



(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 I543635 B

(45)公告日：中華民國 105 (2016) 年 07 月 21 日

(21)申請案號：102146946

(22)申請日：中華民國 102 (2013) 年 12 月 18 日

(51)Int. Cl. : H04R25/00 (2006.01)

G10L21/00 (2013.01)

G06K9/78 (2006.01)

(71)申請人：劉璟鋒(中華民國) (TW)

高雄市鼓山區美術南五街 99 巷 10 號

陳筱涵(中華民國) (TW)

臺南市仁德區忠義三街 24 巷 21 號

(72)發明人：劉璟鋒(TW)；陳筱涵(TW)

(74)代理人：高玉駿；楊祺雄

(56)參考文獻：

CN 1494711A

CN 101300897A

EP 2638708A2

US 7609842B2

US 2002/0135618A1

US 2006/0104454A1

US 2012/0242860A1

審查人員：林宥榆

申請專利範圍項數：17 項 圖式數：5 共 29 頁

(54)名稱

助聽系統與助聽系統之語音擷取方法

(57)摘要

一種助聽系統與助聽系統之語音擷取方法，該語音擷取方法是先根據使用者前方視野影像中之人臉影像分別相對於使用者的距離與方位，分析出最有可能作為講話對象的人臉影像後，再透過波束形成方法，將可能性最高的級別內的人臉影像的方位的語音訊號過濾出來。藉由影像輔助擷取語音訊號的設計，可大幅提高在人群中成功擷取到講話對象之語音的成功率，而有助於降低語音干擾，是一種創新的助聽系統與助聽系統之語音擷取方法設計。

指定代表圖：

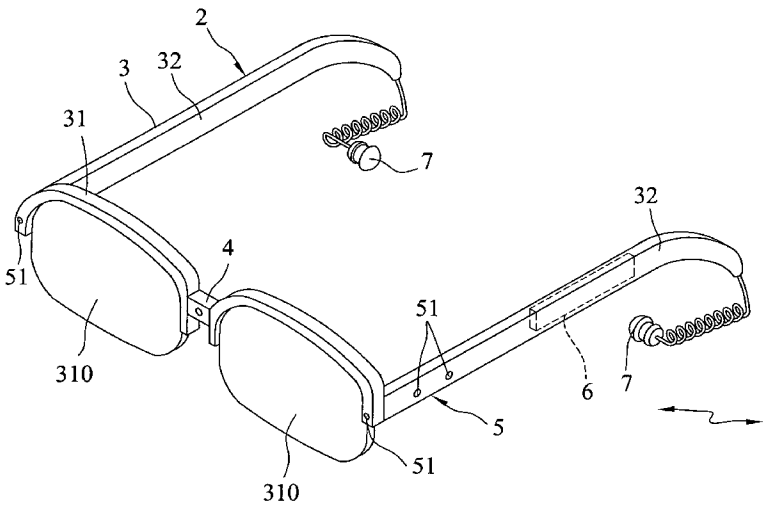


圖1

符號簡單說明：

- 2 . . . 助聽裝置
- 3 . . . 載具
- 31 . . . 前載體
- 310 . . . 鏡片
- 32 . . . 側載體
- 4 . . . 影像擷取器
- 5 . . . 麥克風陣列單元
- 51 . . . 麥克風
- 6 . . . 微處理器
- 7 . . . 聲音輸出器
- 8 . . . 遙控裝置
- 81 . . . 觸控顯示模組

發明摘要

※ 申請案號： 102146946

※ 申請日： 102.12.18

※IPC 分類：

H04R 25/00 (2006.01)

G10L 21/00 (2013.01)

G06K 9/18 (2006.01)

【發明名稱】 助聽系統與助聽系統之語音擷取方法

【中文】

一種助聽系統與助聽系統之語音擷取方法，該語音擷取方法是先根據使用者前方視野影像中之人臉影像分別相對於使用者的距離與方位，分析出最有可能作為講話對象的人臉影像後，再透過波束形成方法，將可能性最高的級別內的人臉影像的方位的語音訊號過濾出來。藉由影像輔助擷取語音訊號的設計，可大幅提高在人群中成功擷取到講話對象之語音的成功率，而有助於降低語音干擾，是一種創新的助聽系統與助聽系統之語音擷取方法設計。

【英文】

【 代 表 圖 】

【本案指定代表圖】：圖（ 1 ）。

【本代表圖之符號簡單說明】：

- | | | | |
|-----------|-------|----------|---------|
| 2 | 助聽裝置 | 5 | 麥克風陣列單元 |
| 3 | 載具 | 51 | 麥克風 |
| 31 | 前載體 | 6 | 微處理器 |
| 310 | 鏡片 | 7 | 聲音輸出器 |
| 32 | 側載體 | 8 | 遙控裝置 |
| 4 | 影像擷取器 | 81 | 觸控顯示模組 |

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：



發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】 助聽系統與助聽系統之語音擷取方法

【技術領域】

【0001】 本發明是有關於一種助聽系統與助聽系統之語音擷取方法，特別是指一種結合影像輔助收音的助聽系統與助聽系統之語音擷取方法。

【先前技術】

【0002】 雞尾酒會效應 (cocktail party effect) 是指人的一種聽力選擇能力，在這種情況下，注意力會集中在某一個人的談話之中，而忽略背景中其他的對話或噪音。使我們可以在吵雜環境中談話。但由於聽損患者的聽力閾值上升，導致這樣的聽覺選擇性傾聽的能力大幅降低，以致於目前的助聽器使用者在面對眾多語言訊號中，如何選取目標語言訊號，而不被其他語言訊號干擾，成了一個很大的挑戰。

【0003】 目前助聽器利用聲音訊號特性不同，雖然可以有效區分語音 (speech) 及噪音 (noise)，而將噪音訊號濾除，但在日常生活環境卻經常同時存在許多語音訊號，且其中大多數語音訊號是患者不想聽的，如雞尾酒會、市場、餐廳等地方，雖然使用方向性麥克風可以降低來自患者後方及側面的聲音訊號，但來自患者前方的聲音訊號中，除

了談話對象的語音外，可能仍然充斥著其他人的講話聲。

【0004】人類的聽力系統本來就是雙耳功能的加成，是利用聲音傳至雙耳的時間及強度差異來進行聲音定位，目前最先進的降低噪音與提高訊雜比的方法，就是雙耳配戴助聽器，應用無線技術可以使兩側助聽器同步處理、互相溝通，這種技術稱為耳對耳(ear-to-ear)通訊，保留訊號的雙耳線索，並以波束形成法(beamforming)技術，過濾出患者正前方的聲音，且提高訊噪比，而達到可在吵雜的環境裡與人面對面交談的目的。

【0005】雖然雙耳通訊與波束成型的技術可以定位聲音的方向，但面對前方有多個語言訊號時，通常無法得知哪個方向的語音訊號是大腦要的，所以只能保留病患最正面的較強語言訊號，以致於在面臨一對二講話、一對三講話，或者是患者前方有很多人在講話的場面時，經常無法順利取得所需的語音訊號。

