

MP3 數位音樂資料的自動化分類

黃群菘

中華大學資訊工程學系

m8902026@chu.edu.tw

劉志俊

中華大學資訊工程學系

ccliu@chu.edu.tw

摘要

隨著多媒體壓縮與網路技術的不斷進步，目前有大量地 MP3 數位音樂資料廣泛地散佈在網際網路環境中。因此，有關音樂資料的內涵式查詢與分類的問題，近年來受到學術界與工業界的高度重視。在本篇論文中，我們提出一種依照歌手來對 MP3 音樂物件進行自動化分類的技術。首先，在 MP3 解碼過程中，我們取出多相位濾波器的係數，並進一步據以計算 MP3 分段演算法所需要的 MP3 特徵值，將每一首 MP3 音樂自動切割為一連串的 MP3 音素。接著我們根據一組訓練用 MP3 歌曲範例建立一個 MP3 音素資料庫。最後，每一個 MP3 音素資料庫中的音素被視為一個辨識子，應用在我們所採行的 k NN 分類器，來對每一首未知歌曲進行歌手分類，自動建立類似 Yahoo 所提供的 MP3 音樂分類網頁。我們亦進行了一組實驗來驗證所提方法的效能，並對結果進行進一步分析與討論。

關鍵字

音樂分類(music classification)、音樂資料庫(music databases)、MP3 分類(MP3 classification)、MP3 資料庫 (MP3 databases)、歌手辨識 (singer identification)、內涵式音樂分類 (content-based music classification)、音樂特徵值擷取 (music feature extraction)

1. 序論

世界上有許多我們亟欲收藏的人類文明，如達文西的蒙娜麗紗的微笑、貝多芬的命運交響曲、電影中的真善美等等，但由於許多現實的原因，使得這些人類資產原件保存不易，因

此我們希望將這些資料數位化，如此一來便可永久保存。其中對於音樂的部份，傳統是採用 CD 光碟片的方式做收藏，優點是收藏容易，音質極佳，缺點是儲存容量過於龐大，因此 MPEG 組織制訂了一種數位音樂檔案的壓縮格式 MP3(MPEG -1 Audio Layer3)[9]，提供了高品質高壓縮比的音訊壓縮技術，於是 MP3 已成為目前數位音樂典藏中最重要且普遍的音樂型態。

由於 MP3 音樂在數位音樂世界中大為盛行，因此在網際網路上有越來越多的 MP3 歌曲可以自由下載，於是 MP3 音樂的收藏與整理變成了一個重要的課題。由於 MP3 的分類技術可以建立有效率的索引結構和提供快速的查詢能力，因此 MP3 的自動分類是目前一項極其重要的研究主題[12][13][19][21][24]。

在眾多的音樂分類方法中，以音樂曲風 (music genres)[31][32]和音樂歌手 (artists) 的分類方法最為常見。以國外知名網站 Yahoo 和 MP3.com 對於音樂的分類為例，如圖 1 所示，Yahoo 網站根據曲風將音樂分為 12 類，如藍調歌曲、鄉村歌曲、爵士音樂等等，而 MP3.com 也以曲風將音樂分成 16 大類。此兩個網站亦提供依歌手名稱的 MP3 音樂分類，如圖 2 所示，Yahoo 網站依照歌手或樂團名稱由 A 到 Z 的字母順序做歌曲的分類。而 MP3.com 網站也有提供相似的分類方法，如圖 3 所示。

MP3 是一種經過赫夫曼編碼法(Huffman coding) 壓縮的音樂檔案，所以我們無法直接

*本論文研究為國科會補助之研究成果，

從 MP3 的原始資料(raw data)中直接找出有涵義的資訊。因此我們必須在 MP3 解碼過程中擷取 MP3 特徵值(MP3 features)，利用 MP3 音樂媒體在能量上變化的特性，將 MP3 音樂切割成一個一個的音素(phonemes)，再計算出每個音素所代表的特徵值。針對這些特徵值再進一步設計相似演算法來進行相似度計算，找出最為接近的歌手分類。本論文的結果可以提供給的音樂搜尋網站以及線上音樂網站作為自動建立音樂分類目錄的工具。



