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

基於多重結構分析聆聽情緒相似度之音樂資訊檢索

A Music Linkage Jukebox based on Multi-Structure Analysis of Music Emotion Similarity

研究生：林芷伊

指導教授：鄭泗東 教授

中 華 民 國 一 百 零 一 年 六 月

基於多重結構分析聆聽情緒相似度之音樂資訊檢索

A Music Linkage Jukebox based on Multi-Structure Analysis of Music Emotion Similarity

研 究 生：林芷伊　　Student：Chih-Yi Lin

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 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年六月

基於多重結構分析聆聽情緒相似度檢索之音樂心情點唱機

學生：林芷伊　　指導教授：鄭泗東

國立交通大學 聲音與音樂創意科技碩士學位學程

摘　　要

作曲家利用音符轉述傳達自己的想法來譜寫音樂作品，藉由音符連續不斷的變化構成音樂的主題，多個音樂主題組合產生一段主要旋律，希望使聆聽者在聆聽此音樂片段時有相似的情緒感受並快速地為聆聽者留下印象深刻、難以忘懷的聆聽經驗。許多音樂情緒分類或辨識的研究將聆聽音樂所產生的情緒感受總結為音樂帶給聆聽者的”心情”。在曲式結構中將許多的音樂主題(主歌與副歌)搭配過門音樂做重複性地些微變化串起來譜成完整的音樂，本論文以樂曲訊號之多重主題結構分析為基礎，提出一套基於聆聽情緒相似之音樂檢索系統，協助聆聽者快速地從音樂資料庫中選擇相似聆聽情緒之音樂檔案，並降低音樂資料多重特徵檢索對記憶體的使用量。本系統主要分為多重主題結構分析、音樂情緒比例分析、音樂情緒檢索等三個部份：首先，利用自相關函數(autocorrelation function)分析多重主題的音樂結構，包括前奏(Intro)、主歌(Verse)與副歌(Chorus)等段落。在音樂情緒比例分析方面，引用 Thayer 提出的情緒模型，將兩百首註有人工標記情緒類別的音樂片段進行特徵萃取與情緒記分，以高斯混合模型(GMM)進行訓練並劃定舒適、哀傷、焦慮與振奮等四個情緒類別的邊界。接著利用此多重主題結構組成的音樂片段做為音樂情緒辨識的測試樣本，計算該音樂所喚起的聆聽情緒比例，最後以距離相似度量測演算法計算任兩段音樂片段之間的情緒相似成分，結果得出並依序列出其聆聽情緒與此檢索音樂片段相似的音樂檔案。系統輸出的使用者介面同時提供此檢索歌曲以及推薦清單中所選歌曲的靜態情緒比例，方便使用者在聆聽歌曲以前快速了解該音樂檔案誘發的聆聽情緒。

A Music Linkage Jukebox based on Multi-Structure Analysis of Music Emotion Similarity

student：Chih-Yi Lin Advisors：Dr. Stone Cheng

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

ABSTRACT

Key melodies are the representative fragments of music which may be the themes that people may easily recall once they heard and that breed a pleasurable and memorable listening experience. This study proposes a music linkage jukebox system that recommends listeners a ranked retrieval list with the proportion of music-induced emotions between the query and music bank collections. 200 music clips with emotion-predefined trained to build up the emotion plane, which demarcates the boundaries of four emotions by Gaussian mixture model. In the system, the multi-theme phrases of musical structure, including the Intro, Verse, and the Chorus are analyzed by autocorrelation function as input test structure, then using feature-weighted scoring algorithms to analyze the ingredients of music emotion with five audio feature sets, which represent the characteristics of the testing music clips. The similarity of emotions between music clips are measured by Euclidean distance algorithms. The outputs of the user-interface not only ranks the resembling music files but also offers a static graph with the proportion of music emotion, which can aid user rapidly in understanding the relationship between music-induced and emotions.

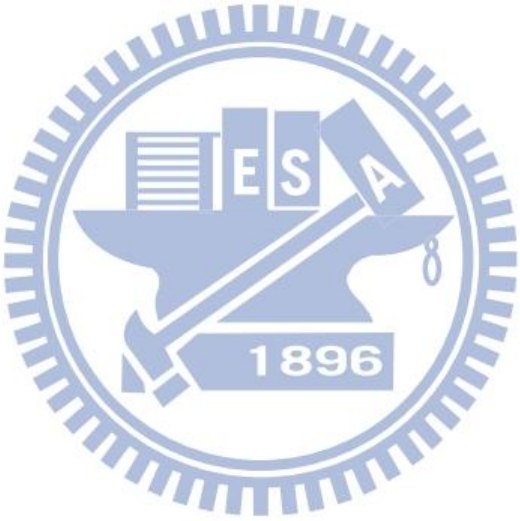
Keywords: Music information retrieval, emotion similarity, music summery, emotion ingredients.

誌　　謝

首先要感謝指導教授鄭泗東老師這兩年來的細心教導與鼓勵，讓我在研究挫折中快速地恢復信心、解決難題；實驗室學長姐(俊傑學長、雲凱學長、于恬學姐、立瑋學長、偉廷學長等)傳授的論文資料與程式資料庫，使我能夠在論文的研究上得到豐富的知識與支援。在這段學習過程中，不但加深了對研究領域的認知與根基，更培養自己在面臨問題時的思考、解決能力，使我得已順利完成畢業碩士論文，並取得碩士學位。

除了老師之外，亦感謝聲音學程的所有好夥伴(小婷、小愛、船長、紀子、哲瑋、小單、姚頭、阿杜、致偉、偉桓、坤廷、楊昕、欣諭)，不管在課業方面還是研究上都不吝嗇的提供我眾多的想法、寶貴的建議與協助，以及實驗室的學長、同學、學弟、學妹們(丞哥、奇穎、阿宏、慧珊、翔翔、婕安、小竹子、歆萍)每天的陪伴與關心，總是叮嚀我要記得吃飯，最後還要感謝一路支持、陪伴我的父親、母親、哥哥、好友們，在我失落無助的時候給予鼓勵與包容，讓我可以繼續奮鬥下去。

目　錄



表目錄

圖目錄

一、 緒 論

1.1 研究動機

音樂有如世上最美的語言，沒有國界、地域或族裔之分，人們藉由音樂來抒發低落的情緒或傳達喜悅的心情，音樂在生活中的重要性也同時地反應在情緒的反應之上，不同的音樂會帶給人們不同的情緒感受，經由細緻的音樂、溫暖的聲音或美妙的韻律把內心深處情感世界特有的激動化為自由自在的自我傾聽，使我們心靈免於壓抑和痛苦。而每個人對音樂的感覺是主觀性的，即使是處於相同的情境也會因為接觸的時代、社會背景及環境的不同而有所不同，更會隨著個人當下情緒的低落、亢奮、愉快而有所變更，因此如何幫助使用者從大量的音樂資料中與多變的情境下快速的有效找出符合自己情緒感受需求的音樂，善用音樂的情感特性釋放情緒、轉換心情成為本研究的主要目的。

隨著資訊科技的發展與通訊技術的進步，使得數位音樂的取得越來越容易，人們不再需要依照音樂專輯編排的順序撥放音樂，當一張唱片撥放完畢時不再需要以手動的方式將唱片包含黑膠唱片、錄音帶、CD 等音樂載體放入音樂播放器才能繼續聆聽音樂，取而代之的是人們每天可以簡單地透過攜便式音樂播放器、智慧型手機的音樂播放軟體等即時性地盡情享受音樂，或經由網路的線上播放系統收聽音樂，數位的聆賞方式取代傳統的使用習慣，這樣的作法提供音樂聆賞者更多創意發揮的空間，可以依照個人喜好或不同的需求自行編排曲目順序聆聽音樂。

雖然網際網路的成熟與數位科技的發展帶來了無窮的便利性與多功能性，其間龐大的音樂檔案數量卻也同時對使用者在整理、管理檔案上產生相當程度的困擾。當使用者在搜尋挑選音樂時，其比對成千上萬首音樂檔案的過長等待時間往往讓使用者無法接受，若能夠將音樂檔案如同文章依據主題利用段落的方式分開，使用者再也不需要將整首音樂檔案從頭聽過才可以瀏覽到所需要的音樂資料，音樂資料分段的好處不止於提供使用者能夠輕易地找到符合所需的音樂資訊，更可以利用分段的結果產生音樂內容的摘要或稱音頻縮略圖(Summary or Audio Thumbnailing)作為以內容為基礎的音樂資料檢索，幫助使用者達到搜尋音樂的目的。音頻縮略圖或音樂內容摘要主要用於概括音樂的

資料，通常是因樂檔案中最讓人印象深刻的音樂片段，其所生成的內容摘要或縮略圖可以幫助我們管理音訊檔案，方便瀏覽或搜尋音頻資料，減輕收聽此類音檔較長部份的問題。對於傳統的檢索方法大多採用關鍵字搜尋的方式，例如：曲名、唱片名稱、演唱者、作曲者、音樂類型、唱片廠牌等等，就音樂資料來說，如果使用者只記得一首歌的某段旋律，而不記得歌名或歌手是誰，就沒辦法找到想要的歌曲了。因此，在這種情形下可以利用音樂內涵式搜尋的方式對這段音樂進行特徵值分析，找出在音樂資料庫中最有可能包含此音樂片段之音樂，如此一來即使我們無法對該音樂下關鍵字，系統也可以依照音樂本身之特性進而完成搜尋的工作。

全球的音樂每天不斷推陳出新，使得音樂資料庫的成長十分驚人，由於音樂屬於時序性的，加上複雜的聲音資料集合，在進行檢索比對的動作通常需要耗費很多的計算時間以及記憶體的用量。為了加快檢索速度，本研究之目的在於發展一套基於多重結構分析聆聽情緒相似度檢索之音樂心情點唱機，針對音樂的內容做主題式的分段來簡化分析過程的複雜度，擷取此多重主題結構的音樂片段來代替完整的音樂作品，並分析此音樂片段的音訊特徵做為音樂資料庫中音樂情緒的指標，如此，在搜尋時只需要比對四種音樂誘發的情緒比例，能大大節省儲存空間外，更能從龐大的資料庫中進行更有效率的查詢。

1.2 系統之理論基礎與相關研究

此章節將介紹本研究內容所涉及之音樂情緒檢索、音樂分段、音樂情緒模型、聆聽音樂誘發的情緒感受以及音訊分析等相關理論基礎與其相關文獻探討。

1.2.1 內涵式音樂資訊檢索

由於多媒體技術的快速進展，音樂創作的普及，使得數位音樂的取得越來越容易，各式各樣的音樂資料變得更加複雜及大量，音樂資料也不再是如同過去以書目資料的形式提供查詢、取得與利用，使用者如何從大量的音樂資料庫中，找出自己喜好的音樂之技術是日益重要的。過去，傳統的檢索方式將音樂資料、組織、分析以作曲者、演奏者、曲風、專輯名稱等項目分門別類，進而建立音樂書目資料庫，使用者依照這些項目雖可

檢索到資料，然而此種檢索的彈性仍然有限。假設當我們聽到某廣播電台播放的音樂片段是自己喜愛的音樂旋律，但對於第一次聽到，不知其曲名、演奏者等詳細資訊時就無法以這些書目資料查得音樂原件，而內涵式音樂資訊檢索即是提供使用者解決此類相關搜尋問題的技術，利用資料本身的特徵去找出使用者想要的資料。

內涵式音樂資訊檢索依據檢索資料類型的不同可分為由(1)符號資料搜尋(Search by symbolic data)和(2)音頻資料搜尋(Search by audiodata)等兩大類[1]。符號資料指的是儲存音樂符號的檔案格式，例如：MIDI、XML，在特徵萃取的過程中可直接取得其音高、節拍、速度、音色的訊息，經由特定演算法的運算找出音樂的旋律、調性、節奏等音樂特徵；音頻資料搜尋(search by audio data)的資料庫則以完成錄音及混音後之聲波波形的檔案格式，經由特定演算法從波形大小中計算出音樂的訊息或聲學特徵後進而得到音高、節拍、速度、音色等音樂特徵，例如：wav、wma、mp3 等都是一般常被用來聆聽的檔案格式。

MIDI 格式儲存多樣性的音樂特徵，記錄了各種音符的強弱、高低、長短等特性來記錄音樂資料，如此多樣性的音樂資訊將有助於音樂搜尋研究上的處理，加上以符號記錄音樂資訊的 MIDI 格式音樂檔案可以很容易地抽取出主旋律特徵[2][3]，所以將 MIDI 音樂檔案轉成音樂序列用來當作音樂檢索的音樂資料是許多內涵式音樂檢索研究者所採取的方式，因此有許多音樂搜尋相關研究是以 MIDI 為主要的音樂格式，[4][5]利用音樂的低階特徵值並同時考慮音樂時序的意義，透過索引結構的建立讓音樂檢索快速地被處理；然而若考慮其現實應用層面的問題，Park 等人[6]以 MP3 格式的音檔萃取多種音樂特徵包含 spectral centroid, spectral rolloff, spectroll flux, zero crossing rates, MFCC 等來表現音樂，接著利用 Feature selection 的方式將原本高維度的特徵值減低，進而利用這些特徵進行音樂搜尋；雖然 mp3 是最常被使用的音樂格式，不過由於 mp3 為失真壓縮，它不像 MIDI 格式般包含了多樣性的音樂特徵，使得如果以 mp3 格式來進行音樂搜尋會有一定程度的困難；Foote[7]這一篇研究中，選擇未經壓縮 wav 的檔案格式，著重在節奏(tempo)和旋律(rhythm)這二種音樂特性來進行研究，並且利用這二種特性來進行

音樂的相似性分析。

1.2.2 音樂分段-音樂主題在音樂資料中扮演的重要角色

創作音樂如同許多文學、藝術創作一樣，在創作過程中就像寫作文一樣，作家先有了靈感再根據寫作原理文章章法：起、承、轉、合的思維模型，有緒論、主題、副題、結論。音樂創作可以被歸納分成靈感、腹稿和動機等三個程序，作曲家在作曲時一樣依靠靈感，有了靈感之後再如同建築師蓋房子一樣擬一個草圖(腹稿)，逐步計劃，注意音樂環節緊湊、前後搭配銜接、決定表現方式，最後再由動機發展出音樂主題。動機是由靈感帶來的最原始主題、旋律的最小單位，創作的關鍵。作曲家遵循音樂曲式結構法則，由動機切入旋律的起點，安排主題的重覆、倒裝、變奏等連續變化，再想出一個或多個副主題拼湊，繼而將許多個主題串起來展成完整優美的音樂。多個主題使樂曲多樣化不單調，但過多的主題容易離題，造成聽覺混淆、缺乏主題的一貫性等問題，好的動機為旋律鋪墊了堅實的根基、發展出鮮明的主題，Lartillot[8]說明因為動機的豐富變化串成音樂主題前後結構的連貫，人們是因為動機的改變而記住音樂的前後結構變化，因此，本研究在此將每個段落假設為同一首音樂中的不同主題，以音樂動機的變化為線索，萃取樂曲中的主題(主歌)和副主題(副歌，一個或多個)作為系統分析的測試音樂片段。

目前已有許多相關音樂分段的研究，以人類聽覺認知對音樂訊號做分析來找出一首歌的重複樣式來代表一整個音樂物件，Cooper and Foote [9][10][11]首先將分析音檔作音頻參數化，統一音檔規格，取樣頻率 22.05KHz，單聲道，以 2048 個取樣點為音框(Frame)單位切割音樂訊號，然後利用一個代表音色之梅爾倒頻譜係數(Mel-scale Frequency Cepstral Coefficients, MFCCs)演算法擷取每一個音框的低階特徵值，接著進行自相關(Autocorrelation)的運算作為該音框所代表的音樂特徵，再以距離演算法來計算音框間的相似度，最後利用相似矩陣記錄音框比對的結果，從相似矩陣的資料中找出重複出現的音樂片段；[12][13][14][15]以人類聽覺認知對音高之敏感度為基礎，提出一種基於音調和重複結構劃分音樂的方法，並以這種重複結構來摘要音樂，[12]萃取如常數品質因數轉換(Constant Q Transform,CQT)，一個為了分析音樂訊號所開發之技術；[13]萃取以色

度(Chroma-Based)為基礎和音調相關(Pitch Related)的旋律特徵；為了減少比對每個音框的運算量、等待時間和記憶體用量，[14][15]提出以色度 Chroma-Based 特徵為基礎的直方圖轉換比對方法，色度直方圖指的是一個特定響度在特定音高種類中，其達到或超過此音高頻率的次數，為了加強找出局部音高的組成部分，利用快速傅立葉演算法(Fast Fourier Transformation, FFT)來計算每個時間點的瞬時頻率，將音高區分成 A A# B C C# D D# E F F# G G#等 12 種類，接著以相關係數演算法(Correlation Calculation)比較任兩個音框所屬的種類是否相同而產生一個二維空間的相似矩陣，進而將相似矩陣上的數據資料轉換映對至色度直方圖以測量音樂資料相似的部分。

1.2.3 音樂情緒模型

古人云：『樂者，心之動也。』亦即，音樂與人們內心的情感有著密切的關係，音樂融合人們的各種情感與情緒體驗；不同的音樂會帶給人們不同的情緒，不同的人對於同一首音樂的感覺與體驗也不盡相同，甚至同一個人對同一首音樂也會因為不同情境有著天壤之別的感受，根據 Music and Emotion[16]一書中的有關音樂與情緒的介紹，將現代的情緒模型分成(1)類別論，也有人稱離散論(discrete or categorical emotion theory)[17][18]和(2)維度論(dimensional model of emotion)[19][20][21]兩種說法。類別論基於基本情緒理論將情緒分門別類，認為情緒有所謂的基本情緒(Basic Emotion)，對每個類別給於一個音樂情緒的形容詞，例如：高興、生氣、悲傷、平靜，強調每個基本情緒之間沒有必然的相關性，是互相獨立的，並不會因為某個情緒稍作變化而影響到另一個情緒的改變；維度論認為情緒應該是連續性的，例如：特定的情緒狀態只是代表一個正向情緒到負向情緒，同一個維度的相反兩極上，或是從快樂到悲傷的連續體中的一個位置。採用幾個心理學研究上的維度(例如：正向度及激昂度)，建立出一個情緒空間，並將情緒以空間中的一點表示。

另外，也有學者將情緒心理學所提到的模型分成兩類：一類為一般通用型[19][20]，舉反心理、生理或認知上可能產生情感相關的情緒反應皆適用；另外一類即是和音樂表達情緒的類別與審美有關，針對音樂所引發的情緒反應模型[17][21]。

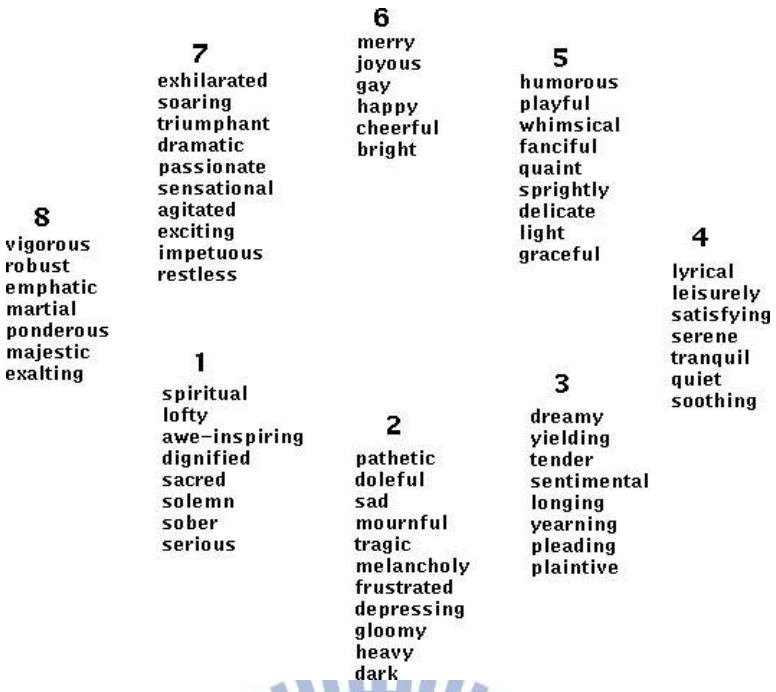


圖 1 Hevner's adjective circle 情緒模型

資料來源：[17]

Hevner[17]是類別論中最常見的情緒模型，也是最早被提出由音樂誘發的情緒模型，Hevner 從音樂學角度提出一個環狀的情緒模型，主要考慮作曲家、演奏家、聽眾的心理感受，於 1936 年設計了一系列音樂引發情緒的實驗，認為音樂本身便隱含著情緒意義，根據不同的音樂結構和音樂表情傳達各種情緒，假設人類聆賞音樂會引發不同的情緒反應，Hevner 透過此實驗，瞭解音樂的聲音和聽者的情緒反應間的關係，藉由實驗結果提出八組情緒相關的形容詞組(adjective group)，如圖 1，每一組代表性的形容詞分別為高貴的(dignified)、傷心的(sad)、悅耳的(dreamy)、平靜的(serene)、優美的(graceful)、快樂的(happy)、使人興奮的(exciting)、強而有力的(vigorous)，在 Hevner 之後陸續有許多學者發表相關研究，Farnsworth[18](1958)為了能從音樂訊號來分析情緒，將其重新細分為十類，而最近期的研究如 Zentner(2008)提出的階層式情緒模型 GEM-9，將四十種情緒依不同的權重彙總成九種情緒，最後再統合成三大類，為了確定每種情緒的主要因素，對與每一種單一的情緒指標給於 1～5 不同的評分。

(1:Not at all,2:Somewhat,3:Moderately,4:Quite a lot,5:Very much)。

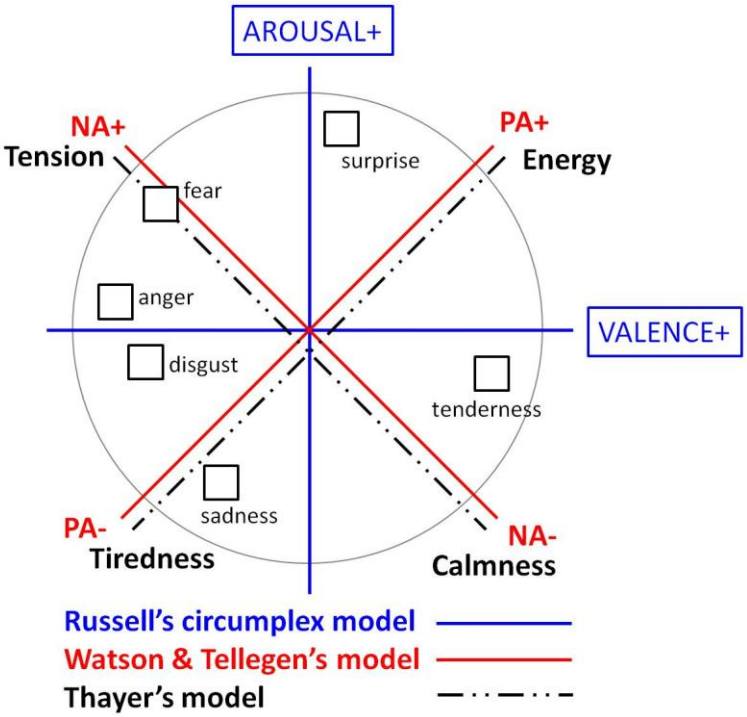


圖 2 各式二維情緒模型比較圖

資料來源：[16]