【發明內容】

【0006】因此，本發明之目的，即在提供一種透過影像輔助擷取語音，而有助於降低語音干擾的助聽系統及助聽系統的語音擷取方法。

【0007】於是，本發明助聽系統，包含一助聽裝置，該助聽裝置包括一影像擷取器、一麥克風陣列單元、一微處理器與至少一聲音輸出器。該影像擷取器可擷取使用者視野前方之視野影像。該麥克風陣列單元包括多個相間隔並可擷取聲音之麥克風。該微處理器是訊號連接於該影像擷

取器與該等麥克風，包括一影像分析模組，及一聲音處理模組，該影像分析模組可分析輸出該視野影像中的所有人脸影像相對於該使用者之距離與方位，並根據該等人臉影像之距離和方位，對該等人臉影像作為講話對象之可能性進行分類，而對應輸出一可能性訊號源分類，該可能性訊號源分類規劃有多個分別具有不同可能性的級別，該聲音處理模組可根據被分類於一可能性最高的級別內的所有人脸影像的方位，以及該等麥克風擷取之聲音訊號，透過波束形成方法過濾出該可能性最高的級別內的所有人脸影像的方位的語音訊號。該聲音輸出器是訊號連接於該微處理器，並可將該微處理器輸出之語音訊號輸出以供聆聽。

【0008】於是，本發明助聽系統之語音擷取方法，包含以下步驟：(a) 將一助聽裝置配置於使用者身上，該助聽裝置包括一影像擷取器與一麥克風陣列單元，該麥克風陣列單元包括多個相間隔之麥克風；(b) 以該影像擷取器擷取該使用者視野前方之視野影像；(c) 使該助聽裝置自該視野影像中辨識出人脸影像，並判斷出該等人臉影像分別相對於該使用者的距離與方位；(d) 使該助聽裝置根據該等人臉影像分別相對於該使用者之距離與方位，對該等人臉影像作為講話對象之可能性進行級別分類，而對應輸出一可能性訊號源分類；及(e) 使該助聽裝置利用該等麥克風感測輸出的聲音訊號，並透過波束形成方法，將步驟(d)之可能性訊號源分類中被分類於一可能性最高的級別內的所有人脸影像的方位的語音訊號過濾出來。

【0009】 本發明之功效：藉由先分析一視野影像中之每一人臉影像，以找出最可能為講話對象的人臉影像後，再利用波束形成方法過濾出最可能為講話對象之方位的語音訊號的設計，可大幅提高在人群中成功擷取到講話對象之語音的成功率，而有助於降低語音干擾。

【圖式簡單說明】

【0010】 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 是本發明助聽系統之一較佳實施例的立體示意圖；

圖 2 是該較佳實施例的功能方塊圖；

圖 3 是本發明助聽系統之語音擷取方法的較佳實施例的步驟流程圖；

圖 4 是該較佳實施例之一耳機裝置擷取之視野影像的示意圖；及

圖 5 是一俯視示意圖，示意說明圖 4 之該等人臉影像所代表之講話對象相對於一視野中心的左右偏離角度。

【實施方式】

【0011】 如圖 1、2、4，本發明助聽系統之較佳實施例，適用於配戴在一聽力受損的使用者頭部，以協助該使用者擷取講話對象之語音訊號。該助聽系統包含一個可供配戴於使用者頭部的助聽裝置 2，及一個與該助聽裝置 2 無線訊號連接而可遙控該助聽裝置 2 之遙控裝置 8。

【0012】 該助聽裝置 2 包括一個供配戴於患者頭部之載具 3，及分別安裝於該載具 3 之一個影像擷取器 4、一個麥克風陣列單元 5、一個微處理器 6 與兩個聲音輸出器 7。在本實施例中，該載具 3 是設計成眼鏡型式，具有一個位於患者雙眼前方且裝設有鏡片 310 之鏡框狀的前載體 31，及兩個左右間隔安裝於該前載體 31 並供配掛於雙耳之鏡腳狀的側載體 32。

【0013】 該影像擷取器 4 是安裝於該前載體 31 之中間區段處，可擷取使用者前方視野的影像，而輸出一視野影像 40。

【0014】 該麥克風陣列單元 5 包括多個麥克風 51，該等麥克風 51 是左右成對地分別安裝外露於該等側載體 32，且前後間隔地分別沿該等側載體 32 長度方向間隔設置，可分別感測周遭聲音而輸出一聲音訊號。實施時，該等麥克風可以採用指向性麥克風、全向性麥克風，或者是兩種麥克風之組合。

【0015】 該微處理器 6 是埋設固定於其中一個側載體 32 中，且訊號連接於該影像擷取器 4 與該等麥克風 51，包括一影像分析模組 61，及一聲音處理模組 62。

【0016】 該影像分析模組 61 可自該視野影像 40 中分析找出所有的人臉影像 401，並進一步分析出該等人臉影像 401 分別相對於配戴此助聽裝置之使用者的距離與方位，及該等人臉影像 401 相對於使用者之左右偏轉角度，並根據該等人臉影像 401 之距離、方位與偏轉角度等參數，分析

出該等人臉影像 401 作為講話對象之可能性。該影像分析模組 61 包括一距離分析單元 611、一方位分析單元 612、一偏轉分析單元 613、一可能性分析單元 614，及一唇形分析單元 615。

【0017】 該距離分析單元 611 於分析該等人臉影像 401 相對於使用者之距離時，其內建有一般人臉影像在各種距離時的參考尺寸基準，會根據該參考尺寸基準來分析評估該等人臉影像 401 之尺寸所對應的距離，例如根據該等人臉影像 401 在該視野影像 40 中所佔畫素面積來進行尺寸分析，並對應輸出該等人臉影像 401 的距離資料。但實施時，分析計算出該視野影像 40 中之該等人臉影像 401 相對於該使用者之距離的方法眾多，不以上述方法為限。

【0018】 配合圖 5，該方位分析單元 612 於分析該等人臉影像 401 分別相對於該使用者的方位時，是以該視野影像 40 之左右向的視野中心 402 為基準，在已分析出該等人臉影像相對於使用者距離後，就可根據該等人臉影像相對於該視野中心 402 之左右間距，而進一步分析出該等人臉影像相對於該使用者視野中心的左右水平偏離夾角，如圖 5 所示之 θ_1 、 θ_2 、 θ_3 ，而得出每一人臉影像相對於使用者之方位，例如位於該視野中心 402 往右偏離 30° 處，並對應輸出該等人臉影像 401 的方位資料。但實施時，分析計算出該視野影像 40 中之該等人臉影像 401 相對於該視野中心 402 之方位的方法眾多，不以上述方法為限。

