

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：95125463

※申請日期：95.7.13

※IPC 分類：

G10L 13/02

一、發明名稱：

歌聲合成方法及其架構

二、申請人：(共1人)

姓名或名稱：古鴻炎

代表人：

住居所或營業所地址：台北市基隆路四段43號資工系

國籍：中華民國

三、發明人：(共2人)

姓名：1. 古鴻炎

2. 廖皇量

國籍：1-2 中華民國

四、聲明事項：

☒ 主張專利法第二十二條第二項 ☒ 第一款或 ☐ 第二款規定之事實，其事實發生日期為：95 年 1 月 17 日。

☐ 申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

☐ 有主張專利法第二十七條第一項國際優先權：

☐ 無主張專利法第二十七條第一項國際優先權：

☐ 主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

☐ 主張專利法第三十條生物材料：

☐ 須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

☐ 不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明係關於一種歌聲合成方法及其架構，特別是指一種以諧波加噪音模型(HNM)作為基礎，針對最大有聲頻率(MVF)的偵測、基頻的偵測、諧波參數的擷取等方面加以改善，並擴增數個用以提升歌聲合成品質的處理方法，包括相位方面的處理，起音、衰減、延持、釋音(ADSR)式的音長分配，加入低頻噪音，以及諧波追蹤等的歌聲合成方法及其架構。

【先前技術】

在人的一生中，聲音跟人類的生活習習相關，從人與人的溝通到欣賞蟲鳴鳥叫，各種的聲音豐富了我們的生活。從二十世紀邁入電腦時代之後，電腦的誕生促成了新的技術革命，已經深深地改變了世界的面貌，利用電腦技術，可以合成各種音樂聲和各種音色，也可以合成現實世界上還未有過的聲音；此外，人們也可以藉由電腦歌聲合成的功能來學習新的歌曲，學習歌譜上的老歌或民謠；而專業的作詞、作曲者，也可藉由電腦歌聲合成，來聆聽、評估自己的作品；然而，目前現有的歌聲合成技術中，由電腦合成出來的歌聲與真人發出的歌聲之間仍有一段差距，因此，如何讓電腦可以合成出音色、音質以及歌唱技巧近似真人的歌聲，仍然是一個問題。

習知之歌聲合成研究中，大致可分類成時域及頻域兩種處理方法；屬於時域的合成方法如：基週同步疊加(Pitch Synchronous Overlap Add, PSOLA)、基週波形時間比例內差法(Time-Proportionated Interpolation of Pitch Waveforms, TIPW)、波表合成法。一般來說，時域上的處理方法，事

先需作的參數分析(如：基週頂點標記)較為簡單、直接，且合成處理的計算量較少。前述之 PSOLA 及 TIPW 兩方法，係源自於語音合成之研究，不過，可用於語音合成的方法，不一定就適合用於歌聲合成，因為歌聲需求的音長、音高變化(音域)是較大的；波表合成法則是一種常被用於作樂器聲音合成的方法，目前並無關於此法用於作歌聲合成的研究報告。

此外，習知在頻域上作參數分析，再據以作樂音(樂器聲、人聲)信號合成的方法，目前已發展出至少三類以上的方法，主要的類別為：(a)加法式合成(additive synthesis)，先產生出各個諧波後再相加；(b)減法式合成(subtractive synthesis)，如：線性預測編碼(Linear Predictive Coding, LPC)之合成方法；(c)頻率調變(frequency modulation)合成，此方法多被應用於樂器聲之合成。除此之外，目前亦有使用弦波模型(sinusoidal model)作歌聲合成者，弦波模型基本上是採取頻域上加法式合成的觀念。

Y. Stylianou 在 1996 年提出的諧波加噪音模型(Harmonic plus Noise Model, HNM)，除了考慮信號中的諧波成分之外，還加入了信號中高頻部份的噪音，使得所合成出之聲音特性可以更接近原始所錄的聲音。

在 HNM 模型中，會逐一對各個聲音信號之音框(frame)作頻譜分析，當分析的是一個有聲的(voiced)信號音框，信號所具有的週期性之特性，可以在頻譜圖上很容易地觀察到：頻譜中曲線的峰值會出現在基頻和其倍數的頻率值上，但是這些峰值的出現並不會含蓋整個頻率域的範圍(0Hz 到取樣率的一半)，而是會分佈在某個頻率值之前。HNM 就是依據聲音信號的這種特性，將信號分解為諧波 $h(t)$ 和噪音 $n(t)$ 兩個部份來表示的模型；對一個

聲音信號中有聲的部份，HNMM 模型會先依信號的頻譜計算出最大有聲頻率 (Maximum Voiced Frequency, MVF) $F_m(t)$ ，最大有聲頻率 $F_m(t)$ 是一個隨 t 時變的值，依據 $F_m(t)$ ，第 t 音框裡頻率值小於 MVF 的頻譜部份，在 HNMM 中視為諧波部份，而頻率值大於 MVF 的頻譜部份，則視為噪音部份，亦即一個頻譜會被分割成譜波與噪音等兩部份。

諧波部份 $h(t)$ ，可以看成是由頻率值有倍數關係的數個弦波所組成，以加法式弦波模型來合成出信號，如公式 1 所示：

$$h(t) = \sum_{k=1}^{K(t)} \alpha_k(t) \cos(2\pi k f_0 t + \phi_k(t)) \quad (\text{公式 1})$$

其中 $\alpha_k(t)$ 和 $\phi_k(t)$ 表示在時間為 t 時，第 k 個弦波的振幅和相位， $K(t)$ 則代表此時諧波部份所包含的弦波個數， f_0 代表基頻。對於有聲音框的頻譜，頻率值大於 MVF 的部份，HNMM 視為是噪音，而對於一個無聲的(unvoiced)音框，則將 MVF 之值視為 0，即整個頻譜都是屬於噪音部份；針對噪音部份，HNMM 將其當成是間隔為 100Hz 的諧波信號，各諧波的振幅值，則由少量的倒頻譜參數代表的平滑頻譜上取樣得到，相位值則以亂數產生。

最後，合成的信號 $s(t)$ 就是由諧波 $h(t)$ 和噪音 $n(t)$ 兩部份的信號值作相加而得到，如公式 2 所示：

$$s(t) = h(t) + n(t) \quad (\text{公式 2})$$

HNMM 模型最初係用來作語音的分析與合成，若將之用於歌聲合成上，則會發現其處理的步驟缺失很多，例如：(a)音節時間長度如何作延長和縮短，以維持聲音的自然度；(b)改變音高後的諧波參數值如何去決定，以保持音色的一致性(保持為同一個人的音色)；(c)各個諧波的相位在改變音高

