The Making of Extreme Vocals Archive

Chin Yun Yu

Abstract

本報告詳述了我製作極端唱腔資料集(Extreme Vocals Archive，簡稱 EVA)的過程。採集了數百段長度不等的吼腔片段後，交由業餘吼腔歌手來標記使用的技巧，最後用 SVM 展示了用此資料集做唱腔分類的結果。

Introduction

在現今 MIR 的研究中，有關人聲方面的研究，都是以有音高的前提去進行；但在一些音樂類型，如 Metal/Metalcore/Hardcore/饒舌等，主唱其技巧的展現並不在旋律，而著重在音色與節奏上的變化。由於這方面的研究並不多，也沒有公開的資料可供研究(多為自行收集)，故整理與製作這類極端唱腔的資料我認為是一個很好的開始。

Data Collecting

我從 [1]這個網站的 Rock/Punk/Metal 分類，選擇了 16 首[1]我認為有嘶吼唱腔的曲子；再從分軌檔中選出有人聲的音軌，使用 Audacity 將其切成一個個段落。最後總共有 565 段長度由 0 至 40 秒不等的人聲片段。

Data Labeling

標記這些資料的類別，在考量有限的時間及題目的發展性，我決定以吼腔使用的技巧當作各片段的類別。與一位業餘的吼腔主唱辛宥炎[2]討論後，總共整理出五種類別：Vocal Fry，False Chord，Vocal Fry Dominated，False Chord Dominated 以及不屬於以上種類的非嘶吼唱腔。Vocal Fry 通常在 Hardcore/Punk/Metalcore 等類型較常使用，給人的感覺音調較高且刺耳；False Chord 聽起來則較低沉與粗厚，常用在 Death Metal 等。Dominated 類別代表兩種唱腔都有使用，但其中一種佔的比例較高。之後交給辛宥炎標記出每個片段最主要的唱腔，忽略有唱腔切換的狀況。最後的類別數量如 Table 1，可以發現唱腔混和的片段佔了約 77%，其中又以 False Chord Dominated 最多，約 53%，相當的不平均。

---

[1] The songs and their bands name are: The Apprehended: 'Still Flyin', Cnoc An Tursa: 'Bannockburn', The Complainiacs: 'Etc', Dark Ride: 'Burning Bridges', Dark Ride: 'Hammer Down', Dark Ride: 'Piece Of Me', Death Of A Romantic: 'The Well', Decypher: 'Unseen', Headwound Harry: 'XXXV', Hollow Ground: 'Ill Fate', Hollow Ground: 'Left Blind', Last Legacy: 'Who's Who In Hell', Storm Of Particles: 'Of Ice And Hopeless Fate', Timboz: 'Pony', Titanium: 'Haunted Age', Wall Of Death: 'Femme'.

[2] 本人有在經營一個 metal youtube channel, https://www.youtube.com/channel/UCu1FDV-RQS8V3s82RuNmOWg.

Table 1. Number of samples in each category.

Category	*Number of samples*
Vocal Fry	40
False Chord	65
Vocal Fry Dominated	135
False Chord Dominated	302
Others	23

Features Extraction

我使用 librosa [2]來將各片段縮減成 feature。Feature 的選擇則是除了 Chroma 相關的 feature 之外都用(假設吼腔和音高沒有相關)，並將時間維度縮減成統計值(平均值、標準差、skewness 等等)。詳細 feature 為 Table 2。

Table 2. Features sets dimension.

Feature set	*Dimension*
MFCC	140
RMSE	7
ZCR	7
Polynomial coefficients	14
Spectral centroid	7
Spectral bandwidth	7
Spectral contrast	49
Spectral flatness	7
Spectral rolloff	7

Looking Inside

在實際使用這些資料前，我認為多觀察一下吼腔的特性對之後的研究會有幫助。從 Figure 1 可以看到，約在 400hz~1000hz 會有很強的能量分布，會把人聲的基頻與泛音壓下去；且該能量分布也有其泛音，但並不是等距分布。或許可以把吼腔的訊號，想成是兩種訊號的疊加：

$$X_{scream} = X_{f0} + \varepsilon_{noise}$$

Figure 2 和 Figure 3 則是將前述取得的 feature 用 Linear Discriminant Analysis(LDA)降成二維後畫出的分布。可以看到大致上 Vocal Fry Dominant 和 False Chord Dominated 是分成兩個區塊的，代表我們取的 feature 有用，可以幫助之後的分類。另外 Vocal Fry 和 Vocal Fry dominated 幾乎疊在一起，不排除可能是標 label 有誤判的情形，又或是 sample 數量太少，使 Vocal Fry 和 Dominated 的差別無法在 LDA 顯示出來。

