



國立中山大學資訊工程學系

碩士論文

印刷樂譜辨識系統

Automatic Recognition of Printed Music Score

研究生：蔡自偉 撰

指導教授：蔣依吾 教授

中華民國 九十三年 七月

摘要

印刷樂譜辨識系統(OMR)可自動辨識光學掃描印刷樂譜，並轉換為電子音樂 MIDI 格式。將一般樂譜數位化有許多好處，例如：表演者可用電腦伴奏來做練習，或建立音樂資料庫且佔用較少記錄空間，還可以針對樂譜做編輯、修改或列印。

不像一般文件，樂譜為二維關係——五線譜上水平方向不同音樂標誌代表不同音長，垂直方向不同位置代表不同音高，而且樂譜中符號大小並不固定，形狀也為動態。由於印刷品質與排版情形不同，而且每次進行光學掃描時所設定解析度也可能有所不同，再加上人為造成樂譜傾斜或雜訊，種種因素使得許多相關研究無法很彈性處理這些問題，或只能在某些限制下進行辨識。

本論文涵蓋了兩個領域知識：一者為影像處理技術，主要使用投影法擷取水平與垂直線段以縮小辨識範圍，再利用型態學法辨識各種音樂符號，另一者為樂理知識，用來提供分析上之輔助，並修正辨識錯誤之處。

本系統分成三大階段，第一階段為前置處理，為避免人為掃描影像之傾斜，首先將影像做角度修正，以利接下來五線譜偵測與移除，移除五線譜是為了使所有音樂符號孤立出來，避免五線譜妨礙辨識。第二階段，針對垂直線段，與非垂直線段音樂符號進行辨識，並組成音樂標誌進行樂理粹鍊。第三階段，將所有音樂標誌記錄成音樂表示語言，再經程式轉換成電子音樂 MIDI 格式，可藉由播放軟體將旋律播放出來。

實驗結果顯示本系統可在短時間內成功辨識各式印刷樂譜，並且沒有解析度之硬性限制。

目 錄

摘要	I
第一章 緒論	1
1.1 前言	1
1.2 基本音樂知識	2
1.3 相關研究	10
1.4 研究概述	18
第二章 理論基礎	20
2.1 正交投影法	20
2.2 型態學法	22
2.3 樂理使用	28
第三章 研究方法及步驟	31
3.1 前置處理	31
3.1.1 傾斜修正	31
3.1.2 五線譜偵測	34
3.2 符號辨識	39
3.2.1 垂直線段音樂符號辨識	40
3.2.2 非垂直線段音樂符號辨識	44
3.2.3 辨識步驟	48
3.2.4 樂理粹鍊	58
3.3 格式轉換	60
第四章 實驗與討論	64
4.1 結果比較	64

4.2 討論.....	71
第五章 結論.....	72
參考文獻.....	73

圖目錄

圖 1.1 譜表.....	2
圖 1.2 譜號 (a) G 譜號 (b) F 譜號 (c) C 譜號.....	3
圖 1.3 終止線與小節線.....	3
圖 1.4 音符.....	4
圖 1.5 音符記法.....	4
圖 1.6 拍號.....	6
圖 1.7 變音記號.....	6
圖 1.8 調號.....	7
圖 1.9 臨時變音記號.....	7
圖 1.10 倚音.....	7
圖 1.11 倚音時值.....	8
圖 1.12 和絃.....	8
圖 1.13 轉位和絃.....	8
圖 1.14 以五線譜樣本偵測五線譜.....	10
圖 1.15 利用 y 軸投影偵測五線譜位置.....	11
圖 1.16 採用霍夫曼轉換偵測小節線與五線譜.....	11
圖 1.17 利用特徵剖面圖辨識符號.....	12
圖 1.18 利用型態學法之骨骼化運算辨識音樂符號.....	13
圖 1.19 利用型態學法辨識音樂符號所建立之各式樣本.....	14
圖 1.20 (a)全音符 (b)全音符之水平片段 (c)全音符之 LAG.....	15
圖 1.21 (a)使用 k-NN 分類法所分類音樂符號之叢集(cluster) (b)以邊界之 相互關係區別臨時記號.....	16
圖 1.22 流程圖.....	19
圖 2.1 相同影像不同尺寸投影圖 (a)大小為 178×206 影像 (b)原影像之	

x 軸投影 (c)原影像之 y 軸投影 (d)原影像放大兩倍(大小為 356×412 影像) (e)原影像放大兩倍後之 x 軸投影 (f)原影像放大兩倍後之 y 軸影投.....	21
圖 2.2 結構元素.....	23
圖 2.3 (a)平移 (b)反射.....	24
圖 2.4 (a)集合 A (b)結構元素 B (c) A 被 B 侵蝕.....	24
圖 2.5 (a)集合 A (b)結構元素 B (c) A 被 B 膨脹.....	25
圖 2.6 A 被 B 開啓.....	25
圖 2.7 A 被 B 閉合.....	26
圖 2.8 骨骼化例子.....	27
圖 2.9 水平位置屬性 — 變音記號與附點位置.....	28
圖 2.10 垂直位置屬性 — 休止符位置.....	28
圖 2.11 寬度屬性 — 小節線與全音符寬度.....	28
圖 2.12 複合符號音樂標誌.....	29
圖 2.13 變音記號規則(G 譜號譜表).....	29
圖 2.14 拍號檢驗 (a)輸入樂譜 (b)辨識結果.....	30
圖 3.1 偵測樂譜最後一組五線譜.....	32
圖 3.2 偵測傾斜角度.....	33
圖 3.3 原圖 3.1 傾斜修正後影像.....	34
圖 3.4 實線箭頭:樂譜圖騰形成投影高峰值 虛線箭頭:未測得譜線.....	35
圖 3.5 音樂符號形成投影高峰值.....	35
圖 3.6 譜表樣本.....	36
圖 3.7 測得之五線譜.....	36
圖 3.8 (a)原圖 3.1 移除譜線後影像 (b)所移除之譜線.....	38
圖 3.9 垂直線段偵測與聲部個數分析.....	40

圖 3.10 符尾及符衍分析.....	42
圖 3.11 小節線位置.....	43
圖 3.12 四分休止符經投影產生垂直線段.....	43
圖 3.13 (a) 還原記號與升記號 (b) 降記號與倚音.....	44
圖 3.14 附點位置.....	47
圖 3.15 辨識流程圖.....	49
圖 3.16 譜號辨識.....	50
圖 3.17 垂直線段擷取(標示為紅色線段)並確認音樂調號與拍號.....	51
圖 3.18 實心符頭辨識與符尾及符衍分析.....	52
圖 3.19 空心符頭辨識.....	53
圖 3.20 垂直線段音樂符號辨識後結果.....	54
圖 3.21 全休止符與二分休止符辨識.....	55
圖 3.22 八分休止符與十六分休止符辨識.....	56
圖 3.23 最後步驟 — 附點辨識.....	57
圖 3.24 圖 3.23 樂理粹鍊.....	59
圖 3.25 程式介面.....	62
圖 3.26 演奏設定.....	62
圖 3.27 播放電子音樂 MIDI 檔.....	63
圖 4.1 測試樂譜(家後).....	65
圖 4.2 PhotoScore 商業軟體辨識圖 4.1 結果.....	66
圖 4.3 本研究程式辨識圖 4.1 結果.....	67
圖 4.4 測試樂譜(親愛的你怎麼不在我身邊).....	68
圖 4.5 PhotoScore 商業軟體辨識圖 4.4 結果.....	69
圖 4.6 本研究程式辨識圖 4.4 結果.....	70
圖 4.7 影像失真而未擷取到垂直線段.....	70

表 目 錄

表 1.1 音符長度.....	4
表 1.2 附點.....	5
表 1.3 休止符.....	5
表 3.1 垂直線段音樂符號.....	41
表 3.2 符頭型態樣板.....	42
表 3.3 四分休止符型態樣板.....	43
表 3.4 降記號與倚音型態樣板.....	44
表 3.5 非垂直線段音樂符號.....	45
表 3.6 譜號型態樣板.....	46
表 3.7 全音符型態樣板.....	46
表 3.8 休止符型態樣板.....	47
表 3.9 附點型態樣板.....	48
表 3.10 拍號型態樣板.....	48
表 3.11 原圖 3.1 樂譜前九小節之音樂語言.....	61



第一章 緒 論

1.1 前言

印刷樂譜辨識系統稱爲 OMR(Optical Music Recognition)，可以自動地辨識光學掃描樂譜，並轉換成電腦可讀格式加以儲存，或轉換爲電子音樂 MIDI 格式播放。一般光學字元辨識系統 OCR(Optical Character Recognition)與 OMR 並不完全相同。不像一般文件 — 印刷字元爲獨立、固定大小，字元與字元間並不會出現相連情形，樂譜爲二維關係 — 五線譜上水平方向不同音樂標誌代表不同音長，垂直方向不同位置代表不同音高，而且樂譜中符號大小並不固定，形狀也爲動態，因此，OCR 相關技術並不一定適合用在樂譜辨識上。

將印刷樂譜轉爲電子音樂格式有許多優點，例如：表演者可用電腦伴奏來做練習，抑或，一個無任何音樂訓練的人可以隨時聆聽一段印刷樂譜旋律，此外，不但可以佔用較少記錄空間，還可以針對樂譜做編輯、修改或列印，也可以用來建立音樂資料庫。

將印刷樂譜數位化通常有三種方式：一種是利用音樂編輯軟體，依照樂譜一個一個音樂符號逐次輸入，這種方法不但費時，而且容易有人爲疏失；另一種則是吟唱或彈奏一段旋律，藉由麥克風或 MIDI 鍵盤等輸入裝置將旋律輸入，再經由軟體轉換成電腦可讀格式，此種方法速度較快，但是硬體需求較高，而且目前技術發展尙未純熟，尤其當和絃或同時多個聲音輸入時，易產生干擾造成辨識錯誤。而光學印刷樂譜辨識系統可以避免上述問題困擾，並且可以減少人工作業時間。

1.2 基本音樂知識

音樂可以依照各種形式記載，讓世界各地人們可以了解、演奏及編寫，因此，樂譜可以說是音樂文字，而國際上最廣為應用之樂譜，即為五線譜。本論文即是針對五線譜之音樂標誌進行辨識研究，因此，在論述本研究方法之前，先介紹文中會提及之基本音樂知識。

五線譜譜表由五條平行線構成，共有五線四間，用來記載各個樂音。把代表樂音之符號記錄在線上或間內，便可表示樂音高低，這些線稱為「五線」，或稱「譜表」。譜表由下方，以第一線、第一間方式稱呼，在五條線之間，可以記下十一個音(圖 1.1)。

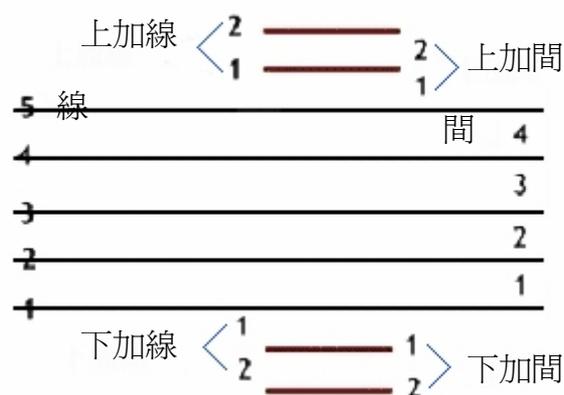


圖 1.1 譜表

一般而言，最常將第二線當做 G 音，這個時候使用 G 譜號，有了這個譜號，所有樂音之位置即確定。G 譜號是「G」字變形(圖 1.2(a))，因為使用在歌曲、小提琴、豎笛、長笛等高音樂器樂譜，所以又名為「高音譜號」或「G 譜表」。另外，F 譜號用於記載低音部樂音，將第四線當作 F 音，為「F」字變形(圖 1.2(b))，常見於男低音、大提琴、伸縮喇叭、低音提琴符低音樂器樂譜，又稱為「低音譜號」或「F 譜號」。為了與高低音部有所區別，還有中音譜號，以第三線作為 C

音，原來是供中音聲部記譜用，但現只用於中提琴、中音長喇叭等樂器的樂譜上，又稱「C 譜號」(圖 1.2(c))。

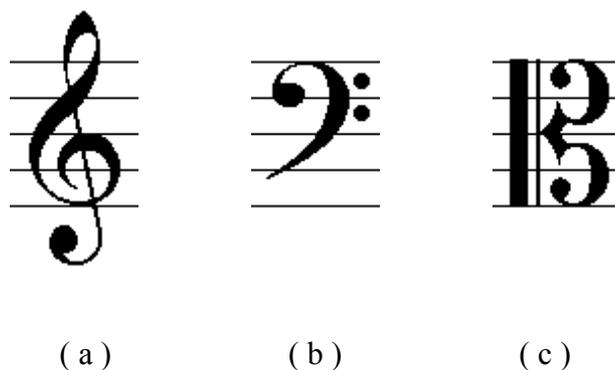


圖 1.2 譜號 (a) G 譜號 (b) F 譜號 (c) C 譜號

五線譜中，為方便節拍分明，分成一個一個音節，這些分隔小節之線，稱為小節線，此外，有時一個段落結束，抑或在一個段落內，但需要轉調時，則畫兩條小節線，但此為兩條細線，以區別終止線。終止線是在樂曲最後，將要結束之處，以兩條豎線並行來表示，其中細者在後，粗者在後，用來表示樂曲終了(圖 1.3)。

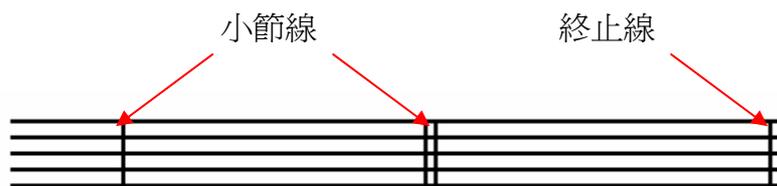


圖 1.3 終止線與小節線

譜表上，用來表示樂音高低及樂音時值之符號，稱為音符，一個音符包括有符頭、符桿與符尾。音符中白橢圓或黑橢圓，稱為「符頭」；附著於符頭上的直線，稱為「符桿」，原則上，符桿長度為三個間；而附著於符桿上之鉤，稱為「符尾」。當音符時值小於八分音符連續出現時，可以直線集結為一拍，此一連結線

稱為「符衍」(圖 1.4)。



圖 1.4 音符

符頭在五線譜第三線以上時，符桿由左邊往下畫；在第二間以下時，就從符頭右邊往上畫；若符頭恰巧記在第三線上時，則向上向下均可(圖 1.5)。



圖 1.5 音符記法

此外，依據音符長度不同，音符寫法也有所不同，如表 1.1 所示。

表 1.1 音符長度

音符名稱	寫法	時值
全音符	○	
二分音符	♪	
四分音符	♩	
八分音符	♪	
十六分音符	♫	

若在音符後面加上一點，表示增加這個音符一半長度，如表 1.1 所示。

表 1.2 附點

附點全音符	
附點二分音符	
附點四分音符	
附點八分音符	
附點十六分音符	

音樂並不一定是連續不停，在一首曲子適當處停頓或休息，可使旋律更生動、更有變化。休止符就是在樂曲中表示聲音停頓之符號。與音符一樣，休止符長度也有不同記法，如表 1.3 所示。

