

行政院國家科學委員會補助  
大專學生參與專題研究計畫研究成果報告

\* \*\*\*\*\*  
\* 計 畫  
\* : 應用調變頻譜分析於樂器音色辨識之研究  
\* 名 稱  
\* \*\*\*\*\*

執行計畫學生： 古珮臻  
學生計畫編號： NSC 99-2815-C-216-007-E  
研究期間： 99年07月01日至100年02月28日止，計8個月  
指導教授： 李建興

處理方式： 本計畫涉及專利或其他智慧財產權，2年後可公開查詢

執行單位： 中華大學資訊工程學系

中華民國 100年03月31日

# 行政院國家科學委員會補助

## 大專學生參與專題研究計畫研究成果報告

\*\*\*\*\*

\* 計畫 \*

\* : 應用調變頻譜分析於樂器音色辨識之研究 \*

\* 名稱 \*

\*\*\*\*\*

執行計畫學生：古珮臻

學生計畫編號：NSC99-2815-C-216-007-E

研究期間：99年7月1日至100年2月底止，計8個月

指導教授：李建興博士

處理方式(請勾選)：立即公開查詢

涉及專利或其他智慧財產權，一年二年後可  
公開查詢

執行單位：中華大學資訊工程學系

中華民國100年3月31日

研究計畫內容：

## 一、摘要

本計劃的目的是透過圖型辨識的方法來分析樂器音色，擬使用調變頻譜分析 (Modulation Spectral Analysis) 來擷取樂器的音色特徵，再以一維主軸分析演算法 (Principal Component Analysis, PCA) 及一維線性分析演算法 (Linear Discriminant Analysis, LDA) 來降低特徵向量的數目並提高辨識率，最後再以分類樂器融合演算法來整合各種不同特徵向量的辨識結果以進一步提高辨識率。

## 二、研究動機與研究問題

現在的樂器種類繁多，形式多樣，各種樂器的音色聲響、表現手法皆獨具特色，隨著網際網路的發展與普遍化，數位音樂檔案越來越多，如何有效率的管理巨大的音樂資料庫，提供使用者搜尋感興趣的樂器音色音樂檔案，是一個很重要的課題，如果能將音樂之屬性先將音樂檔案區分為不同的類型，對於資料庫的管理有很大的幫助，使用者也能依據演奏樂器來瀏覽搜尋感興趣的音樂檔案或音樂片斷，對於這個問題，我們擬採用調變頻譜分析來擷取音樂特徵以辨識不同的樂器音色。

## 三、文獻回顧與探討

Marques 及 Moreno[1]以 LPC 係數，FFT 倒頻譜係數，及 MFCC 係數為特徵，並以 GMM 及 SVM 來辨識八種樂器：風笛、豎笛、橫笛、大鍵琴、風琴、鋼琴、伸縮喇叭及小提琴，正確率 70%。

Eronen[2]比較 LPC 倒頻譜係數、MFCC 係數及其微分值，及各種頻譜和時間域之特徵，以 KNN 來辨識各種樂器或樂器家族，其結果顯示 MFCC 的辨識率最高，其中小提琴及吉他是難辨識的樂器。

Peeter 等人[3]以 MPEG-7 的音訊特徵來分辨樂器的音色相似度，結果顯示將各

種 MPEG-7 的特徵組合起來可以用來評估不同樂器聲音的相似度。

Kaminskyj 及 Czaszejko[4]以 KNN 分類法來辨識獨立聲道的樂器聲音，採用的特徵包括 MFCC、常數 Q 轉換頻譜、振幅封包的平方平均值、頻譜質心及多維音階 MDS 分析軌跡，然後以 PCA 將特徵向量維度降為 710，其辨識率為 93%，而樂器家族的辨識率為 97%。