Juslin and sloboda 一書中[16]將 Russell[19]、Watson[20]和 Thayer[21]等三個情緒模型合併在同一個二維情緒平面上做討論，如圖 2 所示，Russell[19]根據情緒的愉悅程度(valence/pleasure)和激發程度(arousal/activation)兩個軸度來對情緒進行分類，認為各種情緒是以環狀的排列方式分佈在一個二維的向量空間中，如圖 3；Watson 與 Tellegen[20]在 1985 年提出一個以正向情感和負向情感為量測方法的階層式分類綱目(hierarchical taxonomic scheme)，將 Russell[19]所提出之情緒模型的兩軸旋轉 45 度後得到新的軸度：一個結合正價(valence)和高的激發程度(arousal)的正向情感（Positive Affective，PA）維度和一個結合負價(valence)和高的激發程度(arousal)的負向情感（Negative Affective，NA）維度，以此為基礎接著提出一個描述 20 種情緒的分類綱目稱為 PANAS(The Positive and Negative Affect Schedule)的心理模型，圖 4。

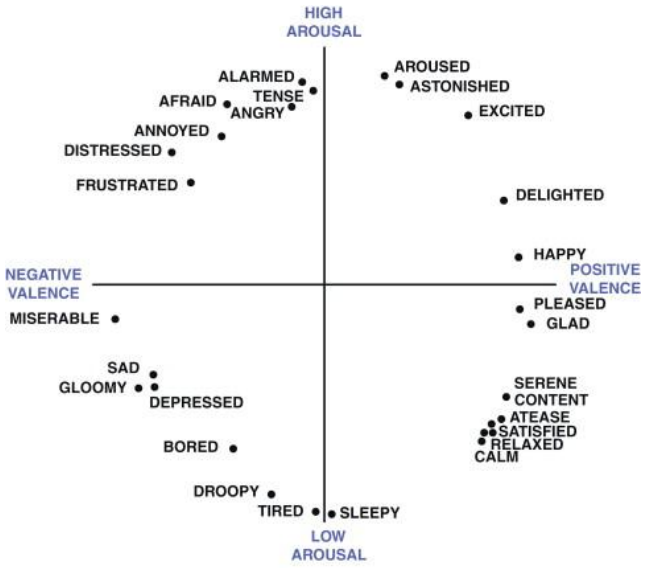


圖 3 Russell's Circumplex Model

資料來源：[19]

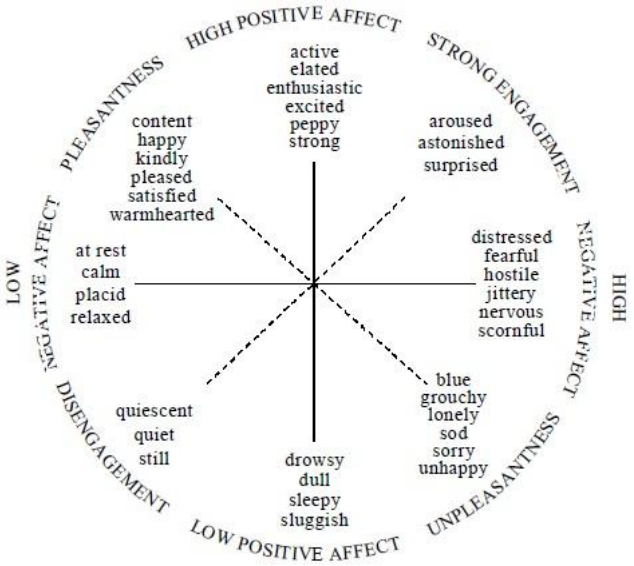


圖 4 Tellegen and Watson Clark 情緒模型

資料來源：[20]

Thayer[21]認為音樂的情感主要是受能量和壓力因素的影響，在 1989 年提出一個以二維空間為基礎的情緒模型（Model of Mood），構成此音樂模型的兩個主要因子：壓力(stress)和能量(energy)，其中壓力指的是快樂(happy)/焦慮(anxious)的程度，能量指的是

平靜(calm)/充滿活力(energetic)的程度，根據壓力和能量對聆賞者所引發的情緒反應再將情緒分成四群，分別為滿足(contentment)、沮喪(depression)、豐富(exuberance)、焦慮(anxious)。由於不同的色彩對於人類的情緒會有各種面向的不同影響，本研究引用 Thayer 提出的情緒模型，分別將四個象限加上不同顏色來表示情緒，滿足(contentment)象限以綠色表示，舒適愉悅、平靜的；沮喪(depression)象限以以藍色表示，讓人憂鬱寡歡、意志消沉的；豐富(exuberance)象限以黃色表示，興奮、生氣勃勃的；焦慮(anxious) 象限以紅色表示，焦躁不安、暴躁、憤怒的，如圖 5，圖中原點解釋為音樂剛開始的前奏，情緒導引及準備的狀態，圖表橫軸為壓力，代表音樂帶給聽者的抽象壓力；縱軸為能量，定義為音樂帶給聽者的抽象能量。例如：音量較大，抑、揚、頓、挫明顯，節奏快速緊湊分明的聲音通常代表音樂的能量較高；反之，音量較平和，拍子緩慢的聲音則表示音樂能量較小；不和諧的和絃或小調的音樂造成的壓力較大，反映出較為沉悶的情緒感受，容易讓聆聽者壓抑的情緒無法釋放等等，然而，這些音訊特徵不同的強弱程度和情緒模型中的兩軸都有直接相對的關係。

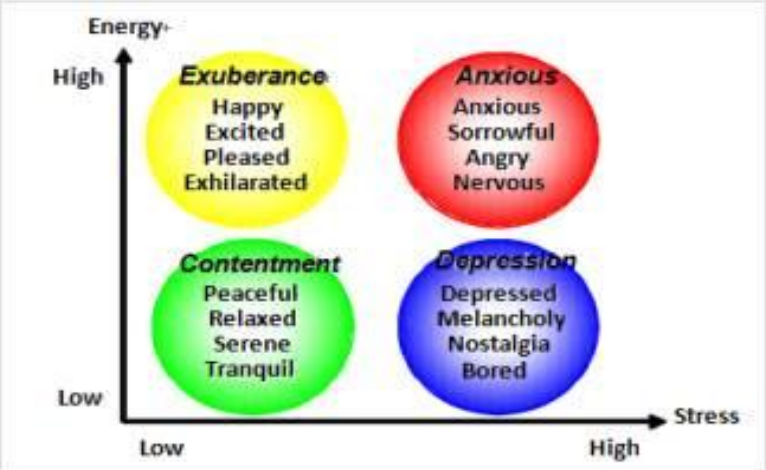


圖 5 Thayer’s 情緒模型

資料來源：[21]

1.2.4 音樂聆賞情緒之心理感受

由於音樂本身的本質是聽覺的媒體，在很多狀況下，人們所感受的音樂聆賞情緒，並不是單一的、彼此無交集的(disjointed)，梅爾認為音樂聆賞是一種動態的過程(dynamic

process)，音樂的理解及欣賞在於人們對音樂的特性之感知(perception)與回應(response)，Clark(1982)則認為情緒有別於情感，相對於情感而言，情緒比較強烈，也比較容易被注意，其中，常見的音樂情緒包含讓人緊張(tension)與鬆弛(repose)、穩定(stability)與不穩定(instability)、模糊(ambiguity)與清晰(clarity)等等，

從音樂的心理層面的觀點看來，音樂心理學是音樂與人的行為、心理的互動關係和規律，音樂能持續不斷且出乎人意料的引起一些緊張感和不穩定的感覺，其引發的情緒反應之影響因素主要歸納為四類：(1) 音樂結構：音的物理屬性(包含音高、音強、音質、時值)與感覺的關係；音程(八度、協和與不協和)對聽覺所產生的結合作用；(2) 演奏者：演奏技巧或表達方式；(3) 聆賞者：包括音樂方面的專業素養、個人偏好、個性、當下心情…等；(4) 背景環境：演出場地、事件等，這些因素皆是影響聆賞情緒的主要項目(Gabrielsson, 2001; Scherer & Zentner 2001)。

音樂演奏的過程中，只用音符無法完整的傳達音樂內容和表現音樂演奏的戲節，因此音樂表情(music expression)便成為在音樂演奏中很重要的因素之一，西方音樂用音符的相對長度和指定音高位置的體系來記譜，或者是運用音樂常見的特性，力度的強弱，速度的快慢、調性的不同，來呈現音樂想表達的情緒或意思。根據不同人物、不同的形象、發生不同的事件會產生不同情緒表現的關係；當音樂演奏時，影響音樂表情最重要的特性是速度以及力度；調性則讓我們在聆聽音樂時會有很明顯的情緒反應，例如表現開朗活潑情緒的音樂，在速度上大都是稍快的，力度上是較弱的，調性可能是屬於大調，表 1 和表 2 歸納歷年來音樂學家和研究學者對音程和音調與其引發的情緒反應之對照關係[22]。

表 1 音程的協合與情緒反應

音程和諧程度	音程	頻率比	情緒反應
協合	同度	1：1	中立
	完全八度	1：2	完美、成就，表現招搖、焦躁，哀悼
	完全五度	2：3	中庸、平靜、欣喜，間帶傷感
	完全四度	3：4	婉約的、哀怨的；滿足、欣喜、顏色、力量、發揚，間帶傷感
不完全協合	大三度	4：5	快樂的、安心的；欣喜、顏色、勇敢、果決、自信、發揚
	大六度	3：5	和悅、力量、勇敢、勝利
	小三度	5：6	忍耐的、順受的；悲傷、愁苦、騷動，另有人認為代表平靜、滿意以及宗教狂熱
	小六度	5：8	愉快的、渴望的；靜穆
不協合	小七度	4：7	哀傷的，悲泣的；疑慮
	大二度	8：9	愉快的 盼望的；帶嚴肅氣
	大七度	8：15	強烈的盼望;騷動 不滿意 驚訝 幻覺
	小二度	16：17	委靡不振；悲傷、痛悼、退讓、焦躁、疑慮
	增四度	8：11	神秘的、厭惡的、反抗的

資料來源：[22]

表 2 調性與情緒的對應

C 大調	和平、高潔、嚴整、樸素	c 小調	溫和、景仰、思慕
D 大調	雄壯、歡樂、充實、華麗	d 小調	勇壯、沉鬱
E 大調	華美、高貴、溫和	e 小調	憂鬱、羞恥
F 大調	柔和、喜悅、平和、充滿	f 小調	暗淡、質樸
G 大調	爽快、熱情、快活、華美	g 小調	沉思、感慨
A 大調	希望、光輝、活潑、熱情	a 小調	柔和、流麗
B 大調	銳利、典雅	b 小調	嚴正、鈍重

資料來源：[22]

1.2.5 音訊特徵萃取

由於音訊資料在多媒體資料當中隨處可見，也扮演著一個重要的特徵，因此音訊資料相關的研究與分析便顯得重要；尤其是基於音訊內涵為主的相關分析更為顯得重要與迫切。一般而言，在音訊資料的內容分析之前，音訊的特徵萃取是首要處理步驟，所謂音訊特徵即為聲音訊號行為模式的一種表現方式，將原始的聲音訊號以量化的方式盡量逼近人耳的感覺感受來代表此音樂的特徵。特徵萃取的分析步驟歸納如下：首先將音訊資料切割成音框單位，針對每個音框中的聲音做特徵分析，產生一組參數，通常包含響度(音量)、節奏、音調，三種影響音樂表情的主要因素，然後在所謂的特徵空間中以統計的方法，將每個音訊檔案做分類。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。

1.2.6 相似度量測

在內涵式音樂資料檢索系統當中，音樂資料的相似度量測是音樂檢索系統能否成功的重要因素之一，相似度量測往往被用來解決使用者無法精確地提出查詢並得到合適的結果，例如：使用者在哼唱時容易出現音調、節拍不符，多音、少音、錯音等情形，因此，計算查詢與音樂資料間的相似度是內涵式音樂資料檢索一項重要的技術，但對於音樂特徵的萃取不同，其適合的相似度演算法也不盡相同。其中，近似字串比對(approximate string matching)演算法和編輯距離演算法(edit distance)常被應用在將音樂的旋律特徵用符號或字串表示的相似度比對上，利用比對計算值的大小來決定兩字串近似的程度，找尋資料庫中相似的音樂檔案。編輯距離定義是兩個字串之間做比對，所需要最少插入(insertion, duplication error)、刪除(deletion, dropout error)、和替代(Transposition error)的數目。

[23]以符號表式 MIDI 格式音樂檔案中音樂資料的主旋律特徵，將其以符號組成一

串時序性的序列，Hsu、Liu 及 Chen[24]把主旋律特徵包含旋律、節奏、和弦等用字串表示，Southampton[25]最早開發 QBH（Query By Humming）系統，將使用者透過麥克風哼唱的音樂資料轉成包含了 U（這個音比前一個音高）、D（這個音比前一個音低）、R（這個音和前一個音相同）的字串來對音樂資料庫搜尋；也有研究學者為了加快檢索速度和使用者的等待時間，透過索引結構的建立讓音樂檢索快速地被處理[26]，不同的索引結構所適合的近似字串比對方式也不相同，不過主要還是以編輯距離的精神為主，插入、刪除、替代這些在編輯距離上的操作正好可以用來處理查詢序列多音、漏音、變調的問題。

除了近似字串比對演算法和編輯距離演算法，距離演算法可以用來解決多個特徵向量高維度的相似度比對，只要兩個要比對的音樂轉換成長度相同的特徵向量，就可以利用這個方式，最常被用的是歐基理德距離演算法，[11]將資料分割成同樣大小的音框，萃取音樂資料中每個音框的低階特徵值，利用距離演算法來計算任兩音框間特徵向量的相似度，其計算結果的距離最小，即為最相近的音樂作品。

二、 音樂多重結構分析

2.1 音樂結構介紹

音樂是由各種音符有次序的安排而流動的聲音藝術，利用聲音來表達作曲者的情感、意志、欲望等內心世界，雖然它是無形的、抽象的、心理的、情感的，但構成音樂形象的聲音是有特定形式的，它是有生命的，作用於人的聽覺，使聆聽者產生一定的聯想，進而在頭腦中形成富有情感的意象，在情緒上受到感染和陶冶。

根據陳文雄音樂與美學-曲式篇一文中的介紹，音樂學所釋義的『形式』，包括兩種不同意義的範圍：(1)音樂曲體（Form In Music），指的是音樂內在的結構形式，用來表達或傳達音樂內容，在特定的時間內各種音樂元素：音色、力度、節奏、旋律等交互作用而產生音樂的輪廓與結構，這些音樂要素不是一群音符的隨便組合，都是經過理性的思維與有秩序的安排所譜寫出來的音樂形式，其中主題(theme)為構成此音樂內容的結構單位之一。(2)音樂曲式（Form of Music）是音樂形式的外表，說明音樂曲子外在結構的規格曲式（Form），用來勾畫音樂形式中不同層次的結構單位分別為樂節、樂句、樂段、段等四種。音樂曲式通常意指音樂史上各時期所使用的曲式，亦即作曲家在譜寫作品時心中設計的一種結構模式，好比建築師的平面構圖一般，例如：巴洛克時期（Baroque,1600-1750）的賦格曲（Fugue）或變奏曲式（Variations）；古典時期（Classical,1750-1820）的奏鳴曲式（Sonata Form）等等，因此，作曲家在作曲時會遵照此種特定之音樂結構的形式加以創作、發揮，稱為音樂形式(music form)。

音樂形式為構成各種歌曲的形式，其構成因素分別為曲體和曲式，包含了音樂理論上的一切原則，即音、音程、音階、調性、節奏、樂句、主題、反覆、變奏、模進等等內在形式，而音樂形式的表達在於應使內容中最重要及次要的部分區分清楚，其中最重要的部分稱為「主題性材料」，主要是指古典音樂中的「主題」或「主題附加部分」（流行音樂中的主歌或副歌部分），而次要部份稱「非主題性材料」，最常見的是樂曲開始的「前奏」、樂曲結束的「尾奏」，及樂曲中間做為連接前後兩個主題用的「過門」或「間奏」。即使各種現代音樂流派紛紛出現，音樂觀和形式觀已經改變，音樂創作趨向自由

化，不再像傳統音樂拘泥於定型的曲式結構，但無論是古典音樂或現代流行音樂，在創作過程中仍然是存有音樂本身的音樂形式和音樂結構的一定規則。

音樂結構存在兩種規則：階層規則(hierarchical rule)和重複規則(repetition rule)。階層規則主要是針對古典音樂，說明音樂物件是以階層方式形成，如從大到小為 movements →sentences→phrases→figures；而同時被應用於流行音樂中的重複規則則是指一段一樣的旋律會重複地出現在音樂物件當中，像是古典樂中的動機(motives)或流行歌曲中的主歌或副歌部分，以上描述主要是以音樂學的角度來描述傳統音樂的音樂形式。

也就是說，以傳統音樂學的概念來說明現代流行音樂，通常，一首普通音樂的音樂結構主要包括了前奏(Intro)、主歌(Verse)、副歌(Chorus)、過門或間奏(Instrument Solo)、尾奏(Ending)等幾個以音樂順序連接而成的部份，組成每個部分的元素大可分為主旋律（Melody）、節奏（拍子 Rhythm）、速度（Tempo，如快板、中板、慢板）、以及襯托主旋律的第二旋律等等音樂元素，但未必每首音樂都齊備上述種種元素。

當我們欣賞一首歌曲的時候，最引人注意且印象深刻的無非就是樂曲中一再重複的音樂片段，也許是駕馭整首歌曲靈魂的主旋律(主歌)，整首歌曲的主要內容，也或許是用來襯托主旋律，與主旋律成對比的第二旋律(副歌)，而主歌和副歌兩部分亦即我們在前面所描述的「主題」部分；所謂的前奏是指歌曲第一主題出現前的音樂，以現代流行音樂來說，前奏是主歌開始前的部分，主要在告訴聆聽者一首歌的開始，同時給予聆聽者在情緒感受的導引和準備，幫助聆聽者進入音樂的內容。無論是什麼的前奏，都應該為樂曲先營造恰當的氣氛，這是十分重要的，音樂前奏的好與壞，影響聆聽者對歌曲的第一印象，正如一篇文章的引子一樣。過門讓音樂段落連接有間歇的效果，間奏則讓音樂段落連接更為自然且順暢[27]。

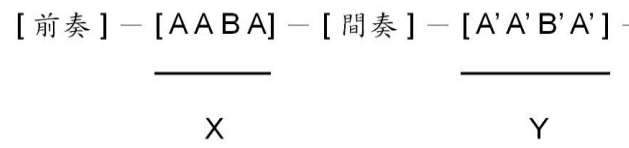


圖 6 流行歌曲的常見曲式

資料來源：[27]

如圖 6，以流行音樂常見的曲式：AABA，其中 A 與 B 分別是兩個不同的音樂段落主題(主歌與副歌)，A'與 B'則是將 A 與 B 作些微的變化。以此例來說，過門發生的地方在 X 或 Y 中的主題段落連接，間奏則發生在不同曲式 X 與 Y 之間段落的連接，此外，在音樂編曲理論中，過門的小節數通常介於一至四小節，而間奏的小節數則介於四至八小節；尾奏通常發生在整首歌曲終結之前，大多有一段作完結的純音樂。較常聽見尾奏的表現方式，如將歌的最後一句重複一次，或重複多次並且漸漸降低聲量至完全沒有聲音（Fade Out）而作完結，以此方式的搭配，不僅使歌曲有收尾的動作，亦使歌曲有前後的呼應和對比，以及讓歌曲有更完美充實的意境與情感。

綜合以上描述，本論文主要針對現代流行音樂作為研究分析，假設所萃取的音檔中存在少部分的古典音樂皆存在重複規則的音樂結構，一首歌的構造主要由前奏，兩段主歌，一段副歌，過門音樂，再來一次的副歌和主歌，以及結尾音樂順序地連接而成的。以音樂內容做主題式的分段，探勘歌曲中的主題性材料，做為系統情緒分析的測試音樂片段，而預設所要萃取音樂片段的部分主要包含前奏、兩段主歌、一段副歌等部分，義即為圖 6 中的 X 部分。

2.2 自相似研究方法(Self-Similarity Analysis)

相似性是音樂檢索、推薦的基礎，本論文參考 Foote[9][10][11]提出的一個基於自相似分析的音樂摘要方法[9][10][11]，該方法將音頻訊號分為固定長度的音框，提取每個音框中 MFCCs 係數作為特徵向量。經由計算任兩特徵向量間的餘弦距離得到一個二維相似度矩陣，最後以累加相似度矩陣各列的數值(Novelty Score)得到具有最大相似度的峰值來找出樂曲中近似重複片段的邊界，以此分段作為一個樂曲最有代表性的部分(摘

要)，並說明在任何時刻音頻訊號中明顯的變化和其的峰值成正比，系統流程圖請參照圖 7，在此我們將針對 Foote's Self-Similarity 方法中的幾個重要步驟做詳細的介紹，首先 2.2.1 音頻參數化 (Parameterization)是訊號分析前做預處理的動作；2.2.2 距離－相似矩陣 (Distance Matrix Embedding)則是將每個音框間做距離運算而得的相似度矩陣；2.2.3 偵測新穎性 (Detecting Novelty)簡單介紹新穎性計分方法(Novelty Score)在此方法中的定義並說明所應用的相關基礎理論－核心相關(kernel correlation)，以及介紹如何利用新穎性計分(Novelty Score)的計算公式測得音頻訊號各個音框間的最大相似值以判斷音樂多重結構的邊界。