圖 1. Yahoo 對 MP3 歌曲依音樂曲風的分類



圖 3. MP3.com 對 MP3 歌曲依歌手的分類

本論文結構說明如下：在第 2 節我們會介紹 MP3 分類系統架構，第 3 節介紹音素資料庫的建立，第 4 節介紹 MP3 音樂的分類，第

5 節說明實驗的過程與結果分析，第 6 節為結論及未來工作。

2. MP3 分類系統架構

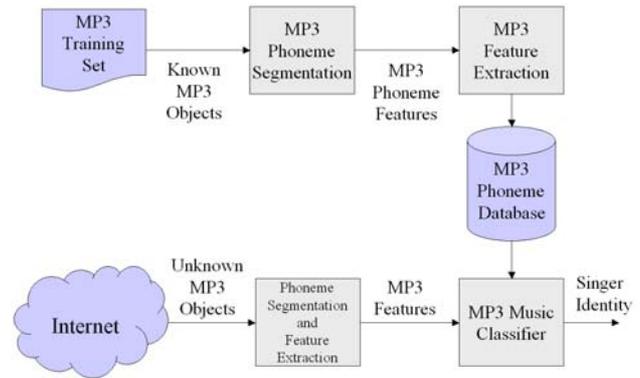


圖 4. MP3 歌曲自動分類系統架構圖

圖 4 是 MP3 歌曲自動分類系統的系統架構圖。分類的方法主要分為兩個步驟：在一開始，我們會先針對 MP3 歌曲以音素為單位作切割的動作，之後從切割下來的音素中找出特徵值。對於這些音素特徵值我們將它們當作 MP3 分類系統的辨識子，並進一步做辨識能力的調整工作。在第二階段，對於網際網路上的 MP3 歌曲，在經過類似第一步驟的切割和特徵值的擷取過程之後，便利用 MP3 音素辨識子來做未知音樂的歌手辨識工作。

3. 音素資料庫的建立

在說明音素資料庫的建立之前，我們必須先瞭解 MP3 的解碼(decoding)的程序 [2][25][27]，才能知道 MP3 音素特徵值的涵義與由來。

對於多媒體做內涵式的查詢和分類之相關研究，都非常強調能夠從壓縮的原始資料中抽出需要的特徵值，以節省特徵值計算的時間。圖 5 是 MP3 解碼的過程圖。MP3 最小的位元流單位是框架(frame)，而每個框架是由 1152 個樣本所組成的，典型的 MP3 歌曲的取樣頻率是 44100 Hz，因此 MP3 歌曲每秒會有 $44100/1152=38.28125$ 個框架。根據 ISO 國際標準組織對於數位音樂的壓縮標準[9]，MP3 的位元流會一個框架接一個框架的解碼為修正

式離散餘弦轉換(MDCT, modified discrete cosine transform)係數。接著將藉由將頻率樣本轉換成時間樣本的动作將 MDCT 係數(576 條頻率線)轉換到對應的 32 個子頻帶(subbands), 而這 32 個子頻帶的係數我們稱之為多相位濾波器係數(polyphase filter coefficients)。最後每個子頻帶會被合成為原始的聲音訊號(PCM audio)。

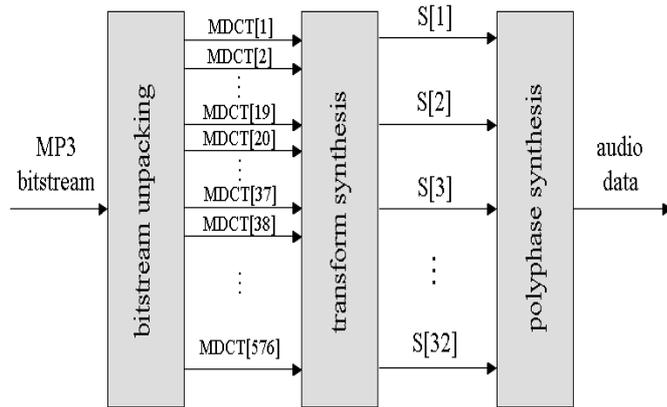


圖 5. MP3 解譯程序

在解譯的過程中, MP3 解碼器會將 576 個 MDCT 係數轉換 32 個子頻帶係數。由於任一個子頻帶的係數都代表了此子頻帶中聲音訊號的能量強度, 對於每一個子頻帶的每一個 MP3 的框架我們做平方的處理。若 $S[i][j]$ 表示第 i 個子頻帶的第 j 個框架, 我們定義第 i 個框架能量係數 PC_i 為

$$PC_i = \sum_{j=1}^{36} (S[i][j])^2 \quad (1)$$

MP3 音素(MP3 phoneme)指的是在五線譜上的一個音符或是樂句中歌手哼唱的一個音節。它可以是歌手哼唱或由音樂樂器的演奏所產生。MP3 音素與音素間變換的特徵是每個框架在於能量上的間隔現象(energy gaps)。舉例來說, 圖 6 是一句歌曲的聲音波形, 我們可以明顯的觀察到在每個音符間能量的改變。因此我們可以將一個框架的 32 個子頻帶的能量 PC_i 加總起來定義為此框架能量(frame energy, FE)[16]。

