於流行音樂中尋找相似唱腔歌手之研究

謝顯孟
國立屏東科技大學
資訊管理系
碩士生
M9756014@mail.npust.edu.tw

吳亞翰
國立屏東科技大學
資訊管理系
碩士生
M9956039@mail.npust.edu.tw

劉寧漢*
國立屏東科技大學
資訊管理系
副教授
gregliu@mail.npust.edu.tw

摘要

雖然音樂檢索已提供許多便利的查詢方式，當記得與查詢對象具相似歌聲的歌手名稱，而忘記查詢對象之人名或歌曲名稱時，則無法獲得想聽的音樂。在目前相關的研究主題中，尚未有提供相似歌聲進行音樂查詢的方式。有鑑於此，本研究提出尋找相似唱腔歌手的研究，藉由歌聲的音色探勘出相似唱腔的歌手，利用符合人耳聽覺敏銳頻率的特徵：經由離散傅立葉轉換後能觀察唱腔變化和共振峰之頻譜統計值，象徵歌手唱腔之特色，藉由歐基里德距離計算歌手唱腔相似度的比較結果，將能提供利用相似歌聲特色之其他歌手人名進行音樂之查詢功能，或進行以相似歌聲為基礎之音樂分類。經本研究實驗證明，所採用之特徵擷取方法，確實有效辨識出歌手唱腔相似的歌曲。

關鍵詞：離散傅立葉轉換、共振峰、歐基里德距離、歌聲相似度

Abstract

Although the existed music retrieval systems provide varies and conveniences query methods for users, there is no system that can retrieve music when the user only remembers the vocal timbre. However, if the user knows the other singer's name that has the similar timbre, the music retrieval system should provide the query way by the similar singer's name. Therefore, this research presents a way to find similar singers based on vocal timbre i.e. by sound of singing to find the other forgot singer with the similar sound. In this research, we adopt two kinds of features to evaluate the difference of the singer, which meet the sensitivity of human hearing. The extracted features include statistics of the spectrums which come from discrete Fourier transform and formants. Moreover, Euclidean distance is used to evaluate the difference of feature of vocal sound. According to the result of experiment, our method can be the basis of the novel retrieval method and music classification by vocal sound.

Keywords：Discrete Fourier Transform, Formant, Euclidean Distance, Vocal Similarity

1. 前言

一般在聽音樂的過程中，較少會注意到正在聽的音樂的歌手名稱或歌曲名稱，但會對音樂的歌聲或旋律有印象，直到想要再次聽喜好的某一首歌曲時，卻忘了歌曲的相關資訊，僅記得歌聲和旋律時，使用者可以使用記憶中的旋律，進行哼唱式查詢，如果無法哼唱出一小片段之旋律或使用者本身之音準不具備哼唱查詢之條件，甚至連旋律都忘記的時候，則無法以哼唱進行查詢，最後只剩記憶中的歌聲聯想到與想聽的歌手有相似唱腔的歌手名稱，卻發現沒有這方面的查詢方式(例如：關鍵字查詢，輸入歌手名稱"林育群"查詢"惠妮休斯頓"之歌曲，卻僅列出林育群之個人所屬歌曲，而無法查詢到惠妮休斯頓之歌曲)，以致於無法找到自己喜好的音樂。而且現行音樂分類或音樂分群的研究之中，沒有針對歌手相似歌聲進行相關研究，所以增加此新式的查詢機制，讓使用者更方便地找到自己需要的音樂是個重要且創新的議題。目前有許多研究學者針對歌者辨識進行研究，雖然在辨識上有相當好的準確率，但仍然有些辨識錯誤的歌手[5]，此辨識錯誤的歌曲中，許多是相似歌聲特徵的歌手所演唱，而這些誤判的歌手在我們的研究議題中，反而具有一定的意義，亦及可做為我們希望提供之另一種查詢方式結果。

歌手之歌聲是個人發聲之習慣，每位歌手都有各自之歌聲腔調，也就是演唱歌曲時的唱腔，常用來比喻歌手唱腔之形容，例如：嗓音渾厚、低沉、高亢明亮等等。針對本研究之唱腔(Vocal Sound)定義，意指歌手之歌聲於歌曲表達內含個人聲音之腔調，例如：張學友於吻別之歌曲，歌聲中有一種傷感且渾厚唱腔之韻

*通訊作者

味，而相似唱腔歌手則是與其他歌手有相似腔調之演唱風格，甚至會誤認為是已知歌手所演唱之歌曲，例如：鄭中基與張學友合作同一首左右為難之歌曲，聆聽時無法輕易地判斷出該片段歌曲之歌聲是屬於誰的歌聲，需仔細聆聽才能分辨出所屬之歌聲。

歌者辨識是從所有歌曲中找出屬於歌手本身之歌曲，如圖 1 (左)，有三位歌手(A、B、C)各五首歌曲，根據歌曲之歌聲特徵，找出相同歌聲或相似歌聲之歌手自己所屬之五首歌曲，本研究相似唱腔歌手，是從所有歌曲中找出與彼此有相似唱腔之歌手，如圖 1(右)，三位歌手(A、B、C)各五首歌曲，根據歌手唱腔之歌聲特徵，找出唱腔特徵最相似之歌曲，進而歸為自己所屬歌曲和相似唱腔之歌曲。

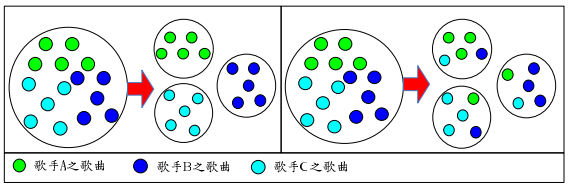

圖 1 歌者辨識(圖左)與相似唱腔歌手(圖右)