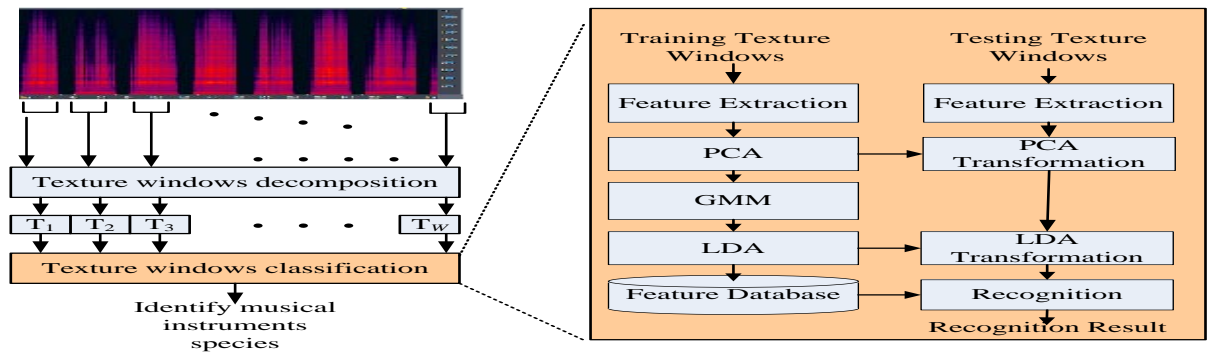
Essid 等人[5]以單一音節的訊號來辨識十種樂器，採用 MFCC、 $\Delta$ MFCC 及各種頻譜特徵(頻譜質心、頻譜頻寬、頻譜對稱性及頻譜平滑度)以 SVM 來辨識，正確率為 76%。

Livshin 及 Rodet[6]以單聲道之音節來辨識 7 種樂器：巴松管、豎笛、橫笛、吉他、鋼琴、小提琴及大提琴，總共計 62 種不同特徵(包括時間域特徵、能量特徵、頻譜特徵、泛音特徵及聽覺特徵等)，然後以漸進式特徵刪除法(GDE)，來選取最佳的 20 個特徵值，再以 LDA 降低特徵維度及以 KNN 來辨識，正確率為 85%，若全部 62 個特徵都使用，其辨識率為 88%。

Benetos 等人[7]擷取 MFCC、MPEG-7 及其他頻譜特徵以辨識 6 種樂器：鋼琴、巴松管、大提琴、橫笛、薩克斯風及小提琴，然後以 branch-and-bound 搜尋法來選取最佳的六種特徵，並以 NMF 來辨識，正確率為 95.2%。

#### 四、研究方法及步驟

在我們所提出之樂器音色辨識系統中，我們先將輸入訊號切成許多紋理視窗來擷取出特徵資料，其紋理視窗切割是以固定時間長度方式。在訓練部分包含有特徵擷取、主軸分析演算法(Principal Component Analysis, PCA)、高斯混合模型分群(Gaussian Mixture Model, GMM)，線性區別分析(Linear Discriminant Analysis, LDA)轉換並儲存分群結果的群中心。而測試部份包含了特徵擷取、主軸分析轉換、線性區別分析轉換和辨識，之後針對紋理視窗做辨識結果的融合來對聲音訊號辨識，圖一為樂器音色辨識流程圖。



圖一 樂器音色辨識系統流程圖

### (一) 紋理視窗(Texture Window)

系統輸入一段樂器聲音訊號後需要對聲音訊號切出許多固定時間紋理視窗，由於樂器有多種彈奏方法擁有一個以上的不同變化，因此將聲音訊號切出許多較少時間紋理視窗，可以取得不同變化特性的特徵。切出數個紋理視窗後，再針對每個視窗取一個個固定長度的音框，然後再針對每個視窗的音框去取得各種特徵。

### (二) 特徵擷取

#### 步驟 1: 取音框(Framing)

將每一個紋理視窗切割成一個一個的音框，大小為 512 訊號樣本，相鄰音框

重疊一半。

#### 步驟 2: 對 Frame 作 DCT

#### 步驟 3: 快速傅立葉轉換(FFT)

將音訊訊號從時域轉換成頻率域

$$X[k] = \sum_{n=0}^{N-1} \tilde{s}[n] e^{-j2\pi \frac{k}{N} n}, \quad 0 \leq k < N$$

