

## 3. 實驗與討論

### 3.1 實驗環境

由於音樂本身屬於藝術的範疇，並沒有所謂絕對正確的標準答案存在，音樂的情緒為音樂所帶給聽者的感受，因此我們採用測試者聆聽測試資料所標記的答案做為標準答案，在此，我們的測試者為具有音樂背景的人。實驗測試平台為 Windows XP (Pentium4 2.4G/512MB)，其他基本設定如下：

#### 3.1.1 古典音樂

- 來源：由網路上蒐集的 MIDI 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中
- 內容：包含巴洛克樂派、古典樂派、浪漫樂派時期的作品
- 屬性：MIDI 包含單軌多軌
- 答案：由兩位鋼琴學齡 10 年以上的專家根據聆聽結果標記四類情緒
- 訓練資料數目：200 首  
(60 首生氣勃勃、32 首焦慮、46 首令人滿足、62 首沮喪)
- 測試資料數目：50 首  
(15 首生氣勃勃、7 首焦慮、10 首令人滿足、18 首沮喪)

#### 3.1.2 流行音樂

- 來源：由網路上蒐集的 KAR 格式歌曲再經程式處理轉為便於比對的格式後存於資料庫中
- 內容：包含近年發行過的中文流行專輯歌曲
- 屬性：多軌
- 答案：由兩位鋼琴學齡 10 年以上的專家根據聆聽結果標記四類情緒
- 訓練資料數目：200 首  
(65 首生氣勃勃、33 首焦慮、40 首令人滿足、62 首沮喪)
- 測試資料數目：50 首  
(12 首生氣勃勃、6 首焦慮、10 首令人滿足、22 首沮喪)

### 3.1.3 歌詞

- 來源：由上述 200 首 KAR 檔案抽取出的歌詞檔案，經處理為便於比對的格式存於資料庫中
- 內容：包含近年發行過的中文流行專輯歌曲
- 屬性：TXT 文字檔
- 答案：由兩位鋼琴學齡 10 年以上的專家根據聆聽結果標記兩類情緒，由於歌詞可能會和歌曲產生不同情緒，在此分為只聽音樂、只看歌詞、同時聽兩者標記。並針對快樂及焦慮兩類標記一組 0~1 之間的答案做為 Fuzzy KNNR 驗證的答案
- 訓練資料數目：200 首(98 首快樂、102 首焦慮)
- 測試資料數目：50 首(18 首快樂、32 首焦慮)



## 3.2 實驗一：針對古典音樂的情緒辨識結果

(1)實驗目的：比較 KNNR、GMM、SVM，三種分類器的效能

(2)資料內容：50 首 midi 檔案

15 首生氣勃勃、7 首焦慮、10 首令人滿足、18 首沮喪

特徵選取出的最佳特徵為：( 速度, 調性 )

(3)參數設定：GMM 最佳 Mixture 數由圖 8 得知為 8；

SVM 最佳參數 Cost 設定為  $2^{-4}$

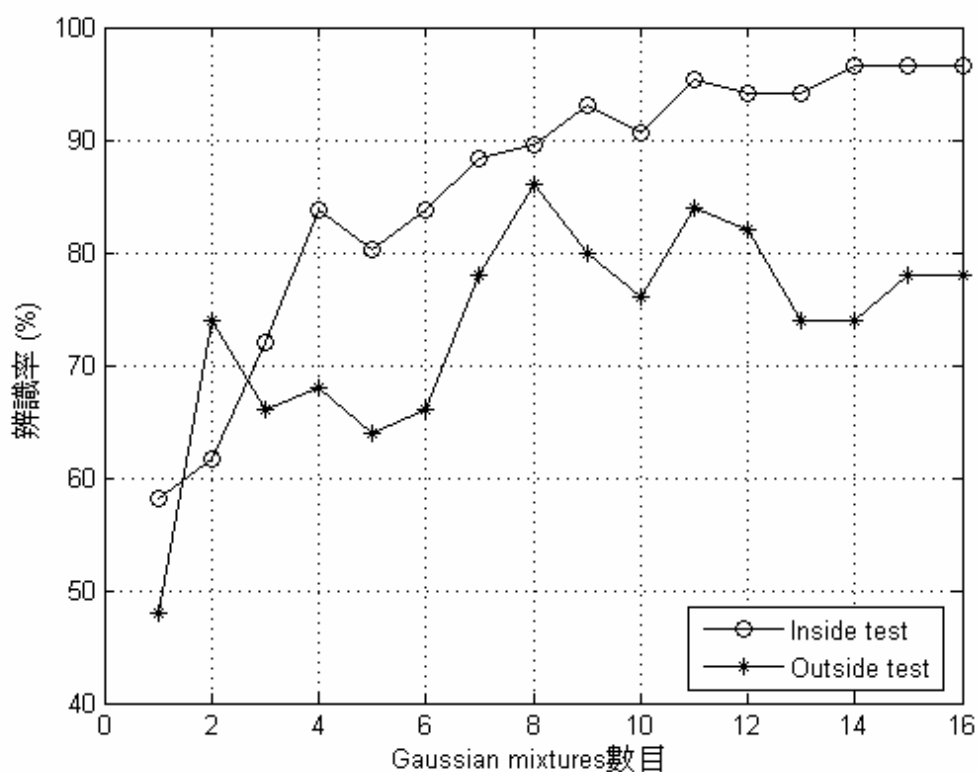


圖 8：辨識率對高斯機率密度函數之個數的關係(古典音樂內容)

