



(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 I456515 B

(45) 公告日：中華民國 103 (2014) 年 10 月 11 日

(21) 申請案號：101125278

(22) 申請日：中華民國 101 (2012) 年 07 月 13 日

(51) Int. Cl. : G06K9/80 (2006.01)

G10L15/06 (2013.01)

G10L17/12 (2013.01)

G06F17/10 (2006.01)

(71) 申請人：國立交通大學 (中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)

新竹市大學路 1001 號

(72) 發明人：宋開泰 SONG, KAI TAI (TW)；簡碩成 CHIEN, SHUO CHENG (TW)；林昭宇 LIN,

CHAO YU (TW)；陳奕彰 CHEN, YI WEN (TW)；陳信宏 CHEN, SIN HORNG

(TW)；江振宇 CHANG, CHEN YU (TW)；吳宜樵 WU, YI CHIAO (TW)

(74) 代理人：李國光；張仲謙

(56) 參考文獻：

US 6964023B2

US 20100158324A1

US 2010/0271507A1

審查人員：李國福

申請專利範圍項數：21 項 圖式數：6 共 30 頁

(54) 名稱

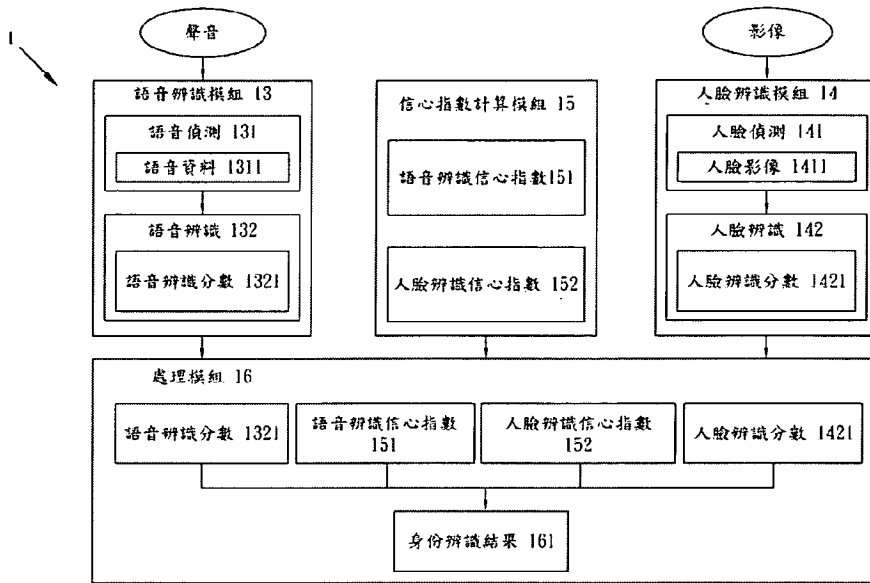
融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人

HUMAN IDENTIFICATION SYSTEM BY FUSION OF FACE RECOGNITION AND SPEAKER RECOGNITION, METHOD AND SERVICE ROBOT THEREOF

(57) 摘要

本發明係揭露一種融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人。此系統採用人臉辨識及語音辨識兩方的結果經過融合計算後產生最終結果。另外，此系統能藉由信心指數評估人臉辨識及語音辨識兩方辨識結果的可信度，若信心指數超出門檻值更納入融合計算。若只有一方的信心指數超過門檻值就只採納一方的結果做為輸出，如果兩方都超出門檻值則使用融合演算法輸出結果。

The present invention discloses a human identification system by fusion of face recognition and speaker recognition, the method and the service robot thereof. The system fuses the results of the face recognition and the speaker recognition. Besides, the system uses the confidence index to estimate the confidence level of the two recognition results. If only one of the confidence indices of the two recognition results reaches the threshold, then only this result is used to act as the output. If the confidence indices of the two recognition results all reach the threshold, then two recognition results are fused to output the final result.



第1圖

- 1 . . . 身份辨識系統
- 13 . . . 語音辨識模
組
- 131 . . . 語音偵測
- 1311 . . . 語音資料
- 132 . . . 語音辨識
- 1321 . . . 語音辨識
分數
- 14 . . . 人臉辨識模
組
- 141 . . . 人臉偵測
- 1411 . . . 人臉影像
- 142 . . . 人臉辨識
- 1421 . . . 人臉辨識
分數
- 15 . . . 信心指數計
算模組
- 151 . . . 語音辨識信
心指數
- 152 . . . 人臉辨識信
心指數
- 16 . . . 處理模組
- 161 . . . 身份辨識結
果



申請日: 101.7.13

IPC分類: G06K 9/80 (2006.01)
G10 L 15/06 (2006.01)
G10 L 17/12 (2013.01)
G06F 17/16 (2006.01)

【發明摘要】

【中文發明名稱】 融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人

【英文發明名稱】 HUMAN IDENTIFICATION SYSTEM BY FUSION OF FACE RECOGNITION AND SPEAKER RECOGNITION, METHOD AND SERVICE ROBOT THEREOF

【中文】

本發明係揭露一種融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人。此系統採用人臉辨識及語音辨識兩方的結果經過融合計算後產生最終結果。另外，此系統能藉由信心指數評估人臉辨識及語音辨識兩方辨識結果的可信度，若信心指數超出門檻值更納入融合計算。若只有一方的信心指數超過門檻值就只採納一方的結果做為輸出，如果兩方都超出門檻值則使用融合演算法輸出結果。

【英文】

The present invention discloses a human identification system by fusion of face recognition and speaker recognition, the method and the service robot thereof. The system fuses the results of the face recognition and the speaker recognition. Besides, the system uses the confidence index to estimate the confidence level of the two recognition results. If only one of the confidence indices of the two recognition results reaches the threshold, then only this result is used to act as the output. If the confidence indices of the two recognition results all reach the threshold, then two recognition results are fused to output the final result.