其中  $N$  為音框大小， $\tilde{s}[n]$  為離散訊號。

步驟 4: 先找到最大能量之音框並以其為中心往左 127 個音框，往右 128 個音框

步驟 5:  $X[k] = \log(1 + |(k)|)$

步驟 6: 對 Frame 作 DCT

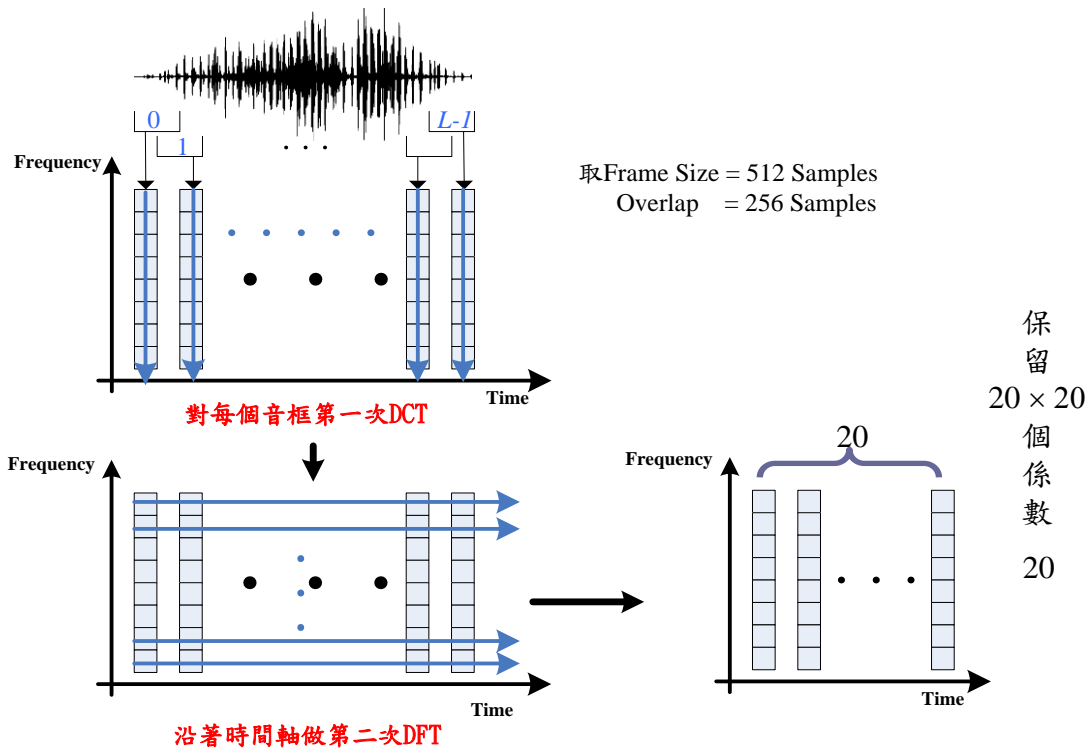
步驟 7: 沿著時間軸作 FFT

快速傅立葉轉換(FFT)

將音訊訊號從時域轉換成頻率域

$$X[k] = \sum_{n=0}^{N-1} \tilde{s}[n] e^{-j2\pi \frac{k}{N} n}, \quad 0 \leq k < N$$

其中  $N$  為音框大小， $\tilde{s}[n]$  為離散訊號。



圖二 樂器音色辨識系統流程圖

### (三) 正規化

令  $x(m)$  為取得不同樂器的所有特徵向量，因為代表不同樂器音色的特徵向量裡的各個特徵值，其數值範圍是不同的，所以將所有特徵值利用正規化在 0 至 1

之間來解決這個問題，計算公式如下：

$$\mathbf{x}'(m) = \begin{cases} 0 & F(m) \leq Q_3(m) \\ \frac{F(m) - Q_3(m)}{Q_{97}(m) - Q_3(m)}, & Q_3(m) < F(m) < Q_{97}(m) \\ 1 & F(m) \geq Q_{97}(m) \end{cases}$$

其中， $\mathbf{x}'(m)$  為正規化後之特徵向量， $Q_3(m)$ 和  $Q_{97}(m)$ 分別代表將第  $m$  個維度所有訓練資料特徵值由小到大排序後所取得的第 97%與第 3%的特徵值。在這裡會將  $Q_3(m)$ 和  $Q_{97}(m)$ 保留下來用於在辨識部分時對測試資料所取得的特徵值做正規劃所用。

#### (四) 主軸分析法 (Principle component analysis, PCA)

主軸分析演算法之主要目的是降低特徵向量之維度，但是降低特徵向量之維度會損失部分資訊，所以我們要如何降低維度後還能保持最大之資訊量，因而不影響辨識之結果，甚至是刪除那些降低辨識率的特徵，而使得辨識率上升，這個問題是 PCA 所要解決的主要課題。

