

# Personalized Music Retrieval Based on Emotions / Situations

李侃儒

Kan Ru Li

國立政治大學資訊科學系  
g9312@cs.nccu.edu.tw

陳良弼

Arbee L.P. Chen

國立政治大學資訊科學系  
alpchen@cs.nccu.edu.tw

## 摘要

本論文的目標為建立一個個人化情緒／情境音樂檢索系統，主要的概念為根據使用者的 feedback 來找出符合該使用者情緒／情境的音樂在 features 上所具備的特性，藉此達到個人化的效果。為了更明確表示出音樂的特性，我們利用統計 features 分布情況的方式來做為音樂的表示法。同時，定義了兩層 features 自動 weighting 的方法來決定每個 feature 在不同情緒／情境下的鑑別度。最後，我們將探討音樂特性與音色對不同的情緒／情境會造成什麼樣的影響，並試著分析音樂與情緒／情境間的關係。

**關鍵詞：**music retrieval、emotion、situation、personalization、music representation、features re-weighting

## 1. 導論

### 1.1 背景

近年來，隨著音樂資料庫的快速成長，以及網路上音樂擁有易取得與低價格的特性，音樂檢索系統也越顯重要。過去，根據不同的應用已經發展出許多不同的音樂檢索系統，例如：由使用者哼唱或輸入一段旋律後查詢音樂 (string matching) [1][2] 等 content-based music retrieval 的方法都有相當詳盡的研究。然而，對於音樂與情緒／情境間關係的研究 (尤其是探討個人化情緒／情境音樂的研究) 卻相當有限。本論文擬提出一個新的個人化音樂檢索方法，將音樂與情緒／情境間的關係做一個結合。

### 1.2 音樂檢索

近年來隨著音樂資料數量快速膨脹，當想要搜尋音樂資料時常會發生只記得部分音樂片段，卻不記得曲名或演唱者的情形，此時想利用曲名或演唱者來進行音樂檢索便顯得相當困難。而對大多數的人而言，記住音樂的旋律是比記住演唱者或曲名更加容易的。許多耳熟能詳的音樂例如兒歌或民謠等，我們都不知道曲名也不知道演唱者，但我們往往能哼上一兩句。顯而易見的，單單以音樂的註解進行音樂檢索顯得不足，也不夠有彈性，此時便發展出了以音樂內容為基礎 (content-based) 的音樂檢索方法。這個方法主要是針對音樂的內容 (例如旋律等) 進行檢索。因此，使用者哼一段音樂，系統便能比對資料庫中所有的音樂，根據音樂片段找出最相近的音樂曲目，而由於音樂資料量相當龐大，因此，這個研究主要的重點就在於如何建立一個好的索引系統[1][2]。

而本論文擬進行的是以情緒／情境為基礎 (emotion/situation-based) 的音樂檢索。傳統的音樂檢索方法是針對記憶中已知的音樂片段來搜尋音樂，但是，在很多時候我們希望找到的是適合特殊情緒／情境的音樂。例如心情不好時我們可能會希望聽一些快樂的音樂，在家中開派對時我們希望播放一些動感的音樂等。與以音樂內容為基礎的音樂檢索方法不同，這階段的主要目的在於了解是什麼樣的音樂特性讓使用者擁有這樣的感覺。雖然一樣是針對音樂的內容進行音樂的檢索，然而，這裡更加強調音樂特性的表示，目的不在於建立索引結構，而是在於瞭解為什麼使用者對特定的音樂會有特定的情緒反應？什麼樣的音樂特性造成這樣的結果？因此，如何更明確地表示出音樂的特性，並根據使用者的喜好進行歸類找出音樂特性就成了這個階段最重要的議題。

近年來有關音樂與情緒的研究主要可分為三類：

- (1) 利用音樂與心理方面所建立的模型進行音樂的分類 [3][8]。關於這方面的研究優點在於有過去的音樂及心理方面相關研究的佐證，但是缺點在於分類法則已經決定了，導致彈性很低，無法加入新的情緒也無法加入新的音樂特徵。
- (2) 根據預先分類好的歌詞資料利用 decision tree 等方式建立歌詞分類法則[4]。這方面的研究主要的重點在於歌詞與情緒間關係的分析上。對於音樂特徵並沒有做相關分析，導致的問題有只用歌詞是否足以判斷音樂情緒？以及對於傳統古典樂等純音樂將沒辦法套用到這裡來定義情緒。
- (3) 根據預先分類好的音樂資料利用 SVM 等方式建立音樂分類法則[7]。使用 SVM 等方法最大的好處是可以直接套用既有的軟體或應用程式。但是缺點在於使用這類分類演算法過程就像一個黑盒子，我們沒辦法調整中間的過程，也很難說出最後得到的分類結果代表什麼意義？為什麼會有這樣的結果？加上這類的研究大多著墨於情緒的分類以及資料的蒐集上，關於學術研究的方法等則較無創新性。