因此本研究對歌手唱腔進行分析，採用可分析唱腔的方法，例如歌手歌聲之音色(Timbre)，音色是象徵人們在對話時，可以分辨其對象之一種方式，於歌聲上也能區別歌手身份。常用於聲音之特徵擷取方法，分別有時域和頻域，時域之優點是可以不考量聲音訊號之變化，即可計算出聲音之特徵，其缺點是不適合計算過於複雜之聲音，例如含有歌聲之流行音樂，而頻域之優點是可以計算過於複雜之聲音，並找出合理之特徵，僅有計算時間較久為其缺點，因為需先將時域經過轉換為頻域，才能進行計算聲音之特徵。頻域可以分析聲音之變化，例如可以觀查其聲音之頻率高低和能量變化大小，從中獲得聲音之特徵。所以，本研究採用符合人耳聽覺感知的離散傅立葉轉換和共振峰的特徵擷取方法，擷取歌手唱腔的特徵值，以獲得相似唱腔的歌曲。

藉著人們有辨識出歌手歌聲的能力，能聽的出不同歌手唱腔的差異，甚至能記下曾經聽過的歌曲歌聲，如 2010 年參加星光大道歌唱比賽的林育群，演唱惠妮休斯頓的歌曲(I Will Always Love You)，被譽為宛如原唱，由此可見，人們會將歌手的歌聲記住，當聽到相似唱腔歌曲時，即可辨識出此歌曲是相似或是相同的歌手。由以上的敘述且以問卷調查之方式，驗證本研究採用之特徵擷取方法的數個特徵值，能有效代表唱腔之特色，能區分不同唱腔並且做為找尋相似唱腔之依據，最後藉由實驗結果證實本研究採用的特徵擷取方法是可行的。

本研究共分為 5 個章節；第 2 章為人聲與歌聲的介紹，並討論歌者辨識與音樂分類，以及比較相似度的相關研究；第 3 章介紹尋找相似唱腔歌手之方法，並實作所提出之方法；第 4 章為歌手相似唱腔的實驗分析與問卷調查的驗證；第 5 章節為結論與未來研究方向。

2. 相關文獻

在此章節，介紹人聲與歌聲的定義，能獲得兩種聲音的意義，然後針對現存歌者辨識的相關研究討論，和已被提出有效分析歌聲的特徵擷取方法，以及目前音樂分類相關的研究與相似度比對的方法。

2.1 人聲與歌聲

人聲的分析常以共振峰(Formants)評估，共振峰是一種特徵擷取方法，能獲得聲音能量最密集區域的頻率，例如：可以藉由發聲的母音(如ㄝㄧㄚㄨㄩ等等)，由共振峰的 F_1、F_2 和 F_3 獲得人聲發聲的頻率。而共振峰的排序以低頻排至高頻，F_1、F_2 和 F_3 為低頻的範圍，通常 F_1、F_2 就可具體表示發聲的母音。人的聲音經過發音訓練後，會存在著一個接近 F_3 的 speaker's Formant，其頻率範圍為 3-4 kHz 之間[1]，但其 speaker's Formant 較不明顯，在專業人士方面(如聲樂老師)，則會有比較明顯的 speaker's Formant，以男性和女性來說，女性的共振峰頻率會大於男性的共振峰頻率，因為女性的聲道較短，能發出較高的音高，而男性則能發出較大的音量，所以在日常生活中，對於人們說話的聲音也會感到聲音的相似性，是由於所發出的頻率相似，以及類似的發聲動作和習慣，藉以判斷說話人的身份是自己認識的人之中擁有相似的人聲。

共振峰也常用來分析歌手的歌聲，且專業歌手有一個特有頻率的歌手共振峰(Singer's Formant)F_3[4]，藉由利用 Perceptual Linear Prediction(PLP)獲得歌手共振峰並分析歌聲[7]。共振峰也是用來分析歌手音色的方法，通常共振峰是基頻(Fundamental Frequency)的倍數，基頻為發聲的基本週期(Fundamental Period)之倒數，象徵為聲音的音高，共振峰能用來分辨音色的好壞，共振峰(F_1、F_2 和 F_3 的頻率)之

間的倍數產生些微的差距，就是歌手獨特唱腔的特色，也是人們能辨識出歌手歌聲的特徵。

2.2 歌者辨識

歌者辨識(Singer Recognition)也可以稱為歌者識別(Singer Identification)，歌者辨識是有些歌曲並未加註辨識歌曲的資訊，則無法辨識出該歌曲的出處，或者是未整理過的音樂資料夾，透過音樂歌曲中的歌聲，以此歌聲的特徵找出與此歌聲特徵相同或相似的歌曲，就是歌者辨識。

為了克服歌者辨識的問題，就有許多研究學者提出不同的方法，Youngmoo 和 Brian[5]提出自動辨識歌手的研究，使用的特徵擷取方法為線性預估編碼(Linear Predictive Coding，簡稱 LPC)和 Warped Linear Prediction(簡稱 WLP)，並以頻率做為特徵值，比較這兩種特徵值的分類效果，LPC 的準確率高於 WLP，但僅有提升一點點。因為 Youngmoo 和 Brian 是以振幅(能量)的頻率為特徵值，較接近背景音樂能量的極點值，並且取的是 12 個特徵值(包含高頻的部分)，因此可能會取到背景音樂的特徵值，而沒有僅取具豐富歌聲的低頻(如共振峰)部分為特徵值，因為人聲或歌聲的頻率在低頻的區域較為顯著，取樣的頻率太高會令辨識的效果成效不彰。

另外，Tsai 和 Wang[9]利用梅爾倒頻譜係數 MFCC(Mel-scale Frequency Cepstral Coefficients)擷取歌聲的特徵值，並且有效地將歌手分類，但他們是利用統計分析的方式，將背景音樂排除後才進行分類的效果，而且 MFCC 的特性是針對高頻的歌聲，僅能分析能量最大的歌聲，而低頻的歌聲就容易被忽略，MFCC 不適合用來做為相似唱腔的特徵方法，因為 MFCC 僅能找出歌聲中音高最高的頻率，相對於高音歌手和中高音歌手的頻率來說，會產生相似的特徵值，無法有效表示歌聲多樣性的特徵。

