

中華大學資訊工程學系
專題期末報告

A Paper Presented to the Project Implementation Course
Department of Computer Science and Information Engineering
Chung Hua University

MP3 國語流行音樂自動斷句
與詞曲同步撥放系統

程弘錡、張智凱、徐詩媛、蕭秀婉
B09702011、B09702103、B09702022、B09702058

指導教授：劉志俊 博士
Advisor: Chih-Chin Liu, Ph.D.

中華民國 100 年 10 月
October 2011

摘要

MP3是目前數位音樂的主流格式，但許多歌曲欠缺詞曲同步資訊。而提供MP3播放流行音樂歌詞同步播放的功能，其詞曲同步資訊是由人工費時編輯而成。為改善此一缺失，本文提出一種根據IEEE1599規範，整合MP3的歌曲訊號，依照MP3樂曲起音點偵測，找出樂曲的起音點，再藉由樂曲分段技術，以呼吸聲偵測技術來進行音樂樂句的自動分段。並利用IEEE1599標準定義的XML語法，配置相對應的歌詞，整合顯示相關異質資訊，儲存自動MP3樂句/曲譜樂句分段以及同步資訊，達到自動化的目的。

在Windows環境下；以C#開發為一套國語流行音樂同步、自動斷句之播程式之上，此播程式能夠在撥放樂曲的同時，同步顯示其對應的歌詞、樂譜以及其他相關的音樂異質資訊，讀取IEEE 1599所整合的MP3 歌曲、曲譜與歌詞，並且解析其中各種異質資訊資料，包含MP3音樂檔案、Midi檔案、同步化歌詞、MusicXML曲譜、音樂斷句狀況……等。系統中包含歌曲的標題、歌手、作詞作曲者等的音樂相關資料，專輯或歌手的圖片，曲譜的資訊，如希望觀賞音樂MV，或是其他相關影音，也可由本系統中直接點選，甚至於撥放MP3音樂時，能夠同步撥放歌詞、音樂樂譜，提供音樂聆聽者一套MP3國語流行音樂自斷斷句以及詞曲同步撥放系統。

為驗證此一程式之改善效率，本文採用的 101 首 MP3 國語流行音樂進行樂理群組偏好法以及呼吸聲偵測分段法測試，計算回覆率以及準確率，並且研究最佳的音樂斷句方式。依據實驗的結果顯示，樂理群組偏好法則有將近 70%的準確率，經過呼吸聲偵測的分段加強驗證後接近於 80%，使用這二種分段法則可改善傳統樂理群組偏好法則的樂句分段方法的效能。

MP3 國語流行音樂自動斷句與詞曲同步播放系統可以讓使用者根據音樂整合檔所記載之內容，瀏覽該首歌曲之各種異質資訊，包含演唱者、作詞者、作曲者、專輯發行日期、專輯封面、歌手寫真、歌手其他作品……等資訊，播放 MP3 檔案時可以藉由同步播放功能，觀看目前歌曲歌詞、查閱樂譜播放資訊、進行樂句分析。

本程式容易受背景音樂干擾，為其缺點。如何在伴奏音樂下偵測到呼吸聲，及如何整合多種不同的樂句的分段點判斷方法，是未來研究探討的方向。

關鍵詞：MP3 斷句、呼吸聲偵測、樂句分段、IEEE 1599、音樂內涵分析。

目錄

摘要.....	0
Chapter 1 前言.....	7
Chapter 2 相關研究.....	11
Chapter 3 系統架構.....	12
Chapter 4 MP3 樂句與 IEEE1599 標準.....	15
Chapter 5 曲譜樂句自斷分段.....	23
Chapter 6 MP3 自動斷句.....	26
Chapter 8 系統實作.....	30
Chapter 9 實驗結果.....	35
參考文獻.....	46
Appendix A - 數位化國語流行音樂 MusicXML 樂譜之建立.....	50
Appendix B - 國語流行音樂樂曲資料庫之建構.....	55
Appendix C - 資料剖析技術.....	60

圖目錄

圖 1 MP3 國語流行音樂自動斷句與詞曲同步撥放系統架構圖	12
圖 2 樂譜中樂句的標註	15
圖 3 IEEE 1599 六層架構.....	16
圖 4 歌詞範例	17
圖 5 IEEE 1599 音樂檔一般格式.....	17
圖 6 專輯封面範例	18
圖 7 曲譜範例	18
圖 8 IEEE 1599 一般層格式描述.....	20
圖 9 樂句事件建立	20
圖 10 MP3 樂句事件同步資訊	21
圖 11 曲譜同步資訊	22
圖 12 音字符串列(NOTE LIST)	30
圖 13 METER 轉換.....	31
圖 14 GROUPER 轉換	31
圖 15 國語流行音樂自動斷句系統的樂句分段結果	32
圖 16 音樂播放模式	33
圖 17 樂句分析模式	33
圖 18 樂譜樂句顯示模式	34
圖 19 流行音樂架構圖	36
圖 20 電腦判斷樂句的片段資訊	37
圖 21 電腦判斷的樂句位置圖	38
圖 22 音樂播放器的歌詞顯示功能	50

圖 23 程式開檔畫面	51
圖 24 歌曲修正方式	51
圖 25 常見問題	52
圖 26 解決方式	53
圖 27 修正完成	53
圖 28 XML 檔案編輯完成	54
圖 29 輸入歌詞於 MUSICXML	55
圖 30 MSDN 之 XMLTEXTREADER	60
圖 31 查詢 XML 元素	61
圖 32 解析 MUSICXML 重要元素	61
圖 33 解析 MUSICXML 結果圖	62

表格目錄

表格 1 一百零一首樂曲之實驗結果統計	42
表格 2 二十一首哼唱歌曲之實驗結果統計	43

Chapter 1 前言

音樂是現代人生活中不可或缺的元素之一，音樂的種類更是包羅萬象，例如：古典樂、流行樂、爵士樂等。隨著電腦興起，音樂資訊儲存媒介從早先黑膠唱片至卡帶、CD，演進為現今數位線上儲存及播放的時代。音樂資訊儲存也不再僅限於音訊訊號，現今國內用戶喜愛的線上音樂服務，如 KKBOX，擁有包含歌曲下載、各種音樂媒體相關資訊查詢、流行音樂歌詞查詢等功能。最新的 MP3 播放程式，如使用者愛用的千千靜聽等，提供包含歌曲試聽、流行音樂歌詞資料庫搜尋、流行音樂與歌詞同步撥放功能。而如 iTunes 等大型數位音樂服務更是提供線上購買歌曲、隨身聽及手機連接、音樂資料庫共享檢視等功能。可見對流行音樂而言，一份完整音樂資料是整合 MP3 聲音訊號、樂譜、歌詞、歌手資料、作曲資料、專輯圖片等多種異質音樂資料。因此，整合各種音樂異質資訊的機制就顯得相當重要。

音樂分析的基本構成單位是樂句(*phrases*)，一首樂曲可視為由一連串樂句所組合而成。以往音樂缺乏自動斷句技術，只能動用人工費時並反覆聆聽方式，一一標記每個樂句的開始與結束時間，才能將音樂訊號、曲譜與歌詞做樂句同步結合。如果能將一首 MP3 數位樂曲自動切割成一連串的樂句，便可自動達到數位樂音與歌詞同步化，將傳統 KTV 歌詞播放系統、流行音樂與歌詞同步撥放功能改為全自動化，減少人工斷句的繁複工作負荷。此外，在學術研究方面，自動偵測一首歌曲中的樂句分段位置，是音樂索引、音樂內涵查詢、音樂曲式自動分類等相關研究所需的基本技術。

以往在樂句分段方面的研究中，最著名的是根據調性音樂生成理(GTTM)[22]中的群組傾向規則(GPR, grouping preference rules)所提出的樂句邊界偵測方法[14]。此方法依據連奏、休止符、長起音點間距、音程變化、力度變化、發音方式改變、以及音長變化來估計發生樂句邊界的可能性。

Temperley 在[31]一書中詳細探討有關樂音群組規則在感知心理學與音樂理論的根據與發展沿革，提出一套樂句結構傾向規則(PSPR, phrase structure preference rules)，藉由考慮樂音群組間隙(gap)、樂句長度、以及韻律對句(metrical parallelism)來計算樂句分段點的可能性。針對 65 首德國民謠的樂句分段實驗結果的回復率為 75.5%，而準確率為 74.6%

在[3][4]等論文中，Cambouropoulos 提出一種稱為區域邊界偵測模型(LBDM, local boundary detection model)的方法來偵測樂句的分段點。LBDM 方法根據音高、起音點間距、以及休止符的變化來估計樂句分段點發生的可能性。實驗結果顯示對 498 個樂句的分段點的偵測可以達到 63%的回覆率與 55%的準確率。而針對蕭邦離別曲的實驗結果顯示，在斷句點發生前第 2 個音有加長的傾向，此規則有助於改善樂句分段點的判讀。

Friberg 等人在[15]一文中提出對樂譜自動進行標點(musical punctuation)的方法。藉由歸納對特定音樂家對音群或是所謂的旋律樣型(melodical gestures)標註逗號的偏好模式，Friberg 提出 13 種規則來自動偵測在最長 5 個音符的音群中是否存在可能的逗號位置(potential comma locations)。此外，文中也試著訓練類神經網路的技術，來進行可能的逗號位置之自動辨識。實驗以 52 首各個時期的古典音樂進行測試，

結果顯示依照 13 種規則的音群標點準確率為 66%，而回復率為 66%。類神經網路的辨識效果則是明顯較差。

在統計語言學的領域中，n 元語法模型(n-gram model)常被用來預測下一個符號是否歸屬於同一個序列之中。在[26]一文中，Pearce 等人利用 n 元語法模型來進行樂句分段點預測，發展出一個稱為 LDyOM 的樂句分段統計模型。此模型利用音高、起音點間距、以及起音-終音間距來訓練出一個樂句預測機率模型。文中比較了 GPR 法則[14][22]、Grouper[31]、LBDM[3][4]與 LDyOM 等樂句分段方法的效能，並以 1705 首德國民謠資料庫的分句資料進行實驗，結果顯示以 Grouper 與 LBDM 方法的效能較佳，但整合四種方法可以得到最佳效能(準確率 87%，回復率 56%，F 量測 66%)。

記憶模型(memory-based model)是自然語言處理領域常用的分析技術。在[2]一文中，Bod 將記憶模型應用於樂句分段。Bod 主張有些樂句是無法使用完形理論(gestalt theory)[32]為基礎的樂句分段方法來進行偵測的。因此，Bod 提出使用樹庫語法(treebank grammar)、馬可夫語法(Markov grammar)、以及資料導向剖析(data-oriented parsing)等三種統計模型來訓練樂句記憶模型。實驗以 5251 首埃森民謠資料庫中的歌曲進行訓練，再以 1000 首歌曲進行測試，結果準確率 76.6%，回復率 85.9%，F 量測 81%。

