

# 即時音頻分析與音高檢測

王駿彥  
資訊工程學系  
國立暨南國際大學  
南投，臺灣  
rafawcy@gmail.com

**Abstract**—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract. (Abstract)*

**Keywords**—component, formatting, style, styling, insert (key words)

## I. 引言

隨著現代數數字音頻技術的快速發展，實時音高檢測在音樂學習、樂器調音和聲音識別等領域的應用越來越廣泛。音高檢測的核心技術之一是快速傅立葉轉換(FFT)，它能夠將時間域轉為頻域訊號，從而提取出音頻信號中的頻率成分。

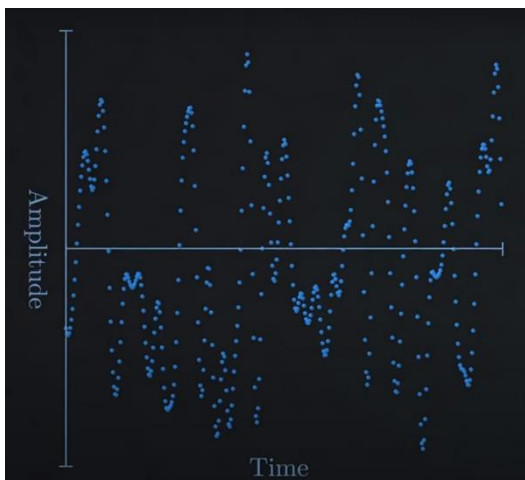
本文提出了一種基於 FFT 和 PyAudio 的實時音高檢測系統。該系統利用 Python 的 PyAudio 庫來捕捉音頻數據，並通過 FFT 進行頻譜分析，實時識別並顯示音符。這依實現不僅適用於音樂愛好者，還可以應用於各種音頻處理和分析系統中。

在接下來的部分，本文將介紹系統的設計與實現，包括信號處理流程、算法細節及實驗結果。通過這些介紹，讀者可以了解如何構建一個高校且準確的實時音高檢測系統。

## II. 快速傅立葉轉換的原理與技術

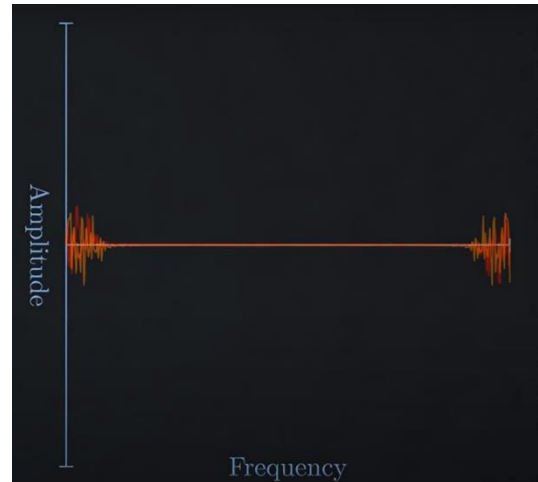
### A. 離散傅立葉轉換 Discrete Fourier Transforms (DFT)