【指定代表圖】 第(1)圖。

【代表圖之符號簡單說明】

- 1：身份辨識系統
- 13：語音辨識模組
- 131：語音偵測
- 1311：語音資料
- 132：語音辨識
- 1321：語音辨識分數
- 14：人臉辨識模組
- 141：人臉偵測
- 1411：人臉影像
- 142：人臉辨識
- 1421：人臉辨識分數
- 15：信心指數計算模組
- 151：語音辨識信心指數
- 152：人臉辨識信心指數
- 16：處理模組
- 161：身份辨識結果

【特徵化學式】

無

【發明說明書】

【中文發明名稱】 融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人

【英文發明名稱】 HUMAN IDENTIFICATION SYSTEM BY FUSION OF FACE RECOGNITION AND SPEAKER RECOGNITION, METHOD AND SERVICE ROBOT THEREOF

【技術領域】

【0001】 本發明係有關於一種身份辨識系統，特別是有關於一種整合人臉辨識及語音辨識，並能夠計算兩者之可靠度，以提高系統之強健性的身份辨識系統。

【先前技術】

【0002】 由於科技的進步，人類對於生活品質的要求也愈來愈高，因此，各種能夠使人類生活更加便利的產品也不斷的被開發出來。其中，服務型機器人的需求量逐年上升，其能夠輔助人類在艱困的環境下工作，也可作為醫療照護用的醫療保健用機器人，或作為兼具娛樂性質及家事服務性質的移動式家用機器人。而服務型機器人的關鍵技術則為身份辨識技術。在另一方面，身份辨識技術也被大量的運用於現今的保全系統。

【0003】 然而，習知技藝之身份辨識系統也存在著許多有待克服的缺點。例如，美國專利第6567775號揭露一種身份辨識系統，其能利用在影片中所有的影像以及聲音做人臉辨識以及語音辨識，最後再進行融合計算。雖然這種方法同時採用人臉辨識以及語音辨識，

能提高身份辨識的正確率，但由於此系統是利用影片做為輸入，因此無法達到即時性。另外，此系統並沒有對人臉辨識以及語音辨識做可信度的評估，因此在複雜的環境之下仍然很容易會出現誤判。

【0004】 而美國專利第2005/0027530號也揭露一種身份辨識系統，其利用隱馬爾可夫模組（Hidden Markov Models，HMM）並融合人臉辨識以及語音辨識。雖然此系統也整合了人臉辨識以及語音辨識，並利用特殊的機制以提高其辨識正確率。然而，此系統仍然沒有對人臉辨識以及語音辨識做可信度的評估，因此在一方無法正常運作的情況之下，此系統則無法運作。

【0005】 另外，習知技藝之身份辨識系統大都僅能應用於固定的地點，其具備預先設置好的辨識環境，因此進行辨識時不會有太大的失誤。然而，當辨識環境會產生變化時，例如應用於服務型機器人，由於服務型機器人並非設置於固定的地點，而是會四處走動，因此辨識環境，如光線及噪音等，也會產生變化。因此，習知技藝之身份辨識系統很容易便會產生誤判或無法辨識的情況。

【0006】 因此，如何提出一種身份辨識系統，不但具有即時性，並且能夠提高其身份辨識的正確率，更能在複雜且變動的環境之下做出正確的身份辨識，即為本發明所欲解決的問題。

【發明內容】

【0007】 有鑑於上述習知技藝之問題，本發明之其中一目的就是在提供一種融合人臉辨識及語音辨識之身份辨識系統，以解決習知技藝之身份辨識系統在辨識環境不佳時容易產生誤判的問題。

【0008】 根據本發明之目的，提出一種融合人臉辨識及語音辨識之身份辨識系統。此系統可包含人臉辨識模組、語音辨識模組、信心指數計算模組及處理模組。人臉辨識模組可擷取影像，並對影像執行人臉偵測以產生人臉影像，再對人臉影像執行人臉辨識，以獲得對應於預設之複數個成員的複數個人臉辨識分數。語音辨識模組可擷取聲音資料，並偵測聲音資料中的語音，以產生語音資料，再對語音資料執行語音辨識，以獲得對應於複數個成員的複數個語音辨識分數。信心指數計算模組可計算人臉辨識的可靠度及語音辨識的可靠度，以產生人臉辨識信心指數及語音辨識信心指數。處理模組可根據複數個人臉辨識分數、人臉辨識信心指數、複數個語音辨識分數及語音辨識信心指數產生身份辨識結果。

【0009】 根據本發明之目的，再提出一種融合人臉辨識及語音辨識之身份辨識方法。此方法可包含下列步驟：利用人臉辨識模組擷取影像，並對影像執行人臉偵測以產生人臉影像，再對人臉影像執行人臉辨識，以獲得對應於預設之複數個成員的複數個人臉辨識分數；藉由語音辨識模組擷取聲音資料，並偵測聲音資料中的語音，以產生語音資料，再對語音資料執行語音辨識，以獲得對應於複數個成員的複數個語音辨識分數；經由信心指數計算模組計算人臉辨識的可靠度及語音辨識的可靠度，以產生人臉辨識信心指數及語音辨識信心指數；以及透過處理模組根據複數個人臉辨識分數、人臉辨識信心指數、複數個語音辨識分數及語音辨識信心指數產生身份辨識結果。

【0010】 較佳地，當人臉辨識信心指數超過預設門檻值時而語音辨識信心指數未超過預設門檻值時，處理模組可根據複數個人臉辨識分數

產生身份辨識結果。

【0011】 較佳地，當語音辨識信心指數超過預設門檻值時而人臉辨識信心指數未超過預設門檻值時，處理模組可根據複數個語音辨識分數產生身份辨識結果。

【0012】 較佳地，當人臉辨識信心指數及語音辨識信心指數均未超過預設門檻值時，處理模組可判定身份辨識結果為無法識別。

【0013】 較佳地，當人臉辨識信心指數及語音辨識信心指數均超過預設門檻值時，處理模組可根據複數個人臉辨識分數、人臉辨識信心指數、複數個語音辨識分數及語音辨識信心指數進行融合計算，以產生身份辨識結果。

【0014】 較佳地，融合計算可由處理模組將各個複數個成員的人臉辨識分數乘上人臉辨識信心指數以產生最終人臉辨識分數，並將各個成員的語音辨識分數乘上語音辨識信心指數最終語音辨識分數，再將對應的最終人臉辨識分數及最終語音辨識分數相加以產生對應於各個成員的身份辨識分數，並根據身份辨識分數產生身份辨識結果。

【0015】 較佳地，信心指數計算模組可根據亮度因子及辨識分數差異性因子來計算人臉辨識信心指數。

【0016】 較佳地，亮度因子可符合下列關係式：

$$C1 = \begin{cases} (G^{avg}) : G^{avg} < T^{low} & ; \\ (1 - G^{avg}) : G^{avg} > T^{high} \\ 1 \cdot T^{low} \leq G^{avg} \leq T^{high} \end{cases}$$

【0017】 其中，C1為亮度因子， T^{high} 為影像之灰階值經正規化至0~1範圍

後之正常亮度之門檻值上限， T^{low} 為影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值下限， G^{avg} 為影像之灰階值經正規化至0~1範圍後的平均值。

【0018】 較佳地，辨識分數差異性因子可符合下列關係式：

$$C2 = \begin{cases} (Y^{1st} - Y^{2nd}) / T, & Y^{1st} - Y^{2nd} < T \\ 1, & Y^{1st} - Y^{2nd} \geq T \end{cases} ;$$