後，如何作同步，以逼近原始錄音的音色特性；(d)合成聲音的頻譜在波谷處均較原始聲音的頻譜低，因為由 HNM 的作法來看，低頻的部份(頻率值小於 MVF 者)是以純粹的諧波表示，高頻的部份(頻率值大於 MVF 者)才使用噪音。由於存在不少的缺失，使得原始的 HNM 模型，不能夠直接拿去作歌聲信號合成的處理。

由此可見，上述習用歌聲合成仍有諸多缺失，實非一良善之設計者，而亟待加以改良。

本案發明人鑑於上述習用歌聲合成所存在的各項缺點，乃亟思加以改良創新，並經多年苦心孤詣潛心研究後，終於成功研發完成本件歌聲合成方法及其架構。

【發明內容】

本發明之目的即在於提供一種歌聲合成方法及其架構，以諧波加噪音模型(HNM)作為基礎，並加以改進和擴充，以合成出更接近自然歌聲的聲音信號。

本發明之次一目的係在於提供一種歌聲合成方法及其架構，除了改進諧波加噪音模型(HNM)之外，更加入歌唱技巧的模擬，以製作出適合具有一字一音發音特色的語言(如：漢語方言、日語、韓語等)所使用之歌聲合成系統。

可達成上述發明目的之歌聲合成方法及其架構，係以諧波加噪音模型(HNM)作為基礎，針對(1)頻譜上諧波與噪音兩部份的分界點(即：最大有聲頻率, MVF)的偵測；(2)基頻的偵測；(3)諧波參數的擷取等加以改善；另外，

在 HNM 模型原先的合成處理步驟之外，並擴增數個用以提升歌聲合成品質的處理方法，包括：(1)相位方面的處理，來讓合成音更能保持原始音的音色特性；(2) ADSR (即：起音 Arrack、衰減 Decay、延持 Sustain、釋音 Release) 式的音長分配，以使伸長或縮短過的音，可以維持自然性；(3)加入低頻噪音，以改進 HNM 在信號低頻部分模型化(modeling)能力的不足；(4)作諧波追蹤，來消除音框間諧波數量的不連續。

1. HNM 分析方法之改進

(1) 最大有聲頻率(MVF)設定

在合成歌聲前需先由人聲錄製各音節的聲音檔，為了避免各音節在錄音時音量的不一致，而使得音量較小的音節受到固定的 MVF 門檻值的限制，導致音量小的音節的諧波數太少，而使合成音失真，因此本發明改以動態的方式來設定 MVF 門檻值。

動態式設定方式係先從一個音節的各個音框裡找出各音框的最大諧波振幅，再找出跨音框的最大值，以該值的 $1/512$ 作為 MVF 的門檻值；然後將 MVF 門檻值和一個音框中各諧波的振幅值作比對，振幅值小於門檻者，視為無聲的頻率，當發生連續 5 個諧波皆為無聲時，最後一個有聲的諧波頻率即為該音框的 MVF，而該 MVF 之前所有的諧波頻率均視為有聲的頻率。

(2) 基頻偵測

為了增加基頻偵測上的準確度，本發明先從音節中延持(Sustain)的部份取出一個音框，並改成使用一種自相關(auto-correlation)函數搭配絕對振幅

差異函數(absolute magnitude difference function, AMDF)的偵測方法，求出該音框的基頻值，作為整個音節中各個音框基頻值的參考。由於每一個音節的信號在分析時，會先經過音節分段的處理，所以可得知延持(Sustain)部份的時間位置。

接著，對於各個音框同樣以自相關函數搭配絕對振幅差異函數(AMDF)的方法，在時域上求出可能的初始基頻值；一個音框若是判斷為有聲，則將該音框信號乘上 Blackman 窗，並於音框後面補上 0，使長度成為 4096 點後，對音框作快速傅利葉轉換(Fast Fourier Transform, FFT)分析得到頻譜；再依時域上求出的初始基頻，在頻譜上該基頻值的附近，搜尋頻譜振幅上的峰值；此峰值點與其前後頻率點的振幅值再作拋物線內插，以求得振幅的最大值，及其相對應的頻率值，亦即基頻。

(3) 諧波參數偵測

求得一個音框的基頻值之後，再依照諧波的倍頻關係，來求得其餘各個諧波所在範圍的局部振幅最大值；為求精確，本發明加入拋物線內插之處理，以求得振幅峰值，及其對應的頻率值與相位值。

此外，顧慮到合成階段進行音高調整時，能夠有較充分的頻譜資訊，以作較準確的頻譜曲線內插，因此，在本發明中，一個音框的高次諧波之偵測，一直偵測到 95%的頻率範圍才停止，並非和原始 HNM 一樣只偵測到最大有聲頻率(MVF)而已。偵測後，將各諧波的頻率、振幅與相位值全都記錄到參數檔中。

2. 提升信號品質的方法

(1) 諧波相位之同步

本發明將每個原始音音框分析得到的相位值都記錄成參數，然後在合成時加以利用，以使合成音保持更多原始音的音色特性。關於相位的處理，包含：

(a) 相位增量控制

為了方便作轉音和抖音等歌唱技巧的模擬，本發明以相位增量的觀念來計算每個信號樣本點的相位值，如公式 3 所示：

$$\phi_k[n] = \phi_k[n-1] + \Delta\phi_k, \Delta\phi_k = \frac{2\pi \cdot f_k}{F_s} \quad (\text{公式 3})$$

當不考慮原始音各音框分析得到的相位值，而依固定的相位增量值，讓相位持續累加下去，則在音框邊界上同一個(第 k 個)諧波的相位自然會連續。但是當考慮原始音各音框分析時得到的相位時，依相位增量累加得到的相位值，就不一定會符合下一個音框的起始相位。為了解決此一問題，本發明將累加的相位值與下一個音框分析時所得的相位值之間的差，平均分配到音框裡的各個樣本點上，以兼顧相位連續性及配合分析時所得的相位值。在取相位的差之前，需先作相位的展開，以使兩者的差在 π 之內，相位展開後的差值 d_k^i 之計算如公式 4 所示：

$$\begin{aligned} M_k^i &= \left\lfloor \frac{1}{2\pi} (\phi_k^i + N \cdot \Delta\phi_k^i - \phi_k^{i+1}) \right\rfloor \\ d_k^i &= \phi_k^{i+1} + 2\pi \cdot M_k^i - \phi_k^i \end{aligned} \quad (\text{公式 4})$$

其中 ϕ_k^i 表示第 i 個音框的第 k 個諧波在分析時得到的相位值， $\Delta\phi_k^i$ 表示第 i 個音框裡第 k 個諧波的相位增量， N 表示音框的長度。