(4)實驗結果：

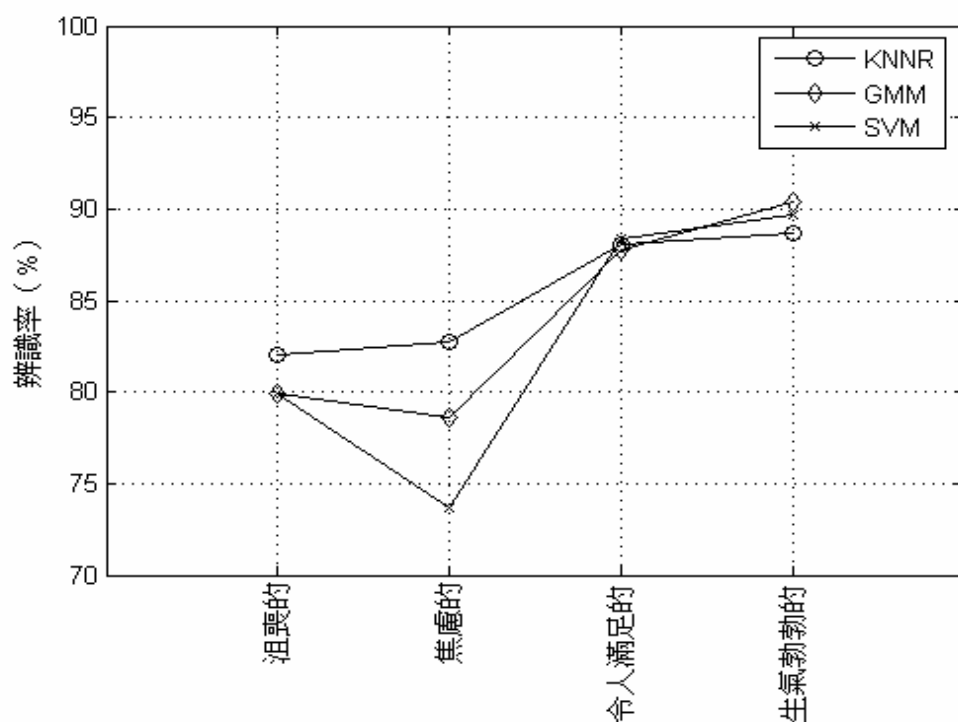


圖 9：三種分類方法對古典音樂分類結果

<b>KNNR</b>	85.56%
<b>GMM</b>	84.34%
<b>SVM</b>	83.16%

表格 9：三種分類器對古典音樂分類的辨識率

#### (4)實驗推論

從實驗結果得知 KNNR 的分類效能最佳，由於我們所選出的特徵為「速度，調性」，但因為人對於情緒感覺會受到環境、心情等因素影響，所以並不是每一首音樂都能符合電腦所歸納出的分類規則，假設大調的音樂會帶給人較快樂的情緒、而小調的音樂會帶給人較悲傷的情緒，從我們的訓練資料中發現也會有小調的歌曲被認為含有快樂的情緒，透過實驗結果，我們認為使用 KNNR 的分類，較

能處理例外情形的發生。

### 3.3 實驗二：針對流行音樂的情緒辨識結果

(1)實驗目的：比較 KNNR、GMM、SVM，三種分類器的效能

(2)資料內容：50 首 kar 檔案

12 首生氣勃勃、6 首焦慮、10 首令人滿足、22 首沮喪

特徵選取出的最佳特徵為：( 速度，鼓組出現的時間 )