【0019】 其中，C2為辨識分數差異性因子，T為複數個人臉辨識分數中最高分數與最低分數的差距門檻值， Y^{1st} 為複數個人臉辨識分數中最高分數， Y^{2nd} 為複數個人臉辨識分數中次高分數。

【0020】 較佳地，人臉辨識信心指數可符合下列關係式：

$$C = C1 \times C2 ;$$

【0021】 其中，C為人臉辨識信心指數。

【0022】 較佳地，信心指數計算模組可根據信心指數曲線來計算語音辨識信心指數，信心指數曲線係符合下列關係式：

$$y = F_R(x) - (1 - F_E(x)) ;$$

【0023】 其中， $F_R(x)$ 為經由事前訓練所獲得的辨識正確分數的累積分佈函數， $F_E(x)$ 係為經由事前訓練所獲得的辨識錯誤分數的累積分佈函數，信心指數曲線之x軸代表經過正規化至0~1後之語音辨識分數，信心指數曲線之y軸則代表經過正規化至0~1後之語音辨識信心指數。

【0024】 根據本發明之目的，又提出一種服務型機器人。此機器人可包含電源供應器及融合人臉辨識及語音辨識之身份辨識系統。此服務

型機器人可利用該融合人臉辨識及語音辨識之身份辨識系統對使用者之身份進行辨識，以決定使用者之使用權限，此融合人臉辨識及語音辨識之身份辨識系統係如同上述。

【0025】 承上所述，依本發明之融合人臉辨識及語音辨識之身份辨識系統、其方法及其服務型機器人，其可具有一或多個下述優點：

【0026】 (1) 本發明可利用即時的資訊來做身份辨識，因此具有即時性。

【0027】 (2) 本發明整合了人臉辨識以及語音辨識，因此能夠有效地提高其辨識的正確率。

【0028】 (3) 本發明能對人臉辨識以及語音辨識做可信度的評估，因此即使在複雜的環境之下也能做出正確的辨識。

【圖式簡單說明】

【0029】 第1圖 係為本發明之融合人臉辨識及語音辨識之身份辨識系統之方塊圖。

第2圖 係為本發明之融合人臉辨識及語音辨識之身份辨識系統之第一實施例之方塊圖。

第3圖 係為本發明之融合人臉辨識及語音辨識之身份辨識系統之第一實施例之流程圖。

第4圖 係為本發明之融合人臉辨識及語音辨識之身份辨識系統之融合辨識實驗結果之統計圖。

第5圖 係為本發明之融合人臉辨識及語音辨識之身份辨識系統之語音辨信心指數算法之示意圖。

第6圖 係為本發明之融合人臉辨識及語音辨識之身份辨識方法之流程圖。

【實施方式】

【0030】 以下將參照相關圖式，說明依本發明之融合人臉辨識及語音辨識之身份辨識系統之實施例，為使便於理解，下述實施例中之相同元件係以相同之符號標示來說明。

【0031】 請參閱第1圖，其係為本發明之融合人臉辨識及語音辨識之身份辨識系統之方塊圖。如圖所示，身份辨識系統1包含語音辨識模組13、人臉辨識模組14、信心指數計算模組15及處理模組16。

【0032】 語音辨識模組13擷取聲音資料，執行語音偵測131以收集聲音資料中之語音2，以產生語音資料1311。此時，語音辨識模組13再利用資料庫（未繪於圖中）儲存的預設的複數個成員的語音樣本對語音資料1311執行語音辨識132，以產生對應於各個成員的語音辨識分數1321。

【0033】 同樣的，人臉辨識模組14擷取影像，並對影像執行人臉偵測141以產生人臉影像1411。而此時，人臉辨識模組14再利用資料庫（未繪於圖中）儲存的預設的複數個成員的人臉樣本對人臉影像1411執行人臉辨識142，以產生對應於各個成員的人臉辨識分數1421。

【0034】 信心指數計算模組15則可根據人臉辨識142及語音辨識132的環境及其它因素，來測定人臉辨識142及語音辨識132的可靠度，以產生人臉辨識信心指數152及語音辨識信心指數151。而處理模組16則可根據各個成員的語音辨識分數1321、語音辨識信心指數151、各個成員的人臉辨識分數1421及人臉辨識信心指數152來產生身份辨識結果161。

- 【0035】由於外在的環境會嚴重影響人臉辨識142及語音辨識132的可靠度。例如，環境的亮度能夠影響人臉辨識142的精確度，因此信心指數計算模組15偵測到環境的亮度過低時，會產生較低的人臉辨識信心指數152。當人臉辨識信心指數152低於預設門檻值時，處理模組16則不採用人臉辨識模組14的辨識結果。
- 【0036】同樣的，若是環境過於吵雜，語音辨識132的可靠度也會降低，因此信心指數計算模組15偵測到環境的噪音過多時，會產生較低的語音辨識信心指數151。當語音辨識信心指數151低於預設門檻值時，處理模組16則不採用語音辨識模組13的辨識結果。
- 【0037】而當人臉辨識信心指數152及語音辨識信心指數151均超過預設門檻值時，表示人臉辨識142及語音辨識132的均具有一定的可靠度。此時，處理模組16則可同時採用語音辨識模組13的辨識結果與人臉辨識模組14的辨識結果以進行融合計算，以產生身份辨識結果161。
- 【0038】值得一提的是，習知技藝之身份辨識系統雖然已經整合了語音辨識及人臉辨識，並且採用二者之辨識結果進行融合計算以提高辨識的精確度。然而，由於習知技藝之身份辨識系統無法判定語音辨識及人臉辨識的可靠度，因此，當語音辨識及人臉辨識有一方處於辨識困難的環境或是兩方都處於辨識困難的環境時，融合辨識結果並無法有效地提高辨識的精確度。
- 【0039】例如，當環境非常吵雜時，語音辨識的結果根本不具備參考價值，若是單純使用融合語音辨識及人臉辨識的辨識結果，反而會使辨識的精確度下降。相反的，本發明提出了信心指數的機制，當

辨識環境不良時，信心指數會降低，當信心指數低於預設門檻值時就不予採用，故本發明能夠有效地降低環境的影響，提高身份辨識系統的強健性。

【0040】 當應用於服務型機器人，環境之變化將持續發生，對於環境的影響更需加以考量。由於機器人之移動造成光線之變化、背光、陰影，或聲源之方向、強度變化等等；對個人化設計之服務型機器人而言，身分辨識錯誤將造成嚴重執行上之差異，甚至不良之後果。