(b) 諧波之間的相位關係

當音高需要調整時，為了維持音色的一致性，新音高的諧波頻率，要

在原始音高的頻譜包絡上內插得到對應的振幅，相位也是如此，但是要在內插前先將相位展開成平滑的相位曲線，以避免在劇烈起伏的曲線上內插出錯誤的相位。

(c) 相對於基頻之相位延遲

在切割音框作分析時，音框的長度不一定剛好是週期長度的整數倍，使得每個音框內，相位測量的基準點會在不同的相對位置上，如圖一所示，音框 $i-1$ 及音框 i 相位測量的基準點 11、12 各離音框 $i-1$ 及音框 i 的左邊界有不同的時間延遲；因此，各諧波分析得到的相位值，必須加上測量基準點的時間延遲所導入的相位延遲，計算公式如公式 5 所示：

$$\phi_{i,k} = \theta_{i,k} + \Delta t_{i,1} \omega_{i,k}, \quad k = 1, 2, \dots, N \quad (\text{公式 5})$$

其中 $\theta_{i,k}$ 表示第 i 音框的第 k 個諧波在音框起始處的相位， $\Delta t_{i,1}$ 表示由基頻所定義的測量基準點的時間延遲， $\omega_{i,k}$ 為第 k 個諧波的角頻率， $\phi_{i,k}$ 表示測量基準點上的基頻同步後的相位。如此讓各諧波之間的相位比較，是在基頻相同的相位位置上進行(即基週同步)，才能維持住諧波之間的真正相位關係。

(2) ADSR 式之音長伸縮

參考電腦音樂裡常用的 ADSR (起音 Arrack、衰減 Decay、延持 Sustain、釋音 Release) 觀念，僅在延持(Sustain)的部份對音長作伸長或縮短，其餘的部份則維持原始長度，以免線性比例調整音長至太短或太長時，合成音聽起來會不自然。

實際使用 ADSR 觀念於人聲時，在信號分析階段，先以人工將一個音

節分成數個區段，接著記錄下這些區段的邊界點上的樣本點時間；在合成階段，根據分析時記錄下的區段邊界樣本點時間，分別施以不同的音長處理方式；以尸丫音為例，如圖二所示，首先界定聲音中無聲部份和有聲部份的邊界，區塊 2 2 即聲音中的有聲部份；再標示出聲音中起音(attack)部份和釋音(release)部份的邊界，區塊 0 20 即為起音部份，區塊 1 21 即為釋音部份。

(3) 加入低頻噪音(Convolution noise)

使用原始的 HNM 方法合成出的聲音，若在頻譜上與原始音作比較，會發現在低頻的部份，合成音的頻譜波谷比原始音的波谷深；以尸丫音為例，如圖三所示，圖三 B 為原始音 31 及合成音 32 之時域的波形；圖三 A 為原始音 31 及合成音 32 裡相同時間位置的音框之頻譜圖，如圖三 A 所示，在低頻的部份，合成音 32 的頻譜波谷比原始音 31 的波谷深，因為在低頻的部份(頻率值小於 MVF)，HNM 是以純粹的諧波表示。為了加強 HNM 在低頻部份模型化(modeling)能力的不足，本發明加入低頻噪音(Convolution Noise, CVNS)，使合成音低頻部份之頻譜，可以較近似於原始聲音低頻部份之頻譜。加入低頻噪音(CVNS)的步驟包含：

(a) Convolution noise (CVNS)分析

先以 HNM 方法依原始聲音的音高、音長，產生出合成的聲音；接著依據 HNM 分析結果得知的有聲部份的位置資訊，分別對原始音、合成音兩個聲音有聲的部份取音框作 FFT 分析，於一較佳實施例中，所取的音框大小為 512 點且重疊一半；分析出兩者的譜頻後再分別作倒頻譜分析，得到倒

頻譜參數，最後將合成聲音的倒頻譜參數減去原始聲音的倒頻譜參數，得到 CVNS 參數。

(b) Convolution noise (CVNS)合成

合成時如果將 CVNS 直接加在合成音對應音框的倒頻譜參數上，會得到低頻過強的吵雜聲音，因此本發明先將 CVNS 作兩項處理：(i)去除諧波影響，將近似基週長度和其倍數位置附近的 CVNS 參數值改設為 0；(ii)設定門檻值，以原先音框信號的各個倒頻譜參數的 0.1 倍作為變動的最大限值。之後，將 CVNS 加至合成音對應音框的倒頻譜參數上，再反轉換回時域，即可得到含有低頻噪音的合成音信號。

(4) 諧波追蹤

在一序列的分析音框裡，最大有聲頻率(MVF)的值突然起伏太大時，可能造成某個音框的諧波數較前後音框多或者是少的情形發生，當這種諧波數不連續太明顯時，在聽覺上會有不自然的感覺。因此本發明進行諧波追蹤，以消除這些短時間的諧波數量之不一致。

當有聲部份所有音框的 HNM 分析完成後，再對這些音框作諧波追蹤，讓相鄰音框的諧波數不會改變得太劇烈。當前後兩個音框中的諧波數不一致時，在接下來的 50ms 中，若找得到可以相接的諧波頻率，則將這 50ms 裡的各個音框，加上此諧波，否則就將該諧波刪除。在作諧波追蹤時，將 MVF 之前的諧波頻率都視為有聲，再者因為音框間的基頻值並不會變化太大，所以大致上相鄰音框間的諧波都會一對一的相連，只有在靠近 MVF 的諧波頻率，可能因為 MVF 差異太大而產生不相連的情況，因此此處之諧波

追蹤，可以看成是對 MVF 的變化加以平滑化。

【實施方式】

以下之實施例係以本發明之歌聲合成方法及其架構所進行之中文歌聲合成。

如圖四之架構圖所示，在進行國語歌聲合成前，先以人聲在錄音室中唸出中文 409 個音節的第一聲發音，然後轉為電腦音檔儲存在檔案儲存裝置 401 中，音檔的取樣率為 22050 Hz，解析度為 16 bits/sample，且一種音節，只需錄一個對應的音檔，一音檔中含有該對應音節的波形；再將該原始音檔依本發明之歌聲合成方法及其架構進行音節分析及歌聲合成。

一、音節分析

首先自檔案儲存裝置 401 中將音檔中之原始音節 4011 的波形讀取出來，由於考慮 ADSR 式之音長伸縮作法，所以先由音節分段模組 402 對各音節作音節分段標記的處理，針對該 409 個音節中有聲和無聲的部份作音節分段標記，將該音節分段成為有聲部份的音框 4021 及無聲部份的音框 4022，再針對該有聲部分的音框 4021 中起音及釋音的部份作音節分段標記，每個音節共有 3 組即 6 個邊界點，記錄在 WAVE 格式檔案的檔頭之內。

