

結合頻譜質心法與梅爾頻率濾波器之音樂聲紋辨識

A Combination of Spectral Centroid And MFCC Feature For Audio Fingerprinting

謝維哲

國立台北科技大學

電腦與通訊研究所

e-mail: t5418030@ntut.edu.tw

蔡偉和

國立台北科技大學

電腦與通訊研究所

e-mail: whtsai@ntut.edu.tw

摘要

隨著網際網路的快速發展，音樂的傳播日益無遠弗屆。當一首歌曲經過轉錄、傳輸以及下載之後，常已與原聲 CD 之訊號差異甚大。如何利用某音樂片段找出資料庫中相同出處的音樂文件乃是一項實用但深具挑戰的問題。本論文利用統計參數模型化方式來分析比對訊號之異同，並提出一種特徵參數，稱為「梅爾頻譜質心倒頻譜係數」(Mel Spectral Centroid Cepstral Coefficients, MSCCCs)以提升系統強健性。經由各種不同「訊號失真」版之參考樣本與測試樣本的交互比對實驗，我們發現 MSCCC 能較傳統聲學特徵，例如「梅爾頻率倒頻譜係數」(Mel Frequency Cepstral Coefficients, MFCCs)表現為佳。

關鍵詞：音樂聲紋辨識，統計參數模型，梅爾頻譜質心倒頻譜係數，梅爾頻率倒頻譜係數。

Abstract

Supported by the rapid progress in computer and network technology, popular music is rapidly becoming one of the most prevalent data types carried by the Internet. As with digitization, a single song can be presented by numerous different-format and different-quality audio data, out of playing/recording, encoding/decoding, and transmission. As a result, how to find a song based on a piece of “distorted” music has become a challenge research problem, named audio fingerprinting. In this study, we propose a robust audio feature, called Mel Spectral Centroid Cepstral Coefficients (MSCCCs), in conjunction with Gaussian mixture modeling technique to deal with this problem. After validating with various distorted music, such as sampling rate change, compression, and noise corruption, we

show that the proposed MSCCCs outperform the conventional features based on MFCCs.

Keywords: Audio Fingerprinting, Gaussian mixture model, Mel Spectral Centroid Cepstral Coefficients, Mel Frequency Cepstral Coefficients.

1. 前言

近年來由於數位化與資料儲存技術的快速發展，使得數位音樂成為網路上傳遞最為頻繁的資料型態之一。然而，隨著音樂的來源越來越廣，想找到特定的音樂卻往往變得越加困難。想像當你開車時，廣播中傳來一段悅耳的旋律，你可能不知道它的歌名，但卻很想知道並希望找到完整的一首歌曲來反覆聆聽欣賞。但問題是你對它的曲名或歌唱者資訊一無所知，且也可能無法準確地哼唱出任何其中的片段，此時你大概會因不知如何是好而感到掃興。也許要等到下次某種情況下再聽到廣播或電視裡播放這首歌時才能知道它的曲名。然而，若此時你用錄音筆或手機將這段悅耳的旋律錄下，之後再將它上傳至電腦或某搜尋引擎網站中查詢，藉由自動辨識技術查出該音樂片段的來源及相關資訊，則可以不需為著應當如何找歌而四處詢問或大費工夫了。

然而現階段的網路搜尋引擎多只著重於文字資料或者圖片的網頁搜尋，很少提供音樂資料的查詢功能。目前常見的線上音樂服務，例如 Napster 等，皆是以分類選單或關鍵字的方式進行音樂查詢，使用者必須輸入歌曲名或歌手名作為查詢依據。這樣的操作方式對於不知任何關鍵字可供參考的使用者來說並不實用，因此發展上述之「以音樂片段找歌」的查詢模式有其必要性。本篇論文探討此種歌曲查詢模式之核心技術，稱為「音樂聲紋辨識」[1]-[3]。

音樂聲紋辨識的目標是：給定一音樂片

段，由一群原聲 CD 訊號的資料庫中找出該音樂片段的來源音檔。然而，由於欲查詢的音樂片段可能經歷了轉錄、檔案壓縮、與資料傳輸等過程，其音質與原聲 CD 訊號常有很大的差異，因此直接比對訊號波形並不可行。本論文利用統計參數化方式來分析比對訊號之異同，並提出一種特徵參數，稱為「梅爾頻譜質心倒頻譜係數」(Mel Spectral Centroid Cepstral Coefficients, MSCCCs)以提升系統強健性。另外，有別於現階段音樂聲紋辨識相關研究均是以原聲 CD 訊號作為比對的參考樣本[3]-[8]，我們進行了各種不同「訊號失真」版之參考樣本與測試樣本的交互比對實驗，例如以較低取樣率之音樂檔作為參考樣本，來測試較高取樣率之音樂片段。其中我們考慮的訊號失真種類有：1) MPEG-II Layer 3 (MP3)編碼與解碼、2) 取樣率變更、3) 加性雜訊污染、4) 時間上的伸縮(快/慢轉)、5) 傳輸通道、與 6) 回音等影響。本論文證實面對各種常見訊號失真時，MSCCC 能較傳統聲學特徵例如「梅爾頻率倒頻譜係數」(Mel Frequency Cepstral Coefficients, MFCCs)[1]表現為佳。