2.3 音樂分類

在音樂分類的相關研究中，音樂特徵擷取方法已能有效地表示音樂的特徵，也從一般的音樂分類延伸至自動音樂分類的研究。分類方法強調於準確率上，加上許多研究學者提出有效的特徵擷取方法，藉以提升分類音樂的準確率。有一些研究提出有利於辨識歌聲或音樂的方法，像是 MFCC[6][8]和 LPC 的特徵擷取方法，MFCC 是取得聲音的能量，以及聲音頻率位於高頻的部分，LPC 是求得聲音能量全極點的特徵方法，這兩種特徵擷取方法通常是人耳不敏銳的頻率範圍，會較適合用於歌者識別。分類器也是研究學者一直在探討的主題之一，為了令分類器達到最佳的準確率，不斷嘗試有效的特徵值，因此有些分類器可以提升音樂分類的準確率，多個分類器可以輔助其它分類器的缺點，常應用於音樂分類的分類器有倒傳遞類神經網路(Back-propagation Neural Network)、k 個最近鄰居演算法(k-Nearest Neighbor Algorithm)，高斯混合模型(Gaussian Mixture Model)等等，另外也有利用決策樹(Decision Trees)和判別分析法(Discriminant Analysis)[10]等等方法，做為分類的方法。

2.4 相似度

比較相似度(Similarity)的研究中，最具代表性的就是歐基里德距離(Euclidean Distance)[2]，也是分群演算法常用來計算距離的一種方法，例如 K-means、SOM、kNN 等等。歐基里德距離常用於影像辨識、語音辨識、音樂推薦、文件推薦、醫學研究等等的領域。

3. 尋找相似唱腔歌手之研究

本研究於流行音樂歌曲中分析歌手之歌聲，利用歌聲之音色(Timbre)找出相似音色之歌手，專業歌手針對不同歌曲會有不同之音色演唱方式，意謂歌曲之演唱風格會隨著歌曲改變而有不同唱腔(Vocal Sound)，我們以不同歌手之歌曲，找出相似唱腔之歌手作為本研究之目的。本研究流程如圖 2：

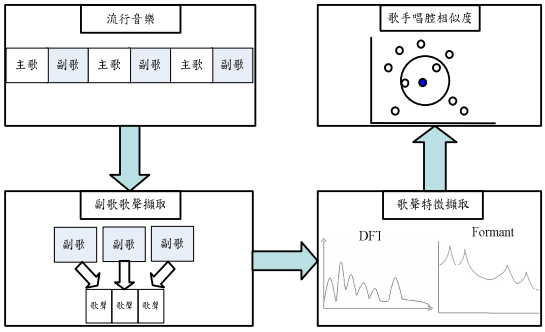

圖 2 歌手唱腔相似度之研究流程

本研究選擇流行音樂之歌曲進行分析，並從歌曲之副歌擷取出歌聲片段，再由本研究採用的兩種特徵擷取方法，分別為離散傅立葉轉

換(Discrete Fourier Transform, DFT)和共振峰(Formants)，獲得歌手唱腔之音色，再從能觀察音色變化之頻譜擷取歌聲之特徵值，再利用歐基里德距離(Euclidean Distance)計算不同歌手唱腔特徵值的距離，由計算出的距離比較不同歌手唱腔的相似度，最後獲得歌手相似唱腔之歌曲的結果。

3.1 副歌歌聲擷取

從 MP3 的音樂檔案，擷取出副歌片段中背景音樂較小，歌聲聲音較大的片段，可以避免或減少過多背景音樂的影響，能讓我們獲得較接近純歌聲的分析，再將擷取出的歌聲轉成 WAV 格式的音訊檔，會較接近原始聲音的資料，以免損失過多的資訊，並保持分析的正確性，會較符合本研究的分析，轉檔設定取樣頻率為 16 kHz、頻道為 1(單聲道)、位元率為 256 kbps、音訊大小為 16 個位元。以張惠妹"記得"歌曲為例，圖 3 (左)是整首"記得"MP3 歌曲轉換成以時間單位所畫出來的圖，圖 3 (右)為本研究擷取出"記得"歌曲所有副歌歌聲片段，經過轉換為 WAV 檔案格式的結果。

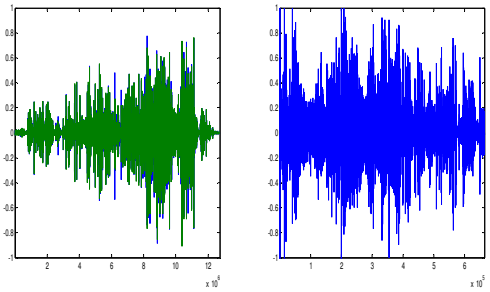

圖 3 整首歌曲(左)與副歌歌聲擷取結果(右)

3.2 歌聲特徵擷取

副歌歌聲擷取後，從歌聲中擷取出唱腔特徵，做為歌手唱腔的特徵值，然而，本研究以副歌歌聲代表整首歌曲的唱腔，需考慮到副歌歌曲的時間長度，能夠有意義地表示歌手在此歌曲的唱腔，因此本研究以離散傅立葉轉換(Discrete Fourier Transform)獲得歌手歌曲之音色，由歌曲之音色表示為歌手之唱腔，且由副歌歌曲之音色獲得歌手唱腔之特徵。接下來需考慮歌曲最具特色之唱腔特徵，以共振峰(Formants)獲得歌手最顯著唱腔之音色，再由音色中擷取歌聲之特徵，因此結合這二種特徵擷取方法，獲得歌聲之音色，做為唱腔之特徵，以下敘述這兩種特徵擷取方法。

3.2.1 離散傅立葉轉換(Discrete Fourier Transform, DFT)

DFT 是一種針對離散式時間序列進行傅立葉轉換的方法，可在有限的序列上進行轉換，將聲音訊號切成多個音框，並以音框為單位進行轉換後得到頻譜，我們以傅立葉分析，從頻譜之音色(Timbre)中分析歌手唱腔變化，且採用人耳聽覺感知範圍的聲音，表示為人耳聽見的聲音變化，做為本研究 DFT 的特徵值。本研究在計算 DFT 之前，我們利用短時矩分析歌聲，將聲音切成多個片段，再加以計算，是為了能獲得歌聲中每個聲音的變化，會先將歌聲的音訊值做前處理，步驟為預強調(設定 α=0.95)、音框化並乘上漢明窗。

經過前處理後，再進行 DFT 的轉換，最後將前處理後的音訊值輸入至 DFT 進行計算，其 DFT 公式如下：

$$F(k)=\sum_{n=0}^{N-1}x(n)e^{-j\frac{2\pi}{N}kn},\quad 0\le k\le N-1 \qquad (1)$$

