強健語者辨識技術研究：一個基於模糊系統的方法

Ing-Jr Ding
Department of Electrical Engineering, National Formosa University
e-mail： ingjr@nfu.edu.tw

摘要

語者辨識乃是音訊(語音)處理領域中極為重要的一項技術，該技術可運用於居家監控或安全防護等一些應用情境中，然而，語者辨識之辨識性能仍無法被使用者所欣然接受，因此，如何積極提昇語者辨識的辨識率是一個極具挑戰性並且值得深入研究的技術課題。模糊理論相關技術已廣泛地被運用於語音辨識領域並得到極佳的成效，然模糊系統於該領域的研究甚少，本論文之研究為基礎於模糊系統之強健語者辨認方法設計，期望能將模糊系統之優異的參數調控機制置入目前主流的語者辨識技術中，增加傳統辨識方法的強健性，以能有效地掌控一個語者辨識系統中的不定參數，並進而使系統在處於極為不利的辨識情況下仍能達到優異的辨識效果。

關鍵詞：語者辨識、高斯混合模型、支持向量機、模糊系統

Abstract

With the ever growing maturity, speaker recognition has found more and more applications in current daily life. Nevertheless, the recognition performance of all speaker recognition systems ever built is undeniably inferior to a human listener. Therefore, how to effectively improve the recognition performance of speaker recognition has been a challenging issue. In this study, the exploitation of fuzzy system (FS) mechanism in the field of speaker recognition will be thoroughly reinvestigated, specifically in the reliable determination of some specific parameters in a speaker recognition system for enhancing system design and performing a satisfactory recognition rate.

Keywords: Speaker recognition; Gaussian mixture model; Support vector machine; Fuzzy system

1. 語者辨識導論

時值資訊科技突飛猛進的今日，在未來的資訊化社會中，人機界面的方式將愈益多樣化且成熟，如語音辨識技術已普及應用於日常生活之中，在手持式嵌入式設備如 PDA、 Smart Phone 或甚至更低階的手機都可見到語音辨識應用植入其中[1]。由於語音之特徵乃為人類身上最為自然的生物特徵之一，並且具備容易產生及擷取等特性，因此除了語音辨識之人機界面應用之外，另一個極具殺手級之應用「語者辨識」，勢必也很可能成為未來主流之自動化辨識服務技術。

語者辨識(Speaker recognition)技術乃是語音處理領域中極為重要之一環，語者辨識的目的是希望能藉由機器以智慧化方式辨認說話者的聲音。說話者之語音訊號輸入至語者辨識系統，該系統即進行該說話者之身份的識別或確認等自動化服務。目前的安全監控(Surveillance)系統應用或門禁系統(Security)應用，仍多以視訊(Video)訊號之分析辨認為主，例如在機場、捷運、或社區等場所中多安裝 IP-Camera 或是較昂貴的錄影設備以擷取視訊的資料而再做後續之身份的判斷。然而，若考慮到隱私權(privacy)之心理因素，則利用視訊模式之身份辨認的方式似乎較為不妥，基於此，利用音訊模式進行語者之聲紋的辨別顯然是一個更恰當的方式語者辨識可提供除了人臉辨識與指紋辨識等以影像為基礎的辨識技術之外的另一種身份判斷的技術。再者，隨著多模式(multi-modality)之模樣判別技術日益受到重視，人體身上的一些重要特徵如人臉、指紋、虹膜等皆可用來做為進行身份的判別，而該類多模式辨識系統多以擷取影像或視訊等生物特徵做為系統在判別上的依據，而若能在該類辨識系統中再加入語音訊號資料以做為在身份確認或身份辨認上的一個輔助工具，則必然能增加辨認系統在判斷上的準確性。因此，語者辨識技術之大量普及化於未來的生活應用似乎已經是一個必然的趨勢。舉凡目前各式的辨識應用軟體，如時下流行之手寫辨識及語音辨識等產品，人們最關心的議題是該類系

統的辨識率，唯有接近人為判斷之辨識準確度才能真正提供生活的高度便利性並最終能為人們所接受。語者辨識技術亦然，辨識性能對於一個語者辨識系統而言是最重要的一項系統性能指標，然而，目前的語者辨識技術的辨識率仍無法達到一個令人滿意的目標，此技術仍然有極大可以改善的空間。基於此，如何發展辨識性能優異的語者辨識系統確實刻不容緩並一直是該學術領域之研究學者所積極投入研發的方向[2]。本研究論文之研究工作的重點在於以模糊系統為基礎，改善語者辨識機制的設計或強化語者辨識演算法，所發展之語者辨識方法為一個具強健式之語者辨識方法，此方法能有效提昇語者辨識系統的辨識性能，並使得辨識系統即使在處於極為不利的辨識情況下亦能達到優異的語者辨識的效果。模糊系統一般較常運用於動態物理系統的設計上，如倒單擺裝置、馬達控制系統、溫度控制系統等[3, 4]，模糊系統實已證實是將複雜動態系統模型化並建立控制器之良好選擇。若能將此模糊系統之優異的參數調控能力亦賦予語者辨識程序，則此語者辨識系統將更具環境適應的特性而能隨時都保持一定的辨識水準。如何將控制領域之模糊系統控制之運用於動態物理系統的設計的概念植入於屬於資通訊相關學門技術領域的語者辨識技術，實為一項極具挑戰且值得深入研究的課題。