而且這些方法主要都是在探討所謂音樂本身的情緒。然而，每個人對音樂的感覺會因為種族、成長背景等各項因素而有所不同，甚至於同一個人在不同的時間也會有不同的感受。因此，一個新的音樂情緒／情境檢索系統應該要能夠根據使用者不同的感受而自動找出最適當的音樂。

### 1.3 系統架構

本論文的系統架構可以圖 1 表示：

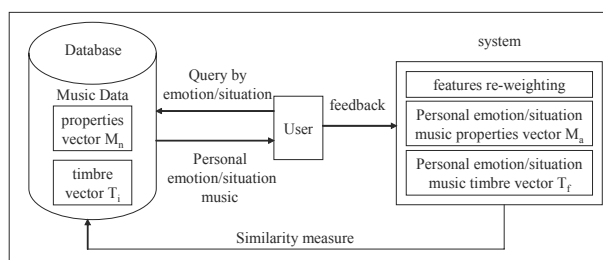


圖 1：系統架構：音樂與情緒／情境

一開始使用者的查詢為某種情緒／情境，系統會根據使用者所選擇的情緒／情境自動挑選出合適的音樂給使用者，而後，使用者可以勾選符合自己感覺的音樂當作 feedback 回傳給系統，系統便可以依據 feedback 找出符合使用者需求的音樂所具備的特性，最後，再依據這些特性重新對音樂資料庫中所有的音樂進行相似度比對，並將相似度較高的音樂回傳給使用者。由於我們可以藉由 feedback 找出符合使用者需求的音樂所具備的特性，因此，對於日後分析使用者的喜好，或了解音樂特性對人的影響等，將會有很大的幫助。

### 1.4 論文結構

在第二節，為了能更明確表示音樂的特性，我們提出一個新的音樂表示法，音樂資料庫中的音樂都將以這種形式來表示，並利用這個表示法進行音樂檢索與相似度計算。第三節則會探討如何利用使用者的 FEEDBACK 做到情緒／情境音樂的個人化。第四節為相關實驗結果及討論。第五節為結論。

## 2. 音樂檢索與相似度計算

如先前所提到的，在檢索情緒／情境音樂時，必須明確表示出音樂的特性。過去在表示一首音樂的特性時，最常使用的方法就是平均值[3][8]，例如我們可以說音樂 M 的平均音高是多少？平均音長是多少？平均音量是多少？希望能藉此找到一些分類規則，例如音量在哪些範圍時分別代表哪些情緒／情境等，然而，平均卻也有著很多缺點：






- (1) 容易被一些較大或較小的值影響而使得平均值的可信度降低。
- (2) 無法明確表示出一首音樂的特性。例如，一首音樂音量，如果以平均來看，那漸升、漸降或平穩的音樂將分不出差異。
- (3) 藉由平均得到的音樂特性不易理解。例如，我們說一首音樂的平均音長是 0.25 秒，平均音量是 75 分貝，對沒有豐富專業知識的人來說，無法理解平均音長 0.25 秒是長是短？而平均音量 75 分貝又會對人的心理造成什麼樣的影響？

所以我們必須建立出一種新的音樂表示法，以期能更明確表現出音樂的特性，因此，我們利用統

計分布情況的方法來表示音樂的 features。而為了得到 features 的分布情況來釐清對於每一種 feature 而言什麼樣的值算大、什麼樣的值算小，我們收集了大量的資料以求得各種 feature 所有可能的出現情形 (domain value)，再以 K-Means (K = 3) 的方式將每一種 feature 分為大、中、小三類 (L、M、S)，如此一來便可將數值資料轉換為類別資料。例如，根據 K-Means 得到音量小於 59 分貝的音符歸類於 S，音量大於 90 分貝的音符歸類於 L，其餘的音符則歸類於 M。例如一首音樂 (假設有十個音符) 的音量原本是 (70, 75, 80, 70, 15, 85, 65, 73, 82, 100)，經過類別轉換後變成 (M, M, M, M, S, M, M, M, M, L)，因此，這首音樂的音量可以用一個 vector  $M = (0.1, 0.8, 0.1)$  來表示，0.1 代表有 10% 的音是屬於類別 L，0.8 代表有 80% 的音是屬於類別 M，0.1 代表有 10% 的音是屬於類別 S。