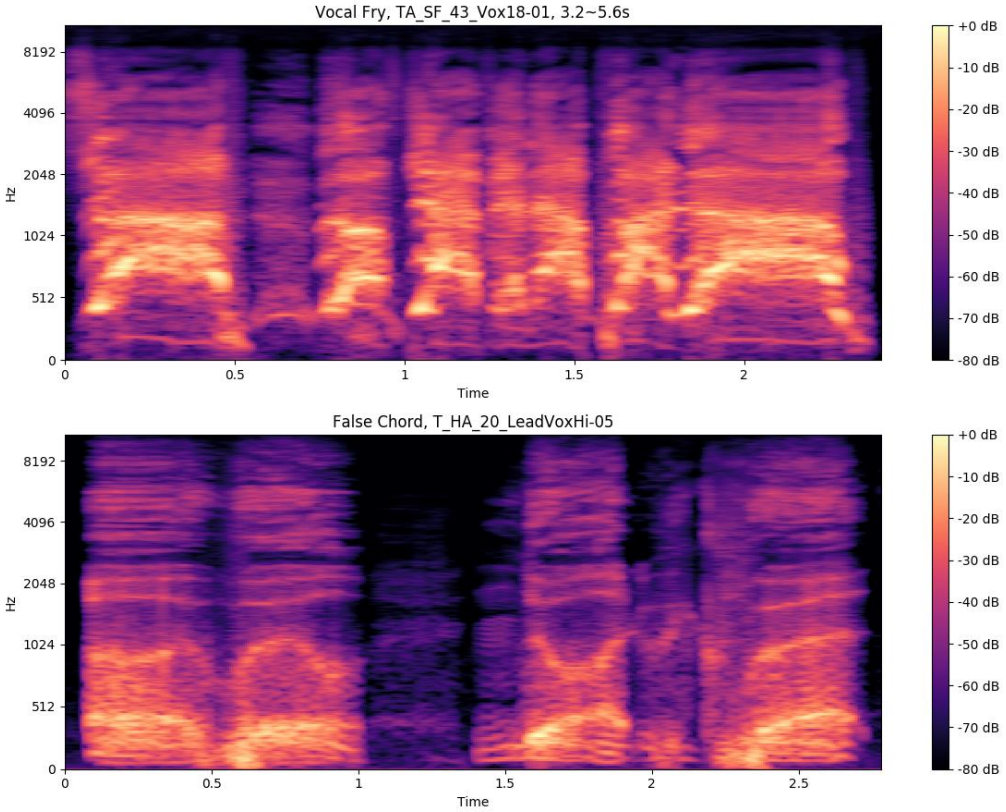

Figure 1. Melspectrogram of Vocal Fry (upper) and False Chord (lower) samples from EVA.

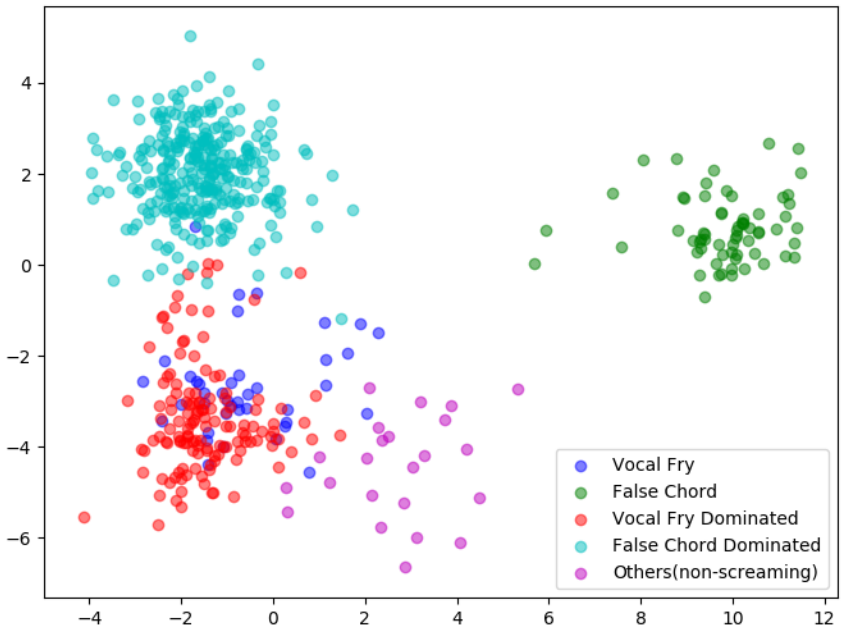

Figure 2. The distribution of each category after reduced the dimension using linear discriminant analysis.