語者辨識依應用情境之不同可再細分為語者識別(Speaker identification)和語者確認(Speaker verification)等兩類應用技術。語者識別為使用者之聲音語料輸入至辨識系統，系統會在所有已經註冊之語者中尋找一位相似度最高之語者以做為該測試語者之識別身份，此類應用問題運用於“Who is him in the system？”之應用情境；語者確認則是辨識系統待測試語者輸入其聲音語料至系統後，系統會判別該測試語者是否即為該語者所宣稱之身份，亦即系統會判別該語者是否屬於已經註冊於系統之語者群體中之一員，和語者識別應用不同的是語者確認應用將僅會實施接受或拒絕語者等兩動作之一，而此類語者確認應用問題運用於“Is him in the system?”之應用情境。

本研究論文擬深入研究如何整合模糊系統之優異的參數調控能力於語者辨認技術領域範疇的語者辨識機制及語者辨識演算法。眾所皆知，語者辨識技術之辨識性能是該技術未來能否全面導入商品化的重要指標，而如何有效提昇語者辨識系統的辨識率至使用者可以接受的程度，一直是該領域之研究學者所積極努力追求的目標。本論文之研究工作為在語者辨識技術領域導入在控制領域中具有傑出表現的模糊系統機制，期許能強化傳統的語者辨識機制或語者辨識演算法的設計，使得辨識系統的辨識性能得以大步改善。

2. 主流之語者辨識技術簡介

目前，語者辨識技術主要可分為三個主流的技術類別，分別是以高斯混合模型(Gaussian Mixture Model, GMM)為基礎的語者辨識技術[5, 6]、以支持向量機(Support Vector Machine, SVM)為基礎的語者辨識技術[7-9]及結合高斯混合模型與支持向量機(Hybrid GMM-SVM)之雙模型的語者辨識技術[10, 11]。高斯混合模型為基礎的語者辨識技術一般常為語者識別應用問題所採用之技術，而語者確認應用問題一般則多採行以支持向量機為基礎的語者辨識技術，結合高斯混合模型與支持向量機之的語者辨識技術，由於該類技術同時兼備高斯混合模型與支持向量機模型等雙模型分類的特性，其一般被運用於須同時實施語者識別及語者確認等兩動作之應用系統。茲就此三個主流類別之語者辨識技術概述如下。

2.1 高斯混合模型為基礎的語者辨識技術

高斯混和模型為一多分類的模型[12]，如圖 1 所示，利用資料庫相對應的聲音資料以訓練某 GMM 模型直至其收斂為止，若系統中有 n 個註冊語者，則須為每個語者皆建立一個 GMM 聲學模型，因此語者識別系統共計須訓練建立 n 個 GMM 模型。在 GMM 模型的設定方面，如高斯混合數的選擇亦必須仔細考量，此部份須配合聲音資料庫的資料量多寡、聲訊訊號的特徵參數特性、及實際的測試等等一併做考慮。而一個完整的 GMM 模型具有三種參數：混和加權值(mixture weight)、平均值向量(mean vector)以及共變異矩陣(covariance matrix)。如下所示：

$$\lambda = \{w_i, \vec{\mu}_i, \Sigma_i\}, \quad i = 1, \cdots, M \tag{1}$$

其中，$w_i, \vec{\mu}_i, \Sigma_i$ 三種參數分別表示混和數加權值、平均值向量、及共變異矩陣，而 M 則是高斯分佈的個數，對於系統中每一語者的聲音皆用一個 GMM 模型來代表它。而對一個 D 維的特徵向量測試音框，其對於某語者之高斯混和模型的相似度分數計算如下所示：

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} w_i b_i(\vec{x}) \tag{2}$$