而由於這樣的作法並沒有考慮到音樂的連續性，所以我們除了使用一般的音樂 features 如 pitch、duration、volume (vol)、beats per minute (bpm) 外，還多考慮了 change of pitch (CP) 及 change of volume (CV) 這兩個 features。其主要是將一首音樂分為很多小節，再來觀察一個小節中 pitch 及 volume 的變化情形。由於音樂一個小節都相當短，變化不會太劇烈，所以我們定義了五種簡單的變化情形：平穩、漸升、漸降、先升後降及先降後升。例如一首 10 個小節的音樂其音量在每個小節的變化情形是：漸升、平穩、平穩、漸降、先降後升、漸降、平穩、先升後降、先降後升、先降後升。這首音樂可以 vector 表示為 (0.1, 0.3, 0.1, 0.2, 0.3)，表示有 10% 的小節是先升後降，有 30% 的小節是先降後升，有 10% 的小節是漸升，有 20% 的小節是漸降，有 30% 的小節是平穩。表 1 列出音樂資料的表示法。

表 1：music vector

	....	vol (L)	vol (M)	vol (S)	CV 	CV 	CV 	CV 	CV 	....
M <sub>1</sub>	....	0.1	0.8	0.1	0.1	0.3	0.1	0.2	0.3	....
M <sub>2</sub>	....	0.1	0.2	0.7	0.1	0.1	0.3	0.3	0.2	....
....										

在確立了音樂的表示方法後，我們希望能根據使用者當前的喜好情況建立出一個 vector  $M_a$ ，使得 vector  $M_a$  與使用者勾選的音樂集合 A 相似度最高 (距離最短)。

$$\text{vector } M_a = \frac{\sum_{i \in A} M_i}{|A|}, \text{ 其中 } |A| \text{ 代表集合 } A \text{ 的音樂數量。}$$






由於單純使用平均的方式得到的 vector  $M_a$  並不能很正確的表示出使用者的喜好，因此，我們將 vector  $M_a$  加上權重，這是由於每個 feature 在不同的情緒/情境中重要度不同，例如，在 happy 的音樂中，聲音大小是一個很重要的判斷要素，而在 peaceful 的音樂中聲音大小的鑑別度可能就不是這麼大，因此，自動調整每個 feature 的 weight 將可以更正確地表示出使用者的喜好情形。

調整每個 feature weight 的方式主要可以分為兩部分：

- (1) 給予鑑別度高的 feature 較高的 weight (vector  $w_f$ )。而計算鑑別度的方式為根據  $M_a$  中各 feature 中類別的分布情況來判斷，當各類別分佈越均勻，代表 feature 不論是大是小影響都不大，所以鑑別度越低，而為了分別計算每個 feature 的分布情況是否均勻，我們利用出現機率最高的減掉出現機率最低的來表示，例如，在表 2 中  $M_a$  的 volume 出現最多的是類別 M，有 50%，而出現最少的是類別 L，有 10%，同樣地， $M_a$  的 CP 出現最多的是類別“漸降”，有 25%，而出現最少的是類別“先升後降”，有 10%，所以 CP 的分布較為均勻，weight 也較低，如表 2 的  $w_f$ 。
- (2) 給予 feature 中鑑別度較高的類別較高的 weight (vector  $w_{f_c}$ )。例如在表 2 中我們可以很清楚地發現到音量是集中出現在類別 M 和類別 S，但是對一首音樂而言，我們並不能確定分布情形高的會出現在 M 或 S 哪一個類別，但我們可以確定類別 L 的分布情況一定比較低。所以當 feature 集中出現在某幾個類別時，只有使用第一部分的 weight 公式是不夠的，而類別鑑別度的計算方式主要是根據使用者勾選的音樂集合 A 等資

料的標準差 (s) 而來的。由於我們的資料都介於 0 到 1 之間，依據標準差 (s) 的定義我們得到的標準差 (s) 將介於 0 到 0.5 之間。所以我們用 0.5 減標準差 (s)，這個值越大代表這個類別一致性越高，所以weight就要比較高，如表 2 的  $w_{f_c}$ 。

表 2：vector  $M_a$  and Features weighting

	vol (L)	vol (M)	vol (S)	CV 	CV 	CV 	CV 	CV 
Vector $M_1$	0.1	0.8	0.1	0.1	0.3	0.1	0.2	0.3
Vector $M_2$	0.1	0.2	0.7	0.1	0.1	0.3	0.3	0.2
Vector $M_a$	0.1	0.5	0.4	0.1	0.2	0.2	0.25	0.25
(max-min)	0.5-0.1=0.4			0.25-0.1=0.15				
$w_f$	0.4/(0.15+0.4)=0.73			0.15/(0.15+0.4)=0.27				
0.5-s	0.5	0.2	0.2	0.5	0.4	0.4	0.45	0.45
$w_{f_c}$ = normalized (0.5-s)	0.56	0.22	0.22	0.23	0.18	0.18	0.20	0.20

