具有累進學習能力之貝氏預測法則在汽車語音辨識之應用

簡仁宗，廖國鴻

國立成功大學資訊工程學系

Page 179 ~ 197

Proceedings of Research on Computational Linguistics Conference XIII (ROCLING XIII)

Taipei, Taiwan

2000-08-24/2000-08-25

具有累進學習能力之貝氏預測法則在汽車語音辨識之應用

Bayesian Predictive Classification with Incremental Learning Capability for Car Speech Recognition

簡仁宗　廖國鴻

國立成功大學資訊工程學系

Email : jtchien@mail.ncku.edu.tw

摘要

在許多的語音辨識應用中，由於週遭環境噪音的影響，使得測試語句與原始語音隱藏式馬可夫模組產生不匹配，導致辨識率明顯的下降。因此，本論文提出以轉換為主之貝氏預測分類器來提昇雜訊語音的辨識率。我們將隱藏式馬可夫模組的平均值向量作轉換並結合貝氏預測分類法則理論，把轉換參數的不確定性導入辨識決策中，其中轉換參數的不確定性用一常態之事前機率密度來表示，因此，我們可以發展出一套以轉換為主的貝氏預測分類器做語音辨識之強健性決策法則。在本論文中另一個重要的特徵，即事前機率累進學習的能力，藉由累進觀測到之測試語料，我們可以連續的更新事前機率密度的統計量，此更新過的統計量可以追蹤到最新環境的統計特性，事前機率之最新統計量將使用在以轉換為主的貝氏預測分類器法則，一個具有強健性決策且能夠累進學習特性的辨識系統即可建立。我們收集了九十公里、五十公里及零公里(引擎開)三種汽車路況的語料，經由實驗的結果，我們的方法在各種路況下辨識率均有明顯的改善。

1. 簡介

在過去這十幾年來，自動語音辨識技術日新月異，進步的相當迅速，而這樣的進步，觸發了語音辨識技術實際應用的動機。例如：語音辨識應用在汽車行動電話之免持撥號系統，將使得行動電話的使用更具人性化，且讓駕駛者多了一份安全保障。可是對於語音辨識的實際應用，常常會遇到不匹配問題，因為語音辨識主要是以樣本比對的技術為基礎，若是語音辨識之應用環境與原始樣本之訓練環境不匹配，將會使得辨識率大幅地降低。而這樣的不匹配可能是來自於週遭的環境噪音、傳輸語音的通道不同、語者不同或訓練樣本模組時模組的不正確等等，而在實際的應用中，影響語音辨識的因素往往是上述多個失真來源的組合。因

此，為了克服語音辨識時不匹配之問題，提出一強健且有效率的補償技術是必要的。

許多學者也提出了許多的方法來解決訓練與測試環境不匹配的問題，一般來說，不匹配的問題可分別在訊號、特徵或模組空間來解決[17]，在此論文中，探討方向著重在模組空間的補償，一般模組補償的方法可分成兩大類：

一、**非參數形式之補償：**此類方法對模組做補償時並不做任何的補償函數形式的假設，例如，Minimax 分類法[14]之主要觀念，假設測試語料的最佳決策參數，會落在所給定模組參數之限定的鄰近範圍內，然後對其決策規則和所對應的參數做調整，以解決不匹配問題。此方法之優點可以克服最差不匹配的情況，但是當鄰近的範圍重疊時，字與字之間的混淆度增加，此方法便無法發揮效用，而且在連續語音辨識方面的應用，此方法將有困難。

二、**參數形式之補償：**此方法先假設一補償函數之形式，然後估測此補償函數中的參數，例如 MLLR（Maximum Likelihood Linear Regression）方法[13]，其假設訓練與測試環境的不匹配可用一線性轉換函數來補償，把原本訓練好之模組參數的平均值部分，做一線性轉換，使得轉換過的參數能較匹配於測試環境，而此轉換參數的估測，可用最佳相似度（Maximum Likelihood, ML）之演算法[7]來估測，然後將估測到的轉換參數嵌入（plug in）辨識的決策中。這樣的方式是假設所估測到的轉換參數是一個能夠代表訓練與測試環境不匹配的正確值，此方法之缺點，在辨識時存在著轉換參數估測錯誤的風險，不能夠很可靠地提昇辨識率。另外一種常見的參數形式之補償是 MAP（Maximum *A Posteriori*）調整方法[6]，其主要是應用最佳事後機率的法則來做調整，利用 EM 演算法[7]來估測參數，此方法也是存在著“參數估測錯誤”的風險，且由於在實際的語音辨識應用中，我們所能獲得的只是訓練好的模組參數與測試語料，而 ML 演算法需要大量的語料才能估測出可靠的參數值，所以此方法亦不適用於線上調整策略

由於在實際的語音辨識應用中，訓練與測試語料之間的不匹配是未知的，所以我們利用 Minimax 分類演算法的優點，把參數的不確定性引入辨識的決策中，在本研究中，我們採用貝氏預測分類器（Bayesian Predictive Classification, BPC）[8][16]並結合轉換為主的調整技術[13]，建立一以轉換為主之貝氏預測分類（Transformation-Based BPC, TBPC）強健決策法則，這裡我們是把轉換參數的不確定性引入辨識的決策中，如此可避免掉 ML 演算法點估測的估測錯誤風險。此外，由於在辨識的應用中，測試環境的統計特性是非固定的（Nonstationary），會隨著時間而改變，因此，我們把 TBPC 和事前機率線上累進學習（Online Prior Evolution, OPE）的策略結合在一起，建立一強健且具學習能力之辨識系統 TBPC-OPE。為了要讓辨識

系統具有累進學習的能力，我們採用近似最佳事後機率演算法[3][10]，此演算法，可提供在 TBPC 決策中轉換參數事前統計量之累進學習機制，從累進觀察到的測試語料中進行累進式學習，使辨識器不斷地追蹤到最新環境的統計特性。所以，我們所提出的辨識系統 TBPC-OPE 不僅具有強健性的決策 TBPC 而且具有線上累進學習的能力 OPE。

2. 以轉換為主之貝氏預測分類法則

為了要加強語音辨識的效能，發展一套以隱藏式馬可夫模組（Hidden Markov Model, HMM）為主的強健性決策法則是必要的。在本論文中我們要把強健統計決策的貝氏預測分類法則和 HMM 參數的線上調整技術結合在一起。因此，我們提出轉換為主的貝氏預測分類法則，以處理對於 HMM 平均值向量之轉換參數的不確定性，而在轉換參數之事前機率的部分，採用線上累進的策略來追蹤最新環境的統計量，如此，具有強健性決策且具學習能力之辨識系統即可建立。

2.1 近似最佳事後機率估測（Approximate MAP Estimation）

在一個人機互動系統中，測試語者的語音資料通常是充裕且以累進方式觀察到，而我們要以這些累進收集到的測試或調整語料來調整已訓練好的非特定語者 HMM 參數以適應測試環境的特性。根據線上轉換演算法[3]，我們可以利用累進收集到的語料來估測轉換參數，然後用此轉換參數來補償測試環境所產生的聲學變化。使用近似最佳事後機率估測[3][10]，我們可以得到

$$(W^{(n)},\eta^{(n)}) = \arg\max_{(W,\eta)} p(W,\eta|\chi^n) = \arg\max_{(W,\eta)} p(\mathbf{X}_n|W,\eta)\cdot p(W,\eta|\chi^{n-1})$$

$$\cong \arg\max_{(W,\eta)} p(\mathbf{X}_n|W,\eta)\cdot p(W,\eta|\varphi^{(n-1)}), \tag{1}$$