Cheng 與 Chew 在[5][6]等論文中提出一種稱為區域最大樂句偵測(LMPD, local maximum phrase detection)的樂句分段方法。LMPD 方法的理論根據是以往在樂句特徵方面的許多研究結果指出，一個樂句的力度變化大致上會呈現出由漸強(crescendo)轉為漸弱(decrescendo)的拱門結構，兩個拱門交界處即是樂句邊界。利用此特性，Cheng 等人定義了樂句

強度(phrase strength)、樂句穩定度(phrase volatility)與樂句典型性(phrase typicality)等衡量公式，來表示某一演奏家在樂句方面的詮釋特性。

以往在樂句分段方面的相關研究，基本上都是基於樂句的群組規則來進行探討。對流行音樂與聲樂等人聲樂曲而言，換氣是樂句的顯著特徵。因此本文提出一種利用換氣聲偵測技術，可針對 MP3 國語流行音樂檔案的音樂自動斷句系統。結合 IEEE 1599[1][2][18]對各種異質音樂資訊的整合能力，可提昇傳統根據樂句群組理論的樂句分段點偵測方法的效能。本文實際建構出一套 MP3 音樂自動斷句系統，具有 MP3 訊號播放，歌詞、樂譜同步[15]顯示功能的程式，提供音樂聆聽者一套更完整的國語流行音樂播程式。

Chapter 2 相關研究

IEEE 1599 標準定義出整合歌曲的異質資訊之 XML 語法與六層架構音樂資訊的描述語法與涵義。六層架構包括最上層的一般音樂描述(general layer)、邏輯層(logical layer)，此兩層架構用以描述各種形式音樂資訊間的參照與同步關係，而結構層(structural layer)描述曲式與結構方面的資訊，譜號層(notational layer)描述樂譜相關的資訊，表演層(performance layer)描述音樂合成資訊，最底層為描述音樂訊號的音訊層(audio layer)。藉由這六層架構，即可讓音樂的資訊做最有效的整合。

我們使用 IEEE 1599 的標準來整合歌曲的 MP3 檔案、MIDI 檔案、動態歌詞、MusicXML 曲譜等異質資訊，利用標準中的脊椎(spines)技術來建構 MP3 樂句、曲譜樂句與歌詞樂句的同步資訊，根據其規範，實做出一套同步播放系統，可以在播放一個樂句的同時，同步呈現其對應之曲譜與歌詞資訊。為了達成 MP3 的自動斷句，本系統分別使用 MP3 樂音的起音點偵測技術以及呼吸聲偵測技術來進行斷點預估。我們取自《Hit101 中文流行鋼琴百大首選》裡的一百首樂曲，以及張智星教授建立之哼唱歌曲資料庫 MIR-QBSH Corpus 中的二十一首哼唱歌曲來進行實驗。

本系統結合 IEEE 1599 系統整合檔案與呼吸聲偵測功能，讓此播放器讀取 IEEE 1599 音樂整合檔的 XML 文件後，會根據演唱者換氣處進行樂句分段，剖析其附註之各種異質音樂資訊，顯示出歌曲的標題、歌手、作曲者、作品名稱以及動態歌詞等關於此樂曲的詮釋資料；專輯的照片等與此歌曲相關的圖片，根據 IEEE 1599 音樂整合檔的參照資訊取出呈現；樂譜的資訊由整合檔的譜號層的樂譜參照取出，能提供使用者一套完整的聽覺與視覺的享受。

Chapter 3 系統架構

本文所提出之 MP3 國語流行音樂自動斷句與詞曲同步撥放系統，是由樂譜資料庫、歌詞資料庫、MP3 歌曲分段、樂曲/樂譜/歌詞同步控制等模組所構成。圖 1 為系統整體架構圖，本文討論的主題範圍在虛線框出的部分。

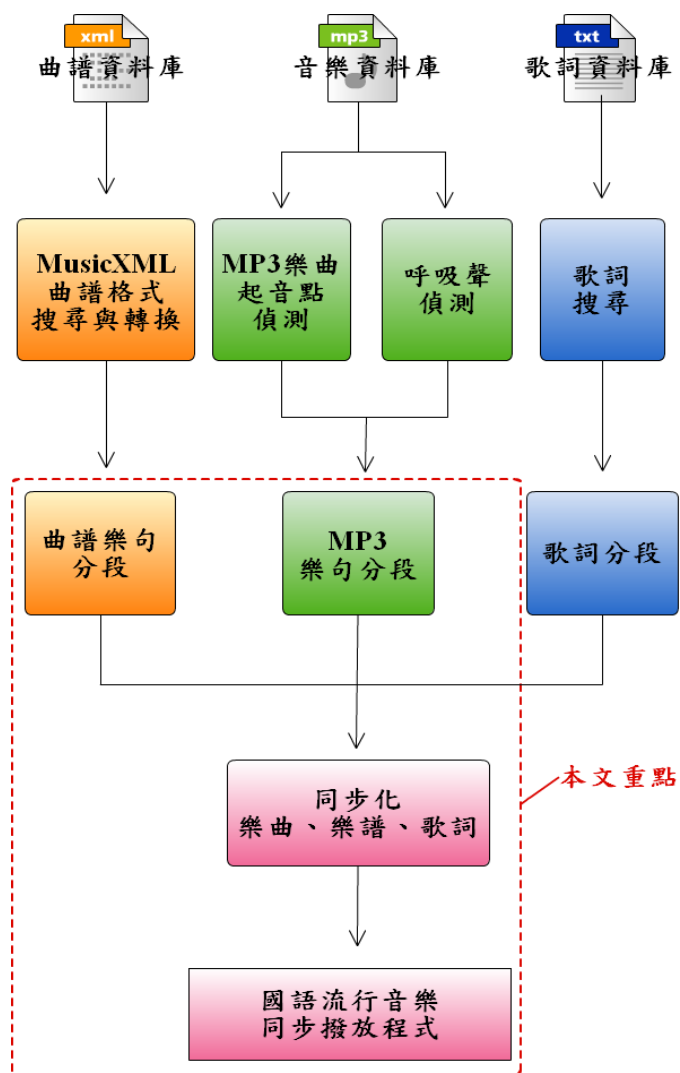


圖 1 MP3 國語流行音樂自動斷句與詞曲同步撥放系統架構圖

- 曲譜資料庫：建造曲譜資料庫是為了整合來自網際網路上各式各樣流行音樂曲譜。曲譜資料庫中部份曲譜是使用五線譜編輯的軟體自行建立而成。由於本文系統同步的單位為樂句，曲譜資訊因此必須使用樂句分段的技術，將一首樂曲資訊切割分成一連串的樂句，進而建立完整的曲譜/樂句資料庫作為音樂查詢的依據。
- 歌詞資料庫：為了建構一個歌詞資料庫，本文收集各種形式的音樂歌詞，除了各大唱片公司發行的紙本歌詞外，主要以網路來源為主。本文整合專門架設歌詞的網站資源，在歌詞自動搜尋方面，採用跟現有的歌詞網站做連結，取得相關的歌詞資訊之後，讓歌詞進行自動斷句。藉由已標記或自動標記的樂句開始時間，來達到歌詞/訊號的同步。之後再把這些同步資訊與 MusicXML 格式樂譜檔案整合，以 IEEE1599 記錄同步資訊。
- MP3 樂句分段：一首 MP3 樂曲、樂譜以及歌詞資訊的同步單位是樂句。本文利用樂句分段的規則，設計出自動計算每一個起音點所會發生樂句分段點的機率值。並根據此機率值的分數大於門限值的設定，將一首 MP3 樂曲切割成一連串的 MP3 樂句，每一段 MP3 樂句包含了多個 MP3 樂音。

- 同步化樂曲、樂譜、歌詞：IEEE1599 標準所定義的 XML 語法，足以整合各種音樂上的曲譜、歌詞或者其他相關之異質資訊(如：演唱者、作曲家、編曲、專輯名稱、發行日期等)，但目前並無任何 MP3 音樂或軟體使用此資訊。本文根據 IEEE1599 規範，整合 MP3 的歌曲訊號，同步化其曲譜及歌詞等內容。同步化主要依照 MP3 樂曲起音點偵測、MP3 樂曲分段此兩大方式來執行。我們利用起音點偵測，找出樂曲的起音點，再藉由樂曲分段技術，配置相對應的歌詞，整合顯示相關異質資訊。
- 國語流行音樂同步撥放程式：本文根據 IEEE1599 規範，實作國語流行音樂同步播程式，此播程式能夠在撥放樂曲的同時，同步顯示其對應的歌詞、樂譜以及其他相關的音樂異質資訊。

Chapter 4 MP3 樂句與 IEEE1599 標準

樂句是音樂分析上的基本單位，如同圖 2 所示，在曲譜中通常以圓滑線標示出一個樂句的範圍。



The image shows a musical score in piano format with Chinese lyrics. A red box highlights a sentence '我的心情是堅固' (My mood is firm) in the first system. A blue box provides a magnified view of this sentence, showing the notes and the chord 'Dm7' above them.

圖 2 樂譜中樂句的標註

本文根據 IEEE 1599 標準的規範，來整合每首 MP3 歌曲、曲譜與歌詞資訊，以及有關 MP3 樂句、曲譜樂句、歌詞樂句的同步音樂資訊。IEEE1599 把音樂分成了六個層面，如圖 3 所示，定義了整合各種樂曲異質資料的 XML 標準與各層次音樂資訊的描述語法，包含最上層的一般層 (general layer) 用以描述音樂的相關資訊、邏輯層 (logical layer) 描述各種形式音符音高等狀況，結構層 (structural layer) 描述歌曲結構的資訊，譜號層 (notational layer) 描述樂譜相關的資訊，表演層 (performance layer) 描述演奏表現的力度、響度、時間等資料、最底層為音訊層 (audio layer)，描述數位音樂訊號。藉由這六層架構，就可以讓音樂的各種資訊做最有效的管理與整合。