表 1.3 休止符

全休止符		相當於 一整小節
二分休止符		相當於 
四分休止符		相當於 
八分休止符		相當於 
十六分休止符		相當於 

各種休止符在譜表上位置為：全休止符在第四線之下，二分休止符在第三線之上，其他休止符大約在譜表中間。有時在同一譜表上，要記兩種以上旋律時，高低可以作上下移動。

音符並不能直接表示拍子個數，而只是提供一種比例關係。音符在一段音樂中代表之拍數，須借助另一個記號：拍號。拍號由上下兩個數字所組成，分母代表這段音樂中以什麼音符為「一拍」，分子代表每小節有幾拍。如圖 1.6 所示，拍號 3 與 8，代表每個小節由三個以八分音符為一拍之節奏所組成，因此，圖中八分音符等於一拍，十六分音符等於半拍。



圖 1.6 拍號

此外，變音記號也是樂譜中常見符號，變音記號種類有：升記號、降記號和還原記號。升記號表示後面跟隨樂音要升高半個音；降記號表示降記號後面跟隨樂音要降低半個音；還原記號則表示將已升高或降低樂音恢復原來音高(圖 1.7)。變音記號可以記在五線譜「線」上和「間」內。但必須對準後面所要表達之樂音，才能使後面樂音有準確音高。寫法上，升號和還原號都有兩條橫線，這兩條橫線要從左下方向右上方傾斜，並且要比兩條豎線畫得粗一些。



圖 1.7 變音記號

變音記號依照位置不同分成兩類，分別稱為調號和臨時變音記號。調號寫在譜號後面，表示在這個調所標到之樂音，不管音高，都要固定升高或者降低。例如：E 大調是 #F、#C、#G、#D，那麼在這個曲調裏所有 F、G、C、D 音都要升高半個音，但是如果遇到臨時變音記號以臨時變音記號為準(圖 1.8)。



圖 1.8 調號

臨時變音記號也叫臨時升降號，表示這個音要臨時改變音高。臨時升降號的效力為一小節，在同一小節之內有同樣音名之樂音都要服從第一個變化音，跟著變化同樣音高，而它前面在就不需要再有再記錄相同符號(圖 1.9)。



圖 1.9 臨時變音記號

五線譜上經常有比正常音符小之八分音符，符桿上帶著一條小斜線，此小音符稱為倚音，是一種裝飾音 (圖 1.10)。



圖 1.10 倚音

倚音在演唱或演奏時，越短促越好，不能占主音太多時間(圖 1.11)。



圖 1.11 倚音時值

多聲部音樂中，按一定音程關係(主要為三度)疊置起來三個或三個以上樂音同時或先後發響稱和絃(圖 1.12)。



圖 1.12 和絃

三度疊置時最低音稱根音，向上依次稱為三度音、五度音、七度音等。根音處於最低位置時稱原位和絃，否則稱轉位和絃(圖 1.13)。按三度關係構成的和絃，合乎泛音自然規律，音響和諧豐富，在多聲部音樂中被廣泛應用。不按三度關係疊置的和絃，若運用適當亦能產生令人滿意和聲效果。



圖 1.13 轉位和絃

本研究最後轉換為電子音樂 MIDI 檔，其全名為「Musical Instrument Digital Interface」，稱為數位樂器介面，運用這種數位介面，能讓不同種類電子樂器透過一組標準資料轉換規則，彼此進行溝通並連結。Roland 公司在 1991 年制定了

General MIDI 標準，簡稱 GM，主要將音源器前 128 種樂器名稱予以固定，並統一音色排列順序。

上述音樂知識包含了大部分記錄五線譜所使用之音樂符號，以及各符號含義，瞭解這些基本概念，除了有助於辨識方法之發展外，亦可在辨識上發生問題時，提供一套分析判斷準則，使辨識結果更為準確。

1.3 相關研究

印刷樂譜辨識自 1960 年代開始，許多方法相繼被提出來[1, 2, 3, 4, 5, 6, 7, 8]。由於音樂符號數量有限，且在譜表上有位置限定，記錄上也有嚴格規則，起初被視為一個單純文件分析問題，事實上，即使不是手寫，而針對印刷樂譜，樂譜辨識仍為一項困難且費時之工作，根據 Blostein 與 Baird[9]之研究指出，主要問題起因於符號分離或不完整所造成連通元件分段上之困難。

五線譜在音樂標誌中扮演一個重要角色，除了定義表示音高之垂直座標系統，也提供水平方向之時間座標系統。在相關研究中，Reed 與 Parker[10]利用五線譜五條譜線厚度相同且等距之特性來辨識五線譜。主要是以垂直搜尋方式，記錄具上述特徵之樣本，取各樣本間角度之平均值，淘汰與平均角度差異過大樣本，其餘樣本則可相互配對偵測出譜線(圖 1.14)。

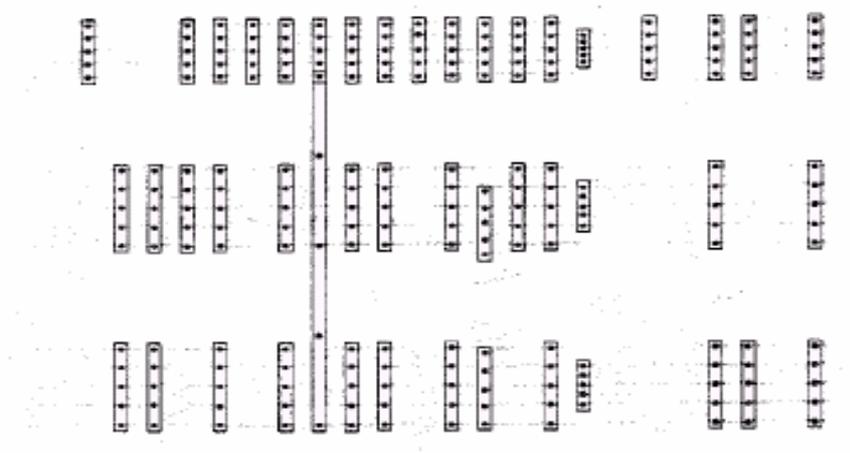


圖 1.14 以五線譜樣本偵測五線譜

Sicard[11]則將整張樂譜作 y 軸投影，利用五線譜譜表在 y 軸上產生明顯高峰值來偵測其位置(圖 1.15)。

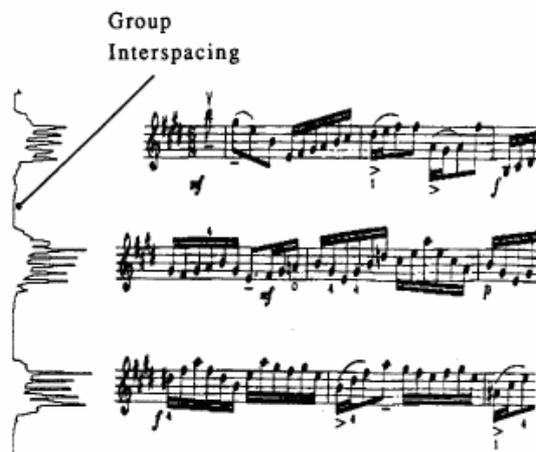


圖 1.15 利用 y 軸投影偵測五線譜位置

另外，Chung[12]採用標準直線表示法，即 $x\cos\theta + y\sin\theta = \rho$ 定義直線，且使用霍夫曼轉換(Hough Transform)偵測小節線與五線譜(圖1.16)。霍夫曼轉換[13]把參數空間細分為累加子單元(accumulator cells)，將直線上各點對應至另一參數空間，使得此直線在參數平面某一累加子單元形成高頻率，藉此累加子單元之參數反推回原直線。



圖 1.16 採用霍夫曼轉換偵測小節線與五線譜

辨識符號方面，Kimura 與 Shridha[14]將臨時記號、休止符、譜號...等音樂符號外圍輪廓與邊界範圍(bounding box)左界正交距離之關係建成特徵剖面圖(character profile)，利用特徵剖面圖相關屬性，如：各位置間相對與絕對極大、極小值...等，來定義該音樂符號(圖 1.17)。

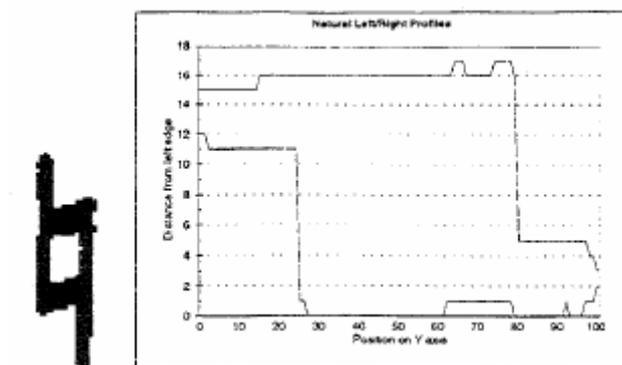
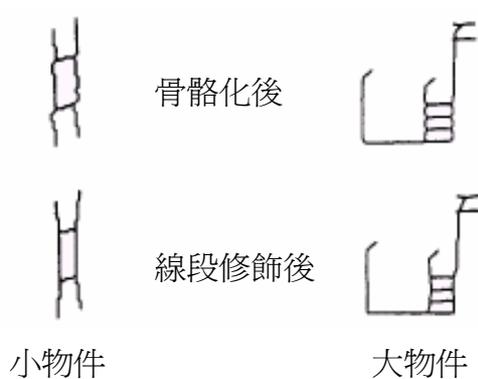


圖 1.17 利用特徵剖面圖

Randriamahefa[15]等人則利用了型態學法中骨骼化運算，辨識之前先將分段後音樂符號骨骼化，並依照其定義小物件及大物件做個別線段修飾(圖 1.18(a))，最後獲得各式音樂符號骨骼圖。將這些骨骼圖之屬性依照相關參數進行辨識，判斷其為何種音樂符號，如：空心符頭骨骼為一小圈圈，且其高度約等於五線譜譜線間距；實心符頭骨骼為一斜率約 45 度之小線段；符桿骨骼斜率介於 80 至 100 度，且其長度大於一定值(圖 1.18(b))...等等。



(a)

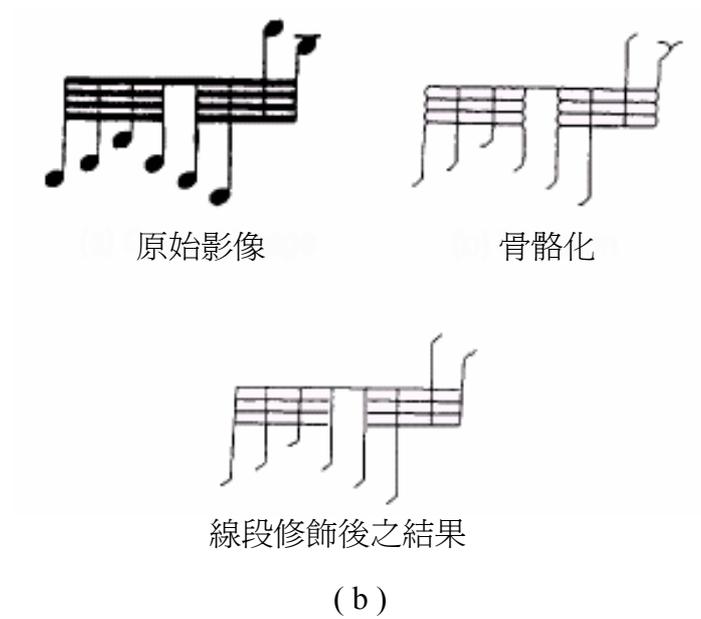
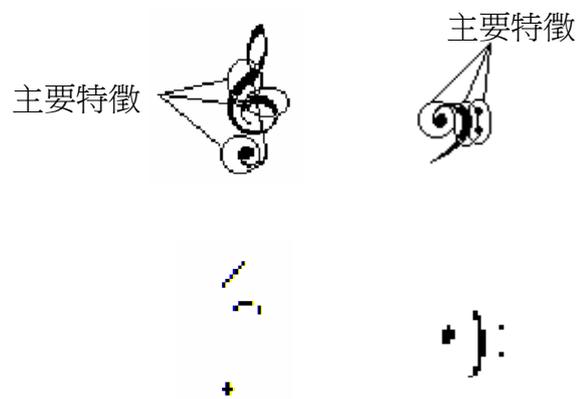


圖 1.18 利用型態學法之骨骼化運算辨識音樂符號

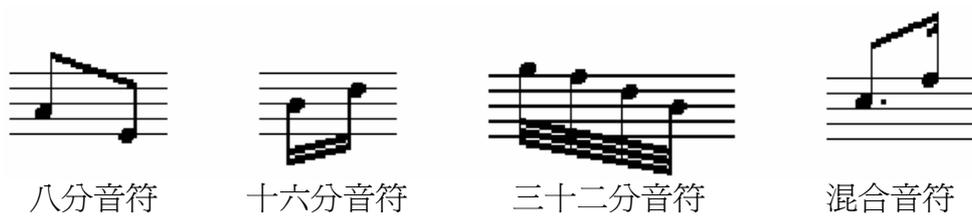
此外，Chung[12]利用型態學法辨識譜號、音符、休止符、升降記號、附點...等音樂符號，針對不同的音樂符號設計不同辨識樣本(圖 1.19)，如遇到無法僅靠型態學法偵測時，再配合符號屬性做判斷(signature analysis)。



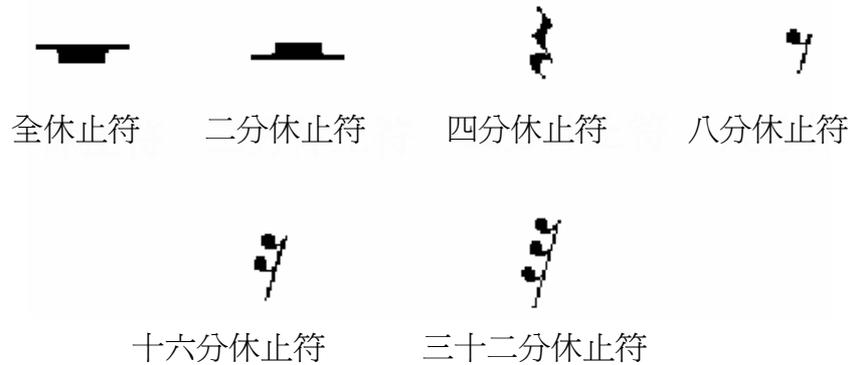
(a) 高低音譜號樣本



(b) 音符樣本



(c) 相連音符樣本



(d) 休止符樣本

圖 1.19 利用型態學法辨識音樂符號所建立之各式樣本

在 Carter 與 Bacon[16]提出之辨識系統中，以列(行)掃描輸入影像，將所有黑色像素水平(垂直)片段(run)視為一節點(node)，並以中心點座標表示此片段。若兩節點重疊且位於相鄰列(行)則連結起來。此方法稱為 LAG(Line Adjacency Graph)，主要用來偵測連通元件，以利音樂符號分段(圖 1.20)。

