tar},σ_{tar})
X_{tar}

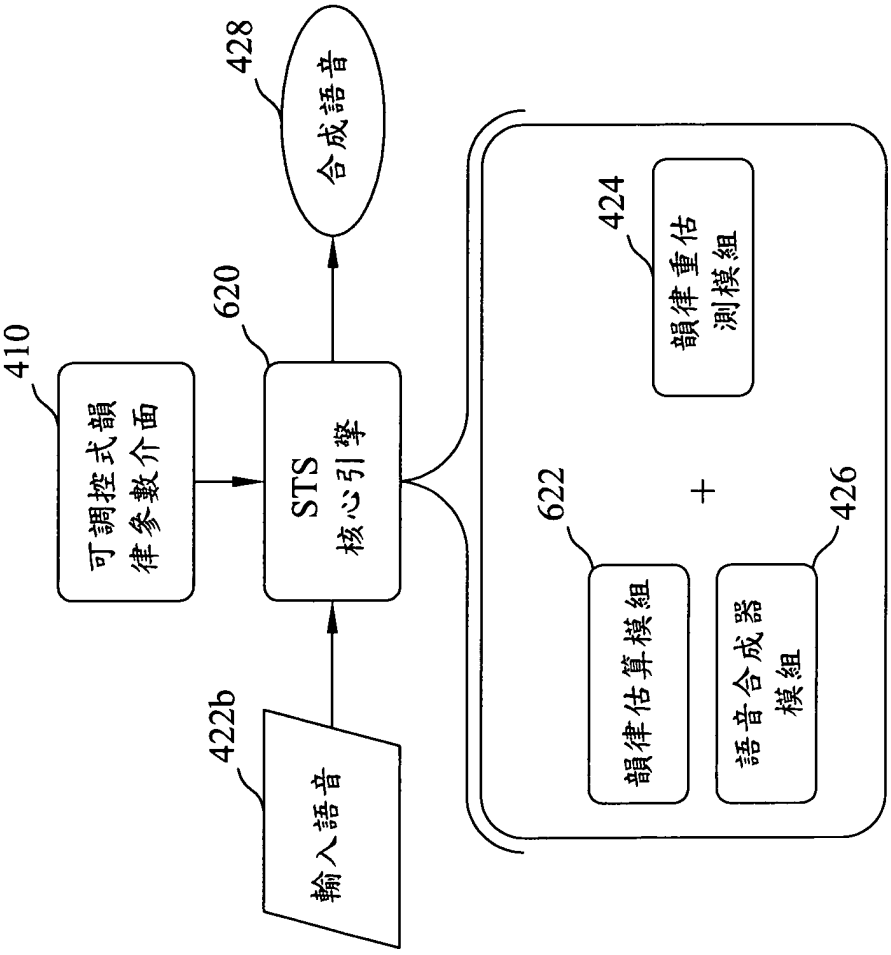
第三圖


400
428
合成語音
420
426
語音合成模組
412
可調控參數組
410
可調控式韻律參數介面
424
韻律重估測模組
$\hat{X}_{tar}$
422
韻律預測或估算模組
X_{src}
STS/TTS的核心引擎
422a
輸入文稿
或
422b
輸入語音