我們平常收錄到的音波並不是完全都是連續的，只是因為每單位時間的能量大小組成的波形，因此，得到的波形並不是連續的，如圖(1)。



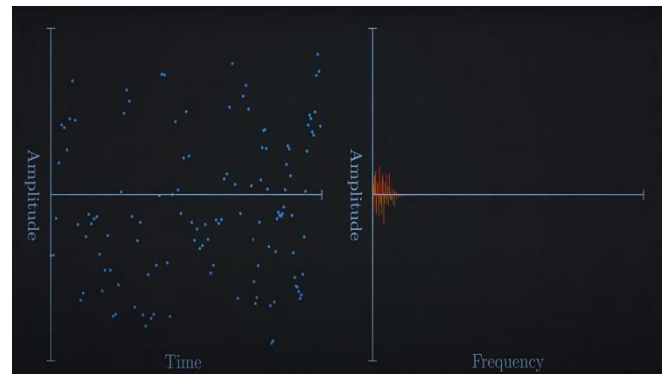
圖(1) 離散型樣本波形

故需要使用離散的傅立葉轉換(DFT)，這樣也可以得到離散且有限的頻譜，如圖(2)，而範圍則通常取決於計算的樣本數及樣本區間。



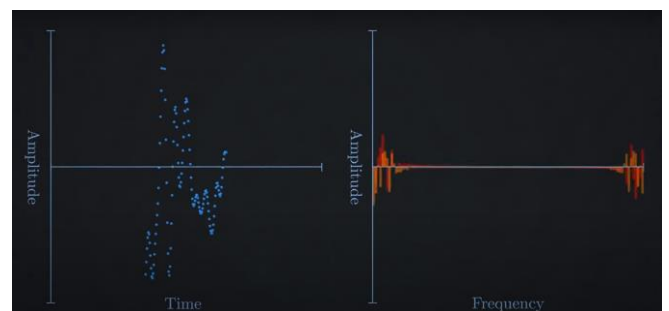
圖(2) 離散傅立葉變換後之頻域圖

樣本區間越長則每單位間距的樣本數越少，能觀測到的最高頻率也就越小，如。



圖(3) 樣本數少之時域及頻域

樣本數減少則降低頻率分辨率，因為無法通過不完整的波形計算出準確的頻率，如圖(4)。



圖(4) 樣本區間小之時域及頻域

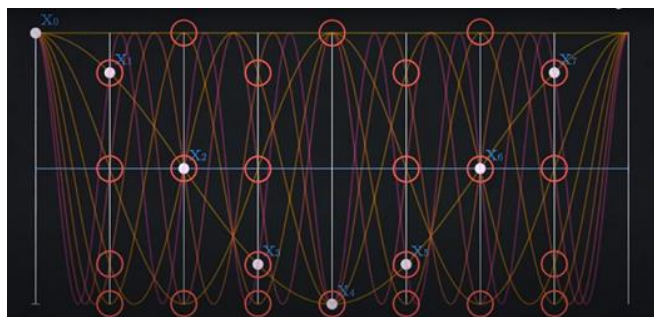
假如一個音波輸入有 8 個樣本數，則表示也會有 8 個頻率區間( $F_0 \sim F_7$ )， $F_0$  表示基頻，計算所有數據點對 x 軸之偏移總和， $F_n$  表示 n 倍的基頻，每次計算出正弦波與餘弦波的偏移量並存放進  $F_n$ ，公式如下面之(1)式。

$$X(k) = \sum_{n=0}^{N-1} e^{-i\frac{2\pi}{N}kn} x(n) \quad k = 0, 1, \dots, N-1 \quad (1)$$

但此方法缺點就是計算量過大，他的時間複雜度為  $O(n^2)$ ，通常的樣本數都是幾萬起跳，計算速度會太慢。

### B. 快速傅立葉轉換 Fast Fourier Transforms (FFT)

FFT 的概念就是以減少重複性的計算來加速 DFT 的轉換方式，以上述章節 II-A 來說，每個 sample 在與  $F_0 \sim F_7$  頻率產生的正弦波與餘弦波做乘法運算時，由於正弦波的週期性，這些不同頻率的波在某些點重和，以圖(5)為例，在中間的樣本點，四個奇數頻率( $F_{1,3,5,7}$ )與偶數頻率( $F_{0,2,4,6}$ )值都相同。原本要計算 8 次的乘法運算只需要計算 2 次就可以得到答案了。



圖(5) 重複運算之樣本點

實踐方法: 把樣本照偶數及奇數分成兩部分，分別對這兩部分的樣本點與  $F_0 \sim F_7$  這 8 個不同頻率波進行乘法運算，對於偶數樣本點部分，對比前四個頻率及後四個頻率( $F_0 \leftrightarrow F_4, F_1 \leftrightarrow F_5, \dots, F_3 \leftrightarrow F_7$ )，頻率波形會在某些點重和，也就是數值相等，而對於奇數樣本點，這些點則是互相抵消，也就是互為相反數，代表對於頻率後半部分無須進行所有的乘法運算，再使用分而治之法，重複分成偶數部分及奇數部分直到僅剩一個樣本點，每次分割都可以利用正弦波的對稱性減半計算次數，這樣可以大大的提升計算效率，計算公式如下面之(2)式。

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

## III. 系統設計與實現

### A. 硬體及軟體配置

- 硬體：
  - 計算機：用於運行音頻處理和分析的程序，具有足夠的計算能力來實時處理音頻信號。
  - 麥克風：用於捕捉環境中的音頻信號，要求具有較高的靈敏度和較低的噪音水平，以保證信號質量。
- 軟體：
  - Python：用於編寫音頻處理和分析的程序。
  - Pyaudio：用於音頻流的實時捕捉和播放。

- Numpy：用於科學計算，特別是快速傅立葉變換 (FFT) 的計算。
- Matplotlib：用於數據的可視化展示。

### B. 系統架構

系統主要包括以下幾種功能：

- 從麥克風中捕捉實時音頻數據。
- 使用 FFT 對音頻數據進行頻譜分析。
- 分析頻譜數據，確定音頻信號的基頻。
- 將基頻轉換為音高並實時顯示出來。

### C. 核心算法

基於傅立葉轉換的音高檢測算法主要步驟如下：

- 音頻信號采集：
  - 使用 PyAudio 庫進行實時音頻信號的采集。通過麥克風捕獲環境中的聲音信號，並將其轉換為數字信號以便進行後續處理。
- 傅立葉變換：
  - 采集到的音頻數據會分成若干幀，每幀包含固定數量的樣本（如 2048 個）。對每一幀數據使用 NumPy 庫中的快速傅立葉變換 (FFT) 函數進行頻譜分析，將時域信號轉換到頻域。
- 頻譜分析：
  - 在頻譜中找到具有最大幅值的頻率，該頻率即為基頻。基頻是最能代表當前音高的頻率。由於 FFT 會生成正頻率和負頻率，我們只需關注正頻率部分。
- 音高顯示：
  - 將計算出的基頻轉換為相應的音高。音高以音符名稱（如 C4、A4）表示，並顯示出來。這一步需要將頻率對應到音高，例如通過公式將頻率轉換為 MIDI 音符號，再根據 MIDI 音符號查找對應的音名。

