

音頻指紋方法研究及實作

黃金本

銘傳大學 資傳系 副教授

e-mail :

hcptw@mail.mcu.edu.tw

陳昱賢、戴斯邨、羅盛宣

銘傳大學 資傳系 學生

e-mail : max60716@yahoo.com.tw,

{e60626, frank15_1}@hotmail.com

摘要

本文在研究音訊有效的表示方法，稱為音頻指紋，並進行實作。將龐大的音樂訊號，使用非常簡短的特徵去表示，將音樂檔案的音頻指紋資料庫建立起；而查尋的音頻剪輯片段，透過他的音頻指紋，可以很快的在資料庫找到對應的音樂檔案。實作結果證實該方法非常有效，且具有下列的特色：1. 強健性；2. 成對獨立性及 3. 高資料庫搜尋效率。

關鍵詞：音頻指紋、強健性、指紋搜索

Abstract

This paper presents an effective audio signal representation called audio fingerprinting and its implementation. The method employs highly short features to represent enormous volume of the audio signals. Those features are used to establish the audio fingerprinting database for music files. During the query, a clip of audio can match with its corresponding music very fast through using a fingerprinting generated by the same method. Experimental results demonstrate effectiveness of the method and show its promising characteristics: 1. Robustness, 2. Pairwise independence, and 3. High database search efficiency.

Keywords: Audio fingerprinting, fingerprinting searching, robustness

1. 前言

音頻指紋是指可以代表一段基於內容的音訊重要聲學特徵的專有識別碼，其主要目的是建立一種有效機制，比較兩個音頻感知質量的數據。注意這裡不是直接比較很大的音頻資料，而是比較其相應較小的音頻特徵(指紋)[1-8]。音頻指紋和其相應的音訊資料例如歌曲名稱、詞曲作者、歌詞等內容一起儲存在資料庫中，並採用指紋做為相對應資料的索引。一般來說，指紋辨識的方法具有以下幾點特性：

1. 強健性（感官(perceptual)類似的不變性）：退化音頻片段應該相似於原來的音頻指

紋片段。

2. 成對獨立性(無碰撞性)：兩個不同的音頻片段，需有不同的感官結果，即必須有不同的指紋。

3. 資料庫搜尋效率：指紋的結構必須有利於在資料庫中快速的搜尋。

對於指紋而言，提取的特徵具有這些的功能可以有效的、快速的辨識信息。離散餘弦變換具有很強的能量壓縮屬性這意味著，大部分信號能量往往集中在少數幾個低頻成分。在去除相關屬性確保每個子指紋可以分別處理，而且經由創造更多的子指紋，可改善他的性能。而多重雜湊(MLH)方法是指子指紋從飛利浦強健混雜的算法(PRH 算法)的頻譜在相鄰的框架及次頻帶之間的差異所產生的。

此方法提供了一種有效的方式來總結音訊頻譜的可鑑別的信息。然而，因為它只能使用相鄰框架的信息，它也可能因為干擾而受到傷害。由於區域雜訊的干擾和音頻指紋搜索的誤導，這些在次指紋的位元可能翻轉。這個問題可以緩解包括更多的框架和執行低通濾波。藉由應用一套低通濾波器，得到的訊息會具有強健的鑑別力，訊息不會因雜訊而改變。這套過濾器被設計成具有正交性，所以輸出對應的特徵，更可以分辨彼此之間的訊號。

2. 研究方法

我們的研究方法為對目前幾個先進的音頻指紋方法[4, 6, 7, 8]，進行系統分析，了解各個方法的優缺點及進行比較，結果發現以地標為基礎的方法[8]具最佳系統性能。因此，針對此方法做詳細的研究，及進行對該方法實作，對系統方塊圖及技術進行詳細的說明。

