I383377

公告本

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）

101年6月27日修正本

※ 申請案號：93121624

※ 申請日期：2004 年 7 月 20 日

※IPC 分類：
G10L 15/24 (2006.01)
H04R 5/04 (2006.01)

一、發明名稱：（中文/英文）

多重感測語音辨識系統及方法

MULTI-SENSORY SPEECH RECOGNITION SYSTEM AND METHOD

二、申 請 人：（共 1 人）

姓名或名稱：（中文/英文）

美商．微軟公司

Microsoft Corporation

代表人：（中文/英文）

艾苹那諾爾 D 巴特萊

EPPENAUER, D. BARTLEY

住居所或營業所地址：（中文/英文）

美國華盛頓州列德蒙微軟路 1 號

One Microsoft Way, Building 8, Redmond, WA 98052-6399, U.S.A.

國　籍：（中文/英文）

美國／U.S.A.

三、發明人：（共 5 人）

姓　名：（中文/英文）

1.黃學東 D/HUANG, XUEDONG D.

2.劉子誠/LIU, ZICHENG

3.張正富/ZHANG, ZHENGYOU

4.辛克萊麥克 J/SINCLAIR, MICHAEL J.

5.亞塞羅亞力詹德/ACERO, ALEJANDRO

國　籍：(中文/英文)

1.美國/USA

2.美國/USA

3.法國/France

4.美國/USA

5.西班牙/Spain

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.；2003 年 7 月 29 日；10/629,278

2.；2003 年 8 月 7 日；10/636,176

☐ 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。

五、中文發明摘要：

本發明組合一傳統音訊麥克風與一額外語音感應器，其基於一輸入提供語音感應信號。該語音感應信號係基於說話者於說話時之動作，例如臉部動作、骨頭振動、喉嚨振動、喉嚨阻抗改變等加以產生。一語音檢測器元件接收來自該語音感應器的輸入並輸出一語音檢測信號，其表示是否該使用者正在說話。該語音檢測器基於該麥克風信號及語音感應信號，產生該語音檢測信號。

六、英文發明摘要：

The present invention combines a conventional audio microphone with an additional speech sensor that provides a speech sensor signal based on an input. The speech sensor signal is generated based on an action undertaken by a speaker during speech, such as facial movement, bone vibration, throat vibration, throat impedance changes, etc. A speech detector component receives an input from the speech sensor and outputs a speech detection signal indicative of whether a user is speaking. The speech detector generates the speech detection signal based on the microphone signal and the speech sensor signal.

七、指定代表圖：

(一)、本案指定代表圖為：第 3 圖。

(二)、本代表圖之元件代表符號簡單說明：

300 語音檢測系統

301 語音換能器

302 捕捉元件

303 音訊麥克風

304 信號處理機

306 語音檢測信號

308 輸出信號

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

九、發明說明：

【發明所屬之技術領域】

本發明係有關於音訊輸入系統。更明確地說，本發明係關於在一多感應換能器輸入系統中之語音處理。

【先前技術】

於很多不同語音辨識應用中，很重要也很講究要有一清楚及原音音訊輸入，以表示予以提供至自動語音辨識系統中之語音。會敗壞輸入至語音辨識系統中之音訊的雜音有兩類型，即為環境雜訊及來自背景語音所產生之雜訊。已經有很多密集工作，用以取消來自音訊輸入之環境雜訊。部份技術已經在音訊處理軟體中可以商業購得，或者，部份已被整合在數位麥克風中，例如通用串列匯流排(USB)麥克風中。

應付有關背景語音之雜訊會有比較多之問題。這是由於有很多不同吵雜環境所造成。例如，當有有關發言者正在群眾中或很多人之間發言時，一傳統麥克風經常會取得不是有關發言者之語音。基本上，在其他人正在發言之環境中，產生自有關發言者的音訊信號會被犧牲掉。

用應付背景語音之先前解決方案為提供一 on/off 開關在一耳機線或手機上。該 on/off 開關被稱為"按下講話"按鈕及使用者需要在說話前按住該按鈕。當使用者按下按鈕時，會產生按鈕信號。該按鈕信號指示該語音信號辨識系統，有關發言者正在說話，或將要說話。然而，部份使用

報告指示此類型之系統並不是使用者所滿意或想要的。

另外，也有很多努力想要以分離開麥克風所拾取之背景出聲者與有關發言者(或前台發言者)。這在無雜音辦公室環境可以合理良好動作，但對於高雜音環境中，並不理想。

於另一先前技術中，來自標準麥克風之信號被組合以來自喉式麥克風之信號。該喉式麥克風藉由量測在發音時於喉嚨間之電阻抗間之變化，而間接記錄喉頭動作。為喉式麥克風所產生之信號被組合以傳統麥克風，並且，產生模組化組合信號之頻譜內容之模型。

一演繹法被用以映圖該吵雜之組合之標準及喉式麥克風信號特性成為一無雜音標準麥克風特性。這是藉由或然最佳過濾法加以估算。然而，喉式麥克風對於背景雜訊相當有免疫性，但喉式麥克風信號之頻譜內容相當有限。因此，使用它以映圖為一無雜音估算特性向量係不夠高度精確。此技術係被說明於 Frankco 等人之用以雜訊強辨識之組合異質感應器與標準麥克風，發表於 2001 年美國佛州奧蘭多之 DARPA RPAR 工坊。另外，穿載一喉式麥克風對使用者來說，增加了額外之不便。

【發明內容】

本發明組合傳統音訊麥克風與一額外語音感應器，其基於一額外輸入提供一語音感應信號。該語音感應信號係基於說話者於說話時之動作，例如臉部動作、骨頭振動、

喉嚨振動、喉嚨阻抗改變等加以產生。一語音檢測器元件接收來自該語音感應器的輸入並輸出一語音檢測信號，以表示是否該使用者正說話。該語音檢測器基於該麥克風信號及語音感應信號，產生該語音檢測信號。

於一實施例中，該語音檢測信號被提供給一語音辨識引擎。該語音辨識引擎提供一辨識輸出，以表示基於麥克風信號之來自音訊麥克風之麥克風信號與來自額外語音感應器之語音檢測信號。

本發明同時也實施為一檢測語音的方法。該方法包含產生一第一信號，其表示以音訊麥克風輸入之音訊；產生一第二信號，表示一使用者為一顏面動作感應器所感應到之使用者表情動作；及基於該第一及第二信號，檢測是否該使用正發言。

於一實施例中，該第二信號包含使用者之頸部之振動或阻抗改變，或者使用者頭或顎振動。於另一實施例中，第二信號包含一影像，以表示使用者嘴部之動作。於另一實施例中，一例如熱敏電阻之溫度感應器被放置在呼吸氣流中，例如在麥克風旁之桿上並感應語音為溫度變化。

【實施方式】

本發明關係於語音檢測。更明確地說，本發明有關於一多感應換能器輸入並產生一輸出信號，以基於所捕獲之多感應輸入表示是否一使用者正在說話。然而，在討論本案之細節之前，本發明可以使用之環境的實施例係加以討

論。

第 1 圖例示一可以實行本發明之適當計算系統環境 100 之例子。計算系統環境 100 係為一適當計算環境例，並不用以限制本發明之使用範圍。該計算環境 100 也不應被解釋為對例示作業環境 100 中所示之元件或其組合之要件。

本發明係可以與各種其他目的或特殊計算系統環境或架構一起操作。可以適當地用於本發明之已知之計算系統、環境、及/或架構之例子包含但並不限定於個人電腦、伺服器電腦、手持式或膝上型裝置、多處理機系統、微處理機為主之系統、機上盒、可程式消費電子、網路 PC、迷你電腦、主機電腦、分散式計算環境其包含上述系統或裝置等等。

本發明可以說明於電腦可執行指令之一般文件中，例如可以為一電腦所執行之程式模組。一般而言，程式模組包含常式、程式、物件、元件、資料結構等等，其執行特定工作或實行特定抽象資料類型。本發明也可以實施於分散式計算環境中，其中，工作係為遠端處理裝置所執行，該等遠端裝置係經由一通訊網路加以鏈結。於一分散式計算環境中，程式模組可以是本地或遠端電腦儲存媒體，其包含記憶體儲存裝置。

參考第 1 圖，一用以執行本發明之例示系統包含一個以電腦 110 形式之一般目的計算裝置。電腦 110 之元件可以包含但並不限定於一處理單元 120、一系統記憶體

1330、及一系統匯流排 121，其連接包含系統記憶體之各種系統元件至處理單元 120。系統匯流排 121 可以為任意類型之匯流排結構，其包含一記憶體匯流排或記憶體控制器、一週邊匯流排、及一本地匯流排，使用任意之匯流排架構。例如，但並不限定於包含工業標準架構(ISA)匯流排、微通道架構(MCA)匯流排、加強 ISA(EISA)匯流排、視訊電子標準協會(VESA)區域匯流排、及稱為 Mezzanine 匯流排之週邊組件互連(PCI)匯流排的架構。