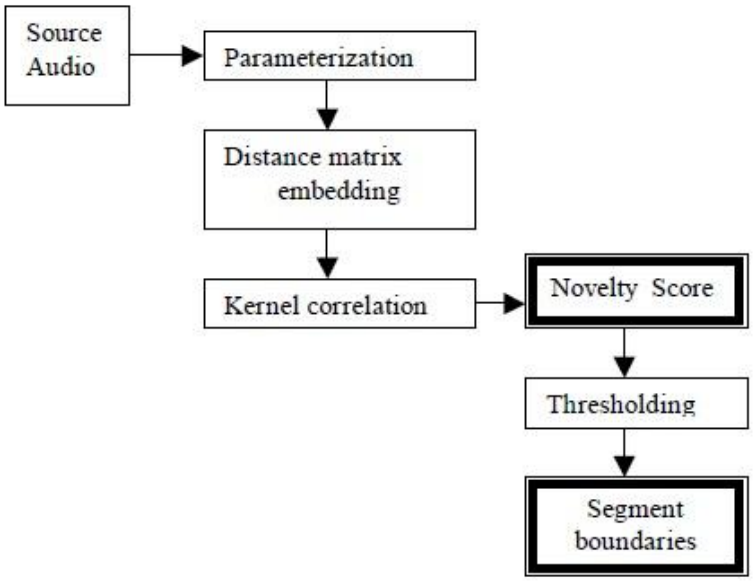


圖 7 Foote's similarity

資料來源：[9][10][11]

2.2.1 音頻參數化 (Parameterization)

Foote 所提出的方法其系統應用相當靈活，可以隨不同的應用加入現有的音頻分析方法，例如：基於人耳對於不同頻率的感受程度，萃取 MFCCs 作為輸入音頻訊號的特徵向量進而達到辨識效果。音頻參數化首要步驟為統一輸入音檔的規格和取樣頻率，針對窗函數的概念對輸入音頻的波形加窗取得獨立的音框，所謂加窗指的是將一段音頻離散時間訊號 $x(n)$，用固定長度的視窗(window)套上去，只看視窗內的訊號，對此視窗內

的訊號作運算，用以求出在此視窗內的音樂特徵。針對不同的應用設計不同的音框長度和重疊長度，音框若太大，就無法抓出音訊隨時間變化的特性；反之，音框若太小，則無法抓出音訊的特性。

2.2.2 距離－相似矩陣 (Distance Matrix Embedding)

完成音頻參數化步驟的音頻訊號會被分割成具連續性同樣大小的音框，每個音框存在獨自代表性的音頻特徵，將每個音框 i 的特徵向量 v_i 和音頻訊號中某個音框 j 的特徵向量 v_j 以距離演算法作時序性的自相似量測，最後，特徵向量間的相似度距離計算結果將產生一個二維空間的相似矩陣 **S**。

在向量空間中判斷兩向量間的距離或稱相似度，有兩種簡單且常用的方式－歐基理德距離(Euclidean Distance)和餘弦相似度(Cosine Similarity)。假設在 L 維空間中存在 v_i 和 v_j 兩特徵向量，則其歐幾里得距離可表示如公式(1)，亦即圖 8 中的距離 D(i , j)，其中 k 表示為音框的索引數目。計算後的數值如果為 0 則表示兩個向量完全相同，而數值越大則代表兩個向量間的相似程度越低。利用歐幾里得距離來度量相似度雖然簡單，但其缺點在於量測結果的單位與程度不明，只能知道距離越小，相似度越高。

$$\mathrm{D}_e\left(v_i, v_y\right)=\sqrt{\sum_{k=1}^{L}\left(v_i(k)-v_j(k)\right)^2} \tag{1}$$

餘弦相似度(Cosine Similarity)如式(2)，以兩組相同基底（Base）與維度（Dimension）向量間的角度（Angle）差距來量測該兩向量間的距離（Distance），其計算結果會介於 0 至 1 之間，當兩個向量間的角度差距越小時，表示該向量間的餘弦角度越小，其計算結果就越接近於 1，也即代表該兩向量相似度越高，反之，其計算結果就越接近於 0，代表該兩向量相似度越低。

一般來說，相似矩陣 S 的最大相似會出現在對角線的方向上，因為每個音框的音訊資料之最大相似部分就是自己本身。

$$\mathrm{D}_c\left(v_i, v_y\right)=\frac{\left\langle v_i, v_j\right\rangle}{\|v_i\|\cdot\|v_i\|}=\sum_{k=1}^{L}\frac{v_i(k)\cdot v_j(k)}{\sqrt{\left(v_i(k)\right)^2\cdot\left(v_j(k)\right)^2}} \tag{2}$$

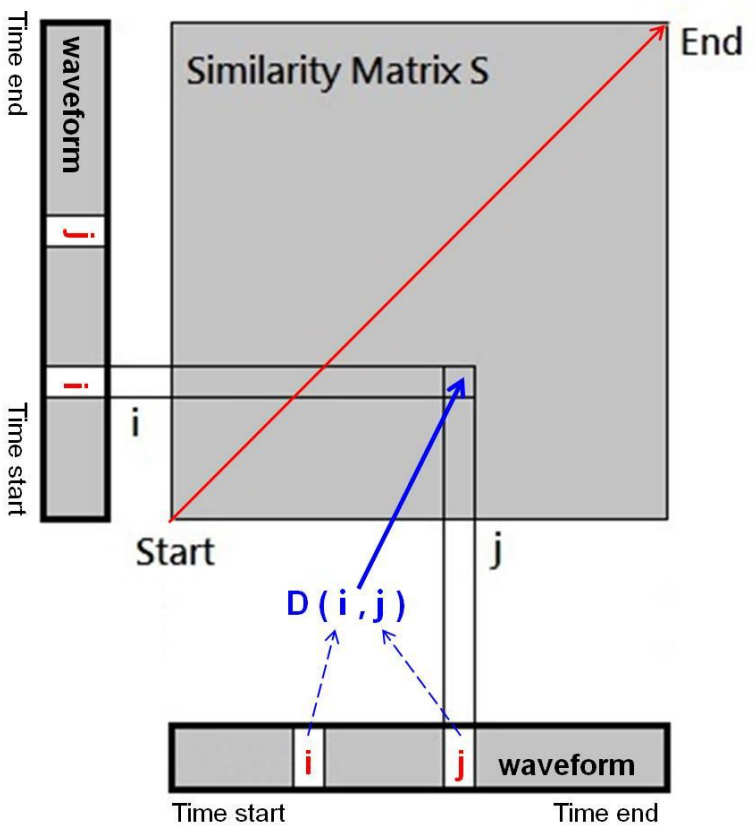


圖 8 基於距離演算法之相似矩陣圖

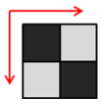
2.2.3 偵測新穎性 (Detecting Novelty)

新穎性(Novelty)在此用來表示音頻訊號顯著的變化點，在討論新穎性之前，我們將針對其用到的相關理論－核心相關(kernel correlation)先做介紹，然後再介紹如何測得新穎性計分(Novelty Score)。

- **核心相關(kernel correlation)**

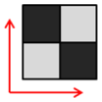
在 Foote's Self-Similarity 方法中，相似矩陣 S 是量測音訊相似度的主要關鍵，為了找出瞬間音符大範圍變化的邊界點，也就是新穎性計分(Novelty Score)，Foote 利用一個看起來像“黑白棋盤”的矩陣來和原本的相似矩陣 **S** 做摺積運算，其中組成棋盤格矩陣的最簡單元素為：一個以主對角線為 1 組成的 2x2 單位同調矩陣(coherence matrix)和一個以反對角線為 1 組成的 2x2 單位不同調矩陣(anti-coherence matrix)，兩個單位矩陣的差即為棋盤的內核心(checkerboard kernel)，如公式(3)中 C' 的第一項和第二項。

$$C' = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3)$$



值得注意的是，在 Foote's Self-Similarity 方法之相似矩陣的運算討論中，為了配合相似矩陣座標軸的起始點，將原本矩陣的座標軸是以左上方向至右下的方式皆改成左下至右上的方式，如公式(4)。

$$C = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4)$$



單位棋盤內核的概念是將一個方形矩陣想像成一個被分割成四等分的正方形，正方形的中心點代表此單位棋盤內核的原點，中心點的左邊和下方依時序性來說代表過去的音框，右邊和上方代表未來的音框，每個小正方形代表一個矩陣元素。式(4)中的第一項是用來量測同一個音框的自相似性程度(self-similarity)，數值越高表示此音框中心點的兩側其音頻訊號之相似性極高；第二項是用來量測橫跨兩個音框的互相似性程度(cross-similarity)，數值越高表示這兩個音框的音頻訊號大範圍幾乎一樣，只有些許的不同。而兩項數值的差就是在這個分法中的重要部份：新穎性計分(Novelty Score)，用來測量訊號本身的相似程度，所得的差值越大，表示此兩個不同音框的訊號彼此非常相似。

棋盤內核的大小可以依照所要分析音頻資料的音框長度自行做調整，小尺寸的棋盤內核用來檢測短時間尺度的顯著改變，如節拍(beats)或音符(notes)；大尺寸的棋盤內核平均短時間尺度所量測的新穎性計分(Novelty Score)，用來檢測較長的音樂結構，如主歌和副歌之間的音樂轉換。大尺寸的棋盤內核構造是一個 2×2 的單位棋盤內核和一個維度為 m×n、構成元素皆為 1 的矩陣做克羅內克積(Kronecker product)運算，克羅內克積定義為兩個任意大小的矩陣間的運算，以符號⊗表示，如果 A 是一個 m×n 的矩陣，B 是一個 p×q 的矩陣，而克羅內克積則是一個 mp×nq 的分塊矩陣，舉例來說：如果要得到一個 4×4 的棋盤內核，就是把 2×2 單位棋盤內核和 2×2 矩陣做克羅內克積運算，如式(5)。利用克羅內克積來改變棋盤內核尺寸大小的優點在於可以保留原本棋盤內核的結構。

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \quad (5)$$

➢ **新穎性計分(Novelty Score)**

由於相似矩陣是將音訊資料切割成音框單位後，任兩音框之間作相似比較而得的數值，所以新穎性計分(Novelty Score)正代表兩個音框間其音頻訊號改變的程度，我們將利用測量而得的新穎性得分作為音樂訊號粗略分段的邊界。

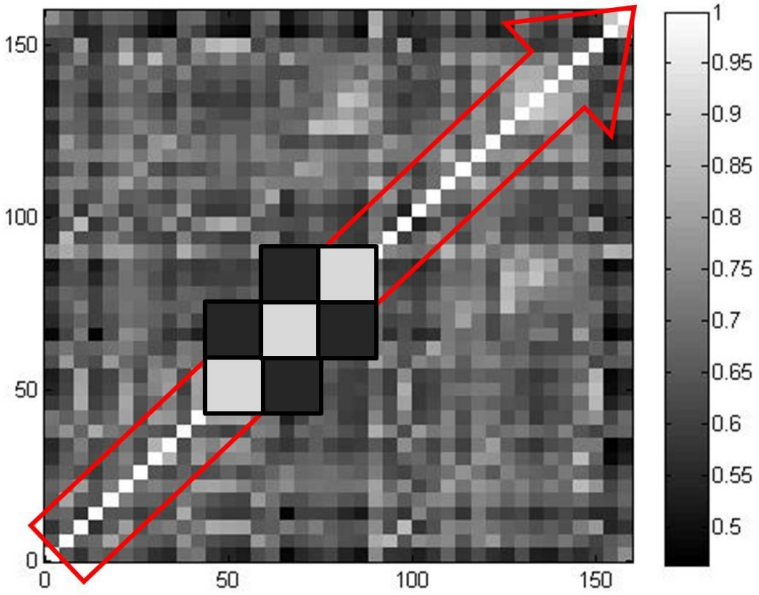


圖 9 新穎性計分的運算概念

如何運算得到新穎性計分呢？做法為想像將一個以單位棋盤內核組成的矩陣，沿著相似矩陣 **S** 對角線的方向滑行，如圖 9。棋盤內核矩陣和相似矩陣 S 中每個元素做乘積運算，最後將所有乘積運算而得的數值加總起來就是新穎性計分(Novelty Score)，如式(6)，其中 $\boldsymbol{C}$ 表示一個棋盤內核矩陣，寬度為 w，中心點(0,0)，中心點兩側分別代表寬度為 $\frac{w}{2}$，以時序性而言，過去的音框和未來的音框；i 則是相對於原始音頻訊號在連續時間索引上的音框數目。為了有效的考慮距離中心點(0,0)在新穎性計分的影響程度，同時避免音框中心點兩側音頻資料組成的不平均所產生的邊緣效應(Edge Effect)，在這裡我們使用加窗概念，利用一個 32×32 高斯徑向基函數的濾波器來平滑棋盤內核矩陣，如圖 10；圖 11 比較原始尚未濾波的棋盤內核(左邊)和經過高斯濾波器平滑後所形成的棋盤內核平面圖(右邊)，其中越接近中心點(0,0)的值越大；反之，越靠近邊緣區域的值

將趨近於 0。

$$N(i)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}C(m,n)S(i+m,i+n) \tag{6}$$

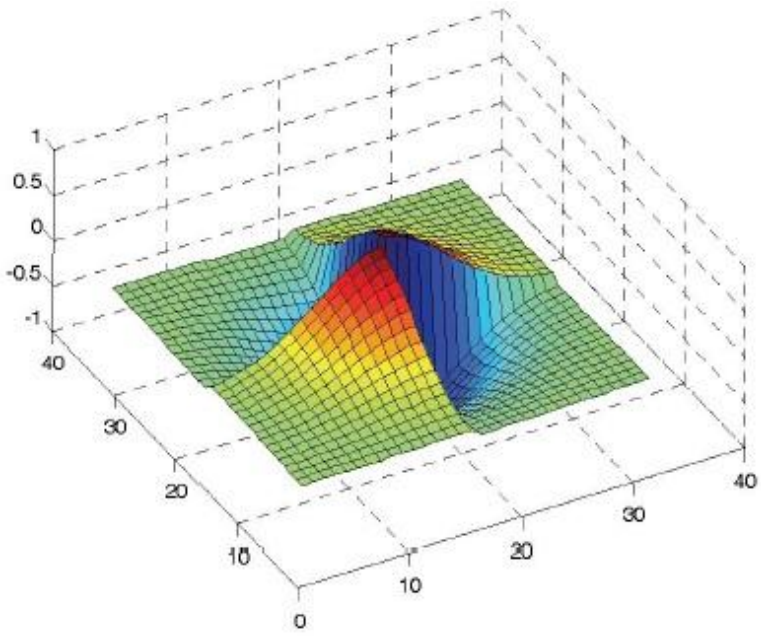


圖 10 32x32 高斯棋盤內核立體圖

資料來源：[28]

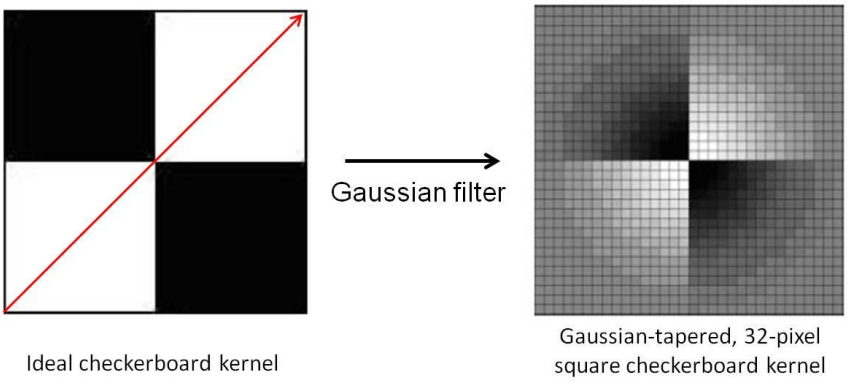


圖 11 32x32 高斯棋盤內核平面圖

資料來源：[29]

三、 音訊分析之方法與原理介紹

3.1 能量頻譜(Power Spectrum)

能量頻譜為一種描述訊號在頻率軸上如何分布的方法，經由快速傅立葉(FFT)的運算後將時間訊號轉換至頻率軸上討論，如式(7)。根據 Parseval 定理，訊號經快速傅立葉轉換後取其振福的平方即為音樂訊號的能量，如式(8)。

$$\underset{k=0,1,2,...,N-1}{X_m[k]} = \sum_{n=0}^{N-1} x_m[n] \times e^{-j\left[\frac{2\pi k}{N}\right]n} \tag{7}$$

$$\underset{k=0,1,2,...,N-1}{P_m[k]} = \left|X_m\left[k\right]\right|^2 \tag{8}$$

其中，x_m[n]為原始音樂訊號，m 為音框索引值，$X_m[k]$為原始訊號經快速傅立葉轉換後的頻譜，$P_m[k]$為訊號的能量頻譜。

3.2 短時距頻譜

當我們在分析聲音時，通常以「短時距分析」（Short-time Analysis）為主，因為音訊在短時間內是相對穩定的。因此，針對已音框化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，配與特定權重的離散傅立葉轉換(Discrete Fourier transform)，其數學定義如下：

$$\underset{k=0,1,2,...,N-1}{S_m[k]} = \sum_{n=0}^{N-1} x_m[n] \times w[n] \times e^{-j\left[\frac{2\pi k}{N}\right]n} \tag{9}$$

$$\underset{k=0,1,2,...,N-1}{f[k]} = \frac{f_s}{N} \times k \tag{10}$$

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視窗函數(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。

圖 12 為單一音框頻譜的圖形，從圖中可以清楚看到音訊在各個頻率的強度大小與分佈。如圖 13 說明連續時間音頻訊號在不同時間各個頻率的強度大小與分佈的頻譜圖。

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。

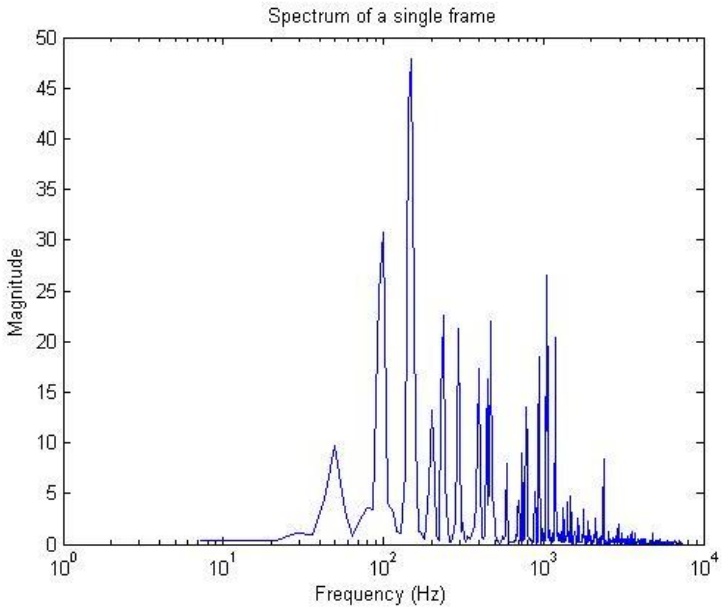


圖 12 單一音框的頻譜圖

資料來源：Mariage Damour.wav_frame#300

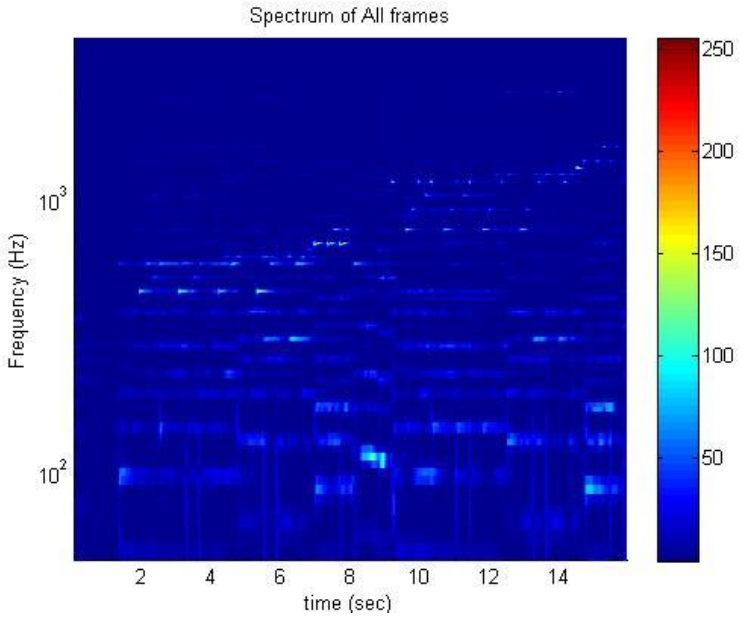


圖 13 連續時間的頻譜圖

資料來源：Mariage Damour.wav

而視窗函數 $w[n]$是用來選取原始音樂訊號某特定部分的實數、且長度有限的序列，常用的視窗函數為矩形視窗(Rectangular window)、漢明視窗(Hamming window)和漢尼視窗(Hanning window)，三種不同視窗之示意圖如圖 14。由於訊號是連續的，如果在傅立業的轉換過程中外加一個矩形窗做取樣，在窗的兩旁會造成訊號的不連續且對於轉換後的訊號兩旁容易產生假訊號，為了降低此問題，使窗內外不會有太劇烈的變化，通常分析時會選擇使用漢明窗或漢尼窗，它具有壓抑短時距訊號的兩端，改善音框訊號在計算頻譜時的邊界效應；保持中間段的特性，使頻譜的數值對比更好。三種視窗之數學定義依序如下。

$$\underset{n=0,1,2,\ldots,N-1}{w[n]} = 1 \tag{11}$$

$$\underset{n=0,1,2,\ldots,N-1}{w[n]} = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \tag{12}$$

$$\underset{n=0,1,2,\ldots,N-1}{w[n]} = 0.5 - 0.5 \times \cos\left(\frac{2\pi n}{N-1}\right) \tag{13}$$

