

國立交通大學

工學院聲音與音樂創意科技碩士學位學程

碩士論文

具時變情緒軌跡介面之自動音樂情緒追蹤系統

Emotion Locus Tracking System for Automatic Mood

Detection and Classification of Music Signals



研究生：傅俊傑

指導教授：鄭泗東 教授

民國九十九年七月

具時變情緒軌跡介面之自動音樂情緒追蹤系統

Emotion Locus Tracking System for Automatic Mood Detection and
Classification of Music Signals

研究生：傅俊傑

Student：Jun-Jie Fu

指導教授：鄭泗東

Advisor：Stone Cheng

國立交通大學

工學院聲音與音樂創意科技碩士學位學程

碩士論文

A Thesis

Submitted to Master Program of Sound and Music Innovative Technologies

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

College of Engineering

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年七月

具時變情緒軌跡介面之自動音樂情緒追蹤系統

摘要

以往對於情緒分類或辨識的研究往往假設整首音樂或是每個音樂段落為單一穩定情緒，並總結為此音樂帶給聽者的”心情”。本研究的所建構的系統以人的聆聽過程為概念，聽者當下的情緒反應應該是受前一段時間所聽到的音樂特徵所影響，這比較接近為一種連續變化的過程。本系統分析音樂與情緒之間的關連模型選用的是由 Thayer 所提出的情緒模型，共分為四大種類的情緒：(1)舒適、(2)哀傷、(3)焦慮、(4)振奮。在系統的訓練模式中，將以大量附有人工標記 Thayer 之情緒類別的音樂片段進行特徵萃取，主要特徵為音量、音樂事件密集度的追蹤、調性的追蹤。使用者當下可能感受到的情緒以情緒平面上以一個紅點(情緒指標)代表，由每個時刻個音樂特徵計算情緒指標於 Thayer 情緒平面上的得分與位移軌跡，最後利用軌跡座標與事先標記好的類別以 GMM 找出各個類別的邊界。在使用者模式中，則是將情緒軌跡位移展示在含邊界的情緒平面上，方便使用者了解各個情緒間的關係，並且增強使用上的經驗感受。

關鍵詞：音樂資訊檢索、音樂情緒辨識、音樂情緒追蹤、特徵萃取、GMM。

Emotion Locus Tracking System for Automatic Mood Detection and Classification of Music Signals

Abstract

As the technology of artificial intelligence and machine learning develops, people are pursuing some applications to interact with computers in a more humanized and personalized way. In the recent years, affective computing for the content-based information retrieval is a very popular research of both image and sound signals. Using emotions as an index for Music Information Retrieval is also a challenge issue for researchers. Music is always plays an important role in people's everyday life, whether they had been experienced professional music education or not, all the different kinds of music exactly could arouse and transmit the different emotional responses of human. Although the emotion is a subjective feeling, the same music might bring different emotions to different persons, but in general, there is still a trend of peoples' emotional responses. Most of the state of art research about music emotion classification or prediction assumes the music is always in a constant emotion or it is in a constant emotion at each possible part. The idea of the system proposed in this paper based on human's listening processing, the emotion response at an instant of time are primary influenced by the features that you heard in the past few seconds, and the emotion responses will be displayed by the moving dot and it's trajectory on the Thayer's emotion plane, this will enhances the listening experiences of listener, finally the trajectory could also be clues for evaluate the average emotion or mood of the music.

Keyword: Music Information Retrieval, Music Mood Recognition, Music Emotion Tracking, Feature Extraction, GMM.

誌謝

本文能順利完成首先要感謝指導教授鄭泗東老師這兩年來的細心栽培，使學程第一屆的我們能夠得到研究與實驗上傳授豐富的知識與支援，同時更分享許多人生上的經驗與道理。在這段學習過程中，不但加深了對研究領域的認知與根基，更提升了自己的想法與眼界，使我得已順利完成碩士學位。

除了老師之外，亦感謝同學于恬、雲凱以及學弟們在研究上能互相討論與協助，更要感謝過去的老師與同學們，沒有他們就不會有現在的我。

最後還要感謝一路陪伴我的父親、母親、哥哥們、姐姐們、女友家怡，不但在生活、心靈與經濟上給予鼓勵、包容與支持，並在沮喪、失落時提供了避風港，讓我可以繼續奮鬥下去。



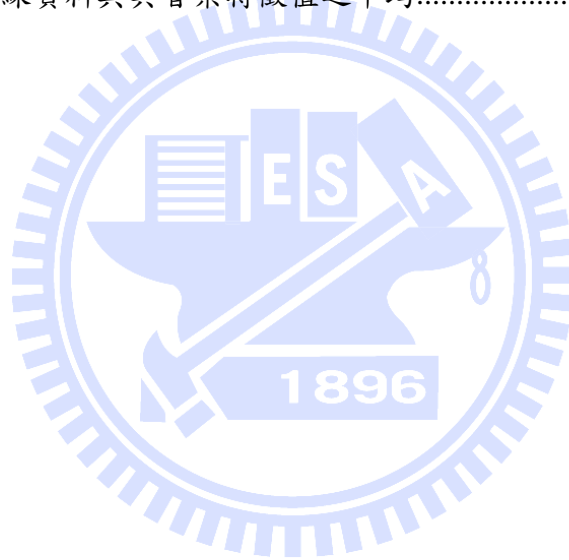
目錄

封面.....	i
摘要.....	iii
Abstract.....	iv
誌謝.....	v
目錄.....	vi
表目錄.....	viii
圖目錄.....	ix
一、緒論.....	1
1.1 研究動機.....	1
1.2 資料檢索與基於內容的音樂資料檢索.....	1
1.3 音訊特徵萃取與分類器.....	2
1.4 音樂與情緒.....	3
1.5 文獻回顧.....	4
二、基本分析方法與原理.....	5
2.1 基於音框的音訊分析.....	5
2.2 短時距頻譜.....	7
2.3 Pitch Class Profile (PCP).....	10
2.4 Gaussian Mixture Model (GMM).....	12
三、音樂與情緒的關連模型.....	16
四、系統架構.....	18
4.1 設計概念.....	18
4.2 系統架構.....	19
4.3 訓練資料.....	22
4.4 音樂事件密集度.....	22
4.5 音量大小.....	28
4.6 音色分析.....	30
4.7 調性分析與和聲和諧程度.....	32
4.8 計分方法.....	37
4.9 情緒邊界的訓練.....	38
五、計算結果分析.....	45
5.1 訓練資料之特徵萃取數據分析.....	45
5.2 系統展示.....	54
六、結論與未來展望.....	59
七、參考文獻.....	60
附錄 A- 訓練資料明細表.....	63
附錄 B- 不同高斯函數個數的 GMM 訓練結果.....	69



表目錄

表 2.2.1 頻譜內容分析方法.....	9
表 4.3.1 各個情緒類別的訓練音樂數量.....	22
表 4.7.1 八度內的音程表.....	36
表 4.8.1 音樂特徵值與 Thayer 情緒模型的對應關係.....	38
表 5.1.1 訓練資料之平均音樂事件密集度.....	45
表 5.1.2 訓練資料之平均對數音量.....	46
表 5.1.3 訓練資料之音色平均值.....	48
表 5.1.4 訓練資料之調性追蹤結果.....	50
表 5.1.5 訓練資料之和諧程度分析結果.....	51
表 5.1.6 各類別訓練資料與其音樂特徵值之平均.....	53



圖目錄

圖 2.1.1	音訊的時域波形.....	5
圖 2.1.2	音訊之音框化分析示意圖.....	6
圖 2.1.3	音框示意圖.....	6
圖 2.2.1	單一音框的頻譜圖.....	7
圖 2.2.2	連續時間的頻譜圖.....	8
圖 2.2.3	Hamming Window.....	8
圖 2.3.1	單一音框的音調層級強度分佈圖.....	11
圖 2.3.2	連續時間的音調層級強度分佈圖.....	11
圖 2.4.1	高斯分部.....	12
圖 2.4.2	混和高斯分部.....	14
圖 3.1.1	Hevner 的音樂情緒模型.....	16
圖 3.1.2	Thayer 的情緒模型.....	17
圖 4.1.1	系統介面示意圖.....	19
圖 4.2.1	系統流程方塊圖.....	20
圖 4.4.1	以音訊頻譜流量進行音樂事件偵測.....	24
圖 4.4.2	頻譜流量音樂事件偵測結果.....	24
圖 4.4.3	音樂事件與節奏關係示意圖.....	25
圖 4.4.4	音樂事件的權重分佈.....	26
圖 4.4.5	貝多芬之月光奏鳴曲-音樂事件密集度.....	27
圖 4.4.6	Muse 之 Stockholm Syndrome-音樂事件密集度變化.....	28
圖 4.5.1	貝多芬之月光奏鳴曲-音量變化.....	29
圖 4.5.2	Muse 之 Stockholm Syndrome-音量變化.....	30
圖 4.6.1	貝多芬之月光奏鳴曲-頻譜質心和頻譜帶寬.....	31
圖 4.6.2	Muse 之 Stockholm Syndrome-頻譜質心和頻譜帶寬.....	32
圖 4.7.1	C 大調調性樣板.....	33
圖 4.7.2	C 小調調性樣板.....	33
圖 4.8.1	計分流程圖.....	35
圖 4.9.1	貝多芬之月光奏鳴曲-情緒軌跡位移.....	38
圖 4.9.2	貝多芬之月光奏鳴曲-情緒軌跡位移所提供的訓練資料.....	39
圖 4.9.3	訓練資料分佈.....	40
圖 4.9.4	類別 1 的高斯分佈.....	41
圖 4.9.5	類別 2 的高斯分佈.....	41
圖 4.9.6	類別 3 的高斯分佈.....	42
圖 4.9.7	類別 4 的高斯分佈.....	42

圖 4.9.8	GMM 分類結果與各類別的邊界範圍.....	43
圖 4.9.9	內部測試辨識率與高斯函數個數的關係.....	44
圖 5.2.1	Zombie-時域波形與樂器編制.....	54
圖 5.2.2	Zombie-第 1 段末之情緒軌跡變化.....	55
圖 5.2.3	Zombie-第 2 段末之情緒軌跡變化.....	56
圖 5.2.4	Zombie-第 3 段末之情緒軌跡變化.....	56
圖 5.2.5	Zombie-第 4 段末之情緒軌跡變化.....	57
圖 5.2.6	Zombie-第 5 段末之情緒軌跡變化.....	57
圖 5.2.7	Zombie-第 6 段末之情緒軌跡變化.....	58



一、緒論

1.1 研究動機

隨著人工智慧和機器學習的技術發展，人們正在追求一種更人性化與個人化的方式來和電腦互動。近年來針對圖像和聲音訊號內容之資訊檢索的情感運算為相當熱門的研究，對於許多學者而言，使用情緒作為音樂資料檢索指標也是一項相當有挑戰性的議題。音樂一直在人們生活中扮演相當重要的角色，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，無論是經由外界各種媒體播放，或是經由哼唱，更甚至是想像，音樂影響了每個人的情緒感受，諸如聽到抒情歌曲會感動、聽到搖滾歌曲會感到激昂...等，如何利用音樂對於人類的情緒影響來辨識與檢索音訊，為本研究的重點。雖然情緒為相當主觀的一種檢索指標，針對不同音樂的情緒也會因人而異，但總結來說還是會有一個大概的趨勢，如大調音樂往往讓人感到快樂與振奮，小調音樂往往讓人感到哀傷和詼諧，大部分的人都可以認同這樣的分類結果。以往對於音樂情緒分類與辨識的相關研究皆假設音樂片段為單一穩定的情緒，實際上人們在聆聽音樂時的情緒變化比較接近一個連續變化的過程，本文嚐試以這樣的概念出發設計一個連續的情緒計分模型，並且將單一和連續變化的情緒之間的關係提出了一種分析的方法。對於音樂訊號的情緒分析，目標是模擬並預測聽者的反應，藉此達到許多人性化的應用以增加生活的便利性，如智慧型手機、PDA 或是 Mp4 Player 等消費性電子的創意互動應用程式或功能。除了商業價值，此項研究也有助於進一步了解音樂本質，實現人工智能對於聽覺的感知能力的模擬或學習，有助於人工智能的發展。

1.2 資料檢索與基於內容的音樂資料檢索

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。如何使電腦有效率且人性化的管理音樂資料庫，此為現今音樂資訊檢索(Music Information Retrieval)的主要研究方向。對於音樂資料庫的應用與管理，以往最基本的音樂資訊檢索例如以文字資訊做為檢索指標，諸如作者，曲目，專輯名稱，音樂類型…等，檔案經由人工的標記即可利用這些資訊對音樂資料庫進行搜尋，並依照使用者需求找出對應的音樂檔案。近期的文獻中，已經越來越多相關研究嘗試以不同的資訊做為檢索指標，其中以基於內容的音樂資訊檢索(Content-based Music Information Retrieval)為多數學者最主要的研究方向。

基於內容的音樂資訊檢索，為利用音樂訊號本身的特徵(features)來對音樂檔案進行檢索的方式。所謂特徵就是指聲音訊號行為模式的一種表現方式，常見的特徵就如一般聽者在聆聽音樂時的感受，例如：音量大小、節奏速度與模式、音色、和聲與旋律…等。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，若能對這些音樂感知上的特徵數值加以檢索，則音訊檔案管理上則相當的多元化且便利，這也是基於內容的音樂資訊檢索目前正被積極研究的原因，其對於電腦音樂檔案可達到人性化的應用或分析方式，例如在廣播的錄音中取出播放音樂的部分、選取特定的音樂檔案的段落進行編輯或剪接、互動式的音樂推薦系統、由哼唱方式搜尋音樂檔案或段落…等。

1.3 音訊特徵萃取與分類器

聲音訊號中存在許多不同的特徵可以描述訊號的狀態或行為，而特徵在時間分佈上有尺度的不同，小的尺度就是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)[1]...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，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。對於特徵萃取後得到的數值，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。比起直接依照經驗法則或實驗結果定義臨界值(threshold value)，另一個更好的方法為利用分類器(Classifier)來找出類別的邊界。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，且對於大部分的情況分類器皆能夠有不錯的分類效能。

1.4 音樂與情緒

屏除特定音樂和個人特殊經驗的連結或是歌詞含意造成情緒上的引導，一般人對於音樂的聆聽感受中最重要的一個關鍵元素就是節奏(rhythm)，旋律(melody)，和聲(harmonic)，音色(timbre)。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，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，由不和諧進行到和諧的聲響在音樂上稱為解決(resolution)，聽者感受會像是得到壓力的紓解。和聲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，組成音彼此頻率比值稱音程(interval)，其不同音程的組合造就不同的和弦的聲響，如大和弦(major chord)給人快樂的感覺，小和弦(minor chord)給人哀愁的感覺。音色的物理意義即是聲音的產生模式，不同的聲源的特殊發聲機制造就每種聲音的獨特性，也是讓人耳分辨各種聲音的不同的主要依據，同時音色也有和諧與不和諧的差異，反映在聲音

頻域的泛音表現，泛音的存在使單獨的聲音猶如一個小的和聲系統，不同頻率間強度比值則會影響其和諧程度。各種音樂特徵與情緒的互相關聯，已有不少心理學家提出不同的模型，並以聽者實驗驗證。將這些模型中較為抽象的維度和量化的特徵值做連結，即可以將其應用於分類器，詳細的音樂情緒關聯模型將在第三章節討論。

1.5 文獻回顧

現今音樂資訊檢索的研究已有不少研究文獻，主要可分為幾個項目：

- (1) 特徵萃取與辨識。著重於新的音訊特徵的探索，或是對於舊有的特徵萃取方法提出改善方式。如音樂調性的辨識[2]，文中提到利用分類器 Hidden Markov Model (HMM) 對古典鋼琴音樂訊號作調性轉換的偵測，用以辨識輸入音樂的調性並探測是否有轉調發生。音樂和弦的辨識[3]，文中提到利用 HMM 來對音樂訊號的和弦做分析與辨識。對於聲音訊號的分割。
- (2) 針對特定目標之分類法的架構與設計。文獻[4]說明基於 GMM 所架構的音樂情緒辨識系統，介紹基本的特徵萃取方法及分析方法。文獻[5]說明利用回歸的分析方式進行音樂情緒辨識。
- (3) 音樂結構的分析。文獻[6]中說明利用音訊中不同音框間相似的程度，來分析音樂段落的結構，並且進一步利用這樣的結果發展出一種音訊分割的方法，利用音訊自我相似(Self Similarity)程度的變化來做音訊分割[7]，最後更利用其來做音訊節奏模式與節奏速度的分析[8]。
- (4) 辨識系統介面的設計。文獻[9]說明其所設計的音樂情緒辨識介面 Mr. Emo，主要以音樂之情緒平面(Emotion plane)的方式呈現，並有多種模式可以與使用者互動之播放系統，如依照使用者在情緒平面上的滑鼠軌跡選擇該軌跡鄰近的音樂播放。文獻[10]說明其架構的線上即時音樂情緒辨識系統 Mood Cloud，以 SVM 預測聽者當下的音樂情緒反應並以條狀圖表示。

二、基本分析方法與原理

2.1 基於音框的音訊分析

第一章節中提到過，聲音的特徵值在時間尺度上有不同的大小，小的特徵在幾毫秒內就可以計算出來，大的特徵通常可由小的特徵的變化或統計來取得，一般人所感知到的音樂特徵屬於大尺度的範圍。關於如何進行特徵值萃取？聲音的本質為空氣的壓力變化所產生的波動，此壓力稱為聲壓，其所產生的波則稱為聲波。實際上聲波為一種類比訊號，是連續不間斷的訊號，利用電聲換能器之偵測則可以將聲波的壓力變化作取樣，改以電壓變化的樣本點記錄於電腦中，此時也會轉換成為一種離散的數位訊號。聲壓隨時間的變化趨勢稱為此音訊的波形，如圖 2.1.1，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)，取樣頻率越高則訊號越接近原始的類比訊號。

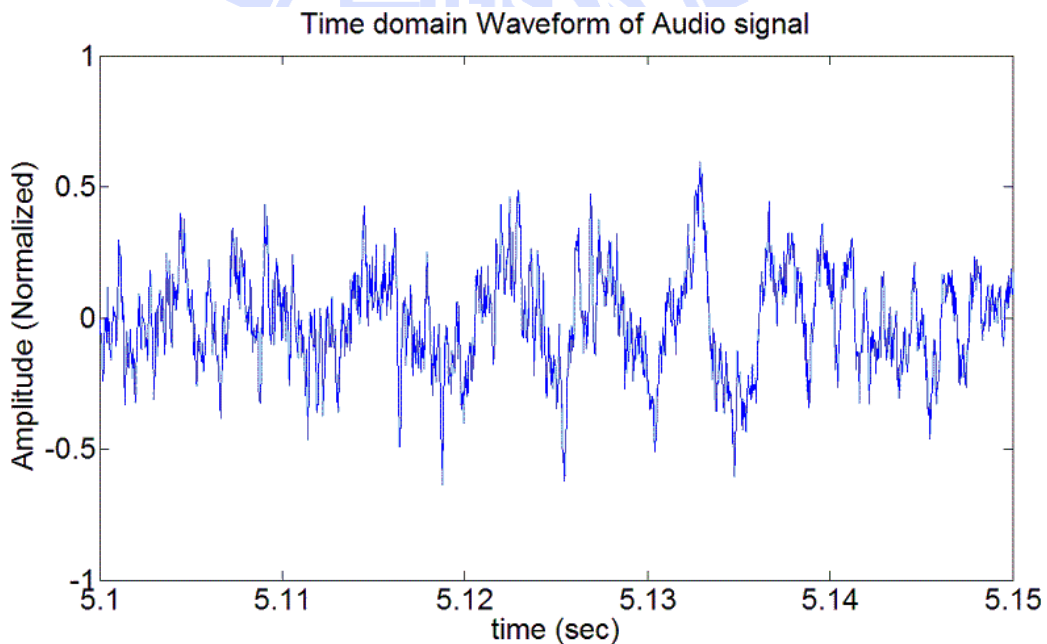


圖 2.1.1 音訊的時域波形

由於音訊的特徵多為隨時間變化的數值，對於音訊分析現在多半採用音框 (frame) 化的分析方式，以一小群鄰近的樣本點代表某一時刻訊號的波形狀態，稱為一個音框，如圖 2.1.2，圖 2.1.3。