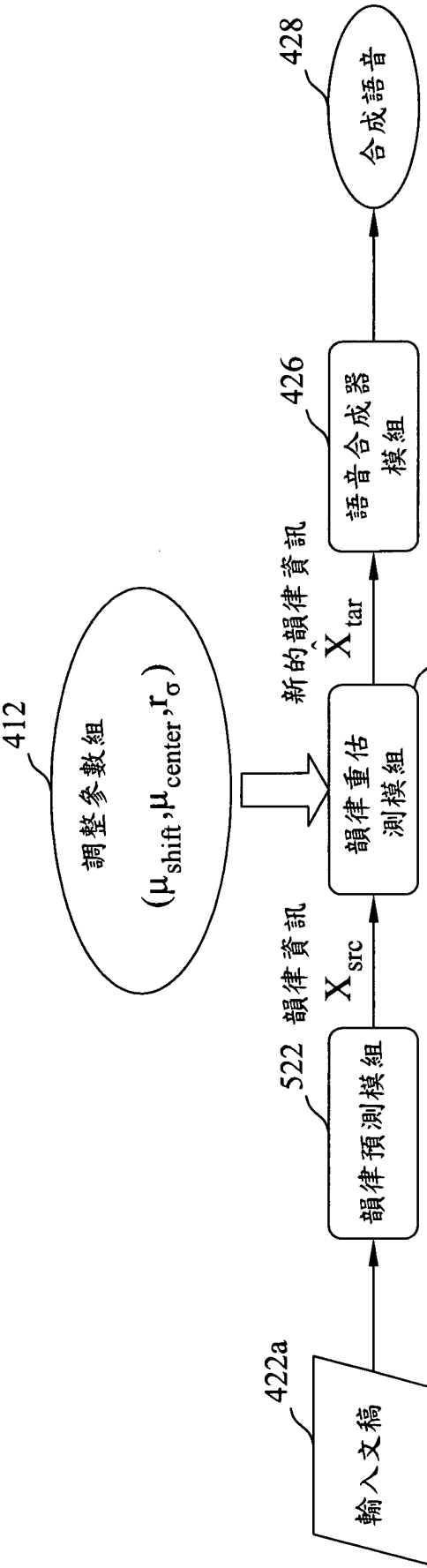
第四圖


410
可調控式韻律參數介面
520
TTS
核心引擎
428
合成語音
422a
輸入文稿
424
韻律重估測模組
522
韻律預測模組
+
426
語音合成器模組

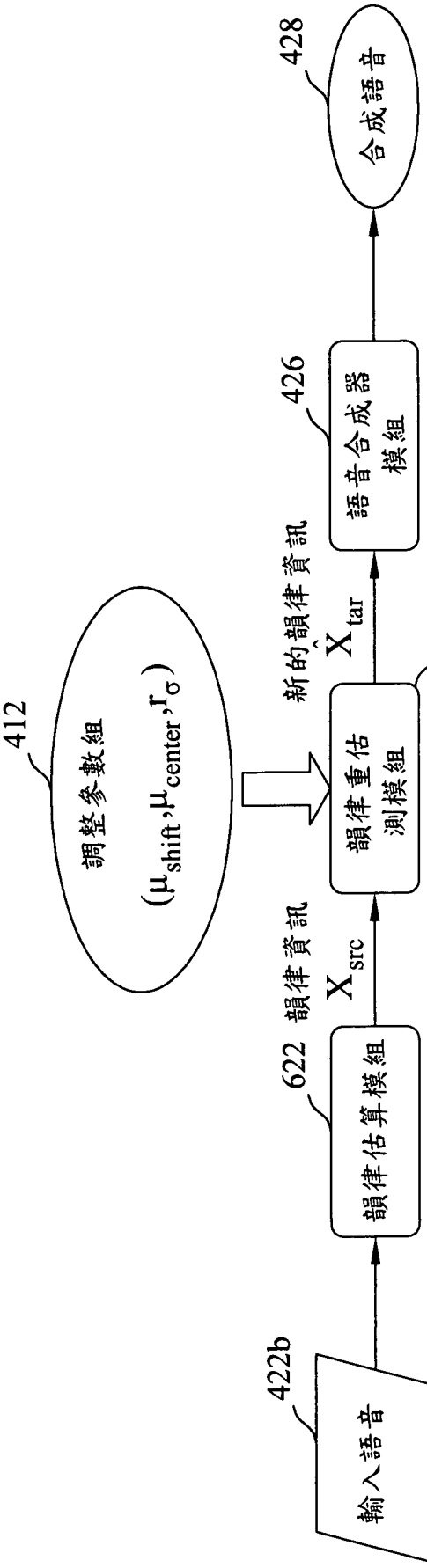
第五圖


410
可調控式韻律參數介面
620
STS
核心引擎
428
合成語音
422b
輸入語音
424
韻律重估測模組
622
韻律估算模組
+
426
語音合成器模組

