

# 使用特徵歌曲對網際網路上的 MP3 音樂物件進行自動化內涵式分類

## Automatic Content-Based Classification of MP3 Music Objects Using Eigensong on the Internet

蔡易行

中華大學資訊工程學系

m8802511@chu.edu.tw

劉志俊

中華大學資訊工程學系

ccliu@chu.edu.tw

### 摘要

MP3 已經成為網際網上最重要的多媒體音樂檔案格式，而目前對 MP3 音樂物件自動化分類的研究非常少。在本篇論文中我們對網路上大量 MP3 音樂物件進行自動化的內涵式分類，我們提出一種利用特徵歌曲 (eigensong) 來進行網路上 MP3 音樂自動分類的演算法，將 MP3 音樂物件分類為三類：古典音樂、男歌手流行歌曲及女歌手流行歌曲。我們並將此演算法應用在我們的網路上的 MP3 音樂資料庫中，其正確率介於 72% ~ 91%。

### 1. 簡介

由於網際網路的快速興起、儲存設備的容量增加以及傳輸、壓縮技術的逐漸成熟，媒體的種類不再侷限於文字，各種能夠表現資料的媒體，如聲音、影像、圖形、動畫等也逐漸地在網路上流傳。這些多媒體物件使得網際網路變得多元化且具有豐富性，衍生出的相關性問題就是多媒體資料的檔案大小是否影響到網路效能以及使用者對接收到的多媒體物件的接受程度，尤其是網際網路上流傳著數以千計的 MP3 音樂，如果將它們全都下載，豈不浪費了網路的資源，並違反了網路的意義，同時我們又不可能聽過所有的 MP3 音樂，單憑歌名是難以判斷出 MP3 音樂的類型，若能使用 MP3 音樂本身的內涵來找出此 MP3 音樂所代表的涵義 (human feeling)，就可以節省許多的找尋時間且能找出最符合使用者須求的 MP3 音樂。面對如此龐大及繁複的資料，除了傳統文字的敘述外，我

們應該可以使用更好的方法去描述以及整理這些多媒體的資料，因為對於多媒體類型的原始檔案，我們很難以傳統的文字搜尋法去做搜尋及比對(諸如現存於網路上的 Yahoo、Kimo 等傳統文數字搜尋引擎)，更難以在原始檔案中找到分類的類別，但是我們可以使用如 Yahoo 上的音樂分類類別當成我們的 MP3 音樂分類目錄。為了能夠更忠實地描述各種媒體的特性，我們針對媒體的內容，使用內涵式的分析，做內涵式特徵值的擷取、辨識，可以適當地對多媒體檔案做更貼切的分類。因此內涵式查詢方法已經廣泛的應用在各種多媒體檔案類型上，如 wave 音樂檔、bmp 圖形檔、MPEG 視訊等等。

多媒體資料的分類除了傳統的文字分類法外，我們可以利用人類知覺的特性來當做分類的依據。想要模擬人類的知覺，就必須對媒體的內涵做分析，建立一個分析辨識的系統，將人類的知覺表現在多媒體資料的特徵值上，然後利用特徵值的變化來當做知覺分類的依據。事實上，針對不同類型的多媒體檔案，例如\*.wav \*.mid...等類型檔案，近年來已有一些專家學者開始進行研究。在這些研究中，我們可以發現到，對於 MP3 音樂檔案類型的內涵式搜尋及分類的研究十分罕見。對於 MP3 歌曲這種新興的多媒體檔案，實在有著研究的必要性。目前對於 MP3 歌曲的分類還停留在人工分類的階段，面對龐大的 MP3 音樂物件，我們勢必尋求一種自動分類的方法來幫助我們完成分類的工作。因此在本篇論文中，我們將針對 MP3 音樂物件的內涵及人類的知覺特性，產生一個 MP3 自動分類的工具，來達成自動分類的目的。