然後以本發明之諧波加噪音模型(HNM)分析模組 403 進行分析，於本實施例中，對於標記為有聲部分的音框 4021，設定該音框長度為 512 點，且音相鄰框重疊 1/2，有聲部分的音框 4021 的參數包括諧波參數 4031 及噪音參數 4032 兩部份；對於標記為無聲部份的音框 4022，則只需記錄噪音參數 4033。噪音參數 4032、4033 的分析方式是，先作 FFT 轉換，對得到的

頻譜再作倒頻譜轉換，取出 30 階的倒頻譜係數作為噪音參數。接著將音節各音框的諧波參數 4031 及噪音參數 4032、4033 經由諧波加噪音模型(HNM)合成模組 404 合成與原始音節相同音高、音長的合成音節 4041；另外，亦將各音框的諧波參數 4031 及噪音參數 4032、4033 儲存在檔案儲存裝置 401 中。

另外，對於較短的無聲子音(如/b, d, g/)，或是由有聲子音、母音開始的音節(如/na, a/)，由於這類音節在信號起始處通常有較不規律的波形變化，而習用的 HNM 模型係以音框為分析單位，這些短暫的波形變化將無法被準確分析出來，因此本發明將這類信號樣本值直接記錄下來，合成時再將這些信號樣本選擇適當的位置，與合成音相接。

接著，將原始音節 4011 與合成音節 4041 以低頻噪音(Convolution noise)分析模組 405 分析出低頻噪音參數 4051，再將該低頻噪音參數 4051 記錄至檔案儲存裝置 401 中的參數檔案內。

低頻噪音分析模組 405 係分別對原始音節 4011 及合成音節 4041 兩個音節有聲的部份取音框作 FFT 分析，於本實施例中，所取的音框大小為 512 點且重疊一半；分析出兩者的譜頻後，再分別作倒頻譜分析以得到倒頻譜參數，為了能準確還原頻譜，本發明保留全部的倒頻譜參數；再將合成音節 4041 的倒頻譜參數減去原始音節 4011 的倒頻譜參數，即可得到低頻噪音參數 4051，之後記錄下每個音框的低頻噪音參數 4051，以供合成時使用。

二、音節合成

如圖五之架構圖所示，合成歌聲前需要先將儲存在檔案儲存裝置 401

中的歌詞檔 4012 以歌詞檔分析模組 406 分析；歌詞檔的範例如表一所示，包含下列幾個部份：曲名、每分鐘拍數以及歌詞序列，其中拍數可以控制整首歌的演唱速度。歌詞又可分為幾個部份：(1)序號，依歌詞數逐漸累加；(2)歌詞字，依據歌詞字到字典檔中找出對應的拼音及參數檔，破音字時，可在歌詞字後放相同發音之另一字，以作為辨別；(3)音高，以音符為單位來記錄，當一個歌詞字為「|」時，視為轉音，即此時的音符音高要用前一音符的歌詞來唱，若音高值為 0 時，則表示為休止符；(4)拍數，根據演唱速度，可以換算出一個音符要唱多久；(5)強弱，用來控制各音符的音量大小。

表一 歌詞檔範例

曲名：青春舞曲			拍數：180 拍/分		
歌詞序列	序號	歌詞字	音高	拍數	強弱
	01	太	B3	1	1
	02	陽	A3	1	1
	03	下	F3#	1	1
	04	山	G3	1	1
	05	明	B3	1	1
	06	朝找	A3	1	1

經過歌詞檔分析模組 406 分析後，取得每個歌詞字對應的音節代號 4061、音高 4062 及音長 4063 資料，再由音節參數取出模組 407 依音節代號 4061 取得對應音節的諧波參數 4031 與噪音參數 4032、4033，然後再由音節參數調整模組 408 依據音高 4062 及音長 4063 資料對音節的各項參數作調整。

為了避免拖拍的問題，音節中有聲的部份需在拍子的起始處就被聽見，因此要對音節的演唱起始時間作調整；由於本發明已對各音節標示有聲與無聲處的邊界，故只需將該有聲與無聲處的邊界移至拍子起始點即可。

為了模擬真人唱歌時換氣的動作，本發明將歌詞的音長保留一部份作為靜音：音符音長大於 1.3 秒時，保留音長的 25%；音長小於 1.3 秒時，保留音長的 17%作為靜音。

在音量方面，由於人類在說話或唱歌的時候，嘴型大小與音量的大小是成正比的，所以音量要根據音節中韻母的不同來調整。

調整後的各項參數再由波形合成模組 409 進行合成音節波形 4091 的合成；一個音符的音高由歌詞檔讀出後，就可決定出新的諧波頻率值，為了維持音色的一致性，對於新的諧波頻率的振幅、相位數值，要在原始的頻譜包絡上，找出新諧波頻率前後相鄰的四個原始諧波的參數，作拉格蘭日內插來得到。

時間長度分配上，依 ADSR 觀念設計音長伸縮規則如下：(1)子音的長度伸縮的倍率限制在 0.6~1.2 倍之間；(2)起音和釋音的長度保持和原音節相同；(3)限制尾音的長度要小於延持部份長度的 1/4；如此，可避免延持 (Sustain) 以外的部份變化過度而不自然。

實際合成時，在歌聲音節的音長範圍內，均勻佈放間隔 200 個樣本點的控制點，控制點上的參數依時間比例由原始音節中的音框參數作拉格蘭日內插得到，控制點之間的信號樣本點上的參數，則由兩邊控制點的參數作線性內插得到，以加快信號合成的速度。

再由低頻噪音處理模組 410 自檔案儲存裝置 401 取出低頻噪音參數 4051，以對該合成音音節波形 4091 作低頻噪音參數 4051 的處理，以產生含有適度低頻噪音的合成音 4101。

該低頻噪音處理模組 410 會將該合成音音節波形 4091 有聲的部份，取重疊 256 點、長度為 512 點的音框作頻譜分析，然後再作倒頻譜分析，以得到該合成音音節波形 4091 的倒頻譜參數；然後，依目前合成音音框在合成音中時間位置的比例，取相對於原始音中相同時間比例的前後兩個音框的低頻噪音參數作線性內插，如公式 6 所示：

$$i = \frac{\text{合成音框起始時間}}{\text{合成音長}} \cdot \frac{\text{原始音長}}{\text{hopsize}}, \text{hopsize} = 512 - 16$$

$$CVNS_i(n) = CVNS_{\lfloor i \rfloor}(n) + \frac{CVNS_{\lfloor i \rfloor + 1}(n) - CVNS_{\lfloor i \rfloor}(n)}{\text{hopsize}} \cdot (i - \lfloor i \rfloor) \quad (\text{公式 6})$$

