

五、發明說明()

本發明一般涉及語音識別系統，尤其涉及用於移動通信設備的語音識別系統。

傳統上已經通過人工操作的鍵盤實現了從人類向機器的信息傳送，其預先假設了具有至少與人類的兩隻手的手指所能舒適自如地跨過的寬度一樣大的尺寸的機器。隨著需求信息輸入的但比傳統的個人機算機更小的電子設備的發展，信息輸入開始採用其它形式，如筆指示 (pen pointing)、觸摸板 (touchpads) 和聲音命令。能夠被筆指示和觸摸板傳送的信息受到設備 (如個人數字助理 (PDAs) 和移動電話) 的顯示性能的限制。因此重點研究用於電子設備的語音識別系統。對於機器而言，機器語音識別方法是基於信號的觀察到的聲學特徵和聲學特徵與話音之間的已知的關係而試圖來解碼語音信號波形的。這種聲學一語音方法是幾乎50年來的研究主題，但是還沒有在實踐上取得大的成功 (參考：Fundamentals of Speech Recognition (語音識別基礎)，L.Rabiner&B.Hjuang,Prentice-Hall)。存在的大量問是，例如，在語音識別技術中公知甚至在語音波形圖中“通常難以區分微弱的清音 (像“f”或“th”) 與默聲，或者微弱的暗 (像“v”或“m”) 與清音或者甚至是默聲”，並且依據非常鄰近的語音單元的同一性，所謂的聲音的共同清晰度 (同前所述) 有大量的聲音的變體存在。在解碼後，試圖通過應用代表

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

經濟部智慧財產局員工消費合作社印製

五、發明說明()

可能與口頭輸入匹配的一系列組的音素的所謂的音素網格來確定聲學—語音方法中的字。網格中的音素的垂直位置是聲學與語音單元的匹配品質的量度(“詞典的訪問”)。但是“對於語音識別的聲學—語音方法的真正問題是關於對於詞典的訪問階段難以得到可靠的音素網格”(同前所述);即,由於任何語音中固有的大量的變體幾乎不可能準確地標示出發音。

在模式識別方法中,給出的語音模式的版本的知識庫被集中起來(“訓練”),並且通過輸入語音模式與知識庫中的語音模式相比來確定最佳的匹配而實現識別。該模式有4個步驟:(1)應用頻譜分析的特徵抽取,(2)模式訓練以產生對於發音類別的參考模式,(3)模式分類來通過測量兩個充分限定的頻譜矢量之間的頻譜“差距”並校準時間以補償說出兩種模式(動態時間偏差DTW)的速率不同來比較未知的測試模式與類別參考模式,及(4)確定邏輯從而使用相似性得分來選擇最佳的匹配。尤其對於步驟(2)和(3)模式識別需要大量的計算,並且用於大量聲音類別的模式識別通常是禁止的。

因此,由於語音的固有的難以預測的變化(包括同音字、字相似性、口音、音級、重音節、語音模式、背景噪音等),依靠人類話音的信息輸入的系統要求相當可觀的信號處理能力和查找表數據庫,以達到最低限度

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 ()

的準確度。主計算機和高級工作站開始接近聲音識別的可接受的水平，但是即便帶有在當前的個人計算機（PC）中可利用的存儲器和計算能力，對於這些機器而言語音識別目前被相當地限制於給出的特定語音命令組。對於帶有比PC小得多的存儲器和處理能力的設備，如PDA、移動電話、玩具、娛樂設備，自然語音的準確識別目前為止是不可能的。例如，一般的聲音起動蜂窩電話通過讀出一個名字然後輸入相關號碼而允許預先設計程序。當用戶隨後讀出這個名字時，蜂窩電話中的微處理器將試圖匹配讀出的名字的聲音模式和儲存的號碼。如已經使用當今的聲音起動的蜂窩電話的人所知道的那樣，有時匹配不準確（由於不一致的發音、背景噪音和缺少處理能力而引發的固有的限制），並且可能僅存儲大約25個號碼。在PDA設備中，對於設備製造者而言，需要進行廣泛的再次設計以實現更限定的聲音識別（例如，當前的PDA不能響應於聲音輸入來搜索數據庫）。

對於聲音輸入拼寫單詞，存在易混淆地設置的問題：{A,J,K},{B,C,D,E,G,P,T,V,Z},{Q,U},{I,Y}和{F,S,X}。通常僅基於發音的微小的關鍵部分而把這些區分出來。由於傳統的識別依賴於在整個發音期間的簡單累積的失真得分（二進制的“是”或“否”），它沒有把重點放在導致差的識別準確度的關鍵部分。顯

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

五、發明說明()

然，應有一種對該關鍵部分進行加權的方法，但是這種方法沒有得到高的識別準確度並且還承擔有大量的計算負擔。

總之，準確快速的聲音識別必須的存儲器和計算還要求增加的耗電量和的複雜操作系統；所有這些都會增加成本。從而，當前的聲音技術由於其重量、耗電量要求、複雜性和成本而對於移動通信設備並不是切實可行的。

因此需要一種準確的語音識別系統，其能夠快速處理更大量的字變化並可在多種不同設備中操作，而沒有計算能力和存儲器要求，也不需要高的耗電量、複雜的操作系統、高的成本和傳統系統的重量，從而聲音信息傳送對於移動電話、PDA、電子玩具、娛樂產品和任何其它的要求通信的設備所用的人對人和人對機器的通信都是切實可行的。本發明向低處理能力的並帶有有限的存儲器存儲能力的電子設備提供準確的語音識別。準確度主要是通過應用獨立的字母數字字符發音來傳送字來實現的，從而克服了詞典訪問的問題。由於獨立的字母數字發音產生在時間區上變化非常小的語音波形並且被一個中頓分開（口述的獨立字符或字母之間的一般的中頓），本發明解決了區分微弱的清音聲音和與共同清晰度相關的同一性的問題。本發明包括一個麥克風、用於產生語音輸入信號的參數表示的前端信號處理器、一個

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

發音數據庫、一個用於比較輸入信號的參數表示與字母發音的參數表示並產生輸入語音與發音數據庫中的字母之間的一系列關聯的字母相似性比較器、一個詞彙數據庫，一個用於比較聚合在一起的大量字母與詞彙數據庫中的字並產生它們之間的一系列的關連的字比較器，和一個用於顯示認可的選擇的字母和字的顯示器。

圖1是用於根據本發明生成個人化的數據庫的系統框圖；

圖2是根據本發明的語音識別系統的框圖；

圖3是根據本發明的LPC前端處理器的框圖；

圖4是根據本發明的字母語音識別系統的框圖；

圖5是根據本發明的由麥克風產生的字母波形示例；

圖6是根據本發明的用於計算總的失真倒頻譜的動態時間偏差初始化流程過程；

圖7是根據本發明的用於計算總的失真倒頻譜的動態時間偏差迭代過程流程；

圖8是根據本發明的用於計算總的失真倒頻譜的相關值的動態時間偏差流程；

圖9是具有把本發明的一個實施例內置其中的蜂窩電話的系統結構框圖；

圖10圖示與已有技術系統ART和Sensory相比的本發明的一個實施例的字識別性能結果。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明()

圖1是本發明的個人化的數據庫100的框圖。麥克風101接收音頻聲音串(例如,字母表的字母或各個漢字符號)並把聲音串轉換為電子波形信號。前端信號處理器102處理波形來產生適合於識別和比較的波形的參數表示。在優選的實施例中,聲音串被線性預測編碼(LPC)處理,產生對於各個字母的參數表示(把冗余信息從波形數據中抽取出來以更清晰地描述各個音頻信號的所謂的“特徵抽取”)。結果,對於英文字母表中的字母是26X26矩陣,其中列中容納各個字符和字母的參數表示,行容納輸入的語音字符或字母。在本發明中,矩陣是由存儲在發音數據庫103中各個用戶對字母的發音構成的“校準模板”。由於聲音輸入由校準模板來校準,在字比較中可避免一般的語音識別不準確(附有一個樣本校準表來作為參考)。

圖2是本發明的優選實施例的框圖。麥克風101接收被傳送到前端信號處理器102以形成參數化的聲音串波形組的一系列的輸入的發音,接著將應用字母比較器201使聲音串波形組與發音數據庫103相比來選擇對各個發音(例如字母)的最佳匹配。作為一個示例,假設名字“Michael”沒有被正確地發音而成為“n-y-d-h-a-b-l”(由於可混淆的字母發音而引起的一些錯誤)。在一個實施例中,字母比較器201接受聲音串並通過相對於發音數據庫103中的所有字母測試示例中的6個字

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明()

母來確定聲音串發音與發音數據庫103中的校準模板之間的“差距”。在另一個實施例中，類似發音的字母（或一些聲音）基於相似性來分組，從而比較更有效。字相似性比較器202比較校準的字母系列與預先記錄的詞彙數據庫203中的所有項。在該示例中，即使字的字母沒能被正確地識別出，由於這裡僅有有限數目的敏感(sensical)字，如“Michael”，準確的字匹配的機會被相當程度地提高。在本發明的優選實施例中，詞彙數據庫203是來自本發明的受讓人VerbalTek, Inc.的可利用的字典數據庫。本發明的另一個實施例有利地利用從摩托羅拉(Motorola)進入詞彙數據庫201的字典數據庫。本發明還有一個實施例利用用戶的地址簿項。本發明期望由用戶對詞彙數據庫203所要求的任何項所構成的字典。例如，用於希望的專用領域（金融、商業、服務業、工藝、學院或諸如法律、醫藥、會計等的所有專業）的專用字可被有利地引入詞彙數據庫203。而且，熟悉本領域的人員可理解本發明期望能有利地用於諸如中文的單音節字。通過把在詞彙數據庫203中預先記錄的波形與輸入的波形相比，可產生一系列的音素組，這些音素可能與口語輸入相匹配，並且產生音素網格。通過基於各個輸入的組合體與詞彙數據庫203中的字之間的接近性對各個輸入的波形指定“得分”來構造網格。“接近性”指數是基於輸入波形與存儲的詞匯波

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

形之間的計算的失真的，從而產生“失真得分”。由於得分是基於字母或字符與個人化的發音數據庫之間的相對準確度的（與傳統的語音識別聲學一語音方法相比）匹配的，音素網格產生95%或更高的字匹配準確度。然後在顯示器204上顯示字的最佳匹配。

