



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I752437 B

(45)公告日：中華民國 111 (2022) 年 01 月 11 日

(21)申請案號：109108337

(22)申請日：中華民國 109 (2020) 年 03 月 13 日

(51)Int. Cl. : G06F3/16 (2006.01)

G10L15/02 (2006.01)

G10L17/26 (2013.01)

(71)申請人：宇康生科股份有限公司 (中華民國) (TW)

臺北市士林區中山北路 5 段 520 號 7 樓

(72)發明人：林海興 LIN, HAI-HSING (TW)；張嘉原 CHANG, CHIA-YUAN (TW)；何冠旻 HO, GUAN-MIN (TW)；陳豫邦 CHEN, YU-BANG (TW)；翁恪誠 WENG, KO-CHENG (TW)；劉峻宇 LIU, CHUN-YU (TW)；林廷容 LIN, TING-JUNG (TW)；曾佳玉 ZENG, JIA-YU (TW)

(74)代理人：高玉駿；楊祺雄

(56)參考文獻：

CN 102272827A

CN 106463113B

US 2018/0348970A

US 2019/0339936A1

審查人員：林琮烈

申請專利範圍項數：7 項 圖式數：8 共 32 頁

(54)名稱

基於至少雙音素的語音輸入操作方法及電腦程式產品

(57)摘要

一種基於至少雙音素的語音輸入操作方法包含，編碼多個參考音素，以定義出多個分別相關聯於多個操作選項的音素標籤，其每一者至少包含選自該等參考音素的第一及第二參考音素；當根據已儲存且與該等參考音素相關聯的語音辨識資料或用戶的個人音素資料且利用語音或聲紋辨識技術確認出收集自用戶的語音信號所含的第一及第二音素分別相似於第一及第二目標參考音素後，自該等音素標籤決定出目標音素標籤，其所含的該第一及第二參考音素分別相同於該第一及第二目標參考音素；及將該等操作選項其中一個與該目標音素標籤相關聯的目標操作選項激活。

指定代表圖：

符號簡單說明：
S20-S28:步驟

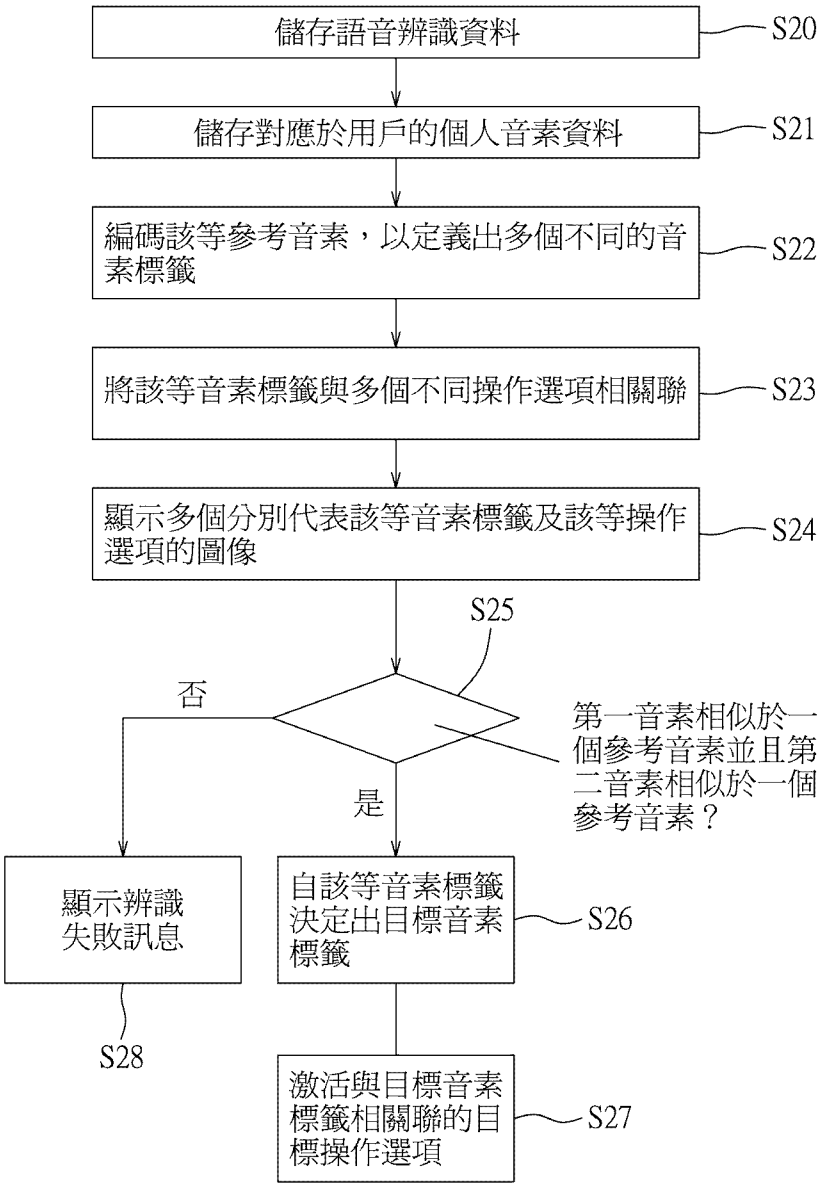


圖 2



公告本

I752437

【發明摘要】

【中文發明名稱】 基於至少雙音素的語音輸入操作方法及電腦程式產品

【中文】

一種基於至少雙音素的語音輸入操作方法包含，編碼多個參考音素，以定義出多個分別相關聯於多個操作選項的音素標籤，其每一者至少包含選自該等參考音素的第一及第二參考音素；當根據已儲存且與該等參考音素相關聯的語音辨識資料或用戶的個人音素資料且利用語音或聲紋辨識技術確認出收集自用戶的語音信號所含的第一及第二音素分別相似於第一及第二目標參考音素後，自該等音素標籤決定出目標音素標籤，其所含的該第一及第二參考音素分別相同於該第一及第二目標參考音素；及將該等操作選項其中一個與該目標音素標籤相關聯的目標操作選項激活。

【指定代表圖】：圖（2）。

【代表圖之符號簡單說明】

S20-S28...步驟

【發明說明書】

【中文發明名稱】 基於至少雙音素的語音輸入操作方法及電腦程式產品

【技術領域】

【0001】 本發明是有關於語音輸入，特別是指一種基於至少雙音素的語音輸入操作方法及電腦程式產品。

【先前技術】

【0002】 語音輸入功能已廣泛用來取代繁冗的手動輸入。在應用上，電腦裝置通常可將收集到用戶所發出的語音經由利用語言模型及聲學模型的語音辨識引擎成功辨識出與該語音有關的字詞、操作指令或應用程式後，該電腦裝置便可將有關的字詞顯示出來，或者執行有關的指令或應用程式。

【0003】 然而，對於構音障礙患者而言，其往往無法發出特定的語音，而所發出語音常有含混不清、沙啞、單調、斷續、發聲音量過大或其他異常的特徵。在此情況下，由於現有的語音辨識技術無法成功辨識構音障礙患者所發出的語音，致使構音障礙患者無法使用現有語音輸入操作方式，例如作為語音溝通設備的平板電腦所提供的語音輸入功能，與外界溝通。