得到對應的低頻噪音參數，再將該低頻噪音參數及合成音的倒頻譜參數依最大限值之條件作相加，如公式 7 所示：

$$\hat{c}_i(n) = \begin{cases} c_i(n) + CVNS_i(n), & \text{if } |CVNS_i(n)| < 0.1 \cdot |c_i(n)| \\ (1 + 0.1 \cdot \text{sign}(CVNS_i(n))) \cdot c_i(n), & \text{otherwise} \end{cases}$$

$$1 \leq n \leq \text{cepstrum order} \quad (\text{公式 7})$$

如此得到修正後的合成音倒頻譜，最後將該倒頻譜逆轉換至時域，即倒頻譜與快速傅利葉轉換之反轉換，如此即可得到含有適度低頻噪音的合成音 4101。

若音節的音長太短，可能發生延持(Sustain, ST)部份的長度過短，使得起音(Attack, AT)和釋音(Release, RL)兩部份幾乎是直接相連，如果起音的結束音量和釋音的起始音量相差太多時，聽起來就會有振幅不連續之喀嚓聲

(click)。因此當合成音的音長過短時，針對起音和釋音兩部份分別偵測各自的振幅最大值，然後將起音的音量調整成和釋音相近。

另外，由於在諧波加噪音模型(HNM)中作轉音模擬時，會因為轉音前後音框的最大有聲頻率(MVF)不變，但因前後音符的基頻不同，而發生較低音高的音框其諧波數量會比較多，為了不使多出來的諧波對聽覺造成影響，因此本發明將諧波數較多的一方在轉音區中向另一方延伸，直到轉音區的邊界，再將其振幅逐漸降為 0，經轉音處理之頻譜圖形如圖六所示。

三、聽測實驗

國語歌聲合成完成後，接著進行主觀的聽測實驗，以評估合成出的歌聲信號之品質。

聽測實驗係以相位同步和 ADSR 音長伸縮作為聽測的因素，因為 ADSR 音長伸縮對自然度具有明顯的影響。實驗分成五種合成方式來互相作比較，分別是：(1)相位未與原始音同步，只作累加；不使用 ADSR，而依線性比例調整音長。(2)相位未與原始音同步，只作累加；使用 ADSR，在延持部份調整音長。(3)相位與原始音同步；不使用 ADSR，而依線性比例調整音長。(4)相位與原始音同步；且使用 ADSR，在延持部份調整音長(即本發明之方法)。(5)使用弦波模型(即不考慮噪音成分)合成歌聲，且最大有聲頻率(MVF)不管任何音節都固定成定值(6,000Hz)，沒有作 ADSR 音長伸縮、相位同步及低頻噪音等處理。

聽測的曲目共有兩首，分別是節奏較快的「青春舞曲」和節奏較慢的「康定情歌」，評估的項目是自然度與清晰度兩項。自然度在於評估合成歌

聲與人類歌聲的接近程度；清晰度在於評估合成出來的歌聲訊號聽起來是否清楚無雜訊，以及咬字的清晰程度。評分的範圍由最高 5 分到最低 1 分，可以打至小數點下一位。

評估的作法是由聽測者分別聽兩首歌曲，每首歌曲都有五種方式合成的音檔，其中，第二種方式(即無相位同步但有作 ADSR 式音長縮放)的合成歌聲，當作自然度與清晰度皆為 3 分的參考音。另外四種方式合成出的歌聲，則以隨機的順序排列讓聽測者試聽，聽測者可以自由選取歌曲中的一段範圍反覆試聽，並與參考音作比較，以比較後的優劣程度來作評分。

試聽者一共有 15 位，評分的平均值如表一所示。在自然度方面，因為「青春舞曲」的節奏較快，所以有、無考慮 ADSR 的方式差異不大，相差約 0.1 到 0.2 分，而有作相位同步的方式，其自然度整體上都較不作相位同步的自然度來得高。在清晰度方面，因為基於相同的合成模型，所以有、無作 ADSR 的影響不大，但有作相位調整者較清晰，可能是受自然度的影響，其中有作 ADSR 式音長分配及相位同步的得分最高，與第五種方式比較，相差約 1 分。

由於「康定情歌」的節奏較慢，因此有作 ADSR 的效果較為顯著，尤其是在音長較長或有轉音的音節，故在自然度上有、無考慮 ADSR 的方式，差異增加至 0.3 到 0.4 分，有作相位同步的方式，其自然度整體上仍然較不作相位同步的自然度來得高。在清晰度方面，表現也與「青春舞曲」的情況類似，其中有作 ADSR 式音長分配及相位同步的仍然是得分最高，與第五種方式比較，同樣相差約 1 分。

表一 聽測實驗之結果

合成方式	合成選項		青春舞曲		康定情歌	
	ADSR	相位	自然度	清晰度	自然度	清晰度
1	X	X	2.9	3.1	2.7	3.0
2	<u>V</u>	<u>X</u>	<u>3.0</u>	<u>3.0</u>	<u>3.0</u>	<u>3.0</u>
3	X	V	3.4	3.4	3.2	3.2
4	V	V	3.6	3.5	3.6	3.4
5	--	--	2.7	2.5	2.6	2.2

本發明所提供之歌聲合成方法及其架構，與前述習用技術相互比較時，更具有下列之優點：

1. 本發明所提供之歌聲合成方法經聽測實驗與其他歌聲合成方法比較後，證實以本方法合成之歌聲在自然度及清晰度上面，均較其他方法有明顯的改進。

2. 本發明之歌聲合成方法除了改進習用諧波加噪音模型(HNM)之缺點之外，並加數項提升歌聲品質之處理方法，所製作出之歌聲合成系統更適合具有一字一音發音特色的語言(如：漢語方言、日語、韓語等)所使用。

上列詳細說明係針對本發明之一可行實施例之具體說明，惟該實施例並非用以限制本發明之專利範圍，凡未脫離本發明技藝精神所為之等效實施或變更，均應包含於本案之專利範圍中。

綜上所述，本案不但在方法上確屬創新，並能較習用歌聲合成方法增進上述多項功效，應已充分符合新穎性及進步性之法定發明專利要件，爰依法提出申請，懇請 貴局核准本件發明專利申請案，以勵發明，至感德便。

