類神經網路於生物測定認證技術及應用之研究(III)—子計劃四：國語語音及語者辨識系統(III)

Mandarin Speech and Speaker Recognition Systems

計畫編號：NSC88-2213-E-009-054
計畫期限：87/8/1 - 88/7/31

主持人:劉啟民　　交通大學資訊工程系副教授

一、中文摘要（關鍵詞：語者辨識，中文語音辨識）

語者辨識可分為語者驗證（speaker verification）與語者認定（speaker identification）。語者驗證是指查核語者的語音是否與宣稱語者相合。語者認定則是由一些候選語者中找出語者身份。語者辨識方法可分為特定語音（Text-Dependent）與非特定語音（Text-Independent）兩種方法。特定語音方法必須有一定的語音內容作為辨識準則；非特定語音方法則無限定語音內容。本計劃將以特定語音探討語者驗證與語者認定方法。

基本上，就語音的模式，就如同語音辨識一樣，Hidden Markov Modeling 以 Doubly Stochastic Process 來模擬語音產生的統計與時間過程對語音處理是相當重要的里程碑。但由於 HMM 以 Maximum Likelihood 方式來訓練與辨識語者往往導致缺乏模式間的分辨性考量考量。因此，以 HMM 模式與類神經網路的分辨模式相配合也導致 HMM 成為類神經網路的一種 Stochastic temporal networks 模式。本子計畫將以 Hidden Markov Model (HMM)為技術基礎探討搭配的在語者辨識系統的應用效能。所探討主題包含五部份：(一)特徵的選取，(二)抗噪音之語音特徵選取，（三）HMM 辨識技術的考量，（四）配合語音辨識系統的語音身份認證、密碼辨識與資料庫查詢系統的設計與實做，（五）配合人像、唇、表情所整合的辨識系統。

英文摘要（Keywords: Speaker Recognition, Hidden Markov Model, Speech Processing）

The objective of this subproject is to study the speaker recognition methods and integrate the methods into the speaker-independent polysyllabic word recognition developed last year. Speaker recognition can be classified as speaker verification and speaker identification. Recognition methods can be considered from text-dependent and text-independent recognition. In this subproject, we consider the speaker verification and speaker identification based on text-dependent recognition. Hidden Markov model is important in representing the time-statistics of speech. However, the training and recognition process using maximum likelihood leads to the weakness in intermodel consideration. Various algorithms have been proposed for binding the HMMs with neural networks. In this subproject, we should consider the applying of HMMs to speaker recognition. There are mainly five topics in this subproject:

1. The selection of speech features.
2. The robust speech feature in noisy environ-ments.
3. The HMM recognition methods
4. The integration of the recognition method with speech recognition for speaker identification, speech access keys, and speech data acquirement.
5. The integration with other subprojects.

二、計畫之緣由與目的

當語者未知且身份有驗證的需求就是語者辨識系統的應用性場合。相對於指紋、眼彩虹等辨識系統，此技術的特殊性除了更方便外，並可藉電話、聲霸卡等透過電話線和網路來做遠距離辨識。藉由語者辨識技術，未來銀行、個人網路資料的存取、與其他服務的管制皆可藉由此種技術來達到。本子計畫目標將研究語者辨識技術，並將此技術整合入前兩年執行國科會計畫所完成的高辨識率非特定語者詞語辨識系統來整體完成語音身份認證、密碼辨識與資料庫查詢。

本年度此部份主要目標將對語音特徵作深入分析，並探討如何在不同環境、不同麥克風、不同傳輸管道下具高辨識率的語音特徵。此部份我們將採相同於背景部份分類，以特徵擷取和特徵間差異計算法、特徵轉換、和摹擬聽覺特徵三方面來來調整所建立系統特徵參數並測試在辦公室內、汽車內、和工作室內的抗噪效能。綜上所言，此部份工作為：

