

# 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：94115348

※ 申請日期：94-5-12

※IPC 分類：

G10L 15/00

## 一、發明名稱：(中文/英文)

用於針對語者無關語音辨識之發音所生合成音訊響應的裝置及方法

APPARATUS AND METHOD FOR SYNTHESIZED AUDIBLE RESPONSE TO AN UTTERANCE IN  
SPEAKER-INDEPENDENT VOICE RECOGNITION

## 二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

英特爾公司/ INTEL CORPORATION

代表人：(中文/英文)

塞門 大衛/SIMON, DAVID

住居所或營業所地址：(中文/英文)

美國加州聖塔克萊拉市密遜大學道 2200 號

2200 Mission College Blvd. Santa Clara, CA, USA

國 籍：(中文/英文) 美國/U. S. A.

## 三、發明人：(共 2 人)

姓 名：(中文/英文)

1. 艾瑞爾 亞朵蘭/ERELL, ADORAM

2. 梅勒 艾瑟/MELZER, EZER

國 籍：(中文/英文)

以色列/ISRAEL

#### 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： 年 月 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：

1. 美國; 2004, 06, 02; 10/857, 848

☐ 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】

本發明係有關於用於針對語者無關語言辨識之發音所生合成音訊響應的裝置及方法。

### 5      【先前技術】

#### 發明背景

一種語者無關語音辨識(SIVR)系統經由將說話語調與預先界定之字彙匹配來辨識該語調內容的意義。舉例言之，於語者無關電話撥號用途，字彙可包括姓名串列。當  
10 使用者說出該字彙中的姓名之一時，該系統辨識該名稱，初始化呼叫與該姓名相關聯之電話號碼。常見，經由將語調與一語音呈現集合個別比對來工作之SIVR系統，可由該詞條之文字呈現自動產生。

為了避免辨識錯誤結果，SIVR應用用途係採用語音驗證技術來通知使用者，已經辨識出詞條，讓使用者可判定  
15 是否繼續進行。經由使用稱作為文字轉語音(TTS)之程序，由所識別之詞條文字自動產生，合成欲播放的口語節段來達成語音辨識。

SIVR程序及TTS程序皆係基於將字符串自動轉換成為  
20 對應之口語組成積木(稱作為音素)之對應順序來達成。但此等轉換方法，後文稱作自轉音(LTP)方法複雜，原因在於於諸如英文之語文中有許多個字母及字母串可有兩個或兩個以上的不同音。例如字母串「ie」於以下三個字的發音各自不同：friend、fiend及lied。可經由將相對大量的記憶體空

間用來儲存轉換規則之綜合集合，而改良選出正確發音的機會。但於埋設式應用用途諸如電話，記憶體屬於先決要件。SIVR發音預測之經濟實作方法，藉統計規則係仰賴產生對應一指定字串之多個可能發音之粗淺發音描述，其中

5 只有若干發音描述可能正確，然後將各個語音呈現與欲辨識的語調內容比對。再度參照前述實例，若使用者說出「friend」，則辨識程序可能嘗試將此語調內容與當字母串「ie」於字friend、fiend及lied發音時所產生之四種語音呈現個別匹配。

10 但此項經濟方法對TTS不可行，TTS的本質必須產生單一發音。結果TTS程序包括耗用大量記憶體的正確發音預測，或節省記憶體但容易產生誤導或甚至產生無法符合使用者預期的荒謬發音之粗糙發音預測。

### 【發明內容】

15 本發明揭露一種方法，包含有下列步驟：選擇一預定字彙之口語元素之多個語音表示法中最密切匹配一語調內容之一種語音表示法，其中該等多個語音表示法包括具有不同可能發音之該等口語元素中之任一者之多種語音表示法；以及根據該等語音表示法中之該一語音表示法，來合

20 成一發聲口語片段。

### 圖式簡單說明

本發明之具體例係藉實施例舉例說明，而非圍限於附圖之各圖，附圖中類似的參考號碼指示對應元件、類似元件或相似元件，附圖中：

第1圖為根據本發明之一具體例，範例語者無關語音辨識系統之示意方塊圖；

第2圖為結合第1圖所示之語音辨識系統之範例行動蜂巢式電話之示意方塊圖；

5 第3圖為增加詞條至第1圖所示語音辨識系統之方法之示意流程圖；

第4圖為使用第1圖所述語音辨識系統響應於語音命令方法之示意流程圖；以及

● 第5圖為範例字圖，顯示儲存於第1圖所述與語者無關  
10 語音辨識系統之字彙對應於不同的口語元素之語音表示法之各個路徑。

須了解為求簡明及清晰說明，附圖顯示之各元件並非必然照比例繪製。例如，部分元件之尺寸可能相對於其它元件誇張呈現以求清晰。

## 15 【實施方式】

詳細說明

● 於後文說明，列舉大量特定細節以供了解本發明。但熟諳技藝人士須了解本發明可無此等特定細節而實作。其它情況下，眾所周知之方法、程序及元件並未說明其細節  
20 俾不致於混淆本發明。

後文若干詳細說明部分係以於記憶體內部於資料位元或二進制數位信號之運算演繹法則或運算之符號表示來呈現。此等演繹法則之說明及呈現可為熟諳資料處理技藝人士用來將其工作本質傳輸給業界其它人員之技術。

於說明書及申請專利範圍，「多數」一詞表示「二或二以上」。

若干本發明之具體例係針對一種語者無關語音辨識(SIVR)系統，該系統使用一種方法其允許使用者經由使用  
5 屬於先前定義之口語元素串列(包括自然語言之字、片語、人名及專有名詞、標新立異之暱稱等)發出語音命令，讓使用者來操作應用程式之功能。

字串表示欲辨識之各個口語元素，若干本發明之具體例包括字轉音(LTP)轉換程序，LTP轉換程序將各個文字呈  
10 現轉換成為可儲存於預先界定之字彙之一或多個可能的語音呈現。