本論文的章節組織如下。第二章描述系統架構並介紹幾種常見聲紋辨識的特徵參數，包括梅爾頻率倒頻譜係數(MFCC)[1][4]、頻譜質心(SC)[4] 和雷尼熵值(RE)[4]，以及介紹本論文所提出之梅爾頻譜質心倒頻譜係數(MSCCC)，另外說明本論文所使用的統計模型系統——高斯混合模型(GMM)[1][4]。第三章描述音樂聲紋辨識實驗細節及結果比較。第四章總結全文。

2. 辨識系統架構

如圖 1 所示為本音樂聲紋辨識系統架構，其運作中包含兩個階段：訓練(training)與測試(testing)。訓練階段之任務在於建立每一首歌曲之聲紋，我們將聲紋表示為參數統計模型。這階段包括兩項模組，一為特徵參數擷取，將雜亂的聲音波形轉成較利於分析的向量序列；另一為統計模型的建立，本論文採用高斯混合模型(Gaussian Mixture Model, GMM)[1][4]。另一方面，測試階段之工作在於辨識使用者之音樂片段為資料庫中的第幾首歌。該階段也包括一項特徵擷取模組，將聲音波形轉成向量序列後，計算該序列相對於各模型的似然率(likelihood)，然後依最大似然率法則[1][4]決定所屬歌曲。其中各模組說明如下。

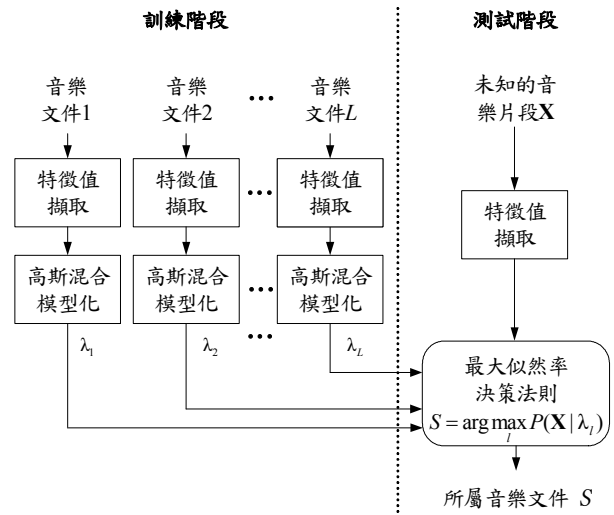


圖 1 音樂聲紋辨識系統架構圖

2.1 特徵參數擷取

本論文採用以音框為基礎(frame-based)之特徵參數。我們測試幾種已知的聲學特徵參數[1][4]，並提出一種更為有效的特徵參數。首先，將音樂訊號經由短時間(short-term)快速傅利葉轉換(Fast Fourier Transform, FFT)成為一串音框頻譜序列。令 $X_{t,j}$ 為第 t 個音框訊號在第 j 個 FFT 頻率刻度上的絕對振幅頻譜(magnitude spectrum)，其中 $1 \leq j \leq J$ ，而 J 為頻率刻度的總數。

2.1.1 梅爾頻率倒頻譜係數(Mel Frequency Cepstral Coefficients, MFCCs)

將各音框的絕對振幅頻譜送入一個「三角頻帶組」(triangular filter banks)，其中頻帶中心即是梅爾刻度值，該刻度值 mel 與頻率 f Hz 的轉換方式為 $\text{mel}(f) = 2595 \cdot \log_{10}(1 + f/700)$ ，而頻帶的寬度為兩相鄰梅爾刻度差。接著計算各頻帶的能量值 $SG_{t,b} = \sum_{j=l_b}^{u_b} |X_{t,j}|^2$ ，其中 l_b 為第 b 個頻帶中最低的 FFT 頻率刻度， u_b 為第 b 個頻帶中最高的 FFT 頻率刻度，然後再將所有頻帶的對數能量值進行離散餘旋轉換(Discrete Cosine Transform, DCT)[1]，以取得倒頻譜係數(cepstral coefficients)[1]，詳細計算方式如下：

$$\mathbf{X}_t = \frac{1}{B} \sum_{b=1}^B \log \left(\sum_{j=l_b}^{u_b} |X_{t,j}|^2 T_b(j) \right) \cos \left(\frac{\pi}{B} (b-0.5) \right) \quad (1)$$

其中 B 為頻帶總數， $T_b(j)$ 為第 b 個頻帶的三角濾波器。

2.1.2 頻譜質心(Spectral Centroid, SC)

