

# 即時性音樂情緒響應追蹤研究

學生：林立緯

指導教授：鄭泗東

國立交通大學 工學院聲音與音樂創意科技碩士學位學程

## 摘 要

音樂為什麼那麼具有吸引力的原因在於它喚起我們的情緒感受，如何從音樂訊號中擷取一套有效的方法識別情緒為本篇研究的目標，本研究認為整首的音樂不會只造成聽者產生單一種穩定的情緒，而是隨著歌曲的行進過程中，有持續的情緒變化，應該是時間性的變化過程，因此本論文提出具時變性的音樂情緒追蹤系統，追蹤音樂內容所喚起聽者的情緒感受，並記錄不同情緒的時間長度比例，比例越高的情緒，聽者的感受也就越長久且強烈。本研究參考音樂心理學的文獻並引用 Thayer 提出的情緒模型，由 200 首已標記情緒的音樂片段之 WAVE 檔中萃取五種特徵，包括音量、音樂事件密集度、音色、調性與音程的和諧度，經過情緒記分以界定在情緒平面上的分佈，用兩種分類器 GMM、BPN 進行訓練，並收錄一百首流行歌曲做為情緒辨識的測試樣本，輸出採三種呈現方式，分別有靜態的情緒追蹤、時間比例與動態的情緒可視化界面，研究最後請 66 名大學生進行共五次的音樂情緒問卷調查做為客觀的結果驗證。

# Tracking the Real-Time Emotional Response of Music Signals

student : Li-wei Lin

Advisors : Dr. Stone Cheng

National Chiao Tung University Master Program of Sound and Music  
Innovative Technologies

## ABSTRACT

The reason for the general appeal of music lies in the emotional feedback that music offers to its listeners. So, how to extract an effective approach from music signal to identify emotions response is the main purpose. This work proposes that music arouse our emotion response turn to the next story with continuous playback, it should be a real-time alteration not a static state. For the reason, research proposes a sequential framework real-time system to tracking emotions evoked by musical signals and recording the rate of each mood total time length. The higher proportion of mood means listeners feeling are more strong and long-term. The research considers psychological factors and cites Thayer's mood model reference. Five feature sets are extracted from the 200 WAV file music clips labeled mood state and applying to 「Emotion Scores Counting」 to define the distribution of sample space. Next, training two classifier (GMM, BPN) and recognize emotion by 100 Pop music testing data. Finally, the system has three methods (static emotion tracking, time-lasted rate and dynamic emotion locus visualization) to exhibit our result. Beside the researchers invite 66 college students attending total of five times music mood questionnaire survey to statistics an objective result as research verification.

## 誌 謝

兩年的時間過得非常的快，從剛入研究所什麼都不懂到現在能夠完成研究，首先最要感謝這兩年來鄭泗東老師的指導，每當遇到問題不得其門而解時，他總是可以適時的對我提供更有建設性方向，在平常的每一次討論中我也總是可以獲得啟發與更新研究的動力與建議，另外我也特別要感謝畢業的學長姐們，所謂：前人種樹，後人乘涼，研究目前的成果必須歸功他們為我們後輩提供的豐富知識，特別是俊傑學長，與他共事時我可以時時從他身上快速的學習，並且教導我，而這次的研究也是在他留下的完整基礎下繼續深入發展，此外還要感謝同學偉廷以及學弟妹船長、布丁、小李，在研究上可以互相激勵與扶持，當遇到瓶頸時還可以一起討論，而平常也能互相吐吐苦水、閒話家常，最後我也謝謝 510 的學長們，由於我們實驗室暫時沒有博士班，因此常常有問題要請教你們，謝謝你們在百忙之中還願意幫我解答問題，當然還有 510 其他同學，都是幫助我能順利完成碩士學位不可或缺的好夥伴。

# 目 錄

摘要.....	i
Abstract .....	ii
致謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
一、緒論.....	1
1.1 研究動機.....	1
1.2 研究背景.....	2
1.2.1 音訊特徵與萃取.....	2
1.2.2 分類器 .....	3
1.2.3 心理情緒反應.....	3
1.3 文獻回顧.....	4
1.4 本篇論文研究之不同處與特點.....	5
二、音訊分析方法與原理介紹.....	6
2.1 短時距傅立葉轉換.....	6
2.2 Pitch Class Profile (PCP).....	7
2.3 Gaussian Mixture Model (GMM) .....	9
2.4 Back Propagation Neural Networks (BPN).....	13
三、音樂與情緒之心理感受.....	16
3.1 音樂特徵與情緒之間的連結.....	16
3.2 音樂誘發的情緒與聽者的心理反應.....	17
3.3 音樂基本情緒模型.....	19
四、系統架構.....	20
4.1 系統流程介紹.....	20
4.2 訓練資料.....	22
4.3 特徵萃取.....	23
4.3.1 音樂事件密集度.....	23
4.3.2 音量大小.....	25
4.3.3 頻譜分佈(音色).....	25
4.3.4 大小調調式分析.....	27
4.3.5 和聲和諧度.....	28
4.4 計分方法.....	29
4.5 資料訓練收集與分類辨識.....	32
4.6 訓練資料訓練結果.....	34

五、音樂情緒可視化.....	36
六、音樂情緒研究驗證.....	39
6.1.1 測試樣本.....	39
6.1.2 範例介紹「Goodbye To Love」.....	39
6.2 測試結果.....	42
6.3.1 問卷調查動機與概要.....	52
6.3.2 問卷調查 曲目一：Dear God.....	53
6.3.3 問卷調查 曲目二：My Heart Will Go On.....	55
6.3.4 問卷調查 曲目三：Goodbye To Love.....	57
6.3.5 問卷調查 曲目四：La Vie En Rose.....	59
6.3.6 問卷調查 曲目五：Nocturne op 9 no 2.....	61
6.4 問卷調查結果討論.....	63
七、音樂情緒之應用.....	64
八、結論.....	67
參考文獻.....	68
附錄資料 A 問卷構面.....	71

## 表目錄

表 3.1.1 音樂元素與情緒指向之關連.....	17
表 4.2.1 訓練資料情緒分類數量統計.....	22
表 4.4.1 不同音樂特徵之間的相對應比例.....	29
表 6.1.1 測試歌曲明細表.....	39
表 6.2.1 測試結果(一).....	42
表 6.2.2 測試結果(二).....	42
表 6.2.3 測試結果(三).....	43
表 6.2.4 測試結果(四).....	44
表 6.2.5 測試結果(五).....	44
表 6.2.6 測試結果(六).....	45
表 6.2.7 測試結果(七).....	45
表 6.2.8 測試結果(八).....	46
表 6.2.9 測試結果(九).....	46
表 6.2.10 測試結果(十).....	47
表 6.2.11 測試結果(十一).....	47
表 6.2.12 測試結果(十二).....	48
表 6.2.13 測試結果(十三).....	48
表 6.2.14 測試結果(十四).....	49
表 6.2.15 測試結果(十五).....	59
表 6.2.16 測試結果(十六).....	50
表 6.3.1 問卷調查結果與研究結果比較表(曲目一).....	54
表 6.3.1 問卷調查結果與研究結果比較表(曲目二).....	56
表 6.3.1 問卷調查結果與研究結果比較表(曲目三).....	58
表 6.3.1 問卷調查結果與研究結果比較表(曲目四).....	60
表 6.3.1 問卷調查結果與研究結果比較表(曲目五).....	62

## 圖目錄

圖 2.1.1 單一音框頻譜分佈 .....	6
圖 2.1.2 漢明窗 Hamming Window .....	7
圖 2.2.1 單一音框的音調層級強度分佈圖 .....	8
圖 2.2.2 連續時間的音調層級強度分佈圖 .....	9
圖 2.3.1 高斯分部.....	10
圖 2.3.2 混和高斯分部.....	11
圖 2.4.1 類神經網絡示意圖 .....	12
圖 2.4.2 人工神經元構造 .....	13
圖 3.1.1 音樂與情緒連結示意圖 .....	16
圖 3.2.1 二維基本情緒模型 .....	19
圖 4.1.1 系統流程方塊圖 .....	20
圖 4.3.1 麥克傑克森之 Billie Jean 原始音樂波形 .....	24
圖 4.3.2 音訊頻譜流量進行音樂事件偵測 .....	24
圖 4.3.3 音樂事件密集程度計算結果.....	24
圖 4.4.1 每個時間點的計分流程 .....	31
圖 4.4.2 衰退函數.....	31
圖 4.5.1 三十秒音檔所產生的情緒軌跡位移結果.....	32
圖 4.5.2 擷取部分位移軌跡所提供的資料 .....	32
圖 4.5.3 訓練資料之情緒樣本分佈.....	33
圖 4.6.1 GMM 分類結果與各類別的邊界範圍 .....	34
圖 4.6.2 兩種分類器訓練結果 .....	35
圖 5.1 情緒可視化之動機 .....	36
圖 5.2 曲目剛開始之情緒軌跡變化截圖 .....	37
圖 5.3 情緒軌跡變化之過程截圖.....	38
圖 5.4 曲目結束後留下之情緒軌跡與時間比例 .....	38
圖 6.1.1 木匠兄妹-Good bye To Love 情緒辨識的結果(GMM) .....	40
圖 6.1.2 不同情緒之累積時間比例(GMM) .....	40
圖 6.1.3 木匠兄妹-Goodbye To Love 情緒辨識的結果(BPN) .....	41
圖 6.1.4 不同情緒之累積時間比例(BPN).....	41
圖 6.2.1 GMM 辨識有效比例 .....	51
圖 6.2.2 BPN 辨識有效比例 .....	51
圖 6.3.1 GEMS-9 問卷調查結果(曲目一).....	53
圖 6.3.2 系統追蹤結果情緒比例(曲目一) .....	53
圖 6.3.3 驗證結果比例對照(曲目一).....	54

圖 6.3.4	GEMS-9 問卷調查結果(曲目二).....	55
圖 6.3.5	系統追蹤結果情緒比例(曲目二).....	55
圖 6.3.6	驗證結果比例對照(曲目二).....	56
圖 6.3.7	GEMS-9 問卷調查結果(曲目三).....	57
圖 6.3.8	系統追蹤結果情緒比例(曲目三).....	57
圖 6.3.9	驗證結果比例對照(曲目三).....	58
圖 6.3.10	GEMS-9 問卷調查結果(曲目四).....	59
圖 6.3.11	系統追蹤結果情緒比例(曲目四).....	59
圖 6.3.12	驗證結果比例對照(曲目四).....	60
圖 6.3.13	GEMS-9 問卷調查結果(曲目五).....	61
圖 6.3.14	系統追蹤結果情緒比例(曲目五).....	61
圖 6.3.15	驗證結果比例對照(曲目五).....	62
圖 7.1	情緒軌跡與時間比例(HAHAHA).....	64
圖 7.2	情緒軌跡與時間比例(My Heart Will Go On).....	65
圖 7.3	情緒軌跡與時間比例(It's My Life).....	65
圖 7.4	情緒軌跡與時間比例(Dear God).....	66

# 一、緒 論

## 1.1 研究動機

音樂為什麼那麼具有吸引力的原因，在於它喚起了聆聽者的情緒感受，為聆聽者提供了情感上的回報(Marcel Zentner, 2008)。音樂一直都扮演著重要的腳色於人們的日常之中，它存在於每一種的文化之中並且發揮著突出的作用性；不論是受過音樂專業教育與否，各式各樣的音樂確實能傳遞與喚起人們各種不同的情緒反應，諸如聽到抒情歌曲會心情沮喪、聽到搖滾歌曲會感到激昂...等。此外根據過去的一份報告之中內容提到，當下流行的休閒活動如：看電視、看電影或閱讀書籍，其中聽音樂佔有最為大量的比重(Rentfrow & Gosling, 2003)。直至今日，人們依舊喜歡在生活中、工作中以音樂相伴。由於現代的音樂訊號易於儲存與記錄，電腦音樂資料的數量與種類相當豐富且普及，然而越來越龐大的電腦音樂資料庫，卻讓使用者在管理與檢索上的困難度增加，而基於音樂資訊內容做為檢索的方式，目前也有相當多不同的方向在進行研究，包括了情緒的檢索、風格派別的檢索、旋律的檢索、音訊相似度的檢索、作曲結構的檢索，其中特別是以情緒做為歌曲檢索的方式，絕對有著高度的開發價值，無論是在醫療環境、教育環境或是諮商環境，善用音樂的情感特性，幫助人們釋放情緒轉換心情(林明穎, 2009)，亦或是日常中生活環境、用餐環境、睡眠環境的氣氛營造，或是視聽廣告、動畫電影、遊戲配樂等等，都可廣為應用與受惠。因此如何從音樂中擷取出對一套方便且準確的方式，精準探覺音樂能引起的情緒反應，那麼應用的效益將是十分可觀。

## 1.2 研究背景

藉由音樂的觸發而不自覺得喚起你我的情緒，這是絕大多數人都曾經歷過的經驗，音樂比言語更能表達深入的情感；既然音樂能引發情緒反應，那麼相類似的音樂就應當能喚起同屬性的情緒，各種媒體的聲音訊號都是經過壓縮最原始波形處理後的資料，當還原回原始聲波後，由原始的波形可擷取許多不同的音訊特徵用以描述訊號的狀態或行為，透過電腦分析這些音樂的特徵即可以對音樂檔案內容有相當程度了解，而不需要請人聽完全部的音樂歌曲內容，或以此發展出許多不同的人性化的應用方式，但特徵萃取後得到的數值，並無法由電腦本身直接判定情緒，為了使電腦能夠自動的來達成我們要求的工作，此時利用分類器(Classifier)來找出各種音訊特徵之間的關連性，並且透過訓練的動作，訓練分類器來建立模式(Model)，讓電腦對萃取後的特徵可以有判斷與分析的能力，本研究即以此概念進行發展並加入與音樂情緒相關的心理學派的觀念，使研究能更加符合人們的實際判斷與感受。

### 1.2.1 音訊特徵與萃取

所謂音訊特徵就是指聲音訊號行為模式的一種表現方式，就如一般聽者在聆聽音樂時的感受一樣，常見於各式各樣的作品中或音樂檔案中，例如：音量大小、節奏速度與模式、音色、和聲與旋律…等。而音訊特徵在時間分佈上有尺度的不同，小的尺度音樂訊號的數值特徵，如：短時距頻譜(Short-time spectrum)與其幾何分部或對比、過零率(Zero Crossing Rate)、平均靜音比率(Average Silence Ratio)…等。大的尺度也就是一般人可以直接感受到的音樂特徵，如：節奏、旋律、調性…等，通常大尺度的特徵可由小尺度的特徵做平均統計或是變化趨勢分析來找出。從訊號中運算得出的訊息推演成為高階特徵中的各種樂理表現，利用不同時間之特徵的變化即可以對音訊做進一步的分析或應用處理，以此概念進行分析特別適合用於現代豐富又多元之多聲部(Polyphonic)的音樂。

### 1.2.2 分類器

利用音訊波形的數值資料可以量化估算這些聽覺上的感知特徵，但特徵萃取後得到的數值，並無法由電腦本身直接判定情緒，因此在分析上需要進一步界定其數值的規範並加以分類，即界定不同的數值大小可能代表的特徵強度與類別，換句話說就是給予電腦一個經驗法則與價值觀的判斷能力。這個能力可以由人給予電腦，但比起直接人們依照自身經驗法則或實驗結果定義臨界值(Threshold value)制定，另一個更好的方法為利用分類器來找出類別的辦法。分類器的作用為將輸入資料以特定方式分類，藉由大量的訓練資料(Training data)將分類器參數達到最佳化，再由訓練的結果界定特徵值和類別之間的關係以對於新的未知資料加以分類。分類器的缺點為大部分的分類器都需要大量的訓練資料才可以找出適當的邊界，這使的分類系統的前置作業較為繁瑣與耗時。其優點是相較於自行定義的邊界，其結果更貼近原始資料庫的分類邊界，本文最後使用了兩種分類器進行測試結果的比較，因為不同的分類器會產生不同的結果，但普遍來說，若是萃取的特徵合適於分類項目且有足夠能力抓出不同分類項目的精隨所在，則對於大部分的情況下分類器皆能夠有不錯的分類效能。

### 1.2.3 心理情緒反應

想要一窺音樂對於人類的情緒變化，就得從最基本心理學的認知討論人類對於情緒的一般性反應。所謂情緒(emotion)，是指個體受到某種刺激所產生的異種身心機動狀態，是一種主觀意識的經驗，且不易自我控制，並且伴隨動機性行為產生的(張春興，2004)，每個人自出生後就擁有許多原始的情緒反應，並且經由後天的環境與教育的影響，發展出一套屬於自己的情緒解讀與情緒反應，因此情緒一詞本身即很難有一明確的定義，而許多學者經過多年長期研究，收集了非常多的人體樣本，進而提出了其獨到情緒模型(Emotion Model)與其模型的兩軸軸度之間的推演關係，用以解釋普遍性的情緒之間的差異以及探討如何能使情緒更為有效的分類方式，與大眾的情緒反應之間的應對關係歸納

的法則，本研究參考相關心理層面因素之研究與文獻，並加入於系統設計理念之中，本文將分為音樂與情緒之間的相互關係以及人們對於音樂的情緒反應態度以及情緒基本模型三個段落加以說明該章節內容。

### 1.3 文獻回顧

本研究的主體內容延續著實驗室過去的研究並加以改良與驗證[1]，整體研究相關的論文可包含數理分析與心理分析等兩部分，在數理分析主要著重於音訊處理(Audio Signal Processing)與分類演算法，例如：文獻[7]提出了新的特徵萃取的演算法計算音樂中節拍的連續性與相對變化大小，文中以平均音訊靜音比(ASR)的平均值與標準差與平均速度用類神經網路進行辨識，文獻[8]著重於特徵萃取以及提出改善原有調性偵測的方法，文獻[9]提及了音樂的段落情緒變化概念，特徵採用了音量特徵、音色特徵、節奏特徵，並利用雙層的 GMM 音樂情緒辨識架構進行分類，文中最後以音量特徵對古典樂曲進行段落標記，每個段落可分屬不同情緒類別，執行完整曲目的情緒追蹤，文獻[10]提出了多重情緒類別(Multilabel)的分類概念，文中測試的音樂不只屬於單一的情緒種類，而是複合式的情緒組合，最後用 SVM 辨識，文獻[11]討論主題是音樂的風格辨識，但風格與情緒之間關聯性很高，且文中收入了許多得音訊特徵都值得作為參考。

在心理分析上主要專注於音樂心理學(Music Psychology)，例如文獻[5]認為音樂是個非常好的誘發情緒反應的媒介，經由大量的問卷分析與複雜的實驗過程，得出了許多我們無法自知的情緒認知，其中有提到主張人的情緒應該是豐富且多元，並且可以同時存在於同一時刻，文獻[12]清楚解釋從音樂表達至人們產生情緒運作模式以及音樂特徵與情緒之間的相互關係，文獻[13]提出泛用的音樂情緒模型定義不同情緒之間的相對定位，模型中收錄了大量的情緒形容詞並且分為八大類別，而文獻[14]則是專注於情緒起伏，文中以情緒軸度中的 Arousal 軸的相位變化，做為古典音樂段落的分配。

#### 1.4 本篇論文研究之不同處與特點

在本篇研究中認為整首的音樂並不是只會造成聽者產生單一穩定情緒，而是隨著歌曲的行進過程中，有起、承、轉、合，高潮迭起等持續的情緒變化，這應該是個時間性的變化過程，因此當歌曲於某種情緒持續的時間越長，就越能夠傳達給聽眾那樣的情緒感受，如 Lie Lu [9] 等人之研究中有類似的概念，過去大部分的音樂情緒辨識與情緒檢所的相關研究中，會分析整段音樂的資訊內容，進而做出情緒歸屬的判定，並以此總結為該音樂帶給聽者的”心情”，忽略了音樂內容中的情緒變化情況，且情緒的判定牽涉了許多主觀的人為因素與個人背景或經驗的影響；因此基於前述的理由，本研究並不刻意去指出該音樂屬於何種情緒的分類，而是建構在即時的情緒追蹤，並搭配情緒可視化界面上持續的追蹤音樂的情緒變化起伏，最後以比例分配方式提供整首曲子於時間上的情緒比例分配情況。

## 二、音訊分析方法與原理介紹

聲音的產生為空氣的壓力變化所產生的波動，或稱為聲波。實際上聲波為一種類比訊號是連續不間斷的訊號，因此會經過取樣而轉換成為離散的數位訊號，而樣本點間彼此的時間間隔之倒數即為此訊號的取樣頻率(Sampling frequency)，對於音訊分析會先取音框(frame)，如此一來便可以對音框內之音訊進行時域(time domain)與頻域(frequency domain)的分析。以下介紹本研究所使用的時頻域分析方法。

### 2.1 短時距傅立葉轉換

取音框後單一音框的聲音訊號，其頻譜可由短時距傅立葉轉換(Short time Fourier transform)計算，短時距傅立葉轉換即為配與特定權重函數的離散傅立葉轉換(Discrete Fourier transform)，單一音框的頻譜的圖形如圖 2.1.1，可以看到特定時間音框寬度內音訊於各個頻率的強度大小與分布，連續的音框化再取短時距傅立葉轉換就可以看到不同時間音訊於各個頻率的強度大小與分布。

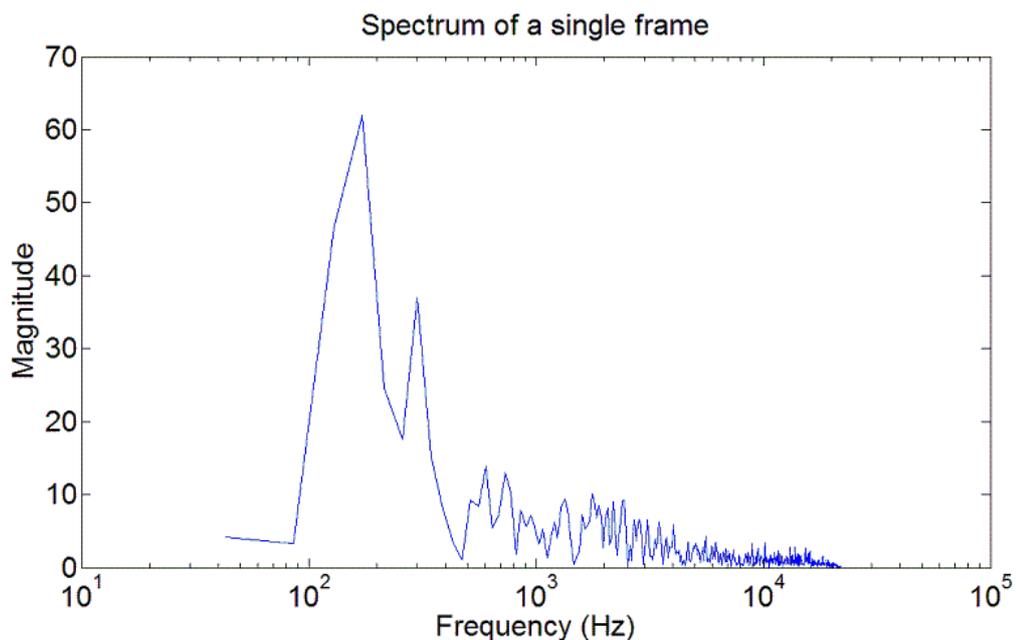


圖 2.1.1 單一音框頻譜分佈

窗函數的用意為改善音框訊號在計算頻譜時的邊界效應，減少失真使頻譜的數值對比性更好，常用的窗函數為漢明窗(Hamming window)或漢寧窗(Hanning window)，研究採用漢明窗圖形如圖2.1.2。

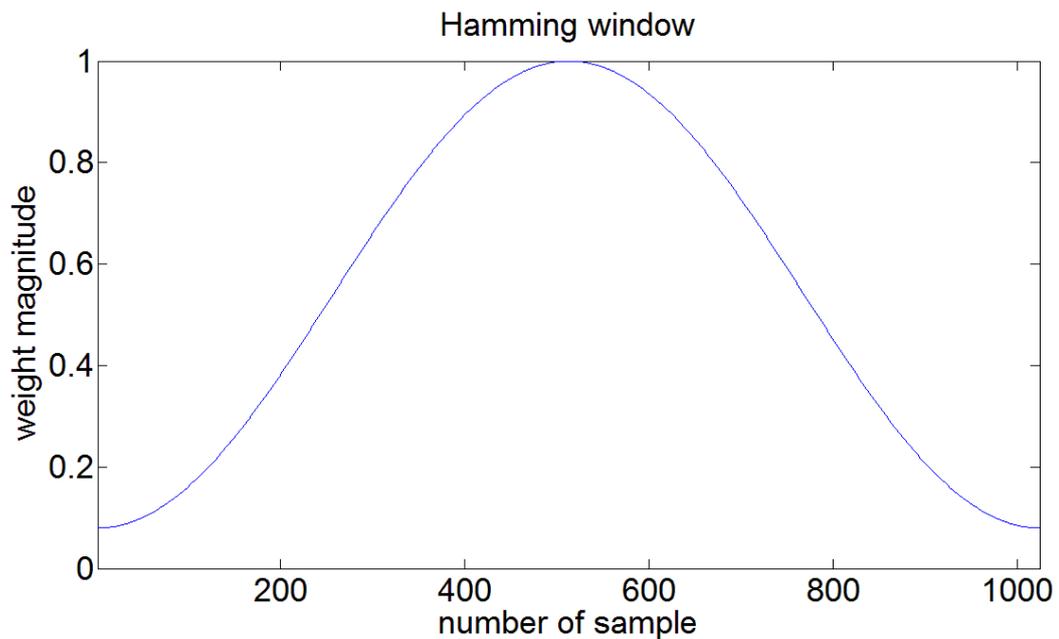


圖 2.1.2 漢明窗 Hamming Window

## 2.2 Pitch Class Profile (PCP)

由短時距傅立葉轉換得到頻譜數值後，可以進一步利用頻譜來計算一般的音樂理論分析上較常用的特徵值音調(Pitch)，音調一般以大寫音文字母 A 到 G 表示。由頻率和半音(semitone)之間的關係式可將頻率換算為音調，再利用音調於倍頻或稱八度(Octave)為相同音調層級的概念，即可將頻譜換算為對應的音調層級(Pitch Class)，如下：

$$P(k) = [24 \times \log_2((f_s / N) \times k / f_1)] \bmod 24 \quad (1)$$

$$PCP[P(k), n] = \sum_{P(k)} |S[k, n]| \quad (2)$$

上式將頻譜數值映射到 24 個音調層級上，因為考量以 12 平均律切割的 12 個音調層級在數值分析應用上不夠準確，故將每個層級中再對半切割，成為 24 個音調層級。第一式中  $k$  為頻域的樣本點數索引， $P(k)$  表示頻域和音調層級空間的對應關係，代表頻域第  $k$  個樣本點之頻率值對應的音調層級， $24\log_2((f_s/N)k/f_1)$  將第  $k$  點的頻率值換算為對應的半音數，再由餘數(mod)方式將倍頻的音調歸為同個音調層級。第二式將頻譜數值轉換到音調層級空間(PCP domain)的表示法，其中  $n$  為音框數的索引， $S[(k=0, 1, \dots, N), n]$  為第  $n$  個音框的頻譜數值， $P(k)$  為音調層級空間的樣本點數索引， $PCP[(P(k)=0, \dots, 23), n]$  則為第  $n$  個音框的音調層級數值，其為頻譜中所有倍頻的相同音調層級的強度加總。對於較為複雜的音訊，如實際的流行音樂，音調層級的表示可以看出音框內的各個的音調層級的強度與和聲架構。以音調層級的表式法，則可以對頻譜套用音樂學理上的分析方式，如音程(Interval)、旋律(Melody)、和弦(Chord)、調性(Mode)……等，各種音樂理論分析或應用。單一音框的音調層級強度分佈如圖 2.2.1，各個時間的音調層級的強度分佈如圖 2.2.2。