【0004】 因此，為了使構音障礙患者能夠利用語音輸入來操作電子裝置或與外界溝通，如何發展出適用於構音障礙患者的語音輸入操作技術遂成為目前重要的議題。

【發明內容】

【0005】 因此，本發明的目的，即在提供一種基於至少雙音素的語音輸入操作方法及電腦程式產品，其能克服現有技術的至少一缺點。

【0006】 於是，本發明所提供的一種基於至少雙音素的語音輸入操作方法，利用一具有語音及聲紋辨識技術的電腦系統來執行，並包含下步驟：(A)儲存與多個彼此不同的參考音素相關聯的語音辨識資料、及對應於一用戶的個人音素資料，該個人音素資料包含由該用戶所發出且分別對應於該等參考音素的多個語音的語音內容；(B)將該等參考音素編碼，以定義出多個彼此不同的音素標籤，其中每一個音素標籤至少包含一選自該等參考音素其中一者的第一參考音素、及一選自該等參考音素其中一者的第二參考音素；(C)將該等音素標籤分別與多個彼此不同的操作選項相關聯；(D)當收集到來自該用戶且至少包含連續的第一音素和第二音素的語音信號後，根據該語音辨識資料且利用語音辨識技術，或者根據該個人音素資料且利用聲紋辨識技術，確認該第一音素是否相似於該等參

考音素其中之一者並確認該第二音素是否相似於該等參考音素其中之一者；(E)當確認出相似於該第一音素的第一目標參考音素、及相似於該第二音素的第二目標參考音素後，根據該第一目標參考音素和該第二目標參考音素，自該等音素標籤決定出一個目標音素標籤，該目標音素標籤所含的該第一參考音素與該第二參考音素分別相同於該第一目標參考音素與該第二目標參考音素；及(F)將該等操作選項其中一個與該目標音素標籤相關聯的目標操作選項激活。

【0007】 本發明的基於至少雙音素的語音輸入操作方法中，每一參考音素為一母音或一音節。

【0008】 本發明的基於至少雙音素的語音輸入操作方法中，每一操作選項為一符號、一字元、一文字內容、一操作指令、一應用程式及一檔案其中一者。當該目標操作選項為一符號時，該電腦系統透過一顯示該符號之方式激活該目標操作選項。當該目標操作選項為一字元時，該電腦系統透過一顯示該字元之方式激活該目標操作選項。當該目標操作選項為一文字內容時，該電腦系統至少透過一顯示該文字內容之方式激活該目標操作選項。當該目標操作選項為一操作指令時，該電腦系統透過一執行該操作指令之方式激活該目標操作選項。當該目標操作選項為一應用程式時，該電腦系統透過一執行該應用程式之方式激活該目標操作選項。當該目標操作選

項為一檔案時，該電腦系統透過一開啟或播放該檔案之方式激活該目標操作選項。

【0009】 本發明的基於至少雙音素的語音輸入操作方法中，在步驟(C)與步驟(D)之間，還包含以下步驟：(G)顯示多個分別代表該等操作選項的圖像，並在該等圖像附近顯示與該等操作選項相關聯的該等音素標籤。

【0010】 本發明的基於至少雙音素的語音輸入操作方法中，當該目標操作選項為一文字內容時，該電腦系統不僅透過顯示該文字內容之方式，還透過一播放一對應於該文字內容的語音內容的方式激活該目標操作選項。

【0011】 本發明的基於至少雙音素的語音輸入操作方法中，該電腦系統包含一用於執行步驟(B)、(C)、(E)、(F)及(G)的使用終端、及一可與該使用終端通訊且用於執行步驟(A)及(D)的辨識伺服器端，還包含以下步驟：在步驟(A)之前，(H)藉由該使用終端，將收集到的該個人音素資料傳送至該辨識伺服器端；在步驟(C)與步驟(D)之間，(I)藉由該使用終端，收集該語音信號，並將一包含該語音信號且有關該用戶的辨識請求送至該辨識伺服器端，以便該辨識伺服器端回應於該辨識請求執行步驟(D)；及在步驟(D)與步驟(E)之間，(J)藉由該辨識伺服器端，當確認出該第一目標參考音素與該第二目標參考音素時，將一含有該第一目標參考音素及該第二目標參考音

素的辨識回覆傳送至該使用終端，以使該使用終端回應於該辨識回覆執行步驟(E)。

【0012】 於是，本發明所提供的一種電腦程式產品儲存在一電腦可讀取媒體，包含多個程式指令，且當一電腦裝置執行該等程式指令時，可完成如以上所述之基於至少雙音素的語音輸入操作方法。

【0013】 本發明之功效在於：由於先儲存了用於語音辨識的語音辨識資料，以及用於聲紋辨識且對應於用戶的個人音素資料，因此不僅對於構音正常的用戶，而且對於構音障礙患者的用戶所發出的有限語音均能被精確地辨識出。此外，基於至少雙音素的編碼方式能定義出相對較多數量的音素標籤，此等音素標籤能被廣泛地應用來與相對較多數量的操作選項建立關聯性。於是，相較於現有利利用相對複雜的語言模型及聲學模型的語音辨識技術，能相對容易且快速地辨識出用戶所發出且含有至少雙音素的語音輸入，以判定出所欲激活的目標操作選項。

【圖式簡單說明】

【0014】 本發明之其他的特徵及功效，將於參照圖式的實施方式中清楚地呈現，其中：

圖 1 是一方塊圖，示例性地繪示用來實施本發明第一實施例的基

於至少雙音素的語音輸入操作方法的電腦系統的一架構；

圖 2 是一流程圖，示例性地說明圖 1 的電腦系統如何實施本發明第一實施例；

圖 3 是一示意圖，示例性地繪示出由該電腦系統的一觸控顯示模組所顯示並含有音素標籤的虛擬鍵盤；

圖 4 至圖 6 是示意圖，示例性地繪示出該電腦系統在不同的使用情況下由該觸控顯示模組所提供並含有音素標籤的顯示視窗；

圖 7 是一方塊圖，示例性地繪示出用來實施本發明第二實施例的基於至少雙音素的語音輸入操作方法的電腦系統的另一架構；及

圖 8 是一流程圖，示例性地說明圖 7 的電腦系統如何實施本發明第二實施例。

【實施方式】

【0015】 在本發明被詳細描述之前，應當注意在以下的說明內容中，類似的元件是以相同的編號來表示。

【0016】 參閱圖1，繪示出的一電腦系統係實施成如智慧型手機或平板電腦的電腦裝置100，其係用來實施本發明第一實施例的基於至少雙音素(Double-phoneme)的語音輸入操作方法，並包含一用於收集來自外部語音的語音收集模組1(例如，麥克風模組)、一用作顯示器和使用操作介面的觸控顯示模組2、一儲存模組3、一