頻譜質心[4]是經由短時間傅利葉轉換後之振幅頻譜的質量中心，也可定義為頻譜的明亮度(brightness)，可作為頻譜型態的估測指標，計算方式如下

$$\mathbf{X}_t = \frac{\sum_{j=l_b}^{u_b} j |X_{t,j}|^2}{\sum_{j=l_b}^{u_b} |X_{t,j}|} \quad (2)$$

另外有許多特徵參數是從頻譜質心延伸出來的，例如頻譜頻寬(Spectral Bandwidth, SB)[4]、頻譜平坦度(Spectral Flatness, SF)[4]、頻譜冠頂因子(Spectral Crest Factor, SCF)[4]等，但在相關論文[4]中指出，這些方法對於音樂聲紋的辨識率提升並無實質的幫助，反而增加其計算的複雜度，故在本論文中不進行討論。

2.1.3 雷尼熵值(Renyi Entropy, RE)

訊號的雷尼熵值[4]是一種量測頻帶中能量分佈的指標，藉此了解頻帶中能量分佈集中與分散的程度，其計算方式如下

$$\mathbf{X}_t = \frac{1}{1-r} \log_2 \left(\sum_{j=l_b}^{u_b} |X_{t,j}|^r \right) \quad (3)$$

其中 r 為可調係數。

2.1.4 梅爾頻譜質心倒頻譜係數(Mel Spectral Centroid Cepstral Coefficients, MSCCCs)

本論文實驗中發現以上三種特徵參數各具其優缺點，因此我們嘗試進行整合以提升辨識效能。其中較可行的整合方式是將 SC 引入 Mel-scale 頻帶分析，我們稱之為梅爾頻譜質心倒頻譜係數。其計算方式如下

$$\mathbf{X}_t = \frac{1}{B} \sum_{b=1}^B \left(\frac{\sum_{j=l_b}^{u_b} (j |X_{t,j}|^2 T_b(j))}{\sum_{j=l_b}^{u_b} |X_{t,j}|^2} \right) \cos \left(\frac{\pi i}{B} (b-0.5) \right) \quad (4)$$

2.2 高斯混合模型(GMM)

高斯混合模型[1][4]是聲學訊號分類中最常見且最成功的模型之一。本論文利用這項技術來建立音樂的聲紋模型。一個高斯混合模型包含若干高斯機率密度函式，每一高斯機率密度函式以期望值 μ_k 與變異量 Σ_k 所描述，其中 k 代表 K 個高斯機率密度函式中之第 k 個，另外含一加權數 w_k 將各高斯機率密度函式加總成為一機率密度。我們將模型參數記為： $\lambda = \{w_k,$

$\mu_k, \Sigma_k | 1 \leq k \leq K\}$ 。這些參數可經由最大化期望值法(Expectation-Maximization, EM)[1][4]估算出，此即訓練階段所須執行之工作。若資料庫中有 L 首歌，則我們產生 L 個模型 $\lambda_1, \lambda_2, \dots, \lambda_L$ 代表這些歌曲的聲紋。

在測試階段，若有一音樂片段之特徵向量序列為 $\mathbf{X} = \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ ，其中每一向量維度為 D ，則我們對各模型 $\lambda_l, 1 \leq l \leq L$ 分別計算似然率：

$$P(\mathbf{X} | \lambda_l) = \prod_{t=1}^T P(\mathbf{X}_t | \lambda_l), \quad (5)$$

其中

$$P(\mathbf{X}_t | \lambda_l) = \prod_{k=1}^K w_k \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|}} \exp \left(-\frac{1}{2} (\mathbf{X}_t - \mu_k)' \Sigma_k^{-1} (\mathbf{X}_t - \mu_k) \right) \quad (6)$$

則根據最大似然率決策法(maximum likelihood decision)[1]， $\mathbf{X}$ 應判斷為

$$S = \arg \max_l P(\mathbf{X} | \lambda_l) \quad (7)$$

3. 音樂聲紋辨識實驗

3.1 資料庫與工作目標

本論文採用五張流行音樂專輯，來當作音樂聲紋辨識的資料。其中包含一張原住民歌手合輯、一張粵語歌手合輯、一張國語歌手合輯，以及兩張包含各類型輕音樂的合輯，每張專輯內含 12~16 首歌曲不等，總共 68 首歌，以求能達到在適當大小的資料庫內，有最多元的流行音樂類型選擇。

我們利用 Windows 作業系統所提供的 Media Player 軟體將專輯光碟中以無失真的 wav 格式轉載存至電腦中，作為本次實驗標準無失真的音樂資料。再從每首歌曲中截取出 10 秒鐘做為測試資料。另外為避免從頭開始截取的內容都是前奏音樂，我們截取每首歌的第 70~80 秒來使用。

3.2 實驗條件

有別於現階段音樂聲紋辨識相關研究[3]-[8]均是以原聲 CD 訊號作為比對的參考樣本，我們進行了各種不同訊號失真版之參考樣本與測試樣本的交互比對實驗，例如以較低取樣率之音樂檔作為訓練模型的參考樣本，來測試較高取樣率之音樂片段。然而，如圖 2a 所示，由於各種類訊號失真版本的比對配對有太多的可能性，我們先考慮單一失真源之若干變化下的測試，例如只考慮不同取樣率之參考樣本與測試樣本的交互比對，如圖 2b。然後固