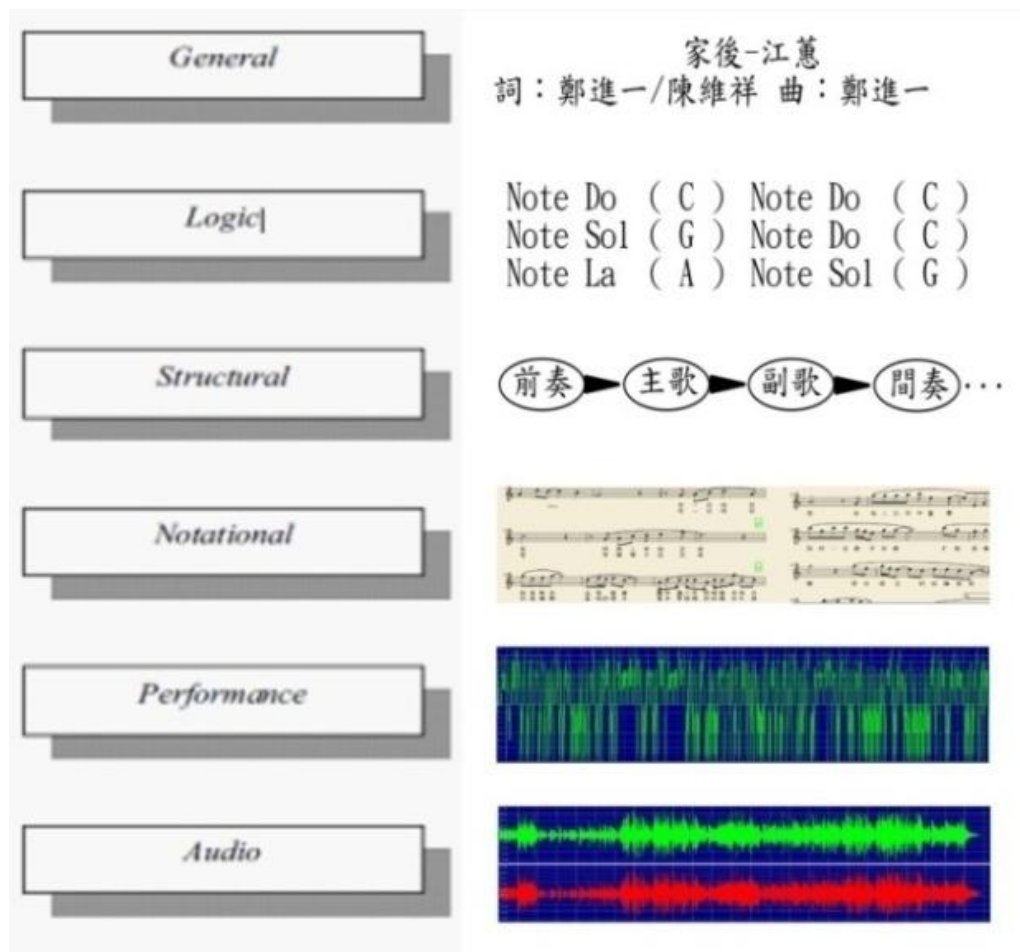


圖 3 IEEE 1599 六層架構

本文以實驗樣本來說明 IEEE1599 的整合架構。以江蕙的「家後」為例，音樂資料可以是「家後」的專輯圖片，或是歌曲 MV 所撥放的影片，因此，一般層中本系統可以描述此歌曲的標題、歌手、作詞者、作曲者、發行日期、曲目編號等相關此歌曲的資料。相關的影片與圖片等檔案的資料如圖 4-7 所示之範例，以<related_files>標記的圖 8 方式整合。

[00:21.58] 有一日咱若老
[00:28.21] 找無人甲咱友孝

圖 4 歌詞範例

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE ieee1599 SYSTEM  
    "http://www.mx.dico.unimi.it/ieee1599.dtd">  
<ieee1599>  
    <general>...</general>  
    <logic>...</logic>  
    <structural>...</structural>  
    <notational>...</notational>  
    <performance>...</performance>  
    <audio>...</audio>  
</ieee1599>
```

圖 5 IEEE 1599 音樂檔一般格式



圖 6 專輯封面範例

家後

鄭進一/陳維祥 鄭進一

有一日咱 若 老 找 無人甲咱 友 孝 我會陪你
 有一日咱 若 老 有 媳婦子兒 友 孝 你若無聊

坐估 椅寮 聽你講少年的時陣 你有外 豪 吃好吃壞 無 計
 拿咱的相片 看卡早結婚的時陣 你外 緣 投 穿好穿醜 無 計

較 怨 天 怨 地 嘛 袂 曉 你 的手 我會 甲 你 牽 條 條 因為
 較 怪 東 怪 西 嘛 袂 曉 你 的心 我會 永 遠 記 條 條 因為

圖 7 曲譜範例

```
<general>
  <description>
    <main_title>家後</main_title>
    <author type="作詞者">鄭進一/陳維祥</author>
    <author type="作曲者">鄭進一</author>
    <work_title>江蕙同名專輯</work_title>
    <date type="發行日期">2001\06</date>
    <genres>
      <genre name="POP" description="台語流行音樂" weight="100"
/>
    </genres>
  </description>
  <related_files>
    <related_filefile_name="家後.mp3" file_format="audio_mp3"
encoding_format="audio_mp3" description="MP3" copyright=""
notes="" />
    <related_filefile_name="家後.jpg" file_format="image_jpeg"
encoding_format="image_jpeg" description="COVER" copyright="唱片
公司" notes="" />
    <related_filefile_name="家後.flv" file_format="video_mpeg"
encoding_format="video_mpeg" description="MV" copyright="唱片公
司" notes="" />
    <related_filefile_name="家後.xml" file_format="text_html"
encoding_format="text_html" description="SCORE" copyright=""
```

```
notes="" />
  </related_files>
</general>
```

圖 8 IEEE 1599 一般層格式描述

本文利用邏輯層來進行詞曲同步化控制。如圖 9 所示，在此層定義了音樂的脊椎(spine)，規範出整首 MP3 樂句的邏輯順序，並且控制各種格式的樂句資訊同步。脊椎是由一連串的音乐事件(events)所組成。我們將一個樂句對應至一個事件，其時間先後排序在 timing 屬性中標註。例如 a2 樂句的 timing 屬性為 1，表示開始時間在距離 a1 樂句出現後一個樂句時出現；而 hpos 屬性標示在曲譜中出現的虛擬空間位置(virtual horizontal dimension)對應順序，同樣使用樂句作為曲譜虛擬空間位置的單位。因為國語流行音樂的樂句結構較簡單，因此一首歌曲中所有樂句時間先後排序，基本會與曲譜虛擬空間位置順序相同。

```
<ieeel599>
  <logic>
    <spine>
      <event id="a1" timing="0" hpos="0" />
      <event id="a2" timing="1" hpos="1" />
      ...
    </spine>
    <los>...</los>
  </logic>
</ieeel599>
```

圖 9 樂句事件建立

本文可以利用音訊層中定義樂句在 MP3 音樂檔案中的起始位置。如圖 10 所示，「家後」歌曲的第一個樂句 a1 與第二個樂句 a2，在 MP3 音樂檔案中的開始時間分別標示為 21 與 28 秒。在播放 MP3 時，若時間有定義事件的標籤，則會觸發相對應的樂句事件，進行樂句同步控制。

```
<audio>
<track file_name="家後.mp3" file_format="audio_mp3"
encoding_format="audio_mp3" md5="">
  <track_indexing>
    <track_eventstart_time="21" end_time="27" event_ref="A1"
description="" />
    <track_eventstart_time="28" end_time="30" event_ref="A2"
description="" />
  </track_indexing>
</track>
```

圖 10 MP3 樂句事件同步資訊

對於曲譜資訊，本文利用譜號層中定義每個樂句的五線譜中邏輯與譜面幾何位置，使其在播放到相對應樂句時，能同步顯示樂句所相對應的曲譜，如圖 11 所示。對於歌曲，歌詞資訊的存在非常重要。利用邏輯層中定義每樂句的歌詞與相對應樂句事件來做到詞曲同步播放，透過樂句事件，控制 MP3 樂句、曲譜、歌詞三者同步顯示。

