

申請日期： <u>90 12 25</u>	案號： <u>90132450</u>
類別： <u>G10L 15/00</u>	

(以上各欄由本局填註)

發明專利說明書

541517

一、發明名稱	中文	語音辨識系統
	英文	
二、發明人	姓名 (中文)	1. 王駿發 2. 王家慶 3. 陳泰龍 4. 張錦展
	姓名 (英文)	1. Jhing-Fa Wang 2. Jia-Ching Wang 3. JTai-Lung Chen 4. Chin-Chan Chang
	國籍	1. 中華民國 2. 中華民國 3. 中華民國 4. 中華民國
	住、居所	1. 台南市大學路一號成功大學電機系 2. 台南市大學路一號成功大學電機系 3. 台南市大學路一號成功大學電機系 4. 台南市大學路一號成功大學電機系
三、申請人	姓名 (名稱) (中文)	1. 國立成功大學
	姓名 (名稱) (英文)	1.
	國籍	1. 中華民國
	住、居所 (事務所)	1. 台南市大學路一號
	代表人 姓名 (中文)	1. 高強
	代表人 姓名 (英文)	1.



本案已向

國(地區)申請專利

申請日期

案號

主張優先權

無

有關微生物已寄存於

寄存日期

寄存號碼

無

五、發明說明 (1)

【發明領域】

本發明係有關於一種語音辨識系統，特別是一種以應用導向超大型積體電路架構(ASIC)之方式設計而成，同時以模組化的設計方式分割為四塊模組之語音辨識系統。

【發明背景】

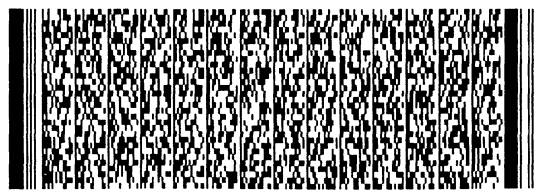
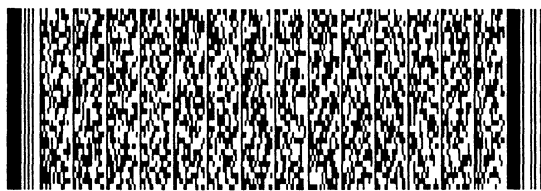
隨著可攜性系統的盛行以及使用語音來做人機的溝通界面日趨重要，將語音處理的演算法實現到硬體上是刻不容緩的。而這些可攜性系統用來處理語音的方式通常以使用單晶片微處理機或數位訊號處理器來達成目的。單晶片微處理機成本低，但運算功能不強難以實現較複雜的語音運算；若使用數位訊號處理器來開發則時間最短、設計彈性度高但成本較高。本發明以應用導向超大型積體電路架構之方式將整個語音辨識的過程(自語音輸入至辨識結果輸出)完整實現，執行速度上比數位訊號處理器來的快，且放到可攜性系統的成本也比數位訊號處理器來的低。

【發明概要】

本發明之主要目的係以應用導向超大型積體電路架構(ASIC)之方式來設計出語音辨識系統，使其執行速度快，且放到可攜性系統的成本也較低。

本發明之另一目的係以模組化的設計方式將該語音辨識系統分割為四塊模組，各模組可自成一個智產元件(IP)，各智產元件可以配合各種不同之商品及應用需求進行設計上的重覆使用，大量縮短上市時間。

為達上述目的，本發明提出一種語音辨識系統，其係以應



五、發明說明 (2)

用導向超大型積體電路架構(ASIC)之方式來設計，輸入為語音訊號，而輸出為辨識的結果，同時以模組化的方式將該系統分為四模組各別進行設計，且該四模組可自成一個智產元件(IP)，該四模組分別為：

系統控制模組，係接受外部訊號控制，進而對內部電路做控制，該外部訊號共有訓練模式及辨識模式兩種；

自相關及線性預測參數模組，其係由自相關參數單元及線性預測參數單元所組成，該兩單元可分別求取自相關參數及線性預測參數，用以將輸入之語音資料轉變成線性預測參數；

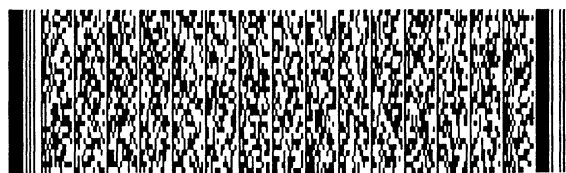
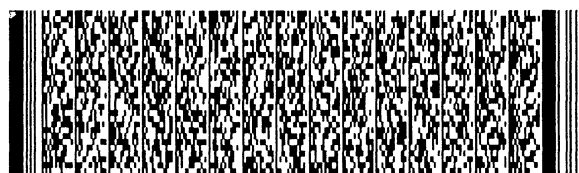
倒頻譜參數模組，其係接受上述線性預測參數，輸出倒頻譜參數，做為語音資料的模型；

辨識模組，係利用動態時間校準(Dynamic Timing Warping, DTW)演算法，係將上述之語音資料的模型做一比對後，輸出辨識分數。

另外，更包括一語音接收模組，係由一組移位暫存器及一組一般性暫存器所構成，其作為外部取樣頻率與內部工作頻率之介面，將接收之語音資料頻率轉換成與內部電路工作頻率一致後，再讀出運算。

另外，更包括一外部記憶體，係用以儲存訓練語音資料的倒頻譜參數。

其中，訓練模式係系統接收訓練語音資料後，通過自相關及線性預測參數模組、倒頻譜參數模組後，將該訓練語音資料的倒頻譜參數存到外部記憶體；而辨識模式係系



五、發明說明(3)

統接收辨識語音資料後，通過自相關及線性預測參數模組、倒頻譜參數模組後，將辨識語音資料的倒頻譜存放在內部暫存器，然後通知辨識模組開始抓取先前存放在外部記憶體的訓練語音資料的倒頻譜參數與內部暫存器的辨識語音資料的倒頻譜參數進行辨識，進而輸出辨識分數，在每一筆的訓練語音資料都得到辨識分數後，交由系統控制模組選擇最佳分數進而輸出辨識結果。