## IV. 困難與挑戰

### A. 實時性要求

實時音高檢測系統需要在極短的時間內完成音頻數據的采集和處理，這對系統的性能提出了很高的要求。

解決方法：採用快速傅立葉變換 (FFT) 減少計算時間。

### B. 噪聲干擾

音頻信號中常常包含各種噪聲，這些噪聲可能會影響音高檢測的準確性。

解決方法：

- 預處理過濾：在進行傅立葉變換前，先對音頻信號進行預處理過濾，去除高頻或低頻噪聲。
- 設置振幅閾值：設置一個合理的振幅閾值，只有當信號的幅度超過該閾值時才進行音高檢測。

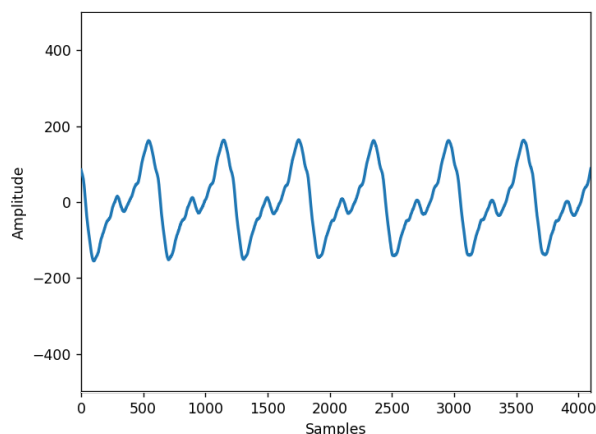
### C. 多頻信號分辨

當音頻信號中存在多個頻率成分時，如何準確地分辨出基頻是一個挑戰。

解決方法：在頻譜中過濾掉一些不可能的頻率成分，例如過高或過低的頻率。

## V. 成果展示

彈奏 D3 圖(如圖(6))，傅立葉轉換會擷取一個波形區間，做完運算後會得到多個頻率值索引，我取振幅影響最大的頻率值索引作為偵測到的頻率基底，也就是輸入波形的主要組成音，如圖(7)。



圖(6) D3 之波形

Frequency:	144.00 Hz	Note:	D3 -0.34
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18
Frequency:	145.35 Hz	Note:	D3 -0.18

圖(7) 預測之頻率

## VI. 結論

通過設計和實現基於傅立葉變換的實時音高檢測系統，成功地驗證了 FFT（快速傅立葉變換）在音頻處理中的

有效性和高效性。該系統能夠準確地捕捉並分析實時音頻信號，並將其轉換為相應的音高顯示出來。

在實驗過程中，我們發現該系統在實時性和準確性方面表現出色。具體來說：

- 實時性：系統能夠在極短的時間內完成音頻數據的采集、處理和分析，達到實時音高檢測的要求。
- 準確性：通過設置合理的頻率範圍和振幅閾值，系統能夠準確地識別出音頻信號中的基頻，並將其轉換為正確的音高。

該實時音高檢測系統具有廣泛的應用前景，不僅適用於音樂愛好者和教育領域，還可以應用於各種音頻處理和分析系統中。例如：

- 音樂教育：為學習者提供實時音高反饋，幫助他們準確地演唱或演奏。
- 音樂創作：幫助音樂製作人和作曲家實時分析音樂作品中的音高變化。
- 聲學研究：用於研究聲學特性和音高辨識的實驗工具。
- 語音識別：提供基於音高分析的輔助信息，提升語音識別系統的準確性。

## REFERENCES

- [1] Kulp, B. D. (1988, January). FFT BIN INTERPOLATION. [http://www.tedknowlton.com/resume/FFT\\_Bin\\_Interp.html](http://www.tedknowlton.com/resume/FFT_Bin_Interp.html)
- [2] Chen, C. (2004). A Study on FFT Architecture Design. <https://ndltd.ncl.edu.tw/cgi-bin/g32/gswweb.cgi/login?o=dnclcdr&s=id=%22092NDHU5442023%22.&searchmode=basic>
- [3] Cochran, W. T., Cooley, J. W., Favon, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G. C., Nelson, D. E., Rader, C. M., & Welch, P. D. (1967). What Is the Fast Fourier Transform? Proceedings of the IEEE, 55(10), 1664-1674. <https://ieeexplore.ieee.org/abstract/document/1447887>
- [4] Lin, V. (2017). 傅立葉轉換到底在轉換什麼? HackMD. <https://hackmd.io/@jkrvivan/B1wHF2lib?type=view>
- [5] 跨元探索. (2022, August 22). 離散信號的傅立葉各式轉換:DFT、IDFT,FFT(FastDFT).Vocus. <https://vocus.cc/article/63030791fd89780001b740ec>