```
<notational>
<graphic instance_group description="國語流行音樂樂譜">
```

```

<graphic instance_file_name="images\家後_page1.tif"
format="image_tiff" position_in_group="1"
spine_start_ref="p4" spine_end_ref="p7"
measurement_unit="pixel">
...
</graphic instance>
<graphic instance_group description="數位樂譜">
<graphic instance_file_name="images\家後_page3.tif" ... >
<graphic instance_file_name="images\家後_page2.tif" ... >
</graphic instance><lyricspart_ref="piano" voice_ref="null">
<syllable start_event_ref="A1" end_event_ref="P1" hyphen="no">
有一日咱若老</syllable>
<syllable start_event_ref="A2" end_event_ref="P2" hyphen="no">
讓找無人甲咱友孝</syllable>
...
</lyrics>

```

圖 11 曲譜同步資訊

Chapter 5 曲譜樂句自斷分段

傳統樂句分段是以心理學的角度來探討音樂群組(musical grouping)的關係。1920年在德國有一群稱作為完形學派(Gestalt school)的心理學家們提出物體群組的相似理論，後來稱為完形理論[32]。完形理論提出音樂群組心理層面的認知方面，包含了三種的基本原則：

- 接近原則(principle of proximity):
時間與空間上彼此之間非常接近的事物，本文傾向於把這些事物組合成群組。
- 相似原則(principle of similarity):
本文傾向於把相似的事物組合成群組。
- 連續原則(principle of continuation):
指受到彼此連續的符號或圖形排列而成整體圖樣的刺激，即使是中空的圖形排列，很容易視為組合成的一種整體圖樣。本文傾向於把形成連續圖案的事物組合為一個群組。

完形學派的心理學家們雖然有提到可將完形理論應用在音樂層面，但是真正將系統化的利用完形理論，提出音樂群組方面理論的研究，是由 Jackendoff 與 Lerdale 兩個人所提出的調性音樂生成理論(GTTM, *A Generative Theory of Tonal Music*)[23]。Cohen 與 Frankland 利用調性音樂生成理論中的群組傾向規則(GPR, grouping preference rules)提出了樂句斷句位置偵測方法之實作[9]。此方法實作可以根據長起音點間距、

發音方式改變、力度變化、音程變化、音長變化、連奏、休止符等用來估算出會發生樂句斷句位置的可能性。

延續調性音樂生成理論的法則，Temperley 在[31]此書中詳細的深討有關於樂音群組的規則，提出了樂句結構傾向規則(PSPR, phrase structure preference rules)，考慮到樂音群組間隙(gap)、韻律對句(metrical parallelism)、樂句長度等來估計出樂句分段點的可能性。本文利用 Temperley 所提出的樂句結構傾向規則，實驗 MP3 音樂及其對應的曲譜拆解成一連串的樂句。樂句結構傾向規則的說明如下：

- 韻律對句法則(Metrical Parallelism Rule)：

樂句的斷句位置大多會發生在於構成韻律結構中產生的對句地方。

Temperley 所提出的樂句結構傾向規則，以第一個法則作用是最為明顯。以流行音樂舉例，樂句斷句位置大多為歌手演唱換氣的地方，而且為了強化情感上的表達，會因此使樂句最後常有休止符或較長音的音符存在。第二法則是為了避免產生太短或太長不合理的樂句斷句，大多數的樂句包含了 6 至 10 個音符，但是還是會有例外，所以第二法則的重要性相較其他法則來的低。第三法則是因為流行音樂樣似詩歌，歌詞與旋律之間會有對仗、反覆的歌詞與旋律會出現，因此把樂句斷句在於形成韻律對句上的位置，比較能符合音樂對仗的特性。

- 樂音群組間隙法則(Gap Rule)：

樂句斷句位置大多會發生在有較長的起音點間隔(inter-onset intervals)、或者是較長的終音點與起音點間隔(offset-to-onset intervals)的地方。

- 樂句長度法則(Phrase Length Rule):

樂句的長度大多數包含八個樂音。以下為此法則計算公式：

$$f_{\text{樂句}}(N) = \text{Min} |(\log_2 N - 3)| \quad (1)$$

N 為樂句所包含的樂音個數。

Chapter 6 MP3 自動斷句

為了達成 MP3 的自動斷句，本文分別使用 MP3 樂音的起音點偵測技術以及呼吸聲偵測技術來進行斷點預估。以下分別對兩種技術做詳細說明。

6.1 MP3 樂音的起音點偵測

本文所使用的起音點偵測特徵值，取自 MP3 解碼過程中的修正式離散餘弦轉換係數(MDCT)[19]。由於一般未經壓縮音樂的起音點偵測技術與 MP3 音樂起音點偵測的差異，在於 MP3 音樂起音點偵測所使用的特徵值必須取自壓縮領域，以降低額外需要的計算量。我們依樂理上八度音程範圍進行能量加總，然後參考 Goto 的起音點判讀準則[12][13]來計算每個音框發生起音事件(*onset event*)的機率，最後參考 Dixon 的平滑化以及峰值挑選技術[8]，進行整合起音點判斷。起音點偵測各步驟說明如下。

- **音框切割與特徵值擷取：**

依據 MPEG 標準[19]，以 44.1KHz 取樣的音樂訊號，每個音框(frame)包含 1152 個樣本。本文在 MP3 解碼過程中取出 MDCT 係數作為起音點偵測特徵值。MP3 音訊每秒有 $44100/1152=38.28125$ 個音框，MDCT 係數每個頻帶的頻率解析度為 38.28125 Hz。

- **八度音程能量加總：**

MP3 既有的頻帶切分方式近似於 E#調($E\#1=38.891$ Hz)。故本文可依照 E#0、E#1、E#2、E#3、...、E#7 的頻率將 MP3 音訊頻帶分為八個

八度音程範圍來進行能量加總統計。

- **起音事件機率計算：**

各八度音程加總後的能量，依照等音量曲線(equal-loudness curves)[20]進行音量感知正規化。起音事件機率的計算方式參考Goto方法設計的精神，規則是發生起音事件的音框，其八度音程能量必須連續兩個音框都呈現上升現象，而兩個音框的八度音程能量之差即為發生起音事件的機率。

- **峰值計算與挑選：**

一首歌曲的起音事件之機率變化曲線，必須經過平滑化處理，來減少起音點誤判的個數。本文使用起音峰值函數來進行平滑化，同時整合各個八度音程的個別判定結果，以及考慮門限值的設定。第 i 個音框的起音峰值函數 $F_{onset}(i)$ 定義如下：

$$F_{onset}(i) = Peaks\left(FE[i, j], M, N, FE_{th}\right) \quad (2)$$

其中 $FE[i, j]$ 為第 i 個音框的第 j 個八度音程能量； FE_{th} 為起音事件門限值； $Peaks()$ 函數為峰值挑選函數，其挑選條件為 $FE[i, j] > FE_{th}$ 且 $FE[i, j]$ 為 $\pm M$ 個音框範圍以及 $\pm N$ 個八度音程範圍內所有音框能量 $FE[i, j]$ 之最大值。 FE_{th} 、 M 與 N 之值的設定由實驗結果調整決定。

Chapter 7 依據呼吸聲偵測 MP3 斷句技術

綜合以往在語音識別以及音樂內涵分析領域的研究成果，呼吸聲偵測的特徵值以梅爾倒頻譜係數(MFCC, Mel-frequency cepstral coefficients) 最為有效。因此，本文亦採用 MFCC 作為呼吸聲偵測的特徵值。由於我們針對 MP3 壓縮格式的數位音樂進行分析，特徵值計算宜在壓縮領域中進行以減輕計算負荷量。在壓縮領域進行 MFCC 特徵值擷取的方式說明如下：

MFCC 特徵值的計算過程大致上是先使用傅立葉轉換將訊號轉換頻率域上，接著按照梅爾尺度對頻帶做切割。接著對每個子頻帶分別計算其能量值，將能量取對數。最後對這些能量對數做離散式餘弦轉換，取得 MFCC 係數。在這一系列的計算過程中，花費時間較多的為傅立葉轉換。對於 MP3 來說，還要再加上把 MP3 解壓縮還原成原始信號的時間。原本計算過程中，做傅立葉轉換的目的是為了將訊號轉為頻率域，而 MP3 的 MDCT 係數本身即為頻率域的資訊。因此本文可以將 MDCT 係數取代傅立葉轉換係數，把原先的傅立葉轉換係數以 MP3 音訊的 MDCT 係數取代。計算過程如下：

首先按照梅爾尺度近似計算公式將 MDCT 係數換算成符合人耳感知特性的梅爾頻譜係數 m_j ，並對其做平滑化的動作：

$$m_j = \begin{cases} MDCT[j], & \text{if } freq(MDCT[j]) \leq 1 \text{ kHz} \\ 2595 \log_{10} \left(1 + \frac{MDCT[j]}{700} \right), & \text{if } freq(MDCT[j]) > 1 \text{ kHz} \end{cases} \quad (3)$$

上述的 $\text{freq}(\text{MDCT}[j])$ 代表的是 MDCT 係數中，第 j 個係數所代表的頻率。接著，計算每個音框的梅爾頻譜係數之能量值，並取對數值，使其符合人耳對音量的感知特性：

$$m_db_j = \log(|m_j|^2) \quad (4)$$

最後透過離散餘弦轉換去除梅爾倒頻譜係數各個維度的相依性，可得到 MP3 音訊在壓縮領域之 13 維的 MFCC 係數。

$$MFCC_j = \sum_{i=1}^{13} m_db_j \cos(j(i - \frac{1}{2}) \frac{\pi}{13}), j=1,2,\dots,13 \quad (5)$$

假設一個音訊樣本 X 可以藉由一連串的 MFCC 特徵向量 M 來表示

$$M = m_1, m_2, L, m_T \quad (6)$$

其中 m_i 為音訊樣本 X 在第 i 個音框的 13 維 MFCC 特徵向量。辨識音訊樣本 X 是否為呼吸聲，此問題亦即在於求在音訊樣本 X 出現 MFCC 特徵向量 M 時，音訊樣本 X 為呼吸聲 B 的機率 $P(B|M)$ 。根據貝氏定理

$$P(B|M) = \frac{P(M|B)P(B)}{P(M)} \quad (7)$$

本文使用 HMM 模型來估算 $P(M|B)$ 之值。考慮到男女呼吸聲的基本差異，我們訓練三種不同的呼吸聲 HMM 模型 H 男生呼吸聲、 H 女生呼吸聲、 H 混和呼吸聲來辨識男生與女生在歌曲中的呼吸聲，以及 H 呼吸與 H 樂音來辨識呼吸與樂音的兩個 HMM 模型。藉由計算 $P(M|H$ 男生呼吸聲)、 $P(M|H$ 女生呼吸聲)、 $P(M|H$ 混和呼吸聲)、 $P(M|H$ 呼吸)、以及 $P(M|H$ 樂音) 之值來判斷音訊樣本 X 為呼吸聲(男生、女生、混和)、呼吸或樂音。

Chapter 8 系統實作

在 Windows 環境下，以 C# 開發出一套具有異質音樂資訊同步及整合的特性的國語流行音樂自動斷句撥放程式。此撥放程式能夠讀取 IEEE 1599 標準的異質音樂整合成的 XML 檔案，直接顯示出音樂的歌曲名稱、演唱者、作曲家、編曲、專輯名稱、發行日期等關於此 MP3 音樂的相關資訊。本文使用 MusicXML 作為統一曲譜儲存格式，部分曲譜由其他五線譜編輯軟體所建立，我們將其統一轉換成 MusicXML 格式後存放在 eXist 曲譜資料庫中。

程式執行樂句結構傾向法則之 MP3 樂句分段需透過音字符串列(Note List)產生，如圖 12，所產生出來的(*.notes)來直接進行 Melisma 的樂句判斷。必需先進行 Melisma 中的 meter 轉換，然後在轉換成 Grouper 樂句(Phrase)。

Note	1200	1500	79
Note	1500	1799	77
Note	1799	2099	76
Note	2099	2400	64
Note	2400	4800	72
Note	6000	6300	79
Note	6300	6599	77
Note	6599	6900	76
Note	6900	7199	67
Note	7199	9600	69
Note	9600	11399	68
Note	11399	11704	67

圖 12 音字符串列(Note List)

執行 Melisma 工具，使用以下指令來完成：“meter 原始檔案名 > 儲存結果檔案名”；其轉換結果如圖 13 所示。

Beat	0	3
Beat	140	0
Beat	315	1
Beat	455	0
Beat	595	2
Beat	735	0
Beat	910	1
Beat	1050	0
Beat	1190	4
Beat	1330	0
Beat	1470	1
Beat	1610	0
Beat	1785	2
Beat	1925	0

圖 13 meter 轉換

最後將 meter 檔案轉換成 grouper 檔案，其命令指令為：“grouper 原始檔案名稱 > 儲存結果檔案名”，即完成 Melisma 的樂句群組分析了，如。

Note	0	315	79
Note	315	595	77
Note	595	910	76
Note	910	1190	64
Note	1190	3605	72
Phrase	4795		
Note	4795	5110	79
Note	5110	5390	77
Note	5390	5705	76
Note	5705	5985	67
Note	5985	8400	69
Phrase	8400		

圖 14 grouper 轉換

透過本文所開發的程式，分析音符時間點、音高等等的資料，即完成國語流行音樂自動斷句系統的樂句分段，結果如圖 15 所示。

音符	開始時間	結束時間	音高	音
Note 0	315	79	Sol	(G)
Note 315	595	77	Fa	(F)
Note 595	910	76	Mi	(E)
Note 910	1190	64	Mi	(E)
Note 1190	3605	72	Do	(C)
Phrase 4795				
此樂句音符數量： 5				
Note 4795	5110	79	Sol	(G)
Note 5110	5390	77	Fa	(F)
Note 5390	5705	76	Mi	(E)
Note 5705	5985	67	Sol	(G)
Note 5985	8400	69	La	(A)
Phrase 8400				
此樂句音符數量： 5				

圖 15 國語流行音樂自動斷句系統的樂句分段結果

本文所開發的程式藉由讀取 IEEE 1599 標準的 XML 音樂整合描述檔案後，會解析出整合 MP3 上的曲譜、歌詞或者其他相關之異質資訊(如：演唱者、作曲家、編曲、專輯名稱、發行日期等)，MusicXML 曲譜檔案會由本文所提出的系統自動斷句技術，進行樂句斷句分析，產生出來的結果會是以圓滑線為樂句單位的方式標記在 MusicXML 樂譜上並顯示出來。使用者可以藉由撥放程式查詢到目前所撥放的 MP3 音樂相關資訊，此外一般撥放器擁有的基本功能(開始、暫停、撥放下一首等功能)，以及音樂播放時能同步顯示當前相對應的歌詞，如圖 16 所示。



圖 16 音樂播放模式

更進階的功能如圖 17 所示，能夠顯示音樂的樂譜、樂句總數、斷句的位置等資訊，提供音樂研究者一個良好的平台，研究各樂句間的狀況。



圖 17 樂句分析模式

樂譜樂句顯示功能如圖 18 所示，能夠讓音樂學者、音樂演奏者直接觀看樂譜，對於音樂分句結構可以更進一步的了解，或做為學習演奏之參考內容。



圖 18 樂譜樂句顯示模式

Chapter 9 實驗結果

本文實驗樣本之樂譜範例，主要取自《Hit101中文流行鋼琴百大首選》裡的一百首樂曲，以及張智星教授建立之哼唱歌曲資料庫MIR-QBSH Corpus (<http://mirllab.org/dataSet/public/MIR-QBSH-corpus.rar>) 中的二十一首哼唱歌曲進行實驗。流行音樂主要包含主歌以及副歌兩個部分，如圖 19 所示。通常副歌的樂句較規則，本文實驗比較全曲與僅含副歌部分的斷句效能。樂句分段效能測試方式採用回覆率 (recall rate) 及準確率 (precision rate) 來呈現，計算公式如下：

$$\text{回覆率} = \left(\frac{\text{系統偵測到之正確樂句數}}{\text{實際樂句數}} \right) * 100\% \quad (8)$$

$$\text{準確率} = \left(\frac{\text{系統偵測到之正確樂句數}}{\text{系統偵測到之總樂句數}} \right) * 100\% \quad (9)$$

其中，實際樂句數為人工辨識樂譜之正確樂句數；系統偵測正確樂句數為系統偵測樂句斷音點與人工判斷樂句斷音點誤差在八分音符以內範圍的樂句個數。



圖 19 流行音樂架構圖

- 實驗流程

使用國語流行音樂自動斷句系統與詞曲同步撥放程式判斷樂句的位置，會與實際音樂樂句不一定相同，我們要檢查電腦判斷的樂句位置是否同等於實際的數據，藉此找出音樂效能測試方式音樂回覆率以及準確率，如圖 20 為系統偵測的國語流行音樂樂曲資料庫的 NO.001 曲目：江蕙-家後 的片段資訊。

