行政院國家科學委員會專題研究計畫 成果報告

壓縮領域中的音樂內涵分析技術
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2221-E-216-050-
執 行 期 間 ：96 年 08 月 01 日至 97 年 07 月 31 日
執 行 單 位 ：中華大學資訊工程學系

計 畫 主 持 人：劉志俊

計畫參與人員：碩士班研究生-兼任助理人員：王鴻文
碩士班研究生-兼任助理人員：江慶涵
碩士班研究生-兼任助理人員：黃宏能
碩士班研究生-兼任助理人員：蕭聖峰
碩士班研究生-兼任助理人員：蔡咏昇
碩士班研究生-兼任助理人員：曾柏嘉
碩士班研究生-兼任助理人員：張俊堂

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 97 年 12 月 02 日

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※

※　　　　　　　　　　　　　　　　　　　　　　※

※　　　壓縮領域中的音樂內涵分析技術　　　※

※　　　　　　　　　　　　　　　　　　　　　　※

※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：　NSC 96-2221-E-216 -050

執行期間：　96 年 08 月 01 日 至 97 年 07 月 31 日

計畫主持人：劉志俊

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：中華大學資訊工程學系

中　華　民　國　97　年　10　月　30　日

行政院國家科學委員會專題研究計畫成果報告

壓縮領域中的音樂內涵分析技術

Content-Based Music Analysis in the Compressing Domain

計畫編號：NSC 96-2221-E-216 -050
執行期限：96 年 08 月 01 日 至 97 年 07 月 31 日

主持人：劉志俊　中華大學資訊工程學系
計畫參與人員：王鴻文、江慶涵、黃宏能、蕭聖峰、蔡咏昇、曾柏嘉、張俊堂 中華大學資訊工程學系

一、中文摘要

隨著網路的快速發展以及多媒體壓縮技術的進步，目前有大量的多媒體資料在網際網路上快速的傳播，所以對於多媒體資料的分類與查詢，顯得日益重要。因此，有關多媒體資料內涵式分析的相關研究，越來越受到學術界的重視。其中數位音樂資料以 MP3 與 AAC 格式最受到大眾的歡迎，但是對於 MP3 與 AAC 格式的音樂內涵式分析的研究並不常見。因此，本計畫針對 MP3 與 AAC 格式的音樂，探討其內涵分析的相關問題，主要研究主題分為三個部份: AAC 電視廣告的偵測與分類、MP3 音樂的和絃分析、以及 MP3 音樂的力度分析等三部分。

關鍵詞：AAC, MPEG-7, AAC commercial segment, AAC commercial classification,

二、緣由與目的

電視廣告的視聽觀眾們隨時在改變，廣告業主難以確定在一個廣播電視節目中，究竟有多少人以及哪一種人在收看他們的廣告，所以廣告業主必須盡可能地讓廣告的內容獲得最大的效益。另一方面，普遍人們對於電視節目所提供的廣告總是難以避免地，而廣告提供給人們的訊息，卻不一定切合人們的需求，甚至基於廣告的重複性，可能對人們產生置入式行銷的行為。但是相對的來說，有一些廣告是受到人們喜愛的、有娛樂效果的或者是實用的，由於電視廣告所產生的影響，可以說是形形色色複雜而多樣，無論在民生、文化、經濟、心理或政治上。一段廣告的訊息切確地決定了人類的思考及行為模式，其影響是深遠地。

各種音訊類型相對於各類的電視廣告內涵，兩者之中是否存在著關聯，這是一個十分有趣的研究，對於廣告在許多學者的研究中，常以對於人類心理涉入性的高低來做分析，但卻缺乏對於廣告音訊內涵的分析。AAC(Advance Audio Coding)為支援多聲道的音訊格式，並在不良的傳輸環境下，對於低頻寬、高音質的需求有優良的表現，是 ISO 制定用以取代立體聲音訊格式，也是 MPEG-4 音訊核心的標準，是現行數位電視標準中的重要音訊格式之一。

當廣告漸漸由傳統的類比廣播走入科技化的數位互動後，人們對於數位化的技術有著嚴苛的需求，如傳輸量的縮減以適應低頻寬的不良環境、影音品質的重現和高效率的應用等等，然而許多廣告中卻是參差著對於使用者無用或有用的資訊，在數位電視廣播中對於這許許多多的廣告內容，針對其音訊內涵用以系統性地分析，獲取廣告相關有益於應用的內涵資訊，則是本計畫的最大初衷。