經由第一種 weighting 方法得到 feature 的鑑別度，再經由第二種 weighting 方法得到該 feature 中各類別的鑑別度後，兩個相乘就可以讓最重要的 feature 中一致性最高的類別擁有最高的 weight，因此，新的 feature weight  $w_i$  (其中 k 表示 vector 的維度)：

$$w_i = w_f * w_{f_c} \quad \sum_{i=1}^k w_i = 1$$

在得到了 vector  $M_a$  與 feature weight  $w_i$  後，音樂與情緒／情境的相似度計算可以 Cosine measure 表示：

$$\text{Sim}_M(M_a, M_n) = \frac{\sum_{i=1}^k w_i * M_a(i) * M_n(i)}{\sqrt{\sum_{i=1}^k w_i * M_a(i)^2} * \sqrt{\sum_{i=1}^k w_i * M_n(i)^2}}$$

其中， $M_n(i)$  代表音樂 vector  $M_n$  中的第 i 個值， $M_a(i)$  代表音樂 vector  $M_a$  中的第 i 個值。

因此，針對使用者的每一種情緒／情境，都可以計算出與音樂資料庫中所有音樂的相似度，相似度愈高的代表愈符合使用者的需求。

### 3. 即時個人化

每一種人類文化都從它的需要、歷史和環境中，尋找到一種音樂風格和表現這種風格的方式 [10]。因此，情緒／情境音樂的個人化是相當重要

的，系統將透過 feedback 逐漸找到符合使用者需求的個人化情緒／情境音樂，藉由使用者的 feedback 來修改上一節中的音樂與情緒／情境相似度公式：

$$\text{Sim}_M(M_a, M_n) = \frac{\sum_{i=1}^k w_i * M_a(i) * M_n(i)}{\sqrt{\sum_{i=1}^k w_i * M_a(i)^2} * \sqrt{\sum_{i=1}^k w_i * M_n(i)^2}}$$

針對使用者勾選的結果，每次都根據 relevant 的音樂重新計算 vector  $M_a$  與 weight  $w_i$ ，逐漸趨近使用者心中所想像之音樂情緒／情境。

而另一方面，為了有效利用 non-relevant 的資訊，並更加接近使用者的需求，我們多加入了一個 feature—音色。在聲音物理特性 (頻率與震幅) 上，影響人類生理及心理情緒因素的三個向量為：音量 (Loudness)、音高 (Pitch) 與音色 (Timbre) [9]。由此可見音色對使用者的喜好影響也是相當大的。

在 midi 中共定義了 128 種樂器，因此我們可以利用一個 128 維度的 vector  $T$  紀錄音樂所使用的樂器，0 代表沒有使用該樂器，1 代表有使用，每次都根據使用者所勾選的結果計算  $T_f = T_{\text{relevant}} - T_{\text{non-relevant}}$ ，與資料庫中的音樂比對計算相似度，相似度公式如下：

$$\text{Sim}_T(T_f, T_i) = \frac{\sum_{i=1}^{128} T_f(i) * T_i(i)}{\sqrt{\sum_{i=1}^{128} T_f(i)^2} * \sqrt{\sum_{i=1}^{128} T_i(i)^2}}$$

而由於音色 vector  $T$  的維度遠大於音樂特性 vector  $M$ ，所以如果直接將兩個 vector 結合成一個，會導致音色的重要度遠大於音樂特性的重要度，因此，新的相似度值 =  $\alpha * \text{Sim}_M(M_a, M_n) + \beta * \text{Sim}_T(T_f, T_i)$ 。因此， $\alpha$  代表音樂特性的重要度， $\beta$  代表音色的重要度，且  $\alpha + \beta = 1$ 。

### 4. 實驗結果

以下將根據不同的情緒／情境來調整  $\alpha$ 、 $\beta$ ，並藉由紀錄每次 feedback 產生的結果，除了觀察音色重要性對各情緒／情境的影響外，也希望驗證藉由 weighted vector  $M_a$  與 vector  $T_f$  是否能逐漸趨近使用者的需求。另一方面，我們也將探討情緒／情境彼此間是否有關聯。這裡所指的情緒是使用者聽了音樂以後的感覺，例如，情緒 happy 指的是使用者聽了音樂以後的感覺是 happy。而情境指的是在某個情境

