

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



以情緒感受為基礎之自動音樂選曲系統

Automatic Music Track Selector Based on Emotional
Appeal from the Similarity of Acoustic Features

研究生：曾于恬

指導教授：鄭泗東 教授

中華民國九十九年六月

以情緒感受為基礎之自動音樂選曲系統

Automatic Music Track Selector Based on Emotional Appeal from the
Similarity of Acoustic Features

研究生：曾于恬

Student：Yu-Tien Tseng

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



Submitted to Master Program of Sound and Music Innovative Technologies
College of Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Master
in

Engineering

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

以情緒感受為基礎之自動音樂選曲系統

學生：曾于恬

指導教授：鄭泗東

國立交通大學工學院聲音與音樂創意科技碩士學位學程

摘要

音樂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：(1)需要花費大量時間整理並管理數位音樂檔案資料庫；(2)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。本文應用內涵式音樂資訊檢索技術開發基於情緒感受之自動化音樂選曲系統以解決上述的兩項問題。研究中以基於音樂理論之聲學特徵萃取音樂訊號之聲學特徵，並以相似度量測演算法計算得出相似的音樂檔案，最後以主觀標記的音樂類型與情緒感受等資料判斷檢索結果是否符合一般認知，並測試所有演算法的檢索效能。在本次研究準備之測試音樂內容之下，音樂類型的測試裡最高檢索效能達 94.17%；在情緒感受的測試裡檢索效能最高有 98.75%。

關鍵字：內涵式音樂資訊檢索、特徵萃取、自動選曲

Automatic Music Track Selector Based on Emotional Appeal from the Similarity of Acoustic Features

Student : Yu-Tien Tseng

Advisor : Stone Cheng

Master Program of Sound and Music Innovative Technologies

College of Engineering

National Chiao Tung University

ABSTRACT

There are two main issues for listeners when they using digital music files. First, manual music classification and management in digital music library take a large amount of time. Second there is no automatic music track selector which is based on emotional appeal in existed audio player. Therefore, in the past there is no such system which can choose music tracks automatically and fits music common sense at the same time. This paper proposes a system which can automatically select music tracks based on emotional appeal to solve the two issues that we have mentioned and this system is an application in content-based retrieval. In this research, the acoustic features are extracted from music signals and based on music theory. After that similarity functions are used to pick the similar music files. Finally evaluation the preferences based on music common sense. Using the music content that we prepare for subjective tests and evaluation methodologies that we design get the maximum average precision in music genre test is 94.17% and in emotion test is 98.75%.

Key words: content-based retrieval, feature extraction, automatic music track selector

誌謝

在交大就讀碩士班將告一段落，這兩年內首先要感謝指導教授鄭泗東博士的諄諄教誨，在研究方面與論文寫作、簡報上的討論與提點使我受益良多；感謝白明憲博士、黃志方博士、曾毓忠博士在教學課堂中開啟知識上的廣度。接著感謝實驗室的學長、同學及學程的同學在研究及課業上的協助以及幫助我快速地適應碩士班的生活，同時也感謝學弟的協助使我論文得以順利完成。此外要感謝朋友的陪伴與加油打氣，謝謝你們在忙碌的工作或研究中，抽出空閒的時間與我談天、假日時一同出遊紓解壓力。最後要感謝家人長久以來的支持與鼓勵，讓我可以沒有後顧之憂的完成這本論文。僅以此文致上我最由衷的感謝。



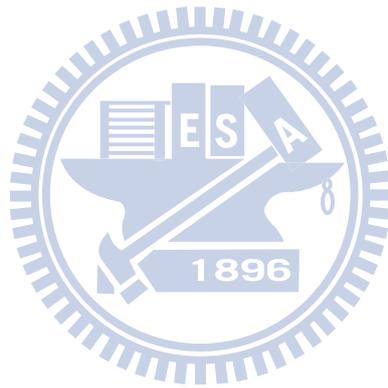
目錄

摘要.....	i
ABSTRACT.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
第一章、緒論.....	1
1.1、研究動機.....	1
1.2、研究方法概論.....	3
1.3、論文貢獻.....	4
1.4、章節概述.....	5
第二章、理論基礎與文獻探討.....	6
2.1、背景知識.....	6
2.2、內涵式音樂資訊檢索.....	7
2.2.1、由符號資料檢索.....	7
2.2.2、由音頻資料檢索.....	8
2.3、自動化歌曲選擇系統.....	9
2.3.1、由文字資料檢索.....	9
2.3.2、由音樂內容檢索.....	10
2.4、音訊前處理.....	11
2.5、特徵萃取.....	11
2.5.1、聲學特徵.....	11
2.5.2、音樂特徵.....	14
2.6、相似度量測.....	15
2.7、音樂心理學.....	16
2.7.1、情緒模型.....	17
2.7.2、音樂與情緒感受.....	18
第三章、研究方法.....	19
3.1、研究假設.....	19
3.2、系統架構.....	20
3.3、資料前處理.....	21

3.4、特徵萃取演算法.....	21
3.4.1、頻譜特徵.....	23
3.4.2、音程特徵.....	25
3.4.3、節奏特徵.....	26
3.4.4、音色特徵.....	27
3.5、相似度量測演算法.....	28
3.6、人機互動分析.....	30
第四章、系統實作及實驗討論	32
4.1、系統實作.....	32
4.1.1、資料標記.....	33
4.1.2、內容分析.....	34
4.2、主觀測試.....	35
4.2.1、音樂類型測試.....	35
4.2.2、情緒感受測試.....	35
4.3、特徵萃取實驗.....	36
4.3.1、訊號處理演算法.....	36
4.3.2、數值統計演算法.....	38
4.4、相似度演算法實驗.....	39
4.5、執行時間比較.....	41
4.6、實驗討論.....	42
4.7、圖形化使用者介面.....	44
第五章、結論與未來展望	46
5.1、結論.....	46
5.2、未來展望.....	46

表目錄

表 1：訊號處理演算法檢索效能比較數據表	37
表 2：數值統計演算法檢索效能比較數據表	38
表 3：相似度量測演算法檢索效能比較數據表	40
表 4：訊號處理演算法運算時間比較	41
表 5：數值統計演算法運算時間比較	41
表 6：相似度量測演算法運算時間比較	41
表 7：系統中各步驟運算時間比較	42



圖目錄

圖 1：研究架構流程圖	4
圖 2：HEVNER 形容詞圈 [47], P249	18
圖 3：細部系統架構流程圖	20
圖 4：資料前處理流程圖	21
圖 5：特徵萃取流程圖	22
圖 6：頻譜特徵萃取流程圖	25
圖 7：音程特徵萃取流程圖	26
圖 8：節奏特徵萃取流程圖	27
圖 9：音色特徵萃取流程圖	28
圖 10：相似度量測流程圖	29
圖 11：系統架構與人機互動分析	31
圖 12：細部系統架構流程圖	33
圖 13：測試音樂資料庫音樂類型分析直條圖	34
圖 14：測試音樂資料庫情緒感受分析直條圖	34
圖 15：音樂類型之檢索效能	35
圖 16：情緒感受之檢索效能	36
圖 17：訊號處理演算法檢索效能比較 (長時距音樂片段分析)	37
圖 18：訊號處理演算法檢索效能比較 (短時距音樂片段分析)	38
圖 19：數值統計演算法檢索效能比較 (長時距音樂片段分析)	39
圖 20：數值統計演算法檢索效能比較 (短時距音樂片段分析)	39
圖 21：相似度演算法檢索效能比較 (長時距音樂片段分析)	40
圖 22：相似度演算法檢索效能比較 (短時距音樂片段分析)	40
圖 23：圖形化使用者介面程式執行完成之畫面	45

第一章、緒論

1.1、研究動機

隨著資訊科技的發展已漸漸改變人們聆聽音樂的習慣。過去從使用黑膠唱片、錄音帶、CD 作為音樂載體，這樣的載體有資料儲存容量上的限制，當一張唱片播放完畢必須動手拿出下一張唱片並放入音樂播放器播放才能繼續聆聽音樂。然而電腦科技的進步，硬碟儲存容量增大，運算速度加快，軟體的進展，越來越多人習慣將實體的音樂 CD 轉換成數位音樂檔案儲存於電腦硬碟中並直接在電腦上播放聆賞，這樣的作法可以一次播放多張專輯唱片，不需要動手更換唱片及播放下一張音樂唱片。數位的聆賞方式改變了舊有的使用習慣，軟體的播放方式也打破專輯的限制，不再只能按照唱片出版實所編排的順序播放音樂，聆賞音樂者可以依照個人喜好或不同的需求自行編排曲目順序，這樣的作法提供音樂聆賞者更多創意發揮的空間。然而當儲存的音樂檔案數量越趨龐大時，在管理上卻容易出現問題，並對使用者產生困擾。

洪元元 [1] 在民國 98 年發表的研究結果歸納出十種現有的線上音樂網站分類架構：分別是(1)語言/國家/地區；(2)日期/年代/時期；(3)音樂廠牌；(4)創作者/演出者個人訊息；(5)載體型態；(6)演奏樂器；(7)音樂類型；(8)行銷需求；(9)情境式分類；(10)俗名分類。在情境式分類下又可分為：(1)節慶；(2)場合/活動；(3)心情；(4)時序；(5)音樂適合對象。

使用者在聆賞音樂的動機分為無特殊動機、自發性動機、誘發性動機。而聆賞音樂的目的則包括休閒娛樂、情感回憶、資訊獲取、教學或學習音樂、應用音樂特性強化行為或情境、社會互動、形象樹立、執行工作任務即滿足從式音樂工作的欲求。

使用在尋求音樂的過程中會遭遇到音樂難以描述的問題。另外音樂需要抽象想像，要如何符合心情、情境、使用者喜好，這些問題在尋求音樂的過程中常會

困擾使用者。

不論是聆賞動機或是尋求音樂的過程，皆會與音樂情境有關。然而要如何精確地描述音樂情境對於未受過音樂訓練的使用者而言有其難處。受過專業音樂訓練者，會用音樂專有名詞來形容音樂，例如：這首曲子是大調的曲式聽起來明快；相對的未受過專業音樂訓練者常用抽象的描述來表達自身的感受，例如：這首歌聽起來輕快、明亮，很適合下午的時候聽。然而這樣的描述方式涉及了聽者自身過去的經驗與文字表達方式。不論何種表達方式對於溝通上不甚便利，也可能會引起誤解。未受過音樂訓練者不了解音樂專有名詞所表達的意涵，抽象的意念在概念的解讀更是困難。

在尋求音樂與聆賞音樂之外，使用者也有個人數位音樂組織的行為。部分使用者採用建立資料夾的方式，自行命名資料夾名稱管理音樂檔案。另外有部分使用者利用音樂管理軟體支援的功能，自行編輯音樂播放清單。在分類上也是依據個人背景建立分類架構，所使用的文字也是相當個人化，例如「好多團」這個分類指的是以合輯方式出版的唱片。

在整理大量的音樂檔案時，必須花費大量時間自行建立系統架構，用語言\地區、音樂類型亦或是情境式的詮釋資料：時間、場合，不論是哪一種分類方法都必須是使用者自行人工分類，這項工作非常辛苦且耗時，因此提出本篇研究針對此問題探討，並提出解決辦法。依照情境式的分類系統，開發出音樂內涵式檢索系統，方便使用者將音樂檔案整理成系統，同時更多元的資訊也方便使用者跟朋友分享音樂或溝通音樂想法。

各種應用整合性的產品推陳出新，消費性電子產品層出不窮，改變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心裡的情緒感受，取而代之的是隨時隨地都可以欣賞到自己所喜歡的音樂，感受音樂所帶給人們的感動，再加上網路的發展，成為另一項熱門的傳播音樂平台。網路的便捷性與個人化的特色，加快了音樂傳遞的速度，網路在音樂上的應用正在蓬勃發展當中，使的音樂作品得以透過更快速與便捷的管道被作曲家發表創作亦以及被聽眾

取得。此外音樂資料庫也成為一個新的研究方向，過去的搜尋方式大多採用關鍵字搜尋的方式，例如：曲名、唱片名稱、演唱者姓名、作曲家姓名、音樂類型、唱片廠牌等等。前述之搜尋方法和聆賞音樂時所感受的情緒並不相符，現在期望用更加符合人性直覺的作法搜尋個人所喜好的音樂作品。

音樂聆賞者在不同的情境下，選曲的偏好都不相同。這涉及兩個層面，其一是不同的情境，其二是個人的音樂喜好。在不同的情緒下，音樂聆賞者對於音樂曲目的選擇與編排也會有不同的決定，例如：開車旅遊時、讀書研究時、假日休閒時以及睡眠前的時機；在睡眠前大多數會選擇平靜舒適的音樂，過於刺激的音樂則不合適；然而開車旅遊時，舒適平靜的音樂卻可能使人精神萎靡，導致駕駛車輛的危險。

另外個人的音樂喜好這個層面受到下列因素的影響：文化環境、社會輿論 [2]；同儕喜好、音樂訓練 [3]；年齡 [4]和其他個人獨特的性格 [5]；情緒感受 [6]。

最後現今的音樂播放程式雖然大多有自動選曲功能，然而其選曲的機制沒有針對音樂內容而設計，自動選曲的機制大多採用隨機的方式。以音樂播放軟體 foobar2000¹ 為例，軟體中提供之自動曲目編排順序以專輯順序隨機、以曲目順序隨機及以檔案儲存位置隨機播放等三種方法編排曲目順序；在選曲方面，必須由只使用者自行輸入，不具備自動選曲的功能。

1.2、研究方法概論

本篇論文研究是內涵式音樂資訊檢索，並選用音訊檔案作為實驗素材。首先收集音樂檔案，建立測試音樂資料庫，詳細討論參考 4.1 節。接著對音訊檔案作資料前處理，詳細的討論參考 3.2 節。資料前處理後對音訊進行特徵萃取，針對音樂的內容建立資料模型，特徵萃取演算法詳細的討論參考 3.3 節。之後根據特徵萃取建立的模型比對歌曲之間的相似程度，依照歌曲相似的程度高低編排曲目順序，相似度演算法詳細的討論參考 3.4 節。對於不同的特徵萃取與相似度演算

¹ Foobar2000, <http://www.foobar2000.org/>

法進行主觀量測，檢查測試的演算法是否符合情緒感受並整理所有測試演算法的效能，主觀量測方法詳細的討論參考 4.2 節，數據整理參考 4.3 節、4.4 節、4.5 節。

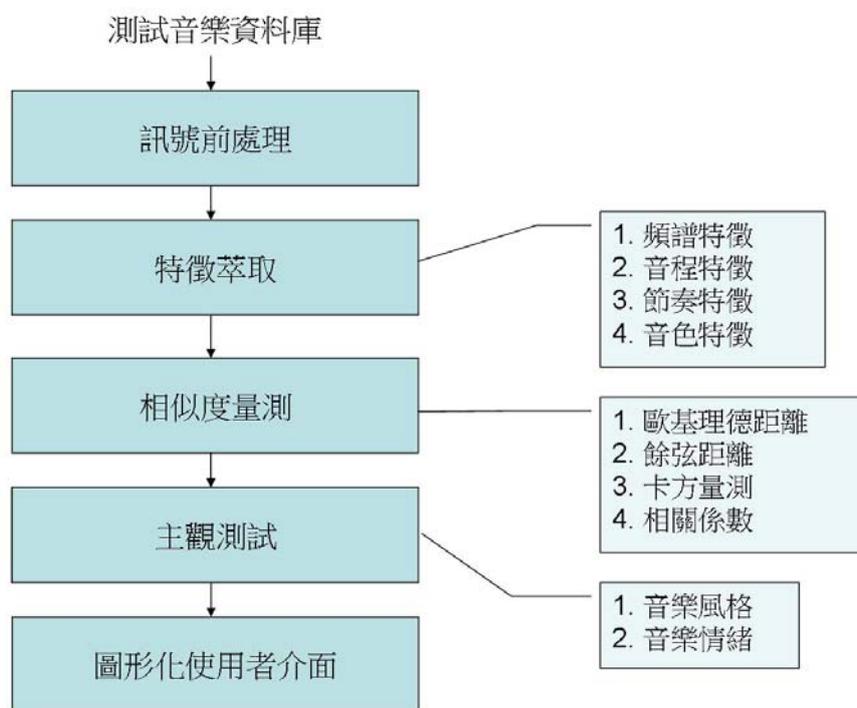


圖 1：研究架構流程圖

1.3、論文貢獻

本篇論文主要是應用內涵式音樂資訊檢索技術，開發自動化音樂曲目選擇與編排系統。目的是為解決在使用者大量數位音樂資料庫中挑選歌曲需要花費大量時間的問題。在傳統的音樂資料檢索系統中大多以文字的格式作為檢索資料，這樣的作法叫不符合聆聽音樂時的直覺感受，因此本文應用內涵是音樂檢索的技術開發音樂資料檢索系統。

此外本篇論文也將探討符合聆聽者之情緒感受的選曲機制，情緒感受屬於人類主觀上的判斷難以有統一的標準，為使研究結果達到人性化的需求，本次研究設計一種以情緒感受為基礎的演算法效能評斷方法。現今音樂播放軟體的自動選曲功能大多採用隨機模式，未能符合聆聽時的情緒變化，因此改良自動選曲方法

是本文主要探討的問題。

1.4、章節概述

本篇論文的第二章將介紹相關背景知識與相關文獻回顧；第三章說明本篇論文的研究方法與測試演算法；第四章討論系統實作、實驗方法和比較實驗結果；第五章總結實驗結果、分析所遭遇到之問題與探討未來改進的方向。