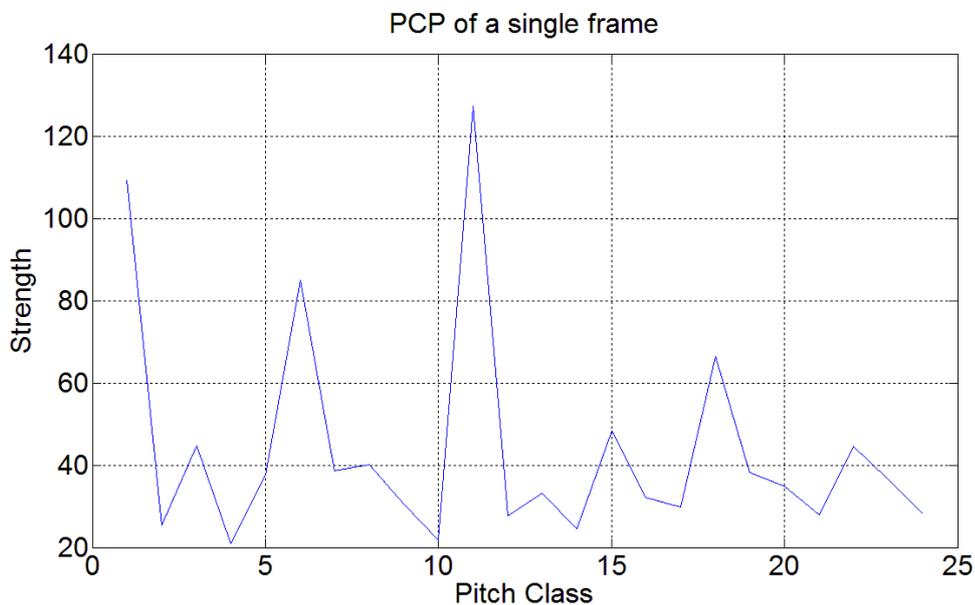


圖 2.2.1 單一音框的音調層級強度分佈圖

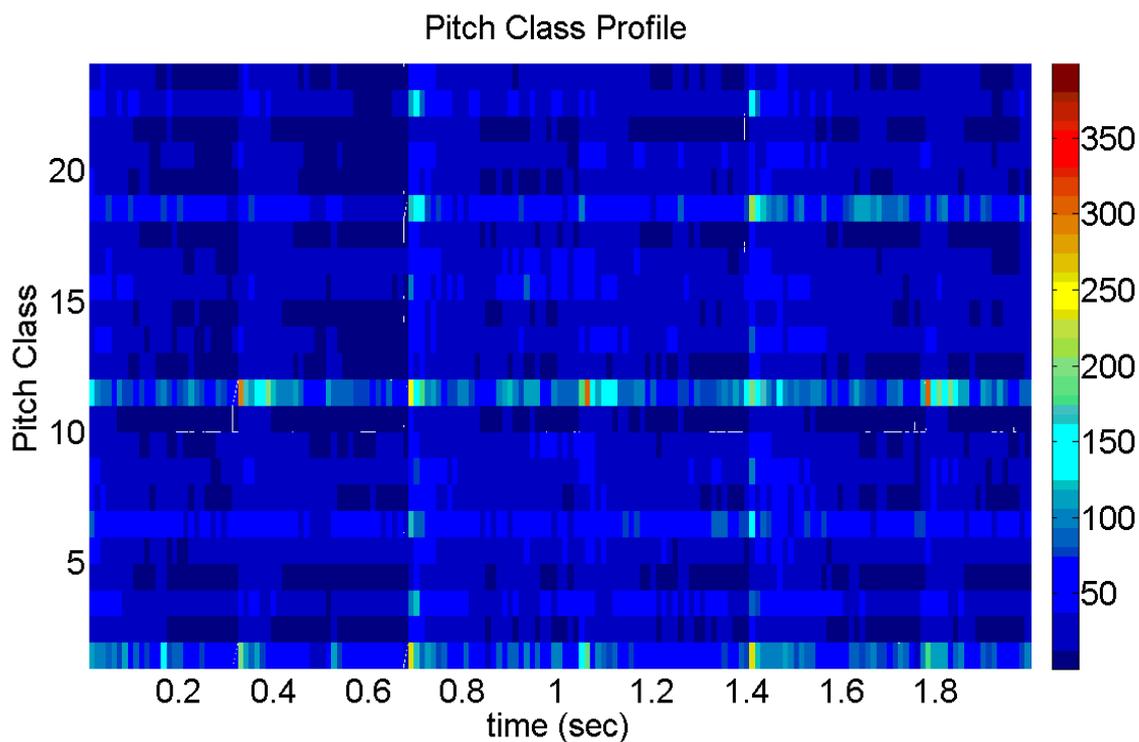


圖 2.2.2 連續時間的音調層級強度分佈圖

### 2.3 Gaussian Mixture Model (GMM)

**基本定義：**假設一個特定種類的量測值如特徵向量，是由一個可以描述該種類分部的機率密度函數(Probability density function)所產生，該機率密度說明該種類各種量測值出現的機率為何。一般而言，高斯分佈的機率密度為一種常見的分佈，一維的狀況下數學表示如下：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2((x-\mu)/\sigma)^2} \quad (3)$$

其高斯分佈曲線(Gaussian distribution Curve)圖形如圖 2.3.1 所示：

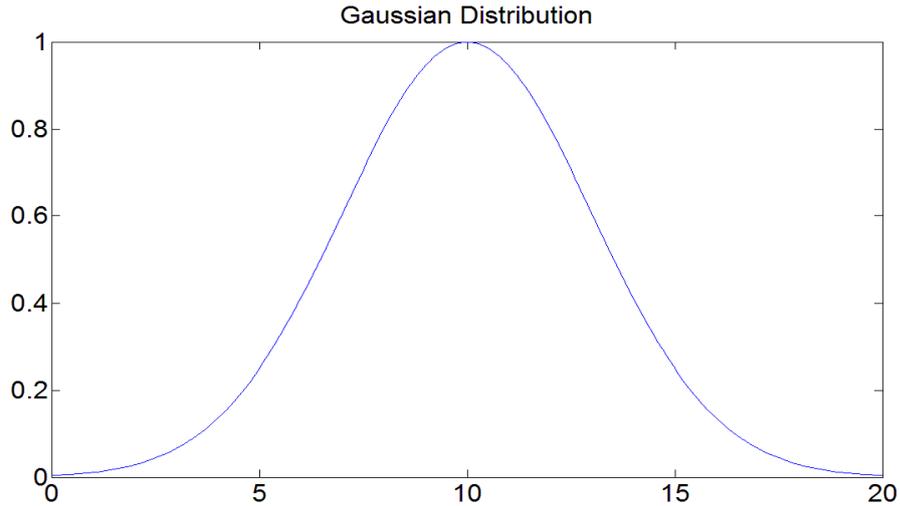


圖 2.3.1 高斯分部

其中有兩個重要的參數， $\mu$  和  $\sigma$ ， $\mu$  為期望值(Expectation value)，其為平均數，位於高斯分佈的中央， $\sigma^2$  稱為變異數(Variance)，而  $\sigma$  為標準差(Standard deviation)，其值的大小和分佈的集中程度有關，值愈小表示越集中。定義如下：

$$\mu \equiv E[x] = \int_{-\infty}^{\infty} xp(x) dx \quad (4)$$

$$\sigma^2 \equiv E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx \quad (5)$$

利用向量和矩陣推廣為高維度表示如下：

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (6)$$

其中  $\boldsymbol{\mu}$  和  $\Sigma$  分別為期望值向量和共變異矩陣(Covariance Matrix)，Covariance 是 Variance 在高維度中的一種推廣，其第  $\mathbf{i}$ - $\mathbf{j}$  個元素代表第  $\mathbf{i}$  維度和第  $\mathbf{j}$  維度的相關性，其值大於零表示正相關，小於零為負相關，等於零代表互相獨立，對角線元素就是變異數，數學定

義如下，同一維的情形，高斯分佈的參數  $\mu$  和  $\sigma$  的值會和其分佈的中心位置和曲線寬度有關。

$$\mu \equiv E[x] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_d] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = \sum_x xP(x) \quad (7)$$

$$\Sigma = \begin{bmatrix} E[(x_1 - \mu_1)(x_1 - \mu_1)] & E[(x_1 - \mu_1)(x_2 - \mu_2)] & \cdots & E[(x_1 - \mu_1)(x_d - \mu_d)] \\ E[(x_2 - \mu_2)(x_1 - \mu_1)] & E[(x_2 - \mu_2)(x_2 - \mu_2)] & \cdots & E[(x_2 - \mu_2)(x_d - \mu_d)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_d - \mu_d)(x_1 - \mu_1)] & E[(x_d - \mu_d)(x_2 - \mu_2)] & \cdots & E[(x_d - \mu_d)(x_d - \mu_d)] \end{bmatrix} \quad (8)$$

**高斯分佈與機率密度函數：**高斯分佈其在統計應用上有許多特殊性質，數據資料若集中在平均數附近，皆可以以高斯分佈做一個近似的分佈模型，因此其為一種良好的統計模型，但是並不是所有的狀況都能以單一高斯分佈描述，GMM 假設某一特定種類的機率密度函數，為數個高斯分佈乘以不同比重的疊加，圖形如圖 2.3.2，數學表示如第(9)式。

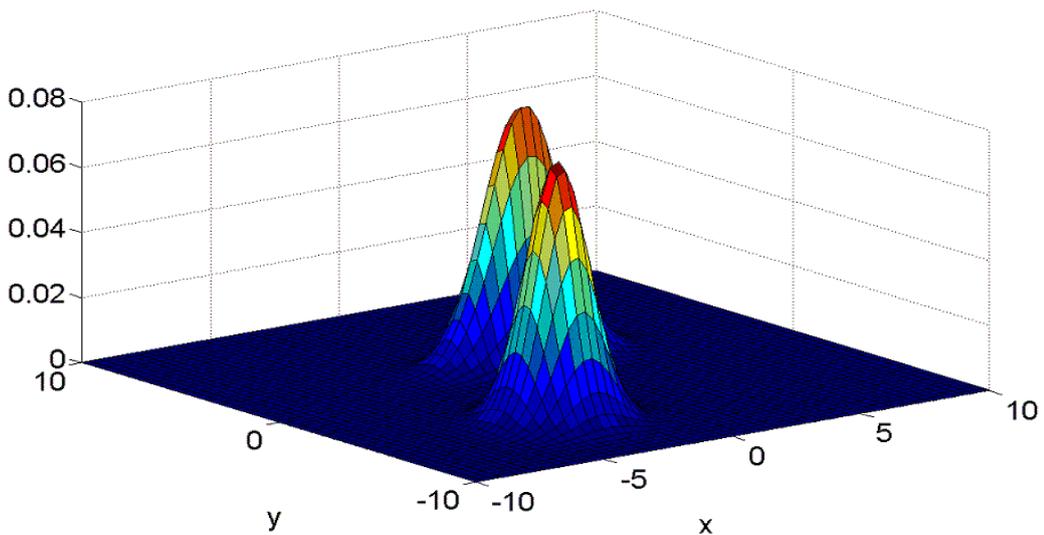


圖 2.3.2 混和高斯分部

$$p(x, y) = \alpha_1 g(x, y; \mu_1, \Sigma_1) + \alpha_2 g(x, y; \mu_2, \Sigma_2) + \alpha_3 g(x, y; \mu_3, \Sigma_3) \quad (9)$$

因此只要知道屬於每個種類的機率密度函數，很容易就可以比較同一個量測值，對應每個種類的機率的大小，進而找出可能性最大的種類為何，但由於一般待測量的數據資料，並沒有辦法得知其實際機率密度函數，找出一進似的機率密度函數的方式如下：

1. 對於每個類別，由一個初始的猜測：給定初始的高斯函數疊加個數，及每個高斯函數的參數，產生一個初始的 GMM。
2. 利用已經設定好的數據，即訓練樣本，利用 GMM 計算分類結果，找出辨識率。接下來的目標就是要使這個辨識率的值增加，辨識率越高代表這個分佈模型越能表示這些訓練樣本。
3. 以微分求極值的方式，由舊有的參數計算出一組新的 GMM 參數。
4. 重複步驟 2~3 疊代，直到辨識率收斂到某一個極值。

此法稱為 Maximum Likelihood Estimation (MLE) 或 Expectation Maximization (EM)，經由反覆疊代，找出一組最佳化的 GMM 參數，當作代表這些數據樣本的機率密度函數。對於未知種類的測試樣本，簡單比較其值對於各個種類的機率值大小(屬於該種類高斯分佈位置的高度)，就可以找出最有可能的種類為何，如此便設計了一個 GMM 分類器。

**註：**疊代過程中並不是一定會收斂到全局最大值(Global Maximum)，也有可能收斂到局部最大值(Local Maximum)。所以並不是所有的數據 GMM 都可以有很好的表現，這和初始設定的參數也都有關係，如高斯函數疊加的數目...等，想要有較好的結果，訓練樣本一定要足夠。

## 2.4 Back Propagation Neural Networks (BPN)

倒傳遞類神經網路模式是目前類神經網路（Artificial Neural Network, ANN）學習模式中最具代表性，也是使用最為廣泛的一種，藉由許許多多的人工神經元相互連結所組成的。其主要的概念是嘗試著模仿人類的神經系統的一種計算系統，藉由許多高度連結的人工神經元所組成，用來處理資訊並對外部的輸入以網路的態勢來回應，但早期 Rosenblatt（1958）發展的感知器僅為單層網路無隱藏層的存在，直到 P Werbos 於 1974 年在其博士論文[15]中提出了隱藏層的學習演算法，實現多層網路的構想才有現在的倒傳遞類神經網路存在。

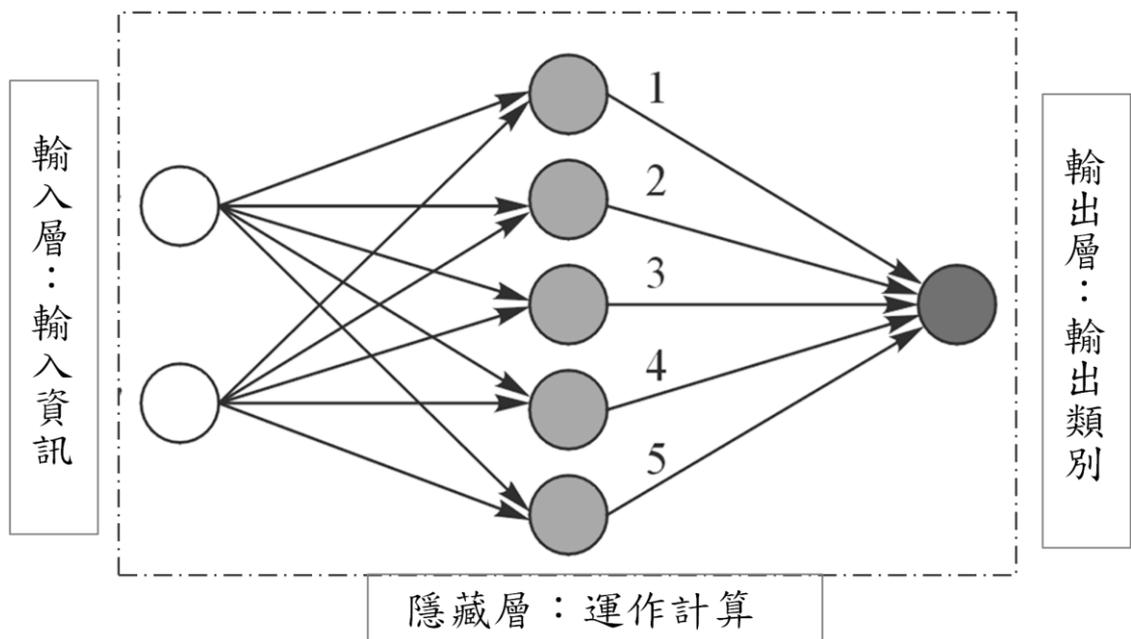


圖 2.4.1 類神經網路示意圖

若把類神經網路視為黑盒子來看的話，則此黑盒子是由許多人工神經元所組成，一般具有三層：輸入層、隱藏層、輸出層。如圖 2.4.1 所示，隱藏層的人工神經元個數可以是一個或二個以上的數目，每個人工神經元可視為一個單獨的處理器，處理器數量若為二個以上其工作方式是以並行方式運作的。隱藏層對輸入訊號的處理，其過程對網路使用者來說是看不見的。

**人工神經元**：一個類神經網路是由許多的人工神經元組成的，人工神經元是類神經網路最基本的單元，又稱為人工神經細胞、類神經元、處理單元、節點等，其內部構造如下圖所示。

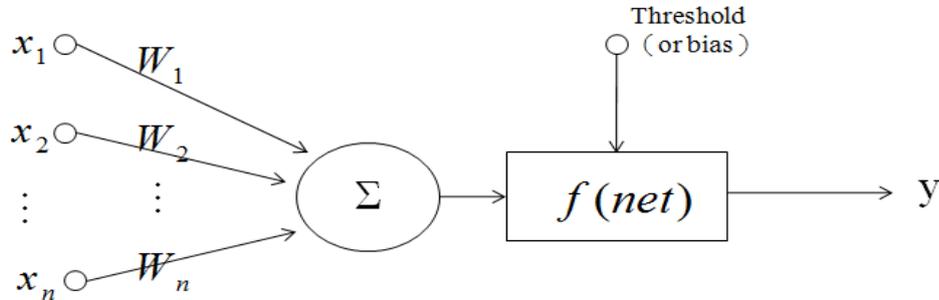


圖 2.4.2 人工神經元構造

參數說明：

$x_n$ ：稱為輸入向量，它相當於生物神經元的樹突（輸入路徑）。

$W_n$ ：稱為加權鍊結值，它相當於生物神經元的神經節強度。

$f(net)$ ：稱為轉換函數，其中的處理包括了對輸入向量的訓練，也稱為活化函數，常使用的函數有：步階函數、雙極值函數、S 形函數等。

Threshold：稱為偏權值或稱閾值。

$y$ ：稱為輸出向量，它相當於生物神經元的軸突（輸出路徑）。

**倒傳遞演算法**：以下說明如何應用倒傳遞演算法於一輸入向量  $x_j$ 、一輸出向量  $o_i$ 、一轉換函數， $f(net_i)$ ，也稱為活化函數公式(11)、與修正網路的鍊結權值  $w_{ji}$ 、 $i$  為該訓練對象的網絡神經元第  $i$  個分量、共有  $m$  個、 $j$  為訓練對象上一層的神經元第  $j$  個分量、總共數量為  $n$  個如公式(11)。

$$net_i = \sum_{j=1}^n w_{ji} x_j - \theta_i \quad (10)$$

$$o_i = f(net_i) \quad (11)$$

$$E = \frac{1}{2} \sum_{i=1}^m (t_i - o_i)^2 \quad (12)$$

由於倒傳遞網路屬於監督式的學習模式，訓練目的在於降低目標向量與輸出向量間的差距，所以通常以誤差函數  $E$  公式(12) 來表示學習品質，其中  $t_i$  為輸出層目標向量的第  $i$  個分量， $o_i$  為輸出向量的第  $i$  個分量。因此網路的學習過程就是將上述的誤差函數最小化，其方法為最陡坡降法，即每當輸入一訓練向量後，就小幅調整網路的鍊結值，調整的幅度正比於誤差函數和鍊結值的敏感程度

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \quad (13)$$

其中  $w_{ji}$  為第  $n-1$  層的第  $j$  個神經元與第  $n$  層的第  $i$  個神經元間的鍊結值。 $\eta$  則為學習速率，亦即每次改變誤差函數的步幅。利用微積分的連鎖律(chain rule)去計算可以得到，最後整理可得下式。

$$\Delta w_{ji} = -\eta (t_j - o_j) f'(net_j) x_i \quad (14)$$

$$w_{ji} = w_{ji} + \Delta w_{ji} \quad (15)$$

此為倒傳遞演算法的關鍵公式。此學習過程通常以一次輸進一個訓練範例的方式進行，直到學習完所有的訓練向量，我們稱為一個學習循環，加權值得學習過程只是將原來的加權值加上下式的改變量，經過反覆數個學習循環直到架構中各個參數值的設定使網絡整體結果達到收斂為止。

### 三、音樂與情緒之心理感受

#### 3.1 音樂特徵與情緒之間的連結

不同的音樂、不同的演奏手法、不同的樂隊編制甚至是不同的場地或節日，皆會營造出不一樣的情緒感受，使聽者有各種多變的感受，但一般認為對於音樂的鋪陳與聆聽感受中最重要的三大關鍵元素就是節奏(rhythm)，旋律(melody)，和聲(harmonic)。節奏的快慢往往會影響聽者的情緒的強度，如充滿憤怒的搖滾樂總是擁有相當快的節奏，悲傷的音樂總是擁有慢的節奏。旋律就像是音樂的外貌，特定的旋律線條通常是使人對音樂產生記憶的關鍵，如大家耳熟能詳的小蜜蜂旋律，貝多芬的命運交響曲…等。和聲則是營造音樂的音樂張力氛圍的重要角色，最主要的表現方式為和弦(chord)，其是由數個音組成並同時發聲的系統，和諧的和聲會帶給人愉悅的感覺，不和諧的和聲則會帶給人緊張的感覺。以上為音樂三大要素，但音樂的元素還有很多的成員，不同的元素都會使聽者產生不同的情緒感受，例如：音量大小、調式類型、音高位置、音色感受....等等，下表 3.1.1 筆者參考了許多文獻研究[ 16, 17, 18, 19, 20 ]並整理出一系列與情緒有關聯性的各種音樂元素以及其對應關係，而本研究收錄了其中五種於第三章節中說明。

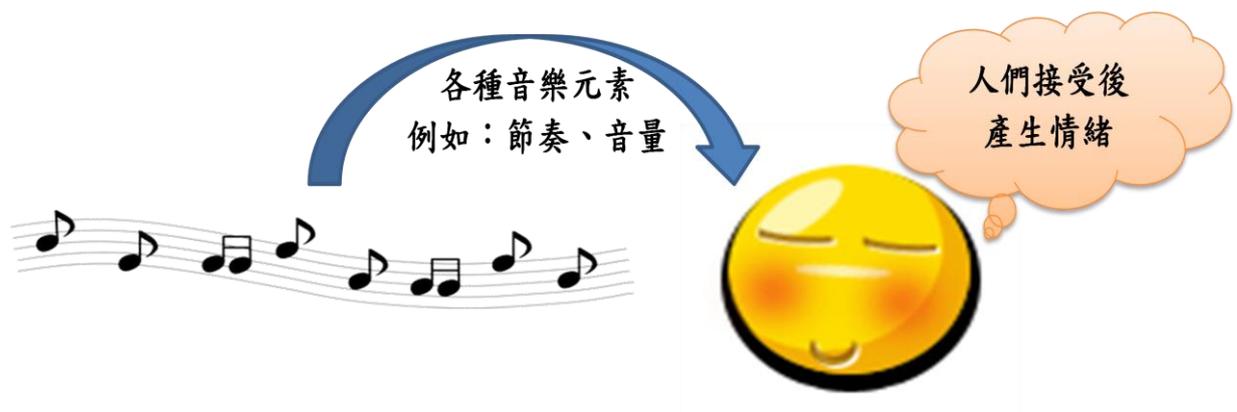


圖 3.1.1 音樂與情緒連結示意圖

表 3.1.1 音樂元素與情緒指向之關連

情緒指向	特徵種類/描述
Active -Positive Ex: happy	Mode Major, Tempo Fast, Harmony Simple, Loudness Loud, Articulation Staccato, Pitch High, Rhythm Flowing
Very Active Positive Ex: excited	Tempo Fast, Loudness Loud, Mode Major, Pitch High, Pitch Variation Large, Harmony Simple, Note Onset Rapid, Pitch Range High, Pitch Contour Up, Articulation Staccato
Very Active Negative	Mode Minor, Loudness Loud, Tempo Fast, Harmony Complex, Note Onset Rapid, Pitch Contour Up, Pitch High, Pitch Range High,
Active-Negative Ex: angry	Mode Minor, Harmony Complex, Articulation Legato, Pitch Variation Small, Tempo Fast
Passive- Negative Ex: sad	Tempo Slow, Articulation Legato, Mode Minor, Harmony Complex, Loudness Soft, Harmony Simple, Pitch Low, Note Onset Slow
Very Passive- Negative Ex: depressed	Loudness Soft, Tempo Slow, Pitch Variation Small, Articulation Legato, Note Onset Slow, Pitch Low
Very Passive Positive	Tempo Slow, Loudness Soft, Articulation Legato, Note Onset Slow, Pitch Low, Pitch Range Low, Pitch Variation Small
Passive-Positive Ex: dreamy	Mode Major, Harmony Simple, Tempo Slow, Articulation Staccato, Loudness Soft, Pitch Contour Up

### 3.2 音樂誘發的情緒與聽者的心理反應

不同的音樂所縈繞出的不同情緒感受，你我通常都能感同身受體會之間的差異，並且也能從受測者的回答中得出答案，但是心理反應是受測者無法控制且無法自知的，有時候音樂可能喚起的是人們的認知、行為、生理情緒這些受測者無法主動提供的主觀要