電腦 110 典型包含各種電腦可讀取媒體。電腦可讀取媒體可以為電腦 110 可讀取之任意媒體，包含揮發性及非揮發性媒體、可移除及非可移除媒體。例如，但並不限定於電腦可讀取媒體可以包含電腦儲存媒體及通訊媒體。電腦儲存媒體包含揮發性及非揮發性，可移除及非可移除媒體，其係可以以任意資訊儲存之方法或技術加以執行，該資訊係例如電腦可讀取指令、資料結構、程式模組或其他資料。電腦儲存媒體包含但並不限定於 RAM、ROM、EEPROM、快閃記憶體或其他記憶體技術、CD-ROM、數位多功能光碟(DVD)、或其他光碟儲存、磁匣、磁帶、磁碟儲存或其他磁儲存裝置、或其他可以用以儲存想要資訊之媒體及可以為電腦 100 所存取之媒體。通訊媒體典型實現為電腦可讀取指令、資料結構、程式模組或其他呈調變資料信號之資料，例如載波 WAV 或其他傳輸機制並包含任意資訊輸送媒介。名稱"調變資料信號"表示一信號，其特徵之一或多數特徵被設定或改變，以在該信號中編碼該資

訊。例如，但並不限定於，通訊媒體包含有線媒體，例如有線網路或直接接線連接、及無線媒體，例如音響、RF、紅外線及其他無線媒體。上述任意組合應包含在電腦可讀取媒體之範圍內。

系統記憶體 130 包含電腦儲存媒體，以揮發性及/或非揮發記憶體，例如唯讀記憶體(ROM)131 及隨機存取記憶體(RAM)132 之形式。一包含有基本常式，以於開機時，協助資訊傳送於電腦 110 內之元件間之基本輸入/輸出系統 133(BIOS)典型儲存於 ROM131 之內。RAM132 典型包含資料及/或程式模組，其係可以為處理單元 120 所立即存取及/或正在操作者。例如，但並不限定於，第 1 圖例示作業系統 134、應用程式 135、其他程式模組 136 及程式資料 137。

電腦 110 可以包含其他可移除/非可移除揮發/非揮發電腦儲存媒體。例如，第 1 圖例示一硬碟機 141，其由非可移除非揮發性磁碟媒體讀取或寫入、一磁碟機 151，其由可移除非揮發性磁碟 152 讀取或寫入、及一光碟機 155，其由可移除非揮發性光碟 156 讀取或寫入，光碟係例如 CD-ROM 或其他光學媒體。其他可以用於例示作業環境中之可移除/非可移除，揮發/非揮發電腦儲存媒體包含但並不限定於磁帶匣、快閃記憶體卡、數位多功能光碟、數位視訊帶、固態 RAM、固態 ROM 等等。硬碟機 141 典型經由一非可移除記憶體介面，例如介面 140 連接至系統匯流排 121，以及，磁碟機 151 及光碟機 155 典型藉由一可移

除記憶體介面，例如介面 150 連接至系統匯流排 121。

上述例示於第 1 圖中之這些裝置及其相關電腦儲存媒體提供了電腦可讀取指令、資料結構、程式模組及其他用於電腦 110 之資料的儲存。於第 1 圖中，例如，硬碟機 141 係被例示為儲存作業系統 144、應用程式 145、其他程式模組 146、及程式資料 147。注意這些元件可以與作業系統 134、應用程式 135、其他程式模組 136 及程式資料 137 相同或不同。作業系統 144、應用程式 145、其他程式模組 146、及程式資料 147 係被給予不同號碼，以作為顯示其為不同拷貝。

一使用者可以例如經由鍵盤 162、一麥克風 163、及一指標裝置 161，如滑鼠、軌跡球或觸控板之輸入裝置，將命令及資訊輸入電腦 110。其他輸入裝置(未示出)可以包含搖桿、遊戲板、衛星碟、掃描器等等。這些及其他輸入裝置係經常經由一使用者輸入介面 160 連接至處理單元 120，該介面係連接至系統匯流排，但也可以為其他介面及匯流排結構所連接，例如一平行埠、遊戲埠或通用串列匯流排(USB)。一監視器 191 或其他類型之顯示裝置也經由一例如視訊介面 190 之介面連接至該系統匯流排 121。除了監視器外，電腦也包含其他週邊輸出裝置，例如喇叭 197 及印表機機 196，其可以經由一輸出週邊介面 195 加以連接。

電腦 110 可以使用邏輯連接至例如一遠端電腦之一或多數遠端電腦，而操作於一網路環境中。遠端電腦 180 可

以為一個人電腦、一手持裝置、一伺服器、一路由器、一網路PC、一同等裝置或其他公用網路節點，並典型包含很多或全部上述有關電腦110所述之元件。示於第1圖之邏輯連接包含一區域網路(LAN)171及一廣域網路(WAN)173，但也可以包含其他網路。此等網路環境係可在辦公室、企業電腦網路、網內網路及網際網路所常見。

當用於LAN網路環境中時，電腦110經由一網路介面或轉接器170連接至LAN171。當用於WAN網路環境中時，電腦110典型包含一數據機172或其他機構，用以在WAN173，例如網際網路上建立通訊。可以內建或外部之數據機172可以經由使用者輸入介面160或其他適當機制連接至系統匯流排121。於一網路環境中，相關於電腦110之程式模組或其部份可以被儲存在遠端記憶體儲存裝置中。例如，但並不限定於，第1圖例示內藏在遠端電腦180上之應用程式185。應了解，所示網路連接係為例示性，其他之用以在電腦間建立通訊鏈路之機構也可以使用。

應注意的是，本發明可以執行在例如有關第1圖所示之電腦系統上。然而，本發明也可以執行在一伺服器、一專用以信息處理之電腦上，或者，在一分散式系統上，其中，本發明之不同部份可以執行在該分散式計算系統之不同部件中。

第2圖例示一本發明可以使用之示範性語音辨識系統之方塊圖。於第2圖中，一喇叭400發聲給麥克風404。為麥克風404所檢測之音訊信號係被轉換為電氣信號，這

些信號被提供至類比至數位(A/D)轉換器 406。

A/D 轉換器 406 將來自麥克風 404 之類比信號轉換為一連串之數位值。於幾個實施例中，A/D 轉換器 406 以 16KHz 及每取樣 16 位元，來取樣該類比信號，藉以建立每秒 32 千位元組之語音資料。這些數位值係被提供給一訊框建構器 407，其隨後，於一實施例中，將這些值收集 25 毫秒訊框，其隔 10 秒後開始。

為訊框建構器 407 所建立之訊框資料被提供給特性抽取器 408，其由每一訊框抽出一特性。特性抽取模組之例子包含用以執行線性預測編碼(LPC)、LPC 導出倒頻譜、預測線性預測(PLP)、稽核模型特性抽取、及梅爾倒頻譜係數(MFCC)特性抽出。注意本發明並不限定於這些特性抽取模組，並且，其他模組也可以在本發明之範圍內使用。

特性抽取模組 408 產生一串流之特性向量，其個別相關於該語音信號之一訊框。此串流特性向量被提供給一解碼器 412，其基於一串流之特性向量、一辭彙 414、一語言模型 416(例如基於一 N-元詞、上下文無關文法、或其混合)、及音響模型 418，指示最像字元序列。用於解碼之特定方法對於本發明並不重要。然而，本發明之態樣包含對音響模型 418 之修改及其使用。

假設字元之最可能順序可以被提供給一選擇信心量測模組 420。信心量測模組 420 指示哪些字元可能為語音辨識器所不當地辨識出。這可以部份基於一二次音響模型(未示出)。信心量測模組 420 然後提供假設字元順序給一輸出

模組，並以識別碼指示哪些字元可能被不當識別出。熟習於本技藝者可以知道信心量測模組420對於本發明之實施並不必要。

於訓練時，一相關於訓練本文426之語音信號係被輸入至解碼器412，並具有訓練本文426之辭彙複本。訓練器424基於訓練輸入，而訓練音響模型418。

第3圖例示依據本發明一實施例之語音檢測系統300。語音檢測系統300包含語音感應器或換能器301、傳統音訊麥克風303、多感應信號捕捉元件302及多感應信號處理機304。

捕捉元件302捕捉來自傳統麥克風303呈音訊信號形式之信號。元件302同時也捕捉來自語音換能器301之輸入信號，其指示一使用者正在說話者。由此換能器所產生之信號可以由各種其他換能器所產生。例如，於一實施例中，換能器為一紅外線感應器，其針對使用者之臉，即嘴巴區域，並產生一信號以表示相對於語音之使用者的臉部動作。於另一實施例中，感應器包含多數紅外線發射器及感應器針對於使用者臉上不同部份。於另一實施例中，語音感應器或多數感應器301可以包含一喉式麥克風，其量測經過使用者喉嚨之阻抗或喉嚨振動。於另一實施例中，該感應器係為一骨頭振動感應麥克風，其位在使用者之面或頭骨(例如顎骨)並感應對應於使用者所產生之語音的振動。此類型之感應器可以放置為與喉嚨接觸，或鄰近或在該使用者之內。於另一實施例中，一例如熱敏電阻之溫度

感應器被放置在呼吸流中，例如在支撐正常麥克風之相同支撐台上。當使用者說話時，所呼出之氣息造成在感應器中之溫度變化，因而，檢測出該語音。這可以藉由傳送一小量穩定狀態電流經該熱敏電阻、將之略微加熱高出室溫加以加強。此呼吸流然後傾向於冷卻該熱敏電阻，這可以藉由在熱敏電阻間之電壓變化加以感應出。於任一情況下，換能器 301 被例示為對背景語音不靈敏但會強烈指示是否使用者正在說話。