(3)參數設定：GMM 最佳 Mixture 數由圖 10 得知為 10；

SVM 最佳參數 Cost 設定為  $2^{-4}$

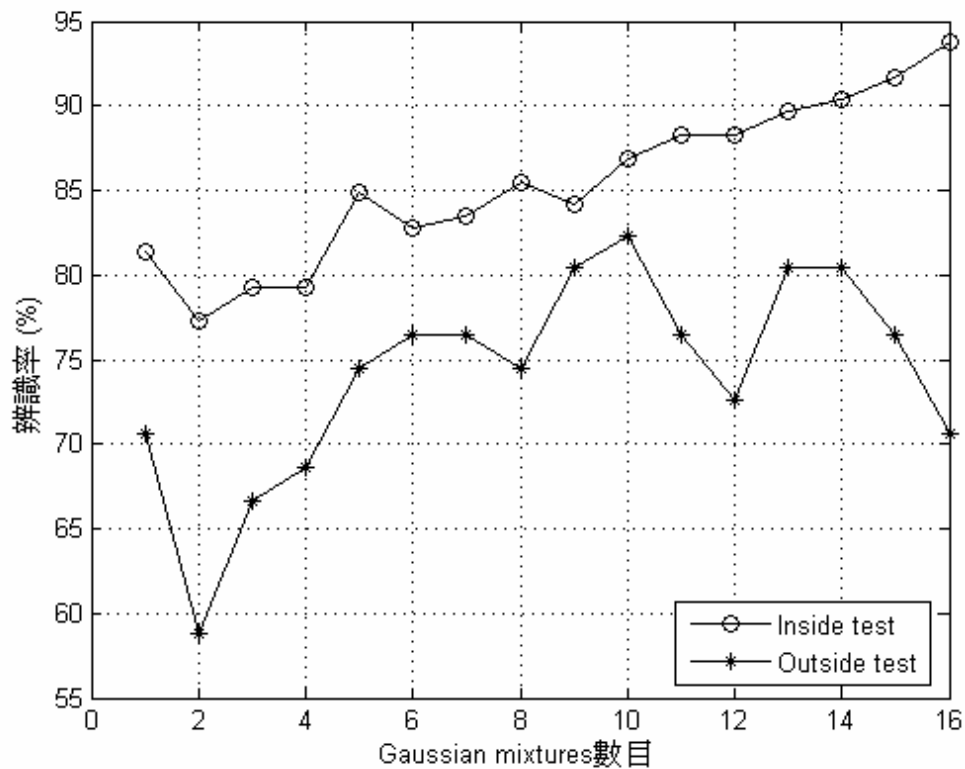


圖 10：辨識率對高斯機率密度函數之個數的關係(流行音樂內容)

(4)實驗結果：

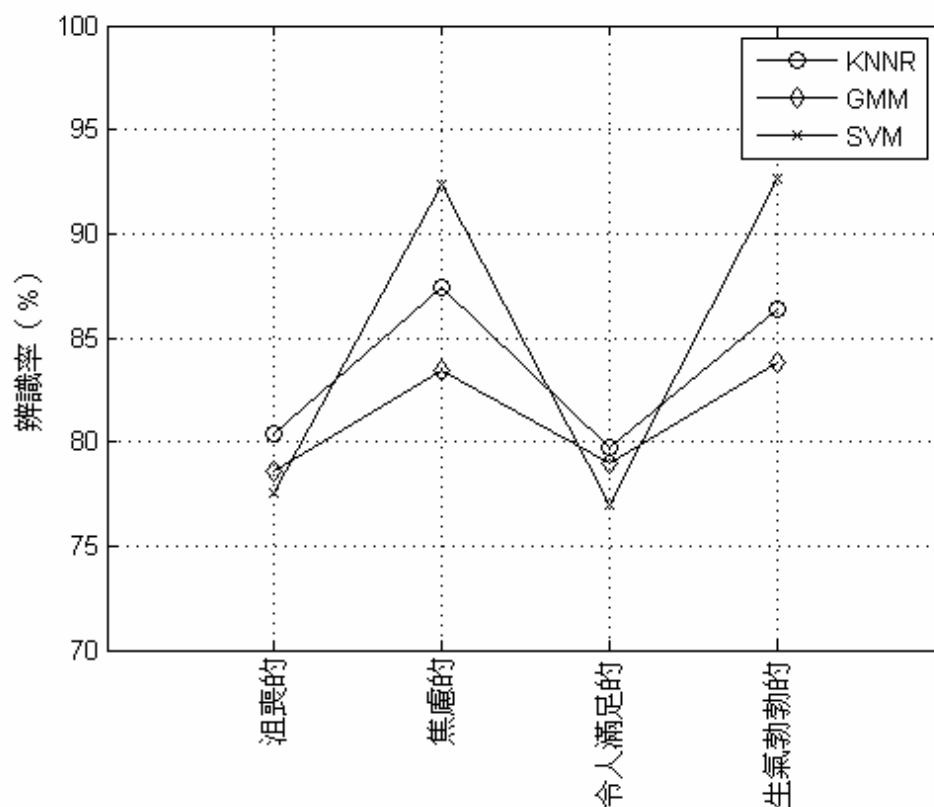


圖 11：三種分類方法對流行音樂分類結果

<b>KNNR</b>	82.44%
<b>GMM</b>	80.47%
<b>SVM</b>	79.11%

表格 10：三種分類器對流行音樂分類的辨識率

#### (4)實驗推論

從實驗結果得知 KNNR 的分類效能最佳，由於我們所選出的特徵為「速度，鼓組出現的時間」，在此我們可以推論出不同時代的音樂，含有的情緒特徵有所不同，其發生的原因可能是因為作曲者受到當時的文化、環境或是作曲的潮流所影響。古典時代的作曲家喜歡用大小調來表現快樂或是憂傷的感情，這種調性的