【圖式之簡單說明】

圖一顯示本發明之架構示意圖；

圖二顯示本發明之詳細架構圖；

圖三顯示語音接收模組之電路架構圖；

圖四顯示決定單元之電路架構圖；

圖五顯示自相關參數單元之電路架構圖；

圖六顯示線性預測單元之電路架構圖；

圖七顯示倒頻譜參數模組之電路架構圖；

圖八顯示處理元件的輸入/輸出界面(I/O)之示意圖；

圖九顯示處理元件之內部示意圖；

圖十顯示100x100動態時間校準之晶格示意圖；

圖十一顯示切分過後的動態時間校準之晶格示意圖；

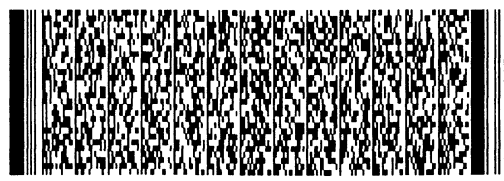
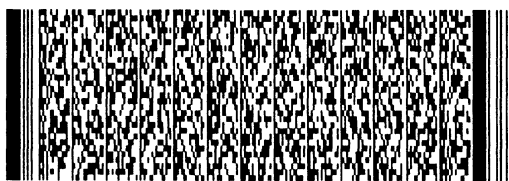
圖十二顯示辨識模組之資料路徑圖；

表一顯示圖十一中前端部份之時序控制。

【主要元件編號】

10 系統控制模組(Control Module)

20 自相關及線性預測參數模組(Autocorrelation&LPC



五、發明說明 (4)

Module)

30 倒頻譜參數模組(Cepstrum Module)

40 辨識模組(DTW Module)

50 外部記憶體(RAM)

60 訓練模式(Training Button)

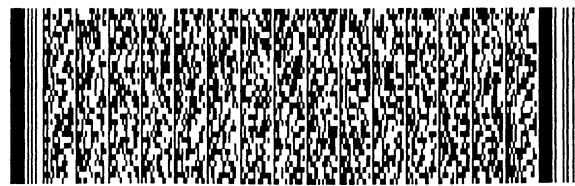
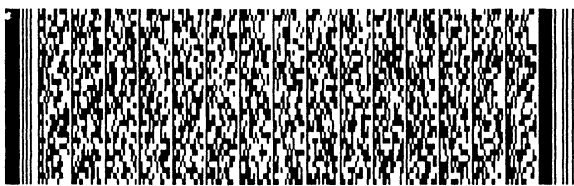
70 辨識模式(Recognition Button)

80 語音訊號(Speech)

90 辨識結果(Recognition Result)

【發明詳細說明】

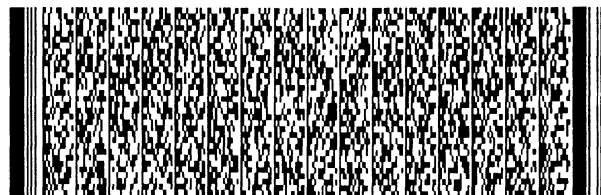
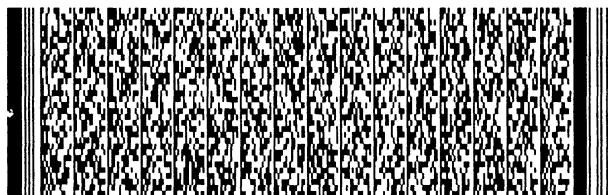
為詳細揭露本發明，以下以一較佳實施例配合圖式做詳細說明。本發明之語音辨識系統，從語音資料輸入麥克風開始，分別要做下列的處理：(1)低通濾波器(Low Pass Filter)，(2)預強調(Pre-emphasis)，(3)取窗處理(Windowing)，(4)自相關係數分析(Autocorrelation Analysis)，(5)端點偵測(Endpoint Detection)，(6)線性預測分析(LPC Analysis)，(7)倒頻譜參數擷取(Cepstrum Extraction)，(8)參考樣本更新(Reference Updating)，(9)動態時間較準(DTW)，(10)決定單元(Decision Logic)。最後而輸出辨識結果(由七段顯示器或LED顯示辨識結果)。整個語音辨識系統用應用導向超大型積體電路架構之方式(ASIC)實現，如圖一所示，係為本發明之架構示意圖，其中輸入為語音訊號(Speech)80，而輸出為辨識的結果(Recognition Result)90，同時以模組化的方式將本發明分為四模組各別進行設計，且該四模



五、發明說明 (5)

組可自成一個智產元件(IP)，各智產元件可以配合各種不同之商品及應用需求進行設計上的重覆使用，大量縮短上市時間；該四模組分別為：系統控制模組(Control Module)10，係接受外部訊號控制，進而對內部電路做控制，該外部訊號共有訓練模式(Training Button)60及辨識模式(Recognition Button)70兩種；自相關及線性預測參數模組(Autocorrelation&LPC Module)20，其係由自相關參數單元及線性預測參數單元所組成，該兩單元可分別求取自相關參數及線性預測參數，用以將輸入之語音資料轉變成線性預測參數；倒頻譜參數模組(Cepstrum Module)30，其係接受上述線性預測參數，輸出倒頻譜參數，做為語音資料的模型；辨識模組(DTW Module)40，係利用動態時間校準(Dynamic Timing Warping, DTW)演算法，係將上述之語音資料的模型做一比對後，輸出辨識分數。另外，還包括一語音接收模組，係由一組移位暫存器及一組一般性暫存器所構成，其作為外部取樣頻率與內部工作頻率之介面，將接收之語音資料頻率轉換成與內部電路工作頻率一致後，再讀出運算；以及一外部記憶體(RAM)50，係用以儲存訓練語音資料的倒頻譜參數。

其中，上述(1)低通濾波器(Low Pass Filter)是放在晶片外部，用離散元件實現。其他四個模組的分工如下：系統控制模組10處理(8)參考樣本更新(Reference Updating)以及(10)決定單元(Decision Logic)，自相關及線性預測參數模組20處理(2)預強調(Pre-emphasis)、



五、發明說明 (6)

(3) 取窗處理(Windowing)、(4) 自相關係數分析(Autocorrelation Analysis)、(5) 端點偵測(Endpoint Detection)以及(6) 線性預測分析(LPC Analysis)，倒頻譜參數模組30處理(7) 倒頻譜參數擷取(Cepstrum Extraction)，辨識模組40處理(9) 動態時間較準(DTW)。

當本發明語音辨識系統接受到不同控制訊號時，系統會根據不同控制訊號情形而有不同的處理模式。共有二種模式，介紹如下：(1) 訓練模式：系統接收訓練語音資料，通過自相關及線性預測參數模組20、倒頻譜參數模組後30，將訓練語音資料的倒頻譜參數存到外部記憶體50。