一般而言，多媒體檔案可以區分成二種型態：影像及音樂。在影像方面，IBM Almaden Research

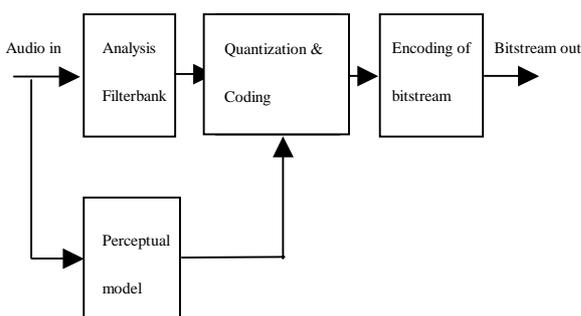
Center, 提出了 QBIC system[7], 這是第一個內涵式的影像資料庫, 主要的觀念是以色彩長條圖 (Color Histogram)、形狀 (Shape) 與紋理 (Texture) 等影像特徵值 (Content Features) 為索引。這個觀念與以往不同在於傳統的查詢是以文字 (Text) 的方式來執行查詢。Carlo Colombo 等人[2]提出以接近人類語意的視覺化方式來查詢, 使用者可以使用人類知覺 (perceptual) 的語意來尋找出符合的圖片。Kyoji Hirata 等人[6]提出將影像分類成多目錄 (multiple categories), 我們由此影像物件產生顏色、形狀等的多重索引指標, 再利用這些索引指標來自動分類。Matthew[1]將臉部影像轉換到特徵值影像, 定義成特徵臉 (eigenface), 例如一個大小為  $M \times M$  的臉部影像, 轉換成  $M \times M$  的矩陣, 此矩陣稱為臉部空間 (特徵臉)。計算特徵臉的步驟如下(1)算出所有臉部影像的平均集合 (2)算出各個影臉部影像與平均集合的差值 (3)得到協方差矩陣(4)計算協方差矩陣的特徵值及特徵向量。使用特徵臉作分類的步驟如下:(1)求出新影像的特徵臉 (2)算出它的權重向量(3)使用最小歐基里德距離算出新影像與預先定義之臉部影像的相似程度。此分類方法最大的好處就是能夠自動計算臉部特徵值的權重。在音樂資料庫方面, [3]提出將一個音樂物件的節奏(rhythm)、旋律(melody)、以及和弦(chords), 當成這個音樂物件的特徵字串, 而且發展出一個稱為 1D-List 的資料結構, 有效率的去與音樂接近字串的比對(approximate string matching)。在接近字串比對演算法(approximate string matching algorithm)中所做的相似度的測量, 是根據音樂理論來設計的。對於 MP3 的內容值查詢而言, 我們不可能輸入整首歌來與資料庫查詢比對, 或是以資料庫中 MP3 的整首歌曲來比對, 我們要將音樂做切割, 分成一段一段的音樂片段來做比對, 這樣才不會花費太多的計算時間。Jonathan Foote[5]中提出了一個 MMI (Maximise Mutual Information) 的方法, 這個方法可以自動的對聲音 (audio) 做分類, 它將聲音做加工, 產生了 13 個的特徵向量

(feature-vector), 將這些特徵向量轉換成樹狀結構, 然後將同一類型的聲音特徵累積成長條圖 (Histogram), 每次要做比對時, 將未知的資料做成長條圖, 然後去跟資料庫的聲音比對, 也就是說去計算兩者之間長條圖的距離, 以這樣的方法可以識別出不同聲音種類。Yasuyuki[13]使用聲音的能量分佈 (energy distribution) 及音量大小 (loudness) 的特徵值來判斷有聲音 (non-silence) 及靜音 (silence) 的分類。他們也提出使用時間密度 (temporal density)、頻寬 (bandwidth) 及中心頻率 (center frequency of subband energy) 的特徵值來做音樂、講話及掌聲的分類。Magne[8]使用模糊理論 (Fuzzy theory) 為基礎, 將分類分為二個步驟: 第一步: 使用 FCM (fuzzy c-means) 探勘出特徵值空間 (feature space) 的結構, 並且將它做特徵值分割 (feature partition)。第二步: 在全體的群集 (cluster membership) 與全體的類別 (class membership) 中解碼出其邏輯關係 (logical relational), 而以此建立模糊關係 (fuzzy relation)。Tong[12]將聲音的分類及查詢分成三個階層: 第一層: 粗略分類層 (coarse-level) - 使用物理特徵: 聲音能量函數 (energy function)、平均越零率 (average ZCR) 及聲音基本頻率組合 (fundamental frequency), 可分類出聲音、音樂、周圍環境聲音及靜音。第二層: 精細分類層 (fine-level) - 使用感官 (perceptual) 特徵: 音色 (Timbre) 及節奏 (rhythm), 利用 HMM 模組將周圍環境聲音細分, 如雨聲、掌聲等。第三層: 直接輸入聲音查詢 - 使用 HMM。目前對 MP3 歌曲音樂檔案類型做內涵式的分析研究非常有限, 在這篇論文中我們提出一種名為特徵歌曲 (eigensong) 的 MP3 音樂內涵式分類方法, 此種方法是以線性代數為基礎, 計算每種不同類型 MP3 歌曲音樂的特徵歌曲, 找出不同類型 MP3 在空間向量中分佈的情形, 進而達到分類的效果。特徵歌曲的分類法與其它的分類法如類神經網路、人工智慧、樣本比對等方法來比較具有下列優點: 計算