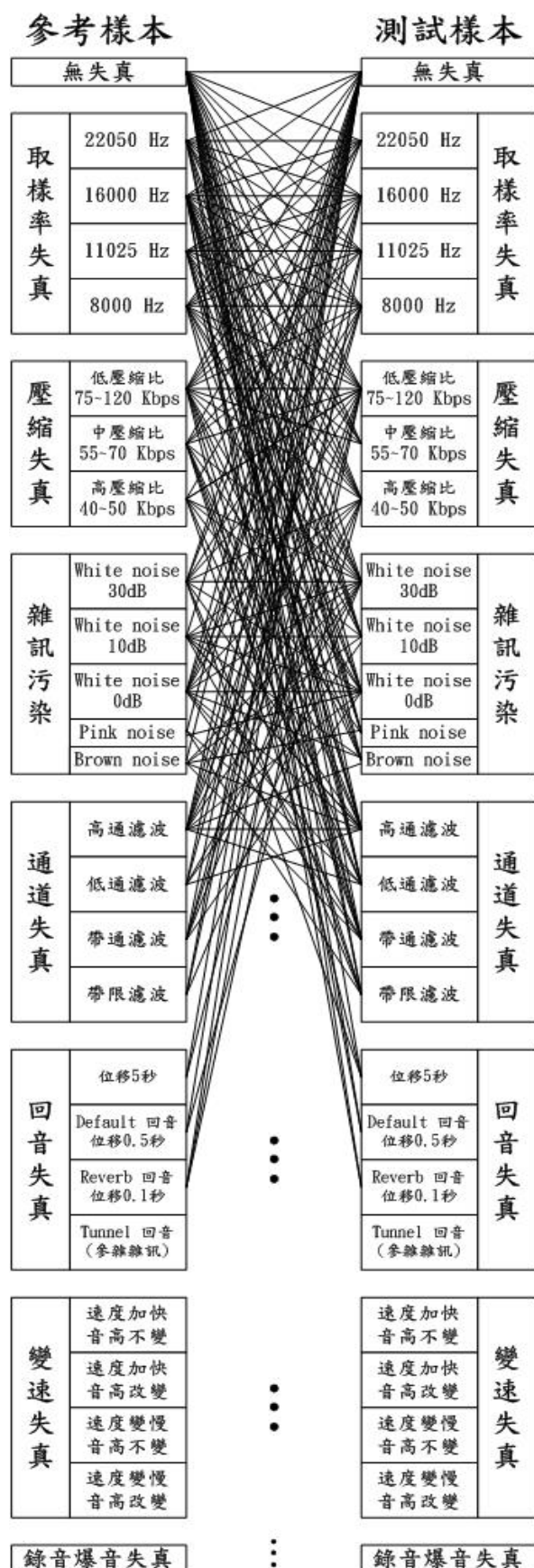


圖 2a 複雜的交互測試圖

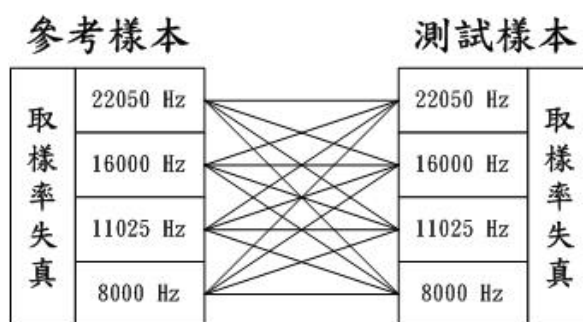


圖 2b 取樣率交互測試圖

定以某種取樣率之資料進行另一種訊號失真情況下的效能評估。另外，為簡單化音樂資料的處理，本論文中的實驗統一採用單聲道 16bit 的 wav 音樂檔格式做測試，並採用 16 個混合高斯來建立高斯混合模型。

3.3 實驗結果分析

表 1 整理出從 8kHz 到 22kHz 的各種常見取樣率配對作測試，其中測試資料的取樣率若與參考樣本的取樣率不同時，先將測試資料進行升取樣(upsampling)或降取樣(downsampling)使之與參考樣本的取樣率一致。然而可預期的是訊號經由升或降取樣之後必定引起失真。

我們可以從表 1 中發現，當測試資料之取樣率與訓練資料之取樣率相同時，辨識率均可達到完美的程度。但當測試資料之取樣率與訓練資料之取樣率不同時，辨識率則明顯地衰退，尤其當測試資料之取樣率低於訓練資料之取樣率時，辨識效能更急遽地衰退。這顯示了訊號由低取樣率升至高取樣率後並無法減低其與高取樣率之參考樣本的不匹配情況(mismatch)。另外我們發現當參考樣本為低取