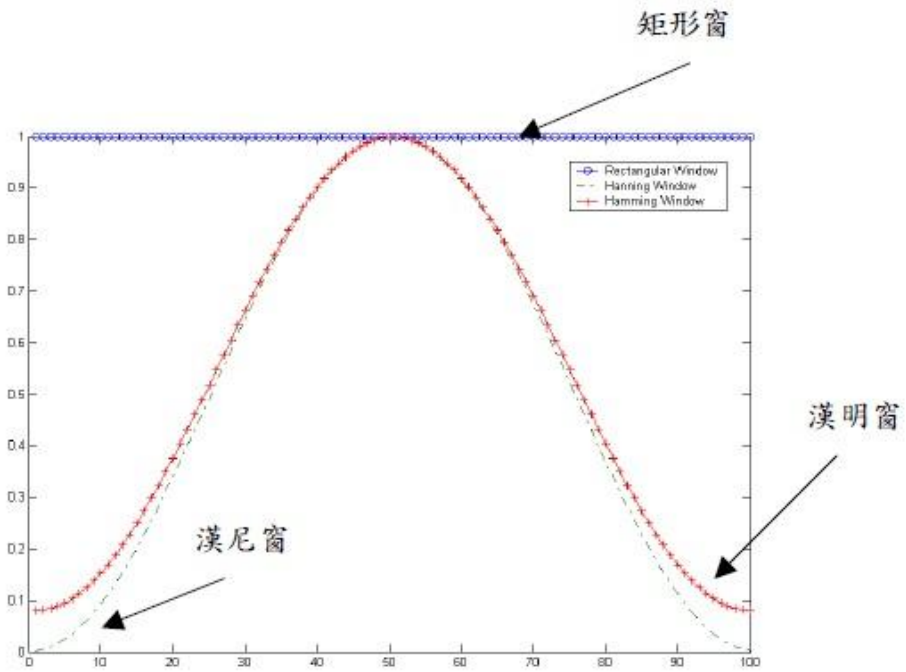


圖 14 三種不同視窗產生的濾波響應圖

3.3 音調層級分析 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音調特徵值(Pitch Class Profile，PCP)，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：

$$P(k) = 24 \cdot \log_2 \left(\frac{f_s}{N} \cdot \frac{k}{f_1} \right) \bmod 24 \tag{14}$$

$$\mathrm{PCP}[P(k), n] = \sum_{P(k)} |S[k, n]| \tag{15}$$

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，

$S[(k=0, 1, \cdots, N), n]$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，$\mathrm{PCP}[(P(k)=0, \cdots, 23), n]$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 15，各個時間的音調層級的強度分佈如圖 16。

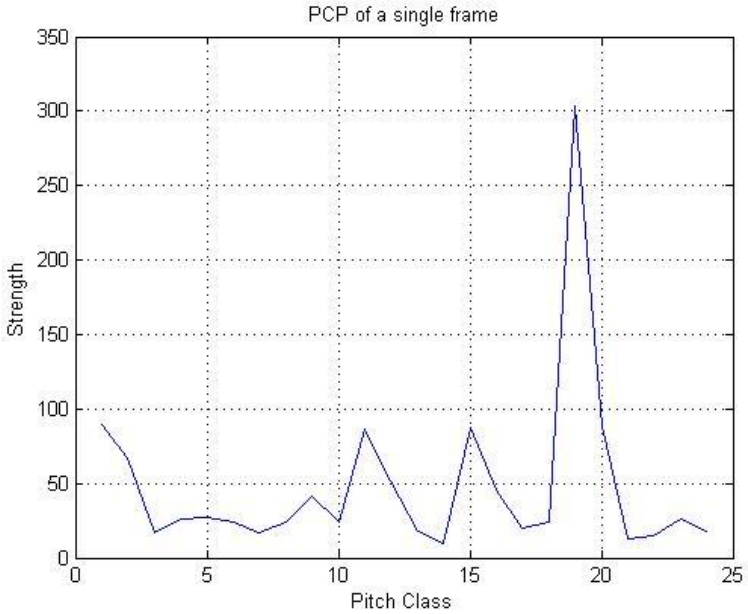


圖 15 單一音框的音調層級強度分佈圖

資料來源：Damour Mariage.wav_frame#300

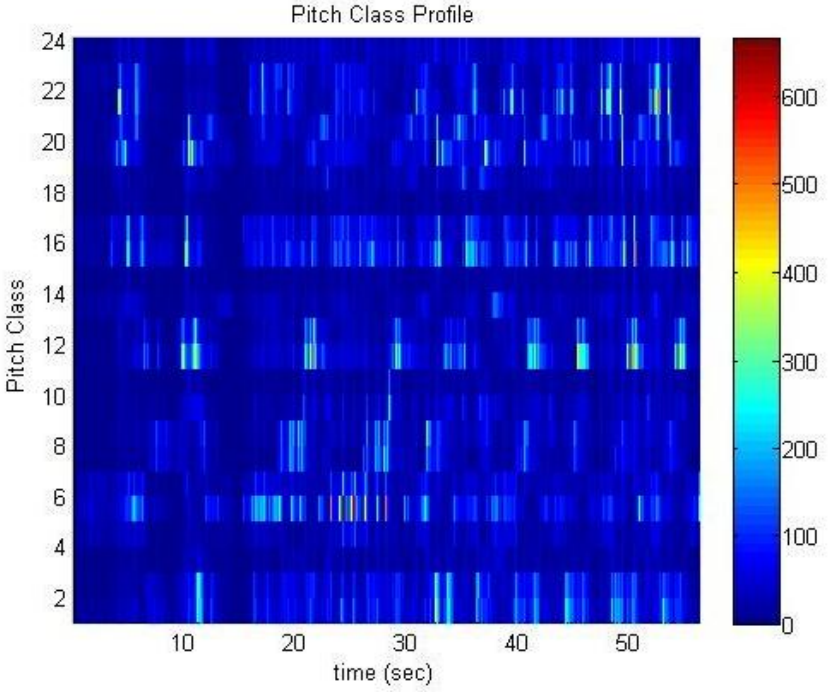


圖 16 連續時間的音調層級強度分佈圖

資料來源：periodmusicDamour Mariage.wav

3.4 高斯混合模型 Gaussian Mixture Model (GMM)

高斯混合模型是單一高斯機率密度函數的延伸，為一種常見的正規分佈。一般在一維的狀況下，高斯機率密度(Probability density function)是用來說明特徵向量 x 在一個特定種類中出現的機率為何，如式(16)為描述特徵向量 x 的機率密度，其分佈圖形如圖 17。

$$p(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{16}$$

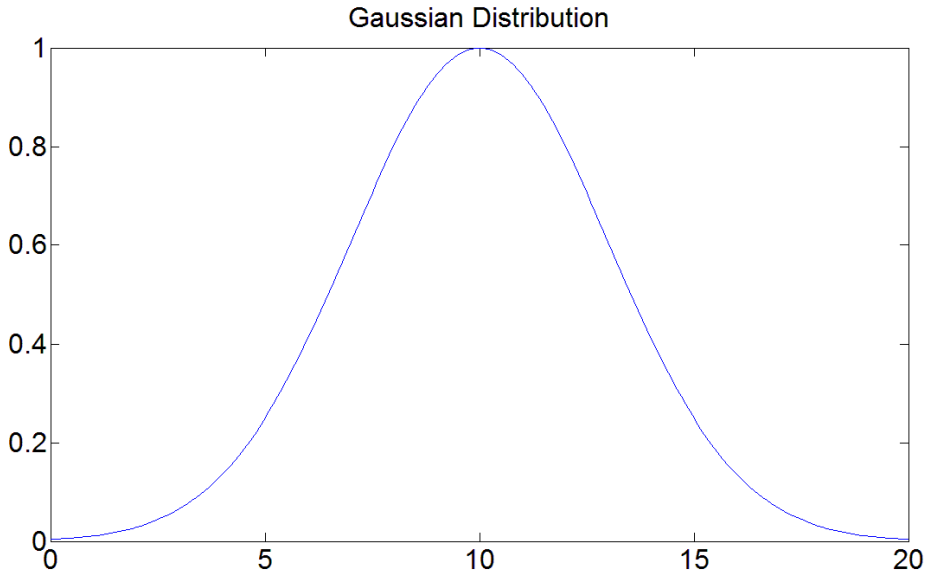


圖 17 高斯分部

其中有 μ 和σ兩個重要的參數，μ 為期望值(Expectation value)，代表密度函數的中心點或平均向量，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：

$$\mu \equiv E[x]=\int_{-\infty}^{\infty} xp(x)\,dx \tag{17}$$

$$\sigma^2 \equiv E[(x-\mu)]=\int_{-\infty}^{\infty}(x-\mu)^2\,p(x)\,dx \tag{18}$$

高斯模型：利用向量和矩陣推廣為高維度的高斯機率密度函數表示如下式(19)：

$$g(x;\mu,\mathrm{C})=\frac{1}{(2\pi)^{d/2}\,|\mathrm{C}|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^{\mathrm{T}}\mathrm{C}^{-1}(x-\mu)\right] \tag{19}$$

其中 μ 和 C 分別為期望值和共變異矩陣(Covariance Matrix)，Covariance 是 Variance 在高維度中的一種推廣，其第 i-j 個元素代表第 i 維度和第 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 和的值會和其分佈的中心位置和曲線寬度有關。

$$\mu \equiv \mathrm{E}[\mathrm{x}] = \begin{bmatrix} \mathrm{E}[x_1] \\ \mathrm{E}[x_2] \\ \vdots \\ \mathrm{E}[x_d] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = \sum_{\mathrm{x}} \mathrm{x}P(\mathrm{x}) \tag{20}$$

$$\mathrm{C} = \begin{bmatrix} \mathrm{E}[(x_1-\mu_1)(x_1-\mu_1)] & \mathrm{E}[(x_1-\mu_1)(x_2-\mu_2)] & \cdots & \mathrm{E}[(x_1-\mu_1)(x_d-\mu_d)] \\ \mathrm{E}[(x_2-\mu_2)(x_1-\mu_1)] & \mathrm{E}[(x_2-\mu_2)(x_2-\mu_2)] & \cdots & \mathrm{E}[(x_2-\mu_2)(x_d-\mu_d)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(x_d-\mu_d)(x_1-\mu_1)] & \mathrm{E}[(x_d-\mu_d)(x_2-\mu_2)] & \cdots & \mathrm{E}[(x_d-\mu_d)(x_d-\mu_d)] \end{bmatrix} \tag{21}$$

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，當所量測的資料 X={x_1,x_2,⋯,x_n}在 d 為空間中的分佈不是橢球狀，就不適合以一個單一的高斯密度函數來描述這些資料點的機率密度函數。此時將採用數個高斯函數的加權平均(Weighted Average)來描述 X 的機率密度，亦即高斯混合模型。如第(22)式為一個二維空間、以三個高斯機率密度函數表示的數學式，其中 $\mathrm{C_j}$ 為各個高斯密度函數的共變異矩陣，而且權重α_1、α_2、α_3 要滿足總和為 1，其分佈圖形如圖 18。

$$p(x,y) = \alpha_1 g(x,y;\mu_1,\mathrm{C}_1) + \alpha_2 g(x,y;\mu_2,\mathrm{C}_2) + \alpha_3 g(x,y;\mu_3,\mathrm{C}_3)$$

$$\mathrm{C}_j = \sigma_j^{\,2}\mathrm{I} = \sigma_j^{\,2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, j = 1,2,3 \tag{22}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

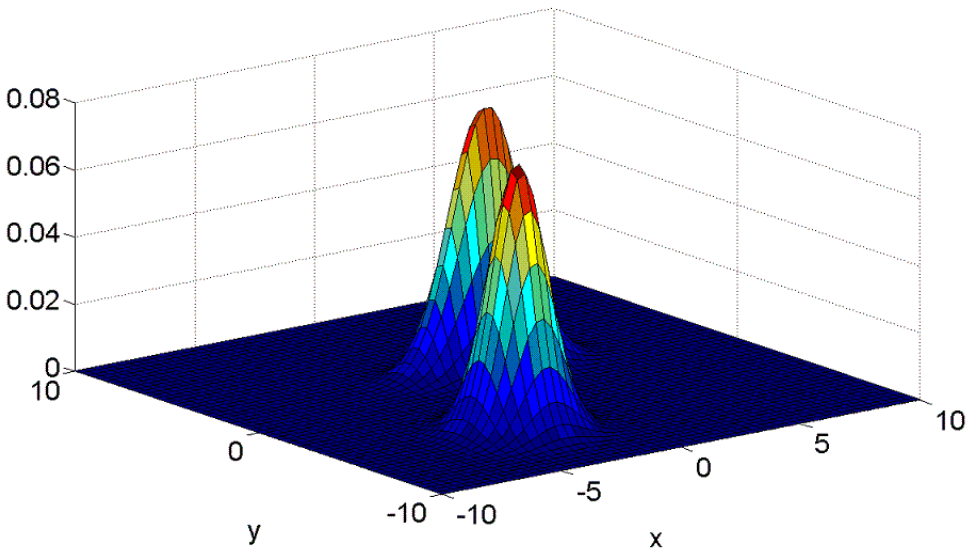


圖 18 混和高斯分部

資料來源[30]

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，找出一近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。

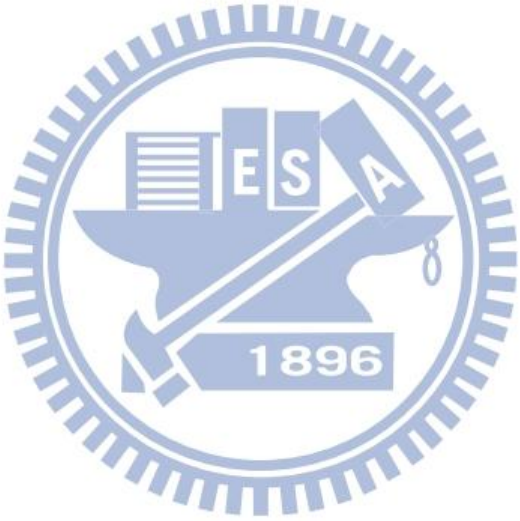
3.以微分求極值的方式，由舊有的參數計算出一組新的 GMM 參數。

4.重複步驟 2~3 疊代，直到辨識率收斂到某一個極值。

此描述的計算方法稱為 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)，也有可能會收斂到局部最大值(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。



四、 研究方法

本章節討論本篇論文的研究方法，第一節提出系統架構流程；第二節說明測試資料預處理步驟－音樂分段；第三節討論音樂多主題架構的情緒分析方法；第四節說明音樂情緒相似度的概念與比對方法。而系統最終輸出之圖形化使用介面－視覺化的音樂自動選曲系統與流程將在第五章音樂心情點唱機再做詳細的介紹。

4.1 系統架構

系統流程如圖 19 所示，系統輸入的音訊資料主要分成訓練資料和測試資料兩大部分，圖中紫色方塊部分為訓練資料和測試資料都需要分析的步驟，包含音訊輸入、特徵萃取、計算能量－壓力情緒分數等；綠色方塊部分代表只針對訓練資料做分析；藍色方塊部分代表測試資料的分析步驟。

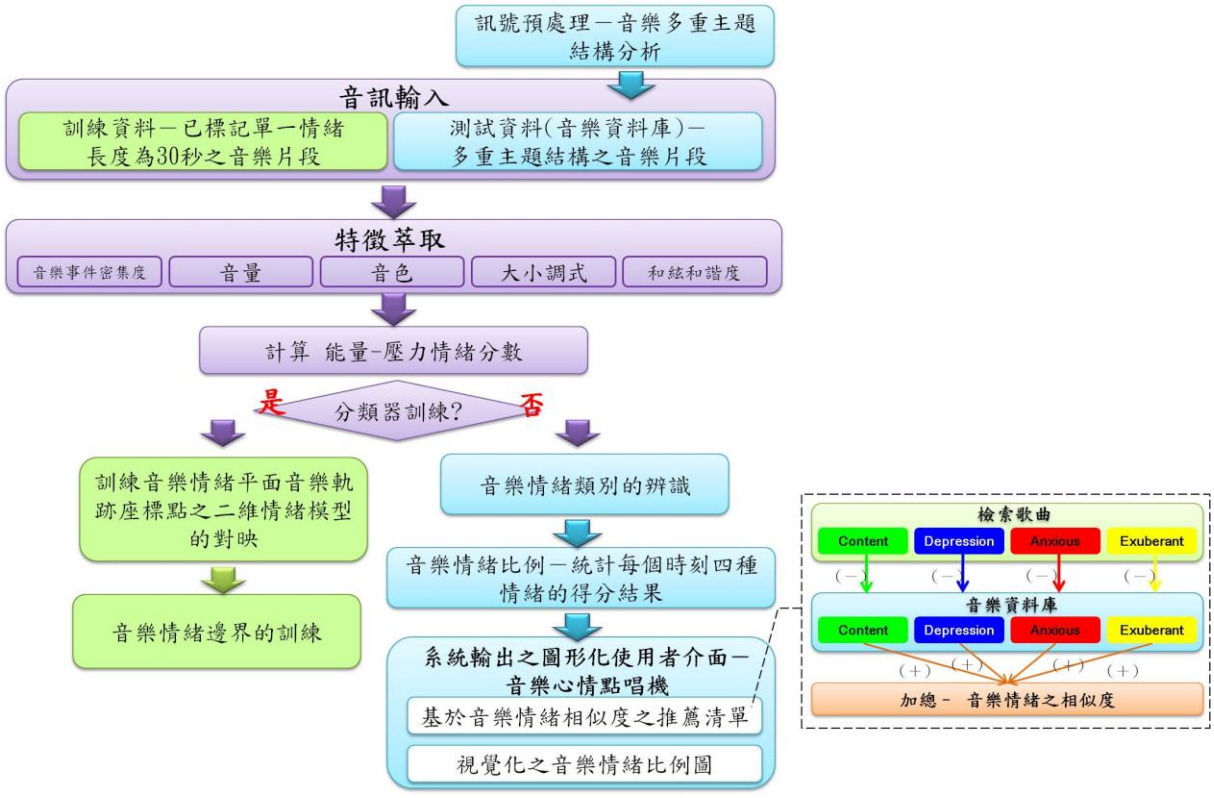


圖 19 系統架構流程方塊圖

訓練資料由兩百首長度為三十秒、已標記單一穩定情緒的音樂片段組成，用於辨識系統測試資料的音樂情緒；測試資料總共有兩百一十首，音訊內容完整、wave 格式的

音樂，包含古典純樂器演奏音樂至流行音樂、電子音樂等各種風格類型，主要用於系統最終輸出的圖性化介面，為系統的音樂資料庫。音樂資料庫的音訊資料首先透過系統的預處理步驟分析多重主題結構的音樂片段，接著萃取並分析這些音樂片段的各種音訊特徵後計算其能量－壓力的情緒得分，最後經由訓練資料界定的情緒邊界辨識在每個時刻所屬的情緒後，事先儲存音樂資料庫中每首歌曲的情緒比例，將被應用於之後系統輸出使用者介面情緒成份相關的即時運算。

4.2 音樂多重主題結構分析

音樂之多重主題結構分析為系統測試資料的訊號預處理步驟，主要是擷取音樂多重主題結構的音樂片段。為了測量以多主題音樂結構為基礎的音樂片段，首先要分析的是樂曲中近似重複片段的週期 t_1,t_2,t_3,t_4 (包含主歌和副歌的近似重複片段)，分別以主歌的近似重複片段週期之邊界和副歌的近似重複片段週期之邊界作為主題性音樂片段的切割點，基於音樂時序性分析所有擷取的切割點來找出樂曲演奏完所有主題結構(第一主題曲式結構)，即將進入第二次重複演奏以前的時間點，亦即間奏部分，由於間奏在音樂結構中主要扮演連接的角色，對於音樂情緒感受的影響並不大，因此我們將間奏部分視為自由性擷取，最後，並將所擷取的時間點作為音樂多重主題結構之週期切割點。音樂多重主題結構之音樂片段的擷取主要著重於主題段落，預設擷取的音樂片段週期如圖 20 的紅色現段，A 代表主歌；B 代表副歌。

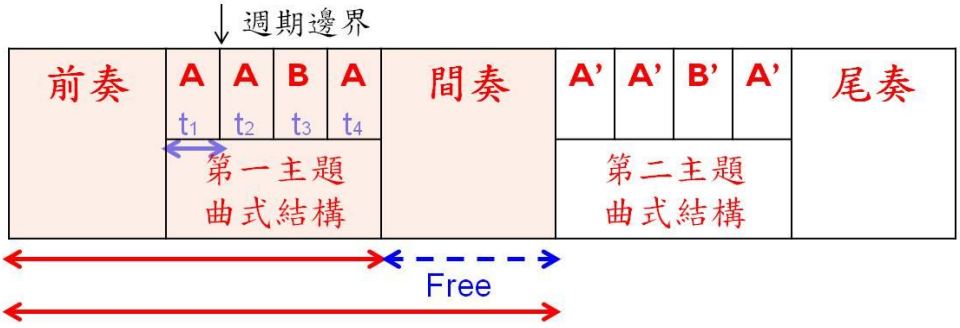


圖 20 說明預設擷取的音樂多重主題結構之週期

音樂多重主題結構之分析方法分成(1)粗略分段(Rough Segmentation)和(2)精細分段(Salient Segmentation)兩個主要步驟,詳細說明如下：

(1) 粗略分段－近似重複片段之邊界檢測

訊號預處理：統一所有要分析的音樂檔案格式為.wav 檔，雙聲道，設定取樣頻率為11025Hz，將音訊檔案切割成固定的音框長度，音框數目依各個音訊檔案的時間長度改變。

特徵萃取：統一設定音訊檔案的所有參數後，首先萃取音頻訊號波形在頻域空間上的頻譜特徵來取代萃取音頻訊號中音樂內容相關的聲音特徵所造成複雜且過大的運算量，常見的如：音高(Pitch)、和弦(Chord)、調性(Tonality)、主音(Key)、rhythm(節奏)、節拍(Tempo)等。考慮重複片段出現的頻率，將每個音框的頻譜振幅值取平方而得的能量頻譜(power spectrum)作為一特徵向量，所謂能量頻譜(power spectrum)定義為一個時間序列的訊號經快速傅立葉(FFT)轉換後振幅的平方值，說明一個時間序列的訊號變化在頻域空間上的能量分步。接著，利用自相關函數(Autocorrelation)計算每個音框的能量頻譜特徵來強調重複片段在時域空間上出現的頻率，如圖 21 頻域上能量分布豐富且明顯的部分代表近似重複片段可能發生的時間點。