因子還是存在於現代的流行音樂，但卻比較不強烈。在我們的實驗中發現現代的流行歌曲受到編曲的影響較大，同樣一首歌曲會因為不同的編曲風格而產生不同的感受。

此外和實驗一一樣，因為人對於情緒感覺會受到環境、心情等因素影響，所以並不是每一首音樂都能符合電腦所歸納出的分類規則，假設在 A 段就編制鼓組的音樂會帶給人較快樂的情緒、而鼓組較晚出現的音樂會帶給人較悲傷的情緒，從我們的訓練資料中也發現到有些曲子鼓組會出現在中間，這類的情形便產生了模糊地帶，透過實驗結果，我們認為使用 KNNR 的分類，較能處理模糊地帶情形的發生

### 3.4 實驗三：歌詞計算方法的辨識結果

- (1) 實驗目的：比較三種歌詞計算方法的效能
- (2) 資料內容：50 個歌詞 txt 檔  
(18 首快樂、32 首焦慮)
- (3) 實驗結果

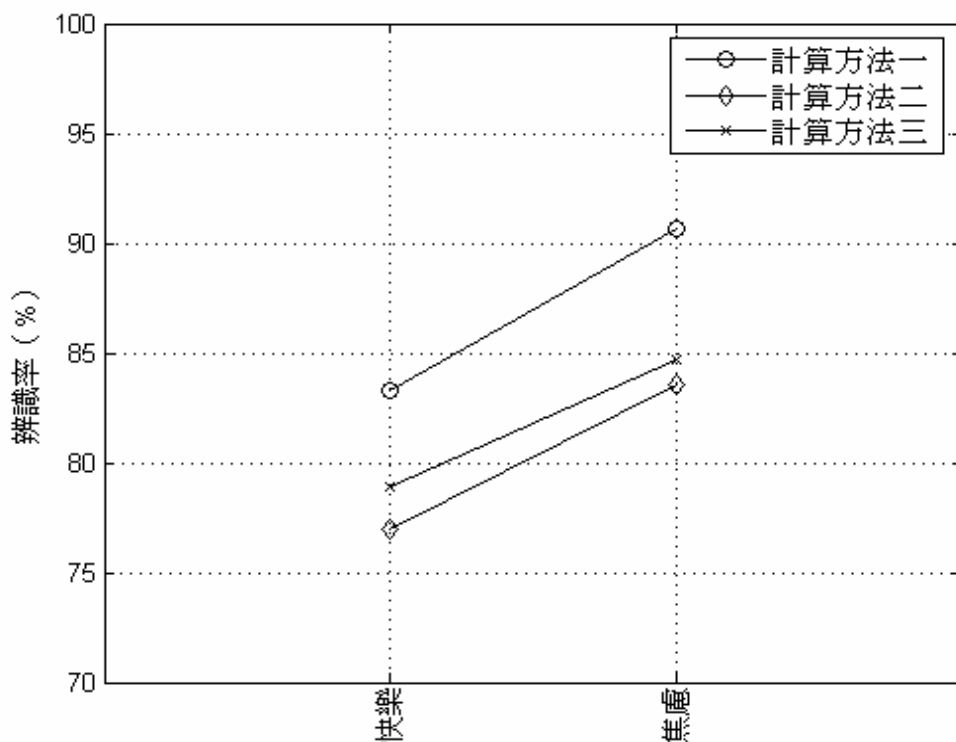


圖 12：三種計算方法對歌詞的分類結果

計算方法一	88.00%
計算方法二	78.00%
計算方法三	82.00%

表格 11：三種計算方法對歌詞的辨識率

#### (4)實驗推論

從實驗結果顯示計算方法一的效能最佳，也就是當我們視每個情緒字彙不是同等重要時會有較好的辨識率。從我們的情緒資料庫統計發現在快樂類別中，詞串「幸福」比「微笑」的詞頻高，而由於我們的資料量不足，因此我們使用人工來建立同義詞表來調整我們的資料庫，在同義詞表中，「快樂」=「開心」=「喜悅」，並採三者最高的詞頻做為最後計算的詞頻，這樣的做法可以解決資料量不足而造成部份應該對應到較高詞頻的詞串。

而計算方法三 TF\*IDF 通常用在文件分類，他可以處理同一個詞出現的多個類別的問題，但在本論文效果並沒辦法彰顯的原因可推論為類別數過少，若是有詞串同時出現於兩類則會被計算成 0，例如：愛情這個詞串同時出現的兩個類別，但其在焦慮類別中含有較高的詞頻，若是使用計算方法三，會將其視為無效的情緒詞。但從詞頻中可以發現作詞者在描述悲傷情緒時，會常用到「愛情」這個詞串，因此該詞被視為完全無效是不客觀的