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\} \quad (3)$$

$$\sum_{i=1}^{M} w_i = 1 \quad (4)$$

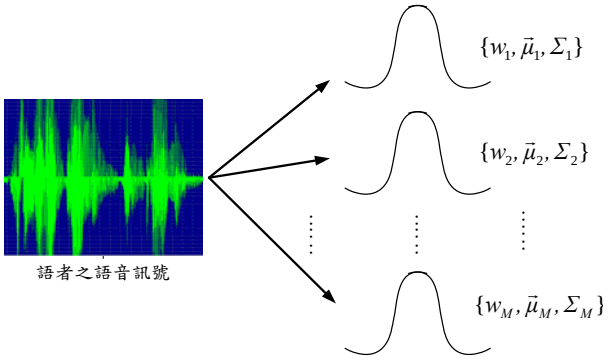

圖 1 運用 GMM 高斯分佈之語者辨識

2.2 支持向量機為基礎的語者辨識技術

SVM 是一種二分類器，由於每一個 SVM 只能分辨兩類資料，因此極適合應用於語者確認之應用情境[13]。假設給定一群訓練語料$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，$y_i$ 為物件 x_i 的類別標籤，y_i 的值為 1 或-1，SVM 的目標在於訓練出一個函數，此函數亦是一個分隔面(Separating hyperplane)：

$$f(x) = w \cdot x + b \quad (5)$$

並使得下列的式子成立：

$$\begin{cases} w \cdot x_i + b \geq 1, & if\ y_i = 1 \\ w \cdot x_i + b \leq -1, & if\ y_i = -1 \end{cases} \quad (6)$$

公式(6)的意義在於，對於所有的訓練資料都能被 $f(x)$分成兩類，其中類別標籤為 1 的一邊，類別標籤為-1 的一邊。明顯地，此時 $f(x)$為一個二分類器，同時也為這兩類訓練資料的分隔平面。但是可以分隔兩類訓練資料的分隔平面有許多條，那要怎麼選擇呢？SVM 的主要概念是在要分類的兩群訓練資料中，求出能產生*最大界線(maximum margin)*的分隔平面。最大界線的概念即是假設所要求的分隔平面和右半邊點的最近距離為 d^+，而和左半邊點的最近距離為 d^-，則最大界線分隔平面為使 $d^+ + d^-$ 成為最大值者。而選取最大界線分隔線的好處就是能儘量減少分類錯誤的結果。

以支持向量機模型為基礎的語者辨識技術即是在 SVM 語者聲學模型訓練建立後，測試語者即可說出語料並對所訓練的 SVM 分隔面進行分邊之動作以決定此語者在已註冊語者的一邊(系統接受)或非經認證註冊語者之一邊(系統拒絕)，如圖 2 所示。

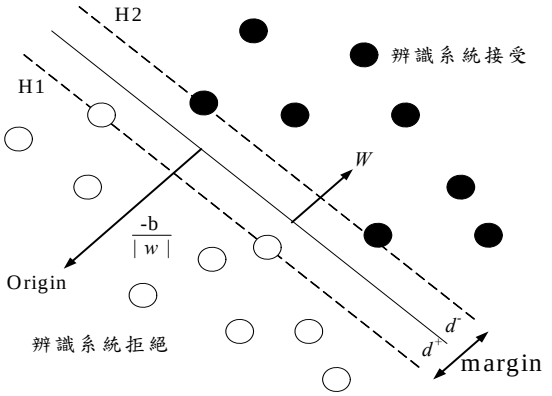

圖 2 運用 SVM 分隔面之語者辨識

2.3 結合高斯混合模型與支持向量機之雙模型語者辨識技術

結合高斯混合模型與支持向量機之語者辨識技術同時兼備高斯混合模型與支持向量機模型等雙模型之分類的特性，該方式一般被運用於須同時實施語者識別與語者確認等兩動作之應用系統，如圖 3 所示。在此混合模型的方式下，語者辨識的程序一般為可先行計算測試語者之語音樣本與辨識系統中之各語者的高斯混合模型的相似度分數值，將數個相似度分數值較高者挑出，若最高相似度分數值的分數明顯大於其它者之分數值，表示該語者極可能為最高相似度分數之 GMM 模型所描述的語者，假設為語者 n，則後續再將代表語者 n 之 SVM 模型挑出以對此測試語者進行語者確認的動作，以決定是否接受此身份的語者或拒絕之。然而，若最高相似度分數值的分數與其餘排名較高的相似度分數值的分數相距甚少，表示該語者尚無法肯定判定是系統中的那一位語者，因此，亦無法再使用 SVM 模型技術進行特定語者的確認，則僅能依據 GMM 相似度分數值的分數資訊進行接受語者或拒絕語者的決策。