在本發明的優選實施例中，把語音波形(其示例在圖5中表示)轉換為參數表示的前端信號處理器利用線性預測編碼(LPC)。LPC尤其適合於本發明，因為(1)LPC對於語音頻譜包絡線的濁音區比對清音區更有效，並且本發明有利地利用了強調字母的有區別的發音並具有自然的中頓的拼寫發音(從而清音區是較為不重要的)，和(2)LPC提供計算起來有效的表示，其考慮了聲道特性(從而允許個人化的發音用小的處理和存儲量來實現)。本發明中LPC的特殊功用在例如LPC自相關方法中圖示出來，其中假設語音分段在給出的間隔之外全部是0(等於用有限的長度的窗口乘以語音信號)。在LPC傳送函數中， $H(z)=S(z)/GU(z)$ ，其中源的增益G從語音信號估測出來，並且定標的源被用作對數字濾波器 $H(z)$ 的輸入，該濾波器被正被產生的語音的聲道參數特性控制。

圖3是根據本發明的優選實施例的LPC前端處理器102的框圖。優選是固定的低級數字系統(通常是一級

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

FIR濾波器)的預加重器301在頻譜上把信號s(n)展平，並且由下面來描述

$$P(z)=1-az^{-1}$$
 (等式1)

這裡 $0.9 \leq a \leq 1.0$ 。在本發明的另一實施例中，預加重器301是具有下面的傳送函數的一級自適應系統

$$P(z)=1-a_nz^{-1}$$
 (等式2)

這裡 a_n 根據預定的自適應標準隨時間(n)而改變，例如 $a_n=r_n(1)/r_n(0)$ ，其中 $r_n(i)$ 是自相關序列的第i個採樣。幀分塊器302用被M個採樣分隔開的相鄰的幀來對N個採樣的幀中的語音信號幀分塊。在本發明的此實施例中，當語音的採樣速率是8kHz時 $N=M=160$ ，相當於它們之間沒有分開的20毫秒的幀。每幀有一個特徵，從而對於一個第二種發音(50幀長)，12個參數代表幀數據，並且產生50X12矩陣(模板特徵組)。開窗器303對每一獨立的幀開一個窗口以在各幀的開始和末尾處最小化信號的不連續。在本發明的優選的實施例中，其中 $M=N$ ，矩形窗口被用於避免在窗口邊界處數據的丟失。自相關器304執行下面給出的自相關

$$r_1(m)=\sum_{n=0}^{N-1-m}x_1(n)x_1(n+m)$$
 (等式3)

這裡 $m=0,1,\dots,p$ ，並且p是LPC分析的級。本發明的優選實施例使用 $p=10$ 但是在本發明的其它實施例中可有利地使用從8到16的p值，並且提高準確度的其它的值也在本發明的期望的範圍內。零自相關是給出的幀

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

經濟部智慧財產局員工消費合作社印製

五、發明說明()

的幀能量。倒頻系數產生器305應用已有技術中已知的Durbin方法把各幀轉換為倒頻系數（對數幅度頻譜的反傅立葉變換，參考下面）。錐形的倒頻開窗器306加權倒頻系數以最小化噪音的效果。選擇錐形的開窗器306來降低低級倒頻系數對整個頻譜傾斜的敏感度以及高級倒頻系數對噪音的敏感度（或其它不希望的易變性）。時間差分器307優選使用近似地正交多項式擬合（在本實施例中，對有限長度的窗口的導數的最小平方估測）產生倒頻系數的一次導數，從而產生處理的信號 $S'(n)$ 。在另一實施例中，可通過時間差分器307應用在已有技術中已知的近似方法來產生二次導數，以提供進一步的語音信號信息，從而改善語音信號的頻譜性能表示。還有一個實施例跳過時間差分器來產生信號 $S''(n)$ 。應理解使用LPC和上述技術的前端信號處理器102的上述描述是為了揭示最佳實施例，前端信號處理的其它的技術和方法可被有利地使用在本發明中。若發音串是拼出的字母或字串，用於匹配發音串的比較技術和方法基本上是類似的，所以下面描述了一種用在比較器201和202的優選實施例中的技術。

在本發明的優選實施例中，參數表示是倒頻系數，輸入的語音與預先記錄的數據庫中的發音通過比較倒頻差距進行對比。輸入的字母或字符（或用於比較器202的在字組合中的字母或字符）產生若干根據相似性分等

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

五、發明說明()

級的候選字母(字)匹配。在把預先記錄的波形與輸入的波形比較時，產生可能與口頭輸入相匹配的一系列的音素組，其在矩陣中排序時產生音素的網格。該網格通過基於各個輸入的組合與詞彙數據庫中的字的接近性來對各個輸入的波形指定“得分”值而排序。“接近性”指數是基於輸入波形與存儲的詞匯波形之間的倒頻差距的，從而產生“失真得分”。由於得分是基於字母（而非整個字）的相對準確（與傳統的語音識別聲學一語音方法相比）匹配的，音素網格產生95%或更準確的字匹配。

圖4表示輸入到字母校準器401的波形參數表示，其中與發音數據庫103相聯系，產生校準矩陣（後附示例）。失真校準器402在下面描述的優選實施例中的倒頻差距的基礎上計算輸入的語音與發音數據庫103中的所有項之間的失真。然後得分計算器403基於預定的標準（如倒頻差距）分配得分並且選擇器404選擇候選字母（字）。在對數幅度與頻率刻度關係上的兩個語音頻譜之間的不同是

$$V(\omega)=\log S(\omega)-\log S'(\omega) \qquad \text{(等式4)}$$

在優選實施例中，為表示兩個語音特徵矢量的不相似性，優選的實施例使用幅度的對數（對於頻率）的平均絕對值，即利用該組平均值，使用頻譜失真（或差距）的對數的均方根(rms)測量

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

$$d(S,S')^p = \int_{-\pi}^{\pi} |V(\omega)|^p d\omega / 2\pi \tag{等式 5}$$

其中當 $p=1$ 時，是頻譜失真的對數的平均絕對值，
當 $p=2$ 時，是頻譜失真對數的rms值。在優選實施例
中，差距或失真測量以信號的複數倒頻譜表示，其被定
義為信號頻譜的對數的傅立葉變換。對於相對於 $\omega=0$ 對
稱的並且對於採樣的數據序列是周期性的能譜，

$\log S(\omega)$ 的傅立葉級數表示為

$$\log S(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega} \tag{等式 6}$$

這裡 $c_n = c_{-n}$ 是倒頻系數。

$$c_0 = \int_{-\pi}^{\pi} \log S(\omega) d\omega / 2\pi \tag{等式 7}$$

$$d(S,S')^2 = \int_{-\pi}^{\pi} |\log S(\omega) - \log S'(\omega)|^2 d\omega / 2\pi = \sum_{n=-\infty}^{\infty} (c_n - c_n')^2 \tag{等式 8}$$

其中 c_n 和 c_n' 分別是 $S(\omega)$ 和 $S'(\omega)$ 的倒頻系數。通
過有限的加和，例如在優選的實施例中是10-30項的加
和，本發明利用截取的倒頻差距。這有效地（意思是相
對小的計算負擔）估算了頻譜差距的對數的rms值。由
於語音信號的可感覺到的響度大約是對數形式，對數頻
譜差距的選擇很適合於鑒別主體的聲音的不同。而且，
低的倒頻系數的易變性主要是由於語音和傳送失真的反
複無常的變化引起的，從而倒頻譜（倒頻差距組）被有
利地選擇來用於失真測量。相同發音的不同的聲學再現
通常是以不同時間速率說出的，從而說出速率變化和持
續時間變化不應促進語言的不相似性得分。動態時間偏
差（DTV）408執行頻譜的動態行為分析以更準確的確

（請先閱讀背面之注意事項再填寫本頁）

裝
訂
線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

定輸入的發音與匹配的數據庫值之間的不相似性。
DTW408通過經把兩個要進行比較的模式之聲學特徵作
柵格映射找到“最佳”路徑而對說出速率波動進行時間
校準和規範化。在優選實施例中，DTW408通過動態設
計最小化不相似性而找到最佳路徑。兩個時間偏差函數
 φ_x, φ_y 分別把語音模式的兩個時間波動指數 i_x, i_y 與公共
時間軸 k 關聯起來，從而

$$\begin{aligned} i_x &= \varphi_x(k) & k &= 1, 2, \dots, T \\ i_y &= \varphi_y(k) & k &= 1, 2, \dots, T \end{aligned} \quad (\text{等式 9})$$

基於偏差函數對將全球模式不相似性定義為經整個
發音的累加的失真：

$$d_\varphi(X, Y) = \sum_{k=1}^T d(\varphi_x(k), \varphi_y(k)) m(k) / M_\varphi \quad (\text{等式 10})$$

這裡 $d(\varphi_x(k), \varphi_y(k))$ 是對於 $x_{\varphi_x(k)} y_{\varphi_y(k)}$ 定義的短
時頻譜失真， $m(k)$ 是非負的加權函數， M_φ 是規範化因
子， T 是正常時序表上兩種語音模式的“正常”持續時
間。路徑 $\varphi = (\varphi_x, \varphi_y)$ 被選擇來用同一性測量整個路徑的
不相似性。在本發明的優選的實施例中，不相似性
 $d(X, Y)$ 被定義為在整個路徑上的 $d_\varphi(X, Y)$ 的最小值，
即：

$$d(X, Y) = \min_{\varphi} d_\varphi(X, Y) \quad (\text{等式 11})$$

在 X 和 Y 是同一字的發音時上述定義是準確的，因
為沿校準路徑最小化累加的失真意味著不相似性是基於
最好的可能的校準來補償說出速率不同而被測量的。已

五、發明說明 ()