```

Note 19495 19810 67 Sol ( G )
Note 19810 20090 69 La ( A )
Note 20090 20405 72 Do ( C )
Note 20405 21315 74 Re ( D )
Note 21315 21595 69 La ( A )
Note 21595 23415 72 Do ( C )
Phrase 24290
此樂句音符數量： 6
Note 24290 24605 67 Sol ( G )
Note 24605 24885 69 La ( A )
Note 24885 25200 72 Do ( C )
Note 25200 25515 74 Re ( D )
Note 25515 26110 76 Mi ( E )
Note 26110 26390 79 Sol ( G )
Note 26390 28210 74 Re ( D )
Phrase 28805
此樂句音符數量： 7
Note 28805 29085 79 Sol ( G )
Note 29085 29400 74 Re ( D )
Note 29400 29715 76 Mi ( E )
Note 29715 31185 79 Sol ( G )
Note 31185 31500 72 Do ( C )
Note 31500 31640 76 Mi ( E )
Note 31640 31815 74 Re ( D )
Note 31815 32095 76 Mi ( E )
Note 32095 33600 72 Do ( C )
Phrase 33600
此樂句音符數量： 9
Note 33600 33740 69 La ( A )
Note 33740 34055 72 Do ( C )
Note 34055 34195 72 Do ( C )
Note 34195 34335 72 Do ( C )
Note 34335 34650 72 Do ( C )
Note 34650 34790 69 La ( A )
Note 34790 34965 72 Do ( C )
Note 34965 35245 72 Do ( C )
Note 35245 35385 81 La ( A )
Note 35385 35700 72 Do ( C )
Note 35700 36015 72 Do ( C )
Note 36015 37800 74 Re ( D )
Phrase 38395
此樂句音符數量： 12
Note 38395 38710 76 Mi ( E )
Note 38710 38990 79 Sol ( G )
Note 38990 39305 76 Mi ( E )

```

圖 20 電腦判斷樂句的片段資訊

如圖 21 所示，以此方式對照樂譜，電腦判斷樂句起音點在於的 9 小節開頭休止符之後以及第 11 小節休止符之後，完全與實際狀況相同，持續測試整首曲目，將可以找出此樂曲的回覆率及準確率。

The image shows two staves of a piano score. The top staff starts at measure 7 with a Dm^7 chord, followed by a G chord. A box labeled 'A C' is above the third measure. The lyrics '有一日咱若' and '有一日咱若' are written below the notes. A black triangle points to the start of the phrase '有一日咱若' in the third measure, with the label '電腦判斷的樂句位置' (Computer-identified phrase position) below it. The bottom staff starts at measure 10 with the lyrics '老老' and '找無 人甲 咱友 孝' and '有媳 婦子 兒友 孝'. A black triangle points to the start of the phrase '找無 人甲 咱友 孝' in the second measure, with the label '電腦判斷的樂句位置' below it. The G chord is indicated above the final measure.

4 *Handwritten signature*

OP : EMI MUSIC PUBLISHING (S.E.ASIA) LTD., TAIWAN

圖 21 電腦判斷的樂句位置圖

● 實驗數據

編號	曲名	實際樂句數	系統偵測之總樂句	系統偵測之正確樂句	系統偵測之錯誤樂句	回覆率	準確率
1	家後	44	35	31	4	70.45%	88.57%
2	征服	38	44	33	11	86.84%	75.00%
3	心動	32	33	15	18	46.88%	45.45%
4	約定	44	31	10	21	22.73%	32.26%
5	活該	48	39	35	4	72.92%	89.74%
6	葉子	37	31	11	20	29.73%	35.48%
7	童話	36	50	11	39	30.56%	22.00%
8	約定	37	36	33	3	89.19%	91.67%
9	後來	48	49	21	28	43.75%	42.86%
10	聽海	49	44	24	20	48.98%	54.55%
11	十年	46	34	31	3	67.39%	91.18%
12	曖昧	35	25	13	12	37.14%	52.00%
13	江南	51	40	22	18	43.14%	55.00%
14	遇見	17	18	11	7	64.71%	61.11%
15	同類	39	28	18	10	46.15%	64.29%
16	唯一	35	34	30	4	85.71%	88.24%
17	勇敢	45	40	32	8	71.11%	80.00%
18	剪愛	41	41	22	19	53.66%	53.66%
19	當你	60	40	27	13	45.00%	67.50%
20	月光	45	33	32	1	71.11%	96.97%
21	候鳥	45	34	25	9	55.56%	73.53%
22	天灰	45	25	21	4	46.67%	84.00%
23	安靜	48	51	30	21	62.50%	58.82%
24	擱淺	46	41	24	17	52.17%	58.54%
25	鋼琴	43	38	21	17	48.84%	55.26%
26	星晴	45	42	27	15	60.00%	64.29%
27	軌跡	57	38	22	16	38.60%	57.89%
28	天空	39	40	35	5	89.74%	87.50%
29	倒帶	51	44	42	2	82.35%	95.45%

30	知足	46	35	7	28	15.22%	20.00%
31	寧夏	42	26	26	0	61.90%	100.00%
32	小薇	33	22	20	2	60.61%	90.91%
33	紀念	36	32	18	14	50.00%	56.25%
34	假如	41	36	28	8	68.29%	77.78%
35	夢一場	52	52	31	21	59.62%	59.62%
36	上弦月	42	37	27	10	64.29%	72.97%
37	我願意	48	37	14	23	29.17%	37.84%
38	如果愛	35	33	11	22	31.43%	33.33%
39	天黑黑	57	46	25	21	43.86%	54.35%
40	我的愛	41	32	26	6	63.41%	81.25%
41	菊花台	54	40	26	14	48.15%	65.00%
42	珊瑚海	56	37	36	1	64.29%	97.30%
43	髮如雪	71	48	40	8	56.34%	83.33%
44	七里香	51	33	21	12	41.18%	63.64%
45	流星雨	36	34	17	17	47.22%	50.00%
46	就是愛	50	32	24	8	48.00%	75.00%
47	聽不到	64	36	29	7	45.31%	80.56%
48	情非得已	49	42	42	0	85.71%	100.00%
49	新不了情	49	33	31	2	63.27%	93.94%
50	理想情人	64	45	27	18	42.19%	60.00%
51	痴心絕對	50	50	33	17	66.00%	66.00%
52	大城小愛	89	84	55	29	61.80%	65.48%
53	眼淚成詩	31	20	19	1	61.29%	95.00%
54	花的嫁紗	65	54	43	11	66.15%	79.63%
55	千年之戀	33	31	24	7	72.73%	77.42%
56	我們的愛	61	54	24	30	39.34%	44.44%
57	一了百了	54	41	33	8	61.11%	80.49%
58	戀人未滿	62	62	38	24	61.29%	61.29%
59	愛很簡單	40	32	23	9	57.50%	71.88%
60	黑色幽默	35	42	8	34	22.86%	19.05%
61	分手快樂	31	24	16	8	51.61%	66.67%
62	除此之外	45	34	26	8	57.78%	76.47%
63	千里之外	34	31	23	8	67.65%	74.19%

64	我只在乎你	43	37	27	10	62.79%	72.97%
65	城裡的月光	40	45	27	18	67.50%	60.00%
66	孤單北半球	28	32	17	15	60.71%	53.13%
67	如果雲知道	50	34	34	0	68.00%	100.00%
68	老鼠愛大米	45	37	30	7	66.67%	81.08%
69	愛一直閃亮	33	30	27	3	81.82%	90.00%
70	至少還有你	47	37	24	13	51.06%	64.86%
71	隱形的翅膀	14	15	0	15	0.00%	0.00%
72	遺失的美好	39	43	23	20	58.97%	53.49%
73	一千年以後	37	30	25	5	67.57%	83.33%
74	給我你的愛	32	35	15	20	46.88%	42.86%
75	那年的情書	36	35	33	2	91.67%	94.29%
76	我會好好的	47	46	36	10	76.60%	78.26%
77	花都開好了	25	31	21	10	84.00%	67.74%
78	愛我還是他	30	29	22	7	73.33%	75.86%
79	如果有一天	43	34	34	0	79.07%	100.00%
80	很愛很愛你	59	38	34	4	57.63%	89.47%
81	從開始到現在	34	29	20	9	58.82%	68.97%
82	我多麼羨慕你	44	43	20	23	45.45%	46.51%
83	一首簡單的歌	41	28	22	6	53.66%	78.57%
84	別在傷口灑鹽	37	26	21	5	56.76%	80.77%
85	檸檬草的味道	38	33	27	6	71.05%	81.82%
86	威尼斯的眼淚	28	29	24	5	85.71%	82.76%
87	他一定很愛你	43	26	26	0	60.47%	100.00%
88	莫斯科沒有眼淚	40	34	27	7	67.50%	79.41%
89	我知道你很難過	45	36	33	3	73.33%	91.67%
90	忠孝東路走九遍	46	32	26	6	56.52%	81.25%
91	月亮代表我的心	40	36	4	32	10.00%	11.11%
92	原來你什麼都不想要	41	40	18	22	43.90%	45.00%
93	當你孤單你會想起誰	44	40	28	12	63.64%	70.00%
94	男人不該讓女人流淚	76	40	37	3	48.68%	92.50%
95	I Believe	30	39	13	26	43.33%	33.33%

96	Lydia	31	31	22	9	70.97%	70.97%
97	Only One	37	37	27	10	72.97%	72.97%
98	Forever Love	54	44	36	8	66.67%	81.82%
99	Kiss Goodbye	40	39	22	17	55.00%	56.41%
100	May I Love You	36	32	28	4	77.78%	87.50%
101	Through The Arbor-Kevin	純鋼琴伴奏					

101首樂曲之實驗結果如表格1所示，實際樂句數與系統偵測之總樂句數無誤差的回覆率57.77%，準確率是68.75%。但部分的系統偵測之正確樂句斷句位置跟實際樂句斷句位置相差不大，對歌詞同步影響不大。本文亦將100首歌曲做了另一個實驗，實際樂句與系統偵測之總樂句兩者相差在八分音符以內的回覆率是62.34%，準確率為74.54%，後者明顯提高。