速度快、簡單易懂、有自有學習能力及對雜訊的較不敏感 (insensitivity)。

本篇論文的編排結構如下:第二節簡介 MPEG 編碼原理,第三節說明 MP3 音樂的特徵值如何擷取,第四節說明如何計算特徵歌曲的方法,第五節說明使用特徵歌曲如何做分類,第六節列出實驗結果及數據,第七節說明未來工作及結論。

## 2. MPEG 音訊編碼



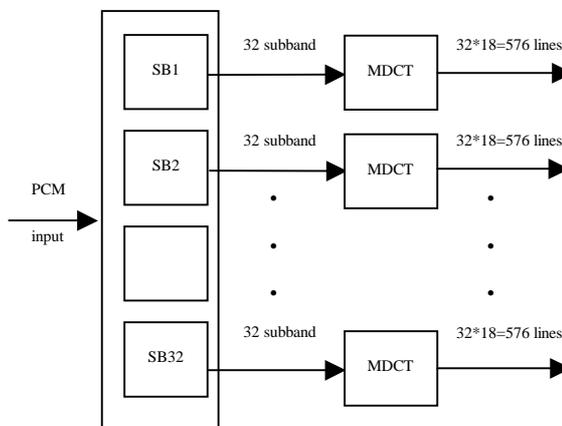
圖一 MPEG 音訊編碼方塊圖。(參考[4])

圖一說明了 MPEG 音訊編碼的流程。如圖所示,其編碼過程包含了四個功能模組:

- 分析濾波器:將輸入的音訊訊號由時間域 (time domain) 分解成 32 個頻率域 (frequency domain) 的子頻率,稱之為子頻率取樣 (subband samples) 或頻率線 (frequency lines)。
- 感官模型:使用心理聽覺 (psychoacoustics) 模型的特性,將人類聽覺範圍以外的音訊訊號做遮蔽 (masked)。
- 量化及識別:將子頻率取樣或頻率線做量化及識別的動作。主要目的是為了要讓在做量化時產生的雜訊會低於遮蔽限制 (masked threshold)。
- 位元流編碼:封裝 MP3 的框架 (frame)。MP3 的框架是由位元流 (bit stream) 所包裝的,位元流包含了量化後的樣本及位元配置資料 (bit allocation information)。

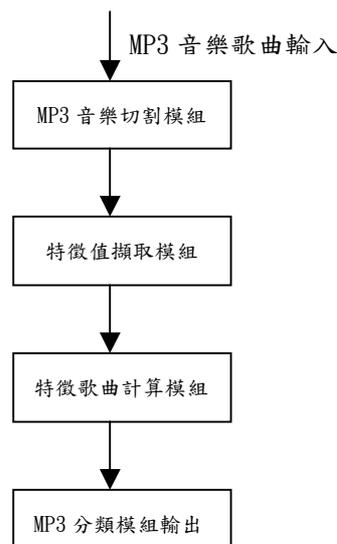
而在 MPEG-1 Layer 3 編碼程序中最重要的一個步驟就是使用修正離散餘弦轉換 (MDCT: Modified Discrete Cosine Transform),其作用在使得抽離出的 32 個子頻帶頻率訊號能夠達到更高的頻率解

析度 (frequency resolution) 及更接近關鍵頻率 (critical band)。而其方法是將每一個子頻帶再加以細分成 18 個子頻帶,因此總共是由  $32 \times 18 = 576$  條的頻率線所組成的,如圖二所示。



