鑑別性事前資訊應用於強健性語音辨識

丁川偉　吳柏樹　簡仁宗

國立成功大學資訊工程學系

{cwting, bswu, chien}@chien.csie.ncku.edu.tw

摘要

在傳統語音辨識系統中，模型的訓練環境與測試環境不匹配(mismatch)**是造成辨識率下降的首要問題，在此議題上，過去文獻已提出許多解決方法，如在語音模型端引入模型參數的不確定性所建立的強健性貝氏預測分類**(Bayesian predictive classification)**法則，或是調整模型於測試環境的調適方法，如最大事後機率**(MAP)**調適以及線性迴歸**(MLLR)**調適，甚至進一步考慮語音模型鑑別性之最小分類錯誤線性迴歸**(MCELR)**調適等方法。其中，貝氏預測分類法則是將模型參數的不確定性**(uncertainty)**適當的引入決策法則以達到決策方法的強健性，而參數不確定性反應了雜訊環境及聲學的變異性，它可由事前機率**(prior density)**來表示，而傳統上貝氏學習則提供了估測並更新參數事前資訊的機制。**

為兼顧決策法則的強健性及鑑別性，本論文提出在貝氏預測分類架構下聲學模型及其事前機率模型之鑑別性訓練及更新，我們使用最小分類錯誤(MCE)**之鑑別性準則來估測模型參數之超參數**(hyperparameter)**，並且提出了兩種更新的方法，其一是直接針對隱藏式馬可夫模型平均值向量參數更新其事前統計量；其二是考慮線性迴歸調整，針對迴歸矩陣之事前資訊在最小分類錯誤準則下做更新。在以汽車噪音雜訊語音資料庫為主的評估實驗中，發現使用更新過後的事前機率可以提昇貝氏預測分類之鑑別性，達成強健性語音辨識效能提升之目的。**

1. 緒論

語音是人與人之間最直接、最自然溝通的方式，隨著科技和語音辨認技術的進步，讓機械聽懂人類的話，不再是遙不可及的夢想。目前實際應用面中，語音辨識的過程仍舊存在著許多問題，最常遇到的，像是訓練環境與測試環境的不匹配問題，因為語音辨識主要是以樣本比對(pattern recognition)**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、或語者不同等，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，強健且有效率的補償技術一直是語音辨識極為重要的研究議題。**

在此研究領域上，已有許多學者提出不同方法來解決不匹配的問題，我們將之大致分為訊號(signal)**空間、特徵參數**(feature)**空間、以及模型參數**(model parameter)**空間三類。在第一種方法中，主要以語音強化**(speech enhancement)**的方式為主，其觀點是將受到環境影響的訊號，透過訊號處理的方式，消減噪音的部份以得到近似乾淨的訊號；第二種方法，與訊號空間的處理觀念類似，都是希望還原原始環境下的特徵參數特性，做特徵參數的補償**(compensation)**；最後一種則是對已經訓練完成的模型參數做處理，其方法可再細分為兩種：其一是利用新環境所得到的少量語料**

將原有之模型參數調適到與新環境接近的方式；另一則是在模型參數中考量其不確定性，以減少新環境中模型變異所造成的影響，進而達到強健性決策的機制。此外在模型的訓練當中，不同模型之間的參數或分佈常會面臨混淆的情況，造成分類錯誤的提升，因此鑑別性(discriminability)的考量也被學者提出引入模型的訓練過程，以期達到更明確之模型並降低分類的錯誤。

在本研究中，主要是在考量參數不確定性的基礎上，希望能夠在鑑別性的分類方法考量下更新其參數的不確定性，以期望進一步達成同時具有鑑別性事前機率的強健性之決策法則。另外在本研究中也將此考量不確定性且具鑑別性的事前機率學習，落實在模型參數的調整，並分為直接對模型參數的調整以及間接對模型參數做調整。在以汽車噪音為主的連續數字語料庫中，都能達到辨識效能的提升。而在本文的編排上，共分為五個小節，除了第一節為緒論外；第二節將簡單介紹模型調適、鑑別式訓練與強健性決策法則之觀念；第三節則為所提出之強健性決策下鑑別性事前資訊學習機制；最後第四、第五兩節則為實驗與結論部份。

2. 文獻探討

2.1 語音調適技術

為了解決在語者或環境所造成的不匹配的變異，學者[4][8]即提出以調適的方法來克服上述所提的問題。調適的精神即在於，以原有的語者獨立聲學模型為初始，運用少量稀疏的特定語者或新環境的調適資料，經由參數的調整，來產生一組新的聲學模型，使它更具有強健性並更適合於測試的語者或環境以達到提升辨識率的效果。調適的技術又可分為特徵向量層面(feature-based adaptation)以及模型層面(model-based adaptation)來調適，在本研究中主要是針對模型層面的調適技術來做說明。

而在模型參數的調適方法中，又可分為模型參數的直接調整與間接調整。在直接調整的技術中，最大化事後機率(maximum *a posteriori*, MAP)[4]為最基本的方法；而間接調整的方法則以線性迴歸轉換為基礎的調適演算法為主。以下將分別介紹之。

2.1.1 直接調整技術

直接調適模型參數的方法，因為是針對模型中各別的參數做調整，所以對於沒有對應調適語料的參數就沒辦法做參數的更新，其示意圖可由 MAP 演算法概念所表示：

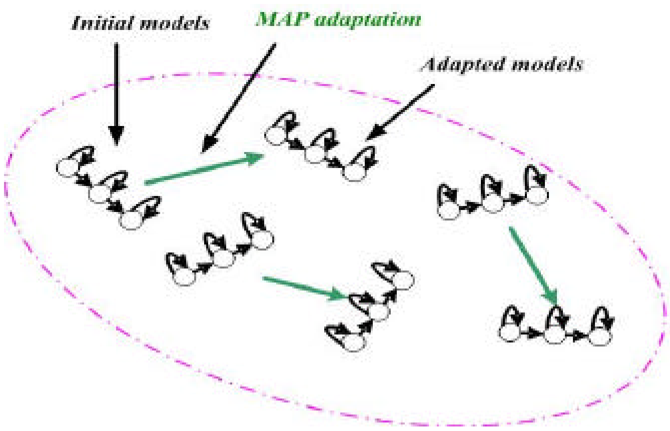

圖 一、最大化事後機率調適方法示意圖

MAP 的基本精神在於結合了事前機率以及調適語料，來估測出新的模型參數。如下式所示，X 代表觀測值，Λ_{MAP} 表示模型參數的最大機率估測值，利用貝氏定理可將事後機率拆解成事前機率與相似度的結合，$g(\Lambda \mid X)$ 及 $g(\Lambda)$ 分別表示參數的事後機率分佈及事前機率分佈，$f(X \mid \Lambda)$ 則是觀測資料 X 的相似度。

$$\Lambda_{MAP} = \arg\max_{\Lambda} g(\Lambda \mid X) = \arg\max_{\Lambda} f(X \mid \Lambda) g(\Lambda) \tag{1}$$

其中在事前機率 $g(\Lambda)$ 的部份，若不考慮事前機率 $g(\Lambda)$ 而將其設為常數函數 (non-informative prior)，則(1)式將會退化成為一般的最大相似度估測法，由此可看出最大化事後機率調適法考慮事前機率的分佈與調整語料的結合，來將事後機率進行最大化的調適。