第二章、理論基礎與文獻探討

本章將討論本篇研究所涉及之理論基礎與相關文獻討論。第一節討論背景知識，主要的內容為多媒體數位資料庫、數碼音樂檔案格式；第二節討論內涵式音樂資訊檢索的相關文獻；第三節針對自動化歌曲選擇系統之文獻探討；第四節討論音訊前處理步驟之相關文獻；第五節針對特徵萃取之相關文獻作討論；第六節討論相似度演算法之相關文獻；第七節討論本篇研究涉及之音樂心理學範疇。

2.1、背景知識

音樂資訊檢索是多媒體資訊檢索中的一支。多媒體資料包括圖片、聲音、影片等。多媒體資訊檢索則是針對大量的多媒體資料有效的找出所需要的資料的技術。而音樂資訊檢索則是針對音樂資料開發智慧型的檢索技術。

數位資料庫以使用者輸入的資料型是分為兩種：以文字為基礎的檢索(text based retrieval)和以內容為基礎的檢索(content based retrieval)[7]。在以文字為基礎的檢索形式中，使用者輸入描述內容的文字作為資料庫中檢索比對資料的依據，常用於查詢音樂內容的文字如：作曲者、曲名、專輯名稱等。在圖書館學中用來描述資料的文字又被稱為metadata，中文譯為後設資料、詮釋資料或元資料。

以內容為基礎的檢索形式中，則沒有特定的輸入方式，通常根據檢索的資料內容設計人機互動界面。以midomi²這個線上音樂搜尋引擎網站為例，依據使用者吟唱或哼唱歌曲中的內容檢索音樂檔案，使用者對著麥克風哼唱歌曲即可查詢歌曲內容。

以內容為基礎的檢索又被稱為內涵式音樂資訊檢索，依據檢索音樂檔案格式的差異，開發搜尋引擎的技術也不相同。另外在檢索音樂的策略上，以內容作為檢索的依據將更加人性化，提供使用者更多的選擇。第二節將詳細討論內涵式音樂資訊檢索系統。

² midomi search for music using your voice by singing or humming, <http://www.midomi.com/>

最後數位音樂檔案格式依照儲存資料的內容可分為兩大類，符號資料與音頻資料。第一種符號資料是記錄音樂符號資料的如音高、調性、節拍、速度等，此類型的檔案格式為：MIDI、XML。另一種音頻資料則是記錄聲波大小的檔案格式，儲存音樂在經過錄音、混音後的聲波資料，這種檔案格式可分為未經壓縮的編碼，例如：wave；以及經資料壓縮的編碼格式，例如：wma、mp3[7]。

2.2、內涵式音樂資訊檢索

內涵式音樂資訊檢索依據檢所資料類型的不同可分為兩大類 [9]：(1) 由符號資料搜尋 (Search by symbolic data)；(2) 由音頻資料搜尋 (Search by audio data)。下面兩小節將分別討論。

2.2.1、由符號資料檢索

符號資料 (symbolic data) 指的是儲存音樂符號的檔案格式，例如：MIDI、XML。在特徵萃取時可直接取得音高、節拍、速度、音色的訊息。再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。此類型現有的資料庫有 C-Brahms、GUIDO/MIR、Orpheus、Probabilistic “Name That Song”、PROMS 等 [9]。

陳若涵 [10]發表之論文探討基於音樂內容的情緒分析與辨識。音訊檔案採用符號資料格式的MIDI檔。用不同的分類器演算法包括：KNNR、GMM、SVM、Fuzzy KNNR，實驗結果顯示KNNR得到最佳的辨識率。特徵選擇方面，在古典樂曲的測試中計算音樂的調性、音高平均和力度平均；在流行音樂中選則以鼓組出現的時間和鼓組密度作為特徵向量。

Yeh 等人[13]則提出個人化之音樂情緒預測系統。採用決策樹演算法依據使用者的背景資料建立音樂預測的模型。音樂特徵則選用平均音高和音調。使用者的背景資料包括：年齡、性別、教育程度、居住地等。音樂特徵則使用平均音高、音高標準差、音程平均、音程標準差、速度、響度、音色、調性等。

Hu等人 [12]提出之論文則使用動態時間校正(dynamic time warping, DTW)演

算法比對相同歌曲之音頻檔案格式和音樂符號檔案格式。特徵萃取的部分則比較四種不同演算法，分別是Chroma、Pitch Histogram、MFCC、Normalized MFCC (NMFCC)。Chroma演算法將頻譜中的頻率直接轉換成最接近的音高單位，並計算每一個音高單位內的能量大小。Pitch Histogram演算法計算出每一個音框中能量最大的頻率作為此音框之代表音高。分別計算各別音框之音高後統計直方圖。MFCC則是將頻譜中頻率的刻度表示成符合人體聽覺的刻度。NMFCC則是正規化後的MFCC向量。實驗結果顯示Chroma演算法萃取出之特徵具有最高的正確率，其次是Pitch Histogram演算法，MFCC演算法則最低。

2.2.2、由音頻資料檢索

由音頻資料搜尋 (search by audio data) 的資料庫則是用紀錄音樂在錄音及混音後之聲波波形的檔案格式，也是一般使用者常用來聆聽的檔案格式，例如：wav、wma、mp3。在這種檔案格式中，無法直接得到音高、節拍、速度、音色等音樂特徵，必須經由特定演算法從波形大小中計算出音樂的訊息；亦或是計算聲學特徵。特徵萃取的演算法將在2.3節中討論。此類型現有的資料庫有 audentify!、CubyHum、Cuidado、Cornell's“QBH”、Shazam、SOMeJB、SoundCompass、Super MBox[9]。

基於不同的音樂內容則會使用不同的特徵作為檢索的依據，Tzanetakis和Cook在2001年提出的研究則是用音頻資料開發出音樂類別的分類系統，所使用之特徵為音色、節奏、音高，即便採用的搜尋資料為音頻資料，仍然自行開發演算法，計算出音樂特徵[13]。Jiang等人[14]也提出音樂類別的分類系統，則直接使用頻譜上的聲學特性做為特徵。這兩篇論文所要找出的音樂訊息相同，然而提出之特徵不全然相同。

在音頻資料的音樂情緒辨識系統方面，大多採用心裡學家研究提出之情緒模型做為情緒分類的依據，例如Tayer提出之情緒模型[46]將人類的情緒反應以二維平面表示。在分類器上則有GMM、回歸統計和系統判別(system identification)的

作法。Lu等人[15]發表從音樂訊號中自動偵測情緒與追蹤之論文，分類器用GMM演算法採用階層式分類架構偵測音樂的情緒類別。在分類器使用的特徵則選用強度特徵、音色特徵、節奏特徵。各項特徵均從波形檔案中萃取得出。另外，情緒追蹤則是將音樂分成不同的片段，分節按照時間順序偵測出其所表示的情緒類別。Korhonen等人[16]則用System identification演算法建立音樂情緒內容的資料模型。

相較於類別式的情緒分類，另一種情緒分類系統則是以連續的型情緒模型做為分類的依據，此這情緒模型將情緒以感受程度的深淺表示，是具有連貫性的。Yang等人[17]則提出採用連續的情緒模情之概念開發出情緒識別系統。情緒不再只屬於單一種情緒類別，而是有感受程度上的差異。論文中採用回歸統計的作法[18]，並將統計完成之結果表示在一個二維平面上。圖像式的表示方法更加符合人類心理的感受程度。選用特徵方面則用PsySound、Marsyas、spectral contrast、DWCH等多個演算法萃取得。Wu和Jeng[19]提出之研究則是用SVM訓練萃取後之音訊資料。一首歌曲有可能屬於不同的情緒類別中，因此將歌曲的情緒類別則用機率方式表示。表示一首歌曲在各個情緒類別中使人感受到的程度高低。

2.3、自動化歌曲選擇系統

自動化歌曲選擇系統指的是依據使用者聆聽音樂的需求，自動篩選音樂內容並編排曲目順序，是內涵式音樂資料檢索技術之應用。下面章節將介紹相關的技術，依照分析內容的資料格式分為由文字資料檢索(text based retrieval)以及由音樂內容檢索(content based retrieval)。

2.3.1、由文字資料檢索

Field等人發表之論文 [20]提出一個音樂曲目選擇與編輯的系統，以metadata作為分析的資料。論文中提出建立個人化的歌曲選擇系統必須蒐集音樂資料及使用者的資料，並且討論如何取得合適的資料。論文中歌曲的資料以情緒表示，情緒被分成四個類別分別為：狂暴的(angry)、冷淡的(chilled)、歡快的(upbeat)、不

屬於前述類別的(non)，每一個情緒類別又分成四個等級以 0 到 3 表示。

Pauws和Eggen發表的論文 [21]同樣的也是處理對於音樂標記的詮釋資料開發個人化之音樂自動選曲系統：PATS (Personalized Automatic Track Selection)。分析之音樂類型為爵士樂。分析的詮釋資料有：曲目名稱、主要演出者、專輯、廠牌、年代、風格、速度、樂器編制、獨奏者、作曲家、製作人、錄音地點、是否為現場表演、節拍等。

2.3.2、由音樂內容檢索

Li和Ogihara發表的論文 [22]研究以聲學特徵為基礎的音樂資訊檢索，音樂資訊檢索的問題主要討論兩個問題：第一問題討論是如何由一個用來當作檢索的音檔搜尋與它聲音相似的音樂檔案；第二個討論的問題則是如何偵測音樂中的情緒。特徵萃取演算法結合小波轉換和由MARSYAS萃取出之音色特徵。相似度量測步驟中則採用歐基理德距離對正規化後之特徵計算歌曲的相似程度。測試的音樂類型為爵士樂與古典樂。

Foote 發表之論文 [23]用聲學特徵做為衡量音樂內容相似度的依據，論文中以MFCC演算法作為聲學特徵，特徵萃取步驟後以自行開發的Q-Tree演算法將特徵向量作數值上的量化，第三步驟將資料庫中量化後的特徵比對相似度，文章中所測試的相似度驗算法為：歐基理德距離和餘弦距離，最後以相似度大小排序，列出與檢索歌曲相似的歌曲。

Lagan和Salomon提出之論文 [24]主要研究問題有二點；首先如何由在距離平面上的軌跡圖形自動產生音樂播放清單；第二研究如何由使用者的回饋的資訊自動產生播放清單。研究結果發現，當分析時加入歌曲的詮釋資料可以提高正確率，因此建議未來的研究當中可以加入合適且可以取得的詮釋資料提高分析的正確率，文章中提及的詮釋資料為音樂類型。特徵萃取用音樂訊號的頻譜表示；資料分群與相似度演算法為：K-means和Earth Mover's Distance[25]。

Meyers發表論文 [26] 此篇論文的動機是缺乏一個符合人性直覺及了解內容

脈絡的自動產生音樂播放清單的工具。由於缺乏具有這樣的工具，因此在音樂數位資料庫中查詢資料會遇到困難。建立一個符合人性直覺及了解內容脈絡的自動產生音樂播放清單的工具的需求是顯而易見的。特徵萃取中分析音樂訊號中之音樂結構的訊息：調性、和聲、音量、節拍和速度。另外也分析歌詞的內容，針對文字的資訊分析。此篇論文分析的訊息包含音樂的內容以及文字的內容。

黃捷發表之論文 [27]同樣的也是從音樂錄音的訊號中分析音樂訊息，計算出內容相似的音樂並排序提供使用者選擇，與其它的文獻的作法不同之處在於此篇研究在特徵萃取之後，利用資料編碼的技術，將在時序下的特徵編碼，在進行關連樣式分析及資料結構索引。

2.4、音訊前處理

在資料量過於龐大時，為節省運算時間增加效率。在特徵萃取前，會先將完整樂曲分段，只取樂曲部分片段萃取其中的特徵。分割音訊時有兩種作法，第一種作法是依照顧定時間長度分割音訊，Xiao等人 [28]發表之研究則是使用此法，分割之時間長度分別為四秒、八秒、十六秒、三十二秒。實驗結果顯示在音樂情緒分類系統中，八秒的長度得到最高的辨識率 83.82%，其次為十六秒 83.08%，三十二秒的分割長度則得到最低的辨識率 68.98%。另一種分割音訊的作法則基於音樂內容做音訊分割 [29] [30]，首先找出原始音訊中重複的片段，接著依據重複的內容片段分割音訊長度。

2.5、特徵萃取

在由音頻資料檢索之內涵式資訊檢索系統中，特徵萃取的目的是量化音樂訊號，作為資料比對的特徵。在這個步驟當中關鍵是找出具有代表性的特徵，以增加資料檢索的正確性。下面將介紹音頻資料的特徵萃取方法：聲學特徵(low level feature)及結合音樂特徵(mid level feature)。

2.5.1、聲學特徵

音樂是一項抽象的藝術，同時也被視為聲音的符號，然而聲音在空氣中傳播

又與各式物理現象有關，在運算上必須將其以數字表示，因此在音樂特徵表示方法上，可分為三種不同階層的表示法，分別為低階特徵(low-level feature)、中階特徵(mid-level feature)、高階特徵(higher-level feature)。

低階特徵用來表示訊號的特徵，無法紀錄音樂中包含的語意訊息，可以從訊號中計算得出之特徵，例如：過零率(zero crossing rate, ZCR)、倒頻譜(Cepstrum)、梅爾倒頻譜係數(Mel-scale Frequency Cepstral Coefficients, MFCC)等等；高階特徵則是將音樂以符號表示，用以傳達語意或情感訊息，這些符號可從 MIDI 檔案中萃取而得，例如：音高表示成 MIDI 音高數字(MIDI note number)、速度以每分鐘中所包含的節拍數目(beats per minute, bpm)表示。

中階特徵則是彌補兩者的缺點，用低階特徵可從訊號中運算得出的訊息將其表示為高階特徵的音樂符號。如此一來即可透過運算從音樂訊號中找出表達音樂中所要傳達的語意訊息。

心理聲學當中提到，人類的大腦只會知覺到聲音有變化的部分，像是振幅的變化、頻率的變化 [31]。另外，也有研究提出音樂中的旋律、節奏、音色皆會影響人們在聽音樂時的情緒反應 [50]，例如：速度快的音樂會使人感到快樂；旋律線向上的音樂會使人朝向正面情緒、反之亦然；旋律中音高最高的音符與音高最低的音符之間的音程越大，則聽者的情緒反應較為激烈；反之音程越小聽者的情緒反應則比較平靜。因此在音樂情緒辨識上，採用人耳與大腦知覺到聲音變化的特性做為音樂情緒辨識系統中的特徵向量。

在音樂理論中所稱的音高(pitch)即是人類心裡對基頻的感受，中央 C (C4)上的 A 音符(A4)一般定為 440Hz。在一般西洋音樂常用的十二平均律中，將八度音平均分成十二的半音。擁有八度音程的兩個不同音高，其基頻為兩倍關係。兩個音程關係半音的音高在頻率上的關係可表示為。不同樂器演奏相同音高時，在頻譜上存在相同的基頻，不同是在高頻帶能量的分布，即不同的泛音組成。

旋律、節奏、和聲是構成音樂的三大要素。在開發音樂資訊檢索系統時，如何從訊號中萃取出音樂感知特徵是一大挑戰。中階特徵即是應用數位訊號處理技

術，從音樂信號中萃取音樂特徵之演算法，模擬人腦在聽到音樂時對於音高、節奏、音色的感知。從前面的討論中得知，音高與頻率有關、節奏與時間有關、音色與頻譜的分布情形有關。因此要從音訊中得到音高或是音色的訊息時，必須先做傅立葉轉換將時域的信號轉換到頻域的信號。

在分析音樂訊號時有兩種分析方法，都是基於離散傅立葉轉換改善頻率的取樣來達到符合音樂信號特性的演算法。Serra[32]發表的研究即是採用聲音合成的概念開發演算法。做聲音合成時將一個聲音的聲波視為由多個不同頻率的正弦波相加組成。因此在分析音樂信號時，也可採用相同的概念，將音樂訊號分成多個正弦波相加，以找出音頻中的音樂特徵。

由Brown和Puckette[33]所提出之常數品質因數轉換(Constant Q Transform, CQT)即是為了分析音樂訊號所開發之技術。前段所提及兩個音程關係半音的音高在頻率上的關係可表示為 $f_1/f_2 = \sqrt[12]{2} \approx 1.059$ ，因此最小的頻率變化量為 $2^{1/12} - 1 \approx 0.059$ 。又基於取樣原理(Nyquist Law)，取樣頻率需大於或等於兩倍的原始訊號頻率，(頻率上的取樣週期)頻率差為 $\delta f = 2^{1/24} - 1 \approx 0.029$ ，即頻率差、波寬或解析度 δf 在各頻率下均為常數，也表示頻率與解析度的比值為常數 $f/\delta f = Q$ ，因此 Q 定義為

$$Q = f/0.029f \approx 34 \quad (1)$$

相當於 1/24-oct 的濾波器組(filter bank)。

在做音樂訊號分析時，頻率解析度必須足夠才能表示出音樂頻率的變化，由前段之敘述，將全頻 f 中的第 k 個頻率取樣表示為

$$f_k = (2^{1/24})^k f_{\min} \quad (2)$$

這裡 f_{\min} 是訊息分析中所需的最低頻率。在 DFT 中，解析度或波寬 δf 定義為取樣頻率(sampling rate, s)

2.5.2、音樂特徵

聲學特徵大多著重於訊號處理技術的討論，然而在分析音樂訊號時，同時考慮音樂理論可以加強特徵的代表性，因此同時考慮聲學特徵與音樂理論的特徵被稱為音樂特徵(mid level feature)。由於音樂特徵種類繁多，且基於不同的內容有適用於不同的演算法，下面將分項介紹所使用之相關技術。