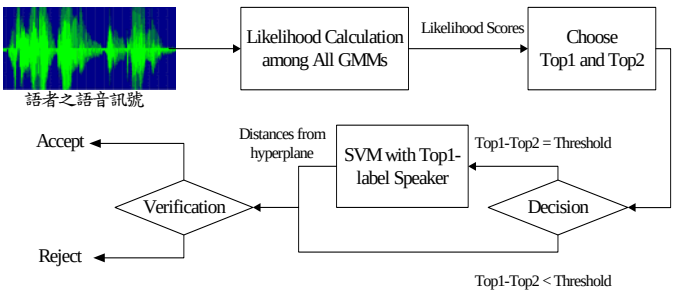

圖 3 運用 SVM-GMM 混合模型之語者辨識

3. 基於模糊系統之強健性語者辨識技術

三類主流之語者辨識技術已如上述，然而，在實際的辨識環境中，經常會出現對於辨識性能較為不利的因素，例如太多背景雜訊聲音的干擾、測試語者的測試語料的數量太少或測試語料的品質太差等，而傳統之以高斯混合模型為基礎的語者辨識技術及以支持向量機為基礎的語者辨識技術因不具備環境適應的能力及對於錯誤容忍的程度太低因而會導致辨識系統的辨識性能無法永遠維持一定程度表現。而對於結合高斯混合模型與支持向量機等兩者模型之語者辨識技術而言，雖然該技術擷取兩類模型辨識技術的優點，但是亦不具環境適應與系統容錯的能力。眾所皆知，模糊系統的特色乃在於強健原系統的設計，使該系統隨時皆能依據系統所處環境之變化，適度地調整校正系統的內部參數，而使得系統皆能維持一定程度的辨識水準。本研究論文之目的乃在設計優異的模糊系統機制於上述主流的語者辨識技術中，希望能將模糊系統之優異的參數調控機制分別置入傳統的語者辨識技術，藉以改善語者辨識機制的設計或強化語者辨識演算法，使得辨識系統即使在處於極為不利的辨識情況下仍能達到優異的語者辨識的效果。再者，此導入模糊系統機制的強健性語者辨識方法又不會增加太多的額外計算量而亦能維持即時運算的能力。

3.1 主流之語者辨識技術與模糊系統機制

目前，語者辨識技術主要可分為三個主流的類別，分別是以高斯混合模型為基礎的語者辨識技術、以支持向量機為基礎的語者辨識技術及結合高斯混合模型與支持向量機等多模型的語者辨識技術。三個主流類別語者辨識技術之概要觀念已如前述，本論文之初步的研究工作是深入且細步地熟稔這些傳統的語者辨識技術，分析各語者辨識技術之辨識機制或辨識演算法的優缺點為何，如何將模糊系統導入某特定語者辨認方法，以改善該方法的缺點並思索亦能將原先方法的優點予以保留。模糊系統導入語者辨識技術的流程可如圖 4 所示，圖 4 所示是一個概括性的階段流程圖。

在階段圖中的第一步即是先選定某特定語者辨識演算法或語者辨識機制，在選定演算法後，即進行此演算法之整個運作的分析，例如在該演算法中之何者特定參數會影響整體辨系統之辨識性能的輸出表現，在此步驟完成影響辨識率之演算方法的特定參數分析之後，即進行第三步之模糊系統調控特定參數，在此步驟中，即是設計一個模糊系統機制以進行演算法之特定參數的調控，在將模糊系統機制整合至該演算法則後，最後一步是模糊系統機制自我調校，此步驟的目的在於改善所設計的模糊系統，使得模糊系統之參數調控性能表現能達到最佳狀況，如此即完成了一個基礎於模糊系統之強健語者辨認方法的設計，並繼續選定下一個特定語者辨識演算法或語者辨識機制而開始另一個循環。