2.1 系統分析

1. 指紋特徵具強健性。
2. 特徵匹配比對具獨立性。
3. 資料庫搜尋效率。

表 1 各方法優缺點比較

	優點	缺點
正規化頻譜子帶	匹配能力強	易受雜訊干擾
離散餘弦轉換的多重雜湊法	能在吵雜的環境下執行，計算效率高	所需記憶體量較大
二進制的音頻指紋識別系統	減少資料庫的大小和資料庫收尋時間	部份信息損失導致性能下降
地標為基礎的方法	匹配能力強，對於有雜訊的環境執行效率也高	對於片段太短的音頻較難搜尋的到

表 2 資料庫搜尋效率比較

	優點	缺點
正規化頻譜子帶	索引清楚	沒有一個有效率的搜尋方法
離散餘弦轉換的多重雜湊法	搜尋比對快	搜尋方法複雜高
二進制的音頻指紋識別系統	具有快速緊湊的資料庫搜尋特色	損失指紋信息，導致比對正確率降低
地標為基礎的方法	搜尋速度快，能顯示歌曲相關資料	運算複雜度中等

針對系統的需求比較個方法的優缺點如表 1 及表 2，這些分析給結果給予選擇使用何種方法

的考量，而以地標為基礎的方法具有非常優越的性能。

上面所提到的三種系統的需求，我們針對各種方法去比較，我們以高、中、低來表示，明顯看出地標為基礎的方法較為優勢如表 3。所以我們以這個方法 進行實作。

表 3 系統需求的比較

	資料庫搜尋效率	指紋特徵具強健性	特徵匹配比對具獨立性
正規化頻譜子帶	高	低	中
離散餘弦轉換的多重雜湊法	中	高	低
二進制的音頻指紋識別系統	高	低	低
地標為基礎的方法	中	高	高

2.2 地標為基礎的方法

每個音頻檔案藉由提取的雜湊值(hash)來表示它的指紋，“資料庫”和“樣本”的音頻檔案都以相同的方法分析。未知的樣本指紋對由音樂資料庫導出的龐大指紋進行匹配(matching)，匹配步驟為先經初步的候選匹配的篩選，接著再經詳細的正確性評估的匹配。

一些指紋特性採用的原則是 1. 指紋在時間上應該的局部化(localized)、2. 位移不變(Translation-invariant)、3. 具強健性及 4. 具足夠的資訊量。時間的局部化建議每個指紋雜湊計算使用音頻採樣點附近的時間，因此較遠的點不會影響雜湊值的計算；位移不變方面是指指紋的雜湊由對應的匹配的內涵所推導出來的，為可重製的且與它在音頻檔案內的位置無關，只要符合時間的局部化且雜湊計算的

數據包含在該檔案內，對一個未知的樣本可能來自原始音頻檔案的任何部分是可以匹配到的。

強健性是指從原始資料庫產生雜湊且可以從它的受到雜訊干擾的音頻版本重製。此外，指紋標記在未知的樣本及資料庫裡的音頻之間應該有足夠高的資訊量(熵值)，以減少虛假標記匹配的機率。若資訊量不足會造成過度及虛假的匹配產生，因而需要更多的後續處理，去剔除這些不對的結果。存在噪音和失真的環境中，若採用太多的資訊量反而會導致產生的指紋標記太脆弱而無法重製 [8]。

為了解決存在高度顯著噪音和失真的強健性問題，研究各種可能的候選特徵及進行實驗，了解那些特徵可以存活於存有噪聲的環境。顯示頻譜圖(spectrogram)的峰值點(peak)，亦稱為地標(Landmark)，在噪聲環境下，具有強健性和近似線性的特性，因此選擇頻譜圖的峰值點為特徵，利用此一特徵為基礎的方法即為以地標為基礎的方法。一個在時間-頻率平面的點，如果它具有比其所有鄰近(圍繞)點的能量都高，即為一個候選峰值點。為了保證對音頻檔案時頻區帶具有合理均勻覆蓋，選擇候選峰值點依據密度為評選的準則。而峰值點在每一時頻區帶也根據它的振幅(Amplitude)的大小去選擇及依據如前面列出的特性最高振幅峰值點在噪音存在的情況下，最可能存活下來的因素，進行適當的調整峰值點的個數。