PCA 是先計算所有訓練資料之特徵向量的平均變異數矩陣之 eigenvalue 及 eigenvector，並以 eigenvector 當作基底來做線性轉換，而 eigenvalue 的大小可以決定其對應之 eigenvector 轉換後之特徵所保留之資訊量大小，eigenvalue 越大表示資料作線性轉換後，特徵的變異數值會越大，而變異數的大小又表示分佈的寬廣，資料分佈越廣表示所保留之資訊量越大，也就是說，以 eigenvalue 值較大之 eigenvector 做為線性轉換之基底，轉換後的特徵分佈範圍會比以 eigenvector 較小的 eigenvector 轉換後的分佈範圍來得大。PCA 之進行步驟如下：

##### 步驟 1: 計算平均向量

$$\mathbf{m} = E[X]$$

其中  $X$  是所有訓練資料之集合， $X = \{\mathbf{x}_i | i = 0, \dots, N\}$ ， $\mathbf{m}$  是所有訓練

資料的平均向量， $N$ 是訓練資料的數量。

步驟 2: 令平均向量為 0

$$\mathbf{x}'_i = \mathbf{x}_i - \mathbf{m}$$

求取平均變異數矩陣,  $\mathbf{C}$

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}'_i (\mathbf{x}'_i)^T$$

求取變異數矩陣  $\mathbf{C}$  的 eigenvalue 及 eigenvector 並將其依 eigenvalue 值由

大至小重新排序

設定臨界值  $T$  (表示所要保留的資訊量程度), 以計算轉換後維度  $d$

$$\sum_{i=1}^d \lambda_i \geq T \times \sum_{i=1}^D \lambda_i$$

其中  $i$  表示第  $i$  大之 eigenvalue,  $D$  為轉換前之維度以所保留之  $d$  個 eigenvector

對所有資料作線性轉換

$$\mathbf{x}_{\text{PCA}} = \mathbf{A}_{\text{PCA}}^T \mathbf{x}'_i$$

其中  $\mathbf{A}_{\text{PCA}}$  為此  $d$  個較大 eigenvector 構成之轉換矩陣。

### (五) 線性區別分析演算法 (Linear Discriminant Analysis, LDA)

線性區別分析演算法之目的是將一個高維度的特徵向量轉換成一個低維度的向量, 並且增加辨識的準確率, 線性區別分析主要處理不同類別間的區別程度而不是用於不同類別之表示方式。線性區別分析演算法的主要精神是要把同類之間的距離最小化, 並且把不同類別之間的距離給最大化, 所以, 必需決定一個轉換矩陣來將維度  $n$  的特徵向量轉換成維度  $d$  的向量, 在這裡  $d \leq n$ , 透過這樣的轉換我們能夠增強不同類別之間的差異性。最常使用的轉換矩陣主要依據 Fisher criterion  $J_F$  來求得:

$$J_F(A) = \text{tr}((A^T S_W A)^{-1} (A^T S_B A))$$



其中， $S_W$ 和  $S_B$ 分別代表的是同類別之散佈矩陣( within-class scatter matrix )和不同類別之散佈矩陣( between-class scatter matrix) ，而同類別之散佈矩陣的公式如下：

$$S_W = \sum_{j=1}^C \sum_{i=1}^{n_j} (\mathbf{x}_i^j - \boldsymbol{\mu}_j)(\mathbf{x}_i^j - \boldsymbol{\mu}_j)^T$$

而  $\mathbf{x}_i^j$  代表在類別  $j$  中的第  $i$  個特徵向量， $\boldsymbol{\mu}_j$  為第  $j$  類的平均向量(mean vector)， $C$  為類別的數目， $n_j$  為類別  $j$  裡的特徵向量個數。而不同類別之散佈矩陣公式如下：

$$S_B = \sum_{j=1}^C n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T$$

其中  $\boldsymbol{\mu}$  為所有類別的平均向量。線性區別分析演算法的目的是要去求出能夠使不同類別之散佈矩陣和同類別之散佈矩陣的比值為最大值轉換矩陣  $\mathbf{A}_{LDA}$  ，而其維度大小為  $n \times d$ ：

$$\mathbf{A}_{LDA} = \arg \max_A \frac{\text{tr}(\mathbf{A}^T S_B \mathbf{A})}{\text{tr}(\mathbf{A}^T S_W \mathbf{A})}$$

此一轉換矩陣，可經由求出  $S_W^{-1} S_B$  的特徵向量(eigenvectors)來得到，而  $\mathbf{A}_{LDA}$  之  $d$  個行向量為前  $d$  個最大特徵值(eigenvalue)值所對應之特徵向量。在此我們是取保留 98% 資訊量作為  $d$  值。