$$FE = \sum_{i=1}^{32} PC_i \quad (2)$$

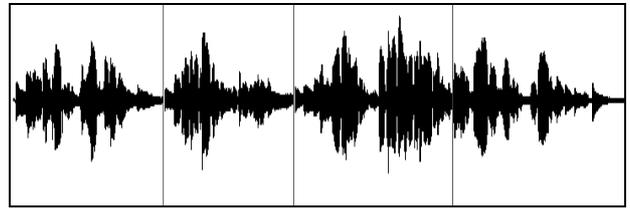


圖 6. 音樂聲音波形

圖 7 顯示一個四個音素的聲音波形。我們可以在兩個音素間發現能量上的間隔, 此間隔稱為 MP3 音素的斷音點(phoneme breaks)。MP3 音樂的斷音指的就是設法自動地從 MP3 歌曲資料中識別出這些能量缺口的位置。

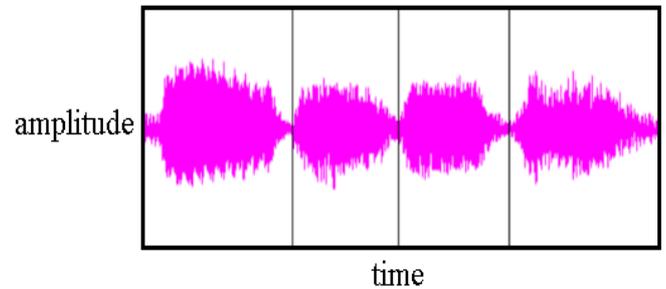


圖 7. 四個音素的聲音波形

音素的聲音波形有四個結構, 也就是起奏(attack)、衰退(decay)、延續(sustain)和釋放(release)四個階段, 一般稱為音素的 ADSR 特性。圖 8 是兩個音素的聲音波形, 其 ADSR 的結構非常明顯, 而能量個間隔就是位於前一個音素釋放的階段和下一個音素起奏的地方。

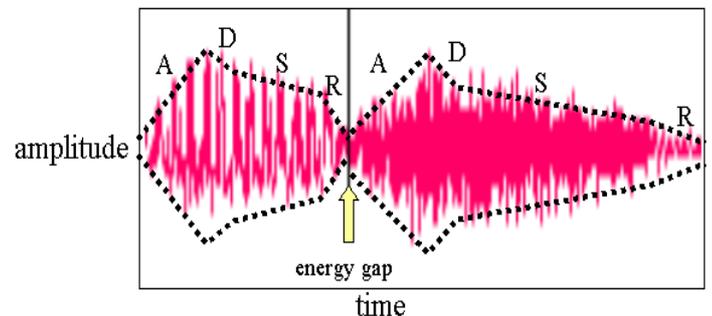


圖 8. 個音素的聲音波形

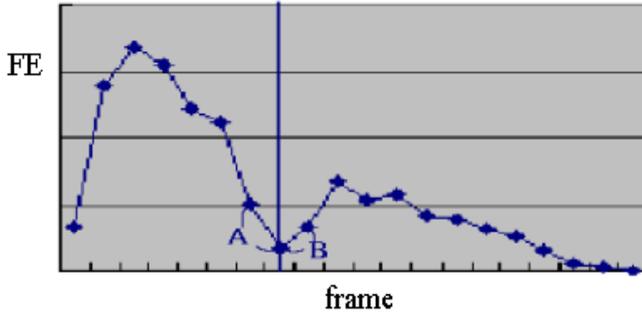


圖 9. 兩個 MP3 音素的 FEs

我們將圖 8 的歌曲轉換成 MP3 檔案的格式並依(2)式取出其特徵值，結果如圖 9 所示。我們發現有兩個特點：首先，能量上缺口的位位置位於 A 和 B 之間，其結果和圖 8 的間隔位置一致，因此我們要切割的位置在上一個音素釋放階段的部分 A 和下一個音素起奏的部分 B。第二，我們可以發現斷點位置的框架能量值(FEs)比一般框架的框架能量值來的低。根據這些特性，我們可以找出 MP3 斷音的方法，此斷音演算法說明如下：

規則 1：假設框架 A、框架 B 和框架 C 是一首 MP3 音樂三個連續的框架，而它們的 MP3 特徵值分別表示為 FE_A 、 FE_B 、和 FE_C ，假如 MP3 音素斷點的位置在於框架 B，則框架 B 的能量低於或相等於框架 A 的能量。因此，我們可以得到

$$\begin{cases} FE_B \leq FE_A \\ FE_B < FE_C \end{cases} \quad (3)$$

$$FE_B < FE_C \quad (4)$$

這是因為框架 A 和框架 B 都是屬於一個音素最後的釋放階段。同樣地，假如框架 B 和框架 C 都是屬於同一個音素的起奏，框架 B 的能量也會低於框架 C 的能量。