運用此原則，一個複雜的頻譜圖，可以藉由調整時頻參數的門檻值來降低找到峰值點的數量，圖1為較密集峰值點的圖，而圖2為較稀疏峰值點的圖。在頻譜圖上所得到的峰值點圖如星空故亦稱該圖為星座圖(Constellation Map)。圖2振幅部份資料已被去除，在建立資料庫及查尋音樂片段的星座圖時，所採用的時頻參數值要一致。

利用星座圖的方法，具有即使在噪音的環境下，匹配的頻譜圖峰值點也不會受到太大的影響的特徵。對匹配音頻片段而言，星座圖上的峰值點模式也是相同的。因此，如果你放置一首資料庫歌曲的星座圖在一條狀圖上，而將幾秒鐘長度的一段音頻樣本(Samples)的星座圖放置在透明的膠片上，然後移動透明的膠片在條狀圖上，若有相當數量的點重疊則表示兩個星座圖對應到相同音訊。

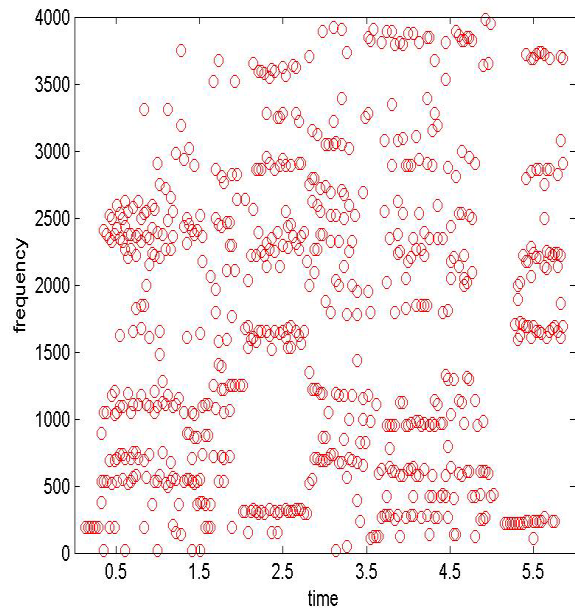


圖 1 具密集的峰值點圖

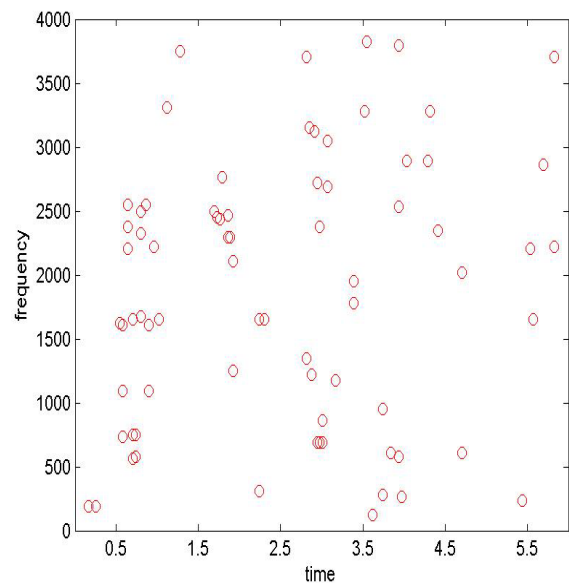


圖 2 具稀疏峰值點圖

2.3 系統架構

本系統架構是對音頻進行提取，得到雜湊值集合，然後儲存起來，再經由使用者輸入片段音頻，進行搜尋及匹配，以得到完整的音樂。本系統是希望使用者讓他的音頻片段去找尋對應的完整的音樂，而搜尋所需的時間希望越短越好。圖3為本系統所需之四大功能方塊為：指紋提取、資料庫搜尋、指紋匹配、資料庫之，以下將針對此四大功能進行解說。