其中，$F(k)$為求出複數的 DFT 係數，n 為音框索引值，N 為音框大小，$x(n)$為經過前處理的音訊值，j 為虛數符號，k 為音框索引值。

求出 DFT 的複數值後，以畢氏定理計算頻譜能量，其畢氏定理公式如下：

$$r(n)=\sqrt{a_n^2+b_n^2} \qquad (2)$$

其中，$r(n)$為 DFT 第 n 個的能量值，a 為 DFT 實數係數，b 為 DFT 的虛數係數。由於人耳聽到的聲音是接近對數的聲音變化，所以將能量再乘上 $20*\mathrm{Log}_{10}$ 轉成分貝(decibels, dB)的單位，以分貝表示人耳聽到的聲音範圍，而人耳聽音樂的範圍約為 20 分貝至 100 分貝，但人耳可以聽到聲音的最小值為 0 分貝，經過 DFT 轉換後得到一個頻譜，如圖 4 所示：

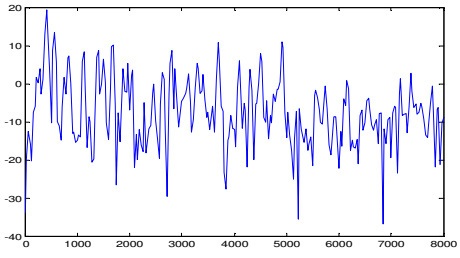

圖 4 張惠妹"記得"歌曲第 92 個音框的頻譜

經過 DFT 轉換後的雙邊頻譜有對稱性，因此我們採用頻譜的一半(如圖 4)，由 0~8000Hz

的頻率象徵歌手唱腔的特徵，本研究採用頻譜所有歌聲做為特徵值，因為每個頻譜的分貝都不相同，會有不一樣的唱腔變化，因此每個頻譜會有 512 個特徵值，取一半後為 256 個特徵值，且發現特徵值有低於人耳聽覺範圍 0 分貝以下的特徵值，每個頻譜都有這樣的特徵值，而且沒有固定的數量，所以我們將 0 分貝以上的特徵值加總，做為人耳聽到一個聲音的成份，可以區別男女歌手在於歌曲唱腔的差異，男歌手通常有較大的音量，女歌手則有較大的音高。但副歌歌聲有非常多的頻譜，我們再對所有頻譜的特徵值取平均，做為整首副歌唱腔的特徵值，但取平均會平分較顯著的特徵值，因此再計算所有頻譜特徵值的標準差，做為整首副歌唱腔的特徵值，最後每一首歌曲會有二個 DFT 的特徵值，分別為 DFT 平均和 DFT 標準差。

3.2.2 共振峰(Formants)

共振峰是聲音能量集中之區域頻率，且能計算出人耳感知敏銳範圍的低頻特徵，較不重視超出人耳感知粗糙的高頻頻率。共振峰於前處理的部分與 DFT 的前處理相同，由前處理過的音訊值做為計算之開始，經過轉換之共振峰頻譜，共振峰的每個頻譜會有多個峰值，一般將多個峰值命名為 F_1、F_2、F_3、…、F_N，共有 N 個共振峰，每個峰值代表不同之頻率與能量，可凸顯聲音之特徵，尤其是低頻之共振峰。本研究利用共振峰求出唱腔之音色(Timbre)，並由音色獲得歌手唱腔之特徵，且以歌聲頻率作為歌手唱腔之特徵值。

以最有效的方式求取歌手唱腔的共振峰，本研究採用接近歌聲能量的方法求出歌聲的包絡線，此方法為線性預測編碼(Linear Predictive Coding, LPC)，求包絡線之前需定義歌手唱腔的全極點模式，其公式如下：

$$H(z)=\frac{G}{1-\sum_{k=1}^{p}a_k z^{-k}} \tag{3}$$

其中 $H(z)$為 DFT 轉後的歌聲頻率成份，G 為常數 1，P 為預測階數，a_k 為線性預測係數。接下來，為了求出線性預測係數，以逆向濾波器(Inverse Filter)進行計算，其公式如下：

$$x(n)=\sum_{k=1}^{p}a_k x(n-k) \tag{4}$$

$x(n)$為經前處理過的音訊值，與 DFT 的前處理相同，p 為預測階數，a_k 為線性預測係數。在計算線性預測係數之前，需先求出自相關係數，自相關係數是以較少的資料點數獲得整個音框訊號之變化情形，計算方法是以音框點數向右移動 k 次，並乘上原始音框再求平均，藉以計算出 k 個係數，其公式如下：

$$r(k)=\frac{1}{N}\sum_{n=0}^{N}x(n)x(n+k),\quad k=0,1,2,\cdots,p \tag{5}$$

$r(k)$為自相關係數，$x(n)$為經前處理過的音訊值，N 為音框大小，k 為自相關係數之個數，p 為預測階數。

接下來，將自相關係數 $r(k)$輸入至 L-D 遞迴演算法(Levinson-Durbin Recursion)計算出 a_k，此計算是以遞迴的方式求出，將計算出來的 a_k 遞迴至 a_k +1 進行計算，自相關係數求出 r(0)=1，因此 $r(0)=a_0=1$，演算法如下所示：

$$E^{[0]}=r(0)$$

$$for \quad k=1 \quad to \quad P$$

$$a_0^{[i-1]}=1$$

$$m_k=\left[r(k)-\sum_{j=0}^{j-1}a_j^{[k-1]}r(k-j)\right]/E^{[k-1]}$$

$$a_k^{[k]}=m_k$$

$$for \quad j=1 \quad to \quad k-1$$

$$a_j^{[k]}=a_j^{[k-1]}-m_k a_{k-j}^{[k-1]}$$

$$end$$

$$E^{[k]}=(1-m_k^2)E^{[k-1]}$$

$$end$$

上式演算法的 p 是預測階數，m_k 為反射係數，最後計算出 a_1、a_2、a_3、...、a_k,共 $_k$ 個線性預測係數，將 a_k 帶回公式 3，需乘上 z 轉換，此轉換以公式 1 經由 DFT 轉換，再代入公式 2 求出能量，完成公式 3 的計算之後，為了表示人耳實際聽到之聲音音量大小之變化，再將求出的 $H(z)$乘上 $20*\mathrm{Log}_{10}$ 算出分貝大小，則可計算出歌手唱腔之包絡線，此包絡線可用來求出共振峰，每個共振峰之峰值都會有一個索引值(index)，我們由索引值計算出頻率，頻率計算公式如下：

$$frequency(x)=\frac{fs}{framesize}\times index \tag{6}$$

其中 *frequency(x)*為第 *x* 個峰值的頻率，*fs* 為取樣頻率，*framesize* 為音框大小，*index* 為峰值的索引值，例如 16000 Hz/512*14=437.5 Hz。圖 5 為張惠妹”記得”歌曲第 92 個頻譜的共振峰：