表 1 取樣率分析實驗

訓練資料之取樣率(Hz)	測試資料之取樣率(Hz)	特徵值辨識率(%)		
		MFCC	SC	RE
22050	22050	100	100	100
	16000	10.3	91.2	79.4
	8000	1.5	77.9	4.4
16000	22050	98.5	100	100
	16000	100	100	100
	11025	61.8	97.1	77.9
	8000	2.94	86.8	2.94
8000	22050	97.1	100	100
	16000	97.1	100	100
	11025	97.1	100	100
	8000	100	100	100

樣率時，反而較適用於各種取樣率的測試資料，尤其當參考樣本取樣率為 8kHz 時，所有高於 8kHz 的測試資料經由降取樣後大都仍可正確辨識出。特別是使用 SC 特徵時，辨識率已達 100%。基於此觀察，我們之後的實驗均固定以 8kHz 作為參考樣本之取樣率。

以下實驗中，我們分別考慮音訊壓縮、雜訊、傳輸通道、回音、播放速度與量化飽和所造成的失真情況。表 2 是壓縮格式對各種特徵值其辨識率的影響分析，灰色背景的表格代表訓練模型的參考樣本與測試資料為一種壓縮格式下辨識率較高的特徵值。我們可以發現，MFCC 與 RE 面對不同條件下的辨識結果很類似，高壓縮比的資料與其他壓縮程度的資料比對率都不高，而 SC 在各種測試條件下都有不錯的表現，本論文提出的 MSCCC 又更進一步改善 SC。

表 3 列出幾種雜訊失真的情況，並同時測試訊雜比高低對辨識率的影響。從中可看出 Brown 雜訊對音樂資料造成失真的程度很小，即使 0dB 的訊雜比，都不影響其辨識率。而 30dB 的訊雜比，也對三種類型的雜訊影響不大。我們觀察在 10dB 訊雜比的條件下，雜訊開始影響辨識率，其中 MFCC 對三種雜訊的辨識率都較佳，但與其他三者差異並不大。

在通道失真模擬測試中，我們取用了四種情況的濾波器，如表 4。我們發現辨識率與用濾波器處理後的失真率，也就是被截掉的頻譜

表 2 壓縮格式(MP3)分析實驗

模型 資料 壓縮率	測試 資料 壓縮率	特徵值辨識率(%)			
		MF- CC	SC	RE	MSC- CC
無壓縮 (無失真 wav 檔)	無壓縮	100	100	100	100
	低壓縮比	100	100	100	100
	中壓縮比	94	98.5	94	98.5
	高壓縮比	4	85.3	3	85.3
低壓縮 (75~120 Kbps)	無壓縮	97	100	97	100
	低壓縮比	100	100	97	100
	中壓縮比	90	98.5	91	100
	高壓縮比	3	85.3	3	90
中壓縮 (55~70 Kbps)	無壓縮	94	72.1	94	96
	低壓縮比	94	72.1	94	96
	中壓縮比	100	100	100	100
	高壓縮比	10	89.7	10	99
高壓縮 (40~50 Kbps)	無壓縮	10	55.9	3	88
	低壓縮比	10	55.9	3	88
	中壓縮比	22	97.1	15	96
	高壓縮比	100	100	100	100

表 3 雜訊失真分析實驗

模型 資料	測試資料		特徵值辨識率(%)			
	訊雜 比(dB)	雜訊 種類	MF- CC	SC	RE	MSC- CC
無失真	無失真	無	100	100	100	100
	30dB	White	98.5	100	98.5	100
		Pink	100	100	100	100
		Brown	100	100	100	100
	10dB	White	42.6	22.1	29.4	32.4
		Pink	83.8	80.9	76.5	82.4
		Brown	100	100	100	100
	0dB	White	2.9	2.9	2.9	1.5
		Pink	14.7	19.1	16.2	20.6
		Brown	98.5	100	98.5	100

表 4 通道失真分析實驗

模型 資料	測試資料		特徵值辨識率(%)			
	通道 種類	失真 率(%)	MF- CC	SC	RE	MSC- CC
無失真	無失真	0	100	100	100	100
	高通濾波	8	2.9	76.5	2.9	75
	帶限濾波	30	1.5	26.5	1.5	32.4
	低通濾波	65	2.9	22.1	4.4	25
	帶通濾波	90	1.5	2.9	1.5	2.9

占整個頻譜圖的比率呈現正相關，表示特徵值的辨識率會受到經過濾波器後的失真程度影響，與濾波器的種類並無直接關連。SC 的辨識率高於 MFCC 與 RE，而本論文提出的 MSCCC 又優於前面三種特徵值。

表 5 為回音失真的實驗結果。我們可看出

表 5 回音失真分析實驗

模型 資料	測試資料		特徵值辨識率(%)			
	回音 種類	位 移 (sec)	MFCC	SC	RE	MSC- CC
無失真	無失真		100	100	100	100
	無	5	100	100	100	100
	Default	0.5	100	100	100	100
	Reverb	0.1	100	100	100	100
	Tunnel (綜合雜訊失真)		82.4	66.2	79.4	75