(2) 辨識模式：系統接收辨識語音資料，通過自相關及線性預測參數模組20、倒頻譜參數模組30後，將辨識語音資料的倒頻譜存放在內部暫存器，然後通知辨識模組40開始抓取先前存放在外部記憶體50的訓練語音資料的倒頻譜參數與內部暫存器的辨識語音資料的倒頻譜參數進行辨識，進而輸出辨識分數。在每一筆的訓練語音資料都得到辨識分數後，交由系統控制模組10選擇最佳分數進而輸出辨識結果90。

本發明之細部結構如圖二所述，茲分述如下：

【語音接收模組】

整體系統包含兩種頻率，一是外部之取樣頻率，一是內部之工作頻率，由於兩者頻率不同，我們設計了一個語音接收模組，其並可當作二種不同頻率的介面。舉例而言，倘若外部語音取樣頻率是8KHz，內部電路的操作頻率

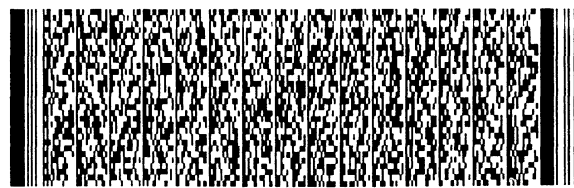
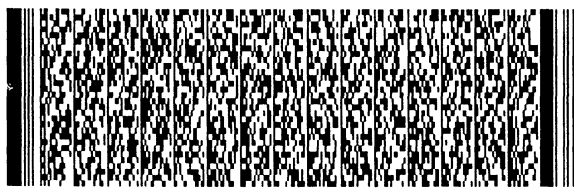


五、發明說明 (7)

是10MHz，則此語音接收模組，一方面以8KHz的頻率接收語音資料，一方面以10MHz的頻率將語音資料讀出運算。該語音接收模組之電路架構如圖三所示。

【系統控制模組】

系統控制模組10，共分為二大部份，一個是外部取樣頻率（以8KHz為例）的有限狀態機（FSM），另一個是內部工作頻率（以10MHz為例）的有限狀態機。二個部份都分別受到外部訊號的控制，進而對內部電路做控制。外部取樣頻率的有限狀態機主要控制語音接收模組。當語音資料收集完成，通知自相關及線性預測參數模組20開始讀取語音資料做運算。內部工作頻率的有限狀態機主要控制外部記憶體50及內部暫存器的讀寫，以及連繫各模組間的溝通。當倒頻譜參數模組30發出完成訊號時，依據操作模式的不同而控制記憶體或暫存器將語音參數寫入：當訓練模式時，將語音參數寫入外部記憶體50而即告完成；當辨識模式時，將語音參數寫入內部暫存器後，通知辨識模組40進行比對動作。當每一筆辨識分數由辨識模組40輸出時，再進而通知決定單元進行運算，最後輸入辨識結果90。決定單元是將每一筆由辨識模組40輸出的辨識分數，做一個最佳分數的判斷。圖四即為決定單元的電路架構，決定單元主要由最小選擇電路及一個計數器構成，當開始訊號啟動時，跟著辨識分數輸入，和目前最小的辨識分數做最小選擇，如果輸入的辨識分數比目前最小的辨識分數還小時，則將輸入的辨識分數寫入A暫存器，並將目前計數器



五、發明說明(8)

的計數值寫入B暫存器，而此處計數器是受開始訊號控制的。當所有的辨識分數輸入後，則B暫存器的值即為辨識結果。

【自相關及線性預測參數模組】

自相關及線性預測參數模組20分為二大單元：自相關參數單元及線性預測參數單元，分別求取自相關參數及線性預測參數。自相關參數單元接收256點的語音資料，輸出十階11個自相關參數。自相關參數單元又分為三大部份：預強調部份、漢明視窗部份及自相關部份，在了解三個部份的觀念後，本發明設計了自相關參數單元的電路架構，如圖五所示，圖中共用到兩個乘法器，上方之乘法器用於將語音資料與漢明視窗相乘，而漢明視窗的值則儲存於漢明視窗唯獨記憶體(Hamming ROM)中。下方之乘法器與一個加法器組合成一組乘法累加電路，可以用循序之方式將十階11個自相關參數依次求出。

線性預測單元係用以求出線性預測參數，其計算有三大部份：累加、除法、參數更新，因此其電路架構如圖六所示，左邊的暫存器A1到A10分別存放自相關參數，右邊的19個位移暫存器則是要存放線性預測參數，下面的暫存器sum則是要累加用，暫存器E是用來存預測誤差的。

【倒頻譜參數模組】

倒頻譜參數模組30係接受線性預測參數，輸出倒頻譜參數。而且面積考量下的情形下，其電路架構如圖七所示，僅運用一組乘法器及加法器，記憶體Lpc_RAM儲存線



五、發明說明 (9)

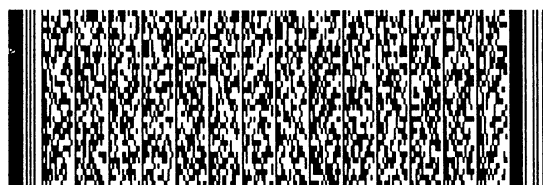
性預測參數，共有10個，每個以16位元儲存，記憶體 Constant_ROM 儲存45個常數，每個以16位元儲存。記憶體 Cep_RAM 儲存倒頻譜參數，共有10個，每個以16位元儲存。

【辨識模組】

辨識模組40採取動態時間校準 (Dynamic Timing Warping, DTW) 演算法，所以設計辨識單元的重心，就放在二部份：處理元件(Processing Element, PE) 設計與控制的方法。

處理元件所要做的事就是將三個方向的距離累加值做一個最小的選擇，然後計算本點的距離值，並加上之而輸出到本點為止的距離累加值，圖八為處理元件的輸入/輸出界面(I/O)示意圖， $tmp(i-1, j)$ ， $tmp(i-1, j-1)$ ， $tmp(i, j-1)$ 為三個方向的距離， U_j 及 R_i 的倒頻譜參數向量距離值則為的本點的距離值，將三個方向的距離累加值做一個最小的選擇，加上本點的距離值之輸出為 $tmp(i, j)$ 。圖九則為處理元件的內部示意圖， $tmp(i-1, j)$ ， $tmp(i-1, j-1)$ ， $tmp(i, j-1)$ 為三個方向的距離，經圖中Minimum(最小值選擇器)，將三個方向的距離累加值做一個最小的選擇， U_j 及 R_i 則輸入倒頻譜參數向量距離值計算器，輸出的本點距離值與最小值選擇器的輸出同時灌入加法器，加法器輸出便是 $tmp(i, j)$ 。