下想要聽的音樂，例如情境study代表現在使用者希望能找些能夠幫助他專心讀書或學習的音樂，而情境sports則是適合在娛樂、運動時聽的音樂。

音樂資料庫中的音樂皆為 midi 檔，有七百三十八首音樂及八位受測者。情緒根據研究音樂與心理方面的專家 Montreal 大學心理系教授 Peretz Isabelle 近年來的分類分為 happy、sad、peaceful、fear 等四種。這四種情緒剛好分布在 Arousal-Valence 的四個角落（圖 2）[5][6]，過多的情緒分類會對使用者造成混淆，所以我們採用了較清楚也較廣為人知的這四種情緒分類。而在情境的分類上，我們先採用了生活中較容易碰到的情境 sports（代表娛樂、運動場合）以及 study（代表讀書、學習場合）來做實驗。實驗方法及步驟如下：

- (1) 使用者選擇希望檢索的音樂類別（情緒— happy、sad、peaceful、fear / 情境— sports、study）。
- (2) 系統會先隨機選出若干首音樂給使用者聆聽。
- (3) 使用者可以勾選出一首以上符合自己需求的情緒 / 情境音樂。此時可以根據使用者勾選的音樂數量除以系統 output 的音樂數量得到 precision。
- (4) 系統會根據使用者的 feedback 建立 vector  $M_a$  以及 vector  $T_f$ ，同時對 features 做 re-weighting。
- (5) 重新與資料庫中的音樂進行相似度計算，同時將分數較高的音樂 output 給使用者。
- (6) 重複(3)、(4)、(5)。

另一方面，為了驗證在情緒 / 情境音樂的檢索上我們的系統能有較好的表現，我們將與音樂檢索軟體 FindSounds Palette (<http://www.findsounds.com>) 來做比較。FindSounds Palette 是目前網路上相當著名的音樂檢索軟體，並被評選為 PC World 的 "Best of the Web"，Yahoo! Internet Life 的 "Incredibly Useful" 及 New York Times 的 "Impressive Technology"。因此我們便利用其音樂相似度檢索的功能來進行個人化的情緒 / 情境音樂檢索。而在實驗方法上與我們的系統的差別在於在步驟(3)時 FindSounds Palette 只能勾選一首音樂當作他的 feedback。雖然使用者只能勾選最符合的一首音樂當作 feedback，但該回合還是會記錄下使用者覺得

符合的音樂數量來估量 precision。

圖 3 到圖 8 為量測的情緒 / 情境音樂 precision，圖中縱軸表示系統的 precision，橫軸表示 feedback 的次數，各曲線代表在不同的音色影響度  $\beta$  下的情況 ( $\alpha + \beta = 1$ )，以及 FindSounds Palette 的表現。