### 3.5 實驗四：歌詞計算方法信心值和歌詞保留率之間的關係

(1)實驗目的：計算 2.4.2 節所提出的歌詞計算方法比例分數差距的信心值和數目之關係。

(2)資料內容：50 個歌詞 txt 檔，透過斷詞，採計算方法一算出快樂與焦慮類別的得分，再轉換成比例分數。

情緒辨識的信心值計算公式如下：

$$\text{信心值} = \frac{\text{大於門檻值且辨識正確的歌曲數目}}{\text{大於門檻值的歌曲數目}} \quad (10)$$

保留率的計算公式如下：

$$\text{保留率} = \frac{\text{大於門檻值歌曲數目}}{\text{全部的歌曲數目}} \quad (11)$$

(3)實驗結果



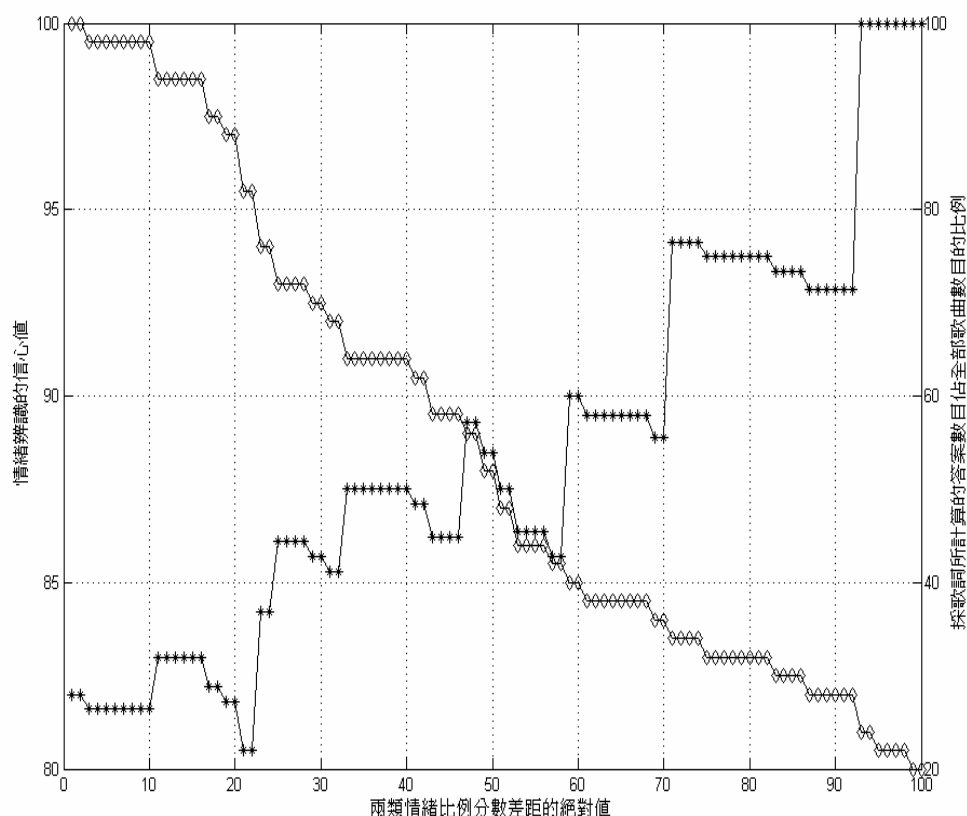


圖 13：用歌詞計算方法的信心值與保留率之關係圖

#### (4) 實驗推論

上升的曲線表示當兩類的情緒比例差異越大時，辨識的信度也會越高，由此可推論當兩類之中的某一類含有較多關鍵詞或是較重要的關鍵詞時，其辨識結果會更具有信度，我們這樣的作法可以解決當兩類分數差異過近而造成辨識效果不佳的狀況。

下降的曲線表示當兩類的情緒比例差異越大時，所保留的歌曲數目也就越少，因為不是每一首歌曲在正面跟負面的情緒詞都會有極大的對比，因此當我們想要用信心值較高的門檻值，相對的，採用歌詞計算出的答案做為辨識結果的數目也會變少。

### 3.6 實驗五：音樂和歌詞標記之比較

(1)實驗目的：比較只聽音樂標記和只看歌詞標記的結果

(2)資料內容：50 首 kar 檔

(18 首快樂、32 首焦慮)

分為只聽音樂標記、只看歌詞標記以及同時聽音樂和歌詞標記

(3)實驗結果

	快樂	焦慮	平均
音樂內容	92.52	82.35	88.02
歌詞內容	84.00	60.00	72.00

表格 12：只聽音樂的標記答案

	快樂	焦慮	平均
音樂內容	94.12	72.73	80.00
歌詞內容	88.24	87.88	88.00

表格 13：只看歌詞的標記答案

	快樂	焦慮	平均
音樂內容	90.00	80.32	86.00
歌詞內容	84.38	77.79	82.00

表格 14: 以歌詞加音樂為標記答案

#### (4)實驗推論

由於一首歌是由音樂和歌詞所結合而成的，但並非每一首歌音樂和歌詞都能表示相同的情緒，我們希望可以透過實驗來驗證我們所設計的計算方法是有效的，因此我們將標記答案分為只聽音樂、只看歌詞、以及同時聽音樂加歌詞三種來標記，以驗證我們的音樂及歌詞的計算方法。