接下來我們就要來推導 100×100 動態時間校準的實際電路，圖十為晶格示意圖，在此有個公式：如果F為音框



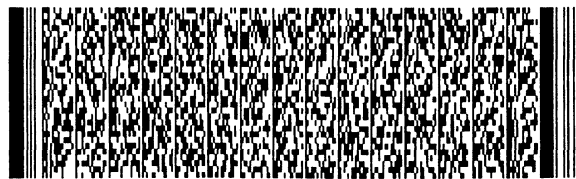
五、發明說明 (10)

數目， W 為扭曲因子(Warping Factor)，則所需要計算的點數為：

$$PEs = F(2W+1) - W(W+1)$$

所以在 100×100 裏，取扭曲因子為15的話，共需要2860次處理元件的計算，這是個很大的數目，所以在控制電路中，需進一步應要去推導規則的控制電路較為適當。首先，將2860點切成的前端部份(0~135)、中間部份(136~2754)、後端部份(2755~2859)三部份，如圖十一所示，前端部份如果將各時間點要計算時，所需要的三方向距離累加值時間點做一表整理時即如表一所示，表一代表各個時間點需要用到左方、左下方、下方哪一方向的距離累加值，及此累加值需從第幾個移位暫存器中獲得，從表一中可觀察得知，前端部份可分為16組，第 n 組有 n 點，在此 n 點裏，第1點只需要下方的距離值累積，用到第 $n-1$ 個移位暫存器的值，第 n 點只需要左方的距離值累積，用到第 n 個移位暫存器的值，其餘則左方的距離值累積用到第 n 個移位暫存器的值，左下方的距離值累積用到第 $2(n-1)$ 個移位暫存器的值，下方的距離值累積用到第 $n-1$ 個移位暫存器的值。中間部份及後端部份也可以採用類似方法尋找出其規則。

圖十二顯示辨識模組的資料路徑，圖中左上方為減法絕對值累加電路，因為要做十階倒頻譜參數的相減取絕對值運算，減法器用以求出本點的距離值，加法器用以將距

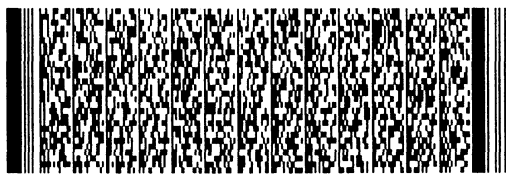


五、發明說明 (11)

離累加值加上本點的距離值。下方為移位暫存器的部份，運用31個暫存器大小之移位暫存器儲存過去的距離累加值。右上方為取左、左下、下三個距離值累積的最小電路，運用3個多工器分別從移位暫存器中選擇出此時需要之值，再經圖中Minimum(最小值選擇器)，將三個方向的距離累加值做一個最小的選擇。

綜上所述，本發明所提供之語音辨識系統，不僅可達預期之實用功效外並且為前所未見之新設計，已符合專利法發明之要件，依法具文申請之。為此，謹請貴審查委員詳予審查，並祈早日賜準專利，至感德便。

以上已將本發明詳細說明，惟以上所述者，僅為本發明之較佳實施例而已，當不能限定本發明實施之範圍，即凡依本發明申請專利範圍所作之均等變化與修飾等，皆應仍屬本發明之專利涵蓋範圍意圖保護之範疇。



四、中文發明摘要 (發明之名稱：語音辨識系統)

本發明提出一個具有訓練模式及辨識模式的完整語音辨識系統，整個系統是以應用導向超大型積體電路架構(ASIC)之方式設計而成，同時以模組化的設計方式將語音訊號處理分割為四塊模組：系統控制模組、自相關及線性預測參數模組、倒頻譜參數模組及辨識模組，各模組可自成一個智產元件(IP)。各智產元件可以配合各種不同之商品及應用需求進行設計上的重覆使用(Design Reuse)，大量縮短上市時間(Time to Market)。

英文發明摘要 (發明之名稱：)



六、申請專利範圍

1. 一種語音辨識系統，其係以應用導向超大型積體電路架構(ASIC)之方式來設計，輸入為語音訊號，而輸出為辨識的結果，同時以模組化的方式將該系統分為四模組各別進行設計，且該四模組可自成一個智產元件(IP)，該四模組分別為：

系統控制模組，係接受外部訊號控制，進而對內部電路做控制，該外部訊號共有訓練模式及辨識模式兩種；

自相關及線性預測參數模組，其係由自相關參數單元及線性預測參數單元所組成，該兩單元可分別求取自相關參數及線性預測參數，用以將輸入之語音資料轉變成線性預測參數；

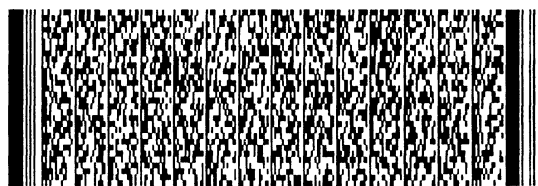
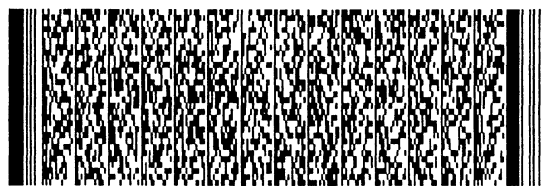
倒頻譜參數模組，其係接受上述線性預測參數，輸出倒頻譜參數，做為語音資料的模型；

辨識模組，係利用動態時間校準(Dynamic Timing Warping, DTW)演算法，係將上述之語音資料的模型做一比對後，輸出辨識分數。

2. 如申請專利範圍第1項所述之語音辨識系統，更包括一語音接收模組，係由一組移位暫存器及一組一般性暫存器所構成，其作為外部取樣頻率與內部工作頻率之介面，將接收之語音資料頻率轉換成與內部電路工作頻率一致後，再讀出運算。

3. 如申請專利範圍第1項所述之語音辨識系統，更包括一外部記憶體，係用以儲存訓練語音資料的倒頻譜參數。

4. 如申請專利範圍第1項所述之語音辨識系統，其中訓練

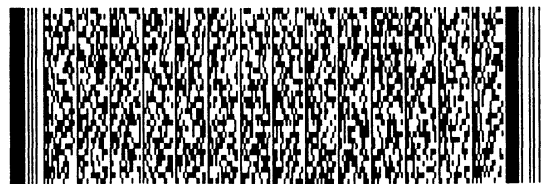
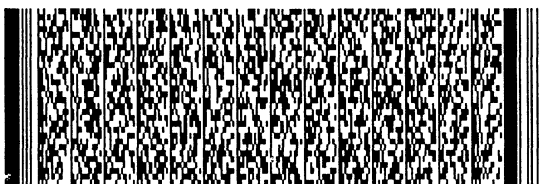