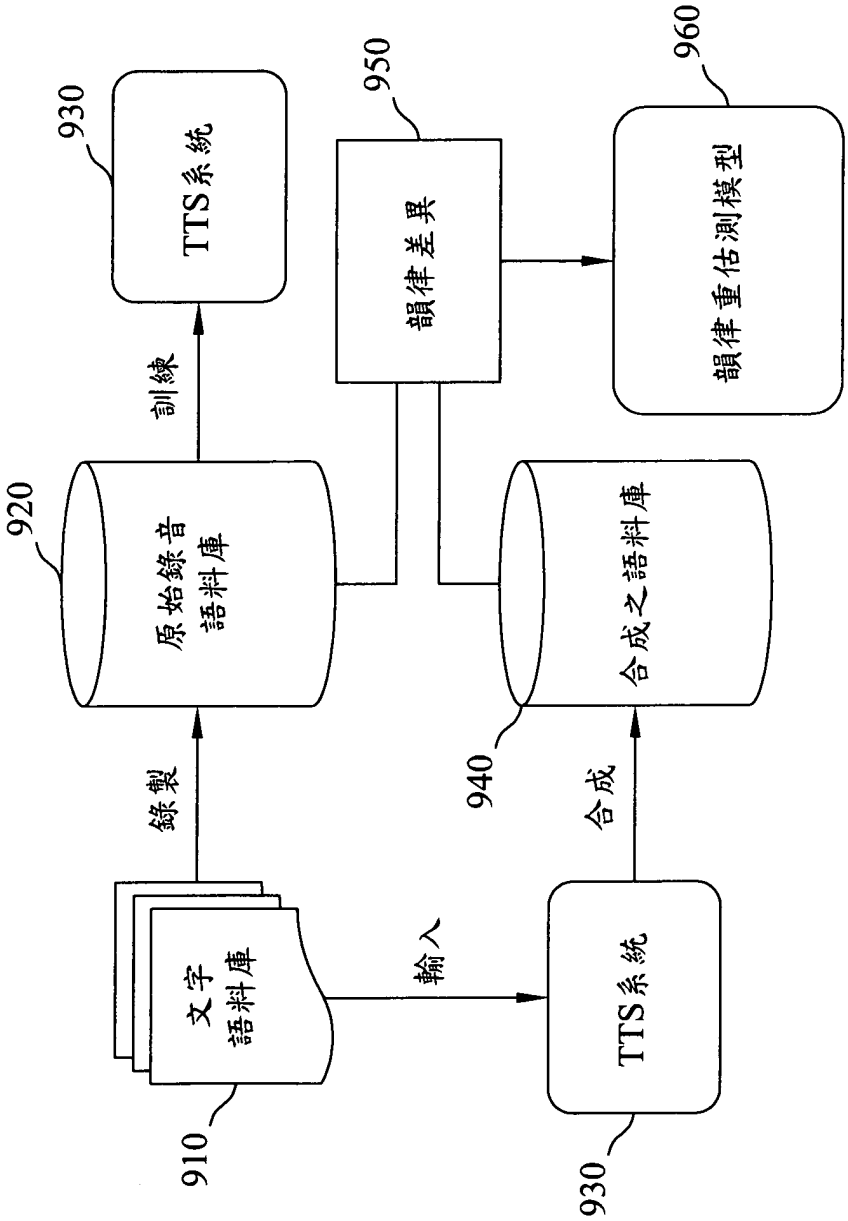
第六圖


422a
輸入文稿
522 韻律預測模組
韻律資訊 X_{src}
412
調整參數組
(μ_{shift},μ_{center},r_σ)
韻律重估測模組
424
新的韻律資訊 $\hat{X}_{tar}$
426
語音合成器模組
428
合成語音