喇叭模組4，以及一電連接該語音收集模組1、該觸控顯示模組2、該儲存模組3和該喇叭模組4的處理單元5。在本實施例中，該處理單元5支援語音及聲紋辨識技術。

【0017】 以下，將參閱圖1及圖2來示例地說明該電腦裝置100如何執行該第一實施例的語音輸入操作方法。大體而言，該語音輸入操作方法可包含以下步驟S20-S28。

【0018】 首先，在步驟S20中，該電腦裝置100將與多個彼此不同的參考音素相關聯的語音辨識資料儲存於該儲存模組3。在本實施例中，每一參考音素可為一母音(Vowel)或音節(Syllable)。更明確地，該電腦裝置100根據收集自多個(構音正常)用戶發出該等參考音素的語音內容並利用例如聲學模型訓練而獲得用於辨識該等參考音素的該語音辨識資料。舉例來說，該等參考音素例如具有四個母音和四個音節，如以下表示1所示。

表1

參考音素							
母音				音節			
a	u	e	o	ha	hu	he	ho

然後，在步驟S21中，通常在一註冊階段，該電腦裝置100儲存對應於一用戶的個人音素資料，該個人音素資料包含由該用戶所發出且分別對應於多個彼此不同的參考音素的多個語音的語音內容。更明確地，在開始語音紀錄之前，該處理單元5例如可使該觸控顯示

模組2顯示該等參考音素，以供該用戶作為有關該等參考音素的發聲的指示，但并不在此限。於是，在語音紀錄期間，該語音收集模組1將收集到由該用戶發出對應於該等參考音素的多個語音的語音內容傳送至該處理單元5，且該處理單元5將該語音內容作為對應於該用戶的該個人音素資料，並將該個人音素資料儲存於該儲存模組3。

【0019】 特別要說明的是，為了紀錄具有可區別性的該等語音以建立符合個人聲紋特性的個人語音資料，該等參考音素的選擇可視該用戶的構音能力而定。換言之，用戶無須發出標準的每一參考音素(即，母音或音節)，只要用戶能發出對應於每一參考音素的語音彼此是可區別的即可。

【0020】 然後，在步驟S22中，該處理單元5將該等參考音素編碼，以定義出多個彼此不同的音素標籤。在本實施例中，每一個音素標籤僅包含一選自該等參考音素其中一者的第一參考音素、及一選自該等參考音素其中一者的第二參考音素。若依照上述表1的範例，該等音素標籤可被定義如以下表2：

表2

音素標籤							
a a	u a	e a	o a	h a a	h u a	h e a	h o a
a u	u u	e u	o u	h a u	h u u	h e u	h o u
a e	u e	e e	o e	h a e	h u e	h e e	h o e
a o	u o	e o	o o	h a o	h u o	h e o	h o o
a h a	u h a	e h a	o h a	h a h a	h u h a	h e h a	h o h a
a h e	u h e	e h e	o h e	h a h e	h u h e	h e h e	h o h e
a h o	u h o	e h o	o h o	h a h o	h u h o	h e h o	h o h o

a hu	u hu	e hu	o hu	ha hu	hu hu	he hu	ho hu
------	------	------	------	-------	-------	-------	-------

【0021】 值得注意的是，在其他實施態樣中，若受限於用戶貧乏的構音能力而導致能發出可區別的參考音素之數量相對較少時，為了可定義出相當數量的音素標籤，該處理單元5定義出的每一音素標籤亦可包含三個或三個以上的參考音素。

【0022】 接著，在步驟S23中，該處理單元5將定義出的該等音素標籤分別與多個彼此不同的操作選項相關聯，並可將該等音素標籤與該等操作選項的關聯關係儲存於該儲存模組3。在本實施例中，每一操作選項可以是一符號、一字元、一文字內容、一操作指令、一應用程式或一檔案，其中的符號、字元和操作指令可以是該觸控顯示模組2所顯示的一虛擬鍵盤所含的任一符號、任一字元和任一操作指令，而其中的應用程式和檔案可以是該觸控顯示模組2所顯示的任一視窗所含應用程式和任一文字、圖形、音頻或多媒體檔案。應注意的是，在實際使用時，該等音素標籤可被應用來與該觸控顯示模組2所提供的不同顯示視窗中的操作選項建立關聯性。

【0023】 之後，在使用該電腦裝置100期間，在步驟S24中，該處理單元5使該觸控顯示模組2當下提供的顯示視窗顯示多個分別代表該等音素標籤及該等操作選項的圖像。更明確地說，將視該觸控顯示模組2當下提供的顯示視窗，會有對應的圖像顯示內容。以下，將就不同使用情況來示例說明。

【0024】 參閱圖3，繪示的示例是一顯示於該觸控顯示模組2提供的一顯示視窗的虛擬鍵盤，此虛擬鍵盤顯示了多個(編輯候選)字元的圖像(如「我」、「好」…等)以及多個分別在該等字元的圖像附近且分別與該等字元相關聯的音素標籤(如「a a」、「a i」…等)、多個(編輯候選)注音符號的圖像(如「ㄅ」、「ㄆ」…等)以及多個分別在該等注音符號的圖像附近且分別與該等注音符號相關聯的音素標籤(如「i a」、「e o」…等)、一(編輯候選)數學符號的圖像(如「>」)以及在該數學符號的圖像附近且與該數學符號相關聯的音素標籤(如「a hu」)和多個(編輯候選)操作指令的圖像(如代表數字鍵切換指令的「123」、代表輸入空格指令的「空格」和代表傳送指令的「傳送」)以及多個分別在該等操作指令的圖像附近且分別與該等操作指令相關聯的音素標籤(如「ha e」、「ha he」、「ha ha」)。

【0025】 參閱圖4，繪示的示例是該觸控顯示模組2提供的「桌面」顯示視窗，其中含有多個分別代表多個不同的應用程式(如「YouTube」、「EVA Facial Mouse」、「米家」…等)的圖像以及多個分別在該等應用程式的圖像附近且分別與該等應用程式相關聯的音素標籤(如「a a」、「a u」、「a e」…等)和多個分別代表多個操作指令(如「我的檔案」、「電話」、「聯絡人」…等)的圖像以及多個分別在該等操作指令的圖像附近且分別與該等操

作指令相關聯的音素標籤(如「o ha」、「ha a」、「ha u」…等)。

【0026】 參閱圖5，繪示的示例是該電腦裝置100在執行「YouTube」應用程式後由該觸控顯示模組2提供的顯示視窗，其中含有多個分別代表多個不同多媒體檔案(如影片1~影片10)的圖像(如圖像1~圖像10)以及多個分別在該等多媒體檔案的圖像附近且分別與該等多媒體檔案相關聯的音素標籤(如「u e」、「e u」、「e e」…等)和多個分別代表多個操作指令(如「首頁」、「發燒影片」、「訂閱內容」…等)的圖像以及多個分別在該等操作指令的圖像附近且分別與該等操作指令相關聯的音素標籤(如「ha a」、「ha u」、「ha e」…等)。