在最大化事事後機率調適準則下，可表示調適後的模型參數如：

$$\hat{m}_{MAP} = \frac{Nt^2}{s^2 + Nt^2}\overline{\mathbf{x}} + \frac{s^2}{s^2 + Nt^2} r \tag{2}$$

其中 N 為調適語料的觀測資料數，$\overline{\mathrm{x}}$ 為調適語料的平均值，其分佈為 $N(\overline{\mathrm{x}}, s^2)$，事前機率則假設為 $N(r, t^2)$ 之機率分佈。

由此可看出，最大化事後機率調適方法結合了事前機率與調適語料所蘊含的資訊。當訊練語料足夠時，可充分將其使用到調適語料以達成良好的調整效能；而在訓練語料不足時，則是由調適語料所運算出來的模型參數平均值向量來主導，然而在此情況下，如果模型複雜度較高或包含較多的參數，則相對而言也必須使用更多的調適語料來估算出每個新的參數值，這也是為何最大化事後機率調適方法，或是直接調整模型參數的方法在調適語料數量的要求上會比較嚴格的原因。

2.1.2 *間接調整技術*

在間接調整模型參數的調適方法中，以線性迴歸轉換為基礎的調適演算法為主，此類方法是以轉換整個模型或是定義的相關狀態分群，所以雖然有部分模型或是狀態沒有對應的調適的語料，但是在同一個分群的，就可以分享使用同一個轉換矩陣做參數的調適轉換，因此在調適語料的數量要求便大為降低，其示意圖可由圖二所表示。

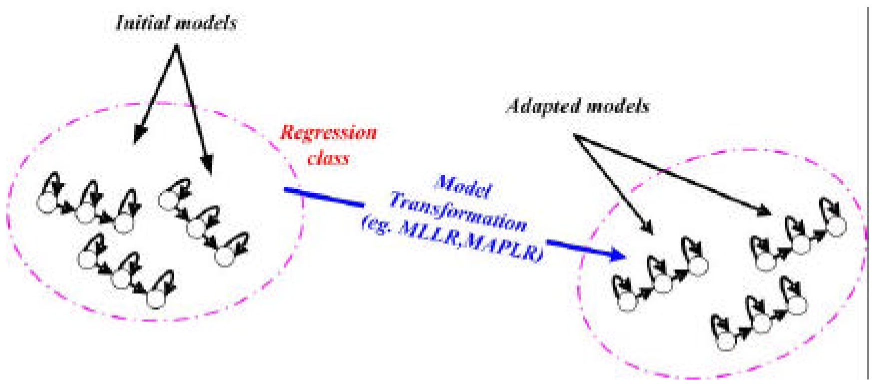

圖 二、迴歸矩陣為基礎的調適方法示意圖

以轉換方式做調適的方法中，最被廣泛使用的即為最大相似度線性迴歸(maximum likelihood linear regression, MLLR)[3][15]調適演算法。其主要為估測同一個馬可夫模型參數群聚(regression class)中的轉換，然後以此轉換對整個群聚內的參數從語者獨立轉換到特定的語者或環境條件下，所以調適語料不需要包含所有模型中的參數，只要在同一個群聚下的轉換有被估測出來，其他同一群聚下的參數就可以一併作調適 而在最大相似度的線性迴歸轉換就是在以最大化相似度的決策法則下做線性迴歸的轉換矩陣估測。

假設在以高斯分佈為狀態機率的隱藏式馬可夫模型中，模型參數 Λ_s，其平均值向量為 m_s 其維度為 $d \times 1$，而經過轉換矩陣 W_s，可以得到調適後的參數 $\hat{m}_s$，其轉換過程我們可以表示為：

$$\hat{m}_s = \mathrm{A}_s m_s + \mathrm{b}_s = \mathrm{W}_s x_\mathrm{s} \tag{3}$$

其中，轉換除了是對於平均值向量的線性轉換外，還加上一個偏差的值，所以整體上我們可以整理出 $x_\mathrm{s} = [1 \;\; m_s^T]^T$ 維度為 $(d+1) \times 1$ 且轉換矩陣為 $\mathrm{W}_s = [A_\mathrm{s} \;\; b_s]$，其維度為 $d \times (d+1)$。調適後的狀態機率為：

$$b_s(x) = \frac{1}{(2p)^{d/2} \left|\Sigma_s\right|^{1/2}} \exp\left\{-1/2\left(\mathbf{x} - \mathrm{W}_{r(s)} x_\mathrm{s}\right)^T \Sigma_s^{-1} \left(\mathbf{x} - \mathrm{W}_{r(s)} x_\mathrm{s}\right)\right\} \tag{4}$$

在整體的訓練中，假設訓練語料為 X，且所有需要估測的轉換矩陣為對應到各別群聚類別的轉換矩陣的集合 $W = \left\{\mathrm{W}_{r(s)}\right\}$，且估測的轉換矩陣必須符合於最大化相似度的法則 $W_{ML} = \arg\max_{W} p(X \mid W, \Lambda)$。由於調適的模型是假設為隱藏式馬可夫模型，經由 EM(expectation maximization)演算法可定義不完整資料對於最大相似度函數的期望值的輔助函數[15]

$$Q(\Lambda, \overline{\Lambda}) = \sum_q p(X, q \mid \Lambda) \log p(X, q \mid \overline{\Lambda}) \tag{5}$$

其中 Λ 為目前的參數估測參數值，而 $\overline{\Lambda}$ 是新的參數估測值。透過對輔助函數中之轉換矩陣參數微分並令其為零可求得轉換矩陣之更新運算式

$$w_n^{ML} = \left(\sum_t \sum_{r=1}^{R} \frac{g_{s_r}(t)}{s_{s_r,i}^2} \mathbf{x}_{t,i} x_{s_r}^T\right) \cdot \left(\sum_t \sum_{r=1}^{R} \frac{g_{s_r}(t)}{s_{s_r,i}^2} x_{s_r} x_{s_r}^T\right) \tag{6}$$

2.2 鑑別性訓練法則

在以鑑別性為主的估測方法，最著名的為最小分類錯誤(minimum classification error, MCE)[13]、最大互斥資訊(maximum mutual information, MMI)[16]等方法為主，近年來又有以支持向量機(Support Vector Machine, SVM)的精神為出發點的最大化邊界(Large Margin)[9]分類法則，主要是以求出使得辨識效果最差的邊界，然後以更新模型參數來拉大此分類錯誤的邊界，以找出最佳的分類邊界。以下我們簡單介紹最小分類錯誤之觀念與訓練方式。最小分類錯誤的鑑別性訓練法

則，是由 Juang et al.[14]所提出來的，在此鑑別性訓練方法中主要分為三個部份，一是定義其鑑別性函數間的錯誤分類量測(misclassification measure)，其二是利用損失函數(loss function)來表示其分類正確與錯誤率，第三步驟是在最小化期望損失(expected loss)的目標下估測模型參數。在統計型語音辨識中，鑑別性函數主要以相似度函數(likelihood function)來代表，而參數估測上主要是以最小化訓練觀測資料的期望的損失為目標，來對應到最小化辨識錯誤率的關係。以下我們開始介紹最小錯誤分類法則中的三個步驟：

1) 錯誤分類量測定義為：

$$d_i(X) = -g_i(X;\Lambda) + \left[\frac{1}{M-1}\sum_{j,j\neq k}\exp\{hg_j(X;\Lambda)\}\right]^{\frac{1}{h}} \quad (7)$$