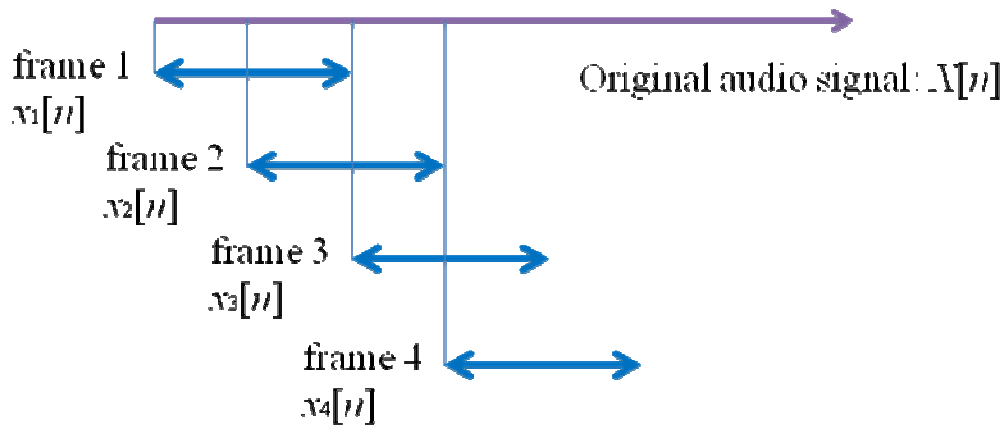


圖 2.1.2 音訊之音框化分析示意圖

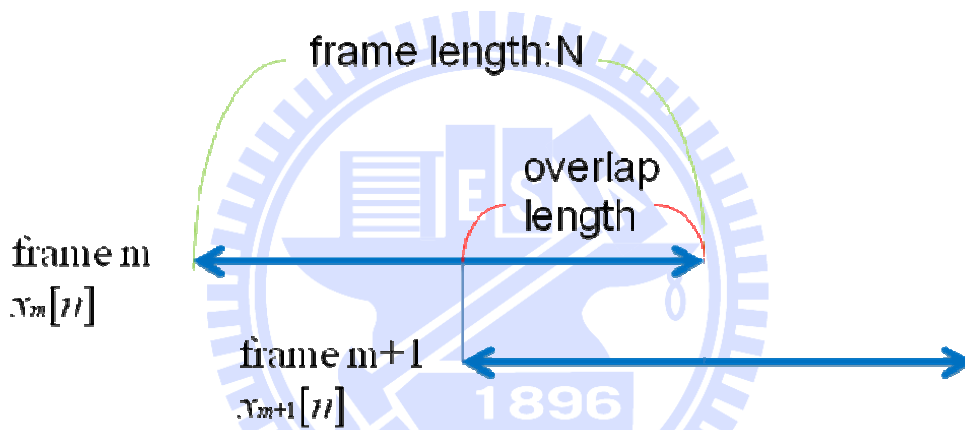


圖 2.1.3 音框示意圖

$X[n]$ 為原始訊號， $x_m[n]$ 為的 m 個音框的訊號， N 為音框樣本點數目或稱音框長度， n 為音框時域樣本點的索引。若定義已音框中心樣本點的時間位置為該音框的時間位置，則音框化的分析方式使得音訊在每個時刻皆可以以一個音框來表示訊號的瞬間狀態，如此一來便可以對音訊進行短時距的分析。對於聲音訊號的分析，通常可以分為時域(time domain)與頻域(frequency domain)的方法，時域方法如音訊波形的震幅與其對應的能量，頻域方法則為音訊波形的頻率成分分佈情形。

2.2 短時距頻譜

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的離散傅立葉轉換(Discrete Fourier transform)，數學定義如下：

$$S_m[k] = \sum_{n=0}^{N-1} x_m[n] \times w[n] \times e^{-j\left[\frac{2\pi k}{N}\right]n} \quad (1)$$

$$f[k] = \frac{f_s}{N} \times k \quad (2)$$

其中 m 為音框數的索引， k 為音框頻域樣本點的索引， $S_m[k]$ 代表第 m 個音框的其對應於頻率 $f[k]$ 的頻譜強度， $w[n]$ 即為每個音框樣本點的對應權重或稱為窗函數(window function)。 $f[k]$ 為音框頻域樣本點所對應的實際頻率值， f_s 為訊號的取樣頻率。單一音框的頻譜的圖形如圖 2.2.1，可以看到音訊於各個頻率的強度大小與分布。連續時間的頻譜圖如圖 2.2.2，可以看到不同時間音訊於各個頻率的強度大小與分布。

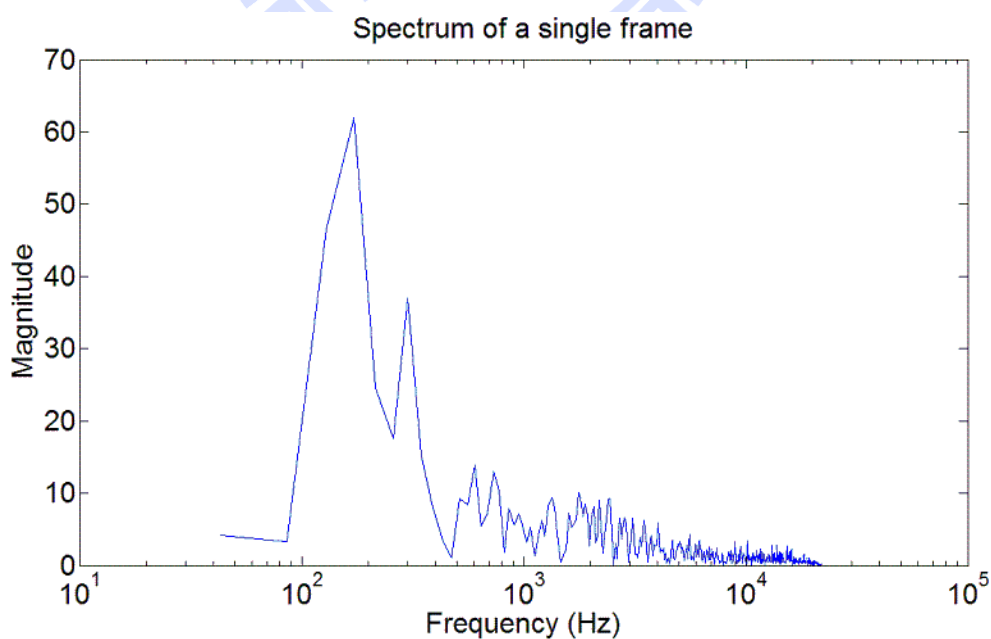


圖 2.2.1 單一音框的頻譜圖

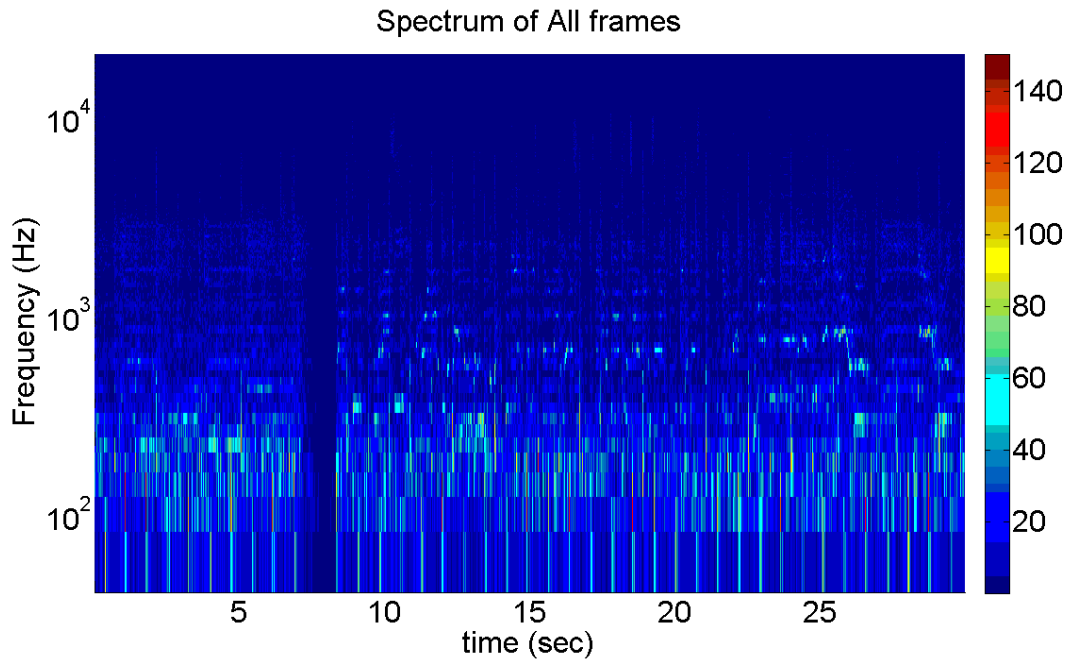


圖 2.2.2 連續時間的頻譜圖

窗函數的用意為改善音框訊號在計算頻譜時的邊界效應，使頻譜的數值對比更好，常用的窗函數為漢明窗(Hamming window)，圖形如圖 2.2.3，數學定義如下。

$$w[n] = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$

$n=0,1,2,\dots,N-1$

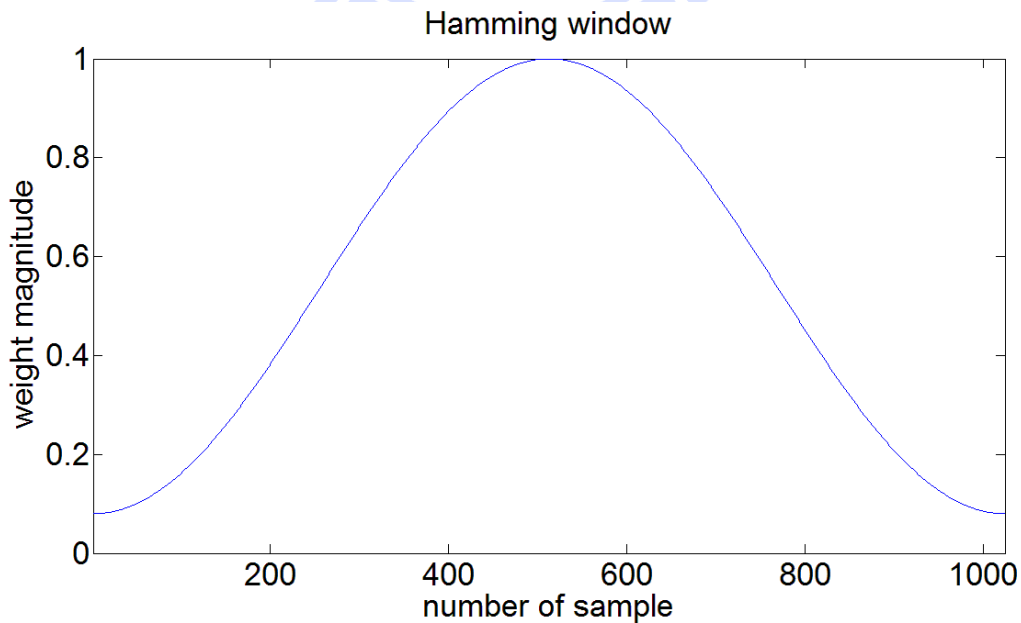


圖 2.2.3 Hamming Window

頻譜的內容和聲音訊號的音色有密切關係，包含聲音訊號的基頻、泛音成分、音高的清晰程度...等，反映在頻譜中各個頻率的強度分布情形。文獻[4]中整理了針對頻譜內容分析的方法如表 2.2.1。

表 2.2.1 頻譜內容分析方法

Spectral Shape Features	Brightness	Centroid of the short-time Fourier amplitude spectrum.
	Bandwidth	Amplitude weighted average of the differences between the spectral components and the centroid.
	Roll off	95th percentile of the spectral distribution.
	Spectral Flux	2-Norm distance of the frame-to-frame spectral amplitude difference.
Spectral Contrast Features	Sub-band Peak	Average of a percent of the largest amplitude values in the spectrum of each sub-band.
	Sub-band Valley	Average of a percent of the lowest amplitude values in the spectrum of each sub-band.
	Sub-band Contrast	The difference between the Peak and Valley in each sub-band.

由表 2.2.1 可以了解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頻譜的分布情形，頻譜對比則是利用八度音(倍頻)的概念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。一種常用的切割方式是由八度做為基準，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，且人耳的頻率響應範圍大約為 20Hz~20kHz，包含了約十個八度，故以八度為單位對於此特定倍數頻率成分的分析較為方便。就音色而言，聲音在不同頻率區塊都有不同的音響效果，高頻的大小代表聲音的明亮程度，低頻則通常代表音訊的蘊含能量、震撼力和飽和感，典型代表如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。

2.3 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)[2]，如下：

$$P(k) = [24 \times \log_2((f_s / N) \times k / f_1)] \bmod 24 \quad (4)$$

$$\text{PCP}[P(k), n] = \sum_{P(k)} |S[k, n]| \quad (5)$$

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， $\text{PCP}[(P(k)=0, \dots, 23), n]$ 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。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 2.3.1，各個時間的音調層級的強度分佈如圖 2.3.2。

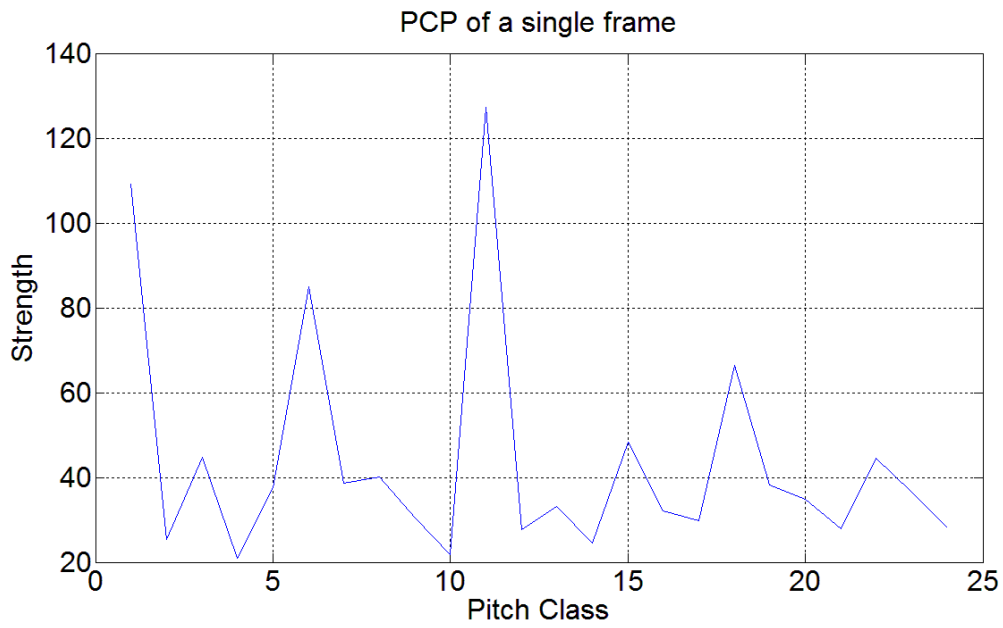


圖 2.3.1 單一音框的音調層級強度分佈圖

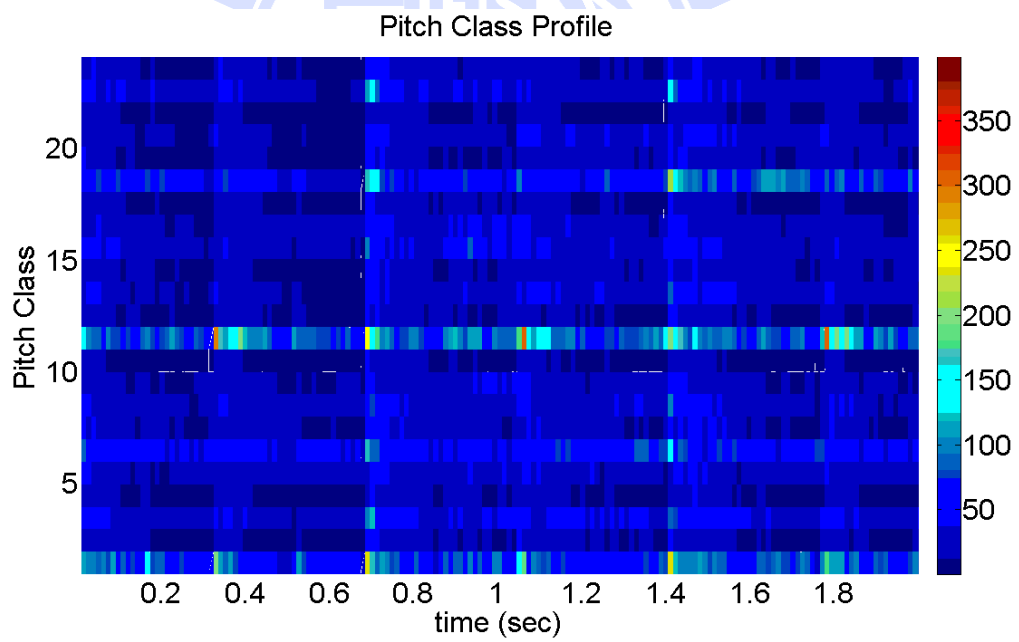


圖 2.3.2 連續時間的音調層級強度分佈圖

2.4 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：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2((x-\mu)/\sigma)^2} \quad (6)$$

其分佈圖形如圖 2.4.1：

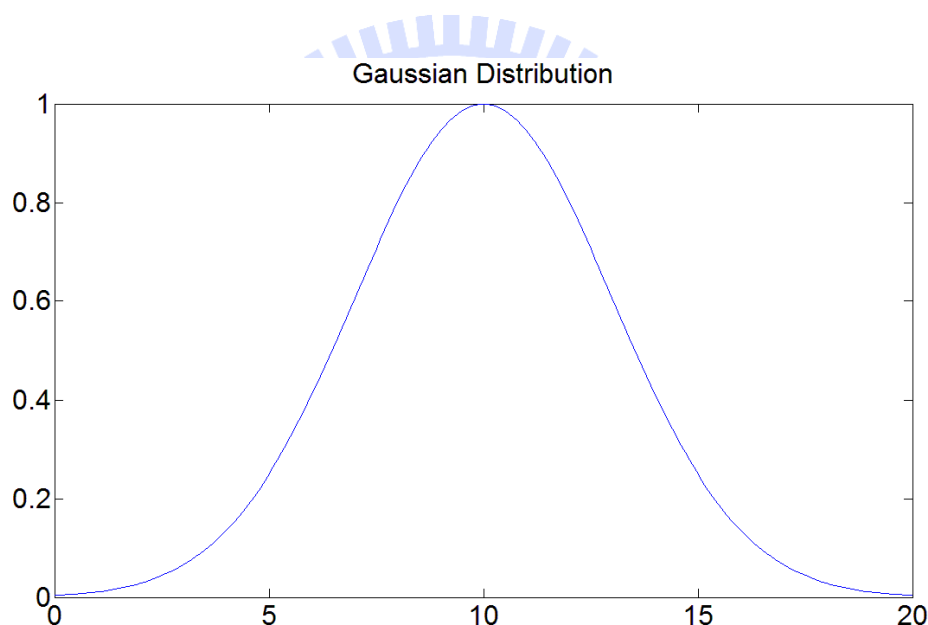


圖 2.4.1 高斯分部

其中有兩個重要的參數， μ 和 σ ， μ 為期望值(Expectation value)，其為平均數，位於高斯分佈的中央， σ^2 稱為變異數(Variance)，而 σ 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：

$$\mu \equiv E[x] = \int_{-\infty}^{\infty} xp(x) dx \quad (7)$$

$$\sigma^2 \equiv E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx \quad (8)$$

利用向量和矩陣推廣為高維度表示如下：

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right] \quad (9)$$

其中 $\boldsymbol{\mu}$ 和 Σ 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 μ 和 σ 的值會和其分佈的中心位置和曲線寬度有關。

$$\boldsymbol{\mu} \equiv E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_d] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = \sum_{\mathbf{x}} \mathbf{x} P(\mathbf{x}) \quad (10)$$

$$\Sigma = \begin{bmatrix} E[(x_1 - \mu_1)(x_1 - \mu_1)] & E[(x_1 - \mu_1)(x_2 - \mu_2)] & \cdots & E[(x_1 - \mu_1)(x_d - \mu_d)] \\ E[(x_2 - \mu_2)(x_1 - \mu_1)] & E[(x_2 - \mu_2)(x_2 - \mu_2)] & \cdots & E[(x_2 - \mu_2)(x_d - \mu_d)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_d - \mu_d)(x_1 - \mu_1)] & E[(x_d - \mu_d)(x_2 - \mu_2)] & \cdots & E[(x_d - \mu_d)(x_d - \mu_d)] \end{bmatrix} \quad (11)$$

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，數學表示如第(12)式，圖形如圖 2.4.2。

$$p(x, y) = \alpha_1 g(x, y; \mu_1, \Sigma_1) + \alpha_2 g(x, y; \mu_2, \Sigma_2) + \alpha_3 g(x, y; \mu_3, \Sigma_3) \quad (12)$$

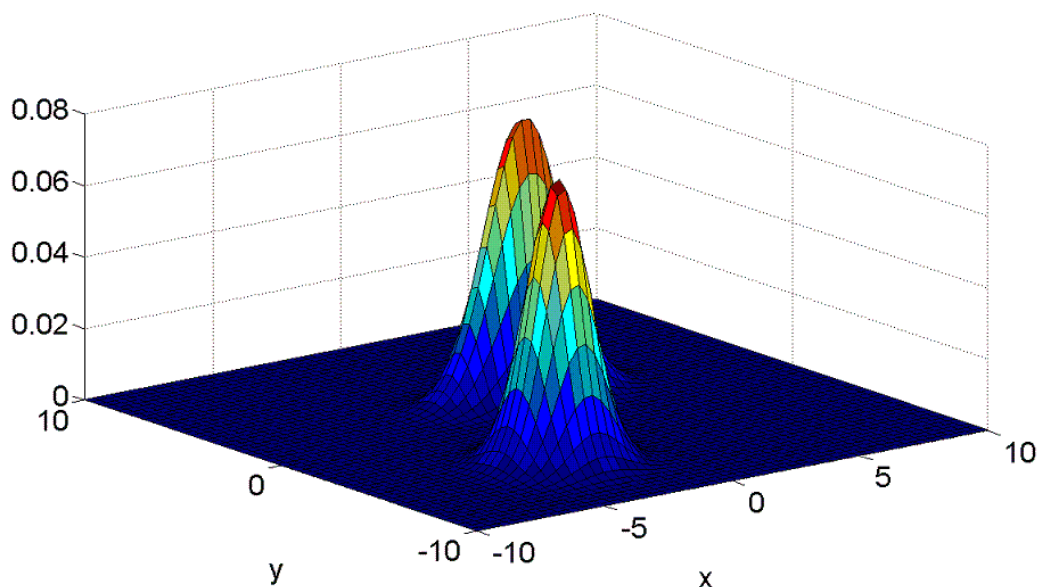


圖 2.4.2 混和高斯分部

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。



三、音樂與情緒的關連模型

關於音樂與情緒的互相聯結，有部分心理學文獻針對其作出討論並提出相關模型，較常被引用的文獻如[11]、[12]、[13]、[14]、[15]。一般而言在音樂與情緒的模型中，抽象的情緒必須和量化的特徵強度有明顯對應的相互關係，如此才可以應用於音樂情緒辨識的系統中。對於種類較多的情緒分類，較無法找出各個種類或形容詞和特徵強度的相互關係，應用上較不方便，如文獻[11]為 Hevner 提出的音樂情緒模型，其為針對音樂與情緒的第一篇相關研究文獻，模型圖可參考圖 3.1.1。共分為八大種類的情緒，各個種類又細分幾種不同的形容詞組，不同形容詞組間的相互關係較不明顯。