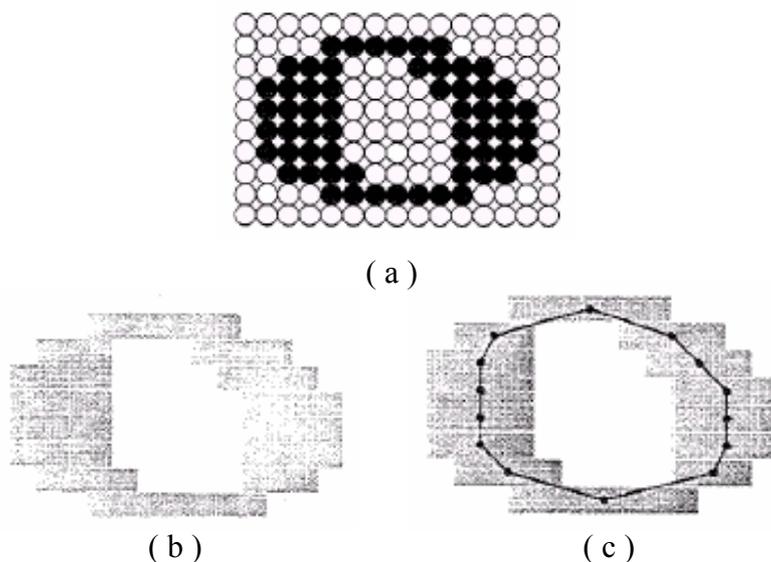
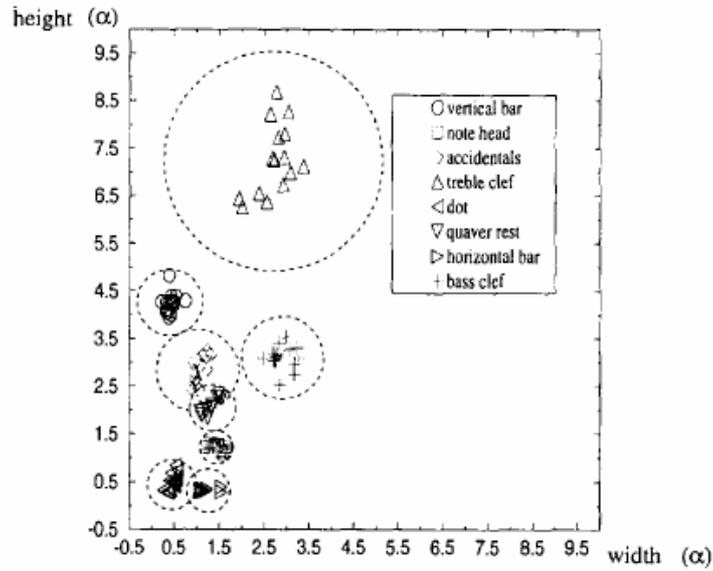


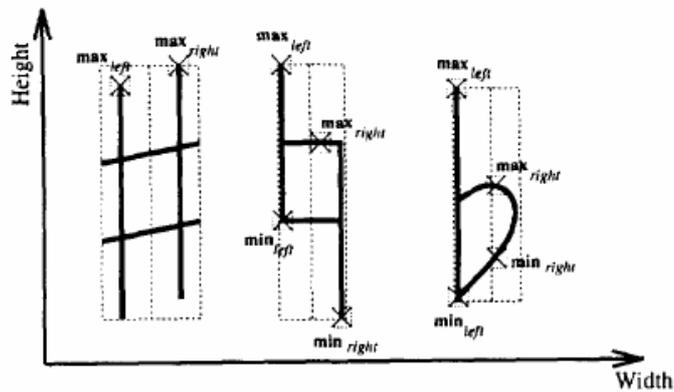
圖 1.20 (a)全音符 (b)全音符之水平片段 (c)全音符之 LAG

印刷樂譜辨識不僅需要影像處理層次技術，更有賴於樂理知識修正[17, 18, 19]。近幾年來，Ng 與 Boyle[20]，以及 Fujinaga[21]皆提出具有「學習」功能之方法，主要是利用 k-NN(k-nearest-neighbor)演算法將音樂符號分類，k-NN 是一種直覺之分類方法，當 $k = 1$ 時，其精神就是在所有資料點中找到一個和測試點距離最近之鄰居點，並以它所屬類別來當成測試點之類別；若將 1-NN 延伸，改用 k 個最近鄰居點投票來決定其所屬類別，則為 k-NN 分類法，其優點是較為公正，尤其當資料點類別越多時，藉由多人投票方式來決定其所屬類別會較具可信度，而缺點是所需花費時間成本較 1-NN 多。

其中，Ng 依照音樂符號之高度與寬度進行分類，如圖 1.21(a)所示， α 為譜線間距與譜線厚度和，再針對各類別辨識屬於何種音樂符號，其中，臨時記號以其上下左右界之相互關係區別出升記號、降記號或還原記號(圖 1.21(b))。



(a)



(b)

圖 1.21 (a)使用 k-NN 分類法所分類音樂符號之叢集(cluster) (b)以邊界之相互關係區別臨時記號

而在 Fujinaga 提出之 AORM(Adaptive ORM)系統中，則對一個音樂符號考慮多個屬性，形成一個集合稱為特徵向量(feature vector)，用來當作 k-NN 分類法之資料點。為提升效率，此系統更進一步對於不同之特徵向量，使用基因演算法(Gene Algorithm)賦予最佳權重。

另一位學者，Rossant[22]採用 template-matching 進行辨識，並將不確定之音樂符號給予一種以上之假設，最後以拍數決定出最有可能之符號，其稱此為 global decision，雖然有效改進了市面上商業軟體之缺點，不過使用時須離線辨識，相對地也造成使用者困擾。

上述研究中仍有許多不同限制，例如：有些對於掃描樂譜解析度要求為 300 dpi 以上，故一頁普通樂譜掃描成二值檔案就需要約 1MB 大小記憶體容量，有些則要求在掃描樂譜時必須要放正，亦即不能有傾斜之情況，否則，就是所提方法時間複雜度極高，實際應用並不方便，而本論文所發展印刷樂譜辨識系統並無以上要求，且可在短時間內得到辨識結果。

1.4 研究概述

由於印刷品質與排版情形不同，因此各類樂譜上音樂符號尺寸與形狀皆不盡相同，而且每次進行光學掃描時所設定解析度也可能有所不同，種種因素都造成辨識上困難，再加上人為因素可能造成樂譜傾斜或雜訊，使得許多相關研究無法很彈性處理這些問題，或只能在某些限制下進行辨識。本論文提出之印刷樂譜辨識系統，除了採用影像處理方法來辨識外，更導入樂理規則判斷，而且所有過程都能夠在動態、彈性和容錯下進行，成功地解決上述問題。

本系統分成三大階段，如圖 1.22 所示，第一階段為前置處理，為免除人為掃描影像之傾斜，首先將影像做角度修正，使樂譜儘可能成為水平，以利接下來五線譜偵測與移除。在偵測五線譜同時，可以求得五線譜譜線間距與譜線厚度，此為辨識階段重要依據，而移除五線譜是為为了使所有音樂符號孤立出來，避免五線譜干擾符號辨識。

第二階段，首先採用投影法區分出垂直線段，所謂垂直線段音樂符號包括符桿、符頭、符尾、小節線、終止線、四分休止符、變音記號、倚音...等，而非垂直線段音樂符號包括譜號、拍號、休止符(四分休止符除外)、附點...等。在辨識過程中，將已辨識出之音樂符號從原圖中移除，使輸入影像資訊由繁亂而單純。將上述辨識出符號組合成音樂標誌，並引入樂理限制與輔助，修正不符合拍數之小節，如此可偵測出全音符及校正辨識錯誤之處，使樂譜完全合乎樂理意義。

最後一個階段，將所有音樂標誌種類、順序、音長、音高...等資訊記錄成音樂表示語言，再經程式轉換成電子音樂 MIDI 格式，藉由播放軟體將旋律播放出來。

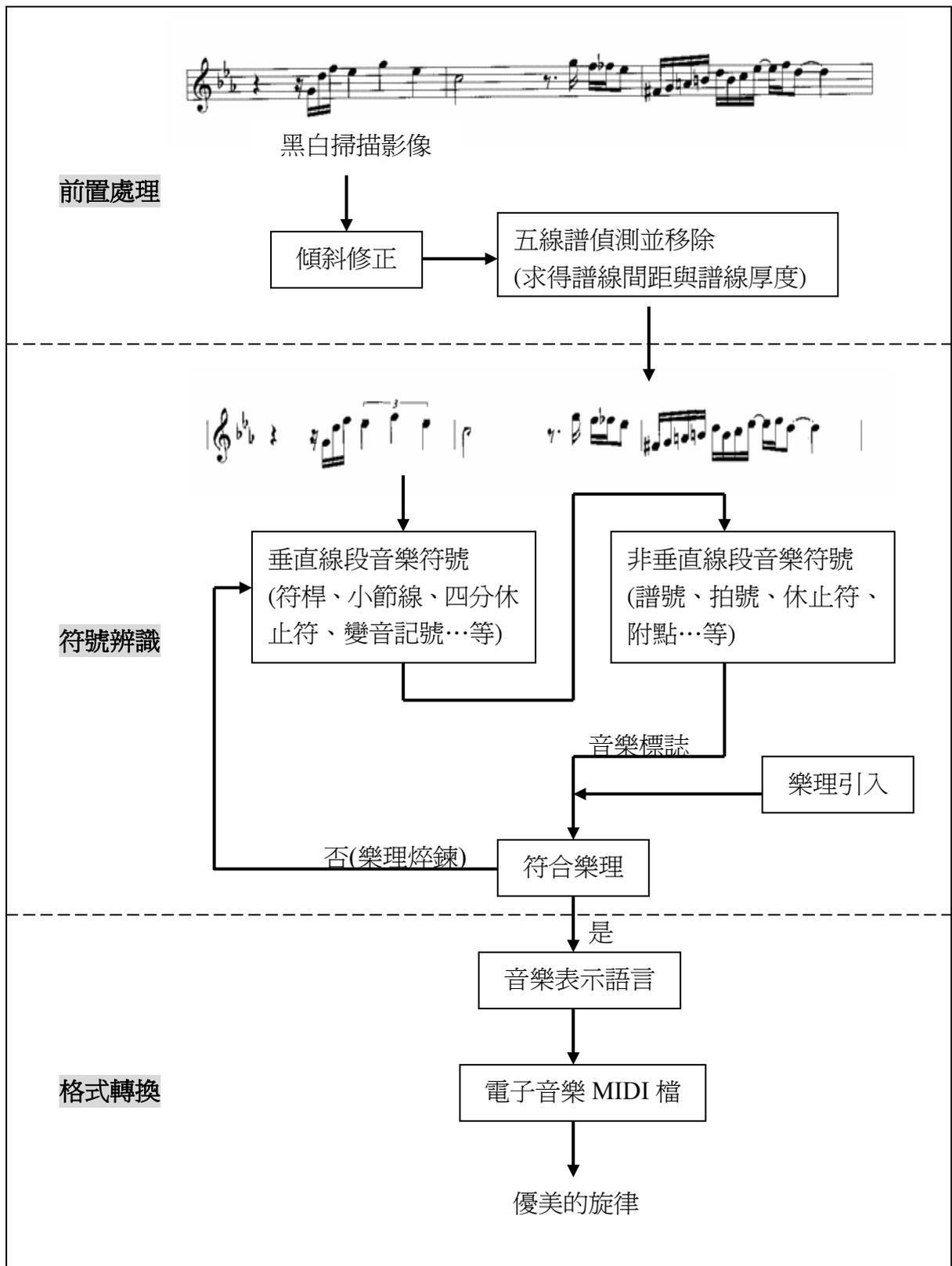


圖 1.22 系統流程圖

第二章 理論基礎

2.1 正交投影法

正交投影法(Projection)常使用在求取傾斜角度，以及水平或垂直線偵測。正交投影法分爲 x 軸投影(x-projection)與 y 軸投影(y-projection)，x 軸投影表示將垂直方向像素投影到 x 軸，因此又稱爲垂直投影(vertical projection)，其公式爲

$$v(j) = \sum_i bp(i, j)$$

其中， i 代表列， j 代表行， $bp(i, j)$ 表該座標像素值爲 0 (黑色)， $v(j)$ 則是累計第 j 行上所有垂直方向像素值爲 0 之點；同理，y 軸投影表將水平方向像素投影到 y 軸，因此又稱爲水平投影(horizontal projection)，其公式爲

$$h(i) = \sum_j bp(i, j)$$

$h(i)$ 則是累計第 i 列上所有水平方向上像素值爲 0 之點。

將一定比例範圍內影像投影到單一軸上，會形成一分布圖，稱爲投影剖面(projection profile)。不同影像所形成之投影剖面通常不相同，但將一影像長寬等比例放大縮小，其投影剖面與原影像分布情形則大致相同(圖 2.1)。

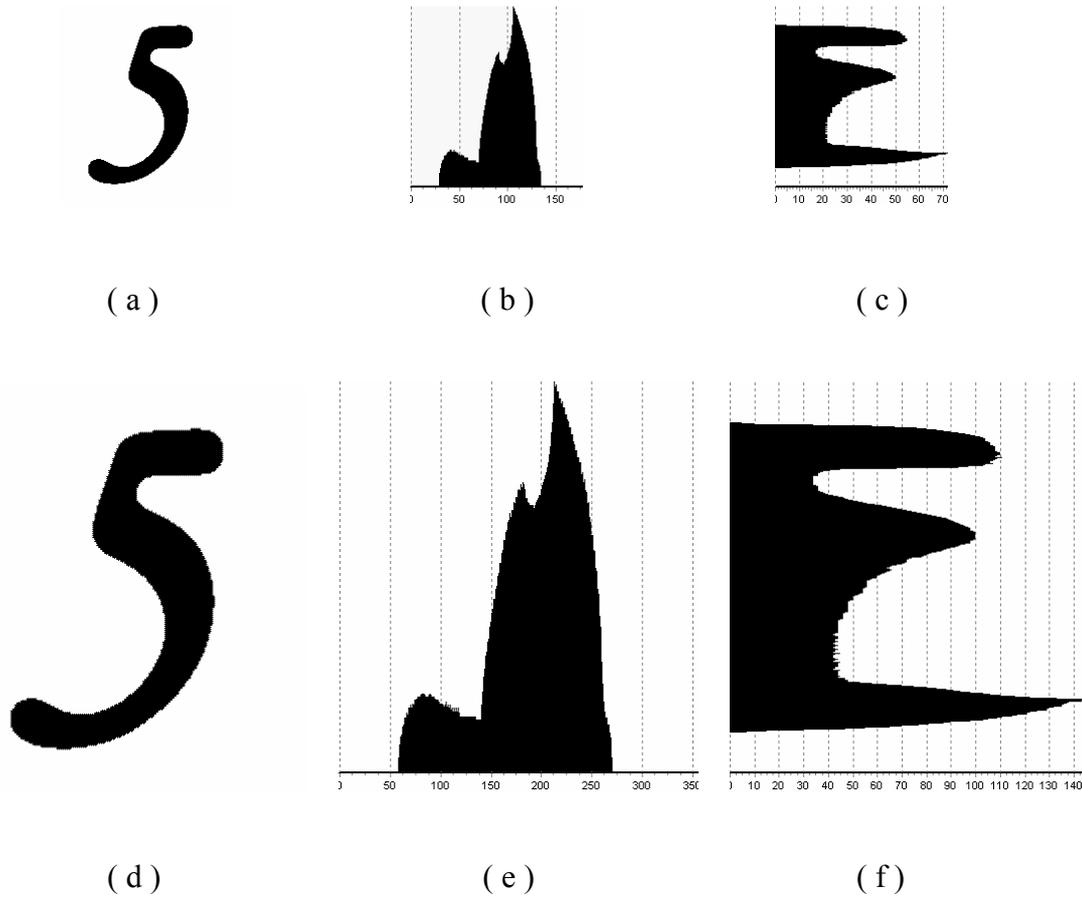


圖 2.1 相同影像不同尺寸投影圖 (a)大小為 178×206 影像 (b)原影像之 x 軸影
 (c)原影像之 y 軸投影 (d)原影像放大兩倍(大小為 356×412 影像)(e)原影像放
 大兩倍後之 x 軸投影 (f)原影像放大兩倍後之 y 軸投影