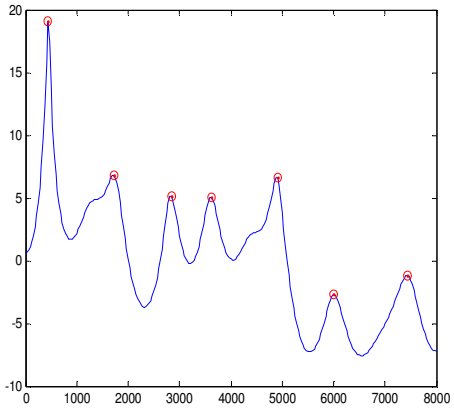

圖 5 張惠妹“記得“歌曲第 92 個音框的共振峰

圖 5 是 0~8000 Hz 的共振峰之歌聲頻譜，其橫軸為頻率，縱軸為分貝，頻譜中的紅色小圓圈為共振峰之峰值，由左邊第一個紅色小圓圈開始命名，第一個為 F_1，第二個為 F_2，…，第七個為 F_7，此頻譜共有 7 個共振峰之峰值，頻譜的頻率越小表示頻率越低，稱為低頻，頻率越大則稱為高頻，此頻譜的 F_1 為低於 1000 Hz 之頻率，是聲音能量最大之峰值，表示唱腔之頻率較集中於 F_1。表 1 為張惠妹”記得”歌曲第 92 個頻譜的共振峰(Formants)索引值(index)和頻率(frequency)。

表 1 第 92 個頻譜共振峰的索引值與頻率

Formants	index	frequency(Hz)
F1	14	437.50
F2	55	1718.75
F3	91	2843.75
F4	116	3625.00
F5	157	4906.25
F6	192	6000.00
F7	238	7437.50

本研究取前三個共振峰分別為 F_1、F_2 和 F_3 的頻率做為歌手唱腔的特徵值，F_1 和 F_2 是一般人說話聲音之頻率特徵值，然而 F_3 是歌手唱腔特有頻率之特徵值[4]，由於副歌歌聲有很多的頻譜，分別都有 F_1、F_2 和 F_3 特徵值，我們各別對所有頻譜 F_1 的特徵值取平均，但對 F_1 的特徵值取平均，會平分掉最顯著的特徵值，因此再計算所有頻譜 F_1 特徵值的標準差，F_2 和 F_3 的特徵值求取方式與 F_1 特徵值相同，每首歌曲之歌手唱腔特徵值，為 F_1、F_2 和 F_3 之平均特徵值與 F_1、F_2 和 F_3 之標準差特徵值，共有六個共振峰的特徵值。

3.3 歌手唱腔相似度

求得歌手唱腔特徵值後，為了比較不同歌手唱腔之間的距離，藉以得知不同歌手唱腔的差異，此差異為歌手唱腔之相似度，我們利用歐基里德距離來進行計算歌手唱腔與其它歌手唱腔之間的距離，歐基里德距離之公式如下：

$$d(x,y)=\sqrt{\sum_{n=1}^{S}(x_n-y_n)^2} \qquad (7)$$

其中 *d(x,y)*表示二維矩陣所求出 *x*, *y* 的距離，*S* 為特徵值的維度，x_n 是目標歌手的特徵值，y_n 是其它歌手的特徵值。求得歌手唱腔與另一位歌手唱腔距離值越小，表示歌手唱腔的相似程度越高，反之，與另一位歌手唱腔距離值越大，表示歌手唱腔的相似程度越低。

接下來以 k 個最近鄰居演算法(k-Nearest Neighbor Algorithm，簡稱 kNN)計算歌手唱腔的相似度，並獲得相似唱腔歌手之結果，kNN 能計算目標對象範圍內相似的 k 個鄰居，因此我們建置一簡易相似唱腔歌手之查詢系統，以獲得查詢相似唱腔歌手之歌曲，音樂樣本數為 56 首，選擇一些可能是相似唱腔歌手的歌曲，例如歌手自己的歌曲，以及隨機挑出的一些歌曲，男歌手 11 位，女歌手 4 位，至少都各一首歌曲，有些二首或多首歌曲，並且以 k=1 做為相似度計算，選擇的歌手是林育群，歌曲名稱為海闊天空(編號 22)，分析出與林育群相似唱腔的歌曲(僅列出最相似的五位歌手，包含查詢的歌曲)，如圖 6 所示：

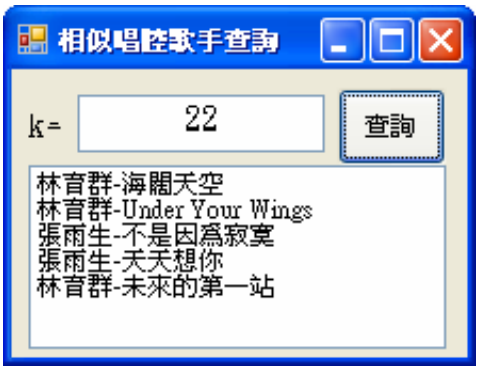

圖 6 林育群之”海闊天空”歌曲查詢結果

4. 實驗分析

本研究收集十位歌手各一首歌曲，做為唱腔分析的樣本歌曲，其中包含林育群與惠妮休斯頓的歌曲(I Will Always Love You)，是較多人認為他們的唱腔是相似的，因此放入本研究的實驗作為驗證對象之一。為了驗證本研究所採用之特徵擷取方法之可行性，以問卷調查之方式獲得相似唱腔歌手之排名。本研究以歐基里德距離計算出歌手唱腔相似距離，並以相似距離進行相似唱腔歌手之排名。由問卷調查的排名結果驗證本研究採用之特徵擷取方法之有效性，以同樣的十首歌曲，統計填問卷者對於歌曲唱腔的相似排名，作為本研究相似唱腔歌手的比較差異。十位歌手名稱與歌曲名稱如表 2，ID 為歌手之代號：