素，此外並非所有人對相同的音樂皆會有情緒上的感受，關於這點常常是每個人自身的不同喜好所致，因此本研究參考文獻[5] (Marcel Zentner, 2008)，並摘錄了其重點做為研究的背景知識。情緒研究若單單憑靠感覺，很容易會對人們心理的認知有所偏差，例如：憤怒、恐懼、厭惡、內疚…等等負面的情緒，上述情緒的產生使人在適應與心情上威脅到了自身的身體與心理的完整性。因此當人們在聽音樂時，往往會變得健忘，並且脫離凡俗，也就是當人們進入如此精神狀態，現實世界中的自身利益與威脅之間的關連性不再強健的存在，因此負向情緒也失去了其效果範圍，所以上述情緒是不該成為音樂情緒研究中的一個選項。同樣如同含義涉及緊張刺激的用語，例如：驚悚、恐懼、害怕，我們當然相信那是音樂最可能令聽者產生的感受，而有些特別的聲音或聲效的確會令聽者恐懼，但在過去的研究中表示，聽者往往會想起恐怖片裡驚悚的片段的配樂與背景，導致聽者害怕的主因並非音樂本身而已，而失去了可信度。人們對於情緒還有另一特別的反應，那就是迴避悲傷(sadness)，從過去的研究統計結果顯示，悲傷的音樂在聽者感受上是不常見的，但有一解釋是：一組悲傷的文字會更為明顯的導致悲傷情緒產生。此外，在日常生活中，悲傷往往是一個大量經驗的累積，人們一般都希望能避免，但現實中存在的卻是當廣播播放悲傷的音樂時，聽者轉台機率卻是低的，如此相反矛盾的情況。而在試驗中，最常見回報含有悲傷成分的用語是憂鬱(melancolique)，也就是說當受測者可以選擇時，大多會選擇憂鬱並將悲傷予以刪除。最後要強調的是情緒絕對不是有顯著地差異或是正反面相互排斥的類型，大部分時間內情緒是以混和的方式表現，在過去提倡的基本音樂情緒類別它們的確能保持彼此間相互和諧存在的一致性，但也是因此無法獲得更細微的情感(Juslin, 1997)。總和來說情緒感受受制於相當多互動的因素例如：音樂特徵、聽者身分背景、表演者的特質、語意或環境因素...等，理論上也應該存在更多的特性使得情緒難以可靠地分辨，但這也使的人們的生活多采多姿。

### 3.3 音樂基本情緒模型

情緒模型是音樂情緒辨識系統中最基本的假設，經由心理學家整理提出的情緒模型，能夠明確的定義出各種不同抽象情緒種類並且分門別類有相似之關係的情緒種類、以及有效地對於普遍的情緒反應皆指出概要定位與訂定軸度關係。而本研究引用的情緒基本模型為 Robert E. Thayer 提出的情緒模型[ 21, 22 ]，此模型分為四個象限，分別為(黃色)激勵人心的、興奮的，(綠色)舒適愉悅的、平靜的，(紅色)焦躁憤怒的、痛徹心扉的，(藍色)使人意志消沉的、憂鬱的。圖中原點解釋為音樂剛開始音量很小、尚未產生情緒的狀態，圖表橫軸為壓力，縱軸為能量。橫軸的定義為音樂帶給聽者的抽象壓力，縱軸的定義為音樂帶給聽者的抽象能量，許多的音樂特徵如：快速緊湊的節奏較大的音量通常代表音樂的能量較高，慢速緩和的節奏較低的音量則表示音樂能量較小，不和諧的背景和聲或小調的調式則會使音樂聽來較為沉悶，令聽者情緒壓抑無法釋放出來的壓迫感，表示音樂造成的壓力較大，這些特徵強度與模型中的橫軸縱軸皆有直接的關係，然而模型指出相對應的關係仍需要以數理分析的方法來進行量化，而如何基於情緒模型的定義來將各種抽象的情緒感受實值計算得出一組量化的結果，此也是是研究中的一大挑戰。

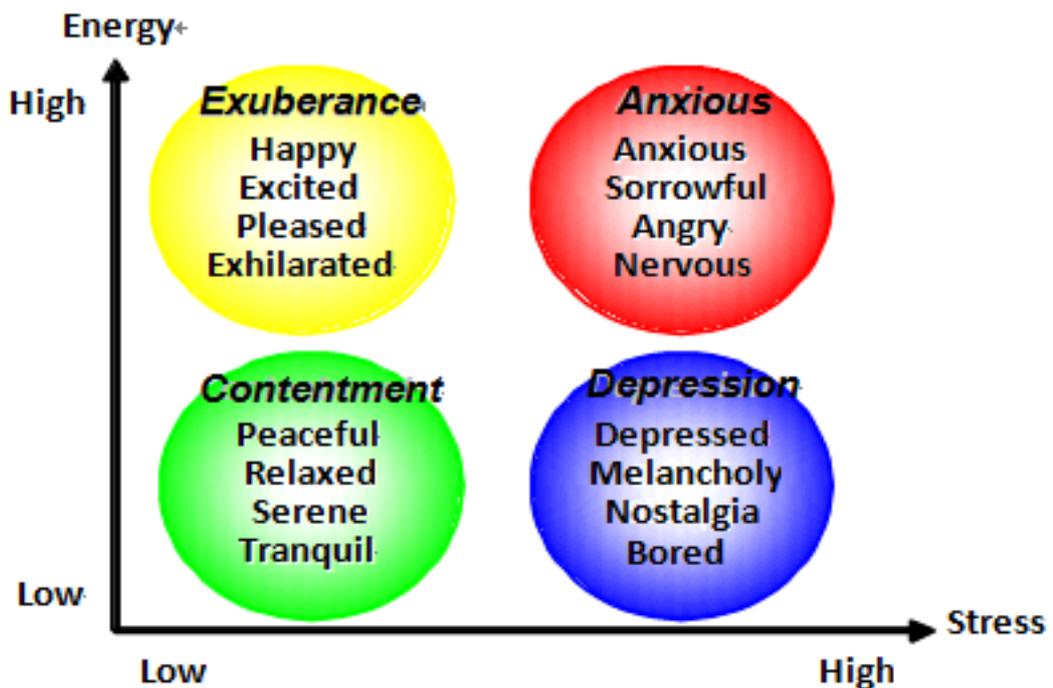


圖 3.2.1 二維基本情緒模型

## 四、系統架構

本章節共包含六個小節，第一小節為流程介紹，接下來五個小節會完整說明音樂的情緒響應分析的步驟，分別為系統流程介紹、訓練資料格式、音訊特徵萃取、計算情緒得分方法、資料訓練收集與分類辨識與最後訓練資料訓練結果，系統使用的設計程式為 MATLAB，研究採用的音訊檔為 WAVE 檔。

### 4.1 系統流程介紹

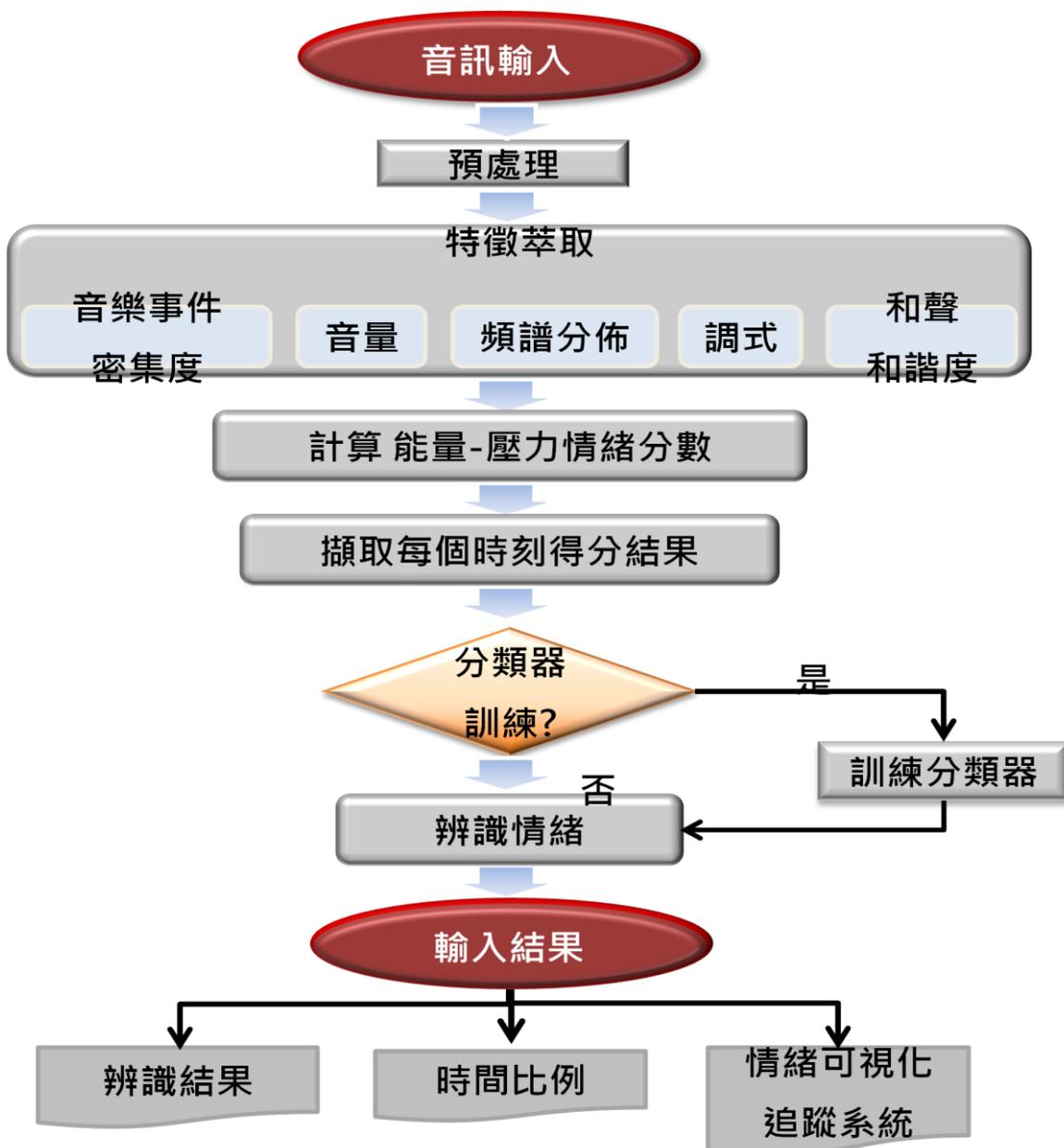


圖 4.1.1 系統流程方塊圖

### (1) 音訊輸入

在系統中，音訊的規格採用 WAVE 格式之音樂檔案，取樣頻率 14700Hz，取樣解析度 16-bits，單聲道，作為基本的訓練與測試之用，訓練音檔採用 30 秒的音樂片段並且附有單一情緒類別的標記，詳細說明請參考 4.2 節。使用者模式中音訊檔案格式與上述相同，而長度則視個別檔案而定。

### (2) 預處理

延續上文所述，要對音訊做進一步的分析，首先將各種音檔皆統一降取樣至 14700hz 以減輕接下來程式運行的負擔，為了表示音訊於每個時刻的狀態，必須先對音訊作音框化的處理，使原始訊號切割為許多等長的音框。音訊音框化的音框長度為 2048 個樣本點，音框重疊長度為 1536 個樣本點。

### (3) 特徵萃取

從經過音框化的每個音框中計算出特徵值大小，主要有五種特徵被萃取，(i)音樂事件密集度、(ii)音量、(iii)頻譜分佈、(iv)調式、(v)和聲和諧程度，詳細方法於 4.3.1 ~ 4.3.5 五個小節中簡單介紹。

### (4) 計算 能量-壓力情緒分數

計分的整體概念是模擬一般人聽音樂的過程，模擬的目標有兩個，分別為當下的情緒得分以及前一段時間至當下所接受到的音樂內容的累積遞延結果，而完整的計分公式與各種特徵對應情緒平面的權重與方法，將於 4.4 節中詳細說明。

### (5) 擷取每個時刻得分結果

經過上章節計算後將獲得呈現移動軌跡一般的連續數值，而我們將取樣後三分之二的結果來做為訓練分類器的訓練資料，從上得到的訓練資料已可大略看出同類的情緒有著類似的軌跡分佈，於 4.5 節中詳細說明。

## (6) 訓練分類器

利用分類器透過訓練的處理將計分結果所得到二維數值加以分類，如此一來可以分辨出不同情緒的種類，詳細內容於4.6節討論。文章於第五章中有比較不同的分類器，並比較產生的結果是否對於系統會產生過大的差異性與並影響性能。

## (7) 輸出結果

輸出的方式有三種，分別為(1)辨識結果、(2)時間比例與(3)情緒可視化追蹤模式，文章採用了七位大家耳熟能詳的著名歌手或樂團，總共收集了一百首大家嚶嚶上口的歌曲，做為測試用的歌曲，許多曲目都是當紅一時的流行歌，希望可以使讀者於閱讀時能更加親切與熟悉，詳細內容於第五章討論。

## 4.2 訓練資料格式

本次的訓練資料總共有 200 個音樂片段，檔案類型為 WAVE 格式，每段長度 30 秒、單聲道、取樣頻率 14700Hz、取樣解析度 16-bits，類型包含不同類型音樂如：流行音樂、搖滾音樂、古典音樂...等。全部皆為多聲部(Polyphonic)的音訊內容，即全為原始 CD 的音訊內容而非由 MIDI 與音源產生的音訊。每個音樂片段皆選定為近似於單一穩定情緒的音樂片段，引用 Thayer 情緒模型中的四種類別：(1)舒適的、(2)哀傷的、(3)焦慮的、(4)振奮的，並由數個聽音測試人員將其標記給予一個與感受相似的音樂情緒類別，各個情緒類別的音樂片段數量可以參考表 4.2.1。

表 4.2.1 訓練資料情緒分類數量統計

情緒:	舒適/平淡	沮喪/沉寂	生氣/焦躁	活潑/動感
數量:	50	50	50	50

### 4.3 音訊特徵萃取

在音訊特徵萃取的部分，本研究主要使用實驗室所留下的資源，共有五種特徵被萃取，(i)音樂事件密集度、(ii)音量、(iii)頻譜分佈、(iv)調式、(v)和聲和諧程度。選擇這些特徵的原因為：根據音樂心理學的研究與日常經驗顯示，它們都相當直接地影響聽者的情緒感受，如節奏速度對應到情緒氛圍的平緩或是激烈，而節奏速度反映在音樂事件的密集程度上。音量大小則是對於當下的情緒感受有增強的作用，音色則可以由音樂在頻譜上的分佈來分析。調式通常有一個趨勢，大調音樂使人感受愉悅，小調音樂則較哀傷與詼諧，不和諧的背景和聲則會使人抑鬱，以下將依序說明。

#### 4.3.1 音樂事件密集度

一般大眾在聽音樂時，規律的音樂事件，如樂器聲，歌聲，為使人感知到節奏的主要訊息，若能找到訊號中的音樂事件端點，即可由端點時間位置的規律做進一步的節奏分析。節奏計算的方式大概為：

1. 音樂事件的端點偵測(Onset detection)。
2. 計算音樂事件的密集度。

將原始訊號音框化後，定義其音框中心樣本點的時間位置即為該音框的時間位置，而該音框的訊號就代表原始訊號於此刻的狀態。音框的頻譜強度數值可由短時距傅立葉轉換(STFT)來計算，代表訊號於該時間各頻率成分之強度，而頻譜流量代表音訊於某一特定時間所有頻率頻譜強度正流量，其算式如(16)式：

$$\text{Spectrum Flux } (m) = \sum_{k=0}^N H(|S_m[k]| - |S_{m-1}[k]|)$$
$$\text{where } H(x) = \frac{x + |x|}{2} \quad (16)$$

上式中第一式為音框  $m$  所對應的時刻下所有頻率的頻譜正流量總和， $N$  為單一音框總樣本點數， $k$  為頻域樣本點數， $S_m[k]$  和  $S_{m-1}[k]$  分別為對應第  $m$  個音框和第  $m-1$  個音框的

頻譜強度，而  $|S_m[k]| - |S_{m-1}[k]|$  為此刻對應到頻率為  $f[k]$  的頻譜強度的流量， $H(x)$  的作用則是篩選出正流量，即只有頻譜強度增加才會被計算在總和中，所有頻率的頻譜強度正流量總和則為當下總頻譜流量，音樂事件強度越強，量值越大，聽起來也越鮮明。實際利用頻譜流量偵測音樂事件偵測的結果如下圖所示：

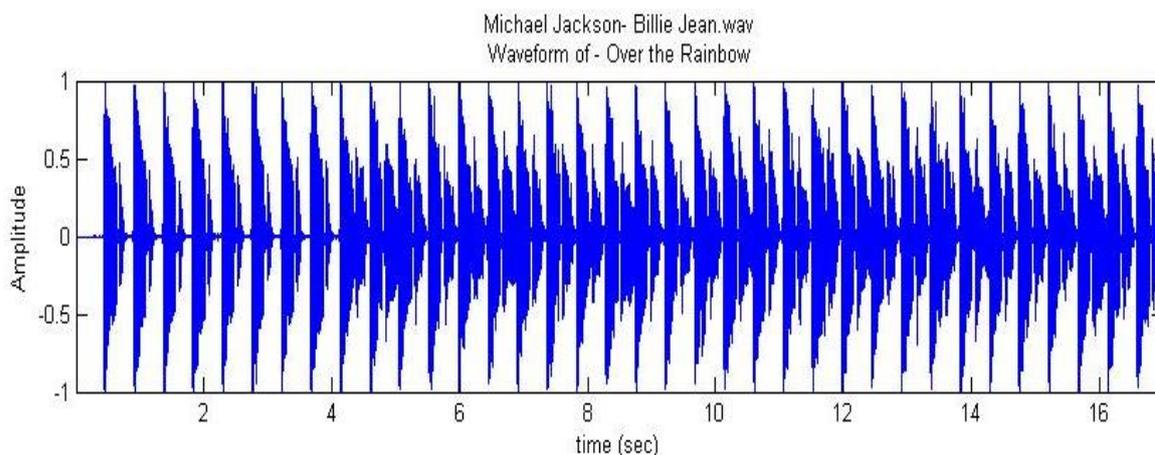


圖 4.3.1 麥克傑克森之 Billie Jean 原始音樂波形

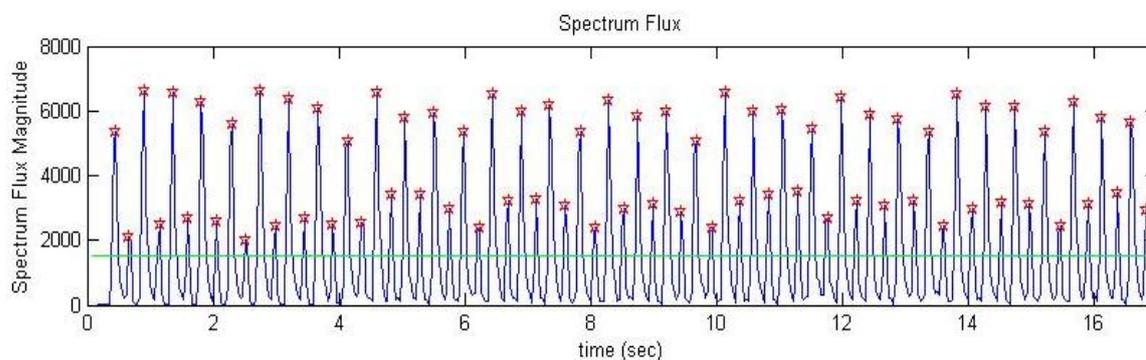


圖 4.3.2 音訊頻譜流量進行音樂事件偵測

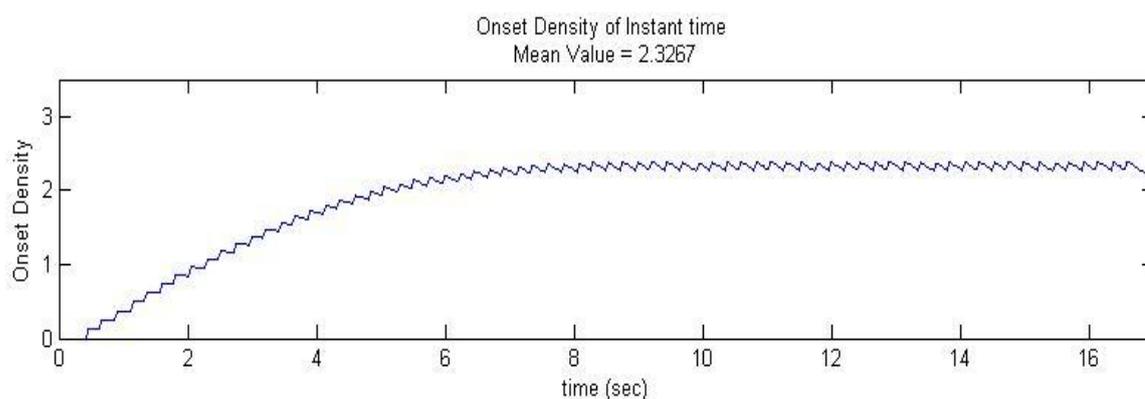


圖 4.3.3 音樂事件密集程度計算結果

上圖 4.3.1 為 0~20 秒的時域波形，圖 4.3.2 則是以公式(16)計算的頻譜正流量，綠色線段為平均流量。可以看到當頻譜流量有峰值出現的時候即對應到一個音樂事件的發生位置。由於一般普遍的音樂型態皆有固定的節奏，在這樣的假設前提下，本系統直接以音樂事件密集程度來代表瞬間的節奏速度，計算結果如圖 4.3.3 所示。

### 4.3.2 音量大小

聲音的大聲或小聲，在人的聽覺感知中稱為該聲音的響度(Loudness)。音量在音樂的表現中也常常和情緒有直接或間接的關聯，其大小或是改變對聽者的情緒具有相當的影響力。如古典音樂的情緒轉折通常伴隨著音量的明顯變化，甚至流行音樂也常用相似的手法安排音樂段落。一般響度可以直接由聲音訊號的音量(Volume)來估測計算。單一音框的音量最簡單的計算方式是計算音框內振幅的絕對值總和如下所述：

$$\text{Volume}_1[m] = \sum_{n=1}^N |x_m[n]| \quad (17)$$

本系統之音量計算方式採用另一種計算方式為計算振幅之平方直總和再取對數單位-分貝(decibel)，實際計算方法如下：

$$\text{Volume}_2[m] = 10 \times \log_{10} \left( \sum_{n=1}^N x_m[n]^2 \right) \quad (18)$$

### 4.3.3 頻譜分佈(音色)

聲音訊號的頻譜分佈內容和音色之間有直接關係，本研究之頻譜分析主要分為兩大項，分別為頻譜的形狀與頻譜的對比。頻譜形狀主要在於分析音訊頻域分布情形，頻譜對比則是將整個頻域切割成為許多子頻帶(Sub-band)，並分析各個子頻帶的相互對比關係。因為一般而言聲音除了基頻，基頻的整數倍或非整數倍分別稱為泛音與 overtone 往

往是造成其獨特音色的原因，其中一種常用的頻帶分析方式是由八度音做為基準，訊號在不同頻率區塊都有不一樣的音響效果，高頻的訊號使聲音聽來明亮，低頻訊號則使聲音聽來充滿震撼力與能量，典型代表如：鼓聲這類節奏強烈的樂器通常位於低頻區塊，電影的爆破效果通常也需要大量的低頻成分。本文以頻譜的質心(Centroid)配合頻帶寬(Bandwidth)的能量積來做分析。頻譜質心計算方式如下式：

$$\text{Spectrum Centroid}[n] = \frac{\sum_{k=1}^{N/2+1} (f[k] \cdot |S_n[k]|)}{\sum_{k=1}^{N/2+1} |S_n[k]|} \quad (19)$$

上式中如同一般質量中心的算法， $n$  為音框數索引， $k$  為頻域樣本點索引，頻譜強度  $S_n[k]$  對應到物體質量，頻域位置  $f[k]$  對應到物體位置。頻譜質心代表頻譜分布的整體質心位置，質心位置越高代表整體的音樂內容高頻偏多或是音域偏高。頻帶寬的計算方式如下式：

$$\text{Bandwidth}[n] = \frac{\sum_{k=1}^{N/2+1} (f[k] - \text{SC}[n]) \cdot |S_n[k]|}{N/2+1} \quad (20)$$

其為每個頻率位置與頻譜質心 SC 的差額配上當前的頻譜強度權重之加權平均，頻帶寬越大代表頻譜能量較為分散，頻帶寬小則代表集中。在樂團形式的音樂中因為配器種類較多所以頻域範圍大，其頻帶寬通常較獨奏曲目大，且質心位置也通常較高，至於兩者對於能量軸度的綜合貢獻，是以頻譜質心乘以頻帶寬來表示之如第(20)式所示。

$$\text{SpectrumEnergy} = \frac{\sum_{k=1}^{N/2+1} (f[k] \cdot |S_n[k]|)}{\sum_{k=1}^{N/2+1} |S_n[k]|} \cdot \frac{\sum_{k=1}^{N/2+1} (f[k] - \text{SC}[n]) \cdot |S_n[k]|}{N/2+1} \quad (21)$$

#### 4.3.4 大小調調式分析

音樂調式是一種相當有識別性的音樂特徵，如小調音樂總是聽起來較為哀傷與詼諧；大調音樂則是聽起來較為快樂與振奮。進行調式追蹤可以一探某特定時間內的音樂對於聽者的情緒感受影響為何。本研究採用 Pitch Class Profile (PCP) 方式，經由短時距傅立葉轉換得到頻譜數值後，可以進一步利用頻譜來計算一般的音樂理論分析上較常用的特徵值音調(Pitch)，而選用 100Hz 至 5000Hz 之間的特定頻率範圍的用意為減少打擊樂器和其他非和聲音訊的干擾。音調一般以大寫音文字母 A 到 G 表示。由頻率和半音(Semitone)之間的關係式可將頻率換算為音調，再利用音調於倍頻或稱八度(Octave)為相同音調層級的概念，即可將頻譜換算為對應的音調層級(Pitch Class)，如下：

$$P(k) = \left[ 24 \cdot \log_2 \left( \frac{f_s \cdot k}{N \cdot f_1} \right) \right] \text{ mod } 24 \quad (22)$$

$$\text{PCP}[P(k), n] = \sum_{P(k)} |S[k, n]| \quad (23)$$

上式將頻譜數值映射到 24 個音調層級上，其中公式(22)中  $k$  為頻域的樣本點數索引， $P(k)$  表示頻域和音調層級空間的對應關係，代表頻域第  $k$  個樣本點之頻率值對應的音調層級， $24 \cdot \log_2 \left( \frac{f_s \cdot k}{N \cdot f_1} \right)$  將第  $k$  點的頻率值換算為對應的半音數，再由餘數(mod)方式將倍頻的音調歸為同個音調層級。第二式將頻譜數值轉換到音調層級空間(PCP domain)的表示法，其中  $n$  為音框數的索引， $S[(k=0, 1, \dots, N), n]$  為第  $n$  個音框的頻譜數值， $P(k)$  為音調層級空間的樣本點數索引， $\text{PCP}[(P(k)=0, \dots, 23), n]$  則為第  $n$  個音框的音調層級數值，其為頻譜中所有倍頻的相同音調層級的強度加總。但因為考量以 12 平均律切割的 12 個音調層級在數值分析應用上不夠準確，故在分析之前將每個層級中再對半切割，成為 2 組共 24 個音調層級，在經由與各組內積值決定用哪一組音調層級，得到如下兩式之結果。

$$\text{simplified PCP} = \text{PCP}(1:2:23) \quad (24)$$

$$\text{or simplified PCP}=\text{PCP}(2:2:24) \quad (25)$$

最後將以此簡化過的 PCP 向量與調性樣板做向量內積，即可以得到某個時刻屬於該調式的機率大小為何，與該調式之調式樣板內積值越大代表此刻出現的音符較接近該調式。計算方法為將第  $n$  個音框以前 8 秒鐘的所有音框和第  $k$  個調式的樣板做內積並計算總和每個音框所對應的調式的機率，最後選出擁有最大內積值的調式作為正確答案，進而得到大小調的結果，並且賦予大小調不同的評價數值。