於一實施例中，元件 302 捕捉來自換能器 301 及麥克風 303 之信號，並將之轉換為數位形式，作為信號取樣之同步化時間序列。元件 302 然後提供一或多數輸出給多感應信號處理機 304。處理機 304 處理為元件 302 所捕捉之輸入信號，並在其輸出提供語音檢測信號 306，其表示是否該使用者正在說話。處理機 304 可以選用地輸出其他信號 308，例如一音訊輸出信號，或例如語音檢測信號，其基於來自各不同換能器之信號，而指示使用者正在說話之或然率。其他輸出 308 將基於予以執行之工作加以變化。然而，於一實施例中，輸出 308 包含一加強音訊信號，其用以語音辨識系統中。

第 4 圖例示多感應信號處理機 304 之其他細節。於第 4 圖所示之實施例中，處理機 304 將參考來自換能器 301 之換能器輸入加以說明，該輸入係為來自接近使用者臉部之紅外線感應器所產生之紅外線信號。當然，可以了解的是，第 4 圖之說明也可以容易地適用至來自一喉嚨感應

器、振動感應器等等之換能器信號。

第4圖顯示該處理機304包含紅外線(IR)為主語音檢測器310、音訊為主語音檢測器312、及其組合語音檢測器314。IR為主語音檢測器310接收為一IR發射器所發射之IR信號並接收為說話者所反射之IR信號，基於該IR信號，而檢測該使用者是否正在說話。音訊為主語音檢測器312接收該音訊信號並基於音訊信號檢測是否該使用者正在說話。來自檢測器310及312之輸出被提供給組合之語音檢測元件314。元件314接收諸信號並基於該兩輸入信號，完成整體評估，判斷是否該使用者正在說話。來自元件314之輸出包含語音檢測信號306。於一實施例中，語音檢測信號306被提供給背景語音移除元件316。語音檢測信號306被用以指示在音訊信號中，何時使用者正在實際說話。

更明確地說，於一實施例中，兩獨立檢測310及312均會產生使用者正在說話之或然率說明。於一實施例中，基於IR輸入信號，IR為主語音檢測器310之輸出為一使用者正在說話的或然率。同樣地，來自音訊為主語音檢測器312之輸出信號係為基於該音訊輸入信號，該使用者正在說話的一或然率。該兩信號然後在元件314被考量，以於一例子中，一二進制決定，以決定是否使用者正在說話。

信號306可以用以進一步在元件316中處理音訊信號，以移除背景語音。於一實施例中，當語音檢測信號306指示使用者正說話時，信號306被簡單用以經由元件316，

提供語音信號給該語音辨識引擎。若語音檢測信號306指示使用者未說話，則語音信號並未經元件316提供給語音辨識引擎。

於另一實施例中，元件314提供語音檢測信號306，作為指示使用者正在說話之或然率之量測值。於該實施例中，音訊信號在元件316中被乘以在語音檢測信號306中所實施之或然率。因此，當使用者正說話之或然率高時，經由元件316所提供給語音辨識引擎之語音信號也具有大振幅。然而，當使用者正說話之或然率低時，經由元件316提供給語音辨識引擎之語音信號之振幅很小。當然，於另一實施例中，語音檢測信號306可以簡單地直接提供給語音辨識引擎，其本身可以決定是否使用者正在說話，並且，基於該決定，如何處理該語音信號。

第5圖示出多感應信號處理機304之另一實施例的細節。除了令多數檢測器以檢測是否一使用者正在說話外，示於第5圖之實施例例示處理機304係呈一單一聯合式語音檢測器320。檢測器320接收IR信號及音訊信號並基於該兩信號，決定該使用者是否正在說話。於該實施例中，特性係首先被個別由紅外線及音訊信號抽取，然後，這些特性被饋送至檢測器320。基於所接收到之特性，檢測器320檢測是否該使用者正在說話並對應地輸出語音檢測信號306。

不管正在使用之系統類型(第4圖所示之系統或第5圖所示之系統)為何，語音檢測器可以使用訓練資料加以產

生及訓練，訓練信號中提供有一吵雜音訊信號與該 IR 信號，及一人工指示(例如一按下說話信號)，其指示是否使用者正在說話。

為了更易說明，第 6 圖顯示一音訊信號 400 及一紅外線信號 402 之繪圖，以大小對時間加以表示。第 6 圖同時也顯示語音檢測信號 404，其指示何時使用者正在說話。當於邏輯高狀態時，信號 404 係被語音檢測器之決定所表示該喇只正在發話。當為邏輯低狀態時，信號 404 表示使用者未在說話。為了決定是否一使用者正在說話並基於信號 400 及 402 產生信號 404，信號 400 及 402 之平均及變化量被週期地計算，例如約每 100 毫秒。平均及變化量之計算係用作為基準平均及變化值，以用以完成語音檢測決定。可以看出，當使用者說話時，較使用者不說話時，音訊信號 400 及紅外線 402 具有較大變化量。因此，當觀看值被處理時，例如每 5 至 10 毫秒，於觀看時之信號的平均及變化(或即變化量)被與基準平均值及變量(或基準變化量)相比。若觀看之值大於基準值，則決定該使用者正說話。若否，則決定該使用者未說話。於一例示實施例中，語音檢測決定基於所觀看值超出基準值一預定臨限而加以完成。例如，在每一觀看中，若紅外線信號並未在基準平均之三個標準偏差內，則被認為使用者正在說話。相同也可以用於音訊信號。

依據本發明之另一實施例，檢測器 310、312、314 或 320 也可以在使用時，例如容許在環境光條件中之變化，

或者，使用者之頭位置的變化，這可能使得光略微改變，而影響IR信號。基準平均及變化量可以每5至10秒再重估，例如使用另一循環時間視窗。這允許這些值被更新，以反映隨著時間上之變化。同時，在基準平均前及變化量係使用移動視窗加以更新，其可以首先決定是否輸入信號對應於使用者正在說話者否。此平均及變化量可以使用信號之部份加以再計算，其對應於使用者未在說話者。

另外，由第6圖中看出，IR信號可以大致進行在音訊信號之前。這是因為一般而言，使用者可以在發聲之前，改變嘴或臉部位置。因此，這允許系統在語音信號可得之前，檢測語音。

第7圖為依據本發明之IR感應器及音訊麥克風之一實施例示意圖。於第7圖中，一頭戴組420被提供有一對耳機422及424，與一延伸桿426。延伸桿426之末端具有一傳統音訊麥克風428與一紅外線收發器430。收發器430可以如所示為一紅外線發光二極體(LED)及紅外線接收器。當使用者在演說時，移動其臉、即嘴時，光由使用者之臉反射，即嘴反射，並且，出現在IR感應器信號會改變，如第6圖所示。因此，可以基於該IR感應器信號，決定該使用者是否正在說話。

應了解的是，雖然第7圖例示一單一紅外線收發器，但本發明也可以想出使用多數紅外線收發器。於該實施例中，有關於由每一紅外線所產生之IR信號之或然率可以分開或同時處理。若被分開處理，則簡單投票邏輯可以用以

決定是否紅外線信號表示說話者正在說話。或者，一或然率模型可以用以基於多數IR信號，來決定是否使用者正在說話。

如上所討論，其他換能器301可以採紅外線換能器以外之形式。第8圖為一頭戴組450之示意圖，其包含頭戴架451，耳機452及454，與一傳統音訊麥克風456，另外，一骨頭感應麥克風。麥克風456及458均可以機械並堅固地連接至頭戴組 451。當臉部骨頭振動傳送經說話者頭骨時，骨頭感應麥克風458將在臉部骨頭之振動轉換為電子語音信號。這些類型之麥克風為已知並可以買到各種大小與形狀之麥克風。骨頭感應麥克風458典型為形成為一接觸麥克風，其係穿載在頭骨頂或耳後(以接觸突起)。骨頭傳導麥克風對於骨頭之振動很靈敏，對於外部聲音來源不靈敏。

第9圖例示多數信號，其包含來自傳統麥克風456之信號460、來自骨頭敏感麥克風458之信號462及相關於一語音檢測器輸出之二進制語音檢測信號464。當信號464為邏輯高狀態時，其表示檢測器已決定該說話者正在發聲。當其於邏輯低狀態時，這對應於說話者未說話。第9圖之信號係當一使用者正穿載第8圖所示之麥克風系統，並具有背景音訊時，由資料被收集之環境所捕獲者。因此，音訊信號460顯示當使用者未說話時之重要動作。然而，骨頭敏感麥克風信號462顯示除了使用者正說話時以外之可忽略信號動作。因此，可以看出，只考量音訊信號460

很困難決定是否使用者正在說話。然而，當使用來自骨頭敏感麥克風時，不論單獨使用或配合音訊信號，均可以容易決定使用者正在說話否。

第 10 圖顯示本發明之另一實施例，其中一頭戴組 500 包含一頭戴件 501、一耳機 502 與傳統音訊麥克風、及一喉式麥克風 506。兩麥克風均機械連接至頭戴件 501 並可以堅固連接至其上。其中有各種不同喉式麥克風可以使用。例如，現行有單元件及雙元件設計。兩者均藉由感應喉嚨的振動並將該等振動轉換為麥克風信號加以動作。喉式麥克風係如所示穿戴於頸旁並藉由一彈性帶或頸帶所固定於定位。當感應元件定位在使用者之喉頭在喉節的一側時，它們可以良好動作。