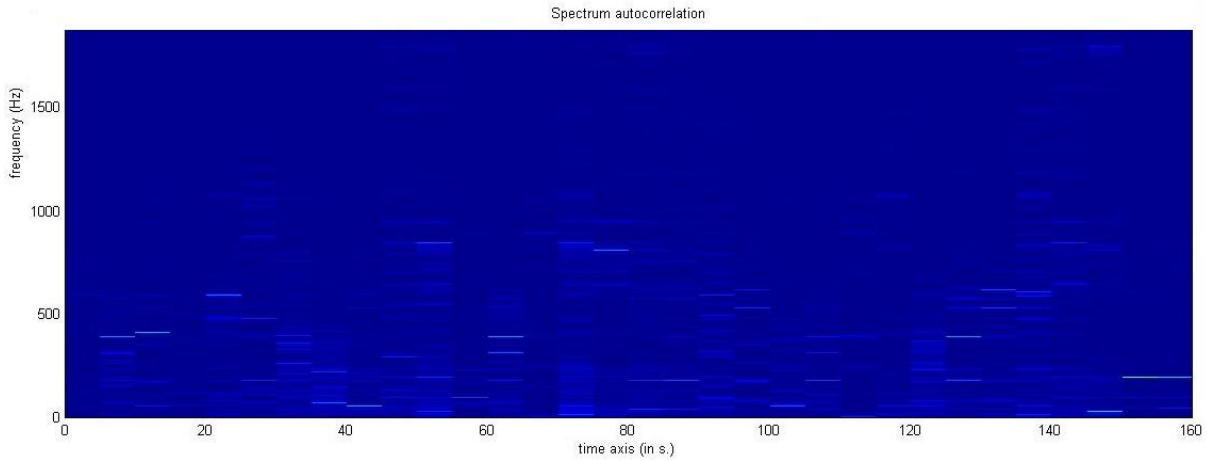


圖 21 利用各個音框之能量頻譜進行自相關函數計算

資料來源：Mariage Damour.wav

相似矩陣：參考 Foote[9][10][11]提出的 Foote's Self-Similarity 方法，首先利用餘弦相似度之距離演算法來計算任兩個音框能量頻譜特徵向量之間的相似度，如圖 22，相似矩陣中顏色越亮代表其相似性越高，主對角線白色部分代表音框自己本身的相似度。從圖 22 中可以由平行主對角線的白色線條或較明亮的方形區塊來判斷歌曲中近似重複片段的部分，其中平行主對角線的線條說明了再次發生的"連續性序列音樂"，而方形區塊表示內部重複出現同種音樂的狀態。基於相似矩陣的對稱性，可以單只針對一個上三角型或下三角型上的資料做分析。

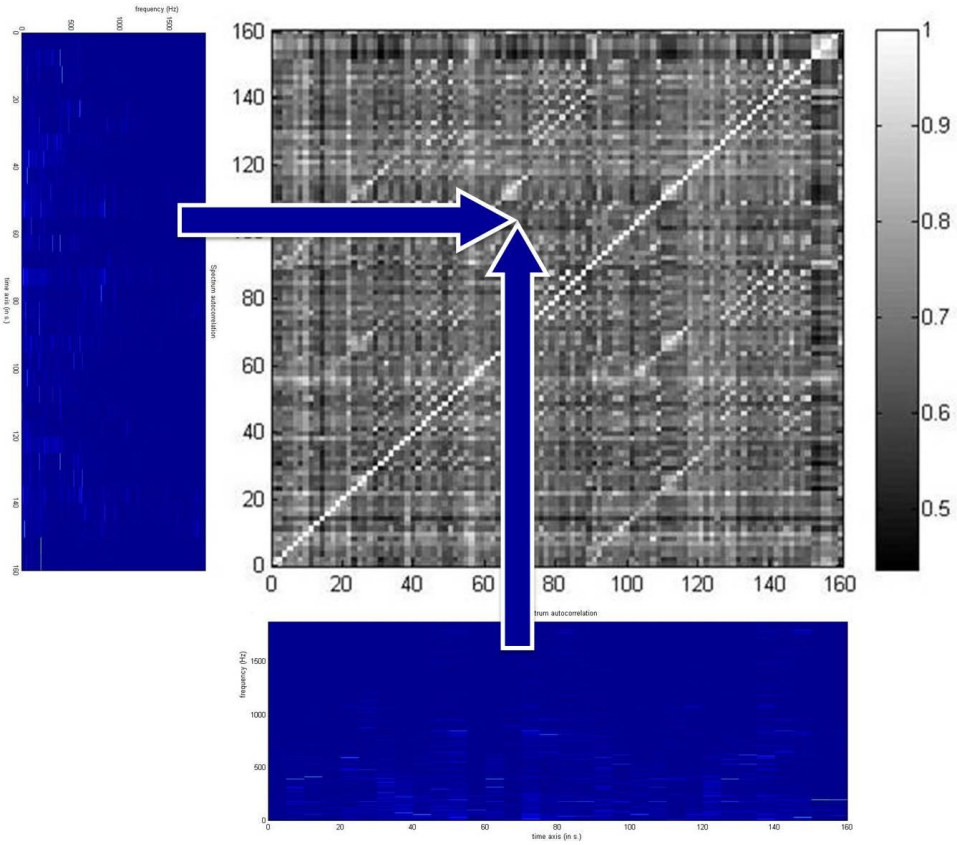


圖 22 以自相關函數計算任兩個音框能量頻譜特徵向量之相似矩陣

音樂分段：如同 Foote's Self-Similarity 的做法，根據公式 23，利用一個 32x32 的棋盤內核矩陣和相似矩陣做各個元素的乘積運算，最後再加總所有音框內的元素乘積數值而得新穎性計分。如圖 23，其下方圖為新穎性計分的峰值圖。

$$N(i)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}C(m,n)S(i+m,i+n) \tag{23}$$

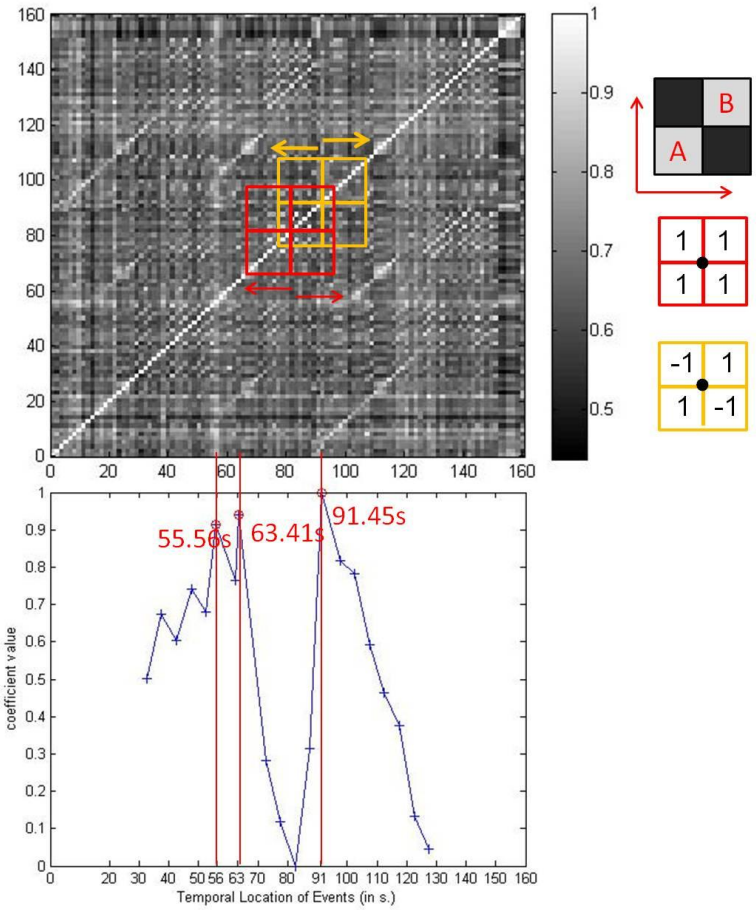


圖 23 相似矩陣和新穎性計分之比對圖

資料來源：Mariage Damour.wav

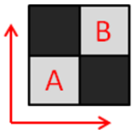
利用已量測的新穎性計分之峰值來判斷近似重複片段之週期切割點，利用前三高的新穎性計分做為各個主題週期的切割點，初步以最高之新穎性計分峰值作為近似重複片段之粗略分段，如圖 23 中之下方圖的 91.45s 處，即為近似重複片段之粗略分段。在此，將針對新穎性計分的峰值(Novelty score = 1)和谷值(Novelty score = 0)分別做介紹：

- **Novelty score = 0**

  根據新穎性計分公式，若將棋盤內核矩陣和一個組成元素皆為 1 的矩陣直接作乘積運算後加總起來，其新穎性計分為等於零，亦即代表圖 23 中下方圖的谷值部分。依據組成棋盤內核的概念來說明組成元素皆為 1 的矩陣，參考公式 24

中相似矩陣的圖式，代表 A 區段音樂的自相似很高，B 區段音樂的自相似很高，且 A 和 B 兩段音樂的相似性也很高。接著，在將此說明對照到圖 23 的下方圖，表示在約 82 秒處，以棋盤內核矩陣的中心點為原點，在過去 72～82s 音框的時序性旋律也會在未來 82～92s 間重複出現。

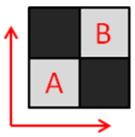
$$N(i)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}C(m,n)S(i+m,i+n)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}\begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}=0 \quad (24)$$



- **Novelty score = 1**

同理，根據新穎性計分公式(25)，若將棋盤內核矩陣和一個其主對角線的組成元素為 1 且反對角線的組成元素為-1 的矩陣直接作乘積運算後加總起來，其新奇計分的結果會趨近於一個大值，代表 A 區段音樂的自相似很高，B 區段音樂的自相似很高，但 A 區段和 B 區段的音樂彼此完全不相似。參考圖 23 下方圖約 91.45s，以棋盤內核矩陣的中心點為原點，在過去 82～92s 音框的旋律和未來 92～102s 間的音框的旋律完全不相似，也許是音樂主題與主題間的轉換邊界點。

$$N(i)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}C(m,n)S(i+m,i+n)=\sum_{m=-\frac{w}{2}}^{\frac{w}{2}}\sum_{n=-\frac{w}{2}}^{\frac{w}{2}}\begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix} \quad (25)$$



(2) 精細分段－音樂多重主題結構之邊界檢測

由於所萃取的音訊特徵為能量頻譜，利用能量頻譜分析而得的相似矩陣來計算新穎性計分的峰值並不能完整地以主題性的結構做分段，無論樂曲的主歌或副歌部分在典型的音樂結構中皆屬於主題性結構，想像作曲家以重複法則依美學原則的安排各種音樂的對比、變化和統一等形式，使得樂曲的組織完整而規律，而演奏家是如何傳達音樂內在的豐富情感，如何透過演奏方式區分兩個都屬於主題性結構的旋律？例如利用改變聲音的大小方式來凸顯不同主題旋律，聲音的大小代表能量的多寡，越大聲的音量代表其訊號的頻譜能量越大，然而，更大的頻譜能量，其存在更多的重複性[31]。因此，根據流

行音樂創作法則和音樂理論，為了進行更精確地檢測音樂多重主題結構之邊界，我們先將新穎性計分的結果分別以峰值發生的時間依序排列得到向量 N_t 和以峰值的計分結果依大小排列得到向量 N_v，並同時做了兩個假設來判斷 N_t 和 N_v，如下：

＜假設一＞

基於音樂創作越趨於自由化，每首歌曲安排重複演奏的形式也不盡相同，一般以演奏完 AABA 形式所組成的第一主題段落後再進入間奏，接著第二主題段落 A'A'B'A' 部份，想像其演奏方式以層層帶入做對比映襯的表演引導我們進入歌曲的主歌和副歌部分，因此我們假設最大新穎性計分的峰值會發生在第二主題段落部分，假設第一主題段落將會發生第二高或第三高的新穎性計分。

＜假設二＞

對於一首長達三～四分(約 240s)的音樂檔案而言，其發生在 40s 附近的音樂有可能只是前奏或主歌的前半段部分，另外，依重複規則將前奏、主歌、副歌等結構做第二次演奏的時間點也不可能發生在 40 秒以前。因此，根據假設一做判斷，將刪除發生在 40 秒以前新穎性計分的峰值。

＜假設三＞

無論主歌或副歌部分，都是由幾個音符依照不同形式的安排而組成的小節，再由小節做變化進而形成主歌或副歌的重複樂句，假設樂句的長度約 10～15 秒，而峰值恰好落於每個棋盤內核矩陣的中心點處。

接著，根據假設來檢測有效峰值的週期邊界，其條件判斷可歸納分為以下二個步驟：

＜步驟一＞

重複檢查、刪除 N_t 向量發生在 40s 以前新奇計分的峰值點，直到新穎性計分峰值點的發生時間不在 40s 以前，然後整理並更新 N_t 向量。

＜步驟二＞

根據假設三，所量測到的時間切割點會恰好落於樂句的中心點，假設最長樂句約為 14 秒。針對已整理並更新後的 N_t 向量繼續作條件判斷，找尋 N_t 向量中第二高的新

奇計分值，並將其所對應的時間點減七秒為 t_t 來代表樂句的起頭。同時也對 N_v 向量中第二高的新穎性計分所對應的時間點減七秒為 t_v。

最後，比較已經過條件判斷做調整的 t_t 和 t_v。若以新穎性計分數值大小依序排列的向量 N_v 中第二大的新穎性計分峰值減七秒的時間點 t_v 還是發生在演奏歌曲六十秒以前或一百四十秒以後，就讓 N_t 向量中第二大的新穎性峰值減 7 秒 t_t 來作為最後精細分段－音樂多重結構之切割時間點。

4.3 多重主題音樂片段的情緒分析

在此為了降低音訊資料分析與計算的複雜度，將選擇已分析音檔中多重主題結構的音樂片段代替整首音樂資料，將此音樂片段分析的情緒比例作為系統測試整首歌曲之音樂情緒縮略圖或音樂情緒摘要，有利幫助加快以音樂情緒比例分類為基礎的視覺化音樂自動選曲系統的檢索速度。另外，為了建立可以確實辨識連續情緒變化的分類模型，在訓練資料方面，選取情緒起伏不大且單一的 30 秒音樂片段，詳細的說明將在接下來 4.3.1 至 4.3.5 小節中探討。

4.3.1 情緒分析之設計概念

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)心理情緒常依個人經歷和記憶改變，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。

4.3.2 訓練資料格式

訓練資料總共有 200 首，音訊內容皆為原始 CD 的音訊內容而非由 MIDI 與音源產生的音訊，檔案類型為 wave 格式的音樂片段，每個音樂片段長度為 30 秒，音樂風格包含流行音樂、搖滾音樂、古典音樂…等各種音樂類型，統一設定相關參數如取樣頻率

14700Hz、取樣解析度 16-bits、單聲道。每個音樂片段之情緒選取的方式採測試人員主觀判斷，全為近似於單一穩定情緒的音樂片段，情緒類別引用 Thayer 情緒模型中的四種情緒成份：(1)舒適的(Content)、(2)哀傷的(Depression)、(3)焦慮的(Anxious)、(4)振奮的(Exuberant)，各個情緒類別的音樂片段數量可以參考表 3。

表 3 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**

訓練資料的情緒在此採用主觀性判斷主要是希望未來本系統在其他使用者的操作下，可以依照使用者自己的情緒感受選取不同的訓練資料來建立系統資料，不僅可以針對個人的使用有較佳分辨效果的分類模型，並藉此大大的降低不同使用者對情緒形容詞主觀的認知落差。

4.3.3 特徵萃取

在進行音訊資料的特徵萃取前，為了表示音訊資料在每個時刻的狀態，必須先對音訊資料做音框化處理，將原始訊號切割為音框長度 2048 個樣本點、重疊長度為 1536 個樣本點的音框，再針對每個音框計算不同音訊特徵的特徵值。

根據音樂心理學的研究與日常經驗顯示，節奏較緊湊的歌曲，通常採用稍快的速度來表現，其所對應到的情緒氛圍屬於較激烈的，而節奏速度是反映在音樂事件的密集程度上；改變音樂的音量大小，對於聆聽者當下的情緒感受有漸層增強的作用；音色透過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，不和諧的背景和聲容易讓人抑鬱，這些都是直接地影響到聆聽者情緒感受的主要音訊特徵，因此，在本研究中主要萃取(i)音樂事件密集度、(ii)音量、(iii)頻譜分佈、(iv)調式、(v)和聲和諧程度等五種特徵進行情緒分析。

- **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，其算式如公式 26：

$$\text{Spectrum Flux } (m) = \sum_{k=0}^{N} H\left(|S_m[k]| - |S_{m-1}[k]|\right)$$
$$\text{where } H(x) = \frac{x + |x|}{2} \tag{26}$$

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：

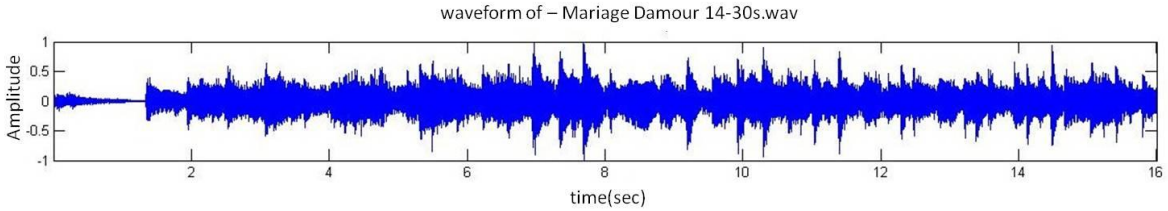


圖 24 音樂片段之原始音樂波形

資料來源：Mariage Damour 14-30s.wav

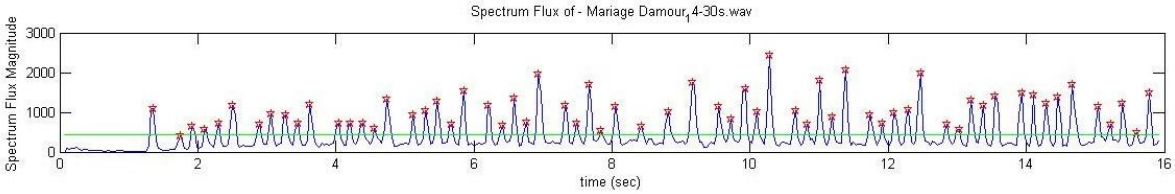


圖 25 音訊頻譜流量進行音樂事件偵測

資料來源：Mariage Damour 14-30s.wav

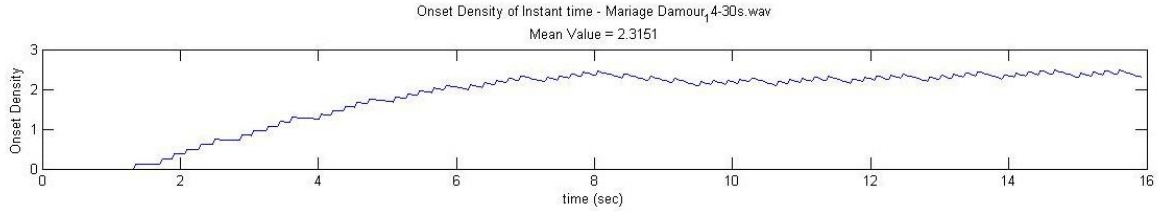


圖 26 音樂事件密集程度計算結果

資料來源：Mariage Damour 14-30s.wav

上圖 24 為 0~20 秒的時域波形，圖 25 則是以公式 26 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 26 所示。

- **響度分析－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：

$$\text{Volume}_1[m] = \sum_{n=1}^{N} |x_m[n]| \tag{27}$$

但由於人的感知系統對於不同頻域音量大小的感受並不相同，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)，亦即本系統採用的計算方式，實際計算方法如下：

$$\text{Volume}_2[m] = 10 \times \log_{10}\left(\sum_{n=1}^{N} x_m[n]^2\right) \tag{28}$$

- **音色分析－頻譜分佈**

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：

$$\text{Spectrum Centroid}[n] = \frac{\sum_{k=1}^{N/2+1} \left(f[k] \cdot |S_n[k]|\right)}{\sum_{k=1}^{N/2+1} |S_n[k]|} \tag{29}$$

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：

$$\text{Bandwidth}[n]=\frac{\sum_{k=1}^{N/2+1}\left(f[k]-\text{SC}[n]\right)\cdot|S_n[k]|}{N/2+1} \tag{30}$$

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(30)式所示。

$$\text{SpectrumEnergy}=\frac{\sum_{k=1}^{N/2+1}\left(f[k]\cdot|S_n[k]|\right)}{\sum_{k=1}^{N/2+1}|S_n[k]|}\cdot\frac{\sum_{k=1}^{N/2+1}\left(f[k]-\text{SC}[n]\right)\cdot|S_n[k]|}{N/2+1} \tag{31}$$

- **調式分析－大小調**

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：

$$P(k)=\left[24\cdot\log_2\left(\frac{f_s}{N}\cdot\frac{k}{f_1}\right)\right]\bmod 24 \tag{32}$$

$$\text{PCP}[P(k),n]=\sum_{P(k)}|S[k,n]| \tag{33}$$

上式將頻譜數值映射到 24 個音調層級上，其中公式(32)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，$24\cdot\log_2\left(\frac{f_s}{N}\cdot\frac{k}{f_1}\right)$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 , …, 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, … , 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。

$$\text{simplified PCP}=\text{PCP}(1{:}2{:}23) \tag{34}$$

$$\text{simplified PCP}=\text{PCP}(2{:}2{:}24) \tag{35}$$

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。

- **調式分析－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。

$$\text{Note1}[n]=\max\left[\text{PCP}(1{:}24,n)\right] \tag{36}$$

$$\text{Note2}[n]=\text{PCP}\left(\text{index of Note1}[n]\pm 12,n\right) \tag{37}$$

$$\text{Dissonance}(n) = \frac{\text{Note2}[n]}{\text{Note1}[n]} \cdot \frac{\text{Note1}[n] + \text{Note2}[n]}{\max\left[\text{PCP(all,all)}\right]} \tag{38}$$

上式(38)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[PCP(all, all)]的比例。

4.3.4 情緒計分方法

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：

$$p_t = \sum_{f=1}^{5}\left[w_x(f) S_f(t)\bar{x} + bias_x + w_y(f) S_f(t)\bar{y} + bias_y \right] \tag{39}$$

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 所述：

表 4 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	2.39	1.09	0.3	0	0
壓力	-0.53	0	0	3.22	3.33

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(39)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[32]，其公式定義如下：

$$ASR = \frac{1}{2N}\sum_{n=0}^{N-1}\left(1 - sgn\left(STE(n) - \rho \times avgSTE\right)\right) \tag{40}$$

$$STE(n) = \sum_{k=0}^{m-1} a_n{}^2(k) \tag{41}$$

$$avgSTE = \frac{1}{N}\sum_{n=0}^{N-1} STE(n) \tag{42}$$

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，*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(41)，$a_n(k)$代表第 n 個音框內第 k 點的音訊振幅之大小數值，*avgSTE* 為 $STE(n)$之平均值如公式(42)，之後再乘以比例參數σ如圖 27；