其量測主要是在於辨識的錯誤率，C_j 為其混淆類別(confusing classes)或稱競爭類別(competing classes)，$g_i(X;\Lambda)$ 是觀測語料 X 對應到正確類別 C_i 的辨識分數，$\left[\frac{1}{M-1}\sum_{j,j\neq k}\exp\{hg_j(X;\Lambda)\}\right]^{\frac{1}{h}}$ 是對應到其他非 C_i 的競爭類別對應到觀測語料 X 的辨識分數，中括號內容的項是對應到 L^h norm，當 $d_k(X) > 0$ 代表發生分類錯誤，$d_k(X) \leq 0$ 代表正確分類。 其中 h 是一個正數，改變 h 及 M 的值，可以改變(7)中具影響力的競爭類別數量，當 $h \rightarrow \infty$，則對應到錯誤辨識的分數為競爭類別中最高的類別：

$$d_i(X) = -g_i(X;\Lambda) + \max_{j,j\neq i} g_j(X;\Lambda) \quad (8)$$

此處類別 C_j 是除了類別 C_i 外，和觀察資料 X 相似度最大的競爭類別。

2) 對於某個觀察資料 X，以損失函數來定義分類器的分類風險，能進一步的表示錯誤分類量測與辨識錯誤率的關係，把錯誤分類量測代入由 sigmoid function 所近似的 0-1 損失函數，即

$$l_i(X;\Lambda) = l\left(d_i(X)\right) \quad (9)$$

其中 sigmoid function 是一個對應到值域範圍為[0,1]的連續性函數且具有可微分的特性，sigmoid function 定義如下 $l_i(d_i) = \frac{1}{1+\exp(-gd_i+q)}$，其中 q 常設為 0，而 g 常設為 1 或大於 1，另外因為當 $d_i(X) < 0$ 時，代表分類正確，所以 $l_i(d_i)$ 會相對於接近於零，代表沒有辨識錯誤的損失；而當 $d_i(X) > 0$ 時，則 $l_i(d_i)$ 會大於零，所以也代表對於 X 有辨識錯誤的損失。

3) 在參數估測上，主要就是尋找能最小化所有觀測資料的期望損失的模型參數，對應於一個觀測資料，其決策上的損失為

$$\ell(X;\Lambda)=\sum_{i=1}^{M} l_i(X;\Lambda)1(X\in C_i) \tag{10}$$

其中$1(\cdot)$為 indicator function，當條件符合時值為 1，否則為 0。在考量所有觀測資料下，整體期望損失為$\ell(\Lambda)=E_X\left[l(X;\Lambda)\right]$，最後，即可利用廣義機率遞減演算法(generalized probabilistic decent, GPD)進行疊代運算以實現 MCE 法則[13]。

$$\Lambda_{t+1}=\Lambda_t-e_t U_t \nabla\ell(X_t,\Lambda)\Big|_{\Lambda=\Lambda_t} \tag{11}$$

2.3 強健性決策法則

以統計為主的決策法則，要實現最佳貝氏分類器必須滿足三點：假設觀測資料的空間Ω_x是已知的、其損失函數是給定的、以及假設真正的機率分佈是已知的。但是在實際應用中建構貝氏決策法則時，我們只有有限的訓練資料，並不能代表整體的觀測資料，而且資料對應的真實分佈並非事前可以得知。為了數學上及計算上方便，真實的分佈往往假設為特定的機率分佈，如高斯分佈，也因為現實的情況與理想上不符合，所以在現實狀況中，設計的決策法則會有 1)少量樣本的失真；2)模型假設的失真；3)測試環境訊號的失真；三種假設錯誤所帶來的失真，導致最佳化的決策法則是無法得到，而其中以第三種測試環境與訓練環境不匹配的問題對語音辨識的衝擊最為嚴重。因此強健性的決策法則被提出，其主要是強調當訓練與測試時具有不一致性的條件差異時，仍然能夠保有一定品質的辨識效率，且能減輕在不匹配的情況下辨識效率的損失。所以如何加強語音辨識系統的強健性是近年來重的要研究議題之一。在強健性的決策法則中主要是考慮參數具有不確定性範圍的隨機性(randomness)而非特定的(deterministic)，其中以貝氏預測分類法則(Bayesian predictive classification, BPC) [6]為基礎的方法會在以下說明。

在貝氏預測決策法則[5][10][11]，是在事前假設的不確定性分佈下以平均的方式考慮所有不確定性中的參數點對決策的影響力，而其參數的不確定性也可以視為其參數的事前資訊。所以此分類器主要是考慮所有可能性的參數點對於給定資料的期望值做為決策的依據。在貝氏預測分類器中，我們首先給定模型參數(Λ,Γ)的事前機率，$p(\Lambda,\Gamma\,|\,j_\Lambda,j_\Gamma)$視為參數的不確定性範圍，且聲學模型參數$\Lambda$和語言模型參數$\Gamma$各自存在於不確定性的範圍$\Omega_\Lambda$和$\Omega_\Gamma$中，當進一步假設聲學模型及語言模型間的獨立性，$p(\Lambda,\Gamma\,|\,j_\Lambda,j_\Gamma)=p(\Lambda\,|\,j_\Lambda)\cdot p(\Gamma\,|\,j_\Gamma)$，所以，其分類器的不確定性是落在參數事前機率分佈下的一個範圍：

$$M_e^*=\left\{p_\Lambda(X\,|\,W),p_\Gamma(W)\,|\,(\Lambda,\Gamma)\sim p(\Lambda,\Gamma\,|\,j_\Lambda,j_\Gamma);\Lambda\in\Omega_\Lambda,\Gamma\in\Omega_\Gamma\right\} \tag{12}$$

其模型參數之不確定性可由圖三示意之。

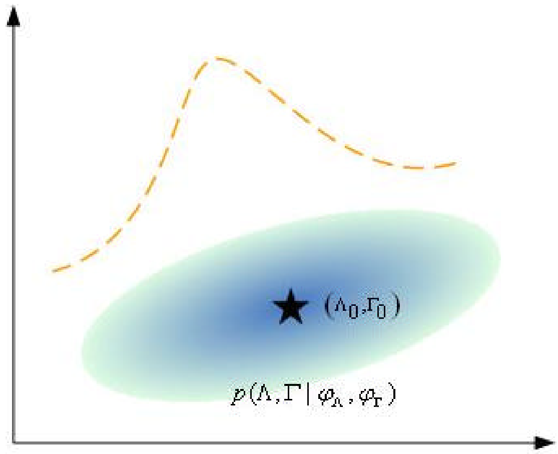

圖 三、貝氏預測法則參數不確定性示意圖

當進一步考慮所有的訓練語料及對應的答案 (X,W) 與給定參數的事前機率 $p(\Lambda,\Gamma\,|\,j_\Lambda,j_\Gamma)$ 和損失函數(loss function)$l(W,W')$**，則其對應的整體決策風險為：**

$$\begin{aligned}\tilde{r}(d(\cdot)) &= E_{(W,X)}E_{(\Lambda,\Gamma)}[l(W,d(X))] \\ &= \sum_{W\in\Omega_W}\int_{X\in\Omega_W} l(W,d(X)\tilde{p}(X\,|\,W)\tilde{p}(W)dX\end{aligned} \tag{13}$$