第 11 圖顯示本發明另一實施例，其中，頭戴組 550 包含一耳內麥克風 552，與一傳統音訊麥克風 554。於第 11 圖所示之實施例中，耳內麥克風 552 係與一耳機 554 一體成型。然而，應注意的是，該耳機也可以由個別元件形成，並與耳內麥克風分開。第 11 圖也顯示傳統音訊麥克風 554 被實施為一貼近式麥克風藉由一延長桿 556 連接至耳內麥克風 552。延長桿 556 可以硬式或可彎式。於頭戴組 550 中，頭戴組之頭戴部份包含耳內麥克風 552 及選用耳機 554，其將頭戴組 550 經由與說話者耳內之磨擦連接而安裝至說話者之頭上。

耳內麥克風 552 感應語音振動，其經由說話者耳導管或經由包圍說話者耳導管之骨頭，或兩者而加以傳送。該

系統以類似於第 8 圖所示之骨頭敏感麥克風 458 之方式，對頭戴組動作。為耳內麥克風 552 所感應之語音振動係為麥克風信號，其係被用於後續處理。

雖然語音感應器或換能器 301 之若干實施例已經加以說明，但可以了解的是，其他語音感應器或換能器也可以使用。例如，電荷耦合裝置(或數位相機)可以以類似於 IR 感應器之方式加以動作。再者，喉頭感應器也可以使用。上述實施例係只以例子方式加以說明。

使用音訊及/或語音感應器信號，作檢測語音之另一技術現在將加以說明。於例示實施例中，一長條圖說明針對在一使用者指定時間內(例如一分鐘內等)，最近訊框之所有變化。對於隨後之每一觀看，該變量被針對所有輸入信號加以計算並比較該長條圖，以決定是否現行訊框表示說話者正在說話否。然後，長條圖被更新。應注意的是，若現行訊框被簡單地插入長條圖中及最舊的訊框被移除，則長條圖只表示使用者正在說話一長時間段時之說話訊框。為了處理此狀況，在長條圖中之說話及未說話訊框數量被追蹤，及長條圖被選擇地更新。若現行訊框被分類為說話，而在長條圖中之說話訊框的數量超出訊框總數之一半以上，則現行訊框並不會被插入該長條圖中。當然也可以使用其他更新技術，此技術係只作例示目的。

本系統可以用於各種應用中。例如，很多現行按下說話系統需要使用者按下並壓住輸入致動器(例如一按鈕)，以進行語音模式互動。使用報告顯示使用者很難滿意如此

之操作模式。同樣地，使用者於按壓硬體按鈕時開始說話，造成在發言開始之啪噠聲。因此，本發明除了按壓說話系統外，也可以簡單地用於語音辨識中。

同樣地，本發明也可以用以移除背景語音。背景語音已經被指明為共同雜訊源，其為電話鈴聲及空調所跟隨。使用上述之語音檢測信號，可以免除很多此背景雜訊。

同樣地，也可以改良變化速率語音編碼系統。因為本發明提供一輸出以指示是否使用者正說話，所以，可以使用更有效之語音編碼系統。因為語音編碼只有在一使用者正說話時才會執行，所以此一系統降低了在音訊會議中之頻寬需求。

可以改良在即時通訊中之地板控制。在傳統音訊會議中喪失之重點為缺少可以通知其他想要發言之音訊會議參與者的機制。這造成了一參與者獨斷一會議的狀況，因為他或她並不知道其他人想要發言之情形。使用本發明，一使用者可以簡單地致動感應器，以指示該使用者想要發言。例如，當紅外線感應器被使用時，使用者簡單地需要移動其臉上肌肉，模仿說話。這將提供語音檢測信號，其指示使用者正在說話，或想要說話。使用喉式或骨頭麥克風，使用者可以簡單地以很軟音調哼出聲，就會再次觸動喉式或骨頭麥克風，表示使用者正在說話或想要說話。

於另一應用中，可以改良用於個人數位助理或小計算裝置，例如掌上型電腦、筆記型電腦或其他類似電腦的電源管理。電池壽命為此等可攜式裝置之主要考量。藉由得

知是否使用者正在說話，指定給數位信號處理之資源需要執行傳統計算功能，及需要以執行語音辨識之資源可以以更有效方式加以分配。

於另一應用中，來自傳統音訊麥克風之音訊信號及來自語音感應器之信號可以被智慧型組合，使得即使背景發聲者與有關發聲者同時發聲時，背景語音可以由該音訊信號中刪除。執行此語音加強能力在某些環境下係特別想要的。

雖然本發明已經針對特定實施例加以說明，但熟習於本技藝者可以了解到，很多在形式及細節上之變化可以在不脫離本發明之精神及範圍下加以完成。

【圖式簡單說明】

第 1 圖為本發明可以使用之一環境之方塊圖。

第 2 圖為本發明可以使用之語音辨識系統之方塊圖。

第 3 圖為依據本發明之一實施例之語音辨識系統之方塊圖。

第 4 及 5 圖例示於第 3 圖中之系統的一部份的兩個不同實施例。

第 6 圖為用於一麥克風信號及一紅外線感應信號，信號大小對時間之表示圖。

第 7 圖為一傳統麥克風與語音感應器之一實施例示意圖。

第 8 圖為一骨頭感應麥克風與一傳統音訊麥克風之示

意圖。

第9圖為分別用於麥克風信號及音訊麥克風信號之信號大小對時間圖。

第10圖為傳統音訊麥克風與喉式麥克風之示意圖。

第11圖為一耳內麥克風與一貼近式麥克風。

【主要元件符號說明】

100	計算系統環境	110	電腦
120	處理單元	121	系統匯流排
130	系統記憶體	131	唯讀記憶體
132	隨機存取記憶體	133	基本輸入/輸出系統
134	作業系統	135	應用程式
136	程式模組	137	程式資料
140	介面	141	硬碟機
144	作業系統	145	應用程式
146	程式模組	147	程式資料
151	磁碟機	152	磁碟
160	使用者輸入介面	161	指標裝置
162	鍵盤	163	麥克風
170	網路介面	171	區域網路
172	數據機	173	廣域網路
180	遠端電腦	185	遠端應用程式
190	視訊介面	191	監視器
195	週邊介面	196	印表機機

197 喇叭

400 喇叭

404 麥克風

406 類比至數位轉換器

407 訊框建構器

408 特性抽取器

412 解碼器

414 辭彙

416 語言模組

418 音響模型

420 量測模組

422 輸出模組

424 訓練器

426 訓練本文

300 語音檢測系統

301 換能器

302 信號捕捉元件

303 音訊麥克風

304 信號處理機

306 語音檢測信號

308 輸出信號

310 紅外線為主語音檢測器

312 音訊為主語音檢測器

314 組合語音檢測器

316 元件

320 聯合式語音檢測器

420 頭戴組

422 耳機

424 耳機

426 延伸桿

428 音訊麥克風

430 收發器

450 頭戴組

451 頭戴件

452 耳機

454 耳機

456 麥克風

458 骨頭敏感麥克風

460 音訊信號

500 頭戴組

501 頭戴件

502 耳機

504 音訊麥克風

506 喉式麥克風

550 頭戴組

552 耳內麥克風

554 音訊麥克風

556 延伸桿

十、申請專利範圍：

1. 一種語音辨識系統，該語音辨識系統包含：

一音訊麥克風，該音訊麥克風基於一感應音訊輸入，輸出一麥克風信號；

一語音感應器，該語音感應器基於由語音動作所產生之一非音訊輸入，輸出一感應器信號；及

一語音檢測器元件，該語音檢測器元件基於該麥克風信號及基於在該感應器信號之一第一特徵中之變化，輸出一語音檢測信號，該語音檢測信號指示一使用者正在說話之一或然率，其中當該使用者正說話時，該感應器信號之該第一特徵具有變化之一第一位準；而當該使用者未說話時，該感應器信號之該第一特徵具有變化之一第二位準，且其中該語音檢測器元件基於該感應器信號之該第一特徵變化之該位準，相對於該第一特徵變化之一基準位準，輸出該語音檢測信號，該第一特徵變化之該基準位準包含在一段給定時間內該特徵之該等第一及該第二位準中之一預定者，在該段給定時間內，該語音檢測器元件更藉由將該語音檢測信號與該麥克風信號相乘以計算一組合信號；及

一語音辨識器，該語音辨識器基於該結合信號來辨識語音以提供一辨識輸出，該辨識輸出基於該組合信號指示在該麥克風信號中之語音，其中辨識語音之步驟包含以下步驟：

增加步驟，基於該語音檢測信號指示該使用者正

在說話之一或然率，將可辨識語音之一可能性增加一定數量；及

減少步驟，基於該語音檢測信號指示該使用者並未說話之一或然率，將可辨識語音之一可能性減少一定數量。

2.如申請專利範圍第 1 項所述之語音辨識系統，其中該基準位準的計算係藉由平均在該時間段內之該第一特徵變化之該位準。

3. 如申請專利範圍第 1 項所述之語音辨識系統，其中該基準位準係在該語音辯識系統操作中被間歇地重新計算。

4. 如申請專利範圍第 3 項所述之語音辨識系統，其中該基準位準係被週期地重新計算，以在一循環時間視窗上，表示該第一特徵的該變化位準。

5. 如申請專利範圍第 3 項所述之語音辨識系統，其中該語音檢測器元件，基於該感應器信號之該第一特徵變化的該位準與該基準位準之一比較，輸出該語音檢測信號，且其中該比較係週期地執行。