以 MP3 格式的音樂資料庫在網際網路以及硬體 MP3 播放器中變得越來越普遍。目前使用者對 MP3 音樂資料庫的查詢，一般都是以文字資料(如：歌名、歌手等)關鍵字作為查詢的依據。由於關鍵字是以人工方式建立索引，容易因為人為的疏忽而導致錯誤，並且浪費大量的資料輸入時間。因此，我們希望能透過 MP3 音樂的特徵來做為查詢的依據，而最好的特徵就是由原始音樂訊號所提供的特徵資訊。在現有許多音樂分類(Music classification)相關研究的方法，大部分都是利用語音辨識的特徵值作為分類的依據，但是卻忽略了音樂本身樂理的特性。因此，我們認為以和弦為基礎音樂內涵式分析，利用和弦這種富含音樂性的特徵可以提供更有效的分類以及查詢的結果。因此，能夠自動判斷 MP3 數位音樂每個樂音的和弦種類，變成一個十分重要的技術。

力度(dynamics)是音樂表現與影響音樂聆賞者心理情緒的重要因素之一。力度變化是音樂家在演奏時的重要表現手法，目的在使音樂展現出不同的情感。例如強的力度多數讓人感覺激昂或緊張，而弱的力度則感覺輕鬆柔和。

作曲家亦會因為力度運用上的不同而創造出各式各樣的曲風，樂曲的演變即是受了力度的強弱變化影響而成。例如巴洛克時期，樂句的強弱交替是以台階式的方式所呈現。到了古典樂派時期，力度則出現漸強或漸弱等更豐富的表示方式。

然而，在目前的音樂內涵相關研究中，一般研究的主題是探討音樂的節奏(rhythm)速度(tempo)與拍子(beat)的自動偵測，而欠缺對力度的深入分析。因此，本論文針對力度的樂理定義為起點，以各種力度在實際演奏錄音中的音強具體表現統計為依據，接著考量不同頻率與音色給人在力度感受方面的改變，來進行感知正規化。進而針對不同作曲家/演奏家的力度分析結果，我們可以建立其音樂力度側寫(dynamics profiles)，根據使用者所選擇特定的力度側寫，我們可以用來辨識一首未知 MP3 音樂中各個音符的力度，自動產生全曲之力度表情符號標記。

三、研究成果

(一) AAC 廣告音訊之內涵式分析

對於 AAC 的廣告音訊分析系統，以下將從幾個步驟加以分解並敘述：

以 DVB 為背景資料，分離電視的數位視訊及音訊，儲存音訊為 AAC 檔案格式，做為音訊取樣樣本。

在廣告音訊上，以典型的音訊特徵及多組 MPEG-7 音訊特徵值公式為基礎，經由程式計算音訊切片內相關的音效特徵值。

依據音訊斷點偵測法做自動音訊分段分析，取得音訊樣本的斷點，並依斷點位置做廣告音訊的音訊分段。

從取得的正確音訊分段，對於廣告分段內容計算其特徵路徑以進行識別及分類，進而建立廣告資料的摘要及索引或其他應用。

隨著通訊與數位壓縮技術的進步，全球的電視廣播已漸漸由類比電視廣播替換成數位電視廣播，在數位電視的廣播界面中，主要可分為衛星、有線及地面廣播三類，其中地面無線廣播藉由基地台運作所產生的邊際效益是目前最大的。目前全球數位地面廣播共分為三種標準，一是美規的 ATSC 標準，於美國、加拿大及韓國採用，另一種是日本獨自採用的 ISDB-T 標準，第三種是歐規的 DVB-T 為台灣及大部分國家所採用。我國亦在 2005 年 11 月的行政院公報修正「數位無線電視電臺技術規範」，其中明訂了數位無線電視電臺的音訊壓縮標準為 MPEG-1、MPEG-2、AC3 或 HE-AAC，其中 HE-AAC 即為改良自 MPEG-2 AAC 標準的高效率音訊壓縮格式。本文所提 AAC 音訊格式取自於部分廠商所開發的數位電視盒音訊，用做廣告音訊資料的取樣來源，自動音訊分析的主要關鍵技術如後文敘述。廣告音訊分析系統架構如圖 1 所示。

