

靜宜大學資訊學院

畢業專題成果報告書



透過樂音生成樂譜之研究

指導教授： 羅峻旗

專題學生： 廖詩涵、繆芯珮、施怡安、張馨元

目錄

壹、前言	2
貳、系統功能	2
參、研究方法與步驟	2
肆、系統畫面展示	6
伍、開發工具及使用環境	7
陸、結論及未來發展	7
柒、附錄	7
◆ 成本分析	7

壹、前言

現今，人工智慧的技術隨著軟硬體技術的發展已日趨成熟，在不同產業中也有了新的應用。目前的市場上有許多人工智慧應用在音樂上，例如人們使用音樂識別軟體來搜尋當下聽到的歌，或是利用語音轉換成文字等等。想要讓人工智慧用於音樂更為廣泛，我們發現到目前在作曲的輔助方面還稍有欠缺。

現今的鋼琴伴奏師會因應實際狀況不同，根據樂隊編制的需求進行多次的試奏後，再去編寫樂譜。我們所看到音樂家整理出來的譜例擁有多個聲部的伴奏，每個聲部的伴奏都需要進行獨立作譜，這些編寫樂譜的過程十分的麻煩且花費大量的時間成本，讓音樂創作和練習音樂的人們，帶來很大的不便。若我們能夠做出一套快速生成樂譜的工具，就能夠有效的解決編寫樂譜所帶來的不便及減少時間成本。

本專題是將聲音辨識的能力拓展到樂器演奏的音高辨識，並產生對應的樂譜。藉著聲音辨識來判斷樂器所彈奏的音高，了解演奏者的詮釋技法和演奏內容，進而將無形的音樂內容轉換成可看見的樂譜，同時也可依據這種方法來判定樂器音高的準確性。以樂器為例，研究一個音高辨識以及對應的樂譜生成方法。

貳、系統功能

1. 以中央位置前後八度音，共 36 個**單音辨識**。
2. 以單音為基礎，加入**和弦辨識**。
3. 將記錄的聲音轉換為可見的數字，**產生簡易樂譜**。

參、研究方法與步驟

本專題是透過樂器聲音頻譜的分析來判別樂器所發出來的音高，而大家也都理解樂器所發出的聲音音高與其聲音基頻有著非常直接的關係，這個關係還有一個特點，就是對於人類而言，音高的感受與基頻是成對數性的。例如中央 C 這個音高的頻率是低八度 C 頻率的兩倍，而高八度 C 的頻率又是中央 C 頻率的兩倍，以此類推。在音樂中所謂的「八度音」，代表的是某個音階與其高音音階的距離，如中央 Do 到高音 Do。我們在樂器中使用的十二平均律就是將這八度音分成十二個半音。

由於上述人類對聲音的感受，這十二個半音的頻率是依等比關係分佈，並不是以等差關係分佈的。基於這些理由，就有公定的標準把樂器音階中所有半音的頻率分別對應到一個編號 p ：

$$p = 69 + 12 \times \log_2 \left(\frac{f}{440} \right)$$

反過來，我們也可以由一個半音的編號得到其基頻 f ：

$$f = 440 \times 2^{(p-69)/12}$$

這個編號規則是以 69 號音也就是中央 A 為基準，其頻率為 440 Hz。由此公式可知，樂器中相鄰兩個半音之間的頻率比為 $12\sqrt{2} \approx 1.05946$ 。由這個公式，我們可以推算出中央 A 前後共三個八度音的頻率如表 1。

表 1: 中央 A 前後三個八度音的頻率

	A	$\flat B$	B	C	C^\sharp	D	$\flat E$	E	F	C^\sharp	G	G^\sharp
O4	220	233	246	261	277	294	311	330	349	370	392	415
O5	440	466	494	523	554	587	622	659	698	740	784	831
O6	880	932	988	1047	1109	1175	1245	1319	1397	1480	1568	1661