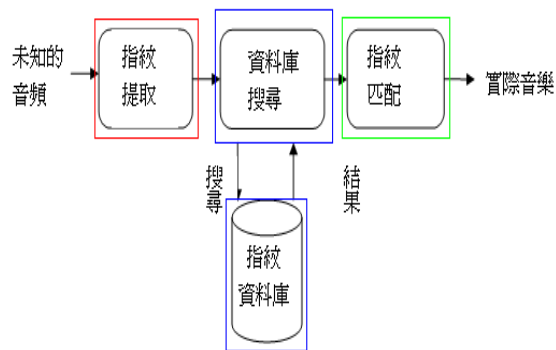


圖 3 系統基本架構

以上所描述的是我們系統基本架構，系統詳細流程如圖4，紅色框框對應的是指紋提取，藍色框框是資料庫及搜尋的部份，綠色框框是指紋匹配的內容，以下將一一介紹。

A. 指紋提取方法

一開始輸入音頻產生所謂的頻譜圖；每個音頻都有自己獨特的頻譜圖，所以利用頻譜圖（時間及頻率所組成）的特徵，以頻譜大小（Amplitude）值較大的點當作地標（Landmark），再利用地標所對應的四個值：開始頻率、結束頻率、開始時間及結束時間當作雜湊值，而這雜湊值就是所謂的指紋。

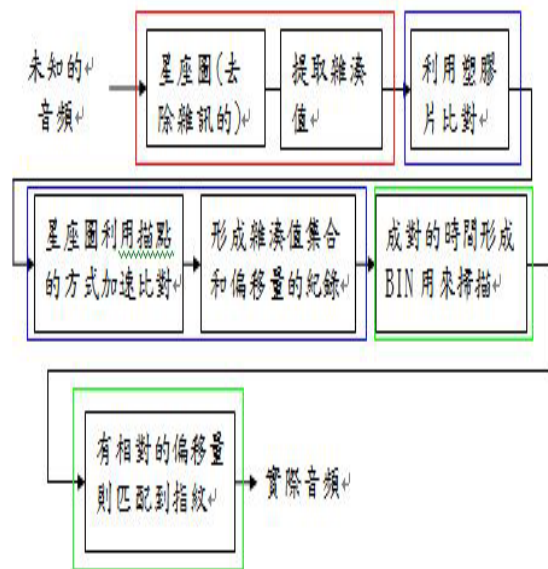


圖 4 系統詳細流程

每個參考的音訊都有數以百計的地標，當要進行搜尋比對時，就參考每個音訊的地標去搜尋相似的雜湊值，例如：音訊裡相對時間的地標、相對的頻率、頻率差和時間差。一但搜尋比對的地標擁有足夠相同音訊的地標時，那就算是有匹配成功了，通常5個匹配值以上的就算是成功了，因為音訊的地標相同是很不容

易的。

B. 資料庫搜尋方法

由於我們重點著重於驗證，所以並沒有特地去建立龐大的資料庫，而是利用 SQAM 測試的音訊檔，當資料庫來進行搜尋，利用上面所述的時間和雜湊值，系統可以很快的搜尋出對應的音訊。

為了執行搜尋，上述指紋步驟是指提取取樣聲音資料而生成一個雜湊的集合，以及時間偏移量紀錄。時間集合內的每一段頻率帶代表一個分散的點集合在查詢樣本和資料庫的聲音資料夾中。假如資料夾匹配，匹配特徵必須有相對的偏移量，在資料夾一開始匹配的時候，一個資料夾的一序列的雜湊必須也在匹配資料夾內有相同的對應的時間序列。

C. 指紋匹配

把要匹配的音頻，讀入程式，會產生兩個值，其中分別是音頻樣本的值及音頻樣本的個數。增加音頻的第一步是利用先前所得到的兩個值去尋找地標，第二步是把地標轉換成雜湊值(20 個位元)儲存起來，這 20 個位元分別是頻率一，頻率差及時間差，而匹配也是利用這雜湊值去完成。