其中 $\tilde{p}(X\,|\,W)$ 為對應到聲學模型的預測分佈密度函數 $\tilde{p}(X\,|\,W)=\int_{\Omega_\Lambda} p(X\,|\,W,\Lambda)p(\Lambda\,|\,j_\Lambda)d\Lambda$；$\tilde{p}(W)$ 為其對應到語言模型的預測分佈密度函數 $\tilde{p}(W)=\int_{\Omega_\Gamma} p(W\,|\,\Gamma)p(\Gamma\,|\,j_\Gamma)d\Gamma$，當上式為 0-1 **的損失函數時，使得上述 $\tilde{r}(d(\cdot))$ 整體風險最小的分類器為：**

$$\tilde{d}(X) = \arg\max_W \tilde{P}(W,X) = \arg\max \tilde{P}(X\,|\,W)\cdot\tilde{P}(W) \tag{14}$$

就是對應到最到最佳的貝氏預測分類器，而其與嵌入式最大事後機率決策法則最大的不同是，其在辨識時考慮了模型參數的不確定性，而非只是單一的模型參數點。

在貝氏預測分類器的研究上，除了在模型的參數上假設其不確定性外，還有假設聲學模型的不匹配可以透過空間轉換的關係做補償[1][2]**，並且假設模型參數間轉換函數的參數具有不確定性 $M_e^* = \{p_\Lambda(X\,|\,W)\,|\,\Lambda = V_{J_\Lambda}(\Lambda_0), J_\Lambda \sim p(J_\Lambda)\}$，其對應的預測密度函數為：**

$$\tilde{p}(X\,|\,W) = \int_{\Omega_\Lambda} p(X\,|\,W,\Lambda = V_{J_\Lambda}(\Lambda_0))\,p(J_\Lambda)d\Lambda \tag{15}$$

其中 $V_{J_\Lambda}(\Lambda_0)$ 是模型參數的轉換函數，而此轉換函數是由 J_Λ 所控制。在此類以轉換函數為主的事前機率，主要是在於轉換模型參數的轉換函數其調整參數上。而在此類分類器中，在設計上必須要先假設事前機率的分佈，以及給定事前機率分佈的超參數，以及如何去計算預測密度函數等問題，而前兩個問題是在實作中假設的問題，而第三個問題是如何近似預測密度函數的問題。

在 BPC **的決策上，當辨識的模型為隱藏式馬可夫模型時，且狀態機率是由混合高斯模型所**

組成，在求解貝氏預測密度函數會包含有狀態序列(state sequence)以及高斯的混合索引(mixture index)的遺失資料(missing data)問題, 因此要對 $\tilde{p}(X \mid W)$ 完整的計算其預測密度函數是複雜度十分高且不可實現的。所以在考慮其簡化的近似方法上是一個研究議題，在文獻中，主要有近似貝氏預測分類器(quasi-Bayes predictive classification, QBPC)、維特比貝氏預測分類器(Viterbi Bayesian predictive classification, VBPC)、以及貝氏預測模型補償(Bayesian predictive density based model compensation, BP-MC)三種方式來近似計算 BPC。為考量在實作上與傳統語音模型作最具效率的連結，本研究採用 BP-MC 的方式近似 BPC 分數。

3. 鑑別性事前資訊訓練方法

在語音辨識的系統中，總是存在著很多不同於訓練環境中的影響，其有可能是來自於環境因素的差異，或者是語者發音特性上不同的差異，甚至於訓練或調適語料不足等等在樣本不足上的問題，這些因素均為降低語音辨識器效能的重要因素之一。有鑑於此，在解決這方面的問題上有在參數訊號端做補償的方法，如隨機向量補償或是鑑別性特徵參數擷取；或是在模型端的補償方法，如第二節中所提的強健性決策方法及調適方法。在貝氏預測分類器中，文獻中主要是以貝氏學習的方法來更新其不確定的事前資訊，此外在文獻中，也有學者提出在隨機向量補償的鑑別性不確定估測方法。在此研究中，我們主要是考慮在語音模型端的不確定性事前資訊的鑑別性調整方法，並且同時考慮了在語音模型參數的不確定性，以及調適轉換矩陣兩種不確定性的鑑別性調整方法。

在我們提出的方法上，以物理意義而言，當在不穩定的環境中估測參數，其一定會隱藏很多估測上的錯誤及失真，所以若不考慮參數的隨機性，而在模型參數上以鑑別性的參數估測方式求出對於訓練資料的最佳分類決策參數，雖然能夠表達出對於訓練或是調整語料的鑑別性，但卻未必能對抗存在於測試環境中參數隨機性所造成的失真影響。所以我們會以考慮參數不確定性的強健性方法，來模擬出此不匹配現象下的參數隨機性。極端而言，對於單一的模型，其不確定性的範圍最好能含蓋到所有存在此模型參數的隨機性空間越好，所以對於不確定性的估測和鑑別性的估測，存在有觀念上互相抗衡之處，因此在估測不確定性及加強鑑別性之間存在於一種抉擇(trade off)的關係。

但是[7]指出，雖然不確定性能夠在存在不確定性的環境中展現出其模型的強健性，當選取的不確定性越大，則相對的會造成不同模型間因為其不確定性有重疊而造成分類決策上的混淆，所以進一步的文獻[1][2][12]中，以貝氏學習的方法在分佈估測(distribution estimation)的精神下，學習對於每個模型而言最能表達其不確定性的事前資訊分佈，然而在貝氏學習訓練下的不確定性也仍舊可能存在於與臨近類別間混淆的情況。所以我們會考慮在不確定性事前資訊上考慮其鑑別性的更新，希望此不確定性的更新除了能夠描述整體的參數不確定性外，也能夠減低與其他類別參數不確定性間的混淆所造成的辨識率下滑，以達成具鑑別性的強健性決策法則的效果。我們並將此觀點落實在模型參數調整的技術當中。

3.1 直接調整

在此方法中，我們主要是以最小分類錯誤來實現我們的事前資訊鑑別性方法，不同於模型參數更

新的方法，我們考量了模型參數的不確定性，而且其不確定性是由一個事前機率密度函數 $p(\Lambda \mid j, W)$ 來表示，且其事前機率密度函數可以是由事前的訓練語料，或是對於測試環境的不確定性有所了解而估測出的事前分佈，而對應於考慮模型參數的不確定性的貝氏預估分類器可以表示為：

$$\tilde{p}(X \mid W) = \log \int_{\Omega} p(X \mid \Lambda, W) p(\Lambda \mid j, W) d\Lambda \tag{16}$$

當我們假設模型為隱藏式馬可夫模型時，其參數為 $\Lambda = \{a_{ij}, c_{ik}, m_{ik}, \Sigma_{ik}, i, j = 1,\ldots, M, k = 1,\ldots K\}$，此模型中共有 M 個狀態，其中 a_{ij} 為狀態轉移機率，每一個狀態內是有 K 個混合數的高斯混合模型，其中 c_{ik} 為第 i 個狀態的第 k 個高斯模型的混合索引，m_{ik} 及 $,\Sigma_{ik}$ 為對應到第 i 個狀態第 k 個高斯模型的平均值向量及共變異矩陣。考量 BPC 近似方法，當我們要計算(16)時，因為會遭遇到狀態序列及混合索引等遺失資料的問題

$$\begin{aligned} \tilde{p}(X \mid W) &= \int_{\Omega} p(X \mid \Lambda, W) p(\Lambda \mid j, W) d\Lambda \\ &= \sum_{s,l} \int_{\Omega} p(X, s, l \mid \Lambda, W) p(\Lambda \mid j, W) d\Lambda \end{aligned} \tag{17}$$