【圖式簡單說明】

圖一為本發明歌聲合成方法及其架構之相位延遲示意圖；

圖二為該歌聲合成方法及其架構之音節分段示意圖；

圖三為該歌聲合成方法及其架構之原始音與合成音中對應音框之頻譜比較圖；

圖四為該歌聲合成方法及其架構之音節分析架構圖；

圖五為該歌聲合成方法及其架構之音節合成架構圖；以及

圖六為該歌聲合成方法及其架構之轉音區諧波數量調整波形圖。

【主要元件符號說明】

- 11 基準點
- 12 基準點
- 2 區塊 2
- 20 區塊 0
- 21 區塊 1
- 31 原始音
- 32 合成音
- 401 檔案儲存裝置
- 4011 原始音節
- 4012 歌詞檔
- 402 音節分段模組
- 4021 有聲部份的音框
- 4022 無聲部份的音框
- 403 諧波加噪音模型分析模組
- 4031 諧波參數
- 4032 噪音參數
- 4033 噪音參數

- 404 諧波加噪音模型合成模組
- 4041 合成音節
- 405 低頻噪音分析模組
- 4051 低頻噪音參數
- 406 歌詞檔分析模組
- 4061 音節代號
- 4062 音高
- 4063 音長
- 407 音節參數取出模組
- 408 音節參數調整模組
- 409 波形合成模組
- 4091 合成音音節波形
- 410 低頻噪音處理模組
- 4101 含有適度低頻噪音的合成音

五、中文發明摘要：

一種歌聲合成方法及其架構，在參數分析階段，係自一檔案儲存裝置中讀取原始音節波形，以一音節分段模組對該音節作音節分段標記，再以一諧波加噪音模型分析模組進行分析，得到該音節中各音框的諧波及噪音參數，再以一低頻噪音分析模組分析出低頻噪音參數；在歌聲合成階段，先將歌詞檔以一歌詞檔分析模組分析，取得每個歌詞字對應的音節代號、音高及音長資料，並由一音節參數取出模組依音節代號取得對應音節的諧波及噪音參數，將資料交給一音節參數調整模組對音節的各項參數作調整，再由一波形合成模組進行歌聲音節波形的合成後，以一低頻噪音處理模組作低頻噪音的加入處理，以產生含有適度低頻噪音的歌聲信號。

六、英文發明摘要：

十、申請專利範圍：

1. 一種歌聲合成方法，包括以下步驟：

步驟一 自一檔案儲存裝置中讀取一錄音檔之原始音節的波形，以一音節分段模組對音節作音節分段標記處理，先針對一音節中有聲和無聲的部份作音節分段標記，將該音節分段成為有聲部份的音框及無聲部份的音框，再針對該有聲部分的音框中起音及釋音的部份作音節分段標記；

步驟二 將該分段標記的音節以一諧波加噪音模型分析模組進行分析，以得到該音節中各音框的諧波參數及噪音參數；

步驟三 將該音節中各音框的諧波參數及噪音參數經由一諧波加噪音模型合成模組合成與該原始音節相同音高、音長的合成音節；

步驟四 自該檔案儲存裝置中讀取歌詞檔，並以一歌詞檔分析模組分析，取得每個歌詞字對應的音節代號、音高及音長資料；

步驟五 由一音節參數取出模組依音節代號取得對應音節的諧波參數與噪音參數，並將資料交給一音節參數調整模組；

步驟六 再由該音節參數調整模組依據該諧波參數、噪音參數、音高及音長資料對音節的各項參數作調整；

步驟七 調整後的各項參數再由一波形合成模組進行合成音節波形的合成。

2. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟三中，對於較短的無聲子音，或是由有聲子音、母音開始的音節，直接記錄其信號樣

本值，再於步驟八合成時將該信號樣本選擇適當的位置，與合成音相接。

3. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟三中，可進一步將該原始音節與該合成音節以一低頻噪音分析模組分析出低頻噪音參數，並將該低頻噪音參數記錄至該檔案儲存裝置中的參數檔案內。
4. 如申請專利範圍第 3 項所述之歌聲合成方法，其中該低頻噪音分析模組係分別對該原始音節及該合成音節中有聲的部份取音框作快速傅利葉轉換分析，分析出兩者的譜頻後，分別作倒頻譜分析以得到倒頻譜參數，再將該合成音節的倒頻譜參數減去該原始音節的倒頻譜參數，即可得到該低頻噪音參數。
5. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟六中，將音節有聲與無聲處的邊界移至該拍子之起始點，以使音節中有聲的部份在拍子的起始處就被聽見，以避免拖拍。
6. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟六中，將歌詞的音長保留一部份作為靜音，以模擬真人唱歌時換氣的動作。
7. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟六中，根據音節中韻母的不同來調整各音節的音量。
8. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟七中，在原始的頻譜包絡上，找出新諧波頻率前後相鄰的四個原始諧波的參數，作拉格蘭日內插，得到新的諧波頻率之振幅及相位數值，以維持音色的一致性。
9. 如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟七中，進行相

位增量控制、諧波之間相位關係、以及相對於基頻之相位延遲等相位的處理，以提升該合成音之音色逼真度。

10.如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟七中，依起音、衰減、延持、釋音(ADSR)觀念設計音長伸縮規則。

11.如申請專利範圍第 10 項所述之歌聲合成方法，其中該音長伸縮規則包含：

(1) 子音的長度伸縮的倍率限制在 0.6~1.2 倍之間；

(2) 起音和釋音的長度保持和原音節相同；

(3) 限制尾音的長度要小於延持部份長度的 1/4。

12.如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟七中，對於音長決定後的音節信號範圍，均勻佈放間隔 200 個信號樣本點的控制點，控制點上的參數依時間比例由原始音節中的音框參數作拉格蘭日內插得到，控制點之間的樣本點上的參數，則由兩邊控制點的參數作線性內插得到。

13.如申請專利範圍第 1 項所述之歌聲合成方法，其中該步驟七中，可進一步由一低頻噪音處理模組自該檔案儲存裝置取出低頻噪音參數，以對合成音音節波形作低頻噪音參數的處理，以產生含有適度低頻噪音的合成音。

14.如申請專利範圍第 13 項所述之歌聲合成方法，其中該低頻噪音處理模組會將該合成音音節波形有聲的部份，取音框作頻譜分析，然後作倒頻譜分析，以得到該合成音音節波形的倒頻譜參數，再依目前合成音音框在

合成音中時間位置的比例，取相對於原始音中相同時間比例的前後兩個音框的低頻噪音參數作線性內插，以得到對應的低頻噪音參數，再將該低頻噪音及合成音兩種倒頻譜參數依最大限值之條件作相加，以得到修正後的合成音倒頻譜，最後將該倒頻譜逆轉換至時域，即倒頻譜與快速傅利葉轉換之反轉換，以得到含有適度低頻噪音的合成音。

15.如申請專利範圍第1項所述之歌聲合成方法，其中該步驟七中，其中該合成音的音長若過短時，針對起音和釋音兩部份分別偵測各自的振幅最大值，然後將該起音的音量調整成和釋音相近。