6. 如申請專利範圍第 5 項所述之語音辨識系統，其中該比

較係較重新計算該基準位準更頻繁執行。

7.如申請專利範圍第 1 項所述之語音辨識系統，其中該音訊麥克風及該語音感應器係被安裝至一頭戴組。

8.一種語音辨識系統，該語音辨識系統包含：

一語音檢測系統，該語音檢測系統包含：

一音訊麥克風，該音訊麥克風基於一感應音訊輸入，輸出一麥克風信號；

一語音感應器，該語音感應器基於由語音動作所產生之一非音訊輸入，輸出一感應器信號；及

一語音檢測器元件，該語音檢測器元件基於該麥克風信號及該感應器信號，輸出一語音檢測信號，該語音檢測信號指示一使用者正在說話之一或然率，其中該語音檢測器元件藉由將該語音檢測信號及該麥克風信號相乘以計算一組合信號；及

一語音辨識引擎，該語音辨識引擎基於該結合信號來辨識語音以提供一辨識輸出，該辨識輸出指示在該感測音訊輸入中之語音；

基於該語音檢測信號指示該使用者正在說話之一或然率，將可辨識語音之一可能性增加一定數量；及

基於該語音檢測信號指示該使用者並未說話之一或然率，將可辨識語音之一可能性減少一定數量。

9.如申請專利範圍第 8 項所述之語音辨識系統，其中該音訊麥克風及該語音感應器被安裝在一頭戴組上。

10. 一種辨識語音的方法，該方法包含以下步驟：

產生第一信號步驟，用一音訊麥克風來產生一第一信號，該第一信號指示一音訊輸入；

產生第二信號步驟，產生一第二信號，該第二信號指示由一臉部動作感應器所感應之一使用者臉部動作；

產生第三信號步驟，基於該等第一及第二信號來產生一第三信號，該第三信號指示該使用者正在說話之一或然率；

產生第四信號步驟，藉由將該使用者正在說話之該或然率與該第一信號相乘，以產生一第四信號；及

辨識語音步驟，基於該第四信號及該語音檢測信號來辨識語音，其中辨識語音之步驟包含以下步驟：

增加步驟，基於該語音檢測信號指示該使用者正在說話之一或然率，將可辨識語音之一可能性增加一定數量；及