表格 1 一百零一首樂曲之實驗結果統計

實驗結果	回覆率平均值	準確率平均值
實際樂句數與系統偵測 之總樂句(無誤差)	57.77%	68.75%
實際樂句與系統偵測 之總樂句(誤差八分音符以內)	62.34%	74.54%

本文針對二十一首哼唱歌曲做實驗，使用 Melisma 的音樂樂句群組分析工具所測試出來的回覆率為 40.7%，準確率是 62.9%。使用的呼吸聲偵測方法的回覆率為 77.8%，準確率為 84%，其中準確率之準度必需在 +/- 50ms(含)以內，如表格 2 所示。

表格 2 二十一首哼唱歌曲之實驗結果統計

實驗結果	回覆率平均值	準確率平均值
實際樂句與系統偵測之總樂句 (相差+/- 50ms(含)以內)	77.8%	84.0%

除第101首樂曲為純鋼琴伴奏，無從判別其準確率以及回覆率外，其餘實驗結果顯示大部分音樂都有優良的準確率以及回復率，少數斷句標準與一般樂譜不太相同，導致準確率以及回復率就會低於平均值，由此可見，音樂使用MP3國語流行音樂自動斷句系統進行分段已經具備良好的斷句功能。

Chapter 10 結論

本文提出一種針對 MP3 數位音樂進行樂句自動分段與呼吸聲偵測的技術，在起音偵測的實驗中，我們使用重覆偵測計算(Rescan Detection Computation)的方法，針對這 21 首哼唱歌曲進行測試比較。使用該測試方法所得到的平均回復率(Recall rate)是 79.26%；精確率(Precision rate)為 83.78%。同時也進行相同歌曲的測試，分別是由不同的人所哼唱的。在這個測試比較中也可以發現，在相同的歌曲由不同人的來演唱也是會有不相同的結果。基本上歌唱技巧的好壞，也會影響我們所做的偵測結果。

在 MP3 樂曲起音點偵測樂句斷句的測試中，本文所採用的 101 首歌曲，使用 Melisma 的音樂樂句群組(Phrase Grouper)分析工具所測試出來的回復率(Recall rate)為 62.3%；精確率(Precision rate)是 74.5%。而本文所使用的呼吸偵測(Breath Detection)方法回復率(Recall rate)為 77.8%；精確率(Precision rate)為 84%。雖然呼吸偵測(Breath Detection)的方法結果比較好，但此方法只能針對有人聲演唱的歌曲來進行，但對於流行音樂來說，已經足夠。

系統同步播放功能結合 IEEE 1599 標準對 MP3 訊號、歌詞與曲譜的整合描述與樂句同步能力，開發了一套可以讀取 IEEE 1599 音樂整合檔的樂句播放器，提供音樂聆聽者一套 MP3 國語流行音樂自斷斷句以及詞曲同步撥放系統，讓使用者根據音樂整合檔所記載之內容，瀏覽該首歌曲之各種異質資訊，包含演唱者、作詞者、作曲者、專輯發行日期、專輯封面、歌手寫真、歌手其他作品……等資訊，播放 MP3 檔案時可以藉由同步播放功

能，觀看目前歌曲歌詞、查閱樂譜播放資訊、進行樂句分析。

至今我們所使用的呼吸偵測方法只能針對有人聲演唱的歌曲來進行，對於流行音樂來說，容易受到背景音樂所干擾，所以如何在伴奏音樂下偵測到呼吸聲是一項高難度挑戰。此外如何整合多種不同的樂句的分段點判斷方法，亦是我們未來要探討的方向。

参考文献

- [1]Baggi, D.L., “An IEEE Standard For Symbolic Music,” *IEEE Computer*, pp.100–102, Nov. 2005.
- [2]Baggi, D.L., A. Baratè, G. Hausand L. A. Ludovico, “A Computer Tool to Enjoy and Understand Music,” in *Proceedings of EWIMT 2005 Integration of Knowledge, Semantics and Digital Media Technology*, pp. 213–217, 2005.
- [3]Bod, R., “Memory-based Models of Melodic Analysis: Challenging the Gestalt Principles,” *Journal of NewMusic Research*, Vol.30, No.3, pp.27–37, 2001.
- [4]Cambouropoulos, E., “The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing,” in *Proceedings of the International Computer Music Conference*, pp.17–22, 2001.
- [5]Cambouropoulos, E., “Musical Parallelism and Melodic Segmentation: A Computational Approach,” *Music Perception*, Vol. 23, No. 3, pp.249–267, 2006.
- [6]Cheng, E. and E. Chew, “A Local Maximum Phrase Detection Method and the Analysis of Phrasing Strategies in Expressive Performances,” In *Proceedings of the First International Conference of the Society of Mathematics and Computation in Music*, 2007.
- [7]Cheng, E. and E. Chew, “Quantitative Analysis of Phrasing Strategies in Expressive Performance: Computational Methods and Analysis of Performances of Unaccompanied Bach for Solo Violin,”

Journal of New Music Research, Vol.37, No.4, pp.325–338,
December 2008.

- [8] Dixon, S. “Onset Detection Revisited,” in *Proc. the 9th International Conference on Digital Audio Effects*, pp 133–137, 2006.
- [9] Frankland, B.W. and A.J. Cohen, “Parsing of Melody: Quantification and Testing of the Local Grouping Rules of Lerdahl and Jackendoff’s A Generative Theory of Tonal Music,” *Music Perception*, Vol.21, No.4, pp.499–543, 2004.
- [10] Friberg, A., R. Bresin, L. Frydén, and J. Sundberg, “Musical Punctuation on the Microlevel: Automatic Identification and Performance of Small Melodic Units,” *Journal of New Music Research*, Vol.27, No. 3, pp.271–292, 1998.
- [11] Good, M., “MusicXML: An Internet-Friendly Format for Sheet Music,” in *XML Conference Proceedings*, 2001.
- [12] Goto, M., “A Chorus-Section Detecting Method For Musical Audio Signal,” in *Proc. 2003 IEEE International Conference on Acoustic, Speech, and Signal Processing*, pp.V-437–440, 2003.
- [13] Goto, M. and Muraoka, Y., “Beat Tracking based on Multiple-agent Architecture — A Real-time Beat Tracking System for Audio Signals,” in *Proc. Second International Conference on Multi-agent Systems*, pp.103 – 110, 1996.
- [14] Gouyon, F., Klapuri, A., Dixon, S., Alonso, M., Tzanetakis, G. Uhle, C., and Cano, P., “An Experimental Comparison of Audio Tempo Induction Algorithms,” *IEEE Transaction on Audio, and*

Speech Processing, 2006.

- [15] Haus, G. and M. Longari, “Time-Based Music Description Approach Based on XML,” *Computer Music Journal*, Vol. 29, No. 1, pp. 70–85, 2005.
- [16] Haus, G. and L. A. Ludovico, “Music Segmentation: An XML-oriented Approach,” in *Lecture Note in Computer Science*, Vol. 3310, pp. 330–346, 2005.
- [17] Haus, G., and A. Pinto, “MX structural metadata as MIR tools,” in *Proceedings of Sound and Music Computing*, Salerno, Italy, 2005.
- [18] *IEEE Std. 1599*, “IEEE Recommended Practice for Defining a Commonly Acceptable Musical Application Using XML,” IEEE, 2008.
- [19] *ISO/IEC11172-3:1993*, “Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s — Part 3: Audio.”
- [20] *ISO 226:2003*, “Acoustics – Normal Equal-Loudness Level Contours,” ISO, 2003.
- [21] Klapuri, A. “Sound Onset Detection by Applying Psychoacoustic Knowledge,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 115–118, 1999.
- [22] Klapuri, A., Eronen, A. J., and Astola, J. T., “Analysis of the Meter of Acoustic Musical Signals” *IEEE Transaction on Audio, Speech, and Language Processing*, Vol. 14, No. 1, pp. 342–355, 2006.
- [23] Lerdahl, F. and R. Jackendoff, *A Generative Theory of Tonal Music*, MIT Press, Cambridge, MA, 1983.

- [24] Ludovico, L. A., “IEEE 1599: a Multi-layer Approach to Music Description,” *Journal of Multimedia*, Vol. 4, No. 1, pp. 9–14, 2009.
- [25] Meier, W., “eXist: An Open Source Native XML Database,” in *Proceedings of NODe Web-and Database-Related Workshops*, 2002.