表 2 十位歌手名稱與歌曲名稱

ID	Singer	Song
A	張學友	忘記你我做不到
B	鄭中基	你的眼睛背叛你的心
C	惠妮休斯頓	I Will Always Love You
D	林育群	I Will Always Love You
E	張雨生	口是心非
F	張信哲	過火
G	張惠妹	記得
H	A-Lin	勇敢的不是我
I	吳宗憲	永保安康
J	文章	紅豆

本研究以 Matlab 7.1 撰寫程式，先將 MP3 轉換成 WAV 檔案格式，且檔案格式設定如表 3 所示：

表 3 音樂檔案格式

音訊檔	設定值
檔案格式	wav
位元率	256 kbps
音訊範例大小	16個位元
頻道	1(單聲道)
音訊取樣速率	16 kHz

表 4 為程式計算歌手唱腔之設定值，其中取樣頻率(fs/sec)取至 WAV 檔案格式之音訊取樣速率之最大值，預測階數為 LPC 之設定階數：

表 4 歌手唱腔分析之設定值

唱腔分析	設定值
取樣頻率(fs/sec)	16 kHz/sec
音框大小(framesize)	512點
音框重疊(overlap)	256點
頻道(channel)	1(單聲道)
預測階數	20階

4.1 驗證歌手唱腔相似度

本章節針對本研究所擷取出之特徵值數量，分別為八個特徵值(DFT 和共振峰)和六個特徵值(僅共振峰)之相似唱腔歌手排名，共兩種方案對問卷調查之相似唱腔歌手排名進行比較，然後由此比較結果選取差距最小之歌手進行實驗，再對兩種數量之特徵值計算相似唱腔歌手之名次差距和平均差距，最後選擇歌手翻唱相同歌曲進行實驗，以驗證歌手唱腔之相似度。

實驗一：八個特徵值與六個特徵值於問卷調查之相似唱腔歌手比較

由本研究八個特徵值之相似唱腔歌手進行排名如表 5，表格內的數字是以縱向排序相似唱腔歌手名次，最相似為第 1 名，次相似為第 2 名，以此類推，最不相似為第 9 名，例如：縱向查 A(張學友)，找對應橫向 B(鄭中基)，數字為 1，表示鄭中基的唱腔為相似的第 1 名，其他歌手的名次也是以同樣的方式獲得。

表 5 本研究八個特徵值之相似唱腔歌手排名

Singer	A	B	C	D	E	F	G	H	I	J
A	-	2	6	5	7	4	3	7	4	4
B	1	-	8	8	8	6	2	6	3	2
C	6	6	-	1	2	1	6	5	6	6
D	7	7	1	-	1	2	7	3	7	7
E	9	8	2	2	-	3	8	4	8	8
F	8	9	3	7	6	-	9	9	9	9
G	2	1	5	4	4	5	-	1	1	3
H	5	5	4	3	3	7	5	-	5	5
I	3	3	7	6	5	8	1	2	-	1
J	4	4	9	9	9	9	4	8	2	-

我們利用問卷調查的方式，獲得人們對於這十位歌手(表 2)相似唱腔排名的結果。最後所收集的問卷數為 17 份，由問卷調查統計相似唱腔歌手排名結果如表 6，表格內的數字是根據問卷者所排序的名次，並加總所有問卷者對同一位歌手的名次，以縱向的方式進行排名，總數最小的最相似排第 1 名，次相似排第 2 名，以此類推，排至總數最大為最不相似的第 9 名，例如：縱向查 A(張學友)，找對應至橫向 B(鄭中基)，數字為 1，因此鄭中基的唱腔在實際人們的聽覺是最相似的第 1 名，其他歌手的名次也是以同樣的方式獲得，如表 6 所示：

表 6 問卷調查之相似唱腔歌手排名

Singer	A	B	C	D	E	F	G	H	I	J
A	-	1	6	7	6	5	7	8	3	3
B	1	-	7	6	4	2	6	6	1	2
C	9	9	-	1	9	9	2	2	9	9
D	8	8	1	-	2	3	3	4	6	6
E	5	5	4	2	-	1	4	3	5	5
F	4	2	5	5	1	-	5	5	4	4
G	6	6	2	3	3	4	-	1	8	7
H	7	7	3	4	7	8	1	-	7	8
I	2	3	9	8	8	7	8	7	-	1
J	3	4	8	9	5	6	9	9	2	-

我們對本研究八個特徵值的相似唱腔名次與問卷分析的相似唱腔名次，取其兩者的差值，做為本研究與實際人們對於相似唱腔的差距，例如 A(張學友)和 B(鄭中基)的差值，計算方式為表 5 縱向的 A(張學友)對應橫向 B(鄭中基)的數字是 1，表 6 縱向的 A(張學友)對應橫向 B(鄭中基)數字為 1，因此表 7 縱向 A(張學友)對應橫向 B(鄭中基)，其差是 0，其餘以此推算，計算出所有的差值後(取絕對值)，再計算某歌手與其他歌手的平均差距，此平均差距為本實驗八個特徵值與問卷調查之誤差差距，平均差距越小表示誤差越小，反之，平均差距越大表示誤差越大，例如：表 7 的平均差距，縱向 D(林育群)與其他九位歌手的相似唱腔排名平均差距為 1.1，其餘以此類推，最後獲得本實驗八個特徵值於相似唱腔歌手之平均差距，如表 7：

表 7 八個特徵值的平均差距

Singer	A	B	C	D	E	F	G	H	I	J
A	-	1	0	2	1	1	4	1	1	1
B	0	-	1	2	4	4	4	0	2	0
C	3	3	-	0	7	8	4	3	3	3
D	1	1	0	-	1	1	4	1	1	1
E	4	3	2	0	-	2	4	1	3	3
F	4	7	2	2	5	-	4	4	5	5
G	4	5	3	1	1	1	-	0	7	4
H	2	2	1	1	4	1	4	-	2	3
I	1	0	2	2	3	1	7	5	-	0
J	1	0	1	0	4	3	5	1	0	-
平均差距	2.2	2.4	1.3	1.1	3.3	2.4	4.4	1.8	2.7	2.2