圖二 MDCT 功能圖

## 3. MP3 分類系統簡介



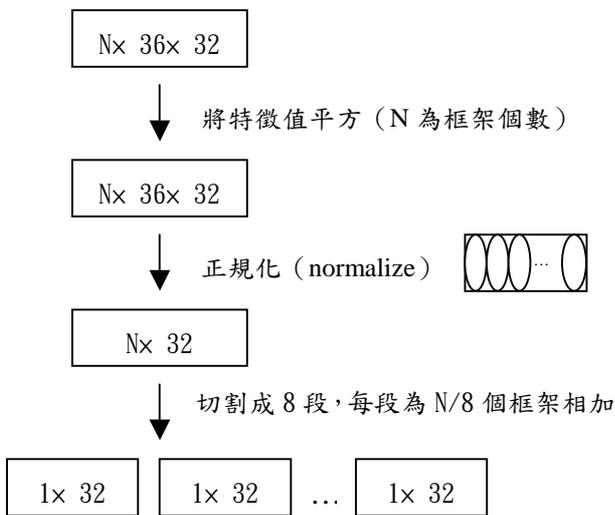
圖三 分類系統架構圖

MP3 分類系統總共有四個模組,分別為 MP3 音樂切割模組、MP3 特徵值擷取模組、特徵歌曲計算模組及 MP3 分類模組,如圖三所示。

### 3.1 MP3 音樂切割模組

在 MP3 檔案標準格式中 [15], 定義了 MP3 的每個框架 (frame) 有 1152 個樣本 ( $32 \text{ subbands} \times 36 \text{ blocks}$ ), 而我們就是以框架的個數來當做切割的

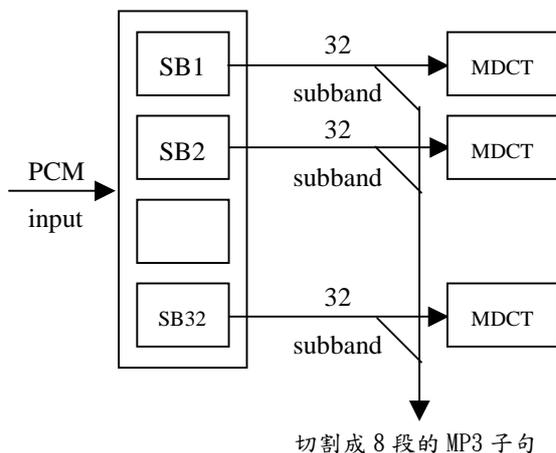
單位。切割方法如圖四所示。對 MP3 音樂做切割的主要目的是為了要讓不同長度的 MP3 可以使用相同的長度單位來比較彼此的相似度。首先，我們的 MP3 有  $N$  個框架，每個框架有 1152 個取樣樣本，也就是說會形成  $N \times 36 \times 32$  的特徵值向量，而我們將每個框架的 1152 個樣本看成是  $36 \times 32$  的矩陣，然後將此矩陣的列 (column) 相加，如此成為僅有 32 個 subbands 值，亦即由  $N \times 36 \times 32$  正規化成  $N \times 32$ 。將 MP3 切割成 8 段的方法就是把每  $\frac{N}{8}$  個的框架相加即可。



圖四 MP3 音樂切割模組

### 3.2 特徵值擷取

我們依據 MP3 編碼的原理來定義 MP3 音樂特徵值，也就是說在本篇論文中我們所擷取的 MP3 音樂特徵值是將聲音訊號由時間域轉換成頻率域的分析濾波器之後所抽離出的 32 個頻率子頻帶。如圖五所示。



圖五 MP3 音樂特徵值 (MSCV) 擷取方塊圖

#### 3.2.1 MP3 音樂特徵值：子頻帶係數向量矩陣

(MSCV: Matrix of Subband Coefficient Vector) 個子頻帶的係數代表了此聲音訊號在此子頻帶的頻率能量強度大小，這 32 個子頻帶係數稱之為多相位濾波器係數向量 (PCV, Polyphase filter Coefficient Vector) [14]。每一句 MP3 音樂都會產生 32 個 PCV 值，所以我們將一句 (sentence) MP3 音樂切割成 8 段的子句 (subsentence)，每一 MP3 音樂子句都有 32 個 PCV 值，所以一句 MP3 音樂會產生  $8 \times 32$  維度的特徵值向量，稱之為子頻帶係數向量矩陣。