當使用者發出語音命令時，系統將使用者的語調內容與字彙中的語音呈現比對，選出最接近匹配辨識使用者發音的特定口語元素。

15 系統可提供使用者對所辨識之口語元素的語音驗證，系統播放出合成之聲學口語片段，使用者可接受該選項或拒絕該選項。後文所述具體例使用之方法特別係針對播放由最接近(最密切)匹配使用者語調內容之特定語音呈現所合成之口語片段。經由允許LTP程序來產生對一指定字串之  
20 多個替代語音呈現，以及選擇最接近(最密切)匹配使用者的發聲內容之發音，本方法當使用指定之處理功率量及記憶體量實作時，可提供更正確合成且更佳發音的語音驗證。相同之LTP模組用於一個完整系統之SIVR元件及文字轉語音(TTS)元件二者之方法之特殊效果，經由減少實作具有可

接受品質之語音辨識系統所需處理功率及所需記憶體容量，也可達成製造成本的下降。

現在參照第1圖，第1圖顯示根據本發明之具體例，SIVR系統控制一應用區塊之範例裝置。後文討論須牢記所述語音辨識系統之各區塊係限於與若干本發明之具體例相關之區塊，所描述之區塊可具有與該等具體例無關之額外功能。

語音控制裝置138有一由SIVR系統100所控制之應用區塊136。語音控制裝置138例如為無線電話、行動蜂巢式電話、地面線路電話、遊戲機、語音控制玩具、個人數位助理器(PDA)、掌上型電腦、筆記型電腦、桌上型個人電腦、工作站、伺服器電腦等。應用區塊136之實例為行動蜂巢式電話之收發器、地面電話之直接存取配置(DAA)、語音控制玩具之馬達與燈具控制區塊、於個人電腦上執行之桌上型公開程式等。SIVR系統100可解譯使用者之語音命令，以及利用命令信號134來發出對應指令給應用區塊136。

SIVR系統100可包括一音訊輸入裝置106、一音訊輸出裝置108、一音訊編碼解碼器114、一處理器120、一輸入裝置122、一顯示器126及一字彙記憶體130。熟諳技藝人士須了解SIVR系統100可於應用區塊136共享部分或全部前述成分區塊。例如處理器120除了於實作SIVR系統100之角色之外可執行或未執行應用區塊136之處理功能；字彙記憶體130可與應用區塊136所使用之儲存記憶體共享或未共享實體記憶體裝置。

音訊輸入裝置106可為轉換器諸如麥克風，轉換器用來將接收得之聲音信號102轉換成為輸入類比音訊信號110。音訊輸入裝置106允許使用者發出語音命令給語音辨識系統。

5 音訊輸出裝置108可為轉換器，諸如揚聲器、頭戴式耳機或耳機等轉換器用來將輸出類比音訊信號112轉換成為被傳輸之聲音信號104。音訊輸出裝置108允許語音辨識系統響應於來自使用者的語音命令而播放一口語節段，作為提供該語音辨識系統所辨識之口語元素之語音驗證。

● 音訊編碼解碼器114可將輸入類比音訊信號110轉換成  
10 為輸入數位音訊信號116，該信號可傳輸至處理器120，且可將處理器120產生之輸出數位音訊信號118轉換成為輸出之類比信號112。

輸入裝置122可為鍵盤、虛擬鍵盤等，來允許使用者載入文數符號串，包括系統隨後可呼叫來辨識之語音命令的  
15 文字呈現；以及載明欲與該等文字呈現個別相關聯之動作，諸如當接收到特定語音命令時，載入欲撥號的電話號碼。輸入裝置122可使用匯流排124指示使用者的選擇給處理器120，匯流排124例如可為通用串列匯流排(USB)介面、個人電腦鍵盤介面或電子產業聯盟(EIA)EIA232串列介面。

● 輸入裝置122也包括手動控制，手動控制允許使用者證實或拒絕來自於命令所得動作，對該系統的控制作請求及選擇。此等控制例如可用來指示即將發出的語音命令，或用來證實或拒絕語音命令之語音驗證，藉此讓系統以對應之動作進行或放棄對應之動作。手動控制可選擇性為分開  
20



手動控制，諸如安裝於汽車之方向盤上的按鈕，其可替代或重複含括於輸入裝置122的手動控制。

顯示器126可為蜂巢式電話液晶顯示裝置(LCD)、個人電腦視覺顯示單元、PDA顯示器等，顯示器126可以視覺方式指示使用者已經使用輸入裝置122所載入之符號，且可視需要來提供其它指示，諸如提醒使用者完成一項程序，且提供所辨識的語音命令之視覺指示。熟諳技藝人士須了解顯示器126可組合指標裝置諸如光筆、手指操作或針筆操作之觸摸式面板、遊戲搖桿、電腦滑鼠、軟鍵、選擇鍵與游標移動鍵的集合等或其組合來額外執行虛擬鍵盤功能，其可替代部分或全部輸入裝置122之功能。處理器120可使用顯示器匯流排128來發送信號給顯示器126。顯示器匯流排128之實例為視訊圖形陣列(VGA)匯流排驅動電腦視覺顯示單元及LCD介面來驅動專有LCD顯示模組。

字彙記憶體130可儲存至少一個語音呈現以及欲對系統將辨識之各個口語元素執行的動作之描述，以及與各個口語元素相關聯之文字呈現。也儲存與使用之音標集合相關聯之聲學模型，諸如隱藏馬可夫(Markov)模型、動態時間包裹樣版等，其為固定，或於應用程式佈署時對使用者口語進行調整。字彙記憶體130例如可為緊密快閃(CF)記憶卡；個人電腦與記憶卡國際協會(PCMCIA)記憶卡；記憶體桿(MEMORY STICK)卡；USB鑰(USB KEY)記憶卡；可電抹除可規劃之唯讀記憶體(EEPROM)；非依電性隨機存取記憶體(NVRAM)；同步動態隨機存取記憶體(SDRAM)；靜態

隨機存取記憶體(SRAM)；整合於微處理器或微控制器之記憶體；雷射光碟唯讀記憶體(CD-ROM)；硬碟；軟碟等。

處理器120可使用記憶體匯流排132將資料寫入字彙記憶體130，以及由字彙記憶體130取還資料，記憶體匯流排  
5 132可為USB、快閃記憶裝置介面、個人電腦與記憶卡國際協會(PCMCIA)卡匯流排等。