1. 特徵擷取和特徵間差異計算法的分析與測試。
2. 特徵轉換的分析與測試。

3. 摹擬聽覺特徵的分析與測試。

三、研究方法與成果

個人的語音特色包含音質、音量、音高、亮度、速度、音調等。而就此類研究主要是選取能表現語者特色的語音特徵；而這些特質，在人體生理上，發聲過程可以聲源與聲道兩部份。其中聲源基本上可以基頻（Pitch）與能量（Energy）兩項參數加以表示；而聲道則可以濾波器輸出 (Filter Bank Output) 、LP 係數、倒頻譜(Cepstrum)等為特徵參數。由於聲源特徵易模仿；而聲道特徵由於每人的聲道特性不同；因此，聲道特色為主要語者特徵考量。聲道的三大共振腔，此三腔特色因人而異，加上舌頭與與齒齦的使用因人而異，為語者辨識的主要生理考量。而聲道特徵在訊號模擬上，又以倒頻譜最常為人使用。本計劃將以此倒頻譜為基礎考量。倒頻譜中又可根據計算複雜度與物理特性分為 Mel-Cepstrum，LPC- Cepstrum，DFT-Cepstrum 等。

聲道特徵又可分長時間特徵(long-term feature)，短時間特徵(short-term feature)。長時間特徵主要得到一與發音內容無關的語音特徵作為辨識基礎。此長時間特徵由於未能有效將人發出不同聲音的聲道特色有效表示出來；因此較難得到高辨識率。而以短時間特徵在國語可以聲韻母為短時間考量單位，此短時間特徵比對結果可累積而得到最後比對結果。此種以聲韻母為特徵單元較有生理意義；但要達此種方法通常要有依自動切割聲韻母方法。本計畫將以短時間特徵為主要研究，長時間特徵為參考。並以上年度所成功發展的聲韻母切割方法可作為此部份研究的基礎。此類研究主要是選取對噪音較不敏感的語音特徵；相對於下一部份語音淨化處理不同的是，此方法並不嘗試將雜音消除，而是由受干擾語音直接求取較不受影響的語音特徵。此方法的優點是不須雜音的統計特色，此特色卻也可能造成未充分利用雜音統計分佈的缺點。文獻上此方面的研究大致可以特徵擷取和特徵間差異計算法、特徵轉換、和摹擬聽覺特徵三方面來介紹。

特徵擷取和特徵間差異計算法主要是探討如何找到較適當的語音表示法或是修改特徵差異的計算公式來達抗噪音的能力。文獻上指出 mel-scaled cepstrum 要比 LPC-cepstrum 具抗噪性，而 cepstrum 又比 DFT 頻譜好。而特徵間差異計算法的研究則指出 cepstrum norm 的計算，以 Euclidean distance 來計算特徵間差異容易受雜音影響，cepstral projection measure 可得較佳的抗噪性。其他 cepstrum 間的差異計算法如 WLR(weighted likelihood ratio), Lifted Cepstral Distance ,RPS(root power sums)和其他語音特徵表示法都在不同應用有其效能。本計畫將以上年度所建立詞語辨識系統，對這些表示法和差異計算法作測試與評估以找到，在中文詞語辨識需求下、最適當的表示法和差異計算法。另、由於對 convolutional noise 有其特色可加以考量。

特徵轉換法主要觀念是去找到一適當的轉換來將特徵參數轉換以得到抗噪性。在文獻中，LDA(Linear Discriminant Analysis) 是找一線性轉換來達抗噪性；另一方面，則嘗試已非線性轉換來達目標。

摹擬聽覺特徵法是根據人耳聽覺的特性來達抗噪性。此特性主要有三：第一點是人耳對不同頻率有不同敏感度，我們可根據此特色將人耳不敏感的語音和雜音去除；第二點是人耳在不同頻率的解析度並不相同，我們可藉此以 Nonuniform Filter Bank 方法來增加 SNR 比。第三點是人耳有一種遮蔽現象(Masking Effects)，也就是存在於一頻域的聲音能導致另一頻域聲音聽不到，我們能藉此將人耳不敏感的語音和雜音去除以利辨識。

四、結論與未來展望

本計畫成果已達成計畫書所提預期成果：

1. 提出具抗噪性的短時間特徵。
2. 提出以 HMM 為主的具抗噪性之語者辨識方法。
3. 提出具抗噪性之語音和語者辨識系統提出各種具語者調適性之語音辨識方法之理論訴求。