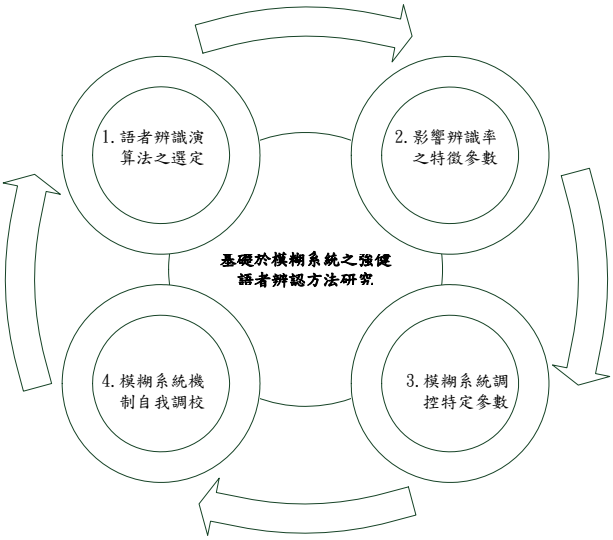

圖 4 模糊系統機制導入於特定語者辨識演算法之分析階段圖

3.2 導入模糊系統機制之以高斯混合模型(GMM)為基礎的語者辨識技術

傳統之 GMM 語者辨識技術在透過已訓練好之 GMM 模型進行語者識別時，運用一個已預先定義長度之決定視窗的 GMM 分類器以進行 GMM 語者模型的相似度判別。此預先定義之決定視窗的時間長度是固定式，亦即在每段固定時間(如每一秒、每兩秒或每五秒等)判定該測試語者所發出之語音片斷與何者 GMM 模型的相似度分數(Likelihood scores)最高，以識別該語者的身份。如同圖 5 所示，對於一個包含 n 個音框且每個音框之時間為 Δt 的決定視窗而言，語者辨識系統決定該語者的身份的決策時間為 $n \cdot \Delta t$ ，此一傳統固定決策時間的判斷機制並未考量測試語料之品質的優劣，亦未考量受測語者在進行語料測試時之背景雜訊聲音的吵雜程度，而齊頭式地將欲用來判別語者的語音音框皆加以考量進去。然而，當背

景雜訊聲音過於吵雜(該段測試聲音的訊雜比太低)，則若將決定視窗的預定長度設定過長，則將累積太多不準確之音框與各 GMM 模型的相似度分數值，勢必極容易造成辨識錯誤的現象；反之，當背景環境幾乎無任何背景聲音(該段測試聲音的訊雜比很高)，則若將決定視窗的預定長度設定太短，則將無法對準確之各音框與各 GMM 模型的相似度分數值全部進行累進加總，多數精準之音框與 GMM 模型間的相似度分數將被迫捨棄，則此現象亦將造成語者辨識系統之辨識率的下降。因此，一個可依情況而定之變動長度方式的決定視窗設計機制將較傳統長度方式的決定視窗更具彈性，而此一彈性的設計方式將期許能使得語者辨識系統之辨識率獲得有效提昇。本研究工作針對如何運用模糊系統之優異的參數調控能力以設計一個具變動長度式之決定視窗的強健 GMM 語者辨認系統進行深入研究，相信本工作之針對此課題的研究成果將對現行的 GMM 語者辨認技術開啟一個全新的研究面向。

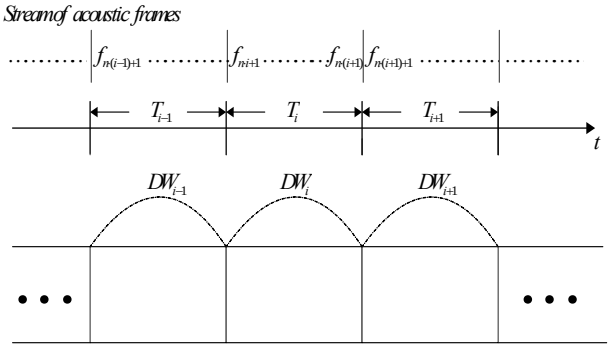

圖 5 傳統 GMM 語者辨識技術之運用固定時間長度的決定視窗(Decision Window, DW)進行模式判別圖

本論文針對此一研究主題的初步構思為運用環境背景吵雜程度做為模糊系統的輸入值，而所設計之模糊系統即會藉由此一輸入值進行判斷，模糊系統之解模糊化之輸出值即為 GMM 語者辨識系統之決定視窗的時間長度，然而，一般在背景環境中之語音訊號的訊雜比之值不容易獲得或已計算得到之訊雜比值的可信賴度低。本研究論文的初步想法為利用極短時距(Short timeslot)之 GMM 模型相似度的差值(Likelihood differences)之資訊以做為判斷決定視窗之給予適當值的依據，擬如圖 6 所示以進行模糊系統機制之導入的研究工作。

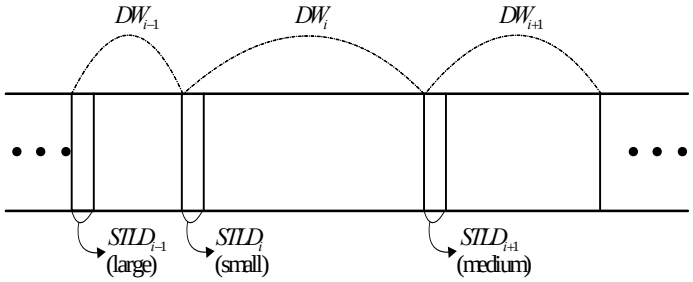

圖 6 擬設計一個具有變動時間長度之決定視窗的 GMM 語者辨識系統示意圖

短時距之 GMM 模型相似度的差異值(Short timeslot likelihood differences, *STLD*)之想法在本質上與訊雜比指標所訴諸之精神極為相仿，兩者皆用以判斷在不特定背景環境中語音訊號之品質的優劣，而 *STLD* 指標相對於訊雜比指標則具有容易取得且計算簡單之優勢。個人在過去的研究中，已成功地運用此 *STLD* 指標於特定音訊事件之偵測的監控應用情境中，在該研究中發現，在背景環境聲音較為乾淨且單純時，此融合 *STLD* 指標的音訊偵測系統可以將決定視窗之長度值調整為較小之值，因此，該系統即可具有即時回應(Real-time response)的特性；而當背景環境聲音較為吵雜且複雜時，此設計又可以彈性地將決定視窗之長度值糾正為較大之值，以累積更多之音框與 GMM 模型比較時之相似度值，此舉則有利於降低誤報率而能大幅提昇系統的辨識率。在該研究中，此用來判斷決定視窗之長度適當值的 *STLD* 指標的定義如下：