A、音高追蹤

音高追蹤(pitch tracking)即是找出音樂中每一個時刻下的音高。音高有多種不同的表示方式或記譜法，為符合運算需要多將音高以音樂數字介面 (Musical Instrument Digital Interface, MIDI) 音高數字 (MIDI note number) 格式表示。MIDI 是一種工業用通訊協定，用於各式電子樂器、電子設備之間信號的傳輸。頻率與 MIDI 數字音高之關係式如下， p 代表音高、 f 代表頻率。

$$p = 69 + 12 \times \log_2 \left(\frac{f}{440} \right) \quad (3)$$

Wu 等人發表的論文 [34] 則是從吵雜的環境中同時追蹤單個音高或兩個音高，與追蹤單個音高不同的是，音訊在經過訊號轉換後使用隱藏馬科夫模型 (hidden Markov model, HMM) 建立資料模型。Ryynanen 和 Klapuri 提出從多音音樂信號中複製出演者旋律的論文中 [35]，同樣的也用 HMM 建立資料模型，在建立模型時同時考慮聲學模型和音樂理論模型，最後輸出音高與休止符的資料序列。

B、旋律萃取

廣義的旋律可定義為一連串連續的音高，音高高低的變化組合成一條旋律線 (melody contour)。旋律是表達音樂作品情感、想法的重要元素之一。由 Emilia Gómez 所發表的論文中整理出常見的估算基頻之演算法 [36]，在時域下的演算法有過零率 (Zero Crossing Rate, ZCR)、自相關係數 (Autocorrelation, ACF)、包絡週期 (Envelope periodicity) 等；頻域下則有倒頻譜 (Cepstrum)、頻譜自相關係數

(Spectrum AC)、小波轉換(wavelet transform, WT)等，然而這些方法僅適用於單聲部的音樂訊號，在多聲部的音樂訊號中無法有效得找出主要旋律之基頻。

從被混音之多音軌音樂訊號中萃取出旋律輪廓(melody contour)，稱之為旋律萃取(melody extraction)或音高追蹤(pitch tracking)。旋律萃取大致可分為兩個步驟，(1)估算基頻；(2)旋律線萃取。另外也有監督式學習與非監督式學習兩種機器學習模式 [37] [38] [39] [40]。Loeffler[41]發表之論文則是討論從多音樂中追蹤音高的方法，則是用高斯混合模型(Gaussian mixture model, GMM)建立資料模型，並且從多音部的音樂中萃取出主要旋律的旋律線條。

C、和弦偵測

和絃在音樂中也是佔有非常重要的地位，常用的和絃有三和絃、五和絃、七和絃。和絃的定義為在同一時刻下同時演奏多個音高的組合，以三和絃為例則是在同一時刻下演奏三個和諧的音高。和聲學是在專門研究和絃的學門，主要在研究如何在同一時間演奏不同音高而獲得和諧或是不和諧的效果。Bello和Pickens[42]提出了利用隱藏馬可夫模型HMM模型計算出音訊中的和絃組合，以監督式的機器學習方法計算隱藏馬可夫模型的參數數值。在特徵萃取方面則採用CQT轉換和chrogram表示法作為特徵。

2.6、相似度量測

在內涵式的資料檢索系統當中，無法精準找到完全相同的資料內容，演算法常用計算相似度之演算法找出相似的資料。下面是文獻中資料檢索系統中常用的相似度演算法。

Harb 和 Chen 提出之論文 [43]討論以範例查詢的資訊檢索系統 (Query by Example Music Retrieval, QEMR) 用來檢所音樂資訊所遭遇的問題與解決辦法。相似的音樂在這篇論文中被定義為在使用者主觀判斷下列出預期期望要聆聽的歌曲清單。聲學特徵為聲音的頻譜。計算相似度時以音訊長度分別計算局部的相似度(local similarity)和總體的相似度(global similarity)。首先計算局部相似度，測

試音檔的長度大約一到兩秒，相似度演算法為KullBack-Leibler (KL) distance。之後計算總體相似度，第一步計算局部相似度矩陣(Local Similarity Matrix, LSM)。第二步計算最小距離(Min Distance, MD)，取LSM中最小三個數值的平均。第三步計算LSM中高頻(1 and 4 KHz)的最小距離(Min Distance for High Frequencies ,MDHF)。第四步驟計算低頻(1 KHz 以下)的最小距離(Min Distance for Low Frequencies ,MDLF)。第五步驟計算LSM中所有數值的平均(Sum Distance ,SD)。最後計算MD, MDHF, MDLF和 SD的平均即為總體相似度。

Qian等人發表之論文 [44]介紹歐基理德距離與餘弦角度距離在影像之內涵式檢索上的應用。在內涵式影像檢索當中，同樣的也是先計算影像中的特徵之後在計算影像間的相似度找出最接近的圖片內容。另外常用的三種相似度演算法：Correlation、Chi-Square、Intersection，各演算法分別的數學定義如下所示。

(1) Correlation:

$$d_{correlation}(H_1, H_2) = \frac{\sum_i H_1'(i) * H_2'(i)}{\sqrt{\sum_i H_1'^2(i) * \sum_i H_2'^2(i)}} \quad (4)$$

where $H'_k(i) = H_k(i) - (1/N) * (\sum_j H_k(j))$, N equals the number of bins in the histogram.

(2) Chi-square:

$$d_{chi-square}(H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \quad (5)$$

(3) Intersection:

$$d_{intersection}(H_1, H_2) = \sum_i \min(H_1(i), H_2(i)) \quad (6)$$

2.7、音樂心理學

音樂心裡學是結合音樂學與心裡學的學門，討論人們聽到音樂後的心理反應。在這篇研究當中，聽到音樂後的情緒反應被當作是判斷檢索資料正確性的依

據，下面將介紹文獻中提出之情緒的心理模型，以及音樂與情緒的關連。

2.7.1、情緒模型

林俊男發表的論文 [45]說明人類聽到聲音信號後，產生感知與意象的過程與探討，並且統整出常用於表達聲音意象的形容詞。論文中指出「感覺」是指人類經由感覺器官例如：視覺、聽覺、嗅覺、味覺、膚覺與平衡覺。感覺器官接受外界的刺激，經由神經傳達到大腦的過程；「感知」在張氏心理學辭典中的定義是由感官以覺知環境中物體存在、特徵及其彼此間關係的歷程。感覺與感知兩者極為相似但仍舊有所差異。生理受到刺激的過程稱為「感覺」，感覺和過去的經驗連結而成組織則稱為「感知」，屬於心理層面。「意象」屬於人類的心理特徵，一項系統的建立涉及到人們感知系統的建立與日常生活經驗的記憶。

在量測情緒的研究當中，大致可分為兩類模型：連續式、類別式。連續式的情緒模型將情緒反應視為連續的，此類的研究以Tayer提出之情緒模型 [46]為例，Tayer情緒平面將人類的情緒反應以二維平面表示，橫軸表示人感受到壓力的高低(valence)，縱軸表示人感受到之能量高低(arousal)。類別式的情緒分類則以Henver 提出的形容詞圈 [47]做為例子。

1936年Hevner 研究人們常用來表達情緒的形容詞，並整理歸類程八大類，如圖二所示。1999年Schubert研究Tayer提出的情緒平面和Henver提出的形容詞圈之間的關連，並將兩個模型整合 [50]。2003年Schubert則是Hevner提出的形容詞圈為基礎，修正形容詞圈中現今較少被使用的形容詞 [48]。本文的研究依照研究的目的，了解人們聽到音樂後的情緒反應，因此必須了解常用於描述音樂的形容詞，以Henver形容詞圈做為研究的基礎。另外，形容詞的翻譯則參考論文 [49]提出之對於Henver形容詞圈的中文翻譯，並做些許修改。本文研究所使用之音樂形容詞中英文對照表可參考附錄一。

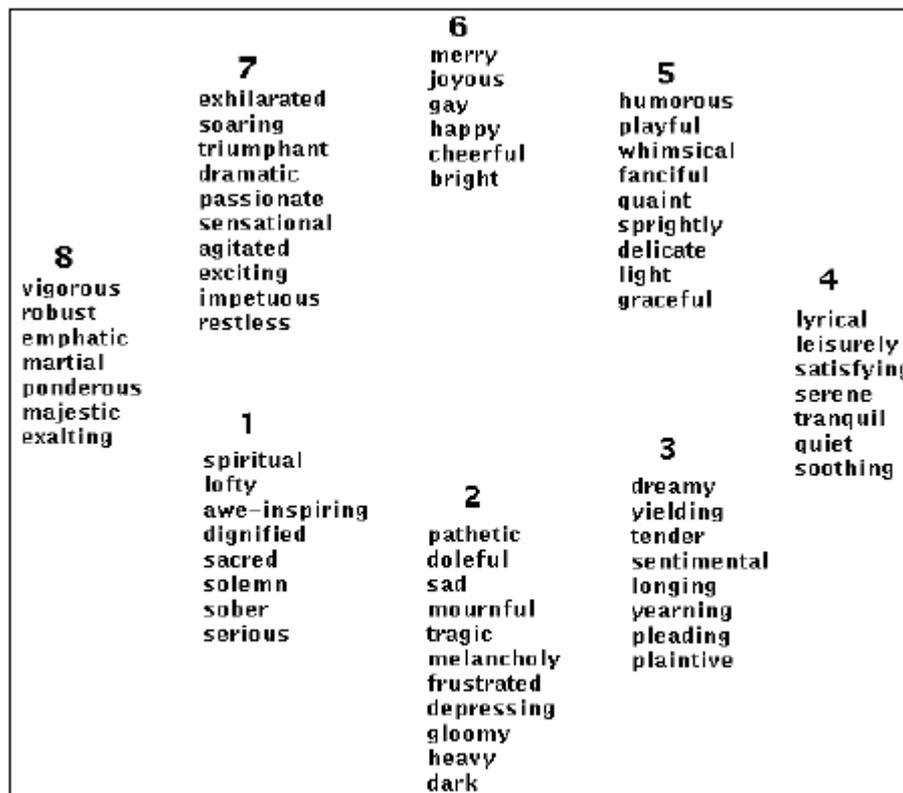


圖 2：Hevner 形容詞圈 [47], p249

2.7.2、音樂與情緒感受

由前段的敘述可知情緒屬於心理層面，屬於較為抽象的概念卻難以將其量化。Emery Schubert提出可量測人類聆聽音樂後的情緒反應，並將量測的結果表示在一個二維的平面上 [50]。這個二維平面橫軸表示正向情緒(valence)的大小，縱軸表示激動(arousal)的程度大小。此篇論文討論第二個議題為各項音樂特徵與情緒反應的關係，例如：節奏輕快的音樂給人較為愉悅的感受，作者完整整理出各種不同的音樂特徵與情緒之間的關聯性並表示在Thayer的情緒平面上。音樂特徵則包括平均音高、音高範圍、旋律方向、調性、主調、音色、和聲、速度、節奏、節拍、顫音、動態等等。以論文中提到之平均音高的例子說明，整首曲子的音高以平均數表示之經由實驗找出最後統計的結果。平均音高較低的曲子給人們的情緒反應呈平靜的；而平均音高較高的曲子則給人的感受是較為激動的情緒。

第三章、研究方法

本章討論本篇論文的研究方法，第一節說明研究假設；第二節提出系統設計與架構流程；第三節討論資料前處理的方法與步驟；第四節討論特徵萃取演算法的方法與步驟及本篇論文測試的演算法；第五節討論相似度量測演算法的步驟與本文所測試的演算法；第六節分析系統的人機互動與流程。

3.1、研究假設

樂曲種類繁多，為方便消費者快速找到自己所喜好的音樂，因此必須將樂曲依照風格分類。由於音樂分類法是一項即具專業的議題，不在本文探討的範圍之內。本文參考Allmusic網站³對於音樂類型之方法，音樂類型分類採用階層式的分類方法，第一層被區分為十一大類：流行及搖滾(Pop/Rock)、爵士(Jazz)、節奏藍調(R&B)、鄉村(Country)、藍調(Blues)、電子樂(Electronic)、拉丁(Latin)、雷鬼(Reggae)、世界(International)、古典(Classical)。

在這篇文章中，在第一層的分類稱為音樂類型(genre)，音樂類型之下的分類被稱為音樂風格(style)。以流行搖滾類型為例，流行及搖滾類型下又被分為：實驗搖滾(Experimental)、舞曲(Dance)、民謠及鄉村搖滾(Folk/Country-Rock)、硬式搖滾(Hard Rock)、流行及搖滾(Pop/Rock)、搖滾樂(Rock & Roll)、輕搖滾(Soft Rock)、車庫搖滾(Garage)、歐洲流行樂(Europop)、其它語言的搖滾樂(Foreign Language Rock)等類別。

在編輯歌曲的播放清單時，大多傾向於選擇類似的音樂接續播放，類似的音樂有三種可能：一是樂曲型態(genre)上的相似，二為樂曲表達的情緒(emotion)相似，最後一種可能為風格(style)上的相近。樂曲型態上的相似指的是樂曲在上層分類的相同；風格上的相近指的是樂曲在下層分類的相近。例如：老鷹合唱團(Eagles)所發表的曲目加州旅社(Hotel California)與披頭四樂團(The Beatles)演唱

³ All Music, <http://www.allmusic.com>

的昨日(Yesterday)這兩首曲子在類型上是一致的，風格上是相近的，因此在編輯音樂播放清單時，將這兩首曲子安排在前後的順序是可以被接受的；反之老鷹合唱團(The Eagles)所發表的曲目加州旅社(Hotel California)與奇異恩典(Amazing Grace)這兩首歌曲在風格上完全不同，前者是搖滾樂，後者為宗教音樂。因此在編輯播放清單時將這兩首歌曲安排在前後順序及是不恰當的編排。另外，女神卡卡(Lady Gaga)演唱的電話(Telephone)與加州旅社(Hotel California)，則是類型相同風格不同的例子。

3.2、系統架構

本次研究的目的為應用內涵式音樂資訊檢索開發自動音樂選曲系統。依照第二章中，文獻對於內涵式音樂資訊檢索步驟的討論本系統分成五大步驟：訊號前處理、特徵萃取、相似度量測以及主觀測試。另外在特徵萃取步驟中又細分為：訊號處理和數值統計；相似度量測步驟中也細分為：局部相似度和總體相似度。所有步驟中所使用之測試的演算法將在後面的章節討論。細部系統架構如圖 3 所示。

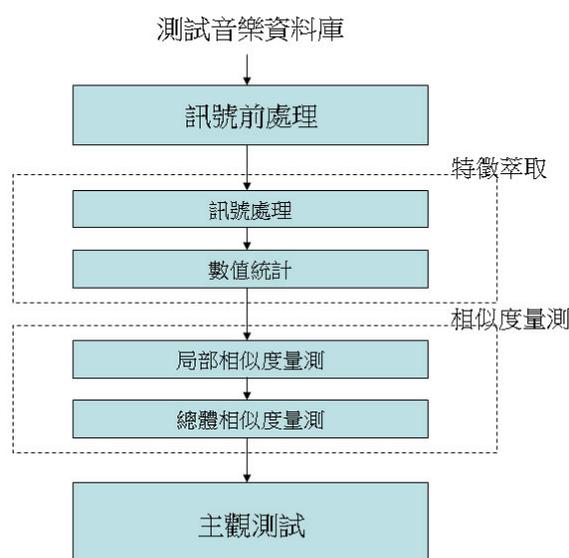


圖 3：細部系統架構流程圖

3.3、資料前處理

首先將雙聲道之音訊轉換成單聲道之音訊；接著降取樣至 22050 Hz；最後將音樂分割成若干音樂片段。分割音樂檔案時採用固定音樂片段數目的想法分割音樂檔案，也就是說每一首歌曲擁有相同數目的音樂片段，每一個音樂片段的時間長度不相同。由於音樂檔案長度不盡相同且此次測試使用歌曲完整的長度，因此選用這樣的想法便於後續資料上的處理。文獻 [28]中提到，樂曲中所表達的情緒會隨著時間改變，相同的情緒停留在短暫的時間裡，因此必須要研究在音樂情緒分析的研究中最合適的音樂片段長度為何。資料前處理演算法細部流程表示在圖 4 中。

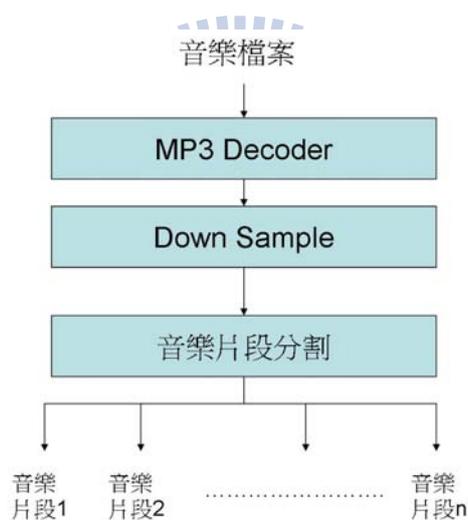


圖 4：資料前處理流程圖

3.4、特徵萃取演算法

本次的研究中測試四種特徵、七種演算法。四種特徵分別以不同的音樂特徵代表，分別測試資料檢所的結果。四種特徵萃取演算法中用到七種演算法，這七種演算法包括三種訊號處理演算法和四種數值統計演算法。特徵萃取演算法細部流程圖表示在圖 5 中，演算法詳細的數學定義參考後面章節的論。

本次研究測試四種特徵萃取演算法：頻譜特徵、音程特徵、節奏特徵、音色

特徵。頻譜特徵主要的目的是測試訊號處理演算法用於分析音樂訊號的效能，頻譜的資料與音樂中的音色、旋律、合聲有關，詳細的步驟與演算法在 3.4.1 節討論。音程特徵則是參考音樂學中音程的定義設計演算法，主要著重在音高變化的訊息，在這裡採用頻率的變化做為代表，詳細的步驟與演算法在 3.4.2 節討論。節奏特徵則是參考文獻 [13]和文獻 [15]提出的方法，以自相關函數(Auto Correlation Function, ACF)及波峰探測(Peak detection)演算法計算音樂中的節奏訊息，詳細的步驟與演算法在 3.4.3 節討論。音色特徵參考文獻 [13]提出之演算法，詳細的步驟與演算法在 3.4.4 節討論。

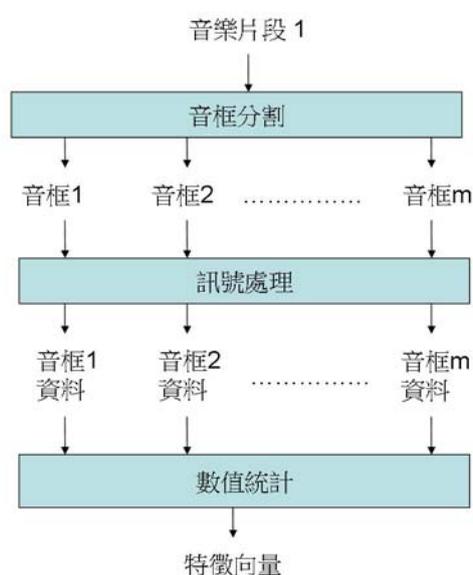


圖 5：特徵萃取流程圖

特徵萃取被分成三個步驟，各別特徵在訊號處理和數值統計步驟中用不同的演算法測試效能，實驗數據參考第四章中針對實驗結果的討論。下面將分別敘述各步驟。

一、依序分別讀入音樂片段之音樂訊號，接著對音樂訊號做音框分割，並對每一個音框的訊號乘上漢明窗函數。漢明窗函數的數學定義表示在式(7)中。

$$\omega(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right), 0 \leq n \leq N \quad (7)$$

音框長度的參數範圍推導如下：

(1) 假設：音樂的速度是 120 bpm (beat per minute)，意思是一分鐘內有 120 拍

則：1 拍 = 0.5 sec.