處理器120例如可為個人電腦中央處理單元(CPU)、筆記型電腦CPU、PDA CPU、數位信號處理器(DSP)、精簡指令集電腦(RISC)、複雜指令集電腦(CISC)或埋設式微控制  
10 器或埋設式微處理器。

處理器120可利用命令信號134而與經過控制之應用區塊136通訊，命令信號134例如可透過實體媒體而傳輸，該等實體媒體諸如為USB、EIA232介面、共享電腦匯流排、微處理器並列埠、微處理器串列埠、或雙埠隨機存取記憶體  
15 體(RAM)介面。當處理器120之資源係介於SIVR系統100與應用區塊136間共享時，命令信號134例如可組成命令位元組集合，係由SIVR系統100之軟體常式進送至屬於應用區塊136之軟體常式之命令位元組集合。

現在額外參照第2圖，其中顯示根據本發明之額外具體  
20 例之範例語音控制行動蜂巢式電話。

語音控制之行動蜂巢式電話150包括SIVR系統100、收發器140及天線142。SIVR系統100可利用命令信號134來控制蜂巢式電話功能。蜂巢式電話150之其它區塊由第2圖刪除，原因在於該等區塊非關所述具體例之語音操作功能。

但熟諳技藝人士了解SIVR系統100可與非關聯語音辨識功能之蜂巢式電話功能，共享部分或全部其組成區塊。舉例言之，音訊輸入裝置106不僅可作為SIVR系統100接收來自使用者之語音命令的手段，同時音訊輸入裝置106也可用來接收欲傳輸至使用者欲通訊之遠方的口語；以及處理器120可額外執行與非關SIVR之蜂巢式電話操作之各方面相關聯的功能。

控制器120之操作與其它系統區塊相關聯之操作於額外參照第3圖及第4圖可更明白了解，其中示意流程圖顯示根據本發明之具體例增加詞條之方法，以及響應於語音命令之方法。

第3圖所示程序200之目的係將一或多對應於新口語元素的語音表示法增加至字彙。於開始時，程序200前進至方塊210，等候使用者定義欲由系統辨識的新口語元素。利用輸入裝置122，使用者可經由載入口語元素之文字表示法於其自然語言拼字中來定義一個新的口語元素，然後按下輸入(ENTER)鍵或進行若干類似的操作來指示文字的載入完成。舉例言之，使用者可載入文字「Stephen」來指示隨後當發出語音命令「Stephen」時欲撥號方的姓名。

當使用者完成表示新口語元素的字串載入時，程序200可前進至方塊220。於方塊220，處理器120將該口語元素文字轉成對應於可辨識之音標或音標群的組成部分。用於前述實例，處理器120可將文字「Stephen」分割成為「s」、「t」、「e」、「ph」及「en」。熟諳技藝人士顯然易知本實例顯示

之再分割僅供方便舉例說明本方法之用，且只表示將文字「Stephen」分成其組成音標或組成音標群的多種替代方式之一，此外，將文字再分成字母群指示數種啟動LTP程序之方式之一。當方塊220完成時，程序200可前進至方塊230。

- 5       於方塊230，處理器120可將使用者所載入之文字表示法經由首先將前述組成部分轉成可能的語音表示法，隨後將語音表示法連鎖並置成為字圖形式來將該文字表示法轉成可能的語音表示法。繼續前述實例，使用卡內基美隆大學(Carnegie Mellon University(CMU))發展的Pronouncing
- 10   Dictionary 0.6版之音標集合，該字典為北美英文之機器可讀取發音字典，可於CMU的網際網路網站取得，將組成部分轉成可能的語音表示法的規則，陳述「e」發成「EH」如同於「Devon」或發成「IY」如同於「demon」；陳述「ph」發音成「F」或「V」；以及陳述「en」發音為如同「encode」
- 15   的「EH N」或如同「seven」的「AH N」。現在參照第5圖，第5圖顯示對應於名字Stephen之範例字圖，其中顯示八個路徑，始於節點400而終於節點402至416。熟諳技藝人士顯然易知字圖可以比第5圖所表示之更加壓縮方式儲存於字彙記憶體130，多個節點可由單一節點替代，多個緣可進入各
- 20   個節點。舉例言之，「F」、「V」、「EH」、「AH」及「N」各有一個節點。始於節點400而終於節點408及412的二路徑係屬於名字Stephen之二正常發音的語音表示法，而其它路徑則屬於通常被視為無效的發音。本例只為口語元素有多於一個可接受的發音的案例，通常根據個別偏好、地區口音

等可接受多種其它發音。於方塊230完成時，程序200前進至方塊240。

於方塊240，程序200等候利用輸入裝置122來規定當系統隨後辨識出對應於載入之文字的語音命令時，欲進行的動作。載明所要求的動作之程序例如係藉單純文字分錄載明，或藉手動驅動分錄載明，其中使用者由顯示器126顯示之串列中選出可能的動作，或二者的組合。於前文說明之實例，使用者指示載入的文字「Stephen」表示撥號Stephen的電話號碼的命令，藉著由一串所顯示的動作中選出「撥號」，然後載入Stephen的電話號碼。方塊240另外可前進程序200流程之方塊210。當使用者完成所規定要求的動作時，程序200前進至方塊250。

於方塊250，處理器120於字彙記憶體130儲存含有口語元素的語音表示法的字圖，連同當辨識之口語元素時對應於採用的動作之說明或指示。字圖可儲存於字彙記憶體130，字圖之儲存方式為其與對先前增加之口語元素產生的字圖共同鏈結，來形成涵蓋全部口語元素之語音表示法的單一字圖。選擇性地，動作之說明或指示可儲存於它處，特別全部的口語元素可關聯單一型動作，差異只在於特定細節。舉例言之，於實作使用語音控制來從事撥號的蜂巢式電話，較佳係由字彙記憶體130中刪除撥號動作的說明或指示，而當辨識出各個口語元素時，只儲存欲撥號的號碼。如同進一步選項，處理器120可於字彙記憶體130儲存文字表示法的本身，如同例如於SIVR系統，要求響應於語音命