有技術中公知動態設計可通過找到最佳路徑來解決如在緊靠上面描述的時序確定問題，意思是這樣使從一個點向另一個點移動的“成本”最低。在本發明的一個實施例中，由於包括在移動中的步驟數目被“if-then”語句確定，時序確定是非同步的。該確定利用了遞歸關係，該遞歸關係允許最佳路徑搜索被遞增地實施並被下面所述的本發明的優選實施例中的算法所執行。用於確定最佳路徑中的下一個點和目的地點的確定規則（“策略”）完全限定了尋求被最小化的成本。引起成本 $\zeta(j,i)$ 的從開始點1向中間點j移動的最佳策略由下式給出：

$$\varphi(1,i)=\min_j[\varphi(1,j)+\zeta(j,i)] \quad (\text{等式12})$$

因此得到從點i到點j的最佳順序的移動和相關的最小成本，

$$\varphi(i,j)=\min_l[\varphi(i,l)+\varphi(l,j)] \quad (\text{等式13})$$

在另一實施例中，順序確定是同步的（對於固定移動數目M確定過程的規律性），相關的最小成本 $\varphi_m(i,1)$ 為

$$\varphi_{m+1}(i,n)=\min_l[\varphi_m(i,l)+\zeta(l,n)] \quad (\text{等式14})$$

其是 used 在本發明的一個實施例中的遞歸關係。

在兩個實施例中，方法有步驟(1)初始化，(2)遞歸，(3)結束和(4)後跟蹤，並且分別如下：

初始化： $\varphi_1(i,n)=\zeta(i,n)$

$$\zeta_1(n)=i, n=1,2,\dots,N$$

五、發明說明 ()

遞歸：
$$\varphi_{m+1}(i,n)=\min_{1\leq l\leq N}[\varphi_m(i,l)+\zeta(l,n)]$$
$$\zeta_{m+1}(n)=\arg\min_{1\leq l\leq N}[\varphi_m(i,l)+\zeta(l,n)]$$
$$n=1,2,\dots,N, m=1,2,\dots,M-2$$

結束：
$$\varphi_M(i,j)=\min_{1\leq l\leq N}[\varphi_{M-1}(i,l)+\zeta(l,j)]$$
$$\zeta_M(j)=\arg\min_{1\leq l\leq N}[\varphi_{M-1}(i,l)+\zeta(l,j)]$$

路徑後跟蹤：最佳路徑 $= (i, i_1, i_2, \dots, i_{M-1}, j)$ ，
這裡 $i_m = \zeta_{m+1}(i_{m+1}) \quad m=M-1, M-2, \dots, 1$
 $i_M=j$ ，

上述算法計算起來有效，從而適合於本發明期望的手持裝置。

圖6，7和8構成用於計算要被比較的模板之間的總的失真的DTW408的優選實施例的流程圖。“差距” $d(i,j)$ (上面等式11)是模板X的第i個特徵與模板Y的第j個特徵之間的失真。圖6描述初始化過程601，其中前面的差距在602處是 $d(0,0)$ 。指數j在603處被增加並且現在前面的差距是在j處的差距 $\text{pre dist}[j]$ ，其等於 $\text{prev dist}[j-1]+d(0,j)$ 。在605，如果j小於模板Y中的特徵數目($j<\text{numY}$)，那麼在606處j被增加並被反饋給604，用於 $\text{pre dist}[j]$ 的新的計算。如果j不大於 numY ，那麼初始化完成，對於總的差距的迭代步驟611如圖7所示開始進行。在612，i被設置為1並且當前的差距($\text{curr dist}[0]$)被計算為 $\text{pre dist}[0]+d(i,$

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 ()

0)。在614，j被設置為1並且導致相關差距d1,d2或d3的可能的路徑被計算為□

$$\text{curr dist}[j-1]+ d(i,j)=d_1$$

$$\text{prev dist}[j]+ d(i,j)=d_2$$

$$\text{prev dist}[j-1]+ d(i,j)=d_3$$

相關的差距的相關值然後在圖8的621和622被測試。如果d3不大於d1且不大於d2，那麼d3是最小值並且curr dist[j]在623處是d3。在對第j個特徵測試為小於Y模板中的特徵數目後，在617處j被增加並被反饋給可能路徑的差距計算，再發生最小化處理。如果d2大於d1並且d3大於d1，那麼d1是最小值並且被設置為curr dist[j]。然後相對於Y模板中的特徵數目再次在626處測試j，在617處j被增加並被反饋來用於遞歸。如果d3大於d2並且d1大於d2，那麼d2是最小值並且被設置為curr dist[j]，並重複同樣的處理使j被增加並被反饋。以這種方式，找到最小差距。如果在626處j大於或等於模板Y中的特徵數目，那麼測試i來看它是否等於模板X中的特徵數目減1。如果i不等於模板X中的特徵數目減1，在618先前的差距被設置為用於j指數（直到numY-1）的當前差距，在616處i被增加並被反饋給613，用於把當前差距設置為先前差距加新的第i個差距，對每個i重複這種處理直到次數j等於模板X中的特徵數目減

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

經濟部智慧財產局員工消費合作社印製

五、發明說明 ()

1。如果i等於模板X中的特徵數目減1，總的失真在628計算為：

$$\text{總失真} = \frac{\text{curr} - \text{dist}(\text{numY} - 1)}{(\text{numY} - \text{numY} - 1)}$$

從而完成找到總的失真的算法。

為實現最佳的準確度，限制時間偏差函數。已有技術中公知甚至小的語音端點錯誤將導致語音檢測準確度上的明顯變差。在被控制的環境中清晰發出的語音中，可達到高的檢測準確度，但是對於一般的應用（如在蜂窩電話中），說話者聲音的反複無常的變化（包括唇部翕動、呼吸、清嗓子等等）和背底噪音使得難以產生準確的端點檢測。如果發音有完好限定的端點（標記模式的開始和結束的幀），相似性比較將更準確。應用於輸入發音的拼寫中的本發明作為字母的通常更準確的發音（與字相比）和在拼寫字時字母之間的通常的中頓而實現了準確度。從而暫時的變化一般被限定於端點區。從而對於時間偏差函數的限制把開始和結束點處的值設置為第一個和最後一個時間函數指數 $i_x=1$ 和 $i_y=T$ 。這些端點限制通過等式(11)根據項 T_x 和 T_y 結合到本發明中

$$M_{\phi}d(X,Y) \cong D(T_x,T_y) = \min_{\phi_x,\phi_y} \sum_{k=1}^T d(\phi_x(k),\phi_y(k))m(k) \quad (\text{等式15})$$

這裡X和Y分別結束於 T_x 和 T_y 。本發明的優選實施例提供一種最適合於本發明的拼寫輸入發音識別系統的動態時間偏差方式。DTW 408利用等式15來沿連接(1,1)和(i_x,i_y)的路徑產生最小的部分積累的失真：

五、發明說明 ()

$$D(i_x,i_y) \cong \min_{\varphi_x,\varphi_y,T'} \sum_{k=1}^T d(\varphi_x(k),\varphi_y(k))m(k) \quad (\text{等式 16})$$

這裡 $\varphi_x(T')=i_x$, $\varphi_y(T')=i_y$, 用限制條件動態設計遞歸變成

$$D(i_x,i_y) = \min_{(i',T')} [D(i_x',i_y') + \zeta((i_x',i_y'),(i_x,i_y))] \quad (\text{等式 17})$$

這裡 ζ 是點 (i_x',i_y') 與 (i_x,i_y) 之間的加權的積累的失真(局部差距) ,

$$\zeta((i_x',i_y'),(i_x,i_y)) = \sum_{l=0}^L d(\varphi_x(T-1),\varphi_y(T-1))m(T-1) \quad (\text{等式 18})$$

Ls 是根據 φ_x 到 φ_y 的從 (i_x',i_y') 與 (i_x,i_y) 的路徑中的移動數目。遞增的失真 ζ 僅沿由各種限制條件限定的路徑被評估，從而最小化過程可有效地在限制條件內解決。但是，不相似的直接推斷的特性也可有利地包括在該方法中。例如，在本發明中，把一幀切分成多個分段來區分易混淆的字母發音，如字母“a”和“f”。可以理解許多不同的限制條件和限制條件的組合在本發明的範圍內。在不同字母的發音中，例如，用於最準確的比較的時間校準不是定義清晰的語言概念，從而在本發明中應用直覺的限制條件。

在操作中，用戶可使用與說話者無關的輸入設定模式，從而使用用於語音識別的預先分組的字母數據庫。在本發明的一個實施例中，菜單選擇允許用戶選擇男聲和女生聲音識別以及語言選擇。發音數據庫103包括用於男聲或女生聲音或不同語言的預先記錄的模板。如果用戶把他／她自己的聲音記錄在他／她選擇的語音中，

五、發明說明()

這將被記錄在發音數據庫103中，從而男／女／語言模板是不必要的。

為生成(“訓練”)個人化的數據庫100，用戶通過向系統以在各個聲音之間帶有0.2秒的間隔的自然的声音口述從"a"到"z"以及從“0”到“9”的記錄字母和數值聲音，從而產生“聲音串”。在本發明的一個實施例中，如果字母繼續向前進，上述的端點檢測方案將檢測不能區分的發音，用戶經顯示器被指令來停止複述並且從開始重複口述。應用對各個字母產生波形的前端信號處理器102來轉換聲音串的各个字母(如圖5所示)。然後波形被分段、分配存儲器中的地址，接著被存儲在存儲器中，從而各個發音被映射到發音數據庫104(把這一處理稱為“貼標籤”)。與字母發音比較器201聯系的發音數據庫103對於英文字母表中的字母形成26X26矩陣，其中列裝有對發音數據庫104中的各个字母的存儲的波形，行裝有用於識別分析的輸入的語音字母(附加有樣本矩陣)。字母發音比較器201比較輸入的發音與列(發音數據庫103)中的所有字母以找到最好的匹配。例如，輸入的字“seat”由用戶拼寫為“s-e-a-t”。前端信號處理器102執行特徵抽取，例如其對每一秒的語音可產生50列的倒頻系數(或矢量)。字母發音比較器201利用動態時間偏差(DTW)比較各个字母的倒頻差距與發音數據庫103中的從a到z的項。

五、發明說明()