從表 12 我們得知若只聽音樂為標記答案，以歌詞所計算出的結果較不具效益，僅有 72%的辨識率，而以只看歌詞做為標記答案時，音樂和歌詞可以得到較平均的辨識率，由此可知當聆聽者只聽音樂時，使用歌詞的計算方法較不具效益，聽者只聽音樂所感受到的情緒會和作詞者所要表達的情緒差異較大，而聆聽者在只聽音樂的時候，也深受我們所取的音樂情緒特徵所影響，約有 88%的辨識率。當只看歌詞時，使用音樂的計算方法卻可以得到和歌詞所計算出來較接近的辨識率，約有 80%的辨識率，而歌詞的計算方法也可達到 88%的辨識率。從實驗中可以推論出當聽者看到歌詞所感受到的情緒會跟這首歌的音樂所要表達的情緒較接近，人對於歌詞所表達的文字情緒主觀性較低。

而表 14 的實驗結果可以看出當我們聽一首歌會同時受到音樂及歌詞的影響，其中又以音樂的特徵影響性大一點，約勝於歌詞 4%的辨識率。

### 3.7 實驗六：歌詞結合音樂計算方法的結果

(1)實驗目的：計算歌詞結合音樂計算方法的結果

(2)資料內容：50 首 kar 檔

(18 首快樂、32 首焦慮)

採用同時聽歌詞和音樂的標記答案

(3)實驗結果：

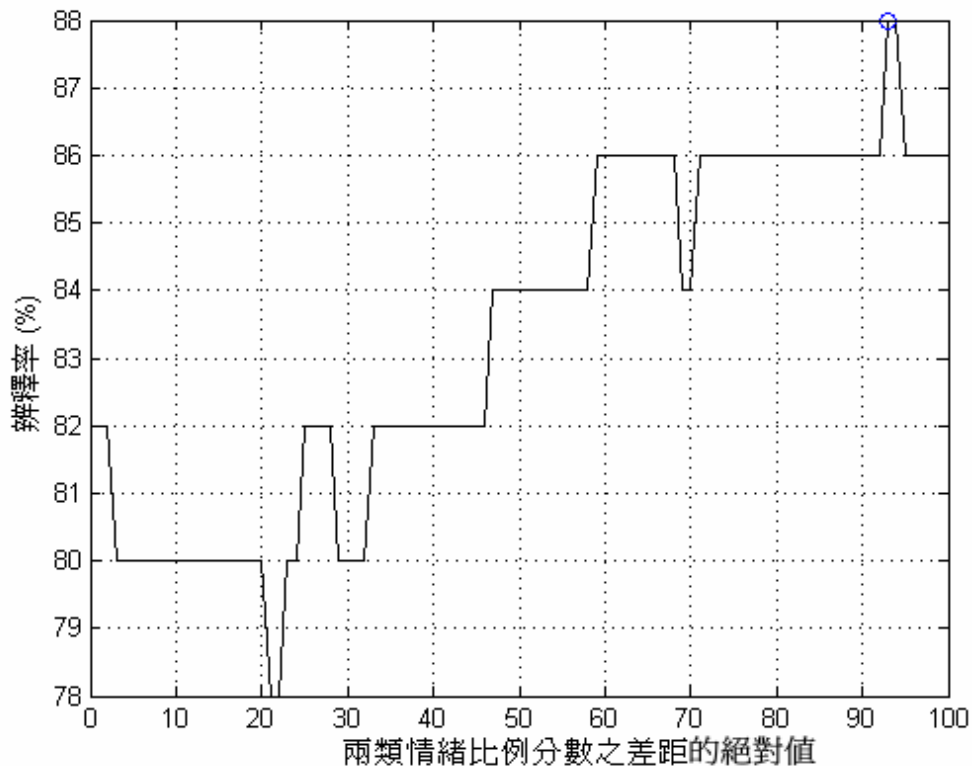


圖 14：歌詞輔助情緒辨識的結果

(4)實驗推論

從圖 12 我們使用窮舉法希望可以找出情緒比例分數差距的最佳門檻值，當兩類情緒比例分數相減大於 0~100 之間的某個值時，我們會認定這首歌透過歌詞計算出的答案比音樂計算出的有效益，因此採用歌詞計算出的答案做為標準答案。以差距 0 為例，當兩類情緒比例分數的差距大於 0 時，我們採用歌詞所算出的答案，也就代表所有的測試資料都採用歌詞所計算出的答案，在此辨識率為 82%；而當差距為 100 時，則表示當兩類情緒比例分數的差距大於 100 時，因此代表

所有的測試資料都採用音樂所計算出的答案，在此辨識率為 86%時代表的意義為兩類情緒比例分數相減出來的值大於 0 時，我們採用歌詞計算出來的結果。

而透過圖 12 我們可以看到當兩類情緒比例分數的差距約在 90 時會出現辨識率的最佳值，約為 88%。從我們的實驗可以發現由音樂辨識情緒已有不錯的效果，但加入歌詞的判斷可以調整音樂主觀上的錯誤，當一首歌詞兩類差距到 60 時，已能將歌詞計算結果提高到和音樂計算結果相同，而差異到 90 時，更能表示這首歌詞有很多某一類的情緒詞或是很重要的關鍵詞，因此我們的系統會更相信差異越大所算出的答案。

