Scream Detection in Heavy Metal Music

Vedant Kalbag
Music Informatics Group
Georgia Institute of Technology, USA
vedant.kalbag@gatech.edu

Alexander Lerch
Music Informatics Group
Georgia Institute of Technology, USA
alexander.lerch@gatech.edu

ABSTRACT

Harsh vocal effects such as screams or growls are far more common in heavy metal vocals than the traditionally sung vocal. This paper explores the problem of detection and classification of extreme vocal techniques in heavy metal music, specifically the identification of different scream techniques. We investigate the suitability of various feature representations, including cepstral, spectral, and temporal features as input representations for classification. The main contributions of this work are (i) a manually annotated dataset comprised of over 280 minutes of heavy metal songs of various genres with a statistical analysis of occurrences of different extreme vocal techniques in heavy metal music, and (ii) a systematic study of different input feature representations for the classification of heavy metal vocals.

1. INTRODUCTION

Vocals in heavy metal music can be very different to those in other styles. Heavy metal vocalists use a variety of techniques, colloquially known as screams or growls, which are produced by modifying the length and shape of the vocal tract [1]. These screamed vocals serve one of two purposes: they are either low and beast-like to accentuate the aggressive, darker themes of heavy metal, or high and screechy, to stand out from the otherwise aggressive sounds of the distorted electric guitar [2]. In this paper we explore methods to detect and classify the type of vocal technique being used by a vocalist.

The automatic identification of different type of vocal techniques in heavy metal could, for instance, inform genre classification systems and aid music recommendation systems based on preference for a specific vocal type. Vocal detection for heavy metal music could also improve vocal extraction as well as (lyrics) transcription for this genre.

Nieto introduced the term 'Extreme Vocal Effects' or EVEs to describe the vocal styles present in heavy metal [3]. These EVEs fall into 3 main categories:

- *Growls*: Growls are common in death metal. They are very noisy and the fundamental frequency is rarely perceived. They are usually loud and produce a high amount of spectral variation [1, 4]
- *Fry Screams*: Fry screams are similar to growls, but are brighter and not as loud. They are produced by a series of irregularly spaced glottal pulses that are induced by inhaling or exhaling [5]
- *Rough Vocals*: Rough vocals are obtained by adding variations in the vocal tract to obtain a harmonically richer spectrum [1, 6]. This is much more common in rock than in metal (e.g., for bands such as *Foo Fighters* and *Breaking Benjamin*).

Figure 1 shows a sample spectrogram for each class. Distinct patterns in the low and mid fry scream can be observed that distinguish them from the other types of screams. The high screams occupy a higher portion of the spectrum as well. It is important to note that, in these examples, the mid fry scream appears to have lower frequency content than the low fry scream. This is because these are examples chosen from different vocalists, and the perceived type of scream varies according to factors discussed in Sect. 3.

Some subgenres of metal also involve sung or 'clean' vocals. In this paper, 'screams' and 'growls' will be used to describe the overall style of distorted heavy metal vocals, and 'clean' will be used to describe sung vocals. The term growl usually refers to the low pitched, rough sounds uttered by animals. Humans occasionally use growl-like voices to express strong emotions. Examples of 'growl' phonations have been seen across the genres of jazz, blues, gospel, samba, country and pop. In ethnic music, the growl is found in *umngqokolo* (the vocal tradition of the Xhosa people), and throat singing (Tuvan and Mongolian) [7]. However, in recent times growls are most strongly associated with metal vocals.

Extreme metal screams can be performed by either inhaling or exhaling which has a noticeable effect on the timbre of the sounds produced. However, in most modern metal, screams are produced by exhaling, and so our work will focus on these types of screams.

The remainder of this paper is structured as follows. After an overview of related work in Sect. 2, a new publicly available dataset is introduced in Sect. 3. We describe several benchmark systems for detection and classification in Sect. 4 and present the corresponding results in Sect. 6. The conclusion in Sect. 7 summarizes the main contributions in gives a brief outlook of future work.