六、申請專利範圍

模式係系統接收訓練語音資料後，通過自相關及線性預測參數模組、倒頻譜參數模組後，將該訓練語音資料的倒頻譜參數存到外部記憶體；而辨識模式係系統接收辨識語音資料後，通過自相關及線性預測參數模組、倒頻譜參數模組後，將辨識語音資料的倒頻譜存放在內部暫存器，然後通知辨識模組開始抓取先前存放在外部記憶體的訓練語音資料的倒頻譜參數與內部暫存器的辨識語音資料的倒頻譜參數進行辨識，進而輸出辨識分數，在每一筆的訓練語音資料都得到辨識分數後，交由系統控制模組選擇最佳分數進而輸出辨識結果。

5. 如申請專利範圍第1項所述之語音辨識系統，其中系統控制模組包括一外部取樣頻率的有限狀態機（FSM）、一內部工作頻率的有限狀態機以及一決定單元；該兩個有限狀態機都分別受到外部訊號的控制，進而對內部電路做控制；該外部取樣頻率的有限狀態機主要控制語音接收模組，當語音資料收集完成，通知自相關及線性預測參數模組開始讀取語音資料做運算；該內部工作頻率的有限狀態機主要控制外部記憶體及內部暫存器的讀寫，以及連繫各模組間的溝通，當倒頻譜參數模組發出完成訊號時，依據操作模式的不同而控制記憶體或暫存器將語音參數寫入；該決定單元係將每一筆由上述辨識模組輸出的辨識分數進行運算後，將最佳分數取出。

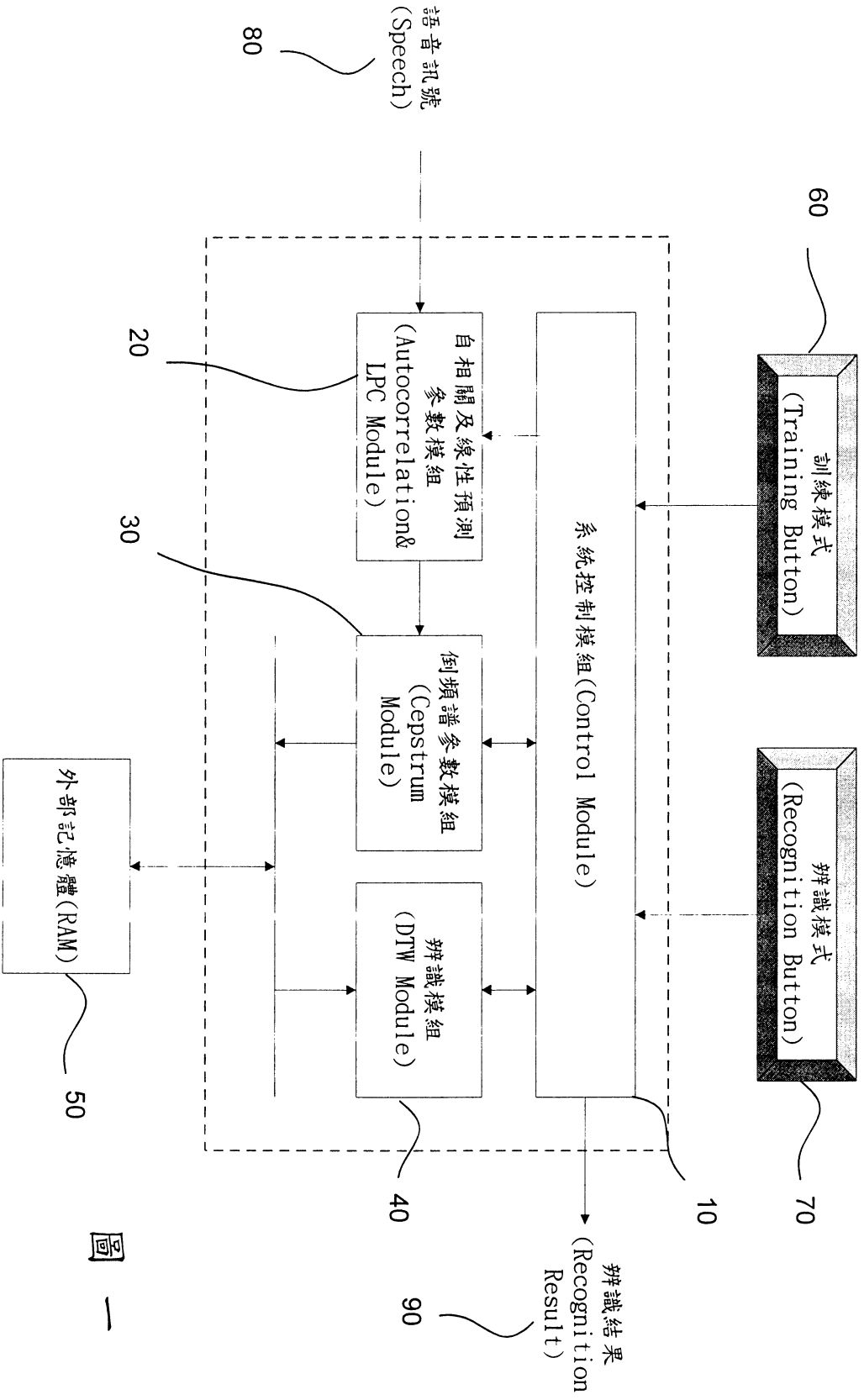
6. 如申請專利範圍第5項所述之語音辨識系統，其中決定單元係由最小選擇電路及一計數器所構成。



六、申請專利範圍

7. 如申請專利範圍第1項所述之語音辨識系統，其中自相關參數單元至少包括兩個乘法器，第一乘法器用於將語音資料與漢明視窗相乘，而漢明視窗的值則儲存於漢明視窗唯獨記憶體(Hamming ROM)中；第二乘法器與一個加法器組合成一組乘法累加電路，可以用循序之方式將自相關參數依次求出。
8. 如申請專利範圍第1項所述之語音辨識系統，其中線性預測參數單元至少包括六組多工器及31個暫存器，其中第1至第10暫存器係分別存放自相關參數，第11至第29暫存器則是存放線性預測參數，第30暫存器則是要累加用，第31暫存器是用來存預測誤差。
9. 如申請專利範圍第1項所述之語音辨識系統，其中倒頻譜參數模組係包括一乘法器、一加法器、一線性預測參數記憶體(Lpc_RAM)、一常數記憶體(Constant_ROM)及一倒頻譜參數記憶體(Cep_RAM)；其中該線性預測參數記憶體(Lpc_RAM)係用以儲存線性預測參數，共有10個，每個以16位元儲存；該常數記憶體(Constant_ROM)儲存45個常數，每個以16位元儲存；該倒頻譜參數記憶體(Cep_RAM)儲存倒頻譜參數，共有10個，每個以16位元儲存。
10. 如申請專利範圍第1項所述之語音辨識系統，其中辨識模組係至少包括一減法絕對值累加電路、31個暫存器、3個多工器及一最小值選擇器。