我們以樂器演奏的聲音作為聲音辨識的依據。因樂器的音高與聲音的基頻有直接的關係—基頻的頻率可決定聲音的音高。要分析樂器的音高，就必須對樂器的聲音進行頻譜分析。而頻譜分析的方法一般都是將聲音的波形資料通過傅利葉轉換得出頻譜。我們將藉由電腦程式進行離散傅立葉轉換，再依據轉換過後所得到的頻譜來判斷聲音的音高。

(1) 聲音數位化

取樣頻率為 44.1KHz，每個樣本以 16 位元做編碼。將聲音訊號切割成百分之一秒的區塊，然後對每一個區塊進行離散傅立葉轉換。

(2) 頻譜分析

將聲音的波形資料通過傅利葉轉換得出頻譜如圖 1。轉換的過程我們將 44.1 KHz 頻譜切割成 2048 個頻率，每個頻率之間的時間隔是 21.553 Hz。樂器的聲音經過傅立葉頻譜轉換之後，我們會得到一個包含 1024 個頻率振幅的向量，根據這個向量，我們會分析其各個頻譜振幅峰值的位置所在，然後再根據這些峰值所對應的頻率，來分析聲音的基頻，以及所衍生出來的泛音。

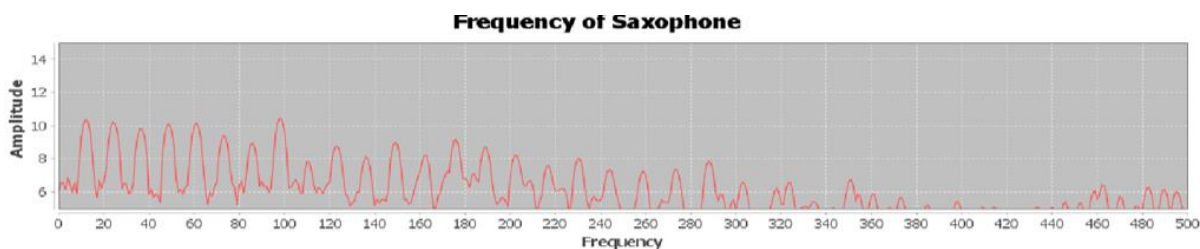


圖 1：頻譜圖

(3) 單音辨識

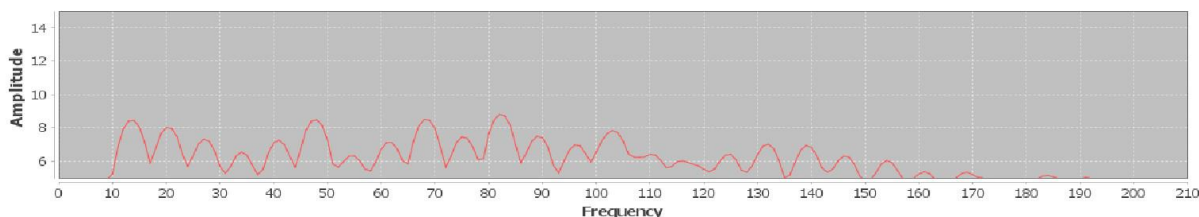


圖 2：低音頻譜峰值

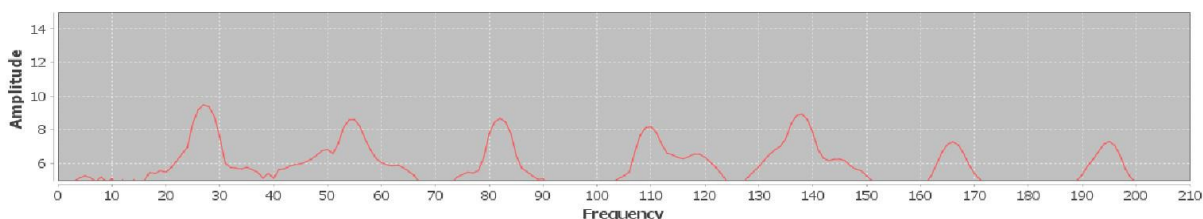


圖 3：高音頻譜峰值

參考圖 2 及圖 3，我們將樂器的聲音經過傅立葉頻譜轉換後可以發現，當樂器發出的聲音的較低的時候，頻譜圖的波峰間隔比較小。而在高音的頻譜圖中，主要峰值的間隔會變大，在這些間隔中會出現一些較不明顯且不規則的小峰值，這些不規則的小峰值會影響我們對音高的判斷。在分析樂器聲音的時候，我們希望能判斷出比較明顯的峰值而忽略那些振幅相對較小較不明顯的峰值。我們使用的演算法大致如下：

