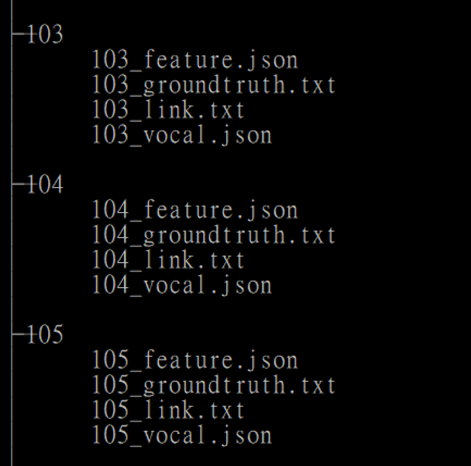
**AIcup歌唱轉譜（Singing transcription）競賽 資料說明**

本dataset包含兩個部分，分別是training set（500首歌曲）與test set（1500首歌曲）。其中，training set裡面的每一首歌都有提供標準答案，但是test set的每一首歌並沒有提供。

training set當中的每一首歌曲，結構大致如下：



在釋出的每一首歌曲當中，都包含兩個txt檔及兩個json檔。第一個txt檔（如103\_link.txt）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（如103\_groundtruth.txt）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，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。這三個數字代表了一個音符的資訊。

在此，我們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 number表示），代表這個音的音高是D3。

至於json檔的部分，第一個json檔（如103\_vocal.json）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。第二個json檔（如103\_feature.json）則包含了更詳盡的一些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, ……],

…

}

要注意的是，feature檔當中，有提供一維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。

至於test set的部分，格式和training set完全一樣，但我們當然不會提供groundtruth.txt這個正確答案的檔案，而是留給參賽者自行預測。