令來顯示文字於顯示器126，或當允許使用者搜尋詞條串列中使用者特別希望修改或刪除的詞條。程序200止於方塊250完成時。

第4圖所示程序300之目的係辨識語音命令且基於語音命令動作。於開始時，程序300前進至方塊320。選擇性地，於開始時，程序300前進至方塊310，等候使用者按下輸入裝置122的開始鍵或類似鍵，或作動分開的手動控制，指示使用者將發出語音命令。隨後程序300前進至方塊320。

然後使用者發出先前使用程序200或其它手段定義的口語元素之一，發出語音命令，該語音命令由音訊輸入裝置106輸入，且被轉成輸入類比音訊信號110。音訊編碼解碼器114將對應於該語調內容的輸入類比音訊信號110轉成輸入數位音訊信號116，其隨後輸送至處理器120。於方塊320，處理器120檢查輸入數位音訊信號116，當處理器120檢測出接收得語調內容時，程序300前進至方塊330。

於方塊330，處理器120搜尋儲存於字彙記憶體130之字圖，找出最匹配所接收的語調內容的音標表示法。當口語元素有多於一種所接受的發音時，不同使用者可能以不同方式發音，或相同使用者於不同的場合可能以不同的方式發音，結果導致處理器120依據該語音命令的發音來選擇不同的字圖途徑。前述實例中，名字Stephen的正常發音係對應於S-T-IY-V-AH-N及S-T-EH-F-AH-N，始於第5圖所示範例字圖節點400及終於節點408及412。若使用者將名字Stephen發音成為S-T-IY-V-AH-N，則處理器120可能選擇始於節點

400而終於節點408的路徑作為最接近匹配所接收得的語調內容的語音表示法。另一方面，若使用者將名字Stephen發音成為S-T-EH-F-AH-N，則處理器120選擇始於節點400而終於節點412的路徑。為求完整，於未找到緊密匹配時，該  
5 程序可選擇性請求使用者重複該語音命令。為求清晰，本選擇性步驟係由第4圖流程圖中刪除。於方塊330完成時，程序300前進至方塊340。

於方塊340，處理器120將所選定的路徑描述的語音表示法轉成口語節段，藉由透過輸出數位音訊信號118傳輸給  
10 使用者，音訊編碼解碼器114可將該口語節段轉換成為類比信號112，以及送至音訊輸出裝置108。選擇性地，處理器120也可於顯示器126顯示對應於所辨識之口語元素的文字表示法，該文字為使用者先前於程序200方塊210執行期間所載入且已經儲存於字彙記憶體130的文字。此外，或替代  
15 顯示文字表示法，處理器120可顯示與該文字相關聯的其它資訊。於方塊340完成時，程序前進至方塊350。

於方塊350，處理器120由字彙記憶體130取還對應於該所辨識之口語元素之預定動作的描述，可利用控制信號134，傳輸對應命令給應用區塊136來初始化該動作。前述  
20 實例中，該實例特別可應用於應用區塊136為蜂巢式電話150的收發器140，處理器120命令收發器140建立其與特定遠端部分的連結。於本特定實例，該命令係撥號先前於程序200將名字Stephen加至字彙記憶體130已經與名字Stephen相關聯的電話號碼。選擇性地，於送出命令至應用

區塊136之前，處理器120首先等候使用者按下輸入裝置122的確認鍵或類似鍵來確認該選擇且來初始化該動作。另一個選擇性動作為處理器120等候一段預定時間，例如約2秒至5秒，於該段時間給予使用者機會來拒絕該選擇，藉按下  
5 輸入裝置122的取消鍵或類似鍵或作動分開的手動控制器來取消該項動作。為求簡明，此等選擇性步驟由第4圖之流程圖說明中刪除。程序300終於方塊350完成時。

於該系統之另一具體例中，將口語元素之文字表示法轉換成為語音表示法，以及決定當辨識各個口語元素時欲  
10 執行的動作之程序，可使用分開裝置排它地執行或額外執行，且可由SIVR系統刪除或未刪除。由SIVR系統刪除本程序，又可去除文字分錄的輸入裝置與顯示器的需求，也可減少所需系統記憶體容量，如此可縮小系統成本、尺寸及複雜度。此種系統之一實例為語者無關之語音控制玩具。

15 本具體例中，口語元素之語音表示法、及由分開裝置產生的欲與口語元素相關聯的動作可於系統製造前或系統製造中預先載入SIVR系統的字彙記憶體；或可於系統製造後或甚至已經佈署後，載入SIVR系統的字彙記憶體。舉例言之，語者無關之語音操作的行動蜂巢式電話可由屬於蜂  
20 巢式電話服務提供業者的伺服器、可由網際網路、可由另一蜂巢式電話、或可由藉纜線連結或無線鏈路連結之電腦下載語音表示法至其字彙記憶體。

於本具體例之變化例中，當辨識各個口語元素時，口語元素的文字表示法以及欲執行的動作可由分開裝置而載



入系統，且可由SIVR系統刪除或未刪除。舉例言之，語音操作的蜂巢式電話、或PDA與行動蜂巢式電話的組合可由藉纜線連結或無線鏈路連結之電腦下載欲撥號的連絡者姓名及電話號碼串列。

- 5        本發明之另一具體例中，口語元素之文字表示法可儲存於字彙記憶體，而當欲辨識語音命令而呼叫該文字表示法時，SIVR系統可於進行中將該字串轉換成為語音表示法。

- 於本發明之又一具體例中，口語元素可連鎖並置來產生單一語音命令。舉例言之，使用者說出口語元素「刪除」，  
10    SIVR系統作語音確認，隨後使用者說出名字「Stephen」，系統作語音確認，然後刪除與該名字「Stephen」相關的詞條。

- 讓處理器120執行本發明之具體例之方法之指令可儲存於裝置138之記憶體(圖中未顯示)，或儲存於電腦可讀取之儲存媒體，諸如軟碟、CD-ROM、個人電腦硬碟、CF記  
15    憶卡、PCMCIA記憶卡、伺服器硬碟、FTP伺服器硬碟、可由網際網路之網頁存取的網際網路伺服器硬碟等。

- 雖然於此處已經舉例說明若干本發明之特色，但熟諳技藝人士顯然易知可做出多項修改、取代、改變、及相當例。因此須了解隨附之申請專利範圍意圖涵蓋落入本發明  
20    之精髓之全部此等修改及變化。