如果輸入的語音持續2秒(100幀)並且語音文件的各個部分是25幀，字母發音比較器201比較倒頻矢量的25列與發音數據庫103中的字母表的26個字母。假設發音數據庫103中的各個字母是25幀長，DTW比較是25X25。由於發音的難以預測的變化、背底噪音和其他因素，字母可被識別為“x-e-k-d”(這些字母的每一個的發音類似於需要的字母，因此被錯誤地“識別”)。在本發明的優選實施例中，在與數據庫中的字母作比較之前，相似的發音的字母被分組在一起，從而搜索更有效。例如，在本發明的優選實施例中的分組重點放在音節的元音聲音上並且被找到來明顯降低相似性計算，從而對手持裝置產生理想化的分組。分組的另一個示例把“a”、“j”和“k”分為同一組，“x”、“s”和“f”分為另一組，“b”、“c”、“d”、“e”和“g”分為又一組。作為失真得分技術的圖示，在“s-e-a-t”示例中，第一字母“s”初始被識別為“x”，從而這裡有一個基於倒頻差距被指定的非零失真得分(例如2.0)；下一個字母“e”被正確地識別，從而得分為0；下一個字母“a”被識別為“k”，其被指定得分1.5；最後一個字母“t”被被識別為“d”，其被指定得分1.0。對於這個字的總的失真得分是4.5。然後失真得分組合起來與詞彙數據庫203中的字相比。但是選擇的候選字母組合起來更清晰(並且“xekd”不

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

五、發明說明()

作為一個字存在)。字相似性比較器202應用該系列字母“x-e-k-d”計算失真得分以基於字母之間的差距用字如下所示產生失真得分:

輸入字	候選字	字母得分	失真得分	相似性%
Xekd	seat	S1+S2+S3 +S4=S	1200	89%
	feat	T1+T2+T3 +T4=T	2380	75%
	heat	U1+U2+U3 +U4=U	4530	68%
	beat	V1+V2+V3 +V4=V	8820	42%

字相似性比較器202利用動態時間偏差(DTW)對各次比較的失真得分進行排序以確定最低的失真得分。在本例中DTW計算僅要求4X4矩陣(反映字的長度),從而實現高的計算效率。與詞彙數據庫203中的字的最接近的匹配(或最大相似性)是“seat”。顯示器204顯示由用戶確認的選擇的字(或各個字母)。可有利地使用任何字母數字顯示裝置,例如液晶顯示器(LCD)。對於在移動電話或PDA中的應用,那麼字母的組合構成字,然後字可與電話號碼匹配或與用於傳送的其他傳送指數匹配。

五、發明說明()

儘管描述的許多示例是基於英語字母表的，應理解它們只是例示，本發明的範圍還包括其它語言，唯一的限制是這種語言是基於可區分的發音的。實際上，由於任何語言的語音識別因其基本上依賴於詞彙數據庫的內容而可通過本發明實現，本發明的一個實施例提供多種語言性能，該發明可對於任何語言來改變。與說出的聲音和詞彙數據庫的相似性比較可通過上述的本發明完成，經用戶口述的聲音來構造發音數據庫104而實現了準確度。

由於它的小的形成因素，本發明的一個實施例允許它集中到用戶的電子設備的現有的操作系統中（例如用於PDA的微軟WindowCE[®]和用於蜂窩電話的ARM7TDMI），從而不需要大範圍的再設計和再改編。本發明的語音識別系統的一個實施例還可以被裝載到諸如蜂窩電話或PDA的裝置的閃存中，從而允許把本發明與現有電子設備進行的簡單快速低廉組合，避免主機裝置的DSP的再設計和再改編。而且語音識別程序可被終端用戶經耦合於閃存的數據端口裝載到存儲器中。這也可通過從因特網下載完成。圖9表示帶有內置其中的本發明的實施例的蜂窩電話的系統結構。閃存901被耦合於還耦合於DSP處理器903的微處理器902，其聯系閃存901和微處理器902執行上述語音識別。只讀存儲器(ROM)裝置904和隨機訪問存儲器(RAM)裝置

（請先閱讀背面之注意事項再填寫本頁）

裝

訂

線

五、發明說明（ ）

905通過提供對於發音數據庫104和詞彙數據庫203的存儲器存儲而服務於DSP處理器903。經麥克風907的語音輸入被編碼器／解碼器（CODEC）906編碼。在DSP處理器903進行語音識別後，語音信號被CODEC 906解碼並被傳送到揚聲器908用於音頻確認。另一種情況是，揚聲器908可以是可視的顯示器。本發明的應用協議接口（API）的一個實施例應用ARM77TDMI作為基礎遵守下面的說明

存儲器應用要求10KB的編碼容量，4KB的緩衝存儲器容量，存儲量（每一模板）是0.5KB。計算的要求是1.9MIPS的語音特徵抽取和每模板0.5MPIS的語音識別。

在通常應用中，本發明允許聲音識別在1-2秒中實現，從而提供與用戶的平滑接口。準確度測試結果一致地為95%水平。本發明的一個實施例的名字（字）識別性能結果在圖10中表示，其中本發明的錯誤率

（VerbalLab）與已有技術系統ART和Sensory進行對比。測試使用24個名字，其中8個（Mickey Mouse, Minnie Mouse, Minnie Driver, Bill Dates, Bill Clinton, Julius Smith, Julia roberts, Bruce Willis）被確認為是含糊的，並且有8個說話者（3女和5男）。對於此次測試中的所有的說話者，本發明有明

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

五、發明說明()

顯的低的錯誤率，錯誤率的所有結果小於測試的其它系統的一半。

尤其這裡應注意本發明對於輸入漢語和韓國語的單音節字符為基礎的字母（或字）是理想的。例如，對於“行動電話”一詞，被作為字符串“xing-dong-dian-hua”直譯出來。各個字是單音節的並且具有它自己的意思（或幾個意思），但是組合起來包含“行動電話”這個唯一的術語。本發明提供各個單音節字母的高準確度的識別（部分由於直接推斷校準），其在組合一起形成一個字詞時由於有限數目的可檢測的選擇而產生更準確的識別。

儘管上面是特定實施例的完全的描述，可使用各種變形、可替換的結構和等同替代體。例如，本發明適合於被分開為發音的任何口語語言、把字母表的字母組合起來發音的字母語言（像英語和俄語）以及發音與字符相關的符號語言（如漢語、韓國語和日語）。而且，為了增強識別準確度，DTW 方案中的限制可從若干可能性中選擇以有利地產生準確的語音識別。因此，上面的描述和圖示不應被看作對本發明的範圍的限制，而發明範圍在後附的權利要求中限定。

四、中文發明摘要 (發明之名稱： 用於移動通信的拼寫語音識別裝置和方法)

一種準確的語音識別系統，包括一個麥克風、用於產生語音輸入信號的參數表示的前端信號處理器、一個發音數據庫、一個用於比較輸入信號的表示與字母發音的參數表示並產生書入語音與發音數據庫中的字母之間的一系列關聯的字母相似性比較器、一個詞彙數據庫，一個用於比較聚合在一起的大量字母與詞彙數據庫中的字並產生它們之間的一系列的關聯的字比較器，和一個用於顯示認可的選擇的字母和字的顯示器。

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

訂

線

英文發明摘要 (發明之名稱：

)
An accurate speech recognition system capable of rapidly processing greater varieties of words and operable in many different devices, but without the computational power and memory requirements, high power consumption, complex operating system, high costs, and weight of traditional systems. The utilization of individual letter utterances to transmit words allows voice information transfer for both person-to-person and person-to-machine communication for mobile phones, PDAs, and other communication devices. This invention is an apparatus and method for a speech recognition system comprising a microphone, front-end signal processor for generating parametric representations of speech input signals, a pronunciation database, a letter similarity comparator for comparing the parametric representation of the input signals with the parametric representations of letter pronunciations, and generating a sequence of associations between the input speech and the letters in the pronunciation database, a vocabulary database, a word similarity comparator for comparing an aggregated plurality of the letters with the words in the vocabulary database and generating a sequence of associations between them, and a display for displaying the selected letters and words for confirmation.

J:\CHING\b08048.doc

六、申請專利範圍

1. 一種語音識別系統，包括：

用於接收聲波並把聲波轉換成電信號的麥克風裝置；

耦合於所述麥克風裝置的用於處理電信號來產生電信號的參考表示的前端信號處理裝置；

用於存儲字母發音的大量參數表示的發音數據庫存儲裝置；

耦合於所述前端信號處理裝置及所述發音數據庫存儲裝置的用於比較電信號的參數表示與字母發音的所述大量參數表示的並響應於預定的準則產生電信號的參數表示與字母發音的所述大量參數表示之間的第一序列關聯的字母相似性比較器裝置；

用於存儲字發音的大量參數表示的詞彙數據庫存儲裝置；

耦合於所述字母相似性比較器及所述詞彙數據庫存儲裝置的用於比較聚集一起的大量的字母發音的參數表示與字發音的所述大量參數表示的並響應於預定的準則產生所述聚集在一起的字母發音的大量參數表示中的至少一個與字發音的所述大量參數表示中的至少一個之間的第二序列關聯的字相似性比較器裝置；及

耦合於所述字相似性比較器裝置的用於顯示所述第一和第二序列的關聯的顯示裝置。

六、申請專利範圍

5. 如申請專利範圍第1項的語音識別系統，其特徵在於所述字母相似性比較器裝置包括：

耦合於所述發音數據庫存儲裝置的用存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示對電信號的參數表示進行校準的字母校準裝置；

用於在電信號的參數表示和存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示上執行動態時間偏差的動態時間偏差器裝置；

耦合於所述字母校準裝置和所述動態時間偏差器裝置的用於計算電信號的參數表示與存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示之間的失真的失真計算裝置；

耦合於所述失真計算裝置的用於響應於預定的準則對所述失真指定一個得分的得分裝置；及

耦合於所述得分裝置的用於選擇存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示中的具有最低的失真的至少一個的選擇裝置。

6. 如申請專利範圍第5項的語音識別系統，其特徵在於所述動態時間偏差器裝置包括用於確定電信號的參數表示與存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示之間的最小倒頻差距的最小化裝置。

(請先閱讀背面之注意事項再填寫本頁)

表

訂

線

六、申請專利範圍

7．如申請專利範圍第1項的語音識別系統，其特徵在於存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示包括漢語的各個字符的發音並且存儲在所述詞彙數據庫存儲裝置中的字發音的所述大量參數表示包括漢語的聚集一起的字串的發音。