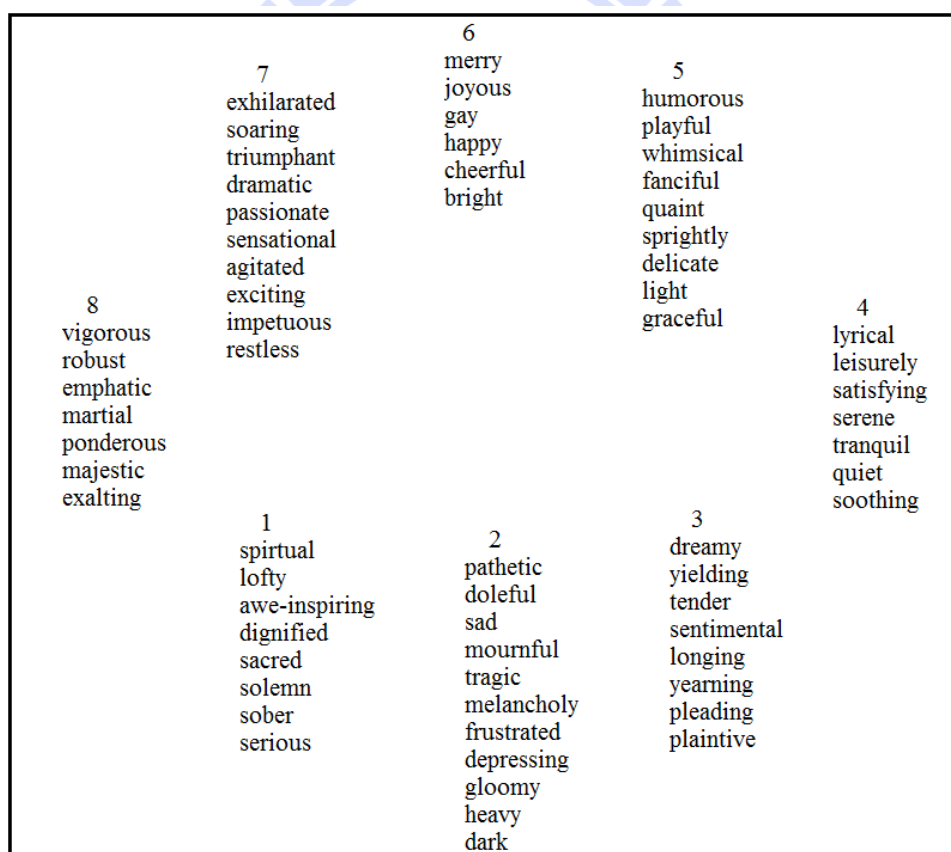


圖 3.1.1 Hevner 的音樂情緒模型

文獻[13]說明由 Thayer 提出的情緒模型，如圖 3.1.2，此模型分為四個象限，分別為(1)激勵人心的，(2)舒適的，(3)焦躁憤怒的，(4)使人消沉的。橫軸的定義為壓力，縱軸的定義為能量。類別只有四種但是對於普遍的情緒反應皆能做出概要的區分。應用於音樂情緒的系統中也相當合適，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，一些簡單的音樂特徵如：節奏、音色、音量，不難發現這些特徵強度和該模型中的橫軸縱軸有相互的關係。如：快速的節奏和較大的音量通常代表音樂的能量較高，慢速的節奏和不和諧的音色則會帶給聽者壓力。

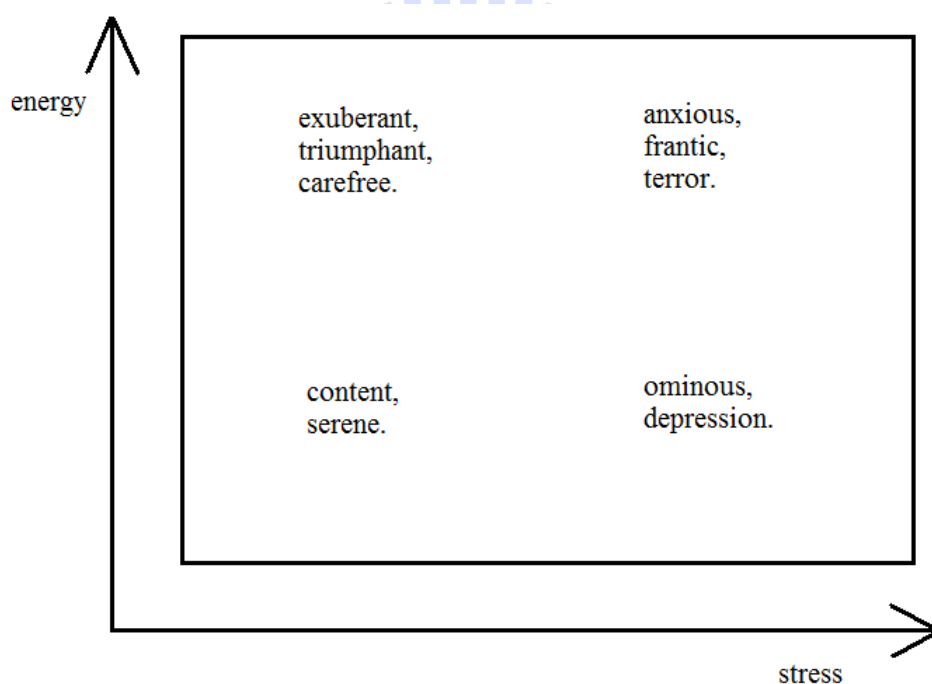


圖 3.1.2 Thayer 的情緒模型

四、系統架構

4.1 設計概念

以往音樂情緒辨識的研究，多半是對於整段音樂訊號做出一個概要性的分類並給予單一標記，如文獻[1]、[4]、[5]、[9]、[16]。其中文獻[4]也有進行情緒追蹤的相關研究，文中假設古典音樂中當音量強度有較大的改變時通常也會對應到情緒的轉折，利用這樣的經驗法則先找出可能的情緒變化段落位置，再假設每一個段落為穩定的情緒並且進行音樂情緒辨識。文獻[10]提出以 SVM 預測聽者當下的情緒反應並以機率的條狀圖表示，不同情緒對應到不同的顏色以增強使用者的使用經驗。有別於前述的相關文獻，本系統以情緒平面上連續的軌跡變化來代表預測的聽者情緒反應，連續的變化過程將更能貼近使用者的情緒反應，而以情緒平面作為預測介面的優點為使用者可以更清楚了解各種不同情緒的相對關係。

考量人類的感知系統，除了聽完音樂後的概要結論，其在聆聽音樂的當下即會有一些相對的情緒反應產生，這比較接近為一個連續變化的過程，聽者的情緒隨著音樂的進行而有相對的改變。為了模擬聽覺感知系統並加強聽者的感受，則需要建構一個連續變化的追蹤系統介面，將隨時間變化的音框訊號進行特徵萃取，計算各個音框的特徵內容是否會造成特殊情緒的累積，如不和諧的音色會造成愉悅感的減低，節奏的速度加快通常會使聽者感到興奮...等。連續的變化的計分過程只考慮當下往前一段時間內所累積的情緒分數，如同一般聽者在聆聽音樂時，某一時刻的情緒通常為離該時刻最接近的一段時間內之音樂內容所影響。若將此情緒變化的軌跡於二維的情緒平面中顯示，則如圖 4.1.1。

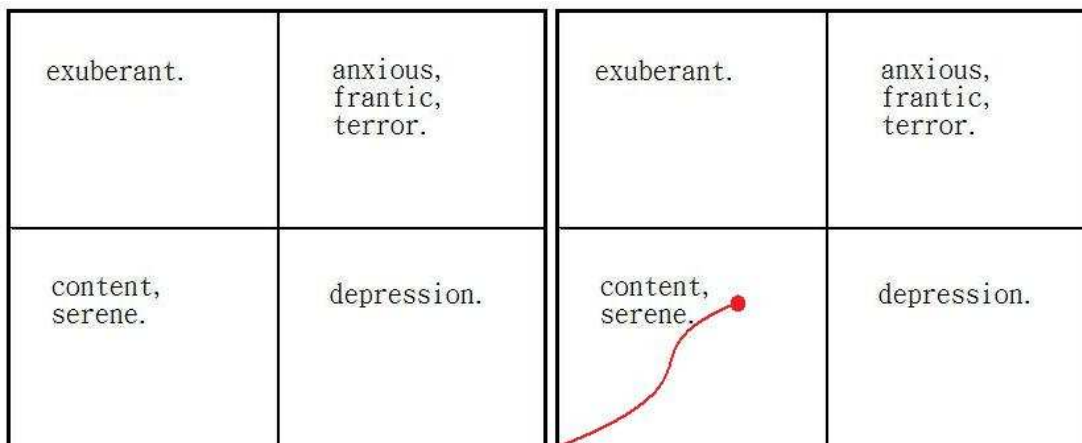


圖 4.1.1 系統介面示意圖

以 Thayer 的二維度情緒模型為例，以點狀指標代表聽者當下的情緒於情緒平面的位置，考慮一開始聆聽音樂並無特殊的情緒，所以聽者的情緒指標初始位置為情緒平面的左下角位置。隨著音樂的進行，聽者受到音樂影響而造成情緒指標的移動，如圖 4.1.1 紅色線段所示。

此外；由於音樂與情緒的相互關係相當複雜，而本研究針對單純音樂內容作為分析對象，因此有基本的假設與前提必須說明：(1) 本研究忽略歌詞的影響，聽者測試只考慮音樂性特徵對聽者的影響。(2) 心理情緒也常和個人經歷和記憶有關，如聽到 Celine Dion 的 My Heart Will Go On 會讓大多數人想到鐵達尼號的電影劇情而對情緒產生了音樂內容以外的影響，所以聽者測試時必須排除這方面的可能，可以利用簡單的問卷盡量使聽者和有相關經驗的音樂分離。

4.2 系統架構

本文系統使用的設計程式為 MATLAB，系統的詳細架構主要分為下列幾個部分，方塊流程圖如圖 4.2.1。

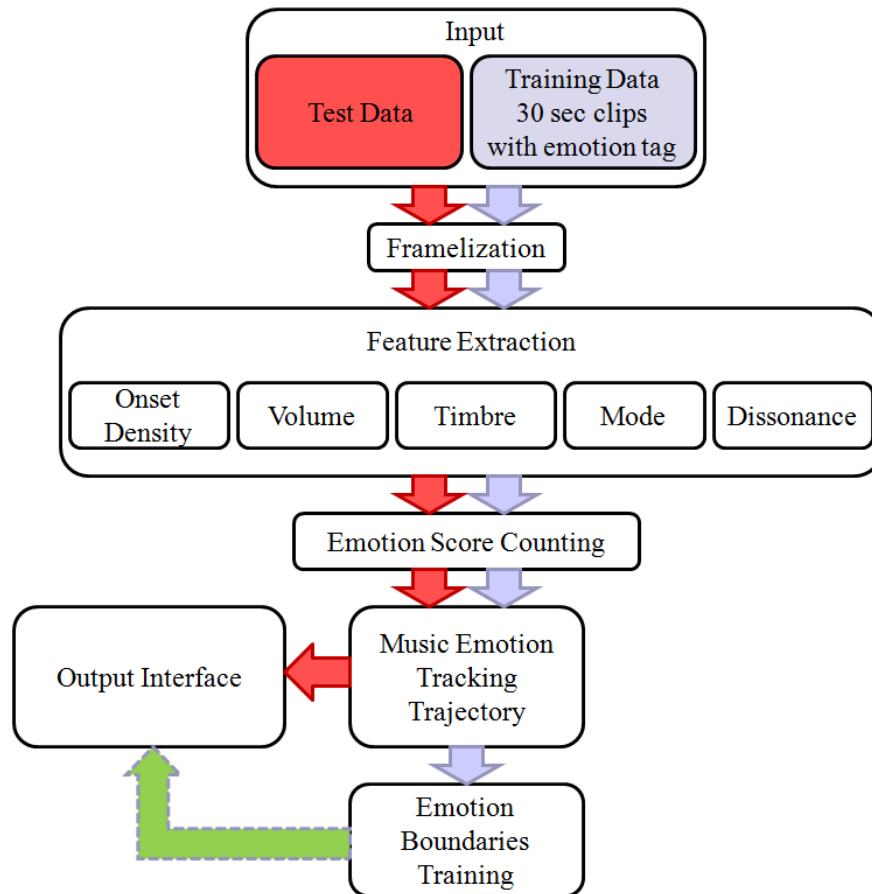


圖 4.2.1 系統流程方塊圖

系統架構主要可以分為訓練模式與使用者模式，其流程分別以淡藍色和紅色表示，詳細流程說明如下：

(1) 音訊輸入

訓練模式中，訓練音訊的規格採用 wave 格式之音樂檔案，取樣頻率 14700Hz，取樣解析度 16-bits，單聲道，作為基本的系統訓練與測試，採用 30 秒為片段的音樂並且附有單一情緒類別的標記，詳細說明請參考 4.3 節。使用者模式中音訊檔案格式相同，長度視檔案而定。

(2) 音框化

如同 2.1 節所述，為了要對音訊做進一步的頻譜分析，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

在特徵萃取的部分，主要可以分為五種特徵的萃取(i)音樂事件密集度、(ii)音量、(iii)音色、(iv)調性、(v)和聲和諧程度。選擇這些特徵的原因為：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。調性通常有一個趨勢，大調音樂使人感受愉悅，小調音樂則叫哀傷與詼諧。如同系統設計概念所述，特徵的計算為各個時間前一段時間內發生的音樂事件所對應的音樂事件密集度、音量、音色、調性與和聲和諧度，詳細方法於 4.4、4.5、4.6、4.7 節討論。

(4) 計算情緒分數增減

本文之音樂情緒的關聯模型選用 Thayer 的情緒模型，類別較少故在計分上較不會有太多模稜兩可的狀況發生。在計分的規則上，仿造一般人聽音樂的過程，當下的情緒為前一段時間內所聽到的音樂內容的累積結果，不同的特徵對應到不同的情緒分數的累加，關於詳細的計分方法將於 4.8 節中討論。

(5) 輸出情緒軌跡圖

最後的情緒分數統計結果將會輸出到情緒平面，以紅點位置代表目前聽者的情緒，並隨著音樂進行展示出聽者情緒位置的改變。

(6) 情緒平面邊界的訓練

由於在情緒模型中，各個情緒並沒有詳細的定義出該情緒的邊界，訓練模式利用分類器的概念將情緒軌跡圖所產生的座標加以分類，如此一來可以定義出不同情緒的邊界。使用者模式中情緒軌跡將會顯示於有邊界關係的情緒平面上，以利使用者了解音樂情緒的變化，往後也可以再更進一步利用情緒軌跡來辨識音樂中各個段落的概要情緒，作出簡易概要的分析，詳細方法於 4.9 節討論。

4.3 訓練資料

本次的訓練資料總共有 192 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，類型包含數種流行音樂、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，是由數個聽音測試人員將其標記給予單一的音樂情緒類別，類型即 Thayer 模型中的四種類別：(1)舒適的、(2)哀傷的、(3)焦慮的、(4)振奮的。各個情緒類別的音樂片段數量可以參考表 4.3.1，詳細的音樂片段明細可以參附錄 A。其中有部分訓練音樂有重複，但其為同一首音樂中較為不同的片段。

表 4.3.1 各個情緒類別的訓練音樂數量

情緒類別	舒適	哀傷	焦慮	振奮
音樂數量	47	48	48	49

4.4 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[17]：

1. 音樂事件的端點偵測 (Onset detection)。
2. 分析端點的位置規律計算節奏資訊。

關於訊號中的端點偵測，有許多不同的方法，但由計算方式區分，主要可以歸類為時域方法和頻域方法。時域方法如振幅或音量的變化[18]。頻域方法如：文獻[19]中說明以不同頻帶的能量變化來進行端點偵測，文獻[20]、[21]則利用相位變化。在端點偵測的任務中，因為時域的方法無法準確分析音高或音色改變的音樂事件端點，近期的文獻中多半都使用頻域的分析方法([20]、[21]、[22]、[24])。

本文針對音樂事件以頻譜的變化來進行探測，接下來將探討如何以頻譜流量 (Spectrum Flux, [28]) 進行端點偵測。

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第二章節中提到的短時距傅立葉轉換來計算，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(13)式：

$$\text{Spectrum Flux } (m) = \sum_{k=0}^N H(|S_m[k]| - |S_{m-1}[k]|)$$
$$\text{where } H(x) = \frac{x + |x|}{2} \quad (13)$$

(13)式中上式為音框 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。實際利用頻譜流量偵測音樂事件偵測的結果如圖 4.4.1：

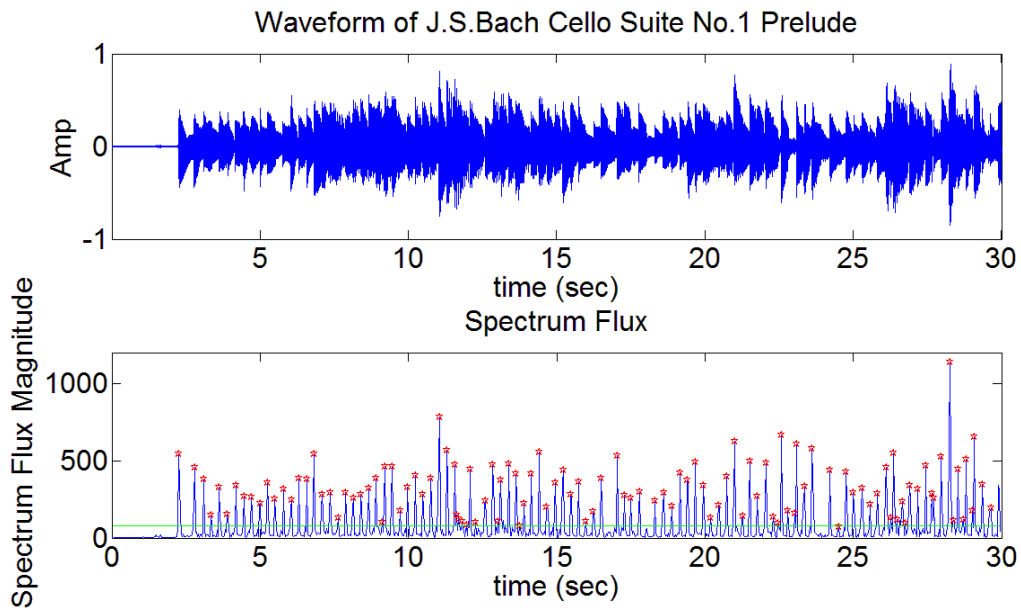


圖 4.4.1 以音訊頻譜流量進行音樂事件偵測

圖 4.4.1 為 J.S. Bach Cello Suite No.1 Prelude 古典吉他演奏版本的音樂事件偵測，上方為 0~30 秒的時域波形，下方則是以公式(13)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。實際與人工標記的音樂事件比較，結果可以參考圖 4.4.2。

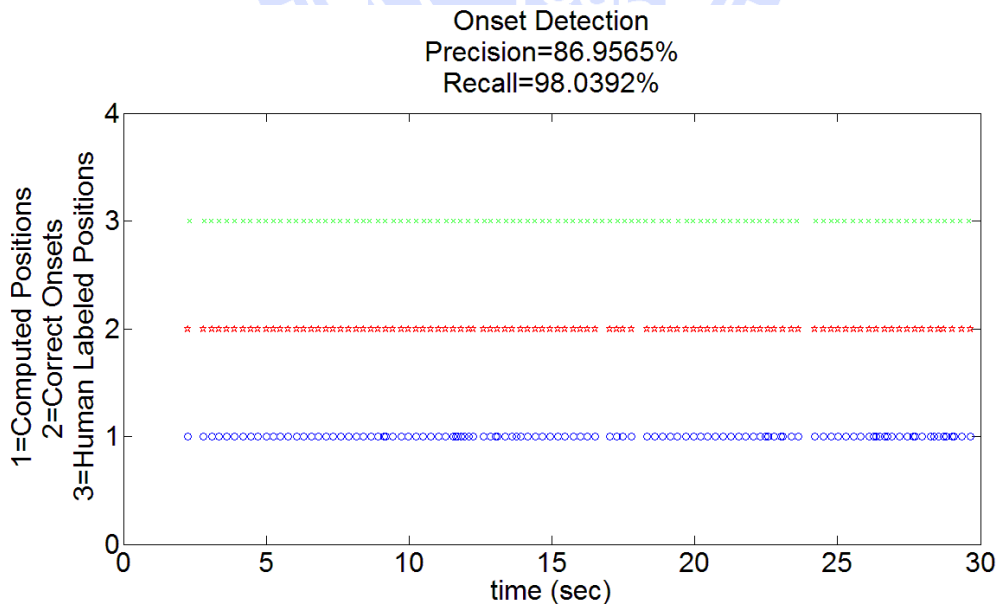


圖 4.4.2 頻譜流量音樂事件偵測結果

圖 4.4.2 中縱軸為 1 的是系統找出的音域事件位置；縱軸為 2 的是系統找出的音樂事件和人工標記的音樂事件的交集，也就是實際被找出的正確音樂事件；縱軸為 3 的是人工標記的音樂事件位置。可以看到大部分的音樂事件位置都可以被準確的找出，達到查準率(Precision)86.96%，查全率(Recall)98.03%，關於查準率與查全率的定義請參考附錄-C。