在我們決定出最佳的轉換矩陣  $\mathbf{A}_{LDA}$  後，我們以  $\mathbf{A}_{LDA}$  將每一正規化(normalized)後之  $n$  維的特徵向量轉換為  $d$  維之向量。令  $\mathbf{f}_j$  為類別  $j$  裡維度為  $n$  的特徵向量，轉換成維度為  $d$  的向量之公式如下：

$$\mathbf{x}_j = \mathbf{A}_{LDA}^T \mathbf{f}_j$$

## (六)高斯混合模型 (Gaussian Mixture Model, GMM)

由於樂器音色相當豐富多變化，因此就算有兩個紋理視窗是從同一種樂器聲音中所切割出來的，所擷取出來的特徵向量也可能會有明顯的不同，所以對於每一種

樂器聲音，我們將使用高斯混合模型(Gaussian mixture model, GMM)來描述，也就是說屬於同一種樂器聲音之不同紋理視窗可以分成幾個小群(高斯分佈)，而屬於同一小群之不同紋理視窗其特徵向量會較相似。

傳統上，使用高斯混合模型於分類辨識，對不同類別之資料需分別建立其高斯混合模型，一般傳統上高斯混合模型之參數預測是使用 EM (Expectation Maximization)演算法。此演算法主要是用來預測高斯混合模型中多變數機率分佈函數之參數值，其目的是要找到最佳之參數 $\Theta$ 使得  $p(X | \Theta)$ 最大，其中  $X = \{\mathbf{x}(t), t = 1, 2, \dots, N\}$  為訓練資料之特徵向量集合， $N$  為訓練資料之數目； $\Theta \equiv \{ p(\theta_r), \mu_r, \Sigma_r | r = 1, 2, \dots, M \}$ ， $p(\theta_r)$ 為在高斯混合模型中第  $r$  個高斯分佈之事前機率(Prior Probability)， $\mu_r$  為平均值向量， $\Sigma_r$  為共變異數矩陣(Covariance matrix)， $M$  為高斯混合模型中高斯分佈之群數。EM 演算法之詳細步驟如下：

#### 步驟 1: 執行 K-Means 演算法

首先，依據高斯混合模型中所指定之高斯分佈群數執行 K-Means 演算法分群，以每群之平均值向量作為每個高斯分佈之平均值向量之初始值，且將共變異數矩陣之初始值設為單位矩陣。

#### 步驟 2: Expectation-Step

對所有資料計算其屬於高斯混合模型中每一高斯分佈之機率比值作為預測值，其公式如下：

$$p(\theta_r | \mathbf{x}(t)) = \frac{p(\theta_r) p(\mathbf{x}(t) | \theta_r)}{\sum_{r=1}^M p(\theta_r) p(\mathbf{x}(t) | \theta_r)}$$

其中

$$p(\mathbf{x}(t) | \theta_r) = \frac{1}{\sqrt{(2\pi)^d \Sigma_r}} \exp\left(-\frac{(\mathbf{x}(t) - \boldsymbol{\mu}_r)^T \Sigma_r^{-1} (\mathbf{x}(t) - \boldsymbol{\mu}_r)}{2}\right)$$

在這裡  $d$  為特徵向量之維度。

### 步驟 3: Maximization-Step

利用步驟 2 所計算之預測值，更新預測之參數值：

權重值：

$$p(\bar{\theta}_r) = \frac{1}{N} \sum_{t=1}^N p(\theta_r | \mathbf{x}(t))$$

平均值向量：

$$\bar{\boldsymbol{\mu}}_r = \frac{\sum_{t=1}^N p(\theta_r | \mathbf{x}(t)) \mathbf{x}(t)}{\sum_{t=1}^N p(\theta_r | \mathbf{x}(t))}$$

共變異數矩陣：

$$\bar{\Sigma}_r = \frac{\sum_{t=1}^N p(\theta_r | \mathbf{x}(t)) (\mathbf{x}(t) - \bar{\boldsymbol{\mu}}_r) (\mathbf{x}(t) - \bar{\boldsymbol{\mu}}_r)^T}{\sum_{t=1}^N p(\theta_r | \mathbf{x}(t))}$$

重覆執行步驟 2~3，直到收斂為止。

## (七) 分類器融合演算法(Classifiers fusion approach)

由於聲音訊號的變化相當多而且豐富，對於不同類別聲音的特性的差異性很難明確的描述出來。該選擇哪一個分類器的辨識結果才能得到最佳辨識效能，或如何整合數個分類器的辨識結果以提高辨識率，是我們的研究方向之一。我們採用分類器的辨識結果的主要方法有投票法。