(2) 再假設：音樂中所使用之最小的拍號為 1/16 拍

則：1/16 拍 = 0.5 * (1/16) = 0.125 sec.

(3) 最後由取樣定理：取樣頻率 $\geq 2 \times$ 訊號中最大頻率

$$\frac{1}{\text{取樣頻率}} \leq \frac{1}{2 \times \text{最大頻率}}$$

$$\text{取樣週期} \leq \frac{1}{2} \times \text{最小週期}$$

所以：音框長度 $\leq 0.125 \div 2$

音框長度 ≤ 0.0625 sec.

二、將時域訊號轉換成頻域訊號。在這個步驟中將測試三種不同的演算法：快速傅立葉轉換(Fast Fourier Transform, FFT)、梅爾倒頻譜系數(Mel-frequency cepstral coefficients, MFCC)、常數因數品質因數轉換(Constant Quality Factor Transform, CQT)[33]。

三、統計所有音框中的資料即為這個音樂片段的特徵向量。在這個步驟中再測試四種不同的統計方法，分別是所有音框在所有頻率帶上能量的平均(AVG)、所有音框中所有頻帶上能量的差值的平均、自相關函數(Auto Correlation Function, ACF)及波峰探測(Peak detection)、頻譜上能量的通量(flux)。

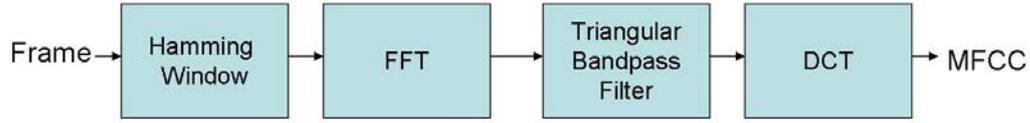
3.4.1、頻譜特徵

訊號處理中的頻譜資料與音樂學中的音色、旋律、和聲皆有相關，在這裡分別計算每一個音樂片段中所有音框的頻譜資料，之後分別計算所有音框中每一個頻率帶下能量的平均，做為這個音樂片段的頻譜特徵向量，頻譜特徵細部流程圖如圖 6 所示。下面列出頻譜特徵中所有測試的演算法之數學定義。

FFT：

$$X[k] = \sum_{j=1}^N x[j] \omega_N^{[j-1][k-1]} \quad (8)$$

MFCC :



Triangular band-pass filter:

$$mel[f] = 2595 \times \log_{10} \left[1 + \frac{f}{700} \right] \quad (9)$$

Discrete Cosine Transform (DCT):

$$C_m = \sum_{k=1}^N \cos \left[\frac{m(k-0.5)\pi}{N} \right] E_k \quad (10)$$

CQT :

$$\delta f = 2^{1/24} - 1 \square 0.029 \quad (11)$$

$$Q = \frac{f}{\delta f} = \frac{f}{0.029f} \square 34 \quad (12)$$

$$X[k] = \frac{1}{N[k]} \sum_{n=0}^{N[k]-1} W[k,n] x[n] \exp \left\{ \frac{-j2\pi Qn}{N[k]} \right\} \quad (13)$$

$$N[k] = \frac{f_s \cdot Q}{f_k}, f_s = \text{sampling rate} \quad (14)$$

$$W[k,n] = \alpha + (1-\alpha) \cos \left(\frac{2\pi n}{N[k]} \right), \text{where } \alpha = \frac{25}{46} \text{ and } 0 \leq n \leq N[k]-1 \quad (15)$$

每一個頻率帶下所有音框能量的平均：

$$E[k] = \frac{\sum_{n=1}^N e[k,n]}{N}, 20 \leq f[k] \leq 20000 \quad (16)$$

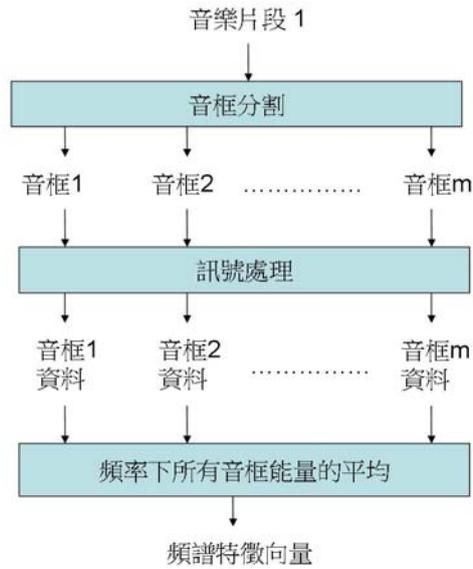


圖 6：頻譜特徵萃取流程圖

3.4.2、音程特徵

音程特徵則是參考音樂學中音程的定義，音程表示兩個音高之間的距離，主要著重在音高變化的訊息，在音程特徵中以頻率的變化做為代表。萃取音程特徵的步驟跟頻譜特徵的步驟只有在數值統計這個步驟中測試的演算法不相同，其餘的步驟皆相同。音程特徵的步驟與流程表示在圖 7 中。頻率下前後音框能量差的數學定義如式(17)。

$$EnergyDiff[n] = FramdData[n] - FrameData[n-1], 1 \leq n \leq N \quad (17)$$

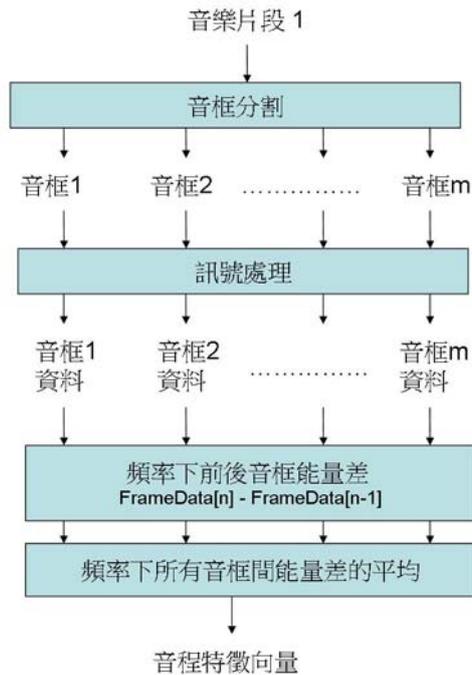


圖 7：音程特徵萃取流程圖

3.4.3、節奏特徵

節奏特徵萃取流程中，在計算完頻譜資料之後，以自相關函數與波峰偵測演算法計算音樂訊號中的節奏訊息。自相關係數用於頻率訊號上之物理意義為各頻率帶出現的頻率，可以類比到音樂理論中對於節奏的定義。節奏特徵細部流程圖如圖 8 所示。下面列出節奏特徵中測試的演算法之數學定義。

自相關係數(ACF)數學式

$$acf[\tau] = \sum_{i=0}^{n-1-\tau} s[i]s[i+\tau] \quad (18)$$

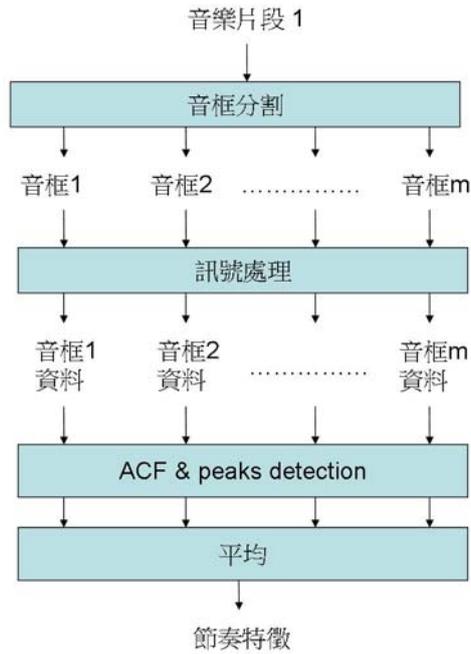


圖 8：節奏特徵萃取流程圖

3.4.4、音色特徵

音色特徵萃取步驟中，在計算完頻譜資料之後，以通量(flux)演算法計算音樂訊號中的音色訊息。通量用於頻率訊號上之物理意義為單位時間下，音框間頻譜能量的變化的總和，可以近似於音色的訊息。音色特徵細部流程圖如圖 9 所示。下面列出音色特徵中測試的演算法之數學定義。

通量(flux)數學定義如下：

$$flux[k] = \sum_{n=1}^N (FrameData[k, n] - FrameData[k, n-1])^2, \quad (19)$$

where $20 \leq f[k] \leq 20000$

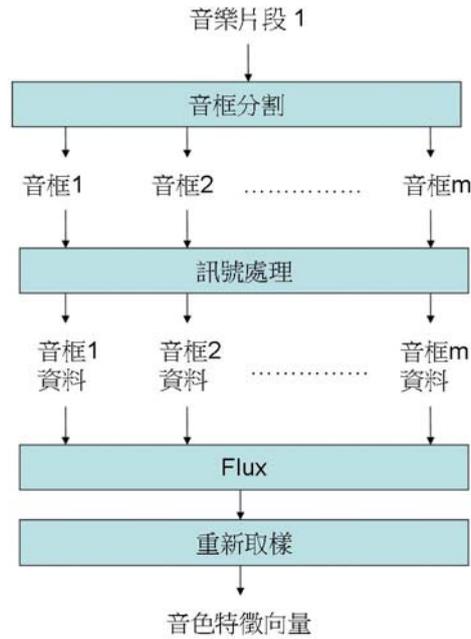


圖 9：音色特徵萃取流程圖

3.5、相似度量測演算法

特徵萃取完成之後，針對特徵向量量測兩首歌曲之間的相似度，依據資料庫中歌曲的相似度即可挑選出內容相似的歌曲並排列出曲目順序。根據文獻 [13] 中發表的實驗結果顯示，使用完整音樂檔案測試檢所結果可以得到最佳的正確率，因此本文也使用完整音樂檔案測試。在量測相似度時，先量測局部特徵(local feature)的相似度，再量測整體特徵(global feature)的相似度，相似度量測詳細的步驟與流程如圖 10 所示。

相似度量測在這篇論文的作法是：首先，分別量測每首曲子間每一個音樂片段的個別距離(distance)；再將每一個音樂片段計算之距離加總即為整首曲子之相似度。在這個步驟當中，測試四種量測相似度的演算法：歐基理德距離(Euclidean Distance)、餘弦距離(Cosine Distance)、卡方測定(Chi-square)、相關係數(Correlation)，各演算法的數學定義在後面的章節討論。量測資料庫中所有歌曲間的相似度後，最後再依據量測之相似度排序，及可得到實驗結果。論文 [19]

中也提到用餘弦距離量測量測歌曲在表達之情緒相似度上可以得當不錯的正確率。

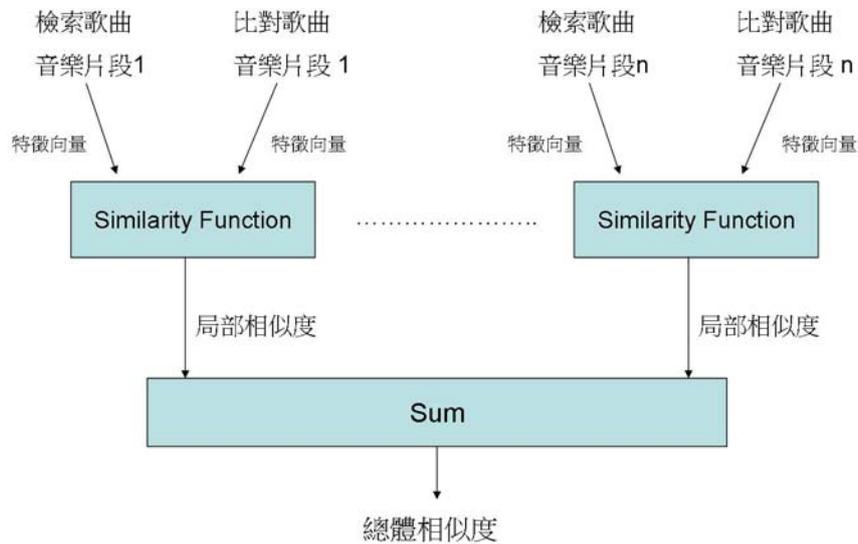


圖 10：相似度量測流程圖

下面將分別說明本次研究中用於量測局部相似度的測試演算法：

A、歐基理德距離

歐基理德距離(Euclidean distance)在幾何的定義為，在向量空間兩個向量間的絕對距離。計算後的數值如果為 0 則表示兩個向量完全相同，而數值越大則兩個向量間的相似程度越低。

$$d_{\text{Euclidean Distance}}(H_1, H_2) = \|H_1 - H_2\| = \sum_{i=1}^N \sqrt{(H_1[i] - H_2[i])^2} \quad (20)$$

B、餘弦距離

餘弦距離(cosine distance)在幾何上的定義是，單位長度下兩個向量的角度關係。計算後的數值如果為 0 則表示完全相同，而數值越大則表示相似程度越低。另外在文字檢索系統中常用餘弦距離演算法搜尋資料。

$$d_{\text{Cosine Distance}}(H_1, H_2) = \frac{H_1 \cdot H_2}{\|H_1\| \|H_2\|} = \sum_{i=1}^N \frac{H_1[i] \cdot H_2[i]}{\sqrt{(H_1[i] H_2[i])^2}} \quad (21)$$

C、卡方測定

卡方測定(chi-square)在機率與統計中的定義是量測兩個變量之間的關聯程度，計算後的數值如果為 0 則表示完全相同，而數值越大則表示變數間的相似程度越低，卡方測定演算法也被用於內涵式影像檢索系統中。

$$d_{\text{chi-square}}(H_1, H_2) = \sum_{i=1}^N \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \quad (22)$$

D、相關係數

相關係數(correlation)在統計上的常用來表示兩個隨機變數之間線性關係的強度和方向。1 表示兩個變數完全相同；-1 表示兩個變數的相關係很低。相關係數演算法也被用於內涵式影像檢索系統中。

$$d_{\text{correlation}}(H_1, H_2) = \frac{\sum_i H'_1(i) * H'_2(i)}{\sqrt{\sum_i H'^2_1(i) * \sum_i H'^2_2(i)}} \quad (23)$$

where $H'_k(i) = H_k(i) - (1/N) * (\sum_j H_k(j))$, N equals the number of bins in the histogram.