### 3.8 實驗七：Fuzzy KNNR 計算結果

(1)實驗目的：使用 Fuzzy Knnr 計算出來的結果與人標記出的答案之比較

(2)資料內容：50 首 kar 檔

(18 首快樂、32 首焦慮)

兩類分別標記一組 0~1 之間的答案，答案為兩人標記的平均

(3)實驗結果：

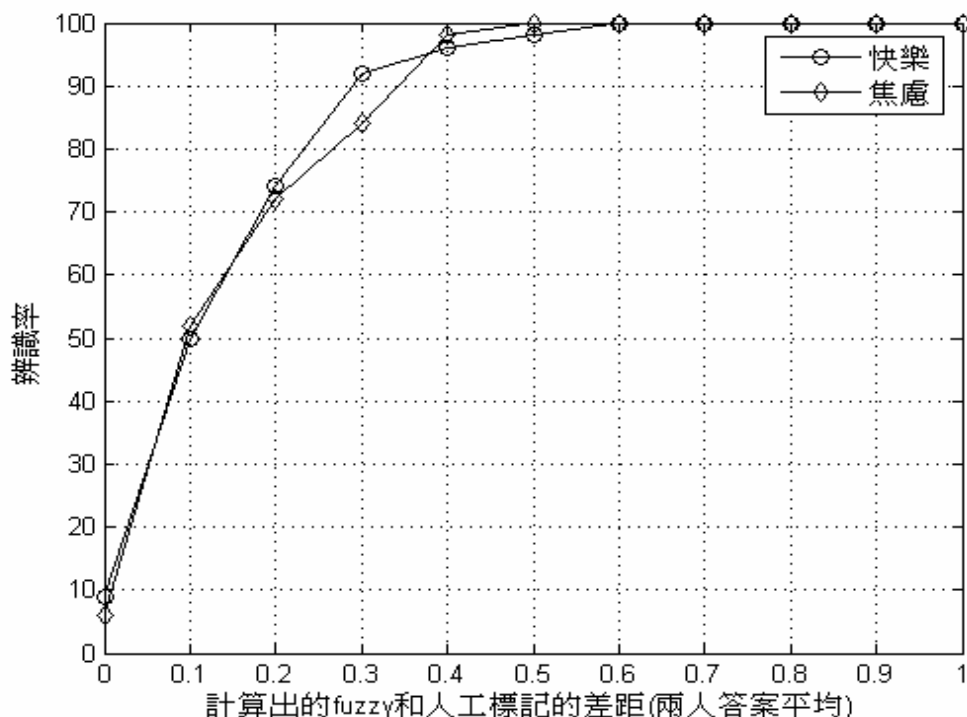


圖 15：Fuzzy KNNR 分類結果

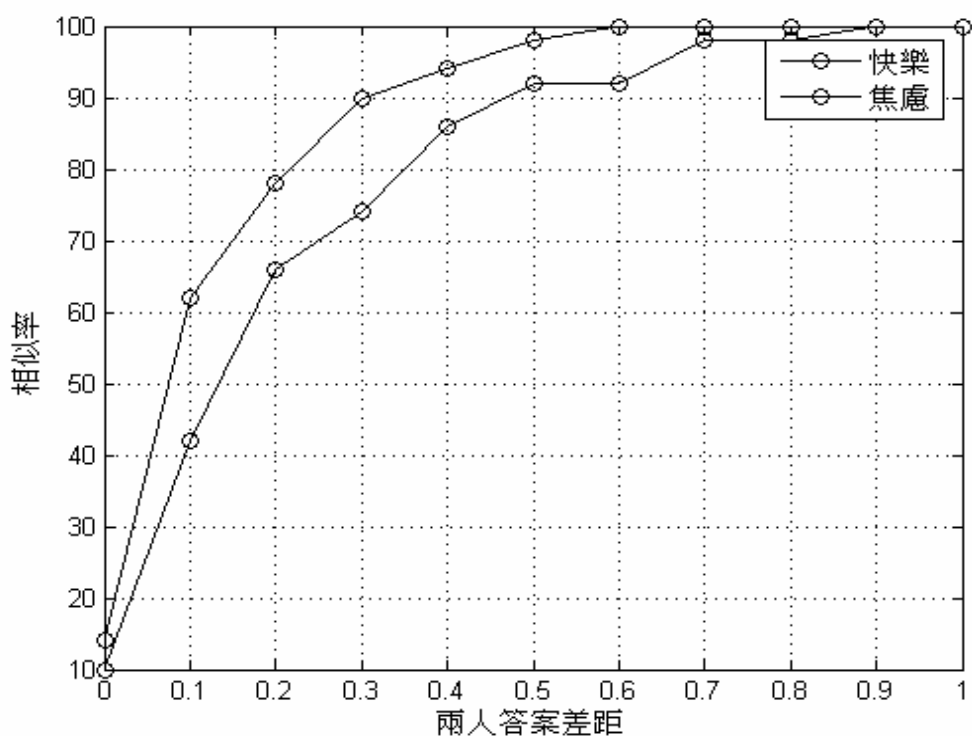


圖 16：人工標記答案差距與相似率之關係

#### (4)實驗推論

從圖 13 所表示的是 Fuzzy KNNR 的分類結果，我們可以由圖 14 看出若是人為標記的選擇已不再是二分法的 0 或 1，而是 0~1 之間的數值，兩個測試者答案的相似率很低，僅只有 15%，但由於答案已從二分法轉為十分法，其人對於答案的評斷也會有一個模糊的界線存在，假設 A 測試者認定這首歌有 0.8 程度的快樂，而 B 測試者認定這首歌有 0.7 程度的快樂，對於情感的判定是可以容許這樣誤差的存在。而傳統的二分法，卻只有 0 或 1 可以選擇，所以假設 A 測試者認定這首歌為快樂其實他可能只有 0.5 的快樂，而 B 測試者認定這首歌為快樂卻有 0.9 的快樂，這樣的作法反而會造成更大的差距，而又或者今天有一首歌他介於快樂和焦慮之間的模糊地帶，用二分法便很容易發生錯誤。

因為我們透過實驗來比較人為標記的相似率，X 軸代表兩人答案的差距，在此我們取差距值的絕對值，若是兩人的答案大於等於 X 軸對應的差距值，我們會認為兩者答案是相似的。

例如：

$X=0.1$

A 測試者認為某首歌有 0.1 的悲傷

B 測試者認為某首歌有 0.2 的悲傷

差距值=0.1

則符合 X

所以我們計算 Fuzzy KNNR 時，也用同樣的方法來比較其和人為的標記，我們可從圖 13 看到當差距值為 0.3 時，已可達到較佳的辨識率，約 88%。而圖 14 當差距值為 0.3 時，卻只有 82% 的相似率，當人為標記約有八成的相似率，我們所使用的方法可以接近九成的辨識率，我們可以推論我們的方法是正確且有效益的，若是能增加我們的測試者，應能得到更準確的結果。