【0027】 參閱圖6，繪示的示例是該電腦裝置100在執行一社群溝通應用程式後由該觸控顯示模組2提供的一顯示視窗，其中含有多個分別代表多個文字內容的圖像(如「我需要幫忙」、「我要小便」…等)以及多個分別在該等文字內容的圖像附近且分別與該等文字內容相關聯的音素標籤(如「a a」、「a i」…等)和多個分別代表多個操作指令的圖像(如「清除」、「傳送和發音」、「儲存」…等)以及多個分別在該等操作指令的圖像附近且分別與該等操作指令相關聯的音素標籤(如「u ha」、「u hi」、「u he」…等)。

【0028】 在本實施例中，在此情況下，該用戶可根據該觸控顯示模組2當前的顯示視窗所含的操作選項以及與其相關聯的音素標

籤，發出與所欲操作選項相關聯的音素標籤有關的第一音素及第二音素。然而，在其他實施態樣中，若每一音素標籤含有三個或三個以上的參考音素時，用戶則必須發出與所欲操作選項相關聯的音素標籤有關的多個音素，其數量應與每一音素標籤所含的參考音素的數量一致。

【0029】 於是，當該處理單元5接收到由該語音收集模組1收集到來自該用戶且包含連續的第一音素和第二音素的語音信號時，在步驟S25中，該處理單元5可以根據該儲存模組3儲存的該語音辨識資料且利用語音辨識技術，或者根據該儲存模組3儲存的該個人音素資料且利用聲紋辨識技術，確認該第一音素是否相似於該等參考音素其中之一者並確認該第二音素是否相似於該等參考音素其中之一者。若該處理單元5確認出相似於該第一音素的第一目標參考音素、及相似於該第二音素的第二目標參考音素時，則流程進行步驟S26。否則，流程進行步驟S28。特別要說明的是，在實際執行步驟S25時，該處理單元5例如可以先根據該語音辨識資料且利用語音辨識技術來執行該第一音素與該第二音素的確認操作，並在無法成功確認時，再利用該個人音素資料且利用聲紋辨識技術來執行該第一音素與該第二音素的確認操作，但并不在此限。

【0030】 在步驟S26中，該處理單元5根據該第一目標參考音素及該第二目標參考音素以及該儲存模組3儲存的該關聯關係，自該

等音素標籤(即，在步驟S24中當前顯示視窗所含的音素標籤)決定出一個目標音素標籤。該目標音素標籤所含的該第一參考音素與該第二參考音素分別相同於該第一目標參考音素與該第二目標參考音素。

【0031】 接著，在步驟S27中，該處理單元5根據該儲存模組3儲存的該關聯關係將該等操作選項(即，在步驟S24中當前顯示視窗所含的操作選項)其中一個與該目標音素標籤相關聯的目標操作選項(亦即所欲操作選項)激活。

【0032】 當該處理單元5確認出該第一音素與該第二音素其中一者與該等參考音素其中每一者均不相似時，在步驟S28中，該處理單元5使該觸控顯示模組2顯示辨識失敗訊息。於是，該用戶可在重新發出含有與所欲操作選項相關聯的語音標籤有關的第一音素與第二音素後，該電腦裝置100重新執行步驟S25的操作直到能確認出相似於該第一音素與該第二音素的參考音素為止。

【0033】 以下，將進一步就該目標操作選項的實際形式來示例地說明該處理單元5如何激活該目標操作選項的方式。

【0034】 若該目標操作選項為一符號(如圖3所示的虛擬鍵盤中的數學符號「>」)時，該處理單元5透過一使該觸控顯示模組2顯示該符號在一編輯區(圖未示)的方式來激活該目標操作選項。相似地，若該目標操作選項為一字元(如圖3所示的虛擬鍵盤中的字元

「我」)時，該處理單元5透過一使該觸控顯示模組2顯示該字元的方式來激活該目標操作選項。若該目標操作選項為一文字內容(如圖6所示的顯示視窗中的「我需要幫忙」)時，該處理單元5不僅透過一使該觸控顯示面板2顯示該文字內容在一溝通紀錄區(如圖6所示的溝通紀錄區)方式，還透過一使該喇叭模組4播放一對應於該文字內容的語音內容的方式來激活該目標操作選項。若該目標操作選項為一操作指令(如圖4所示的顯示視窗中的操作指令「電話」)時，該處理單元5透過一執行該操作指令的方式(如使該觸控顯示模組2從原本的桌面顯示視窗切換至與「電話」有關的顯示視窗)來激活該目標操作選項。若該目標操作選項為一應用程式(如圖4所示的顯示視窗中的應用程式「YouTube」)時，該處理單元5透過一執行該應用程式的方式來激活該目標操作選項，並使該觸控顯示模組2從原本的顯示視窗(如圖4所示)切換至與該應用程式相關的顯示視窗(如圖5所示)。若該目標操作選項為一檔案時，該處理單元5透過一開啟或播放該檔案之方式來激活該目標操作選項。

【0035】 值得注意的是，上述第一實施例的語音輸入操作方法可被編程為一包含多個程式指令的電腦程式產品，並將該電腦程式產品儲存在一電腦可讀取媒體(例如，該儲存模組3)。當該電腦裝置100執行該等程式指令時，該電腦裝置100可完成如以上所述基於至少雙音素的語音輸入操作方法。

【0036】 參閱圖7，繪示出的另一電腦系統不僅包含作為使用終端的該電腦裝置100，還包含一辨識伺服器端200。該電腦裝置100與該辨識伺服器端200協同來實施本發明第二實施例的基於至少雙音素(Double-phoneme)的語音輸入操作方法。在本實施例中，該辨識伺服器端200可經由一通訊網路300與該電腦裝置100通訊，並支援語音及聲紋辨識技術。

【0037】 以下，將參閱圖7及圖8來示例地說明該電腦系統如何執行該第二實施例的語音輸入操作方法。大體而言，本實施例的語音輸入操作方法為上述第一實施例的語音操作方法的變化實施例，並可包含以下步驟S80-S91。

【0038】 首先，在步驟S80中，該辨識伺服器端200預先儲存有該語音辨識資料。

【0039】 在步驟S81中，在註冊階段，該電腦裝置100經由該通訊網路300，將該語音收集模組1收集到由一用戶發出對應於多個參考音素的多個語音的語音內容傳送至該辨識伺服器端200。

【0040】 然後，在步驟S82中，該辨識伺服器端200將來自該電腦裝置100的該語音內容作為對應於該用戶的個人音素資料並將其儲存。值得一提的是，在實際使用時，該辨識伺服器端200亦可用作有關個人音素資料的雲端伺服器，並廣為收集且儲存大量其他用戶(如構音異常的用戶)的個人音素資料，並將此大量資料進一步利用

人工智慧的分析或進行機器學習可獲得作為用於特殊語音(如構音異常用戶所發出的語音)辨識的語音資料庫(圖未示)。

【0041】 接著，該電腦裝置100依序執行步驟S83至步驟S85。由於該電腦裝置100在步驟S83至步驟S85的操作細節分別相同於上述步驟S22至步驟S24(圖2)的所有操作細節，故在此不再贅述。