2.2 型態學法

以科學之觀點，型態學(Morphology)[23, 24]為生物學一個分支，通常表示對動植物形狀和結構之研究；數學上，型態學(Mathematical Morphology)為一有效擷取影像分量之工具，特別是針對形狀之表示和描述；利用在影像處理方面，型態學意指分析影像幾何結構之方法，例如：邊界、輪廓、凸形封包...等。

數學型態學是以集合理論作為表達，集合用來表示影像中物體形狀，例如：一二值影像中所有黑色像素集合是此影像之完整描述。在二值影像中，上述集合是二維指數空間一部份，其中，集合中每一個元素是一個二維空間向量變數值組，代表影像中座標為 (x, y) 之黑色像素點。

型態學濾波(Morphological filter)為基本型態操作組合，主要運算有侵蝕(erosion)、膨脹(dilation)、閉合(closing)、開啓(opening)、骨骼化(skeletonization)、細化(thinning)...等，較標準線性濾波更適合使用在形狀分析上。大體來說，型態操作為經由一具有某種形狀與尺寸之影像，交互作用於原始影像而轉換成另一影像，此具有某種形狀與尺寸之影像，稱為結構元素(structure element)(圖 2.2)，在原始影像上，具有與結構元素尺寸形狀完全相同之幾何特性部分將被保存，其餘則被消滅。因此，型態操作可以將影像資料簡化成只存留具有某種特性，與此特性不相關部分則不存在。

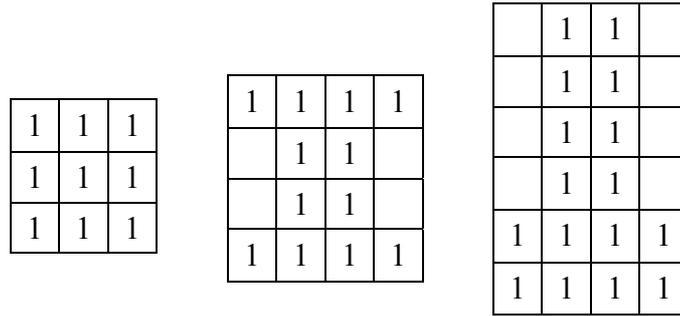


圖 2.2 結構元素

定義在尤拉(Euclidean)環境上之數學型態學稱為尤拉型態學(Euclidean morphology)，而定義在數位環境上之數學型態學則稱為數位型態學(digital morphology)。一般來說，除了前者使用於連續訊號處理，後者使用於數位訊號處理，其餘方面，兩者關係相當近似。雖然實際操作多在數位環境，但藉由熟諳由拉環境數學型態學，有助於了解這些型態操作如何應用在理論及實務上。

由於本研究之輸入影像為二值影像，因此，所有討論皆在二維空間上。首先，介紹平移(translation)與反射(reflection)兩種運算基本定義，再敘述一些主要型態運算。

基本定義

給一影像 A ，將 A 平移點 x ，以 $A_{\vec{x}}$ 表示，定義為

$$A_{\vec{x}} = \{ \vec{a} + \vec{x} \mid \forall \vec{a} \in A \} = A + \vec{x} \quad (\text{圖 2.3(a)})$$

給一影像 B ，將 B 反射，以 \hat{B} 表示，定義為

$$\hat{B} = \{ -\vec{b} \mid \forall \vec{b} \in B \} \quad (\text{圖 2.3(b)})$$

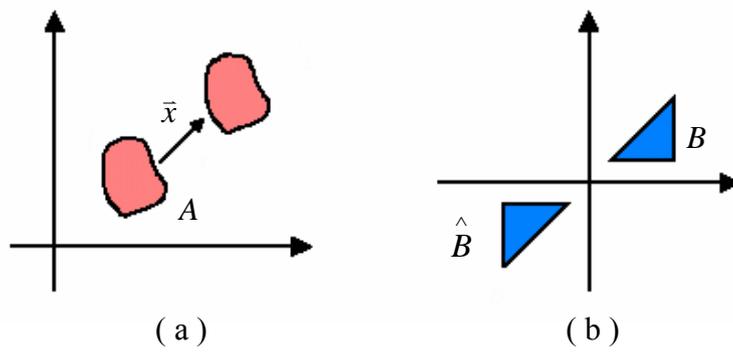


圖 2.3 (a)平移 (b)反射

侵蝕與膨脹

假設 A 和 B 是 Z^2 之集合， A 被 B 侵蝕，以 $A \ominus B$ 表示，定義為

$$A \ominus B = \{ \bar{x} \mid (B)_{\bar{x}} \subseteq A \}$$

意即被 B 侵蝕的 A 是 B 平移 \bar{x} 之後包含在 A 中所有 \bar{x} 的集合(圖 2.4)。

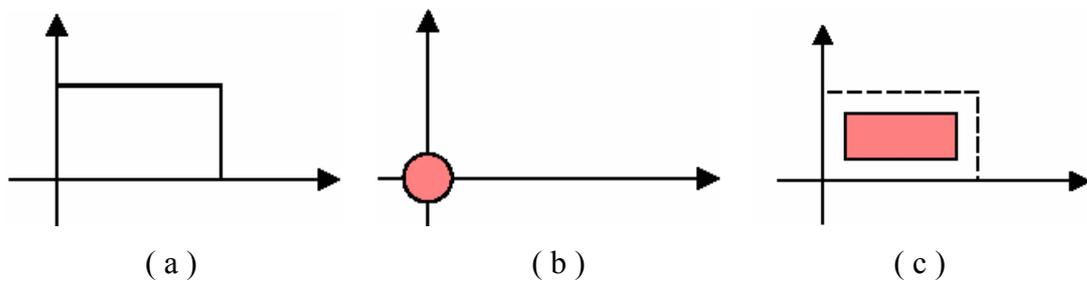


圖 2.4 (a)集合 A (b)結構元素 B (c) A 被 B 侵蝕

而 A 被 B 膨脹，以 $A \oplus B$ 表示，定義為

$$A \oplus B = \left\{ \vec{x} \mid (\hat{B})_{\vec{x}} \cap A \neq \phi \right\}$$

意即所有可以使 \hat{B} 在平移 \vec{x} 後與 A 至少重疊一個非零元素的集合(圖 2.5)。

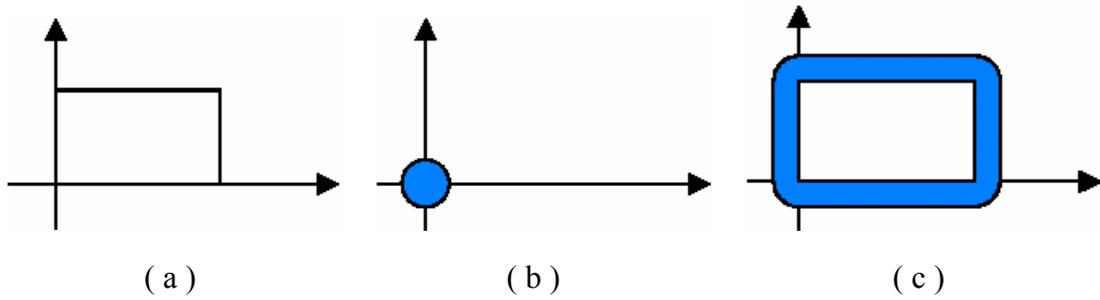


圖 2.5 (a)集合 A (b)結構元素 B (c) A 被 B 膨脹

開啓與閉合

將侵蝕與膨脹分別加以組合，可以得到不同效果，若將影像先侵蝕後再膨脹，稱之為開啓，反之則稱為閉合。

集合 A 被結構元素 B 開啓，記為 $A \circ B$ ，定義為

$$A \circ B = (A \ominus B) \oplus B \quad (\text{圖 2.6})$$

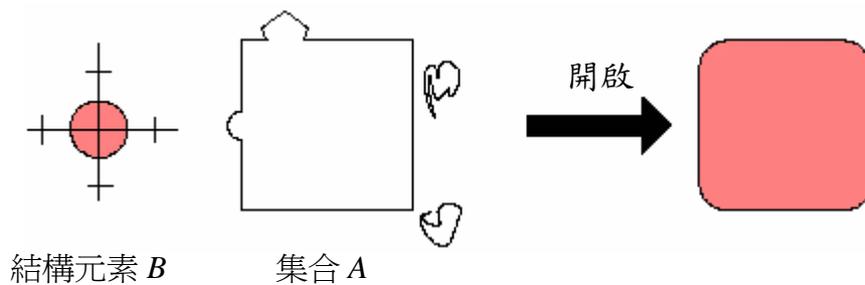


圖 2.6 A 被 B 開啓

集合 A 被結構元素 B 閉合，記為 $A \cdot B$ ，定義為

$$A \cdot B = (A \oplus B) \ominus B \quad (\text{圖 2.7})$$

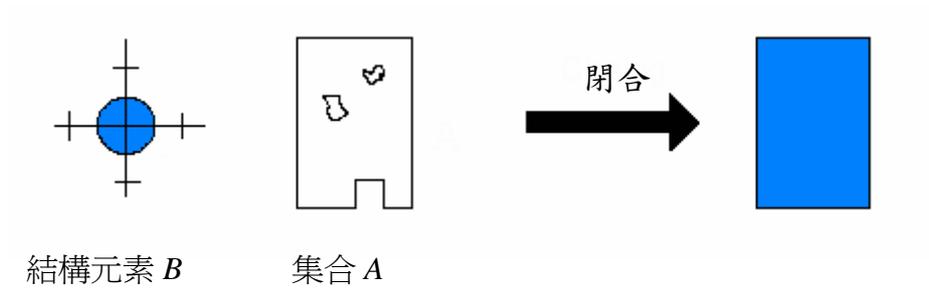


圖 2.7 A 被 B 閉合

骨骼化

集合 A 的骨骼化可以侵蝕與閉合運算來表示，假設 $S(A)$ 為 A 之骨骼，則

$$S(A) = \bigcup_{k=0}^K S_k(A)$$

其中

$$S_k(A) = \bigcup_{k=0}^K \{(A \ominus kB) - [(A \ominus kB) \circ B]\}$$

而 B 為結構元素，「 \circ 」為開啓運算， $(A \ominus kB)$ 表 A 之連續 k 次侵蝕，即

$$(A \ominus kB) = ((\dots(A \ominus B) \ominus B) \dots) \ominus B \quad \text{進行 } k \text{ 次}$$

K 是在 A 侵蝕為空集合之前的最後迭代步驟，意即

$$K = \max \{k \mid (A \ominus kB) \neq \phi\}$$

圖 2.8 為一骨骼化例子，第一行表示原集合 A (上方) 與集合 A 以結構元素 B 進行兩次侵蝕，因為 A 再一次被侵蝕即為空集合，因此 $K=2$ 。而第二行表示第一行之集合以 B 進行一次閉合運算，第三行則為第一行與第二行之差集。第四行包括兩部分骨骼與最後結果(下方)。此外，由此例可得知依此方法獲得之骨骼不保證連通。

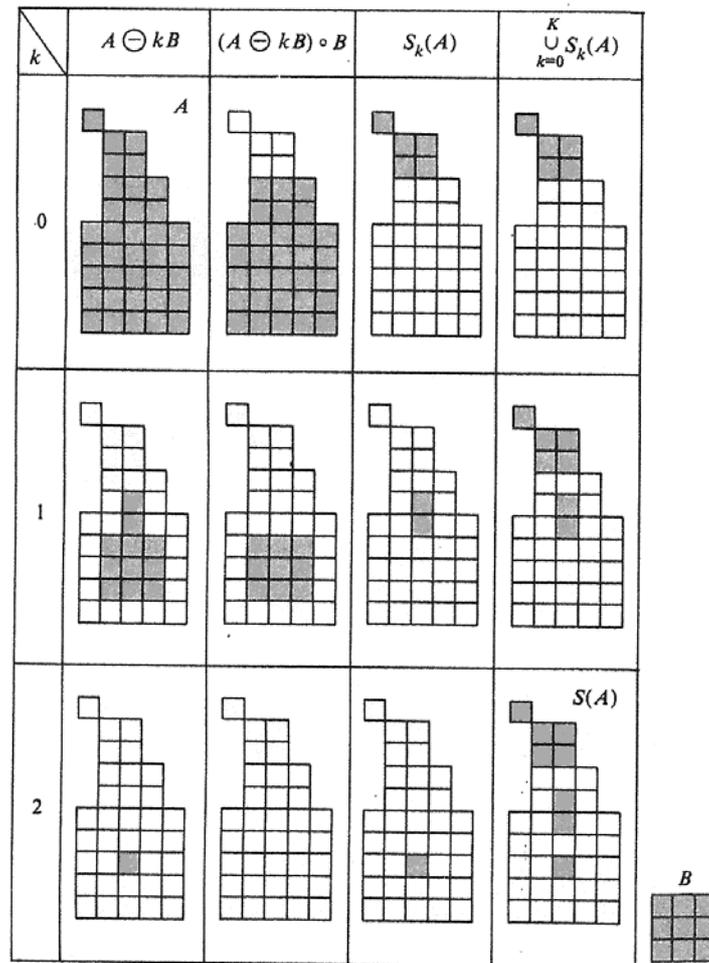


圖 2.8 骨骼化例子

2.3 樂理使用

印刷樂譜辨識需要適時藉以樂理輔佐，以避免或修正辨識錯誤。針對各局部音樂符號辨識，本研究引入下列樂理：1)水平位置屬性。例如：臨時變音記號只會出現在譜號後方或音符前方，附點只跟隨在音符或休止符後方(圖 2.9)...等；2)垂直位置屬性。例如：全休止符在第四線下方，二分休止符在第三線上方，而八分休止符則在第二到第四線間(圖 2.10)...等；3)長度屬性。例如：二分音符符桿長度大於三倍譜線間距與譜線厚度和，G 譜號長度遠大於譜表寬，C 譜號長度遠等於譜表寬...等；4)寬度屬性。例如：小節線寬度小於一定值，全音符寬度大於實心與空心符頭寬度(圖 2.11)...等。