目前節奏萃取又分為節奏標記(tempo induction)([23]、[24])和節奏追蹤(beat tracking) ([18]、[20]、[22])兩種研究方向，前者將音訊檔案視為有穩定節奏的訊號，目標是找出其穩定的節奏速度值；後者是針對節奏速度較不穩定的音訊計算其節奏速度的變化。由於本系統是考慮各個時間的音樂情緒變化，節奏追蹤為較適當的計算方式。以往分析節奏速度時，最常遇到的一大爭議就是判定節奏速度的客觀性，常見的狀況如：相同一首雙數拍的音樂，或許多數人會和人工定義的標準答案以相同的拍度(BPM - Beats Per Minute)打拍子，但有些人會以兩倍或以二分之一倍的拍速來跟著音樂打拍子。這是節奏追蹤的效能評估一直以來最大的一項爭議[24]。

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，如圖 4.4.3 所示：

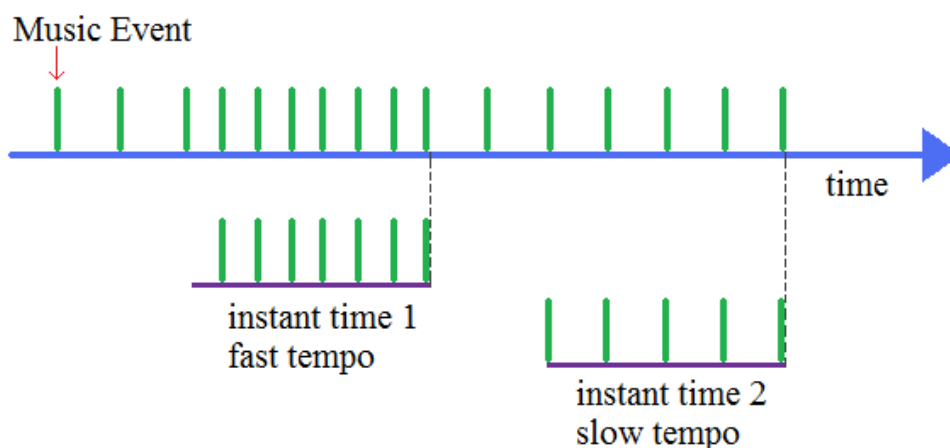


圖 4.4.3 音樂事件與節奏關係示意圖

音樂事件與節奏關係示意圖如圖 4.4.3，藍色線段為音樂的進行方向，綠色線段代表音樂事件的端點位置，虛線分別為兩個不同的聆聽音樂的當下音框時刻，紫色線段代表固定的回溯時間長度，可以看到不同時間的當下回溯前一小段時間為當下節奏速度感的主要依據，此即為本系統針對節奏速度的計算所使用的概念。在客觀看待每一個音樂事件的情況下，所有的音樂事件可以以 0 和 1 來表示，0 表示無音樂事件，1 表示有音樂事件。但考慮人在聆聽音樂的過程中，最主要影響當下的節奏感的應該是當下往前一段時間內的音樂事件，而非經過太久和未來未發生的音樂事件。所以本系統利用當下時刻以前的一小段時間的音樂事件密集程度代表當下的節奏速度，並且給予各個音樂事件距離當下的時間長度不同的權重，離當下越遠的音樂事件的權重越小，反之離當下越近的音樂事件權重越大。這個權重函數本系統簡單以半個 Hamming Window 來表示，如圖 4.4.4。

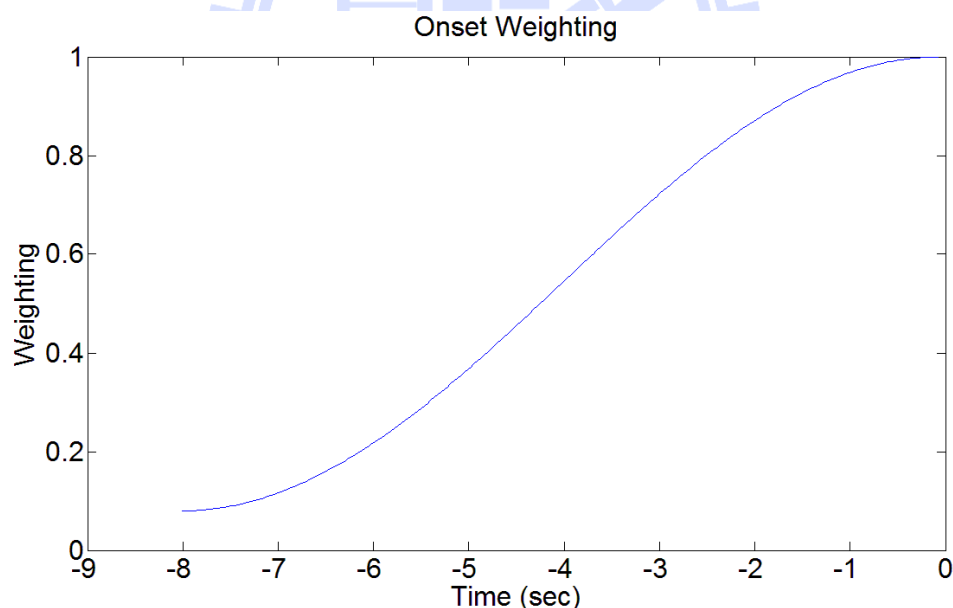


圖 4.4.4 音樂事件的權重分佈

圖 4.4.4 中可以看到，x 軸座標代表其和當前的音框時間的差額，即當下往前 8 秒鐘所有時刻的權重比例。總結上面敘述，任一個當下的節奏速度可以利用音樂事件的密集程度表示為下式：

$$\text{Onset Density}[n] = \sum_{m=\# \text{ of frames of past 8 seconds}}^n \text{Onset}[m] \cdot w[m] \quad (14)$$

其中 n 為當前的音框索引， m 為音框索引， $\text{Onset}[m]$ 為代表各個音框有無音樂事件的函數，當某個音框時刻有音樂事件則為 1 反之為 0， $w[m]$ 即為 Hamming Window 型式的權重函數。實際計算結果可以參考圖 4.4.5 與圖 4.4.6，分別為貝多芬的月光奏鳴曲與 Muse 的 Stockholm Syndrome 之音樂事件密集程度變化。

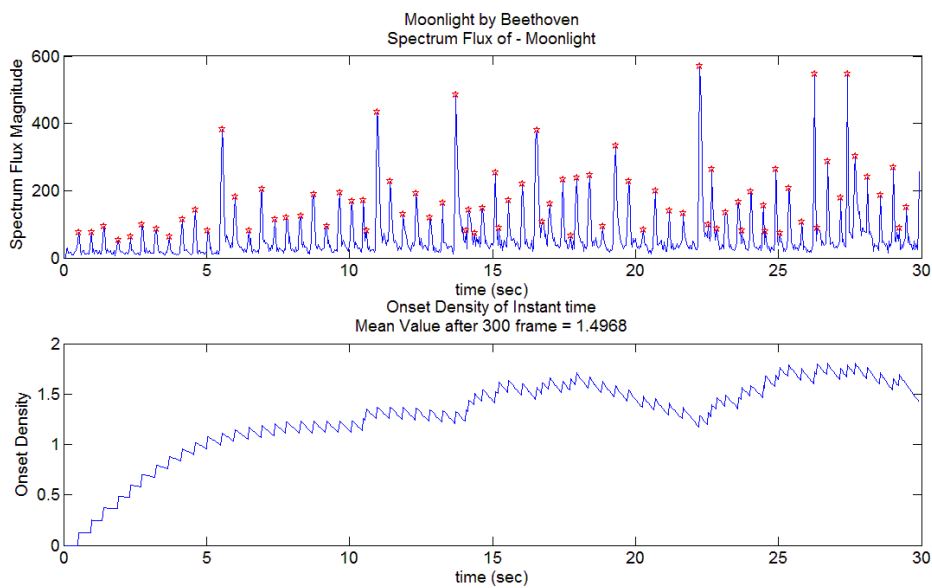


圖 4.4.5 貝多芬之月光奏鳴曲-音樂事件密集度

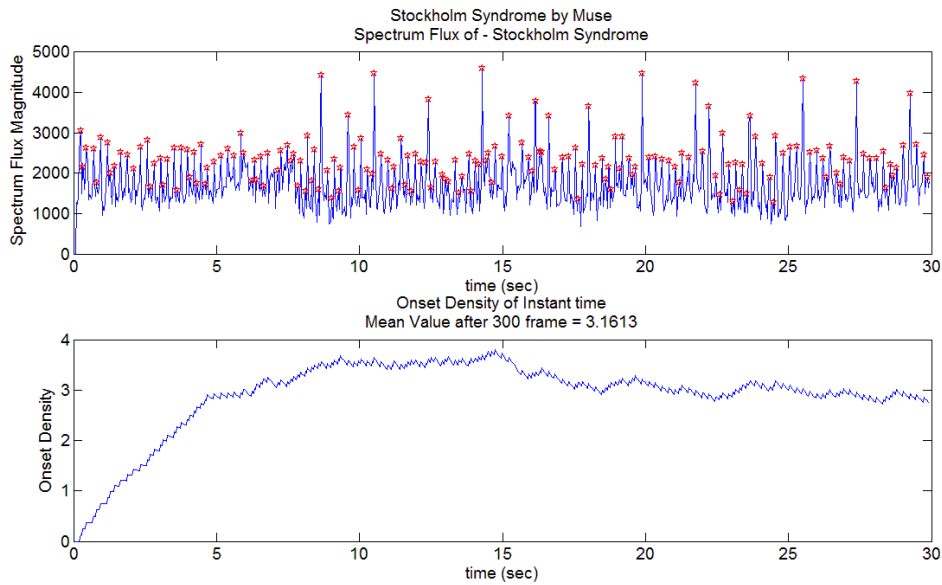


圖 4.4.6 Muse 之 Stockholm Syndrome-音樂事件密集度變化

由圖 4.4.5 與圖 4.4.6 可以看到，音樂事件的密集程度可以表示出音樂內容的節奏速度大小之不同。這樣的概念有別於以往的節奏追蹤系統，以音樂事件密集程度來代替節奏追蹤的最大優點為其排除了節奏的倍數差異的難題。考慮一個常見的情況：一首慢拍速的音樂，在音樂進行中樂手突然以較短促且快速的音符獨奏，這一瞬間聽者的節奏速度感受應該是相當強烈的，但以往的節奏標記或追蹤系統則傾向音樂段落前後整體速度不變而較無法反映出這樣的段落，而且倍數拍的差異會造成系統無法辨識出其音樂實際上到底是快還是慢，這些問題對於情緒辨識與預測上是一個相當大的阻礙，若需應用節奏速度為特徵來幫助情緒辨識，節奏速度需要是絕對的節奏速度而不能容許倍數差異。另外因為本系統的目標是找出當下的情緒反應而非讓電腦依照節奏進行互動，以音樂事件的密集程度代表音樂當下的節奏速度是一個替代節奏追蹤的可行辦法。訓練資料 192 首音樂片段的平均音樂事件密集度與各個類別的平均的計算結果請參考第五章。

4.5 音量大小

聲音的大聲或小声，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。單一音框的音量計算方式如文獻[27]所述，最簡單的是計算音框內振幅的絕對值總和，如下：

$$\text{Volume}_1[m] = \sum_{n=1}^N |x_m[n]| \quad (15)$$

但由於人的感知系統在不同頻率時感知上的音量大小並不相同[26]，另一種為計算振幅之平方直總和再取對數單位-分貝(decibel)，較接近人的音量感知，如下：

$$\text{Volume}_2[m] = 10 \times \log_{10} \left(\sum_{n=1}^N x_m[n]^2 \right) \quad (16)$$

本系統之音量計算方式採用第二種計算方式，實際計算結果可以參考圖 4.5.1 和 4.5.2。

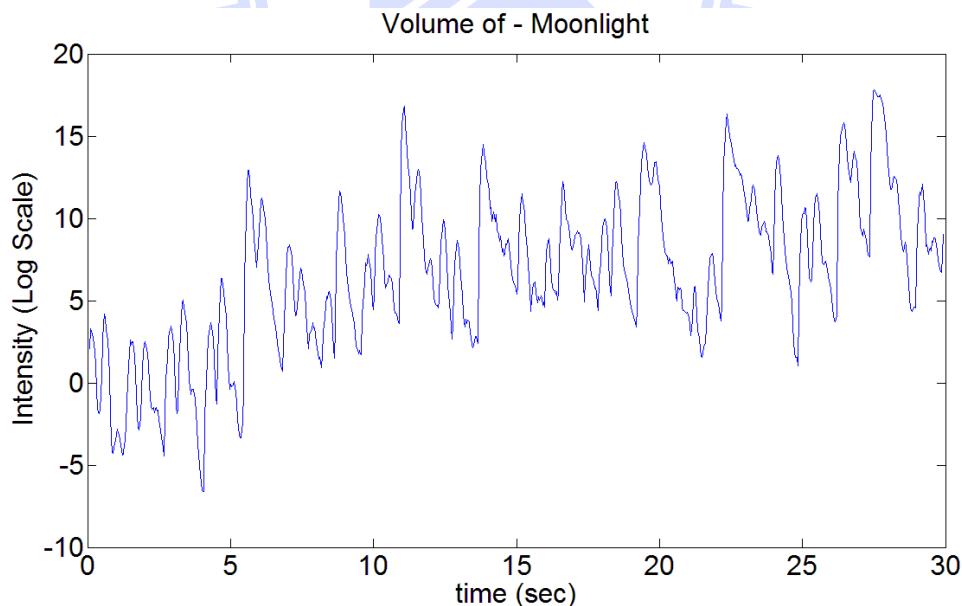


圖 4.5.1 貝多芬之月光奏鳴曲-音量變化

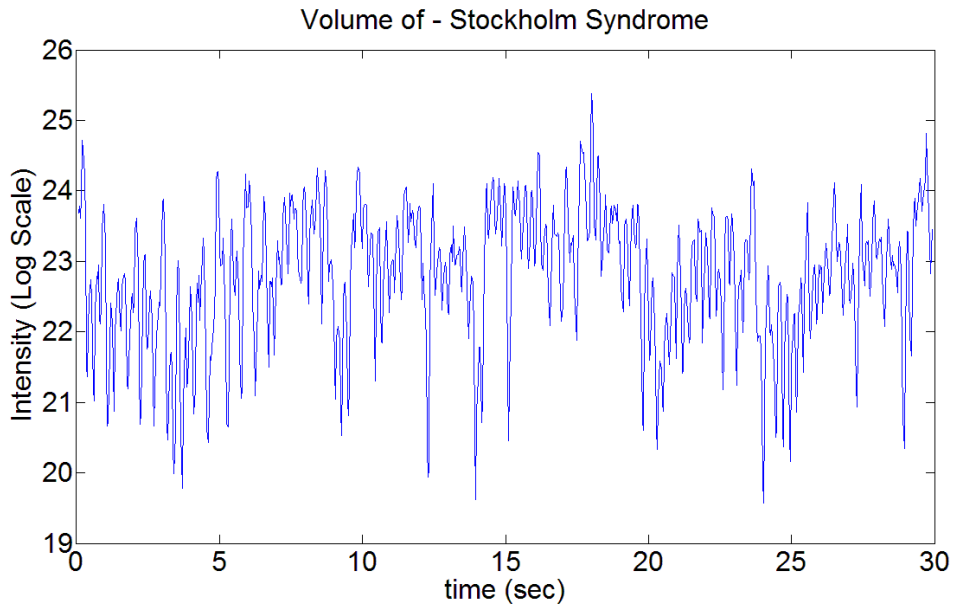


圖 4.5.2 Muse 之 Stockholm Syndrome-音量變化

4.6 音色分析

音色分析的基本方法如表 2.2.1 所述，本文嘗試以頻譜的質心(Centroid)配合頻帶寬(Bandwidth)來做分析。頻譜質心計算方式如下式：

$$\text{Spectrum Centroid}[n] = \frac{\sum_{k=1}^{N/2+1} (f[k] \cdot |S_n[k]|)}{\sum_{k=1}^{N/2+1} |S_n[k]|} \quad (17)$$

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：

$$\text{Bandwidth}[n] = \frac{\sum_{k=1}^{N/2+1} (f[k] - \text{SC}[n]) \cdot |S_n[k]|}{N/2+1} \quad (18)$$

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，如圖 4.6.1 和圖 4.6.2，圖中綠色線段為整體平均值，對於能量的貢獻方式如第(19)式，可以簡單以頻譜質心乘以頻帶寬表示。

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

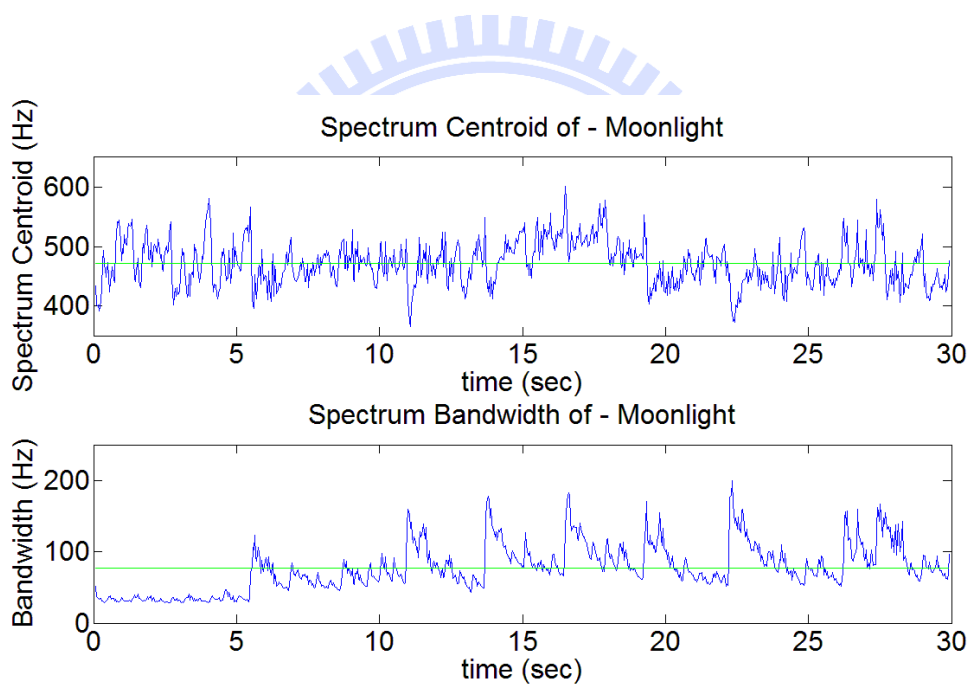


圖 4.6.1 貝多芬之月光奏鳴曲-頻譜質心和頻譜帶寬

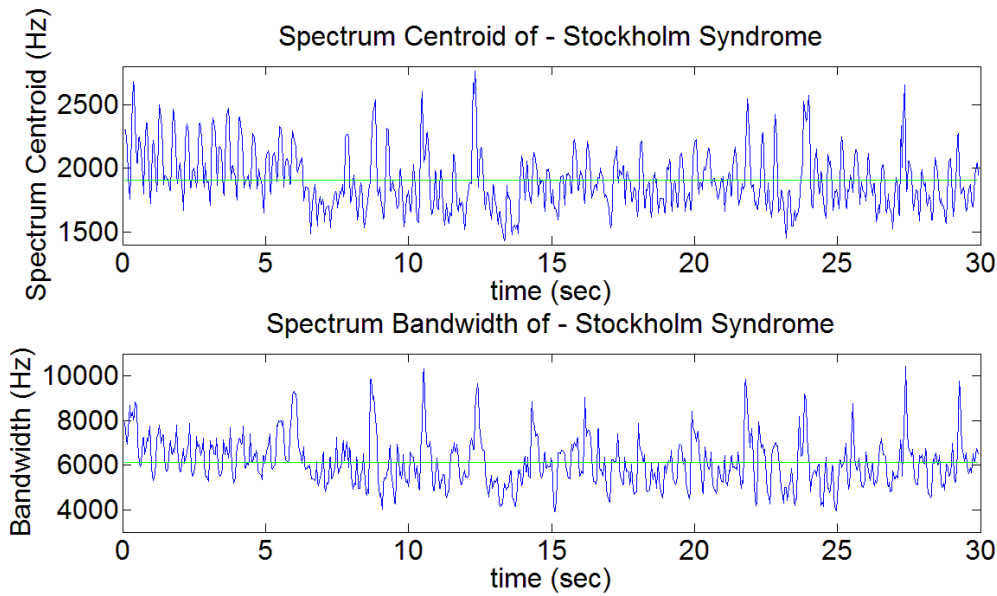


圖 4.6.2 Muse 之 Stockholm Syndrome-頻譜質心和頻譜帶寬

4.7 調性分析與和聲和諧程度

一般而言音樂調性也為一種相當具代表性的音樂特徵，旋律和和聲的變化造就了一個音樂的整體感，在音樂理論上稱為調性。調性在一般的音樂中皆可以得到一些概要性的分析，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。進行調性的追蹤可以知道前一小段時間內的音樂內容對於聽者的情緒感受影響大概為何。計算短時距頻譜並且轉換其中 100Hz 至 5000Hz 之間的頻譜為音調層級表示法有助於音樂調性的分析，選用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。文獻[29]以平均率分析音樂調性中各個調性的音符出現時間比例，比例的多寡即為該調性的音符權重，各種不同的音符權重分配即為該調性的 12 音符的調性樣板，如圖 4.7.1 和圖 4.7.2：