經由計算出(表 7)八個特徵值之平均差距，我們再計算六個特徵值之差距和平均差距，並計算出總平均差距，作為本實驗整體平均差距，計算結果如表 8。

表 8 歌手之總平均差距

特徵值數量	總平均差距
八個特徵值	2.4
六個特徵值	2.556

實驗二：歌手不同歌曲於其他歌手歌曲之實驗

由表 8 總平均差距之數據得知，八個特徵值之整體歌手平均差距小於六個特徵值之整體歌手平均差距，表示八個特徵值獲得誤差較小之相似唱腔歌手。從表 7 之平均差距得知，林育群是差距最小的，因此我們拿 D(林育群)的其它歌曲，和原來的十首歌曲(表 2)進行實驗，因此挑 D(林育群)的 9 首歌曲，分別為 1(累格)、2(I Will Always Love You)、3(Amazing Grace)、4(It's My Time)、5(My Heart Will Go On)、6(Under Your Wings)、7(一個人生活)、8(未來的第一站)和 9(海闊天空)，與另外九位歌手各一首歌曲進行實驗分析，表格內的數據為相似唱腔排名，以縱向的方式排名，最相似為第 1 名，次相似為第 2 名，以此類推，最不相似為第 17 名。

表 9 林育群的九首歌曲和其他九位歌手一首歌曲之實驗

Singer\Song	D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9
D-1	-	10	6	3	4	3	4	3	3
D-2	16	-	13	11	16	16	11	15	16
D-3	9	8	-	4	12	9	3	7	10
D-4	4	4	3	-	9	6	1	6	5
D-5	11	17	16	16	-	5	16	12	6
D-6	5	16	7	9	2	-	10	2	1
D-7	6	3	1	2	11	7	-	5	8
D-8	3	13	4	5	3	2	6	-	2
D-9	1	15	8	6	1	1	7	1	-
A	13	9	10	13	13	13	13	13	13
B	10	12	12	14	10	12	14	10	12
C	14	1	5	7	15	14	5	14	14
E	15	2	14	8	14	15	9	16	15
F	17	5	17	17	17	17	17	17	17
G	8	7	9	10	8	10	8	9	9
H	2	6	2	1	7	4	2	4	4
I	7	11	11	12	5	8	12	8	7
J	12	14	15	15	6	11	15	11	11

由表 9 的實驗來看，在前三名之中有一些是林育群(1~9)自己的歌曲，以縱向的 1(累格)來看，可以發現林育群的 2(I Will Always Love You)是第 16 名，雖然都是林育群所演唱之歌曲，此數據表示兩首歌曲之唱腔不相似。接下來，分析橫向的歌手 H(A-Lin)，她的唱腔和林育群的唱腔有七首排在前四名，這說明林育群的歌曲唱腔相似於 A-Lin 的唱腔，如以林育群找相似唱腔歌手的前四名，則會找出歌手 A-Lin。

實驗三：兩種特徵值數量於相似唱腔歌手之名次差異

由於我們希望找出符合人們聽覺認知最相似唱腔之歌手，因此再對相似唱腔歌手的名次進行比較，分別為第 1 名、第 2 名、…、至第 9 名，計算出各個名次的平均差距，以找出最符合人們聽覺認知之相似唱腔歌手。表 6 是符合人們聽覺的相似唱腔排名，因此以表 6 的名次做為比較的依據，計算出與本研究兩種特徵值數量的平均差距，其計算結果如圖 7 所示：

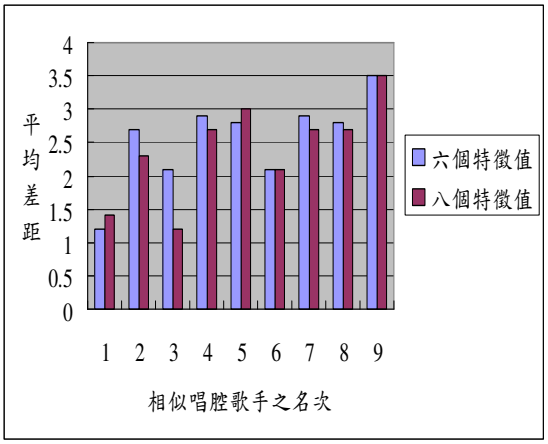

圖 7 八個特徵值和六個特徵值於問卷名次之平均差距

圖 7 中，八個特徵值在於前三名能獲得優於六個特徵值之相似唱腔歌手，兩種特徵值數量皆在第一名有低於 1.5 之平均差距。

實驗四：歌手翻唱歌曲之評估

接下來，針對歌手翻唱其他歌手之歌曲進行驗證本研究之特徵值，以本研究八個特徵值對不同歌手且相同歌曲名稱進行實驗分析，亦拿五首被翻唱歌曲之歌手組合，五首歌曲為 a(張惠妹)和 b(張宇)之(1)趁早歌曲、c(張惠妹)和 d(林俊傑)之(2)記得歌曲、e(張雨生)和 f(楊培安)之(3)大海歌曲、g(范瑋琪)和 h(康康)之(4)到不了歌曲、i(惠妮休斯頓)和 j(林育群)之(5)I Will Always Love You 歌曲，其中，惠妮休斯頓和林育群翻唱之歌曲名稱皆為 I Will Always Love You(原唱為 Dolly Parton)，其餘為前者原唱，後者為翻唱，例如：原唱 a(張惠妹)歌曲名稱為(1)趁早，翻唱者為 b(張宇)，其餘以此類推。表 10 之數字為縱向排序歌手之相似唱腔名次，由最相似排為第 1 名，次相似為第 2 名，以此類推，最不相似為第 9 名。