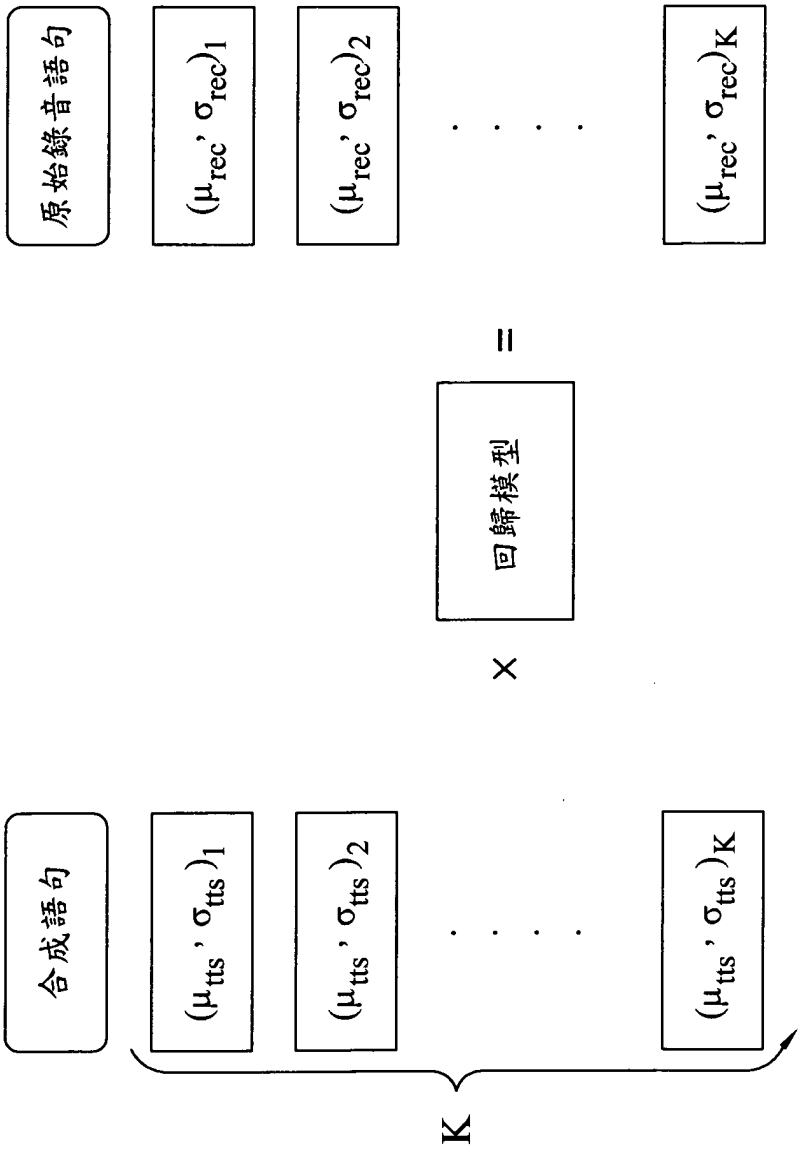
第七圖


輸入語音
422b
韻律估算模組
622
韻律資訊
X_{src}
韻律重估測模組
424
新的韻律資訊
$\hat{X}_{tar}$
語音合成器模組
426
合成語音
428
調整參數組
$(\mu_{shift}, \mu_{center}, r_{\sigma})$
412