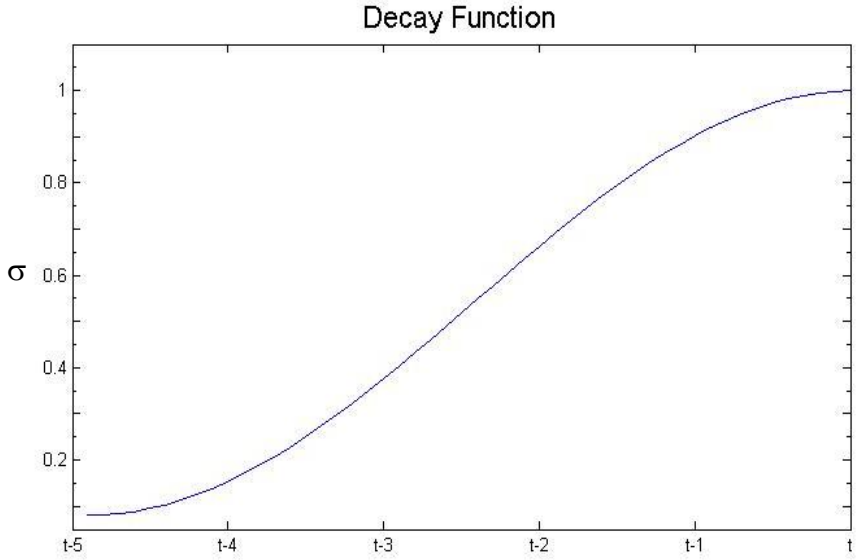


圖 27 衰退函數

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 *ASR* 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 28 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 σ 之合，例如圖 28 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。

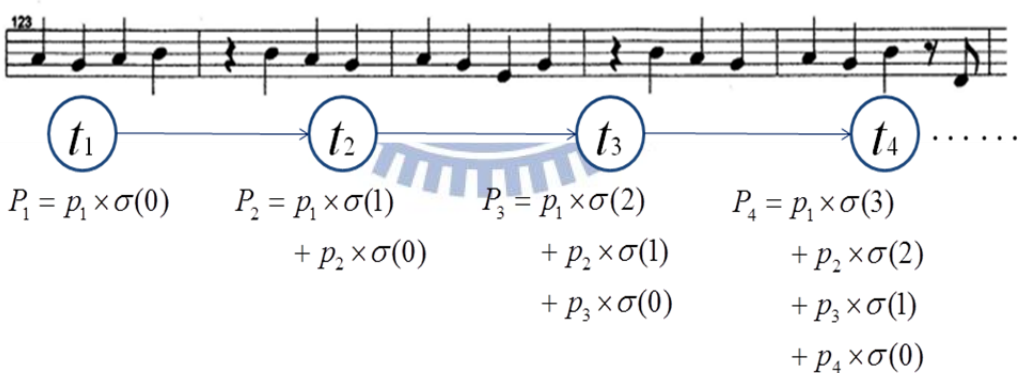


圖 28 每個時間點的計分流程

4.3.5 音樂情緒比例

本系統將訓練資料經由情緒計分公式計算後所得到的情緒計分模型對應到 Thayer 情緒模型的座標平面上會產生一個情緒位移的軌跡，根據不同的音樂特徵強度會得到相對的加減分和位移結果，但是並不存在明確的邊界可以用來界定不同情緒之間的差異，

因此，本系統使用 GMM 分類器依據訓練資料的軌跡分布來定義各個情緒邊界的範圍，這樣的設計概念可以幫助系統利用已訓練好的情緒邊界來辨別新的輸入音樂資料在每個時刻的可能情緒為何，最後依時間長度比例統計整首歌曲在每個時間所屬的情緒指標來作為系統中的情緒靜態比例。

接下來將詳細介紹情緒邊界的訓練和辨識結果之取得方式，如圖 29，說明訓練資料情緒軌跡位移的取得方式，經由情緒計分公式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，因此訓練資料的取得為情緒軌跡位移後半段的部分，也就是說，研究採用的訓練音檔時間長度為三十秒，但在實際分析中只採用音檔後二十秒產生的結果，而每間隔幾個音框擷取一個軌跡作標作為訓練資料，如圖 30。

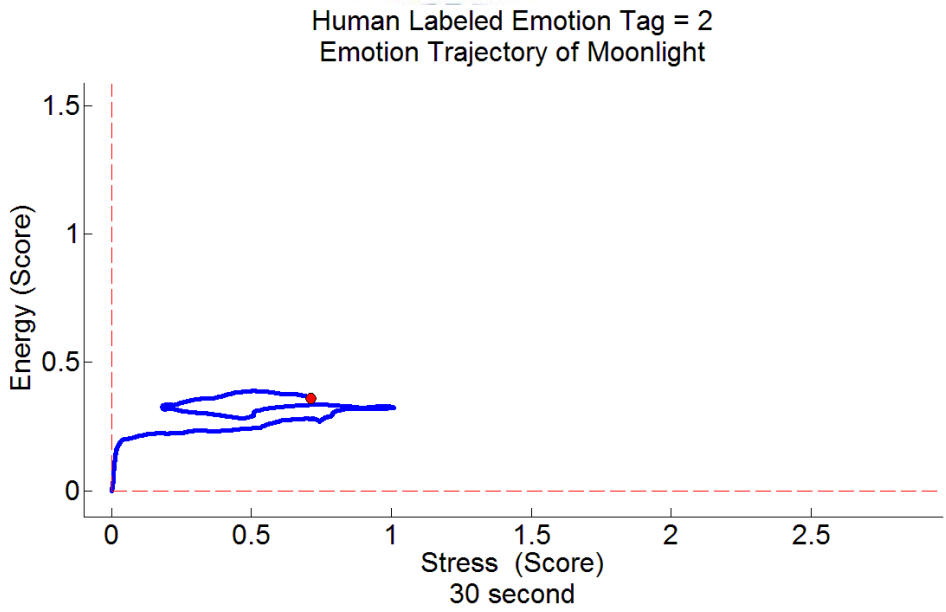


圖 29 貝多芬之月光奏鳴曲-情緒軌跡位移

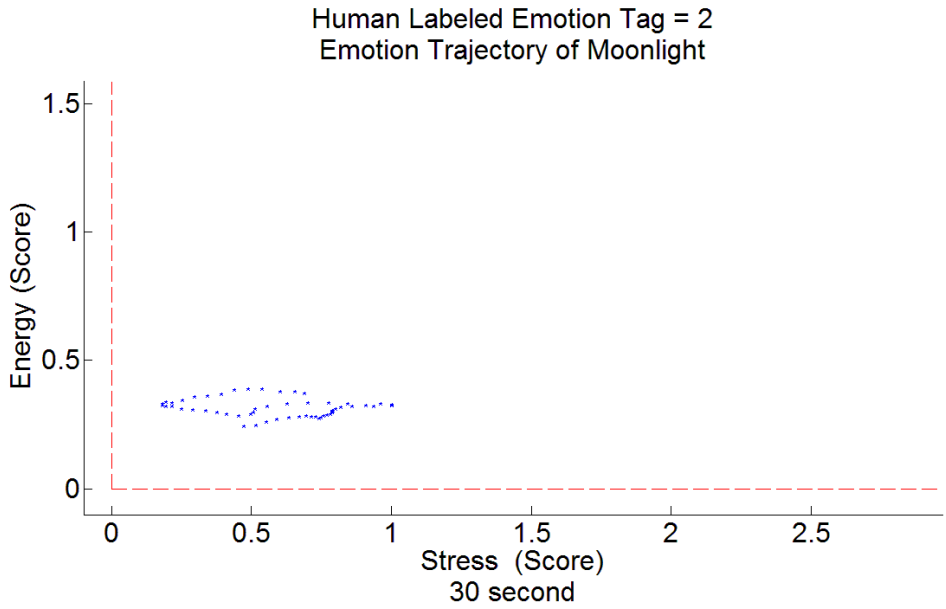


圖 30 貝多芬之月光奏鳴曲-情緒軌跡位移所提供的訓練資料

系統經過200首已標籤單一情緒的訓練音檔經過情緒計分結果的情緒模型之座標對應與軌跡的採樣過程後，將得到所有音樂片段情緒軌跡座標點的分佈圖，如圖 31，雖然不同類別的訓練資料雖然有部分重疊，但大致上可以看到其分佈方式如最初假設一樣，類似 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。

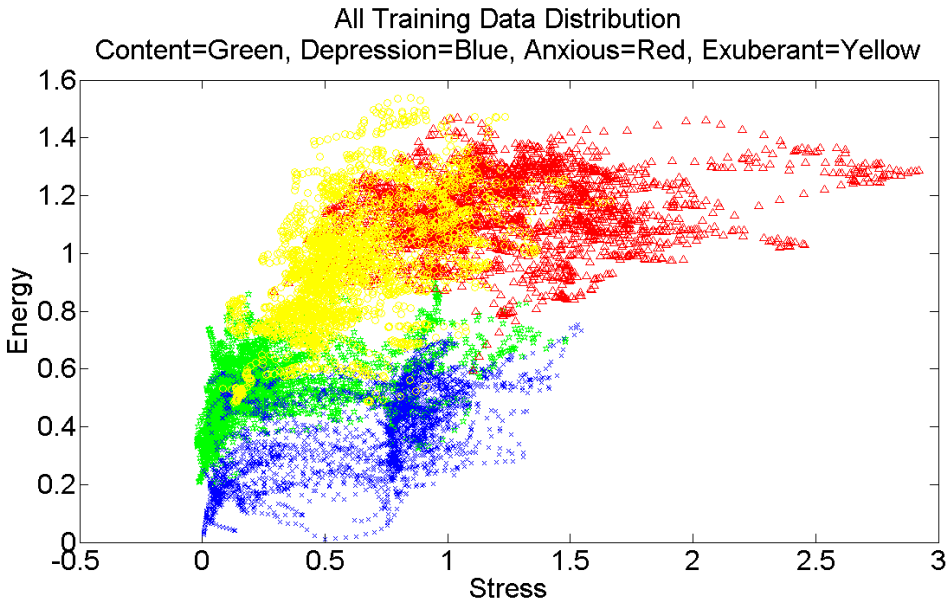


圖 31 訓練資料之情緒樣本分佈

最後，經過四個情緒類別的 GMM 訓練與邊界範圍的計算，可以得到各個類別的機率函數密度(PDF)和等高線的分布情形，最後在將四個情緒類別的 PDF 視為由數個高斯函數疊加的分佈，試找出四個情緒類別之機率函數密度的交叉範圍，即為各個情緒類別的邊界，如圖 32。

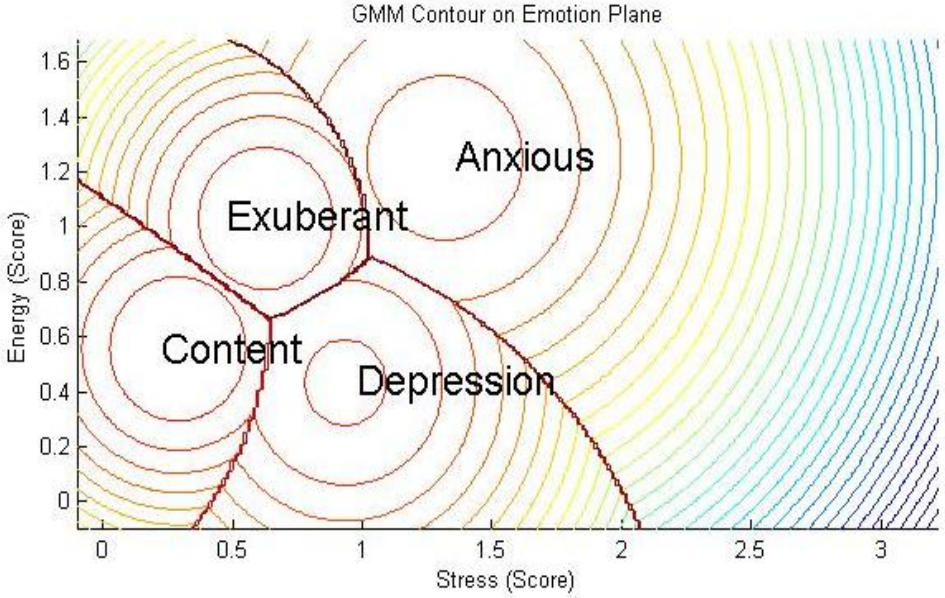


圖 32 GMM 分類結果與各類別的邊界範圍

在系統的測試模式下，當輸入新的音訊資料時，根據情緒計分公式的計算會得到一個新的情緒計分軌跡座標點，系統對照情緒邊界範圍來判別每個軌跡座標點所屬的情緒類別後，再統計四個情緒類別各自累積的時間點，最後經由加總四個情緒類別累積的時間來計算整首音樂內容各自情緒累積的時間作為系統分析的靜態情緒比例。

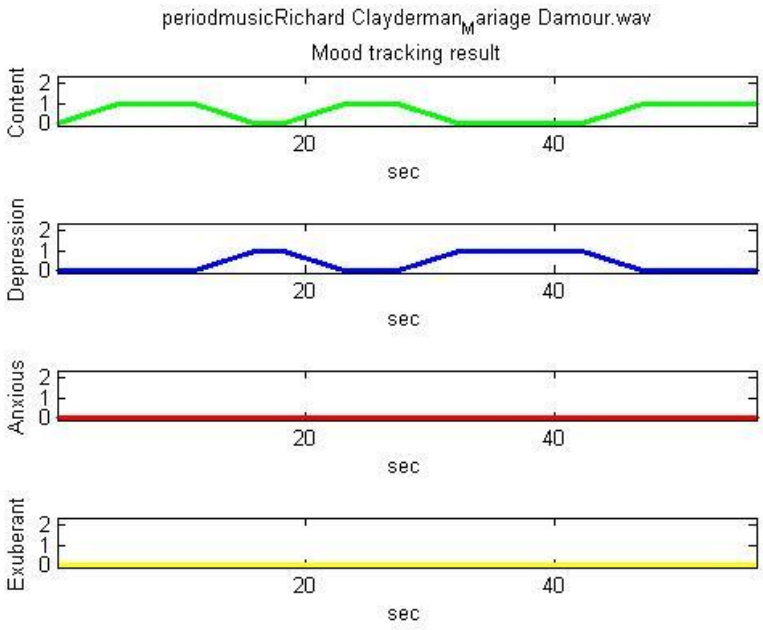


圖 33 情緒類別辨識知結果

以理查克萊德門鋼琴演奏的歌曲－夢中的婚禮(Mariage Damour)為例，原本曲長為兩分四十一秒的歌曲在經過多重主題結構分析的步驟後，被分割成五十七秒長的主題性音樂片段，將此音樂片段輸入系統測試音樂情緒的成份，辨識每秒中音樂片段所屬的情緒類別，如上圖 33 所示，對於長達五十七秒的音樂片段而言，其音樂情緒縮略圖組成的情緒類別為 Content 和 Depression，最後再分別統計時間長度各個情緒類別的時間比例就是系統測量的靜態情緒比例。

4.4 音樂情緒之相似度量測

經由前小節得到歌曲四個情緒的累積時間比例後，再針對任兩首音樂的靜態情緒比例直接以歐基理德距離演算法計算情緒相似度，如圖 34，其計算概念類似音框的相似度量測方法，先將各自情緒比例相減後再平方的距離表示每個情緒類別的相似度，最後再

加總四個情緒類別的相似度來代表兩首歌曲的情緒相似度，亦即整體相似度。

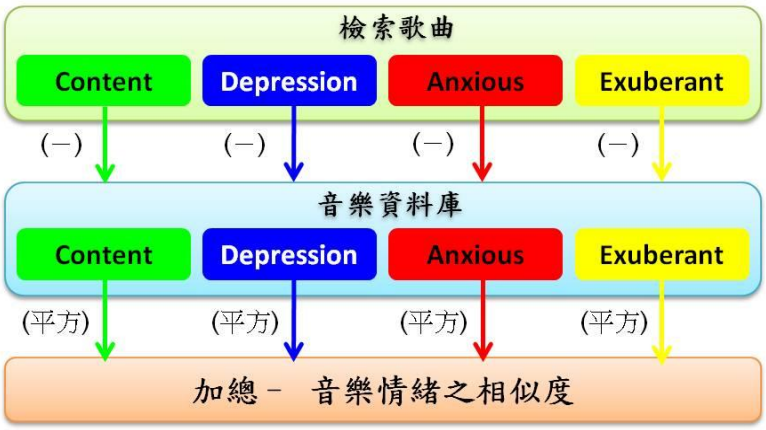


圖 34 情緒相似度之分析概念

對於過去計算歌曲相似度的相關研究中，往往都是先分析每個音框的特徵向量後，再針對特徵向量量測任兩首歌曲之間的局部相似度，最後再量測整體相似度。這樣的分析過程總是占用相當大的記憶體用量，不是造成分析上歌曲長度的限制就是多特徵萃取的選擇，在此，本系統用來計算四個情緒比例的分析方法可以大幅降低電腦的記憶體空間，亦可有效地縮短檢索歌曲情緒比對過程的等待時間，因此我們預期可以藉此方法幫助本研究視覺化音樂推薦系統的管理與應用更為實用。

五、 音樂情緒點唱機

隨著資訊科技的發展，電腦技術的進步，使得數位音樂的取得越來越容易，如何讓使用者從大量的音樂資料中與多變的情境條件下，找出使用者自己喜好的音樂是一件困難的事情，因此音樂自動選曲系統成為主要發展中的應用服務之一。

在建立與管理個人音樂資料庫方面，音樂聆賞者在使用傳統的音樂推薦系統聆賞音樂時，最常遇到兩種問題無非是現存的音樂播放軟體中不存在以情緒為依據之自動選曲功能，因此在過去的聆賞經驗中，個人化需求與差異性行銷的趨勢下，尚未有符合一般認知的音樂自動選曲系統，傳統的音樂自動選曲系統並無法滿足每位使用者，如果能推翻傳統以曲風、樂團或專輯、演唱者及音樂相關資料等作為關鍵字的音樂資料檢索方式，加入基於音樂情緒內容的檢索方式來幫助使用者直覺地依照情緒感受的分類方式來整理大量的音樂檔案，不僅僅提供使用者收聽音樂的訴求，讓使用者自行挑選欲播放的歌曲清單，更幫助使用者可以輕易的獲得新的音樂資訊，解決自行搜尋音樂的困擾。

本研究之目的在於發展一套基於多重結構分析聆聽情緒相似度檢索之音樂情緒點唱機，以音樂情緒比例分類為基礎的視覺化音樂自動選曲系統，根據使用者所選的音樂檔案，利用已標籤好情緒比例的訓練音檔進行音樂資料分析，追蹤檢索音檔之音樂內容喚起聆聽者的情緒感受，記錄在每個時間不同情緒的時間長度，嘗試推薦使用者具有相似情緒成份的音樂。

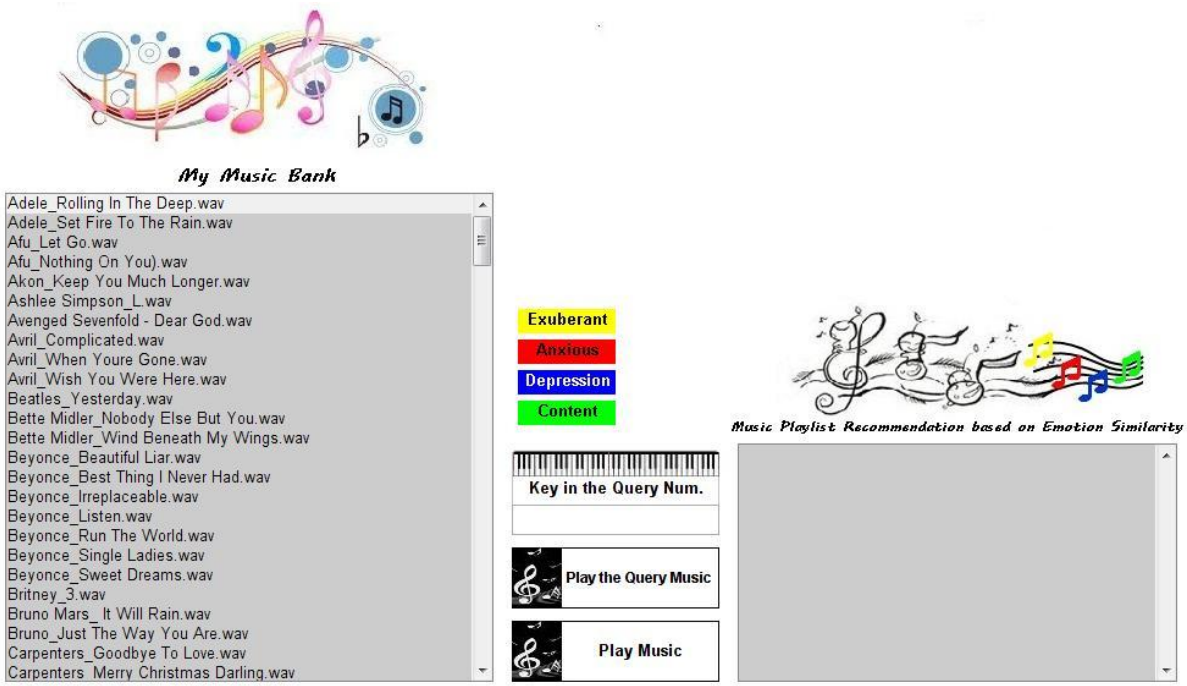


圖 35 音樂情緒點唱機之使用者介面

5.1 圖形化使用者介面

視覺化音樂自動選曲系統如圖 35，為系統之圖形化使用者初始介面，介面總共分為(1)音樂資料庫(My Music Bank)、(2)播放音樂資料庫之歌曲按鈕鍵(Play Music)、(3)欲檢索歌曲數目之輸入(Key in the Query Num.)、(4)音樂情緒點唱機之音樂推薦清單(Music Playist Recommendation based on Emotion Similarity)以及(5)播放音樂推薦清單之歌曲按鈕鍵(Play the Query Music)等五大部分。詳細的系統操作步驟介紹如下：

首先，根據第三章所介紹的理論將測試資料輸入系統分析後而得的最終介面，其中音樂資料庫共 210 首測試歌曲，其中包含各種風格的英文流行歌曲，古典音樂等，作為此系統圖形化使用者介面的音樂資料庫選單和系統事先分析儲存的情緒比例。透過這個介面使用者可以從音樂資料庫選單中清楚地瀏覽每首音樂的情緒比例，對於未曾聆聽過的歌曲，使用者可以很直接地透過即時運算四種情緒元素組成的圓餅圖快速了解歌曲可能誘發的情緒感受，同時也提供使用者多方面的選擇，若使用者想要聆聽與音樂資料庫中具有相同情緒組成的歌曲時，使用者可以自行輸入要檢索相關情緒類型的歌曲數目，

藉由系統即時運算、比對音樂資料庫中的音樂情緒而產生音樂情緒之推薦清單；在使用者選取音樂、了解所選歌曲之情緒組成元素的同時，系統也設計兩個音樂播放鍵讓使用者選擇聆聽音樂資料庫中的檢索歌曲或清單中推薦的相似情緒比例的歌曲，此種分開的音樂播放鍵設計提供使用者更人性化的操作方式，不用在選取歌曲後就一定要聆聽完整的音樂，等待整首歌曲撥放完畢後才可以再選擇其它類型的音樂，使用者可以單只透過此音樂推薦系統了解歌曲的情緒組成部份，並沒有硬性規定聆聽與否。