8．如申請專利範圍第1項的語音識別系統，其特徵在於存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示包括韓國語的各個字符的發音並且存儲在所述詞彙數據庫存儲裝置中的字發音的所述大量參數表示包括韓國語的聚集一起的字串的發音。

9．如申請專利範圍第1項的語音識別系統，其特徵在於存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示包括日語的各個字符的發音並且存儲在所述詞彙數據庫存儲裝置中的字發音的所述大量參數表示包括日語的聚集一起的字串的發音。

10．如申請專利範圍第1項的語音識別系統，其特徵在於存儲在所述發音數據庫存儲裝置中的字母發音的所述大量參數表示包括法語的各個字符的發音並且存儲在所述詞彙數據庫存儲裝置中的字發音的所述大量參數表示包括法語的聚集一起的字串的發音。

11．一種字母相似性比較器，包括：
用於接收電信號參數表示的裝置；

經濟部智慧財產局員工消費合作社印製

六、申請專利範圍

用於存儲大量的字母發音參數表示的發音數據庫存儲裝置；

耦合於所述接收裝置和所述發音數據庫存儲裝置的用存儲在所述發音數據庫存儲裝置中的所述大量的字母發音參數表示對電信號參數表示進行校準的字母校準裝置；

用於在電信號參數表示和存儲在所述發音數據庫存儲裝置中的所述大量的字母發音參數表示上執行動態時間偏差的動態時間偏差器裝置；

耦合於所述字母校準裝置和所述動態時間偏差器裝置的用於計算電信號參數表示與存儲在所述發音數據庫存儲裝置中的所述大量的字母發音參數表示之間的失真的失真計算裝置；

耦合於所述失真計算裝置的用於響應於預定的準則對所述失真指定一個得分的得分裝置；及

耦合於所述得分裝置的用於選擇所述大量的字母發音參數表示中的具有最低的失真的至少一個的選擇裝置。

12. 一種電通信裝置，包括：

用於接收聲信號並從那裏產生電信號的麥克風；

耦合於所述麥克風的用於編碼和解碼電信號的邊碼解碼器；

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

耦合於所述編碼解碼器的用於處理電信號從而產生電信號的參數參數表示的信號處理器；

耦合於所述信號處理器的用於存儲數據庫並使其中的第一扇區區存儲大量字母發音參數表示且使其中的第二扇區存儲大量字發發音參數表示的數據庫存儲單元；

耦合於所述信號處理器及所述數據庫存儲單元的用於比較電信號的參術表示與所述數據庫存儲單元的所述第一扇區中的所述大量字母發音參數表示的第一比較器；

耦合於所述第一比較器的用於響應於預定的準則選擇至少一個所述的大量字母發音參數表示的第一選擇器；

耦合於所述信號處理器及所述數據庫存儲單元的用於比較聚集在一起的字母發音參數表示與在所述數據庫存儲單元的所述第二扇區中的所述大量字發音參數表示的第二選擇器；

耦合於所述第二比較器的用於響應於預定的準則選擇至少一個所述的大量字發音參數表示的第二比較器；及

耦合於所述第一和第二選擇器的用於顯示所述大量選擇的字母發音參數表示的至少一個的及用於顯示所述大量字發音參數表示的至少一個的顯示器。

(請先閱讀背面之注意事項再填寫本頁)

訂
線

六、申請專利範圍

13．如申請專利範圍第12項的電通信裝置，其特徵在於存儲單元在所述數據庫存儲單元的所述第一扇區中的所述大量字母發音參數表示響應於參數表示的相似性被分組。

14．如申請專利範圍第12項的電通信裝置，其特徵在於所述第一比較器響應於所述數據庫存儲單元的所述第一扇區中的所述大量字母發音參數表示校準電信號的參數表示。

15．如申請專利範圍第12項的電通信裝置，其特徵在於所述數字信號處理器計算倒頻系數以產生電信號的參數表示、大量字母發音參數表示和所述大量字發音參數表示。

16．如申請專利範圍第12項的電通信裝置，其特徵在於所述第一比較器利用動態時間偏差產生電信號的參數表示與所述大量字母發發音參數表示之間的比較。

17．如申請專利範圍第16項的電通信裝置，其特徵在於所述第一比較器利用倒頻差距來比較電信號的參數表示與所述大量字母發音參數表示之間的比較。

18．如申請專利範圍第12項的電通信裝置，其特徵在於所述第二比較器利用動態時間偏差來產生所述聚集在一起的大量字母發音參數表示與所述大量字發音參數表示之間的比較。

六、申請專利範圍

19．如申請專利範圍第16項的電通信裝置，其特徵在於所述第二比較器利用字母發音序列來比較所述聚集在一起的大量字母發音的參數表示與存儲在所述數據庫存儲單元中的所述大量字發音參數表示。

20．如申請專利範圍第18項的電通信裝置，其特徵在於所述第二比較器利用倒頻差距來比較所述聚集在一起的大量字母發音的參數表示與存儲在所述數據庫存儲單元中的所述大量字發音參數表示之間。

21．一種識別語音發音信號的的方法，包括步驟：

- (a) 參數化大量字母發音；
- (b) 存儲所述大量參數化的字母發音；
- (c) 參數化大量字發音；
- (d) 存儲所述大量參數化的字發音；
- (e) 接收聲波；
- (f) 把聲波轉換成電信號；
- (g) 參數化電信號；
- (h) 比較所述參數化電信號與所述存儲的大量參數化字母發音；
- (i) 響應於預定的參數相似性準則選擇至少一個所述的存儲的大量參數化字母發音；

(請先閱讀背面之注意事項再填寫本頁)

訂
線

六、申請專利範圍

(j) 顯示所述選擇的至少一個所述的存儲的大量參數化字母發發音；

(k) 把所述選擇的至少一個所述的存儲的大量參數化字母發音聚集來形成參數化的字；

(l) 比較所述參數化的字與所述存儲的大量參數化字發音；

(m) 響應於預定的參數相似性準則選擇至少一個所述的存儲的大量參數化字發音；及

(n) 顯示所述選擇的至少一個所述的存儲的大量參數化字發發音。

22．如申請專利範圍第21項的方法，其特徵在於步驟(b)還包括響應於預定的參數化的字母發音相似性準則把所述大量參數化字母發音分組。

23．如申請專利範圍第21項的方法，其特徵在於步驟(h)還包括應響應於預定的準則方法用所述大量參數化的字母發音校準所述大量參數化的電信號的步驟。

24．如申請專利範圍第21項的方法，其特徵在於步驟(a)包括利用倒頻系數來參數化所述大量參數化字母發音。

25．如申請專利範圍第21項的方法，其特徵在於步驟(c)包括利用倒頻系數來參化所述大量參數化字發音。

六、申請專利範圍

26．如申請專利範圍第21項的方法，其特徵在於步驟(g)包括利用倒頻系數來參數化電信號。

27．如申請專利範圍第21項的方法，其特徵在於步驟(h)包括利用倒頻差距來比較參數化的電信號與所述大量參數化字母發音。

28．如申請專利範圍第21項的方法，其特徵在於步驟(I)包括利用倒頻差距來比較參數化的字與所述大量參數化字發音。

29．如申請專利範圍第21項的方法，其特徵在於步驟(I)包括利用字母發音序列來比較參數化的字與所述大量參數化字發音。

30．如申請專利範圍第21項的方法，還包括在步驟(a)之前說出大量字母發音的步驟。

31．如申請專利範圍第30項的方法，還包括在所述說出大量字母發音的步驟後，區分說話者是男還是女的步驟。

32．如申請專利範圍第30項的方法，還包括在所述說出大量字母發音的步驟後，區分說話的字母是否基本上清晰的步驟。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