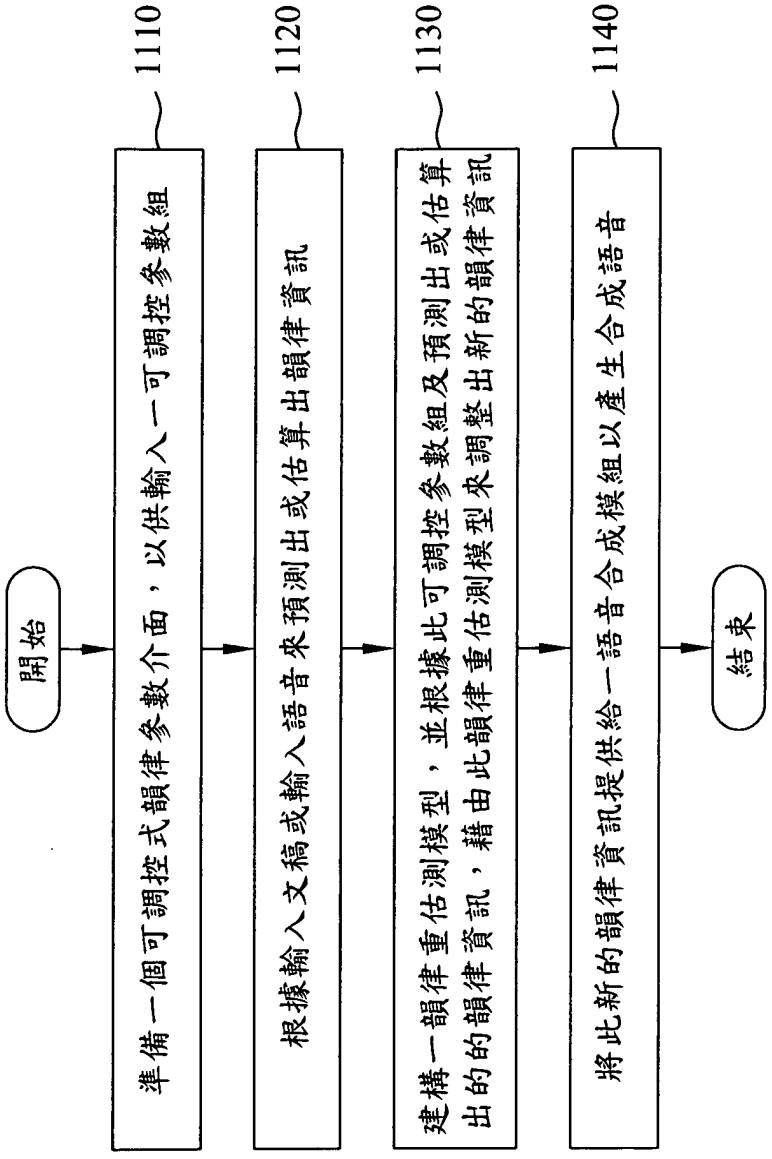
第八圖


930
TTS系統
950
960
韻律差異
韻律重估測模型
訓練
920
原始錄音
語料庫
合成之語料庫
940
錄製
合成
輸入
文字
語料庫
910
TTS系統
930

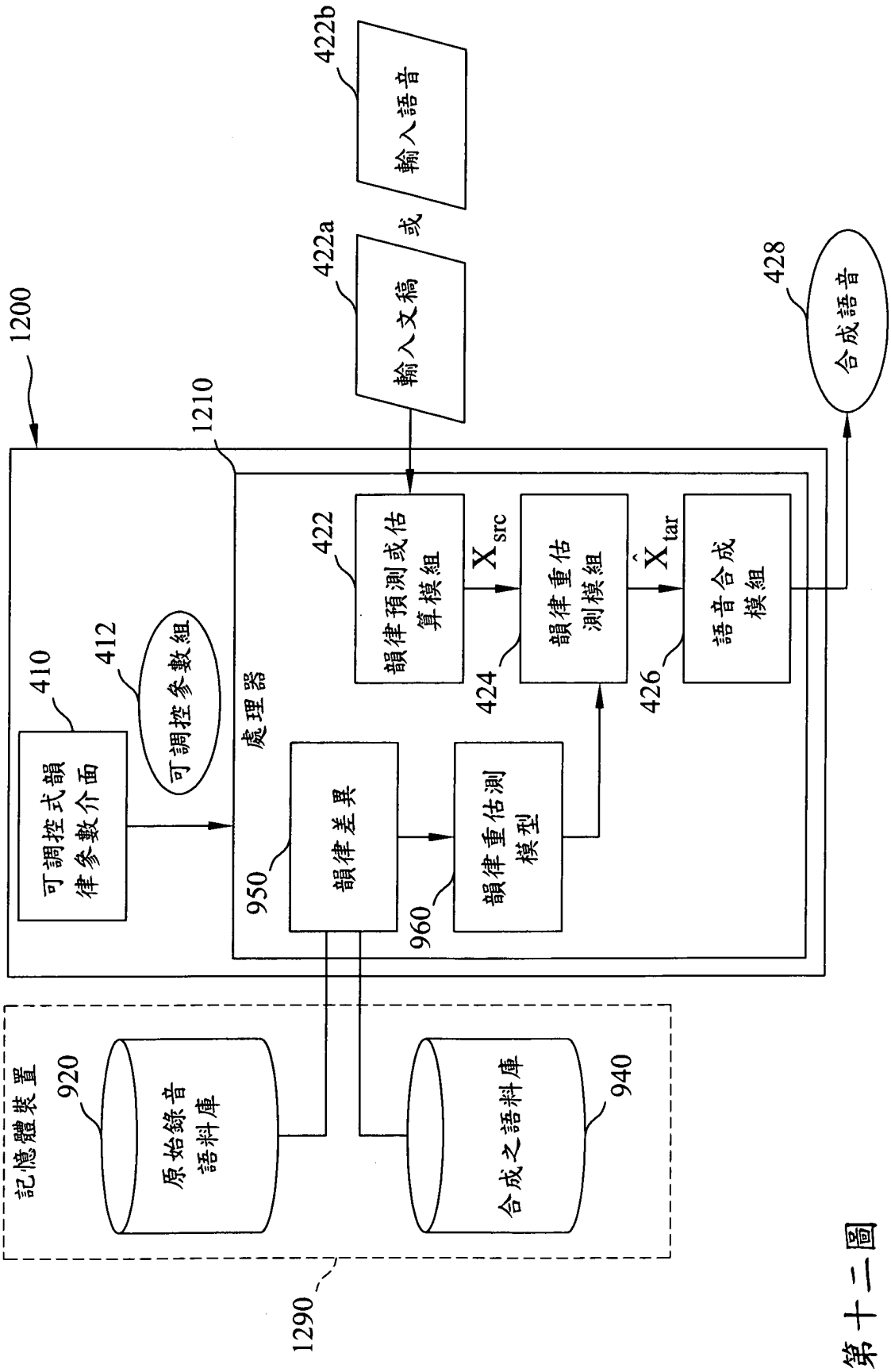
第九圖


合成語句
$(\mu_{tts}, \sigma_{tts})_1$
$(\mu_{tts}, \sigma_{tts})_2$
$(\mu_{tts}, \sigma_{tts})_K$
K
×
回歸模型
=
原始錄音語句