#### 4.3.5 和聲和諧度

和聲即為不同音程之間的比值，使人感受到不同的和諧程度。和諧的和聲使人的感受為正面的情緒，反之不和諧的和聲讓人感覺到負面的情緒，研究中以每個音框之 PCP 向量最大值(即最顯著的音)與其增四/減五度音程比例為一個不和諧程度的指標。

$$\text{Note1}[n] = \max[\text{PCP}(1:24,n)] \quad (26)$$

$$\text{Note2}[n] = \text{PCP}(\text{index of Note1}[n] \pm 12, n) \quad (27)$$

$$\text{Dissonance}(n) = \frac{\text{Note2}[n]}{\text{Note1}[n]} \cdot \frac{\text{Note1}[n] + \text{Note2}[n]}{\max[\text{PCP}(\text{all}, \text{all})]} \quad (28)$$

上式(28)中 Note1[ $n$ ]為第  $n$  個音框中能量最強的音，Note2[ $n$ ]則為 Note1[ $n$ ]所對應的增四減五度，可以由 Note1[ $n$ ]所對應的 PCP 維度索引再加/減 12 可以得到，Dissonance[ $n$ ]則為第  $n$  個音框的不和諧程度，其計算方式為 Note1[ $n$ ]和 Note2[ $n$ ]的比例乘以 Note1[ $n$ ]和 Note2[ $n$ ]能量和整體最大值  $\max[\text{PCP}(\text{all}, \text{all})]$ 的比例。

#### 4.4 計算情緒得分方法

在系統的計分概念中包含了(1)當下情緒感受得分以及(2)隨時間的情緒遞延兩個概念：當下情緒感受得分為聽者從聽到音樂後接受到的情緒效果，也是聽到聲音瞬間最直接的感受；時間的情緒遞延為自前一小段時刻至當下瞬間這段時間的綜合聆聽感受，隨著時間的流逝，人們的情緒感受也會隨之累積並且持續醞釀。本研究設計了一套簡單的方法來模擬上述聆聽音樂的過程感受，當下情緒得分的公式定義如下：

$$P_t(x, y) = \sum_{f=1}^5 \left[ w_x(f) S_f(t) \bar{x} + bias_x + w_y(f) S_f(t) \bar{y} + bias_y \right] \quad (29)$$

$P_t$  定義為聽者當下的情緒得分，是一個二維的函數分別為  $x$  維度和  $y$  維度，由前述 5 種特徵強度分別配與特定權重再加上一個偏移值來決定，公式中  $f$  為特徵值的索引，分別代表五種不同的特徵， $S_f(t)$  為  $t$  時刻第  $f$  種特徵的強度， $w_x(f)$ 、 $w_y(f)$  則分別對應到 Thayer 情緒模型的兩軸中的能量軸和壓力軸，音樂特徵值之間互相的變化比例與對應的情形，本研究中經過多次實驗調整給予的配給權重如表 4.4.1 所述：

表 4.4.1 不同音樂特徵之間的相對應比例

特徵：	音樂事件密集度	音量	頻譜分佈	調式	和聲和諧度
能量：	<b>2.39</b>	<b>1.09</b>	<b>0.3</b>	-	-
壓力：	<b>-0.53</b>	-	-	<b>3.22</b>	<b>3.33</b>

偏移值  $bias_x$ 、 $bias_y$  設置的目的是為了補足五種特徵經過計算再乘以權重之後卻仍達不到理想的加總效果，特別是現今流行音樂作曲創意無限、曲調自由，跳脫舊有的譜曲習慣規則，因此為補足在壓力軸度上頻域分析的不足之處以及壓抑能量軸度的過度伸張，所以在公式(29)中添加一個由分析歌曲內容得到的參考值來加以控制啟動的偏移值，

因此重點即在於如何控制電腦自動判斷出哪些歌曲符合上述的缺陷，而本研究解決的方法為對每首歌曲計算出它所有時間的訊號靜音率  $ASR$ (Audio Silence Ratio)之後再取平均作為參考值，系統將根據該  $ASR$  的平均值自動判斷是否啟動偏壓，因此筆者根據經驗設置了一個參考值的臨界數量作為系統判斷的依據，當歌曲經計算得出的參考值低於預設的臨界數量時，系統即在計分公式中啟動偏移的機制，該參考值  $ASR$  公式計算參考自文獻[3]，其公式定義如下：

$$ASR = \frac{1}{2N} \sum_{n=0}^{N-1} \left( 1 - \text{sgn} \left( STE(n) - \rho \times \text{avg}STE \right) \right) \quad (30)$$

$$STE(n) = \sum_{k=0}^{m-1} a_n^2(k) \quad (31)$$

$$\text{avg}STE = \frac{1}{N} \sum_{n=0}^{N-1} STE(n) \quad (32)$$

第一式中大寫  $N$  是單位時間內之音框總數，小寫  $n$  為音框索引值， $\text{sgn}$  函數是邏輯函數，輸入大於 0 輸出 1 小於 0 輸出 -1，邏輯函數中的  $STE(n)$  為第  $n$  個音框的短時距能量 (shot-time energy) 之值如公式(31)， $a_n(k)$  代表第  $n$  個音框內第  $k$  點的音訊振幅之大小數值， $\text{avg}STE$  為  $STE(n)$  之平均值如公式(32)，之後再乘以比例參數  $\rho$  如圖(4.4.2)；而研究中實際運作情況為設定每 0.25 秒做為一個時間單位，每時間單位內分配有  $N$  個音框，每個音框點數  $m$  為 512 點，最後經過計算得到整首歌曲的所有  $ASR$  訊號，本研究採取對所有的  $ASR$  取平均數來做為系統判斷是否啟動偏移的準則，若低於設定的平均  $ARS$  臨界值則偏移量啟動加入計分，若不低於設定則不啟動，不過研究過程中發現到，當該歌曲的音量過大、能量過強的時候，會使這個方法失效，甚至在計分時會使結果產生完全錯誤的結果，因此系統對於音量過大之音樂，必須先行過濾出來，不要啟動偏移量，以減少錯誤的結果發生而降低辨識成功率。

經由上述過程計算出的情緒得分還要考量隨時間流逝所產生的遞延效果，整個過程如圖 4.4.1 所示，其中的  $t$  代表不同的時間點，而每個時間點的完整情緒得分  $P_t$  為時間  $t$  新增的當下情緒得分  $P_t$  在加上過去時間的情緒得分  $P_{t-n}$  乘以衰退函數  $\sigma$  之合，例如下圖中第三秒的得分為  $p_3$  再加上  $p_2$ 、 $p_1$  乘上各自的衰退函數之總合，其中衰退函數代表隨時間流逝，前段情緒的感受逐步的被當下的感受取代，最後經計算得到的二維數值將呈現如移動軌跡般的連續效果。

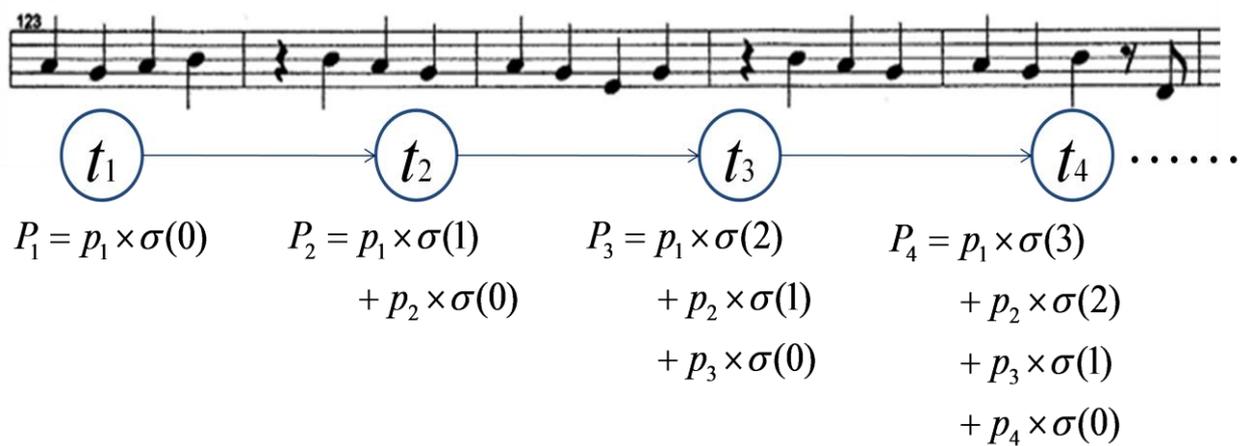


圖 4.4.1 每個時間點的計分流程

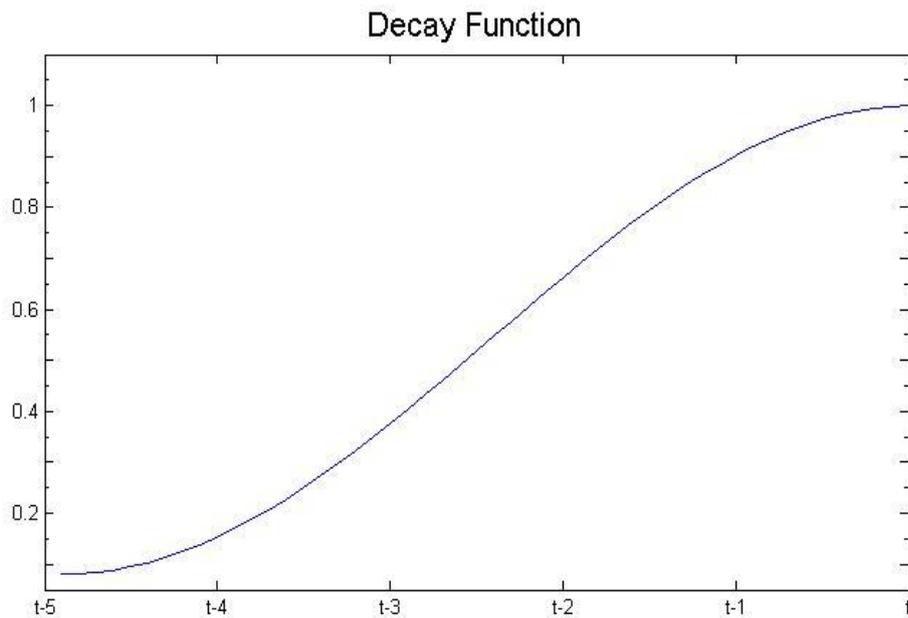


圖 4.4.2 衰退函數

#### 4.5 資料訓練收集與分類辨識

經由上章節得到每個時間的情緒得分，不同的特徵強度會使聽者情緒感受產生相對的加減分和位移，研究採用的訓練音檔時間長度為三十秒，本系統的計分模型所得到的情緒得分對應到 Thayer 情緒模型的座標位移，其計算結果如圖 4.5.1 所示：

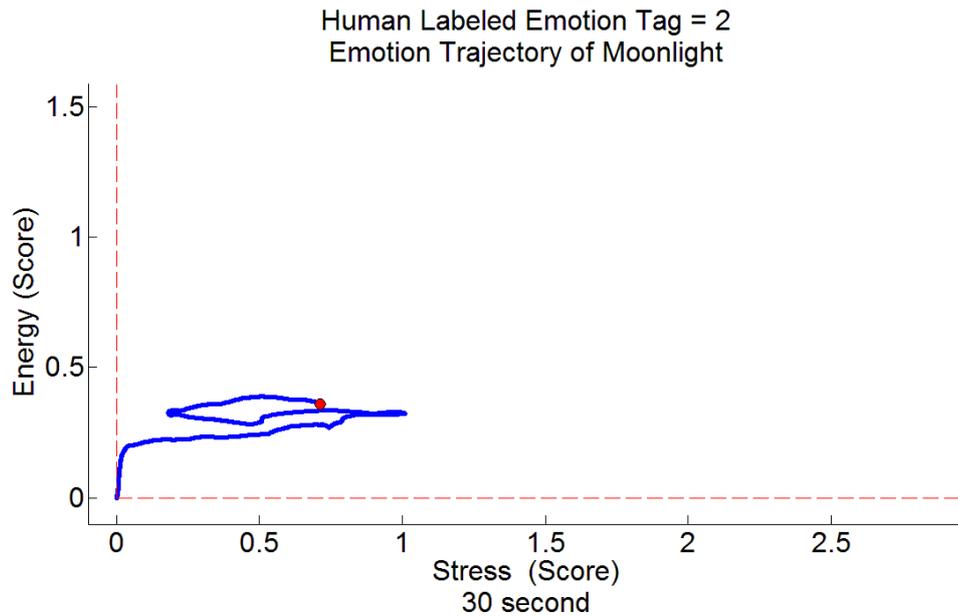


圖 4.5.1 三十秒音檔所產生的情緒軌跡位移結果

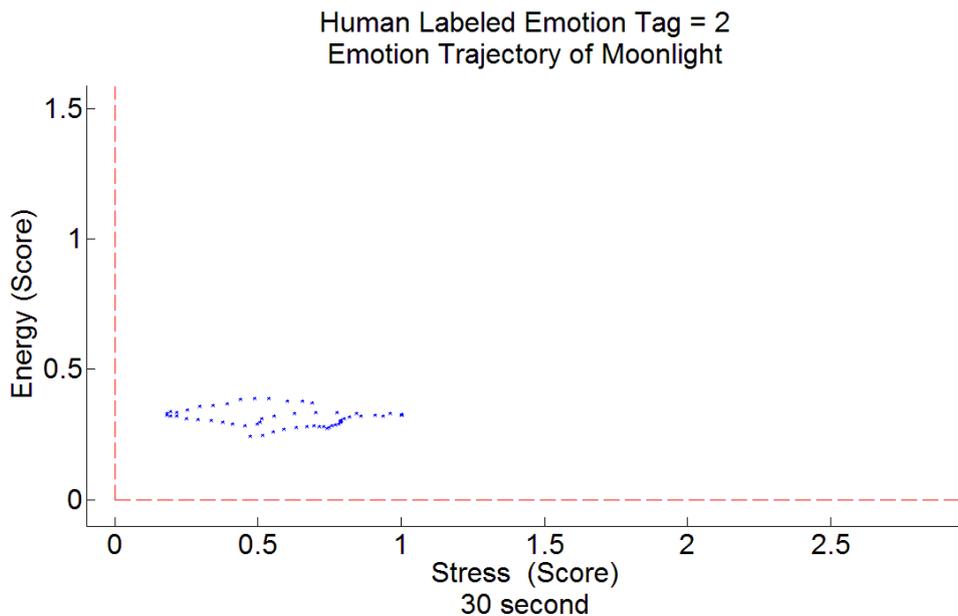


圖 4.5.2 擷取部分位移軌跡所提供的資料

由於情緒軌跡位移在剛開始計分的時候會有一段得分累積的時間，雖然每個音檔內容皆已事先挑選情緒感受一致的片段，但研究在實際運作中只採用音檔後二十秒產生的結果，而每間隔十個音框取一個軌跡座標做為訓練資料，而得到如上圖 4.5.2 之結果。

當全部的200首經作者標記過的音樂片段都經過如上的計分與採樣過程，最終將得到所有的音樂片段所遺留下個軌跡座標點，如下圖4.5.3所示，不同顏色代表著不同的情緒，綠色代表舒適的音樂情緒軌跡，藍色代表哀傷的音樂情緒軌跡，紅色代表焦慮的音樂情緒軌跡，黃色代表振奮的音樂情緒軌跡，當使用的音樂不同或標記的情緒不同，最後遺留下的軌跡分佈也會有些許不一樣，不同類別的訓練資料雖然有部分重疊，但大致上已可由目測看出各種情緒的樣本之間的群聚分佈情況，且與研究採用的情緒模型的假設亦有若干相似。

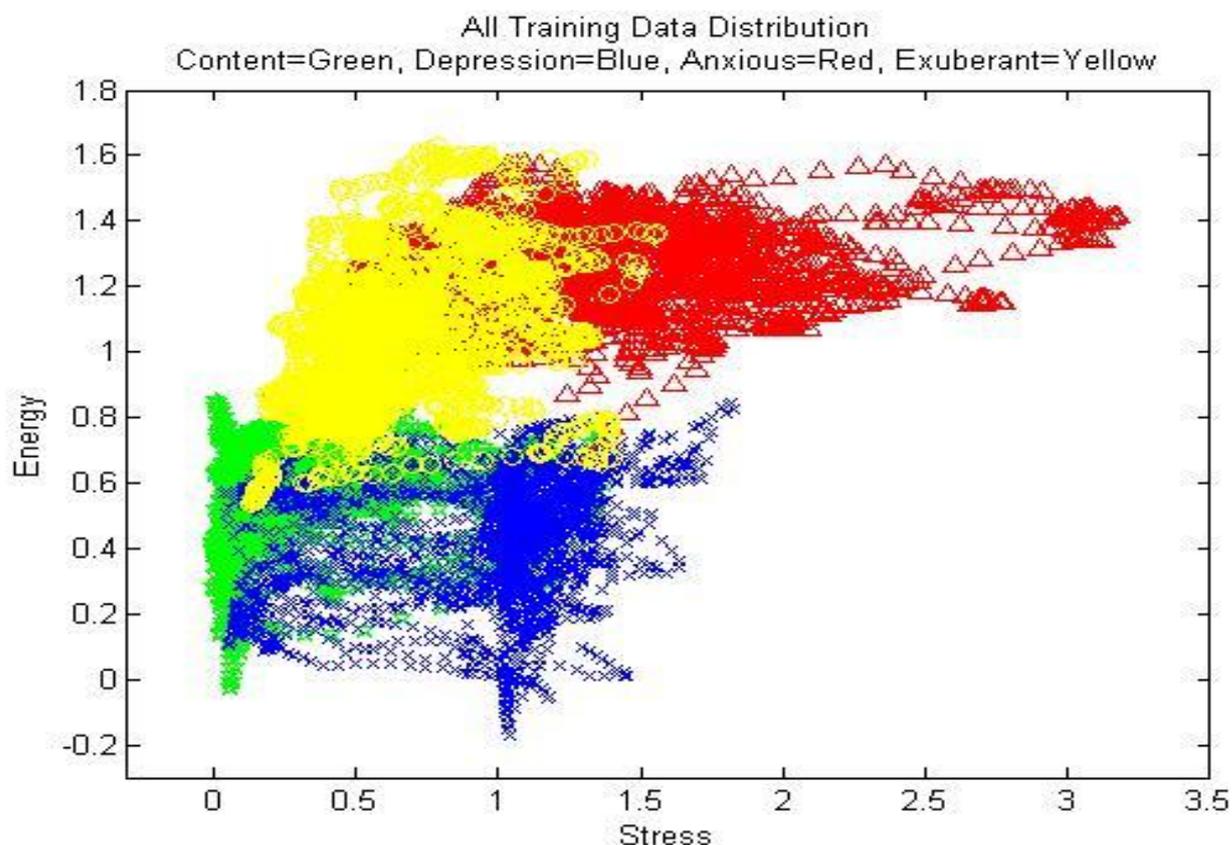


圖 4.5.3 訓練資料之情緒樣本分佈

## 4.6 訓練資料訓練結果

由圖 4.5.3 所示，確實能大概的看出不同情緒的資料其分佈的範圍有所不同，接下來我們利用分類器讓電腦學習如何自動的分辨音樂內涵的情緒，不同的分類器其果不盡然會相同，本研究分別使用兩種分類器 GMM(Gaussian mixture model)以及 BPN(Back Propagation Neural Networks)，來進行分類訓練，並於下章節文章比較兩種分類器情緒辨識的結果。GMM 可以依據訓練資料的分部，找出各個情緒邊界的範圍，不同情緒的資料有各自的分佈狀況，將每個情緒分類的原始 PDF(Probability Density Function)視為由數個高斯函數疊加的分布，找出其可能的 PDF 外貌為何，再從四組情緒的機率密度中找出等高線的邊界。經過 GMM 訓練過程後可以得到如圖 4.6.1 的結果：

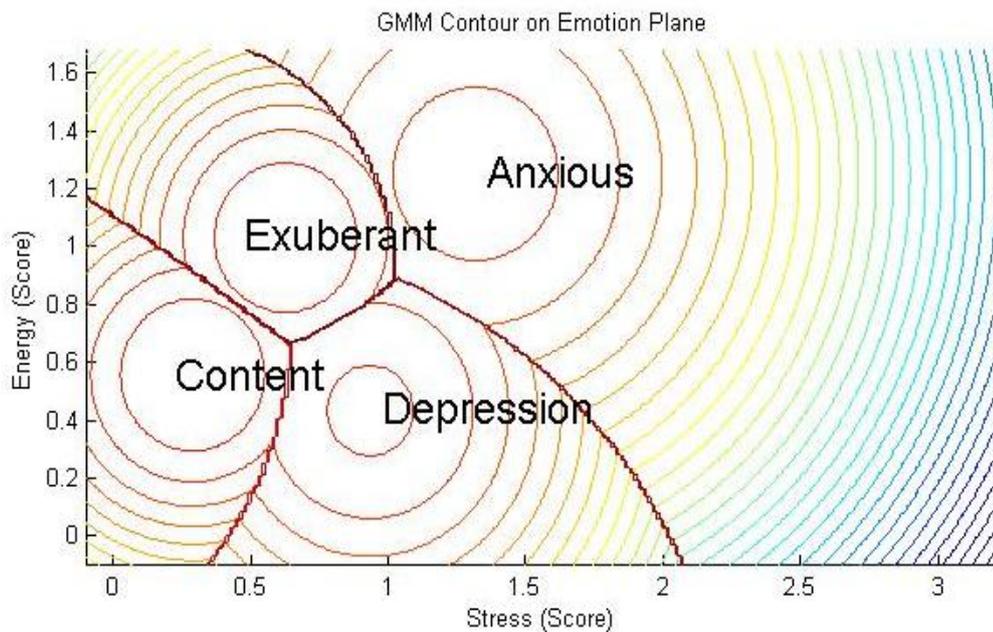


圖4.7.1 GMM分類結果與各類別的邊界範圍

BPN內的神經元間互相連結，擁有許多與人類大腦相同功能的特性，它藉著訓練過程來調整類神經網路中各神經元間的連結強弱(也就是調整加權值)，亦即模擬各層神經元將所要隱含的知識放入神經元間的「連結權重」。完成訓練後的網路，透過調整完成

的加權值，可藉著隱含其中的知識去判斷將來輸入資料的屬性，從而決定輸出的結果為何，最後兩種分類器對於不同情緒訓練的分類結果的內部測試如下圖所示：

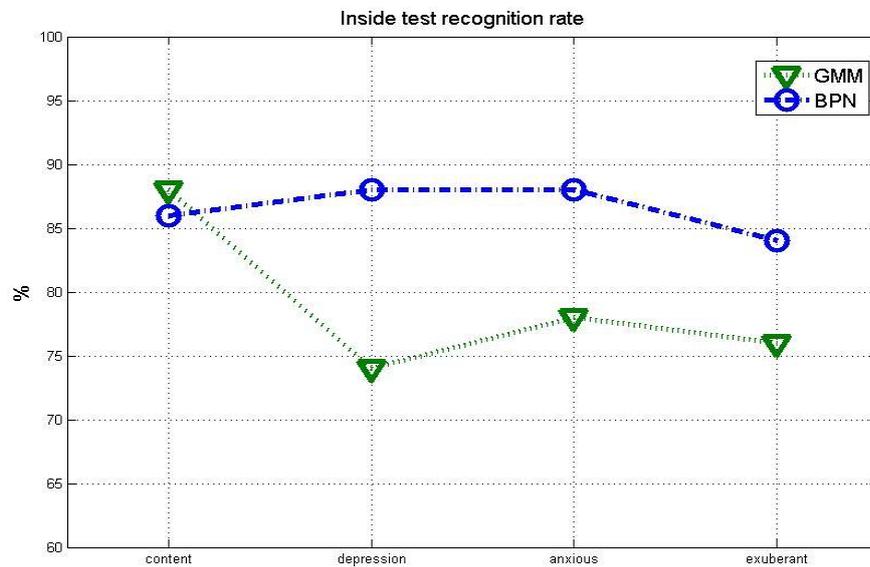


圖4.6.2 兩種分類器訓練結果

藉由大量已標記不同情緒類別的訓練資料(Training Data)所建立的情緒樣本分佈結果來訓練分類器使分類器參數達到最佳化，以對於新的未知音樂資料可以用訓練好的分類器所建立的模型加以進行情緒辨識，而系統情緒辨識結果的輸出方式有三種，分別為(1)辨識結果、(2)時間比例與(3)情緒可視化追蹤模式，本文之第六章節將實際利用市面上流行的歌曲，來討論我們的研究方法是否能有效得區分區別出每個測試音樂的它的情緒特性與情緒間的時間比例分配，並且作為本研究的成果驗證。