單純的回音失真對四種特徵參數所獲得的結果並無差異。但若回音失真的同時又參雜了雜訊失真等其他類型的失真時，則由表中最後一列中可看出 MFCC 對於雜訊失真的強健性略勝 SC 和 RE，而本論文提出的 MSCCC，其辨識率則是介於 SC 與 MFCC 之間。

表 6 顯示播放速度變化所造成之失真分析結果。我們可發現各種特徵參數在此失真情況下並不會有辨識率衰退的情況。

最後，我們考慮實際的使用情況，將資料庫的 68 個歌曲片段，經過電腦連接高品質的揚聲器播放，並以 PDA 重新錄音成 8kHz 的 wav 檔，藉此討論錄音時可能伴隨的各樣失真情況。本實驗共錄製四次，每次的音量大小不同，對錄音檔所造成的爆音(因量化飽和)程度也不同。表 7 中爆音率的定義，是將錄音檔中所有爆音的取樣點數，除上全部的取樣點數。另一方面，我們嘗試將含有爆音的音框去除，不輸入模型的似然率計算。

實驗結果顯示，本論文中提出的 MSCCC 其總體表現優於其他三種特徵值。但將含有爆音的音框去除的方法，對改善辨識率並無幫助。

表 6 變速失真分析實驗

模型資料	測試資料	特徵值辨識率(%)			
	速度	MFCC	SC	RE	MSC-CC
無失真	無失真	100	100	100	100
	加快一倍	100	100	100	100
	減慢為 80%	100	100	100	100

表 7 錄音檔爆音失真分析

模型資料	測試資料		特徵值辨識率(%)			
	取樣點爆音率(%)	處理	MFCC	SC	RE	MSC-CC
無失真	無失真		100	100	100	100
	0	無	33.8	42.6	41.2	54.4
	0.0875		70.6	58.8	60.3	67.6
	0.8845		52.9	51.5	58.8	66.2
	3.3110		33.8	42.6	44.1	55.9
	0.0875	去爆音	69.1	57.4	60.3	66.2
	0.8845		47.1	51.5	52.9	64.7
	3.3110		38.2	33.8	30.9	36.8

3.4 特徵值的辨識結果比較

在 3.3 節當中，我們已針對各種特徵值做了許多失真的測試與分析。為方便從單一圖表中，直接觀察出各種特徵值的優劣及其整體的強健性，本節取出若干最具代表性的失真情況進行交互測試。而此種交互測試的分析有助於了解若系統建立時並沒有無失真的音樂資料來訓練模型時，則利用已失真的音樂資料來建立聲紋模型之效能如何。表 8 列出了幾個代表性的失真測試，測試資料皆為標準的 8kHz、16bit、單聲道 wav 檔。

將表 8 中包含無失真共 17 種的音樂資料，分別取四種特徵值，各做一對一的交互測試，結果如表 9 到表 12，其中縱軸代表建立模型的樣本資料種類，而橫軸代表進行測試的資料種類。

我們計算出表 9 到表 12 每一列的平均值後，將其整理於表 13，比較 17 種失真情況下，平均辨識率最高的特徵值用灰色背景表示出來，可以很明顯的看出本論文提出的新方法 MSCCC，其整體的強健性高於其他三者。另外，我們可以看出 White 雜訊與帶限濾波(失真率 30%)所造成的失真影響辨識率最鉅，而目前欲克服此種問題的方式只能儘量將測試資

表 8 代表性的失真情況

No.	失真種類	失真程度
0	無	無
1	MP3 壓縮	壓縮比(75~120 kbps)
2	MP3 壓縮	壓縮比(55~70 Kbps)
3	MP3 壓縮	壓縮比(40~50 Kbps)
4	White 雜訊	30dB SNR
5	White 雜訊	10dB SNR
6	通道失真	高通濾波(失真率 8%)
7	通道失真	帶限濾波(失真率 30%)
8	Reverb 回音	回音位移 0.5 秒
9	Tunnel 回音	綜合雜訊失真
10	變速失真	速度加倍, 音高不變
11	錄音檔(有爆音)	爆音率 0.0875% (爆音點數/所有點數)
12	錄音檔(有爆音)	爆音率 0.8845% (爆音點數/所有點數)
13	錄音檔(有爆音)	爆音率 3.3110% (爆音點數/所有點數)
14	錄音檔(無爆音)	無爆音, 但相對 SNR 較高
15	取樣率失真	16000 Hz→8000 Hz
16	取樣率失真	11050 Hz→8000 Hz

料的失真情況納入參考樣本中，使模型得以匹配測試資料。

4. 結論

本論文提出了一個新的特徵值參數萃取方法 MSCCC，經過了幾種常見的失真測試，其中包含了實際的錄音失真測試，證實此方法比起之前常用的其他方法(如 MFCC、RE、SC)，在音樂聲紋辨識上有更佳的效能，在最後的失真音樂資料交互比對實驗當中，也驗證了 MSCCC 的強健性。