$$STLD = \left| \sum_{i=1}^{m} \log f(x_i \mid \lambda_1) - \sum_{i=1}^{m} \log f(x_i \mid \lambda_2) \right| \quad (7)$$

在公式(7)中，λ_1 與 λ_2 分別是特定音訊事件聲音的 GMM 聲學模型及背景環境聲音的 GMM 聲學模型，$f(x_i \mid \lambda_1)$ 與 $f(x_i \mid \lambda_2)$ 分別是第 i 個音框 x_i 對於模型 λ_1 與模型 λ_2 等高斯分佈的相似程度分數值。此 *STLD* 指標之值即為一個模糊控制器的輸入值，而經過模糊系統之運算後，即產生一個合適的決定視窗之值(Window Length, *WL*)，該模糊控制器的模糊規則庫之設計為：

規則 1: If *STLD* is small,
Then *WL* is big,
規則 2: If *STLD* is big,
Then *WL* is small.

然而，在語者辨識的情境中，此用來決定一個決定視窗的長度值之依據的 *STLD* 指標勢必要

做適當的修正或重新定義以因應不同的應用情境。在一般特定音訊事件偵測的情境中，GMM 聲學模型計有二者，一為特定聲音事件聲音的 GMM 模型，另一為背景環境聲音的 GMM 模型，而若在語者確認的新應用情境中，GMM 模型的數量可能不只為二，一般可行的做法是分為三類：安全身份語者的 GMM 模型、不安全身份語者的 GMM 模型與背景環境聲音的 GMM 模型；再者，又若在語者識別的另一新應用情境中，GMM 模型的數量可能又更多，倘若是一個可識別 *N* 個語者的語者識別系統，則 GMM 模型的數量將有 *N* 個，因此，在語者確認及語者識別的應用中，*STLD* 指標分別該如何做相對應的修正，及所擬定之運用類似 *STLD* 指標參數以設計一個具有變動時間長度之決定視窗的 GMM 語者辨識系統的模糊系統該如何設計，又模糊規則該如何建構、模糊規則數的數量該為何、模糊歸屬函數又應該如何定義、該選用 Sugeno 方式還是 Mamdani 方式以進行解模糊化動作及依據此想法所構建的模糊機制在實際語者辨識系統的運作成效如何等，實有待本論文之後緒研究進行。

3.3 整合模糊系統機制於以支持向量機為基礎的語者辨識技術

個人在過去的研究中已將模糊系統與支持向量機兩者模型合併使用於語音辨識領域中之貝氏語者調適類別的 MAP 調適演算法[14]，該設計方法之理念如圖 7 所示：

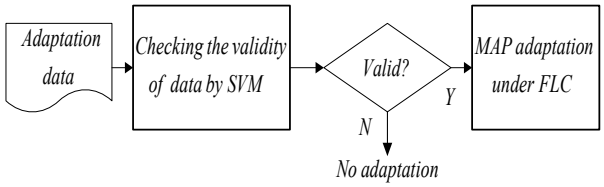

圖 7 混合 SVM 與 FLC 之 SVM-FLC 雙模機制圖

在圖 7 中，SVM 模型可做為一個新語者欲進行語者調適時之調適語料的一個確認機制，該設計之 SVM 模型可針對調適語料的品質進行分類，僅選取品質較佳之語料以做為 MAP 語者調適程序時之調適語料，而品質較差之語料，因恐甚而降低語者調適之性能，寧可捨棄此類語料而不做調適。在進行 MAP 語者調適程序時，設計了一個模糊控制器以額外加入了調適語料數量之考量因子，當調適語料數量不足時，此控制器會適當地將 MAP 演算法的調適速度減緩，以確保不調差語音模型，又當調適語料數量充足時，控制器則會加快 MAP 演算法的調適速度，以充份運用調適語料而能達到快速調適之功效。

本論文後緒於此課題的研究工作將以上述之 SVM-FLC 雙模機制於貝氏語者調演法研究做為基礎，再深入研究如何設計並整合一個合適的模糊系統機制於以支持向量機為基礎的語者辨識技術，本論文於此方向之研究初擬定為整合模糊系統於 SVM 語者辨識系統以進行語者模糊分類(Fuzzy SVM Classification by Fuzzy Systems) 的研究，本研究之不同於傳統之模糊分類(或稱軟性分類, Soft Clustering)方法在於本研究為運用模糊系統暨控制器機制以導入傳統之硬性分類(Hard Clustering)方式，而此方式之特色在於將模糊系統之優異的參數調控能力整合至樣本分類的應用問題中。