在（1）式中，先前累積之測試語料的事後機率密度函數 $p(W,\eta|\chi^{n-1})$ 用一最相近的事前機率密度函數 $p(W,\eta|\varphi^{(n-1)})$ 來取代，其中 $\varphi^{(n-1)}$ 會根據累積觀察到的測試語料 χ^{n-1} 不斷的更新。在本質上，W 和 η 是互相獨立的，因此，（1）式的估測可分解下列兩個步驟

$$W^{(n)} = \arg\max_{W} p(\mathbf{X}_n|W,\eta)\cdot p(W), \tag{2}$$

$$\eta^{(n)} = \arg\max_{\eta} p(\mathbf{X}_n|W^{(n)},\eta)\cdot p(\eta|\varphi^{(n-1)}), \tag{3}$$

其中 $\chi^n = \{\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n\}$ 是累進觀測到之測試語料，這裡假設每個語音音框互相統計獨立且

機率分佈相同，n 是測試語料的索引，$W^{(n)}$ 是第 n 句未知語料 $\mathbf{X_n}$ 的內容或字串列，$\eta^{(n)}$ 是第 n 句語料對應的轉換參數，$\varphi^{(n-1)}$ 是由前 n-1 句語料 χ^{n-1} 累進得到的事前機率參數（代表環境之統計量），$p(W)$ 是字串列的事前機率，也就是語言模型，$p(\eta|\varphi^{(n-1)})$ 是轉換參數 η 的事前機率密度函數。近似最佳事後機率估測法中的第一個步驟即執行（2）式，估測出對於現在語料 $\mathbf{X}_n$ 之最可能的語料內容 $W^{(n)}$，在這個步驟中，轉換參數 η 假設已被萃取出且被嵌入最佳事後機率解碼器（MAP decoder）的辨識系統中。在得到最可能的語料內容 $W^{(n)}$ 之後，第二個步驟，利用（3）式估測出最佳的轉換參數 $\eta^{(n)}$。基於此最佳事後機率估測演算法，當給定事前機率初始的統計量 $\varphi^{(0)}$，我們可以藉由測試語料 $\mathbf{X}_1$ 和其所對應最可能之語料內容 $W^{(1)}$ 代入式子（3），估測出轉換參數 $\eta^{(1)}$，同時 $\varphi^{(0)}$ 會更新變為 $\varphi^{(1)}$，此更新過的事前機率統計量 $\varphi^{(1)}$ 可用在下一次參數 $(W^{(2)},\eta^{(2)})$ 的估測。以此類推，我們可累進地估測出參數 $(W^{(1)},W^{(2)},\cdots,W^{(n)})$ 和 $(\eta^{(1)},\eta^{(2)},\cdots,\eta^{(n)})$，而此遞迴的演算法可使我們能夠累進地去追蹤最新環境的統計特性。因此可知，在本論文中辨識系統所採用的學習策略是*累進*而且是*非監督式（unsupervised）*的。

2.2 貝氏預測分類的應用

在傳統的辨識調整方法中，轉換參數 η 的估測可用最佳相似度法[13][17]、最佳事後機率法[1][6]或近似貝氏[3][10]估測法得到，而此估測到的轉換參數再嵌入最佳事後機率的辨識器中，然後辨識出最可能的語料內容。這樣的調整方法是在辨識的過程中，把點估測（Point Estimate）之轉換參數假裝為一個正確的值，而嵌入其聲學模組相似度的計算中，因此，隱含了點估測錯誤的風險，造成辨識的錯誤。所以，在本論文中提出轉換為主之貝氏預測分類法則，其主要的觀念是把轉換參數視為隨機變數並將其機率分佈考慮在聲學模組相似度的計算中，將轉換參數的不確定性平均起來取代點估測的方式，如此，可避免點估測錯誤的風險。在本研究中，我們用 $\tilde{p}_\eta(\mathbf{X}_n|W)$ 來取代（2）式中的相似度 $p(\mathbf{X}_n|W,\eta)$，（2）式可改寫成

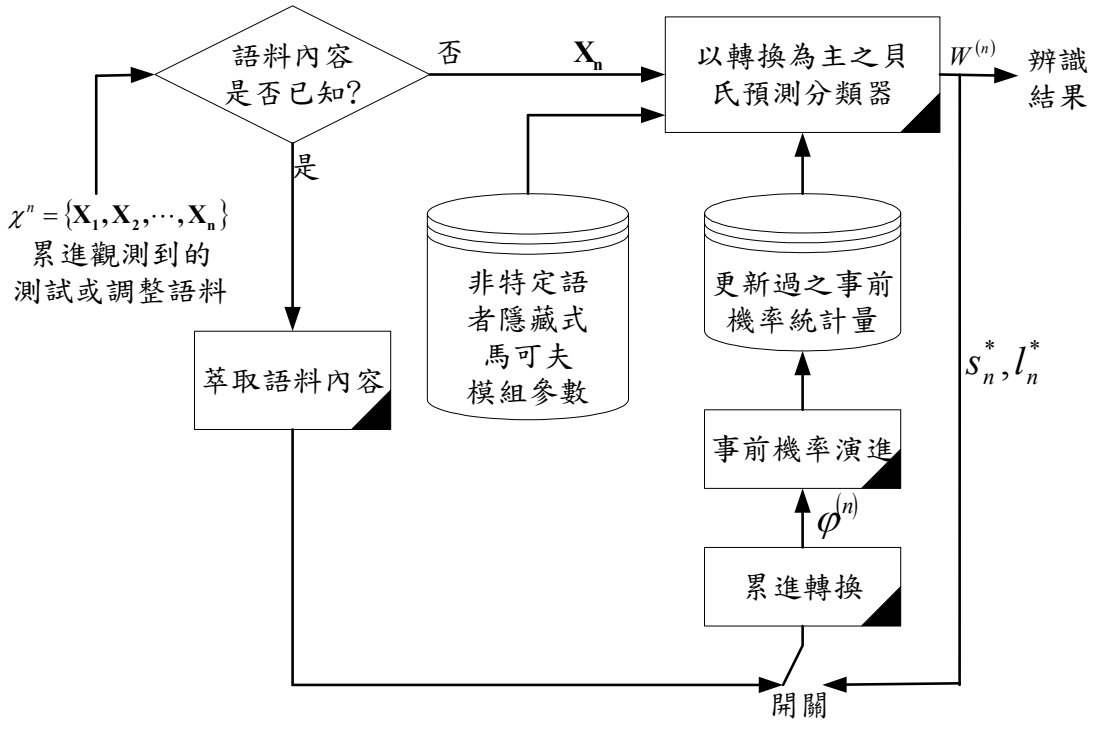
$$W^{(n)} = \arg\max_W \tilde{p}_\eta(\mathbf{X}_n|W)\cdot p(W), \tag{4}$$