3.6、人機互動分析

人機互動的分析對於設計人機介面的設計有很重要的影響，依照分析的結果了解在實做系統中人機介面所要包含的功能內容為何。人機互動分析與系統流程表示在圖 11 中，下面將依序說明各步驟間的關係。

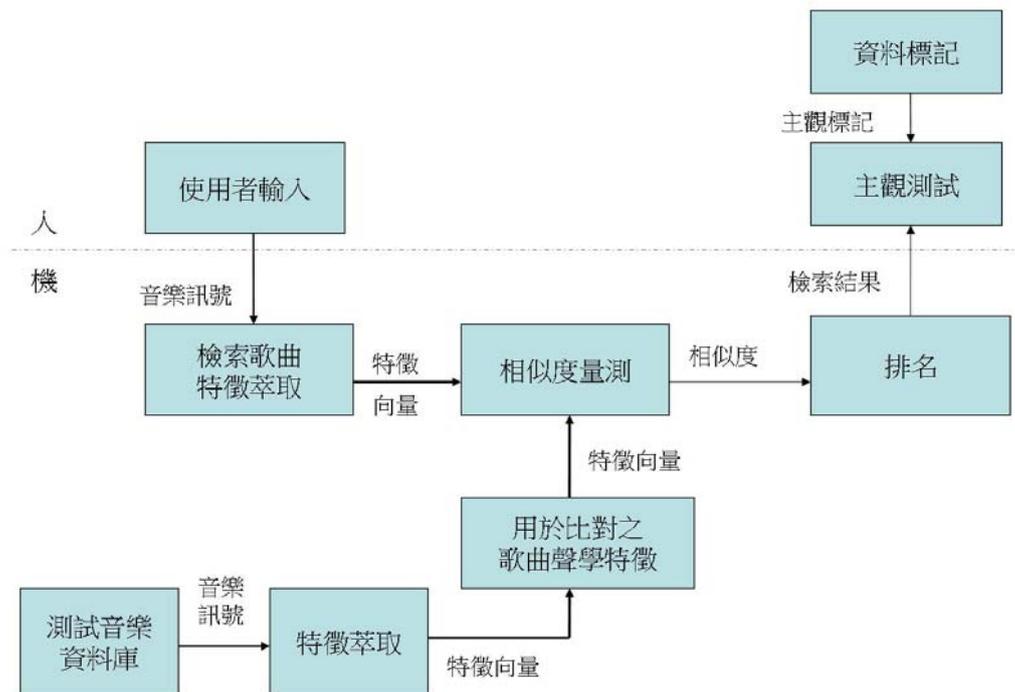


圖 11：系統架構與人機互動分析

- 一、**使用者輸入**：由使用者依照各人喜好或需求自行判斷和選曲，輸入一首用於系統檢索的歌曲，輸入系統作訊號分析。
- 二、**檢索歌曲特徵萃取**：對使用者輸入的檢索歌曲之音樂訊號做特徵萃取，輸出的資料為特徵向量。
- 三、**測試音樂資料庫**：蒐集音樂資料庫，作為與檢索歌曲比對之音樂資料。測試音樂資料庫在輸入音樂資料後同樣地，也必須針對音樂訊號作訊號處理，稱為特徵萃取，特徵萃取後之各項特徵數值以向量格式表示。
- 四、**相似度量測**：比對檢索歌曲的特徵向量與資料庫中用其它歌曲的特徵向量。
- 五、**排名**：將相似度量測計算完成之數據依據依照相似程度排序。
- 六、**主觀量測**：排序後之結果，依據音樂資料標記的主觀資訊判斷檢所資料的正確性，以判斷演算法的效能。
- 七、**資料標記**：以受測者的主觀判斷標記音樂檔案的內容所屬之音樂風格以及情緒類型。

第四章、系統實作及實驗討論

在這個章節中第一節討論系統實作的步驟與測試的音樂檔案格式和內容；第二節討論主觀測試的方法及實驗結果；第三節說明特徵萃取演算法的實驗以及針對實驗結果與理論比較的討論；第四節說明相似度演算法的實作以及針對實驗結果與理論比較的討論；第五節討論音樂片段的時間長度與檢索效能的關連；第六節比較系統中各步驟與測試演算法的執行時間。

4.1、系統實作

本段將探討如何測試不同演算法及效能，目的是為找出最接近人類聆聽音樂感知的演算法。細部的測試系統流程如圖 12 所示。首先準備測試音樂資料庫 (testing music database)，從作者個人數位音樂檔案蒐藏中隨機挑選 161 首歌曲。隨機的方式為挑選每張音樂專輯中曲目順序為三的歌曲。音樂檔案蒐集完成後對音樂檔案的內容進行主觀資料的標記。音樂檔案格式為 Mp3。後面章節將分別敘述在資料標記、音樂訊號前處理、特徵萃取、相似度量測、主觀測試所用到的演算法與效能比較。

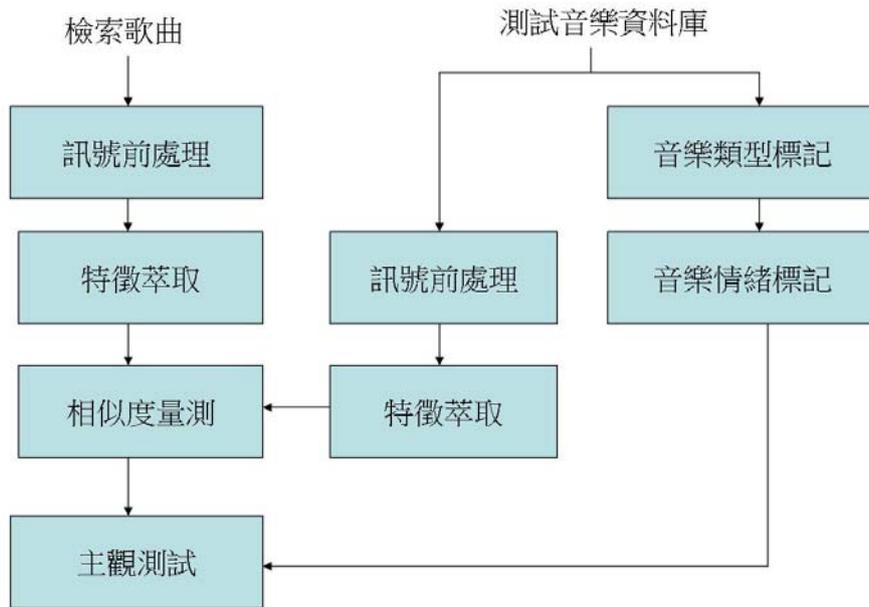


圖 12：細部系統架構流程圖

4.1.1、資料標記

主觀資料標記的目的是為了解人類對於音樂內容的主觀認知。本文中音樂內容的分類有兩大主題：音樂類型與情緒感受。音樂類型的分類本文參考Allmusic網站⁴之分類方法，音樂類型分類採用階層式的分類方法，第一層被區分為十一大類：流行及搖滾(Pop/Rock)、爵士(Jazz)、節奏藍調(R&B)、鄉村(Country)、藍調(Blues)、電子樂(Electronic)、拉丁(Latin)、雷鬼(Reggae)、世界(International)、古典(Classical)，本文以這十一大類作為音樂類型分類的音樂內容主觀標記。另一組主觀標記的資料為情緒感受，情緒感受的分類參考論文 [47]提出的形容詞圈，學者將常用於形容音樂情緒的形容詞整理列表並將形容詞分成八大類。受測者在聽完音樂後，從形容詞圈中提到之形容詞中找出主觀判斷上與個人情緒感受最接近的形容詞標記音樂檔案，依照受測者標記的形容詞即可找出情緒感受的分類。本次測試所用的形容詞中英文對照表參考附錄，測試資料庫中所有的曲目與主觀標記資料也可參考附錄。

⁴ All Music, <http://www.allmusic.com>

4.1.2、內容分析

測試音樂資料庫經過主觀資料標記後，可針對音樂內容分析資料，作為演算法效能評斷的參考。測試音樂資料庫中包含的歌曲類型有流行搖滾(52%)、爵士樂(32%)、鄉村音樂(10%)、電子樂(3%)、節奏藍調(0.4%)、電影配樂(0.9%)、拉丁(0.7%)，如圖 13 所示。

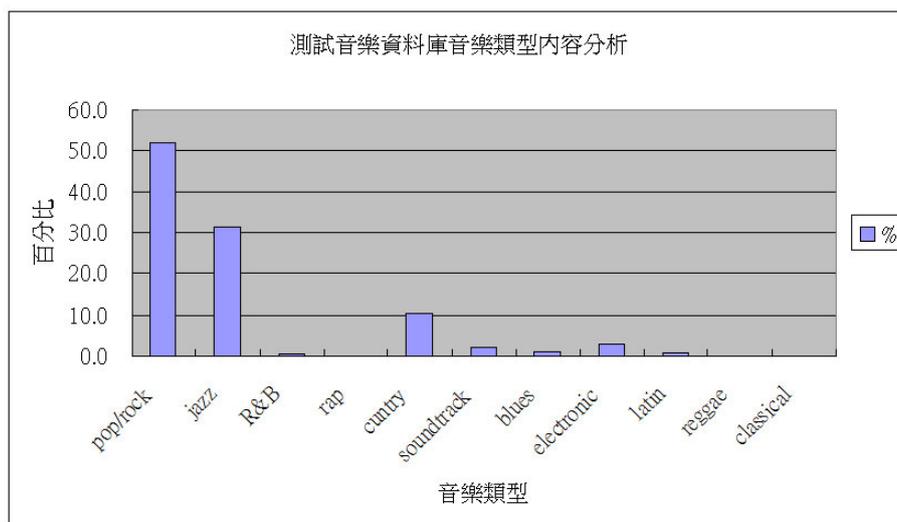


圖 13：測試音樂資料庫音樂類型分析直條圖

情緒感受的類別分析中，測試音樂資料庫中包含的情緒感受屬於第一類占 1.9%、第二類占 14.9%、第三類占 19.9%、第四類占 14.3%、第五類占 21.7%、第六類占 9.9%、第七類占 11.2%、第八類占 6.2%，如圖 14 所示。

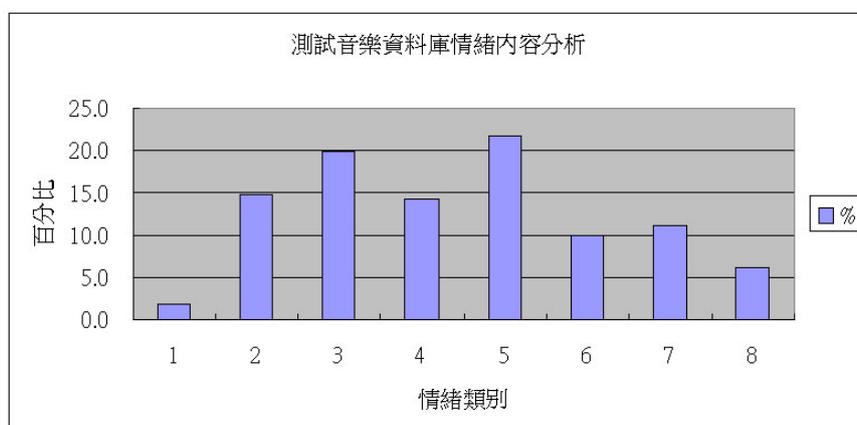


圖 14：測試音樂資料庫情緒感受分析直條圖

4.2、主觀測試

本文以查準率(precision)測試檢索效能，查準率的數學式如式(24)。查準率常用於評斷搜尋引擎的效能。分母為抓回音樂檔案總數，也是在相似度量測後的排序前 N 名的曲目；查準率的分子為抓回來的相關音樂檔案數目，相關音樂檔案的判斷分為音樂類型相關與情緒感受相關，詳細內容將在後面章節討論。

$$\text{Precision} = \text{Relevant Music Files Retrieval} / \text{Total Retrieval Music Files} \quad (24)$$

(Total Retrieval Music Files = 5, 10, 15, 20, 40)

4.2.1、音樂類型測試

在音樂類型測試裡，所有相關音樂檔案(total retrieval music files)被定義為與檢索歌曲類別的歌曲，因此依序檢查每一首歌曲的音樂類型，如果它與檢索歌曲的音樂類型相同則得到一分；反之與檢索歌曲的音樂類型不同則沒有得到分數。圖 15 以節奏特徵和餘弦距離這組演算法作為一個例子，可以看出在音樂類型的測試中，抓回的音樂檔案數目在 10 以上，查準率在圖上有一個明顯的下降，因此在設計搜尋引擎時建議的抓回音樂檔案總數要小於 10。

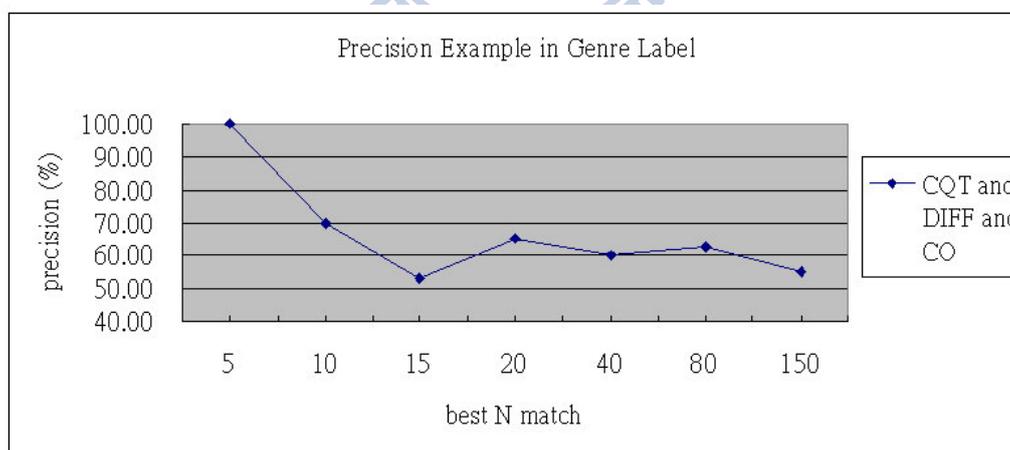


圖 15：音樂類型之檢索效能

4.2.2、情緒感受測試

由於在聆聽音樂時，不同歌曲的情緒變化是可以被接受的，因此在這裡所有相關音樂檔案(total retrieval music files)的定義是，與前一首歌曲的情緒類別差異

為 2 的歌曲，如式(25)表示。如果它與前一首歌曲的情緒類型差小於或等於 2 則得到一分；反之與前一首歌曲情緒類別的差大於 2 則沒有得到分數。圖 16 中的例子是音色特徵與歐基理得距離。

$$|EmotionLabel(n+1) - EmotionLabel(n)| \leq 2 \quad (25)$$

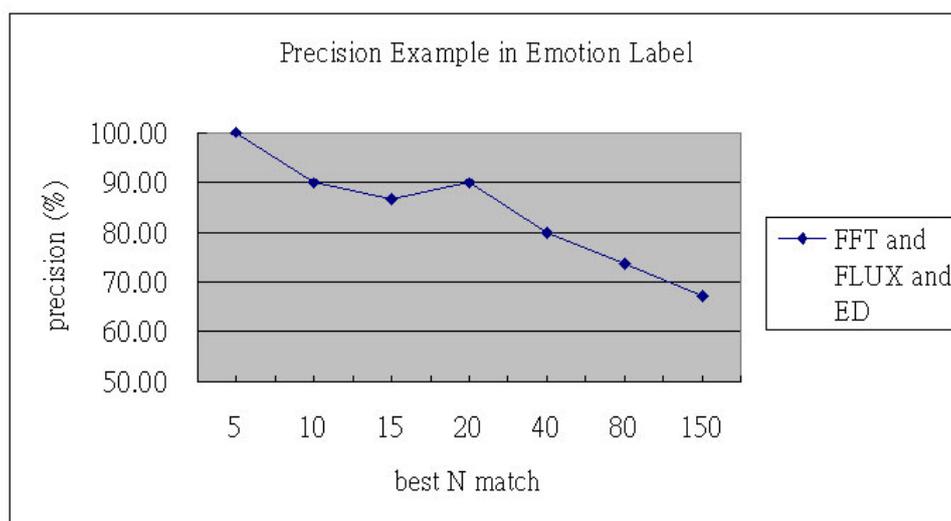


圖 16：情緒感受之檢索效能

4.3、特徵萃取實驗

特徵萃取演算法分成兩部分討論：訊號處理演算法、數值統計演算法。由第三章的討論可知，頻譜特徵測試三種訊號處理演算法，分別是：FFT、MFCC、CQT；頻譜特徵所用的數值統計演算法為 AVG；音程特徵所用的數值統計演算法為 DIFF；節奏特徵所用的數值統計演算法為 ACF；音色特徵所用的數值統計演算法為 FLUX。下面將分別討論實驗結果。

4.3.1、訊號處理演算法

在這個小節討論三種訊號處理演算法，分別在不同的音樂內容中的檢索效能。三種演算法分別是：FFT、MFCC、CQT；三種音樂內容分別是：音樂類型、情緒感受、音樂類型和情緒感受的平均。此外這個實驗的目的是比較三種訊號處理演算法的效能，因此數值統計和相似度量測演算法必須相同。在這個實驗中，

數值統計演算法為平均頻譜特徵(AVG)。檢索效能以平均查準率表示，實驗數據與圖表列在下面文章中。訊號處理演算法分別在三種內容下的檢索結果列在表 1 中；圖 17 和圖 18 分別是訊號處理演算法在分析音樂片段的長度在不同時距下之檢索效能的比較圖；實驗結果的討論在 4.6 節說明。

	音樂類型		情緒感受		類型和感受	
	長時距	短時距	長時距	短時距	長時距	短時距
FFT	82.4	83.33	41.46	56.46	61.93	69.9
MFCC	45.97	57.19	73.75	70.83	59.11	64.01
CQT	70.1	71.77	45.94	45.83	58.02	58.8

表 1：訊號處理演算法檢索效能比較數據表

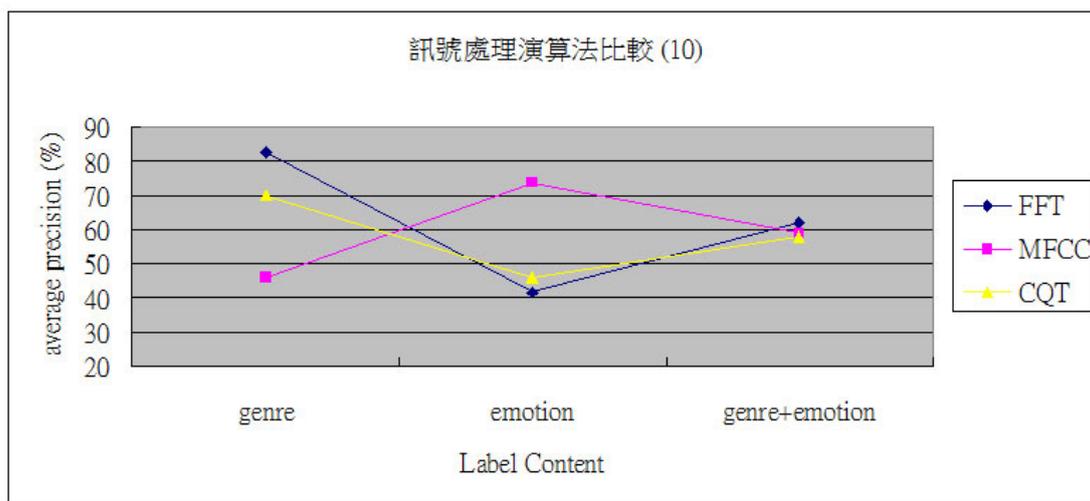


圖 17：訊號處理演算法檢索效能比較 (長時距音樂片段分析)