【0041】 請參閱第2圖，其係為本發明之融合人臉辨識及語音辨識之身份辨識系統之第一實施例之方塊圖。如圖所示，身份辨識系統2包含麥克風21、聲音變化偵測器28、語音辨識器23、語音辨識信心指數計算器25、攝影機22、人臉偵測器29、人臉辨識器24、人臉辨識信心指數計算器27、處理器26。身份辨識系統2利用麥克風21擷取聲音資料。而聲音變化偵測器（Voice Activity Detection, VAD）28則接收聲音資料，並執行語音偵測281收集其中的語音資料2811。語音辨識器23則再利用資料庫（未繪於圖中）儲存的預設的複數個成員的語音樣本對語音資料2811執行語音辨識232，以產生對應於各個成員的語音辨識分數2321。語音辨識信心指數計算器25則可根據事前訓練而得之信心指數曲線251來產生語音辨識信心指數252。

【0042】 語音辨識信心指數的計算方式可經由事前訓練，利用大量實驗的結果得到辨識正確分數的累積分佈函數，以及辨識錯誤分數的累積分佈函數，並綜合兩者之結果求得信心指數曲線251。信心指數曲線251可以利用下列的關係式來表示：

$$y = F_R(x) - (1 - F_E(x)) ;$$

【0043】 其中， $F_R(x)$ 係為經由事前訓練所獲得的辨識正確分數的累積分佈函數， $F_E(x)$ 係為經由事前訓練所獲得的辨識錯誤分數的累積分佈函數，信心指數曲線251之x軸代表經過正規化至0~1後之語音辨識分數2321，信心指數曲線之y軸則代表經過正規化至0~1後之語音辨識信心指數252。因此，在獲得各個成員的語音辨識分數2321後，語音辨識信心指數計算器25即可根據信心指數曲線251產生對應的語音辨識信心指數252，以估測語音辨識232的可靠度。

【0044】 同樣的，身份辨識系統2利用攝影機22擷取影像。人臉偵測器29則對此影像執行人臉偵測291，以產生人臉影像2911。人臉辨識器24再利用資料庫（未繪於圖中）儲存的預設的複數個成員的人臉樣本對人臉影像2911執行人臉辨識242，以產生對應於各個成員的人臉辨識分數2421。

【0045】 其中，在執行人臉辨識242的過程中，影像會先經過前處理，以減少影像在不同照度、不同大小及臉部旋轉角度等情況的差異性，並經過正規化的調整，以得到標準的影像形式。而在演算法的部份，利用主成份分析法（Principal Component Analysis，PCA）對輸入的臉部影像進行資料維度的化減，以減少辨識所需要的資料量。最後可利用RBF（Radial Basis Function）類神經網路對化減後的資訊進行辨識。RBF類神經網路輸出的結果會被正規化至0~1，即為輸入影像相對於資料庫每位成員的辨識分數，當輸入影像與資料庫每位成員相似度愈高則分數愈高。計算出

來的辨識分數即可進行信心指數的計算。人臉辨識信心指數計算器27則可根據亮度因子271及辨識分數差異性因子273來計算人臉辨識信心指數272。

【0046】 亮度因子271可利用下列關係式來表示：

$$C1 = \begin{cases} (G^{avg})^2 \cdot G^{avg} < T^{low} & ; \\ (1-G^{avg})^2 \cdot G^{avg} > T^{high} \\ 1, T^{low} \leq G^{avg} \leq T^{high} \end{cases}$$

【0047】 其中，C1為亮度因子271， T^{high} 為所擷取影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值上限(可為0.6)， T^{low} 為所擷取影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值下限(可為0.4)， G^{avg} 為所擷取影像之灰階值經正規化至0~1範圍後的平均值。

【0048】 而辨識分數差異性因子273則可以利用下列關係式來表示：

$$C2 = \begin{cases} (Y^{1st} - Y^{2nd}) / T \cdot Y^{1st} - Y^{2nd} < T & ; \\ 1, Y^{1st} - Y^{2nd} \geq T \end{cases}$$

【0049】 其中，C2為辨識分數差異性因子273，T為人臉辨識分數2421中最高分數與最低分數的差距門檻值， Y^{1st} 為人臉辨識分數2421中最高分數， Y^{2nd} 為人臉辨識分數2421中次高分數。也就是說，當各個成員的人臉辨識分數2421差異性愈高時，人臉辨識242的可靠度也愈高，相反的，當各個成員的人臉辨識分數2421差異性愈低時，人臉辨識242的可靠度也愈低。

【0050】 因此，人臉辨識信心指數272則可以利用下列關係式來表示：

$$C = C1 \times C2 \quad ;$$

【0051】 其中，C為人臉辨識信心指數272。因此，在獲得各個成員的人臉

辨識分數2421後，人臉辨識信心指數計算器27即可根據亮度因子271及辨識分數差異性因子273產生人臉辨識信心指數272，以估測人臉辨識242的可靠度。當然，前述僅為舉例，人臉辨識信心指數計算器27及語音辨識信心指數計算器25更可根據其它因素來決定人臉辨識信心指數272及語音辨識信心指數252，以估測兩者的可靠度。

【0052】 得到了人臉辨識信心指數272、各個成員的人臉辨識分數2421、語音辨識信心指數252、各個成員的語音辨識分數2321，處理器26則可以進行信心指數與門檻值的比較，結果可以分為下列三種：

【0053】 (1) 只有一方的信心指數超過門檻值。

【0054】 (2) 兩方的信心指數都超過門檻值。

【0055】 (3) 沒有任何一方的信心指數超過門檻值。

【0056】 當人臉辨識信心指數272及語音辨識信心指數252中僅有一方信心指數超過門檻值，這表示有一方辨識環境較差，其結果不具參考價值。因此，處理器26僅會採取超過門檻值的結果來做為身份辨識結果261。而如果兩方的信心指數都低於門檻值，這表示兩方的辨識環境都不佳，兩方的結果都不具參考價值。因此，處理器26則判定為無法辨識，並輸出無辨識結果。

【0057】 而當兩方的信心指數都超過門檻值時，這表示兩方的結果應都具參考價值，此時處理器26則進行融合計算，將各個成員的人臉辨識分數2421乘上人臉辨識信心指數272，並將各個成員的語音辨識分數2321乘上語音辨識信心指數252，並將兩者的結果相加。

每個資料庫的成員均有各自的人臉辨識分數2421及語音辨識分數2321，將這二個辨識分數乘上對應的信心指數後相加，即可得到最後的辨識分數。當得到最後的辨識分數時，選取最高分即可以得到最後的身份辨識結果261。表1則提供了一個融合計算的範例。