所以我們進一步以音框為單位的補償方法來近似此貝氏預測分類器，在此論文中，我們假設只考慮隱藏式馬可夫模型的平均值向量 m_{ik} 的不確定性，其他參數則視為固定不改變。而且假設一個模型中的每一個狀態均存在一個平均值向量，其不確定性事前機率為可定義為如下之高斯分佈

$$\begin{aligned} p(m_{ik} \mid j_{ik}, W) &= \frac{1}{\sqrt{2p}\,|t_{ik}|^{d/2}} \exp \frac{-1}{2} \left\{ (m_{ik} - m_{ik})^t t_{ik}^{-1} (m_{ik} - m_{ik}) \right\} \\ &= N(m_{ik}; m_{ik}, t_{ik}) \end{aligned} \tag{18}$$

其中 $j_{ik} = \{m_{ik}, t_{ik}\}$ 為第 i 個狀態第 k 個混合高斯平均值向量 m_{ik} 的不確定性事前機率超參數。把不確定的影響以音框為單位作補償並且假設其共變異矩陣為對角化矩陣 $\Sigma_{ik} = diag\{s_{ikd}^2\}$ 且平均值矩陣的共變異矩陣超參數也為對角化矩陣 $t_{ik} = diag\{t_{ikd}^2\}$ 時，所以每一個狀態內的每一個混合高斯其貝氏預測密度函數可以分維度各別計算為

$$\begin{aligned} \tilde{f}_{ikd}(x_d) &= \int f(x_d \mid q_{ikd}) p(q_{ikd} \mid j_{ikd}) dm_{ikd} \\ &= N(x_d, m_{ikd}, g_{ikd}^2) \end{aligned} \tag{19}$$

其中 $g_{ikd}^2 = s_{ikd}^2 + t_{ikd}^2$，且對於上式的各別狀態中個別補償過後的高斯混合模型為 $\tilde{p}_i(\mathbf{x}) = \sum_{k=1}^{K} w_{ik} \tilde{f}_{ik}(\mathbf{x})$。在導入鑑別性決策法則上，在此我們的鑑別性函數是定義為貝氏預測密度函數，當給定觀測資料為 $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ 後其鑑別性函數可以寫為下式：

$$\begin{aligned} \tilde{g}(X; W, l, j_q) &= \log \tilde{P}_R(X \mid W) = \log \left[p_{\hat{s}_0} \prod_t a_{\hat{s}_{t-1}\hat{s}_t} w_{\hat{s}_t \hat{l}_t} \tilde{f}_{\hat{s}_t \hat{l}_t}(\mathbf{x}_t) \right] \\ &= A_{\hat{s}\hat{l}}^* + \sum_t \sum_d \log w_{\hat{s}_t \hat{l}_t} \tilde{f}_{\hat{s}_t \hat{l}_t}(x_{td}) \end{aligned} \tag{20}$$

其中假設 $\hat{s}$ 及 $\hat{l}$ 為對應於正確字串的最佳狀態序列及最佳混合索引序列。且在最小分類錯誤決策法則中，可定義出分類錯誤量測與對應於超參數的期望損失函數 $\ell(X;j)=\frac{1}{M}\sum_{i=1}^{M} l_i(X;j)$，使用廣義梯度遞減的方法對模型平均值向量的超參數做參數更新，對於微分量先做符合於參數限制的事前轉換 $m_{ikd} \rightarrow \tilde{m}_{ikd}=m_{ikd}/g_{ikd}$，$t_{ikd} \rightarrow \tilde{t}_{ikd}=\log t_{ikd}$ 後，對於 $\tilde{m}_{ikd}(n+1)=\tilde{m}_{ikd}(n)-e_n \left.\frac{\partial l_i(X_n;j)}{\partial \tilde{m}_{ikd}}\right|_{\tilde{m}_{ikd}=\tilde{m}_{ikd}(n)}$，則對應於超參數的平均值微分量為：

$$\frac{\partial l_i(X_n;j)}{\partial \tilde{m}_{ikd}}=-g l_i(d_i)(1-l_i(d_i))\sum_{t=1}^{T} d(\hat{s}_t-i)d(\hat{l}_t-k)\left(\frac{x_{td}}{g_{ikd}}-\tilde{m}_{ikd}\right) \quad (21)$$

其中 $d(\cdot)$ 是一個 Kronecker delta function 同理於對超參數的變異數更新量 $\tilde{t}_{ikd}(n+1)=\tilde{t}_{ikd}(n)-e_n \left.\frac{\partial l_i(X_n;j)}{\partial \tilde{t}_{ikd}}\right|_{\tilde{t}_{ikd}=\tilde{t}_{ikd}(n)}$ 其對應的微分量推導為：

$$\frac{\partial l_i(X_n;j)}{\partial \tilde{t}_{ikd}}=-g l_i(d_i)(1-l_i(d_i))\sum_{t=1}^{T} d(\hat{s}_t-i)d(\hat{l}_t-k)\left[\left(\frac{x_{td}-m_{ikd}}{g_{ikd}}\right)^2-1\right]\left(\frac{t_{ikd}^2}{g_{ikd}^2}\right) \quad (22)$$

，並且對上式之微分量參數轉回原參數空間即可完成鑑別性的超參數更新法則。

得到具有不確定性資訊補償的模型後，進一步，我們在最小分類誤差的鑑別性法則下，給定部分的調整語料來對其不確定性的事前資訊分佈做調整，所以此調整出來的模型會是在以鑑別性為主軸下所調整出來且不同於以貝氏更新方法的貝氏預估分類器，而且因為我們主要是針對於參數的不確定性資訊做調整，所以進一步來說，我們的方法是以強化模型間的鑑別性下的模型參數的不確定性調整，在此研究中我們將此方法命名為 DBPC(discriminative BPC)。

3.2 間接調整

在環境的不匹配狀況下除了以直接對模型參數的調整方法外，還有以間接的調整方法，如以迴歸矩陣為基礎的方法，而模型參數的直接調適，可以細部的針對每一個狀態或是混合高斯做調整，但是當我們的調整語料有所不足時，直接調適則可能遭遇資料量不完全或資料量不足的問題而使得調整效能無法充分彰顯，所以另一個方法是以群聚的方式用轉換的概念做批次調整，這樣可以彌補單獨調整時資料不完整或是緩和調適資料量不足的問題，然而在以轉換矩陣為主的調整方法中，也主要是以特定的(deterministic)的轉換矩陣方法做調整，但是當考慮環境變異性及少量調整資料的問題情況下，轉換矩陣也有其不確定性，而在 LRBPC(linear regression based BPC)中首度以考量轉換矩陣也存在有不確定性的情況下，達成一個具有考慮了轉換矩陣不確定性的決策法則，進一步的，我們是假設當轉換矩陣的不確定性在調整上是以鑑別性的條件下來做學習，以達成在模型參數的轉換中經由鑑別性的轉換矩陣不確定性調整，而達成轉換後的參數也具有鑑別性的不確定行映射。