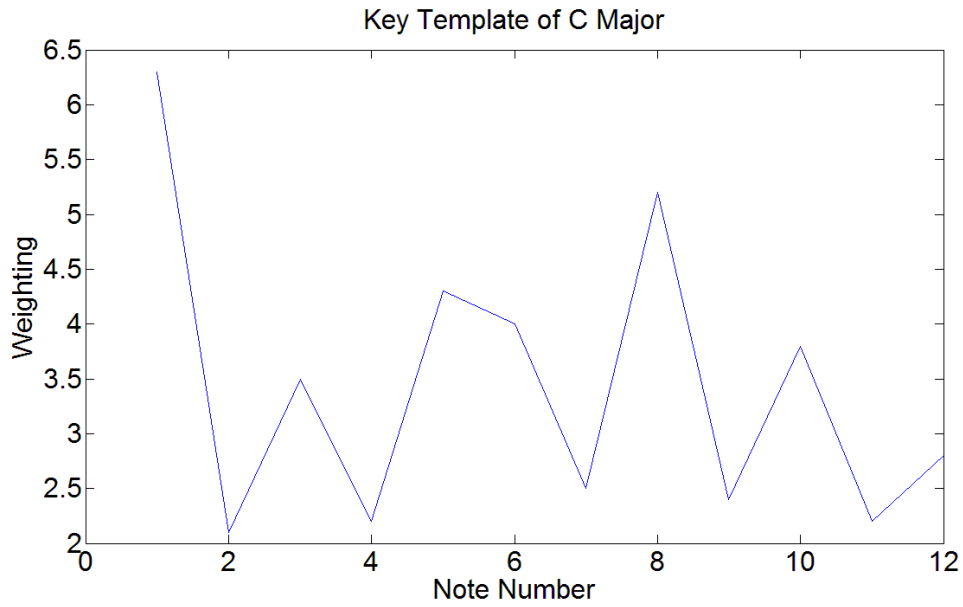


圖 4.7.1 C 大調調性樣板

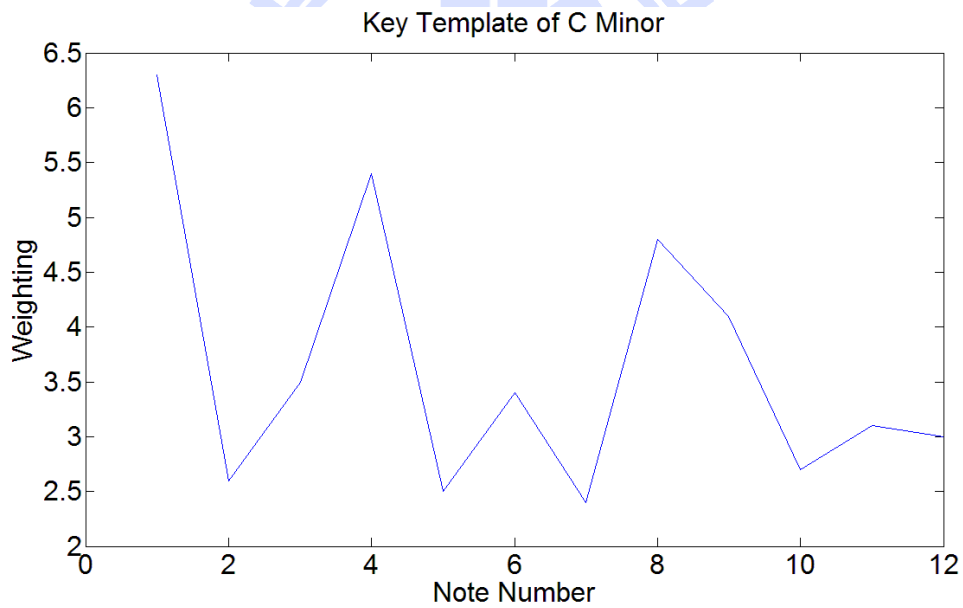


圖 4.7.2 C 小調調性樣板

圖 4.7.1 和 4.7.2 中橫軸代表平均律中由 C 至 B 的 12 個音名，注意這邊只針對音高作分析而不將同音異名分開討論；縱軸代表音符出現的比例強度或是權重。將權重位置橫向移動即可以得到 12 個大調與 12 個小調的調性樣板，總共 24 種調性樣板。文獻[2]也以類似的方式製作調性樣板，但是此篇將 KEY 與 MODE 分

成兩個層級來探討。實際計算利用 2.3 節所述音調層級表示法，先以下列兩種樣板進行調音：

$$\begin{aligned} \text{Tuning1} &= [1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0] \\ \text{Tuning2} &= [0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1] \end{aligned} \quad (20)$$

比較所有音框之 24 維的 PCP 向量和上述兩種樣板的內積總和值大小，即可知道音訊內容之音準比較接近為 Tuning1 還是 Tuning2，利用此結果簡化 PCP 向量成為 12 維的向量，即對應十二平均律中的 12 個不同的音調層級：

$$\begin{aligned} \text{if } \sum_m \text{PCP}_m \cdot \text{tuning 1} &> \sum_m \text{PCP}_m \cdot \text{tuning 2} \\ \text{simplified PCP} &= \text{PCP}(1:2:23) \\ \text{else } \sum_m \text{PCP}_m \cdot \text{tuning 1} &< \sum_m \text{PCP}_m \cdot \text{tuning 2} \\ \text{simplified PCP} &= \text{PCP}(2:2:24) \end{aligned} \quad (21)$$

上式中 m 為音框索引， PCP_m 為第 m 個音框的 PCP 向量。以簡化的 PCP 向量與調性樣板作向量內積即可以得到某個時刻屬於該調性的機率大小為何，與該調性調性樣板內積值越大代表此刻出現的音符較接近該調性。在音樂的表現中，音符可以同時出現與先後出現，因此對於音調層級的分析也需要考慮在某個時間往回一段時間內，例如：當下以前的 8 秒以內的所有音調層級變化。每個音框所對應的調性的機率，以當下往前 8 秒內的所有音框之調性機率總和大小來判別，如下式：

$$\text{KeyProb}[k, n] = \sum_{m=\# \text{ of frames of past 8 seconds}}^{m=n} \text{PCP}_m \cdot \text{KeyTemplate}[k] \cdot w[m]$$

$$\text{Key}[n] = \max(\text{KeyProb}[1:24, n]) \quad (22)$$

其中 n 、 m 為音框的索引， n 為當前的音框， m 則為由過去 8 秒到當前的音框 n 的所有音框， k 為調性維度的索引。KeyProb[k, n] 為第 n 個音框中第 k 個調性的機率值，計算方法為將第 n 個音框以前 8 秒鐘的所有音框和第 k 個調性的樣板做內積並計算總和，Key[n] 為 n 個音框的調性，由 KeyProb[1:24, n] 之最大值求得。如同音樂事件密集度的分析，配與各個音框的音調層級表示法適當權重 $w[m]$ (同 4.4 節中所述)，減低離現在時刻較遠的音調層級的比例，最後選出擁有最大內積值的調性作為正確答案。實際計算如圖 4.7.3：

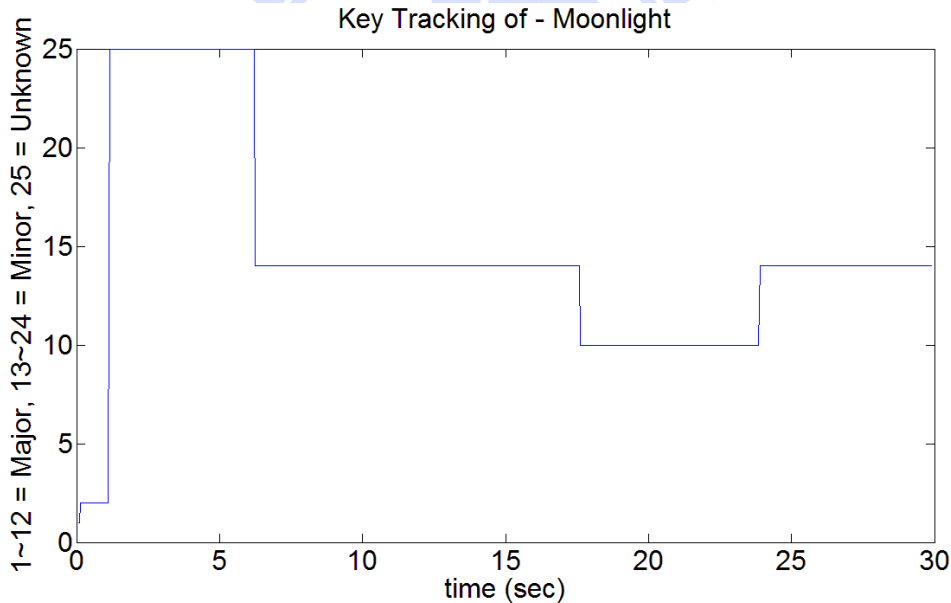


圖 4.7.3 貝多芬之月光奏鳴曲-調性追蹤

圖 4.7.3 顯示貝多芬之月光奏鳴曲第一樂章前 30 秒的音訊內容的調性追蹤結果，原曲為 #C 小調，可以看到調性追蹤結果前 5 秒由於音符太少與音量太小而無法計算，隨後判斷為 14 (#C 小調)。但本研究中只需要知道調性為大調或是小調，此外在音量太小的時候應設定為不確定，以避免分析到無意義的數值，即三種輸出：大調、小調、無法辨識，所有訓練資料的調性追蹤結果分析請參考第五章。

一般在聆聽音樂時很容易就可以發現，不同的和聲帶給聽者不同程度的和諧感，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，因此和聲為一種相當清楚明瞭和情緒有直接關聯的特徵。對於和聲的分析，最普遍的方法為聲音頻譜的分析方式，因為和聲通常反映在其頻譜強度分佈模式。在音樂理論分析上，學者常用音程的概念來探討兩個聲音的和諧與否，音程的定義為兩音之間的音高上的距離或差異，而音程的命名分為兩個部分：(1)音程的距離或稱度數(2)形容音程的形容詞。音程的度數的計算方法為兩音的音名之間經過的音名數目，如 A 和 C 其中間共經過 A、B、C 三個音名，所以 C 為 A 的三度音。反之 A 為 C 的六度音，因為 C 到 A 共經過 C、D、E、F、G、A 六個音名。其中一度、四度、五度、八度定義為完全音程，二度、三度、六度、七度則為大小音程。音程的形容詞關係到實際上兩音之間的半音差，種類有完全、大、小、增、減。完全、增或減用來修飾完全音程，大和小則用來修飾大小音程，詳細的定義如表 4.7.1。

表 4.7.1 八度內的音程表

音程名稱	完全一度	小二度	大二度	小三度	大三度	完全四度	增四或減五度	完全五度	小六度	大六度	小七度	大七度	完全八度
半音差額	0	1	2	3	4	5	6	7	8	9	10	11	12

不同音程即代表不同的音高頻率比，對應到不同的和諧程度。最早由 Pythagorean 提出震動的弦長為簡單比例即為和諧，如八度音的 2 比 1，五度音的 3 比 2。後期 Helmholtz 也提出兩音是否和諧是關鍵在於拍音(beat)的產生，最後的結論和 Pythagorean 的結果幾乎相同。音程的概念在音樂表現中相當的重要，目前普遍的音樂簡單可以分解成為伴奏和聲部分和主要旋律部分，各個音符不論是同時或是先後的發生皆有所謂的音程關係，但可惜目前並無相關研究指出各種音程與其

進行或變化對於情緒的影響為何，因此在應用上還存在許多問題。本文中嘗試以最簡單的和聲和諧程度分析，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，如下式。

$$\begin{aligned} \text{Note1}[n] &= \max(\text{PCP}(1:24,n)) \\ \text{Note2}[n] &= \text{PCP}(\text{index of Note1}[n] \pm 12, n) \\ \text{Dissonance}(n) &= \frac{\text{Note2}[n]}{\text{Note1}[n]} \cdot \frac{\text{Note1}[n] + \text{Note2}[n]}{\max(\text{PCP}(\text{all}, \text{all}))} \end{aligned} \quad (23)$$

上式中 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。

4.8 計分方法

在系統計分的方法中，本文設計了一套計分方法簡易的模擬聆聽音樂的過程，其概念為聽者當下的情緒受到前一小段時刻的聆聽感受影響，而隨時間流逝，情緒感受也會隨之釋放。定義聽者當下的情緒位置 P_t 表示於情緒平面上，而 P_t 為過往的聆聽過程的累加，前述 5 種特徵強度分別配與特定權重 w 決定了當下的情緒得分 p_t 如第(24)式所述， f 為特徵值的索引，分別代表 5 種不同的特徵， $S_f(t)$ 為 t 時刻第 f 種特徵的強度， $w_x(f)$ 、 $w_y(f)$ 分別為情緒平面之 x 軸和 y 軸對應各個特徵的權重關係。

$$p_t = \sum_{f=1}^5 \left[w_x(f) S_f(t) \bar{x} + w_y(f) S_f(t) \bar{y} \right] \quad (24)$$

整個計分過程如圖 4.8.1， P_t 的累加過程中，而每個時刻的累加分數 p_t 再乘以衰退函數 σ ，代表隨時間衰退，即情緒的釋放。

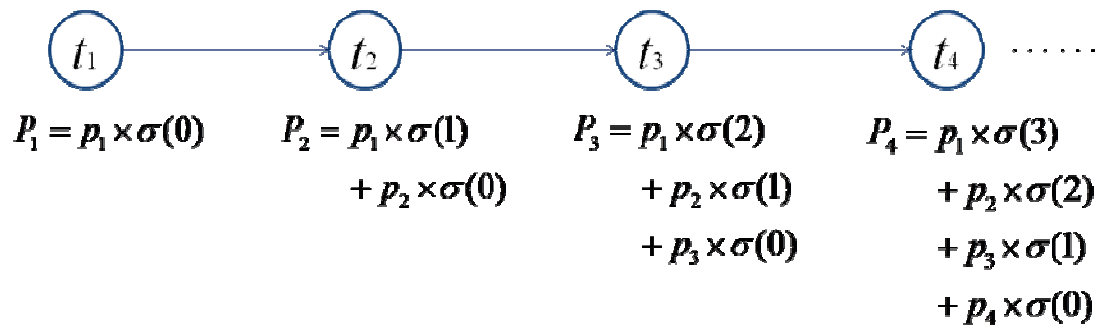


圖 4.8.1 計分流程圖

關於連續時間的音樂特徵值變化與 Thayer 情緒模型的兩軸對應的情形，目前並無文獻做詳盡的分析，本研究中依照計算結果調整配給的權重如表 4.8.1：

表 4.8.1 音樂特徵值與 Thayer 情緒模型的對應關係

	音樂事件密集度	音量	音色	調性	和聲和諧度
能量	0	0	0	1	0.91
壓力	1	0.4	0.14	0	0

4.9 情緒邊界的訓練

本系統中，不同的特徵強度會使聽者情緒指標產生相對的加減分和位移，但是並無明確邊界可以界定出不同情緒之間的邊界。利用本系統的計分模型所產生的情緒位移座標，可以使用 GMM 來訓練出各個情緒邊界的範圍，這樣的概念除了可以使使用者了解音樂的變化和情緒變化的關係，也可利用於整體最終概要情緒的辨識。新的測試音樂的情緒指標位移即可利用已經訓練好的邊界來辨別每一段落的可能情緒為何。由於實際計算後的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，訓練資料的取得為情緒軌跡位移後半段的部分，且為隔數

個音框取一個軌跡座標做為訓練資料，如圖 4.9.1 和圖 4.9.2：

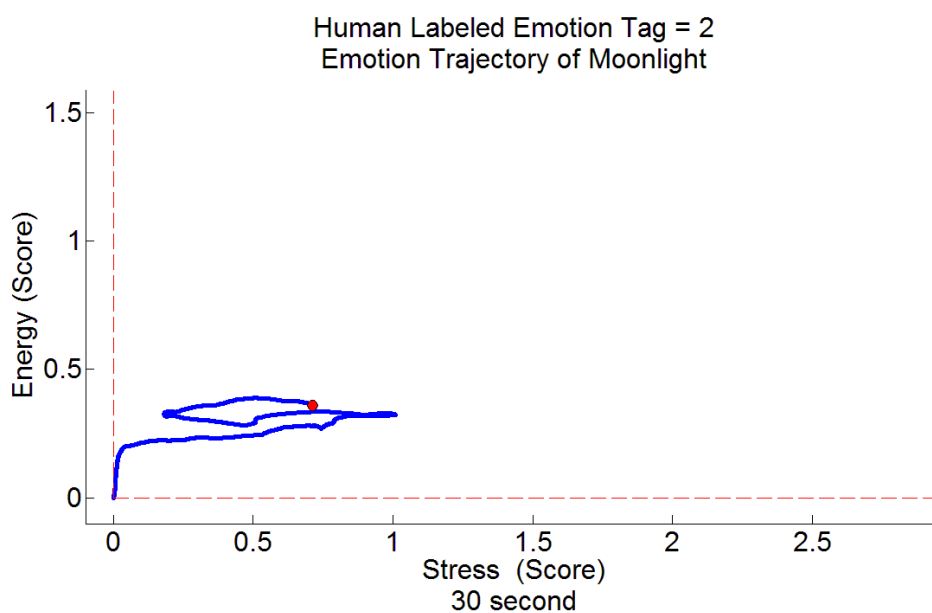


圖 4.9.1 貝多芬之月光奏鳴曲-情緒軌跡位移

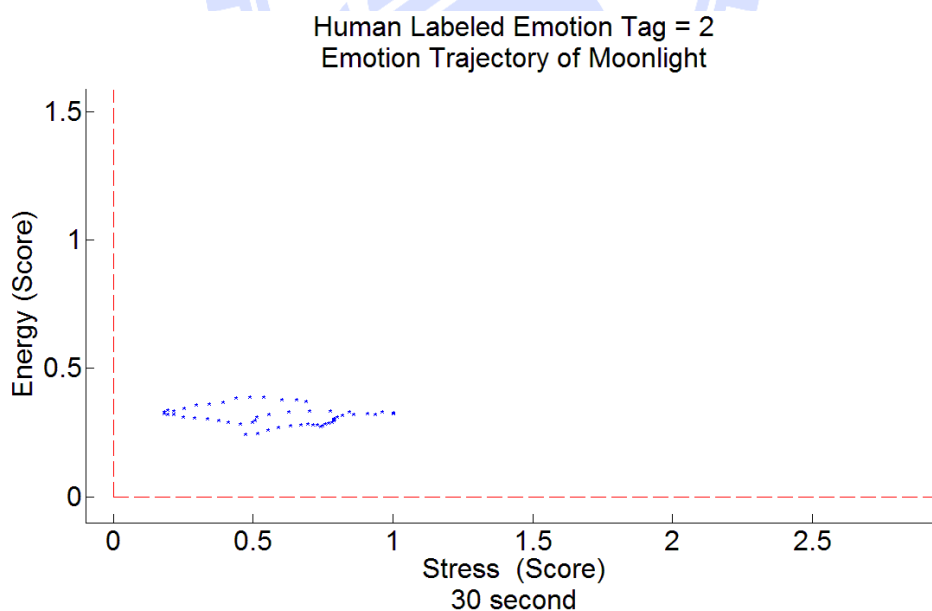


圖 4.9.2 貝多芬之月光奏鳴曲-情緒軌跡位移所提供的訓練資料

圖 4.9.1 和圖 4.9.2 說明了訓練資料的取得方式，得到數點軌跡座標後即代表此類別(圖例中為類別 2-哀傷)的訓練資料，經過 192 首音樂片段的訓練資料採集後的

資料分佈如圖 4.9.3：

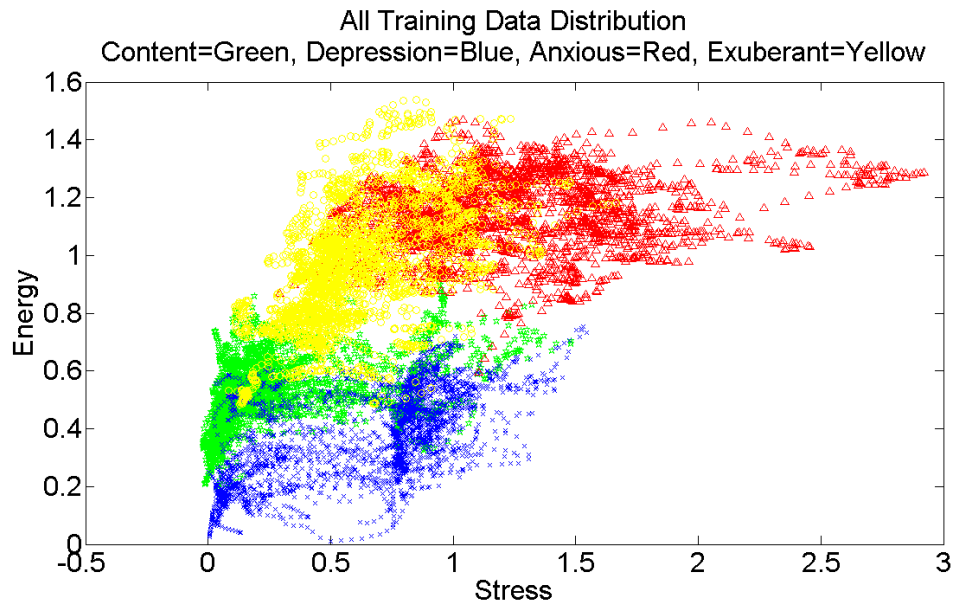


圖 4.9.3 訓練資料分佈

圖 4.9.3 中可以看到，不同類別的訓練資料雖然有部分重疊，但大致上可以看到的分佈方式如一開始的假設，即 Thayer 情緒模型中的分佈情形。綠色代表舒適的音樂情緒軌跡，集中在情緒平面上壓力與能量皆比較小的區域；藍色代表哀傷的音樂情緒軌跡，集中在情緒平面上壓力較大但能量小的區域；紅色代表焦慮的音樂情緒軌跡，集中在情緒平面上壓力較大且能量也大的區域；黃色代表振奮的音樂情緒軌跡，集中在情緒平面上壓力較小但能量大的區域。經過各個類別的 GMM 訓練與邊界範圍的計算，可以看到各個類別的 GMM 訓練結果，如圖 4.9.4、圖 4.9.5、圖 4.9.6、圖 4.9.7，最終 GMM 之邊界分佈情形如圖 4.9.8。