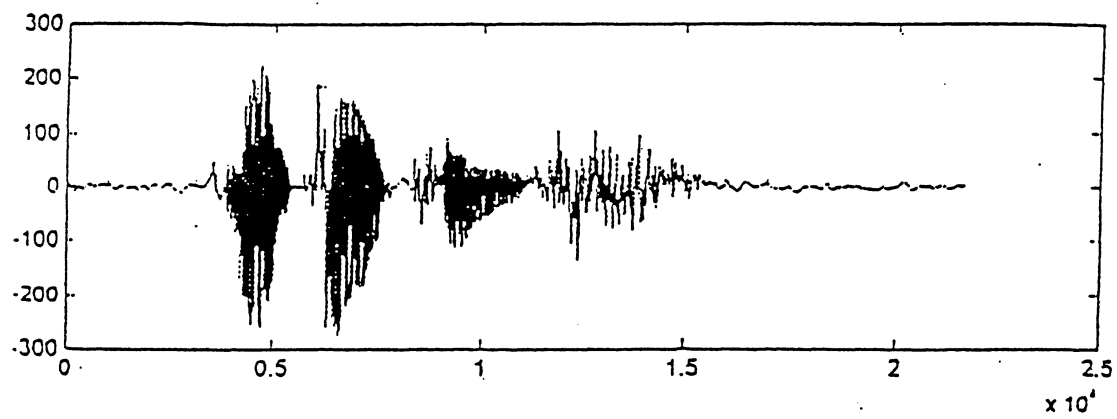


圖 5

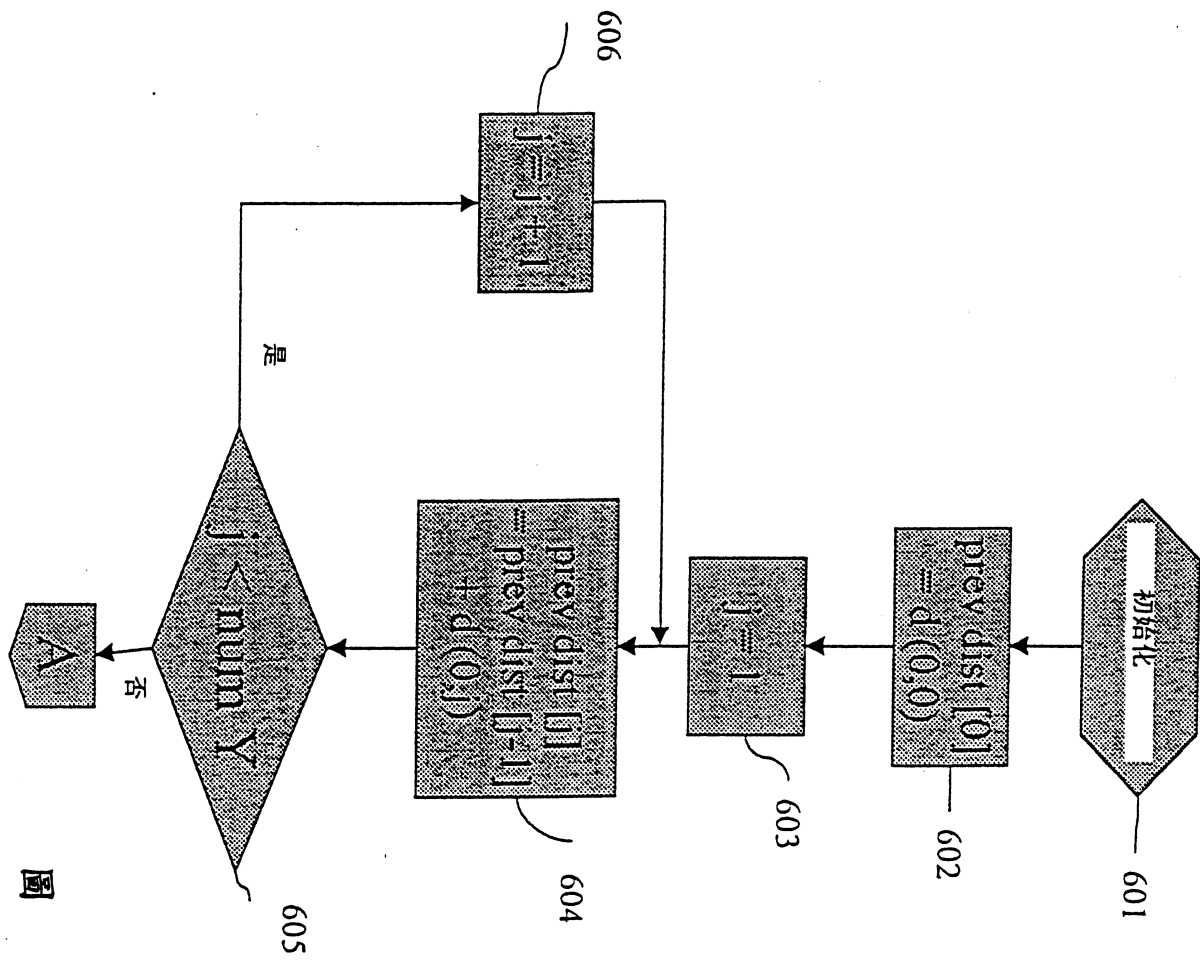


圖 6

公告本

90.11.08

申請日期	90.3.22
案 號	90106813
類 別	G10L15/00

A4 90 11 8
C4

504663

(以上各欄由本局填註)

發 明 專 利 說 明 書		
一、發明 名稱	中 文	用於移動通信的拼寫語音識別裝置和方法
	英 文	
二、發明 創作人	姓 名	(1) 詹姆士·潘 (2) 金潤 (3) 喬瑟芬·張 (4) 陳俊彥
	國 籍	(1) 美國 (2) 韓國 (3) 美國 (4) 美國
三、申請人	住、居所	(1) 美國加利福尼亞州94305史丹福市艾斯康迪多村94D (2) 美國加利福尼亞州94304帕洛奧圖市#302奧克克瑞克大道1540號 (3) 美國加利福尼亞州95131聖荷西市梅貝瑞巷1235號 (4) 美國加利福尼亞州94555費蒙特市魁北克村公園5393號
	姓 名 (名稱)	美商維博泰克有限公司
	國 籍	美國
	住、居所 (事務所)	美國加利福尼亞州95134聖荷西贊克路2890號209室
	代 表 人 姓 名	詹姆士·潘

J:YUEH/C05291

裝

訂

線

91. 7. - 8 修正
年 月 日
A8
B8
C8
D8
補充

六、申請專利範圍

2．如申請專利範圍第1項的語音識別系統，其特徵在於所述前端信號處理裝置包括：

用於把所述麥克風裝置產生的電信號在頻譜上展平的預加重器裝置；

耦合於所述預加重器裝置的用於把電信號分成帶有由M個採樣分開的相鄰的幀的N個採樣的幀的幀分塊裝置；

耦合於所述幀分塊裝置的用於對各個幀開窗口的開窗裝置，

耦合於所述開窗裝置的用於使幀自相關的自相關裝置；

耦合於所述自相關裝置的用於把各幀轉換成倒頻系數的倒頻系數產生裝置；及

耦合於所述倒頻系數產生裝置的用於加權倒頻系數的錐形開窗裝置，從而產生聲音波的參數表示。

3．如申請專利範圍第2項的語音識別系統，其特徵在於所述前端信號處理裝置還包括耦合於所述錐形開窗裝置的用於產生倒頻系數的一次導數的時間差分裝置。

4．如申請專利範圍第2項的語音識別系統，其特徵在於所述前端信號處理裝置還包括耦合於所述錐形開窗裝置的用於產生倒頻系數的二次導數的時間差分裝置。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