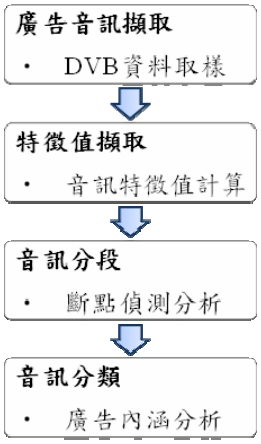

圖 1 廣告音訊分析系統架構圖

(二) MP3 數位音樂中的和弦識別

在判斷和弦之前我們必須先分析樂音聲音波形的四個結構，也就是起奏(attack)、衰退(decay)、延續(sustain)、釋放(release)，一般也稱為樂音的 ADSR 結構，由於和弦是由三個或三個以上不同樂音組合而成，因此，我們先針對和弦基本構成的樂音進行 MDCT 係數的分析。表 1 表示各種音高與其基頻頻率落在 MDCT 分析主頻帶之對照表。由表 1 我們可以看出在低音的部分(如 C3, Db3, D3)，樂音基頻頻率幾乎都落在相同的 MDCT 頻帶係數上，因此不能由單一的 MDCT 係數來區分低音的部分。因此，我們利用樂音的和諧結構特性，以及其泛音頻率來輔助辨識出正確的樂音。

表 1 各種音高與其基頻頻率落在 MDCT 分析主頻帶之對照表

	C	Db	D	Eb	E	F	F#	G	Ab	A	Bb	B
3	4	4	4	5	5	5	6	6	6	6	7	7
4	7	8	8	9	9	10	10	11	11	12	13	13
5	14	15	16	17	18	19	20	21	22	23	25	26
6	28	29	31	33	35	37	39	41	44	46	49	52
7	55	58	62	65	69	73	77	82	87	92	103	109

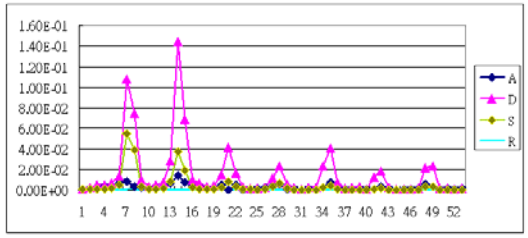

圖 2 樂音 C5 的 ADSR 結構對應 MDCT 係數能量分佈圖

從圖 2 的能量分佈我們可以看出樂音在衰退期的能量佔了絕大部分，且其和弦組成頻率比較少受雜訊影響(和弦組成頻率 SNR 比較高)。因此，我們認為在計算 MDCT 係數時只

計算衰退的部分會得到較準確的結果。由實驗數據我們可以得到在能量最高峰前的框架屬於起奏的部分，在最高峰往下算 2 倍起奏框架長度部分屬於衰退的部分。

首先，我們將要分析的單一樂音的衰退期，利用五個八度中同一家族的每一個樂音所對應到的 MDCT 頻帶係數能量相加。然後比較在 12 組樂音家族中哪一個樂音家族的能量最高，能量最高的即是該樂音所落的家族，如圖 3 所示。接著我們再進一步分析該樂音是在同一家族中的哪一個。若 x 為該樂音基頻所對應的 MDCT 頻帶係數，則我們判斷在 x/2 與 x/4 基頻的位置，是否有能量出現，我們就可以決定出該樂音屬於哪一個八度。