情緒可是化系統的螢幕介面(如下圖 5.1 所示)中共可分成四種情緒的區塊，由左而右分別為 (Exuberant 黃色)激勵人心的、興奮的，(Content 綠色)舒適愉悅的、平靜的，(Anxious 紅色)焦躁憤怒的、痛徹心扉的，(Depression 藍色)使人意志消沉的、憂鬱的，圓圈越密集代表該區塊之情緒樣本數量機率密度越大，此外由於情緒本身並無邊界可言，因此系統利用分類器分類的結果，使情緒可以有數量化的邊界定，而當系統運行中時，介面上的線段端點(如下圖綠色線段頂端之圓點)則會隨著音樂的播放而持續移動並留下移動後的軌跡並且變換成代表該區塊情緒的顏色(如下圖狀況中 Content 的代表色為綠色)，線段端點會根據程式計算出來的結果，即時的指出歌曲內容於當下之時間所營造出的情感狀態，下圖為歌曲剛開始的時候，而本研究假設在還沒聽到音樂時，聽者的心情理應還是平靜無特別感受的狀況，系統介面之起點從畫面最左下角的座標(0,0)開始，線段的端點會即時的反應經程式計算後的結果，並移動至適當的情緒區塊，也由於前述之理由，系統需要一小段時間給他累積記分，因此設定於五秒後才開始記錄各種情緒的累積時間。

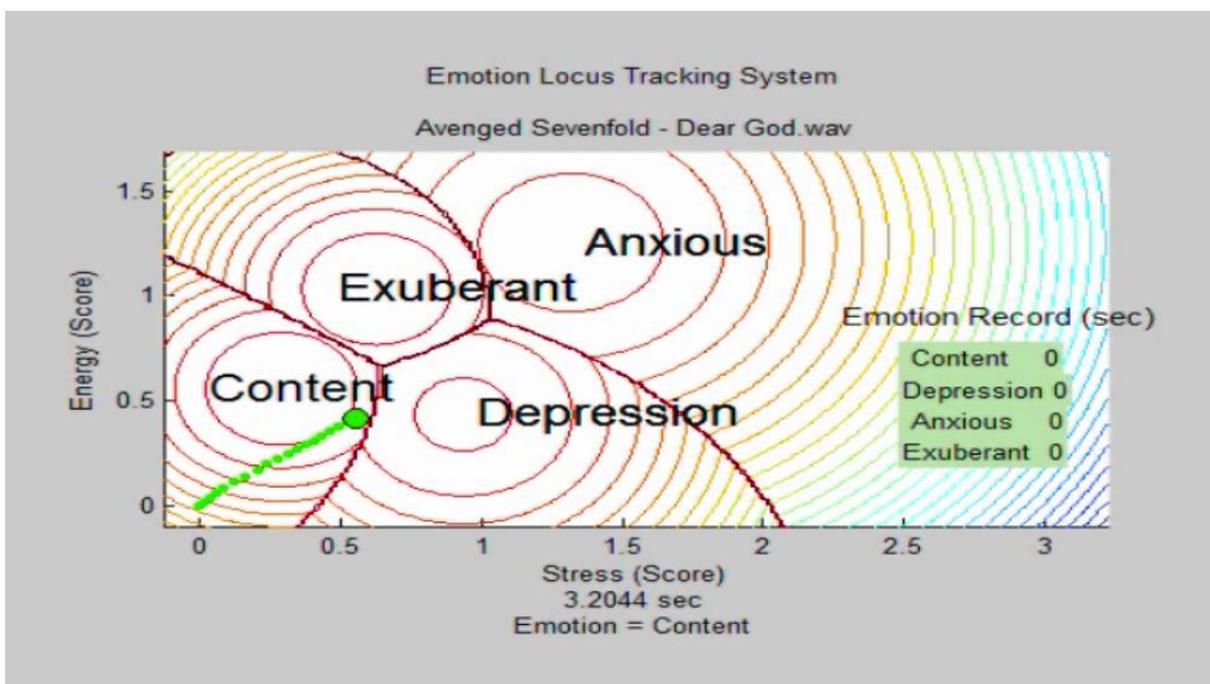


圖 5.2 曲目剛開始之情緒軌跡變化截圖

接著系統將保留線段端點(藍點)所經過的情緒軌跡如下圖 5.3 所示，最後當歌曲即將結束時，系統此時會顯示出歌曲各種的情緒累積時間比例於畫面之中，如圖 5.4 所示：

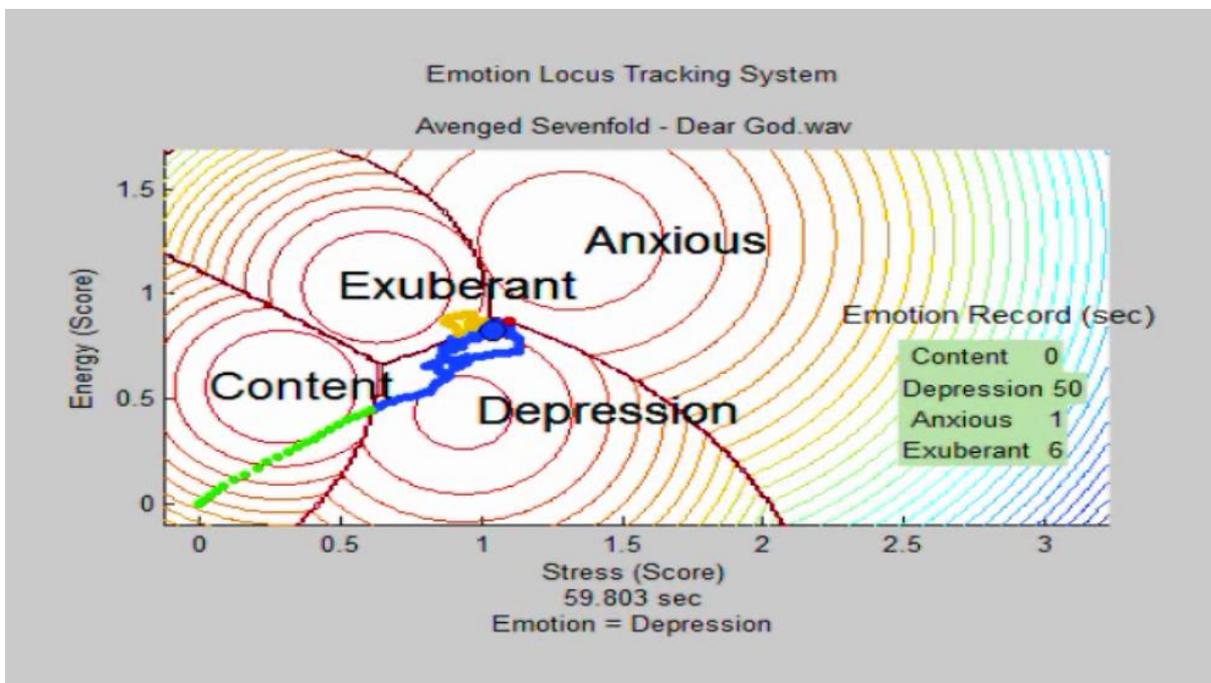


圖 5.3 情緒軌跡變化之過程截圖

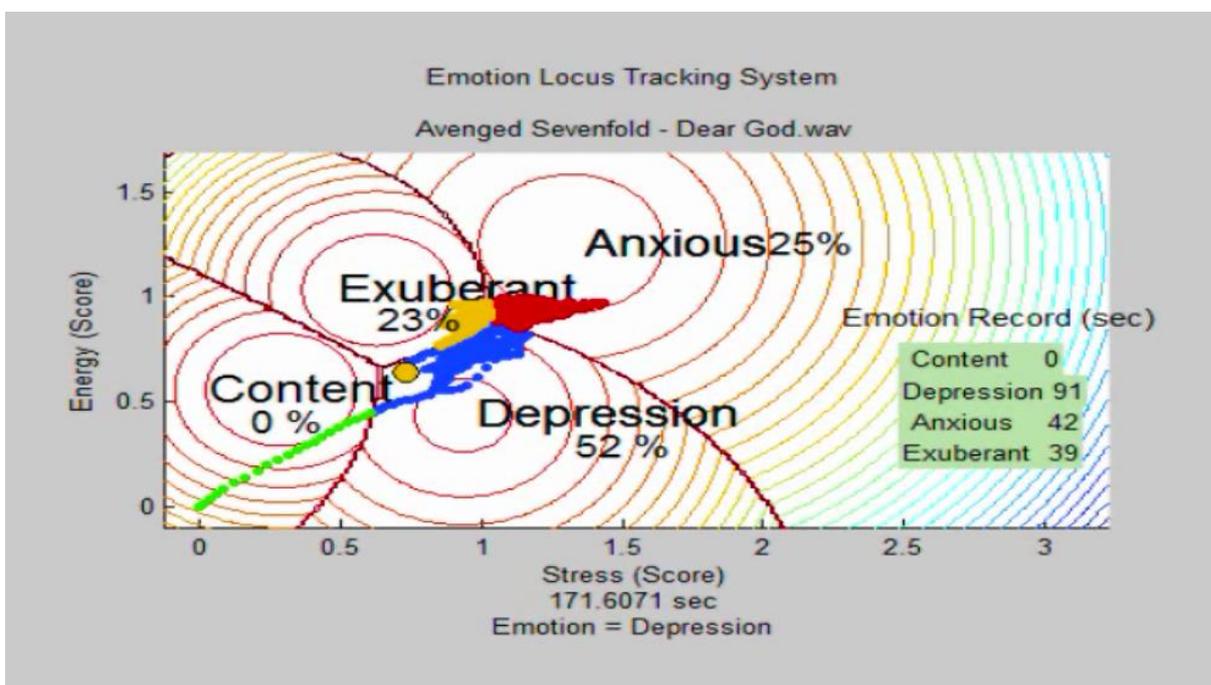


圖 5.4 曲目結束後留下之情緒軌跡與時間比例

## 六、音樂情緒研究驗證

為了測試系統是否真的能夠有效地識別歌曲營造出來的情緒氛圍，研究用收錄的一百首測試歌曲並以兩種分類器分別進行測試(章節 6.1~6.2)，而第二階段則實際請聽者進行問卷調查(章節 6.3~6.4)，研究以公正與公開的方式，將所有測試的歌曲、辨識結果、問卷結果等資訊表列於各章節內容中，供同領域中的研究者可做為比較與參考的數據。

### 6.1.1 測試樣本

筆者採用了八位著名歌手或樂團，每位歌手都是紅極一時的大明星，收集了共一百首的流行歌曲，其中許多都是大家膾炙人口的歌曲，即使沒直接聽到歌曲，但馬上就能快速勾起過去聆聽記憶中的情緒感受，希望讓讀者閱讀文章本章節時能更加熟悉與親切，並且有切身的感受而非只是閱讀一排不具有特殊感受的數字報告。

表 6.1.1 不同音樂特徵之間的相對應比例

歌手名單	數量
木匠兄妹合唱團(The Carpenters)	15
少女時代(SNSD)	13
女神卡卡(Lady Gaga)	12
聯合公園(Linkin Park)	12
披頭四(Beatles)	12
阿巴合唱團(ABBA)	9
席琳狄翁(Celine Dion)	13
安雅(ENYA)	12

### 6.1.2 範例介紹「Goodbye To Love」

我們對於音樂的情緒定義為：「整首的音樂並不是只會造成聽者產生單一種穩定的情緒，而是隨著歌曲的行進有著不同的變化過程」，因此該首歌曲於某種情緒持續的時間越長，就越能夠傳達給聽眾那樣的情緒感受歌曲共長 236 秒，這首歌曲作者為木匠兄妹，從開始到結束，大部分的時間皆是屬於比較舒適、安詳的情緒，但當歌曲進入尾聲

時，系統辨識為高興、興奮，對照原本歌曲，此時的歌手開始拉高了音量，並加入更多的樂器伴奏，使歌曲聽來更顯活潑輕快。

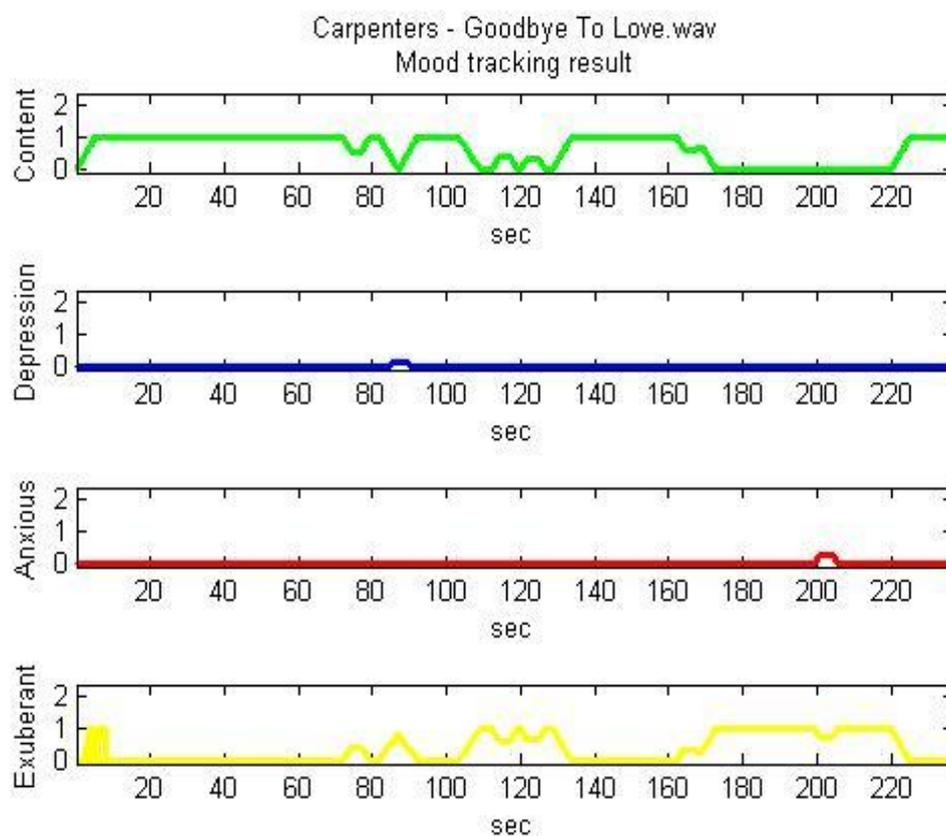


圖 6.1.1 木匠兄妹-Goodbye To Love 情緒辨識的結果(GMM)

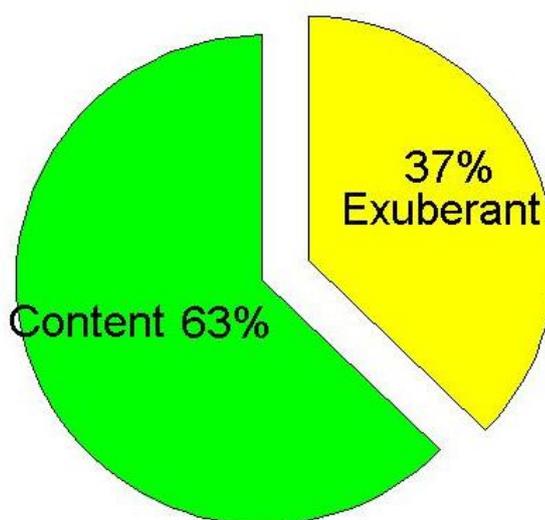


圖 6.1.2 不同情緒之累積時間比例(GMM)

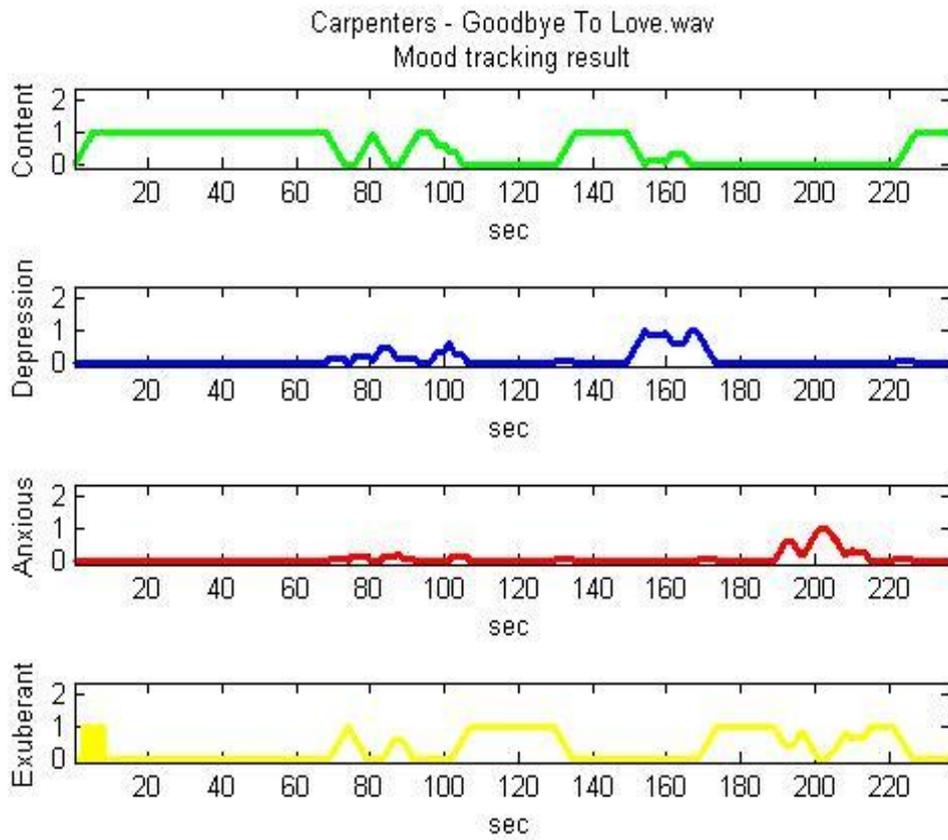


圖 6.1.3 木匠兄妹-Goodbye To Love 情緒辨識的結果(BPN)

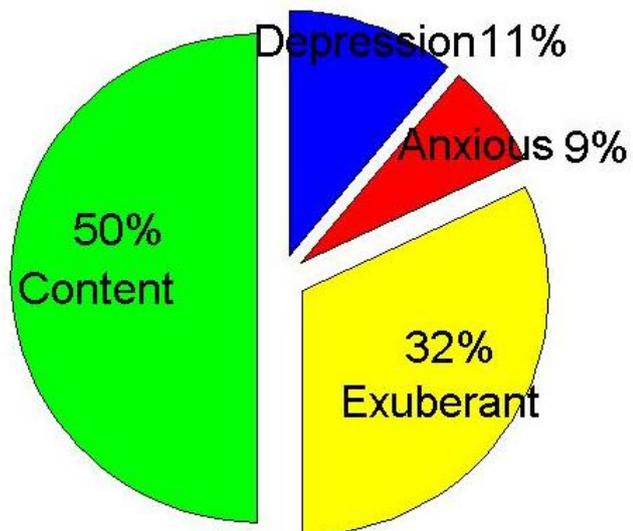


圖 6.1.4 不同情緒之累積時間比例(BPN)

## 6.2 測試結果

表 6.2.1 測試結果(一)

演唱者：木匠兄妹合唱團(The Carpenters)				
分類器：GMM				
歌曲名稱(共 15 曲)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Yesterday Once More	77.15%	13.24%	0.00%	9.61%
Superstar	22.58%	52.52%	12.89%	12.01%
Rainy Days And Mondays	53.69%	34.19%	3.92%	8.19%
Back In My Life Again	90.21%	1.91%	0.00%	7.88%
Ticket To Ride	20.37%	79.63%	0.00%	0.00%
Goodbye To Love	66.37%	0.00%	0.00%	33.63%
Bless The Beasts And Children	71.89%	16.73%	0.00%	11.38%
It's Going To Take Some Time	95.25%	4.75%	0.00%	0.00%
There's A Kind Of Hush	72.09%	25.80%	0.00%	2.11%
Sweet, Sweet Smile	68.65%	20.88%	1.80%	8.67%
I Won't Last A Day Without You	89.48%	0.00%	0.00%	10.52%
For All We Know	6.84%	86.41%	0.00%	6.75%
Touch Me When We're Dancing	89.79%	0.00%	0.00%	10.21%
Sing	72.74%	17.37%	0.00%	9.88%
Please Mr. Postman	30.63%	19.85%	14.64%	34.88%

表 6.2.1 測試結果(二)

演唱者：木匠兄妹合唱團(The Carpenters)				
分類器：BPN				
歌曲名稱(15)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Yesterday Once More	56.70%	34.81%	2.74%	5.74%
Superstar	9.51%	70.58%	5.74%	14.17%
Rainy Days And Mondays	34.37%	50.81%	4.35%	10.47%
Back In My Life Again	64.36%	26.24%	4.94%	4.45%

Ticket To Ride	20.14%	79.12%	0.74%	0.00%
Goodbye To Love	50.24%	10.62%	6.84%	32.31%
Bless The Beasts And Children	60.01%	25.60%	4.91%	9.47%
It's Going To Take Some Time	89.00%	9.25%	0.47%	1.29%
There's A Kind Of Hush	61.78%	37.62%	0.61%	0.00%
Sweet, Sweet Smile	34.18%	55.83%	9.02%	0.97%
I Won't Last A Day Without You	69.78%	16.33%	2.84%	11.05%
For All We Know	15.00%	82.26%	1.69%	1.05%
Touch Me When We're Dancing	76.10%	8.94%	6.27%	8.70%
Sing	10.42%	18.60%	16.47%	54.51%
Please Mr. Postman	56.70%	34.81%	2.74%	5.74%

表 6.2.3 測試結果(三)

演唱者：少女時代(SNSD)				
分類器：GMM				
歌曲名稱(13)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Boyfriend	2.83%	2.93%	8.14%	86.11%
ECHO	5.75%	0.00%	14.71%	79.54%
Gee.wav	5.77%	3.14%	29.43%	61.67%
HAHAHA	5.33%	0.00%	0.00%	94.67%
run devil run	11.75%	16.31%	29.66%	42.28%
visual dreams	8.79%	0.00%	3.37%	87.84%
way to go	9.50%	0.00%	2.97%	87.54%
COOKY	9.23%	0.00%	5.74%	85.02%
Mrtaxi	6.64%	0.00%	39.57%	53.79%
Genie	23.59%	0.00%	6.69%	69.72%
chocolate love	21.41%	0.00%	5.13%	73.46%
oh	15.75%	0.00%	0.00%	84.25%
hoot	4.77%	0.00%	13.38%	81.84%

表 6.2.4 測試結果(四)

演唱者：少女時代(SDSN)				
分類器：BPN				
歌曲名稱(13)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Boyfriend	1.22%	1.95%	5.59%	91.24%
ECHO	14.99%	1.36%	1.75%	81.90%
Gee.wav	4.66%	0.90%	9.36%	85.09%
HAHAHA	0.00%	0.00%	0.00%	100.00%
run devil run	0.42%	0.13%	10.38%	89.07%
visual dreams	1.43%	0.17%	16.89%	81.51%
way to go	4.08%	3.29%	1.24%	91.39%
COOKY	3.45%	0.16%	11.04%	85.35%
Mrtaxi	1.39%	19.85%	23.05%	55.70%
Genie	2.37%	0.18%	34.24%	63.21%
chocolate love	3.45%	0.40%	1.00%	95.15%
oh	5.17%	0.54%	0.61%	93.67%
hoot	3.94%	0.50%	9.13%	86.44%

表 6.2.5 測試結果(五)

演唱者：女神卡卡(Lady Gaga)				
分類器：GMM				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Bad Romance	41.83%	3.82%	2.54%	51.82%
Starstruck	24.76%	0.00%	0.00%	75.24%
Paparazzi	5.28%	3.30%	44.15%	47.27%
Boys Boys Boys	10.76%	0.00%	29.34%	59.90%
Love Game	5.41%	6.83%	45.35%	42.41%
Money Honey	16.71%	0.00%	10.35%	72.93%
Poker Face	20.40%	0.00%	0.00%	79.60%
Just Dance	34.22%	2.04%	2.53%	61.21%
Fuck You	15.15%	6.02%	27.21%	51.62%
Dirty Ice Cream	11.15%	0.00%	0.00%	88.85%
Monsta	7.69%	5.27%	43.69%	43.35%
Optimist	11.55%	0.00%	79.15%	9.30%

表 6.2.6 測試結果(六)

演唱者：女神卡卡(Lady Gaga)				
分類器：BPN				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Bad Romance	8.81%	1.44%	2.03%	87.72%
Starstruck	4.99%	14.99%	35.08%	44.94%
Paparazzi	1.90%	6.67%	4.01%	87.42%
Boys Boys Boys	5.45%	8.68%	2.37%	83.50%
Love Game	11.85%	3.68%	11.59%	72.89%
Money Honey	2.42%	3.29%	67.37%	26.91%
Poker Face	9.22%	0.30%	35.06%	55.42%
Just Dance	3.14%	6.87%	45.40%	44.59%
Fuck You	11.52%	0.99%	52.47%	35.01%
Dirty Ice Cream	8.39%	1.43%	17.79%	72.40%
Monsta	1.71%	1.72%	15.98%	80.58%
Optimist	5.86%	9.83%	38.28%	46.02%

表 6.2.7 測試結果(七)

演唱者：聯合公園(Linkin Park)				
分類器：GMM				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Papercut	6.10%	0.00%	76.66%	17.24%
One Step Closer	14.49%	0.00%	61.40%	24.12%
With You	10.39%	0.00%	55.64%	33.97%
Points Of Authority	11.97%	0.15%	72.42%	15.46%
Crawling	17.17%	10.77%	34.35%	37.71%
Runaway	9.30%	0.00%	62.14%	28.56%
By Myself	10.68%	1.88%	58.56%	28.88%
In The End	13.12%	6.61%	36.17%	44.10%
Place For My Head	21.43%	0.00%	50.42%	28.14%
Forgotten	3.90%	0.00%	61.38%	34.72%
Cure For The Itch	29.61%	2.25%	17.21%	50.94%
Pushing Me Away	8.21%	1.85%	57.83%	32.11%

表 6.2.8 測試結果(八)

演唱者：聯合公園(Linkin Park)				
分類器：BPN				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Papercut	2.90%	0.70%	70.83%	25.57%
One Step Closer	17.20%	17.75%	10.57%	54.48%
With You	7.22%	1.46%	56.57%	34.75%
Points Of Authority	0.29%	0.16%	94.29%	5.26%
Crawling	5.77%	1.32%	83.69%	9.22%
Runaway	9.07%	1.06%	63.43%	26.44%
By Myself	7.25%	3.03%	82.59%	7.13%
In The End	7.04%	12.52%	57.12%	23.32%
Place For My Head	5.98%	1.61%	74.42%	17.99%
Forgotten	2.15%	7.05%	65.55%	25.26%
Cure For The Itch	10.10%	12.08%	59.01%	18.81%
Pushing Me Away	22.51%	0.74%	58.36%	18.40%

表 6.2.9 測試結果(九)