$(\mu_{rec}, \sigma_{rec})_1$
$(\mu_{rec}, \sigma_{rec})_2$
$(\mu_{rec}, \sigma_{rec})_K$

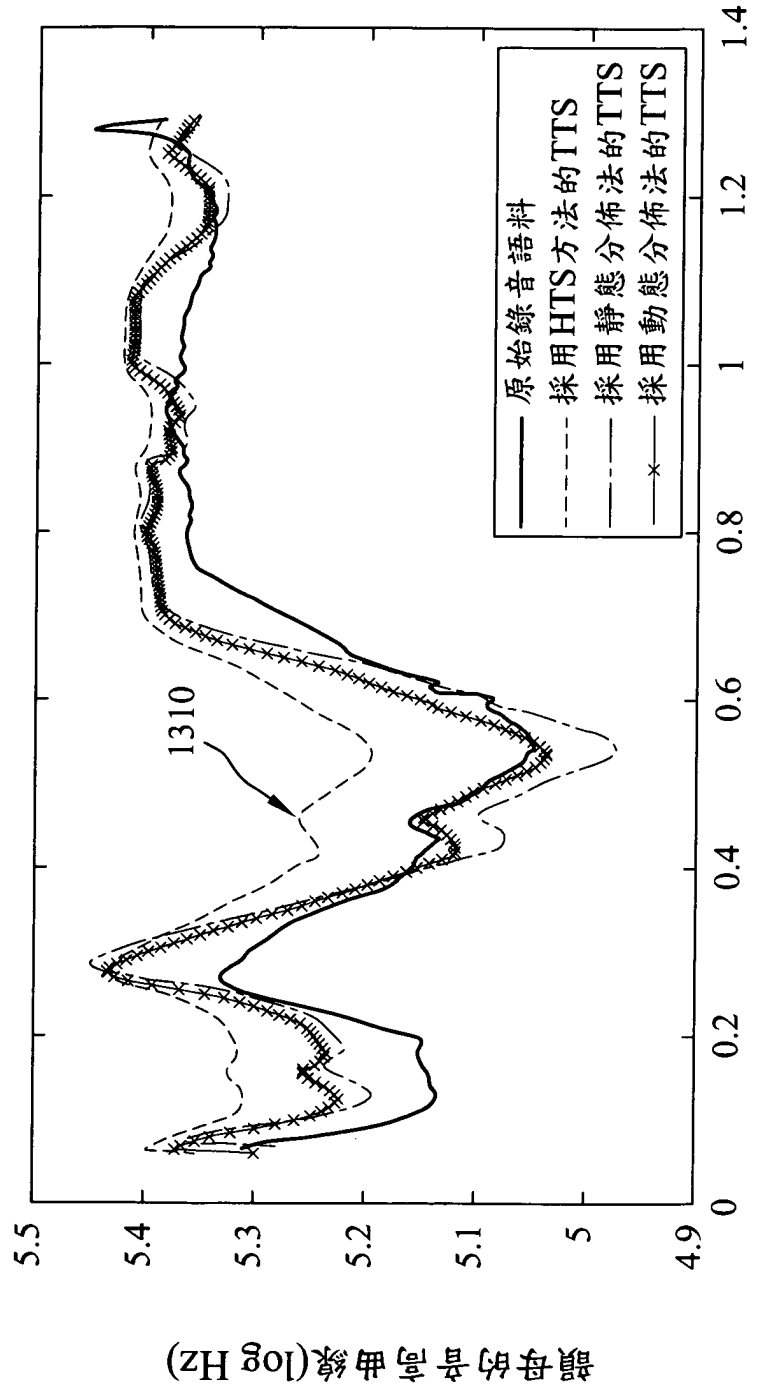
第十圖


開始
準備一個可調控式韻律參數介面，以供輸入一可調控參數組
1110
根據輸入文稿或輸入語音來預測出或估算出韻律資訊
1120
建構一韻律重估測模型，並根據此可調控參數組及預測出或估算出的的韻律資訊，藉由此韻律重估測模型來調整出新的韻律資訊
1130
將此新的韻律資訊提供給一語音合成模組以產生合成語音
1140
結束