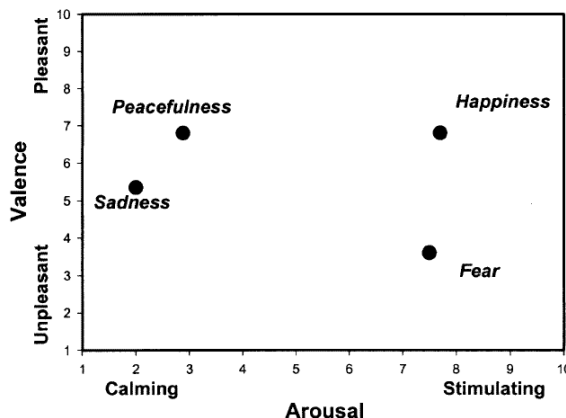


圖 2：Arousal-Valence 圖中 happy、sad、peaceful、fear 的分布情況

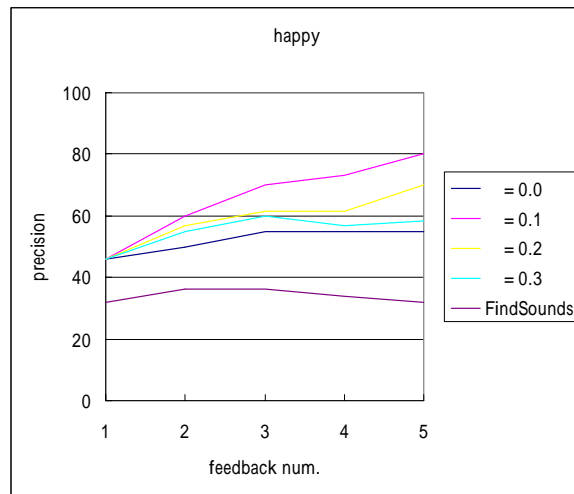


圖 3：情緒 "happy" precision

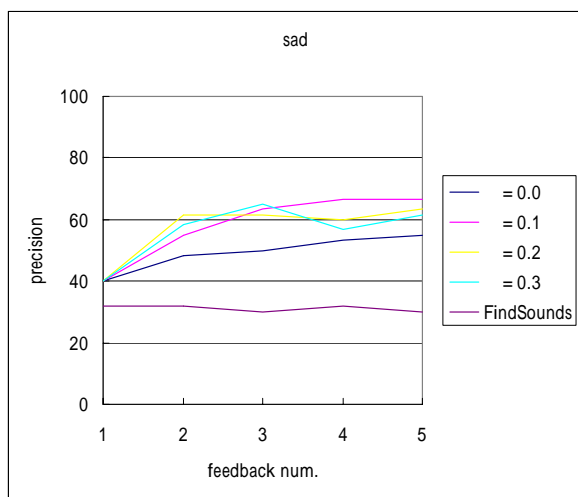


圖 4：情緒 “sad” precision

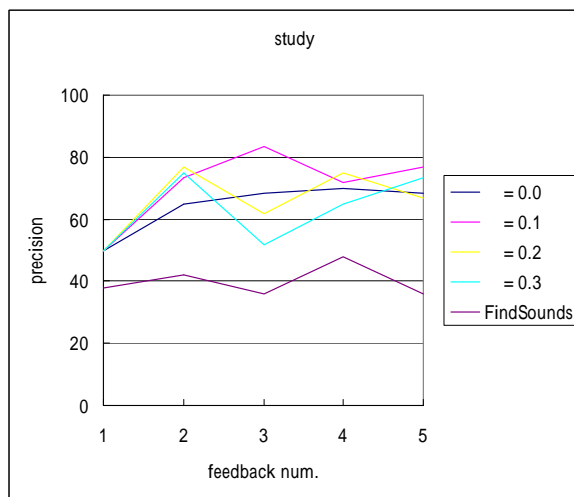


圖 7：情境 “sports” precision

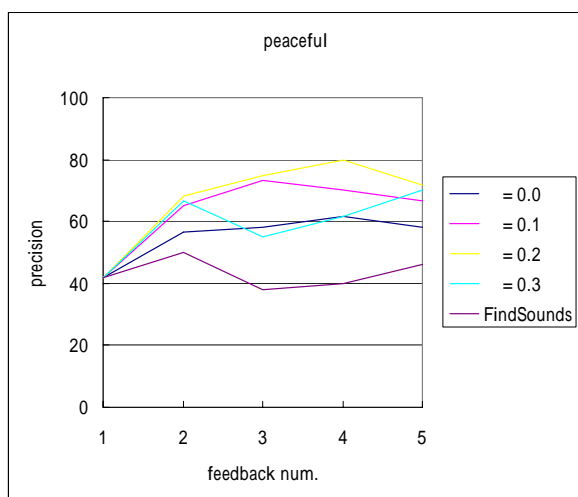


圖 5：情緒 “peaceful” precision

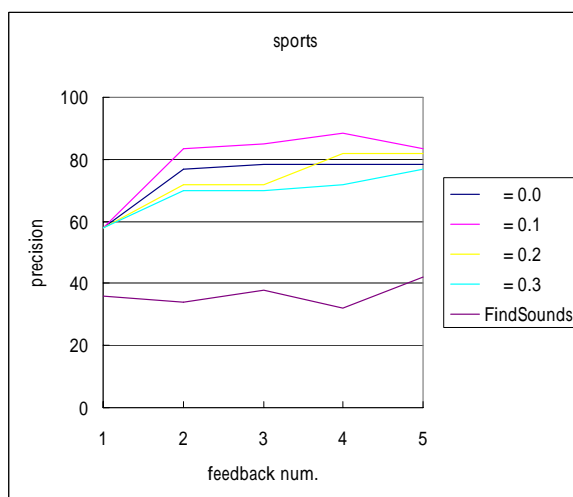


圖 8：情境 “study” precision

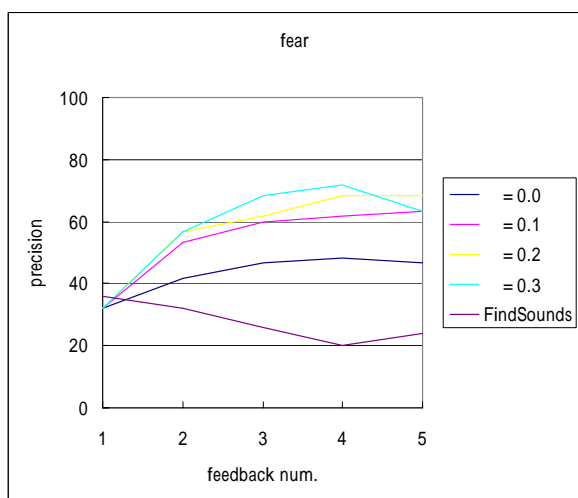


圖 6：情緒 “fear” precision

根據以上的資料我們可以歸納出一些結論：

- (1) 與 FindSounds palette 相比，我們的方法在檢索情緒／情境音樂上有相當好的效果。
- (2) 當  $\beta$  的 weight 等於零時，就代表完全不考慮音色的影響，可以很明顯地看出比起有加入音色資訊的資料正確性較低，這點也符合先前提到的音色對人的影響。
- (3) 當  $\beta$  的 weight 越來越高時，在 feedback 的過程中可能會導致正確率降低，原因主要是由於當我們把音色的重要度逐漸調高時，會慢慢忽略音樂本身的特性，而導致系統會以樂器為優先去進行音樂檢索。
- (4) 對於情緒 happy 和 sad 以及情境 sports 來說， $\beta$  的 weight 低一點比較好。但對於其他情緒或情

境而言， $\beta$  的 weight 高一點比較好。主要原因在於音樂本身特性的鑑別度高低，一般而言，happy 的音樂通常節奏較快而且為大調，而 sad 的音樂通常節奏較慢而且為小調，但是對於例如 peaceful 或 fear 的音樂則比較沒有音樂上的相關研究指出什麼樣的音樂特性會讓人產生這樣的感覺[6]。對於這種音樂特性鑑別度較低的音樂類型就必須藉助音色的幫助，例如，小提琴的聲音優美。大提琴給人親切感。而豎琴擁有精緻纖美和輕柔如流水般的聲音[10]。因此當音樂特性鑑別度較低時，音色就會顯得較為重要。