演唱者：披頭四(Beatles)				
分類器：GMM				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Happy Together	51.06%	17.34%	1.99%	29.62%
Can't Buy Me Love	100.00%	0.00%	0.00%	0.00%
YESTERDAY	58.92%	37.77%	0.00%	3.31%
Imagine	100.00%	0.00%	0.00%	0.00%
I Want To Hold Your Hand.	6.22%	1.19%	30.50%	62.09%
Something	18.81%	55.78%	25.41%	0.00%
Hey Jude	59.89%	0.00%	0.00%	40.11%
Let It Be.	82.56%	0.00%	0.00%	17.44%
Good Morning	48.77%	0.00%	2.09%	49.13%
Love Me Do	52.69%	0.00%	0.00%	47.31%
Stand By Me	72.57%	0.00%	0.00%	27.43%
yellow submarine	67.58%	0.00%	0.00%	32.42%

The Night Before	1.1%	29.3%	51.5%	7.92%
Lucy In The Sky With Diamonds	75.22	6.31%	0.00%	18.38

表 6.2.10 測試結果(十)

演唱者：披頭四(Beatles)				
分類器：BPN				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Happy Together	41.58%	23.35%	3.90%	31.17%
Can't Buy Me Love	100.00%	0.00%	0.00%	0.00%
YESTERDAY	36.59%	48.74%	7.28%	7.39%
Imagine	89.74%	10.26%	0.00%	0.00%
I Want To Hold Your Hand.	4.02%	2.57%	23.59%	69.82%
Something	65.90%	6.15%	2.42%	25.52%
Hey Jude	30.54%	28.17%	4.28%	37.01%
Let_It_Be.	60.41%	13.39%	4.10%	22.11%
Good Morning	13.20%	21.51%	8.88%	56.41%
Love Me Do	8.78%	20.47%	16.31%	54.44%
Stand By Me	29.83%	8.14%	8.88%	53.15%
yellow submarine	30.19%	21.47%	4.87%	43.46%
The Night Before	9.67%	6.54%	9.42%	74.02%
Lucy In The Sky With Diamonds	59.05%	26.60%	2.71%	11.64%

表 6.2.11 測試結果(十一)

演唱者：ABBA				
分類器：GMM				
歌曲名稱(9)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
I Do, I Do, I Do	6.86%	0.00%	0.00%	93.14%
Loving You	44.38%	37.50%	10.07%	8.04%
Lay All Your Love On Me	25.17%	8.01%	9.11%	57.71%
Money Money Money	26.83%	37.07%	11.99%	24.12%

Dancing Queen.	39.42%	0.00%	4.64%	55.94%
Honey Honey	26.55%	2.44%	13.25%	57.75%
I Have A Dream	84.16%	9.34%	1.66%	4.83%
Mamma Mia	33.78%	4.24%	0.00%	61.99%
I saw you walking in the rain	13.85%	5.04%	20.27%	60.85%

表 6.2.12 測試結果(十二)

演唱者：ABBA				
分類器：BPN				
歌曲名稱(9)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
I Do, I Do, I Do	4.12%	0.29%	0.11%	95.48%
Loving You	20.25%	51.04%	8.35%	20.36%
Lay All Your Love On Me	4.70%	7.34%	7.47%	80.49%
Money Money Money	11.74%	48.72%	12.26%	27.28%
Dancing Queen.	7.53%	0.45%	7.62%	84.39%
Honey Honey	8.29%	1.37%	3.39%	86.95%
I Have A Dream	79.92%	14.81%	2.73%	2.54%
Mamma Mia	10.99%	10.68%	5.22%	73.10%
I saw you walking in the rain	6.62%	9.12%	22.78%	61.48%

表 6.2.13 測試結果(十三)

演唱者：席琳狄翁(Celine Dion)				
分類器：gmm				
歌曲名稱(13)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
My Heart will Go On	20.99%	62.45%	16.56%	0.00%
Think Twice	62.81%	4.24%	0.00%	32.95%
My Love	55.21%	10.49%	8.80%	25.50%
The Power of Love	46.38%	13.56%	3.57%	36.49%
Tell Him	57.75%	26.65%	0.00%	15.60%
Falling Into You	15.40%	51.47%	25.08%	8.05%
All By Myself	22.02%	53.02%	20.80%	4.16%

Immortality	81.68%	6.50%	0.00%	11.82%
Beauty And The Beast	18.13%	51.81%	27.73%	2.33%
There Comes A Time	17.38%	15.21%	54.49%	12.93%
Seduces Me	67.78%	21.38%	0.00%	10.85%
The First Time Ever I Saw Your Face	22.26%	68.59%	9.15%	0.00%
Goodbyes	21.63%	46.90%	30.14%	1.33%

表 6.2.14 測試結果(十四)

演唱者：席琳狄翁(Celine Dion)				
分類器：bpn				
歌曲名稱(13)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
My Heart will Go On	7.78%	72.57%	7.96%	11.69%
Think Twice	45.55%	2.66%	5.50%	46.30%
My Love	57.01%	3.72%	4.46%	34.80%
The Power of Love	21.97%	21.03%	6.63%	50.37%
Tell Him	36.56%	38.25%	2.91%	22.28%
Falling Into You	15.66%	11.38%	19.06%	53.90%
All By Myself	10.83%	65.78%	13.02%	10.37%
Immortality	70.02%	11.90%	4.59%	13.49%
Beauty And The Beast	7.24%	66.12%	12.79%	13.84%
There Comes A Time	9.01%	28.66%	41.89%	20.44%
Seduces Me	53.88%	28.78%	8.88%	8.45%
The First Time Ever I Saw Your Face	8.84%	82.49%	2.69%	5.97%
Goodbyes	46.67%	10.62%	10.24%	32.47%

表 6.2.15 測試結果(十五)

演唱者：安雅(ENYA)				
分類器：gmm				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Orinoco Flow	98.38%	0.00%	0.00%	1.62%
Aniron	20.64%	73.15%	6.21%	0.00%

Storms in Africa	92.62%	0.00%	0.00%	7.38%
Caribbean Blue	100.00%	0.00%	0.00%	0.00%
Wild Child	100.00%	0.00%	0.00%	0.00%
Water Shows the Hidden Heart	26.34%	73.66%	0.00%	0.00%
Anywhere Is	49.81%	22.17%	15.22%	12.79%
Cursum Perficio	33.42%	59.25%	7.33%	0.00%
Amarantine	61.25%	38.75%	0.00%	0.00%
Trains and Winter Rains	81.87%	3.75%	0.00%	14.38%
Boadicea	35.75%	64.25%	0.00%	0.00%
May It Be	26.94%	67.58%	0.00%	5.48%

表 6.2.16 測試結果(十六)

演唱者：安雅(ENYA)				
分類器：bpn				
歌曲名稱(12)	各種情緒所佔的時間比例			
	Content	Depression	Anxious	Exuberance
Orinoco Flow	38.15%	26.42%	16.55%	18.88%
Aniron	5.77%	94.23%	0.00%	0.00%
Storms in Africa	56.94%	17.37%	10.53%	15.16%
Caribbean Blue	23.81%	22.64%	24.76%	28.79%
Wild Child	89.45%	10.55%	0.00%	0.00%
Water Shows the Hidden Heart	10.26%	89.74%	0.00%	0.00%
Anywhere Is	31.46%	58.56%	6.25%	3.73%
Cursum Perficio	13.15%	86.85%	0.00%	0.00%
Amarantine	52.52%	46.08%	0.59%	0.81%
Trains and Winter Rains	64.36%	22.91%	4.37%	8.36%
Boadicea	31.14%	68.86%	0.00%	0.00%
May It Be	11.51%	84.04%	3.09%	1.36%

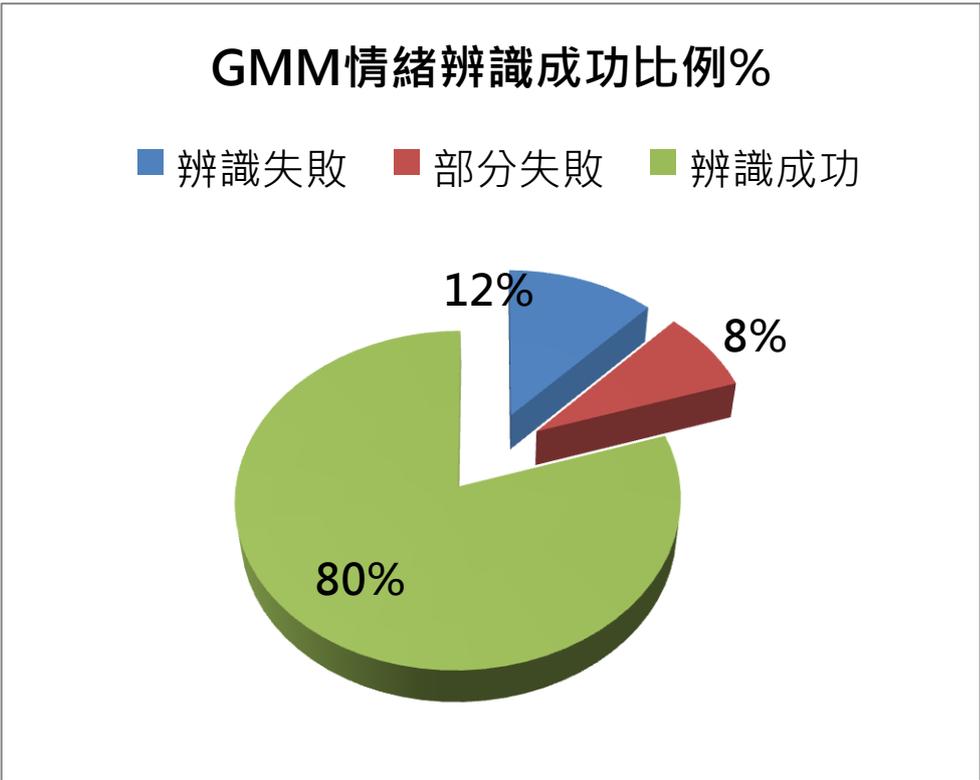


圖 6.2.1 GMM 辨識有效比例

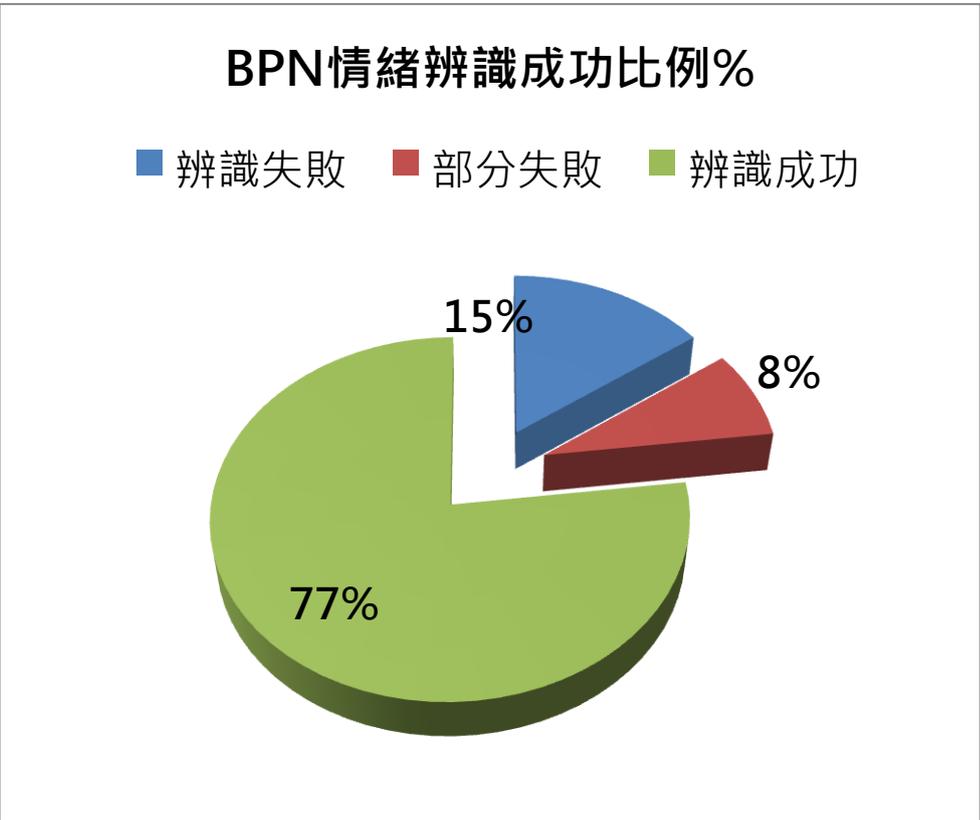


圖 6.2.2 BPN 辨識有效比例

以上為總數為一百首經過標記情緒的測試樣本音樂，經過統計兩種分類器的效能，結果如上圖 6.2.1 與圖 6.2.2 所示，圖中辨識成功代表系統使用的方法能夠有效的識別歌曲的最主要情緒，並且比重與聽起來的感受也相當接近，辨識失敗代表無法成功判斷歌曲的主要情緒是什麼，而部分失敗則是代表系統主要情緒的判定比重不合乎聽起來的感受，當最大比重的情緒判斷出現大量的誤差時，相對的第二情緒的判定也通常都會有問題，因此將上述情況發生時歸類於部分失敗，結果顯示以 GMM 的成功率較佳可達 80% 的成功率，而 BPN 則略低僅 77%。

### 6.3.1 問卷調查動機與概要

第一階段是對不同的音樂進行測試，可以驗證本研究對於不同音樂的辨識效能的廣泛度，但情緒終究是人類才有的感覺，需要大量的人來驗證成效並給予最直接的驗證數據，如此才是真正驗證研究的結果是否真能反應大部分人共同的感受，因此第二階段，我們針對本校中大學部音樂與科技通識課的同學進行問卷調查，共計 66 名學生(43 男、23 女)，由於是通識課因此學生來自各種不同科系，包含人文、商業、資訊、工程等各種不同的背景，在課堂剛開始大家就座後，讓他們聽準備好的樣本音樂，問卷設計參考自文獻(3)，這篇文獻主要歸納出了九種藉由歌曲喚醒聽者的情緒感受，這九種情緒是符合人們的情緒反應且具有鑑別性的情緒，試驗中會要求受測者在聆聽音樂後，就他個人的對於這首歌曲聆聽後的感受，以李克特量尺(likert scale)對每種情緒給予從 1(not at all) 至 5(very much)個別給分，另外問卷於標示有此次測試歌曲的曲名、歌手、曲風等資訊，最後在交由我們統計。該試驗最後得到了許多寶貴且客觀的數據，下頁我們將問卷調查數據與我們的研究得到的結果放在一起比較之間的差異與驗證。

### 6.3.2 問卷調查 曲目一：Dear God

演唱者：Avenged Sevenfold

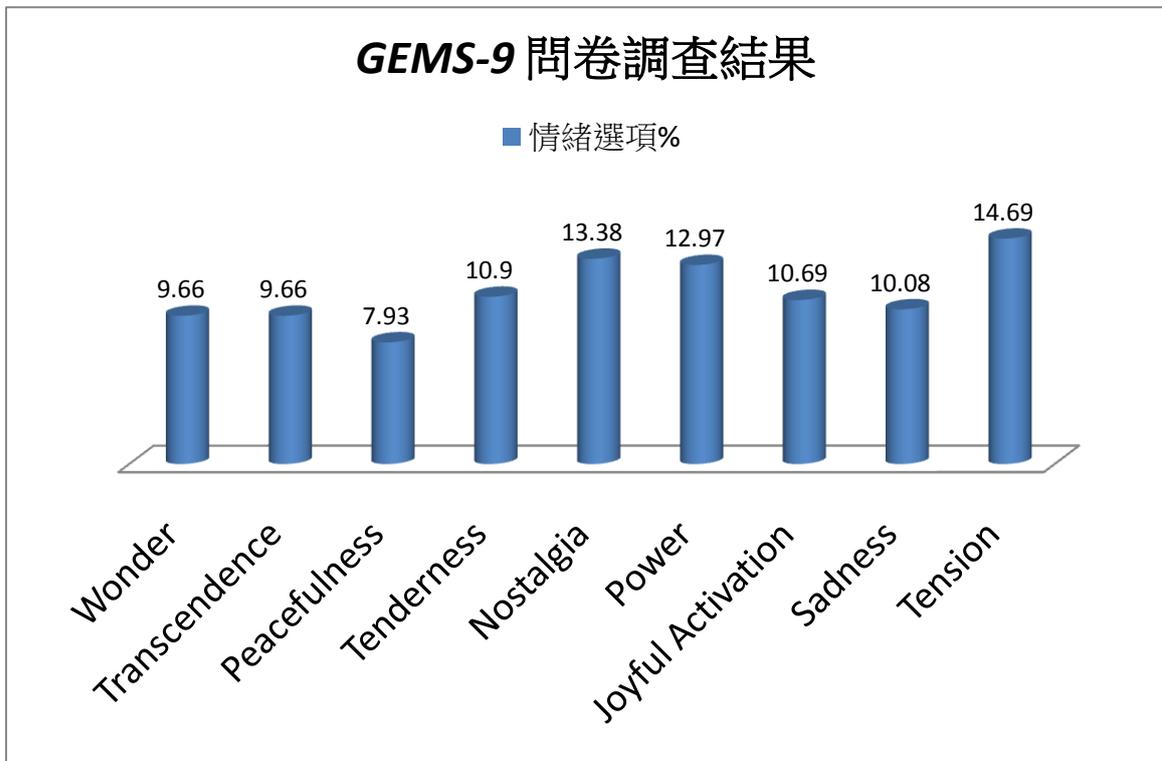


圖 6.3.1 GEMS-9 問卷調查結果(曲目一)

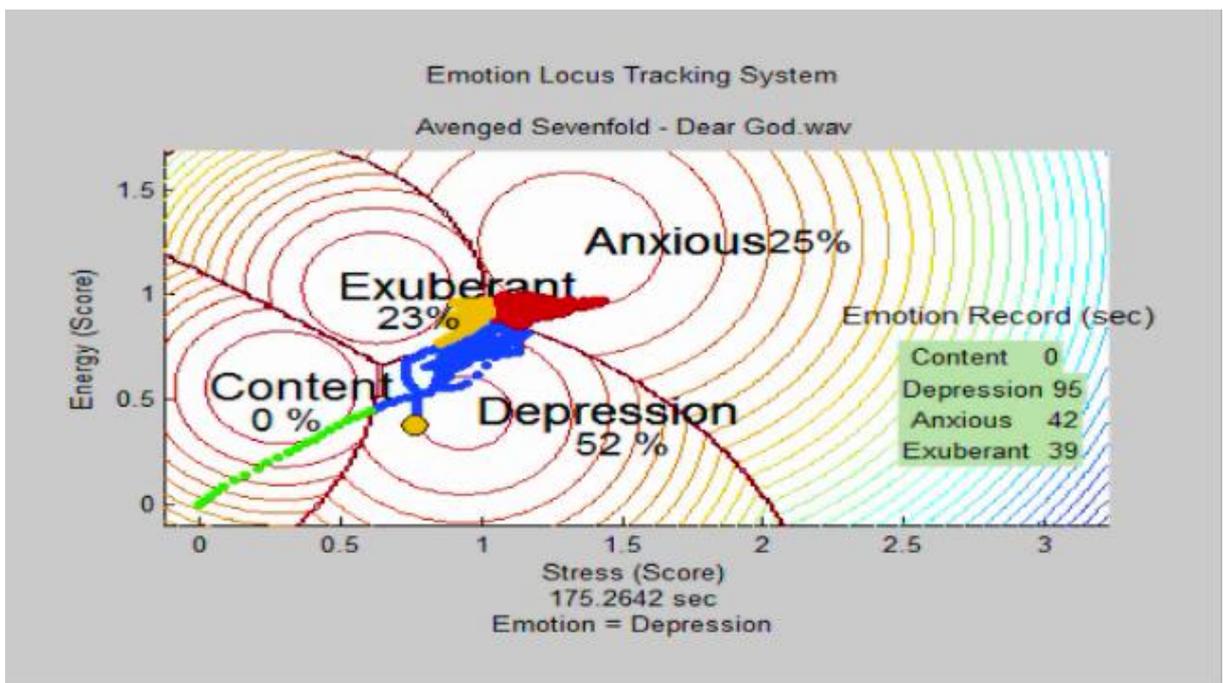


圖 6.3.2 系統追蹤結果情緒比例(曲目一)

表 6.3.1 問卷調查結果與研究結果比較表(曲目一)

類組	GEMS-9 問卷結果%	研究結果%
A類 <b>SUBLIMITY</b>	Wonder 9.66 Transcendence 9.66 Peacefulness 7.93 Tenderness 10.90 Nostalgia 13.38 合計： <b>51.55%</b>	Content : 1 & Depression : 52  合計： <b>53%</b>
B類 <b>VITALITY</b>	Power 12.97 Joyful Activation 10.69 合計： <b>23.66%</b>	Exuberance <b>23%</b>
C類 <b>UNEASE</b>	Sadness 10.08 Tension 14.69 合計： <b>24.77%</b>	Anxious <b>25%</b>

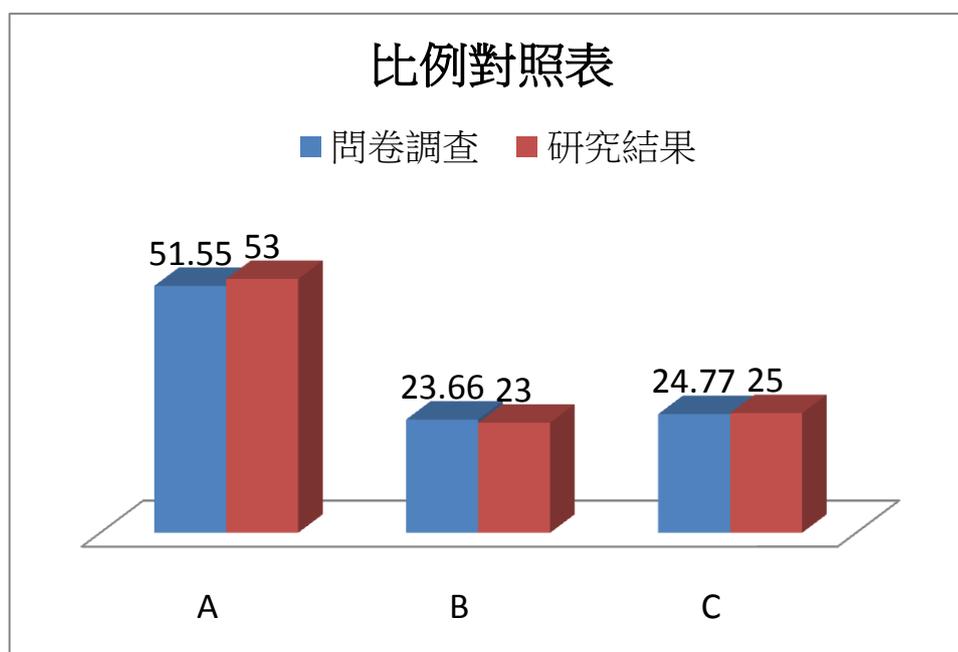


圖 6.3.3 驗證結果比例對照(曲目一)

### 6.3.3 問卷調查 曲目二：My Heart Will Go On

演唱者：Celine Dion

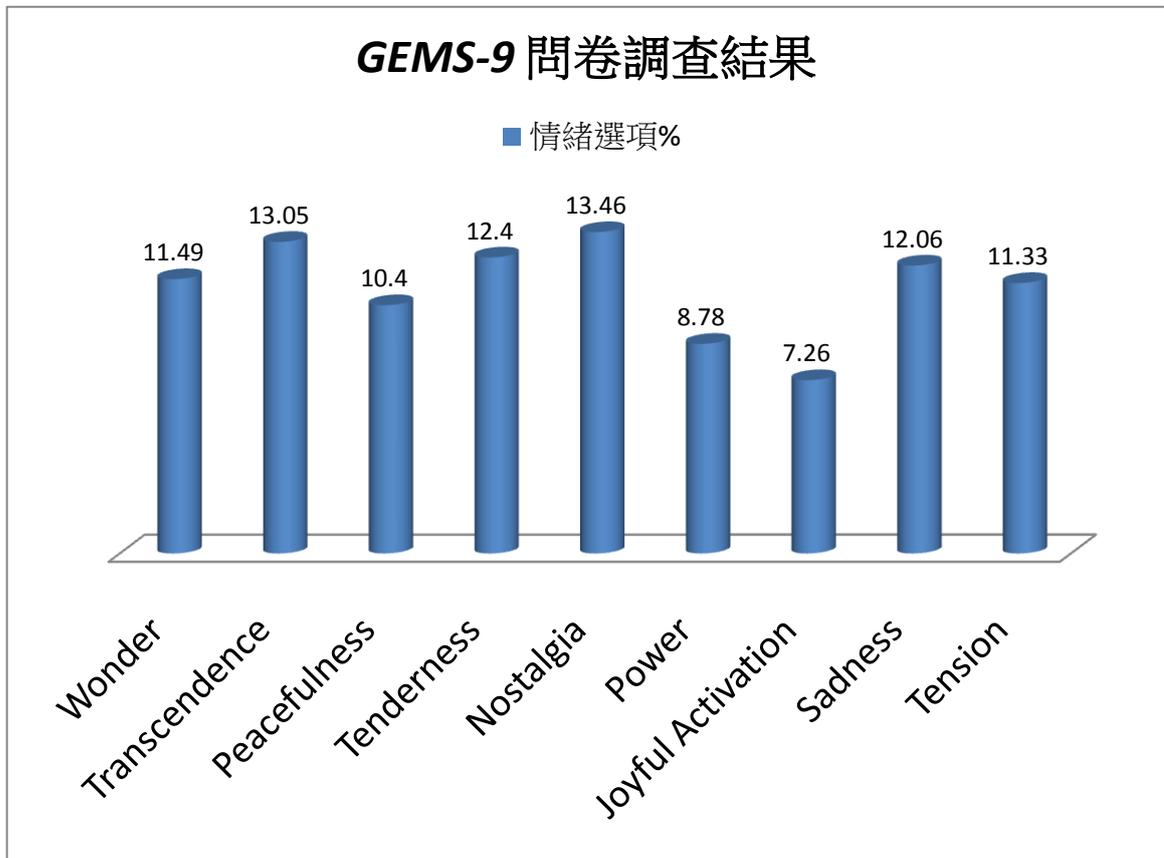


圖 6.3.4 GEMS-9 問卷調查結果(曲目二)