其中相似度 $\tilde{p}_\eta(\mathbf{X}_n|W)$ 是由下列的積分得到

$$\tilde{p}_\eta(\mathbf{X}_n|W) = \int p(\mathbf{X}_n|W,\eta)p(\eta|\varphi^{(n-1)},W)d\eta.$$

$$= E\left\{p(\mathbf{X}_n|W,\eta)|\varphi^{(n-1)}\right\} \tag{5}$$

觀察第（5）式，我們可知以轉換為主的貝氏預測分類法是把轉換參數的不確定性引入辨識時相似度的計算，此預測機率分佈（Predictive Distribution）函式 $\tilde{p}_{\eta}(\mathbf{X}_n|W)$ 為轉換參數分佈中相似度函式 $p(\mathbf{X}_n|W,\eta)$ 的期望值，因此 TBPC 可視為是把轉換參數的不確定性平均化，這是和 ML 點估測不一樣的地方。所以，很明顯地，我們所提出的演算法其好處在於將轉換參數 η 的事前機率同時應用在第（4）和（5）式 TBPC 的辨識決策中而且也應用第（3）式累進轉換的技術裡，使事前機率具有累進學習的能力。因此，基於近似最佳事後機率估測法（3）（4），我們可以建立具有強健決策與累進學習能力的辨識系統，整個系統的策略如圖一所示。輸入端是累進觀測到的測試或調整語料 $\chi^n=\{\mathbf{X_1},\mathbf{X_2},\cdots,\mathbf{X_n}\}$ ，若輸入語料 $\mathbf{X_n}$ 的內容未知則為非監督式的學習，走“否”路線，此時亦需要 SI HMM 參數和上一次更新過之事前機率統計量，然後輸入以轉換為主之貝氏預測分類器，辨識出結果 $W^{(n)}$ 並得到最佳的狀態和混合數序列 s_n^*,l_n^* ，利用 s_n^*,l_n^* 的資訊，經過累進轉換處理可得到一組更新過的轉換參數事前機率統計量 $\varphi^{(n)}$ ，此更新過的統計量將用在下一句語料的 TBPC 決策，如此循環以達到累進式的學習，同理，對於監督式的學習（“是”路線），因為已經有相對應的語料內容，所以和非監督式的差別在於能夠提供較正確的資訊供事前機率統計量的累進，不過在一般較實用的語音辨識應用中，環境的學習均為非監督式。



圖一、TBPC-OPE 之辨識系統架構圖

3. 貝氏預測相似度量測

在上個章節中討論到以轉換為主之貝氏預測分類法則的應用，在這個章節我們將詳細地解說此強健性的決策法則如何推導出來。

3.1 以轉換為主之調整（Transformation-based Adaptation）

考慮一具有 L 個狀態 K 個混合數的隱藏式馬可夫模組 $\lambda=\{\lambda_i\}=\{\omega_{ik},\mu_{ik},r_{ik}\}$，$i=1,\cdots,L$，$k=1,\cdots,K$，我們定義對於語料 $\mathbf{X}_n$ 中觀測音框 $\mathbf{x}_t^{(n)}$ 之狀態觀測機率為一多變數高斯（Gaussian）機率密度函式

$$p(\mathrm{x}_t^{(n)}|\lambda_i)=\sum_{k=1}^{K}\omega_{ik}N(\mathrm{x}_t^{(n)}|\mu_{ik},r_{ik})$$

$$=\sum_{k=1}^{K}\omega_{ik}(2\pi)^{-d/2}|r_{ik}|^{1/2}\exp\left[-\frac{1}{2}(\mathbf{x}_t^{(n)}-\mu_{ik})^T r_{ik}(\mathbf{x}_t^{(n)}-\mu_{ik})\right] \tag{6}$$

其中 ω_{ik} 為混合數之權重且 $\sum_{k=1}^{K}\omega_{ik}=1$，而 $N(\mathbf{x}_t^{(n)}|\mu_{ik},r_{ik})$ 是具有 d 維度之平均值向量 μ_{ik} 和 $d\times d$ 精確矩陣（變異數矩陣之反矩陣）的高斯機率密度函式。當以轉換為主的調整被應用來克服環境的不匹配時，我們將會使用一些特定環境的語料 χ^n 估測出參數為 $\eta^{(n)}=\{\eta_c^{(n)}\}$ 的轉換函式 $G_{\eta^{(n)}}(\cdot)$，將分群過之隱藏式馬可夫模組參數做調整，使得調整過的模組參數能適用在新的辨識環境。在本論文中，我們考慮隱藏式馬可夫模組之平均值向量，加上一偏差向量 $\mu_c^{(n)}$ 以達到調整的目的，因此，轉換函式將被定義為

$$\lambda^{(n)}=G_{\eta^{(n)}}(\lambda)=\{\omega_{ik},\mu_{ik}+\mu_c^{(n)},r_{ik}\} \tag{7}$$

其中 c 代表轉換群組的索引，而具有索引 i 和 k 的隱藏式馬可夫參數假設是屬於第 c 轉換群組，亦即 $\lambda_{ik}\in\Omega_c$。因此，當給予隱藏式馬可夫模組參數 λ_{ik} 和轉換參數 $\eta_c^{(n)}=\mu_c^{(n)}$，則 $\mathbf{x}_t^{(n)}$ 的相似度函式表示為

$$p(\mathbf{x}_t^{(n)}|\lambda_{ik},\eta_c^{(n)})\propto|r_{ik}|^{1/2}\exp\left[-\frac{1}{2}(\mathbf{x}_t^{(n)}-\mu_{ik}-\mu_c^{(n)})^T r_{ik}(\mathbf{x}_t^{(n)}-\mu_{ik}-\mu_c^{(n)})\right]. \tag{8}$$

此外，在我們提出的架構中另一個重要的課題是事前機率密度函式的選擇。基於統計學上機率密度分佈之共軛特性（Conjugate Prior）[4]和式子（5）的計算考量上，我們選擇一多變數高斯機率密度函式來當轉換參數的事前機率密度函式，定義如下

$$p(\eta_c^{(n)}|\varphi_c^{(n-1)})=p(\mu_c^{(n)}|\nu_c^{(n-1)},\kappa_c^{(n-1)})\propto$$

$$|\kappa_c^{(n-1)}|^{1/2}\exp[(\mu_c^{(n)}-\nu_c^{(n-1)})^T\kappa_c^{(n-1)}(\mu_c^{(n)}-\nu_c^{(n-1)})] \tag{9}$$

其中 $\varphi_c^{(n-1)}=\left(\nu_c^{(n-1)},\kappa_c^{(n-1)}\right)$ ，代表累積（*n-1*）個語句對應之轉換參數事前機率統計量， $\nu_c^{(n-1)}$ 和 $\kappa_c^{(n-1)}$ 分別為此高斯函式之平均值向量和精確矩陣。