### 【圖式簡單說明】

第1圖為根據本發明之一具體例，範例語者無關語音辨識系統之示意方塊圖；

第2圖為結合第1圖所示之語音辨識系統之範例行動蜂

巢式電話之示意方塊圖；

第3圖為增加詞條至第1圖所示語音辨識系統之方法之示意流程圖；

第4圖為使用第1圖所述語音辨識系統響應於語音命令  
5 方法之示意流程圖；以及

第5圖為範例字圖，顯示儲存於第1圖所述與語者無關語音辨識系統之字彙對應於不同的口語元素之語音表示法之各個路徑。

### 【主要元件符號說明】

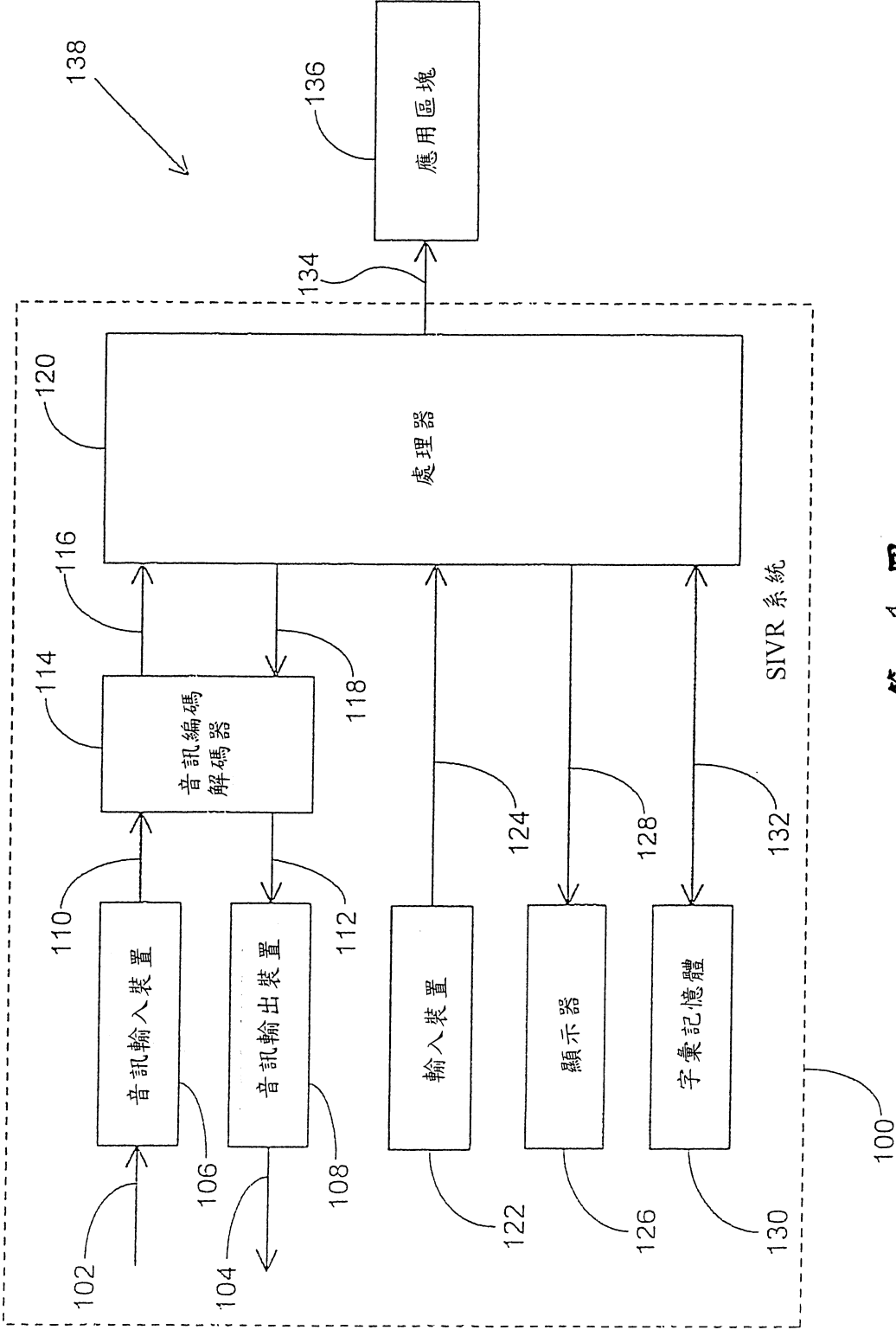
|                        |                  |
|------------------------|------------------|
| 100…語者無關語音辨識系統，SIVR 系統 | 130…字彙記憶體        |
| 102…聲音信號               | 132…記憶體匯流排       |
| 106…音訊輸入裝置             | 134…命令信號         |
| 108…音訊輸出裝置             | 136…應用區塊         |
| 110…輸入之類比音訊信號          | 138…語音控制裝置       |
| 112…輸出之類比音訊信號          | 140…收發器          |
| 112…輸出之數位化音訊信號         | 142…天線           |
| 114…音訊編碼解碼器            | 150…語音控制之行動蜂巢式電話 |
| 116…輸入之數位化音訊信號         | 200…程序           |
| 120…處理器                | 210-250…方塊       |
| 122…輸入裝置               | 300…程序           |
| 124…匯流排                | 310-350…方塊       |
| 126…顯示器                | 400-416…節點       |
| 128…顯示器匯流排             |                  |