2. RELATED WORK

While there exists, to the best knowledge of the authors, no previous work on the automatic categorization of heavy metal vocals, one related field is the detection of screams in

Copyright: © 2022 Vedant Kalbag et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

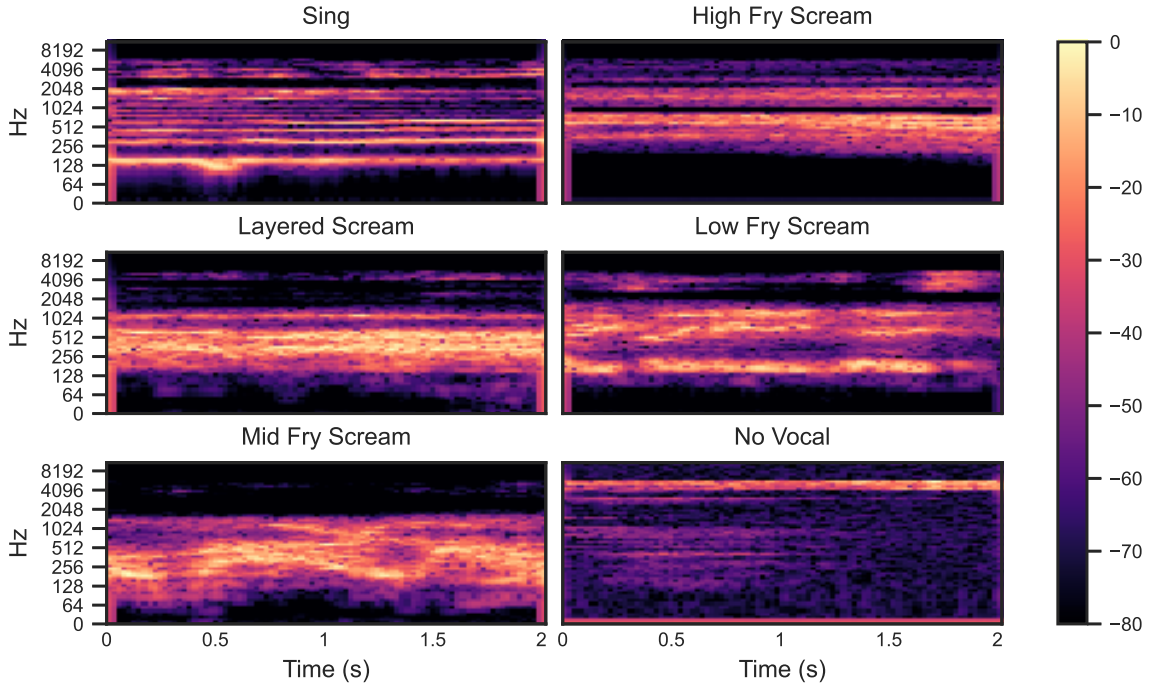

Figure 1. Example spectrogram representation of different screams.

urban environments.

In the previous section, we introduced different types of screams in metal music. Here, we will discuss past work with scream detection in general, followed by related work on screamed vocals in heavy metal music.

Prior work in detecting screams aims at the detection and localization of screams in urban sound, the detection of screams in subways, scream and shout recognition in noise, and scream detection for home applications. Various approaches were taken to achieve these tasks. Huang et al. used Mel Frequency Cepstral Coefficients (MFCCs) and Support Vector Machine (SVM) to classify blocks of audio captured from a microphone array into the two classes scream and non-scream [8]. Rabaoui et al. use a one class SVM to classify a sound into 9 categories, including scream, gunshot, explosion, door slam or dog barks, using features such as spectral centroid, spectral roll-off, zero-crossing rate, MFCCs and Linear Predictive Coding Coefficients (LPCCs) [9]. They also included the first and second derivatives of these features, but determined that they were not helpful in improving performance. Lafitte et al. used a deep neural network approach with MFCCs to detect shouted voice/screams in subway trains, and classify audio into shout, conversation and noise [10]. Other work in detecting screams in noise also uses MFCC and spectral entropy features with GMM classifiers to achieve this task [11–13]. The best performing of these methods was able to achieve equal error rates (EERs) of 0.3% and 0.8% under 0dB and -5dB signal to noise ratio (SNR) conditions. This approach, while useful in identifying screams in noisy conditions, cannot be translated well to detecting screams in music since the noise added was that of subway stations, trains and air conditioners.