### (1) 投票法(Majority Voting)

對於同一輸入聲音訊號，每個分類器辨識結果有可能不同，一般而言，若多數分類器

的辨識結果都相同，則其可信度較高，因此投票法觀念以最多相同樂器的辨識結果為

最終的辨識結果。

## 五、實驗結果

在實驗中所使用之樂器音色資料，總共有 10 種樂器如(表一)，取樣頻率為 44100 Hz，其訓練資料之樂器錄音與測試資料之樂器錄音皆為在不同人下使用不同演奏方法之錄音。

我們分別使用不同頻率之樂器聲資料來做訓練與測試，其中訓練資料有 130 及測試資料有 259 個檔案，在表一中顯示每種樂器聲音所切割之音段與紋理視窗的個數。接著對每一紋理視窗取特徵，分類正確率( $CA_{TW}$ )計算的方式定義為：

$$CA_{TW} = \frac{N_{CA}}{N_{TW}} \times 100$$

其中  $N_{CA}$  為辨識正確的紋理視窗個數， $N_{TW}$  為總共紋理視窗個數。

我們在實驗過程我們使用各種高斯混合模型分群數來比較各種辨識結果，可以看到在紋理視窗的辨識結果，表二為 GMM=5，表三為 GMM=10，表四為 GMM=15，表五為 GMM=20，表六為 GMM=25，表七為 GMM=30，表八為 GMM=35，表九為 GMM=40，表十為 GMM=45，表十一為 GMM=50，表十二 紋理視窗最佳分類結果，最佳分類正確率為 73.13%。

表一 樂器各個文理視窗個數

Musical instruments	Training Data	Testing Data	Files
Name	$N_{TW}$	$N_{TW}$	Number
大提琴	753	1508	24
小提琴	912	1728	27
小號	633	1262	36
中提琴	719	1376	24
長笛	355	708	16
喇叭	659	1305	42
單簧管	477	956	24
電吉他	858	1793	24
鋼琴	1056	2024	24
薩克斯	294	588	18

表二 GMM=5 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	68.89%	68.89%	68.89%	69.12%	69.18%
0.96	68.96%	69.01%	69.16%	69.13%	69.47%
0.97	69.07%	69.06%	69.13%	69.01%	69.24%
0.98	69.25%	69.17%	69.35%	69.34%	69.32%
0.99	69.58%	69.99%	70.03%	70.06%	70.00%

表三 GMM=10 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	71.18%	71.47%	71.51%	71.25%	71.41%
0.96	71.04%	71.16%	71.14%	71.16%	71.18%
0.97	70.87%	71.20%	71.29%	71.10%	71.17%
0.98	71.01%	71.23%	71.42%	71.34%	71.52%
0.99	71.57%	71.69%	71.40%	71.35%	71.69%

表四 GMM=15 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	71.91%	72.03%	71.93%	72.12%	72.08%
0.96	71.98%	72.03%	71.90%	72.11%	72.21%
0.97	72.46%	72.52%	72.42%	72.37%	72.77%
0.98	72.52%	72.41%	72.61%	72.80%	72.61%
0.99	73.20%	73.09%	72.86%	72.79%	73.13%

表五 GMM=20 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	72.24%	72.04%	71.97%	71.77%	71.70%
0.96	70.65%	70.43%	70.46%	70.42%	70.18%
0.97	72.25%	72.39%	72.55%	72.30%	72.21%
0.98	70.95%	70.95%	70.86%	70.64%	70.55%
0.99	71.33%	71.07%	70.71%	70.45%	70.54%

表六 GMM=25 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	71.50%	71.43%	71.26%	71.23%	71.07%
0.96	72.86%	72.53%	72.36%	72.36%	72.38%
0.97	70.73%	70.55%	72.39%	70.40%	70.16%
0.98	70.90%	70.50%	70.46%	70.00%	69.83%
0.99	70.95%	70.97%	70.69%	70.51%	70.22%

表七 GMM=30 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	72.06%	72.20%	71.64%	71.88%	71.62%
0.96	72.00%	71.77%	71.81%	71.78%	71.70%
0.97	70.61%	70.36%	70.12%	70.00%	69.72%
0.98	70.12%	69.81%	69.81%	69.67%	69.29%
0.99	70.92%	70.72%	70.14%	69.99%	70.03%