## 五、中文發明摘要：

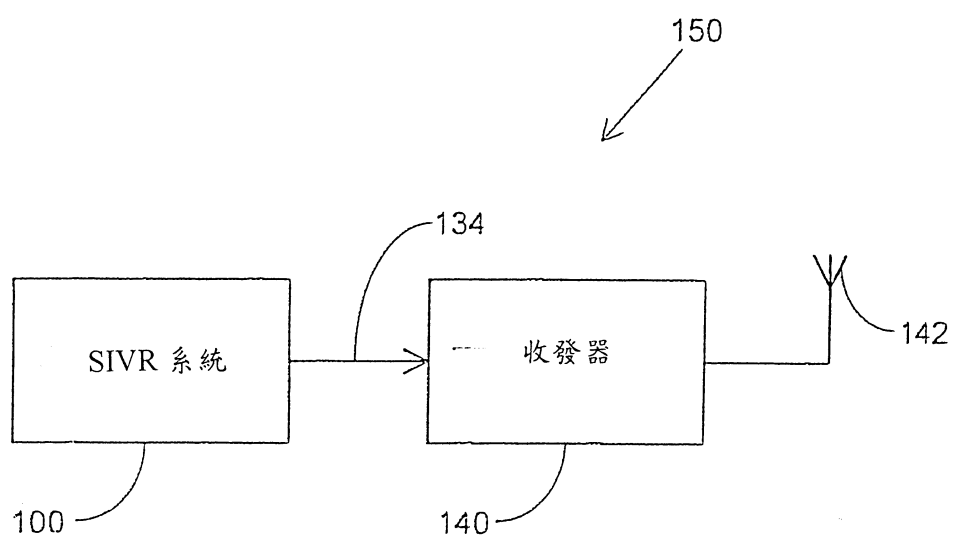
當語者無關語音辨識(SIVR)系統辨識出一說話語調匹配屬於一預先界定字彙之一口語元素的語音表示內容時，該SIVR系統可播放合成口語片段，作為讓使用者驗證該語調內容被正確辨識的手段。當於該字彙中之一口語元素有多於一項可能的發音時，該系統可選擇最接近匹配該使用者語調內容的聲音，以及播放對應於該特定語音表示內容之一合成口語片段。

## 六、英文發明摘要：

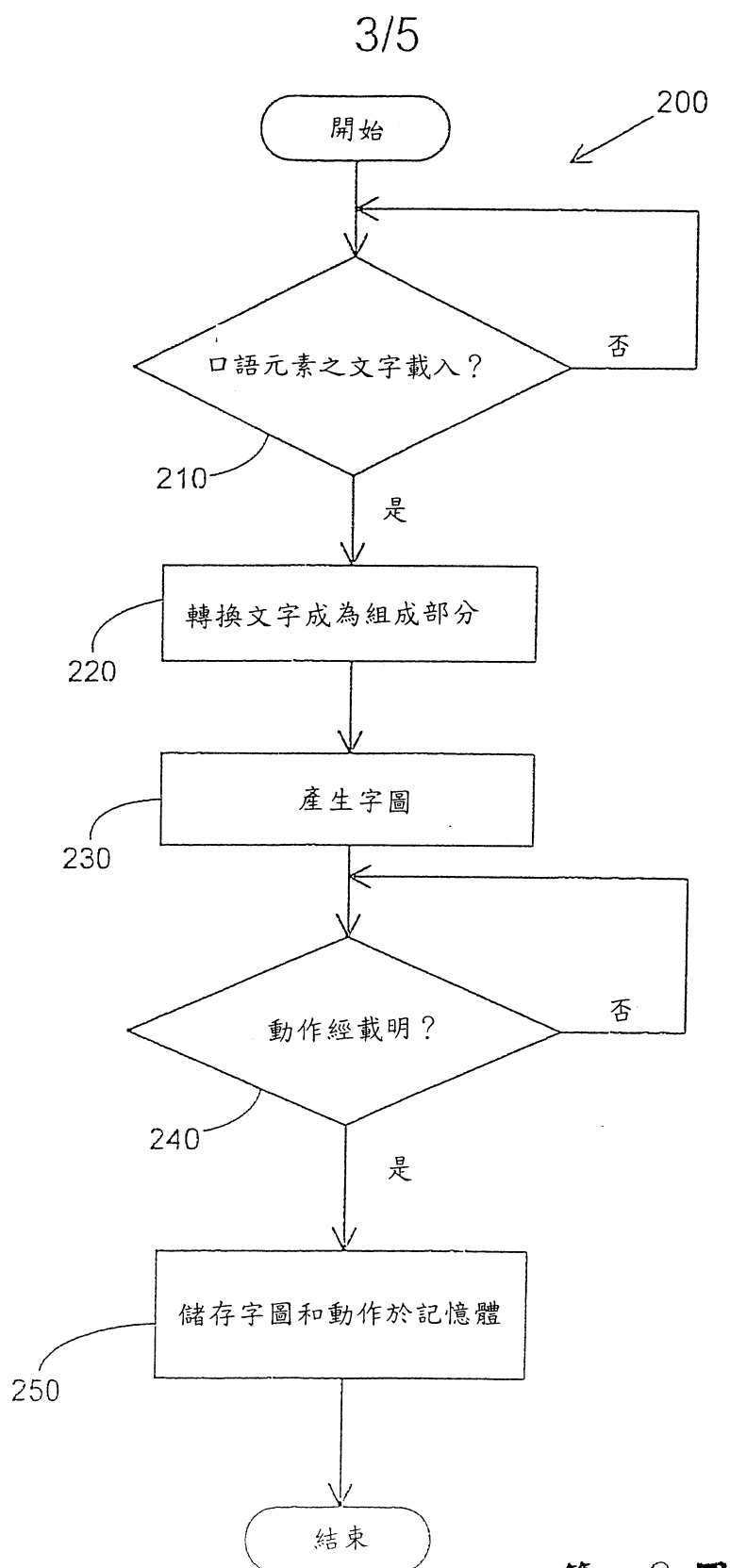
When a speaker-independent voice-recognition (SIVR) system recognizes a spoken utterance that matches a phonetic representation of a speech element belonging to a predefined vocabulary, it may play a synthesized speech fragment as a means for the user to verify that the utterance was correctly recognized. When a speech element in the vocabulary has more than one possible pronunciation, the system may select the one most closely matching the user's utterance, and play a synthesized speech fragment corresponding to that particular representation.