【0019】 該偏轉分析單元 613 於分析該等人臉影像 401

之偏轉角度時，主要是分析該等人臉影像 401 相對於使用者正面的左右偏轉角度，並對應輸出該等人臉影像 401 的偏轉角度資料。

【0020】該可能性分析單元 614 內建有一用以分析該等人臉影像 401 作為講話對象之可能性的可能性分類基準，該可能性分類基準主要是參考一般人面對面進行講話時的講話習慣與禮儀，分別就人臉影像 401 相對於該使用者之距離遠近、左右偏離角度大小，以及左右偏轉角度大小設定有不同之運算權重比。

【0021】一般而言，人與人的講話距離大致可區分為親密距離、社交距離、禮儀距離與公共距離等四種，其中，親密距離約在 0.5 公尺內，通常用於夫妻、情侶、父母、子女或知心好友等重要人物間；社交距離則在 0.5~1.5 公尺範圍，伸手可握到對方的手，但不易觸碰到對方的身體；禮儀距離則大約在 1.5~3 公尺範圍，適用於表示敬重，以及會議與儀式場合；公共距離則大於 3 公尺，適用於一般演講者與聽眾間，以及不相識的人。基於此，越親密或越重要的講話對象通常會越接近使用者，且會越接近使用者之視野中心，同時會盡量面向使用者。

【0022】因此，該可能性分析單元 614 在分析該等人臉影像 401 作為講話對象之可能性時，是以距離為優先考量，距離越近，可能性越高，然後再考量方位，越接近該視野中心 402 者，表示越接近使用者視野中心，作為談話對象的可能性越高，最後，再考量該等人臉影像 401 之左右

偏轉角度，偏轉角度越小者，表示該人臉影像 401 越是正面朝向使用者，作為談話對象的可能性越高。此外，距離使用者越近之運算權重比越大，距離越遠則相對越小；越接近該視野中心 402，也就是越接近該使用者之視野中心者的運算權重比越大，越偏離該視野中心 402 的運算權重比越小；偏轉角度越小，表示該人臉影像 401 越是正面面對使用者，其運算權重比越大，反之越小。

【0023】該可能性分析單元 614 於分析該等人臉影像 401 作為講話對象之可能性時，會根據該可能性分類基準所設定之運算權重比，以及該等人臉影像 401 分別相對於使用者之距離、方位與偏轉角度等參數，對該等人臉影像 401 進行作為講話可能性之統計分析，並對應輸出一可能性訊號源分類，該可能性訊號源分類依據作為講話對象之可能性高低，分設有多個不同可能性級別，例如最高可能性級別、中度可能性級別，及低可能性級別等，會將該等人臉影像 401 之各參數的運算權重比統計分析後，依據各可能性級別所設定之可能性門檻限制，以及該等人臉影像 401 之可能性分析結果，將該等人臉影像 401 分別分類於對應的可能性級別中。

【0024】該唇形分析單元 615 可於該可能性分析單元 614 分析輸出該可能性訊號源分類後，對被分類有人臉影像 401 之可能性最高的級別中的所有人脸影像 401 進行唇形影像分析，嘴唇有開合變形即表示該人臉影像 401 可能是在講話，並對應嘴唇有開合變形之每一個人臉影像 401 分

別輸出一講話訊號。例如先判斷是否有人臉影像 401 被分類於最高可能性級別，若有，則對該最高可能性級別內之所有人臉影像 401 進行唇形影像分析，若無人臉影像 401 被分類於該最高可能性級別中，則接著判斷是否有人臉影像 401 被分類於該中度可能性級別，若有，則對該中度可能性級別內的所有人臉影像 401 進行唇形分析。若中度可能性級別也無人臉影像 401 時，則停止判斷，表示當下視野影像 40 中的所有入作為講話對象的可能性非常低。

【0025】該聲音處理模組 62 可接收處理該等麥克風 51 感測輸出之聲音訊號，例如進行類比/數位轉換、降噪處理等音源訊號常見之處理作業。該聲音處理模組 62 包括一語音觸發單元 62，及一語音定位單元 621。該語音觸發單元 62 可接收分析該等麥克風 51 感測輸出之使用者正面的聲音訊號中是否含有語音訊號，若分析出含有語音訊號時，表示使用者正面有人在講話，則觸發啟動該影像擷取器 4 開始擷取影像，並觸發該影像分析模組 61 開始進行講話對象之可能性分析。

【0026】該語音定位單元 621 會被該唇形分析單元 615 輸出之所有講話訊號驅動，而透過波束形成（beamforming）方法，擷取每一講話訊號所對應之人臉影像 401 的方位的聲音訊號，也就是排除其他方位的聲音訊號，然後，再進一步透過語音提取方法，自所得到之聲音訊號中將不需要之噪音降噪濾除，並將語音訊號過濾放大輸出，就可得到嘴唇有開合變形之人臉影像 401 的語音訊號，並將取得

之所有語音訊號輸出至該等聲音輸出器 7。

【0027】 由於透過波束形成 (beamforming) 方法擷取特定方位之聲音訊號，以及自擷取之聲音訊號中過濾出語音訊號，並對雜音進行降噪過濾處理等都是習知技術，且方法相當多，亦非本發明之改良重點，因此不再詳述。

【0028】 該等聲音輸出器 7 為可供塞裝於使用者耳中，可將該聲音處理模組 62 擷取放大之語音訊號輸出以供患者聆聽。

【0029】 該遙控裝置 8 可透過無線通訊技術與該助聽裝置 2 進行無線通訊與資料傳輸，可接收顯示該助聽裝置 2 之視野影像 40，所述無線通訊可以是 WiFi、ZigBee、NFC（近場通訊，Near Field Communication）或藍芽（Bluetooth）等，但不以上述類型為限。

【0030】 該遙控裝置 8 包括一用以顯示所接收之該視野影像 40 之觸控顯示模組 81，及一遙控模組 82。該觸控顯示模組 81 可供觸控點選其顯示之視野影像 40 中的人臉影像 401，該遙控模組 82 會對應該觸控顯示模組 81 被點選之每一個人臉影像 401 產生一點選訊號，該遙控裝置 8 會將所有點選訊號無線發送至該助聽裝置 2，以控制該助聽裝置 2 運作。

【0031】 該語音定位單元 621 可被該遙控裝置 8 所傳送之每一點選訊號驅動，而擷取放大輸出各點選訊號所對應之人臉影像 401 之方位的語音訊號，並將語音訊號傳送至該等聲音輸出器 7 輸出。

【0032】該遙控裝置 8 為一般常見之行動裝置，例如智慧型手機、平板電腦或筆記型電腦等，該遙控模組 82 是以 APP 模式程式化建構於該遙控裝置 8，可於被啟動時，使該遙控裝置 8 與該助聽裝置 2 無線訊號連結。但實施時，該遙控裝置 8 類型不以上述行動裝置類型為限，且該遙控模組 82 設置於該遙控裝置的型式不以 APP 方式為限，也可透過電子元件與軟體程式建構而成。