Most work related to heavy metal vocals focuses on the physiology of screamed vocals [7, 14–16], their spectral properties [17], and exploratory acoustic feature analyses [7, 18]. There has been limited work on detecting and classifying the types of vocals present in heavy metal. Nieto uses k-means clustering to group different vocal styles into the three classes *Growl*, *Fry Scream*, and *Roughness* [1]. The dataset used consisted of labeled recordings of the 6 vocalists' screams. While this work was successful at grouping similar classes together, it could not predict the type of EVE present. Due to a lack of data with start and end times of vocal events annotated, a sliding window approach similar to Huang, where the scream detection algorithm is applied to every block in a sliding window to determine the start and end times of a scream [8] could not be implemented, and hence identifying when a scream occurs, or identifying what different kinds of screams are present within one file were not possible.

3. DATASET

Currently, there exists no publicly available dataset with annotated vocals for heavy metal. To enable this study, as well as to facilitate future research on this topic, we present the newly created *Metal Vocal Dataset* (MVD). This dataset consists of 57 songs from 34 bands and 47 albums. The list of songs can be found in the appendix. Most of these songs were released during the last two decades, since use of vocal effects beyond Mid Fry screams has increased in this period.

A playlist containing all the songs present in the dataset

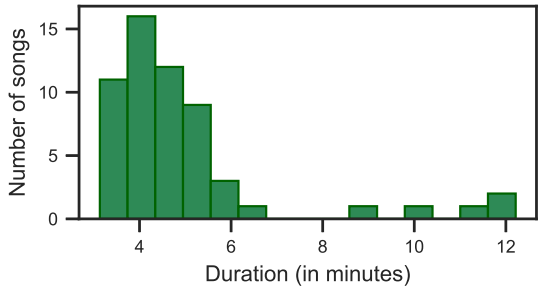

Figure 2. Distribution of dataset based on song length in minutes.

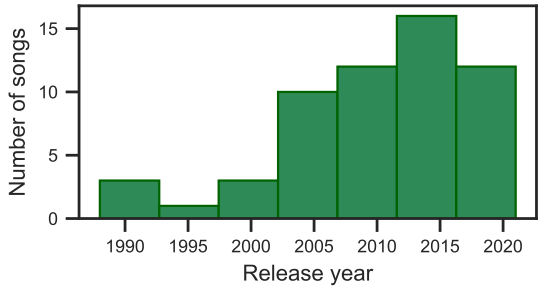

Figure 3. Distribution of dataset based on release year.

was created. [1] The distribution of the songs selected for the dataset based on the year of release is shown in Figure 3.

The annotations have been released under the MIT license and are available online. [2] The audio files themselves are not included, but can be retrieved using a script provided in the repository.

3.1 Data Selection

The songs selected were from genres such as death metal, groove metal, progressive metal, black metal, and metal core. The traditional subgenres of death metal, black metal and groove metal were included as they contain mostly one class of screams (mid fry screams), while modern subgenres such as metal core and progressive metal were chosen since a wide variety of vocal effects are used in these genres. The songs were selected with the aim to capture a wide variety in vocal styles and are listed in a playlist. [1]

3.2 Dataset Statistics

The distribution of the songs selected for the dataset based on the year of release is shown in Fig. 3. The increase for more recent years reflects the increased use of vocal effects beyond mid fry screams.