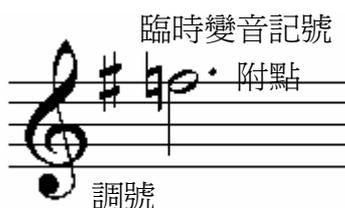


圖 2.9 水平位置屬性 — 變音記號與附點位置

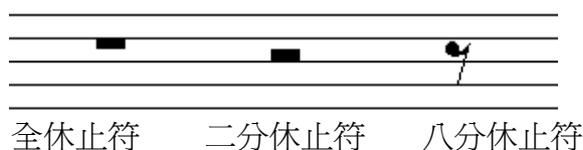


圖 2.10 垂直位置屬性 — 休止符位置



圖 2.11 寬度屬性 — 小節線與全音符寬度

辨識過程中，對於音樂標誌重建也需要使用到樂理。當多個符號鄰近出現時，各符號間是否存在關係與其相互意義需進一步考量，例如：複合符號之音樂標誌(圖 2.12)；或者，出現臨時變音符號後，同一小節之內所有同樣音名之樂音都要服從第一個變化音(圖 2.13)...等。

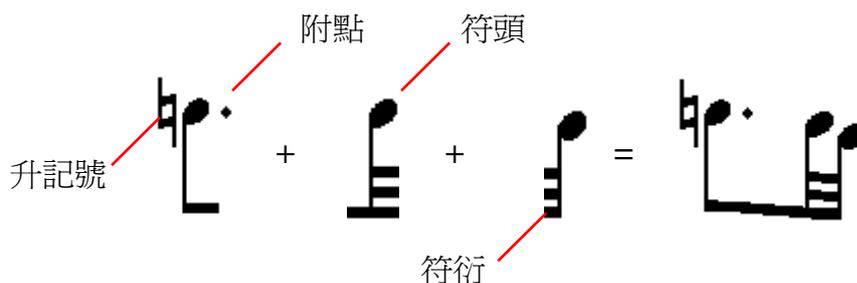


圖 2.12 複合符號音樂標誌

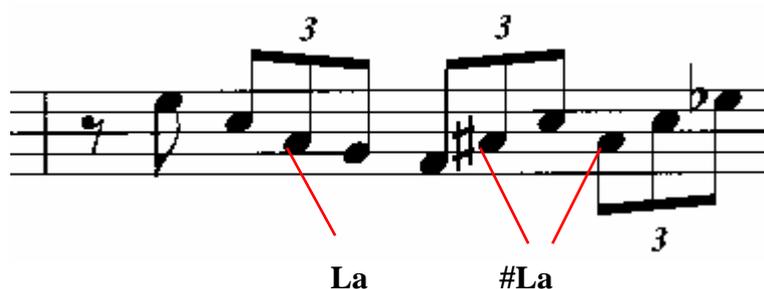
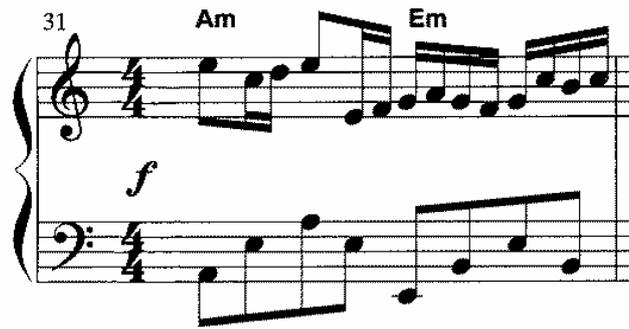
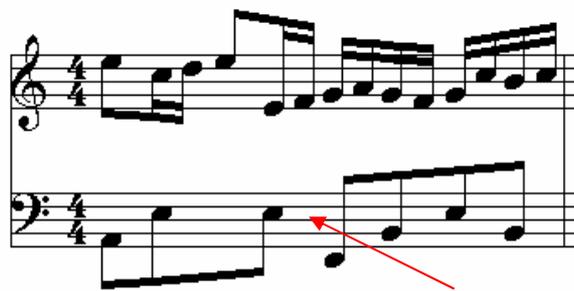


圖 2.13 變音記號規則(G 譜號譜表)

辨識完成後，在格式轉換之前，為使樂譜辨識結果完全合乎樂理，因此，需要對辨識獲得之音樂標誌做總體性檢驗。本研究藉由拍號規定，判斷每一小節是否與拍數吻合，用以偵測出錯誤之處，再採取適當補救措施。如圖 2.14 所示，拍號可檢查出此小節辨識結果遺漏一個八分音符。



(a)



(b)

圖 2.14 拍號檢驗 (a)輸入樂譜 (b)辨識結果

本研究在符號辨識時，引入符號特有屬性；標誌重建時，考慮各符號間關係；小節修正時，使用拍號規則，藉由這些樂理，提高辨識準確性。

第三章 研究方法及步驟

3.1 前置處理

印刷樂譜可能由於人為疏忽或硬體裝置本身因素造成掃描影像傾斜，導致辨識準確性降低，因此在辨識之前，須先將所擷取數位樂譜進行傾斜修正。此外，五線譜與部分音樂符號重疊，為避免辨識階段產生干擾，本系統將先行將五線譜移除。

3.1.1 傾斜修正

樂譜傾斜角度偵測方法為擷取同一譜線上兩端點斜角。一般而言，樂譜上方通常附有圖騰、插畫或記有歌名、作曲者、演奏速度...等資訊，因此，以樂譜最後一組五線譜第一線作為傾斜角度之偵測，而不以其餘譜線為考慮，除可避免上述干擾，亦可增快執行速度。

首先，由下而上一列一列計算輸入樂譜之黑色像素，當投影值數次超過高峰且連續低峰次數大於閾值(threshold)則判定擷取到最後一組五線譜，如圖 3.1 所示(歌譜選自 Canon[25])。

The image shows a musical score for piano, consisting of five systems of staves. Each system includes a treble clef staff and a bass clef staff. Chord annotations are placed above the treble staff. The systems are numbered 43, 46, 49, 52, and 55. The final system (55) is highlighted with red dashed lines and a red arrow pointing to the beginning of the staff.

System 43: Chords C, G, Am, Em, F, C.

System 46: Chords F, G, C, G, Am, Em.

System 49: Chords F, C, F6, G, C, G.

System 52: Chords Am, Em, F, C, F6, G.

System 55: Chords C, G, Am, Em, F, C.

圖 3.1 偵測樂譜最後一組五線譜

將此區間上下一些距離作 x 軸投影，擷取此五線譜兩端內側為左右界，如圖 3.2 藍色線段所示，搜尋最小投影值中最近左右界 x 座標 $x1$ 、 $x2$ ，最小投影值代表此 x 座標軸上無音樂符號存在，而最近左右界為最接近譜線兩端點。由投影下界往上，擷取與譜表第一線交點 $(x1, y1)$ 、 $(x2, y2)$ 為參考點(圖 3.2)。



圖 3.2 偵測傾斜角度

樂譜依兩參考點所形成之角度作校正，圖 3.3 為圖 3.1 樂譜順時針 0.32 度傾斜修正後影像。



圖 3.3 原圖 3.1 傾斜修正後影像

3.1.2 五線譜偵測

首先，偵測可能譜線。將修正樂譜進行 y 軸投影形成一剖面圖，擷取超過高

峰值位置，連續出現高峰值起點與終點判定為同一線段上下界，記錄所有線段上下界 y 座標。估算相鄰線段上下界間距與各線段上下界間距，分別累計各數值出現次數，頻率最高者訂定為樂譜之譜線間距與譜線厚度，譜表寬度即為四倍譜線間距與五倍譜線厚度和，且此兩參數為後續階段定義音樂符號尺寸之基準。

再者，過濾不當線段。上述高峰值可能由於樂譜含有歌詞、圖騰，或音樂符號集結而造成，如圖 3.4(歌譜選自 My Funny Valentine[26])實線箭頭與圖 3.5(歌譜選自 Spanish Nights[27])箭頭所示。因此，分析線段上下相鄰間距是否至少其一為倍數譜線間距或線段厚度是否為譜線厚度，可剔除不符合邏輯之線段。



圖 3.4 實線箭頭：樂譜圖騰形成投影高蜂值
虛線箭頭：未測得譜線

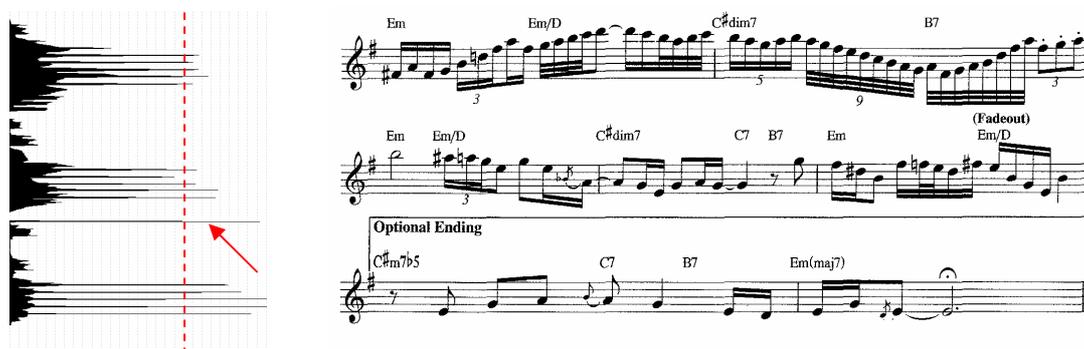


圖 3.5 音樂符號形成投影高蜂值

最後，補救未測得譜線。將前述移除剩餘線段依上下相鄰間距歸類出每一組譜表，判定所缺譜線位置，若位於第二、三、四線則依照譜線間距補齊，如圖 3.4 上方虛線箭頭所示，否則，取 y 軸投影值較大者決定所缺為第一線或第五線，如圖 3.4 下方二虛線箭頭所示。

五線譜左側可能記有演奏樂器名稱，為避免干擾辨識，須偵測出每組五線譜譜表左右界，先依據譜線間距與譜線厚度建立一寬度為譜線間距之譜表樣本(圖 3.6)，檢驗測得之五線譜位置，擷取五線譜起點與終點，如圖 3.7(歌譜選自紅綵妹妹[28])所示，圖中五線譜為依照測得之 y 座標與左右界描繪。

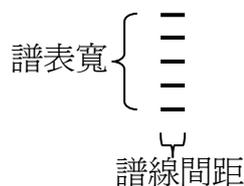


圖 3.6 譜表樣本

(5) 紅綵妹妹
The Red Ribbon Gal

緞造民歌 Suiyuan Folksong 羅炳良編 arr. Daniel LAW

Con Spirito

The musical score is presented in a system with four staves. The top two staves are for Tenors I and II (Tcn. I and Tcn. II), and the bottom two are for Baritone (Bar.) and Bass. The music is in 4/4 time with a key signature of two flats. The lyrics are written in Chinese characters below the vocal staves. The score includes dynamic markings like *Con Spirito* and *f* (forte). The lyrics are: 紅綵妹妹 嗯啾啾，長得好那麼 嗯啾啾，櫻桃小口 嗯啾啾，一點點那麼 嗯啾啾。

圖 3.7 測得之五線譜

移除譜線時，為避免破壞音樂符號，考慮譜線 x 座標上下各二倍譜線厚度黑色像素量，若超過譜線厚度值則與予保留，否則移除該 x 座標譜線，圖 3.8(a) 為圖 3.1 移除譜線後影像，圖 3.8(b) 為所移除之譜線。此一方法除與譜線相切之全音符與二分音符外，大部分音樂符號皆可獲得保全，如圖 3.8(a) 所示。

43 C G Am Em F C

46 F G C G Am Em

49 F C F⁶ G C G

52 Am Em F C F⁶ G

55 C G Am Em F C

全音符被破壞

二分音符被破壞

二分音符被破壞

(a)



(b)

圖 3.8 (a)原圖 3.1 移除譜線後影像 (b)所移除之譜線

3.2 符號辨識

音樂符號依屬性不同區分為垂直線段音樂符號與非垂直線段音樂符號，根據一定步驟進行辨識，並隨時將已辨識出之音樂符號清除，最後，須以樂理知識修正，使辨識結果符合樂理意義。

符號辨識主要採用投影法與型態學法。擷取垂直線段時，對前置處理階段測得之五線譜上下各譜表寬度做 x 軸投影，擷取出長度超過 1.8 倍譜線間距與譜線厚度和之線段，如圖 3.9 所示，紅色部分為偵測到之垂直線段，紫色部分為測得之五線譜，藍色部分為每組五線譜投影上下界。

以型態學法進行辨識時，將預先建好之型態樣板參照輸入樂譜之譜線間距與譜線厚度，依一定比例調整為適當尺寸，進而與輸入影像進行核對。定義調整尺寸後型態樣板 k 為 M^k ， B^k 為其黑色像素總數，與欲比對影像 I 之圖像於 (x, y) 之相互關係 C^k 為：

$$C^k(x, y) = \frac{1}{B^k} \sum_{(i,j) \in M^k} M^k(i, j) * I(x+i, y+j)$$

$$M^k(i, j) = \begin{cases} 1 & \text{黑色像素,} \\ 0 & \text{白色像素,} \end{cases}$$

$$I(x+i, y+j) = \begin{cases} 1 & \text{黑色像素,} \\ 0 & \text{白色像素。} \end{cases}$$

當 $C^k(x, y)$ 愈趨近於 1.0 表示圖像與該型態樣板愈相似，實作上則給定一閾值以區分相近與不相似。

音樂依聲部個數可分為單聲部(monophony)與多聲部(polyphony)。多聲部樂譜中，同一聲部之每組譜表最前方小節線會連接起來，有些更在此小節線前加上大括號或中括號，因此，擷取五線譜左界之垂直線段並分析其長度，若無垂直線段或長度為譜表寬時為單聲部音樂(圖 3.5); 否則，其長度為兩倍或三倍譜表寬(圖 3.9)，並可依此關係決定聲部個數。

The diagram shows five musical staves with vertical red line segments at the beginning of each staff, indicating the start of a musical phrase. The segments are grouped by brackets on the left, with labels indicating the instrument and the number of staff widths the segment spans:

- Violin 1:** Labeled "二倍 譜表寬" (2x staff width).
- Violin 1:** Labeled "三倍 譜表寬" (3x staff width).
- Viola 1:** Labeled "三倍 譜表寬" (3x staff width).
- Viola 2:** Labeled "三倍 譜表寬" (3x staff width).
- Violoncello:** Labeled "二倍 譜表寬" (2x staff width).