然而並不是所有的 MP3 音素都會有 ADSR 的結構。雜訊可能會破壞了這種特性。因此為了降低雜訊所造成的干擾，我們改取連續 n 個框架的框架能量 FE' 的和來替代原本的框架能量 FE ，做為我們新的 MP3 的特徵值。

$$FE' = \sum_{i=1}^n FE_i \quad (5)$$

此處的 FE_i 代表第 i 個框架的 MP3 特徵值，而 FE' 是相對應的新的 MP3 特徵值，根據實驗，我們發現當 $n = 4$ 的時候，我們會得到最佳的結果。由於音素的斷點位置都發生在較低的能量處，因此我們也使用下述規則增加我們切割的準確率。

規則 2：假設框架 A、框架 B 和框架 C 是一首 MP3 音樂三個連續的框架，而它們的 MP3 特徵值分別表示為 FE_A 、 FE_B 、和 FE_C ，而 MP3 音樂的框架平均能量為 FE_{avg} ，則框架 B 是 MP3 音素斷點的位置，則

$$k \cdot \text{Max}(FE_A, FE_B, FE_C) < FE_{avg} \quad (6)$$

此處的 $\text{Max}(FE_A, FE_B, FE_C)$ 指的是在 FE_A 、 FE_B 、和 FE_C 中最大的框架能量； k 是介於 0 到 1 之間的常數。公式(6)的涵意為音素斷點必須發生在能量較低的區域。

在前面我們曾提到，MDCT 係數和多相位濾波器係數(PC)都可以當作 MP3 的特徵值。在這篇論文裡面，我們採用 MDCT 係數 MC_i 來當作我們的特徵值

$$MC_i = (\text{MDCT}[i])^2, i = 1, 2, \dots, 576 \quad (7)$$

$$\text{FMCV} = (MC_1, MC_2, \dots, MC_{576}) \quad (8)$$

每個 MP3 框架都可以用一個 576 維度的特徵值向量(FMCV)來代表。由於 MP3 音素是由一連串 MP3 框架所組成的，因此我們可以定義 n 個框架的 MP3 音素的音素特徵值(PMCV)

$$\text{PMCV} = (\sum_{i=1}^n MC_1, \sum_{i=1}^n MC_2, \dots, \sum_{i=1}^n MC_{576}) \quad (9)$$

在經過感知正規化(perceptual normalization)之後，便可以將每個音素特徵值儲存起來以建立分類用的音素資料庫。

4. MP3 音樂的分類

如前所述，MP3 分類的第一個階段就是去建構一個已知歌手的歌曲的 MP3 音素資料庫。由於歌手所哼唱的不同音素的數目是有限的，且每個歌手有不同的音色，因此每個歌手也擁有他們自己獨特的音素集合。因此未知

MP3 歌曲的音素可以和音素資料庫內相同歌手的相似音素間產生關連性。如此一來，在分類的第二個階段就是將 MP3 音素資料庫內的每個音素當作是辨識子(*discriminator*)來識別每一個未知的 MP3 歌曲。

由於每個音素的辨識能力或獨特性是不相等的。因此為了去衡量每個音素的對其歌手唱歌的獨特性，我們將和音素 f 的音素特徵值距離最短的另一個歌手所哼唱的音素 f' 的音素特徵值間的距離，定義為音素 f 的辨識半徑 γ_f

$$\gamma_f = \text{Min}(\text{dist}(PMCV_f, PMCF_{f'})) \quad (9)$$

這裡的 $PMCV_f$ 和 $PMCF_{f'}$ 是分別代表音素 f 和音素 f' 的音素特徵值；而 $\text{dist}(PMCV_f, PMCF_{f'})$ 則是 $PMCV_f$ 和 $PMCF_{f'}$ 間的歐基理德距離。

圖 10 顯示在特徵空間中三個歌手的 18 個音素特徵值。我們可以看到 f_1 的辨識半徑明顯的比 f_2 大許多，這表示音素 f_1 的音色是非常獨特的且也表示 f_1 具有較佳的辨識能力。因此我們可以說由於 f_1 可以辨識出三個相同歌手的音素，所以 f_1 是一個較好的辨識子。相對的，我們可以說 f_2 是不好的辨識子由於它無法正確地辨識出相同歌手的任何音素。