【0033】如圖 2、3、4 所示，以下接著說明本發明助聽系統之語音擷取方法的步驟：該語音擷取方法包含以下步驟：

【0034】（一）於使用者頭部配設該助聽裝置 2。將該眼鏡型式之載具 3 配置於使用者頭部，而當作一般眼鏡使用，並將該等聲音輸出器 7 分別塞裝於兩耳中。

【0035】（二）分析辨識語音。當使用者正面無語音訊號的環境時，該影像擷取器 4 與該影像分析模組 61 不會運作，僅該麥克風陣列單元 5 會持續感測周遭聲音訊號，該語音觸發單元 62 會持續接收分析該等聲音訊號內容，當該語音觸發單元 62 分析出來自使用者正面之聲音訊號含有語音訊號時，該語音觸發單元 62 便會觸發啟動該影像擷取器 4 與該影像分析模組 61。

【0036】（三）該影像擷取器 4 被觸發啟動而擷取該使用者視野前方之視野影像 40。

【0037】（四）分析視野影像 40 中之人臉影像 401 作為講話對象的可能性。當該影像擷取器 5 開始擷取輸出該視

野影像 40 時，該影像分析模組 61 會開始分析該視野影像 40，並於分析出該視野影像 40 中含有人臉影像 401 時，該距離分析單元 611、該方位分析單元 612 與該偏轉分析單元 613 會分別開始進一步分析每一個人臉影像 401 相對於使用者距離、方位與偏轉角度。然後，該可能性分析單元 614 會根據該可能性分類基準，統計分析出該視野影像 40 中之所有人臉影像 401 作為講話對象之可能性的可能性訊號源分類。

【0038】步驟（五）分析人臉影像 401 之嘴唇是否開合變形。在該可能性分析單元 614 分析輸出該可能性訊號源分類後，該唇形分析單元 615 會接續分析被分類有人臉影像 401 之可能性最高的級別內之所有人臉影像 401 的嘴唇是否開合變形，並對應嘴唇有開合變形的人臉影像 401 分別輸出一講話訊號。當無人臉影像 401 被分類於該高度可能性級別時，會接著分析該中度可能性級別之人臉影像 401 的嘴唇開合，以便擷取被分類於該中度可能性級別之人臉影像 401 方位的語音訊號。

【0039】步驟（六）過濾輸出嘴唇有開合變形的人臉影像 401 之方位的語音訊號。該語音定位單元 622 會被該講話訊號驅動，而透過波束形成方法過濾出該講話訊號對應之人臉影像 401 的方位的語音訊號。

【0040】步驟（七）遙控點選預定收音之講話對象。若使用者想要自己選擇要聆聽之講話對象時，可啟動該遙控裝置 8 之遙控模組 82，使該遙控裝置 8 開始與該助聽裝置

2 配對連結，並接收顯示出該助聽裝置 2 之該視野影像 40，並使該遙控裝置 8 於其顯示之該視野影像 40 中之人臉影像 401 被點選時，對應被點選之該人臉影像 401 發送一點選訊號至該助聽裝置 2。

【0041】步驟（八）過濾輸出被遙控點選之人臉影像 401 之方位的語音訊號。在該助聽裝置 2 與該遙控裝置 8 訊號連結後，該助聽裝置 2 會優先接受該遙控裝置 8 之控制。該聲音處理模組 62 會被該點選訊號驅動，透過波束形成方法過濾出被點選之人臉影像 401 之方位的語音訊號，讓使用者自己聆聽自己選擇之講話對象的語音。使用者點選人臉影像 401 時，可點選多個，或者是以圈選方式圈選出一個區域範圍內的多個人臉影像 401，而可選擇聆聽多位講話對象的講話聲。

【0042】透過上述語音擷取方法設計，可方便使用者自行選擇要透過該影像分析模組 61 來分析出最可能為講話對象之人臉影像 401 後，自動過濾輸出前述最可能為講話對象之人臉影像 401 方位的語音訊號，或者是要透過該遙控裝置 8 自行點選要過濾輸出之人臉影像 401 方位的語音訊號，可方便使用者根據當下講話環境的需求，自行選擇較合適的語音擷取方式，以準確獲得欲進行對話之對象或欲聆聽對象之語音訊號。

【0043】在本實施例中，該影像分析模組 61 於對影像擷取器 4 擷取之視野影像 40 中的該等人臉影像 401 分析輸出該可能性訊號源分類後，會再進一步配合分析最高可能性

級別或中度可能性級別內之人臉影像 401 之嘴唇是否開合變形的技術，使得該聲音處理模組 62 可準確地擷取出可能正在講話之人臉影像 401 之方位的語音訊號，而可大幅提高在多人環境中取得最可能是講話對象的語音訊號的成功率。

【0044】 但實施時，並不以設置該唇形分析單元 615 為必要，也就是說，在分析輸出該可能性訊號源分類後，不需再進行嘴唇變形分析，若有人臉影像 401 被分類於最高可能性級別，表示於該最高可能性級別中的這些對象都是相當重要的，為了避免遺漏某位對象的講話內容，所以可改成直接擷取輸出該最高可能性級別內之所有人臉影像 401 之方位的語音訊號，或者是在無人臉影像 401 被分類於最高可能性級別時，直接擷取輸出該中度可能性級別之所有人臉影像 401 之方位的語音訊號，若所有人臉影像 401 都被分類於低可能性級別時，則不對該等人臉影像進行語音擷取，此時，該耳機裝置 2 可進入一般常見之全面性收音模式，或者是僅擷取使用者前方之聲音的指向性收音模式，由於此兩模式為習知助聽器收音技術，因此不再詳述。

【0045】 此外，實施時，在分析輸出該可能性訊號源分類時，也可不考慮人臉左右偏轉之偏轉角度的運算權重比，僅透過相對於使用者之距離與方位，就足以有效提高在多人環境中取得最可能為講話對象的語音訊號的成功率。

【0046】 再者，實施時，該影像擷取器 4 不以設置在該

載具 3 之前載體 31 的左右向中心位置為必要，也可設置在該前載體 31 左右側，或者是改設在該載具 3 之其它部位。

【0047】 綜上所述，透過該麥克風陣列單元 5 所能提供之波束形成技術，並以該影像擷取器 4 之視野影像 40 輔助判斷最可能之講話對象的設計，使得該助聽裝置 2 可透過該視野影像 40 中之每一人臉影像 401 相對於使用者距離、方位與偏轉角度，先自該視野影像 40 中分析出最可能為講話對象的人臉影像 401 後，再進一步利用波束形成方法，過濾出最可能為講話對象之人臉影像 401 之方位的語音訊號，可大幅提高在人群中成功擷取到講話對象之語音的成功率，且可進一步配合辨識人臉影像 401 之嘴唇是否開合變形，更準確地自最可能為講話對象的多位對象中過濾輸出正在講話者的語音訊號，可改善目前助聽裝置 2 使用上的缺點。並可進一步配合該遙控裝置 8 可供使用者自行選擇要擷取語音之對象的設計，可方便使用者透過點選顯示於該觸控顯示模組 81 之視野影像 40 中的人臉影像 401，選擇有興趣聆聽之對象所在方位的語音，可明顯提高助聽系統的功能性與使用方便性，是一種創新的助聽裝置 2 與助聽系統設計。因此，確實能達成本發明之目的。