### 3.3 計算特徵歌曲

計算特徵歌曲的方法有四個步驟。第一步，首先將一句的 MP3 歌曲切割成 8 段，每一段的 MP3 有 32 個子頻帶 (subband)，也就是說一句 MP3 歌曲  $S$  是一個  $8 \times 32$  的特徵值矩陣，或是說  $S$  為一個 256 維度的向量。第二步，算出 MP3 歌曲  $S_1, S_2, \dots, S_M$  ( $M$  為歌曲數量) 的平均值及差異值，MP3 歌曲平均值及差異值的定義分別為

$$S_{avg} = \frac{1}{M} \sum_{n=1}^M S_n$$

$$D_i = S_i - S_{avg}, i = 1, 2, \dots, M$$

第三步，算出協方差矩陣 (covariance matrix)

$$C = \frac{1}{M} \sum_{n=1}^M D_n D_n^T$$

$$= A A^T, A = [D_1, D_2, \dots, D_M]$$

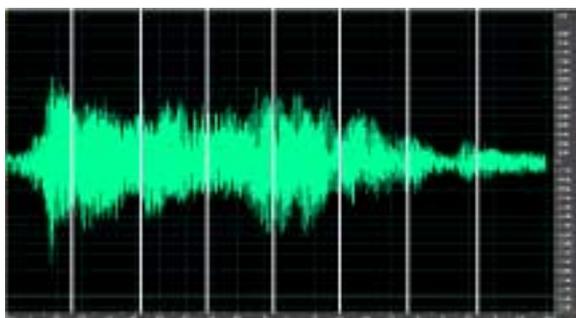
第四步，算出  $C$  的特徵值及特徵向量。

圖六為女歌手劉若英所唱的歌曲“後來”中的一個 MP3 歌曲，將它解碼後的时间域波形圖。



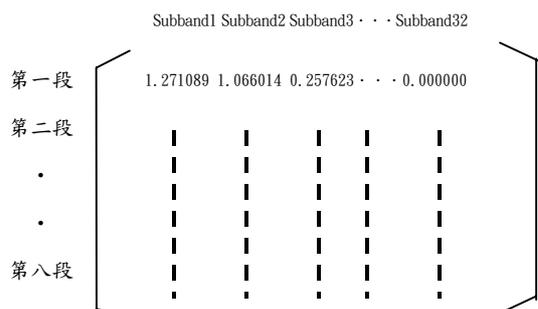
圖六 MP3 歌曲“後來”的時間域波形圖

以下是以此句 MP3 歌曲來計算特徵歌曲的例子。如圖七所示，我們將此句 MP3 歌曲以框架 (frame) 個數切割成八段，每段的長度為  $\frac{\text{框架總數}}{8}$ 。



圖七 以 frame 個數切割 MP3 歌曲

然後分別算出這 8 段 MP3 的 PCV 值，所以我們得到了 MSCV 特徵值矩陣，如圖八所示。

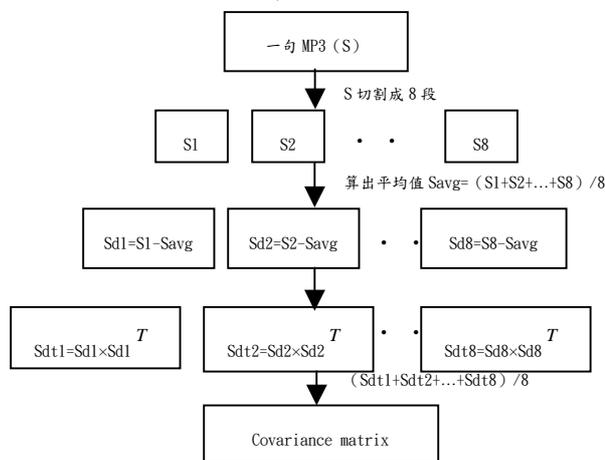


圖八 MP3 特徵值矩陣 (MSCV)

以同樣的方法分別算出其它 MP3 歌曲的 MSCV 特徵值矩陣之後，可以依上述的定義算出 MP3 歌曲的平均值及差異值，再以此算出協方差矩陣，最後算出協方差矩陣的特徵值及特徵向量，即為所求的特徵歌曲。

在數學觀點來看，我們希望能夠在 MP3 歌曲特徵值

矩陣中找到其分佈的主要方向，或者說就是找出協方差矩陣的特徵向量，因為這些特徵向量可以將 MP3 歌曲特徵值的變化叢聚在一起，因此我們定義它們為特徵歌曲。其計算流程圖如圖九所示。



圖九 特徵歌曲計算過程

### 3.4 使用特徵歌曲如何做分類

我們定義的特徵歌曲就是由上節所計算出的協方差矩陣之特徵向量。也就是所如果我們有  $N \times N$  維度的 MP3 音樂特徵值矩陣，那麼就會計算出  $N^2 \times N^2$  的協方差矩陣，如此，我們會得  $N^2$  個維度為  $N^2$  的向量，以此  $N^2$  個特徵值來算出代表描述此 MP3 音樂特徵的特徵向量，以這些特徵向量為基底 (basis)，算出不同類型 MP3 的係數。