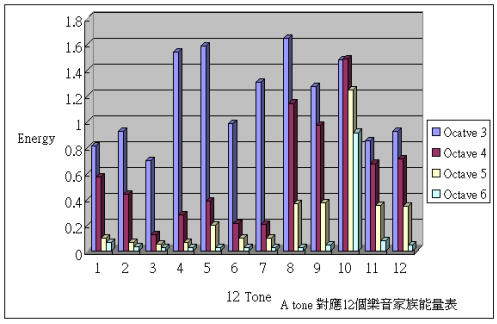

圖 3 12 組樂音家族的能量分佈統計

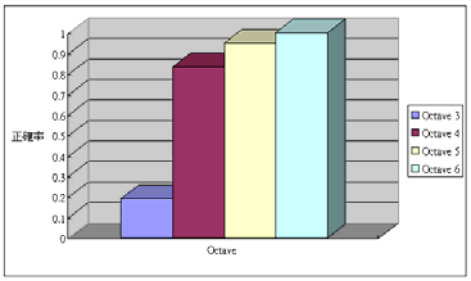

圖 4 音高與識別正確率

根據[17]我們知道 MDCT 係數具有線性的特性。因此，和弦的 MDCT 係數必定和其構成音的 MDCT 係數有線性的關係。我們可以將此線性關係寫成(1)式，其中 Xi 表示第 i 個構成音的 MDCT 係數向量、Ci 表示第 i 個構成音的加權係數。

$$MDCT_{chord} = \sum_{i=1}^{N} X_i C_i^T \qquad (1)$$

根據 MDCT 的線性關係，我們以 C5 和弦為例，統計其 MDCT 係數能量分佈，結果如圖 5 所示。我們可以看出 C5 和弦能量幾乎等於{C5+E5+G5}三個組成音的能量和。因此我們利用先前提到的單一樂音判斷方式，設定一個適當的能量門檻值，當所對應 MDCT 係數能量加總大於門檻值時，我們就認定該樂音為和弦的構成音之一，再利用和弦組成的規則就可以得到正確的和弦。

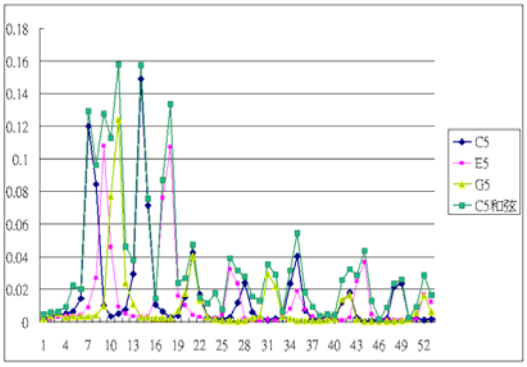

圖 5 C5 和弦及其構成音{C5,E5,G5}的 MDCT 係數能量分佈

(三) MP3 數位音樂中的力度分析

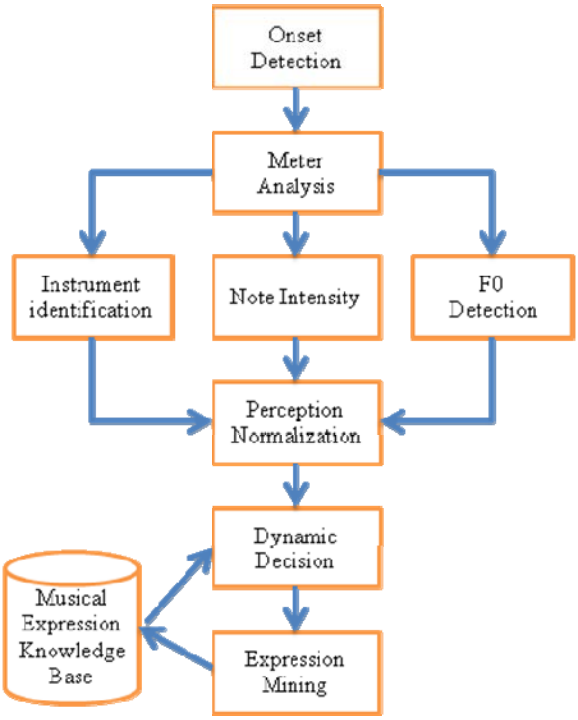

圖6 力度分析系統整體架構圖

本文所提出之力度表現自動偵測系統之整體架構如圖 6 所示。首先，我們起音點偵測(onset detection)技術將一首 MP3 音樂切分為一連串的樂音(notes)。接著我們利用以往在樂句分析方面所提出的方法[25]，將一連串的樂音群組化形成一連串的樂句。由於一般樂譜的力度符號大多是以一個小節(meter)或半個小節為單位來標記，樂句與小節的對應關係仍需進一步深入探討，本文以人工方式對樂句分析的結果選取小節的邊界。