3.2 轉換為主貝氏預測分類法則之實現

在以轉換為主之貝氏預測分類法則中，對於一個未知之測試語料的決策是根據一預測機率分佈來決定，其計算是藉由導入轉換參數的不確定性而得到，然而式子（5）中的預測機率密度函式的計算是相當困難的。Huo *et al.*[8]對積分使用 Laplace 的方法去近似得到預測機率密度函式的值。他們也考慮隱藏式馬可夫模組 Missing Data 問題的特性來計算預測機率密度函式。因此，將觀測語料 $\mathbf{X}_n$ 之狀態和混合數序列 $(\mathbf{s}_n,\mathbf{l}_n)$ 整合進預測機率密度函式的計算中，產生下列式子[12]

$$\tilde{p}_\eta(\mathbf{X}_n|W)=\sum_{\mathbf{s}_n,\mathbf{l}_n}\int p(\mathbf{X}_n,\mathbf{s}_n,\mathbf{l}_n|W,\eta)p(\eta|\varphi^{(n-1)},W)d\eta$$

$$\cong\max_{\mathbf{s}_n,\mathbf{l}_n}\int p(\mathbf{X}_n,\mathbf{s}_n,\mathbf{l}_n|W,\eta)p(\eta|\varphi^{(n-1)},W)d\eta \tag{10}$$

其中所有可能狀態和混合數序列 $(\mathbf{s}_n,\mathbf{l}_n)$ 之機率總和，藉由代入最相似的狀態和混合數序列來趨近，此即所謂的 Viterbi 貝氏預測分類（Viterbi BPC）。在[12]中，一近似 Viterbi 貝氏演算法被提出並用來搜尋最佳未知的狀態和混合數序列 $(\mathbf{s}_n^*,\mathbf{l}_n^*)$ ，並且找出近似的預測機率密度函式。事實上，此 Viterbi 貝氏搜尋演算法並不能精確地完成 Viterbi 貝氏預測分類法則[12]。因此，另一種近似預測機率密度函式的方法，是直接地計算每一個觀測音框的預測機率密度函式，取代傳統使用的高斯機率密度函式。此預測機率密度函式可視為一補償過的觀測機率密度函式，並且將其應用到式子（2）之 MAP 決策中，如此便能近似貝氏預測分類法則。在[12]，此方法稱為 BP-MC （Bayesian Predictive Density Based Model Compensation）。在他們的實驗中， BP-MC 和 Viterbi BPC 在效能上互相媲美。所以，在本論文中採用 BP-MC 的方法實現式子（4）（5）以轉換為主之貝氏預測分類法則。使用這個方法，式子（6）中的狀態觀測機率密度函式被修改為

$$\tilde{p}(\mathbf{x}_t^{(n)}|\lambda_i)=\sum_{k=1}^{K}\omega_{ik}f(\mathbf{x}_t^{(n)}|\lambda_{ik}) \tag{11}$$

其中對於單一音框之補償過的觀測機率密度函式 $f(\mathbf{x}_t^{(n)}|\lambda_{ik})$ 被定義為

$$f(\mathbf{x}_t^{(n)}|\lambda_{ik})=\int p(\mathbf{x}_t^{(n)}|\lambda_{ik},\eta_c)p(\eta_c|\varphi_c^{(n-1)})d\eta_c \tag{12}$$

因此，我們將藉由把（11）式之預測機率密度函式嵌入 MAP 辨識器中進而建立以轉換為主之貝氏預測分類法則。此法則的重點即在推導（12）式中貝氏預測相似度量測（Bayesian Predictive Likelihood Measure, BPLM）。

3.3 貝氏預測相似度量測（BPLM）之推導

在式子（12）貝氏預測相似度量測中，我們把式子（8）和（9）代入式子（12）化簡即可得到

$$f(\mathbf{x}_t^{(n)}|\lambda_{ik}) \propto |r_{ik}|^{1/2} |\kappa_c^{(n-1)}|^{1/2} \int \exp\left\{ -\frac{1}{2}\left[(\mathbf{x}_t^{(n)} - \mu_{ik} - \mu_c)^T r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik} - \mu_c) \right.\right.$$
$$\left.\left. + (\mu_c - \nu_c^{(n-1)})^T \kappa_c^{(n-1)} (\mu_c - \nu_c^{(n-1)}) \right] \right\} d\mu_c \tag{13}$$

由於式子（13）中指數的部分可應用統計書[4]上的定理得到下列的等式

$$-\frac{1}{2}\left[(\mathbf{x}_t^{(n)} - \mu_{ik} - \mu_c)^T r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik} - \mu_c) + (\mu_c - \nu_c^{(n-1)})^T \kappa_c^{(n-1)} (\mu_c - \nu_c^{(n-1)}) \right]$$
$$= -\frac{1}{2}\left\{ (\mu_c - \overline{\mathbf{m}})^T (\kappa_c^{(n-1)} + r_{ik}) (\mu_c - \overline{\mathbf{m}}) \right.$$
$$\left. + (\kappa_c^{(n-1)} + r_{ik})^{-1} \kappa_c^{(n-1)} r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)}) (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)})^T \right\} \tag{14}$$

其中

$$\overline{\mathbf{m}} = (\kappa_c^{(n-1)} + r_{ik})^{-1} \left[\kappa_c^{(n-1)} \nu_c^{(n-1)} + r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik}) \right] \tag{15}$$

因此（13）式可以整理成

$$|r_{ik}|^{1/2} |\kappa_c^{(n-1)}|^{1/2} \exp\left[-\frac{1}{2} (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)})^T (\kappa_c^{(n-1)} + r_{ik})^{-1} \kappa_c^{(n-1)} r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)}) \right]$$
$$\times |\kappa_c^{(n-1)} + r_{ik}|^{-1/2} \int |\kappa_c^{(n-1)} + r_{ik}|^{1/2} \exp\left[-\frac{1}{2} (\mu_c - \overline{\mathbf{m}})^T (\kappa_c^{(n-1)} + r_{ik}) (\mu_c - \overline{\mathbf{m}}) \right] d\mu_c \tag{16}$$

在（16）式中包含了一個高斯機率密度函數的積分項，其積分結果為 1 ，因此我們得到

$$f(\mathbf{x}_t^{(n)}|\lambda_{ik}) \propto |r_{ik}|^{1/2} |\kappa_c^{(n-1)}|^{1/2} |\kappa_c^{(n-1)} + r_{ik}|^{-1/2}$$
$$\exp\left[-\frac{1}{2} (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)})^T (\kappa_c^{(n-1)} + r_{ik})^{-1} \kappa_c^{(n-1)} r_{ik} (\mathbf{x}_t^{(n)} - \mu_{ik} - \nu_c^{(n-1)}) \right] \tag{17}$$

本實驗我們假設 r_{ik} 和 $\kappa_c^{(n-1)}$ 均為對角化矩陣，我們可推導出

$$f(\mathbf{x}_t^{(n)}|\lambda_{ik}) \propto |\kappa_c^{(n-1)} + r_{ik}|^{1/2}$$

$$\exp\left[-\frac{1}{2}(\mathbf{x}_t^{(n)}-\mu_{ik}-\nu_c^{(n-1)})^T(\kappa_c^{(n-1)}+r_{ik})(\mathbf{x}_t^{(n)}-\mu_{ik}-\nu_c^{(n-1)})\right] \tag{18}$$