在 MLLR 中，其主要是找出一組特定的轉換矩陣 $\hat{R}$ 然後透過此轉換矩陣來做參數的轉換以達成調適的目的，所以其決策法則主要是給定原來的模型參數及轉換矩陣下轉換參數決策：

$$\hat{W} = \arg\max_{W} P(X \mid W, \Lambda, \hat{R}) \tag{23}$$

但是當我們進一步考慮其轉換矩陣 $g(R \mid j)$ 也有其不確定資訊時，我們可以線性迴歸貝氏預測分類器的方法平均化的考慮轉換矩陣的不確定性資訊：

$$\tilde{P}_R(X \mid W, \Lambda) = \int p(X \mid W, R, \Lambda) g(R \mid j) dR \tag{24}$$

所以在此我們可以將其線性迴歸貝氏預測分類器，取代(24)中以嵌入(plug-in)特定迴歸矩陣的方法來表示：

$$\hat{W} = \arg\max_{W} \tilde{P}_R(X \mid W, \Lambda) \tag{25}$$

所以針對第 i 個狀態第 k 個混合高斯的平均值轉換可以表示為 $\hat{m}_{ik} = A_c m_{ik} + B_c = R_c x_{ik}$，其中下標 c 代表的是其高斯所對應到的群聚類別(regression class)，且轉換矩陣 $R_c = [B_c \;\; A_c]$ 且延伸的平均值向量為 $x_{ik} = [1 \;\; m_{ik}^T]^T$，當我們進一步假設其轉換矩陣 R_c 為單變數的迴歸矩陣時，則其中 A_c 矩陣為對角化的矩陣 $A_c = diag\{a_{cl}\}$，則代表每個維度的轉換可以各自計算，所以對應到第 d 維的平均值可以表示為 $\hat{m}_{ikd} = a_{cd} m_{ikd} + b_c$ [15]。在此簡化的轉換矩陣中我們另外定義其另定義 $q_c = [a_{c1}, ..., a_{cd}, b_{c1}, ..., b_{cd}]^T$，且考慮以音框為單位的補償型式則

$$\tilde{f}_{ik}(\mathbf{x}_t) = \int f(\mathbf{x}_t \mid q_c, m_{ik}, \Sigma_{ik}) g(q_c \mid j_c) dq_c \tag{26}$$

其中 $f_{ik}(\mathbf{x}_t)$ 為原來 HMM 模型中第 i 個狀態第 k 個混合高斯，而 $\tilde{f}_{ik}(\mathbf{x}_t)$ 為經過以音框為單位的補償後的機率模型, 當 HMM 的共變異矩陣為對角化矩陣時 $\Sigma_{ik} = diag\{s_{ikd}^2\}$, 則 $f_{ik}(\mathbf{x}_t)$ 可以拆開為每個維度各別考慮，其中 $q_{cd} = [a_{cd} \;\; b_{cd}]^T$。假設 $g(q_{cd} \mid j_{cd})$ 為高斯分佈且不同維度間的轉換參數假設為獨立，則對應到第 d 維的轉換矩陣參數可以表示 $j_{cd} = \{m_{q_{cd}}, \Sigma_{q_{cd}}\}$，所以對應到第 d 維的轉換參數事前機率為：

$$\begin{aligned} g(q_{cd} \mid j_{cd}) &= g(a_{cd}, b_{cd} \mid m_{q_{cd}}, \Sigma_{q_{cd}}) \\ &= \frac{1}{2p} \left|\Sigma_{q_{cd}}\right|^{-1/2} \exp\left\{\frac{-1}{2}\left(q_{cd} - m_{q_{cd}}\right)^T \Sigma_{q_{cd}}^{-1} \left(q_{cd} - m_{q_{cd}}\right)\right\} \end{aligned} \tag{27}$$

其中 $m_{q_{cd}} = \left[m_{a_{cd}} \;\; m_{b_{cd}}\right]^T$ 為第 c 個轉換矩陣的第 d 維的參數的平均值超參數，而

$$\Sigma_{q_{cd}} = \begin{bmatrix} s_{a_{cd}}^2 & s_{a_{cd}b_{cd}}^2 \\ s_{a_{cd}b_{cd}}^2 & s_{b_{cd}}^2 \end{bmatrix} \tag{28}$$

為此維度下轉換參數的共變異矩陣。將(27)式帶回其(26)式整理其積分後可以得到，其線性迴歸

貝氏預測密度函數為：

$$\tilde{f}_{ik}(x_{td}) = N\left(x_{td}; \hat{m}_{x,ikd}, \hat{s}^{\,2}_{x,ikd}\right) \tag{29}$$

$$\hat{s}^{\,2}_{x,ikd} = s^{\,2}_{b_{cd}}\left(1+\frac{s^{\,2}_{a_{cd}}}{s^{\,2}_{b_{cd}}}\right)^2 + m^2_{ikd}\left(s^{\,2}_{a_{cd}} - \frac{s^{\,4}_{a_{cd}b_{cd}}}{s^{\,2}_{b_{cd}}}\right) + s^{\,2}_{ikd} \tag{30}$$

$$\hat{m}_{x,ikd} = m_{a_{cd}} m_{ikd} + m_{b_{cd}} \tag{31}$$

在導入鑑別性決策法則上，同前子小節，鑑別性函數為

$$\begin{aligned}\tilde{g}(X;W,\Lambda,j_q) &= \log \tilde{P}_R(X\mid W,\Lambda) = \log\left[p_{\hat{s}_0}\prod_t a_{\hat{s}_{t-1}\hat{s}_t} w_{\hat{s}_t\hat{l}_t} f_{\hat{s}_t\hat{l}_t}(\mathbf{x}_t)\right] \\ &= A^*_{\hat{s}\hat{l}} + \sum_t\sum_d \log f_{\hat{s}_t\hat{l}_t}(x_{td})\end{aligned} \tag{32}$$

其中 $A^*_{\hat{s}\hat{l}}$ 為最佳狀態序列及最佳混合數序列所對應的狀態轉移機率的分數。則所對應的分類錯誤量測函數 $d(X;\Lambda,j_q) = -\tilde{g}(X;W_i,\Lambda,j_q) + \tilde{G}(X;\Lambda,j_q)$ 同樣代入 sigmoid function 函數成為損失函數後，利用廣義梯度遞減的方法可對於轉換矩陣參數的不確定性，並且對於參數做符合限制的預先轉換 $j \rightarrow \tilde{j}_c$ 則超參數的更新通式為

$$j_c(n+1) = j_c(n) - e_n \left.\frac{\partial l(X_n;\Lambda,j_q)}{\partial j_c}\right|_{j_c = j_c(n)} \tag{33}$$

上式 j_c 為對應到第 c 個迴歸矩陣的超參數，另外分維度各別微分 $j_{cd} = \left\{m_{q_{cd}}, \Sigma_{q_{cd}}\right\}$ 可得

$$\frac{\partial l(X_i;\Lambda,j_q)}{\partial j_{cd}} = a l(d)(1-l(d))\left\{-\frac{\partial \tilde{g}(X;W_i,\Lambda,j_q)}{\partial j_{cd}} + \frac{\partial \tilde{G}(X;\Lambda,j_q)}{\partial j_{cd}}\right\} \tag{34}$$