而根據我們從實驗結果得到的情緒／情境音樂 features 分布情況（表 3、表 4、表 5、表 6）分析後，可以將情緒／情境中的 happy／sports 與 peaceful／study 分布情況差異較明顯的 features 分別整理成表 7 與表 8，希望能藉此了解情緒與情境是否有相關性。例如，聽起來是 happy 的音樂是否就很適合在 sports 時聆聽？聽起來是 peaceful 的音樂是否就是能夠幫助使用者 study 的音樂？以下我們將來探討 happy／sports 與 peaceful／study 是否會呈現出相同的特性。

表 3：duration、pitch 分布情況

	D 短	D 中	D 長	P 低	P 中	P 高
sports	0.87	0.08	0.04	0.48	0.36	0.16
happy	0.86	0.11	0.02	0.44	0.40	0.16
study	0.29	0.44	0.27	0.37	0.52	0.10
peaceful	0.43	0.31	0.26	0.29	0.50	0.21

表 4：volume、BPM 分布情況

	V 弱	V 中	V 強	BPM 慢	BPM 中	BPM 快
sports	0.05	0.21	0.74	0.00	0.01	0.99
happy	0.03	0.22	0.75	0.00	0.08	0.92
study	0.22	0.59	0.19	0.00	1.00	0.00
peaceful	0.21	0.55	0.25	0.69	0.31	0.00

表 5：type of change of pitch 分布情況

	CP 先升後降	CP 先降後升	CP 漸升	CP 漸降	CP 平穩
sports	0.04	0.02	0.53	0.19	0.22
happy	0.05	0.03	0.52	0.28	0.13
study	0.02	0.02	0.29	0.32	0.35
peaceful	0.05	0.03	0.31	0.26	0.34

表 6：type of change of volume 分布情況

	CV 先升後降	CV 先降後升	CV 漸升	CV 漸降	CV 平穩
sports	0.09	0.16	0.28	0.32	0.15
happy	0.12	0.21	0.22	0.37	0.07
study	0.13	0.14	0.31	0.30	0.11
peaceful	0.10	0.11	0.26	0.28	0.25