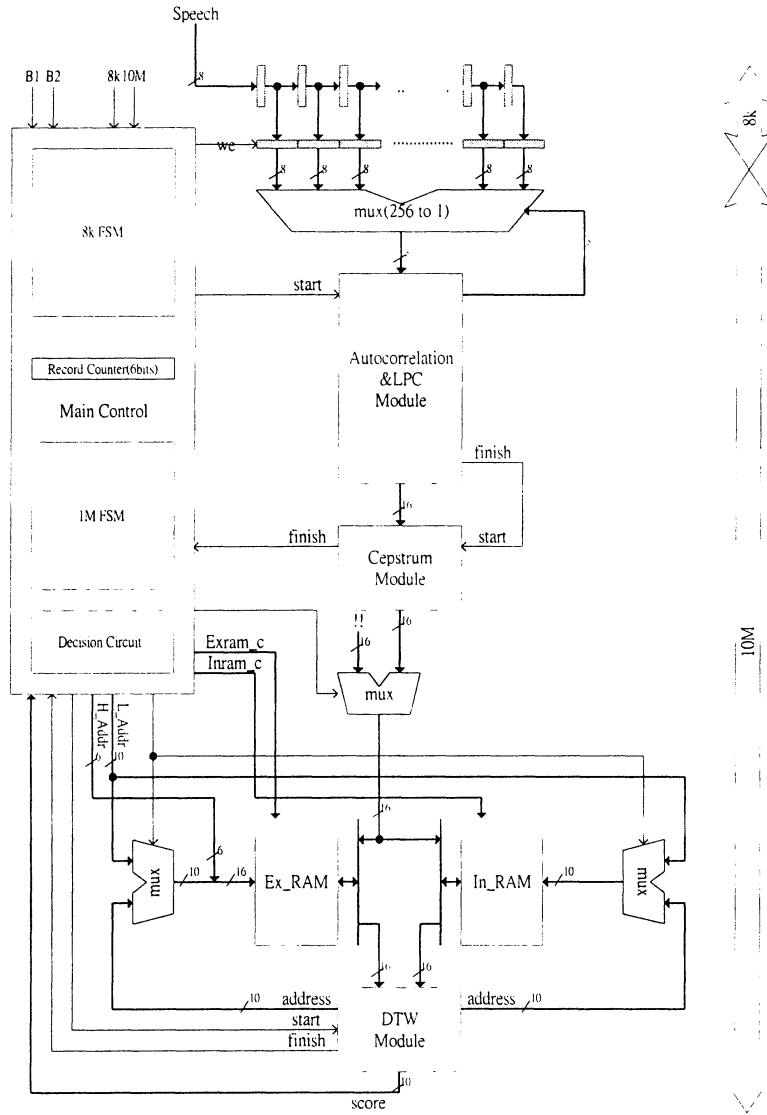


圖 二

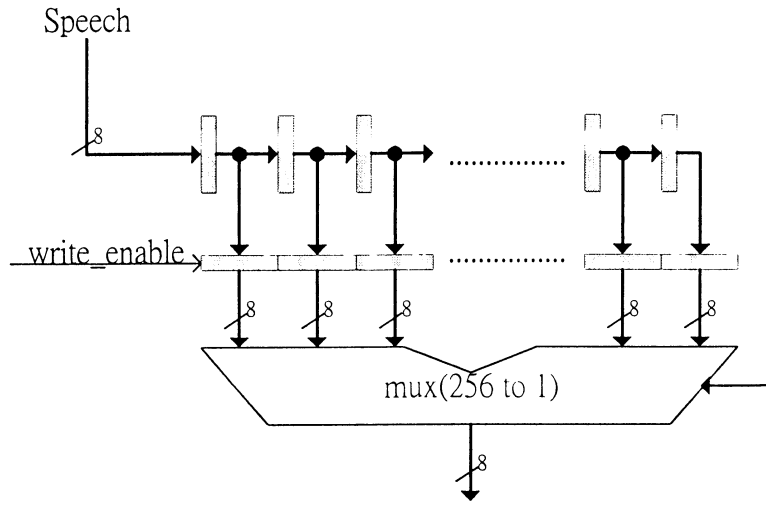


圖 三

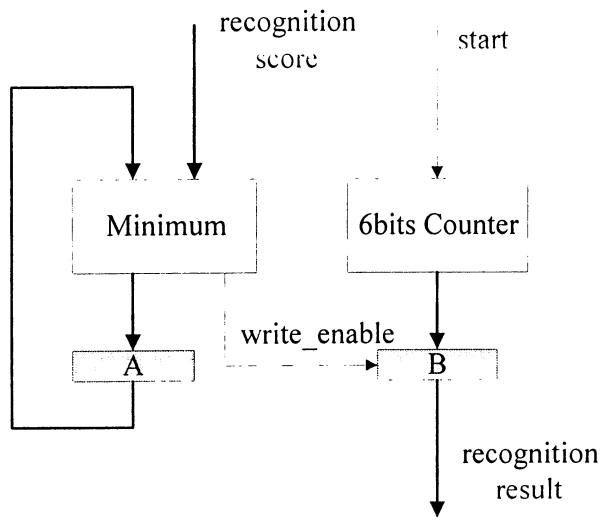


圖 四

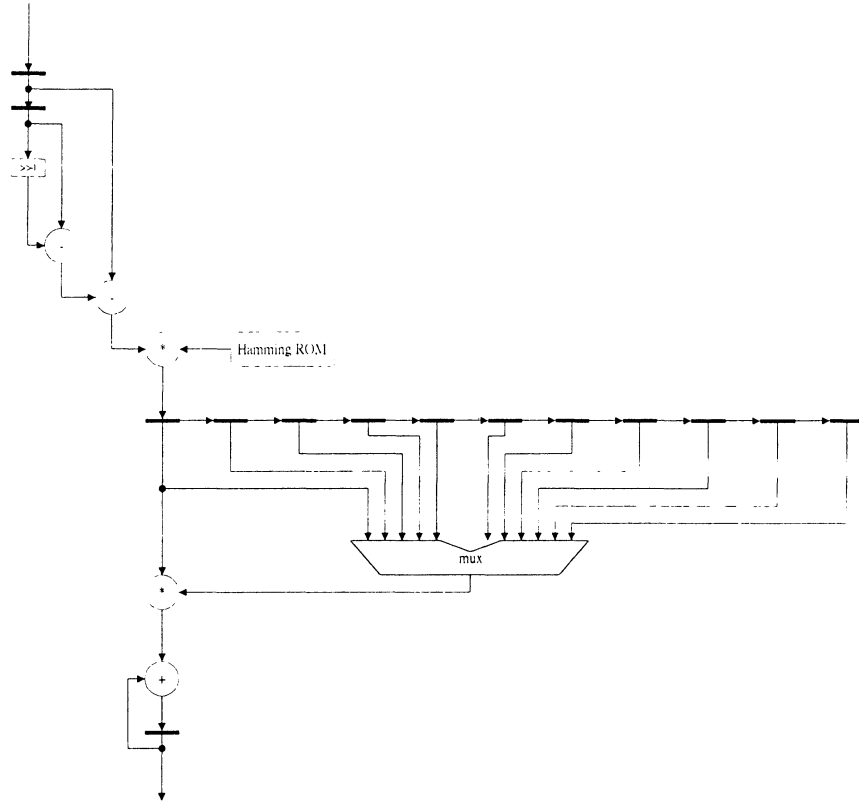