此結果將應用在辨識的決策中，我們可以觀察到此貝氏預測相似度量測為高斯分佈，與原本未補償的高斯分佈之差別在於（18）式把事前機率密度函式的平均值向量 $\nu_c^{(n-1)}$ 和精確矩陣 $\kappa_c^{(n-1)}$ 分別加到隱藏式馬可夫模組中的平均值向量 μ_{ik} 與精確矩陣 r_{ik} ，因此在此決策中引入轉換參數的事前機率統計量，可以使系統的決策更具強健性。

4. 事前機率之累進學習

在我們提出的方法中，除了強健的 TBPC 決策之外，另一個重要的特點為事前機率累進學習的能力。在[9]中，闡述以 BPC 為基礎之語音辨識系統，事前機率密度函式佔了相當重要的地位，他們評估出事前機率統計量的變化對於 BPC 的效能具有關鍵性的影響。因此，我們把累進學習之技術應用到本論文所提出的 TBPC 決策之事前機率統計量，使得辨識系統能不斷地追蹤到最新環境的統計特性，讓辨識系統更具強健性。

4.1 Viterbi 近似法

在式子（3）近似最佳事後機率估測法中，第 n 個累進轉換參數 $\eta^{(n)}$ 的估測是藉由最大化 $p(\mathbf{X}_n|W^{(n)},\eta)$ 和 $p(\eta|\varphi^{(n-1)})$ 的乘積而得到，其中 $p(\mathbf{X}_n|W^{(n)},\eta)$ 為目前語料 $\mathbf{X}_n$ 之觀測機率密度函式， $p(\eta|\varphi^{(n-1)})$ 為 η 之事前機率密度函式且其參數 $\varphi^{(n-1)}$ 由觀測語料 χ^{n-1} 累進學習得到。因為 TBPC 決策，在辨識出最佳 $W^{(n)}$ 的過程中， $\mathbf{X}_n$ 所對應之最相似（most likely）狀態和混合數序列 $(\mathbf{s}_n^*,\mathbf{l}_n^*)$ 亦可同時產生，所以在（3）式參數估測的演算法中，其 missing data 的問題可用下列的 Viterbi 近似法來克服

$$\eta^{(n)} \cong \arg\max_{\eta} R(\eta) = \arg\max_{\eta} p(\mathbf{X}_n,\mathbf{s}_n^*,\mathbf{l}_n^*|W^{(n)},\eta)p(\eta|\varphi^{(n-1)}) \tag{19}$$

第 c 類的 $R(\eta_c)$ 可以推導出以下的高斯機率函式

$$R(\eta_c) \propto \left|\kappa_c^{(n)}\right|^{1/2}\exp\left[-\frac{1}{2}\left(\mu_c-\nu_c^{(n)}\right)^T\kappa_c^{(n)}\left(\mu_c-\nu_c^{(n)}\right)\right] \propto p(\mu_c|\varphi_c^{(n)}) \tag{20}$$

其中參數統計量 $\varphi_c^{(n)}=\left(\nu_c^{(n)},\kappa_c^{(n)}\right)$ 表示為

$$\nu_c^{(n)} = \left(\kappa_c^{(n-1)}+\sum_{i,k\in\Omega_c}c_{ik}r_{ik}\right)^{-1}\cdot\left(\kappa_c^{(n-1)}\nu_c^{(n-1)}+\sum_{i,k\in\Omega_c}c_{ik}r_{ik}\overline{\mathbf{b}_c}\right) \tag{21}$$

$$\kappa_c^{(n)} = \kappa_c^{(n-1)} + \sum_{i,k\in\Omega_c} c_{ik} r_{ik} \tag{22}$$

其中

$$c_{ik} = \sum_t \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k) \tag{23}$$

$$\overline{\mathbf{b}_c} = \frac{\sum_t \sum_{i,k\in\Omega_c} (\mathbf{x}_t^{(n)} - \mu_{ik})\delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k)}{\sum_t \sum_{i,k\in\Omega_c} \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k)} \tag{24}$$

從（21）（22）式可知事前機率密度函式之參數 $\varphi_c^{(n-1)}$ 經觀測語料 $\mathbf{X_n}$ 辨識後更新為 $\varphi_c^{(n)} = \left(\nu_c^{(n)}, \kappa_c^{(n)}\right)$，因此我們的重點在於建立起可重複產生的高斯事前/事後機率對，這符合統計學上共軛分佈之特性，也就是當隨機變數其事前機率分佈為高斯分佈，則對於任何樣本量和樣本值，此隨機變數之事後機率分佈也是高斯分佈[4]，而在（19）式所得到的轉換參數 μ_c 之事後機率可表示為 $p\left(\mu_c \middle| \varphi_c^{(n)}\right)$，所以新的參數 $\left(\nu_c^{(n)}, \kappa_c^{(n)}\right)$ 可視為更新過之轉換參數事前機率的統計量，並將應用在下一句測試語句 $\mathbf{X}_{n+1}$ 的 TBPC 辨識。因此，線上累進學習能夠累進地提供強健性 TBPC 決策所需要轉換參數不確定性的最新知識。

4.2 初始參數之估測

由於我們所提出的 TBPC 決策法則導入轉換參數事前機率的統計量 $\left(\nu_c^{(n-1)}, \kappa_c^{(n-1)}\right)$，而對於此統計量我們所採取的策略為累進式的學習，所以，當第一測試語料區塊 $\mathbf{X}_1$ 進入 TBPC 決策時，此時我們需要一初始值 $\left(\nu_c^{(0)}, \kappa_c^{(0)}\right)$ 提供 TBPC 決策。基本上，此初始值對於未知的轉換參數要能提供足夠的知識，如此才能夠累進地產生可靠的參數值 $\left(\nu_c^{(n)}, \kappa_c^{(n)}\right)$。為了要讓此初始值能夠充分地反應出轉換的物理意義，我們從大量的訓練語料中估測出此初始值，估測的方式如下所示

$$\nu_c^{(0)} = \frac{\sum_t \sum_{i,k\in\Omega_c} \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k)(\mathbf{x}_t - \mu_{ik})}{\sum_t \sum_{i,k\in\Omega_c} \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k)} \tag{25}$$

$$\kappa_c^{(0)} = \frac{\sum_t \sum_{i,k\in\Omega_c} \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k) r_{ik}}{\sum_t \sum_{i,k\in\Omega_c} \delta(s_{n,t}^* - i)\delta(l_{n,t}^* - k)} \tag{26}$$

其中 $\mathbf{x}_t$ 代表訓練語料中音框 t 的特徵向量，也就是我們在 Segmental K-Means 訓練的最後

一次遞迴中，先找出訓練語料所對應的 HMM 參數。 $\nu_c^{(0)}$ 的產生為所有訓練語料與其所對應屬於轉換類別 c 之 HMM 參數 μ_{ik} 之偏差（bias）的平均值，$\kappa_c^{(0)}$ 為屬於 c 類別中精確矩陣的平均值，因此 $\left(\nu_c^{(0)},\kappa_c^{(0)}\right)$ 可充分地代表轉換參數的初始統計量。