a. 對於每個頻率 f_i 而言 ($0 \leq i < 1024$)，設定其鄰近頻率的範圍。如果是頻譜最低頻的部分(頻譜中靠左邊的頻率，第一個峰值尚未判斷出來的時候)，我們設定其鄰近頻率範圍為 1 到 $2f_i - 1$ 。如果第一個峰值已經出現，且其峰值所在頻率為 f_b ，則我們設定 f_i 的鄰近頻率範圍為 $f_i - f_b + 1$ 到 $f_i + f_b - 1$ 。

b. 對於 f_i ，判斷自己振幅是否為鄰近頻率振幅中的最大值，如果是，則設定 f_i 為峰值的所在($p_i = 1$)，否則設定 $p_i = 0$ 。如果 f_i 是第一個出現的峰值所在，則我們暫時假設這個頻率 f_i 就是樂器聲音的基頻，此時我們設定基頻頻率 $f_b = f_i$ 。

c. 重複回到步驟 a 直到所有頻譜頻率都做出是否為峰值所在的判斷。

以上的演算法是找出聲音的基頻所在。在第一個峰值出現前，我們就以兩倍頻率作為判斷是否為峰值的範圍。當第一個峰值出現後，我們先假設其為聲音的基頻所在，那麼未來峰值出現的位置應該會在基頻的倍數上，所以我們判斷峰值的範圍必須規範在距離一個基頻的範圍以內，避免較高的峰值取代了相鄰的峰值，而遺失原本的峰值資訊。使用以上的演算法，我們可以準確的推算出頻譜中相對振幅較高的那幾個峰值頻率。

(4) 和弦辨識

參考圖 4~圖 6，原以為和弦會規律的依照單音的頻率產生出對應的頻率位置，經過多次的測試之後發現，當三個單音一起按的時候，和弦會產生出一個不規則的新頻率如圖 7。所以我們採用了人工學習的方式，去記錄每一個和弦所產生的頻率，並辨識