表1人臉辨識與語音辨識的融合範例

	信心指數	A	A*	B	B*	C	C*	D	D*
人臉辨識	0.8	0.30	0.240	0.45	0.360	0.20	0.160	0.05	0.040
語音辨識	0.9	0.40	0.360	0.50	0.450	0.01	0.009	0.09	0.081
加成結果			0.600		0.810		0.169		0.121

【0058】如表1所示，身份辨識系統共有A、B、C、D這四個人的資料，人臉辨識及語音辨識的信心指數分別為0.8及0.9。而A、B、C、D的數值分別為人臉辨識與語音辨識的數值。A*、B*、C*、D*的數值的分別代表信心指數乘上辨識結果的數值。

- 【0059】 在雙方的信心指數均高的情況之下，經由融合計算，B獲得最高分，因此身份辨識結果為B。而在實際的應用上，可對應各個資料庫的成員設置個別的辨識器，各個辨識器專門用來對特定的成員進行辨識，並產生相應的辨識分數。
- 【0060】 請參閱第3圖，其係為本發明之融合人臉辨識及語音辨識之身份辨識系統之第一實施例之流程圖。本流程圖舉例說明了本發明之融合辨識方法。
- 【0061】 在步驟S31中，身份辨識系統執行人臉辨識及語音辨識以獲得人臉辨識結果、人臉辨識信心指數、語音辨識結果、及語音辨識信心指數，並進入步驟S32。
- 【0062】 在步驟S32中，身份辨識系統判斷人臉辨識信心指數及語音辨識信心指數是否超過預設門檻值？若僅有一方之信心指數超過預設門檻值，則進入步驟S321；若雙方之信心指數均超過預設門檻值，則進入步驟S322；若雙方之信心指數均未超過預設門檻值，則進入步驟S323。
- 【0063】 在步驟S321中，身份辨識系統選擇信心指數超過預設門檻值的辨識結果作為身份辨識結果。
- 【0064】 在步驟S322中，身份辨識系統整合雙方之辨識結果以執行融合計算，並進入步驟S3221。
- 【0065】 在步驟S3221中，身份辨識系統輸出融合計算結果。
- 【0066】 在步驟S323中，身份辨識系統判定無法辨識。
- 【0067】 請參閱第4圖，其係為本發明之融合人臉辨識及語音辨識之身份

辨識系統之融合辨識實驗結果之統計圖。在本實施例中，採用9個人並各取20張人臉影像以及超過20秒的聲音資料做為訓練資料庫。並採用9人各取10張人臉影像以及10筆超過1秒的聲音做為測試資料庫。測試用的人臉圖像經過各種亮度的調整，測試用聲音資料庫則是混入不同大小的噪音，以設計出15種不同情境下的實驗。如圖所示，本發明之融合辨識的辨識率能夠有效地提升。

【0068】 請參閱第5圖，其係為本發明之融合人臉辨識及語音辨識之身份辨識系統之語音辨信心指數算法之示意圖。如同前述，本發明之語音辨識信心指數的計算方式可經由事前訓練，利用大量實驗的結果得到辨識正確分數的累積分佈函數 $F_R(x)$ ，以及辨識錯誤分數的累積分佈函數 $F_E(x)$ 。信心指數曲線251可以利用下列的關係式來表示：

$$y = F_R(x) - (1 - F_E(x)) ;$$

【0069】 與上述函數對應的圖表則如第5圖所示。在一實施例中，主要可以利用三種估計方法來決定分數的取得，包含通用背景模型（UBM）正規化、最大值正規化以及幾何平均正規化，其中UBM即是考量到應用環境所建立的模型，可有效判斷噪音所在，進而決定分數的高低。語音辨識信心指數也是在分析辨識分數間的差距，當第一高分與其餘分數差距愈大，則其可信度愈高。

【0070】 儘管前述在說明本發明之融合人臉辨識及語音辨識之身份辨識系統的過程中，亦已同時說明本發明之融合人臉辨識及語音辨識之身份辨識方法的概念，但為求清楚起見，以下仍另繪示流程圖詳細說明。

- 【0071】 請參閱第6圖，係為本發明之融合人臉辨識及語音辨識之身份辨識方法之流程圖。
- 【0072】 在步驟S61中，利用人臉辨識模組擷取影像，並對該影像執行人臉偵測以產生人臉影像，再對人臉影像執行人臉辨識，以獲得對應於預設之複數個成員的複數個人臉辨識分數。
- 【0073】 在步驟S62中，藉由語音辨識模組擷取聲音資料，並偵測聲音資料中的語音，以產生語音資料，再對語音資料執行語音辨識，以獲得對應於複數個成員的複數個語音辨識分數。
- 【0074】 在步驟S63中，經由信心指數計算模組計算人臉辨識的可靠度及語音辨識的可靠度，以產生人臉辨識信心指數及語音辨識信心指數。
- 【0075】 在步驟S64中，透過處理模組根據複數個人臉辨識分數、人臉辨識信心指數、複數個語音辨識分數及語音辨識信心指數產生身份辨識結果。
- 【0076】 本發明之融合人臉辨識及語音辨識之身份辨識方法的詳細說明以及實施方式已於前面敘述本發明之融合人臉辨識及語音辨識之身份辨識系統時描述過，在此為了簡略說明便不再重覆敘述。
- 【0077】 綜上所述，本發明可利用即時的資訊來做身份辨識，因此具有即時性。本發明整合了人臉辨識以及語音辨識，因此能夠有效地提高其辨識的正確率。本發明能對人臉辨識以及語音辨識做可信度的評估，因此即使在複雜的環境之下也能做出正確的辨識，可因應服務型機器人工作於變化之環境，或因其本身之移動造成辨識條件變化(如距離改變、背光、陰影等等狀況)之情況下，仍可達

到辨識的效果，提升辨識準確性與穩定度。可見本發明在突破先前之技術下，確實已達到所欲增進之功效，且也非熟悉該項技藝者所易於思及，其所具之進步性、實用性，顯已符合專利之申請要件，爰依法提出專利申請，懇請貴局核准本件發明專利申請案，以勵創作，至感德便。

【0078】 以上所述僅為舉例性，而非為限制性者。任何未脫離本發明之精神與範疇，而對其進行之等效修改或變更，均應包含於後附之申請專利範圍中。

【符號說明】

【0079】 1、2：身份辨識系統
13：語音辨識模組
131、281：語音偵測
1311、2811：語音資料
132、232：語音辨識
1321、2321：語音辨識分數
14：人臉辨識模組
141、291：人臉偵測
1411、2911：人臉影像
142、242：人臉辨識
1421、2421：人臉辨識分數
15：信心指數計算模組
151、252：語音辨識信心指數
152、272：人臉辨識信心指數
16：處理模組