5. 實驗

5.1 語料庫

我們準備了兩組語料庫[1]，一組是以近距離麥克風之方式錄下的*乾淨語料*，這組語料總共有 1400 句中文連續數字，包含了 70 位男生 70 位女生，每個人發音 10 句，其中 1000 句 50 男 50 女用來訓練乾淨之非特定語者 HMM 參數，另外 400 句 20 男 20 女用來做測試，另一組是*噪音語料庫*，此組語料庫為實際汽車環境下以遠距離麥克風方式錄得，此組語料包含 5 位男生 5 位女生發音的中文連續數字語料，每人分別在車速 0 公里（怠速路況）錄製 10 句，50 公里（市區路況）錄製 20 句，90 公里（高速公路路況）錄製 30 句，其中每位語者在不同路況下取出 5 句做為調整語料，其餘的做為測試語料。所使用的汽車為 TOYOTA COROLLA 1.8（錄製 2 男 2 女）和 YULON SENTRA 1.6（錄製 3 男 3 女），使用的錄音設備為 MD 隨身聽，型號為 MZ-R55，錄音採用遠距離錄音麥克風，型號為 SONY ECM-717 ，麥克風與語者間的距離約 50 公分，錄音時冷氣開啟，窗戶緊閉，所錄下的語音經由音效卡轉成數位音檔。此兩組語料庫取樣頻率均為 8 kHz ，以 16 bit 的方式儲存，連續數字長度為 3 至 11 個數字長。

為了瞭解汽車噪音語料庫中噪音程度，我們從錄得的噪音語料計算出兩種汽車在不同路況下的信號雜訊比（SNR），由於乾淨語音訊號並無法精確地從噪音語料中還原得到，因此，我們假設乾淨語音信號與背景噪音為互相獨立，故 SNR 的計算方式定義為

$$SNR(dB)=10\log_{10}\left(\frac{\sigma_{noisy\ speech}^2-\sigma_{noise}^2}{\sigma_{noise}^2}\right) \tag{27}$$

其中 $\sigma_{noisy\ speech}^2$ 為噪音語料的變異數， σ_{noise}^2 為背景噪音的變異數，SNR 以分貝（dB）為單位。由於使用兩種不同等級的汽車錄音，因此我們分別算出其所錄製語料的 SNR ，結果列於表一，我們可看到，車速越快所產生的噪音程度越高，SNR 值越低，另一方面，汽車等級越高，所產生的噪音程度越低，SNR 值越高。

信號雜訊比 SNR（dB）			
	YULON	TOYOTA	平均
0 公里	5.63	10.3	7.96
50 公里	-6.53	0.34	-3.1
90 公里	-10.14	-3.77	-6.96
乾淨語料	25.1		

表一、不同環境語料之信號雜訊比

5.2 辨識架構與基本辨識結果

在本論文的實驗中，語音特徵參數包含 12 階 LPC derived cepstrum、12 階 delta cepstrum、1 階 delta log energy 和 1 階 delta delta log energy，共 26 個維度。語音模組 HMM 的規格部分，每個數字用 7 個狀態來代表，每一語句前端插入一個背景雜訊狀態，後端亦插入一個背景雜訊狀態，而語句中數字與數字之間插入一個選擇性背景雜訊狀態，此三個背景雜訊狀態均不相同，所以我們的 HMM 共有 73 個狀態，每個狀態的混合數為 4 個。

在實驗部分，我們首先實現出一組在乾淨環境、0 公里車速、50 公里車速和 90 公里車速的基本實驗結果，實驗數據以數字錯誤率 （digit error rate, DER）來表示，乾淨環境下的測試語料為乾淨語料庫中的 400 句測試語料，而 0 公里、 50 公里和 90 公里車速下的測試語料分別為噪音語料庫中的 50 句、150 句和 250 句（每人在各公里車速分別有 5、15 和 25，共有 10 人）測試語料，在基本系統實驗中，語音辨識器是使用乾淨非特定語者 HMM 參數，不做任何的補償，實驗結果列於表二

測試語料	**數字錯誤率 DER（%）**
乾淨語料	10.6
0 公里噪音語料	25.61
50 公里噪音語料	54.97
90 公里噪音語料	62.33

表二、乾淨語料、0、50、90 公里語料之基本實驗結果

5.3 轉換類別個數與數字錯誤率之關係

在以轉換為主的貝氏預測分類法則中，辨識率會受轉換參數類別的個數影響，如果能夠

適當地增加類別的個數，將可提昇辨識率。這裡我們做了個簡單的實驗，利用人工把非特定語者 HMM 參數分為兩類，第一類為背景雜訊狀態類，第二類為非背景雜訊狀態類，因此我們需要估測出兩類的事前機率統計量做 TBPC 的決策。其實驗結果為表三，在表中列出使用 1 個和 2 個轉換類別個數的數字錯誤率，轉換參數之事前機率統計量是從測試語料中採非監督式的學習。由於本論文所提出的策略為累進式的學習，學習的效能會受測試語料的測試順序影響，因此我們將各個公里語料隨機產生 10 組測試順序，然後再將 10 組測試結果取平均，列出實驗結果。我們可看到使用 2 個轉換類別個數的數字錯誤率明顯降低。因此，在以下的實驗中，轉換類別個數均採用兩個。

測試語料 ＼ 類別個數	1	2
0 公里	14.32	12.53
50 公里	39.94	36.24
90 公里	51.65	46.32

表三、TBPC-OPE 中轉換類別個數與數字錯誤率之關係

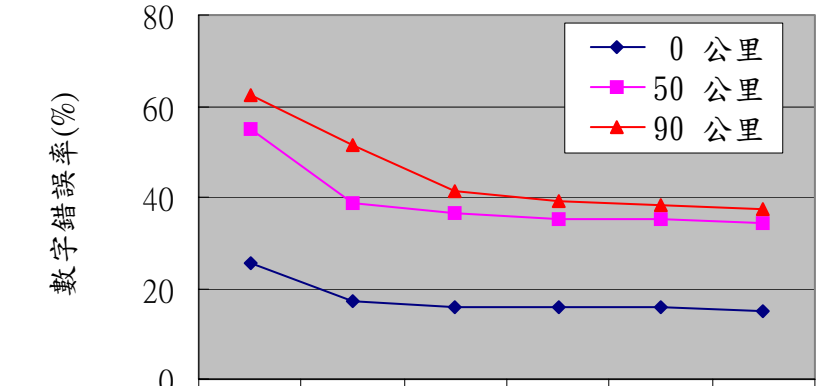
5.4 調整語料量與數字錯誤率之關係

為了瞭解調整語料與數字錯誤率之關係，我們利用監督式學習的方式，首先利用調整語料，以累進學習的方式估測轉換參數之事前機率統計量，然後把估測出的統計量應用在測試階段的 TBPC 決策，在測試階段並不做事前機率統計量的累進。所使用的調整語料為噪音語料庫中的調整語料，在各個車速下每人均有 5 句調整語料，在實驗中，我們考慮語者不匹配的問題，因此把每個語者的調整語料分開做監督式學習，然後把由調整語料累進學習到的事前機率統計量，應用在測試階段的 TBPC 決策中，在測試階段並不做事前機率統計量的更新。實驗結果列於圖三，由圖中可看出 TBPC 只需少量的調整語料，即使僅僅一句，就可明顯地降低數字錯誤率，例如 50 公里之語料，數字錯誤率降低了 30% 。當調整語料累進時，數字錯誤率呈現遞減之趨勢，我們可看到當使用完 5 句調整語料語料後，在各個公里車速下之語料均有 40 % 數字錯誤率的降低，由此可知，以轉換為主的貝氏預測分類法則加上事前機率累進式的學習，能夠累進地提昇辨識率。