【0048】 惟以上所述者，僅為本發明之較佳實施例而已，當不能以此限定本發明實施之範圍，即大凡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】

【0049】

2	助聽裝置	611.....	距離分析單元
3	載具	612	方位分析單元
31	前載體	613	偏轉分析單元
310	鏡片	614	可能性分析單元
32	側載體	615	唇形分析單元
4	影像擷取器	62	聲音處理模組
40	視野影像	621	語音定位單元
401	人臉影像	62	語音觸發單元
402	視野中心	7	聲音輸出器
5	麥克風陣列單元	8	遙控裝置
51	麥克風	81	觸控顯示模組
6	微處理器	82	遙控模組
61	影像分析模組		

【生物材料寄存】

國內寄存資訊【請依：寄存機構、日期、號碼順序註記】

國外寄存資訊【請依：寄存國家、機構、日期、號碼順序註記】

【序列表】(請換頁單獨記載)



申請專利範圍

1. 一種助聽系統，包含一助聽裝置，該助聽裝置包括：

一影像擷取器，可擷取一使用者前方之視野影像；

一麥克風陣列單元，包括多個相間隔且可擷取聲音之麥克風；

一微處理器，訊號連接於該影像擷取器與該等麥克風，包括一影像分析模組，及一聲音處理模組，該影像分析模組可分析輸出該視野影像中的所有人脸影像相對於該使用者之距離與方位，並根據該等人臉影像之距離和方位，對該等人臉影像作為講話對象之可能性進行分類，而對應輸出一可能性訊號源分類，該可能性訊號源分類規劃有多個分別具有不同可能性的級別，該聲音處理模組可根據被分類於一可能性最高的級別內的所有人脸影像的方位，以及該等麥克風擷取之聲音訊號，透過波束形成方法過濾出該可能性最高的級別內的所有人脸影像的方位的語音訊號；及

至少一聲音輸出器，訊號連接於該微處理器，並可將該微處理器輸出之語音訊號輸出以供聆聽。

2. 如請求項 1 所述的助聽系統，其中，該影像分析模組包括一可分析出該視野影像中之該等人臉影像相對於使用者之距離遠近的距離分析單元、一可分析出該等人臉影像相對於使用者之方位的方位分析單元，及一內建有一可能性分類基準的可能性分析單元，該可能性分類基準內建有用以分析運算出該可能性訊號源分類時之人

臉影像之距離與方位的運算權重比，該可能性分析單元是根據每一人臉影像之距離與方位的運算權重比，分析統計輸出該可能性訊號源分類。

3. 如請求項 2 所述的助聽系統，其中，該影像分析模組還包括一可分析出該等人臉影像相對使用者左右偏轉之偏轉角度的偏轉分析單元，該可能性分類基準還內建有用以分析運算出該可能性訊號源分類時之偏轉角度的運算權重比，且該影像分析模組之該可能性分析單元是根據該等人臉影像之距離、方位與偏轉角度的運算權重比，分析統計輸出關於該等人臉影像之該可能性訊號源分類。
4. 如請求項 2 或 3 所述的助聽系統，其中，該影像分析模組還包括一唇形分析單元，該唇形分析單元可分析一分類有至少一人臉影像且可能性最高之級別內的所有人臉影像的嘴唇是否有開合變形，且會對應每一個嘴唇有開合變形之人臉影像輸出一講話訊號，該聲音處理模組可被該講話訊號觸發，而過濾出該講話訊號對應之人臉影像之方位的語音訊號。
5. 如請求項 2 或 3 所述的助聽系統，其中，該助聽系統還包含一與該微處理器無線訊號連接之遙控裝置，該遙控裝置包括一可顯示該微處理器無線發送之該視野影像的觸控顯示模組，及一遙控模組，該觸控顯示模組可供觸控點選該視野影像中之該等人臉影像，該遙控模組可於該觸控顯示模組顯示之一人臉影像被點選時，輸出一

對應該被點選之人臉影像的點選訊號，該遙控裝置可將該點選訊號無線發送至該微處理器，該聲音處理模組會被該點選訊號驅動，而透過波束形成方法過濾出該點選訊號代表之人臉影像之方位的語音訊號。

6. 如請求項 5 所述的助聽系統，其中，該遙控裝置為智慧型手機、平板電腦或筆記型電腦，該遙控模組是程式化安裝建構於該遙控裝置。
7. 如請求項 5 所述的助聽系統，其中，該遙控裝置與該微處理器間是透過 WiFi、ZigBee、藍芽或 NFC 無線通訊方式進行無線通訊。
8. 如請求項 2 或 3 所述的助聽系統，其中，該聲音處理模組包括一可於分析該等麥克風接收輸出之聲音訊號中含有語音訊號時，觸發啟動該影像擷取器開始擷取影像之語音觸發單元。
9. 如請求項 1 所述的助聽系統，還包含一可配戴於使用者頭部之載具，該載具包括一左右延伸且位於使用者臉部前側之前載體，及二左右間隔地安裝於該前載體並可供配戴於雙耳之側載體，該影像擷取器是安裝於該前載體，該等麥克風是分別安裝於該等側載體。
10. 如請求項 9 所述的助聽系統，其中，部分麥克風是左右間隔地安裝於該前載體。
11. 一種助聽系統之語音擷取方法，包含以下步驟：
 - (a) 將一助聽裝置配置於使用者身上，該助聽裝置包括一影像擷取器與一麥克風陣列單元，該麥克風陣列

單元包括多個相間隔之麥克風；

(b)以該影像擷取器擷取該使用者視野前方之視野影像；

(c)使該助聽裝置自該視野影像中辨識出人臉影像，並判斷出該等人臉影像分別相對於該使用者的距離與方位；

(d)使該助聽裝置根據該等人臉影像分別相對於該使用者之距離與方位，對該等人臉影像作為講話對象之可能性進行級別分類，而對應輸出一可能性訊號源分類；及

(e)使該助聽裝置利用該等麥克風感測輸出的聲音訊號，並透過波束形成方法，將步驟(d)之可能性訊號源分類中被分類於一最高可能性的級別內的所有人臉影像的方位的語音訊號過濾出來。

12. 如請求項 11 所述的助聽系統之語音擷取方法，其中，步驟(d)是使該助聽裝置根據人臉影像偏離該視野影像之一視野中心的偏離角度來定義出相對該使用者之方位，且針對人臉影像相對於該使用者之距離大小，及方位分別給予不同之運算權重比，並根據該等人臉影像之距離與方位分別對應的運算權重比，分析統計輸出關於該等人臉影像作為談話對象之可能性的可能性訊號源分類。

13. 如請求項 12 所述的助聽系統之語音擷取方法，其中，步驟(c)還使該助聽裝置分析辨識該等人臉影像相對

使用者左右偏轉的偏轉角度，步驟（d）是使該助聽裝置針對人臉影像之偏轉角度大小分別給予不同之運算權重比，且根據每一人臉影像之距離、方位與偏轉角度分別對應的運算權重比，分析統計輸出該可能性訊號源分類。