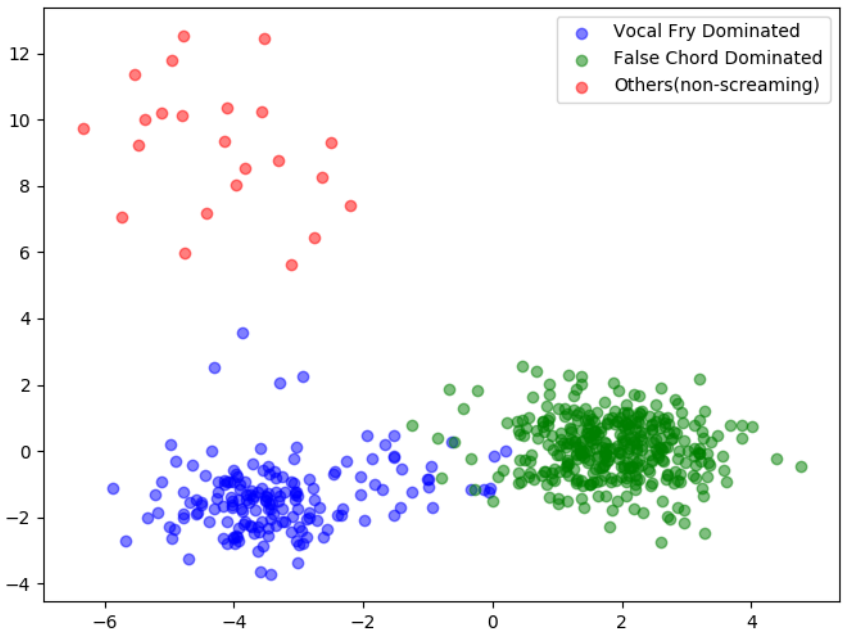

Figure 3. The distribution of each class after merged Vocal Fry to Vocal Fry dominated and merged False Chord to False Chord dominated, and reduced the dimension using LDA.

Usage: Vocal Techniques Classification

有了上述的 feature，就可以使用傳統機器學習的 SVM 模型來分類。我使用了 sklearn [3]的 SVC，kernel 為線性，切出 70%的資料用來訓練，其他做測試。因為資料分布不均，所以有依照類別數量比例乘上權重，結果為 Table 3。由於純 Vocal Fry 和 False Chord 的數量太少，所以多加了一個如 Figure 3 方式合併類別的結果。嘗試不同 feature 組合後，最適合的組合是 MFCC/spectral bandwidth，平均準確率在 80%上下，表示 spectral bandwidth 在辨別唱腔方面有指標性。

Table 3. Classification results.

Features set	*Accuracy (original label)*	*Accuracy(merged label)*
MFCC	74.97%	75.82%
MFCC/rolloff	71.94%	81.43%
MFCC/contrast	71.82%	75.18%
MFCC/flatness	72.93%	74.54%
MFCC/centroid	77.66%	81.08%
MFCC/bandwidth	78.77%	**86.0%**
MFCC/bandwidth/zcr	**79.0%**	**86.0%**
MFCC/bandwidth/rmse	**79.0%**	**86.0%**
All	69.94%	73.70%

Discussion

雖然在唱腔辨識上的準確率還不錯，但這樣的方式忽略太多時間上的資訊。從 Figure 1 可以觀察到，這些吼腔的頻譜是有相似 pattern 的，或許可以用一些深度學習的做法(如 CNN)，來善用這些特性做辨別。與之相對的，label 也可以加入時間因素，例如某段時間是 Vocal Fry、某段是 False Chord 這樣，做更詳細的 labeling。又或是加入 F0 的資訊，讓這個資料集也可以用在 F0 的偵測上。

Conclusion

實驗的結果顯示吼腔技巧分類還有很多進步的空間，希望這樣的資料集能幫助之後相關領域的研究。

參照

[1] M. Senior, "The 'Mixing Secrets' Free Multitrack Download Library," [Online]. Available: http://www.cambridge-mt.com/ms-mtk.htm. [Accessed 25 6 2018].

[2] B. McFee, M. McVicar, S. Balke, C. Thomé, C. Raffel, D. Lee, … A. Holovaty, "librosa/librosa: 0.6.1," 24 5 2018. [Online].

[3] Pedregosa et al., "Scikit-learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825-2830, 2011.