匹配分成四個步驟，第一步把要測試的音頻片段的地標找出，第二步利用這些地標轉換為雜湊值(20 個位元)，第三步把測試音頻的雜湊值跟資料庫裡的雜湊值集合進行匹配，即可找到相對應的音頻。匹配到的會形成分散圖而這個圖非常稀疏，由於特殊的雜湊值是由一對組合值形成的。指紋匹配圖如圖 5。

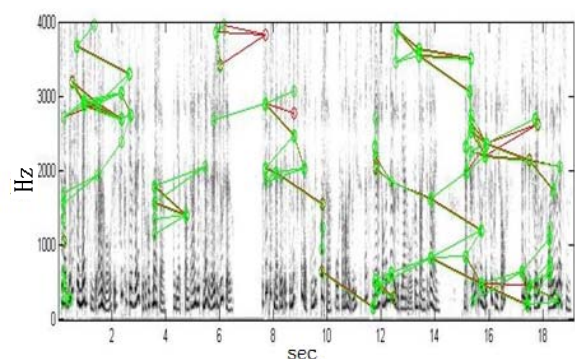


圖 5 指紋匹配圖

D. 資料庫

系統直接在MatLab裡以變數的性質建立音樂檔案當作資料庫，而此資料含有許多不同特性的音訊，例如：男高音、女低音等等…。所使用到的資料有：音訊ID、音樂名稱、WAV

檔案。

3. 實作結果

本章提供實作結果，實驗環境為 Pentium IV, 3.06-GHz 電腦，MATLAB 7.0 軟體。採用由 Sound Quality Assessment Material 網站所提供的音訊檔案，檔名如表 5，來評估專研音頻指紋方法的性能，音訊內容說明如下。為了評估提出方法對抗雜訊能力，加入高斯雜訊使其得到適當的信號雜訊比(信雜比, Signal to noise ratio, SNR)利用提出的方法進行比對。

信雜比(SNR, dB)定義為:

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right)$$

這裡 P_{signal} ， P_{noise} 個別為音頻信號及雜訊的功率。

表 5 測試音訊檔案序號

音訊檔案序號	音訊檔案名稱
1	bass47_1.wav
2	gspe35_2.wav
3	harp40_1.wav
4	horn23_2.wav
5	quar48_1.wav
6	sopr44_1.wav
7	spfe49_1.wav
8	spff51_1.wav
9	spfg53_1.wav
10	spme50_1.wav
11	spmf52_1.wav
12	spmg54_1.wav
13	trpt21_2.wav
14	vioo10_2.wav
15	frer07_1.wav
16	gspe35_1.wav

3.1 GUI 介面

圖形介面(GUI)是藉圖形友善的展示系統性能，使用者可以在交談的模式下執行各項功能。因此，利用 GUI 介面，可以不經由敘述檔或輸入指令群，只要利用滑鼠選擇參數及執行指令即可。在這種情況下，使用者不必知道程式的內容或其執行的方式，可立即獲得所要的答案。提供的圖形介面如圖 6，GUI 元件包括測試音頻的選單、試聽按鈕、開始匹配鈕、離

開紐等等。



圖 6 介面

3.2 匹配結果

輸入測試音頻後，經由匹配過程，會在視窗裡顯示所匹配到那首歌的 ID、歌曲名稱及簡介。如下所示:

ID

7

spfe49_1.wav:為一外國女性話語，語言是英文

上面所顯示出的 ID 編號是依照表 5 去做編排的，顯示出匹配到的音頻後，會播放出所匹配到的那一首音樂。之後會顯示出兩個圖，圖 7 是匹配到的音頻的頻譜圖(含有地標雜湊值)，圖 8 則是測試音頻的頻譜雜湊圖(紅色的圈代表的是地標，綠色的圈代表的是匹配到的地標)。