#### 3.4.1 分類基本原理

本篇論文中我們所使用到的分類依據的基本原理就是以線性代數中矩陣運算的特徵值為基礎，而特徵值的主要應用就是在找出一個系統的線性差異方程式。其基本定義為：

$$Ax = \lambda x$$

A 為  $n \times n$  矩陣

$\lambda$  稱為特徵值或是 A 的獨特值

$x$  稱為特徵向量或是 A 的獨特向量

我們分類方法就是在 MP3 音樂特徵值矩陣中找出特徵值及特徵向量，並以它們為基底，算出每一 MP3 音樂以此為基底的特徵空間中的座標。而分類的方法就是去比對 MP3 歌曲的座標，一般來說，相同類型的 MP3 歌曲其座標會叢聚在一起，所以只要比對此新的 MP3 特徵值之基底與原先 MP3 特徵值

基底的相似性即可將它們分類。也就是說假如我有特徵向量  $U$ ，且  $x$  為 MP3 音樂特徵值矩陣，則基底係數  $C$  的定義為：

$$C = U^{-1}x$$

例如： $u_1 = (3,2)^T$ ， $u_2 = (1,1)^T$ ， $x = (7,4)^T$ 。

找出以  $u_1$ 、 $u_2$  為基底的  $x$  坐標。

$$U = (u_1, u_2) = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}$$

所以

$$c = U^{-1}x = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 7 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

$$\therefore x = 3u_1 - 2u_2$$

## 4 實驗結果

在本節中，我們將介紹實驗環境設定(取樣樣本)、實驗結果及實驗結果的分析。

### 4.1 實驗環境設定 MP3 音樂歌曲取樣樣本

在 MP3 音樂歌曲取樣樣本方面，如表一所示，在我們的 MP3 流行歌曲資料庫中準備了三種不同類型的 MP3 音樂歌曲：古典音樂(classical)、男歌手流行歌曲(male)及女歌手流行歌曲(female)，每種類型各取了 200 句，所以共有 600 句的 MP3 音樂。

MP3 類型	歌曲數量
古典	200
女歌手	200
男歌手	200

表一 MP3 資料庫中之 MP3 種類

### 4.2 實驗步驟及結果

首先將三種類型 MP3 音樂共 600 句的音樂全部混合在一起，算出其特徵向量，並以此特徵向量為基底。然後我們將未知(unknow)類型的 MP3 算出其特徵值矩陣，並將此特徵值矩陣以先前所算出的特徵向量為基底，做基底轉換的動作，如此我們可以

得到以特徵向量為基底的座標。我們將求得的這 256 個座標值  $X_i$  以統計的方式算出其平均值(mean)及變異數(variance)。

$$\mu = \log \left( \frac{\sum_{i=1}^{256} |X_i|}{256} \right)$$

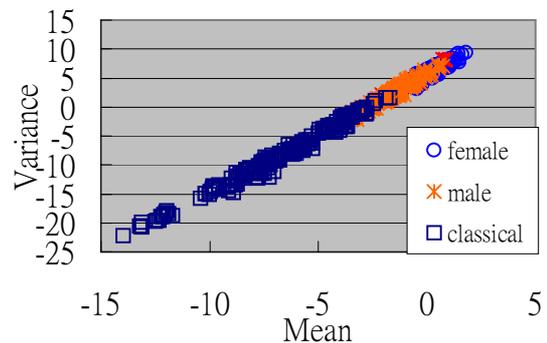
$$\delta = \log \left( \frac{\sum_{i=1}^{256} (|X_i| - \mu)^2}{255} \right)$$