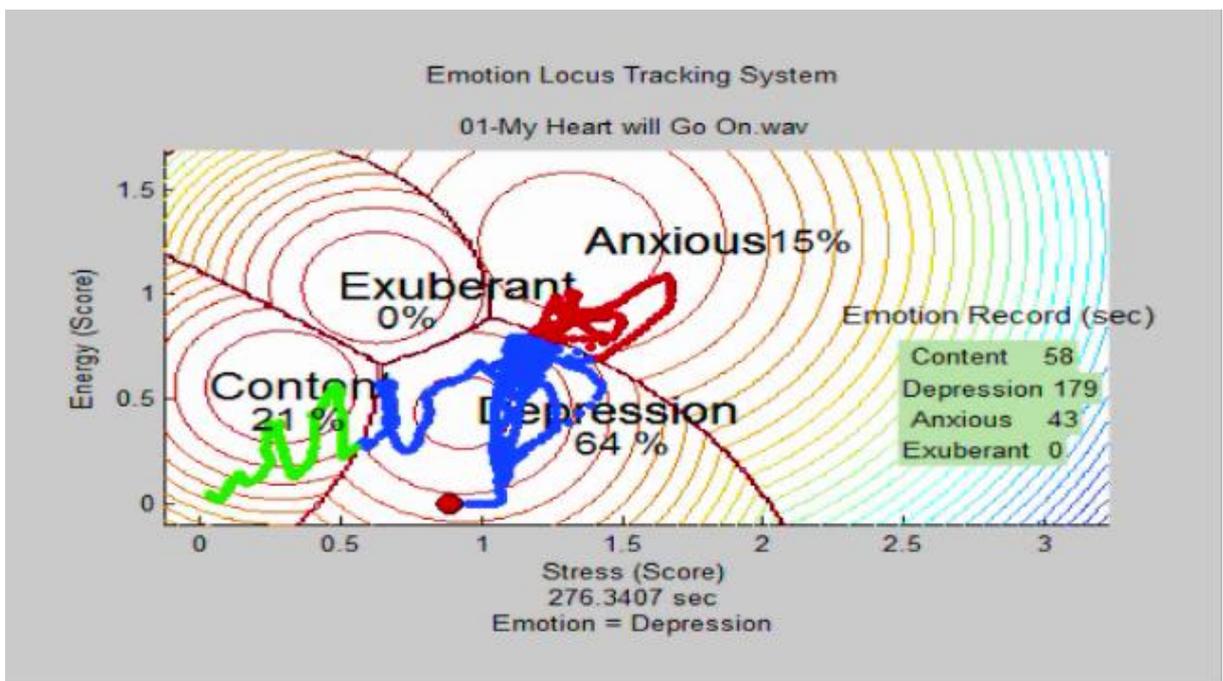


圖 6.3.5 系統追蹤結果情緒比例(曲目二)

表 6.3.2 問卷調查結果與研究結果比較表(曲目二)

類組	GEMS-9 問卷結果%	研究結果%
A類 <b>SUBLIMITY</b>	Wonder 11.49 Transcendence 13.05 Peacefulness 10.40 Tenderness 12.40 Nostalgia 13.46 合計： <b>60.58%</b>	Content : 21 & Depression : 64  合計： <b>85%</b>
B類 <b>VITALITY</b>	Power 8.78 Joyful Activation 7.26 合計： <b>16.06%</b>	Exuberance <b>0%</b>
C類 <b>UNEASE</b>	Sadness 12.06 Tension 11.33 合計： <b>23.39%</b>	Anxious <b>15%</b>

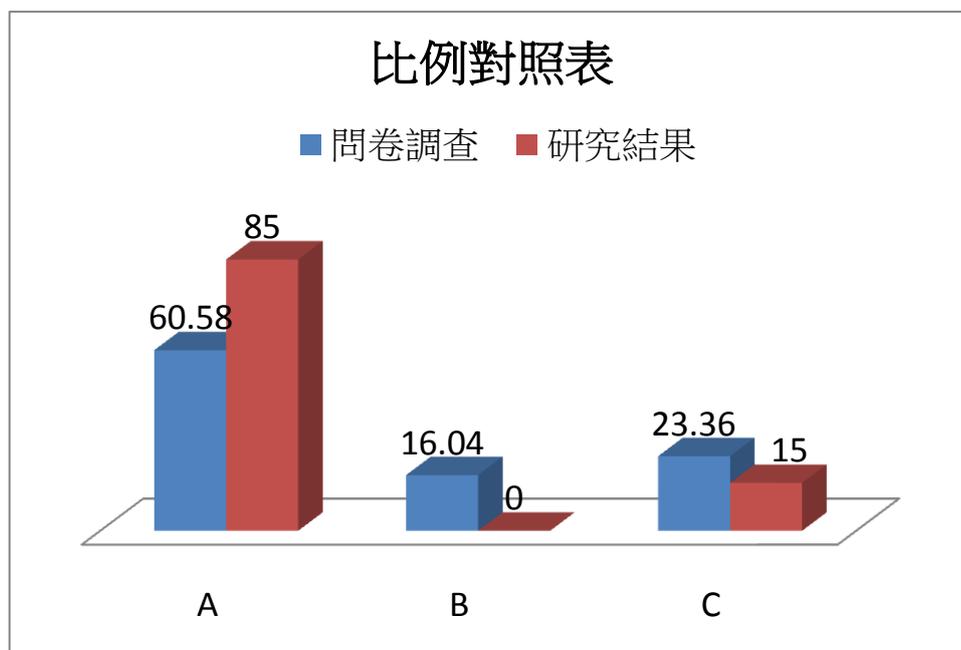


圖 6.3.6 驗證結果比例對照(曲目二)

### 6.3.4 問卷調查 曲目三：Goodbye To Love

演唱者：Carpenters

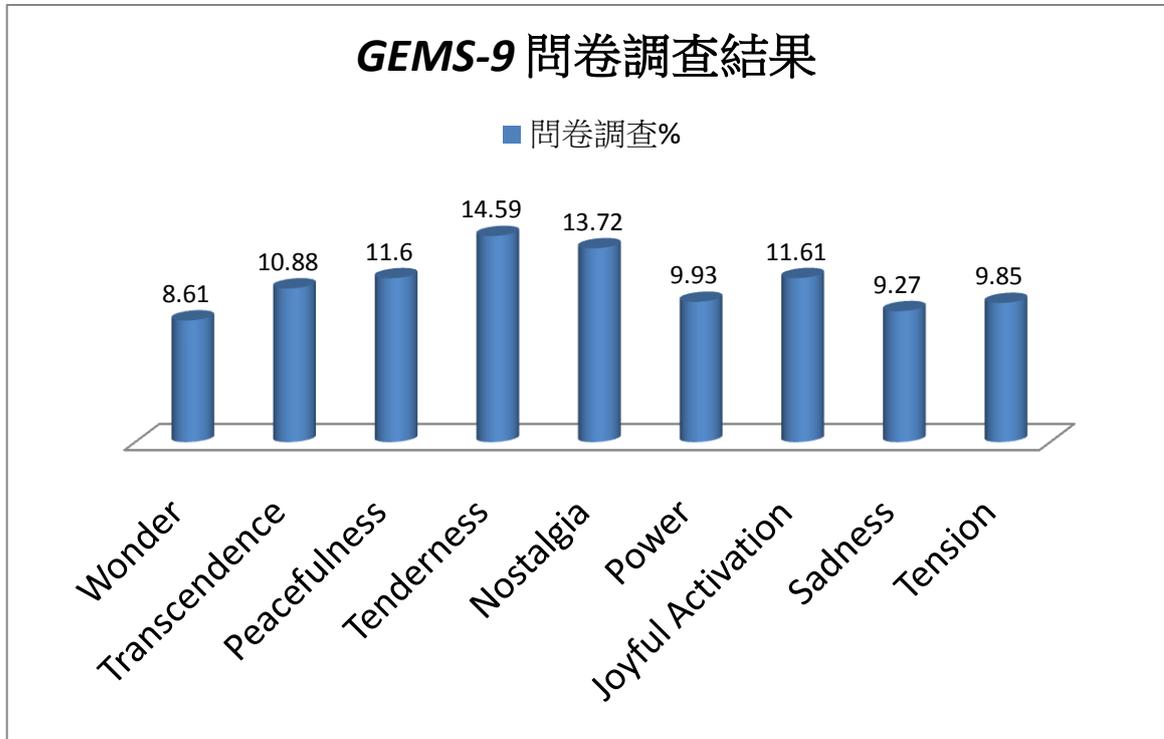


圖 6.3.7 GEMS-9 問卷調查結果(曲目三)

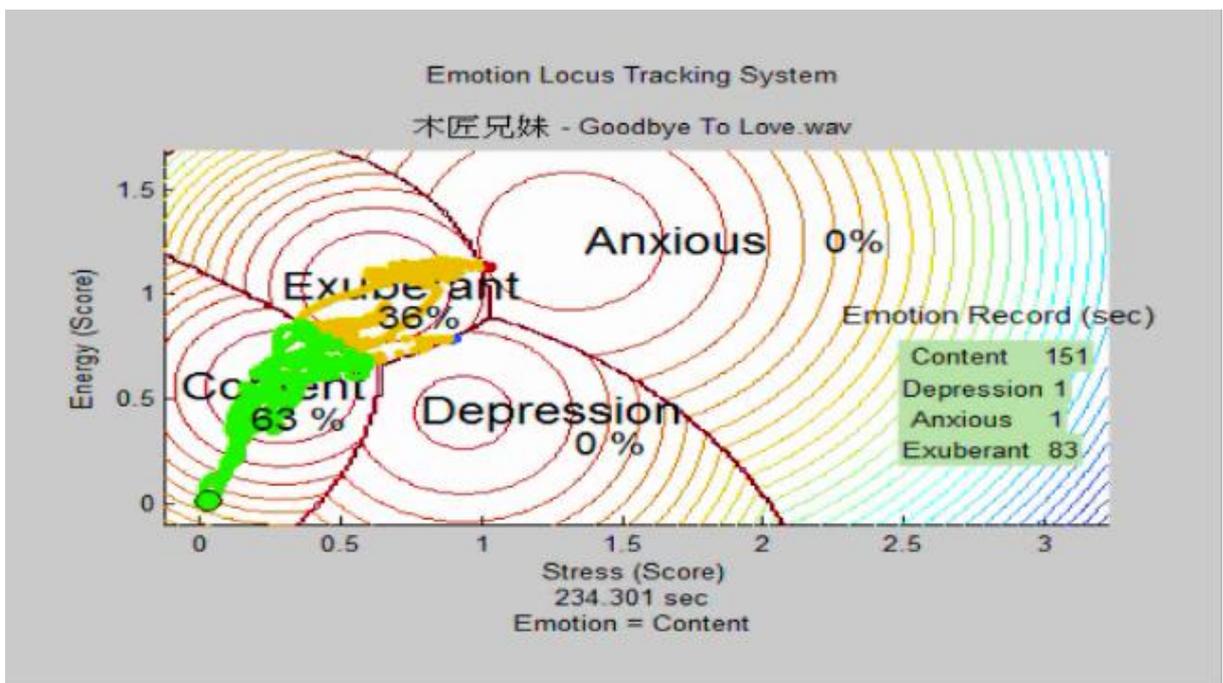


圖 6.3.8 系統追蹤結果情緒比例(曲目三)

表 6.3.3 問卷調查結果與研究結果比較表(曲目三)

類組	GEMS-9 問卷結果%	研究結果%
A類 <b>SUBLIMITY</b>	Wonder 8.61 Transcendence 10.88 Peacefulness 11.60 Tenderness 14.59 Nostalgia 13.73 合計： <b>59.27%</b>	Content : 1 & Depression : 52  合計： <b>62%</b>
B類 <b>VITALITY</b>	Power 9.93 Joyful Activation 11.61 合計： <b>21.4%</b>	Exuberance <b>36%</b>
C類 <b>UNEASE</b>	Sadness 9.27 Tension 9.85 合計： <b>19.12%</b>	Anxious <b>1%</b>

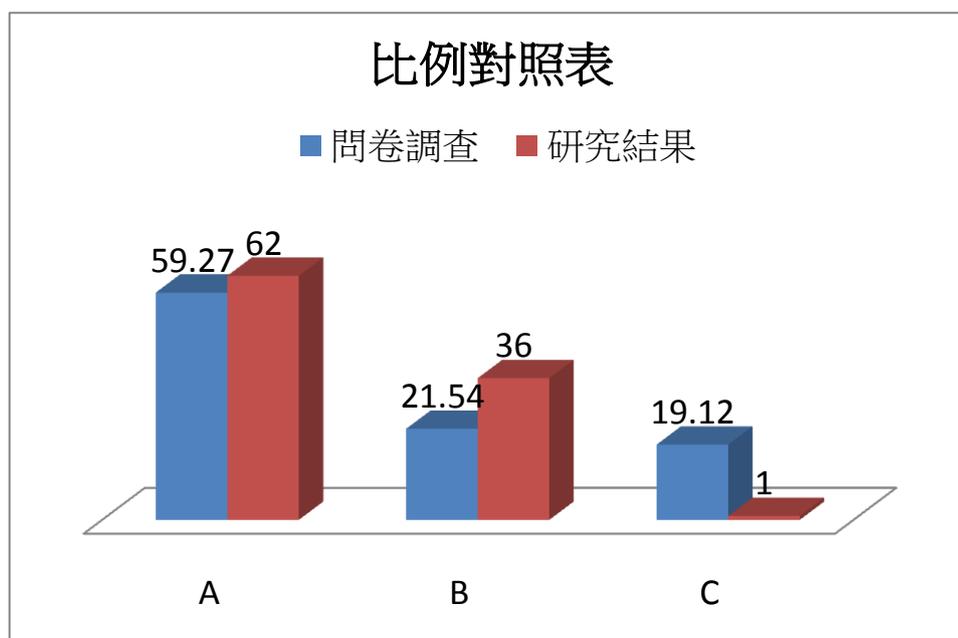


圖 6.3.9 驗證結果比例對照(曲目三)

### 6.3.5 問卷調查 曲目四：La Vie En Rose

演唱者：Edith Piaf

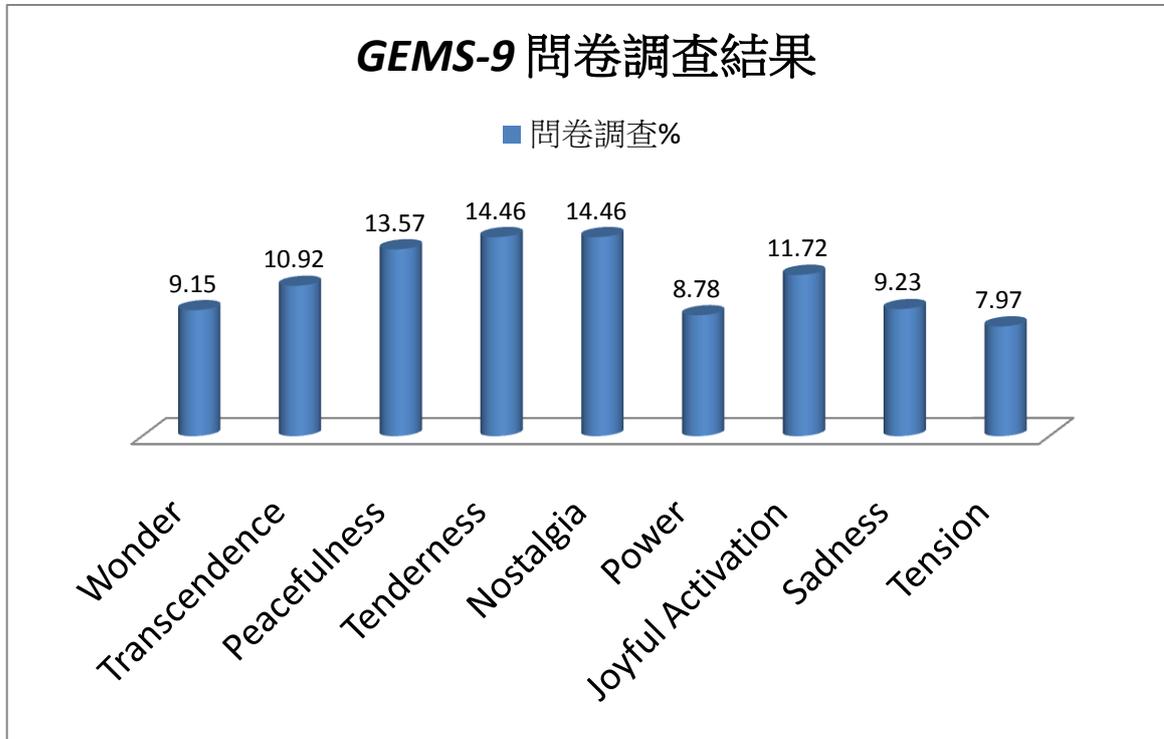


圖 6.3.10 GEMS-9 問卷調查結果(曲目四)

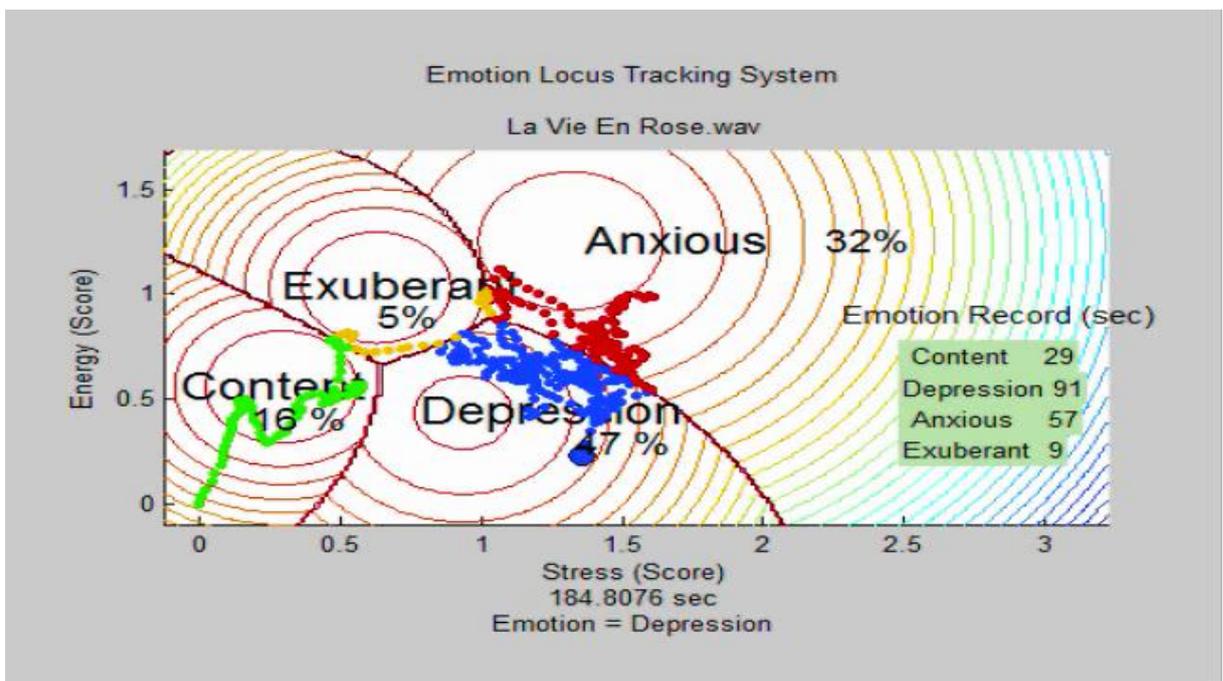


圖 6.3.11 系統追蹤結果情緒比例(曲目四)

表 6.3.4 問卷調查結果與研究結果比較表(曲目四)

類組	GEMS-9 問卷結果%	研究結果%
A類 <b>SUBLIMITY</b>	Wonder 9.15 Transcendence 10.92 Peacefulness 13.57 Tenderness 14.46 Nostalgia 14.46 合計： <b>62.57%</b>	Content : 1 & Depression : 52  合計： <b>63%</b>
B類 <b>VITALITY</b>	Power 8.78 Joyful Activation 11.72 合計： <b>20.52%</b>	Exuberance <b>36%</b>
C類 <b>UNEASE</b>	Sadness 9.27 Tension 9.85 合計： <b>17.27%</b>	Anxious <b>1%</b>

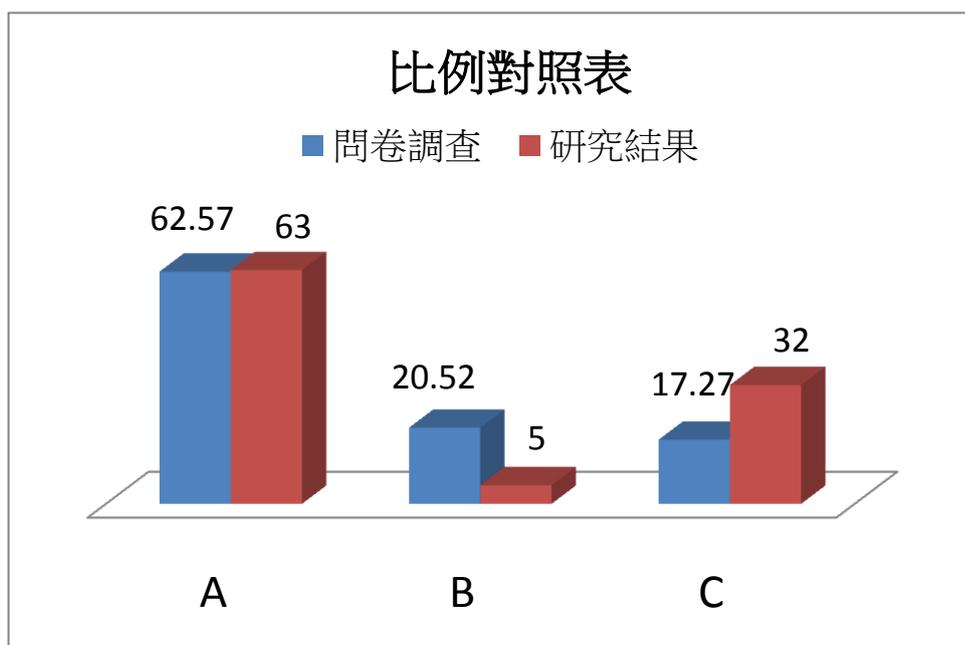


圖 6.3.12. 驗證結果比例對照(曲目四)

### 6.3.6 問卷調查 曲目五： Nocturne op 9 no 2

演唱者：Chopin

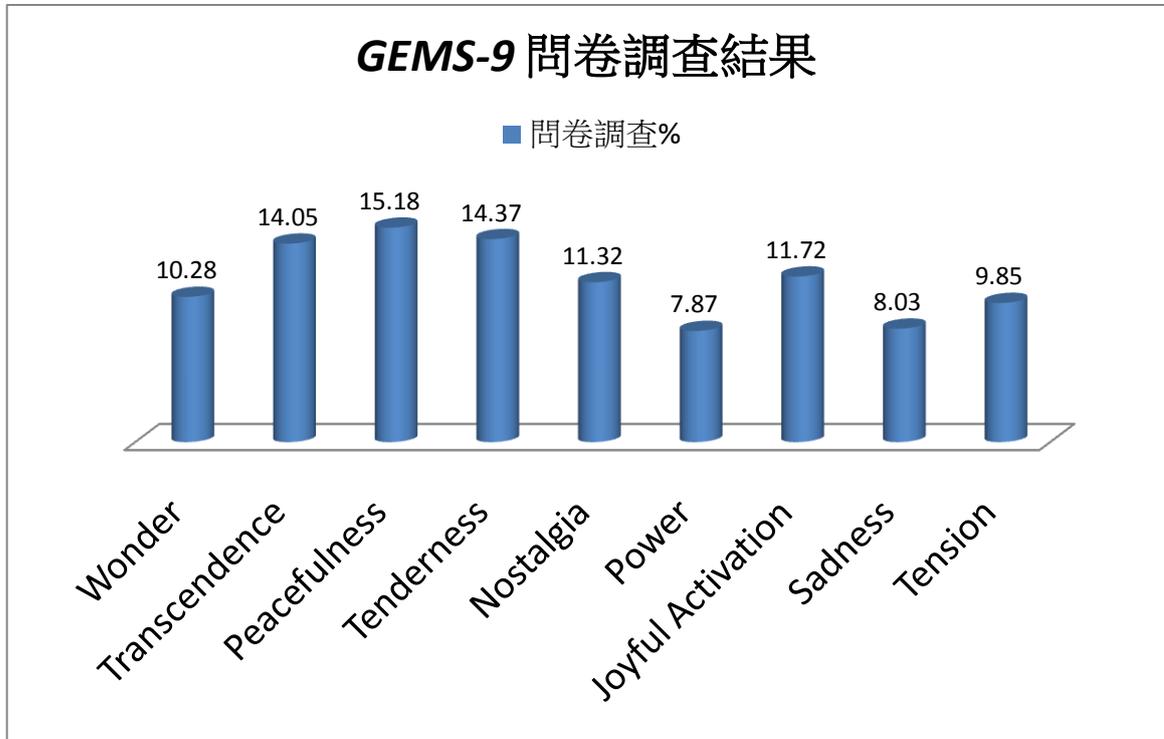


圖 6.3.13 GEMS-9 問卷調查結果(曲目五)

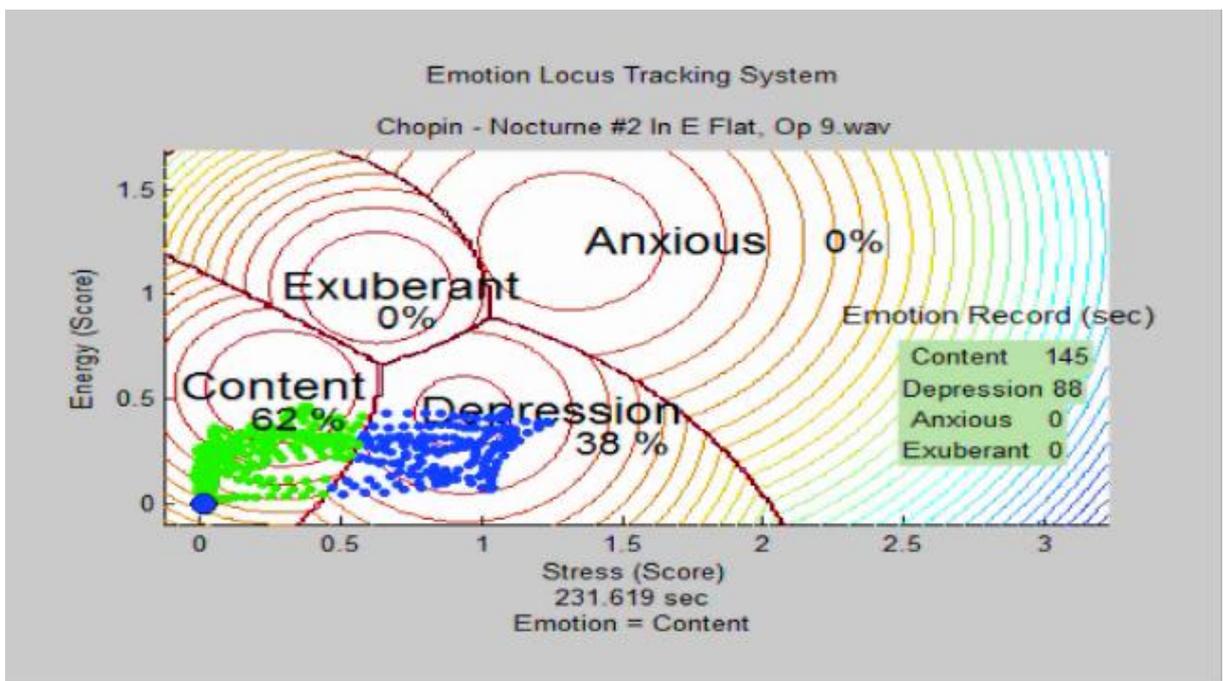


圖 6.3.14 系統追蹤結果情緒比例(曲目五)