14. 如請求項 11、12 或 13 所述的助聽系統之語音擷取方法，還包含一步驟（f）使該助聽裝置分析一分類有至少一人臉影像且可能性最高之級別內的所有人臉影像的嘴唇是否有開合變形，並對應嘴唇有開合變形之人臉影像分別輸出一講話訊號，步驟（e）是使該助聽裝置僅輸出該講話訊號對應之人臉影像之方位的語音訊號。
15. 如請求項 11、12 或 13 所述的助聽系統之語音擷取方法，還包含一步驟（g）與一步驟（h），該步驟（g）是使該助聽裝置將該視野影像傳送顯示於一遙控裝置，並使該遙控裝置於其顯示之該視野影像中之至少一人臉影像被點選時，對應被點選之該人臉影像發送一點選訊號至該助聽裝置，該步驟（h）是使該助聽裝置於收到該點選訊號時，透過波束形成方法，過濾出該點選訊號所對應之人臉影像之方位的語音訊號。
16. 如請求項 11 所述的助聽系統之語音擷取方法，其中，步驟（b）該助聽裝置是於該等麥克風接收之聲音訊號中含有語音時，才啟動該影像擷取器開始擷取影像。
17. 如請求項 11 所述的助聽系統之語音擷取方法，其中，步驟（a）之耳機裝置具有一可供配戴於使用者頭部之