圖 五

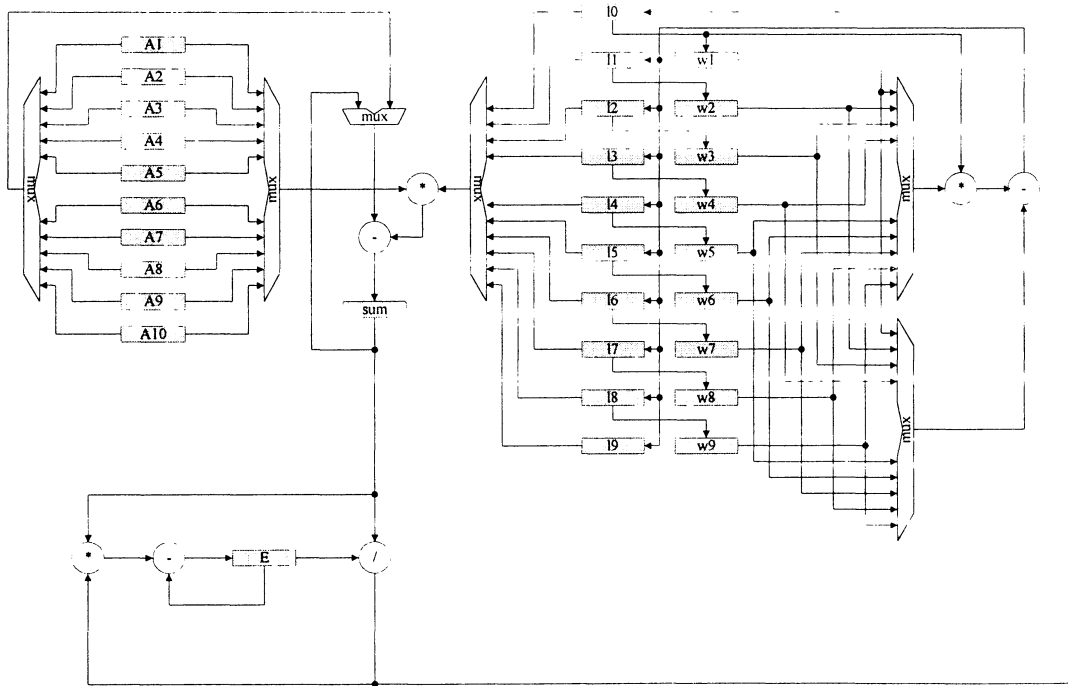


圖 六

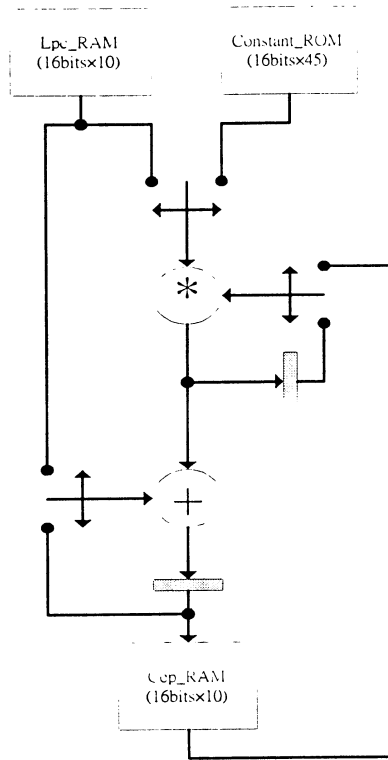


圖 七

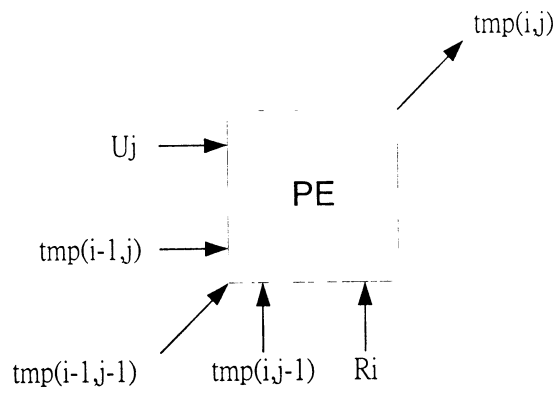


圖 八