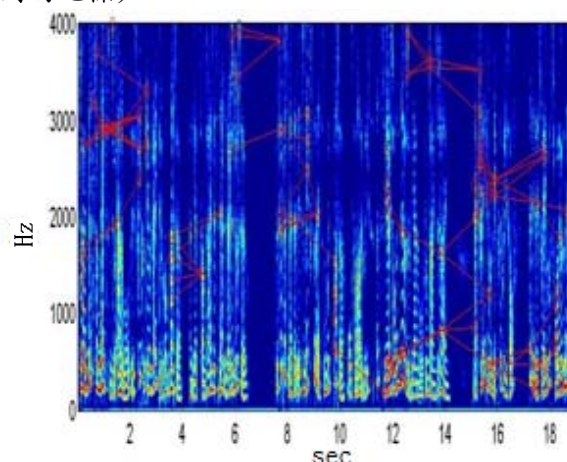


圖 7 頻譜圖

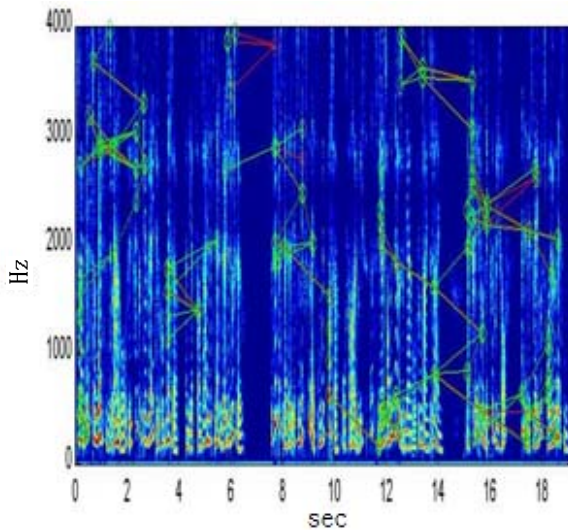


圖 8 匹配的頻譜圖

接下來是音頻加雜訊的匹配結果，以 spfe49_1.wav 這首歌為例，用 Matlab 內建的高斯雜訊加在測試音頻裡，圖 9 是加了高斯雜訊使信雜比為 20dB 時，所匹配出來的圖，圖 10 是加了高斯雜訊使信雜比為 13dB 時，所匹配出來的圖，圖 11 是加了高斯雜訊使信雜比為 0dB 時，所匹配出來的圖。實作結果證實這個方法在信雜比大於 0dB，可以正確的找出對應的音頻。

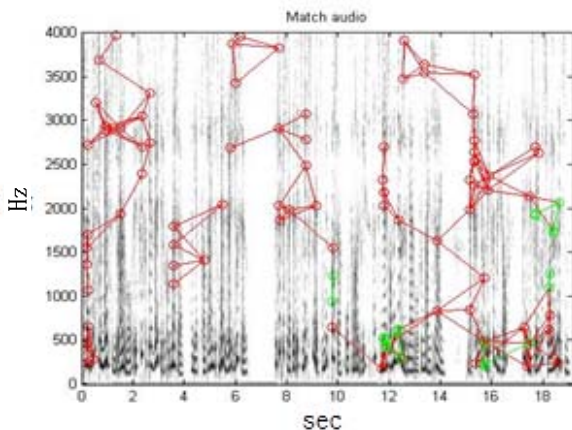


圖 9 SNR 為 20dB 時匹配的頻譜圖

4. 結論

本文研究成果，針對音訊去找音頻的指紋，此方法可以提供未來在音訊搜尋的發展，這樣的研究可以提供使用者更方便、快速的找尋音樂。使用頻譜圖可以清楚瞭解音頻檔案的頻率和時間關係，進而仔細觀察音訊匹配的相同點與相異點，可以更了解音訊的內涵。目前建置的不同的音訊檔的資料庫，對實作的之方法進行驗證，可以獲得快速搜尋及正確的匹配相對應的音訊檔。而且有雜訊存在時，當其信雜比