- 161、261：身份辨識結果
- 21：麥克風
- 22：攝影機
- 23：語音辨識器
- 24：人臉辨識器
- 25：語音辨識信心指數計算器
- 251：信心指數曲線
- 26：處理器
- 27：人臉辨識信心指數計算器
- 271：亮度因子
- 273：辨識分數差異性因子
- 28：聲音變化偵測器
- 29：人臉偵測器
- S31~S32、S321~S323、S3221、S61~S64：步驟

【主張利用生物材料】

無

【發明申請專利範圍】

- 【第1項】 一種融合人臉辨識及語音辨識之身份辨識系統，係包含：
- 一人臉辨識模組，係擷取一影像，並對該影像執行一人臉偵測以產生一人臉影像，再對該人臉影像執行一人臉辨識，以獲得對應於預設之複數個成員的複數個人臉辨識分數；
 - 一語音辨識模組，係擷取一聲音資料，並偵測該聲音資料中的語音，以產生一語音資料，再對該語音資料執行一語音辨識，以獲得對應於該複數個成員的複數個語音辨識分數；
 - 一信心指數計算模組，係計算該人臉辨識的可靠度及該語音辨識的可靠度，以產生一人臉辨識信心指數及一語音辨識信心指數；
- 以及
- 一處理模組，係根據該複數個人臉辨識分數、該人臉辨識信心指數、該複數個語音辨識分數及該語音辨識信心指數產生一身份辨識結果；
- 其中，該信心指數計算模組係根據一亮度因子及一辨識分數差異性因子來計算該人臉辨識信心指數。
- 【第2項】 如申請專利範圍第1項所述之融合人臉辨識及語音辨識之身份辨識系統，其中當該人臉辨識信心指數超過一預設門檻值時而該語音辨識信心指數未超過該預設門檻值時，該處理模組則根據該複數個人臉辨識分數產生該身份辨識結果。
- 【第3項】 如申請專利範圍第2項所述之融合人臉辨識及語音辨識之身份辨識系統，其中當該語音辨識信心指數超過該預設門檻值時而該人

臉辨識信心指數未超過該預設門檻值時，該處理模組則根據該複數個語音辨識分數產生該身份辨識結果。

【第4項】 如申請專利範圍第3項所述之融合人臉辨識及語音辨識之身份辨識系統，當該人臉辨識信心指數及該語音辨識信心指數均未超過該預設門檻值時，該處理模組則判定該身份辨識結果為無法識別。

【第5項】 如申請專利範圍第4項所述之融合人臉辨識及語音辨識之身份辨識系統，其中當該人臉辨識信心指數及該語音辨識信心指數均超過該預設門檻值時，該處理模組則根據該複數個人臉辨識分數、該人臉辨識信心指數、該複數個語音辨識分數及該語音辨識信心指數進行一融合計算，以產生該身份辨識結果。

【第6項】 如申請專利範圍第5項所述之融合人臉辨識及語音辨識之身份辨識系統，其中該融合計算係由該處理模組將各個該複數個成員的該人臉辨識分數乘上該人臉辨識信心指數以產生一最終人臉辨識分數，並將各個該成員的該語音辨識分數乘上該語音辨識信心指數以產生一最終語音辨識分數，再將對應的該最終人臉辨識分數及該最終語音辨識分數相加以產生對應於各個該成員的一身份辨識分數，並根據該身份辨識分數產生該身份辨識結果。

【第7項】 如申請專利範圍第1項所述之融合人臉辨識及語音辨識之身份辨識系統，其中該亮度因子係符合下列關係式：

$$CI = \begin{cases} (G^{mz})^2 \cdot G^{mz} < T^{low} \\ (1-G^{mz})^2 \cdot G^{mz} > T^{high} \\ 1 \cdot T^{low} \leq G^{mz} \leq T^{high} \end{cases}$$

；

其中，CI為該亮度因子， T^{high} 為該影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值上限， T^{low} 為該影像灰階值經正規化

至0~1範圍後對應於正常亮度之門檻值下限， G^{avg} 為該影像之灰階值經正規化至0~1範圍後的平均值。

【第8項】 如申請專利範圍第7項所述之融合人臉辨識及語音辨識之身份辨識系統，其中該辨識分數差異性因子係符合下列關係式：

$$C2 = \begin{cases} (Y^{1st} - Y^{2nd}) / T, & Y^{1st} - Y^{2nd} < T \\ 1, & Y^{1st} - Y^{2nd} \geq T \end{cases} ;$$

其中，C2為該辨識分數差異性因子，T為該複數個人臉辨識分數中最高分數與最低分數的一差距門檻值， Y^{1st} 為該複數個人臉辨識分數中最高分數， Y^{2nd} 為該複數個人臉辨識分數中次高分數。

【第9項】 如申請專利範圍第8項所述之融合人臉辨識及語音辨識之身份辨識系統，其中該人臉辨識信心指數係符合下列關係式：

$$C = C1 \times C2 ;$$

其中，C為該人臉辨識信心指數。

【第10項】 如申請專利範圍第9項所述之融合人臉辨識及語音辨識之身份辨識系統，其中該信心指數計算模組係根據一信心指數曲線來計算該語音辨識信心指數，該信心指數曲線係符合下列關係式：

$$y = F_R(x) - (1 - F_E(x)) ;$$

其中， $F_R(x)$ 係為經由事前訓練所獲得的辨識正確分數的累積分佈函數， $F_E(x)$ 係為經由事前訓練所獲得的辨識錯誤分數的累積分佈函數，該信心指數曲線之x軸代表經過正規化至0~1後之該語音辨識分數，該信心指數曲線之y軸則代表經過正規化至0~1後之該語音辨識信心指數。