圖 3.9 垂直線段偵測與聲部個數分析

3.2.1 垂直線段音樂符號辨識

本研究所辨識垂直線段音樂符號如表 3.1 所示。

表 3.1 垂直線段音樂符號

種類	名稱	寫法
音符	實心符頭	●
	空心符頭	○
	符桿	
	符尾), 9, 9
	符衍	—, ≡, ≡≡
小節線	小節線	
	複小節線	
	終止線	
休止符	四分休止符	≡≡≡
變音記號	升記號	#
	降記號	b
	還原記號	♮
裝飾音	倚音	♪

音符辨識部分，針對垂直線段兩側，將表 3.2 之符頭型態樣板依比例自動調整高度為譜線間距與二倍譜線厚度和之尺寸進行辨識。實心符頭方面，在五線譜移除影像做辨識，待實心符頭偵測完畢，擷取符頭寬度值；空心符頭方面，由於

前置處理之五線譜移除階段破壞掉部分空心符頭，因此對於空心符頭需在傾斜修正，尙未移除五線譜之影像做辨識，且進一步確認與符頭相切之矩形範圍內，其黑色像素數量小於一定值。若偵測到實心或空心符頭，表示此垂直線段為符桿，將依 y 座標值分析其音高。

表 3.2 符頭型態樣板

名稱	寫法	形態樣板
實心符頭	●	●
空心符頭	○	○

符尾及符衍位於附有實心符頭符桿之符頭另一端，對此類符桿左右兩側作 y 軸投影，偵測其連續黑色像素片段個數以決定音符時值。待所有符尾辨識完畢，考慮相鄰符桿之符尾，若其左右成對出現則判定為符衍(圖 3.10)。

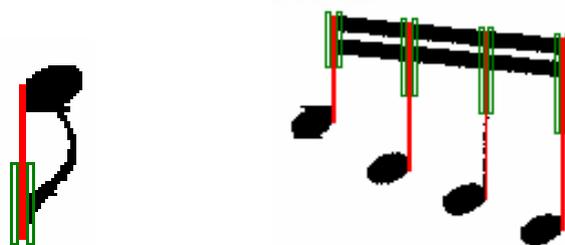
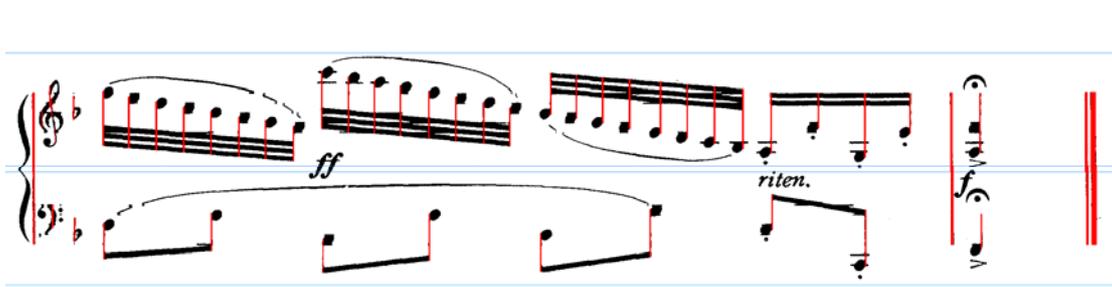


圖 3.10 符尾及符衍分析

小節線為左右界無其他音樂符號之狹長形狀音樂符號。對於單聲部樂譜，小節線兩端點座落於譜表第一線及第五線，而多聲部樂譜之小節線下端點位於第一線或該投影下界，上端點則位於第五線或該投影上界，如圖 3.9 與圖 3.11 所示。當兩小節線緊密出現時，則依寬度關係判定為複小節線或終止線(圖 3.11)。



(a)



(b)

圖 3.11 小節線位置

四分休止符經投影分析，中心部份會產生一處於第二線上方與第四線上方之垂直線段(圖 3.12)，因此，系統在進行四分休止符辨識時，將自動調整表 3.3 型態樣板之高為二倍譜線間距與三倍譜線厚度和，且針對上述垂直線段 x 座標位置進行之偵測。

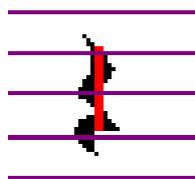


圖 3.12 四分休止符經投影產生垂直線段

表 3.3 四分休止符型態樣板

名稱	寫法	形態樣板
四分休止符		

調號緊鄰譜號且順序有一定規則，降記號調號按照「高完全四度」關係，依序為「Si Mi La Re Sol Do Fa」，圖 3.7 為降 A 大調；升記號調號音程按照「低完全四度」關係，依序為「Fa Do Sol Re La Mi Si」，圖 3.9 為 G 大調。臨時變音記號必出現在音符左側，且須與音符同音高，其中還原記號與升記號由兩短垂直線段組成，但還原記號左邊垂直線段在右邊垂直線段上方，升記號則相反(圖 3.13(a))；若只有一垂直線段可能為降記號或倚音，此時以表 3.4 型態樣板進行分析(圖 3.13(b))。

表 3.4 降記號與倚音型態樣板

名稱	寫法	形態樣板
降記號	b	˘
倚音	♪	•



圖 3.13 (a) 還原記號與升記號 (b) 降記號與倚音

3.2.2 非垂直線段音樂符號辨識

本研究所辨識非垂直線段音樂符號如表 3.5 所示。

表 3.5 非垂直線段音樂符號

種類	名稱	寫法
譜號	G 譜號	
	C 譜號	
	F 譜號	
音符	全音符	
休止符	全休止符	
	二分休止符	
	八分休止符	
	十六分休止符	
附點	附點	
拍號	數字 2	
	數字 3	
	數字 4	
	數字 6	
	數字 8	
	數字 9	

譜號有三種：G 譜號、C 譜號和 F 譜號。G 譜號主要特徵為第一線下方實心圓圈，C 譜號為第一線到第五線兩粗細垂直線段，C 譜號第四線與第三、第四間之實心圓圈，本系統依其特徵所建立型態樣板如表 3.6。

表 3.6 譜號型態樣板

名稱	寫法	型態樣板
G 譜號		
C 譜號		
F 譜號		

由於部分全音符在五線譜移除階段被破壞，因此亦需在傾斜修正，尙未移除五線譜之影像做辨識。全音符寬度大於實心符頭寬度，高度為譜線間距與二倍譜線厚度和，且方向不同於空心符頭，為左上往右下傾斜，型態樣板如表 3.7 所示。

表 3.7 全音符型態樣板

名稱	寫法	型態樣板
全音符		

休止符中，全休止符與二分休止符型態相同，惟全休止符上方緊貼第四線，而二分休止符下方緊貼第三線；八分休止符位於第二線到第三間，第三間有一實

心黑點；十六分休止符位於第一線到第三間，第二間與第三間皆有一實心黑點，表 3.8 為本系統依上述特徵所建立之型態樣板。

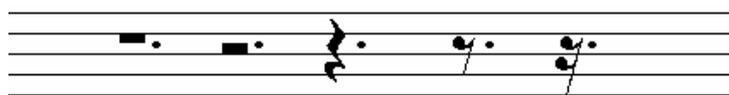
表 3.8 休止符型態樣板

名稱	寫法	型態樣板
全休止符		■
二分休止符		■
八分休止符		◡
十六分休止符		◡、◡

附點出現在「線」上音符時，附在符頭右上方，出現在「間」上音符，則附在符頭右方(圖 3.14(a))，而休止符之附點一律記錄在第三間休止符右側(圖 3.14(b))，型態樣板如表 3.9 所示。



(a)



(b)

圖 3.14 附點位置

表 3.9 附點型態樣板

名稱	寫法	型態樣板
附點	.	.

拍號跟隨在調號後且只出現在各聲部第一組譜表，第一線到第三線記錄分母，代表音值，第三線到第五線記錄分子，代表拍數，辨識時系統自動將表 3.10 型態樣板高度調整為二倍譜線間距與譜線厚度和。

表 3.10 拍號型態樣板

名稱	寫法	型態樣板
數字 2	2	2
數字 3	3	3
數字 4	4	4
數字 6	6	6
數字 8	8	8
數字 9	9	9

3.2.3 辨識步驟

光學掃描樂譜經前置處理階段定位出譜表，獲得音樂符號 y 軸位置大致範圍，至於各類音樂符號實際分布，須在雜亂資料中漸漸釐清出來，因此，除辨識方法外，辨識步驟亦為影響成果一重要因素。本研究辨識流程圖如圖 3.15 所示。

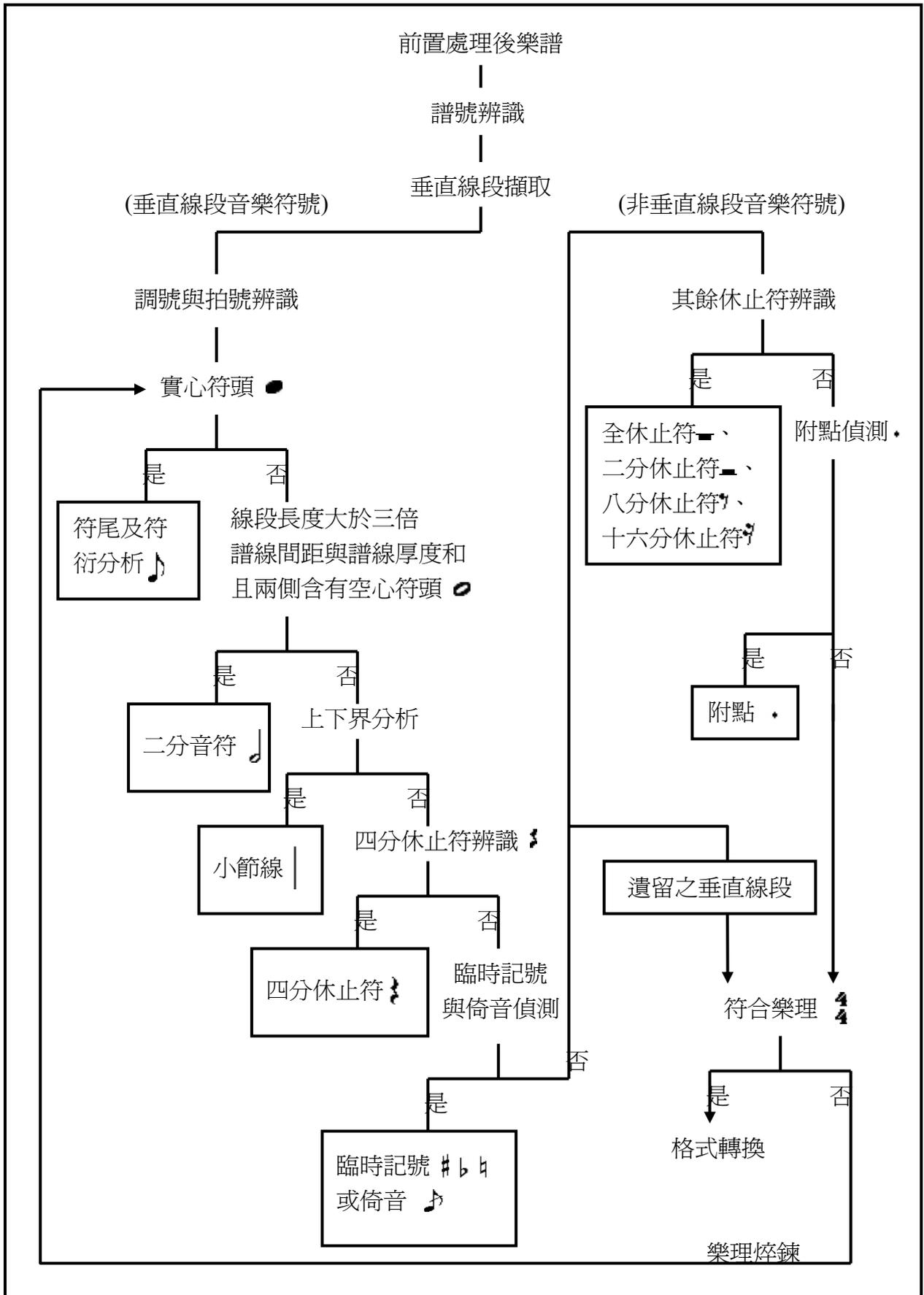
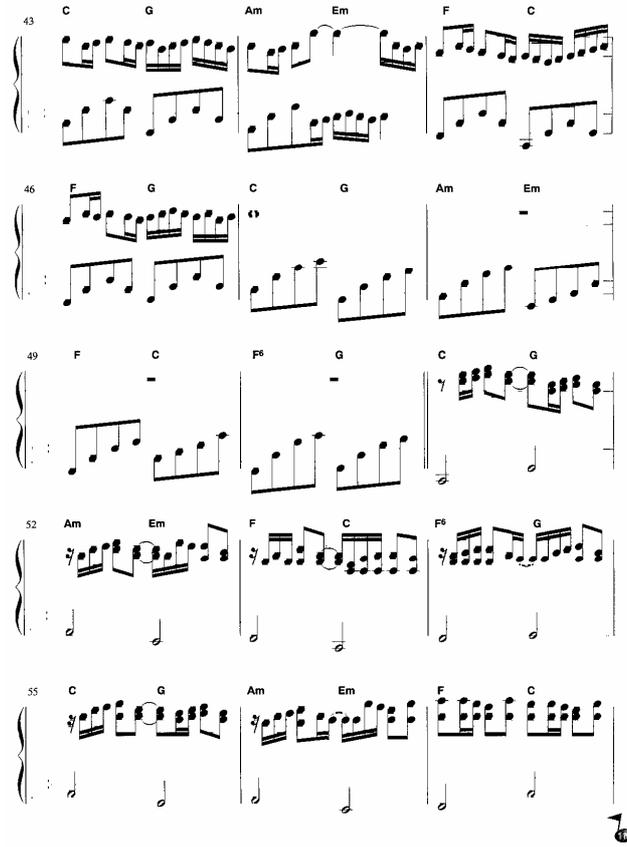
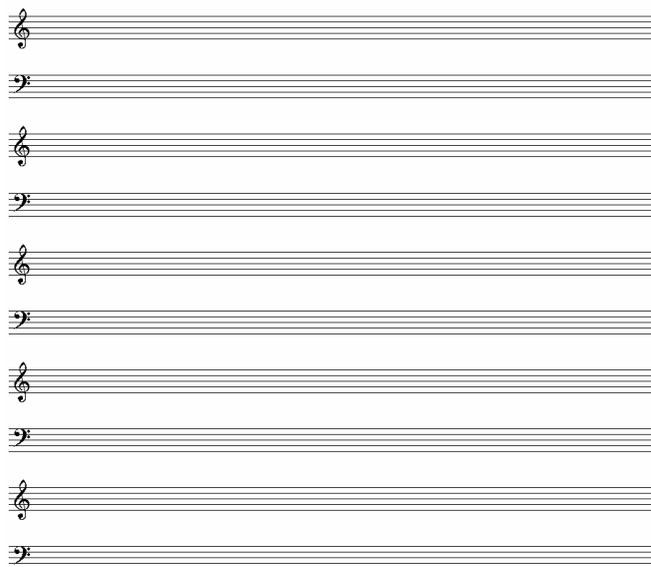


圖 3.15 辨識流程圖

由於譜號、調號與拍號資訊影響整張樂譜大部分音符音高、旋律...等，且譜號含有許多黑色像素叢集處，容易在投影時產生高鋒值造成辨識上不必要之混淆，因此，首先辨識出譜號並清除之(圖 3.16)。



(a)



(b)

圖 3.16 譜號辨識

緊接著對譜表以投影法擷取出垂直線段 x 軸位置，此時，可決定聲部個數，並在譜號右側確認音樂調號與拍號(圖 3.17)。

The image displays a musical score snippet with five systems of music. Each system consists of a grand staff (treble and bass clefs) and a line of chord symbols above it. The chord symbols are: C, G, Am, Em, F, C (System 1); F, G, C, G, Am, Em (System 2); F, C, F#, G, C, G (System 3); Am, Em, F, C, F# (System 4); C, G, Am, Em, F, C (System 5). Vertical red lines are drawn through the score, marking the boundaries of musical phrases. These lines are positioned at the beginning of each measure and at the end of each phrase. The score is numbered 43, 46, 49, 52, and 55 at the beginning of each system. The key signature is one flat (B-flat), and the time signature is 4/4. The score ends with a double bar line and a repeat sign.