大於 0 dB 以上，可以獲得正確匹配的結果。

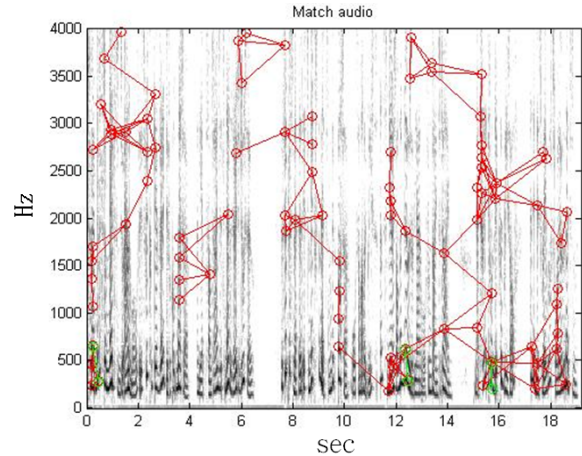


圖 10 SNR 為 13dB 時匹配的頻譜圖

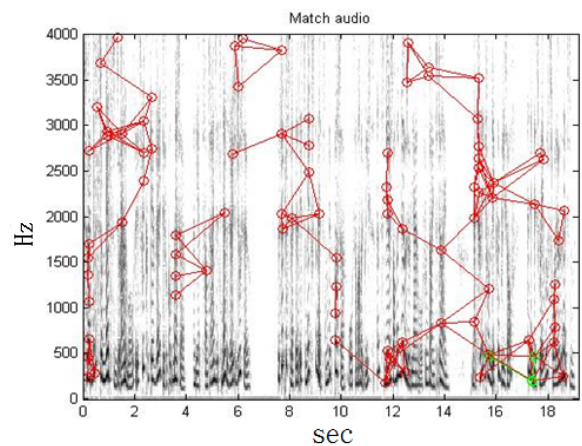


圖 11 SNR 為 0dB 時匹配的頻譜圖

參考文獻

- [1] Allamanche E., Herre J., Hellmuth O., Bernhard Fröbach B. and Cremer M., "AudioID: Towards Content-Based Identification of Audio Material," in *100th AES Convention, Amsterdam, The Netherlands*, May 2001.
- [2] Fragoulis D., Rousopoulos G., Panagopoulos T., Alexiou C. and Papaodysseus C., "On the Automated Recognition of Seriously Distorted Musical Recordings," *IEEE Transactions on Signal Processing*, Vol.49(4), pp.898-908, April 2001.
- [3] Herre J., Allamanche E., Hellmuth O., "Robust Machine Of Audio Signals Using Spectral Flatness Features," 2010/1/3
- [4] Kim S., Yoo C. D., "Boosted Binary Audio Fingerprint Based on Spectral Subband Moments," in *Proc. ICASSP*, pp. 241-244, Hawaii, 2007.

- [5] Logan B., "Mel Frequency Cepstral Coefficients for Music Modeling," in *Proc. of the International Symposium on Music Information Retrieval (ISMIR) 2000*, Plymouth, USA, October 2000.
- [6] Liu Y., Yun H. S., and Kim N. S., "Audio Fingerprinting Based on Multiple Hashing in DCT Domain," *IEEE Signal Processing Letters*, Vol. 16, no. 6, June 2009.
- [7] Seo J. S., Jin M., Lee S., Jang D., Lee S., and Yoo C. D., "Audio Fingerprinting Based on Normalized Spectral Subband Moments," *IEEE Signal Processing Letters*, Vol. 13, no. 4, April, 2006.
- [8] Wang A. L.-C., "An Industrial-Strength Audio Search Algorithm,"
URLs: <http://www.shazam.com>, <http://ismir2003.ismir.net/presentations/Wang.PDF>