3.4 模糊系統於語者辨識程序之模型重估與細部元件改良研究

模糊系統一般運用於動態物理系統的設計上，如倒單擺裝置、馬達控制系統、溫度控制系統等，模糊系統實已證實是將複雜動態系統模型化並建立控制器之良好選擇。在模糊系統之設計中有兩個階段；首先是從已知的非線性系統方程式，由區段非線性或者局部近似的觀念來建構模糊模型。此種方法所有模糊子空間均具有幾乎相等地空間。此類架構稱作平衡結構。另一方法是直接使用系統輸入-輸出的訓練資料；傳統學習演算法是根據所建構模糊模型的誤差，而沒有考慮訓練資料的分佈情形。因此其結構在學習以後，尤其當訓練資料具有大誤差點時，將會成為不平衡結構。第二階段是模糊控制器設計。在模糊控制設計方法中，有使用平行分佈補償(parallel distributed compensation, PDC)的概念，其是對個別地模糊規則設計控制器，然後以線性矩陣不等式(linear matrix inequality, LMI)來檢查其穩定性。眾所周知，當模糊規則的數目很大時，要找到合適控制器可能變得很困難。此外，在大多數模糊控制分析方面，總是考慮系統模糊模型已知的情況，而並沒有探討當從不同建構模糊模型技巧得到的模型，用來直接設計控制器，其對系統性能的影響分析。

本研究論文欲模型化而建立模糊系統的目標是語者辨識機制或語者辨識演算法，此一目標為一個程序(process)，而非傳統之動態物理系統(physical system)的設計，因此，在系統

建模的階段具有相當大之研究發展的空間。本研究工作預計先利用眾多語者的大量語料做為模型學習時之訓練資料，設計一學習演算法以能有效地將語者辨識程序模型化，而若此方法所建立的模糊模型的運作成效不顯著，本研究預期將再嘗試尋找語者辨識程序的非線性系統方程式，利用局部線性系統的觀念將語者辨識程序建模。在此建模方式下，語者辨識程序將被分解(decomposition)為一個子系統的集合，亦即語者辨識程序是由一些子程序所形成的一個集合，因而看似複雜的整個語者辨識程序即可在一些較有規律之小程序上進行分析。這些小程序的每一個皆是一個線性的子系統，其表示整個語者辨識程序的某段區域性(local)之具規律的行為，而在此線性子系統上有其輸入資料與輸出資料(輸入資料經此線性子系統的分析後產生輸出資料)。一個模糊規則(或推論)即代表一個線性子系統，而整個語者辨識程序的輸出值即是每個線性子系統的輸出結果之組合，而組合的方式可以利用簡單的線性結合(linear combination)的方法。在此機制下，語者辨識程序之模糊度(fuzziness)即可用一種近似係數調控的方式加以考量。

在將語者辨識程序系統化建模之後，本研究將繼續再進行模糊控制器之設計與分析工作，在此階段，研究工作尤重於控制器之細部元件的分析與改良，以達成最佳化語者辨識系統之系統性能的目標。本論文擬將語者辨識程序建立模糊系統並進而對該系統設計控制器的研究工作如圖 8 所示：