綜合以上所述，此音樂自動選曲系統除了提供使用者更方便、更直覺化且更多元化的聆聽內容外，更讓使用者能夠輕易地透過簡單的情緒比例圖式找到符合所需的音樂資訊，唱出屬於自己心情的音樂，系統執行完成之圖形化使用者介面如圖 36。

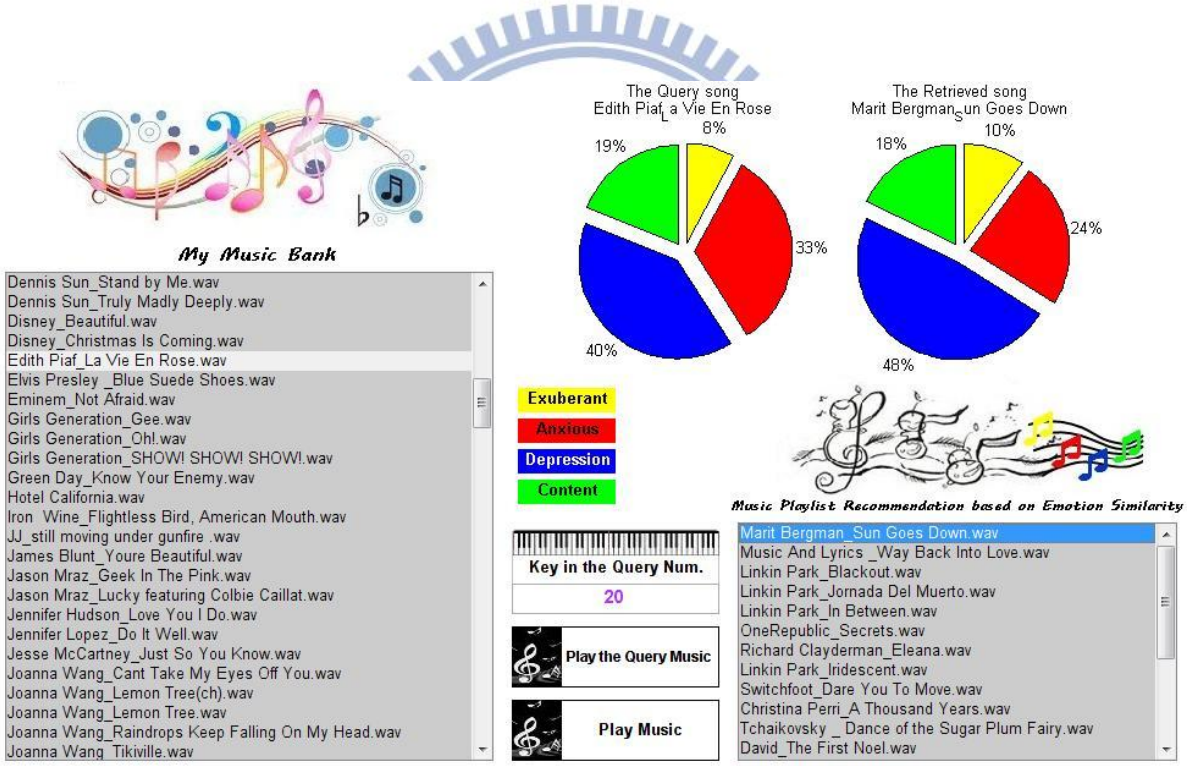


圖 36 系統執行完成後的最終圖形化使用者介面

六、 實驗結果分析

6.1 音樂多重主題結構之擷取結果

為了檢測音樂多重主題結構之音樂片段擷取結果，參考查準率和查全率的公式設計一個測量音樂多重主題結構之萃取結果精確度的公式(43)，在此定義為片段相似度，系統分析音樂片段的長度和手動標記之音樂週期的各自片段的根號長度相乘為分母，比對系統分析音樂片段的長度和手動標記的音樂週期的重疊長度為分子，兩個相除作為系統量測音樂多重主題結構的精確度，所得的相似度分數越高表示系統擷取的結果越精準。

$$Segment\ Similarity = \frac{\text{length of overlapped measure of } v_i \text{ and } u_i}{\sqrt{\text{length } v_i} \times \sqrt{\text{length } u_i}} \quad (43)$$

v_i : period of manual extraction

u_i : period of auto extraction

6.1.1 檢測準確度

接下來將探討如何檢測音樂多重主題結構之結果準確度。對流行歌曲來說，同首歌曲的音樂特徵會隨著不同演唱者的表演方式、聲音語言做改變；古典音樂也會隨著不同的演奏者和演奏樂器而有所改變，因此為了能夠精準檢測音樂多重主題結構之結果準確度，主要分為四個測量方向，詳細介紹如下：

(1) 同一首歌曲，不同語言，不同演唱者改編翻唱的歌曲，如經典代表的電影主題曲“新不了情”，共有五個翻唱版本，每個版本經由人工手動標記的週期也不盡相同，其檢測結果如下表 5。

表 5 歌曲“新不了情”各自版本之多重主題結構的片段相似度結果和總體準確度

Song / Version	Manual	Program	Precision%
新不了情_原唱版	0~127s	0~100s	89%
新不了情_翻唱版本 1	0~103s	0~102s	99%
新不了情_翻唱版本 2	0~138s	0~112s	90%
新不了情_翻唱版本 3	0~125s	0~124s	99%
新不了情_翻唱版本 4	0~125s	0~130s	98%
Totally Precision in Average			95%

(2) 同一首歌曲，相同演唱者、演唱方式，但不同語言，如韓國女子團體主唱的“NOBODY”，分別為韓文版和中文版，表 6。

表 6 歌曲“NOBODY”之多重主題結構的片段相似度結果和總體準確度

Song/Version	Manual	Program	Precision%
Nobody/Chinese version	0~87s	0~118s	86%
Nobody/Korean version	0~87s	0~107s	90%
Totally Precision in Average			88%

(3) 同一首歌曲,不同演唱者,不同的表達語言,如男歌手 Robbie Williams 主唱的“Better man”英文版，由女歌手林憶蓮翻唱成中文版，表 7。

表 7 歌曲“Better man”之多重主題結構的片段相似度結果和總體準確度。

Song/Version	Manual	Program	Precision%
Better Man/Chinese version	0~153s	0~117	87%
Better Man/English version	0~134s	0~133s	99%
Totally Precision in Average			**93%**

(4) 古典音樂，不包含任何聲音語言，存樂器演奏的音樂，如下表 8

表 8 古典歌曲之多重主題結構的片段相似度結果和總體準確度

Classical Song/composer	Manual	Program	Precision%
Nocturne/Chopin	0~88s	0~58s	81%
Mariage Damour/ Paul De Senneville	0~57s	0~56s	99%
Ballade Pour Adeline/ Paul de Senneville	0~54s	0~67s	90%
Pour Elise/Beethoven	0~100s	0~102s	99%
Alla Turca/Mozart	0~124s	0~137s	95%
Totally Precision in Average			**93%**

6.1.2 結果討論

綜合以上各種類型歌曲的總體準確度，含配樂之多重主題結構之總體準確度為 93%；而不含配樂之多重主題結構之總體準確度為 94%，可以很明顯的看出音樂曲式 AABA 在第二次重複表現之前之間奏萃取與否的影響並不大,與我們一開始假設要萃取音樂片段的條件相呼應，可以將間奏部分視為自由片段。

6.2 音樂情緒心理分析調查

聆賞音樂的情緒感受是主觀的，聆聽環境、聆聽者當下的心情、成長背景...等等都是影響情緒感受的主要因素，因此為了驗證系統分析的結果是否具客觀性，有效的推薦使用者搜尋相似情緒的音樂檔案，本研究參考文獻[33]設計一份以心理學家 ZENTNER 提出的情緒模型 GEM-9 做為音樂情緒比例的問卷調查(附錄一)，同時也藉此問卷調查將數理演算分析的音樂情緒和心理學家建立的音樂情緒模型做比對，分析其驗證結果是否互相吻合。

6.2.1 問卷調查

問卷調查對象共計 66 位一年級到四年級的大學生(43 男、23 女)，來自於各系所包含人文、商業、資訊、工程等不同的背景，詳細受測者資料可見附錄二。問卷調查過程採不事先告知聆聽者將進行一個關於音樂情緒辨識問卷的填寫，只是單純針對聆聽音樂做簡單介紹，讓聽者在沒有預設立場下先聆聽歌曲以後再填寫問卷，為了降低聆聽者對問卷上的情緒形容詞在認知上有落差的情形，問卷上也針對各個情緒形容詞標有統一的解釋和翻譯。

表 9 問卷調查之測試音樂

演奏者	歌曲名稱	原曲長度	音樂片段長度
Celine Dion	My Heart Will Go On	00:04:41	00:01:22
Carpenters	Goodbye To Love	00:03:56	00:01:20
Chopin	Nocturne op 9 no 2	00:04:08	00:02:01
Debussy	Claire De Lune	00:04:20	00:01:33
Edith Piaf	La Vie En Rose	00:03:05	00:02:05
Avenged Sevenfold	Dear God	00:06:34	00:01:35

6.2.2 問卷調查與實驗結果分析

問卷調查的測試音樂如上表 9，各測試歌曲的計分統計及情緒比例圖形請參考附錄三，問卷調查中提供的使用音檔為完整長度的音樂內容，而系統使用的音檔為主題性的音樂片段，其設計概念主要提供使用者於選擇聆聽音樂前一個音樂情緒縮略圖的概要代表，如同每本書籍的封面文字簡短介紹一樣，給於聆聽者一個簡單的情緒類別參考。

在此我們針對幾首測試音樂問卷調查的結果與系統分析的實驗結果做分析比對與討論，詳細介紹如下：

- **問卷調查，測試音樂一：My Heart Will Go On，請參考表 10。**

表 10 My Heart Will Go On 之結果分析

類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	11.49	73.27%	Content 27% Depression 73% 合計：100%
	Transcendence	10.8		
	Peacefulness	13.05		
	Tenderness	12.40		
	Nostalgia	13.46		
	Sadness	12.07		
B 類 VITALITY	Power	8.78	16.04%	Exuberance 0%
	Joyful	7.26		
C 類 UNEASE	Tension	11.33	11.33%	Anxious 0%

My Heart Will Go On 一曲以笛聲演奏作為主旋律，輔以電子氛圍長音的表現方式，加上女主唱哀傷的腔調，即使聽不懂英文歌詞的聆聽者亦可感受那浩瀚的悲愴情感，這也許是導致問卷中超然(Transcendence)的得分偏高，懷舊、憂鬱(Nostalgia)的情緒最高的因素之一。

這首曲子經過系統的情緒分析，主要表現在 Depression 的情緒上，超然的感覺也讓 Content 占有不少比例，比較問卷調查和系統分析的結果，可以很明顯的看出其共通點都是以 A 類 SUBLIMITY 的情緒占大部分，因此系統分析的音樂情緒縮略圖在此音樂的分析中是具有客觀性的參考價值。

由於這首歌是鐵達尼號(Titanic)電影主題曲，因此在問卷調查結果是否受聆聽者喚起電影劇情的記憶的影響，也是一個值得探討的議題

- **問卷調查，測試音樂二：Dear God**

問卷調查的結果和實驗結果比較如下表 11。

表 11 Avenged Sevenfold - Dear God 結果分析

類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	9.66	61.65%	Content 32% Depression4 % 合計：36%
	Transcendence	7.93		
	Peacefulness	9.66		
	Tenderness	10.90		
	Nostalgia	13.38		
	Sadness	10.8		
B 類 VITALITY	Power	12.97	23.66%	Exuberance 58%
	Joyful	10.69		
C 類 UNEASE	Tension	14.69	14.69%	Anxious 6%

由上表 11 中的問卷調查結果可以知道，懷舊、憂鬱(Nostalgia)和動力(Power)和激動(Tension)這三個情緒的比例較高，在 Dear God 這首歌之中大致可以分兩個段落，主歌觸人心弦的滑音吉他和鋪層的主唱旋律都是比較抑鬱的，在副歌激昂的歌聲中再一口氣爆發出來，後面一段破音吉他加進來的橋段再接木吉他的平靜，最後在漸漸進到吉他獨奏的橋段。歌曲的編曲安排非常巧妙，一般聆聽者的情緒會漸漸的被帶領到最後的高潮，因此在問卷上會顯示這三種情緒較多。

但在系統的情緒分析步驟中，由於分析的音檔為音樂片段，演奏長度約為一分三十五秒，切割時間點恰好落於副歌激昂爆發結束的時候，因此系統在 A 類 SUBLIMITY 的情緒比例就會少於問卷測試的結果。然而以客觀性來討論系統分析情緒類別的結果，其音樂情緒縮略圖的分析結果也如同問卷調查結果，將整首歌曲三個主要情緒懷舊、憂鬱(Nostalgia)和動力(Power)和激動(Tension)提供給聆聽者參考。

➢ **問卷調查，測試音樂三：Nocturne op 9 no 2**

最後，針對古典音樂 Chopin - Nocturne opus 9 no 2 的問卷結果如下，

表 12 Chopin - Nocturne opus 9 no 2 的結果分析

類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	10.29	73.27%	Content 37% Depression 63% 合計：100%
	Transcendence	14.06		
	Peacefulness	15.18		
	Tenderness	14.38		
	Nostalgia	11.33		
	Sadness	8.03		
B 類 VITALITY	Power	7.87	19.59%	Exuberance 0%
	Joyful	11.72		
C 類 UNEASE	Tension	7.14	7.14%	Anxious 0%

在古典音樂中的音樂內容和流行的歌曲很不相同，比如這首蕭邦的鋼琴獨奏曲，整首歌曲中只有鋼琴的演奏，在音樂內容上的能量普遍偏低，因此問卷調查的情緒比例以 A 類 SUBLIMITY 低能量的情緒感受佔大部分，對照系統分析的結果就如預期一樣，提供聆聽者有效的情緒類別參考價值，亦即 Content 和 Depression 為歌曲的主要情緒類別。

七、 音樂情緒之應用

作曲者藉由音樂傳達想法、情感，而對聆聽者而言，最具影響力的媒介即是音樂，能夠穿越時空，跨越語言進入心靈深處(Premuzic and Furnham, 2007)，同時反應在聆聽者的情緒感受上，本系統提出以音樂情緒比例分類為基礎的視覺化自動選曲系統有別於傳統以音樂風格分類的音樂推薦系統，嘗試提供符合使用者需求或心情的音樂，因此本系統舉凡在音樂資訊檢索、醫療環境、教育環境、或是大眾傳播...等等的應用都有重要的幫助。各種應用環境的詳細介紹如下：

a. 醫療環境-音樂心理治療

當有些話無法用言語表達，有些情緒找不到宣洩的出口，或是當語言或肢體表現無法成為溝通的管道，音樂就成了最佳的媒介，成功的音樂活動可使人們增強社交能力。音樂心理治療透過音樂當媒介，治療者可以直接碰觸到當事人的內心情緒，達到情感的宣洩、心靈的撫慰、強大的支持等等目的，卻不需要當事人掏心挖肺的把自己的傷痛講出來。在(Canadian Association for Music Therapy, 1992) 文獻中定義所謂的音樂治療是「有技巧地使用音樂來恢復、維持、改善病人的生理、心理與情緒問題」。音樂治療的療程屬於「個別化」的療程，相對的，音樂情緒感受也是很主觀的，例如同樣一段音樂，你的文化背景不一樣，生長的過程裡面如果你有接受過這個音樂跟沒有接受過這個音樂，這段音樂也會帶給人們不一樣的感受，如果聆聽者本身對此音樂不甚了解，音樂無法和他產生共鳴，這個音樂療程對他的心理療程就不再存在意義。本系統分析各種音訊特徵來識別整首歌曲的音樂情緒比例，幫助治療師依照各個個案選擇不同情緒比例的合適歌曲，例如：音樂治療用於改善腦傷患者的情緒，腦傷病人之所以會產生情緒問題，不光是有可能因為傷到了腦部司掌情緒的部份，也可能是因為對自己的病況感到焦躁、憂鬱，針對此情形，醫療人員可以讓病人選擇聆聽一首符合自己心情的歌，讓他藉由歌聲來舒發情感，激勵其求生的意志，同時系統列出的音樂情緒推薦清單也提供醫療人員更多有關病患的情緒資訊，醫療人員可透過系統分析病患所選的音樂情緒比例來探知其心理狀態，針對其情緒困擾源來對症下藥，對於一些防衛心較強的精神疾患，亦可藉由

此系統挑選類似情緒比例的歌曲幫助病患改善情緒，在心理學上，更是一種投射作用。

b. 教育環境-兒童音樂教學

近年來專家紛紛發表音樂是開發幼兒右腦最直接也是最有效的方法之一。原因是音樂可以深入大腦的中樞神經，直接刺激右腦α學習腦波，此α學習腦波會把幼兒不一定聽的懂的訊息，自然吸入大腦中並逐漸累積，而且自然地流露在思想及行動中，激發孩子正面的個性及氣質，當α學習腦波活躍時，對孩子的反應力及專注力都有非常重大的幫助。對於 0~6 歲的小孩子，如果能給予有計劃的音樂教育，經常聆聽音樂接受聲音的刺激，α腦波將更形活潑，運作也隨著增強，就能提高人類本能的學習能力。本研究提出的視覺化音樂情緒推薦系統，依照四種鮮明顏色來區分表示情緒比例，提供使用者依照教學情境選擇各種不同情緒比例的歌曲來幫助兒童腦力的開發，協助孩子學習認知音樂概念包含音樂情緒及遵從指令的技能；藉由視覺化的音樂情緒推薦系統幫助家長或老師快速地瀏覽並了解音樂資料庫中的音樂。

舉凡有心理上困擾，希望透過音樂得到協助，讓自己心靈恢復平衡，獲得喜悅與滿足感或兒童心智、語言發展遲緩、自閉、過動、專注力不足、情緒控管等等病患皆可搭配此音樂情緒推薦系統選擇符合使用者情境與情緒因素的音樂做音樂治療。

c. 大眾傳播-廣播電台之應用

廣播媒體與我們日常生活息息相關，對於處於現代生活忙碌的我們，廣播媒體比電視的當日新聞、節目及報紙的昨日訊息更為迅速與親切，無論是在走路、開車或搭乘大眾傳播等上班途中只要轉開收音機或手機程式即可收聽播音員透過麥克風所播放出來的音樂、節目與即時訊息，由於其便利性吸引多數聽眾。本系統建立基於情緒比例分類的音樂資訊檢索方便播音員無論在選擇節目性質的背景音樂亦或心情談天節目的配樂都可以輕鬆篩選音樂，對於音樂廣播電台更是效用無窮，例如：電台 DJ 可以在午茶時刻選擇播放情緒較快樂之輕音樂類型的歌曲；午夜時刻選擇情緒較平靜舒緩的心靈音樂的歌曲等等，電台 DJ 皆可以透過此音樂情緒推薦系統輕鬆的選曲適合的歌曲。

d. 日常生活中的應用

除了醫療環境、教育環境與大眾傳播等的應用以外，此系統亦可提供我們日常生活的普遍需求。當我們想要聆聽以某種情緒占大部分的歌曲卻怎麼想破頭也無法找到適合的歌曲時；當我們想要找尋搭配各種活動的背景音樂；當我們今天心情不好想要聆聽氣氛 high 一點的音樂;當我們早上醒來想要聽一些振奮人心的音樂…等等各種情境皆可透過此系統依情緒比例分析找尋適合的歌曲。

對於系統未來的應用層面也是相當廣泛，隨著觸控面板、手機…等各種整合性電子產品的推陳出新，開發技術的發展，已經大幅改變聆賞者過去聆賞音樂的方式，聆聽音樂不再被限制在音響前面才可以享受音樂帶給我們心裡的情緒感受，取而代之的是隨時隨地都可以聆賞享受到自己所喜歡的音樂，感受音樂所帶給人們的感動。如果能將本系統的操作平臺從電腦程式移植於手機、音樂播放器等各種觸控面板或者將本系統加入電腦的音樂播放軟體中都是相當值得研究開發的方向。

在建立與管理個人音樂資料庫方面，音樂聆賞者在使用數位音樂檔案聆賞音樂時，最常遇到兩種問題無非是現存的音樂播放軟體中不存在以情緒為依據之自動選曲功能，因此在過去的聆賞經驗中，尚未有符合一般認知的自動音樂選曲系統。如果能推翻傳統以曲風、樂團或專輯、演唱者及音樂相關資料等作為關鍵字的音樂資料檢索方式，加入基於音樂情緒內容的檢索方式來幫助使用者直覺地依照情緒感受的分類方式來整理大量的音樂檔案，一定更增加其便捷性也更符合人性直覺的作法。

八、 結論

8.1 論文貢獻

本篇論文主要是應用內涵式音樂資訊檢索技術建構一個具時變情緒比例的音樂內容搜尋引擎，以音樂多重結構概括描述完整的音樂檔案，作為該歌曲的音樂縮略圖(音樂摘要)，目標為能夠快速有效地找出符合一般認知及情緒感受的音樂資訊檢索演算法，以音樂理論為基礎設計特徵萃取-情緒計分的分析公式，建立音樂資料的情緒指標數值，同時為了解決在使用者大量數位音樂資料庫中挑選歌曲，各個音框做相似比對所需要花費大量時間的問題，本研究以四種音樂情緒比例作為檢索的音樂情緒標籤值，最後以相似度量測演算法計算音樂資料中四種音樂情緒相似的內容，進而產生與音樂資料庫中符合相同、類似情緒組成的音樂推薦清單。

另外，在傳統的音樂資料檢索系統中大多認為音樂只會造成單一種穩定的聆聽情緒反應，過去以單一種情緒分類音樂，這樣的作法無法符合聆聽音樂時隨著音樂高潮迭起的直覺感受，因此本研究認為聆聽音樂的情緒是隨著歌曲的行進，有起、承、轉、合等持續的情緒變化，每首歌曲誘發情緒的比例只是在於不同的分配情況。

8.2 結論

在音樂多重結構分析的測試實驗中，綜合四種測試方法的總體準確度為 93%，結果顯示藉由系統分析的音樂片段週期符合一般人們對於音樂多重結構之音樂片段週期的認知與感受；同時為客觀的驗證系統情緒分析的成效，更採用問卷方式之調查結果做比對，結果顯示對於整首歌曲情緒組成最多的情緒元素系統確實可以以音樂縮略圖(音樂摘要)分析之情緒概括整首歌曲且準確判斷，另外，或許是調查對象本身的主觀因素影響下或不同音樂族群的愛好者對於歌曲的解讀有不一樣的看法，調查比對結果尚存小部分之結果差異，但由於本研究系統的測試資料可以由使用者自行定義的歌曲來組成的音樂資料庫，每個使用者可以輸入自己感受的情緒音樂片段至系統訓練進而建立各自專屬的情緒分類模型，這樣的做法也可以大大降低主觀感受的認知落差。

九、 參考文獻

1. R. Typke, F. Wiering, and R. C. Veltkamp. ‘‘A Survey of Music Information Retrieval Systems.’’ In DAFX-05: Proceedings of the 8th Int. Conference on Digital Audio Effects, pages 153–160, 2005.
2. H. Shih, Shrikanth S. Narayanan ‘‘Automatic Main Melody Extraction from Midi Files with a Modified Lempel-Ziv Algorithm.’’ Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, pp. 9-12, 2001.
3. Zhao Fang, Wu Yadong, ‘‘Melody Extraction Method from Polyphonic Midi Based on Melodic Features.’’ Compute Engineering, Publishing House of Journal of Computer Engineering, pp. 165-167, Beijing, China, 2007.
4. Bin Cui, Jialie Shen , Gao Cong , Heng Tao Shen and Cui Yu, “Exploring Composite Acoustic Features for Efficient Music Similarity Query. ” In Proc. of the Multimedia, Santa Barbara, California, USA, October 23-27, 2006.
5. Y.H. Tseng, ‘‘Music Indexing and Retrieval for Digital Music Libraries.’’ Proceedings of The First International Workshop on Intelligent Multimedia Computing and Networking (in The Fifth Joint Conference on Information Sciences), Atlantic City, NJ USA, Vol. 2, pp.533-536, Feb. 27 to Mar. 3, 2000.
6. Kyu-Sik Park, Won-Jung Yoon, Kang-Kue Lee, Sang-Heon On and Ki-Man Kim, “MRTB Framework: A Robust Content-Based Music Retrieval and Browsing.” In IEEE Transactions on Consumer Electronics , Vol.51 , No.1 , February 2005.
7. J. Foote , Matthew Cooper ,‘Unjung Nam , “Audio Retrieval by Rhythmic Similarity.”, In Proc. of Institute Research Coordination Acoustics Music (IRCAM) , 2002.
8. O. Lartillot, “Discovering Musical Patterns through Perceptive Heuristics.” Proc. of International Symposium on Music Information Retrieval, ISMIR'03, 2003.
9. M. Cooper, and J. Foote, “Automatic Music Summarization via Similarity Analysis.”, In

Proc. Int. Conf. Music Information Retrieval, 2002.