【0042】 之後，當該電腦裝置100的該處理單元5接收到由該語音收集模組1收集到來自該用戶且包含連續的第一音素和第二音素的語音信號時，該電腦裝置100經由該通訊網路300，將一包含該語音信號且有關該用戶的辨識請求傳送至該辨識伺服器端200(步驟S86)。

【0043】 然後，該辨識伺服器端200在接收到來自該電腦裝置100的該辨識請求時，可以根據已儲存的該語音辨識資料且利用語音辨識技術，或者根據已儲存且對應於該用戶的該個人音素資料並利用聲紋辨識技術(又或者根據上述用於特殊語音辨識的語音資料庫且利用語音辨識技術)，確認該第一音素是否相似於該等參考音素其中之一者並確認該第二音素是否相似於該等參考音素其中之一者(步驟S87)。若該辨識伺服器端200確認出相似於該第一音素的第一目標參考音素、及相似於該第二音素的第二目標參考音素時(即，成功辨識)，則流程進行步驟S88。否則，流程進行步驟S91。

【0044】 在步驟S88中，該辨識伺服器端200經由該通訊網路

300，將一含有該第一目標參考音素及該第二目標參考音素的辨識回覆傳送至該電腦裝置100。

【0045】 然後，在步驟S89中，該電腦裝置100的該處理單元5根據該辨識回覆所含的該第一目標參考音素及該第二目標參考音素以及該儲存模組3儲存的該關聯關係，自該等音素標籤決定出一個目標音素標籤。該目標音素標籤所含的該第一參考音素與該第二參考音素分別相同於該第一目標參考音素與該第二目標參考音素。

【0046】 接著，該電腦裝置100的該處理單元5，相似於上述步驟S27(圖2)，根據該儲存模組3儲存的該關聯關係將該等操作選項其中一個與該目標音素標籤相關聯的目標操作選項(亦即所欲操作選項)激活(步驟S90)。

【0047】 當該辨識伺服端200確認出該第一音素與該第二音素其中一者與該等參考音素其中每一者均不相似時(即辨識失敗)，在步驟S91中，該辨識伺服端200經由該通訊網路300，將辨識失敗訊息傳送至該電腦裝置100。於是，該電腦裝置100的該處理單元5可將來自於該辨識伺服端200的該辨識失敗訊息顯示於該觸控顯示模組2，以供該用戶觀看。於是，該用戶可在重新發出含有與所欲操作選項相關聯的語音標籤有關的第一音素與第二音素後，該電腦系統重新執行步驟S86與步驟S87的操作直到能確認出相似於該第一音素與該第二音素的參考音素為止。

【0048】 綜上所述，由於先儲存了用於語音辨識的語音辨識資料，以及用於聲紋辨識且對應於用戶的個人音素資料，因此不僅對於構音正常的用戶所發出的語音，而且對於構音障礙患者的用戶所發出的有限語音均能被精確地辨識出。此外，基於至少雙音素的編碼方式能定義出相對較多數量的音素標籤，此等音素標籤能被廣泛地應用來與相對較多數量的操作選項建立關聯性。於是，相較於現有利用相對複雜的語言模型及聲學模型的語音辨識技術，能相對容易且快速地語音辨識出用戶所發出且含至少雙音素的語音輸入，以判定出所欲激活的目標操作選項。故確實能達成本發明的目的。

【0049】 惟以上所述者，僅為本發明之實施例而已，當不能以此限定本發明實施之範圍，凡是依本發明申請專利範圍及專利說明書內容所作之簡單的等效變化與修飾，皆仍屬本發明專利涵蓋之範圍內。

【符號說明】

【0050】

100...電腦裝置

1...語音收集模組

2...觸控顯示模組

3...儲存模組

4...喇叭模組

5...處理單元

- 200...辨識伺服端
- 300...通訊網路
- S20~S28...步驟
- S80~S91...步驟

【發明申請專利範圍】

【第1項】一種基於至少雙音素的語音輸入操作方法，利用一具有語音與聲紋辨識技術的電腦系統來執行，並包含以下步驟：

(A)儲存與多個彼此不同的參考音素相關聯的語音辨識資料、及對應於一用戶的個人音素資料，該個人音素資料包含由該用戶所發出且分別對應於該等參考音素的多個語音的語音內容；

(B)將該等參考音素編碼，以定義出多個彼此不同的音素標籤，其中每一個音素標籤至少包含一選自該等參考音素其中一者的第一參考音素、及一選自該等參考音素其中一者的第二參考音素；

(C)將該等音素標籤分別與多個彼此不同的操作選項相關聯；

(D)當收集到來自該用戶且至少包含連續的第一音素和第二音素的語音信號後，根據該語音辨識資料且利用語音辨識技術，或者根據該個人音素資料且利用聲紋辨識技術，確認該第一音素是否相似於該等參考音素其中之一者並確認該第二音素是否相似於該等參考音素其中之一者；

(E)當確認出相似於該第一音素的第一目標參考音素、及相似於該第二音素的第二目標參考音素後，根據該第一目標參考音素和該第二目標參考音素，自該等音素標籤決定出一個目標音素標籤，該目標音素標籤所含的該第一參考音素與該第二參考音素分別相同於該第一目標參考音素與該第二目標參考音素；及

第 1 頁，共 4 頁(發明申請專利範圍)

(F)將該等操作選項其中一個與該目標音素標籤相關聯的目標操作選項激活。

【第2項】如請求項1所述的基於至少雙音素的語音輸入操作方法，其中，在步驟(A)中，每一參考音素為一母音或一音節。

【第3項】如請求項1所述的基於至少雙音素的語音輸入操作方法，其中：

在步驟(C)中，每一操作選項為一符號、一字元、一文字內容、一操作指令、一應用程式及一檔案其中一者；及

在步驟(F)中，當該目標操作選項為一符號時，該電腦系統透過一顯示該符號之方式激活該目標操作選項，

當該目標操作選項為一字元時，該電腦系統透過一顯示該字元之方式激活該目標操作選項，

當該目標操作選項為一文字內容時，該電腦系統至少透過一顯示該文字內容之方式激活該目標操作選項，

當該目標操作選項為一操作指令時，該電腦系統透過一執行該操作指令之方式激活該目標操作選項，

當該目標操作選項為一應用程式時，該電腦系統透過一執行該應用程式之方式激活該目標操作選項，及

當該目標操作選項為一檔案時，該電腦系統透過一開啟或播放該檔案之方式激活該目標操作選項。

【第4項】如請求項3所述的基於至少雙音素的語音輸入操作方法，在步驟(C)與步驟(D)之間，還包含以下步驟：

(G)顯示多個分別代表該等操作選項的圖像，並在該

第2頁，共4頁(發明申請專利範圍)

等圖像附近顯示與該等操作選項相關聯的該等音素標籤。

【第5項】如請求項4所述的基於至少雙音素的語音輸入操作方法，在步驟(F)中，當該目標操作選項為一文字內容時，該電腦系統不僅透過顯示該文字內容之方式，還透過一播放一對應於該文字內容的語音內容的方式激活該目標操作選項。