定義 $\Omega_c = \{i,k\}$ 為屬於第 c 個群聚類別的狀態混合高斯索引，然後對於鑑別性函數中的迴歸矩陣的超參數平均值及超參數偏差平均值微分量為

$$\frac{\partial \tilde{g}(X;W_i,\Lambda,j_q)}{\partial \tilde{m}_{a_{cd}}} = -\sum_t\sum_{ik\in\Omega_c} V_t(i,k)\left[\left(\frac{m_{ikd}m_{a_{cd}} + m_{b_{cd}} - x_{td}}{\hat{s}_{x,ikd}}\right) m_{ikd}\right] \tag{35}$$

$$\frac{\partial \tilde{g}(X;W_i,\Lambda,j_q)}{\partial \tilde{m}_{b_{cd}}} = -\sum_t\sum_{ik\in\Omega_c} V_t(i,k)\left[\frac{\left(m_{ikd}m_{a_{cd}} + m_{b_{cd}} - x_{td}\right)}{\hat{s}_{x,ikd}}\right] \tag{36}$$

轉換矩陣的超參數變異數及偏差值超參數變異數的微分量可以更新如下

$$\frac{\partial \tilde{g}(X;W_i,\Lambda,j_q)}{\partial \tilde{s}_{a_{cd}}} = \sum_t\sum_{ik\in\Omega_c} V_t(i,k)\left[\frac{\left(u_{ikd}m_{q_{cd}} - x_{td}\right)^2}{\hat{s}^{\,2}_{x,ikd}} - 1\right]\cdot\frac{m^2_{ikd}s^{\,2}_{a_{cd}}}{\hat{s}^{\,2}_{x,ikd}} \tag{37}$$

$$\frac{\partial \tilde{g}(X;W_i,\Lambda,j_q)}{\partial \tilde{s}_{b_{cd}}} = \sum_t\sum_{i,k\in\Omega_c} V_t(i,k)\left[\frac{\left(u_{ikd}m_{q_{cd}} - x_{td}\right)^2}{\hat{s}^{\,2}_{x,ikd}} - 1\right]\cdot\frac{s^{\,2}_{b_{cd}}}{\hat{s}^{\,2}_{x,ikd}} \tag{38}$$

其中 $V_t(i,k)$ 為對應到時間點 t 第 i 個狀態第 k 個高斯發生的事後機率且 $u_{ikd} = [m_{ikd}\ \ 1]$，在以

使用 Viterbi 演算法做最佳狀態及混合數序列的取代時 $V_t(i,k)$ 可以用 Kronecker delta function 取代，並且將對應的前處理轉回原參數空間，所以在此部分，我們提出一個以鑑別性法則來更新其迴歸矩陣超參數的統計量，並命名為 DLRBPC(discriminative LRBPC)，以達成依據給定的調適語料鑑別性的學習其轉換矩陣不確性的範圍。

4. 實驗

4.1 語料庫

連續數字語音資料庫

訓練語料庫共有 1000 句中文連續數字，包括 50 位男生、 50 位女生，每個人發音 10 句，每句語音長度為 3 至 11 個數字長，所有語料皆為麥克風錄音，語音訊號取樣頻率為 8kHZ，語音訊號量化析度為 16 位元。另外我們使用汽車噪音環境下之語料庫(CARNAV98)中的五十公里噪音之語料作為測試語料。其中共有 10 位語者，每人各錄製十五句語料。

汽車噪音語料庫

此組語料庫為實際汽車環境下以遠距離麥克風方式錄得, 如表二所示, 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，在五十公里語料中有 150 句測試語料，在九十公里語料中有 250 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 取樣頻率均為 8 k Hz ，以 16 bit 的方式儲存，連續數字長度為 3 至 11 個數字長。其中每種不同速度下的訊號雜訊比為：

表 一、不同環境語料之信號雜訊比

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

4.2 語音模型與參數設定

實驗主要是以連續密度隱藏式馬可夫模型(continuous density hidden Markov model, CDHMM)做為 baseline 系統的架構，在辨識噪音連續數字語料時，由於每個狀態所分配到的音框數不固定，所以不限定混合數數量，由程式自動產生，但最大混合數限制為 4 個。在連續數字 HMM 的定義中，每個數字模型固定為 7 個狀態，並且加入語句前後及中間三個靜音狀態，總狀態個數為 73 個。在語音特徵參數求取部份，每個特徵參數皆為 26 維度，其中包括了 12 階的 MFCC，12 階

的 delta MFCC ，1 階的 log energy 以及 1 階的 delta log energy。

4.3 實驗結果

在實驗的部分，我們分別討論 DBPC 及 DLRBPC 的實驗評估在汽車噪音語料庫下的方法評估。在此實驗，我們主要為評估在乾淨環境下所訓練的模型在 50 公里及 90 公里汽車噪下的辨識率，我們首先列出乾淨語料訓練的初始模型在未調適前對於噪音語料的辨識率如下：

表 二、測試環境對於模型調適前的辨識率

語料環境	50km	90km
辨識率	39.21%	32.89%

- **DBPC 之評估實驗**

在此實驗中，初始的語者獨立模型是來自 1000 句的乾淨語料，我們主要是針對環境不匹配下的調適。並且提供 5 至 15 句的調整語料，並且以 2 句為單位。在此實驗中，我們主要是以 MAP 調整及 BPC 的不確定性調整下和我們提出的以最小化分類錯誤法則更新不確定性的方法做比較，以下分別是五十公里與九十公里的汽車噪音環境下的辨識效果。

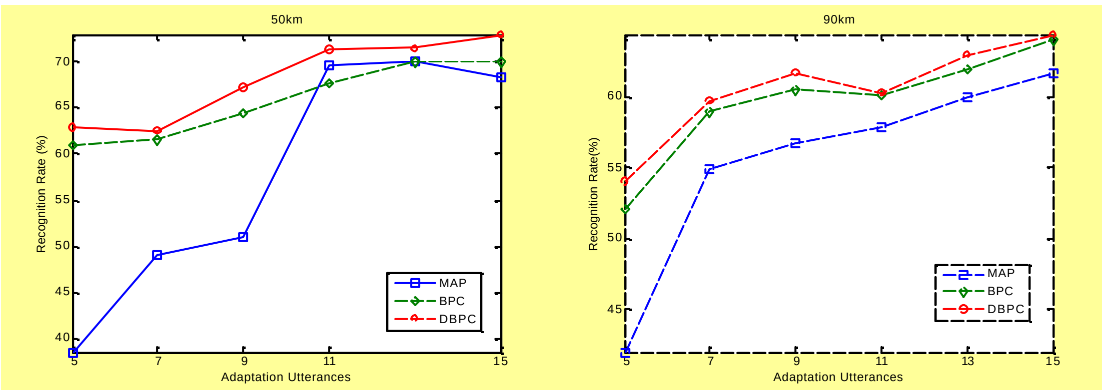

圖 四、DBPC 在五十公里與九十公里噪音語料調適結果

其中發現，雖然有考慮貝氏更新的 BPC 可以在調整資料量不足，以及環境不穩定下表現出相較於的最大事後機率調整的強健性，但是當進一步考慮鑑別性的更新不確定性時，則可以再些微的提升整體的辨識率。

- **DLRBPC 之評估實驗**

在此實驗中，我們首先以 1000 句的連續數字乾淨語料來訓練出一組語者獨立的模型，為了能更實驗在不匹配的環境中調適的效率，我們分別在 50 公里及 90 公里的噪音語料下做辨識，辨識前我們各取其環境下一至五句的語料當成是調適語料，在方法評估上，我們與 MLLR 及 MCELR 和我們提出的鑑別性迴歸貝氏預測分類器做比較。在我們方法中的事前資訊，是以汽車噪音中的 0 公里的 50 句調適語料分 10 位語者來估測出語者相依迴歸矩陣，然後取其樣本平均值及樣本共變異數當成初始的不確定性的事前資訊。以下是在五十公里與九十公里下的調適實驗結果：