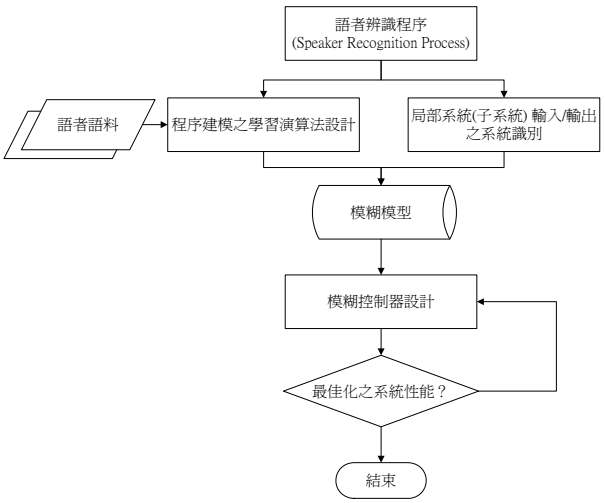

圖 8 語者辨識程序之最佳化模糊系統暨控制器的發展流程圖

如何將模糊邏輯控制之運用於動態物理系統的設計的概念植入於屬於資通訊相關學門技術領域的語者辨識技術，實為一項極具挑戰且值得深入研究的課題。屬於資通訊相關領域的技術若能導入自動控制學門領域之精湛模糊控制理論，相信不僅對於資通訊領域的技術而言能有一番新的面貌，而對於自動控制學門領域而言亦能有一個嶄新的研究面向。就資通訊領域的技術而言，運用模糊系統機制乃是更智慧化將一個程序/流程(procedure)施行自動化控制，而如何將領域專家的知識注入於模糊控制機制之中實則是控制器之設計的重點。就模糊控制在語者辨識程序的機制或演算法之參數調校上的設計而言，仍有多項技術上的議題需詳加考量與突破，其將關係到一個語者辨識系統的性能是否能一直維持令人滿意的狀況。後緒研究工作將就組成模糊控制器之各個元件分別進行效能的改善。

4. 結論

語者辨識技術雖說已是一項相當便利的人機介面操作模式，然其辨識系統的辨識準確率一直是一項具挑戰性的課題。未來若要讓此技術能全面普及化，則必不容許有太多的錯誤判斷(false alarm)。因此，如何發展出一個辨識性能優異且實用性高之語者辨識機制暨辨識演算法一直是語音處理相關領域之研究學者目前正積極努力的目標。本論文之研究工作乃為基礎於模糊系統之語者辨識機制暨語者辨識演算法的設計，該設計將模糊系統之優異的參數調控機制分別置入主流的語者辨識技術中，以能有效地掌控一個語者辨識系統中的不定參數，並進而使系統在處於極為不利的辨識情況下仍能達到優異的辨識效果。所設計的方法已具備優之辨識性能的特色，相信必能讓語者辨識技術在產品化的軌道上更向前邁進一大步。

參考文獻

[1] Rabiner, L. R., “The Power of Speech,” ***Science***, Vol. 301, pp. 1494-1495, 2003.

[2] Wang, J. -C., Yang, C. -H., Wang, J. -F. and Lee, H. -P., “Robust Speaker Identification and Verification,” ***IEEE Computational Intelligence Magazine***, Vol. 2, No. 2, pp. 52-59, 2007.

[3] Kermiche, S., Saidi, M. L., Abbassi, H. A. and Ghodbane, H., “Takagi-Sugeno Based

Controller for Mobile Robot Navigation," ***Journal of Applied Science***, Vol. 6, pp. 1838-1844, 2006.

[4] Zimmermann, H. -J., ***Fuzzy Set Theory and Its Applications***, 3rd ed., Kluwer Academic Publishers, 1996.

[5] Burget, L., Matejka, P., Schwarz, P., Glembek, O. and Cernocky, J., "Analysis of Feature Extraction and Channel Compensation in a GMM Speaker Recognition System," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 15, No. 7, pp. 1979-1986, 2007.

[6] Kenny, P., Boulianne, G., Ouellet, P. and Dumouchel, P., "Speaker and Session Variability in GMM-Based Speaker Verification," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 15, No. 4, pp. 1448-1460, 2007.

[7] McLaren, M., Vogt, R., Baker, B. and Sridharan, S., "Data-Driven Background Dataset Selection for SVM-Based Speaker Verification," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 18, No. 6, pp. 1496-1506, 2010.

[8] Ferras, M., Leung, C. -C., Barras, C. and Gauvain, J. -L., "Comparison of Speaker Adaptation Methods as Feature Extraction for SVM-Based Speaker Recognition," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 18, No. 6, pp. 1366-1378, 2010.

[9] Campbell, W. M., Campbell, J. P., Gleason, T. P., Reynolds, D. A. and Shen, W., "Speaker Verification Using Support Vector Machines and High-Level Features," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 15, No. 7, pp. 2085-2094, 2007.

[10] You, C. H., Lee, K. A. and Li, H., "GMM-SVM Kernel with A Bhattacharyya-Based Distance for Speaker Recognition," ***IEEE Transactions on Audio, Speech, and Language Processing***, Vol. 18, No. 6, pp. 1300-1312, 2010.

[11] You, C. H., Lee, K. A. and Li, H., "An SVM Kernel with GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition," ***IEEE Signal Processing Letters***, Vol. 16, No. 1, pp. 49-52, 2009.

[12] Rabiner, L. and Juang, B. -H., ***Fundamentals of Speech Recognition***, Prentice-Hall, USA, 1993.

[13] Burges, C. J. C., "A Tutorial on Support Vector Machines for Pattern Recognition," ***Data Mining and Knowledge Discovery***, Vol. 2, No. 2, pp. 121-167, 1998.

[14] Ding, I. –J., "A Hybrid SVM-FLC Approach to MAP Speaker Adaptation for Speech Recognition," ***ICIC Express Letters - An International Journal of Research and Surveys***, Vol. 5, No. 2, pp. 349-354, 2011.