5.5 不同貝氏預測分類器之比較

在這個部份的實驗，我們實現 Surendran 與 Lee 所提出的貝氏預測分類器[19]和 Jiang

與 Huo 所提出的貝氏預測分類器[11]，實驗結果比較於表四，



	0	1	2	3	4	5
0 公里	25.61	17.28	15.74	16.04	15.74	15.12
50 公里	54.97	38.48	36.51	35.26	34.95	34.12
90 公里	62.33	51.47	41.3	39.04	38.22	37.47

調整語料之句數

圖三、TBPC-OPE 中調整語料量與數字錯誤率之關係

測試語料＼方法	Baseline	Jiang and Huo	Surendran and Lee	TBPC-OPE
乾淨語料	10.6	8.51	8.4	7.47
0 公里	25.6	18.61	15.43	12.53
50 公里	55	49.83	38.38	36.24
90 公里	62.3	60.25	50.91	46.32

表四、不同 BPC 方法之數字錯誤率之比較

由實驗數據可看出，我們所提出的方法 TBPC-OPE 對於每種測試語料，數字錯誤率均比其他方法低，雖然 Jiang 與 Huo 所提出的貝氏預測分類器之辨識結果比基本系統好，但由於其所考慮的是非特定語者 HMM 參數中平均值向量的不確定性，所以當訓練與測試語料的不匹配程度增大時，利用 BPC 決策找出的最佳狀態和混合數序列 $(\mathbf{s}_n^*,\mathbf{l}_n^*)$ 的資訊常常不是很可靠，而其方法對於每一個 HMM 狀態和混合數的平均值向量均需一組事前機率統計量，當要利用少量觀測語句且不可靠的資訊 $(\mathbf{s}_n^*,\mathbf{l}_n^*)$ 來更新很多組的事前機率統計量，此時將有可能造成事前機率統計量的更新錯誤，導致辨識率的提昇有限，從乾淨、0 公里、50 公里、90 公里測試語料之辨識結果，即可得到驗證。對於 Surendran 與 Lee 所提出的貝氏預測分類器，

其考慮的是轉換參數的不確定性。轉換類別之個數在實驗中我們設定為 2 個，因此將能利用 Viterbi 演算法找出較佳的 $(\mathbf{s}_n^*,\mathbf{l}_n^*)$ 來估測出較可靠之轉換參數事前機率統計量，應用在第二次辨識時的 TBPC 決策，但由於其 TBPC 使用到的轉換參數事前機率統計量是利用當時輸入的測試語料估測出來的，並沒有累進式學習的能力，因此其數字錯誤率高於我們提出的 TBPC-OPE。在這個部分的實驗中，另一個值得注意的是乾淨測試語料的辨識，由於 BPC 把參數的不確定性引入辨識的決策中，因此 BPC 也能克服訓練模組時模組的不正確及語者的差異，由實驗結果得知， BPC 決策不僅能應用在噪音環境下做語音辨識，在乾淨的環境中也能提昇辨識率。

在語音辨識的應用中，辨識時間亦是重要之考量之一，因此我們也列出辨識時間的比較。實驗中我們所使用的設備為 Pentium III 450 及 256 Mega RAM，實驗結果列於表五，這裡我們列出每句語料所花的平均辨識秒數，數據顯示我們所提出的 TBPC-OPE 所花的辨識時間比其他兩種 BPC 方法少。由於 Surendran 和 Lee 所提出的方法需要兩次的辨認且須估測轉換參數事前機率的統計量，因此所花的辨識時間為最久。然而， Jiang 和 Huo 所提出的方法，其只須一次的辨識並對每個狀態每個混合數的事前機率統計量做估測與更新，因此所花的辨識時間比 Surendran 和 Lee 的方法少，而我們所提出的 TBPC-OPE 之辨識時間比 Jiang 和 Huo 的方法少，這是因為 TBPC-OPE 需要更新轉換參數事前機率統計量只有兩類，所以 TBPC-OPE 的辨識時間比 Jiang 和 Huo 的方法少，只比基本系統的辨識時間多一點點。

方法	Baseline	Jiang and Huo	Surendran and Lee	TBPC-OPE
辨識時間（秒／句）	0.138	0.245	0.382	0.192

表五、不同 BPC 法則之辨識時間比較

經由實驗結果發現我們所提出的 TBPC-OPE 降低的數字錯誤率最多，所花的辨識時間最短，因此 TBPC-OPE 是相當地具有潛力的方法。

6. 即時展示系統

為了實際評估我們所提出的演算法效能，我們實作了一套即時展示系統，直接在線上做語音辨識，此展示系統並不是直接開車在車上做展示；而是先錄製一段汽車背景雜訊，用喇叭放出來，以模擬實際的汽車噪音環境，而麥克風的位置是以免持式遠距離麥克風為主，大

約和語者相距 25 到 35 公分之間，線上錄下一段語音後做即時語音辨識。在展示系統中列出了三種方法的辨識結果：第一是基本系統的結果，在這個部分並不使用任何的補償技術，第二為 Jiang 與 Huo 所提出的貝氏預測分類器之辨識結果，第三為我們所提出的 TBPC-OPE 之辨識結果。

6.1 背景模組

由於在噪音環境下的中文連續語音辨識並不能達到百分之百的正確，而為了提昇噪音環境下的切音結果，我們在系統中提供可線上訓練背景模組的選項，當此選項勾選時，在辨識時的背景雜訊狀態均用此線上訓練出來的背景模組來取代。背景模組的訓練方式，會依實際錄得的聲音長短訓練出不同混合數個數的背景模組，如 32 或 16 個混合數。所使用的訓練演算法是用向量量化（Vector Quantization, VQ）演算法[15]。