16.如申請專利範圍第1項所述之歌聲合成方法，其中該步驟七中，該合成之歌聲中若有轉音的部分，需將轉音前後音框中，諧波數較多的一方在轉音區中向另一方延伸，直到轉音區的邊界，並且將其振幅逐漸降為0。

17.一種歌聲合成之架構，包括：

一檔案儲存裝置，係儲存各種檔案及參數；

一音節分段模組，係自該檔案儲存裝置中讀取一錄音檔之原始音節的波形，並對該原始音節作音節分段標記處理，先針對一音節中有聲和無聲的部份作音節分段標記，將該音節分段成為有聲部分的音框及無聲部分的音框，再針對該有聲部分的音框中起音及釋音的部份作音節分段標記；

一諧波加噪音模型分析模組，係對該分段標記的音節作分析，以得到一音節中各音框的諧波參數及噪音參數；

一諧波加噪音模型合成模組，係依據該音節中各音框的諧波參數及噪音

參數，合成與該原始音節相同音高、音長的合成音節；

一低頻噪音分析模組，係依據該原始音節與該合成音節中有聲部份的音

框作分析，以得到一低頻噪音參數；

一歌詞檔分析模組，係自該檔案儲存裝置中讀取一歌詞檔，並分析該歌

詞檔，以取得每個歌詞字對應的音節代號、音高及音長等資料；

一音節參數取出模組，係依據該音節代號取得對應音節的諧波參數與噪

音參數，並將資料交給一音節參數調整模組；

一音節參數調整模組，係依據該諧波參數、噪音參數、音高及音長資料

對音節的各項參數作調整；

一波形合成模組，係依據該音節參數調整模組調整後的各項參數，進行

合成音節波形的合成；

一低頻噪音處理模組，係自該檔案儲存裝置取出對應的低頻噪音參數，

以對該合成音節波形作低頻噪音參數的處理，而產生含有適度低頻

噪音的合成音。

18.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該諧波加噪音模型

合成模組對於較短的無聲子音，或是有聲子音、母音開始的音節，會直

接記錄其信號樣本值，再於合成時將該信號樣本選擇適當的位置，與合

成音相接。

19.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該低頻噪音分析模

組係分別對該原始音節及該合成音節中有聲的部份取音框作快速傅利葉

轉換分析，分析出兩者的譜頻後，分別作倒頻譜分析以得到倒頻譜參數，

再將該合成音節的倒頻譜參數減去該原始音節的倒頻譜參數，以得到該低頻噪音參數。

20.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該音節參數調整模組會將一音節有聲與無聲處的邊界移至該拍子之起始點，以使該音節中有聲的部份在拍子的起始處就被聽見，以避免拖拍。

21.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該音節參數調整模組會將歌詞的音長保留一部份作為靜音，以模擬真人唱歌時換氣的動作。

22.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組會在原始的頻譜包絡上，找出新諧波頻率前後相鄰的四個原始諧波的參數，作拉格蘭日內插，得到新的諧波頻率之振幅及相位數值，以維持音色的一致性。

23.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組會進行相位增量控制、諧波之間相位關係、以及相對於基頻之相位延遲等相位的處理，以提升該合成音之音色逼真度。

24.如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組會依起音、衰減、延持、釋音(ADSR)觀念設計音長伸縮規則。

25.如申請專利範圍第 24 項所述之歌聲合成之架構，其中該音長伸縮規則包含：

(1) 子音的長度伸縮的倍率限制在 0.6~1.2 倍之間；

(2) 起音和釋音的長度保持和原音節相同；

(3) 限制尾音的長度要小於延持部份長度的 1/4。

26. 如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組對於音長決定後的音節信號範圍，均勻佈放間隔 200 個信號樣本點的控制點，控制點上的參數依時間比例由原始音節中的音框參數作拉格蘭日內插得到，控制點之間的樣本點上的參數，則由兩邊控制點的參數作線性內插得到。
27. 如申請專利範圍第 17 項所述之歌聲合成之架構，其中該低頻噪音處理模組會將該合成音音節波形有聲的部份，取音框作頻譜分析，然後作倒頻譜分析，以得到該合成音音節波形的倒頻譜參數，再依目前合成音音框在合成音中時間位置的比例，取相對於原始音中相同時間比例的前後兩個音框的低頻噪音參數作線性內插，以得到對應的低頻噪音參數，再將該低頻噪音及合成音兩種倒頻譜參數依最大限值之條件作相加，以得到修正後的合成音倒頻譜，最後將該倒頻譜逆轉換至時域，即倒頻譜與快速傅利葉轉換之反轉換，以得到含有適度低頻噪音的合成音。
28. 如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組在該合成音的音長過短時，會針對起音和釋音兩部份分別偵測各自的振幅最大值，然後將該起音的音量調整成和釋音相近。
29. 如申請專利範圍第 17 項所述之歌聲合成之架構，其中該波形合成模組在該合成音具有轉音屬性時，會將轉音前後音框中，諧波數較多的一方在轉音區中向另一方延伸，直到轉音區的邊界，再將其振幅逐漸降為 0。