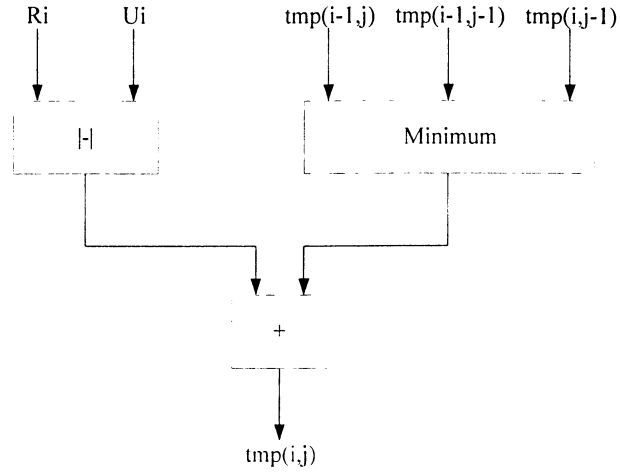


圖 九

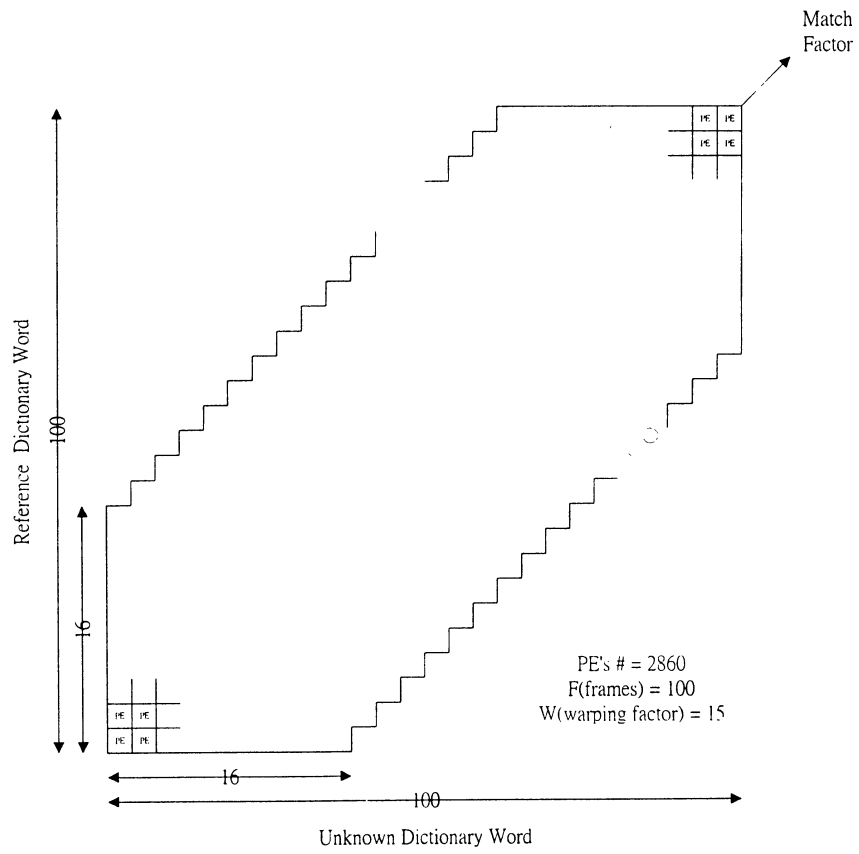


圖 十

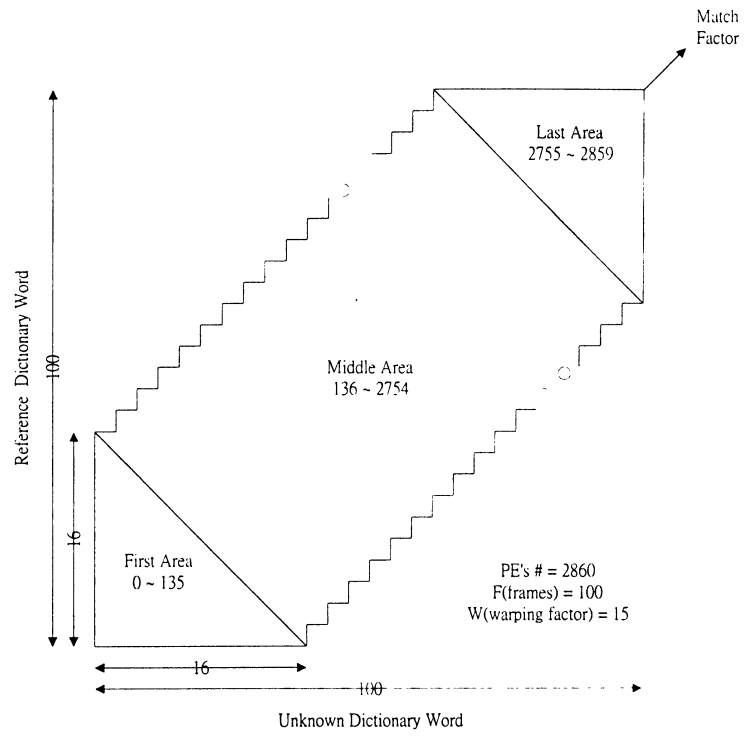


圖 十一

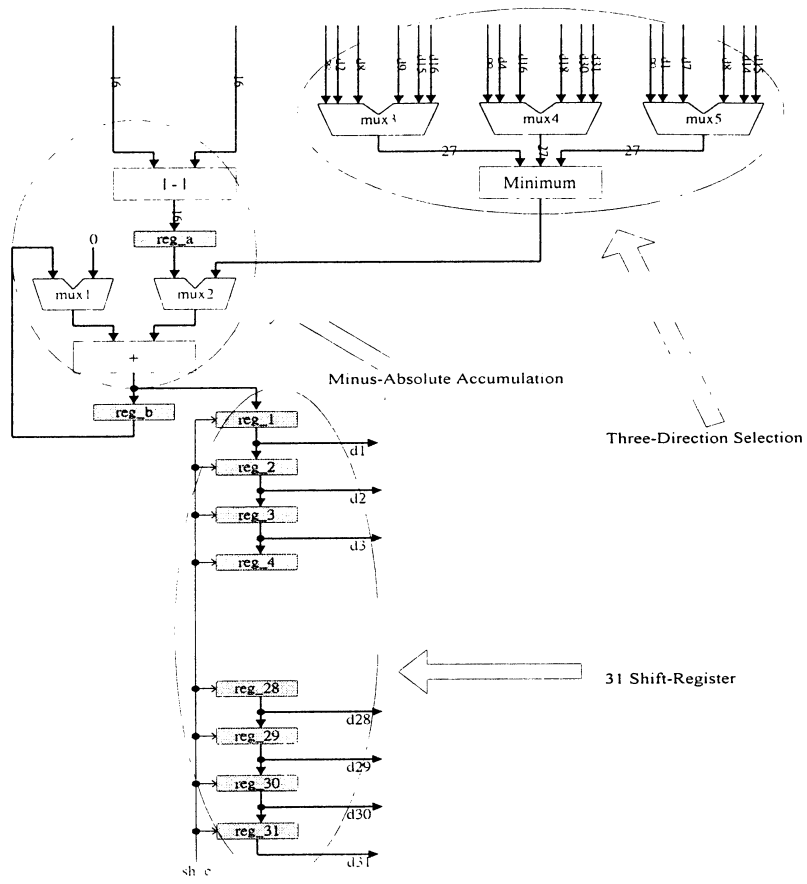


圖 十二