5. 致謝

這項研究之部分經費來自國科會研究計畫編號 NSC96-2221-E-027-097-MY3。

參考文獻

- [1] Huang, X., Acero, A., and Hon, H. W., *Spoken Language Processing*, Prentice Hall, 2001.
- [2] Cano, P., Batle, E., Kalker, T., and Haitsma, J., "A Review of Algorithms for Audio Fingerprinting," *IEEE Workshop on Multimedia Signal Processing*, pp. 169-173, 2002.
- [3] Haitsma, J. and Kalker, T., "A Highly Robust Audio Fingerprinting System," *Proc. 3rd Int. Symp. Music Information Retrieval*, pp. 144-148, 2002.
- [4] Ramalingam, A. and Krishnan, S., "Gaussian Mixture Modeling of Short-Time Fourier Transform Features for Audio Fingerprinting," *IEEE Transactions on Information Forensics and Security*, 2006.
- [5] Sukittanon, S. and Atlas, L., "Modulation Frequency Features for Audio Fingerprinting," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1773-1776, 2002.
- [6] Lu, C. S., "Audio Fingerprinting Based on Analyzing Time-Frequency Localization of Signals," *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 174-177, 2002.
- [7] Herre, J., Hellmuth, O., and Cremer, M., "Scalable Robust Audio Fingerprinting Using MPEG-7 Content Description," *IEEE Workshop on Multimedia Signal Processing*, pp. 165-168, 2002.
- [8] Burges, C.J.C., Platt, J.C., and Jana, S., "Distortion Discriminant Analysis for Audio Fingerprinting," *IEEE on Speech and Audio Processing*, Vol. 11, No. 3, pp. 165-174, 2003.

表 9 MFCC 失真測試實驗

		測試資料之失真總類編號																	平均辨
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	識率(%)
參考樣本之失真種類編號	0	100	100	94	4	99	43	3	1	100	82	100	71	53	44	34	100	100	66
	1	97	100	90	3	97	43	3	3	99	79	99	69	56	46	32	97	97	65
	2	94	94	100	10	93	29	1	0	94	68	94	68	53	57	26	94	94	63
	3	10	10	22	100	4	3	1	1	10	6	10	4	4	4	4	10	10	13
	4	99	99	88	1	100	44	1	3	99	82	99	65	51	43	34	99	99	65
	5	56	54	34	3	63	100	3	1	56	41	54	10	7	4	12	56	54	36
	6	3	3	3	1	4	3	100	3	3	4	3	3	3	3	3	3	3	9
	7	3	3	1	1	3	1	1	100	4	6	4	4	3	3	4	3	3	9
	8	100	99	94	10	99	47	3	3	100	94	100	72	53	44	29	100	100	67
	9	91	90	72	4	90	41	6	3	97	100	93	53	34	26	21	91	90	59
	10	100	100	94	6	99	43	3	1	100	84	100	66	51	47	26	100	100	66
	11	62	60	60	3	56	7	3	3	57	29	54	100	100	84	91	62	63	53
	12	53	51	47	3	50	10	1	1	53	28	53	100	100	100	90	53	54	50
	13	41	40	53	1	37	7	3	1	32	21	37	84	99	100	31	41	43	39
	14	29	29	32	1	31	6	1	1	29	15	29	94	85	40	100	29	28	34
	15	100	100	94	1	99	44	3	1	100	82	100	72	53	50	32	100	100	67
16	97	97	91	4	94	41	3	1	96	78	97	68	51	44	26	97	100	64	

表 10 SC 失真測試實驗

		測試資料之失真總類編號																		平均辨
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	識率(%)	
參考樣本之失真種類編號	0	100	100	99	85	100	22	76	26	100	66	100	59	59	51	43	100	100	76	
	1	100	100	99	85	100	21	75	26	100	74	100	59	57	47	43	100	100	76	
	2	72	72	100	90	68	1	34	10	62	29	69	66	63	56	59	72	72	59	
	3	56	56	97	100	54	3	29	6	53	25	56	65	69	62	60	56	56	53	
	4	100	100	97	81	100	24	76	26	100	71	100	50	51	44	37	100	100	74	
	5	22	22	12	4	25	100	29	3	24	28	19	7	7	4	4	22	22	21	
	6	72	72	54	44	72	22	100	7	72	54	69	43	37	31	25	72	72	54	
	7	10	10	6	4	9	3	3	100	10	6	10	3	3	3	3	10	10	12	
	8	100	100	94	75	100	18	75	28	100	82	100	57	57	47	38	100	100	75	
	9	76	76	47	38	76	25	51	22	87	100	71	29	32	28	24	76	76	55	
	10	100	100	93	81	100	21	75	25	100	72	100	56	51	43	37	100	100	74	
	11	46	46	76	78	37	3	28	1	37	13	37	100	100	100	99	46	46	52	
	12	38	38	69	75	35	4	22	4	31	13	34	100	100	100	91	38	38	49	
	13	21	21	43	53	18	1	10	3	19	6	15	93	100	100	63	21	21	36	
	14	40	40	71	71	37	1	25	3	34	18	41	100	97	93	100	40	40	50	
	15	100	100	100	85	100	21	76	28	100	69	100	59	56	51	41	100	100	76	
	16	100	100	99	84	100	25	76	31	100	68	100	57	53	51	41	100	100	76	