十一、圖式：

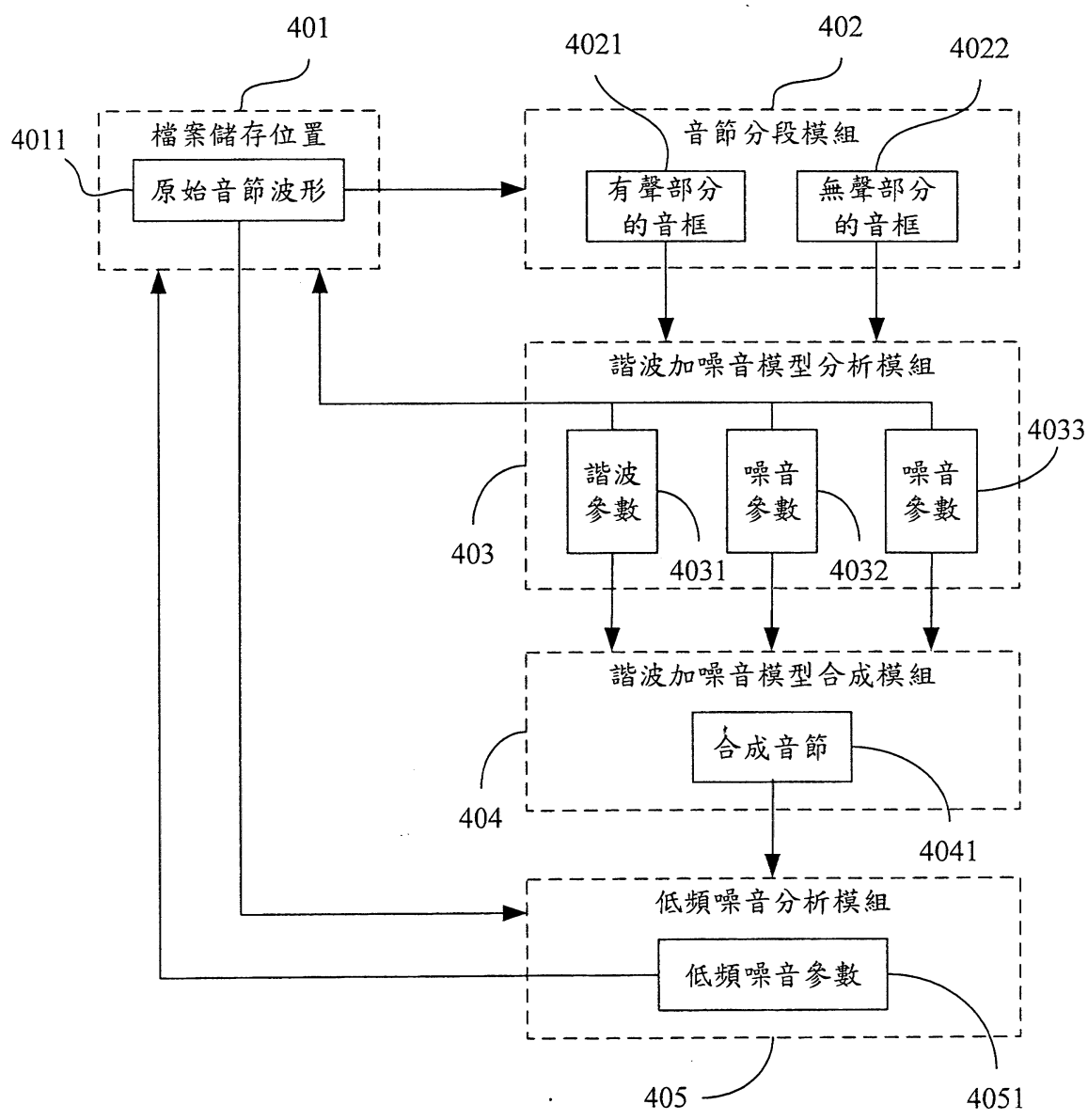


圖 四

十一、圖式：

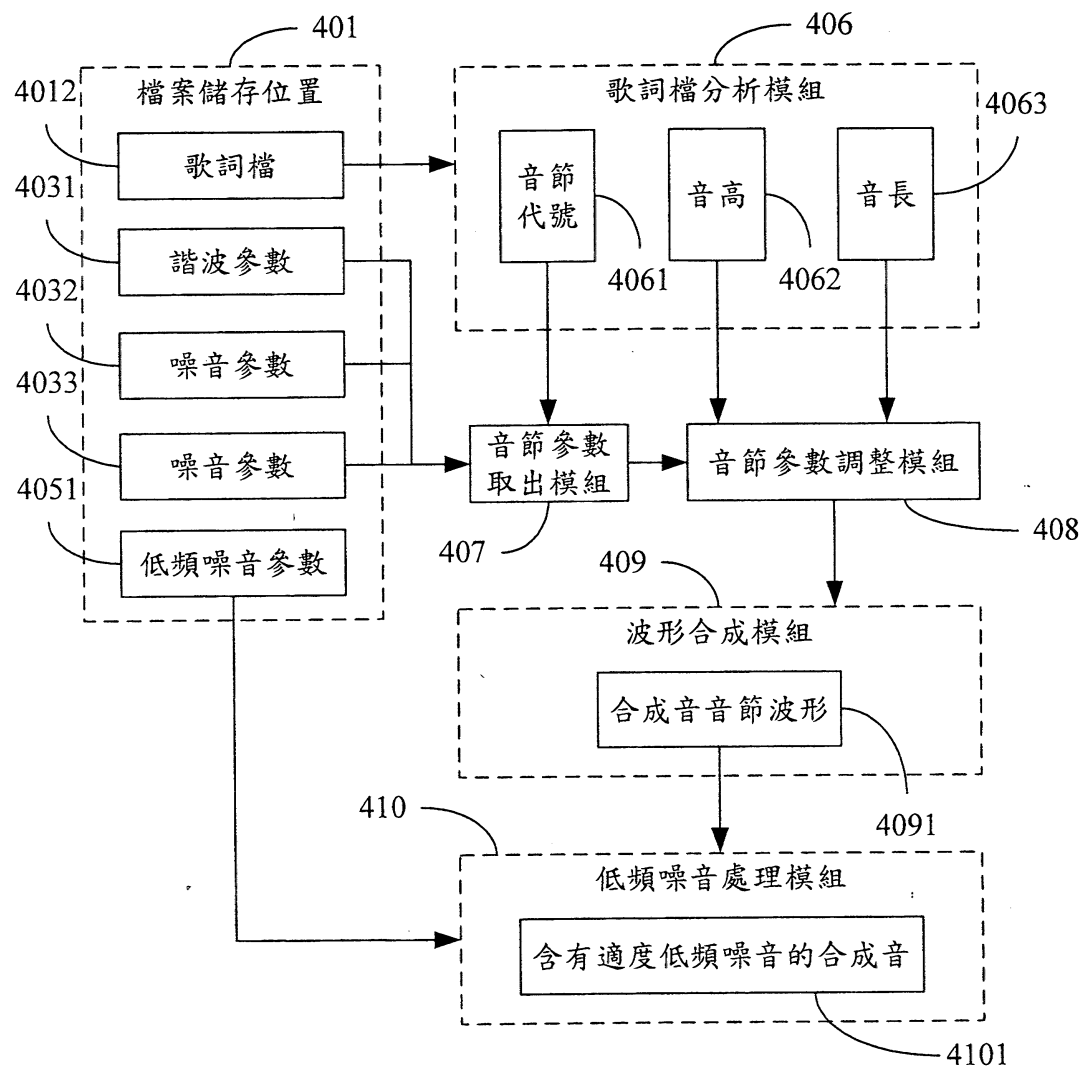


圖 五

七、指定代表圖：

(一)本案指定代表圖為：第(五)圖。

(二)本代表圖之元件符號簡單說明：

- 401 檔案儲存裝置
- 4012 歌詞檔
- 4031 諧波參數
- 4032 噪音參數
- 4033 噪音參數
- 4051 低頻噪音參數
- 406 歌詞檔分析模組
- 4061 音節代號
- 4062 音高
- 4063 音長
- 407 音節參數取出模組
- 408 音節參數調整模組
- 409 波形合成模組
- 4091 合成音音節波形
- 410 低頻噪音處理模組
- 4101 含有適度低頻噪音的合成音

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：