圖 3.17 垂直線段擷取(標示為紅色線段)並確認音樂調號與拍號

垂直線段擷取後，由特徵最顯著之實心符頭開始辨識，其黑色像素厚度為譜線間距與二倍譜線厚度和，為其他垂直線段音樂符號所沒有，此外，若出現和絃，同一符桿上將有數個符頭。待實心符頭偵測完，針對這些符桿分析符尾及符衍(圖 3.18)。

(a)

(b)

圖 3.18 實心符頭辨識與符尾及符衍分析

二分音符亦可能出現和絃，但由於二分音符不附有符衍與其它音符集結，所以與實心符頭之符桿不同，不會出現較短符桿，因此，辨識空心符頭時，為增加辨識準確度，須屏除較短垂直線段(圖 3.19)。

43 C G Am Em F C

46 F G C G Am Em

49 F C F⁶ G C G

52 Am Em F C F⁶ G

55 C G Am Em F C

(a)

(b)

圖 3.19 空心符頭辨識

垂直線段若為小節線或四分休止符所形成，上下界會處於固定位置，再加以其型態特徵分析，不易造成含糊不清而辨識為其他音樂符號，而剩餘垂直線段，較短者可能為變音記號或裝飾音，將其辨識出來並在輸入影像上清除(圖 3.20)。

上述辨識步驟完成後，遺留之垂直線段可能為未辨識出之小節線、符桿或其他音樂符號，這部份將留待樂理粹鍊階段再行進一步確認。

(a)

(b)

圖 3.20 垂直線段音樂符號辨識後結果

至於未形成垂直線段之音樂符號，藉著其特有位置、尺寸、型態，先辨識特徵較明顯之全休止符與二分休止符(圖 3.21)，再對八分休止符與十六分休止符偵測(圖 3.22)。

Figure 3.21 illustrates the identification of whole and half rests in musical notation. The top section shows five systems of musical notation, each with chord symbols above the notes. The rests are highlighted with a vertical line segment. The bottom section shows a full musical score with multiple staves and rests.

圖 3.21 全休止符與二分休止符辨識

43 C G Am Em F C

46 F G C G Am Em

49 F C F# G C G

52 Am Em F C F# G

55 C G Am Em F C

Detailed description: This block shows a guitar score with five systems. Each system consists of a treble clef staff with a melody line and a bass clef staff with a bass line. Above each system, chord names are written in black text. The chords are: System 1: C, G, Am, Em, F, C; System 2: F, G, C, G, Am, Em; System 3: F, C, F#, G, C, G; System 4: Am, Em, F, C, F#, G; System 5: C, G, Am, Em, F, C. The notation includes various rhythmic values and accidentals.

(a)

Detailed description: This block shows a guitar score with five systems, focusing on rhythmic notation. Each system consists of a treble clef staff and a bass clef staff. The notation includes various rhythmic values, rests, and accidentals. The first system shows a complex rhythmic pattern in the treble staff. The second system shows a similar pattern. The third system shows a pattern with a slash and a vertical line, indicating a specific rhythmic figure. The fourth and fifth systems show patterns with slashes and vertical lines, indicating specific rhythmic figures.

(b)

圖 3.22 八分休止符與十六分休止符辨識

由於附點尺寸較小，可能因樂譜印刷模糊或掃描失真等因素與雜訊產生模稜兩可，因此本研究將其留至最後一步驟，此時所有音樂符號完成且已清除，樂譜資訊最為單純，圖 3.23 為完成附點辨識最後影像。

(a)

(b)

圖 3.23 最後步驟 — 附點辨識

本研究在辨識階段中並未偵測全音符，因此在下一小節樂理淬鍊階段中，將針對於不合樂理之小節(measure)，以及尚未進行之全音符辨識做處理，詳細情形闡述如下。

3.2.4 樂理淬鍊

樂譜在每一小節中，所有音符總拍數必須符合拍號，本階段將以此樂理檢驗上一小節辨識結果，針對未辨識到之音樂符號，如圖 3.23 實線箭頭所示，或辨識錯誤之處，如圖 3.23 虛線箭頭所示，進行樂理淬鍊。此階段所作辨識將對輸入樂譜傾斜修正後之影像處理，辨識時，則依拍數誤差調整上一小節辨識時所訂型態相似閾值。

當小節總拍數不足時，針對可能誤失之音樂符號偵測，若辨識到音符時，則需同時確認是否存在符尾、臨時記號或附點...等相關音樂符號；當小節總拍數超過拍號時，分析是否遺漏小節線，否則調高致使拍數增加之符號相似閾值，如：空心符頭、全休止符、二分休止符、附點...等，且調低致使拍數減少之閾值，如：符尾，再進行一次偵測。

經上述程序數次修正後，若小節總拍數仍小於拍號時，於該小節末端加入所缺拍數休止符，使旋律誤差在一小節內。圖 3.24 為圖 3.23 經樂理淬鍊之結果。

43 C G Am Em F C

46 F G C G Am Em

49 F C F⁶ G C G

52 Am Em F C F⁶ G

55 C G Am Em F C

Detailed description: This block shows a guitar score with five systems of music. Each system consists of a treble clef staff with a melody line and a bass clef staff with a bass line. Above each system, a set of chords is provided. The chords are: System 1: C, G, Am, Em, F, C; System 2: F, G, C, G, Am, Em; System 3: F, C, F⁶, G, C, G; System 4: Am, Em, F, C, F⁶, G; System 5: C, G, Am, Em, F, C. The music is in a key with one flat (F major or D minor) and a 4/4 time signature. A double bar line is present at the end of the fifth system, followed by a repeat sign.

(a)

Detailed description: This block shows a piano score with five systems of music. Each system consists of a treble clef staff and a bass clef staff. The music is in a key with one flat (F major or D minor) and a 4/4 time signature. The first system shows a complex melody in the treble and a bass line. The second system continues the melody and bass line. The third system features a more active treble line with many sixteenth notes and a steady bass line. The fourth and fifth systems show a similar pattern of active treble lines and steady bass lines. The score ends with a double bar line.

(b)

圖 3.24 圖 3.23 樂理粹鍊

3.3 格式轉換

光學樂譜經過本系統前置處理與符號辨識階段已辨識出各音樂標誌，但由於轉換成音樂語言時，僅記錄音符與休止符本身音高、時值...等資訊，因此需先將各音樂符號其音樂涵義詮釋出來，例如：音符音高須同時考慮調號與臨時變音記號，音符、休止符時值會因附點或符尾有無而不同。

本論文使用之音樂語言為 Hoos 與 Hamel 所研發 Guido 音樂標誌格式[29]，主要記錄方式如下，表 3.11 節錄原圖 3.1 樂譜前九小節轉換成音樂語言之結果。

音符：音名 變音記號 音階 * 時值 複點，如：d#2*1/4. ， c-1*3/8 ...等，

休止符：_*時值 複點，如：_*1/4.，_*1/2. ...等，

其餘：\符號名稱<“參數”>，如：\clef<“c”>，\meter<“4/4”>，\key<“D”> ...等，

複合：1) 和絃：{音符 1,音符 2,音符 3, ...}，如：{c0*1/4, e0*1/4, g0*1/4} ...等，

2) 倚音：\grace<時值>(音符 1 音符 2...)，如：\grace<32>(f0*1/4) ...等。

表 3.11 原圖 3.1 樂譜前九小節之音樂語言

```

{[ \clef<"treble"> \key<0> \meter<"4/4">
e2*1/8 c2*1/16 d2*1/16 e2*1/8 d2*1/16 c2*1/16 d2*1/16 h1*1/16 c2*1/16
d2*1/16 e2*1/16 d2*1/16 c2*1/16 h1*1/16 | c2*1/8 a1*1/16 h1*1/16 c2*1/8
g2*1/8 g2*1/4 g2*1/16 c2*1/16 h1*1/16 c2*1/16 | a1*1/8 c2*1/16 h1*1/16
a1*1/8 g1*1/16 e1*1/16 g1*1/16 e1*1/16 e1*1/16 f1*1/16 g1*1/16 a1*1/16
h1*1/16 c2*1/16 | a1*1/8 c2*1/16 h1*1/16 c2*1/8 h1*1/16 a1*1/16 h1*1/16
c2*1/16 d2*1/16 c2*1/16 h1*1/16 c2*1/16 a1*1/16 h1*1/16 | c2*1/1 |
_*4/4 | _*4/4 | _*4/4 \bar<2> _/8 {e2*1/16, c2*1/16} {f2*1/16, d2*1/16}
{g2*1/8, e2*1/8} {e2*1/8, c2*1/8} {e2*1/8, c2*1/8} {h1*1/16, g1*1/16}
{c2*1/16, a1*1/16} {d2*1/8, h1*1/8} {h1*1/8, g1*1/8} ],
[ \clef<"bass"> \key<0> \meter<"4/4">
c0*1/8 g0*1/8 c1*1/8 g0*1/8 g-1*1/8 d0*1/8 g0*1/8 d0*1/8 | a-1*1/8
e0*1/8 a0*1/8 c0*1/16 d0*1/16 e0*1/16 f0*1/16 e0*1/16 d0*1/16 e0*1/4 |
e-1*1/8 c0*1/8 f0*1/8 c0*1/8 h-2*1/8 g-1*1/8 c0*1/8 g-1*1/8 | f-1*1/8
c0*1/8 f0*1/8 c0*1/8 g-1*1/8 d0*1/8 g0*1/8 d0*1/8 | c0*1/8 g0*1/8
c1*1/8 e1*1/8 g-1*1/8 d0*1/8 g0*1/8 h0*1/8 | a-1*1/8 e0*1/8 a0*1/
c1*1/8 e-1*1/8 g-1*1/8 h-1*1/8 e0*1/8 | f-1*1/8 c0*1/8 f0*1/8 a0*1/8
c0*1/8 e0*1/8 g0*1/8 c1*1/8 | f-1*1/8 d0*1/8 a0*1/8 c1*1/8 g-1*1/8
d0*1/8 g0*1/8 h0*1/8 \bar<2> c-1*1/2 g-1*1/2 ]]

```

本論程式介面如圖 3.25 所示，視窗分割成兩框架，上方框架為傾斜修正後影像，執行辨識時會圖示出每個步驟對樂譜所做處理，下方框架同時記錄偵測得到之音樂符號或標誌。辨識完成時，程式自動顯示樂譜資訊，包括調號、傾斜修正角度、譜表數、聲部個數、譜線間距、譜線厚度與符頭寬度。而進行格式轉換之前，可調整演奏速度和樂器，當辨識樂譜無記載拍號或辨識錯誤時，亦可人為設定以利樂理粹鍊(圖 3.26)。



圖 3.25 程式介面

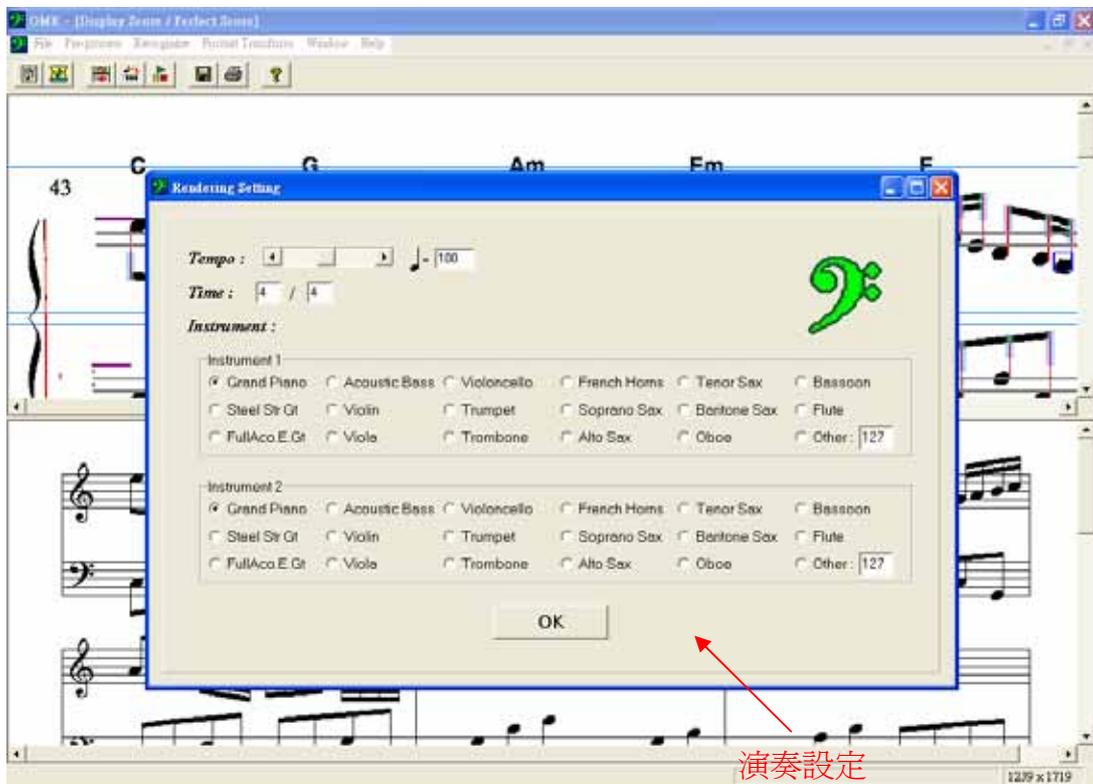


圖 3.26 演奏設定

最後，將 Guido 音樂標誌格式轉換成電子音樂 MIDI 檔是以 Martin 與 Hoos 撰寫之 gmntomidi 程式[30]執行，此電子音樂檔可使用任何電子音樂合成器播放出來。本論文將上述程式鑲在本程式中，故自載入樂譜後，由辨識至播放均可在本程式下直接執行(圖 3.27)。