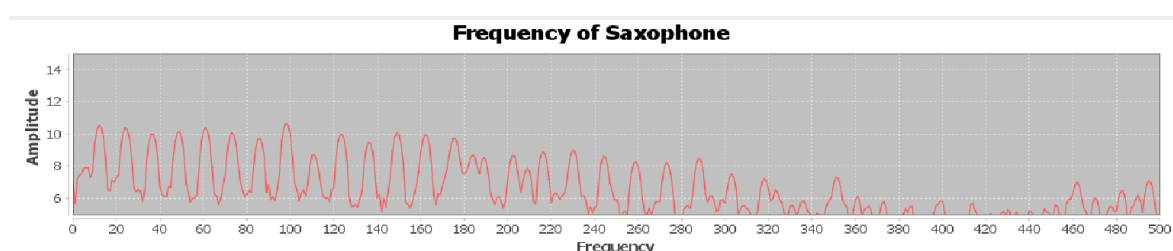


圖 4: 中央 Do 的頻率

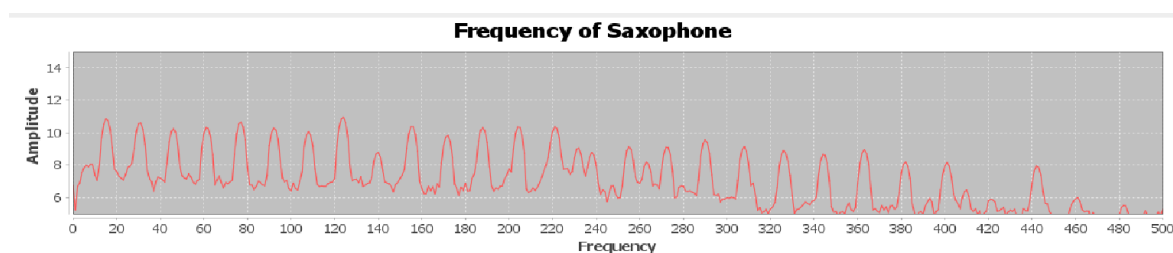


圖 5: 中央 Mi 的頻率

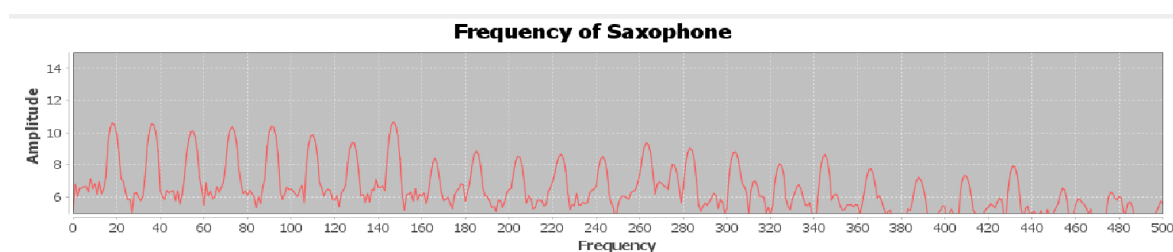


圖 6: 中央 Sol 的頻率

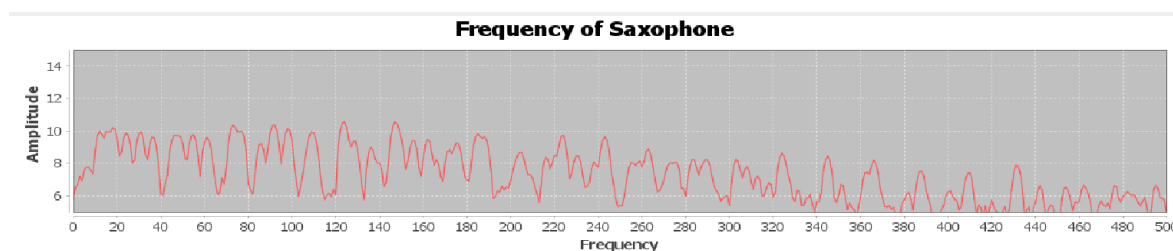


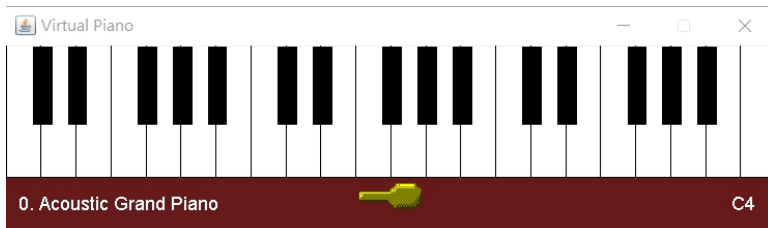
圖 7: 和弦 Do Mi Sol

(5) 產生簡易樂譜

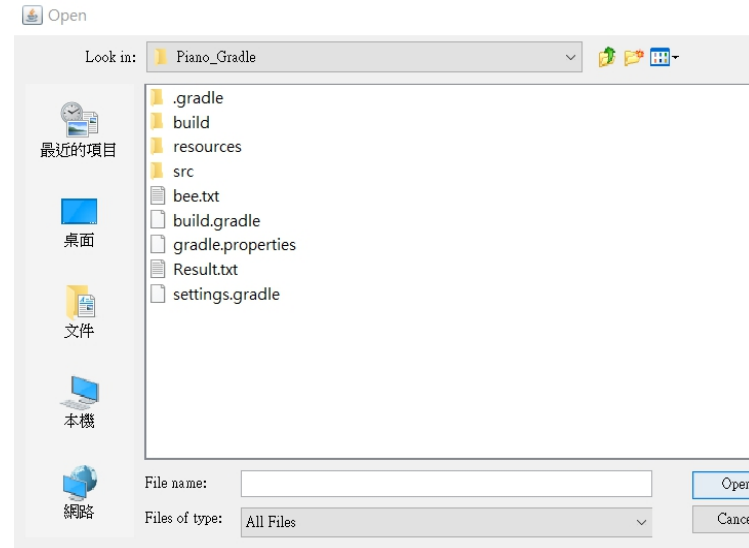
記錄使用者所彈奏的單音及和弦，轉換成簡易的數字樂譜，並顯示畫面。

肆、系統畫面展示

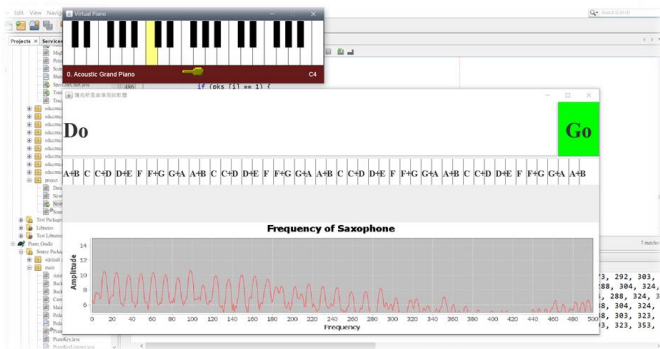