另一個影響音素辨識能力的是相同歌手相鄰音素的數目。假如一個音素有較多相鄰的相同歌手音素，則這個音素就的確代表了這個歌手的一種典型聲音。因此我們可以將落於相同歌手辨識半徑內的音素數目定義為該音素的相同歌手相鄰音素出現頻率 ω_f 。

舉例來說，圖 11 顯示了兩個有相同辨識半徑的辨識子 f_3 和 f_4 。然而 f_3 的出現頻率為 6 而 f_4 的出現頻率只有 1，因此這表示 f_3 比 f_4 有較強的辨識能力，也就是說 f_3 比較能代表此歌手的一種典型聲音。

辨識半徑是計算和其最相鄰的不同歌手的音素的距離。最為相鄰的音素不一定必須是某一個歌手所演唱的相同的音符或音節，只要它們是由不同的歌手所哼唱的話就符合條件。

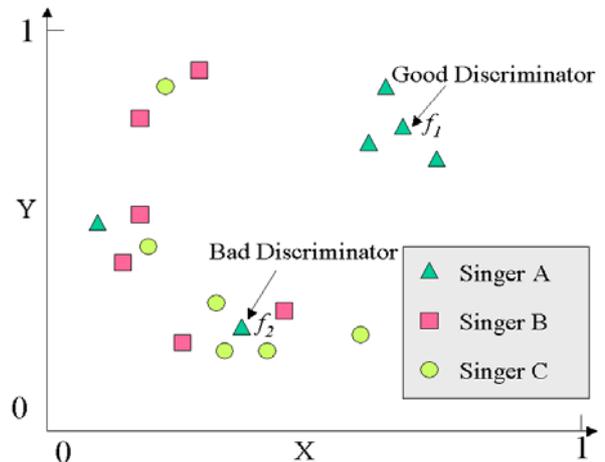


圖 10. 好的和壞的辨識子

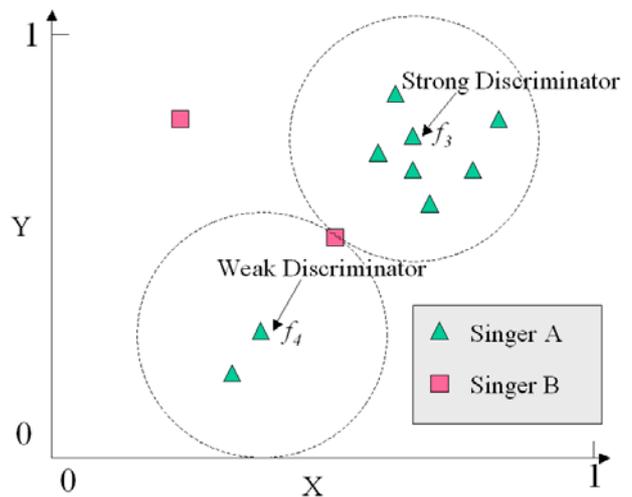


圖 11. 強辨識力與弱辨識力的辨識子

在音素資料庫中的每一個辨識者的出現頻率是可以動態調整的。為了得到音素資料庫中每一個辨識子的出現頻率，我們將每一個辨識子一開始的出現頻率設定為零，之後每個辨識子再逐一加入第二組相同歌手的訓練音素，假如這個音素的距離是落在某一辨識子的辨識半徑內，則此辨識子的出現頻率便增加 1。

音素資料庫內的每一個辨識子皆須經過辨識半徑 γ_f 和出現頻率 ω_f 的計算。最後我們可以將辨識子的辨識函數 $D(f)$ 定義為

$$D(f) = \gamma_f \cdot \log_2(\omega_f + 1) \quad (10)$$

我們採用 k NN 的分類方法做未知歌曲的分類工作[6]。對於未知的 MP3 歌曲，一開始

先分割成一連串的音素。為了效能的考量，我們擷取 MP3 歌曲前 N 個音素來做分類。對每個音素，我們和音素資料庫中的每個辨識子做比對的工作，並且找出前 k 個通過相似門檻值的最相近的音素。換句話說，我們根據 $k*N$ 個相近音素的相似度總和來決定未知歌曲的歌手為何。

5. 實驗

我們隨機選擇 10 位男歌手和 10 位女歌手來作實驗。對每位歌手我們收集 30 首不同的歌曲，其中 10 首歌曲用來建立我們的音素資料庫，10 首歌曲拿來調整音素辨識子，10 首歌曲拿來做為 MP3 分類系統的實驗歌曲。因此，在我們的實驗中總共用到 600 首歌曲。