表 6.3.5 問卷調查結果與研究結果比較表(曲目五)

類組	GEMS-9 問卷結果%	研究結果%
A類 <b>SUBLIMITY</b>	Wonder 10.28 Transcendence 14.05 Peacefulness 15.18 Tenderness 14.37 Nostalgia 11.32 合計： <b>65.27%</b>	Content：62 & Depression：38  合計： <b>100%</b>
B類 <b>VITALITY</b>	Power 7.87 Joyful Activation 11.72 合計： <b>19.54%</b>	Exuberance <b>0%</b>
C類 <b>UNEASE</b>	Sadness 8.03 Tension 7.14 合計： <b>15.17%</b>	Anxious <b>0%</b>

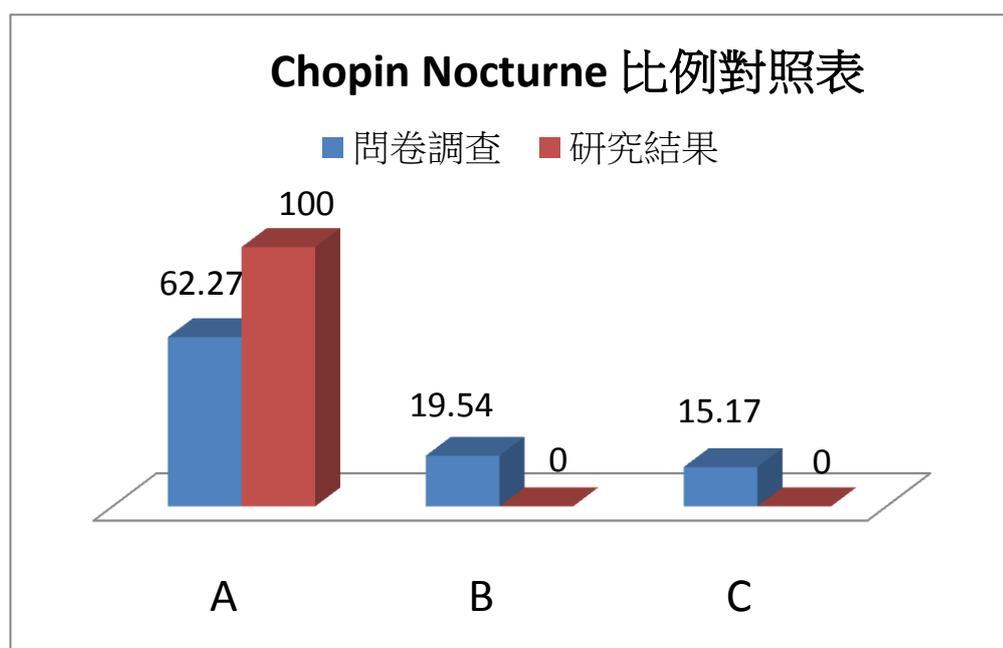


圖 6.3.15.驗證結果比例對照(曲目五)

#### 6.4 問卷調查結果討論

本研究對於音樂的情緒定義為：「整首的音樂並不是只會造成聽者產生單一種穩定的情緒，而是隨著歌曲的行進有著不同的變化過程」，因此該首歌曲於某種情緒持續的時間越長，就越能夠傳達給聽眾那樣的情緒感受，所以採用時間累計比例方式做為情緒辨識的結果，而不刻意去指出該音樂完整屬於何種情緒的分類，這樣的方法可以減輕當歌曲營造出來的情緒不顯著或歌曲本身主觀性太強烈時，系統與使用者之間認知不同造成的磨擦。另外人們對於不同音樂的專注力都不一樣，每個人對情緒的主觀認知也不盡相同，本試驗為得到客觀的數據，測試者為各領域的年輕大學生(19~25歲)，得到的是綜合性的結果，而由試驗提供五組驗證案例，評分高的情緒表示感受最直接強烈，本系統利用時間的累計方式皆可成功的判斷出歌曲造成人們最主要的情緒感受；深入討論每個案例的各種情緒比例分配，主要情緒類別較準確，次要情緒類別系統結果與問卷結果尚存有落差，系統的情緒分佈較問卷結果集中，造成此結果的原因除了本研究之系統雖然採用時間累計比例方式做為辨識結果，然而一首歌曲即使情緒有起有伏，但過程仍不至於是變化劇烈到每種情緒都有涵蓋，但人的情緒感受卻不是那麼單純，並且在考量到每個人都有不同的喜好或反應，音樂的情緒追蹤就更為難以有百分之百的準確判斷。

## 七、音樂情緒之應用

音樂一直都扮演著重要的腳色於人們的日常之中，各式各樣的音樂能傳遞與喚起人們各種不同的情緒反應，因此若在日常生活中，皆能善用音樂的情感特性，來營造各種想要的環境氣氛，那麼生活的環境將會更加美妙，例如：在幼童遊戲的場所就適合利用能帶給人快樂、興奮感覺的音樂，讓小孩子會更快樂，當在辦派對的時候，就適合用些能讓情緒更加高亢激昂的音樂，讓與會的朋友更容易擺脫牽掛，玩得更加盡興，而與情人用餐的環境就適合些能營造出浪漫、感性的氣氛的音樂，睡眠的環境適合用些能營造出方便入睡的音樂，利用音樂將情緒帶入日常的生活環境中並且幫助人們營造出想要的情緒感受與是合的環境氣氛，除了生活中，無論是用於醫療環境、教育環境或是諮商環境等等，都很適合用音樂來碩造出理想的環境感受。

### 環境一：慶生會

當要幫小孩辦生日會時，就適合能營造歡樂、興奮、有活力的音樂，讓所有的孩童都玩得更快樂。範例歌曲：HAHAHA

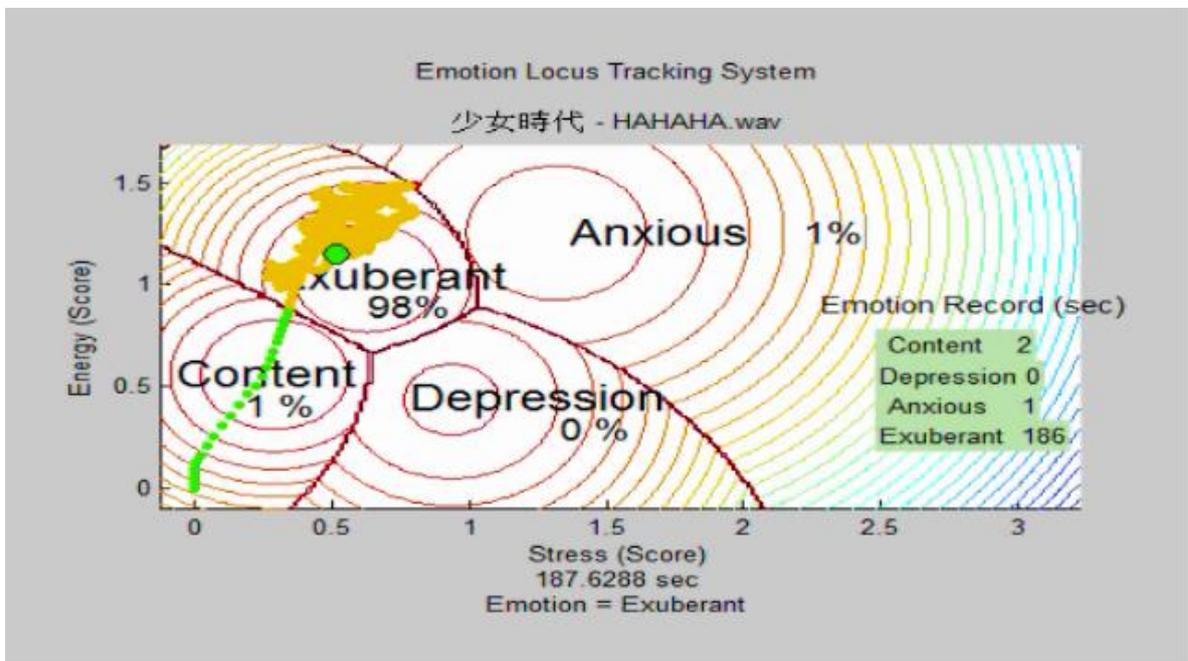


圖 7.1 情緒軌跡與時間比例(HAHAHA)

### 環境二：動人場景

出差多日的情人回來，家中準備了豐盛的燭光晚餐，此時再搭配上一些歷久彌新的動聽情歌，整體氣氛的營造相當的浪漫。範例歌曲：My Heart Will Go On

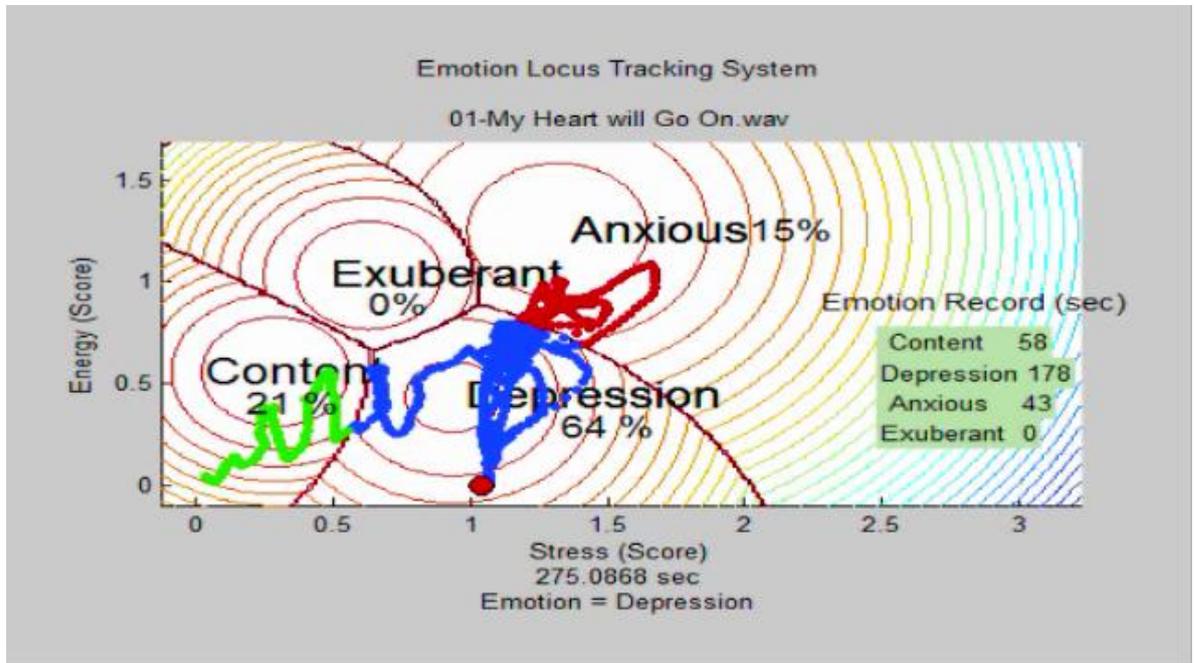


圖 7.2 情緒軌跡與時間比例(My Heart Will Go On)

### 環境三：舉辦派對

當舉辦派對或舞會剛開始時，就適合選用一些聽起來激動亢奮、生氣焦躁的歌曲讓與會的來賓都能盡情釋放出他的熱情。範例歌曲：It's My Life

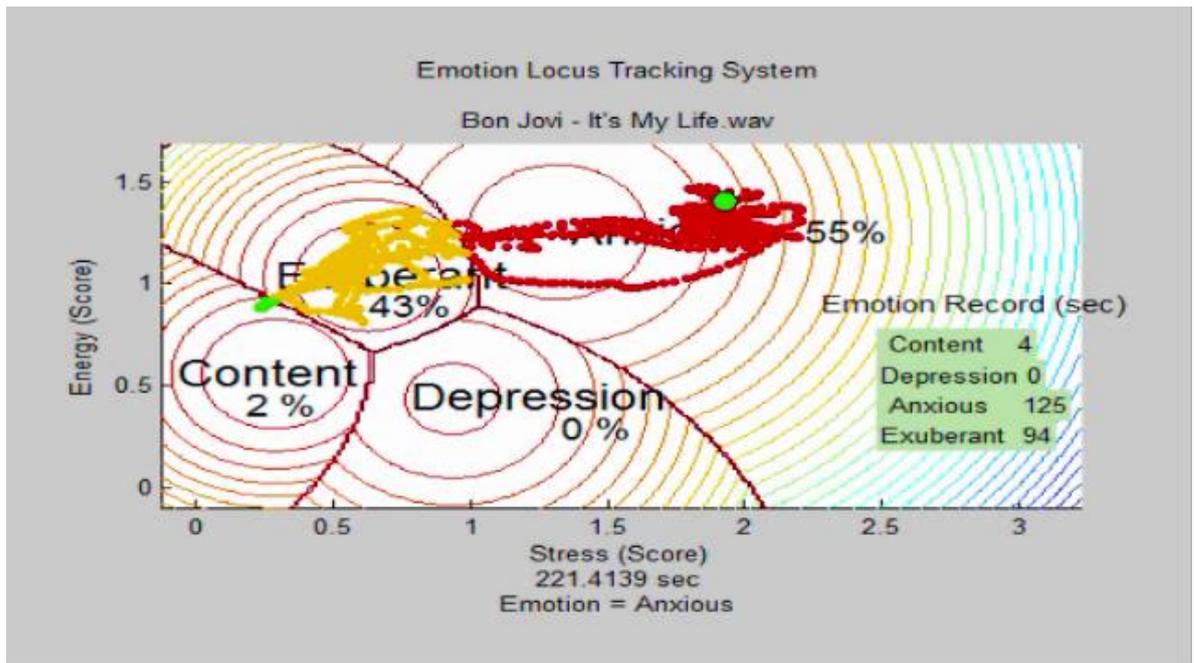


圖 7.3 情緒軌跡與時間比例(It's My Life)

環境四：離別場景

替離職的同事或朋友送行或宴會結束散場時，就適合用一些熱鬧但是略帶悲情的歌曲，來營造出曲終人散時，離情依依的氣氛。範例歌曲：Dear God

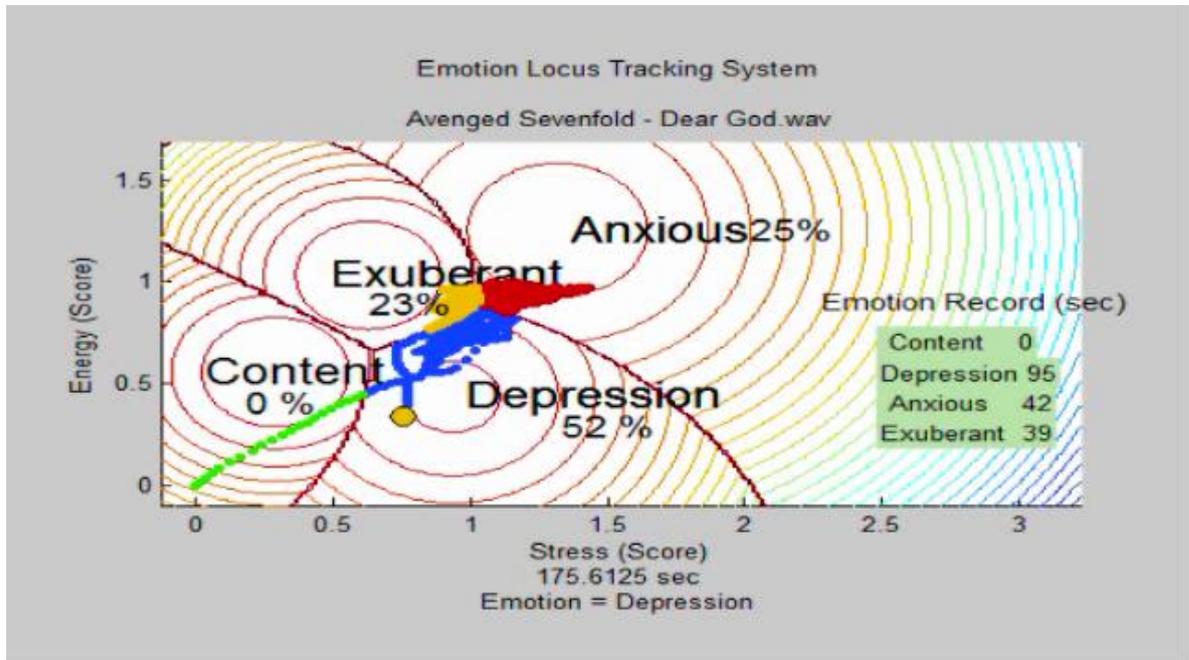


圖 7.4 情緒軌跡與時間比例(Dear God)

## 八、結論

此研究建構了一個具時變情緒軌跡的音樂情緒追蹤系統，音樂樣本為 WAVE 檔，備有共 200 個每個 30 秒已標記情緒的音樂片段做為訓練音檔，另收錄 100 首完整的國外流行音樂歌曲做為測試樣本；研究的情緒模型中有四種類型的情緒分別為(1)舒適的 (2)沮喪的 (3)焦躁的 (4)激勵人心的。由訓練音檔中萃取出五種的特徵，包括音量、音樂事件密集度、頻譜分佈、調式與音程的和諧度，再由系統給予情緒計分，計分後的結果數值會呈現連續如軌跡一般的分佈，最後收集所有訓練音檔提供的資料利用分類器訓練出分類模型，其中 GMM 對測試樣本的辨識成功率可達 80%，BPN 的成功率也有 77%。研究提供了靜態的情緒追蹤結果與時間比例以及動態的情緒可視化的方式共三種方法來呈現研究結果。為更客觀的來驗證研究成效，另採用問卷方式進行驗證，調查對象為本校同學共 43 男 23 女，調查結果顯示，系統對於成分最多之情緒確實可以準確判斷，但對於其他第二線的情緒的判定，或許是在調查對象本身的主觀因素影響下，與調查結果尚有差異存在，另外，我們尚未討論不同的音樂類型對於系統造成的影響，例如：古典樂與流行音樂或西洋樂曲與中式戲曲，不同音樂族群的愛好者想必對於歌曲的解讀也會有不一樣的想法，在以大學生做為樣本的調查與研究比對之下也顯示出，古典音樂的情緒辨識結果明顯不如流行音樂來的準確。

## 參考文獻

1. 傅俊傑，「具時變情緒軌機介面之自動音樂情緒追蹤系統」，國立交通大學，碩士論文，民國九十九年
2. 曾于恬，「以情緒感受為基礎之自動音樂選曲系統」，國立交通大學，碩士論文，民國九十九年
3. 林明穎，「音樂與情緒關係定位之研究」，國立台灣師範大學教育心理與輔導學系，碩士論文，民國九十八年
4. 張春興，現代心理學，台灣東華書局股份有限公司，台北，2004
5. M Zentner, D Grandjean, KR Scherer, “Emotions evoked by the sound of music: Characterization, classification, and measurement,” *Emotion*, Vol. 8, No. 4, pp. 494–521, 2008
6. Rentfrow, P. J., Gosling, S. D., “The do re mi’s of everyday life. The structure and personality correlates of music preferences,” *Journal of Personality and Social Psychology*, 84, 1236–1256, 2003
7. Feng, Y., Zhuang, Y., Pan, Y., “Music information retrieval by detecting mood via computational media aesthetics,” *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence* Washington, DC, USA 2003
8. W. Chai and B. Vercoe, “Detection of key change in classical piano music”, *ISMIR*, London, 2005
9. L.Lu, D.Liu, and H.-J.Zhang, “Automatic mood detection and tracking of music audio signals,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol 14 No 1,; pp. 5-18, January 2006
10. K Trohidis, G Tsoumakas, G Kalliris, “Multilabel classification of music into emotion,” *ISMIR*, 2008
11. George Tzanetakis, and Perry Cook, “Musical Genre Classification of Audio Signals” *IEEE Transactions on, Speech, and Processing*, vol 10, No. 5, July 2002

12. Juslin, P.N. 1997 “Emotional Communication in Music Performance: A Functionalist Perspective and Some Data.” *Music Perception* vol 14 No 4, pp 383–418
13. K. Hevner. “Experimental studies of the elements of expression in music.” *The American Journal of Psychology*, vol 48 No2 pp:246–268, 1936
14. Steven R. Livingstone, Emery Schubert and Janeen D. Loehr, “Emotional arousal and the automatic detection of musical phrase boundaries”, *International Symposium on Performance Science* ISBN 978-94-90306-01-4 2009
15. P. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, Cambridge, MA, 1974
16. SR Livingstone, AR Brown, R Muhlberger “Influencing the Perceived Emotions of Music with Intent” In, Innocent, T. (Ed.) *Third Iteration*, Melbourne: CEMA, pp. 161-170
17. R Bresin, A Friberg “Emotional Coloring of Computer-Controlled Music Performances” *Computer Music Journal*, vol 24 No. 4, pp:44-63, MIT Press. Winter 2000
18. Meyers, O.C., “A Mood-Based Music Classification and Exploration System,” MS Thesis, Massachusetts Institute of Technology (MIT), USA 2007
19. J. A. Russell, “A circumplex model of affect”, *Journal of Personality and Social Psychology*, vol 39 No. 6, pp:1161–1178, 1980
20. SR Livingstone, AR Brown, “Dynamic Response: Real-Time Adaptation for Music Emotion” *Second Australasian Conference on Interactive Entertainment*, Sydney, Australia, 2003
21. R. E. Thayer, “The origin of everyday moods: managing energy”, Oxford University Press, New York, 1996
22. Thayer, R.E. “The Biopsychology of Mood and Arousal,” Oxford University Press, NY, 1989
23. C. Duxbury, M. Sandler, and M. Davies, “A hybrid approach to musical note onset detection,”. *Digital Audio Effects Conf. (DAFX, '02)*, pp. 33–38, Hamburg, Germany, 2002

24. Anders Friberg, Vittorio Colombo, “Generating Musical Performances with Director Musices” *Computer Music Journal*, vol 24 No 3, pp. 23–29, 2000
25. Fabien Gouyon and Perfecto Herrera, “A beat induction method for musical audio signals,” in *Proc. WIAMIS Special session on Audio Segmentation and Digital Music*, 2003
26. Hanna Järveläinen, Vesa Välimäki, “Audibility of initial pitch glides in string instrument sounds,” *Proceedings of the International Computer Music Conference*, vol 17-23 ,pp:282–285, Havana, Cuba, September 200.
27. Andre Holzapfel and Yannis Stylianou, “Beat tracking using group delay based onset detection,” in *MIREX at 7th International ISMIR 2008 Conference*, 2008
28. V. Kandia and Y. Stylianou, “Detection of clicks based On group delay,” *Accepted in Canadian Acoustics*, 2008
29. Simon Dixon, “Onset detection revisited,” In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, Sept. 18–20, 2006
30. K Lee, “Automatic chord recognition from audio using enhanced pitch class profile”, *Proceedings of the International Computer Music 2006*.
31. Dixon, S. “An interactive beat tracking and visualisation system,” *International Computer Music Conference*, pages 215–218, San Francisco CA
32. Jyh-Shing Roger Jang, “Audio Signal Processing and Recognition,” available at the links for on-line courses at the author's homepage at: <http://www.cs.nthu.edu.tw/~jang>.
33. Jyh-Shing Roger Jang, “Speech and Audio Processing Toolbox,” available from the link at the author's homepage at: <http://www.cs.nthu.edu.tw/~jang>.
34. 陳若涵, 「以音樂內容為基礎的情緒分析與辨識」, 國立清華大學, 碩士論文, 民國九十五年
35. Yi-Hsuan Yang, Homer H Chen, *Music Emotion Recognition*, CRC Press, London, 2010

音樂情緒分析: 以數理方式分析由音樂內容喚起聽者的情緒。

曲名: **La Vie\_En\_Rose**

演唱者: **Edith Piaf**

曲風: **Bossa nova, Pop.**

When providing your ratings, please describe how the music you listen to makes you feel (e.g., this music makes me feel sad). Do not describe the music (e.g., this music is sad) or what the music may be expressive of (e.g. this music expresses sadness). Keep in mind that a piece of music can be sad or can sound sad without making you feel sad. Please rate the intensity with which you *felt* each of the following feelings on a scale ranging from 1 (*not at all*) to 5 (*very much*).

	1	2	3	4	5	1	2	3	4	5
	Not at all	Somewhat	Moderately	Quite a lot	Very Much					
1	<b>Wonder (驚異, 讚歎): Filled with wonder, Dazzled, Moved</b>									
2	<b>Transcendence(超然; 靈性): Fascinated, Overwhelmed, Feelings of spirituality</b>									
3	<b>Power(動力的; 喜悅的; 狂歡的): Strong, Triumphant, Energetic</b>									
4	<b>Tenderness(溫和; 親切): Tender, Affectionate, In love</b>									
5	<b>Nostalgia(懷舊之情): Nostalgic, Dreamy, Melancholic(憂鬱的)</b>									
6	<b>Peacefulness(平靜的; 寧靜的): Serene, Calm, Soothed</b>									
7	<b>Joyful Activation(高興的, 充滿喜悅的): Joyful, Amused, Bouncy</b>									
8	<b>Sadness(悲哀, 悲傷): Sad, Sorrowful</b>									
9	<b>Tension(激動的): Tense, Agitated, Nervous</b>									

Zentner, M., Grandjean, D., & Scherer, K. R., Geneva Emotional Music Scale (GEMS-9), Emotion, 2008

## 附錄資料 A 問卷構面