1. 模擬鋼琴



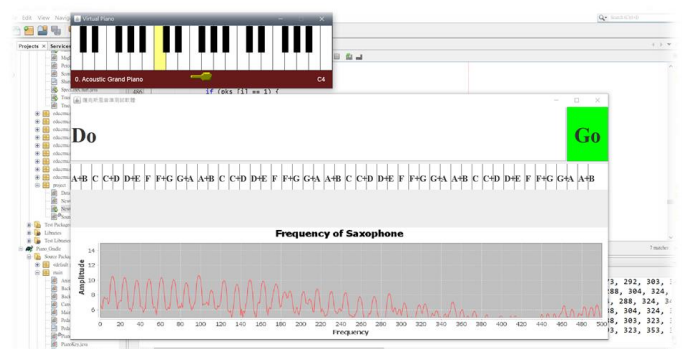
2. 選擇樂譜檔案



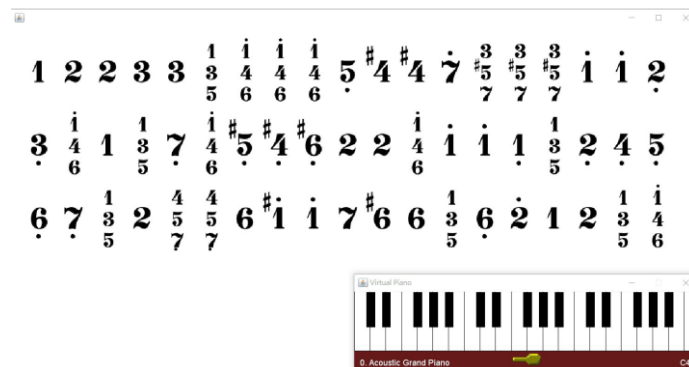
3.單音辨識



4.和弦辨識



5. 產生簡易樂譜



伍、開發工具及使用環境

- 開發工具

Netbeans

- 使用環境

Windows7/10

陸、結論及未來發展

1. 將簡易樂譜轉為正式的簡譜來閱讀。
2. 加上音的長短變化，辨識每個音符的節拍。
3. 使用機器學習的演算法，來幫助音高辨識更加的準確。

柒、附錄

- 成本分析

- 經費預算需求表（執行中所需之經費項目單價明細）

項 目 名 稱	說 明	單 位	數 量	單 價	小 計	備 註
				臺幣(元)	臺幣(元)	
電腦	專案之進行	部	1	26000	26000	由系上 實驗室提供
共 計					26000	