載具，該耳機裝置是配置於使用者頭部。



圖式

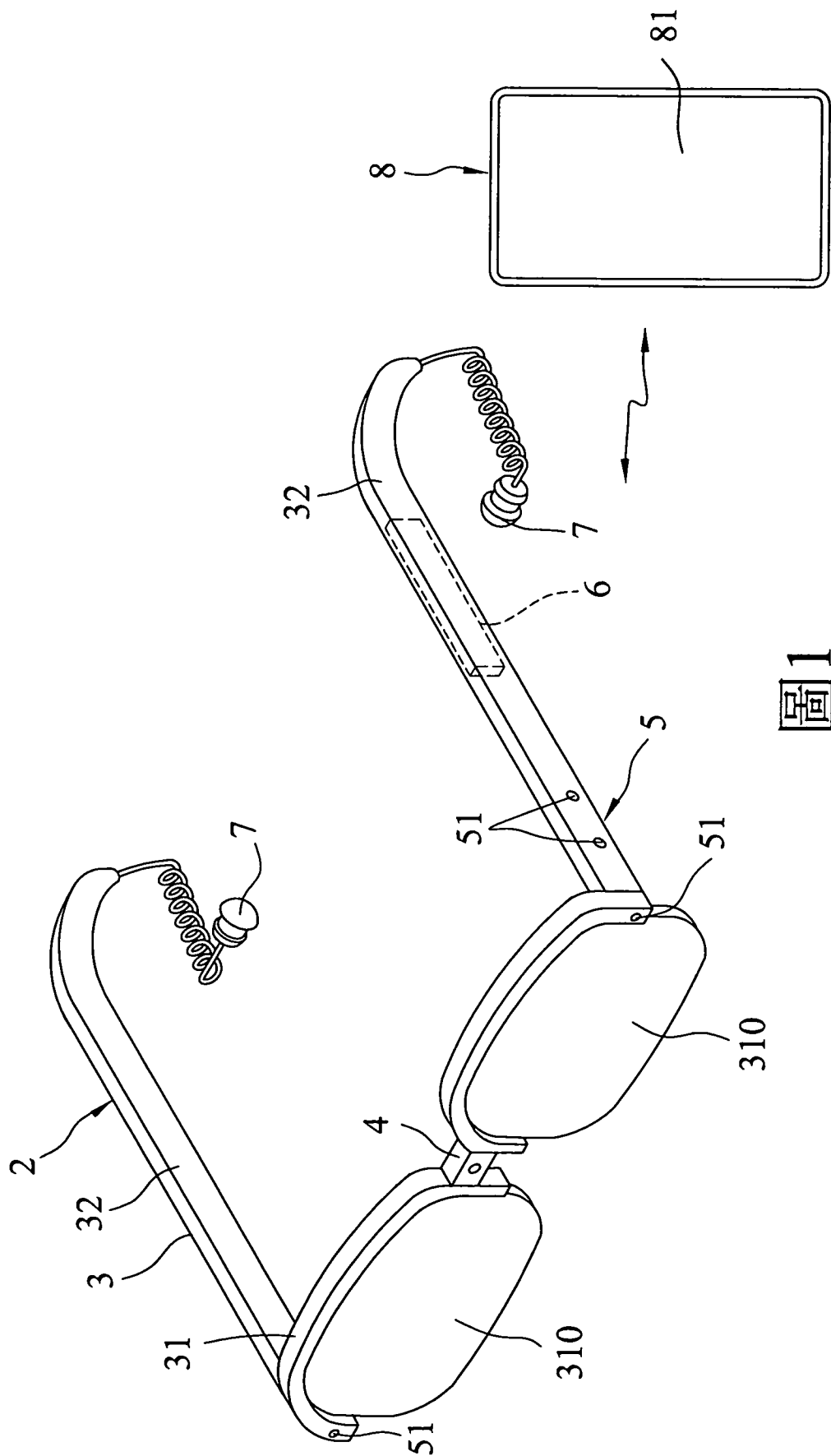


圖1

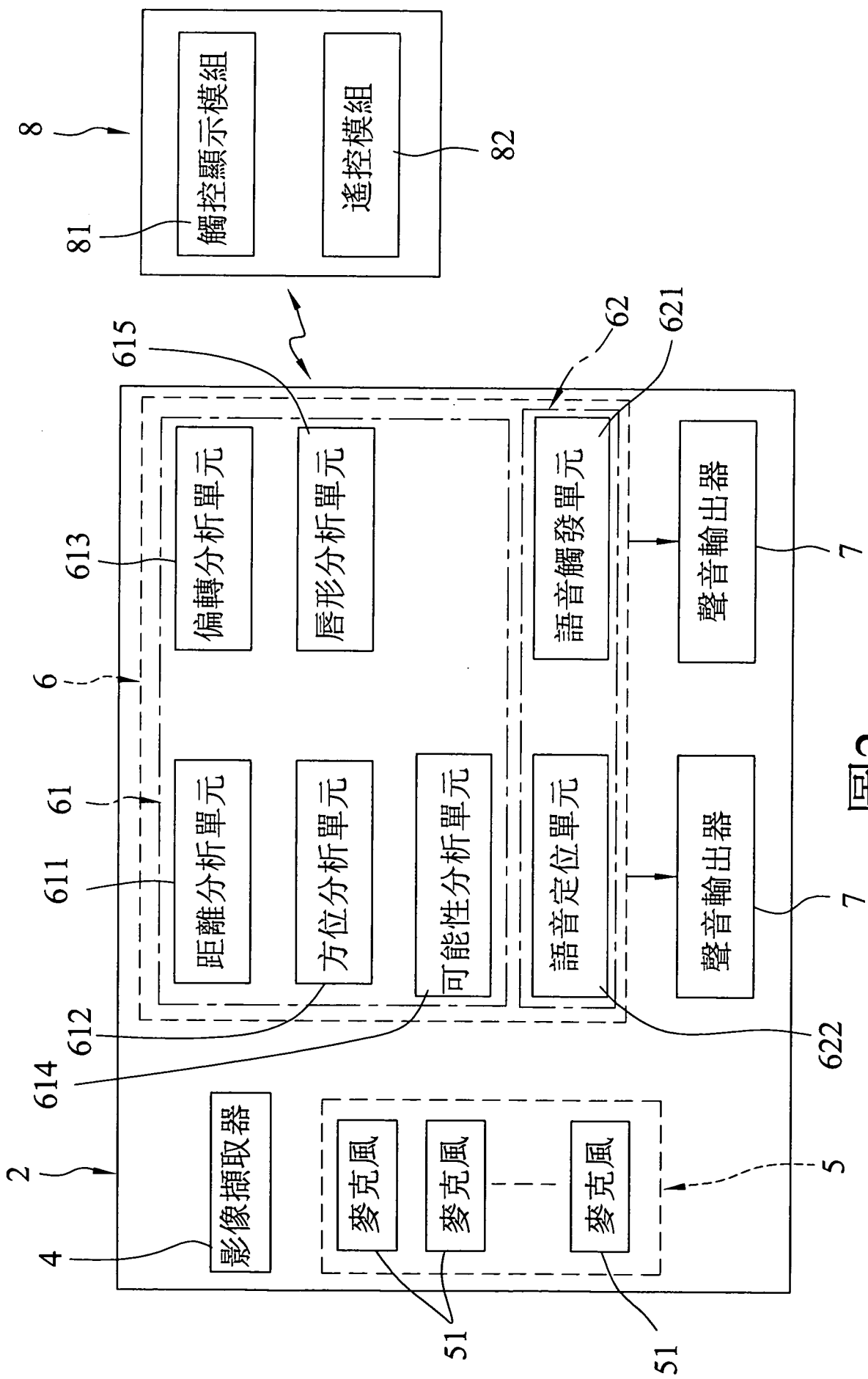


圖2