我們針對影響實驗結果的三個因素來做分析。首先是 k NN 中的 k 值對於實驗的影響，再來是相似門檻值的設定，最後是各個歌手間相似度的分析。

首先我們第一個實驗便是要去找出 k NN 分類方法中最佳的 k 值設定。如表 1 所示，當 k 值設定介於 50 到 100 時會有較佳的結果。因此我們在針對 10 到 100 的區間作進一步分析，結果如圖 12 所示，當 k 值為 80 時，會有最佳的結果(此時相似門檻值設定為 0.2)。

表 1. 各種 k 值設定下的分類準確率

k	1	50	100	500	1000	10000
Precision	10%	70%	70%	50%	20%	10%

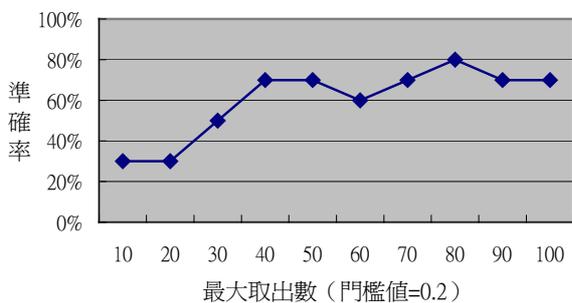


圖 12. k 值對分類準確率的影響

第二個實驗測試相似門檻值的影響。如圖 13 所示，當門檻值設定為 0.01 時，沒有發現

任何一個相似的辨識結果。此時分類如同亂數隨機分類，準確率為 $1/20 = 5\%$ 。在另一方面，當門檻值設定為 1 時，任何一個相似的结果皆會被接受。根據我們的實驗，當門檻值設定為 0.2 時，我們會得到最佳的結果。

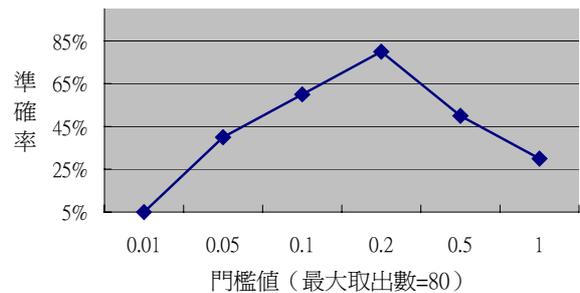


圖 13. 門檻值的對分類準確率的影響

最後是 MP3 分類系統的歌手間相似度的分析。圖 14 所示是男女歌手分類準確率的比較圖，從圖中我們可以發現，男歌手的平均辨識率比女歌手高出 9 個百分比。而圖 15 則是所有歌手的準確率比較。從圖中我們發現，在男歌手裡頭，游鴻明的辨識率最高，原因可能是因為游鴻明的聲音獨特，有較佳的辨識結果，而張宇的辨識率最差，原因可能是因為張宇唱歌咬字不清，因此造成 MP3 歌曲音素切割的準確率不高造成辨識結果不佳。在另一方面，女歌手中張惠妹的辨識率最高，原因是因為張惠妹的聲音咬字清楚，歌聲高亢有力而特殊；而莫文蔚的辨識率便較差，原因可能是因為她的歌聲比較大眾化，比較沒有特色。

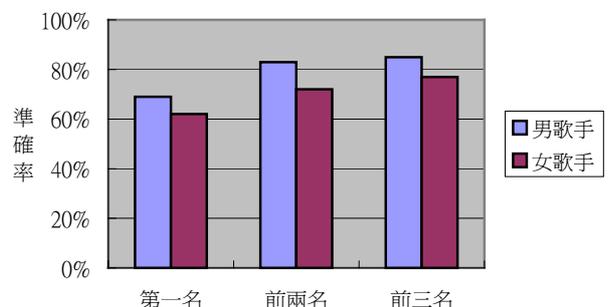


圖 14. 門檻值對男女歌手分類準確率的影響

表 2. 男歌手分類結果交叉分析

來源 分類為	伍思凱	周華建	林志炫	張宇	張雨生	張學友	游鴻明	熊天平	齊秦	劉德華
伍思凱	8	0	0	1	0	0	0	0	1	0
周華建	0	7	0	2	2	0	0	0	0	1
林志炫	1	0	7	0	0	1	1	1	0	0
張宇	0	0	0	6	0	0	0	0	0	0
張雨生	0	3	1	0	7	0	1	1	0	0
張學友	0	0	0	0	0	7	0	0	0	0
游鴻明	0	0	0	1	0	2	8	0	0	2
熊天平	0	0	1	0	0	0	0	6	3	0
齊秦	0	0	1	0	0	0	0	2	6	0
劉德華	1	0	0	0	1	0	0	0	0	7