第 1 圖

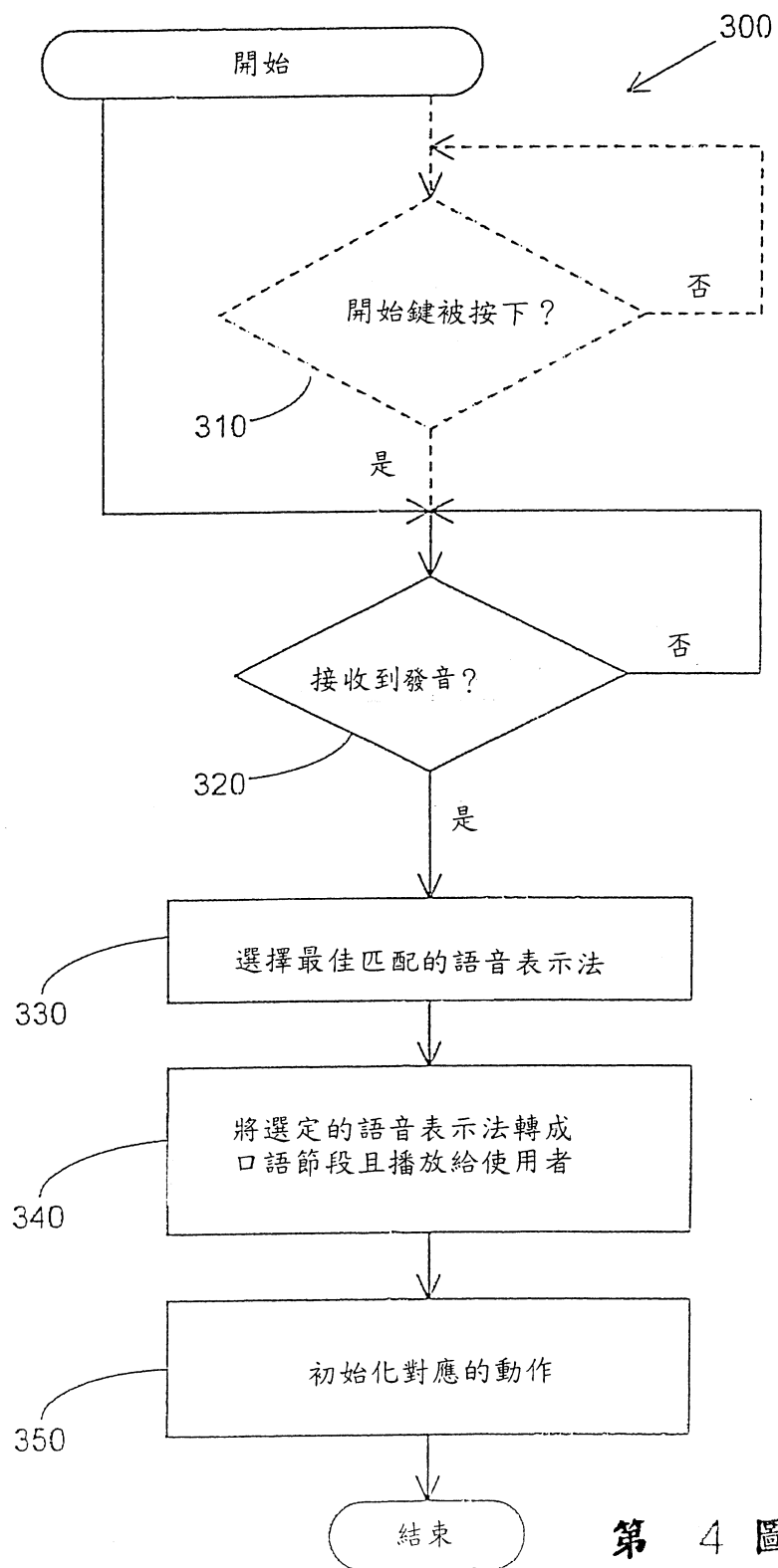


第 2 圖

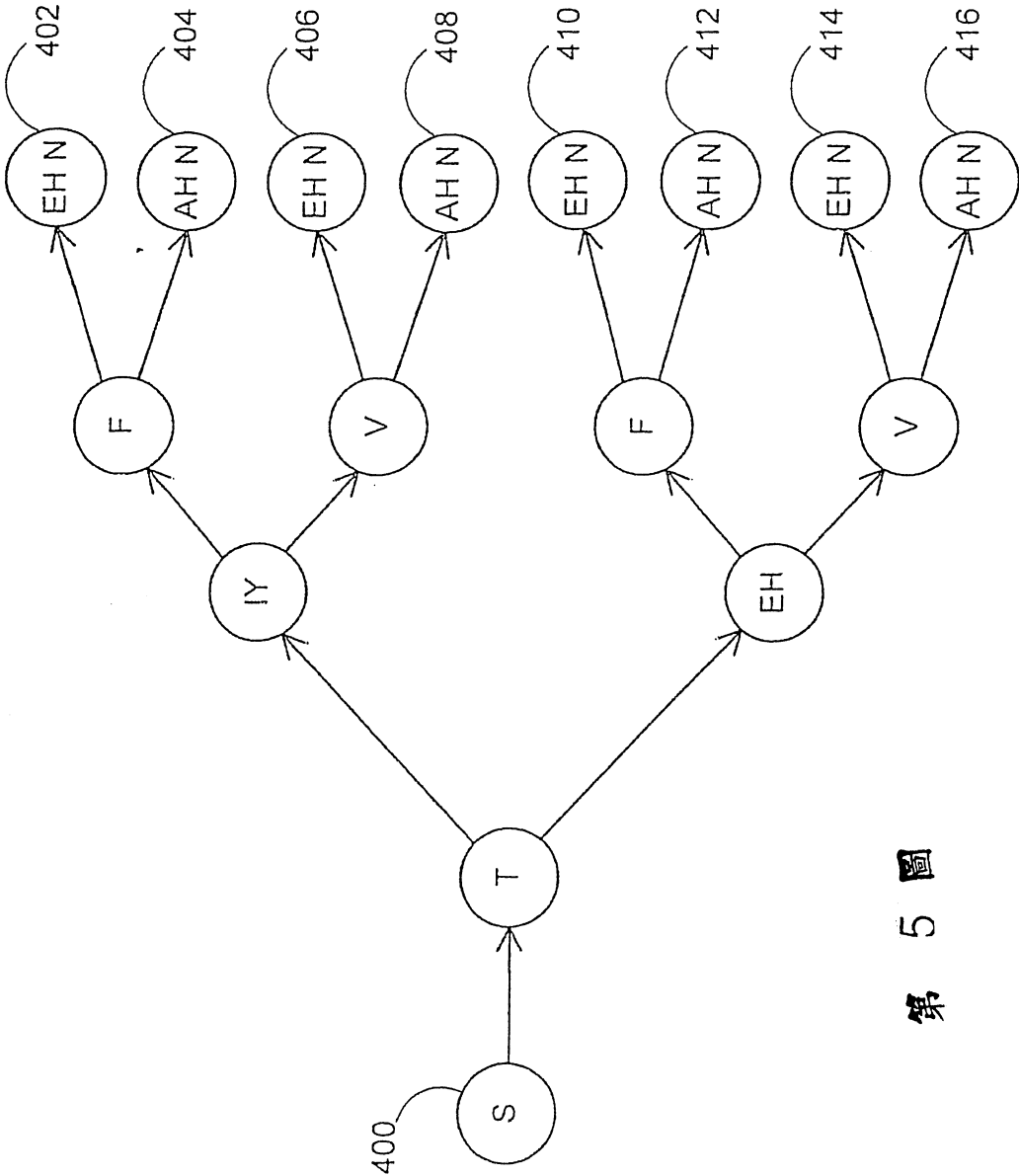


第 3 圖

4/5



第 4 圖



第 5 圖



七、指定代表圖：

(一)本案指定代表圖為：第 ( 4 ) 圖。

(二)本代表圖之元件符號簡單說明：

300…程序

310-350…方塊

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

## 十、申請專利範圍：

第94115348號申請案申請專利範圍修正本

95.07.10.

1. 一種用於語者無關語音辨識之方法，該方法包含有下列步驟：

5           選擇一預定字彙之口語元素之多個語音表示法中最密切匹配一語調內容之一種語音表示法，其中該等多個語音表示法包括具有不同可能發音之該等口語元素中之任一者之多種語音表示法；以及

10           根據該等語音表示法中之該一語音表示法，來合成一發聲口語片段。

2. 如申請專利範圍第1項之方法，進一步包含有下列步驟：  
儲存該等語音表示法。

15           3. 如申請專利範圍第1項之方法，進一步包含有下列步驟：  
由該等口語元素之文字表示法產生該等語音表示法。

4. 如申請專利範圍第1項之方法，進一步包含有下列步驟：  
將識別由該等該等語音表示法中最密切匹配該語調內容之該一語音表示法所表示之該口語元素的資訊顯示出來。

20           5. 如申請專利範圍第1項之方法，進一步包含有下列步驟：  
進行與該等口語元素之一相關聯之一預定動作。

6. 如申請專利範圍第2項之方法，其中儲存該等語音表示法之步驟進一步包含儲存該等語音表示法成一字圖。

7. 一種語者無關語音辨識裝置，該裝置包含：

一處理器，該處理器係用來選擇一預先界定字彙之口語元素之多個語音表示法中與對應於一語調內容之一輸入數位化語音信號之一部分最密切匹配之一語音表示法，其中該等多個語音表示法包括具有不同可能發音之該等口語元素中任一語元素之多種語音表示法；以及該處理器係另用來根據該等語音表示法中之該一語音表示法，而合成一輸出數位化語音信號。

8. 如申請專利範圍第7項之裝置，進一步包含：

用來儲存該等語音表示法之一記憶體。

9. 如申請專利範圍第8項之裝置，其中該記憶體係用來儲存該等語音表示法成為一字圖。

10. 如申請專利範圍第7項之裝置，其中該處理器係用來由該等口語元素之文字表示法產生該等語音表示法。