減少步驟，基於該語音檢測信號指示該使用者並未說話之一或然率，將可辨識語音之一可能性減少一定數量。

11.如申請專利範圍第 10 項所述之方法，其中產生該第二信號之步驟包含以下步驟：

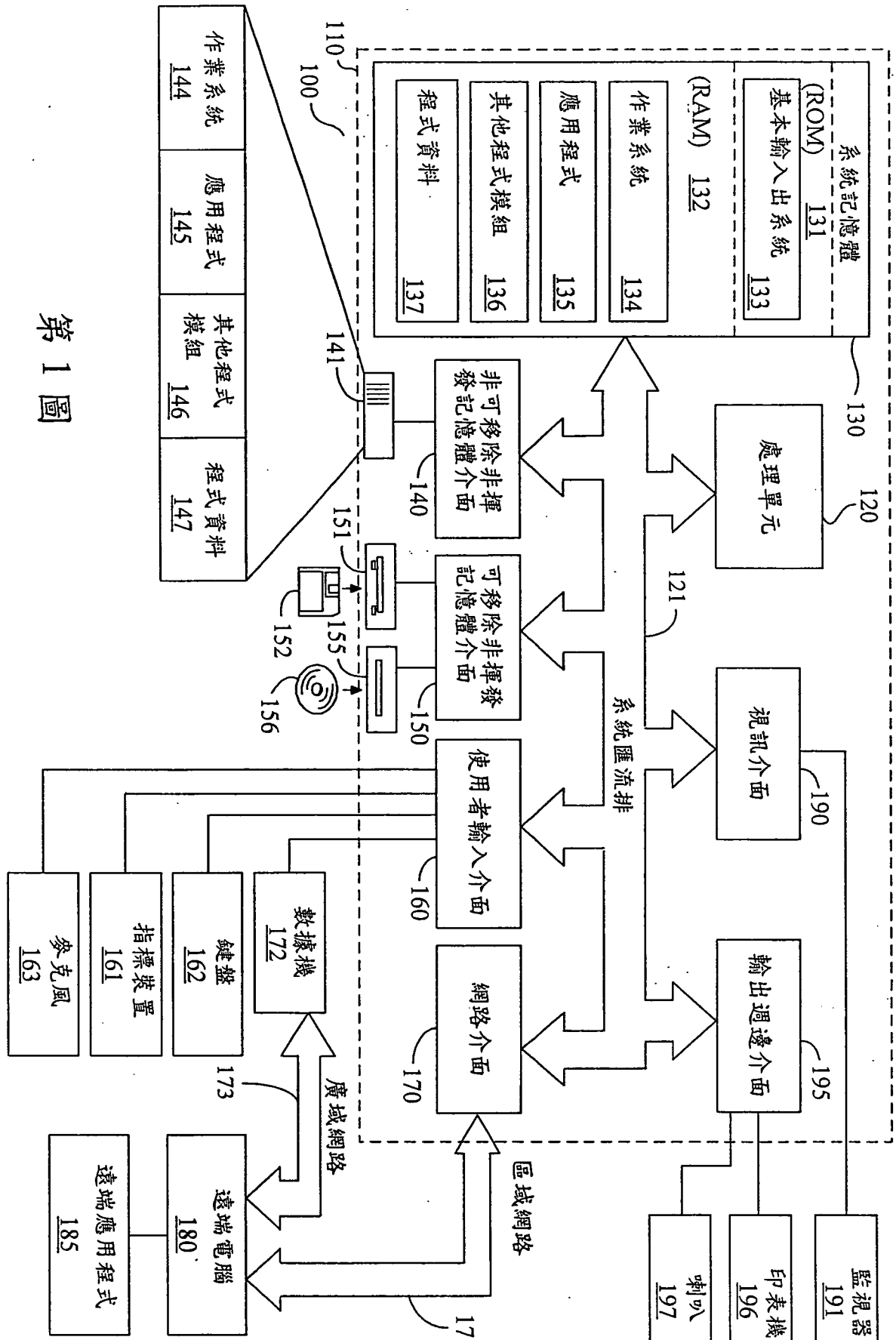
感應步驟，感應該使用者之顎部及頸部中之一者的振動。

12.如申請專利範圍第 10 項所述之方法，其中產生該第二信號之步驟包含以下步驟：

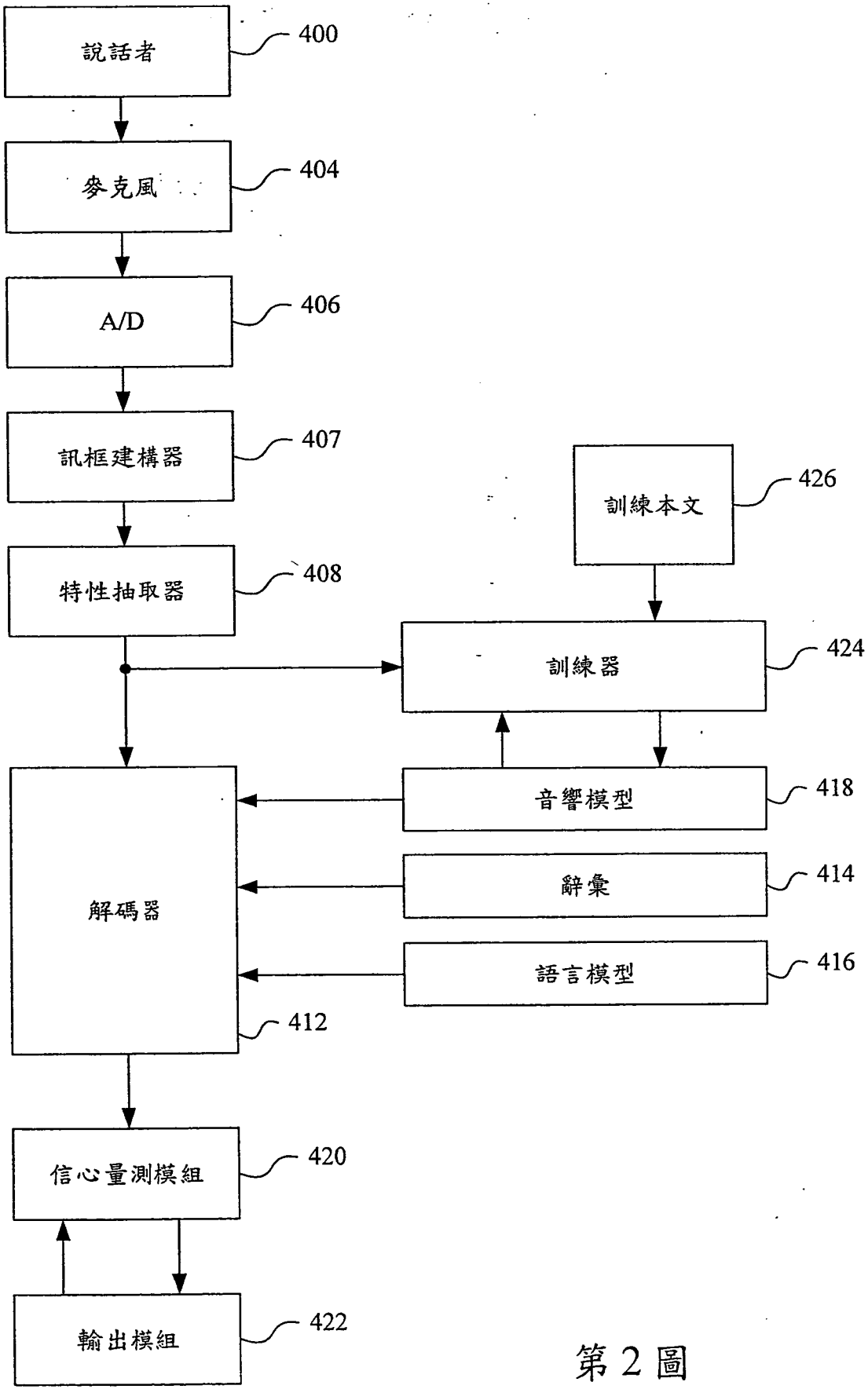
感應步驟，感應一影像，該影像指示該使用者嘴部之動作。

13.如申請專利範圍第 10 項所述之方法，更包含以下步驟：

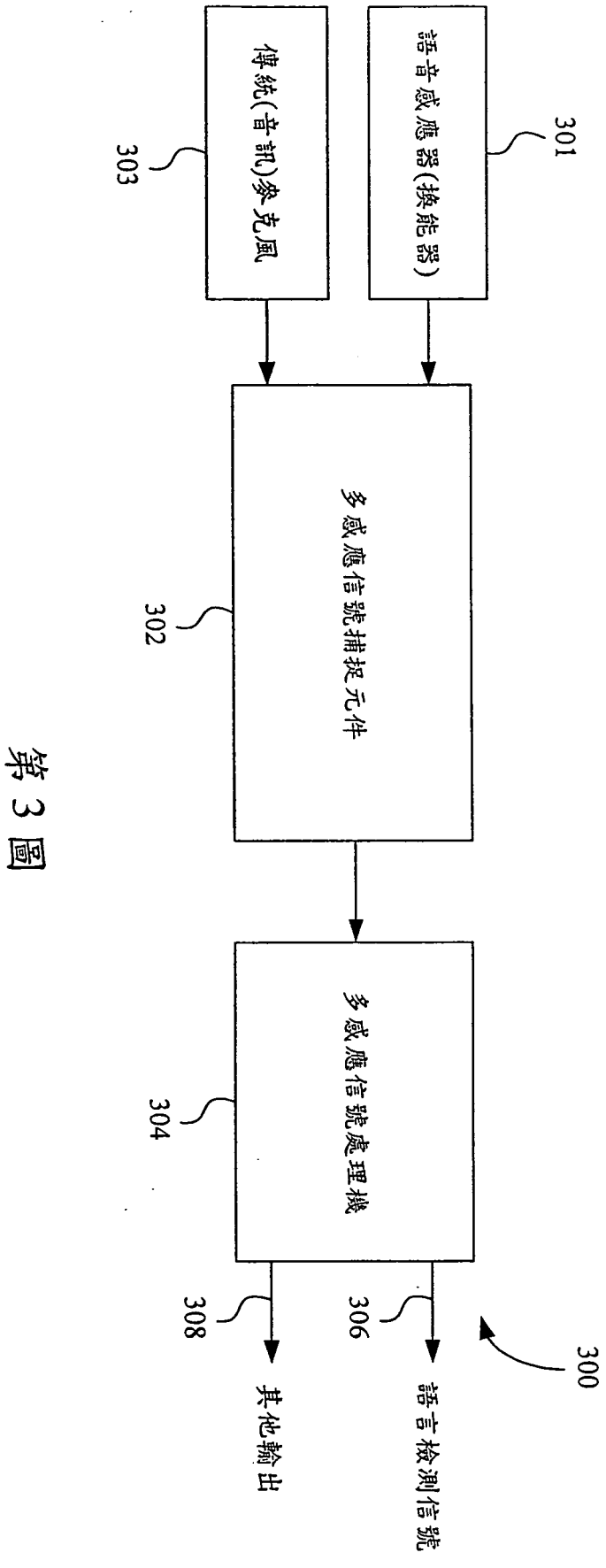
提供步驟，基於檢測是否該使用者正在說話，提供一語音檢測信號。


系統記憶體
(ROM) 131
基本輸入出系統 133
(RAM) 132
作業系統 134
應用程式 135
其他程式模組 136
程式資料 137
處理單元 120
130
100
110
系統匯流排 121
非可移除非揮發記憶體介面 140
可移除非揮發記憶體介面 150
使用者輸入介面 160
網路介面 170
視訊介面 190
輸出週邊介面 195
監視器 191
印表機 196
喇叭 197
141
151
152
155
156
作業系統 144
應用程式 145
其他程式模組 146
程式資料 147
數據機 172
鍵盤 162
指標裝置 161
麥克風 163
173
廣域網路
區域網路
171
遠端電腦 180
遠端應用程式 185

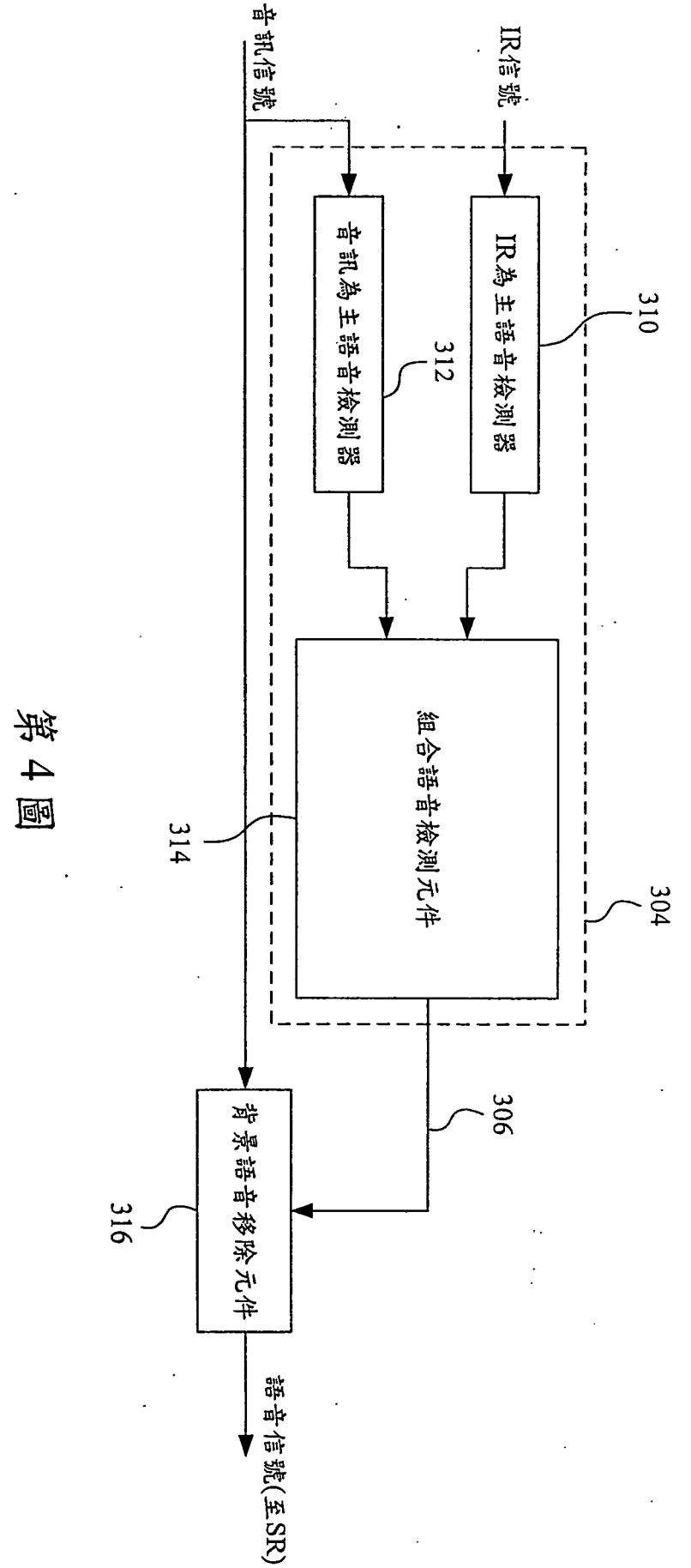
第 1 圖


說話者
400
麥克風
404
A/D
406
訊框建構器
407
特性抽取器
408
訓練本文
426
訓練器
424
音響模型
418
辭彙
414
語言模型
416
解碼器
412
信心量測模組
420
輸出模組
422