【第6項】如請求項5所述的基於至少雙音素的語音輸入操作方法，該電腦系統包含一用於執行步驟(B)、(C)、(E)、(F)及(G)的使用終端、及一可與該使用終端通訊且用於執行步驟(A)及(D)的辨識伺服器端，還包含以下步驟：

在步驟(A)之前，(H)藉由該使用終端，將收集到的該個人音素資料傳送至該辨識伺服器端；

在步驟(C)與步驟(D)之間，(I)藉由該使用終端，收集該語音信號，並將一包含該語音信號且有關該用戶的辨識請求傳送至該辨識伺服器端，以便該辨識伺服器端回應於該辨識請求執行步驟(D)；及

在步驟(D)與步驟(E)之間，(J)藉由該辨識伺服器端，當確認出該第一目標參考音素與該第二目標參考音素時，將一含有該第一目標參考音素及該第二目標參考音素的辨識回覆傳送至該使用終端，以使該使用終端回應於該辨識回覆執行步驟(E)。

【第7項】一種電腦程式產品，儲存在一電腦可讀取媒體，包含多個程式指令，且當一電腦系統執行該等程式指令時，可完成如請求項1至5其中任一項所述之基於至少雙音素的語音

輸入操作方法。

【發明圖式】

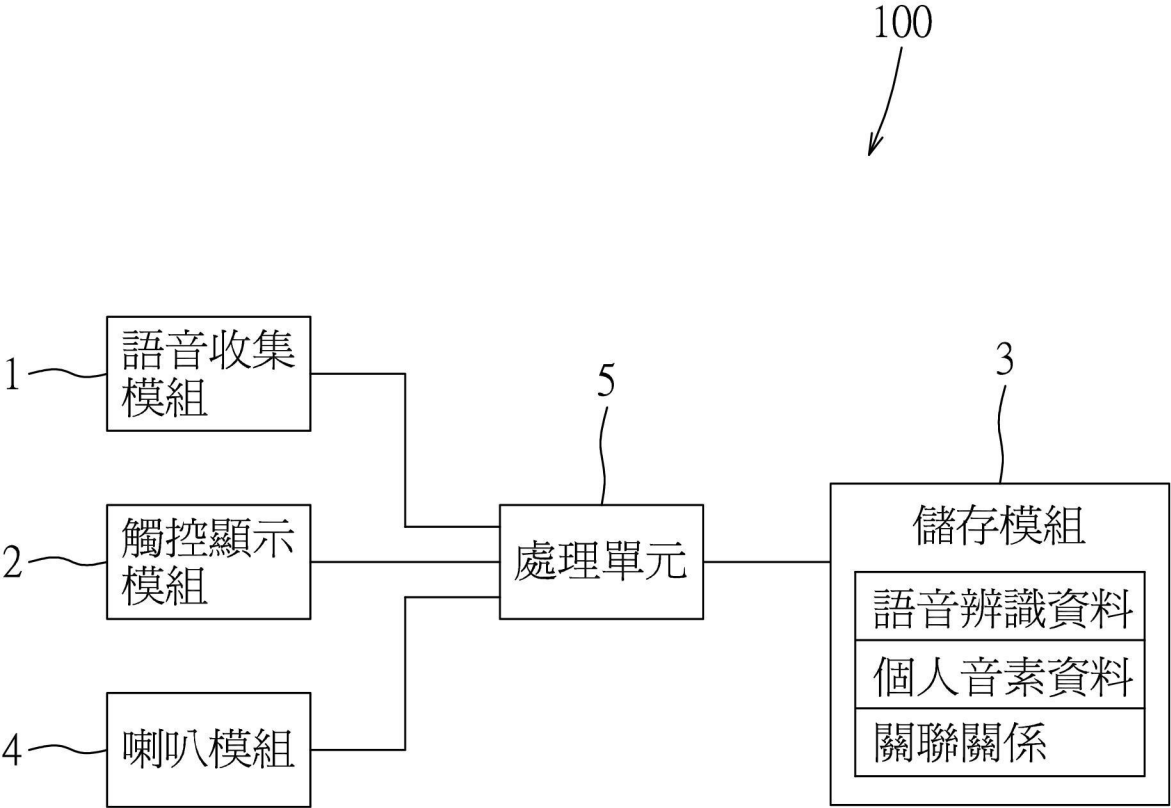


圖 1

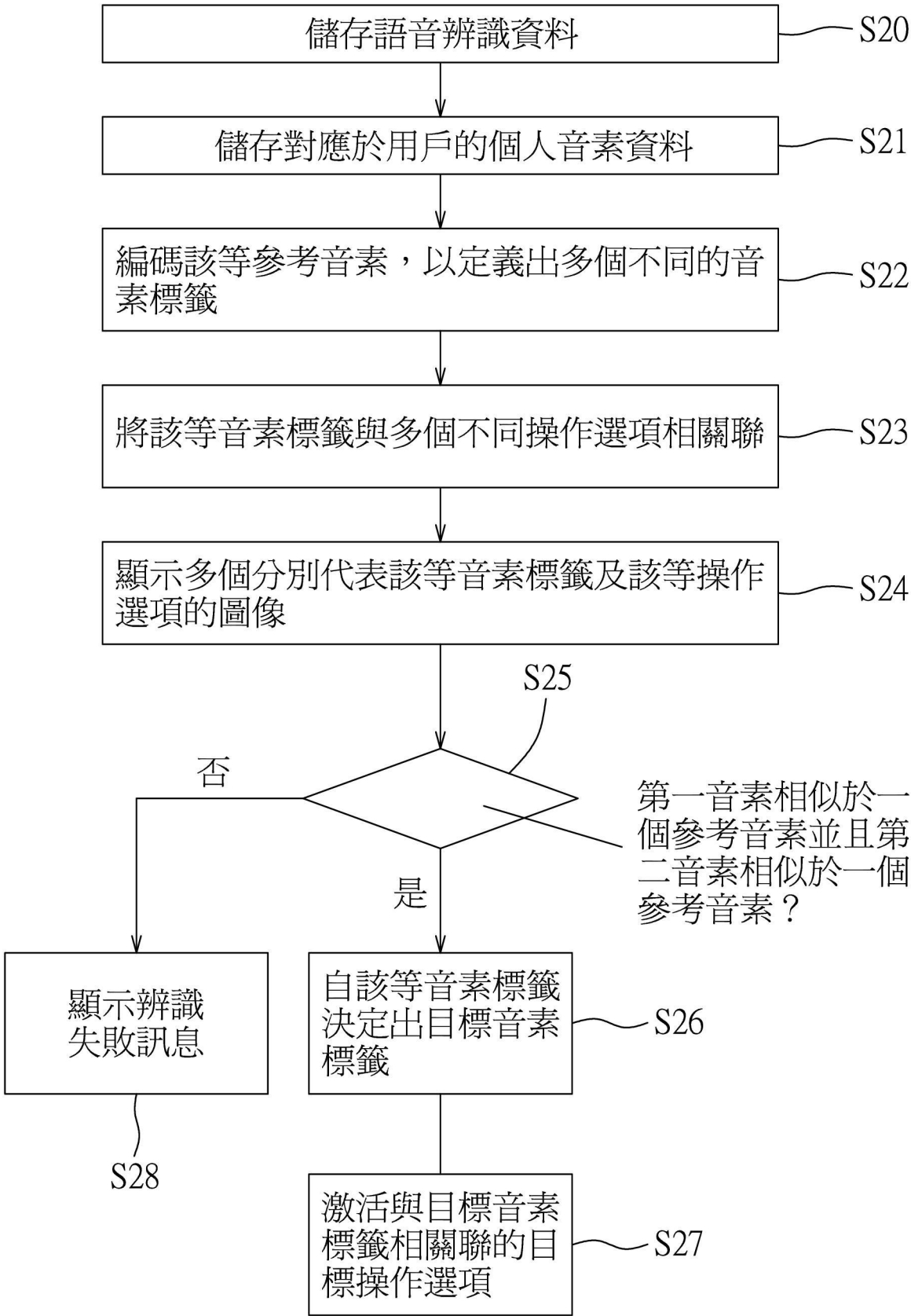


圖 2

aa 我		ai 好		ae 這		ao 但		au 妳		a he 有		a ha 不		a hu >	