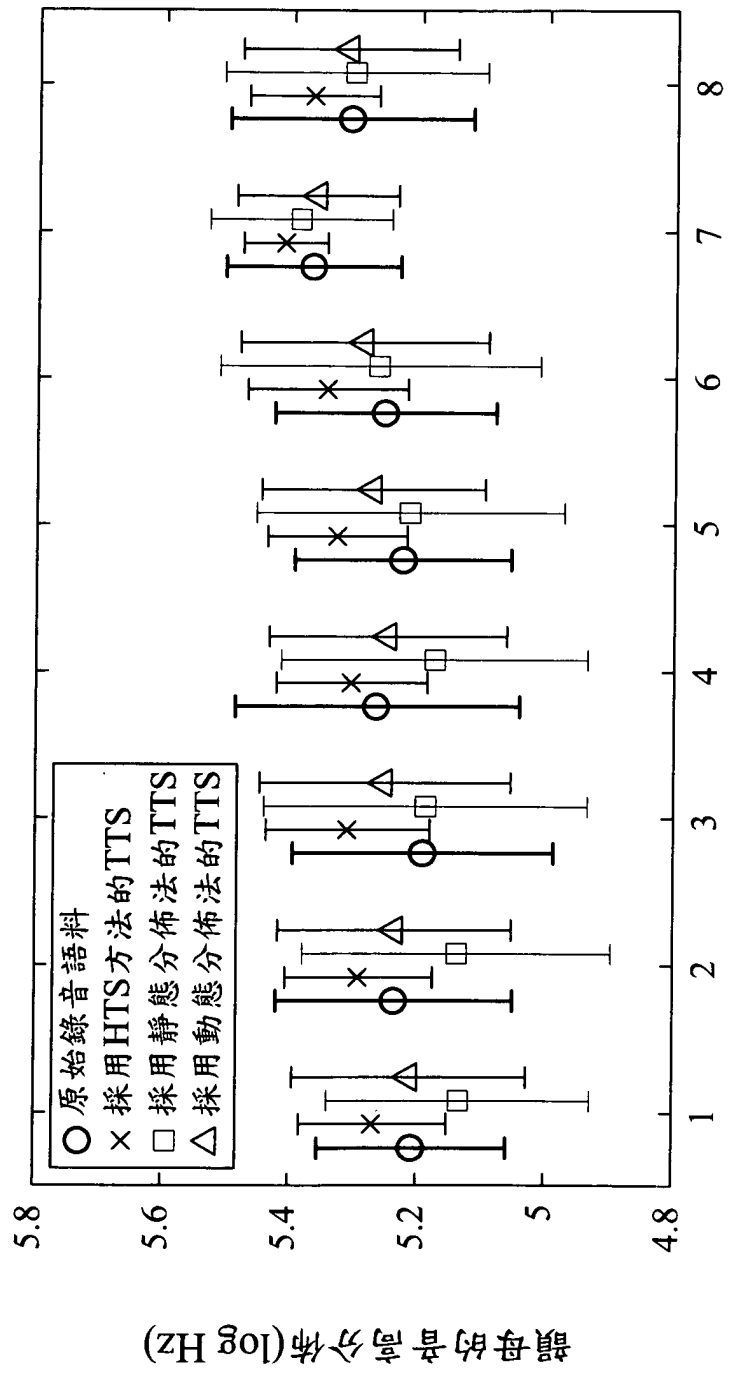
第十一圖


1200
1210
422b
輸入語音
或
422a
輸入文稿
428
合成語音
422
韻律預測或估算模組
X_{src}
韻律重估測模組
$\hat{X}_{tar}$
語音合成模組
424
426
410
可調控式韻律參數介面
412
可調控參數組
處理器
950
韻律差異
960
韻律重估測模型
記憶體裝置
920
原始錄音語料庫
940
合成之語料庫
1290