【第11項】 一種融合人臉辨識及語音辨識之身份辨識方法，係包含下列步驟：

利用一人臉辨識模組擷取一影像，並對該影像執行一人臉偵測以

產生一人臉影像，再對該人臉影像執行一人臉辨識，以獲得對應於預設之複數個成員的複數個人臉辨識分數；

藉由一語音辨識模組擷取一聲音資料，並偵測該聲音資料中的語音，以產生一語音資料，再對該語音資料執行一語音辨識，以獲得對應於該複數個成員的複數個語音辨識分數；

經由一信心指數計算模組計算該人臉辨識的可靠度及該語音辨識的可靠度，以產生一人臉辨識信心指數及一語音辨識信心指數；

透過一處理模組根據該複數個人臉辨識分數、該人臉辨識信心指數、該複數個語音辨識分數及該語音辨識信心指數產生一身份辨識結果；以及

利用該信心指數計算模組根據一亮度因子及一辨識分數差異性因子來計算該人臉辨識信心指數。

【第12項】 如申請專利範圍第11項所述之融合人臉辨識及語音辨識之身份辨識方法，更包含下列步驟：

利用該處理模組在當該人臉辨識信心指數超過一預設門檻值時而該語音辨識信心指數未超過該預設門檻值時，根據該複數個人臉辨識分數產生該身份辨識結果。

【第13項】 如申請專利範圍第12項所述之融合人臉辨識及語音辨識之身份辨識方法，更包含下列步驟：

利用該處理模組在當該語音辨識信心指數超過該預設門檻值時而該人臉辨識信心指數未超過該預設門檻值時，根據該複數個語音辨識分數產生該身份辨識結果。

【第14項】 如申請專利範圍第13項所述之融合人臉辨識及語音辨識之身份辨識方法，更包含下列步驟：

經由該處理模組在當該人臉辨識信心指數及該語音辨識信心指數

均未超過該預設門檻值時，判定該身份辨識結果為無法識別。

【第15項】 如申請專利範圍第14項所述之融合人臉辨識及語音辨識之身份辨識方法，更包含下列步驟：

由該處理模組在當該人臉辨識信心指數及該語音辨識信心指數均超過該預設門檻值時，根據該複數個人臉辨識分數、該人臉辨識信心指數、該複數個語音辨識分數及該語音辨識信心指數進行一融合計算，以產生該身份辨識結果。

【第16項】 如申請專利範圍第15項所述之融合人臉辨識及語音辨識之身份辨識方法，更包含下列步驟：

透過該處理模組將各個該複數個成員的該人臉辨識分數乘上該人臉辨識信心指數以產生一最終人臉辨識分數，並將各個該成員的該語音辨識分數乘上該語音辨識信心指數以產生一最終語音辨識分數，再將對應的該最終人臉辨識分數及該最終語音辨識分數相加以進行一融合計算，以產生對應於各個該成員的一身份辨識分數，並根據該身份辨識分數產生該身份辨識結果。

【第17項】 如申請專利範圍第11項所述之融合人臉辨識及語音辨識之身份辨識方法，其中該亮度因子係符合下列關係式：

$$C1 = \begin{cases} (G^{avg})^2 \cdot G^{avg} < T^{low} & ; \\ (1 - G^{avg})^2 \cdot G^{avg} > T^{high} \\ 1 \cdot T^{low} \leq G^{avg} \leq T^{high} \end{cases}$$

其中，C1為該亮度因子， T^{high} 為該影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值上限， T^{low} 為該影像之灰階值經正規化至0~1範圍後之正常亮度之門檻值下限， G^{avg} 為該影像之灰階值經正規化至0~1範圍後的平均值。

【第18項】 如申請專利範圍第17項所述之融合人臉辨識及語音辨識之身份辨

識方法，其中該辨識分數差異性因子係符合下列關係式：

$$C2 = \begin{cases} (Y^{1st} - Y^{2nd}) / T, Y^{1st} - Y^{2nd} < T & ; \\ 1 \cdot Y^{1st} - Y^{2nd} \geq T & \end{cases}$$

其中，C2為該辨識分數差異性因子，T為該複數個人臉辨識分數中最高分數與最低分數的一差距門檻值，Y^{1st}為該複數個人臉辨識分數中最高分數，Y^{2nd}為該複數個人臉辨識分數中次高分數。

【第19項】 如申請專利範圍第18項所述之融合人臉辨識及語音辨識之身份辨識方法，其中該人臉辨識信心指數係符合下列關係式：

$$C = C1 \times C2 & ;$$

其中，C為該人臉辨識信心指數。

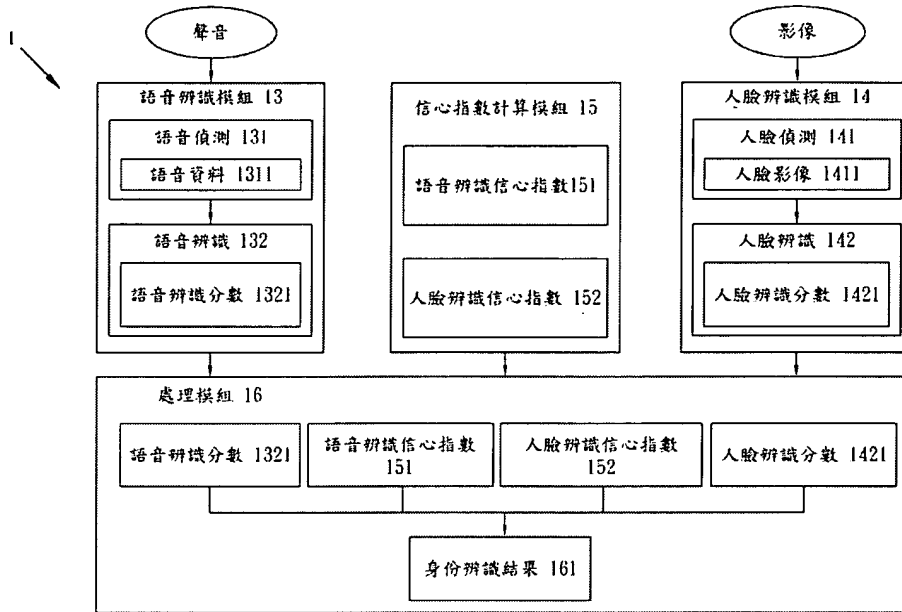
【第20項】 如申請專利範圍第19項所述之融合人臉辨識及語音辨識之身份辨識方法，其中該信心指數計算模組係根據一信心指數曲線來計算該語音辨識信心指數，該信心指數曲線係符合下列關係式：