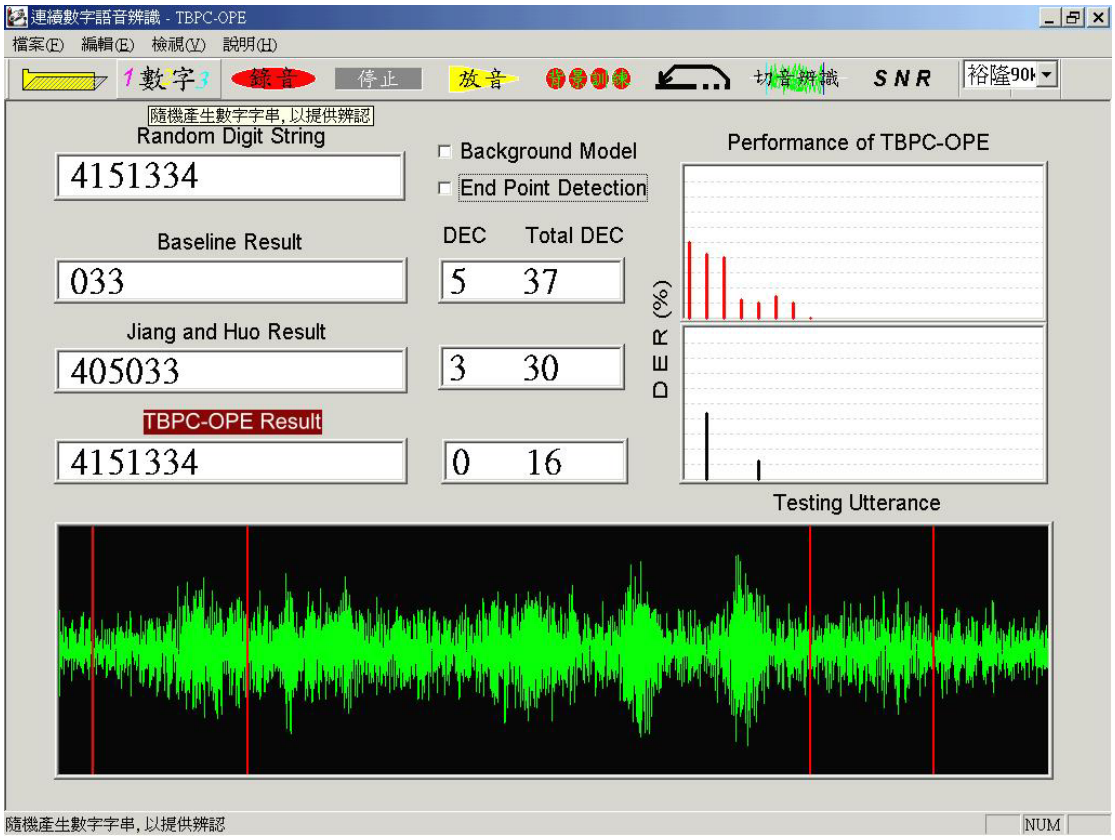
6.2 即時展示系統介面

如圖四所示，此即為我們所發展出的展示系統介面。所使用的工具為 Visual C++ 6.0 ，工作平台為 Microsoft Windows 2000。所使用的音效卡為創巨公司生產的，可支援全雙工，能在撥放噪音的同時進行錄音，因此我們把噪音種類的撥放選項加入展示系統的介面，讓使用者可以更方便選擇所要模擬的車速路況。因為在展示時，噪音撥放的音量大小會影響辨識的效能，因此在展示系統的介面中，我們提供可線上計算 SNR 的功能選項，以瞭解系統測試時的噪音強度。由於本系統做連續數字的辨識，因此我們可利用電腦隨機產生一組長度為 7 至 10 個數字字串，使用者就唸此字串，然後把每個方法得到的辨識結果和隨機產生的字串做比較，即可算出每個方法辨識的數字錯誤個數 DEC（digit error count）。為了要能客觀地顯示出 TBPC-OPE 的效能，我們也列出累積至目前的總數字錯誤個數 Total DEC，另一方面，為了顯示出 TBPC-OPE 累進學習演算法的效能，我們畫出 TBPC-OPE 每一句和每三句的數字錯誤率 DER（%）。

7. 結論

在本論文中，我們為噪音環境下的語音辨識提出一具有線上累進學習能力的 TBPC 強健性決策，由於 TBPC 決策把轉換參數的不確定性導入辨識的決策中，因此在噪音環境下的辨識效能會比傳統用 ML 或 MAP 做點估測（point estimate）的方法更具強健性。由於在 TBPC

的決策中使用轉換參數的事前機率統計量，而累進學習的策略能讓此統計量不斷地追蹤最新環境的統計特性，因此使得 TBPC 的決策更具可靠性。從監督式學習的實驗中，我們可觀察到 TBPC 強健性的決策，運用少量的調整語料，即使僅僅一句，就能降低在噪音環境下的數字錯誤率，而且數字錯誤率隨著調整語料的增加而遞減，更驗證了我們所提出的線上累進學習策略。在不同 BPC 應用的比較實驗中，由於我們所提出的是以轉換為主的 TBPC ，且採用累進學習的策略，因此，不論在辨識率或辨識時間方面均比 Surendran 與 Lee 所提出的貝氏預測分類器和 Jiang 與 Huo 所提出的貝氏預測分類器，具有較佳的辨識效能。另外一個值得注意的是， TBPC-OPE 在乾淨的測試語料中也能降低數字錯誤率，也就是克服在訓練模組階段時模組的不正確及語者差異所產生的不匹配問題。



圖四、展示系統之介面

參考文獻

[1] 簡仁宗 ，林敏順 (1999), “音框同步之雜訊補償方法在汽車語音辨識之應用”, 第十二屆計算語言學研討會, pp. 239-251.

[2] C. Chesta, O. Siohan and C.-H. Lee, “Maximum *a posteriori* linear regression for hidden Markov model adaptation”, *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, vol. 1, pp. 211-214, 1999.

[3] J.-T. Chien, “Online hierarchical transformation of hidden Markov models for speech recognition”, *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 656-667, November 1999.

[4] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, 1970.

[5] A. P. Dempster, N. M. Laird and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *J. Royal Statist. Society (B)*, vol. 39, pp. 1-38, 1977.

[6] J. L. Gauvain and C.-H. Lee, “Maximum *a posteriori* estimation for multivariate Gaussian mixture observations of Markov chains”, *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 291-298, 1994.

[7] X. D. Huang, Y. Ariki and M. A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, 1990.

[8] Q. Huo, H. Jiang and C.-H. Lee, “A Bayesian predictive classification approach to robust speech recognition”, *IEEE Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 1547-1550, 1997.

[9] Q. Huo and C.-H. Lee, “A study of prior sensitivity for Bayesian predictive classification based robust speech recognition”, *IEEE Proceedings of International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, vol 2, pp. 741-744, 1998.

[10] Q. Huo and C.-H. Lee, “On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate”, *IEEE Transactions Speech and Audio Processing*, vol. 5, no. 2, pp. 161-172, March 1997.

[11] H. Jiang, K. Hirose and Q. Huo, “Improving Viterbi Bayesian predictive classification via sequential Bayesian learning in robust speech recognition”, *Speech Communication*, vol. 28, no. 4, 1999.

[12] H. Jiang, K. Hirose and Q. Huo, “Robust speech recognition based on a Bayesian prediction approach”, *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, 1999.

[13] C. J. Leggetter and P. C. Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models”, *Computer Speech and Language*, vol. 9, pp. 171-185, 1995.

[14] N. Merhav and C.-H. Lee, “A minimax classification approach with application to robust speech recognition”, *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 90-100, 1993.

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

[16] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, UK, 1996.

[17] A. Sankar and C.-H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition", *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 190-202, 1996.

[18] K. Shinoda and T. Watanabe, "Speaker adaptation with autonomous control using tree structure", *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, pp. 1143-1146, 1995.

[19] A. C. Surendran and C.-H. Lee, "Predictive adaptation and compensation for robust speech recognition", *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, vol. 2, pp. 463-466, 1998.