表 7：情緒／情境 - happy／sports 的關聯

	bpm	Change of pitch	Change of volume
Happy	快	漸降	較有變化
Sports	更快	較平穩	較平穩

表 8：情緒／情境 - peaceful／study 的關聯

	duration	bpm	Change of volume
Peaceful	較短	較慢	較平穩
Study	普通	普通	較有變化

happy／sports 與 peaceful／study 這四種情緒／情境都是屬於一個正向的情緒（Valence 較高），只有程度上（Arousal）有所差別，而從表中我們可以發現到 sports 的音樂比 happy 的音樂節奏更快更明顯，而在 pitch 以及 volume 的變化上 happy 的音樂變化性較為豐富。而 peaceful 比起 study 的音樂音長較短而節奏較慢，而且 pitch 以及 volume 的變化較為單調。因此可以推論 Arousal 的高低順序為 sports > happy > study > peaceful，這個結論與我們所認知的現實情況也相當符合，sports 時往往需要一個節奏較強的音樂，但是通常變化性不大，例如聲音可能就是一直維持在大聲。happy 的音樂節奏雖然沒有像 sports 般完全集中在快，但還是屬於一個

較快節奏的音樂，而且音樂的變化性也比較大。到了 study 的音樂節奏就開始明顯變慢，但是音樂的變化性還是較高。最後，peaceful 的音樂節奏就已經完全放慢了，而音樂的變化性也變得較小，除了在 pitch 上有些變化外，volume 已經是集中在小聲區而變化較小了。

## 5. 結論

在本篇論文中，我們根據使用者的feedback建立起weighted vector  $M_a$ 與vector  $T_f$ 來表示該名使用者當下的音樂喜好，藉此找出音樂資料庫中符合使用者需求與喜好的音樂。另一方面，由於我們使用features的分布情況與重要度來表示使用者的音樂喜好，因此我們可以藉由觀察系統中紀錄的使用者音樂喜好來幫助我們了解為什麼該名使用者會喜歡這樣的音樂（表3、表4、表5、表6、表7、表8）。對於一些較難分類的音樂將有很大的幫助，例如，我們通常只知道快樂的音樂大多是大調，而悲傷的音樂較常為小調，然而對於其他情緒／情境甚至更多類別的音樂而言，我們是很難了解其分類特性的。但是藉由觀察系統中紀錄的使用者音樂喜好，將有助於我們了解音樂特性對於使用者感受的影響。同時，由於是使用features的分布情況來表示音樂資料，所以未來在features expansion上也會變得相當容易，只需要將新的features加入音樂vector中就可以了。

在未來研究方面，除了可以探討更多音樂與情緒／情境的關係；也可以運用在音樂治療之中，針對不同的病症與病患進行資料的分析。同時，也可以與心理系方面合作，除了可以蒐集更多相關領域的研究資料外，也可以利用生理測定儀等儀器來測定使用者生理反應以獲得一個更客觀的實驗結果，並可加以觀察是否能利用音樂幫助使用者抒發情緒、營造情境，讓使用者能更有效率地達成心中的目標。

## 參考文獻

[1] A. Chen, M. Chang, J. Chen, J. L. Hsu, C. H. Hsu, and Spot Y. S. Hua, 2000. "Query by Music Segments:

An Efficient Approach for Song Retrieval." *Proceedings of the IEEE International Conference on Multimedia and Expro*, 873-876.

[2] C. C. Liu, J. L. Hsu, and A. Chen, 1999. "An Approximate String Matching Algorithm for Content-Based Music Data Retrieval," *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, 451-456.

[3] D. Liu, L. Lu, and H.J. Zhang, 2003. "Automatic Mood Detection from Acoustic Music Data." *International Symposium on Music Information Retrieval*.

[4] D. Yang and W.S. Lee, 2004. "Disambiguating Music Emotion Using Software Agents." *International Symposium on Music Information Retrieval*, 218-223.

[5] N. Gosselin, I. Peretz, M. Noulhiane, D. Hasboun, C. Beckett, M. Baulac, and S. Samson, 2005. "Impaired recognition of scary music following unilateral temporal lobe excision." *Brain*, 128, 628-640.

[6] S. Khalifa, I. Peretz, B. Jean-Pierre, and R. Manon, 2002. "Event-related skin conductance responses to musical emotions in humans." *Neuroscience Letters*, 328, 145-149.

[7] T. Li and M. Ogihara, 2003. "Detecting emotion in music," *International Symposium on Music Information Retrieval*, 239-240.

[8] Y. Feng, Y. Zhuang, and Y. Pan, 2003. "Music Information Retrieval by Detecting Mood via Computational Media Aesthetics," *IEEE/WIC International Conference on Web Intelligence*, 235-241.

[9] 游恆山編譯(1990), Philip G. Zimbardo 原著, 心理學, 五南圖書.

[10] 孟憲福主譯, 殷于涵審校(2004), S. Sadie and A. Latham 原著, 劍橋音樂入門 (The Cambridge Illustrated Guide of Music), 果實出版, 城邦文化發行.