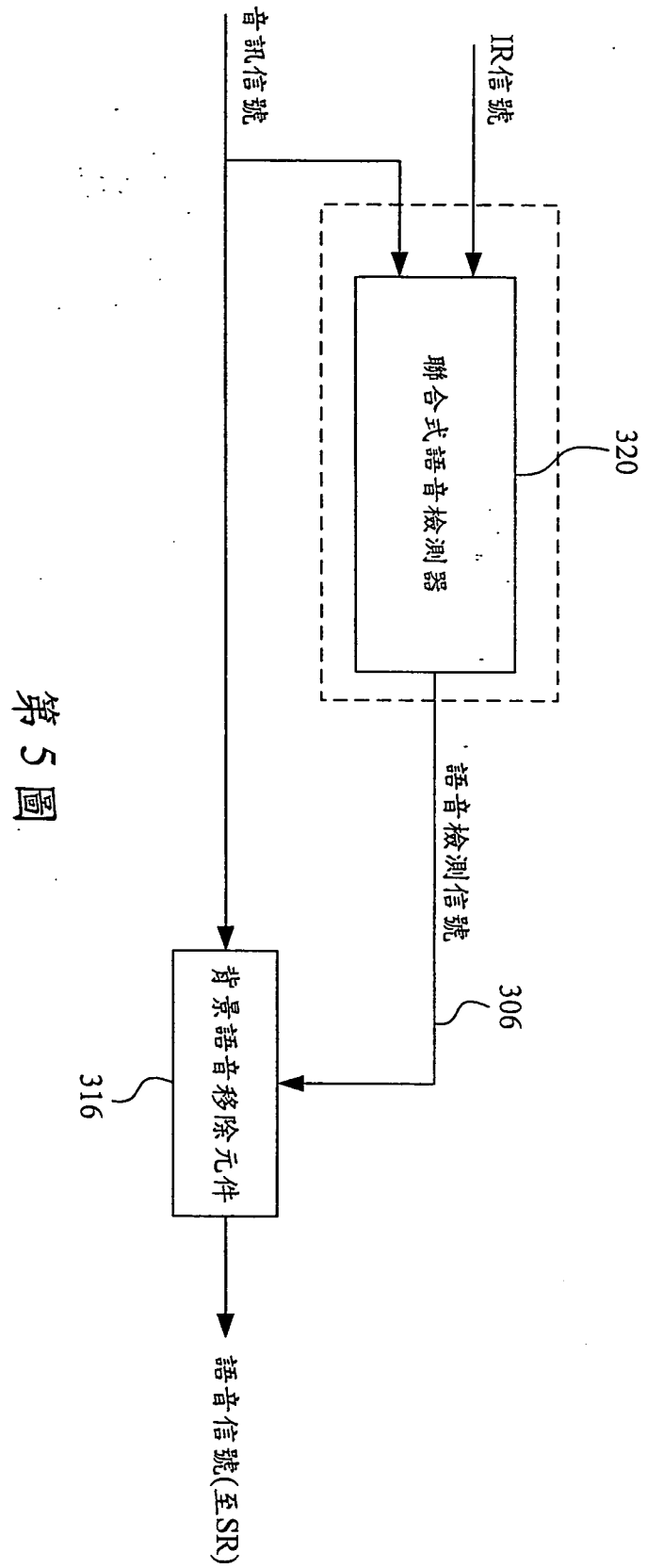
第 2 圖


語音感應器(換能器)
301
傳統(音訊)麥克風
303
多感應信號捕捉元件
302
多感應信號處理機
304
語言檢測信號
306
其他輸出
308
300

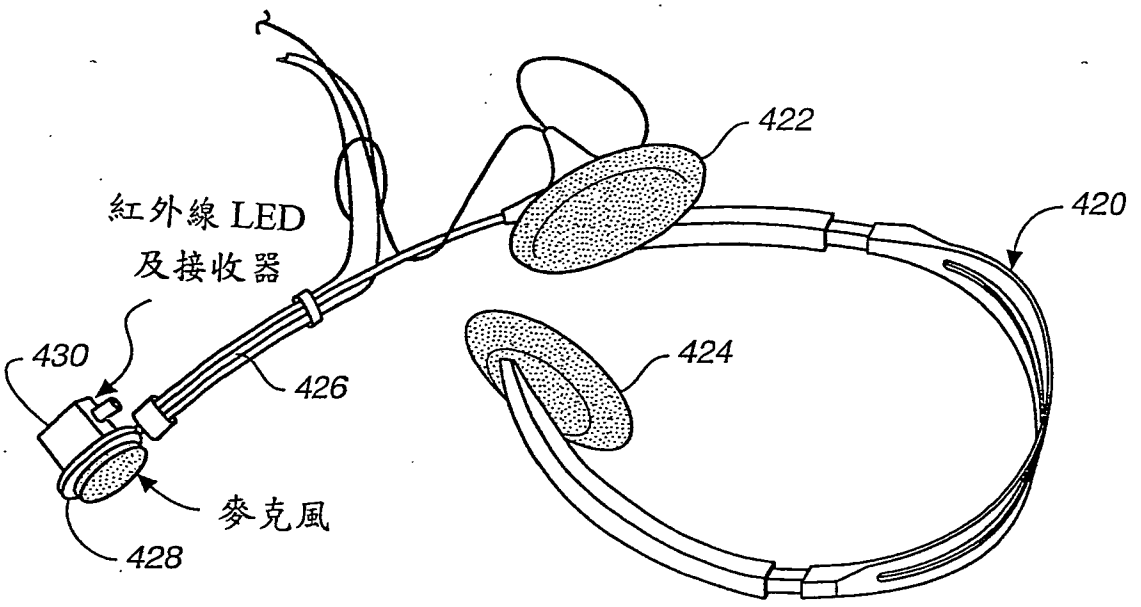
第 3 圖


IR信號
音訊信號
IR為主語音檢測器
音訊為主語音檢測器
組合語音檢測元件
背景語音移除元件
語音信號(至SR)
310
312
304
314
306
316