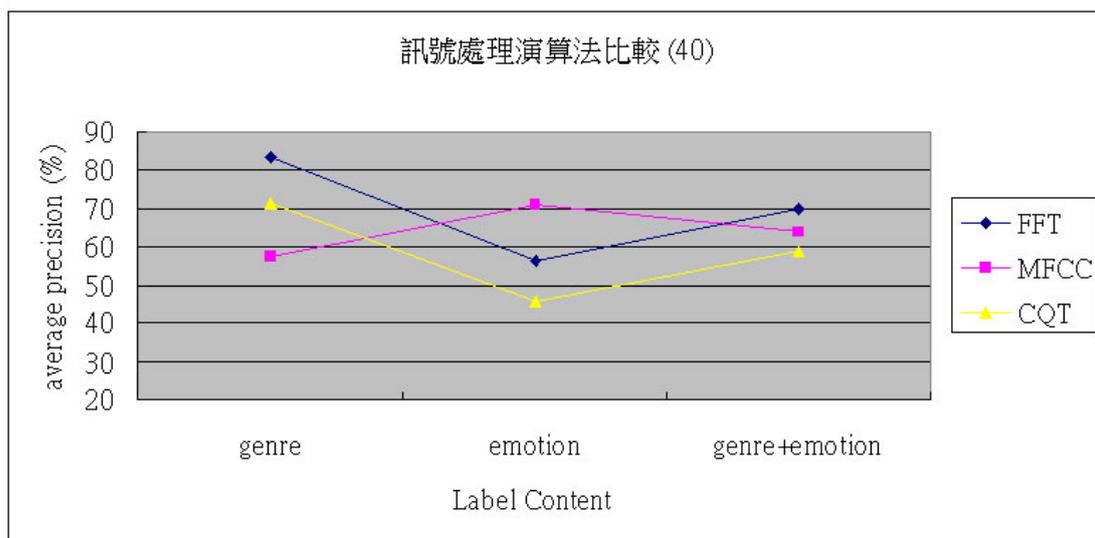


圖 18：訊號處理演算法檢索效能比較 (短時距音樂片段分析)

4.3.2、數值統計演算法

在這個小節討論在四種特徵萃取演算法中所採用的四種數值統計演算法，分別在不同的音樂內容中的檢索效能。四種數值統計演算法分別是：在頻譜特徵中的音框平均(AVG)、在音程特徵中的音框差(DIFF)、節奏特徵中的自相關係數和波峰偵測(ACF)、在音色特徵中的頻譜通量(FLUX)；三種音樂內容分別是：音樂類型、情緒感受、音樂類型和情緒感受的平均。檢索效能以平均查準率表示，實驗數據與圖表列在下面文章中。數值統計演算法分別在三種內容下的檢索結果列在表 2 中；圖 19 和圖 20 分別是數值統計演算法在分析音樂片段的長度在不同時距下之檢索效能的比較圖；實驗結果的討論在 4.6 節說明。

	音樂類型		情緒感受		類型和情緒	
	長時距	短時距	長時距	短時距	長時距	短時距
頻譜特徵	66.16	70.76	53.72	57.71	59.69	64.24
音程特徵	61.46	62.4	74.03	64.48	67.74	63.44
節奏特徵	69.34	70.59	61.35	53.44	65.35	62.01
音色特徵	47.08	40.52	60.83	68.75	53.96	54.64

表 2：數值統計演算法檢索效能比較數據表

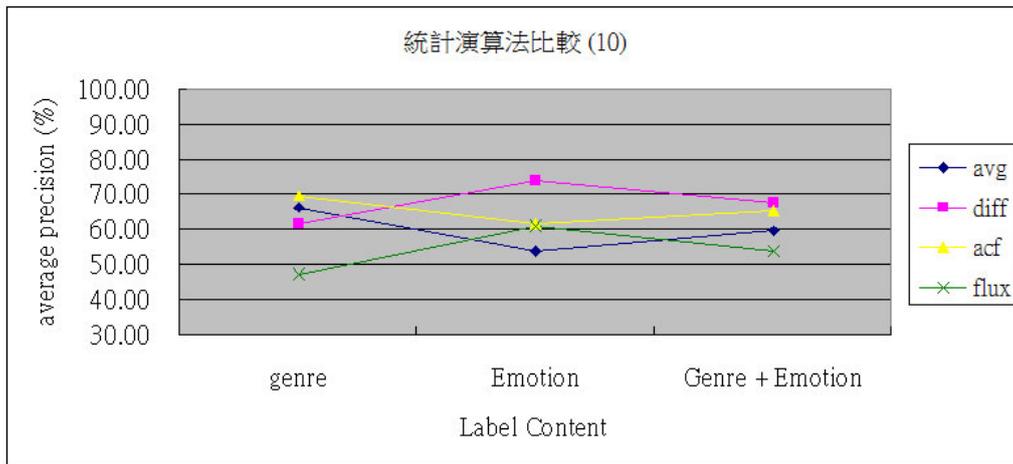


圖 19：數值統計演算法檢索效能比較（長時距音樂片段分析）

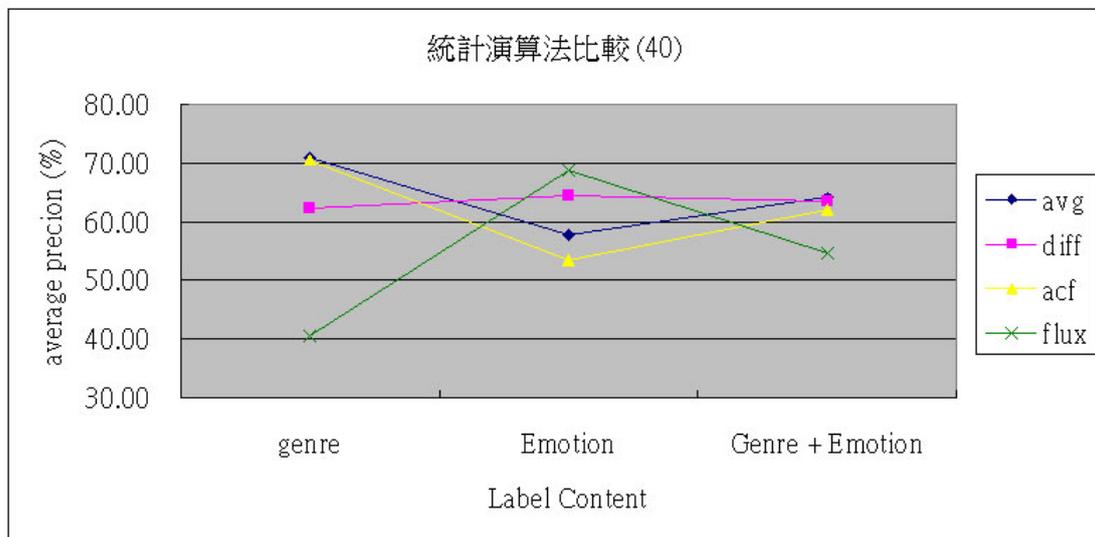


圖 20：數值統計演算法檢索效能比較（短時距音樂片段分析）

4.4、相似度演算法實驗

在這個小節討論四種量測相似度之演算法，分別在不同的音樂內容中的檢索效能。四種演算法分別是：歐基理德距離(ED)、餘弦距離(CD)、相關係數(CO)、卡方測定(CS)；三種音樂內容分別是：音樂類型、情緒感受、音樂類型和情緒感受的平均。在這個實驗中以前面四種特徵的平均查準率表示檢索效能，訊號處理演算法為 FFT，實驗數據與圖表列在下面文章中。

相似度演算法分別在四種內容下的檢索結果列在表 3 中；圖 21 和圖 22 分別

是相似度量測演算法在分析音樂片段的長度在不同時距下之檢索效能的比較圖；實驗結果在 4.6 節討論。

	音樂類型		情緒感受		類型與情緒	
	長時距	短時距	長時距	短時距	長時距	短時距
ED	58.92	57.64	65.9	62.78	62.41	60.21
CD	60.1	62.4	60.59	63.02	60.35	62.71
CO	61.81	64.31	60.42	58.85	61.11	61.58
CS	63.61	59.93	62.12	59.72	62.86	59.83

表 3：相似度量測演算法檢索效能比較數據表

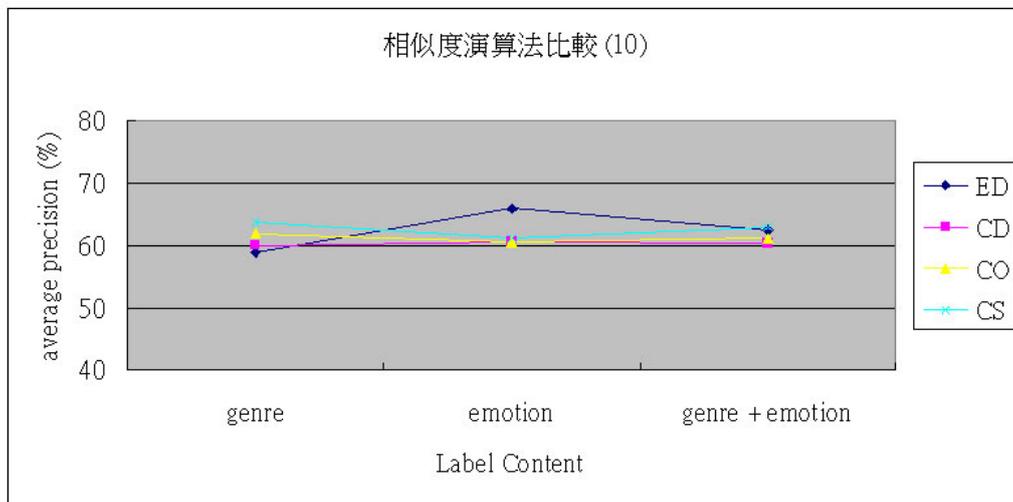


圖 21：相似度量測演算法檢索效能比較 (長時距音樂片段分析)

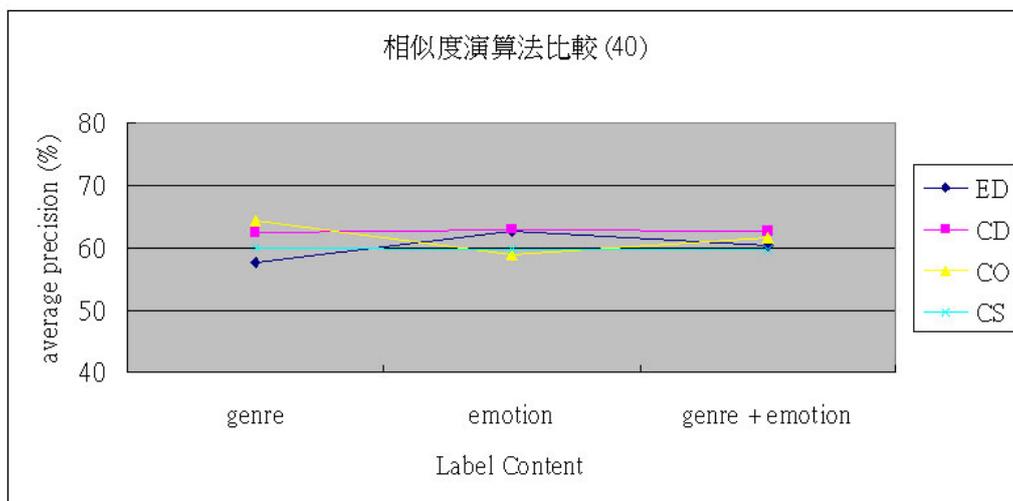


圖 22：相似度量測演算法檢索效能比較 (短時距音樂片段分析)

4.5、執行時間比較

本次實驗所用之機器規格為 2.16G 的 CPU 和 3G 的 RAM。作業系統為 windows VISTA。所使用的軟體是 MATLAB。以測試音樂資料庫的前三首歌曲的資料做小規模的測試比較各步驟與各演算法的運算時間。

首先測試訊號處理演算法，在這裡測試三種訊號處理演算法用在計算頻譜特徵所花費的時間長短，相似度演算法為歐基理德距離，數據表示在表 4 中。

演算法	執行時間 (s)
FFT	56.217
MFCC	74.406
CQT	68.2795

表 4：訊號處理演算法運算時間比較

接著測試數值統計演算法，在這裡測試四種數值統計演算法用在計算四種特徵所花費的時間長短，在這次的測試中，所有特徵中之訊號處理演算法皆為 FFT，相似度演算法為歐基理德距離，數據表示在表 5 中。

演算法	執行時間 (s)
頻譜特徵 (avg)	55.4107
音程特徵 (diff)	80.1601
節奏特徵 (acf)	143.8077
音色特徵 (flux)	55.7189

表 5：數值統計演算法運算時間比較

第三測試相似度量測演算法，在這裡測試四種相似度量測演算法用在比對音樂片段所花費的時間長短，在這次的測試中頻譜特徵為測試的演算法，數據表示在表 6 中。

演算法	運算時間 (s)
歐基理德距離(ED)	0.0949
餘弦距離(CD)	0.1214
相關係數(CO)	0.0615
卡方測定(CS)	0.0787

表 6：相似度量測演算法運算時間比較

最後測試系統中三個步驟分別花費的執行時間，前處理的步驟包含：mp3檔案的解碼⁵、降取樣、音樂片段分割；特徵萃取演算法以節奏特徵作為代表，訊號處理演算法為CQT；相似度演算法以歐基理德距離為例子。數據表示在表7。

演算法	執行時間 (s)
前處理	22.0193
特徵萃取	486.7898
相似度	0.0949

表 7：系統中各步驟運算時間比較

4.6、實驗討論

A、主觀測試

(1) 音樂類型測試

在音樂類型的測試裡，檢索效能最高的演算法是 CQT 與音程特徵及餘弦距離的相似度量測在長時距的音樂片段分析之下，平均查準率達 94.17%。最低的檢索效能的演算法為 MFCC 與音色特徵及餘弦距離的相似度量測在短時距的音樂片段分析之下，平均查準率為 12.50%。所有演算法的實驗數據可參考附錄。

(2) 情緒感受測試

在情緒感受的測試裡，檢索效能最高的演算法是 MFCC 與頻譜特徵及餘弦距離的相似度量測在長時距的音樂片段分析之下，平均查準率達 98.75%。最低的檢索效能的演算法為 FFT 與頻譜特徵及卡方測定的相似度量測在長時距的音樂片段分析之下，平均查準率為 32.08%。所有演算法的實驗數據列在附錄中。

B、特徵萃取演算法

特徵萃取演算法分成訊號處理演算法和數值統計演算法兩部分討論。在訊號處理演算法的討論中，以訊號處理演算法在四種特徵萃取演算法下的平均查準率討論實驗結果並歸納分析。在數值統計演算法的討論中，以三種訊號處理演算法

⁵ mp3read and mp3write for Matlab: <http://labrosa.ee.columbia.edu/matlab/mp3read.html>

的平均查準率討論實驗結果並歸納分析。

(1) 訊號處理演算法

在音樂類型的測試裡，FFT 演算法有最好的效能，其次是 CQT，而 MFCC 則是最低。在情緒感受的測試裡，MFCC 演算法有最好的效能，而 FFT 和 CQT 則依據不同的音樂片段長度而在排名上有所改變。最後一個測試同時檢查音樂類型和情緒感受，在這個測試裡，長時距的音樂片段分析中三個演算法的效能沒有明顯的差距；短時距的音樂片段分析裡，三種訊號處理演算法效能的由高至低分別是 FFT、MFCC、CQT。

從理論推導中可以知道，FFT 和 CQT 這兩種演算法都是從離散傅立葉轉換 (DCT)開發而成，這兩個演算法的差別在頻率上的取樣不同以及運算時間上的差異。FFT 在頻率上的取樣間隔為等差，CQT 在頻率上的取樣間隔為等比。CQT 是針對分析音樂訊號所開發的演算法，在這次的實驗中 CQT 演算法的平均效能雖然不是最高，但是在音樂類型的測試中，CQT 在音程特徵中得到總體最高的查準率。另外，實驗結果也得到 FFT 這個演算法同時具有快速以及準確的特性。

最後一個測試的演算法 MFCC 原是針對語音訊息所開發的演算法，在訊號分析中只留下語音的頻段，就理論推導而言這樣的作法也許會將音樂中音色的訊息除去。然而就這次的實驗結果來看，MFCC 用於分析音樂訊號中的情緒訊息，而在音樂類型的分析上就不合適。

(2) 數值統計演算法

在音樂類型的測試裡，節奏特徵所用的演算法(acf)和頻譜特徵所用的演算法(avg)分別在長時距音樂片段分析和短時距音樂片段分析的實驗中得到最高的查準率，音色特徵(flux)得到最低的查準率。在情緒感受的測試裡，音色特徵(flux)和音程特徵(diff)則分別在長時距音樂片段分析和短時距音樂片段分析的實驗中得到最高的查準率，而頻譜特徵(avg)和節奏特徵(acf)則分別在長時距音樂片段分析和短時距音樂片段分析的實驗中檢索效能最低。

就實驗結果來看，音程特徵和音色特徵的演算法在數學定義上都是以相鄰音

框間的數值差距來定義特徵向量，所以音樂上情緒感受在數學上以相鄰音框間的數值變化來表示可以得到較高的檢索效能。而頻譜特徵和節奏特徵則是以獨立音框內的絕對數值大小來定義特徵向量，因此從實驗數據中可得在分析音樂類型時以獨立音框內的絕對數值大小來表示特徵向量，可以得到比較好的檢索效能。