表 3. 女歌手分類結果交叉分析

來源 分類為	王菲	辛曉琪	孫燕姿	張惠妹	梁詠琪	莫文蔚	許茹芸	萬芳	趙詠華	劉若英
王菲	6	2	1	0	1	2	3	1	1	2
辛曉琪	0	6	1	0	0	1	0	1	0	1
孫燕姿	0	0	6	1	0	0	0	0	2	0
張惠妹	0	0	1	7	0	1	0	0	0	1
梁詠琪	1	2	1	0	7	1	0	0	0	0
莫文蔚	0	0	0	0	0	5	0	0	0	0
許茹芸	3	0	0	0	0	0	7	0	0	0
萬芳	0	0	0	0	0	0	0	7	0	0
趙詠華	0	0	0	0	0	0	0	0	6	0
劉若英	0	0	0	2	2	0	0	1	0	6

6. 結論

本文提出一種利用 MP3 音素特徵值的相似性對未知 MP3 音樂進行自動分類的技術。本技術可以應用在網際網路搜尋引擎(如 www.yahoo.com) 或 MP3 音樂提供網站(如 www.mp3.com) 建立 MP3 分類目錄之用。此外，本文提出之技術並可進一步應用在音樂內涵分析(music content analysis)與音樂資料探勘

(music data mining)等研究主題中。在未來我們的研究工作大致有：

1. 加大 MP3 資料樣本的數目：因為當歌手的資料樣本越多，我們對於該歌手的音素就能有更完整的收集。
2. 尋找其它的特徵值：加入其它的特徵值，來增加辨識的結果。

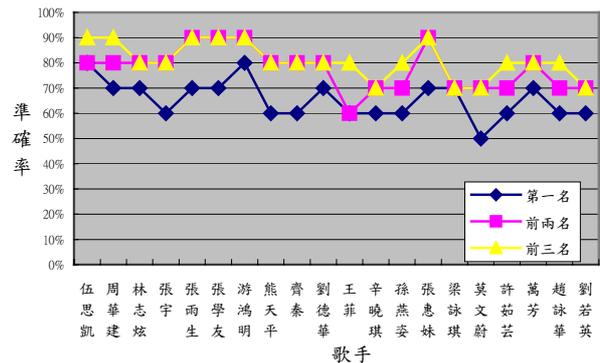


圖 15. 門檻值對每一歌手分類準確率的影響

7. 參考文獻

- [1] Bakhmutova, V., V. D. Gusev, and T. N. Titkova, "The Search for Adaptations in Song Melodies," *Computer Music Journal*, Vol. 21, No. 1, pp. 58-67, Spring 1997.
- [2] Brandenburg, K., and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *Journal of the Audio Engineering Society*, Vol. 42, No. 10, Oct 1994, pp. 780-792.
- [3] Chen, J. C. C. and A. L. P. Chen, "Query by Rhythm: An Approach for Song Retrieval in Music Databases," *In Proc. of 8th Intl. Workshop on Research Issues in Data Engineering*, pp. 139-146, 1998.
- [4] Chou, T. C., A. L. P. Chen, and C. C. Liu, "Music Databases: Indexing Techniques and Implementation," *in Proc. IEEE Intl. Workshop on Multimedia Data Base Management Systems*, 1996.
- [5] Foote, J., "Content-Based Retrieval of Music and Audio", *Multimedia Storage and Archiving systems II, Proc. SPIE, Vol.3229*, pp. 138-147.
- [6] Fukunaga, K., *An Introduction to Statistical Pattern Recognition*, San Diego, CA, Academic Press, 2nd ed., 1990.
- [7] Ghias, A., Logan, H., Chamberlin, D., and Smith, B. C., "Query by Humming: Musical Information Retrieval in an Audio Database," *in Proc. of Third ACM International Conference on Multimedia*, pp. 231-236, 1995.
- [8] Hsu, J.L., C.C. Liu and A.L.P. Chen, "Discovering Non-Trivial Repeating Patterns in Music Data," *IEEE Transactions on Multimedia*, Vol. 3, No. 3, pp. 311-325, 2001.