經濟部智慧財產局員工消費合作社印製

91.7.1 修正
年 月 日
補充

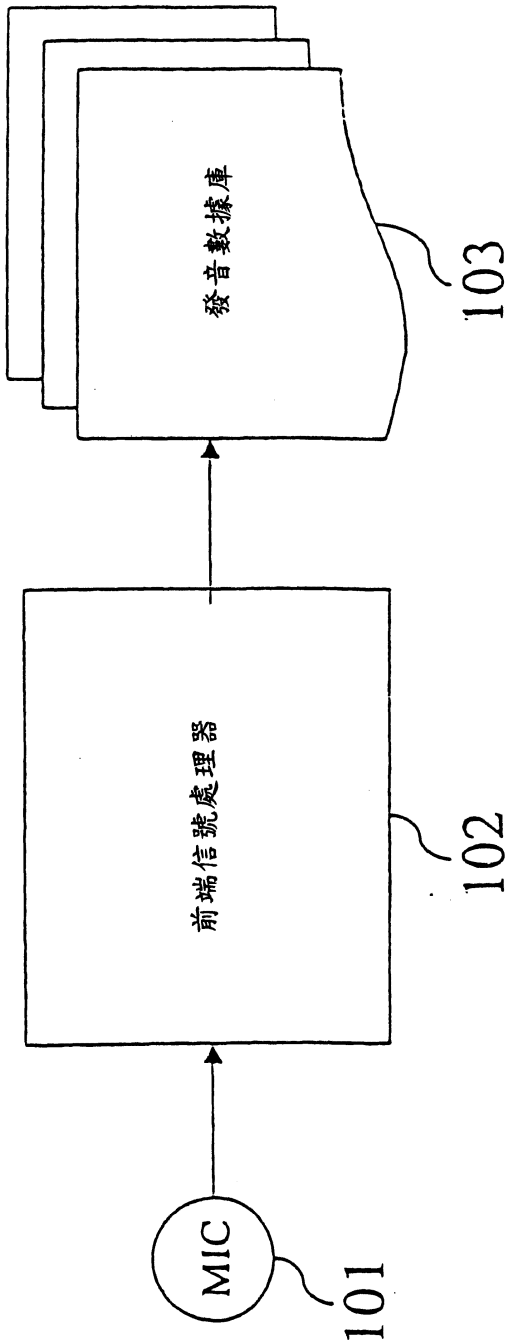


圖 1

正
補
充

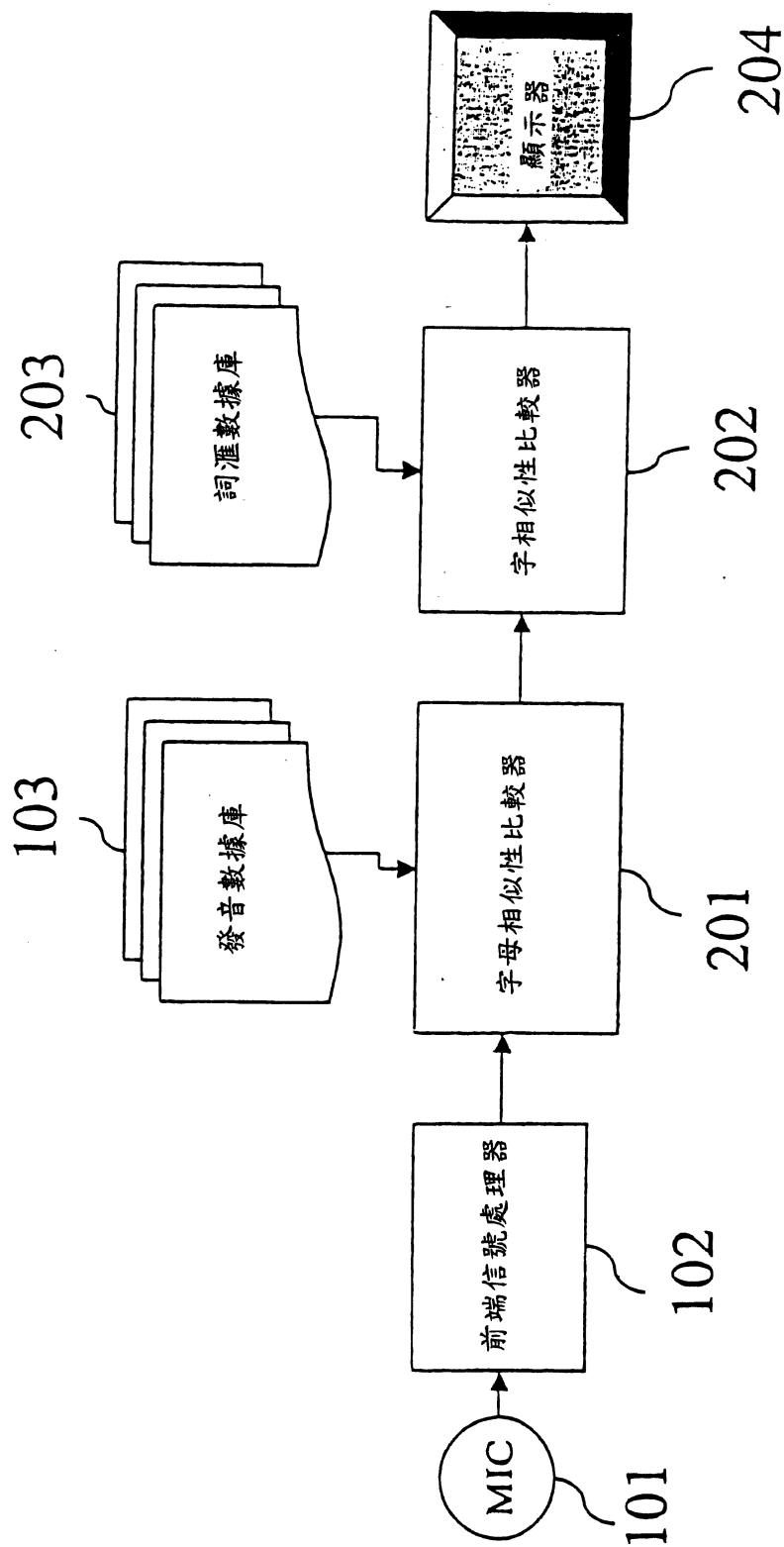


圖 2

91. 7. -8

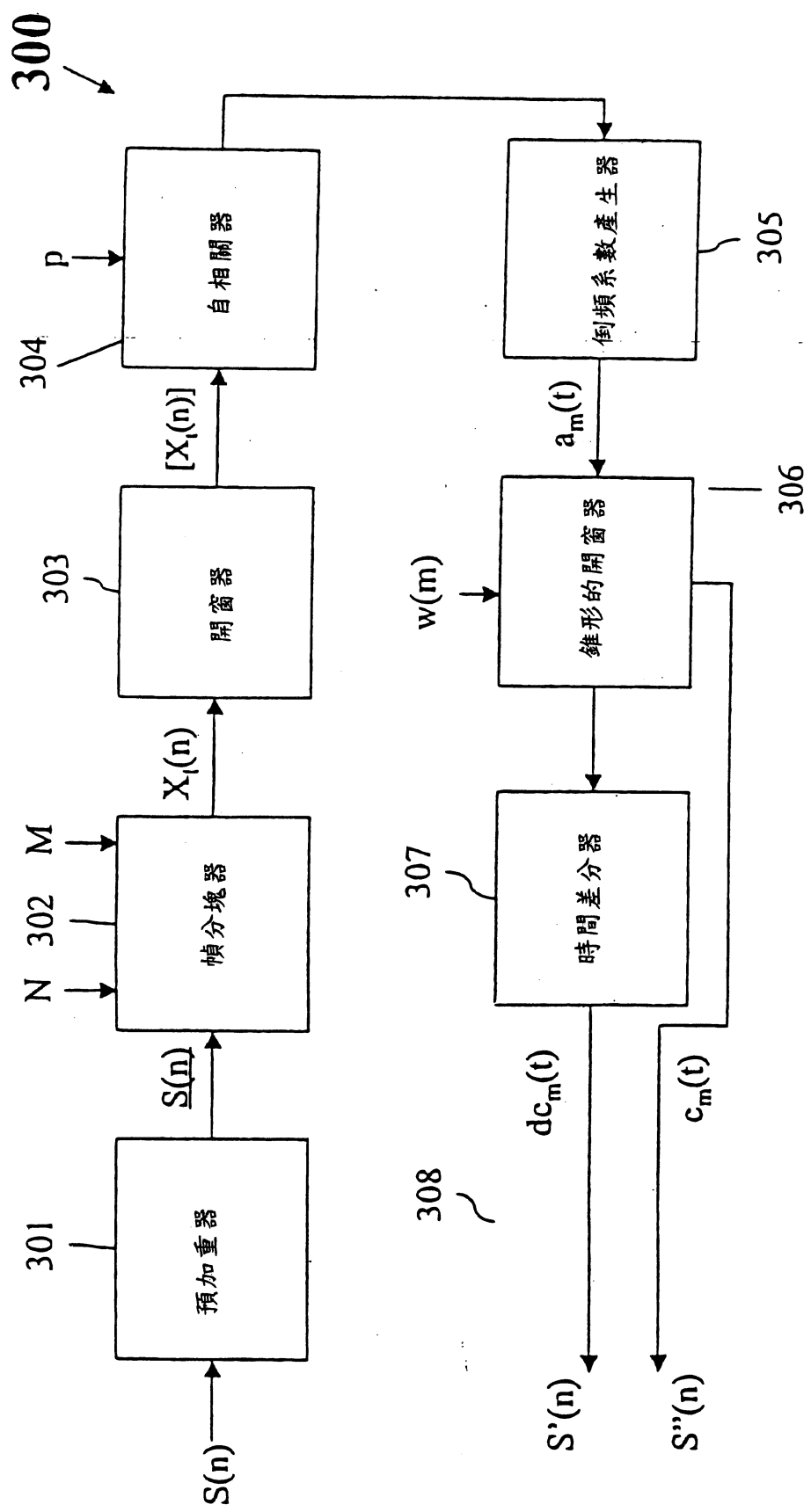


圖 3

91. 7. 48