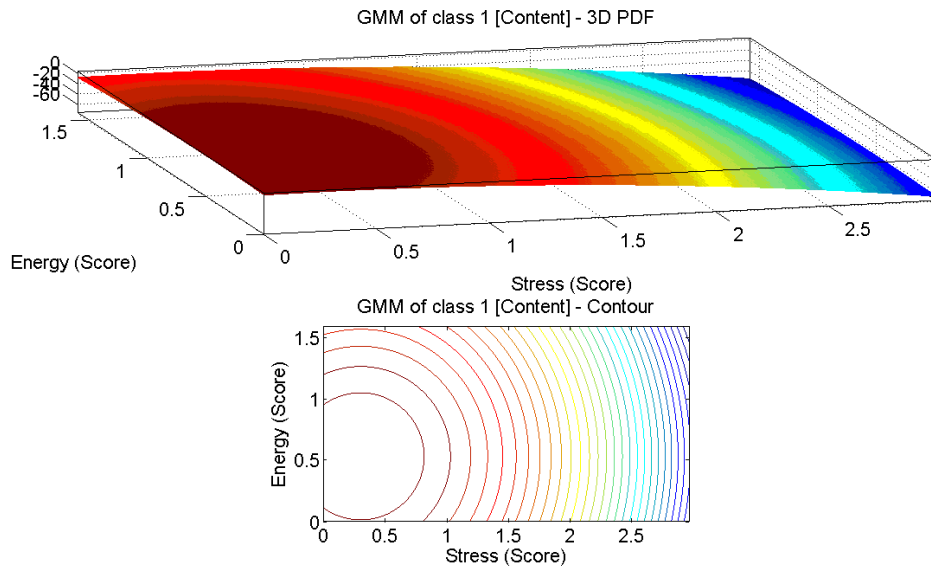


圖 4.9.4 類別 1 的高斯分佈

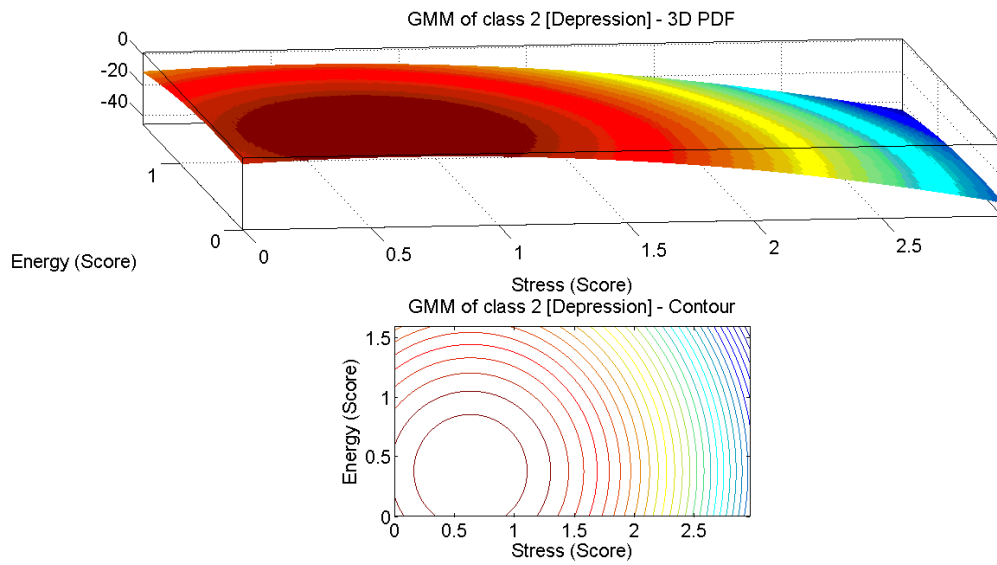


圖 4.9.5 類別 2 的高斯分佈

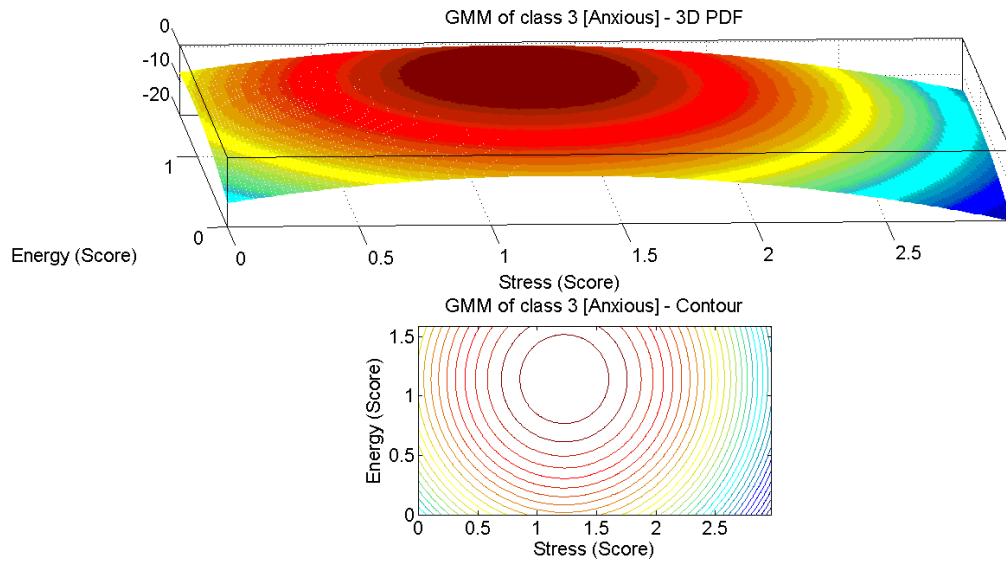


圖 4.9.6 類別 3 的高斯分佈

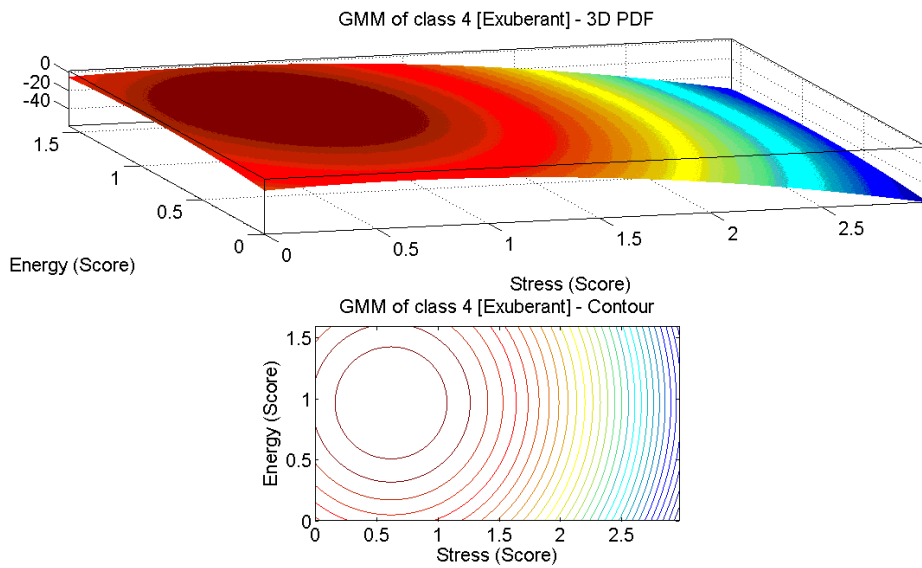


圖 4.9.7 類別 4 的高斯分佈

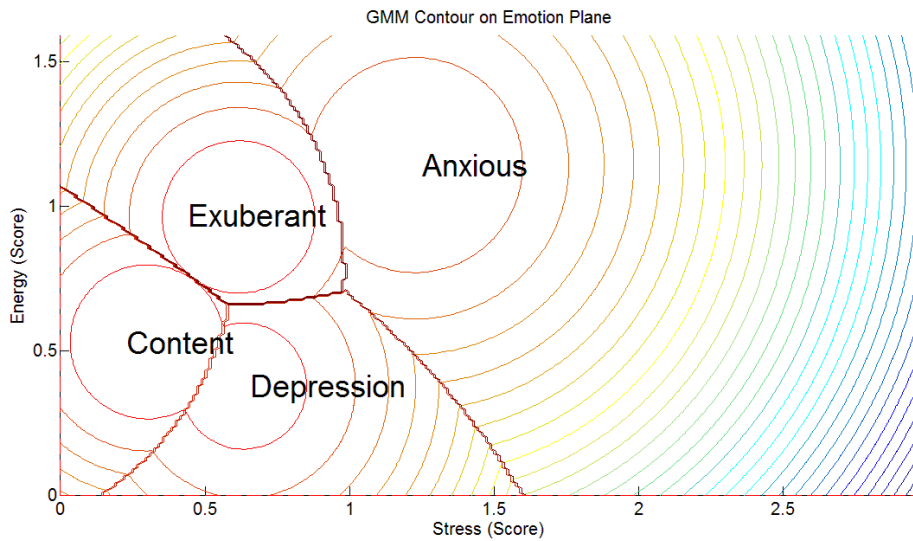


圖 4.9.8 GMM 分類結果與各類別的邊界範圍

由圖 4.9.4、圖 4.9.5、圖 4.9.6、圖 4.9.7 可以看到各個類別的 GMM 訓練結果與分佈情形，圖上方為立體的機率函數密度(PDF)，下方則為等高線分佈情形。圖 4.9.8 則為計算 4 個類別之機率函數密度的交叉範圍，即各個類別的邊界範圍。可以看到情緒類別的分佈大致如 Thayer 的情緒模型。最終得到內部測試(Inside Test)的辨識率為 70.731%，以上為高斯函數個數為 1 個的分類結果，高斯函數個數為 2 和 4 時的分類結果可以參考附錄-B。隨著預設之高斯函數個數的不同，內部測試辨識率的變化可以參考圖 4.9.9，圖 4.9.9 中可以看到辨識率隨著高斯函數個數的增加到達約 82%的辨識率極限。

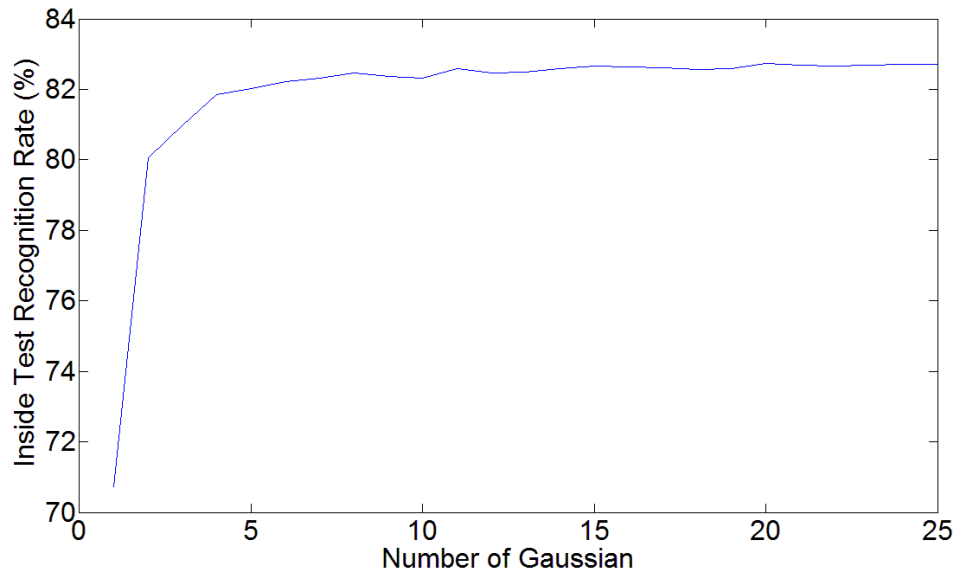


圖 4.9.9 內部測試辨識率與高斯函數個數的關係



五、計算結果分析

5.1 訓練資料之特徵萃取數據分析

依照第 4.4 節所述之音樂事件密集度計算方法，計算 192 首音樂片段 30 秒平均的音樂事件密集度可以參考表 5.1.1。

表 5.1.1 訓練資料之平均音樂事件密集度

#	舒適	哀傷	焦慮	振奮
1	1.8317	3.0496	3.2409	3.2936
2	1.8453	1.1819	3.0641	3.0282
3	1.7024	1.9434	2.8976	3.3936
4	1.9103	0.8110	2.5484	2.7766
5	1.3115	1.0173	3.0286	3.0385
6	1.4537	1.3003	2.8617	2.6636
7	2.5231	1.3672	2.4955	2.9833
8	2.1873	0.3458	2.7683	3.9182
9	2.2359	1.9967	2.9789	3.2996
10	2.6977	2.0073	3.0476	2.3755
11	1.9870	1.5381	4.1978	2.3457
12	2.0313	2.0497	4.0223	2.5309
13	1.9722	1.4118	3.3700	3.2153
14	1.4844	1.8707	2.4319	2.9252
15	2.8222	2.5172	2.9471	3.2434
16	1.7246	2.9798	3.1930	3.7821
17	2.0055	2.2104	3.2477	2.9349
18	2.6601	1.4994	3.2585	3.0436
19	2.4175	1.0035	2.8901	1.9382
20	1.8883	1.4006	3.2712	2.4206
21	1.6350	1.0909	3.3324	2.6651
22	1.8200	1.8398	4.0271	2.5650
23	2.5509	2.6897	3.5103	2.6325
24	1.7254	1.2864	3.3009	2.6048
25	1.4723	0.8537	3.8805	2.8579
26	1.4943	1.2465	3.1613	3.5842
27	2.2960	0.9182	4.3277	3.2044

28	2.0225	1.4023	3.1613	2.3740
29	1.9296	1.0613	3.2772	2.3306
30	2.1345	2.6120	3.1451	3.2506
31	2.2991	1.6045	3.0251	2.9204
32	2.2690	2.1549	3.1813	2.8675
33	2.4904	0.8318	2.9729	3.0190
34	2.2830	1.1181	3.0601	3.0593
35	2.4925	1.4968	3.1231	2.9120
36	2.0861	2.0409	3.2701	2.4453
37	2.1394	2.1507	2.6639	2.8007
38	1.5446	1.8224	3.6848	3.6573
39	1.9343	1.8949	2.9260	3.0403
40	1.6680	2.2407	2.7026	3.6291
41	2.0609	1.9488	3.1341	3.2465
42	2.7015	2.0783	2.5851	2.3675
43	1.6739	1.3878	2.8928	2.7410
44	2.1175	1.5612	2.9717	3.5653
45	1.9161	2.5493	3.1005	4.0751
46	2.0311	1.4914	3.0500	3.5884
47	2.2657	1.9852	2.9760	2.4725
48		2.0291	2.7041	3.2618
49				2.8129
平均值	2.0371	1.6852	3.1439	2.9735

由表 5.1.1 中可以看到各個的音樂片段的平均音樂事件密集度，詳細的曲目與曲名可以參考附錄 A，最下方紅色粗體為類別的整體平均值，結果顯示舒適與哀傷的音樂之平均音樂事件密集度相較於焦慮與振奮之音樂事件密集度小，顯示出普遍舒適與哀傷的音樂傾向擁有較慢的節奏速度，而焦慮與振奮的音樂傾向擁有較快的節奏速度。

關於音量特徵值的計算方法如 4.5 節所述，訓練資料之對數音量特徵值與平均值可以參考表 5.1.2。

表 5.1.2 訓練資料之平均對數音量

#	舒適	哀傷	焦慮	振奮
---	----	----	----	----

1	11.1771	3.2003	22.1426	21.6997
2	14.1486	8.5754	22.3818	21.6981
3	11.7757	11.5506	17.2224	20.9144
4	15.9214	-5.8588	22.4159	21.0332
5	9.7617	2.7862	21.3015	22.8668
6	12.5124	0.3592	22.7900	19.4383
7	16.8226	16.8014	22.3679	21.3074
8	19.4260	3.0423	23.1911	15.2963
9	16.0838	10.6259	23.2835	16.1981
10	17.8882	21.6207	23.0234	16.0541
11	14.1631	10.7957	18.4686	15.6414
12	17.3166	16.0223	20.4099	18.3693
13	17.4973	4.9014	21.7834	21.2597
14	14.1788	7.0042	21.2062	13.8898
15	14.6220	14.0049	21.9476	12.6168
16	13.0088	2.2947	19.8094	15.5411
17	15.8966	6.4080	22.0125	15.7500
18	17.5812	8.7185	20.0361	16.1465
19	19.4984	6.5264	22.4953	15.9583
20	16.9691	9.6697	22.5226	20.3851
21	17.1701	-7.4158	22.4138	21.0579
22	14.0770	2.6182	21.0081	21.3479
23	16.7856	12.1148	22.1412	21.5731
24	7.1502	2.4250	22.3334	22.4234
25	13.0659	0.0229	22.9650	20.7070
26	5.7126	6.6791	22.7809	20.3404
27	8.3592	4.5940	23.0280	19.2323
28	11.8356	11.3021	18.8045	20.7587
29	7.4672	5.9531	22.3850	21.8499
30	12.4491	8.6593	22.4061	19.8079
31	10.8851	-5.6521	19.5560	23.0822
32	11.1367	0.1423	20.8583	21.8483
33	12.6500	-3.7572	22.0959	24.5151
34	9.6689	11.9493	19.9754	21.7224
35	12.5085	6.6621	18.2893	22.2493
36	16.6456	14.0207	21.1261	17.7154
37	18.3787	14.2418	20.6834	20.9020

38	9.0470	13.7707	21.0761	21.6081
39	15.0709	11.2558	21.8314	20.6228
40	14.6818	12.0924	19.3953	19.8355
41	19.3894	10.1160	21.4420	18.6932
42	13.5982	16.4096	20.3383	8.9868
43	15.8667	14.1034	19.9857	13.4193
44	9.9621	14.5853	20.9755	10.7957
45	10.1850	6.2713	21.4743	10.3363
46	17.6062	-0.0349	21.6069	12.4264
47	21.1600	12.5759	22.0873	21.1665
48		13.2968	20.9263	21.7429
49				20.4166
平均值	14.0162	7.6677	21.3500	18.8418

由表 5.1.2 中可以看到訓練資料之對數音量平均值，整體而言舒適與哀傷的音樂音量較小，焦慮與振奮的音樂音量則較大，這說明在情緒平面中音量對於其能量為一種具代表性的特徵。

關於第 4.6 節中所提出的音色分析方法，其訓練資料之音色平均值可以參考表 5.1.3。

表 5.1.3 訓練資料之音色平均值

#	舒適	哀傷	焦慮	振奮
1	1.4212	0.2659	15.4051	25.7307
2	2.2700	0.6281	12.3567	16.7706
3	2.7127	1.7557	9.7837	17.8149
4	4.1202	0.0731	26.6246	17.7583
5	1.5529	0.2099	16.9406	28.2290
6	2.2920	0.1559	18.1537	14.7561
7	3.2612	0.5060	17.8911	21.5183
8	5.9301	0.8958	22.1035	3.6282
9	3.3491	0.6366	28.7267	4.5481
10	4.6121	2.9886	26.6908	2.6187
11	2.6926	1.5296	14.5962	7.6497
12	1.8882	2.1236	14.3366	15.2733
13	4.2018	0.1253	21.6638	22.2530

14	2.2236	0.5823	24.5277	4.3081
15	0.8668	1.6120	27.3120	3.5418
16	1.3183	0.2590	12.8288	7.0121
17	3.4633	0.9274	22.1055	5.9315
18	6.6403	0.8477	9.4153	4.8118
19	7.9069	0.2170	18.9842	5.9432
20	3.3993	0.1133	24.8059	10.2530
21	3.4185	0.0890	24.2745	15.7012
22	2.3080	0.3015	10.5256	15.9907
23	1.5620	0.9974	21.6654	12.3760
24	1.8000	1.3946	24.6790	12.5709
25	3.0762	0.2516	11.8103	12.8962
26	1.8112	0.5377	17.3964	14.8959
27	1.1641	0.5554	17.3987	16.1092
28	2.1226	3.5942	9.4717	12.5228
29	1.2326	0.7060	15.7415	18.0220
30	1.9873	1.0005	15.9678	19.6877
31	1.2825	0.1143	17.2056	26.9205
32	2.5267	0.3527	16.3533	18.6180
33	4.2524	0.0987	26.3106	30.5681
34	1.4311	0.3657	15.2839	16.6761
35	2.1687	0.0533	11.9919	17.7639
36	4.9618	1.0731	17.4845	10.6104
37	7.8112	2.9058	21.1661	15.2444
38	1.0507	0.8735	19.5681	20.3399
39	3.4247	0.8157	23.3713	15.6932
40	0.7927	2.2783	17.0305	12.4274
41	1.6508	3.3209	22.8337	16.2166
42	2.0779	1.6586	21.4935	5.2941
43	1.0330	2.4021	20.9250	12.8026
44	1.9731	0.7515	20.3074	5.4818
45	2.7237	0.3066	19.9300	2.0090
46	3.5456	0.1020	27.4096	2.7074
47	2.3809	1.9329	25.6676	16.8107
48		5.3492	20.8787	18.8112
49				13.0903
平均值	2.8020	1.0549	19.1541	13.6573

表 5.1.3 顯示音色之計算結果，其整體對於音樂之能量差異較為顯著，結果相同於音樂事件密集度與音量，焦慮和振奮的音樂之平均音色特徵值明顯比舒適和哀傷的音樂來的大。

依照第 4.7 節所述之調性追蹤方法，取出現次數最多的調性為其整體計算結果，所有訓練資料的調性計算結果可以參考表 5.1.4，其中 0 為無法辨識，1 代表大調，2 代表小調。

表 5.1.4 訓練資料之調性追蹤結果

#	舒適	哀傷	焦慮	振奮
1	1	2	2	1
2	1	1	1	1
3	1	1	2	1
4	1	0	2	1
5	1	1	1	1
6	1	1	1	1
7	1	2	1	1
8	1	2	1	1
9	1	2	2	1
10	1	2	1	1
11	2	2	1	1
12	1	2	1	1
13	1	2	1	1
14	1	2	1	1
15	1	2	1	1
16	1	2	1	1
17	1	2	2	1
18	2	2	2	1
19	1	1	1	1
20	2	2	1	1
21	2	0	1	1
22	1	1	2	1
23	1	2	1	1
24	1	0	1	1
25	1	0	1	1