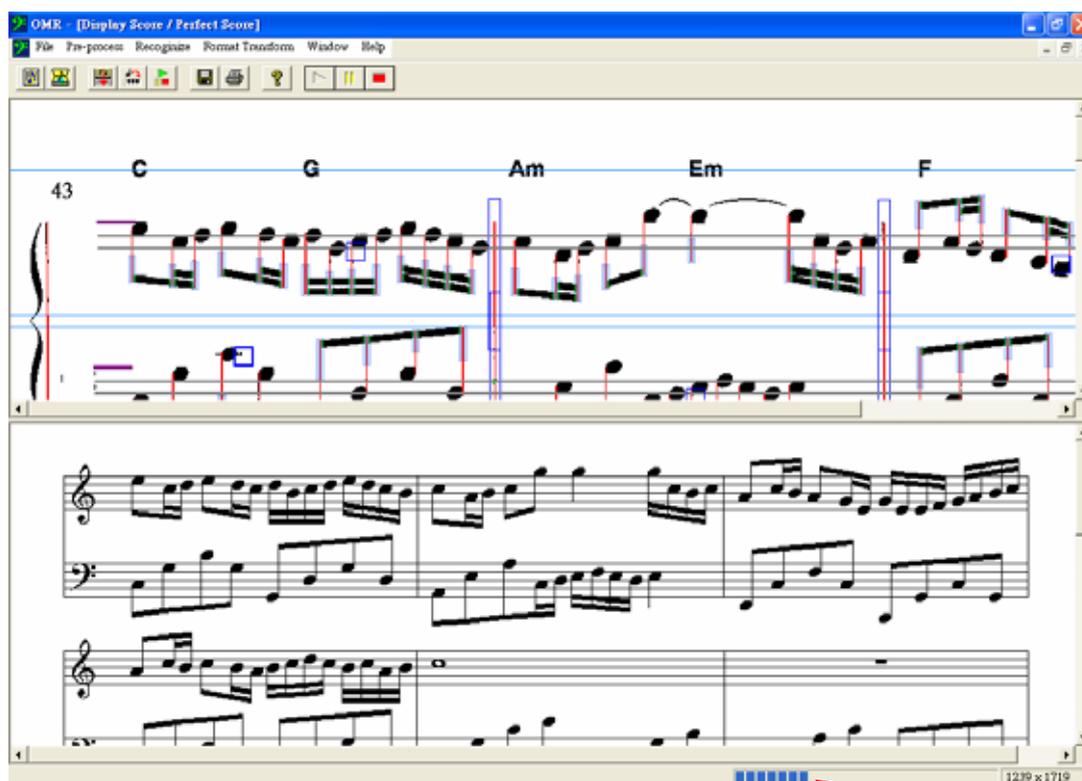


圖 3.27 播放電子音樂 MIDI 檔

播放進行中

第四章 實驗與討論

4.1 結果比較

儘管目前已有許多 OMR 相關商業軟體，例如：PIANOSCAN、NOTESCAN、MIDISCAN、PhotoScore、SmartScore 與 SharpEye...等等，但是仍然無一具有完美辨識能力，因此使用者通常被要求在高解析度、樂譜形式一定...等限制下，才有可能產生近乎百分之百辨識度。此外，由於商業軟體須兼顧速度與辨識能力，因此通常不會確認與修正辨識結果，而僅是快速將輸入樂譜轉換成音樂表示語言以供儲存或電子音樂檔播放出來，至於辨識錯誤部份，需藉助軟體所附編輯功能自行將其改正過來。

選擇以 PhotoScore 商業軟體為測試對象，此軟體可自其公司網站[31]免費取得試用版本，並以圖 4.1(樂譜選自家後[32])作為輸入影像，解析度設為此軟體要求下限 200dpi。比較此商業軟體與本研究所提方法之辨識結果，如圖 4.2 與圖 4.3 所示，其中，圖 4.2 為商業軟體辨識結果，而圖 4.3 為本研究程式辨識結果。

28 **Em7** **Dm7**
 少年 跟你 跟甲 老 人 情 世 事 已 經 看 透
 就 跟 你 跟 甲 老 人 人 情 世 事 嘛 已 經 看 透

31 **Dm7** **Cmaj7** **B \flat /C** **C**
 透 透 有 有 啥 人 比 你 卡 重 要
 透 透 有 有 啥 人 比 你 卡 卡 重 重 要 要

34 **Fmaj7** **Em7** **3**
 阮 的 一 生 獻 乎 恁 兜 才 知 幸 福 是 吵 吵 鬧

37 **Em7** **Dm7**
 鬧 等 待 返 去 的 時 陣 若 到 我 會 會
 鬧 等 待 返 去 的 時 陣 若 到 我 你 會

40 **C** **1. Dm7**
 讓 你 先 走 走 因 為 我 會 唔 甘
 讓 我 先 走 因 為 我 會 唔 甘



圖 4.1 測試樂譜(家後)

程式自動判斷之拍號(正確為 4/4)

符號辨識錯誤

受歌詞影響而辨識錯誤

音高錯誤

(少辨識出降記號)

圖 4.2 PhotoScore 商業軟體辨識圖 4.1 結果



圖 4.3 本研究程式辨識圖 4.1 結果

使用此商業軟體辨識前，僅管已將辨識條件設為最嚴謹，包括歌詞、文字、連奏符號...等皆不做辨識考慮，但是結果仍然無法成功避開歌詞干擾，如圖 4.2 虛線箭頭所示，此外，右方實線箭頭為符號辨識錯誤之處，左方實線箭頭為其自動判斷之拍號，正確為 4/4；而本論教程式所得結果並未產生錯誤。

再以相同條件辨識另一樂譜，如圖 4.4(樂譜選自親愛的你怎麼不在我身邊 [33]) 所示，其測試結果呈現於圖 4.5(此商業軟體辨識)與圖 4.6(本研究程式辨識)。

親愛的你怎麼不在身



~ 演奏重點提示 ~

本曲主要練習左手add2的分解和弦及琶音式的分解和弦應用，並注意轉指的地方。

詞：姚 謙
 曲：陳 小 霞
 唱：江 美 琪

♩ = 80

1 **F** **Em7**

mf

4 **Em7** **Dm7**

7 **F/G** **G** **C**
 這裡的空氣

10 **C** **Am**
 很新鮮 這裡的小吃 很特別

90

圖 4.4 測試樂譜(親愛的你怎麼不在我身邊)

The image displays four systems of musical notation, each consisting of a treble clef staff and a bass clef staff. The notation includes various notes, rests, and accidentals. A red arrow points to a note in the second system of the fourth system, with the text "少辨識出八分休止符" (少辨識出八分休止符) written above it.

圖 4.5 PhotoScore 商業軟體辨識圖 4.4 結果



圖 4.6 本研究程式辨識圖 4.4 結果

分析辨識錯誤之處，圖 4.5 箭頭指示出少一八分音符，此小節與樂譜拍號不符，證明此商業軟體並未進行辨識結果驗證，而圖 4.6 箭頭所指處指示出少一八分音符，原因為原影像該處失真而未擷取到垂直線段，圖 4.7 為其放大圖，此因素導致整個音符未被辨識出來。另外，該小節後方之八分休止符為樂理粹鍊時所強制加入，用來使旋律誤差達到最小。

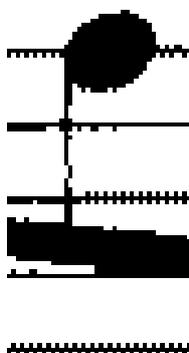


圖 4.7 影像失真而未擷取到垂直線段

4.2 討論

除了上小節提及情況會使本研究程式產生錯誤外，下列情形亦可能導致辨識結果與輸入樂譜不符：1)小節線未偵測到或多偵測得，使樂理粹鍊時產生錯誤分析；2)前置處理階段之五線譜偵測未能正確擷取得譜表位置；3)對於音符音高判斷錯誤，或和絃中某個樂音偵測遺漏，本研究無法進行修正；4)樂譜中含有其它文中未提及之音樂符號。

若欲加強本研究辨識能力，方法可朝三方向著手：1)改善辨識方法，尤其音樂符號連通性被破壞之問題；2)提高輸入影像品質，包括取適當解析度掃描、儘量避免傾斜、樂譜印刷清晰...等；3)引入更多樂理限制，如：各音樂符號相對位置、常用和弦之音程關係...等。

第五章 結 論

樂譜辨識主要為一對音樂符號編碼問題。由商業軟體之辨識情形可知，相關技術上有許多改善空間。

本研究提出之方法分成三大階段，首先為前置處理，針對掃描影像之傾斜做角度修正，以利五線譜偵測與移除；在偵測五線譜同時，可以求得五線譜譜線間距與譜線厚度，此為辨識階段重要依據。辨識階段，採用投影法擷取具垂直線段之音樂符號，亦可縮小辨識範圍，對於非垂直線段音樂符號則採型態法偵測；接著，引入樂理，將音樂符號組合成音樂標誌與進行樂理淬鍊。最後，將所有音樂標誌記錄成音樂表示語言，可供建立音樂資料庫，或用較少記錄空間，亦可再經由程式轉換成電子音樂 MIDI 格式，藉由播放軟體將旋律播放出來。

本研究程式在硬體設備與程式皆未最佳化之環境下進行測試，剔除五線譜偵測失敗情形，對於一張 A4 尺寸、150dpi 解析度且包含十組譜表之二值樂譜影像，可在一分鐘內獲得辨識結果，證明本研究提出之方法具可行性。

參考文獻

- [1] D. Pruslin, “Automatic recognition of sheet music,” Sc.D. Dissertation, Massachusetts Institute of Technology, 1966.
- [2] M. Eden, “Other pattern recognition problems and some generalizations,” *Recognizing Patterns*, edited by M. Eden and P. Kolars, MIT Press, 1968.
- [3] D. S. Prerau, “Computer pattern recognition of printed music,” *AFIP Conference proceedings*, Vol.39 1971 fall joint computer conference, pp. 153-62, 1971.
- [4] A. Andronico and A. Ciampa, “On automatic pattern recognition and acquisition of printed music,” *Proceedings of the ICMC*, pp. 245-78, 1982.
- [4] N. P. Carter, “Automatic recognition of printed music in the context of electronic publishing,” Ph.D. thesis, University of Surrey, 1989.
- [5] S. Inokuchi, and H. Katayose, “Computer and music,” *Journal of the IEICE*, 73(9), pp. 965-7, 1990.
- [6] H. Miyao and Y. Nakano, “Head and stem extraction from printed music scores using a neural network approach,” *Proceedings of the 3th ICDAR*, pp. 1074-9, 1995.
- [7] K. C. Ng and R. D. Boyle, “Recognition and reconstruction of primitives in music scores,” *Image and Vision Computing*, 14(1), pp. 39-46, 1996.
- [8] T. Hori, S. Wada, T. Howzan, S. Y. D. E. Kung and W. Bastiaan Kleijn, “Automatic music score recognition/play system based on decision based neural network,” *1999 IEEE Third Workshop on Multimedia Signal Processing*, Cat. No.99TH8451, pp. 183-4, 1999.
- [9] D. Blostein, and H. S. Baird, “A critical survey of music image analysis,” H. S. Baird, H. Bunke, and K. Yamamoto, editor, *Structured Document Image Analysis*. Springer-Verlag, 1992.
- [10] K. T. Reed and J. R. Parker, “Automatic computer recognition of printed music,” *Proceedings of the ICPR*, pp. 803-7, 1996.
- [11] E. Sicard, “An efficient method for the recognition of printed music,”

Proceedings of the 11th IAPR, pp. 573-6, 1992.

- [12] Y. C. Chung, "Recognition of Printed Sheet Music Using Hough Transform And Morphology Operations," Master thesis, National Taiwan Normal University, Taiwan, 1995.
- [13] R. O. Duda and P. E. Hart, "Use of the Hough transform to detect lines and curves in pictures," *Commun. ACM*, vol 15 No 1, pp. 11-15, 1972.
- [14] F. Kimura and M. Shridha, "Handwritten numerical recognition based on multiple algorithms," *Pattern Recognition*, 19(1), pp. 1-12, 1986.
- [15] R. J Randriamahefa, J. P. Cocquerez, C. Fluhr, F. Pépin, and S. Philipp, "Printed music recognition," *Proceedings of ICDAR*, pp. 898-901, 1993.
- [16] N. P. Carter, R. A. Bacon, "Automatic Recognition of Printed Music," In H. S. Baird, H. Bunke, and K. Yamamoto, editor, *Structured Document Image Analysis*. Springer-Verlag, Heidelberg, 1991.
- [17] B. Modayur, "Music score recognition – a selective attention approach using mathematical morphology," Technical Report, Electrical Engineering Department, University of Washington, Seattle, 1996.
- [18] B. Couasnon, B. Rétif, "Using a grammar for a reliable full score recognition system," *Proceeding of ICMC*, Ban., Canada, pp. 187–194, 1995.
- [19] D. Blostein, L. Haken, "Using diagram generation software to improve diagram recognition: a case study of music notation," *IEEE Trans. on PAMI*, 21(11), pp. 1121–1135, 1999.
- [20] K. C. Ng and R. D. Boyle, "Recognition and reconstruction of primitives in music scores," *Image and Vision Computing*, 14(1), pp. 39-46, 1996.
- [21] I. Fujinaga, "Adaptive Optical Music Recognition," Ph. D. thesis, McGill University, 1996.
- [22] F. Rossant, "A global method for music symbol recognition in typeset music sheets," *Pattern Recognition Letters*, 23(10), pp. 1129-41, 2002.
- [23] J. Serra, "Image Analysis and Mathematical Morphology," Academic Press, London, 1982.

- [24] J. Serra, "Image Analysis and Mathematical Morphology: Theoretical Advances," Academic Press, London, 1988.
- [25] 紀華麟, "流行輕鬆彈," 樂匠音樂出版社, p. 113, 2003.
- [26] 施富然, "Old Jazz 鋼琴精選 No.1," 卓著出版社, p. 33, 2003.
- [27] Kenny G, "Spanish Nights," p. 29, Hal Leonard, 2002.
- [28] 羅炳良, "香港聲樂作品集," 基督教文藝出版社, p. 67, 1999.
- [29] H. H. Hoos, and K. A. Hamel, "The GUIDO music notation format version 1.0, Specification Part 1: Basic GUIDO," Technical Report TI 20/97, Technische Universität Darmstadt, 1997.
- [30] L. Martin, and H. H. Hoos, "gmn2midi version 1.0," Computer Program, 1997.
<<http://www.informatik.tu-darmstadt.de/AFS/GUIDO/>>.
- [31] Neuratron Limited, "PhotoScore Demo Version," Commercial Package, 2003.
<<http://www.neuratron.com/>>.
- [32] 紀華麟, "流行輕鬆彈," 樂匠音樂出版社, p. 8, 2003.
- [33] 紀華麟, "流行輕鬆彈," 樂匠音樂出版社, p. 90, 2003.