表 11 RE 失真測試實驗

		測試資料之失真總類編號																	平均辨
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	識率(%)
參考樣本之失真種類編號	0	100	100	94	3	99	29	3	1	100	79	100	60	59	44	41	100	100	65
	1	97	97	91	3	96	28	3	3	97	76	97	57	56	34	37	97	97	63
	2	94	94	100	10	90	18	3	1	94	54	94	68	65	46	53	94	94	63
	3	3	3	15	100	1	1	1	1	3	3	3	1	1	1	1	3	3	9
	4	97	97	85	3	100	32	1	1	97	76	97	56	53	29	44	97	97	63
	5	44	44	19	1	53	100	0	1	40	31	44	3	1	1	7	44	44	28
	6	4	4	4	1	4	3	100	1	4	4	4	6	4	4	4	4	4	10
	7	4	4	4	3	4	1	4	100	4	3	4	4	4	4	4	4	4	10
	8	100	100	93	6	99	34	3	3	100	88	100	65	62	37	46	100	100	67
	9	91	91	62	6	91	28	3	3	93	100	88	32	35	13	25	91	91	56
	10	100	100	93	3	99	31	3	1	100	78	100	57	56	37	37	100	100	64
	11	66	63	63	4	53	7	1	1	57	24	56	100	100	93	99	66	66	54
	12	56	56	59	3	49	10	1	1	50	26	54	100	100	100	88	56	56	51
	13	26	26	41	3	18	4	3	3	26	16	22	84	99	100	51	26	26	34
	14	44	44	50	4	47	9	1	1	41	25	40	99	93	59	100	44	44	44
	15	100	100	94	3	99	28	3	1	100	81	100	60	60	37	43	100	100	65
16	100	100	94	3	99	24	3	1	100	79	100	63	62	35	41	100	100	65	

表 12 MSCCC 失真測試實驗

		測試資料之失真總類編號																		平均辨
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	識率(%)	
參考樣本之失真種類編號	0	100	100	100	90	100	32	75	32	100	75	100	68	66	56	54	100	100	79	
	1	100	100	100	93	100	28	78	32	100	78	100	66	66	59	57	100	100	80	
	2	96	96	100	99	96	16	63	25	96	62	96	81	78	65	74	96	96	78	
	3	88	88	96	100	88	6	51	16	84	47	87	82	76	79	74	88	88	73	
	4	100	100	100	90	100	32	76	34	100	76	100	63	65	49	51	100	100	79	
	5	37	37	26	15	40	100	35	3	40	35	35	12	10	7	7	37	37	30	
	6	72	72	68	49	74	34	100	7	68	50	71	43	40	26	32	72	72	56	
	7	7	7	7	6	7	3	4	100	6	7	7	3	3	3	3	7	7	11	
	8	100	100	100	88	100	35	76	37	100	84	100	71	79	62	60	100	100	82	
	9	76	78	60	46	82	40	59	18	81	100	79	46	40	34	37	76	76	60	
	10	100	100	100	88	100	32	74	31	100	75	100	65	68	57	49	100	100	79	
	11	71	69	82	82	72	1	51	7	69	32	69	100	100	100	97	71	69	67	
	12	71	71	81	84	68	4	53	6	66	40	69	100	100	100	100	71	69	68	
	13	43	43	50	57	43	4	31	4	46	26	41	96	100	100	85	43	41	50	
	14	66	66	79	79	65	1	50	4	60	35	68	100	100	96	100	66	65	65	
	15	100	100	100	87	100	28	76	34	100	75	100	68	65	53	51	100	100	79	
	16	99	99	99	88	99	26	76	34	99	75	99	68	63	53	51	99	100	78	

表 13 平均辨識率比較

		特徵值參數			
		MFCC	SC	RE	MSCCC
		平均辨識率(%)	平均辨識率(%)	平均辨識率(%)	平均辨識率(%)
參考樣本之失真種類編號	0	66	76	65	79
	1	65	76	63	80
	2	63	59	63	78
	3	13	53	9	73
	4	65	74	63	79
	5	36	21	28	30
	6	9	54	10	56
	7	9	12	10	11
	8	67	75	67	82
	9	59	55	56	60
	10	66	74	64	79
	11	53	52	54	67
	12	50	49	51	68
	13	39	36	34	50
	14	34	50	44	65
	15	67	76	65	79
	16	64	76	65	78