在完成樂音與小節切分之後，我們接著計算每個樂音的音強(intensity)。由於 MP3 音樂的基本編碼單位為音框(frames)，而每秒鐘典型的 44.1kHz 取樣頻率的 MP3 音樂包括 38.28125 個音框，我們可以計算每一個樂音 n 的平均音強 $I_{Arg}(n)$

$$I_{Arg}(n) = \sum_{i=1}^{k} \sum_{j=0}^{575} MDCT^2[i,j]/k \qquad (2)$$

其中樂音 n 包含 k 個音框，MDCT[i,j]表示第 i 個音框的第 j 個頻帶之修正式離散餘弦轉換係數。

而樂音 n 的最大音強 $I_{Max}(n)$ 為

$$I_{Max}(n) = \mathrm{Max}(\sum_{j=0}^{575} \mathrm{MDCT}^2[i,j]) \quad (3)$$

其中 i =1 , 2 ,.., k，代表第 i 個音框，公式(3)亦即樂音 n 的 k 個組成音框中，音強最大的音框之音強。

由於每種樂器 ADSR(Attack, Decay, Sustain, Release)波封特性不同，樂音 n 的最大音強 $I_{Max}(n)$ 可能需要平滑化處理。所以我們採取連續 f 個音框平均音強的最大值來修正

$$I'_{Max}(n) = \mathrm{Max}(\sum_{i=0}^{f-1} \sum_{j=0}^{575} \mathrm{MDCT}^2[i+l,j]) \quad (4)$$

其中 i=1,2,…,k，表示每個樂音的力度實際感受會受到音高與樂器音色的差異而改變，所以我們需要透過樂器識別(instrument identification)以及基頻偵測(F0 detection)技術來辨識出每個樂音之音高與音色。本文所使用之基頻偵測技術請參考[25]一文。

得到每個樂音的音強、音高與音色之後，我們需根據音高與音色進行力度的感知正規化，以修正聽者的實際力度感受。感知正規化方式在下一節中說明。

由於力度記號的標示一般是以一個小節或半個小節為主，本文目前以一小節力度辨識為單位。我們對各種樂派、作曲家、演奏家的實際演奏錄音，參考其樂譜上的力度記號，統計其 6 種力度(pp、p、mp、mf、f、ff)的分割切點。以感知正規化的樂音力度參考力度分割點統計，決定每個樂音的力度強度，再彙總判斷出一個小節的 6 種力度強弱。

由於力度表現是音樂演奏表情詮釋的主要手法之一。我們未來將依作品、作曲家、演奏家、樂派等劃分，進一步分析其力度表現的關聯法則，深入解讀其力度表現慣用手法。

四、計畫成果自評

本計劃目前的研究成果至已發表期刊論文一篇[6] ，研討會論文四篇[5][7][8][9] ，另有兩篇期刊論文投稿中。

五、參考文獻

[1] 吳智偉、劉志俊, “支援 MPEG-7 之電影 AC-3 環場音效內涵描述工具,” *2005 數位生活與網際網路科技研討會*, 2005.

[2] 吳智偉、劉志俊, “AC-3 環場音效與電影劇情關聯之資料探勘模型,” *第三屆數位典藏技術研討會*, 2004.

[3] 葉億真、劉志俊, “音效資料的內涵式分類及其在電影資料庫的應用,” *第二屆數位典藏技術研討會*, 2003.

[4] 鄭煒平、劉志俊, “網際網路電影資料庫之音效自動分段索引系統,” *第六屆網際網路應用與發展研討會*, 2005.

[5] 江慶涵、劉志俊, “MP3 數位音樂中的和弦自動識別,” 二〇〇六數位生活科技研討會, 2006.

[6] 王鴻文、劉志俊, “AC-3 電影音效的內涵式自動分段,” *Journal of Information Technology and Applications*, Vol. 2, No. 1, pp. 43-53, 2007.

[7] 許鸚南、邱繼正、李一欣、江慶涵、劉志俊, “MP3 數位音樂之和弦偵測及其 3D 舞者動畫呈現,” 二〇〇七數位生活科技研討會, 2007.

[8] 蕭聖峰、劉志俊, “MPEG-2 AAC 電影音效的內涵式自動分段,” *NCS 2007 全國計算機會議*, 2007.

[9] 蔡咏昇、劉志俊, “MP3 音樂的力度自動偵測與表現分析,” *二〇〇八數位生活科技研討會*, 2008.