$$y = F_R(x) - (1 - F_E(x)) & ;$$

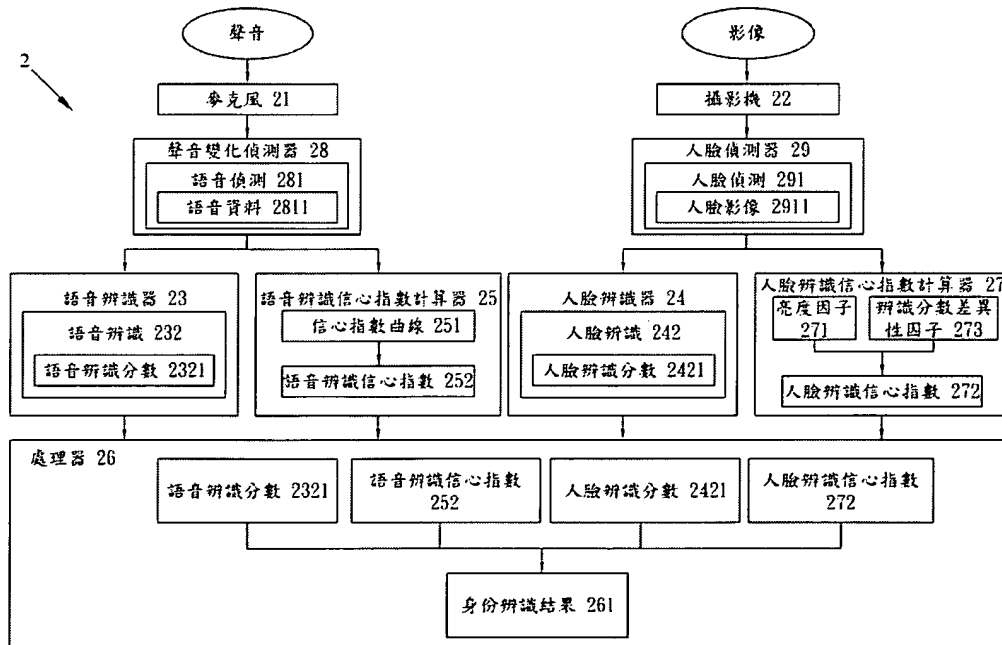
其中，F_R(x)係為經由事前訓練所獲得的辨識正確分數的累積分佈函數，F_E(x)係為經由事前訓練所獲得的辨識錯誤分數的累積分佈函數，該信心指數曲線之x軸代表經過正規化至0~1後之該語音辨識分數，該信心指數曲線之y軸則代表經過正規化至0~1後之該語音辨識信心指數。

【第21項】 一種服務型機器人，係包含一電源供應器及一融合人臉辨識及語音辨識之身份辨識系統，該服務型機器人係利用該融合人臉辨識及語音辨識之身份辨識系統辨識一使用者之身份，以決定該使用者之使用權限，該融合人臉辨識及語音辨識之身份辨識系統係如申請專利範圍第1項至第10項中之任一項之所述。

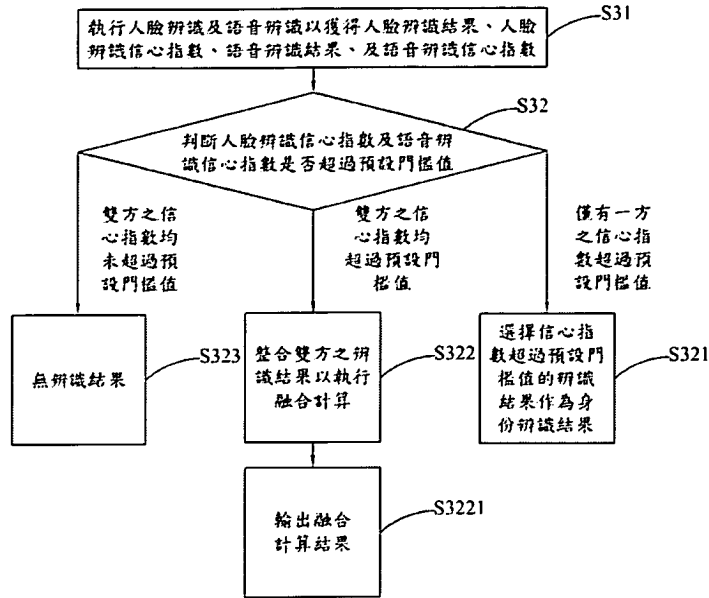
【發明圖式】



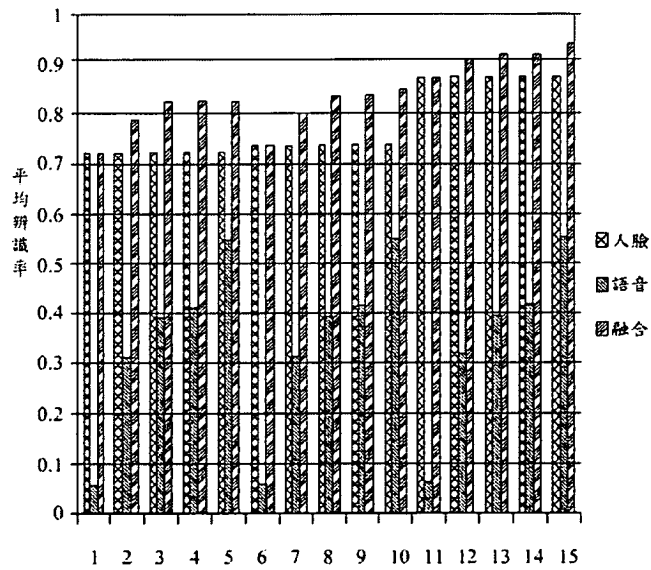
第1圖



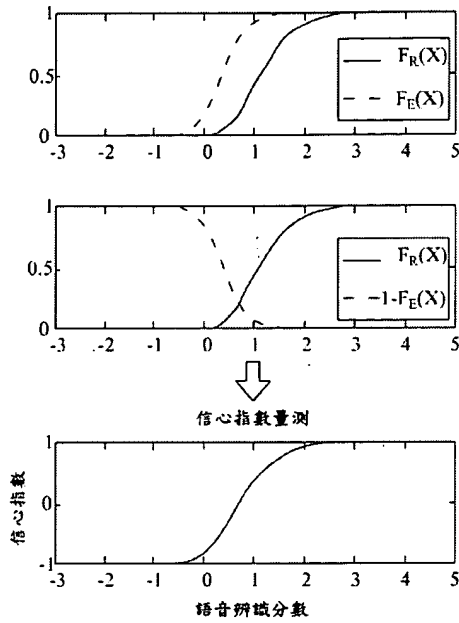
第2圖



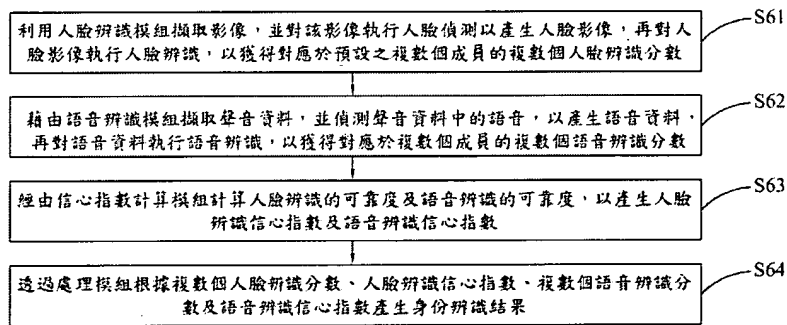
第3圖



第4圖



第5圖



第6圖