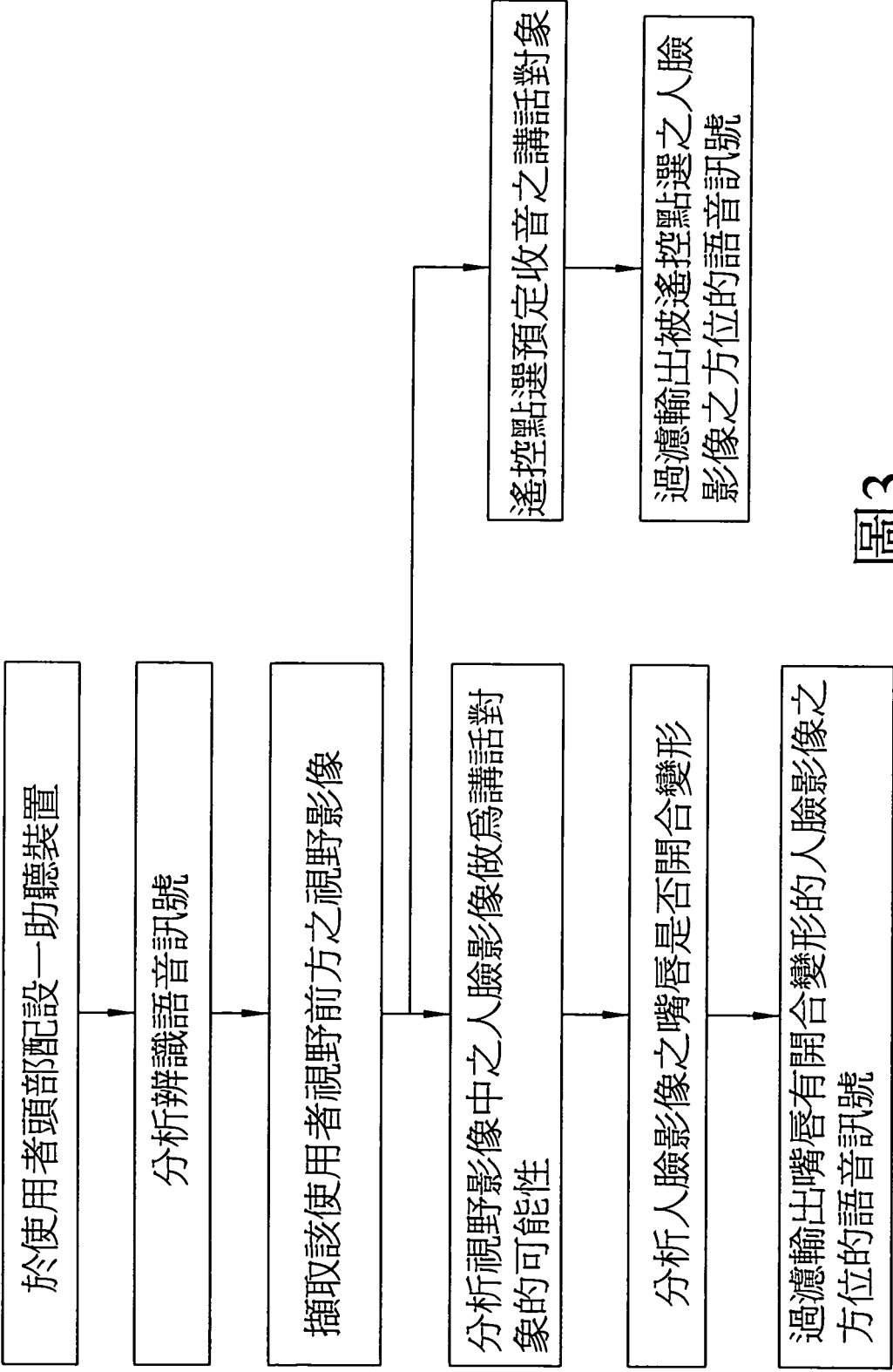


圖3

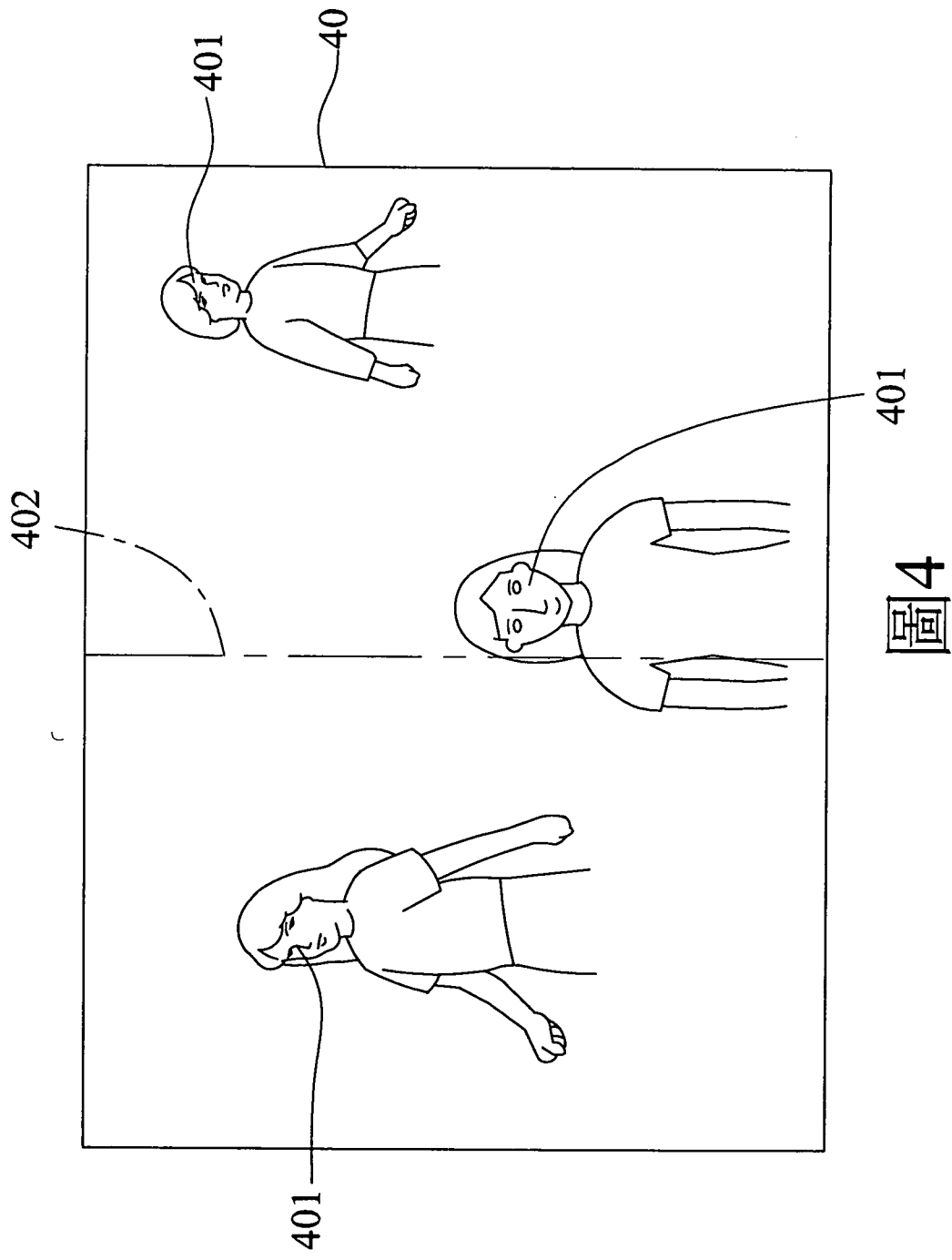


圖4

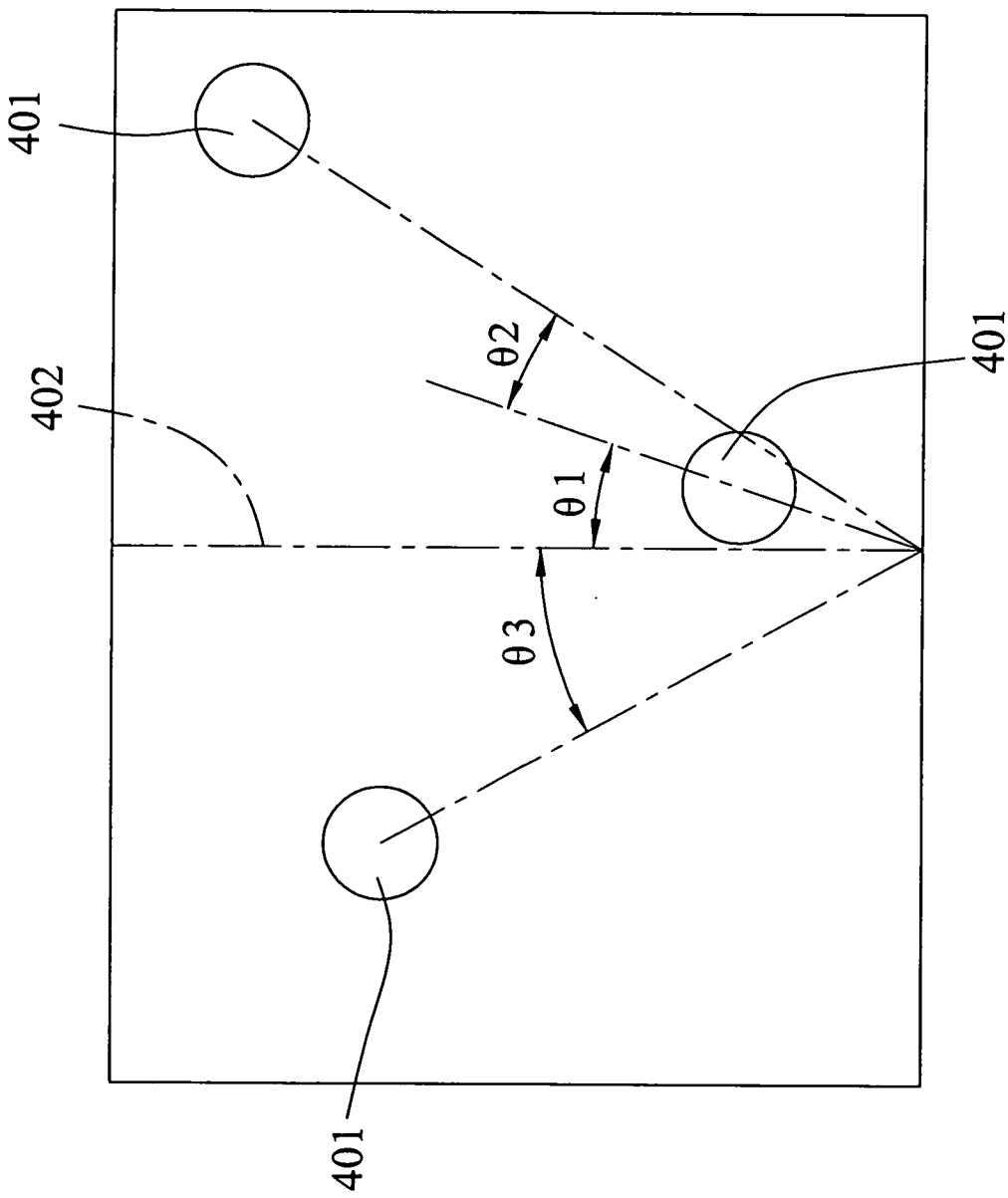


圖 5