第十二圖


原始錄音語料
採用HTS方法的TTS
採用靜態分佈法的TTS
採用動態分佈法的TTS
1310
5.5
5.4
5.3
5.2
5.1
5
4.9
0
0.2
0.4
0.6
0.8
1
1.2
1.4
時間長度(秒)
韻母的音高曲線(log Hz)

第十三圖


○ 原始錄音語料
× 採用HTS方法的TTS
□ 採用靜態分佈法的TTS
△ 採用動態分佈法的TTS
5.8
5.6
5.4
5.2
5
4.8
1
2
3
4
5
6
7
8
韻母的音高分佈(log Hz)
句子的號碼

第十四圖

5.6
5.4
5.2
5
4.8
0
0.5
1
1.5
2
時間長度(秒)
韻母的音高曲線(log Hz)
採用HTS方法的TTS
外國腔調
類似機器人

第十五圖

四、指定代表圖：

(一)本案指定代表圖為：第（四）圖。

(二)本代表圖之元件符號簡單說明：

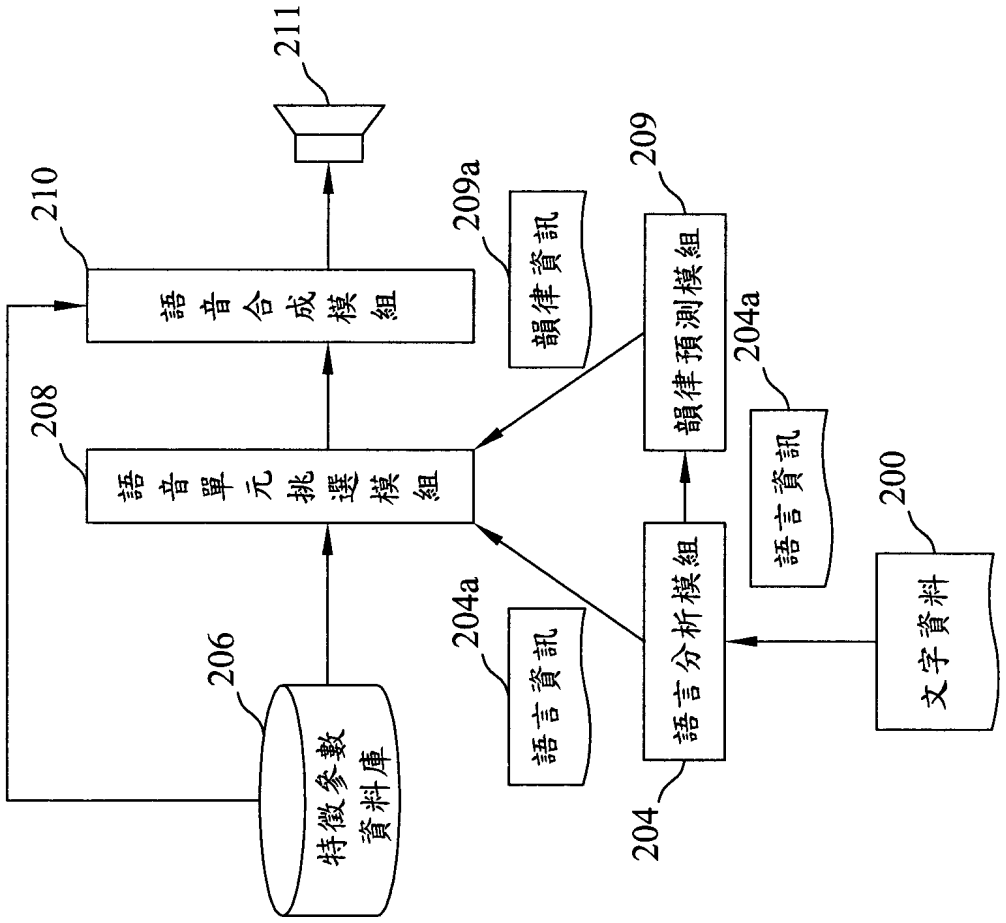
400 韻律重估測系統	410 可調控式韻律參數介面
412 可調控參數組	420 STS/TTS 的核心引擎
422 韻律預測或估算模組	422a 輸入文稿
422b 輸入語音	424 韻律重估測模組
426 語音合成模組	428 合成語音
X_{src} 韻律資訊	$\hat{X}_{tar}$ 調整後的韻律資訊

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

[S]

[S]

99145318



第二圖