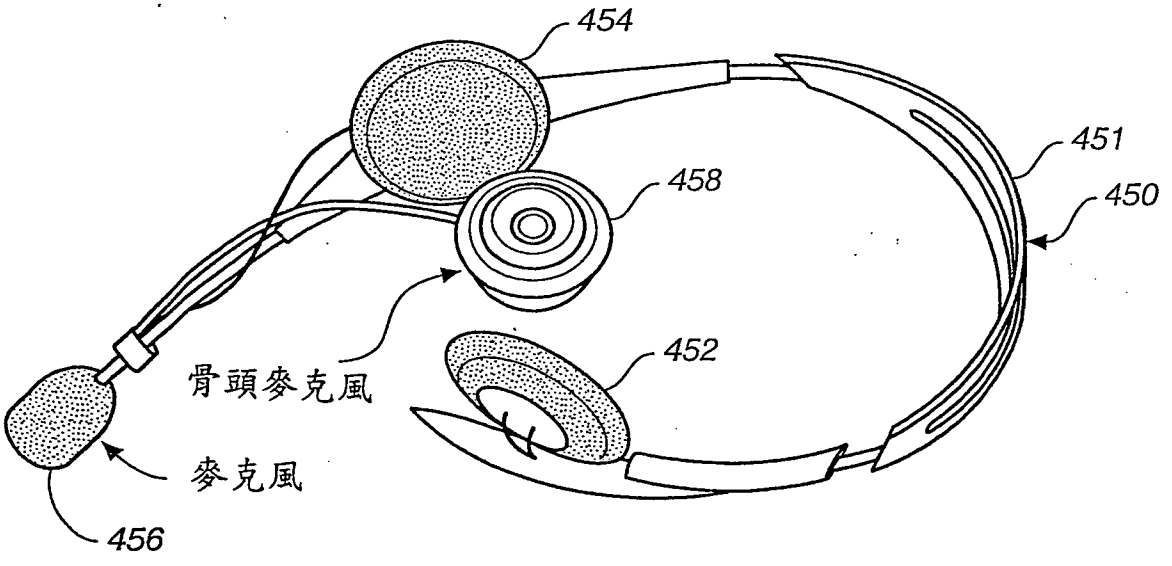
第4圖



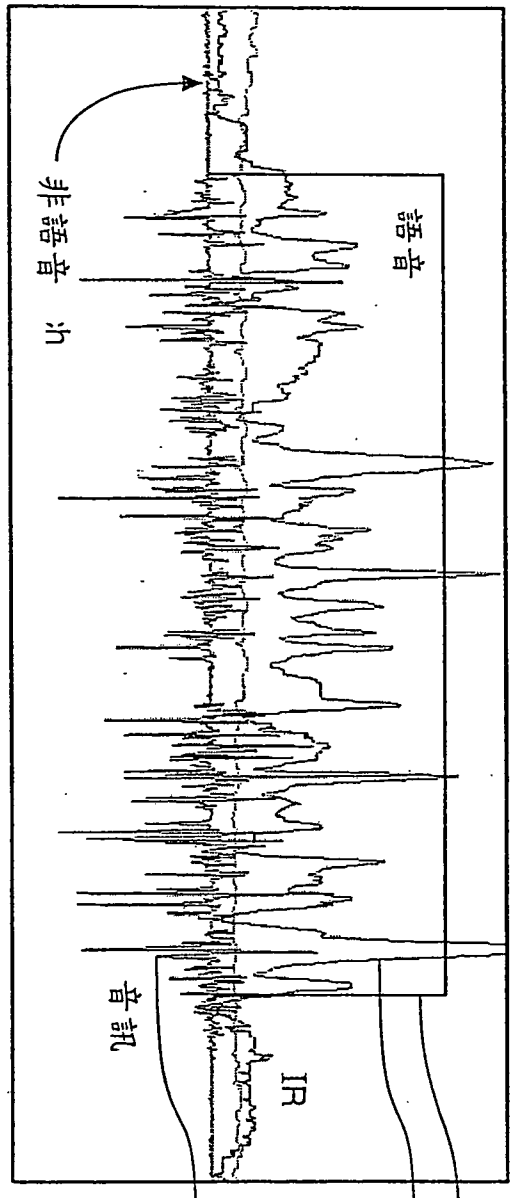
第 5 圖


422
420
紅外線 LED
及接收器
430
426
424
麥克風
428

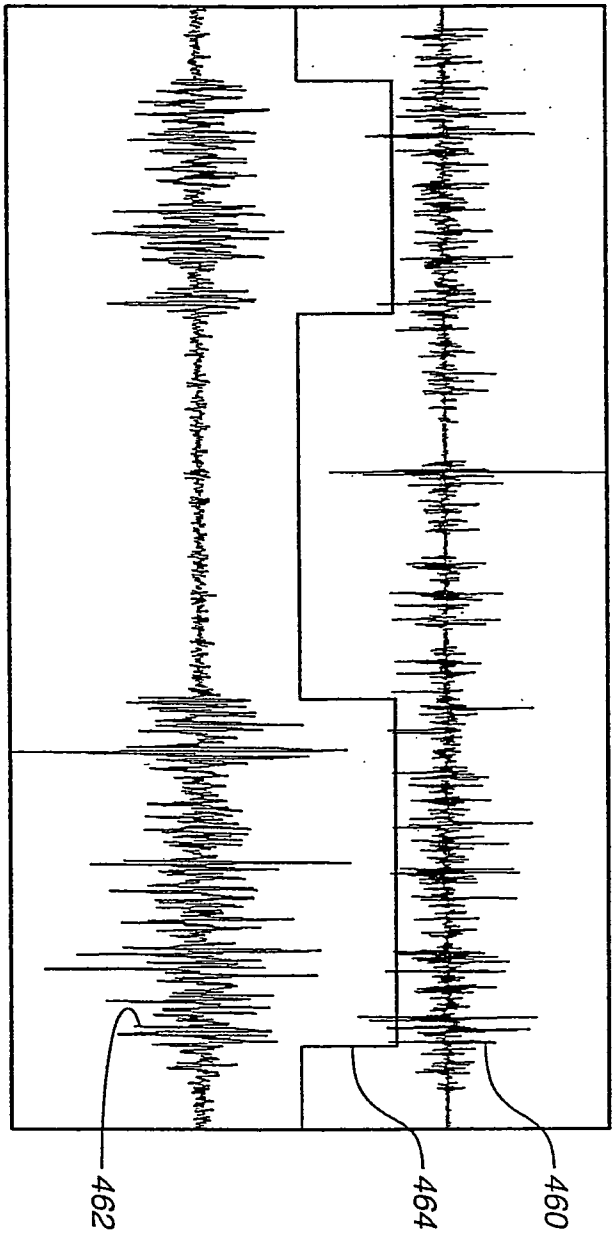
第 7 圖


454
451
458
450
骨頭麥克風
452
麥克風
456

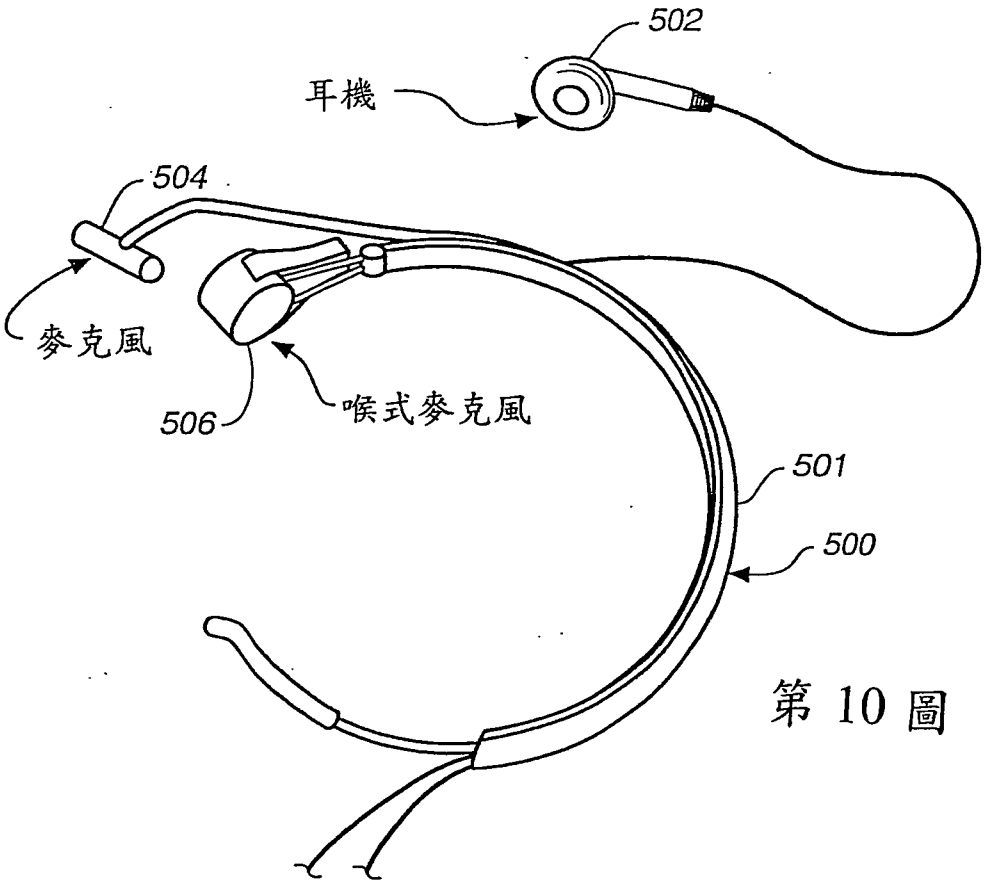
第 8 圖



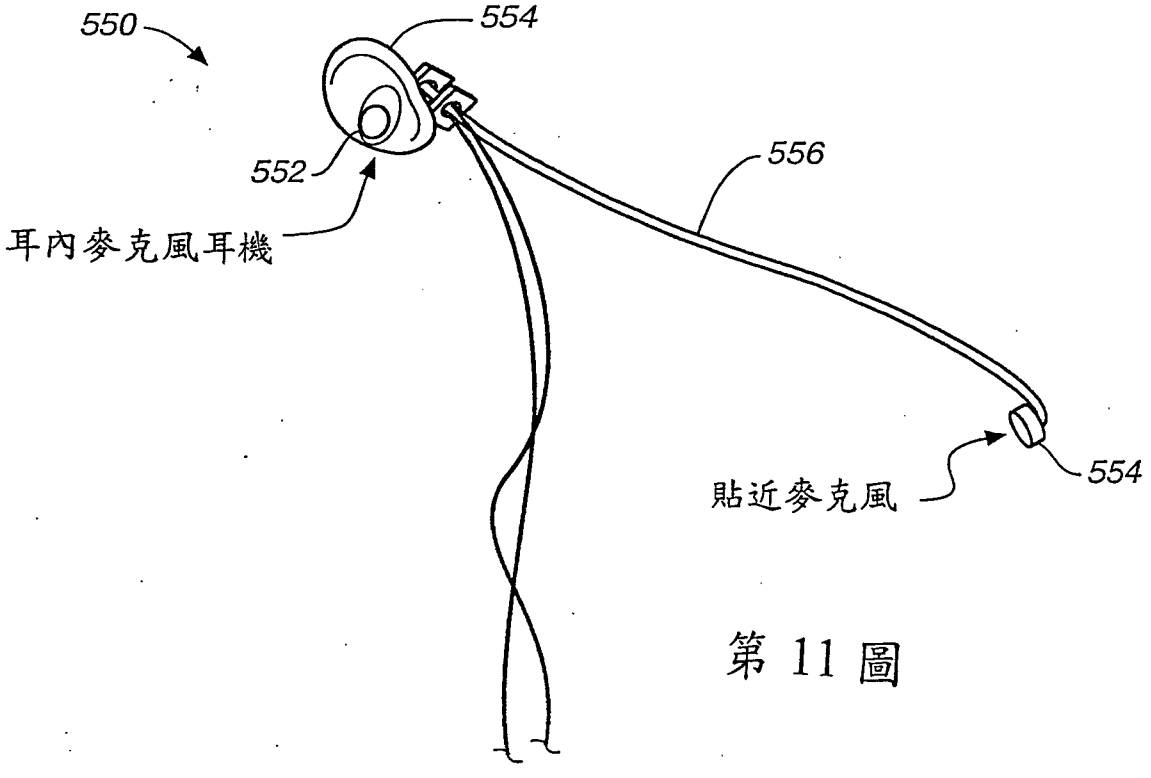
第 6 圖



第 9 圖


502
耳機
504
麥克風
506
喉式麥克風
501
500

第 10 圖


550
554
552
耳內麥克風耳機
556
貼近麥克風
554

第 11 圖