**AIcup歌唱轉譜（Singing transcription）競賽 提供dataset說明**

**Dataset介紹**

本dataset為AIcup歌唱轉譜競賽當中，第一波釋出的dataset，共計包含500首歌曲，包含中文、日文、韓文、英文等語言的流行歌。

在釋出的每一首歌曲當中，包含兩個txt檔及兩個json檔。第一個txt檔當中包含一個Youtube連結，指向該首歌曲的原始影片。

第二個txt檔為該歌曲正確的轉譜結果，是主辦方找人標記的檔案。檔案中包含若干行，每一行都包含三個數值，以空格分隔，每一行都代表一個音的資訊，從左到右分別為那個音的onset, offset及pitch值，前兩個數值為浮點數，第三個數值則是正整數。在此，保證每一個音符都不會產生重疊，且前面一行的音符，在時間順序上，必定在後面一行的音符的前面。

26.379167 26.596908 50

舉例而言，上面這一行取自1\_groundtruth.txt的第一行，代表dataset當中第1首歌曲的第一個音。這一個音的起始時間是第26.379167秒，結束時間是26.596908秒，而這個音的音高則是50（以midi值表示），代表這個音的音高是D3。

vocal.json檔則是我們提供的pitch tracking的結果，以每0.032秒為一個frame，frame之間沒有overlap，給出每個frame的音高值，供參賽者參考。

最後，feature.json是我們提供的一些audio analysis常用的feature，同樣以0.032秒為一個frame，frame之間沒有overlap。資料格式為：

{

“time”: values,

“vocal\_pitch”: values,

“feature\_name1”: values,

“feature\_name2”: values,

…

}

其中提供一維time表示該frame對應到的實際時間，並且也包含一維vocal\_pitch，其pitch的數值和vocal.json相同。

**Pitch tracking結果的產生方式**

在此，我們首先使用一個以u-net的架構訓練出的人聲分離（Singing voice separation）網頁，分離出人聲，並且只使用這一部份，進行pitch tracking，求出json檔。在此使用的pitch tracking方法，是基於[1]當中所描述的方法，稍作修改而成。若參賽者對詳細的過程有疑問，可以再向我們提出。

**Feature 結果的產生方式**

Feature的抽取同樣針對經過人聲分離後的人聲，使用的函式庫為pyAudioAnalysis[2]，該函式庫共可抽取34種feature，我們提供了除了MFCC外的21種。

**Dataset架構**

在本資料夾底下，有500個資料夾，資料夾的名稱為歌曲的編號，從1編到500。每個資料夾當中，有兩個txt檔及一個json檔。假設歌曲編號為a，則Youtube連結的txt檔檔名為a\_link.txt，正確的轉譜結果的txt檔檔名為a\_ groundtruth.txt，而json檔的檔名則訂為a\_vocal.json和a\_feature.json。舉例而言，在資料夾「1」當中，有1\_groundtruth.txt、1\_link.txt、1\_vocal.json和1\_feature.json共四個檔案。為了方便參賽者查找，我們有提供一個簡易的python程式traverse.py，可以讀取所有的檔案。

**參考資料（Reference）**

1. J.-C. Chen, J.-S. Roger Jang: “TRUES: Tone Recognition Using Extended Segments,” ACM Transactions on Asian Language Information Processing, No. 10, Vol. 7, Aug 2008.（原論文推薦使用AMDF，但我們提供的pitch tracking的結果是使用ACF，並搭配該論文中的DP方法所求得的）
2. <https://github.com/tyiannak/pyAudioAnalysis>