$X$  為特徵值矩陣

$\mu$  為平均值

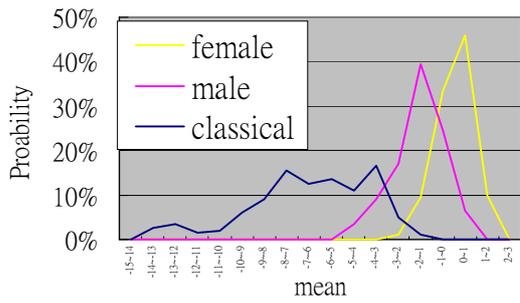
$\delta$  為變異數

我們將 600 句的 MP3 歌曲樣本，把它們放入我們的 MP3 分類系統中，使用特徵歌曲分類演算法，如此我們得到的結果如圖十所示。

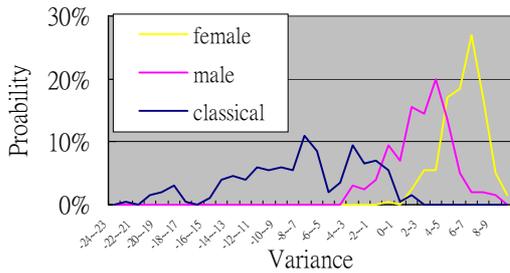


圖十 使用特徵歌曲分類結果圖

圖十一為我們 MP3 資料庫中所有歌曲平均值的機率分佈圖，由觀察可知男、女聲的特徵值分佈情形，並不十分明顯，容易造成分類錯誤的狀況，而為了解決分類不精確的問題及增加分類的正確率，我們加入變異數當成 MP3 分類的依據，變異數的機率分佈圖如圖十二所示。

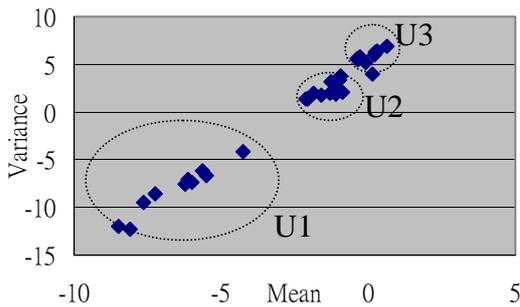


圖十一 平均值機率分佈圖



圖十二 變異數機率分佈圖

圖十三為輸入樣本以外的未知類型 (unknown type) MP3，共輸入了三十句 MP3 歌曲，我們由圖得知，可將這三十首 MP3 分成三類，我們將它命名成 U1、U2 以及 U3，同時我們可以說 U1 是古典樂類型的 MP3 歌曲，U2 為男歌手所唱的 MP3 歌曲，U3 為女歌手所唱的 MP3 歌曲。

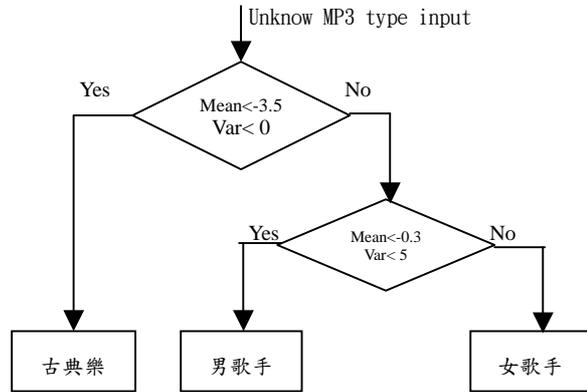


圖十三 未知類型 MP3 分類結果

### 4.3 實驗分析

我們發現不同類型的 MP3 音樂，所得到的平均值及變異數大小有著不同的差異，所以平均值及變異數大小可以當成我們 MP3 分類的依據。若平均值小於

-3.5 及變異數小於 0，則此 MP3 音樂為古典音樂類型，若平均值大於 -3.5 而小於 -0.3 且變異數小於 5，則此 MP3 音樂為男歌手所唱的音樂類型，若平均值大於 -0.3 且變異數大於 5 的 MP3 音樂為女歌手所唱的音樂類型。其分類樹狀結構圖如圖十四所示。



圖十四 MP3 分類樹狀結構圖

我們現在將我們 MP3 流行歌曲資料庫中的 600 句 MP3 使用特徵歌曲分類演算法來運算，並且使用上述的 MP3 分類樹，可得到表二的結果。由實驗結果得知古典樂的分類結果正確率為  $\frac{182}{200} = 91\%$ ，男歌手所唱的流行歌曲 MP3 的分類結果正確率為

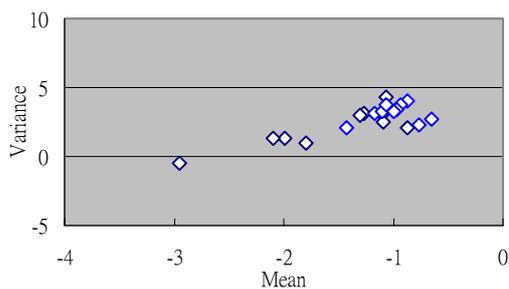
$\frac{167}{200} = 83.5\%$ ，女歌手所唱的流行歌曲 MP3 的分