表八 GMM=35 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	72.56%	72.54%	72.53%	72.31%	72.21%
0.96	69.46%	69.39%	69.29%	69.22%	68.47%
0.97	71.87%	71.53%	71.47%	71.48%	71.38%
0.98	71.03%	71.14%	70.63%	70.39%	70.07%
0.99	66.15%	65.78%	65.60%	65.29%	64.58%

表九 GMM=40 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	68.35%	68.04%	67.83%	67.70%	67.41%
0.96	68.83%	68.48%	68.09%	67.52%	67.41%
0.97	68.63%	68.35%	68.18%	67.93%	67.38%
0.98	68.33%	67.89%	67.71%	67.31%	66.98%
0.99	68.42%	68.27%	68.00%	67.79%	67.38%

表十 GMM=45 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	64.27%	63.66%	63.18%	63.07%	62.98%
0.96	64.87%	64.70%	64.05%	63.91%	63.80%
0.97	64.67%	64.22%	63.69%	63.30%	63.07%
0.98	65.22%	64.83%	64.55%	64.30%	64.10%
0.99	65.97%	65.71%	65.50%	65.22%	64.79%

表十一 GMM=50 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	67.58%	67.36%	67.10%	66.85%	66.52%
0.96	70.18%	70.17%	69.95%	70.00%	69.64%
0.97	70.01%	69.86%	69.72%	69.57%	69.80%
0.98	64.76%	64.72%	64.27%	63.99%	63.53%
0.99	67.33%	67.23%	67.28%	66.70%	66.45%

表十二 根據 Testing Data 紋理視窗最佳分類結果

PCA=0.99,GMM=15,LDA=0.99

	大提琴	小提琴	小號	中提琴	長笛	喇叭	單簧管	電吉他	鋼琴	薩克斯
大提琴	1071	127	30	199	9	8	27	8	13	16
小提琴	63	1251	20	267	32	16	48	2	17	12
小號	23	37	807	49	51	177	49	0	46	23
中提琴	129	220	36	874	46	25	23	3	8	12
長笛	7	22	35	24	517	50	20	0	9	24
喇叭	16	6	95	18	32	1033	24	9	38	34
單簧管	23	72	62	34	26	49	580	1	72	37
電吉他	56	6	3	9	0	69	35	1541	74	0
鋼琴	51	34	54	43	12	44	13	31	1736	6
薩克斯	30	22	57	49	45	50	36	7	14	278

對於同一輸入聲音訊號，我們採用投票法來辨識，投票法觀念以最多相同樂器的辨識結果為最終的辨識結果，表十三為GMM=5，表十四為GMM=10，表十五為GMM=15，表十六為GMM=20，表十七為GMM=25，表十八為GMM=30，表十九為GMM=35，表二十為GMM=40，表二十一為GMM=45，表二十二為GMM=50，表二十三 紋理視窗最佳分類結果，最佳分類正確率為73.13%。

表十三 GMM=5 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	89.96%	90.35%	89.58%	89.58%	89.58%
0.96	88.80%	90.73%	90.73%	89.96%	90.35%
0.97	88.80%	90.73%	89.96%	90.73%	90.73%
0.98	89.19%	89.19%	89.58%	90.73%	90.73%
0.99	90.73%	91.51%	90.73%	90.73%	90.35%

表十四 GMM=10 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	90.35%	91.51%	92.28%	91.89%	92.28%
0.96	91.51%	91.51%	91.89%	90.73%	92.28%
0.97	91.51%	90.35%	90.73%	91.12%	91.89%
0.98	89.96%	88.80%	89.58%	90.73%	89.96%
0.99	90.73%	91.51%	90.73%	91.89%	91.12%

表十五 GMM=15 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	93.44%	93.05%	93.44%	93.44%	93.05%
0.96	92.66%	93.05%	93.82%	93.82%	92.66%
0.97	<b>94.21%</b>	93.82%	93.82%	92.66%	93.05%
0.98	91.51%	90.73%	91.89%	91.51%	91.89%
0.99	92.28%	91.89%	92.28%	92.66%	92.66%

表十六 GMM=20 之實驗結果

LDA PCA	0.95	0.96	0.97	0.98	0.99
0.95	91.89%	89.96%	90.35%	90.73%	89.58%
0.96	91.89%	92.28%	93.05%	93.05%	93.44%
0.97	91.12%	91.51%	92.28%	91.51%	91.89%
0.98	93.05%	91.51%	91.89%	90.73%	89.58%
0.99	91.51%	91.12%	91.51%	91.12%	90.35%