26	1	2	1	1
27	2	2	2	1
28	1	2	1	1
29	1	2	1	1
30	1	2	1	1
31	1	0	1	1
32	1	0	2	1
33	1	0	1	1
34	1	2	1	1
35	1	2	1	1
36	2	1	1	1
37	2	2	1	1
38	1	2	1	1
39	1	2	1	1
40	1	2	1	1
41	1	2	1	1
42	1	2	1	1
43	2	2	2	1
44	2	2	1	1
45	1	2	1	1
46	1	0	1	1
47	1	2	1	1
48		1	1	1
49				1

由表 5.1.4 中可以看到調性追蹤之平均結果，舒適與振奮的音樂幾乎都傾向為大調歌曲，而哀傷與焦慮的音樂通常則為小調歌曲較多，這也是一般普遍的音樂與調性關係，而不和諧音程增四減五度的數據統計結果可以參考表 5.1.5。

表 5.1.5 訓練資料之和諧程度分析結果

#	舒適	哀傷	焦慮	振奮
1	0.4556	0.7011	4.4689	2.5507
2	0.5362	1.2691	3.7851	2.1385
3	0.3142	0.7688	3.1090	2.4167
4	0.4681	0.4377	4.2564	2.7829

5	0.4096	0.8502	4.1514	3.4622
6	0.3944	0.3234	3.2638	3.8624
7	0.6759	0.4186	6.3860	3.0210
8	1.6674	0.5212	6.2323	1.7298
9	1.0154	0.9128	4.7833	1.8237
10	1.3673	2.7275	4.4142	2.2462
11	0.7452	0.8049	5.1671	2.3958
12	1.5518	1.4541	4.4239	3.5116
13	1.0089	0.3753	4.5558	4.1352
14	0.7272	0.4663	5.1374	2.1863
15	0.5933	0.6122	5.1324	3.9764
16	0.3217	0.5267	3.1404	2.6031
17	0.5921	1.1389	3.4469	2.0713
18	0.7938	1.0249	3.3106	1.7601
19	0.9328	0.4711	4.4258	1.9210
20	0.4158	0.5207	6.0743	2.1236
21	0.8971	0.4305	4.9116	2.4727
22	0.7795	1.1426	6.6892	3.2070
23	1.0442	0.4635	5.5233	2.4223
24	0.4005	0.3686	4.3575	2.6569
25	0.3833	0.2948	5.0354	2.1424
26	0.4694	0.3640	3.0843	4.1329
27	0.8418	0.2936	8.5227	2.6381
28	1.0519	0.4695	2.7681	2.7080
29	0.6206	0.4727	4.6290	2.9710
30	1.1174	0.5737	4.3022	4.2966
31	0.7463	0.6759	6.4949	4.9700
32	0.7088	0.5666	3.9750	4.9140
33	1.7482	1.0201	5.6837	3.9918
34	1.2975	0.3931	4.7702	4.2126
35	1.1338	0.4528	3.3851	3.0404
36	1.8177	0.8104	3.8944	2.5391
37	3.1139	1.4968	6.3978	2.5680
38	0.5041	0.8900	3.3943	5.1607
39	0.7532	0.3888	4.8565	2.2790
40	0.8854	0.7857	4.5005	2.5526
41	0.8867	0.4542	3.6244	2.9924

42	0.8764	0.5265	5.9096	1.2326
43	0.3206	1.3460	6.5724	1.9739
44	0.5746	0.4490	7.1224	0.8171
45	0.5394	0.3871	4.8675	2.4009
46	1.1379	1.1282	6.4303	1.7809
47	0.6631	1.2376	5.1569	3.7426
48		1.9663	3.9997	2.4535
49				3.3135
平均值	0.8574	0.7536	4.8026	2.8429

由表 5.1.5 中可以看到，舒適與哀傷的音樂平均不和諧音較少，焦慮的音樂則是最多，結果顯示此項特徵並不能有效突顯出其對於 Thayer 情緒模型之壓力軸的對應關係，反而凸顯出能量軸的關係。目前關於音樂之聽覺壓力的相關研究缺乏，並沒有任何相關文獻提出音樂之壓力如何分析，本研究嘗試以各時刻最顯著音的不和諧音程(增四/減五)著手分析，結果顯示並沒有辦法有效的計算出抽象的音樂壓力數值。

以上為各個音樂片段特徵值明細與類別整體平均值，各個類別的特徵值平均整理可以參考表 5.1.6。

表 5.1.6 各類別訓練資料與其音樂特徵值之平均

類別 特徵值	舒適	哀傷	焦慮	振奮	情緒平面 對應關係
音樂事件 密集度	2.0371	1.6852	3.1439	2.9735	能量
對數音量	14.0162	7.6677	21.3500	18.8418	能量
音色	2.8020	1.0549	19.1541	13.6573	能量
調性	大調居多	小調居多	大小調各半	皆為大調	壓力
和聲和諧度	0.8574	0.7536	4.8026	2.8429	壓力

由表 5.1.6 可以看到各種音樂特徵值對於各個類別之音樂的差異，結果顯示對於

能量的分析，音樂事件密集度、音量與音色三種特徵值皆有相同的趨勢，最大直皆為焦慮類型的音樂，振奮類型的音樂次之，舒適的音樂第三，而哀傷的音樂為最小。在壓力的分析上，調性追蹤對於焦慮和哀傷的音樂判定為小調居多，舒適與振奮的音樂則以大調為主。而和聲和諧度無法突顯出壓力的差異，反倒是和能量相關的特徵值有相同的趨勢，結果顯示和聲和諧度並無法準確表示壓力的差異。

5.2 系統展示

本節展示系統計算情緒轉折較為明顯的音樂之結果，本次範例歌曲為英國樂團 Breed 77 翻唱之 Zombie，原唱為 The Cranberries(小紅莓樂團)。選擇本音樂作為範例之原因為此曲段落分明，圖 5.2.1 為此曲的時域波形，並附有各個段落的樂器編制。

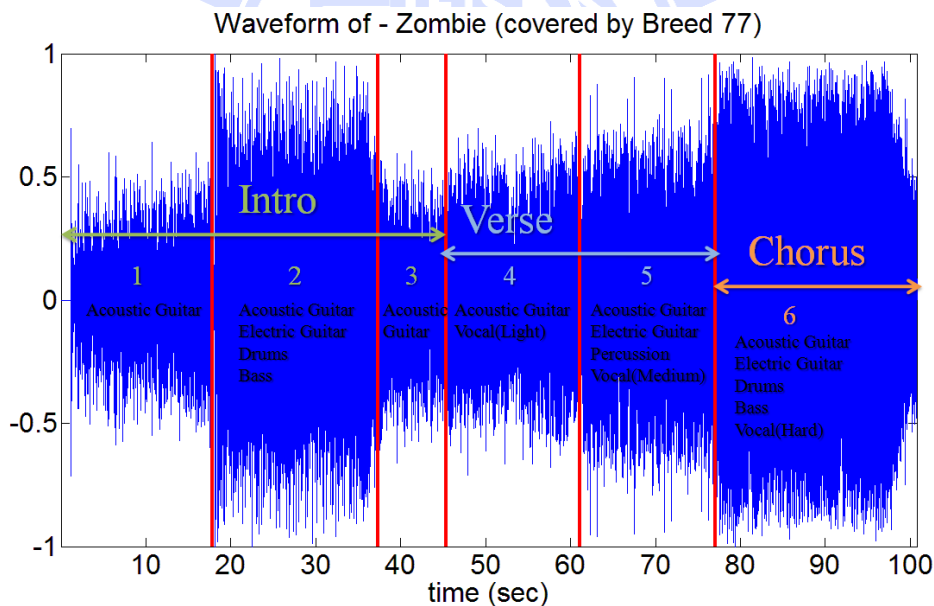


圖 5.2.1 Zombie-時域波形與樂器編制

圖 5.2.1 可以看到 Zombie 從 0 至 100 秒之時域波形與樂器編制，總共有 6 個明顯的段落，段落 1 到 3 為前奏的部分，第 1 段為木吉他刷和弦節奏，第 2 段加入了電吉他、鼓與貝士，第三段類似於第 1 段。段落 4、5 為主歌的部分，第 4 段為木吉他刷和弦節奏與音量適中的歌唱，第 5 段加入了電吉他、打擊樂器且歌唱

聲音變的稍微高亢一些。第 6 段為副歌的部分，較前段加入了鼓與貝士，歌唱部分變成激烈的嘶吼。

系統介面於各個段落尾端時的軌跡可以參考圖 5.2.2 至圖 5.2.7。可以看到第 1 段的情緒較為舒適偏向振奮；第 2 段則因為鼓和電吉他的聲音使情緒偏向焦慮與振奮；第 3 段回到舒適區且前奏結束；第 4 段為主歌的部分較為舒適；第 5 段末準備進入副歌情緒偏向振奮與焦慮；第 6 段副歌進行中，情緒高漲位於焦慮區域。系統實際動畫請至交通大學聲音與音樂創意科技碩士學位學程網頁觀賞。

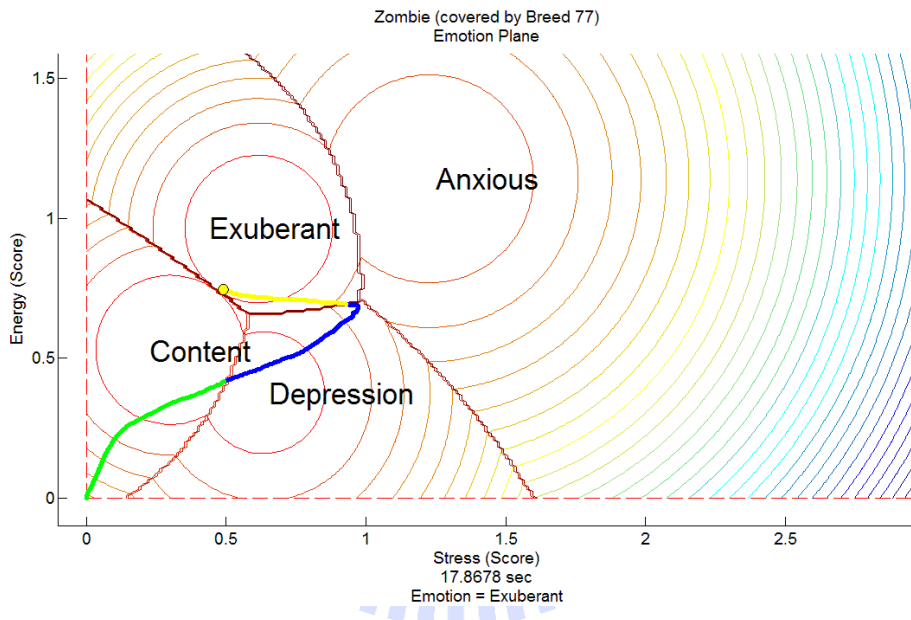


圖 5.2.2 Zombie-第 1 段末之情緒軌跡變化

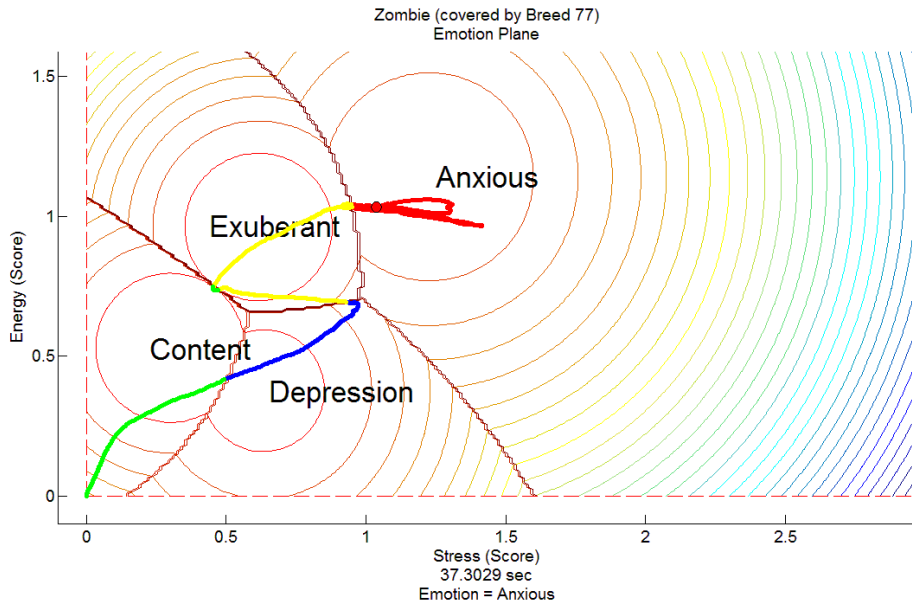


圖 5.2.3 Zombie-第 2 段末之情緒軌跡變化

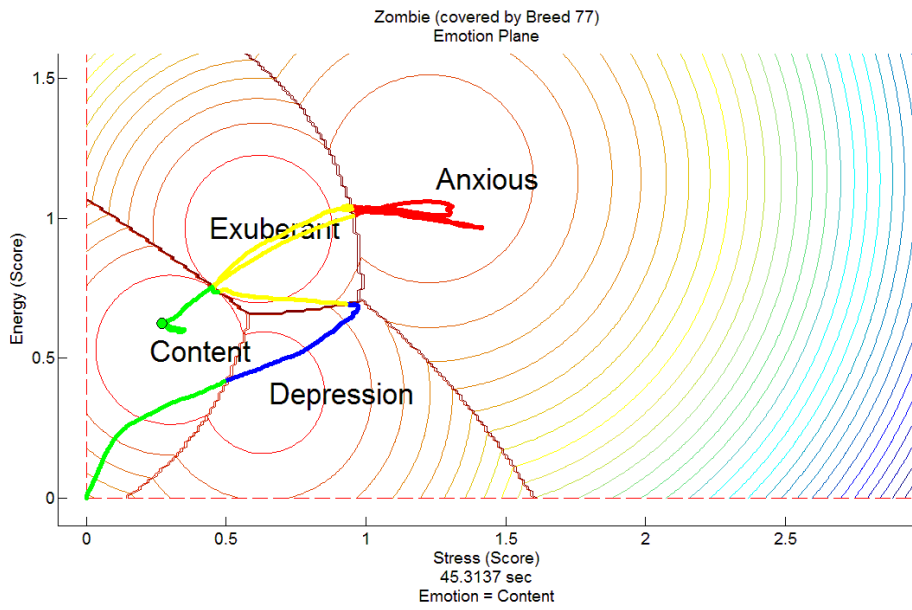


圖 5.2.4 Zombie-第 3 段末之情緒軌跡變化

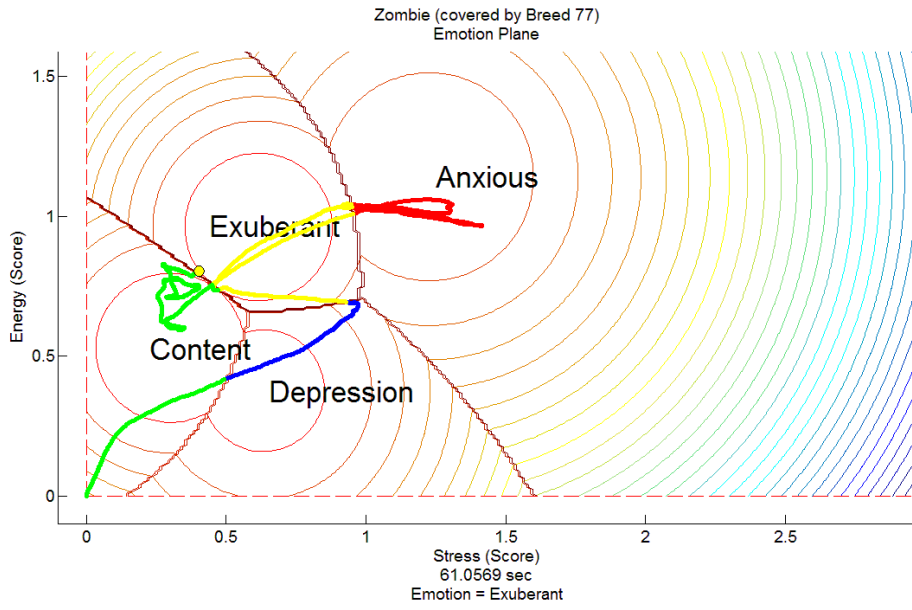


圖 5.2.5 Zombie-第 4 段末之情緒軌跡變化

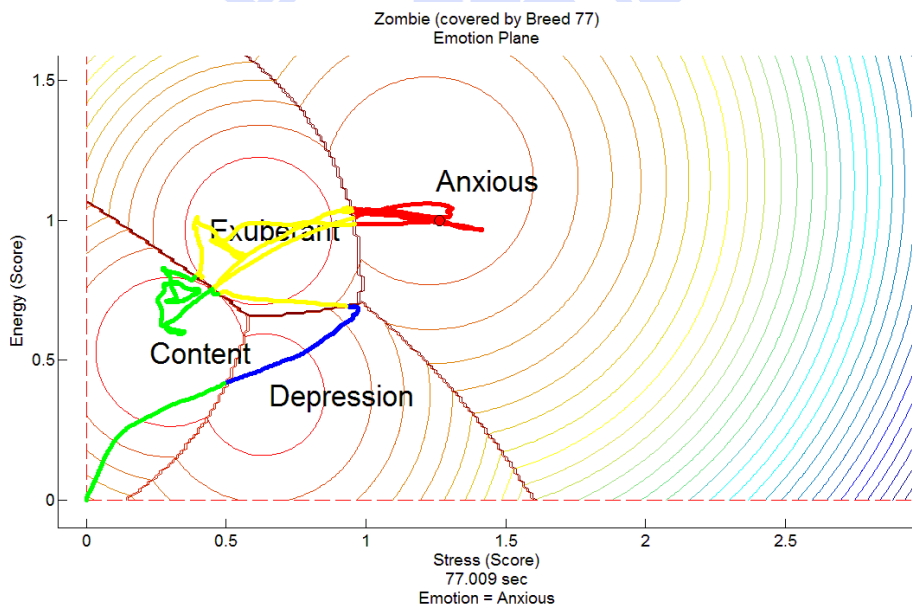


圖 5.2.6 Zombie-第 5 段末之情緒軌跡變化

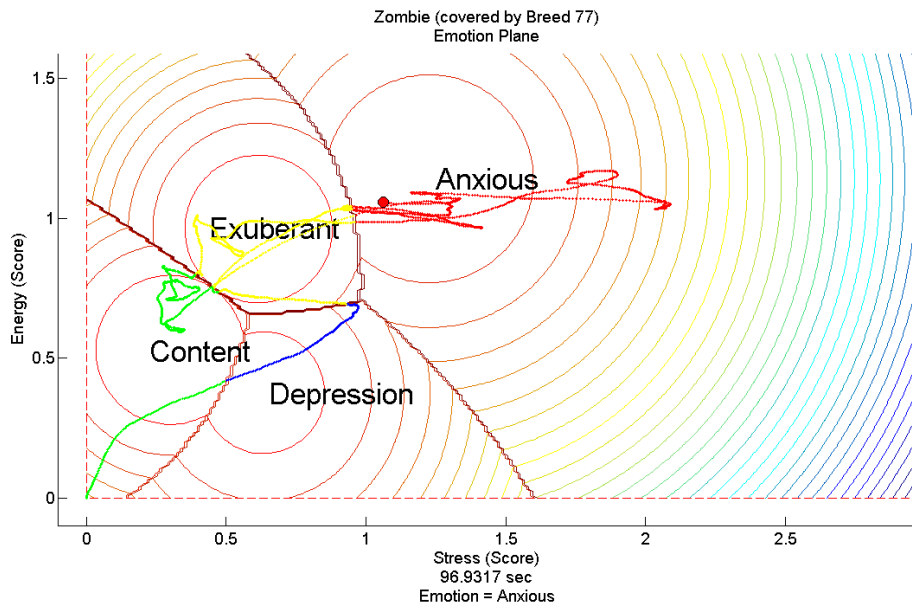


圖 5.2.7 Zombie-第 6 段末之情緒軌跡變化



六、結論與未來展望

在往後的系統中，將會考量更多足以影響情緒的音樂特徵，並且實作於智慧手機應用程式上，配合即時的音樂情緒軌跡互動與概要情緒辨識功能，將帶給使用者更深刻的音樂聆聽經驗。目前也有學者正在進行腦波與音樂情緒的相關研究，並使用類似於 Thayer 的情緒模型，往後若能得到更具信服力的研究結果，將可以利用其結果對本系統進行最佳化，使的各個特徵值對應的情緒分數能夠更精準並且貼近人類的感知。

配合及時的情緒的變化，人與電腦的創意互動也將增加相當大的可能性，例如在舞台劇中，可以配合音樂情緒變化來控制舞台燈光的變化，使觀眾經歷更有趣的鑑賞經驗。在音樂檢索的相關應用中，反向的應用目前也逐漸有學者投入研究，如利用節奏追蹤系統來控制燈光與舞台表演者互動，往後或許能依照該情緒的特有模式來進行自動作曲，即自動作出帶有特定情緒的音樂，這也是相當值得研究的方向。

七、參考文獻

- [1] 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.
- [2] W. Chai and B. Vercoe., “Detection of key change in classical piano music,” In ISMIR, London, 2005.
- [3] Kyogu Lee, “Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile,” In Proc. Intl. Computer Music Conf. (ICMC), New Orleans, USA, 2006.
- [4] 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, 14(1): 5–18, January 2006.
- [5] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su y Homer H. Chen. “Music Emotion Classification: A Regression Approach,” In Proc. IEEE Int. Conf. Multimedia Expo 2007 (ICME'07), Beijing, China, pp. 208-211.
- [6] Foote, J., “Visualizing music and audio using self-similarity,” In Proc. ACM Multimedia 99, Orlando, FL, pp 77-80, 1999.
- [7] Foote, J., “Automatic Audio Segmentation using a Measure of Audio Novelty,” in Proc. ICME 2000.
- [8] Foote, J. and Uchihashi S., “The beat spectrum: a new approach to rhythm analysis,” In Proceedings of the 2nd International Conference on Multimedia and Expo (ICME'01), IEEE, 2001.
- [9] Y.-H. Yang et al, “Mr. Emo: Music retrieval in the emotion plane,” Proc. ACM MM, 2008, accepted.
- [10] Cyril Laurier and Perfecto Herrera, “Mood Cloud: A Real Time Music Mood,” Music Technology Group Universitat Pompeu Fabra, 2008.
- [11] Hevner, K., “Experimental studies of the elements of expression in music,” American Journal of Psychology, 48, 248-68, 1936.