ia ㄣ		ii ㄉ		ie ㄨ		io ㄟ		iu ㄣ		i ha •		i hu ㄣ			
eo ㄣ		eu ㄣ		e he ㄣ		e ha ㄣ						oo ㄟ			
o wo ㄣ		o ha ㄣ		o hu ㄣ		u a ㄣ								u oa ㄣ	
u hu ㄣ		he a ㄣ		he i ㄣ		he e ㄣ		he o ㄣ						ha a ㄣ	
ha e 123						ha he 空格								ha ha 傳送	

圖 3

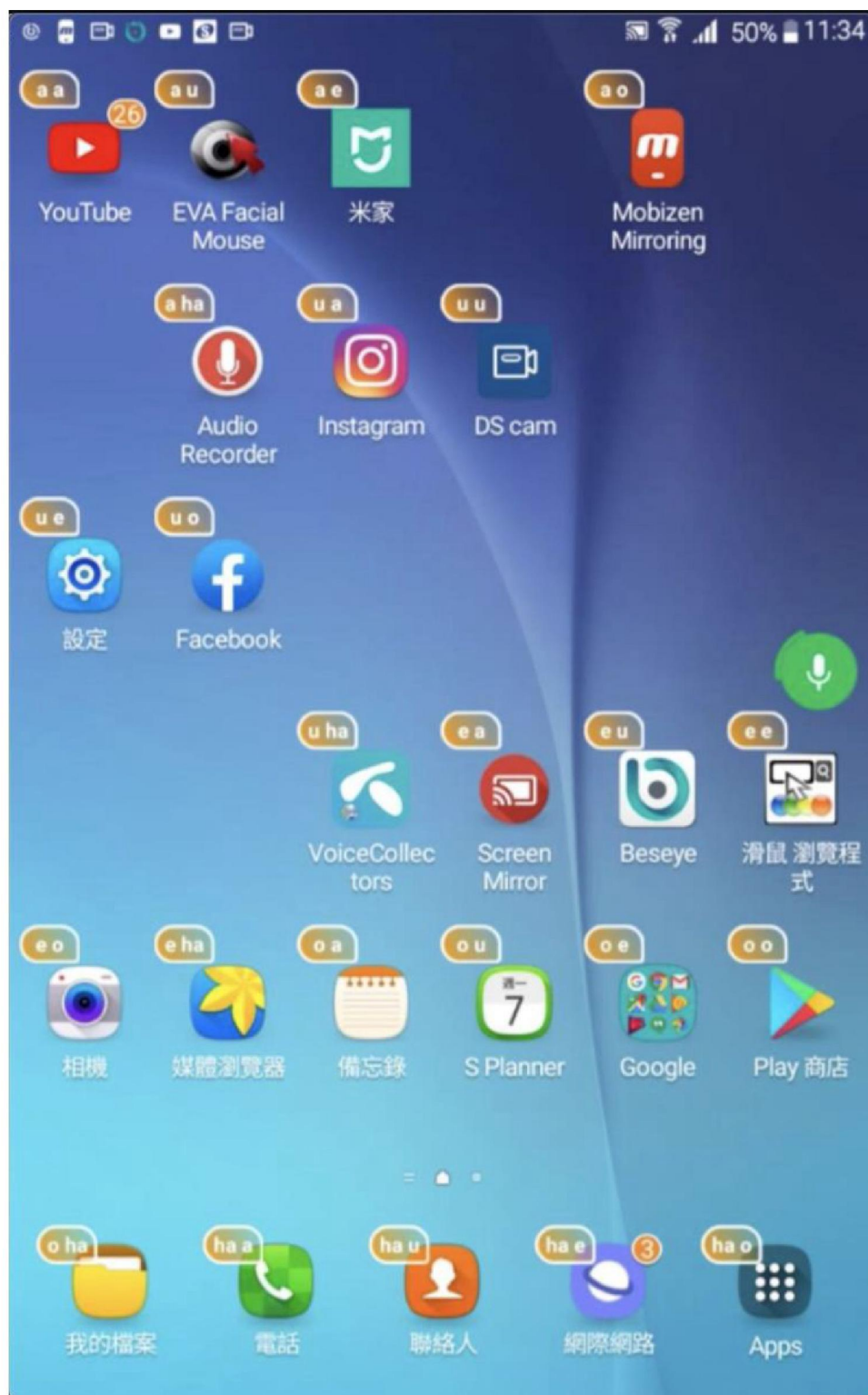


圖 4



圖 5

搜尋	ua 我要幫忙 ai 我要小便 ae 我要大便 ao 我累了 au 我要躺下 u ha 清除				
群組A	ia 救命啊 ii 我要吃東西 ie 我要喝水 io 我吃飽了 iu 我要洗頭				
群組B	ea 我要洗臉 ei 我要洗手 ee 我要洗澡 eo 我要去走走 eu 我需要按摩				
群組C	oa 我要看電視 oi 我覺得熱 oe 我覺得冷 oo 我覺得痛 ou 我覺得冷				
...	群組名稱(群組B)				
加入成員	溝通紀錄區				
	uo ↑ uu ↓				
編輯區					
		u hi 傳送和發音		u he 儲存	u ho 錄音