11. 如申請專利範圍第10項之裝置，進一步包含：

用來允許載入該等文字表示法之一輸入裝置。

12. 如申請專利範圍第7項之裝置，進一步包含：

一顯示器，

其中該處理器係欲於該處理器上把識別由該等語音表示法中最密切匹配該語調內容之該一語音表示法所表示之該口語元素的資訊顯示出來。

13. 如申請專利範圍第7項之裝置，其中該處理器係欲啟始與該等口語元素之一相關聯之一預定動作。

14. 一種語音操作之行動蜂巢式電話，包含：

一收發器；

一天線；以及

一處理器，該處理器係用來選擇一預先界定字彙之口語元素之多個語音表示法中與對應於一語調內容之一輸入數位化語音信號之一部分最密切匹配之一語音表示法，其中該等多個語音表示法包括具有不同可能發音之該等口語元素之任一語音元素之多個語音表示法；以及該處理器係用來根據該等語音表示法中之該一語音表示法，而合成一輸出數位化語音信號。

5

15. 如申請專利範圍第14項之語音操作之行動蜂巢式電話，進一步包含：

10

用來儲存該等語音表示法之一記憶體。

16. 如申請專利範圍第15項之語音操作之行動蜂巢式電話，其中該記憶體係用來儲存該等語音表示法成為一字圖。

15

17. 如申請專利範圍第14項之語音操作之行動蜂巢式電話，其中該處理器係用來由該等口語元素之文字表示法產生該等語音表示法。

18. 如申請專利範圍第17項之語音操作之行動蜂巢式電話，進一步包含：

20

用來允許載入該等文字表示法之一輸入裝置。

19. 如申請專利範圍第14項之語音操作之行動蜂巢式電話，其中該處理器係欲啟始與該等口語元素之一相關聯之一預定動作。

20. 如申請專利範圍第19項之語音操作之行動蜂巢式電

話，其中該預定動作進一步包括命令該收發器來與一特定遠方對象建立連結。

21. 一種包含電腦可讀儲存媒體之物品，該電腦可讀儲存媒體上儲存有指令，而該等指令當由一處理器執行時會導致進行下列動作：

選擇一預定字彙口語元素之多個語音表示法中最密切匹配一語調內容之一種語音表示法，其中該等多個語音表示法包括具有不同可能發音之該等口語元素中之任一者之多種語音表示法；以及

- 根據該等語音表示法中之該一語音表示法，來合成一發聲口語片段。

22. 如申請專利範圍第21項之物品，其中該等指令進一步導致進行下列動作：

儲存該等語音表示法。

23. 如申請專利範圍第21項之物品，其中該等指令進一步導致進行下列動作：

儲存該等語音表示法成為一字圖。

24. 如申請專利範圍第21項之物品，其中該等指令進一步導致進行下列動作：

- 由該等口語元素之文字表示法產生該等語音表示法。