C、相似度演算法

在相似度演算法的實驗中,四種相似度演算法在檢所效能上沒有明顯的差異，因此推斷相似度演算法對於檢所效能沒有絕對的影響。另外推論總體相似度演算法的設計也許對於檢所效能會有影響以及特徵萃取演算法對於檢所效能具有絕對的影響力。

D、分析音樂片段的時距長短

以實驗數據來看，分析的音樂片段時距長短對於實驗結果沒有絕對的影響，特徵萃取演算法的影響大過於音樂片段時距長短的影響。在四種特徵萃取演算法裡，可依據不同的音樂內容判斷最合適的時距長度或由實驗結果歸納而得。

E、運算時間

在運算時間方面，整體系統中以特徵萃取這個步驟所需要花費的時間最長，其次是訊號前處理，相似度比對的步驟所需要花費的時間最少。各別演算法的部分，三種訊號處理演算法花費的時間由高至低分別為 MFCC、CQT、FFT；數值統計演算法花費的時間由高至低分別為節奏特徵、音程特徵、音色特徵、頻譜特徵。

4.7、圖形化使用者介面

由前面的實驗結果挑選出檢索效能最高的演算法作為音樂篩選器的演算法，音樂篩選器一的演算法是 FFT、頻譜特徵和卡方測定；音樂篩選器二的演算法是 MFCC、頻譜特徵和餘弦距離；音樂篩選器三的演算法是 CQT、音程特徵和餘弦距離。

系統操作的步驟如下：首先在視窗左上角的下拉式選單中，使用者依據個人

的喜好或需求選擇檢索歌曲。點選完成後，歌曲的相關訊息會被顯示在下方的欄位中。接著使用者可以依據個人的喜好選擇音樂篩選器，以及輸入所需要的曲目數量。按下「GO」按鍵則開始比對檢索歌曲與資料庫中的資料，比對完成後檢索的結果列在右邊的空格裡。最後按下「Play」按鈕呼叫電腦預設的音樂播放程式播放音樂檔案聆聽音樂。程式執行完成之畫面如圖 23 所示。

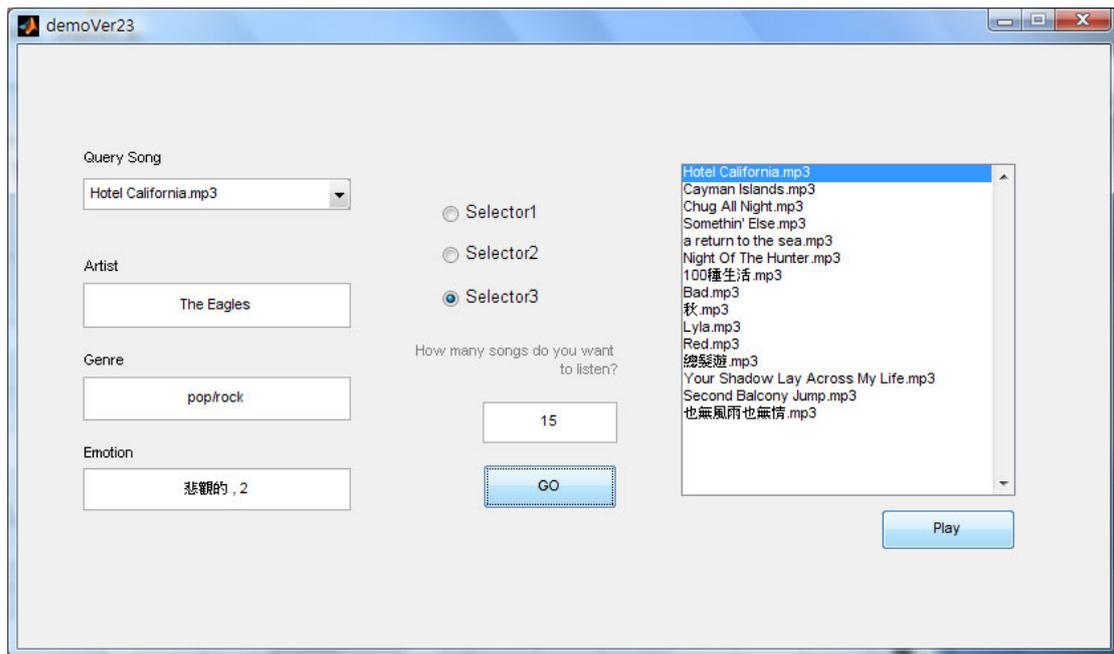


圖 23：圖形化使用者介面程式執行完成之畫面

第五章、結論與未來展望

在這個章節中第一節總結本篇論文的研究方法及實驗結果；第二節討論實驗方法未來建議改良的方向及系統應用。

5.1、結論

本篇論文應用內涵式音樂資訊檢索技術開發音樂內容的搜尋引擎，目標為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演算法，取建立資料的數值特徵；以相似度量測演算法計算音樂資料中相似的內容。在特徵萃取中測試四種演算法；在相似度以局部相似度量測和總體相似度量測計算資料的相近程度，並且測試四種相似量測演算法。最後以人為標記的資料測試各演算法的檢索結果是否符合一般對於音樂內容的認知及情緒感受。

本次的實驗結果在音樂類型的測試裡，最高的檢索效能達 94.17%，其演算法是 CQT 與音程特徵及餘弦距離的相似度量測在長時距的音樂片段分析之下；在情緒感受的測試裡，檢索效能最高的演算法是 MFCC 與頻譜特徵及餘弦距離的相似度量測在長時距的音樂片段分析之下，平均查準率達 98.75%。在運算時間方面，整體系統中以特徵萃取這個步驟所需要花費的時間最長。

並且從實驗結果中歸納出以下的結論。在音樂類型的測試裡，特徵萃取演算法的以單獨音框內的絕對數值大小設計演算法可以得到較高的檢索效能；在情緒感受的測試裡，以相鄰音框間的數值變化設計演算法可以得到較高的檢索效能；而在同時符合相同音樂類型及連續的情緒變化這兩項條件的測試中，四種特徵萃取演算法的效能差異不大。在相似度演算法的實驗中，不同的相似量測演算法對於檢索效能的影響差異不多。另外分析之音樂片段的時間長度對於檢索效能的影響不大，特徵萃取演算法對於檢索效能的影響最多。

5.2、未來展望

從實驗結果中得到以下的推論，建議未來的研究以及應用方法，分別針對系

統中各步驟討論。

- A、特徵萃取的演算法中將多項特徵結合，使相似度的判斷上可以同時符合多項條件提升檢索效能。另外也可以針對不同的音樂內容修改特徵萃取的演算法，使其對於音高或是節奏的判斷更為準確，減少雜訊的影響。最後也可以針對運算速度的提升做改進。
- B、在相似度量測的計算中，改良局部相似度和總體相似度的演算法，也許可以提升檢索效能。
- C、在人機互動方面，由於情緒反應沒有統一的標準，因此在人機互動方面建議加入使用者回饋機制，使系統有更完整的資訊用於判斷音樂的內容是否符合使用者的認知。另外在輸入檢索資料的部分，在輸入檢索歌曲的同時也加入詮釋資料的輸入，使系統在判斷歌曲相似度上有更多的資料提升檢索效能。
- D、在系統未來應用方向上，將平臺移植於手機、MP3 播放器上。另外也可將此項功能加入電腦的音樂播放軟體中或線上數位音樂資料庫的檢索上。



參考文獻

- [1] 洪元元，從使用者聆賞歷程探討線上音樂分類架構，碩士論文，國立台灣大學圖書資訊學研究所，民國98年7月。
- [2] Furman, C.E., and Duke, R.A. (1988). Effects of majority consensus on preferences for recorded orchestral and popular music. *Journal of Research in Music Education*, 36, 4, 220-231.
- [3] Geringer, J.M. (1982). Verbal and operant music listening in relationship to age and musical training, *Psychology of music*(special issue), 47-50.
- [4] Holbrook, M.B., and Schindler, R.M. (1989). Some exploratory findings on the development of musical tastes. *Journal of Consumer Research*, 20, 119-124.
- [5] Rubin, D.C., Rahhal, T.A., and Poon, L.W. (1998). Things learned in early adulthood are remembered best. *Memory & Cognition*, 26, 1, 3-19.
- [6] 許游雅，國小學童音樂偏好與情緒感受之探究，碩士論文，國立屏東教育大學音樂學系，民國98年7月。
- [7] Ze-Nian Li, Mark S. Drew. *Fundamentals of Multimedia*. Prentice Hall United States ed edition. November 1, 2003
- [8] Sonnenschein D. . *Sound design : the expressive power of music, voice, and sound effects in cinema*. Studio City, CA : Michael Wiese Productions, 2001.
- [9] 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.
- [10] 陳若涵，以音樂內容為基礎的情緒分析與辨識，碩士論文，國立清華大學資訊系統與應用研究所，民國九十五年六月。
- [11] Y. Zhuang et al. Building a Personalized Music Emotion Prediction System, *PCM 2006, LNCS 4261*, pp. 730 - 739, 2006.
- [12] Hu, N., Dannenberg, R., and Tzanetakis, G. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (New Paltz, NY, Oct. 19--22)*. IEEE Computer Society Press, Piscataway, NJ, 2003, 185--188.
- [13] G. Tzanetakis, P. Cook. Music Genre Classification of Audio Signals. *IEEE transactions on audio, speech, and language processing*, vol.10, no. 5, July 2002。
- [14] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, "Music type classification by spectral contrast features," in *Int. Conf. Multimedia Expo.*, vol. 1, 2002, p. 113–116.
- [15] L. Lu, D. Liu, H. J. Zhang. Automatic Mood Detection and Tracking of Music Audio Signals. *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14, No. 1, pp. 5-18, January 2006。

- [16] M. D. Korhonen, D. A. Clausi, and M. E. Jernigan, "Modeling emotional content of music using system identification," *IEEE Trans. Syst., Man., Cybern.*, vol. 36, no. 3, pp. 588–599, Jun. 2006.
- [17] Y. H. Yang, Y. C. Lin, Y. F. Su and H. H. Chen. "A Regression Approach to Music Emotion Recognition". *IEEE transactions on audio, speech, and language processing*, vol.16, no. 2, February 2008 .
- [18] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "Music emotion classification: A regression approach," in *Proc. IEEE Int. Conf. Multimedia and Expo. 2007 (CME'07)*, Beijing, China, pp. 208-211.
- [19] T. L. Wu and S. K. Jeng. Probabilistic estimation of a novel music emotion model. In *14th International Multimedia Modeling Conference*. Springer, 2008.
- [20] Field, A., Hartel, P. and Mooij, W. 2001 Personal DJ, an Architecture for Personalised Content Delivery. *International World Wide Web Conference Proceedings of the 10th international conference on World Wide Web*, Hong Kong, 1-7.
- [21] S. Pauws and B. Eggen. PATS: Realization and user evaluation of an automatic playlist generator. In *ISMIR*, 2002.
- [22] Li, T. and Ogihara, M. 2004 Content-based music similarity search and emotion detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, 705-708.
- [23] Foote, Jonathan. Content-Based Retrieval of Music and Audio, in C.-C. J. Kuo et al., editor, *Multimedia Storage and Archiving Systems II*, Proc. of SPIE, Vol. 3229, pages 138-147, 1997.
- [24] B. Logan and A. Salomon 2001. A music similarity function based on signal analysis. In *Proc IEEE Intl Conf on Multimedia and Expo*.
- [25] B. Logan and A. Salomon. A content based music similarity function. Technical report, Compaq Cambridge Research Laboratory, June 2001.
- [26] Owen Craigie Meyers. A Mood-Based Music Classification and Exploration System. Master thesis. Program in Media Arts and Sciences, MASSACHUSETTS INSTITUTE OF TECHNOLOGY. June 2007.
- [27] 黃捷，運用樣式所引技術之高效性內涵式音樂檢索，碩士論文，國立成功大學資訊工程研究所，民國九十七年七月。
- [28] Z. Xiao, E. Dellandréa, W. Dou, L. Chen, What is the Best Segment Duration for Music Mood Analysis ?, *International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 17-24. 2008.
- [29] R. M. Parry and I. Essa, "Feature weighting for segmentation," in *Proc. 5th Int. Conf. Music Information Retrieval*, 2004, pp. 116–119.
- [30] 陳彥傑，應用於音樂資料的快速重複片斷找尋演算法：以人類感知為依據的

方法，碩士論文，國立成功大學電機工程學系，2004

- [31] Fastl, H., E. Zwicker, et al. (2007). "Psychoacoustics Facts and Models." 3rd. from <http://dx.doi.org/10.1007/978-3-540-68888-4>.
- [32] Xavier Serra and Julius Smith III, Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition, *Computer Music Journal*, Vol. 14, No. 4, Winter 1990.
- [33] Brown, J.C. and Puckette, M.S. (1992). "An Efficient Algorithm for the Calculation of a Constant Q Transform", *J. Acoust. Soc. Am.* 92, 2698-2701. (1992)
- [34] M. Wu, D. Wang, and G. J. Brown, "A multi-pitch tracking algorithm for noisy speech," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 1, pp. 369–372, Orlando, Fla, USA, May 2002.
- [35] M. P. Ryynnen and A. P. Klapuri, "Transcription of the singing melody in polyphonic music," in *Proc. 7th International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, 2006.
- [36] Gómez, Emilia, Klapuri, Anssi and Meudic, Benoît(2003). 'Melody Description and Extraction in the Context of Music Content Processing', *Journal of New Music Research*,32:1,23 — 40
- [37] M. Marolt, "Audio melody extraction based on timbral similarity of melodic fragments," in *Proc. EUROCON 2005*, Nov. 2005.
- [38] Gómez, Emilia, Klapuri, Anssi and Meudic, Benoît(2003)'Melody Description and Extraction in the Context of Music Content Processing',*Journal of New Music Research*,32:1,23 -40
- [39] A. S. Durey and M. A. Clements, "Features for melody spotting using hidden Markov models," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, 2002, pp. 1765-1768.
- [40] K. Dressler, "Extraction of the melody pitch contour from polyphonic audio," in *Proc. 6th International Conference on Music Information Retrieval*, Sept. 2005. MIREX05 extended abstract, available online <http://www.musicir.org/evaluation/mirex-results/articles/melody/dressler.pdf>.
- [41] D. B. Loeffler. Instrument Timbres and Pitch Estimation in Polyphonic Music. Master's thesis, Georgia Institute of Technology, April 2006.
- [42] J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signals," in *Proceedings of the International Symposium on Music Information Retrieval*, London, UK, 2005.
- [43] Hadi Harb, Liming Chen, A Query by Example Music Retrieval Algorithm, *Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services WIAMIS03*, pp 122-128, World Scientific publications,

April 9-11, Queen Mary, University of London, UK

- [44] Gang Qian , Shamik Sural , Yuelong Gu , Sakti Pramanik. Similarity between Euclidean and cosine angle distance for nearest neighbor queries, Proceedings of the 2004 ACM symposium on Applied computing, March 14-17, 2004, Nicosia, Cyprus [doi>10.1145/967900.968151]
- [45] 林俊男，人工生音信號意象感知評價之研究，碩士論文，國立雲林科技大學工業設計研究所，民89。
- [46] R. E. Thayer. The Biopsychology of Mood and Arousal, New York, Oxford University Press, 1989.
- [47] K. Hevner. Experimental Studies of the Elements of Expression in Music. American Journal of Psychology, 48: 246-68. 1936.
- [48] Emery Schubert. Update of the Hevner Adjective Checklist. Perceptual and Motor Skills, 96:1117-1122, 2003.
- [49] 莊雅量，應用音樂性聲音訊號傳遞訊息屬性的可能性研究—以行動電話之“聽聲辨人”為例，碩士論文，國立交通大學應用藝術研究所，2001。
- [50] Schubert, E. Measurement and Time Series Analysis of Emotion in Music. Diss. University of New South Wales, 1999.
- [51] X. Hu. Music and Mood: Where Theory and Reality Meet. iConference, Feb, 2010.
- [52] Jyh-Shing Roger Jang, "Speech and Audio Processing Toolbox", available from the link at the author's homepage at "<http://mirllab.org/jang>".