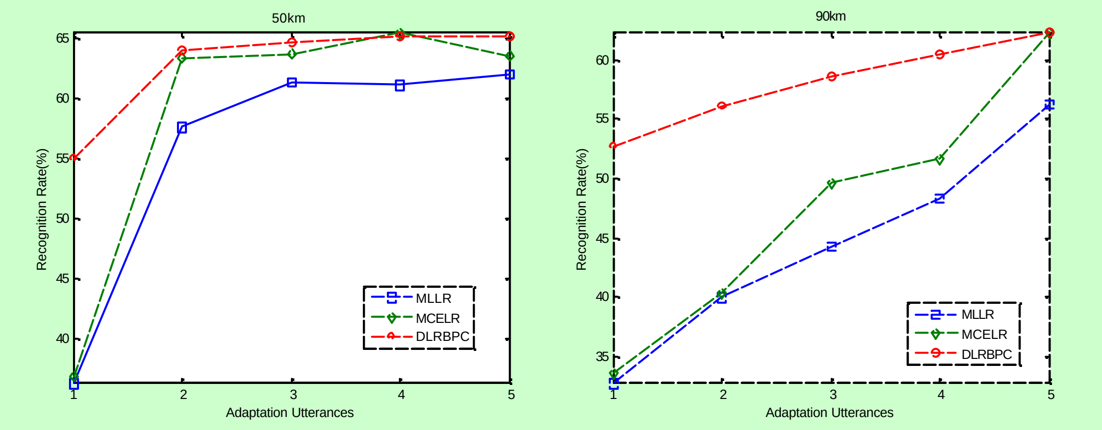

圖 五、DLRBPC 在五十公里與九十公里噪音語料調適結果

在此實驗中, 我們曾首先在乾淨語料下以語者為單位對語者獨立的模型估測各別語者的迴歸矩陣，然後取其樣本平均值及樣本共變異數為我們迴歸矩陣的事前不確定性，但是其辨識效果不如我們預期，其可能原因是在乾淨語料中求出的為反應出語者變異的迴歸矩陣事前資訊，不同於描述噪音環境變異下的不確定性，所以我們以汽車噪音中的 0 公里的 50 句調適語料來估測我們不確定性的事前資訊。此外在此實驗中，我們提出的方法，在資料量不足的情況下還能夠保有穩定的辨識水準，主要是在於不確定性的描述及更新能夠符合於測試環境中的不匹配情況且也能代表估測迴歸矩陣的事前資訊。在較複雜的九十公里語料中，因為噪音的問題比 50 公里下嚴重所以參數的不確定性影響將更加劇烈，同時在少量調整語句下，又會有估測錯誤問題，所以明顯的同時考慮鑑別性及不確定性的迴歸矩陣估測可以達成較穩定的辨識效果。

5. 結論與未來展望

在本研究當中，我們提出以鑑別性的法則更新語音模型中不確定事前資訊的評估方法，不同於傳統的鑑別性訓練與調適方法針對模型參數調整，我們考慮在模型參數的超參數來調整，希望在保有維持參數的不確定性的資訊同時，還能兼顧其具有鑑別性的特性，達到更穩鍵之語音辨識效能。在初步的實驗中，我們發現在貝氏預測分類法則中導入對不確定性的鑑別性更新，在噪音環境下，有其提升辨識率的空間，嚴格來說，如同於第二節所言，當初始的不確定性資訊能夠描述參數間不匹配的關係時，則鑑別性的貝氏分類器可以進一步提升辨識率。

在辨識的計算量上，因為我們是使用於 BPMC 的近似架構下，所以整體的辨識流程是融合於傳統的辨識器中，不會有額外的計算量負擔。但是在參數更新上，我們是以最小化分類錯誤為鑑別性法則的出發點，其參數更新的計算量是在求取梯度微分量上及疊代更新的計算，所以主要的計算差異上是在鑑別性的超參數更新上。

在未來研究上，是否可以考慮其他的鑑別性函數來分析不一樣的鑑別性函數下對於不確定性的影響，另一方面，在我們現行所實現的方法，主要是假設在初始環境與調整後的環境間的不確定性都是同為高斯分佈的假設下，但也許這個通道的不確定性會是其他的機率分佈，如何分析模型參數與環境不匹配的參數對合關係，也是可以再進一步分析探討的問題。

參考文獻

[1] J.-T. Chien, "Linear regression based Bayesian predictive classification for speech recognition," *IEEE Trans. Speech and Audio Processing,* vol. 11, pp. 70-79, July 2002.

[2] J.-T. Chien and G.-H. Liao "Transformation-based Bayesian predictive classification using online prior evolution," *IEEE Trans. Speech and Audio Processing*, vol. 9, pp. 399-410, May 2001.

[3] M. J. F. Gales and P. C. Woodland, "Mean and variance adaptation within the MLLR Framework," *Computer Speech and Language,* Vol. 10, pp. 249-264, 1996.

[4] J.-L. Gauvain and C-H. Lee, "Maximum *a posteriori* estimation for multivariate Gaussian mixture observation of Markov chains", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 291-298, April 1994.

[5] Q. Huo and C.-H. Lee, "Robust speech recognition based on adaptive classification and decision strategies," *Speech Communication*, vol. 34, pp. 175-194, 2001.

[6] Q. Huo and Chin-Hui Lee, "A Bayesian predictive classification approach to robust speech recognition," *IEEE Trans. Speech And Audio Processing*, vol.8, no.2, 2000.

[7] Q. Huo and C.-H. Lee, "A study of prior sensitivity for Bayesian predictive classification based robust speech recognition," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 2, 1998, pp. 741-744.

[8] Q. Huo and C.-H. Lee, "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate," *IEEE Trans. Speech Audio Processing,* vol. 5, pp. 161-172, Mar.1997.

[9] H, Jiang, X. Li and C Liu, "Large margin hidden Markov models for speech recognition," *To appear in IEEE Trans. Audio,Speech and Language Processing,* 2006.

[10] H. Jiang and Li Deng, "A robust compensation strategy for extraneous acoustic variations in spontaneous speech recognition," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 1, January 2002.

[11] H. Jiang and Li Deng, "A Bayesian approach to the verification problem: applications to speaker verification," *IEEE Trans. on Speech and Audio Processing,* vol. 9, vo. 8, pp. 874-884, 2001.

[12] 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, pp. 313-326, 1999.

[13] B.-H. Juang, W. Hou and C.-H. Lee, "Minimum classification error rate Methods for Speech Recognition", *IEEE Trans. Speech and Audio Processing,* vol. 5, no. 3 , pp. 257-265, May 1997.

[14] B.-H. Juang, and S. Katagiri, "Discriminative learning for minimum error classification", *IEEE Trans. Speech and Audio Processing*, vol. 40, no. 12 , pp. 3043-3054, Dec 1992.

[15] C. J. Leggeter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language*, pp. 171-185, 1995.

[16] Y. Normandin, R. Cardin and R. De Mori, "High-performance connected digit recognition using maximum mutual information estimation," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 299-311, 1994.