## 3.9 錯誤分析

### 3.9.1 音樂內容

1. MIDI 輸入單位有誤：從網路上收集到的 MIDI 檔案大多不是以採譜為目的而製，很多 MIDI 檔案是以聆聽為考量，因此在節拍單位常出現和原曲單位不一致的情況。例如：節拍為 100 的 4 分音符可被記錄為 50 的 8 分音符，所以當我們在萃取節拍特徵時，便會產生錯誤。

### 3.9.2 歌詞內容

1. 有隱藏涵義：許多歌詞都有隱藏的涵義，例如下一次戀愛<sup>6</sup>這首歌；

還記得不停流著眼淚

對天空喊著

我恨你的那年冬季

雖然當時心情還清晰

關於恨早隨他遠去

---

<sup>6</sup> 出自本文的測試資料庫



其作詞人是要描述揮別過去的戀情，重新開始期待下一次的戀愛，應該是正向的情緒，但是就文句來看，負向的情緒字彙卻比正向強烈，因此會發生判斷錯誤。

2. 歌曲和歌詞相反：約佔 20%。例如五月天的志明與春嬌，歌曲部分表現的是生氣勃勃情緒，但是歌詞描述是分手的情緒，兩者產生矛盾。
3. 情緒詞彙分佈情況與假設不同：例如梁靜茹的分手快樂，歌詞中根據斷詞會大量出現「分手」、「快樂」造成正面和負面的情緒分數相近，而發生分類錯誤的情況。

### 3.9.3 人為與資料庫問題

1. 人對情緒的感覺不同，二分法約有 20%的差異。

由於心情會影響一個人聆聽一首歌的感覺，因此同一首歌對於不同的人，可能會產生不同的答案。
2. 人對情緒感覺若以程度來標記差異性更大。

因為以程度來標記會從二分法轉為十分法，從簡轉繁，會造成更大的差異。
3. 人對情緒的感受，有時會受到音樂的影響，有時則受到歌詞的影響，較難找出一個平衡點。

並非每一首歌曲在歌詞及音樂上都能帶給聽者同樣的情緒，因此要找出一個方法可以滿足每一首在音樂及歌詞上情緒相反的方法，會有一定的困難度，這也是因為有時音樂和歌曲產生相反是因為人為主觀的判斷。
4. 人為標記的人數不足，答案較不客觀。

由於標記歌曲需要花費一定的時間，因此我們只採用兩個人標記的答案做為判斷，但有時候會發生兩人答案有極大差異的情況，所以應該增加更多的人為標記，才能更有效的評估我們的方法。
5. 音樂及歌詞資料庫的資料量不足

而目前音樂和歌詞資料庫的資料量不足，特別是歌詞資料庫，需要更多的情

緒詞串以及由更多資料統計出來的詞頻才会有更準確的判斷。