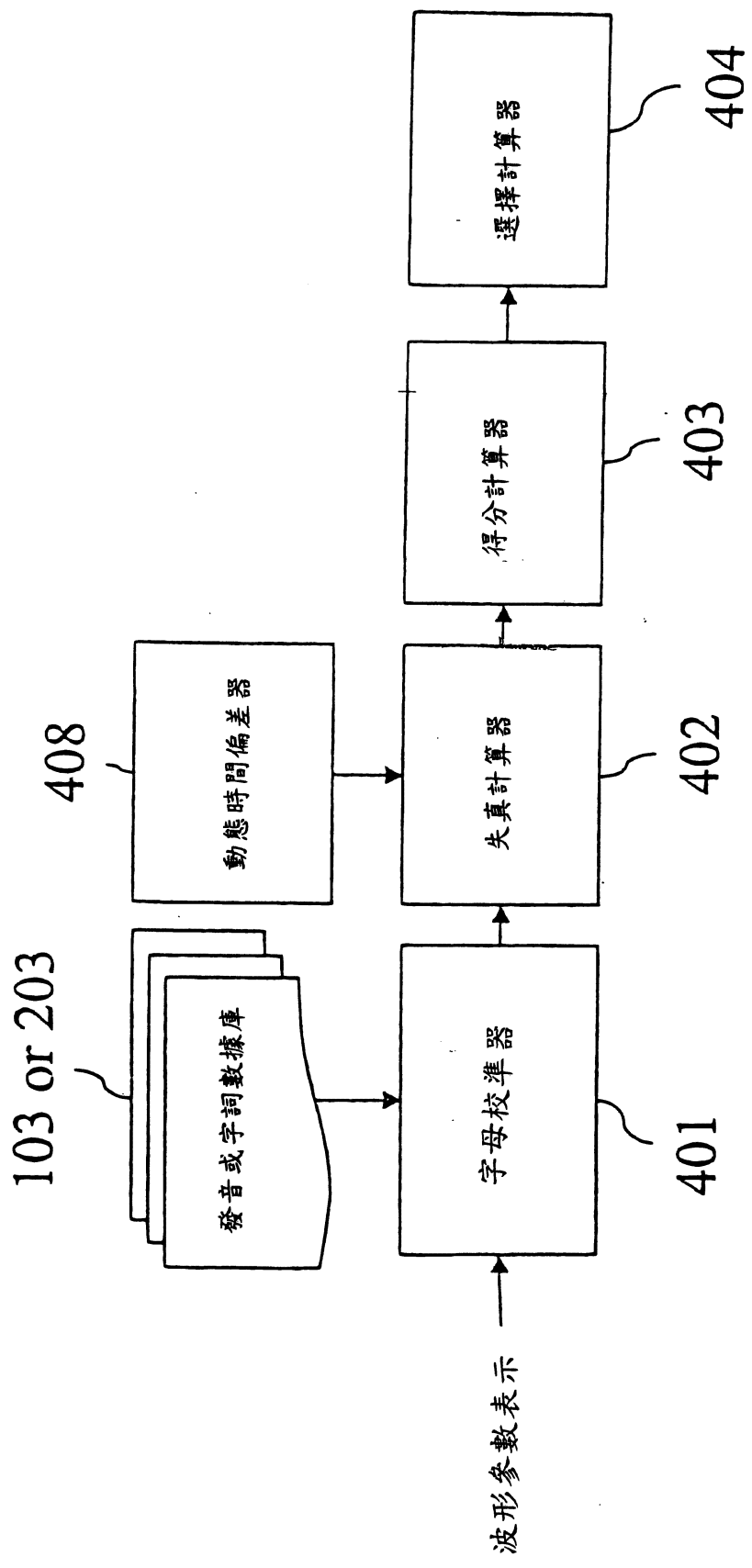


圖 4

修正
年月日 補充

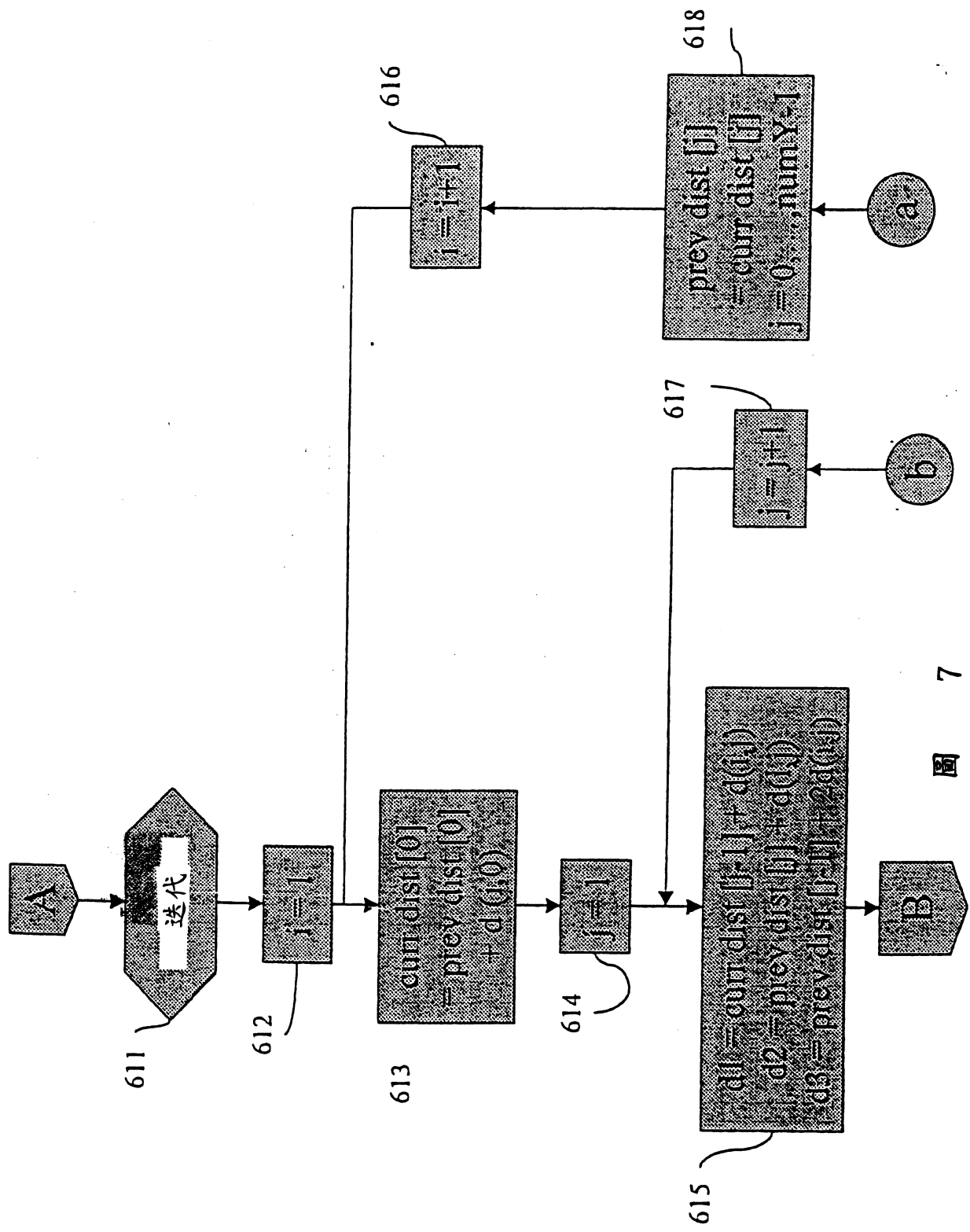


圖 7

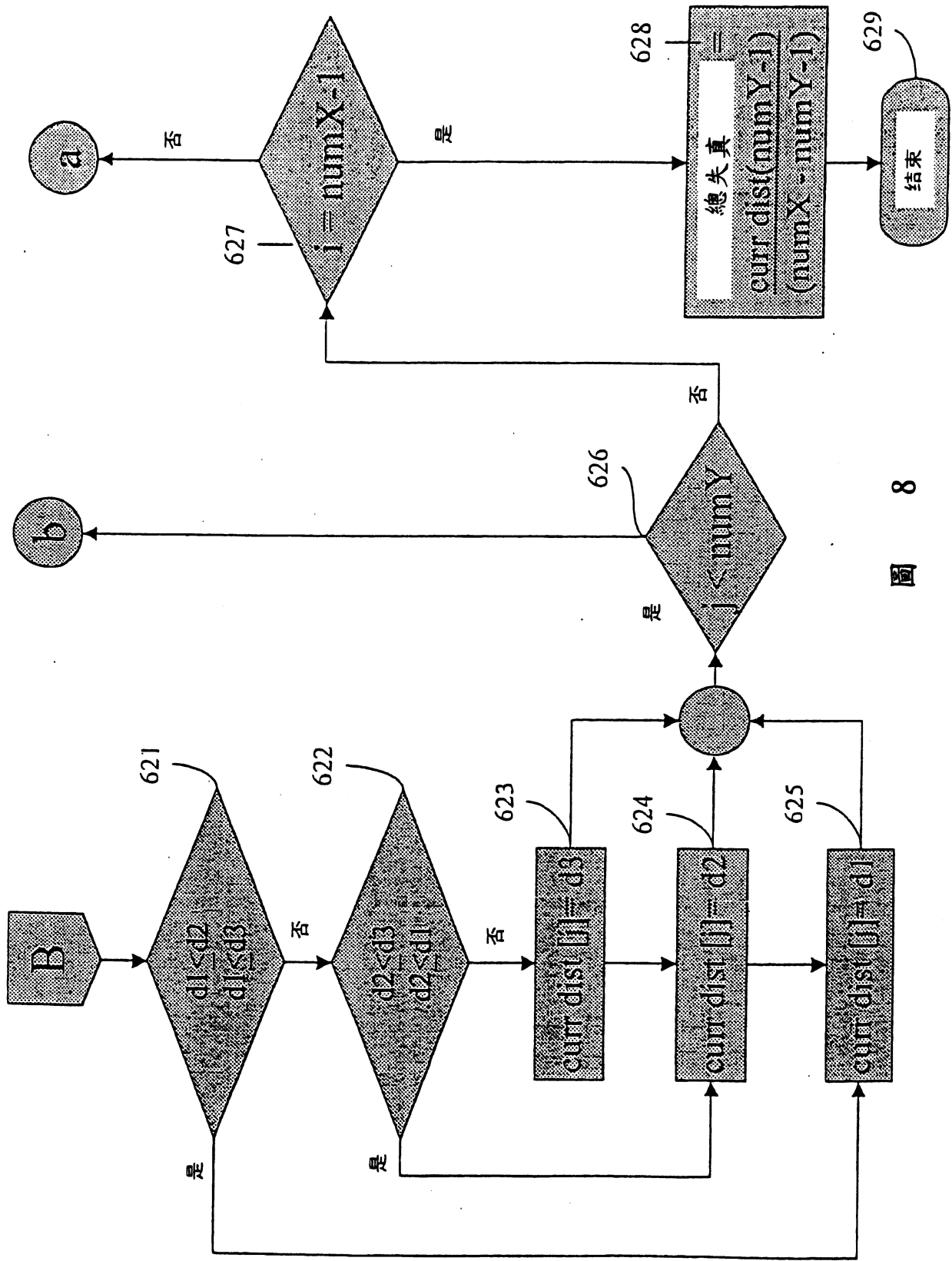


圖 8

91.7. - 8
補正

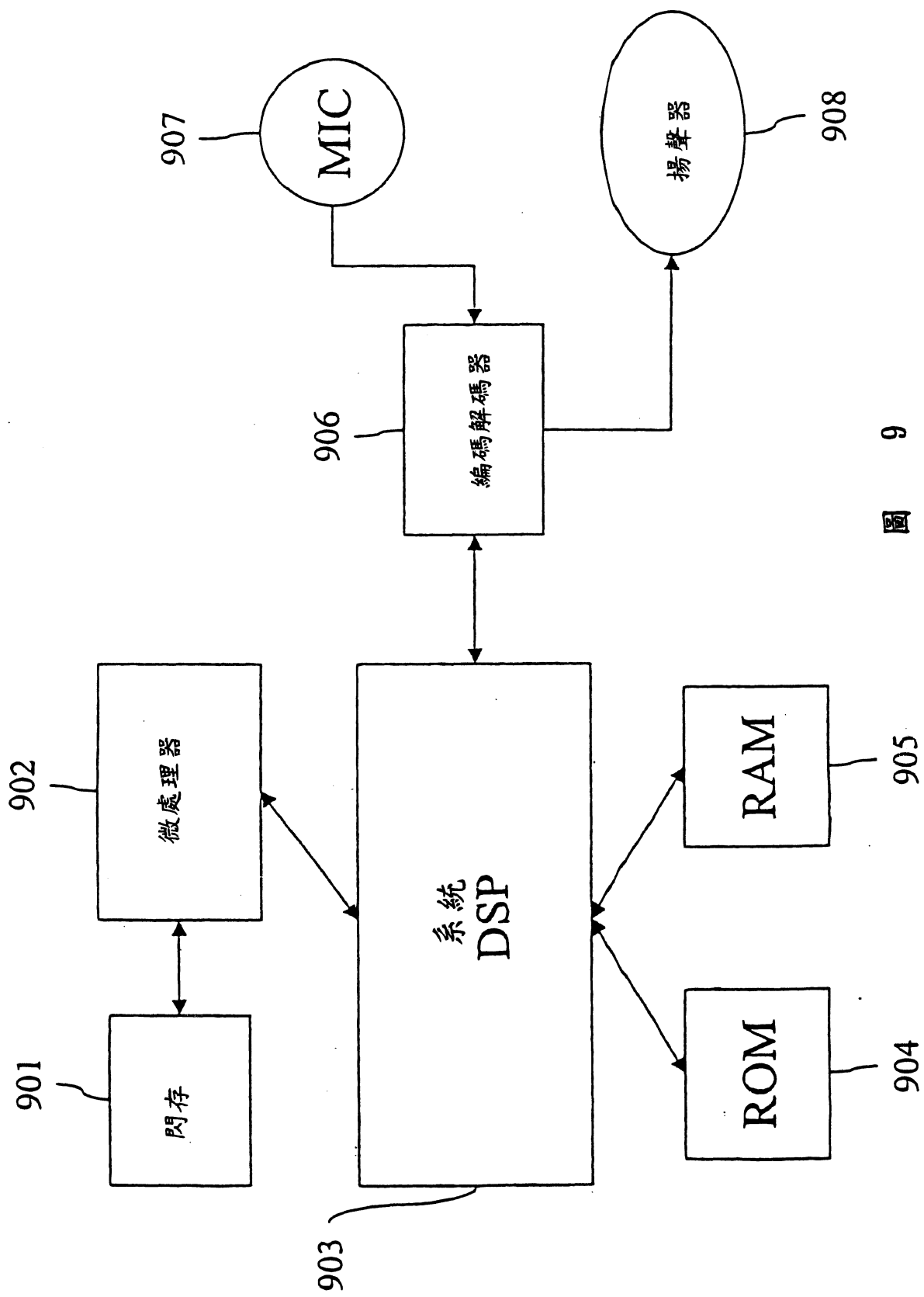


圖 9



圖 10