表十七 GMM=25 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	91.12%	91.51%	91.12%	90.35%	91.51%
0.96	91.51%	91.12%	90.73%	90.35%	91.12%
0.97	91.12%	91.51%	89.96%	90.73%	90.35%
0.98	91.12%	91.12%	90.35%	91.12%	90.35%
0.99	91.12%	91.12%	90.35%	90.35%	89.96%

表十八 GMM=30 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	89.58%	90.35%	90.73%	89.96%	91.12%
0.96	90.73%	90.35%	89.58%	89.58%	91.51%
0.97	90.73%	90.73%	89.96%	89.96%	88.80%
0.98	88.80	88.80	89.58%	89.19%	88.80
0.99	90.73%	89.96%	89.19%	89.58%	88.80%

表十九 GMM=35 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	90.73%	91.12%	90.35%	89.96%	89.96%
0.96	88.80%	87.64%	88.03%	87.64%	88.03%
0.97	89.58%	89.58%	88.80%	88.42%	89.58%
0.98	90.73%	91.51%	89.96%	89.19%	88.80%
0.99	85.71%	86.10%	84.94%	85.33%	83.78%

表二十 GMM=40 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	89.19%	88.42%	87.64%	87.64%	84.56%
0.96	86.87%	87.64%	88.42%	88.03%	87.26%
0.97	89.19%	87.64%	87.64%	87.26%	87.64%
0.98	87.26%	87.64%	86.87%	86.87%	84.94%
0.99	87.26%	86.49%	84.94%	83.78%	83.01%

表二十一 GMM=45 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	85.71%	84.94%	82.24%	83.01%	83.01%
0.96	83.40%	84.94%	83.78%	81.47%	79.92%
0.97	82.63%	81.85%	80.69%	81.08%	79.54%
0.98	84.17%	82.63%	82.63%	80.31%	80.69%
0.99	82.24%	83.01%	81.08%	80.69%	80.69%

表二十二 GMM=50 之實驗結果

LDA \ PCA	0.95	0.96	0.97	0.98	0.99
0.95	86.87%	87.64%	87.26%	85.71%	86.49%
0.96	89.58%	88.80%	89.58%	89.19%	87.26%
0.97	87.64%	87.64%	87.26%	86.87%	86.49%
0.98	82.24%	81.85%	82.24%	83.01%	80.69%
0.99	85.71%	86.10%	85.33%	84.94%	84.17%

表二十三最佳分類結果平均正確率為 94.21%

PCA=0.97, GMM=15, LDA=0.95

Musical instruments Name	Testing Number	Correct	Error	正確率
大提琴	24	24	0	100.00%
小提琴	27	26	1	96.30%
小號	36	35	1	97.22%
中提琴	24	22	2	91.67%
長笛	16	16	0	100.00%
喇叭	42	41	1	97.62%
單簧管	24	22	2	91.67%
電吉他	24	21	3	87.50%
鋼琴	24	22	2	91.67%
薩克斯	18	15	3	83.33%

## 六、參考文獻

- [1] J. Marques and P. Moreno, "A study a of musical instrument classification using Gaussian mixture models and support vector machines", Compaq Comput. Corp., Tech. Rep. CRL99/4, 1999.
- [2] A. Eronen, "Comparison of features for musical instrument recognition", in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust., 2001, pp. 19-22.
- [3] G. Peeters, S. McAdams, and P. Herrera, "Instrument sound description in the context of MPEG-7", in Proc. Int. Comput. Music Conf., 2000, pp. 166-169.
- [4] I. Kaminskyj and T. Czaszejko, "Automatic recognition of isolated monophonic musical instrument sounds using kNNC", J. Intell. Inf. Syst., Vol. 24, No. 2/3, pp. 199-221, Mar. 2005.
- [5] S. Essid, G. Richard, and B. David, "Efficient musical instrument recognition on solo performance music using basic features", in Proc. Audio Engineering Society 25th Int. Conf., London, U.K., 2004.
- [6] A. A. Livshin and X. Rodet, "Musical instrument identification in continuous recordings", in Proc. 7th Int. Conf. Digital Audio Effects, 2004, pp. 222-226.
- [7] E. Benetos, M. Kotti, and C. Kotropoulos, "Musical instrument classification using non-negative matrix factorizatoion algorithms and subset feature selection", in Proc. ICASSP, 2006, Vol. V, pp.221-224.