10. J. Foote, “Visualizing Music and Audio using Self-Similarity.”, Proceedings of the seventh ACM international conference on Multimedia.
11. M. Cooper, J. Foote, ‘‘Summarizing Popular Music via Structural Similarity Analysis.’’ IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2003.
12. M. Wang, L. Lu, and H.J. Zhang, “Repeating Pattern Discovery from Acoustic Musical Signals.”, Advances in Machine Learning and Cybernetics, Lecture Notes in Computer Science, pp.249-257.
13. W. Chai, “Music Thumbnailing via Structural Analysis.” Proceedings of ACM Multimedia Conference , pp.223-226.
14. L. Xiao, J. Zhou, “Using Chroma Histogram to Measure the Perceptual Similarity of Music.”, ICME, pp.1317-1320.
15. A. Tian, W. Li, L. Xiao, D. Wang, J. Zhou, and T. Zhang, “Histogram Matching for Music Repetition Detection.”, ICME, pp.662-665, 2009.
16. Patrik N. Juslin and John A. Sloboda (Eds.), ‘‘Handbook of Music and Emotion: Theory, Research, Applications.’’ Oxford: Oxford University Press, Ch08.indd, 2010.
17. K. Hevner, “The Affective Value of Pitch and Tempo in Music.” The American Journal of Psychology, vol. 49, no. 4, pp. 621-630, Oct. 1937.
18. Farnsworth., R. Paul, ‘‘The Social Psychology of Music.’’ The Dryden Press, 1958.
19. J.A. Russell, “A Circumplex Model of Affect.” Journal of Personality and Social Psychology Vol.39, No.6, pp.1161-1178, 1980.
20. D. Watson and A. Tellegen, ‘‘Toward a Consensual Structure of Mood.’’ Psychol. Bull. 98,219–235, 1985.
21. Thayer, R. E., ‘‘The Biopsychology of Mood and Arousal.’’ Oxford University Press, 1989.

22. 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.

23. A. Uitdenbogerd, and J. Zobel. “Manipulation of Music for Melody Matching.” In Proceedings of ACM Multimedia 98, September 11-15, 1998.

24. J.L. Hsu, C.C. Liu, and A.L.P. Chen. “Discovering Non-trivial Repeating Patterns in Music Data.”, IEEE Transactions on Multimedia, 2001.

25. A. Ghias, J. Logan, D. Chamberlain, B. C. Smith, “Query by Humming-Musical Information Retrieval in an Audio Database.”, ACM Multimedia, San Francisco, 1995.

26. W. Lee, and A.L.P. Chen. “Efficient Multi-Feature Index Structures for Music Data Retrieval.” In Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases, 2000.

27. 董信宗，“流行音樂組曲之電腦音樂編曲”，政治大學資訊科學系碩士論文，民國 96 年

28. J.B. Li, S.N. He ,H. Zheng , Z.X. Niu, “Representative Excerpts Extraction from Music.” Communications, Circuits and Systems Proceedings, 2006 International Conference on , vol.1, no., pp.158-161, June 2006.

29. B.L. Smith, “A Comparison and Evaluation of Approaches to the Automatic Formal Analysis of Musical Audio.” A thesis Submitted to McGill University, Canada, 2010.

30. 李永剛，“實用歌曲作法”，全音樂譜出版社，民國 71 年

31. L. Lu, H. Zhang, “Automated Extraction of Music Snippets”, Proceedings of the eleventh ACM international conference on Multimedia, 2003.

32. 董德文，“曲式分析與寫作應用作曲理論”，花山文藝出版社

33. MIR toolbox, https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

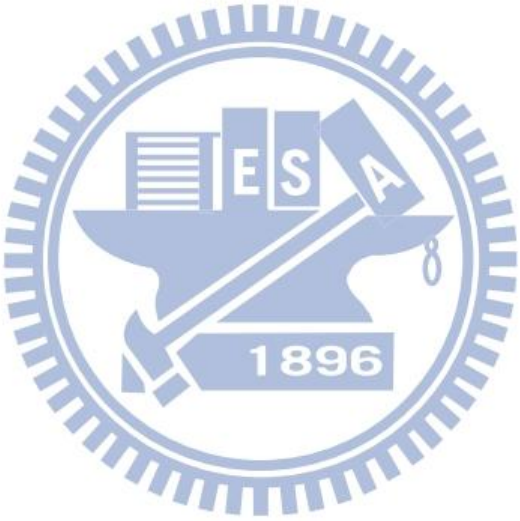
34. Olivier Lartillot, “MIRtoolbox1.3.3 User’s Manual.”, Finnish Centre of Exce!ence in

Interdisciplinary Music Research, University of Jyväskylä, Finland, June, 2011.

35. Olivier Lartillot, Petri Toiviainen ,“A Matlab Toolbox for Musical Feature Extraction From Audio.”, the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, 2007.

36. Jun-Jie Fu, “Emotion Locus Tracking System for Automatic Mood Detection and Classification of Music Signals.” A thesis Submitted to National Chiao Tung University, Taiwan, 2010.

37. Li-Wei Lin, “Tracking the Real-Time Emotional Response of Music Signals.” A thesis Submitted to National Chiao Tung University, Taiwan, 2011.



附錄一　音樂情緒分析之問卷範例

姓名:_____ 學號:_____ 系所年級:_____

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:

演唱者:

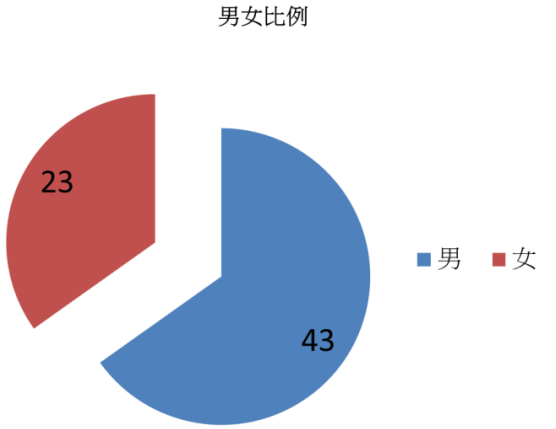
曲風:

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*).

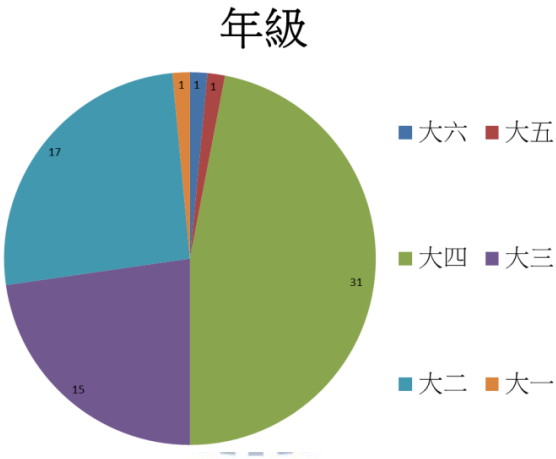
	1 Not at all　2 Somewhat　3 Moderately　4 Quite a lot　5 Very Much	1	2	3	4	5
1	**Wonder (驚異; 讚歎):** Filled with wonder, Dazzled, Moved					
2	**Transcendence(超然; 靈性):** Fascinated, Overwhelmed, Feelings of spirituality					
3	**Power(動力的; 喜悅的; 狂歡的):** Strong, Triumphant, Energetic					
4	**Tenderness(溫和;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

附錄二　問卷調查之受測者資料

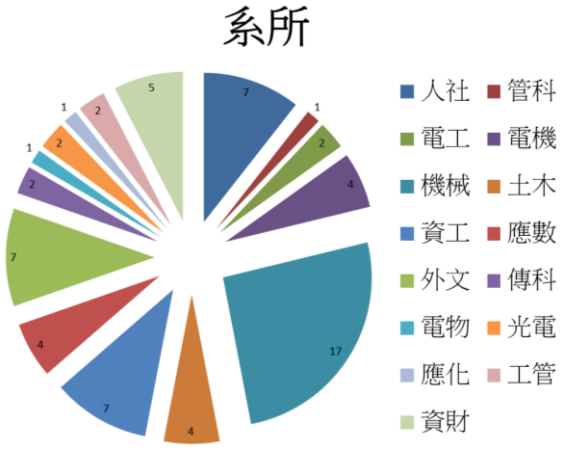
1. 男女分布



2. 年級分布



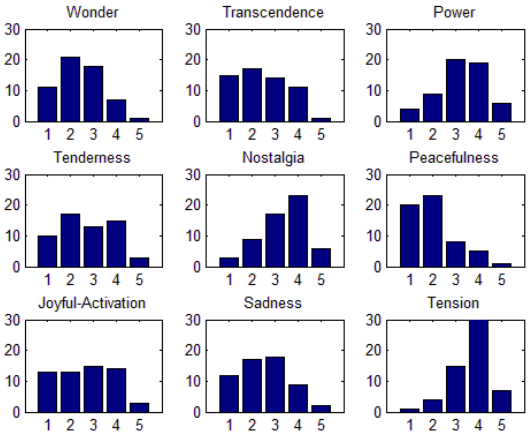
3. 系別分布



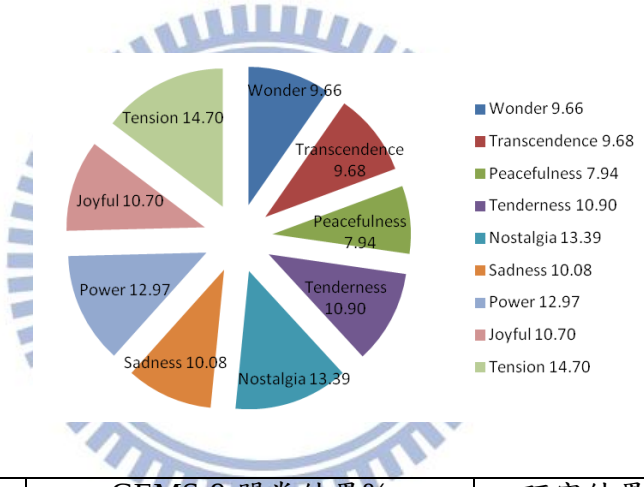
附錄三　測試音樂之問卷調查結果

＜Avenged Sevenfold　－　Dear God＞

- 問卷計分統計



- 問卷情緒比例

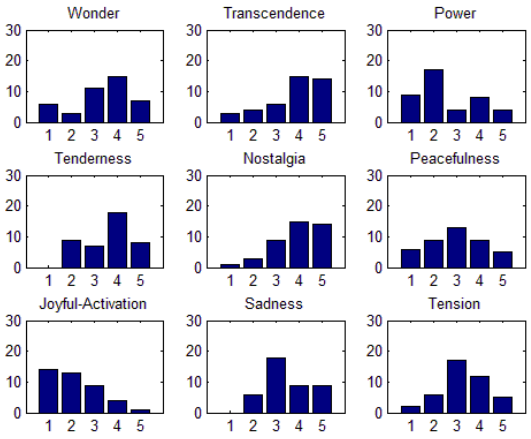


- 結果分析

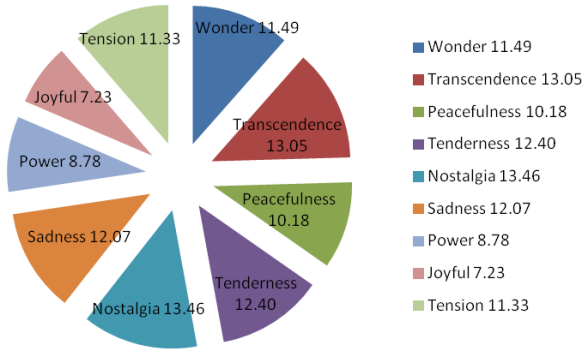
類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	9.66	61.65%	Content 32% Depression4 % 合計：36%
	Transcendence	9.68		
	Peacefulness	7.94		
	Tenderness	10.90		
	Nostalgia	13.39		
	Sadness	10.08		
B 類 VITALITY	Power	12.97	23.67%	Exuberance 58%
	Joyful	10.70		
C 類 UNEASE	Tension	14.70	14.70%	Anxious 6%

＜**Celine Dion — My heart will go on**＞

- 問卷計分統計



- 問卷情緒比例

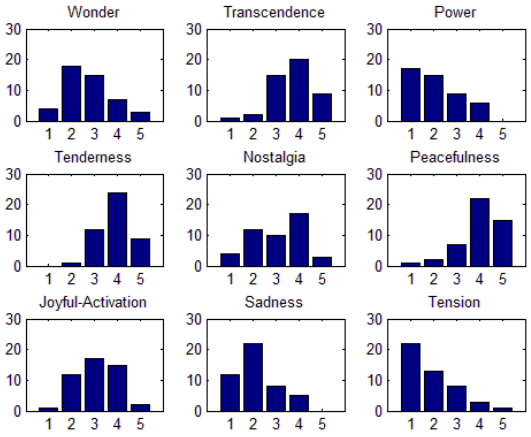


- 結果分析

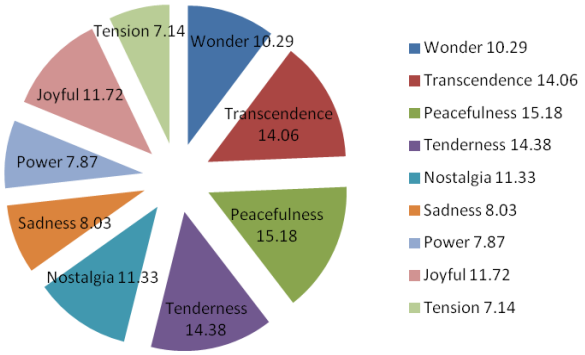
類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	11.49	72.65%	Content 27% Depression 73% 合計：100%
	Transcendence	13.05		
	Peacefulness	10.18		
	Tenderness	12.40		
	Nostalgia	13.46		
	Sadness	12.07		
B 類 VITALITY	Power	8.78	16.01%	Exuberance 0%
	Joyful	7.23		
C 類 UNEASE	Tension	11.33	11.33%	Anxious 0%

＜**Chopin — Nocturne opus 9 no 2**＞

- 問卷計分統計



- 問卷情緒比例

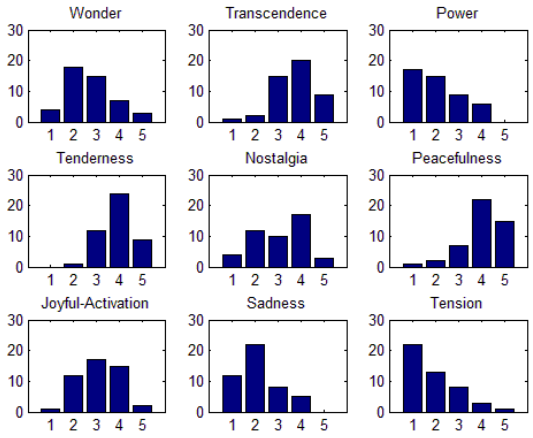


- 結果分析

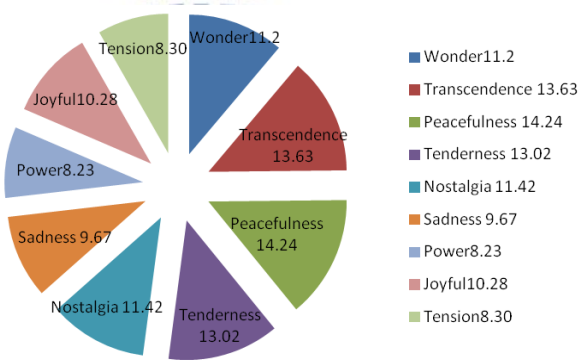
類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	10.29	73.27%	Content 37% Depression 63% 合計：100%
	Transcendence	14.06		
	Peacefulness	15.18		
	Tenderness	14.38		
	Nostalgia	11.33		
	Sadness	8.03		
B 類 VITALITY	Power	7.87	19.59%	Exuberance 0%
	Joyful	11.72		
C 類 UNEASE	Tension	7.14	7.14%	Anxious 0%

＜**Debussy － Claire de lune**＞

- 問卷計分統計



- 問卷情緒比例

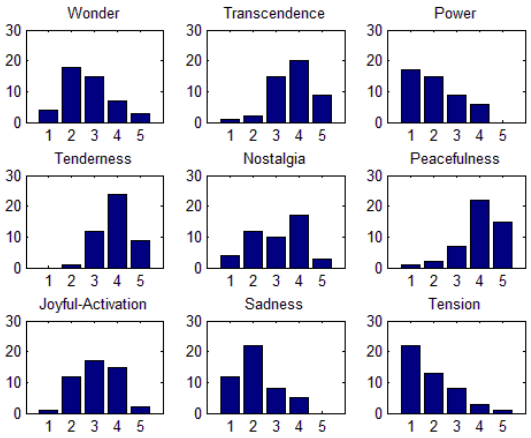


- 結果分析

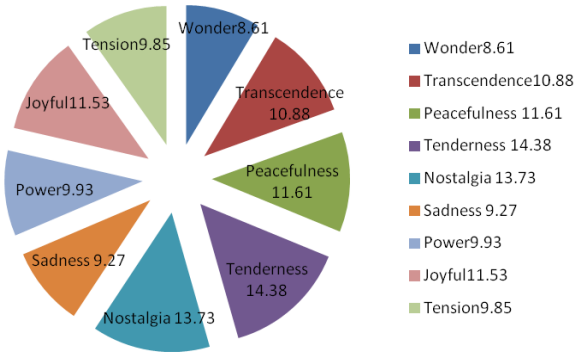
類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	11.20	73.18%	Content 100%
	Transcendence	13.63		
	Peacefulness	14.24		
	Tenderness	13.02		
	Nostalgia	11.42		
	Sadness	9.67		
B 類 VITALITY	Power	8.23	18.51%	Exuberance 0%
	Joyful	10.28		
C 類 UNEASE	Tension	8.30	8.30%	Anxious 0%

＜**Carpenters － Goodbye to Love**＞

- 問卷計分統計



- 問卷情緒比例

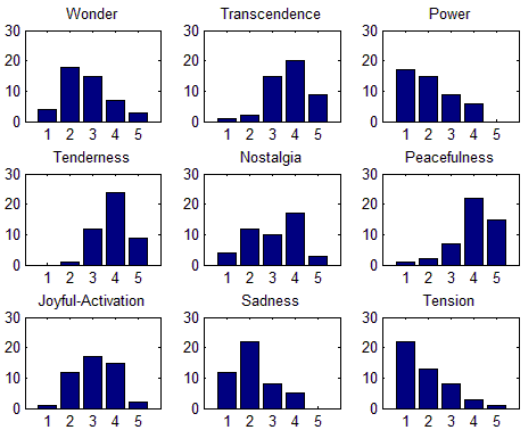


- 結果分析

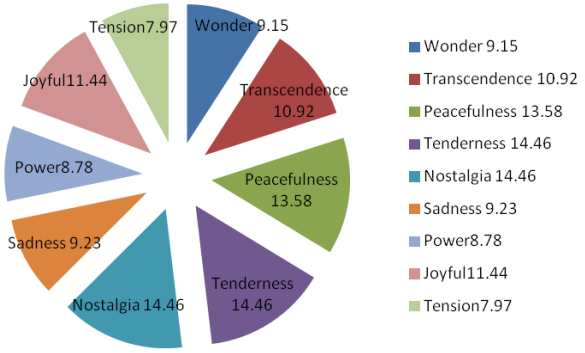
類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	8.61	68.48%	Content 86% Depression 14% 合計：100%
	Transcendence	10.88		
	Peacefulness	11.61		
	Tenderness	14.38		
	Nostalgia	13.73		
	Sadness	9.27		
B 類 VITALITY	Power	9.93	21.46%	Exuberance 0%
	Joyful	11.53		
C 類 UNEASE	Tension	9.85	9.85%	Anxious 0%

＜Edith Piaf － La Vie En Rose＞

- 問卷計分統計



- 問卷情緒比例



- 結果分析

類組	GEMS-9 問卷結果%			研究結果%
A 類 SUBLIMITY	Wonder	9.15	71.80%	Content 19% Depression 40% 合計：59
	Transcendence	10.92		
	Peacefulness	13.58		
	Tenderness	14.46		
	Nostalgia	14.46		
	Sadness	9.23		
B 類 VITALITY	Power	8.78	20.22%	Exuberance 8%
	Joyful	11.44		
C 類 UNEASE	Tension	7.97	7.97%	Anxious 33%