表 10 歌手翻唱歌曲之實驗

Singer(Song)	a (1)	b (1)	c (2)	d (2)	e (3)	f (3)	g (4)	h (4)	i (5)	j (5)
a (1)	-	6	1	2	3	4	2	2	6	5
b (1)	8	-	6	9	6	8	6	5	4	4
c (2)	2	4	-	3	4	6	3	4	5	6
d (2)	1	8	2	-	7	5	1	3	7	7
e (3)	4	1	4	5	-	3	5	6	1	3
f (3)	6	9	9	6	5	-	7	9	3	2
g (4)	3	7	3	1	8	7	-	1	8	9
h (4)	5	2	5	4	9	9	4	-	9	8
i (5)	7	3	7	8	1	2	9	8	-	1
j (5)	9	5	8	7	2	1	8	7	2	-

由表 10 得知，林育群之唱腔相似於惠妮休斯頓為第 1 名，反之，惠妮休斯頓之唱腔相似於林育群則為第 2 名，雖然惠妮休斯頓之唱腔最相似張雨生之歌曲大海，與之前實驗歌曲不同，之前歌曲名稱為口是心非(如表 5)，且問卷調查之數據顯示，林育群之相似唱腔歌手第 2 名為張雨生(如表 6)，因為是林育群模仿惠妮休斯頓之歌曲唱腔，所以林育群會較近似於惠妮休斯頓之唱腔。在於張惠妹和張宇之趁早歌曲方面，兩位歌手之唱腔並不相似，因為本研究是以歌曲之唱腔，非以歌曲之背景音樂為分析對象，因此可以明顯看出兩者唱腔之差異。

5.結論與未來研究方向

相似唱腔歌手可以藉由 DFT 與共振峰所擷取出的特徵值，表示歌手的唱腔，也能有效地辨識歌手唱腔的相似程度。本研究使用二種特徵擷取方法，能擷取出歌手豐富的唱腔特徵，主要在於歌手歌聲具有獨特的音色和音質，令歌手的唱腔可以被辨識出來，然而，當歌手之歌曲唱腔並不具特色之特徵時，則會出現與異性相似或不具相似唱腔歌手之結果。

本研究擷取之副歌歌聲，能辨識出歌手唱

腔之特徵，例如中低音歌手和高音歌手之唱腔，前者能量頻率較低，後者能量頻率較高，即可分析出兩者之差異。如果女歌手之歌聲頻率較低，則可能與男歌手有相似之唱腔，反之，男歌手之歌聲頻率較高，則會與女歌手有相似之唱腔，因此以歌手張雨生之歌曲唱腔，有一些是以高音演唱歌曲，經過分析張雨生之唱腔，會相似高音之女歌手(如惠妮休斯頓)，而較不與中低音之女歌手相似。

流行音樂之數量仍然持續增加中，如能將此研究延伸至自動音樂分類，做為音樂檢索之索引，對喜好聆聽音樂之使用者會有更多選擇，藉以獲得自己喜好歌手之相似唱腔歌曲，例如，提供相似唱腔歌手之查詢，可供選擇不同唱腔歌手之歌曲。音樂推薦方面，導入本研究方法，可推薦相似唱腔之歌曲，讓使用者聽到更多元化歌手之歌曲。

如能增加更多不同歌手和歌曲，就能提供更多相似唱腔歌手的音樂，在音樂副歌歌聲擷取部分，提出或採用有效的方法加以自動化擷取(例如，高斯混合模型(Gaussian Mixture Models, 簡稱 GMM)進行歌聲與背景音樂之分類)，可以充分獲得歌手唱腔之特徵，再加上自動分類機制，分類相似唱腔歌手之歌曲，亦可讓整個研究方法自動地找出相似唱腔歌手之歌曲。自動化之方法有利於未來的相關研究，加上利用更有效之人耳聽覺感知範圍的特徵擷取方法，如 Perceptual Log Area Ratio(PLAR) [3]，相信能運用在大型音樂資料庫上，並且提供更多相似唱腔歌手之歌曲。

誌謝

本研究承國科會專題研究計畫(編號：NSC98-2200-E-020-004 及 NSC98-2221-E-020-025-MY2)補助經費，謹此誌謝。

參考文獻

[1]. Bele, I. V., "The Speaker's Formant," ***Elsevier, Journal of Voice***, Vol. 20, No. 4, pp. 555-578, 2006

[2]. Cheng, C. C. and Hsu, C. T., "Content-Based Audio Classification with Generalized Ellipsoid Distance," ***Proceedings of the Third IEEE Pacific Rim Conference on Multimedia(PCM)***, pp. 328-335, 2002

[3]. Chow, D. and Abdulla, W. H., "Robust Speaker Identification Based on Perceptual Log Area Ratio and Gaussian Mixture Models," ***Proceedings of 8th International Conference on Spoken Language Processing(INTERSPEECH)***, pp. 1761-1764, 2004

[4]. Cook, P. R., "Identification of Control Parameters in an Articulatory Vocal Tract Model with Applications to the Synthesis of Singing," ***PhD Thesis, Stanford University, USA***, 1991

[5]. Kim, Y. E. and Whitman, B., "Singer Identification in Popular Music Recordings Using Voice Coding Features," ***Proceedings of 3rd International Conference on Music Information Retrieval***, pp. 164-169, 2002

[6]. McKinney, M. F. and Breebaart J., "Features for Audio and Music Classification," ***Proceedings of 4rd International Conference on Music Information Retrieval(ISMIR)***, pp. 151-158, 2003

[7]. Millhouse, T. J., and Clermont, F., "Perceptual Characterisation of the Singer's Formant Region: A PreLiminary Study," ***Proceedings of 11th Australiasian International Conference on Speech Science and Technology***, pp. 253-258, 2006

[8]. Pye, D., "Content-based Methods for the Management of Digital Music," ***IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP)***, Vol. 4, pp. 2437-2440, 2000

[9]. Tsai, W. H. and Wang, H. M., "Automatic Singer Recognition of Popular Music Recordings via Estimation and Modeling of Solo Vocal Signals," ***IEEE Transactions on Speech and Audio Processing***, Vol. 14, No 1, pp. 330-341, 2006

[10]. West, K. and Cox, S., "Features and Classifiers for the Automatic Classification of Musical Audio Signals," ***Proceedings of 5rd International Conference on Music Information Retrieval(ISMIR)***, pp. 1-6, 2004