- [26] Pearce, M. T., D. Müllensiefen, and G. A. Wiggins, “A Comparison of Statistical and Rule-based Models of Melodic Segmentation,” in *Proceedings of the International Computer Music Conference*, 2008.
- [27] Schaffrath, H., *The Essen Folksong Collection in the Humdrum Kern Format*, In: D. Huron (Ed.), Menlo Park, CA: Center for Computer Assisted Research in the Humanities, 1995.
- [28] Scheirer, E. D., “Tempo and Beat Analysis of Acoustic Musical Signals,” *J. Acoust. Soc. Amer.*, Vol. 103, No. 1, pp. 588 – 601, 1998.
- [29] Sommaruga, L. and D. Baggi, “Towards a Semantic Web Based Model for the Tonal System in Standard IEEE 1599,” *Journal of Multimedia*, Vol. 4, No. 1, pp. 40–45, 2009.
- [30] Tagliolato, P., “Music Structure Representation: a Formal Model,” in *Proceedings of Sound and Music Computing*, Marseille, France, 2006.
- [31] Temperley, D., *The Cognition of Basic Musical Structures*, MIT Press, Cambridge, MA, 2001.
- [32] Wertheimer, M., “Untersuchungen zur Lehre von der Gestalt II,” *Psychologische Forschung*, Vol. 4, No. 1, pp. 301–350, 1923.

Appendix A - 數位化國語流行音樂 MusicXML

樂譜之建立

從傳統的黑膠唱片，演進到錄音帶、CD，實體越來越精緻化，進而到近年來的 MP3，演變成不需要實體也可直接在電腦撥放。傳統的 MP3 播放器，好比說千千靜聽以及 winamp，如圖 22 所示，只提供聲音訊號及歌詞秀的功能，卻沒有以曲譜、歌詞、音樂訊號同步作分段的特性，這即是音樂研究者所需要的樂句分段功能；對音樂愛好者來說，在撥放音樂同時，也希望專輯照片、歌手照片、歌詞等異質資訊一併顯示，而目前的撥放器亦沒有自動化配置歌詞的方式，所以我們開發出一套整合音樂異質資訊的國語流行音樂自動斷句與詞曲同步撥放系統，建置 MusicXML 是為了讓歌曲有良好的儲存方式，讓這本系統可以應用在音樂動態歌詞的製作，藉由軟體偵測，便能夠自動產生字幕，提供使用者更好的音樂饗宴。



圖 22 音樂播放器的歌詞顯示功能

步驟一、 掃描樂譜

我們先將樂譜掃描存入電腦，存檔成 TIF 檔後，利用 PhotoScore Ultimate 來偵測與修改樂譜，如圖 23 所示，首先將照片打開。



圖 23 程式開檔畫面

步驟二、 人工校正

<1>. 目前這個檔案偵測到四個錯誤，第一個到第四個錯誤是樂譜尾端的的線沒有偵測到。修改方式為：先使用滑鼠右鍵，選擇 Barline 的第一個線條，如圖 24，就可以完成。



圖 24 歌曲修正方式

<2>. 我們以另外一個例子來說明其它容易發生的錯誤，如圖 25，現在軟體偵測到有三個二分音符，和原圖的三個全音符不同。

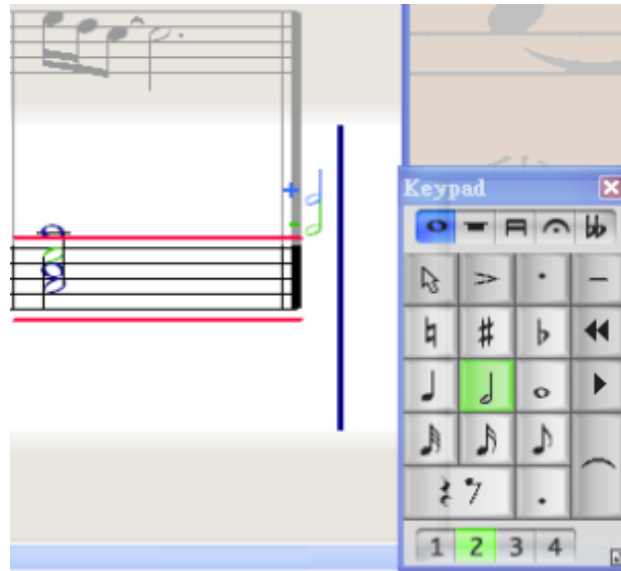


圖 25 常見問題

<3>. 首先將綠色的(Keypad 最底排的數字 2)二分音符改成藍色的主旋律(數字 1)。

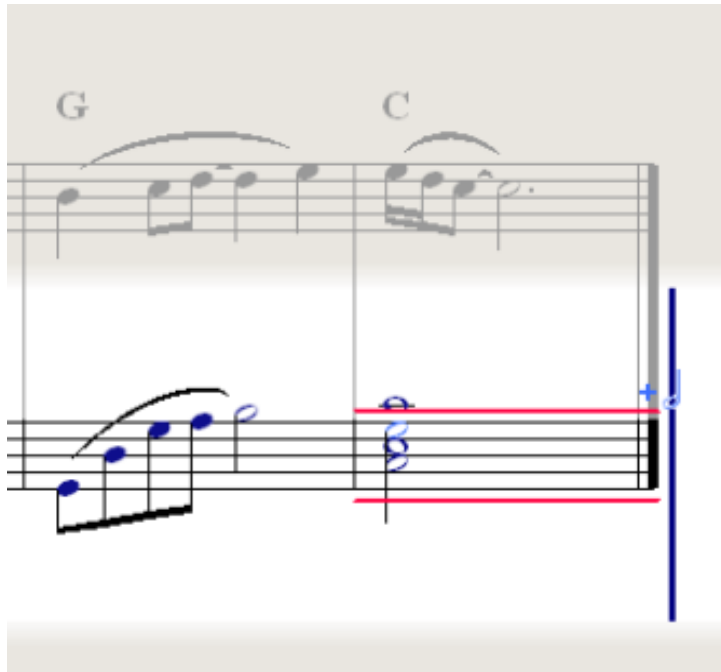


圖 26 解決方式

<4>. 之後將二分音符修改成全音符即完成，如圖 27。



圖 27 修正完成

步驟三、 數位樂譜完成

將檔案存檔成 xml 檔，如圖 28

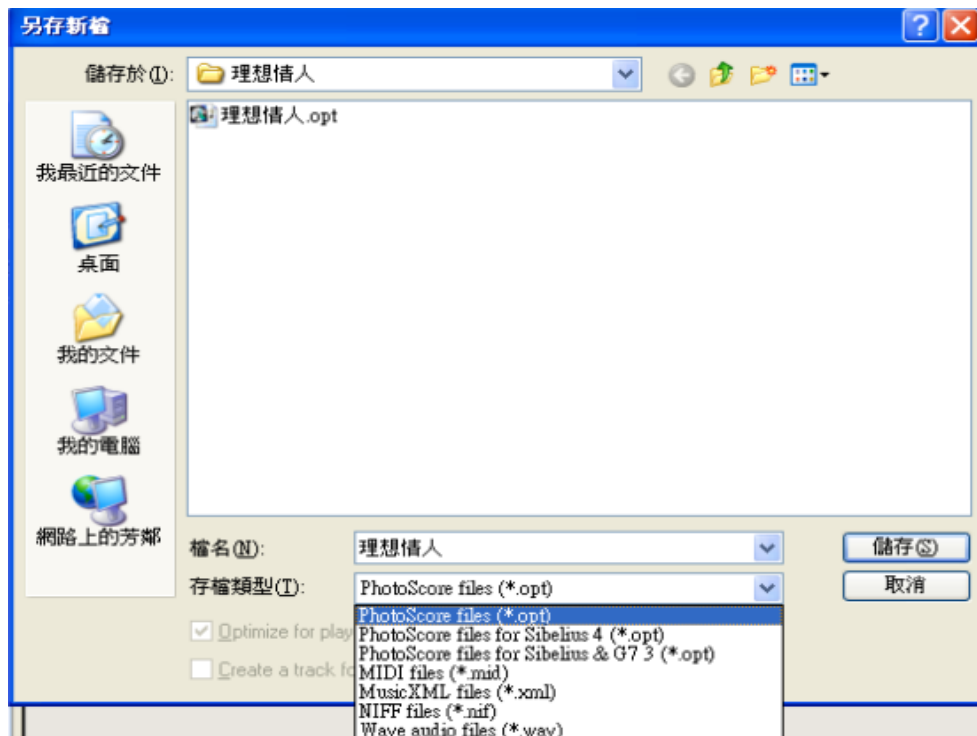


圖 28 XML 檔案編輯完成

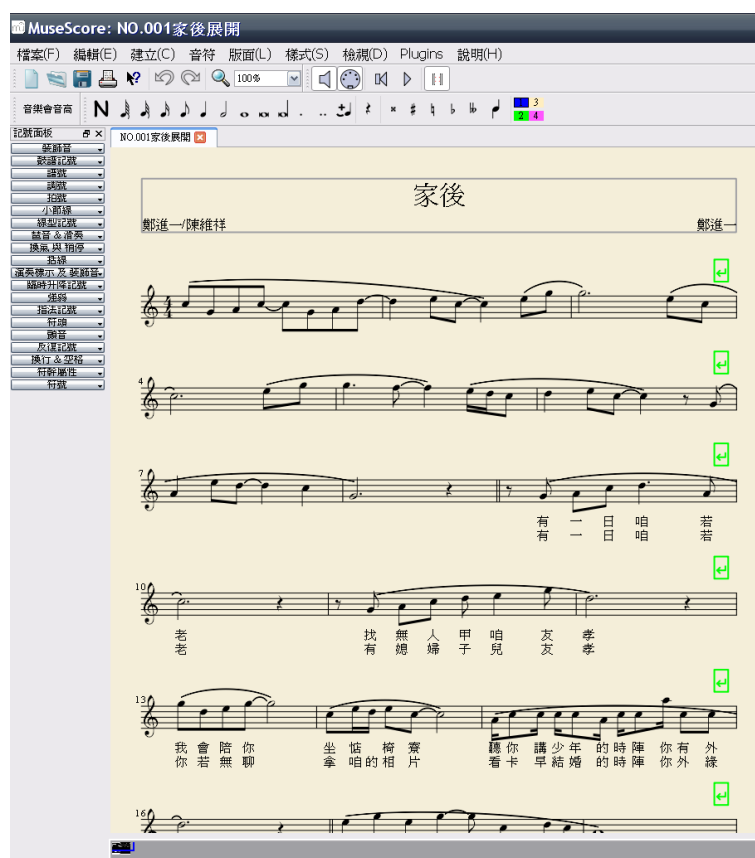
目前已經納入 101 首樂曲的 xml 資料於資料庫之中，這對於斷句的判斷是非常重要的，前置作業，我們實驗測試中已經完成了對這 101 首檔案所進行的測試，詳細實驗結果於 Appendix D。

Appendix B – 國語流行音樂樂曲資料庫之建構

國語流行音樂樂曲資料庫是能讓主程式連接，並且搜尋內部樂曲，讓撥放程式能夠撥放內部樂曲、以及歌詞，目前已納入 101 首歌曲的 MP3 數位音訊、歌詞，未來可以將更多樂曲新增至此資料庫中，讓此撥放程式擁有更多內容，能夠撥放更多樂曲。

實作方法

先將樂譜完整的加入作詞家、作曲家、曲名等資訊，再加入歌詞於 MusicXML，如圖 29，最後在將 MP3 數位音訊放入詞曲資料庫中。



The screenshot shows the MuseScore interface with the title 'NO.001家後展開'. The score is for the song '家後' by 鄭進一 (陳維祥). The lyrics are: 有二日咱若老 找無人甲咱友孝 有媳婦子兒友孝 我會陪你坐位椅寮 聽你講少年的時陣 你若有外緣 你若無聊 拿咱的相片 看卡早結婚的時陣 你外緣.

圖 29 輸入歌詞於 MusicXML

樂曲清單：

歌曲 編號	曲名	作詞	作曲	演唱
1	家後	鄭進一/陳維祥	鄭進一	江蕙
2	征服	袁惟仁	袁惟仁	那英
3	心動	林夕	黃韻玲	林曉培
4	約定	姚若龍	陳小霞	周蕙
5	活該	王武雄	詹凌駕	信樂團
6	葉子	陳曉娟	陳曉娟	阿桑
7	童話	光良	光良	光良
8	約定	光良	光良	光良
9	後來	施人誠	玉城千春	劉若英
10	聽海	林秋離	涂惠元	張惠妹
11	十年	林夕	陳小霞	陳奕迅
12	曖昧	姜憶萱/顏壘軒	小冷	楊丞琳
13	江南	李瑞洵	林俊傑	林俊傑
14	遇見	易家揚	林一峰	孫燕姿
15	同類	易家揚	李偲菘	孫燕姿
16	唯一	王力宏	王力宏	王力宏
17	勇敢	陳嘉文/夏木	洪敬堯	張惠妹
18	剪愛	林秋離	涂惠元	張惠妹
19	當你	張思爾	林俊傑	王心凌
20	月光	Sugiyama Kouichi	橋本淳	王心凌
21	候鳥	方文山	周杰倫	S. H. E
22	天灰	施人誠/馮士哲	馮士哲	S. H. E
23	安靜	周杰倫	周杰倫	周杰倫
24	擱淺	宋健彰	周杰倫	周杰倫
25	鋼琴	許常德	桑田佳祐	范逸臣
26	星晴	周杰倫	周杰倫	周杰倫
27	軌跡	黃俊郎	周杰倫	周杰倫
28	天空	衛斯理/小米	衛斯理	蔡依林
29	倒帶	方文山	周杰倫	蔡依林
30	知足	阿信	阿信	阿信
31	寧夏	李正帆	李正帆	梁靜茹
32	小薇	阿弟	阿弟	黃品源

33	紀念	姚謙	Tanya Chua	蔡健雅
34	假如	姚若龍	Jun Hae Sung	信樂團
35	夢一場	袁惟仁	袁惟仁/許華強	那英
36	上弦月	方文良	方文良	許志安
37	我願意	姚謙	黃國倫	王菲
38	如果愛	姚謙	金培達	張學友
39	天黑黑	廖瑩如/April	李思菘	孫燕姿
40	我的愛	小寒	林毅心	孫燕姿
41	菊花台	方文山	周杰倫	周杰倫
42	珊瑚海	方文山	周杰倫	周杰倫 /LARA
43	髮如雪	方文山	周杰倫	周杰倫
44	七里香	方文山	周杰倫	周杰倫
45	流星雨	鄔裕康	平井堅	F4
46	就是愛	梁鴻斌	周杰倫	蔡依林
47	聽不到	阿信	阿信	梁靜茹
48	情非得已	張國祥	湯小康	庾澄慶
49	新不了情	黃鬱	鮑比達	萬芳
50	理想情人	艾利絲	小冷	楊丞琳
51	痴心絕對	蔡伯南	蔡伯南	李聖傑
52	大城小愛	王力宏/陳鎮川 /K. Tee	王力宏	王力宏
53	眼淚成詩	林夕	郭寶龍	孫燕姿
54	花的嫁紗	崔巖	崔巖	王心凌
55	千年之戀	F. I. R. /謝宥慧/林志 年	F. I. R.	F. I. R.
56	我們的愛	F. I. R. /謝宥慧	F. I. R.	F. I. R.
57	一了百了	武雄	詹凌駕	信樂團
58	戀人未滿	施人誠	鍾興民	S. H. E
59	愛很簡單	娃娃	陶喆	陶喆
60	黑色幽默	周杰倫	周杰倫	周杰倫
61	分手快樂	姚若龍	郭文賢	梁靜茹
62	除此之外	阿怪	陳達偉	范逸臣
63	千里之外	方文山	周杰倫	周杰倫
64	我只在乎你	慎芝	Takashi Miki	鄧麗君
65	城裡的月光	陳佳明	陳佳明	許美靜
66	孤單北半球	Benny C.	方文良	林依晨

67	如果雲知道	季忠平&許常德	季忠平	許茹芸
68	老鼠愛大米	楊臣剛	楊臣剛	王啟文
69	愛一直閃亮	瑞業	鄧智彰	羅美玲
70	至少還有你	林夕	Davy Chan	林憶蓮
71	隱形的翅膀	王雅君	王雅君	張紹涵
72	遺失的美好	姚若龍	黃漢青	張紹涵
73	一千年以後	李瑞洵	林俊傑	林俊傑
74	給我你的愛	顏壘軒	TANK	TANK
75	那年的情書	姚謙	黃國倫	江美琪
76	我會好好的	伍佰	伍佰	王心凌
77	花都開好了	施人誠	左安安	SHE
78	愛我還是他	娃娃/陶喆	陶喆	陶喆
79	如果有一天	易齊	郭文賢	梁靜茹
80	很愛很愛你	施人誠	玉城千春	劉若英
81	從開始到現在	李焯雄	Oh Seok Jun/You Hae Jun	張信哲
82	我多麼羨慕你	姚謙	張洪量	江美琪
83	一首簡單的歌	王力宏/陳鎮川	王力宏	王力宏
84	別在傷口灑鹽	鄔裕康	郭子	王豫民
85	檸檬草的味道	李焯雄	李德菘	蔡依林
86	威尼斯的淚	鄭中庸	永邦	永邦
87	他一定很愛你	李志清	李志清	阿杜
88	莫斯科沒有眼淚	許常德	伍樂城	Twins
89	我知道你很難過	胡如虹	葉良俊	蔡依林
90	忠孝東路走九遍	鄔裕康	郭子	動力火車
91	月亮代表我的心	孫儀	湯尼	鄧麗君
92	原來你什麼都不想要	鄔裕康	郭子	張惠妹
93	當你孤單你會想起誰	莎莎	施盈偉	張棟樑
94	男人不該讓女人流淚	王中言	黃國倫	蘇永康
95	I Believe	阿怪	金亨錫	范逸臣
96	Lydia	F. I. R.	F. I. R.	F. I. R.
97	Only one	李卓雄	黃韻玲	順子
98	Forever Love	王力宏/十方/何啓宏 /于景雯	王力宏	王力宏
99	KISS GOODBYE	王力宏	王力宏	王力宏
100	May I Love You	施人誠	Jung Yeon Jun	張智成

101	Through The Arbor-Kevin	Kevin Kern 鋼琴演奏曲
-----	-------------------------	------------------

納入的 101 首歌曲以及歌詞，使用歌曲 MP3 分段將可以達到國語流行音樂自動斷句系統與詞曲同步撥放系統的主要功能，建置國語流行音樂樂曲資料庫，是為了讓程式有更多的實驗數據，也可以讓這本系統可以應用在音樂動態歌詞的製作。

Appendix C - 資料剖析技術

本附錄將說明如何使用 C#.NET 中 `XmlTextReader` 類別，並從檔案讀取 Extensible Markup Language (延伸式標記語言, XML)。 `XmlTextReader` 提供使用者 XML 的直接剖析和 Token 處理，並且實作 XML 1.0 規格以及符合全球資訊網協會 (W3C) 之 XML 規格的命名空間 (Namespace)，提供了對 XML 快速、Token 化資料流存取。



圖 30 MSDN 之 `XmlTextReader`

設計目標為讀取國語流行音樂樂曲資料庫的 MusicXML 檔案，對音樂愛好者來說，在撥放音樂同時，也希望歌詞等資訊一併顯示，所以我們開發出一套整合音樂異質資訊的國語流行音樂自動斷句與詞曲同步撥放系統，解析 MusicXML，讓這本系統可以應用在音樂動態歌詞的製作，藉由軟體偵測，便能夠自動產生字幕，提供使用者更好的音樂饗宴。

實作方法：

以 C# 查詢 XML 文件中的元素方法如圖 31 所示，開啟檔名為 "NO.001 家後展

開.Xml"，解析此 XML 文件，使用迴圈 while 方式，查詢"NO.001 家後展開.Xml" 文件中所有元素。