- [9] ISO/IEC 11172-3:1993, "Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s — Part 3: Audio."
- [10] Kosugi, N., Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima, "Music Retrieval by Humming," in *Proceedings of IEEE PACRIM'99*, pp. 404-407, 1999.
- [11] Kosugi, N., Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima, "A Practical Query-By-Humming System for a Large Music Database," in *Proc. ACM Multimedia*, 2000.
- [12] Lambrou, T. *et al.*, "Classification of Audio Signals Using Statistical Features on Time and Wavelet Transform Domains," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 3621-3624, 1998.
- [13] Li, S. Z., "Content-Based Audio Classification and Retrieval Using the Nearest Feature Line Method," *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 5, pp. 619-625, Sept. 2000.
- [14] Liu, C. C., A. J. L. Hsu, and A. L. P. Chen, "Efficient Theme and Non-Trivial Repeating Pattern Discovering in Music Databases," in *Proc. of IEEE Intl. Conf. on Data Engineering*, pp. 14-21, 1999.
- [15] Liu, C. C., A. J. L. Hsu, and A. L. P. Chen, "An Approximate String Matching Algorithm for Content-Based Music Data Retrieval," in *Proc. of IEEE Intl. Conf. on Multimedia Computing and Systems*, 1999.
- [16] Liu, C. C., and Wei-Yi Kuo, "Content-Based Segmentation of MP3 Music Objects," in *Proc. of the Workshop on the 21st Century Digital Life and Internet Technologies*, 2001.
- [17] Liu, C. C. and Po-Jun Tsai, "Content-Based Retrieval of MP3 Music Objects," in *Proc. of the ACM Intl. Conf. on Information and Knowledge Management (CIKM 2001)*, 2001.
- [18] Liu, Z. *et al.*, "Audio Feature Extraction and Analysis for Scene Classification," in *Proc. IEEE First Workshop on Multimedia Signal Processing*, pp. 343-348, 1997.
- [19] Liu, Z. and Q. Huang., "Classification of Audio Events in Broadcast News," in *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 364-369, 1998.
- [20] Lu, G.J. and T. Hankinson, "A Technique Towards Automatic Audio Classification and Retrieval," in *Proc. IEEE Intl. Conf. on Signal Processing*, Vol. 2, pp. 1142-1145, 1998.
- [21] Lu, G.J. and T. Hankinson, "An Investigation of Automatic Audio Classification and Segmentation," in *Proc. IEEE Intl. Conf. on Signal Processing*, Vol. 2, pp. 776-781, 2000.
- [22] Martin, K. D., and Y. E. Kim, "2pMU9. Musical instrument identification : A pattern-recognition approach," in *the 136th meeting of the Acoustical Society of America*, October 13, 1998.
- [23] Melih, K., and R. Gonzalez, "Audio Retrieval Using Perceptually Based Structures", in *Proc. of IEEE International Conference on Multimedia Computing and system*, pp 338-347, 1998.
- [24] Moreno, P.J. and R. Rifkin, "Using The Fisher Kernel Method for Web Audio Classification," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 2417-2420, 2000.
- [25] Noll, P., "MPEG Digital Audio Coding," *IEEE Signal Processing Magazine*, Vol. 14, No. 5, pp. 59-81, Sept. 1997.
- [26] Painter, T. and A. Spanias, "Perceptual Coding of Digital Audio," *Proceedings of the IEEE*, Vol. 88, No. 4, pp. 451-515, April 2000.
- [27] Pan, D., "A Tutorial on MPEG/Audio Compression," *IEEE Multimedia Magazine*, Vol. 2, No. 2, pp. 60-74, Summer 1995.
- [28] Rolland, P. Y., G Raskinis, and J. G. Ganascia, "Musical Content-Based Retrieval: an Overview of the Melodiscov Approach and System," In *Proc. ACM Multimedia 99*, pp. 81-84, 1999.
- [29] Smith, G., H. Murase, H. Kashino, "Quick Audio Retrieval Using Active Search", in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 3777-3780, 1998.
- [30] Tsai, Po-Jun and Chih-Chin Liu, "An MP3 Search Engine on the Internet", in *Proc. of 2000 Workshop on Internet & Distributed Systems*, Vol. 1, pp. 18-27, 2000.
- [31] Tzanetakis, G., G. Essl, and P. Cook, "Automatic Musical Genre Classification of Audio Signals," in *Proc. Int. Symposium on Music Information Retrieval (ISMIR)*, Bloomington, Indiana, 2001.
- [32] Tzanetakis, G., and P. Cook, "A Framework for Audio Analysis Based on Classification and Temporal Segmentation," in *Proc. EUROMICRO Conf.*, Vol. 2, pp. 61-67, 1999.
- [33] Wold, E., T. Blum, D. Keislar, and J. Wheaton, "Contented-Based Classification, Search, and Retrieval of Audio", *IEEE Multimedia Vol. 3, No. 3*, pp. 27-36, Fall 1996.
- [34] Zhang, T. and C.-C.J. Kuo, "Hierarchical Classification of Audio Data for Archiving and Retrieving," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 3001-3004, 1999.
- [35] Zhang, T. and C.-C.J. Kuo, "Audio Content Analysis for Online Audiovisual Data Segmentation and Classification," *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 4, pp. 441-457, May 2001.