- [12] Lazarus, R.S., "Progress on a cognitive-motivational-relational theory of emotion," *American Psychologist*, 46, 819-834, 1991b.
- [13] Thayer, R.E. "The Biopsychology of Mood and Arousal," Oxford University Press, NY, 1989.
- [14] A. Tellegen, D. Watson and L.Clark., "On the dimensional and hierarchical structure of affect," *Psychological Science*, 1999.
- [15] R. Reisenzein and W. Schonpflug, "Stumpf 's Cognitive-Evaluative Theory of Emotion," *American Psychologist*, 1992.
- [16] Meyers, O.C., "A Mood-Based Music Classification and Exploration System," MS Thesis, Massachusetts Institute of Technology (MIT), USA, 2007.
- [17] 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.
- [18] Dixon, S. (2001b). "An interactive beat tracking and visualisation system," In Proceedings of the International Computer Music Conference, pages 215–218, San Francisco CA. International Computer Music Association.
- [19] C. Duxbury, M. Sandler, and M. Davies, "A hybrid approach to musical note onset detection," in Proc. Digital Audio Effects Conf. (DAFX,'02), Hamburg, Germany, 2002, pp. 33–38.
- [20] Andre Holzapfel and Yannis Stylianou, "Beat tracking using group delay based onset detection," in MIREX at 7th International ISMIR 2008 Conference, 2008.
- [21] V. Kandia and Y. Stylianou, "Detection of clicks based On group delay," Accepted in *Canadian Acoustics*, 2008.
- [22] Simon Dixon, "Mirex 2006 audio beat tracking evaluation: Beatroot," in MIREX at 7th International ISMIR 2006 Conference.
- [23] 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.
- [24] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, and C. Uhle, "An experimental comparison of audio tempo induction algorithms," *IEEE Trans.*

Speech and Audio Proc., vol. 14, no. 5, 2006, to appear.

- [25] Hanna Järveläinen, Vesa Välimäki, “Audibility of initial pitch glides in string instrument sounds,” Proceedings of the International Computer Music Conference, Havana, Cuba, 17-23 September 2001, 282–285.
- [26] Thomas D. Rossing, Richard F. Moore, Paul A. Wheeler, “*The science of the sound*” 3rd Ed.
- [27] 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>.
- [28] 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.
- [29] Krumhansl, C.L., “*Cognitive Foundations of Musical Pitch*,” Oxford Psychological Series, no. 17, Oxford University Press, New York, 1990.



附錄 A- 訓練資料明細表

#	曲名	作者/演出者	人工標記 情緒類別
1	Over the Rainbow	藤田惠美	1
2	Over the Rainbow	藤田惠美	1
3	Don't know Why	Norah Jones	1
4	Don't know Why	Norah Jones	1
5	Melody Fair	藤田惠美	1
6	Melody Fair	藤田惠美	1
7	Depois do Natal	小野麗莎	1
8	Depois do Natal	小野麗莎	1
9	All My Loving	藤田惠美	1
10	Jingle Bell Rock	小野麗莎	1
11	White Christmas	小野麗莎	1
12	White Christmas	小野麗莎	1
13	White Christmas	小野麗莎	1
14	Paz Azul	小野麗莎	1
15	Paz Azul	小野麗莎	1
16	The Christmas Song	小野麗莎	1
17	The Christmas Song	小野麗莎	1
18	Boas Festas	小野麗莎	1
19	Boas Festas	小野麗莎	1
20	Um Anjo Do Ceu	小野麗莎	1
21	Um Anjo Do Ceu	小野麗莎	1
22	The Wee Small Hours of the Morning	小野麗莎	1
23	The Wee Small Hours of the Morning	小野麗莎	1
24	Leaving on a Jet Plane	藤田惠美	1
25	Leaving on a Jet Plane	藤田惠美	1
26	Silent Night	小野麗莎	1
27	Not Going Anywhere	Keren Ann	1
28	Not Going Anywhere	Keren Ann	1
29	Polly	Keren Ann	1
30	Polly	Keren Ann	1

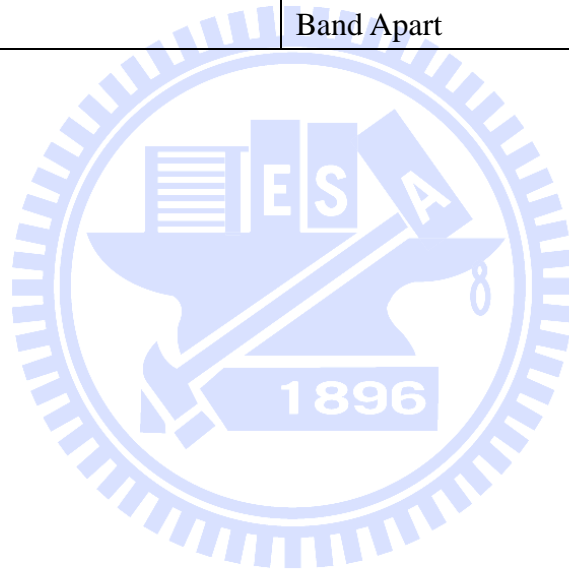
31	Sit on the Sun	Keren Ann	1
32	Right Now and Right Here	Keren Ann	1
33	Right Now and Right Here	Keren Ann	1
34	Seventeen	Keren Ann	1
35	Seventeen	Keren Ann	1
36	Stephanie Says	Keren Ann	1
37	Stephanie Says	Keren Ann	1
38	In My Life	藤田惠美	1
39	In My Life	藤田惠美	1
40	Moonglow	Noon	1
41	Moonglow	Noon	1
42	I Like to Recognize the Tune	John Pizzarelli	1
43	Close to You	Noon	1
44	Mountain Greenery	John Pizzarelli	1
45	Moon River	Fried Pride	1
46	Moon River	Fried Pride	1
47	Fun Out of Love	Noon	1
48	沒有你的世界	河野伸	2
49	我的煩惱	河野伸	2
50	9 Crimes	Damien Rice	2
51	A Summer's Day	久石讓	2
52	A Summer's Day	久石讓	2
53	Resphoina	久石讓	2
54	記得	鄭智薰	2
55	Symmetry	Mew	2
56	The Following Day	Jacques Stotzem	2
57	All I Need	Radiohead	2
58	Red Sky	David Gilmour	2
59	Red Sky	David Gilmour	2
60	The Hill	Marketa Irglova	2
61	The Hill	Marketa Irglova	2
62	The Hill	Marketa Irglova	2
63	View of Silence	久石讓	2
64	View of Silence	久石讓	2
65	Le dessous des cartes	Jacques Stotzem	2
66	Accidental Babies	Damien Rice	2
67	空房間	The Daydream	2

68	Summer	久石讓	2
69	Summer	久石讓	2
70	Vent D'Ailleurs	Jacques Stotzem	2
71	Gloomy Sunday	Loreena McKennitt	2
72	Sleep Don't Weep	Damien Rice	2
73	Secret Garden	Jacques Stotzem	2
74	Secret Garden	Jacques Stotzem	2
75	Walking in the Air	藤田惠美	2
76	Secret Garden	Jacques Stotzem	2
77	Legend of Wind	久石讓	2
78	Fragile Dream	久石讓	2
79	Fragile Dream	久石讓	2
80	Paun Haung	久石讓	2
81	寂靜	鄭智薰	2
82	Moon light	Beethoven	2
83	Who am I	FF7 OST	2
84	End of May	Keren Ann	2
85	By the Cathedral	Keren Ann	2
86	Ending Song	Keren Ann	2
87	Ending Song	Keren Ann	2
88	Do What I Do	Keren Ann	2
89	Do What I Do	Keren Ann	2
90	Walk Real Slow	Keren Ann	2
91	The Morning After	Keren Ann	2
92	New Born	Muse	2
93	Space Dementia	Muse	2
94	Dark Shines	Muse	2
95	Paranoid Android	Radiohead	2
96	15 Steps	Radiohead	3
97	15 Steps	Radiohead	3
98	15 Steps	Radiohead	3
99	Muscle Museum	Muse	3
100	Cave	Muse	3
101	Cave	Muse	3
102	Cave	Muse	3
103	Showbiz	Muse	3
104	Showbiz	Muse	3

105	Showbiz	Muse	3
106	Assasain	Muse	3
107	Assasain	Muse	3
108	Assasain	Muse	3
109	Uno	Muse	3
110	Uno	Muse	3
111	City of Delusion	Muse	3
112	City of Delusion	Muse	3
113	City of Delusion	Muse	3
114	Sober	Muse	3
115	Sober	Muse	3
116	Sober	Muse	3
117	Knights of Cydonia	Muse	3
118	Knights of Cydonia	Muse	3
119	Knights of Cydonia	Muse	3
120	Stockholm Syndrome	Muse	3
121	Stockholm Syndrome	Muse	3
122	Stockholm Syndrome	Muse	3
123	Thoughts of a Dying Atheist	Muse	3
124	Thoughts of a Dying Atheist	Muse	3
125	Thoughts of a Dying Atheist	Muse	3
126	New Born	Muse	3
127	New Born	Muse	3
128	New Born	Muse	3
129	New Born	Muse	3
130	Space Dementia	Muse	3
131	Space Dementia	Muse	3
132	Hypermusic	Muse	3
133	Hypermusic	Muse	3
134	Hypermusic	Muse	3
135	Citizen Erased	Muse	3
136	Citizen Erased	Muse	3
137	Micro Cuts	Muse	3
138	Micro Cuts	Muse	3
139	Dark Shines	Muse	3
140	Paranoid Android	Radiohead	3
141	Electioneering	Radiohead	3

142	Electioneering	Radiohead	3
143	Electioneering	Radiohead	3
144	Sky Sky Sky	DEPAPEPE	4
145	Sky Sky Sky	DEPAPEPE	4
146	Canon	DEPAPEPE	4
147	Star	Youha	4
148	Star	Youha	4
149	Selfish Jean	Travis	4
150	Selfish Jean	Travis	4
151	Tall Fiddler	Tommy Emmanuel	4
152	Tall Fiddler	Tommy Emmanuel	4
153	When Your Mind's Made Up	Glen Hansard	4
154	Day Tripper	Beatles	4
155	Day Tripper	Beatles	4
156	I Can Do Better	Avril Lavigne	4
157	Electric de Chocobo	FF7 OST	4
158	Electric de Chocobo	FF7 OST	4
159	Fiddle de Chocobo	FF7 OST	4
160	Fiddle de Chocobo	FF7 OST	4
161	Highwind	FF7 OST	4
162	Ride on	FF7 OST	4
163	Ride on	FF7 OST	4
164	Kurumi	Mr. Children	4
165	Kurumi	Mr. Children	4
166	masa Ume	Spitz	4
167	masa Ume	Spitz	4
168	Spring Song	Spitz	4
169	Spring Song	Spitz	4
170	Summer Paeade	DEPAPEPE	4
171	Wherever You Will Go	The Calling	4
172	Wherever You Will Go	The Calling	4
173	New World	L'arc en Ciel	4
174	New World	L'arc en Ciel	4
175	Stay Away	L'arc en Ciel	4
176	Stay Away	L'arc en Ciel	4
177	Free Eat	L'arc en Ciel	4
178	Free Eat	L'arc en Ciel	4

179	Star Chaser	Spitz	4
180	Star Chaser	Spitz	4
181	Am I Wry No	Mew	4
182	Eurostar	T Square	4
183	Eurostar	T Square	4
184	Fightman	T Square	4
185	Bird land	Weather Report	4
186	Bird land	Weather Report	4
187	The Man in the Green Shirt	Weather Report	4
188	Jerry's Breakdown	Peppino D'Agostino	4
189	Tennessee Rag	Chet Atkins	4
190	Higher	The Band Apart	4
191	When You Wish Upon a Star	The Band Apart	4
192	Band Apart	Band Apart	4



附錄 B- 不同高斯函數個數的 GMM 訓練結果

高斯函數個數：2

Inside-Test 辨識率：80.0595%

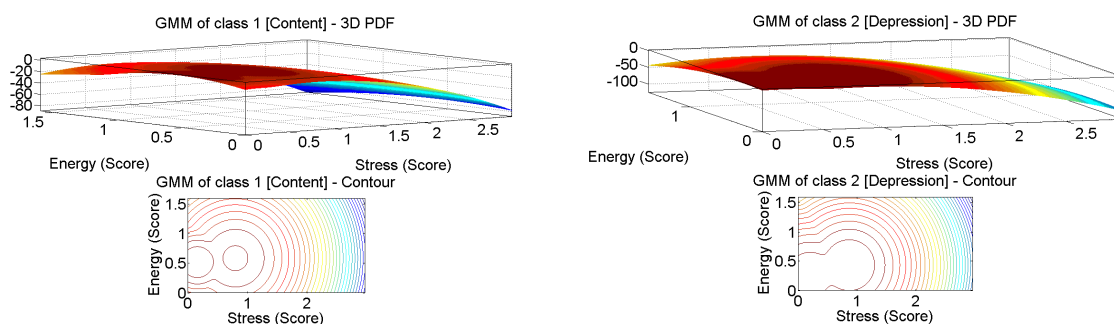


圖 B.1 高斯函數個數 2-類別 1&2 分佈

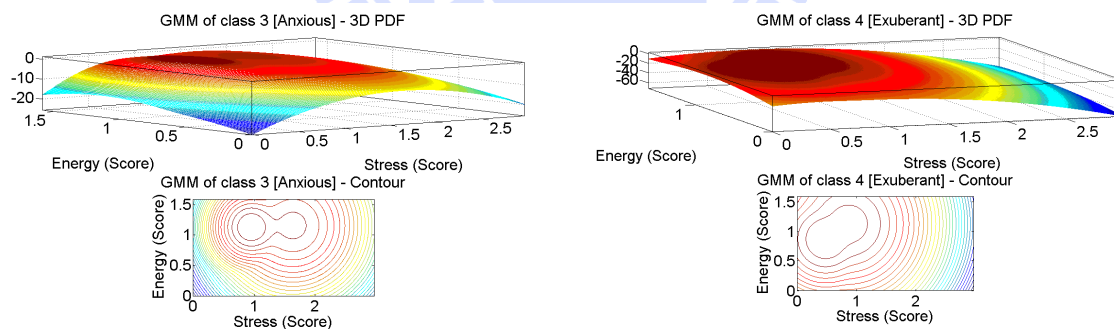


圖 B.2 高斯函數個數 2-類別 3&4 分佈

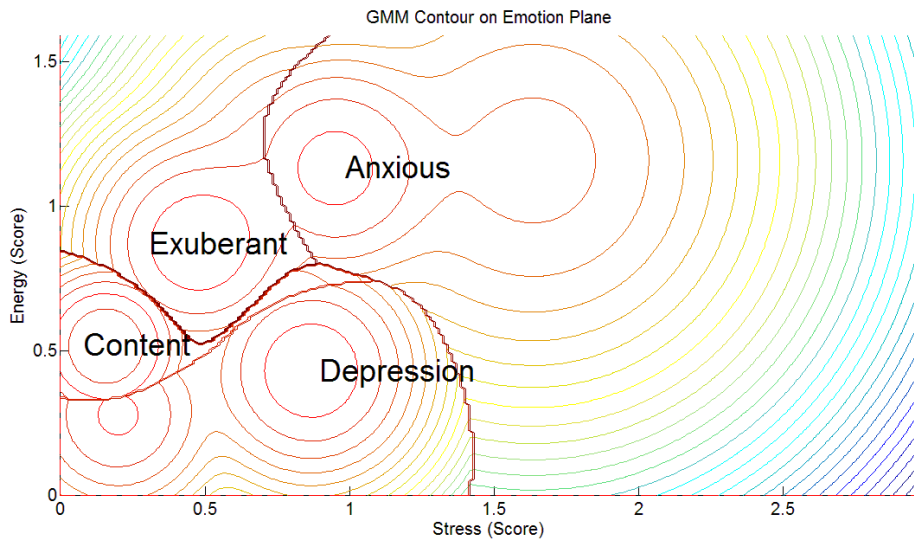
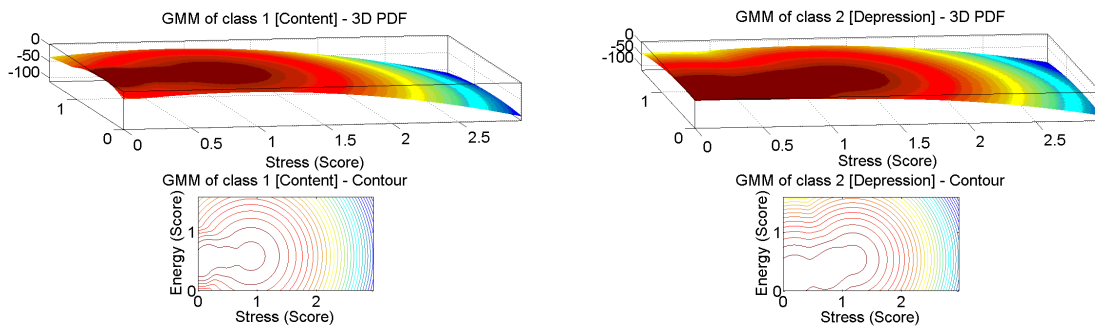


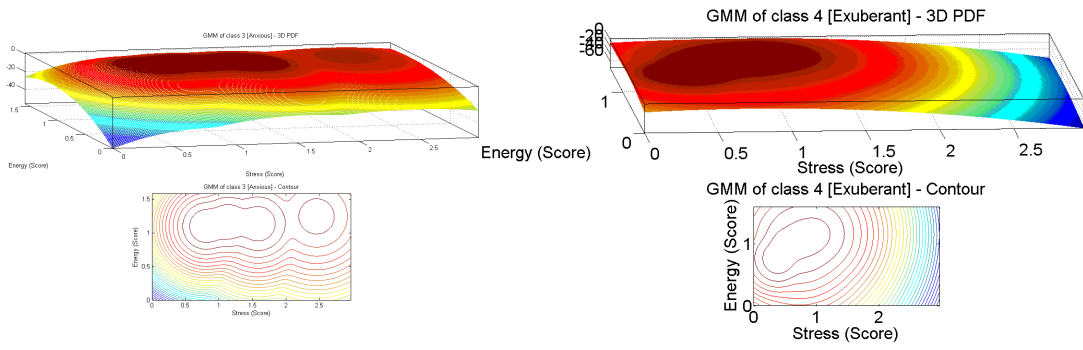
圖 B.3 高斯函數個數 2-所有類別邊界

高斯函數個數：4

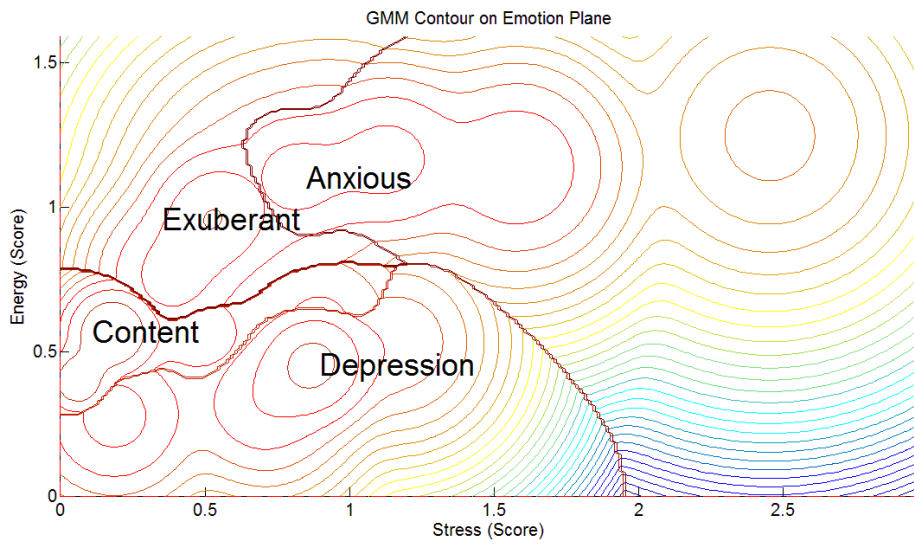
Inside-Test 辨識率：81.8452%



圖B.4 高斯函數個數4-類別1&2分佈



圖B.5 高斯函數個數4-類別3&4分佈



圖B.6高斯函數個數4-所有類別邊界

附錄 C- 查準率與查全率之數學定義

在評估資料檢索的方式中，常利用查準率(Precision)與查全率(Recall)來衡量系統的效能，以下為其數學定義：

$$\text{Precision} = \frac{\text{Relevant Documents Retrieved}}{\text{Total Retrieved Documents}}$$

$$\text{Recall} = \frac{\text{Relevant Documents Retrieved}}{\text{Total Relevant Documents}}$$

其中 Relevant Documents Retrieved 為檢索系統所找到正確的資料數目，Total Retrieved Documents 為檢索系統總共回傳的資料數目，Total Relevant Documents 則為被檢索的資料庫中實際上符合檢索條件的資料數目。也就是說查準率為評估系統回傳資料的正確比例，而查全率為評估系統實際上能找到的正確資料數目。