```
XmlTextReader reader = new XmlTextReader("NO.001家後展  
開.Xml");  
while (reader.Read())  
{  
...  
}
```

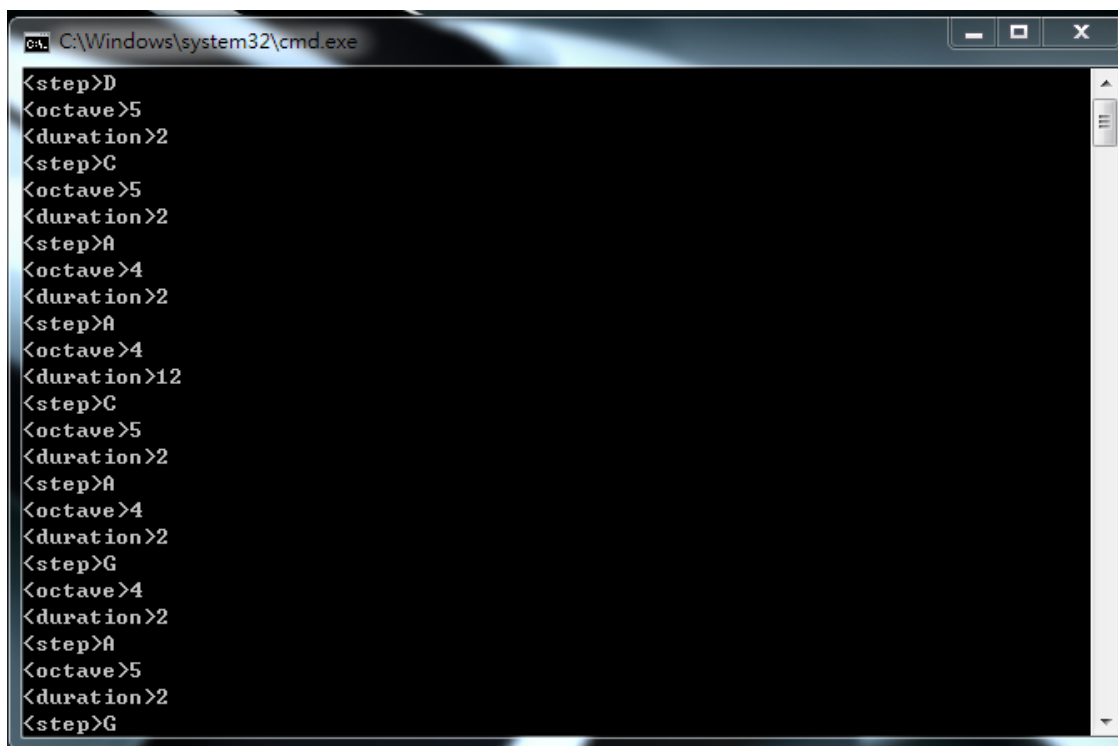
圖 31 查詢 XML 元素

透過 XML 解析，其名稱為找尋 MusicXML 中重要的元素。step 代表音符音高、octave 代表音符位於第幾個音樂八度中、duration 代表音符持續的時間。

```
case XmlNodeType.Element: // The node is an element.  
if (reader.Name == "step" || reader.Name == "octave" ||  
reader.Name == "duration")  
{  
a = reader.Name;  
reader.Read();  
Console.WriteLine("< " + a + ">" + reader.Value + "\n");  
}  
break;
```

圖 32 解析 MusicXML 重要元素

解析出來的 MusicXML 範例內容由圖 33 所示，可以提供本系統 MusicXML 的所有元素內容，解析此樂譜，進而達成國語流行音樂自動斷句與詞曲同步撥放系統的功能。



```
C:\Windows\system32\cmd.exe
<step>D
<octave>5
<duration>2
<step>C
<octave>5
<duration>2
<step>A
<octave>4
<duration>2
<step>A
<octave>4
<duration>12
<step>C
<octave>5
<duration>2
<step>A
<octave>4
<duration>2
<step>G
<octave>4
<duration>2
<step>A
<octave>5
<duration>2
<step>G
```

圖 33 解析 MusicXML 結果圖