圖6

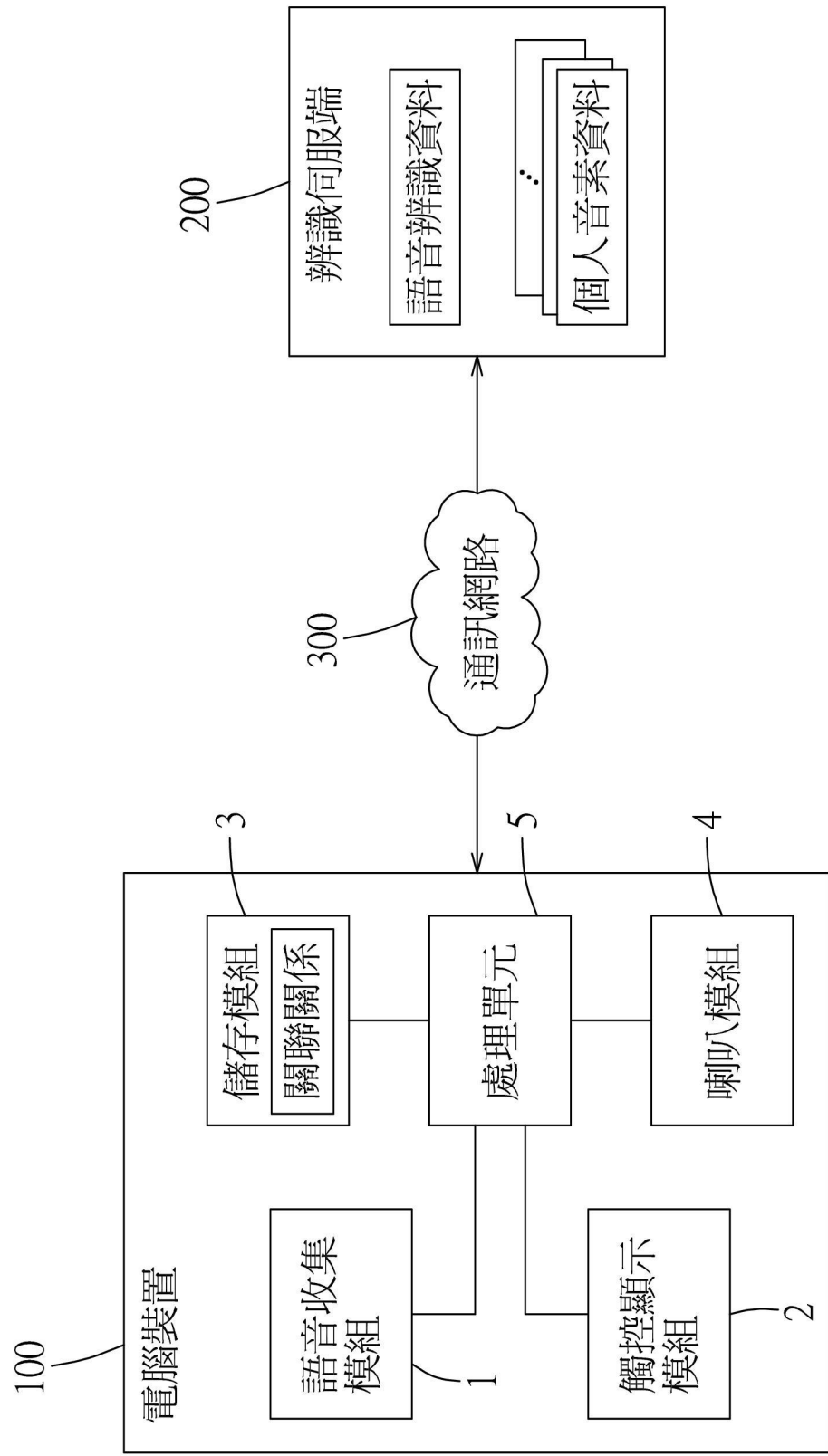


圖7

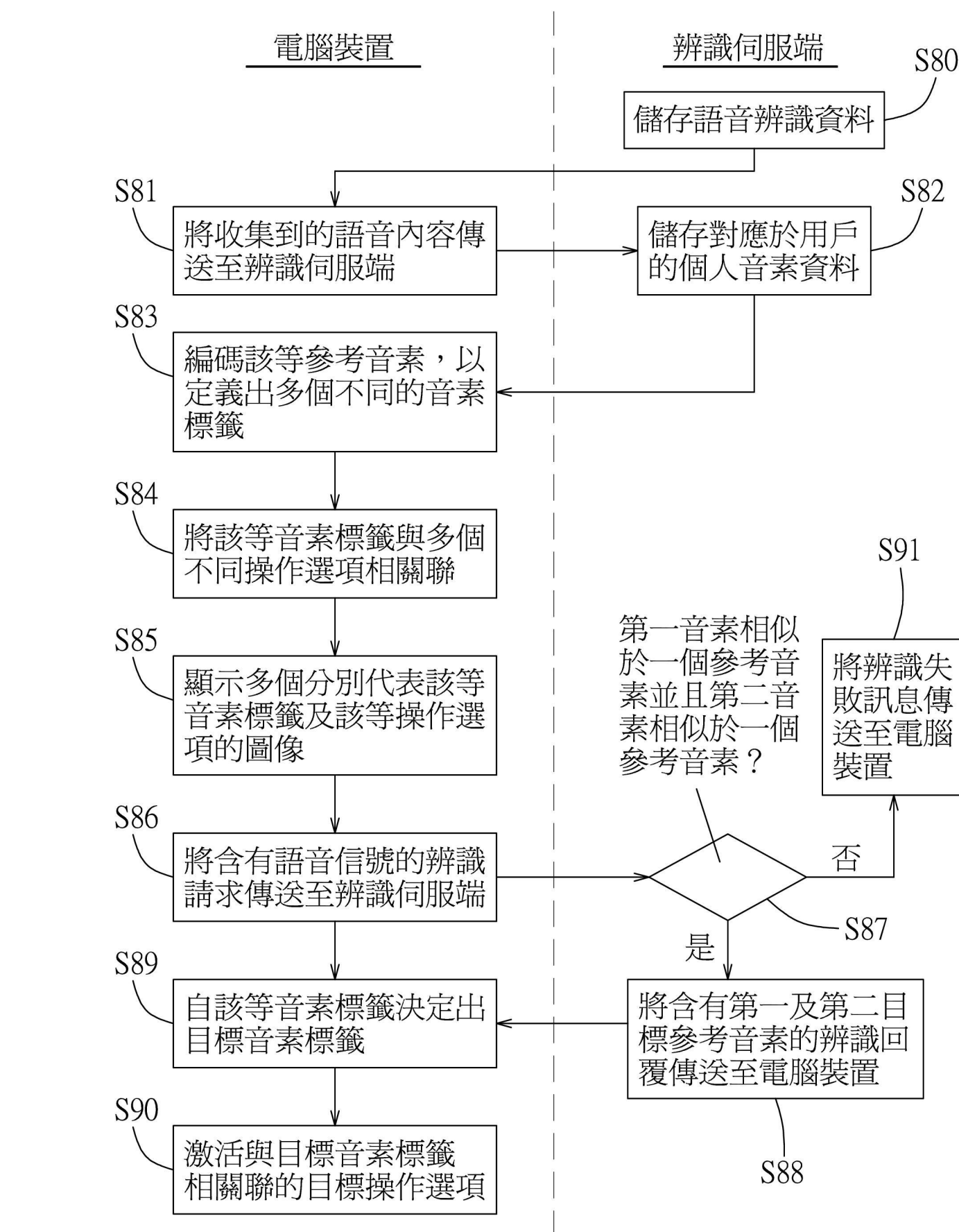


圖 8