附錄

音樂情緒標記形容詞中英對照表

Adjective List			
英	中	英	中
spiritual	崇高純潔的	dreamy	夢幻般的
lofty	崇高的	yielding	柔順的
awe-inspiring	激勵人心的	tender	溫柔的
dignified	尊貴的		
scared	神聖的	sentimental	多愁善感的/感情用事的
solemn	隆重的	longing	孤獨的
		yearning	思念的/渴望的
sober	冷靜的	pleading	懇求的
serious	嚴肅的	plaintive	悲傷的/哀愁的
		lyrical	感情豐富的
pathetic	可憐的		
doleful	悲傷的	leisurely	悠閒的
sad	傷心的	satisfying	舒適的/和藹可親的
mournful	悲觀的	serene	平靜的/穩重的
tragic	悲慘的	tranquil	平靜的/安寧的
melancholy	憂鬱的	quiet	安靜的
frustrated	挫敗的	smoothing	柔順的
depressing	壓抑的		
gloomy	陰暗的	humorous	幽默的
heavy	沉重的	playful	玩樂的
dark	黑暗的	whimsical	異想天開的/古怪的
		fanciful	幻想的
sprightly	生氣勃勃的/輕快的		
delicate	敏銳的/細緻的	graceful	優雅的
light	明快的	quaint	古雅的
merry	歡樂的	impetuous	狂暴的/猛烈的
joyous	樂趣的	restless	煩躁的
gay	有朝氣的	vigorous	精力充沛的
happy	快樂的	robust	喧鬧的/結實的
cheerful	興高采烈的		
bright	明亮的	emphatic	有力的
exhilarated	興奮的	martial	軍隊的/軍事的
		ponderous	笨重的
soaring	飛翔的	majestic	威嚴的
triumphant	勝利的	exalting	頌揚的/讚揚的
dramatic	戲劇性的		
passionate	激情的	sensational	引起轟動的
exciting	興奮的	agitated	激動的

測試音樂資料庫曲目清單與標記內容

#	Artist	Filename	Genre	Adjective	Emotion
1	盧廣仲	100 種生活.mp3	pop/rock	leisurely	4
2	黃建為	5 Strings.Mp3	pop/rock	sprightly	5

3	Stars	Ageless Beauty.mp3	electronic	vigorous	8
4	王若琳	As Love Begins To Mend.mp3	pop/rock	yielding	3
5	Doug Munro, Mariano Mangas	Autumn Leaves.mp3	jazz	graceful	5
6	方大同	Bad.mp3	pop/rock	quaint	5
7	Placebo	Battle for the Sun.mp3	pop/rock	impetuous	7
8	Mr. Scruff	Beyond.mp3	electronic	whimsical	5
9	Miles Davis	Blue in Green.mp3	jazz	tranquil	4
10	Charles Mingus	Boogie Stop Shuffle [Unedited Form].mp3	jazz	quaint	5
11	Beirut	Brandenburg.mp3	country	gloomy	2
12	The Red Hot Chili Peppers	Buckle Down.mp3	pop/rock	robust	8
13	Selfkill	Cake on Body.mp3	pop/rock	plaintive	3
14	順子	Can't Get Enough.mp3	pop/rock	passionate	7
15	Kings of Convenience	Cayman Islands.mp3	pop/rock	serene	4
16	The Eagles	Chug All Night.mp3	pop/rock	exalting	8
17	Club 8	Close to Me.mp3	pop/rock	bright	6
18	Blur	Crazy Beat.mp3	pop/rock	exciting	7
19	The Cardigans	Daddy's Car.mp3	pop/rock	light	5
20	Bebo Valdés	Danzas Cubanas, for piano~No. 1.mp3	jazz	delicate	5
21	Billie Holiday	Darn That Dream.mp3	jazz	tranquil	4
22	Quincy Jones	Desafinado.mp3	jazz	joyous	6
23	Hank Mobley	Dig Dis.mp3	jazz	sprightly	5
24	Fourplay	Eastern Sky.mp3	jazz	soaring	7
25	熊寶貝	Feel.mp3	pop/rock	mournful	2
26	The Velvet Underground	Femme Fatale.mp3	pop/rock	yearning	3
27	Herbie Hancock	Firewater.mp3	jazz	leisurely	4
28	Tizzy Bac	For the Way I Live.mp3	pop/rock	doleful	2
29	Marcus Miller	Free to Be.mp3	jazz	exhilarated	7
30	Wynton Marsalis Quartet	Free.mp3	jazz	smoothing	4
31	Amy Winehouse	Fuck Me Pumps.mp3	jazz	playful	5
32	DEPAPEPE	Great Escape.mp3	easy listening	happy	6
33	DEPAPEPE	Hachiroku.mp3	easy listening	light	5
34	Sarah Vaughan	He's My Guy.mp3	jazz	tender	3
35	Radiohead	High and Dry.mp3	pop/rock	delicate	5
36	Zee Avi	Honey Bee.mp3	country	leisurely	4
37	The Eagles	Hotel California.mp3	pop/rock	mournful	2
38	Tizzy Bac	How About.mp3	pop/rock	depressing	2
39	The Innocence Mission	I Never Knew You from the Sun.mp3	country	sentimental	3
40	Belle & Sebastian	If She Wants Me.mp3	pop/rock	gay	6
41	The Cardigans	In The City.mp3	pop/rock	cheerful	6
42	The Eagles	In the Afternoon.mp3	pop/rock	delicate	5
43	Graham Coxon	In the Morning.mp3	pop/rock	lyrical	4
44	Alison Krauss	It Wouldn't Have Made Any Difference.mp3	country	yielding	3
45	Zion Lockwood	Jazzy June.mp3	electronic	quaint	5
46	Marcus Miller/Lalah Hathaway	La Villette.mp3	jazz	lyrical	4
47	順子	Lex Yeux Ouverts.mp3	pop/rock	yielding	3

48	The Eagles	Life In The Fast Lane.mp3	pop/rock	agitated	7
49	Janis Ian	Life Is Never Wrong.mp3	country	gay	6
50	Sergio Mendes	Like a Lover.mp3	electronic	sprightly	5
51	Carpenters	Looking For Love.mp3	pop/rock	doleful	2
52	Richard Hawley	Love of My Life.mp3	pop/rock	sentimental	3
53	昊恩家家	Love, Love, Love.mp3	pop/rock	yearning	3
54	Louis Armstrong	Loveless Love.mp3	jazz	happy	6
55	Oasis	Lyla.mp3	pop/rock	vigorous	8
56	Stereophonics	Madame Helga.mp3	pop/rock	impetuous	7
57	Kings of Convenience	Me in You.mp3	pop/rock	serene	4
58	K 歌情人電影原聲帶	Meaningless Kiss.mp3	pop/rock	sentimental	3
59	Dave Weckl Band	Mesmer-Eyes.Mp3	jazz	whimsical	5
60	Jacob Young	Minor Peace.mp3	jazz	plaintive	3
61	Armstrong/Ella Fitzgerald/Louis Armstrong	Moonlight in Vermont.mp3	jazz	quiet	4
62	Tizzy Bac	Multiply.mp3	pop/rock	pathetic	2
63	George Benson	My Latin Brother.mp3	jazz	merry	6
64	Johanna Grussner	My Ship.mp3	jazz	satisfying	4
65	Bob Dylan	My Wife's Home Town.mp3	pop/rock	gloomy	2
66	Cannonball Adderley	Mystified (aka Angel Face).mp3	jazz	light	5
67	Piano Magic	Night Of The Hunter.mp3	pop/rock	melancholy	2
68	Sonny Rollins	Nishi.mp3	jazz	graceful	5
69	DEPAPEPE	Old Beach.mp3	easy listening	leisurely	4
70	The Triplets of Belleville 電影原聲帶	Opening Theme.mp3	stage & screen	light	5
71	Frente!	Ordinary Angels.mp3	pop/rock	bright	6
72	The Eagles	Out Of Control.mp3	pop/rock	impetuous	7
73	João Gilberto	Para Machuchar Meu Coracao (To Hurt My Heart).mp3	jazz	satisfying	4
74	Corinne Bailey Rae	Put Your Records On.mp3	r&b	joyous	6
75	巴奈	Rainy Day Hu Da Nan 下雨天.mp3	country	yearning	3
76	Belly	Red.mp3	pop/rock	gloomy	2
77	咖啡王子一號店	Sad thing.mp3	stage & screen	doleful	2
78	Daria Toffali	Samba de uma nota so.mp3	jazz	gay	6
79	Dexter Gordon	Second Balcony Jump.mp3	jazz	soaring	7
80	陳綺貞	Sentimental Kills.mp3	pop/rock	restless	7
81	Lifehouse	Sky Is Falling.mp3	pop/rock	mournful	2
82	Cannonball Adderley	Somethin' Else.mp3	jazz	exciting	7
83	胡德夫	Standing on my land.mp3	country	solemn	1
84	The Who	Substitute.mp3	stage & screen	vigorous	8
85	Radiohead	Subterranean Homesick Alien.mp3	pop/rock	heavy	2
86	9 Lazy 9	Sunday Monday.mp3	electronic	sprightly	5
87	Dave Brubeck Quartet	Take Five.mp3	jazz	delicate	5
88	Bebel Gilberto	Tanto Tempo [Peter Kruder Remix].mp3	latin	fanciful	5

89	Weather Report	Teen Town.mp3	jazz	exhilarated	7
90	Stanley Clarke	The Dancer.mp3	jazz	light	5
91	Brian Chiu	The Portrait of J.mp3	jazz	smoothing	4
92	Richard Bona/Susheela Raman	Tiki.mp3	jazz	quiet	4
93	Sixpence None The Richer	Too Far Gone.mp3	pop/rock	sad	2
94	黃小楨	Toy.mp3	pop/rock	plaintive	3
95	New Order	Turn my way.mp3	pop/rock	sober	1
96	Malia	Two fugitives in the night.mp3	jazz	sentimental	3
97	Hank Mobley	Up a Step [-].mp3	jazz	smoothing	4
98	王若琳	Vincent.mp3	pop/rock	sad	2
99	The Calling	Wherever You Will Go.mp3	pop/rock	vigorous	8
100	Coldplay	White Shadows.mp3	pop/rock	emphatic	8
101	Oasis	Who Feels Love.mp3	pop/rock	whimsical	5
102	Joanna	Yesterday.mp3	country	pleading	3
103	Booker Ervin	You Don't Know What Love is.mp3	jazz	passionate	7
104	Doves	Your Shadow Lay Across My Life.mp3	pop/rock	restless	7
105	Metropolitan Jazz Affair	Yunowhathislifeez [Motorcity Mix].mp3	jazz	whimsical	5
106	Peter Broderick	a glacier.mp3	ambient	melancholy	2
107	The Cardigans	a good horse.mp3	pop/rock	passionate	7
108	Piano Magic	a return to the sea.mp3	pop/rock	depressing	2
109	Piano Magic	bad patient.mp3	pop/rock	longing	3
110	Lifehouse	blind butt.mp3	pop/rock	plaintive	3
111	Zac Efron	it takes two.mp3	stage & screen	pleading	3
112	Beat Radio	memoir of a lightning bolt.mp3	pop/rock	agitated	7
113	Dido	never want to say it's love.mp3	pop/rock	sentimental	3
114	Peter Broderick	null.mp3	ambient	spiritual	1
115	Peter Broderick	with the notes in my ears.mp3	pop/rock	tranquil	4
116	Tizzy Bac	you'll see.mp3	pop/rock	yearning	3
117	魏如萱	一起去旅行[Let's go travelling].mp3	pop/rock	dreamy	3
118	黃建為	也無風雨也無情.mp3	country	serene	4
119	圖騰樂團	太陽.mp3	pop/rock	sentimental	3
120	巴奈	巴奈 19.mp3	country	light	5
121	黃小楨	巴奈流浪記.mp3	pop/rock	pathetic	2
122	蕭煌奇	心情像風箏.Mp3	pop/rock	plaintive	3
123	陳綺貞	月的夜.mp3	pop/rock	plaintive	3
124	張雨生	永遠都像才認識.mp3	pop/rock	cheerful	6
125	陳建年	甘苦的外籍船員.mp3	country	yearning	3
126	Tizzy Bac	田納西恰恰.mp3	pop/rock	dark	2
127	陳奕迅	多少.mp3	pop/rock	sad	2
128	929	早起去爬山.mp3	country	playful	5
129	夾子電動大樂隊	有大官.mp3	pop/rock	robust	8
130	這位太太	而我[眼已垂落].mp3	pop/rock	fanciful	5
131	這位太太	你明知道可是裝作不知道.mp3	pop/rock	plaintive	3
132	陳珊妮	你是我的酒.mp3	pop/rock	graceful	5
133	巴奈	我.mp3	country	mournful	2
134	Finn	改變自己.mp3	country	quaint	5
135	50TA(狩野英孝)	ラップ〜東京寄り道メロディ〜.mp3	pop/rock	joyous	6
136	脫拉庫	雨.mp3	pop/rock	doleful	2

137	陳建年	孩子與妳 我的天堂.mp3	blues	tender	3
138	陸家駿	秋.mp3	jazz	serene	4
139	自然捲	風和日麗(男生版).mp3	pop/rock	leisurely	4
140	929	飛翔.mp3	country	yielding	3
141	旺福	夏夕夏景.mp3	pop/rock	light	5
142	陳建年&巴奈	捆綁.mp3	country	plaintive	3
143	彭靖惠	海市蜃樓.mp3	pop/rock	depressing	2
144	圖騰樂團	起風.mp3	pop/rock	light	5
145	張懸	迷惑.mp3	pop/rock	yearning	3
146	舞曲大閱兵 7	國境之南.mp3	pop/rock	cheerful	6
147	Tizzy Bac	淺色的那條.mp3	pop/rock	emphatic	8
148	張懸	最愛你哭泣時候的眼睛.mp3	pop/rock	longing	3
149	小虎隊	喜歡.mp3	pop/rock	lyrical	4
150	愛吃飯合作社	焚風(演奏曲).mp3	country	quaint	5
151	自然捲	買.mp3	pop/rock	humorous	5
152	盧廣仲	愛情, 習作.mp3	pop/rock	sprightly	5
153	熊寶貝	煙.mp3	pop/rock	robust	8
154	熊寶貝	解藥.mp3	pop/rock	plaintive	3
155	卡奇社	遊園驚夢.mp3	pop/rock	bright	6
156	范曉萱 & 100% 樂團	電影.mp3	pop/rock	dramatic	7
157	1976	蒼白之舞.mp3	pop/rock	depressing	2
158	陳本瑜	輕輕.mp3	pop/rock	lyrical	4
159	彭郁雯的爵士 狂草	嬉水 Bathing in the Stream.mp3	jazz	graceful	5
160	張雨生	總髮遊.mp3	pop/rock	agitated	7
161	舞曲大閱兵 6	寶貝 - oa 張懸.mp3	pop/rock	happy	6

實驗數據總表

	Algorithm			Genre		Emotion		Genre & Emotion	
	特徵萃取		相似度	長時距	短時距	長時距	短時距	長時距	短時距
1	FFT	AVG	ED	82.92	84.17	37.08	55.00	60.00	69.58
2	FFT	AVG	CD	81.25	81.25	32.92	55.83	57.08	68.54
3	FFT	AVG	CO	80.00	79.58	63.75	58.75	71.88	69.17
4	FFT	AVG	CS	85.42	88.33	32.08	56.25	58.75	72.29
5	FFT	DIFF	ED	36.67	50.42	97.50	70.00	67.08	60.21
6	FFT	DIFF	CD	60.83	33.75	47.92	55.42	54.38	44.58
7	FFT	DIFF	CO	50.42	37.08	60.42	55.00	55.42	46.04
8	FFT	DIFF	CS	70.42	58.33	65.00	78.75	67.71	68.54
9	FFT	ACF	ED	77.50	83.75	52.08	41.25	64.79	62.50
10	FFT	ACF	CD	77.50	77.08	59.17	40.00	68.33	58.54
11	FFT	ACF	CO	77.50	77.08	59.17	42.92	68.33	60.00
12	FFT	ACF	CS	82.50	79.58	49.58	57.08	66.04	68.33
13	FFT	FLUX	ED	65.83	45.42	70.42	91.67	68.13	68.54

14	FFT	FLUX	CD	56.25	69.58	51.67	57.08	53.96	63.33
15	FFT	FLUX	CO	64.17	55.83	59.17	70.83	61.67	63.33
16	FFT	FLUX	CS	66.67	57.50	70.83	62.50	68.75	60.00
17	MFCC	AVG	ED	47.08	57.08	60.42	66.25	53.75	61.67
18	MFCC	AVG	CD	39.17	55.83	98.75	69.17	68.96	62.50
19	MFCC	AVG	CO	51.67	58.75	62.08	73.33	56.88	66.04
20	MFCC	AVG	CS	50.83	57.08	62.92	74.58	56.88	65.83
21	MFCC	DIFF	ED	40.00	67.50	86.25	63.33	63.13	65.42
22	MFCC	DIFF	CD	46.67	78.75	90.42	76.67	68.54	77.71
23	MFCC	DIFF	CO	57.50	81.67	72.92	73.33	65.21	77.50
24	MFCC	DIFF	CS	64.17	83.75	84.17	37.50	74.17	60.63
33	MFCC	ACF	ED	65.42	85.00	49.58	36.67	57.50	60.83
34	MFCC	ACF	CD	60.00	84.17	56.25	47.92	58.13	66.04
35	MFCC	ACF	CO	60.00	84.17	55.00	53.75	57.50	68.96
36	MFCC	ACF	CS	41.25	55.00	70.83	62.08	56.04	58.54
37	MFCC	FLUX	ED	37.50	27.08	63.75	73.75	50.63	50.42
38	MFCC	FLUX	CD	27.50	12.50	65.42	75.00	46.46	43.75
39	MFCC	FLUX	CO	30.42	49.58	61.67	66.25	46.04	57.92
40	MFCC	FLUX	CS	35.00	21.67	71.67	61.67	53.33	41.67
25	CQT	AVG	ED	73.33	73.33	47.92	42.50	60.63	57.92
26	CQT	AVG	CD	61.67	68.75	52.08	48.33	56.88	58.54
27	CQT	AVG	CO	69.17	71.67	34.58	41.67	51.88	56.67
28	CQT	AVG	CS	76.25	73.33	49.17	50.83	62.71	62.08
29	CQT	DIFF	ED	68.75	38.75	87.50	69.58	78.13	54.17
30	CQT	DIFF	CD	94.17	89.17	62.92	87.50	78.54	88.33
31	CQT	DIFF	CO	74.58	72.08	67.08	42.92	70.83	57.50
32	CQT	DIFF	CS	73.33	57.50	66.25	63.75	69.79	60.63
41	CQT	ACF	ED	67.50	57.50	65.00	58.33	66.25	57.92
42	CQT	ACF	CD	76.67	55.00	74.17	65.00	75.42	60.00
43	CQT	ACF	CO	76.67	55.00	74.17	77.08	75.42	66.04
44	CQT	ACF	CS	69.58	53.75	71.25	59.17	70.42	56.46
45	CQT	FLUX	ED	44.58	21.67	73.33	85.00	58.96	53.33
46	CQT	FLUX	CD	39.58	42.92	35.42	78.33	37.50	60.63
47	CQT	FLUX	CO	49.58	49.17	55.00	50.42	52.29	49.79
48	CQT	FLUX	CS	47.92	33.33	51.67	52.50	49.79	42.92