類結果正確率為  $\frac{144}{200} = 72\%$ 。而我們發現要將流

行歌曲中的 MP3 精確的分類成男歌手及女歌手並不太容易，因為男歌手在音域上的表現會比女歌手來的穩定，所以假如我們輸入的女歌手其聲音聽起來似男生 (man-like)，則會誤分類成男歌手，如圖十五所示。

MP3 類型	分類正確數	誤判數
古典樂	182	18
男歌手	167	33
女歌手	144	56

表二 使用特徵歌曲分類結果



圖十五 使用特徵歌曲分類似男聲 MP3 結果

## 5 未來展望及結論

在本篇論文中，我們參考了特徵臉的演算法，來做 MP3 音樂類型的分類，特徵臉是一項應用在臉部辨識中十分成熟的技術。我們引用其方法將 MP3 音樂自動分為古典音樂、男歌手所唱的歌及女歌手所唱的歌等三種類型。根據實驗，分類成功率達到 72% 以上

對來未來的展望，我們會依據 MP3 音樂的聲音感覺 (audio-visual) 做分類的類型，將分類類型區分為抒情的音樂、憂鬱的音樂、高興的音樂等等，並且將此系統應用在網際網路上，建置一個如 Yahoo 的 MP3 搜尋分類目錄，成為網際網路上唯一的 MP3 情緒 (emotion) 搜尋引擎。此外，在心理醫療方面，本方法可以當成心理醫師在診療過程的輔助工具，幫助病患針對不同的病情，找到適合的心理治療音樂。

## 6 參考文獻

[1]A.T. Matthew, and P.A. Pentland, "Face Recognition Using Eigenface", *Computer Vision and Pattern Recognition, 1991.Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991.

[2]C. Colombo, A. Del Bimbo, and P. Pala," Semantics in Visual Information Retrieval", *IEEE Multimedia, Vol.6, No.3, July-September,1999.*

[3]C.-C. Liu, J.-L. Hsu, and A.L.P. Chen," An Approximate String Matching Algorithm for Content-Based Music Data Retrieval", *IEEE Multimedia Computing and Systems, Vol.1*, July 7-11, 1999.

[4]G. Robbert van der Waal, K. Brandenburg and G. Stoll, "Current and future standardization of high-quality digit audio coding in MPEG", *Applications of Signal Processing to Audio and Acoustics, 1993, Final program and paper Summaries. 1993 IEEE Workshop on*, 1993.

[5]J. Foote," A Similarity Measure for Automatic Audio classification", *Proc.AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora. Stanford, March 1997.*

[6]K. Hirata, S. Mukherjea, W.-S. Li, and Y. Hara,"Integrating Image Matching and Classification for Multimedia Retrieval on the Web", *IEEE International Conference on Multimedia Computing and Systems*, June 7-11, 1999.

[7]M. Flickner *et al*, "Query by Image and Video Content: The QBIC System", *IEEE Computer*,28(9),sep,1995.

[8]M. Setnes and R. Babuska, " Fuzzy Relational Classifier Trained by Fuzzy Clustering", *IEEE TRANSACTIONS ON SYSTEMS,MAN,AND CYBERNETICS-PART B: CYBERNETICS, VOL. 29, NO. 5, OCTOBER 1999.*

[9]P. Noll," MPEG Digit Audio Coding", *IEEE Signal Processing Magazine Volume : 145*, Sept. 1997.

[10]P. Noll, "High Quality Audio For Multimedia : Key Technologies and MPEG Standards", *Global Telecommunications Conference, 1999. GLOBECOM'99, Volume : 4*, Dec. 1999.

- [11]T.J. Robinson,” The K-D-B-Tree: A Search Structure for Large Multidimensional Dynamic Indexes“, *Proc. ACM SIGMOD Conf*, 1981.
- [12]T. Zhang and C.-C. Jay Kuo, ” Hierarchical Classification of Audio Data for Archiving and Retrieving”, *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on Volume:6*, 1999.
- [13]Y. Nakajima *et al*, ”A fast audio classification from MPEG coded data”, *Acoustics, Speech, and Signal Processing, 1999 Proceeding.*, 1999 IEEE International Conference.
- [14]蔡伯俊, 劉志俊, ”網際網路上 MP3 內涵式搜尋引擎”, *Proceedings of 2000 Workshop on Internet & Distributed Systems, Vol 1*, pp 18-27.
- [15]ISO/IEC 11172-3:1993, “Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s – Part 3:Audio.”