There are a total of 281.6 min of audio across the 6 classes (including the 'no vocal' class). The class distribution in the dataset is visualized in Fig. 4 based on the total time annotated in seconds. The Mid Fry scream is the largest part

[1] https://tinyurl.com/metal-vocal-dataset-playlist
[2] https://github.com/VedantKalbag/metal-vocal-dataset

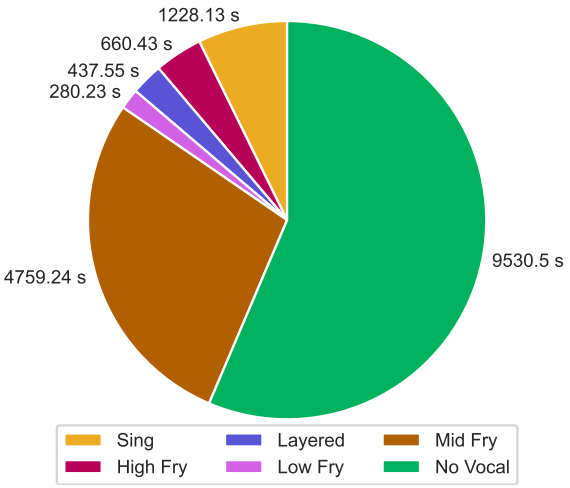

Figure 4. Distribution of dataset based on total time per class.

of the dataset; although the songs were selected carefully to contain all different classes, the Mid Fry scream is most prevalent in modern metal music.

3.3 Data Split

The data was split into 3 subsets for training, testing, and validation. This was done after division of audio files into 1 second blocks as described further in Sec. 4. Since the class distribution was heavily skewed towards blocks labeled 'no vocals', the dataset was undersampled to balance out classes. All classes that had more samples than the class with minimum samples were undersampled to the nearest thousand, for both the 3-class as well as the 6-class problem.

The data is accompanied by a recommended split into the subsets train, validation, and test (approx. 70:15:15). The data was split such that no band's songs are present in both the training and test/validation sets. Undersampling was applied before the split to balance the class distribution, as undersampling after the split would lead to considerably smaller test and validation sets. The blocks were first divided into an approx. 70:30 split, ensuring that no band was present in both subsets. This split at a band level was done to avoid overfitting any one vocalist/band and hence giving false results. The 30% split was then divided into two equal subsets at random. This was done because when restricting one band to be in either the test or validation set only drastically reduced the size of these sets, and would render them useless. In addition, a recommended split with imbalanced class distribution containing all data is provided as well.

3.4 Annotation Methodology

Since most screams in modern metal are variations of a fry scream, we have focused on these for our dataset. The variations are caused by a change in the shape and length of the vocal tract, where lengthening the vocal tract makes

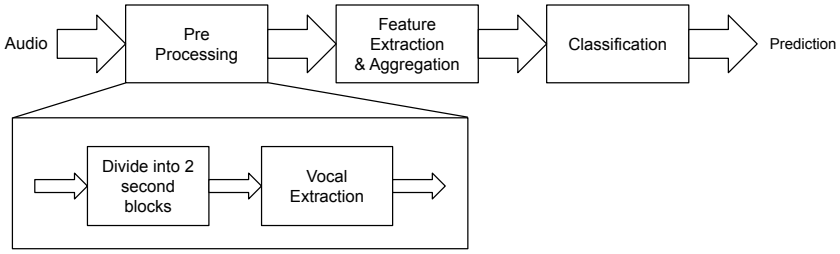

Figure 5. Block diagram of benchmark systems.

the scream sound lower, and vice versa. We have defined 3 fry scream categories based on the perceived sound: High, Mid, and Low. Thus, the vocal events were annotated with the following class labels: *Sing*, *High Fry* scream, *Mid Fry* scream, *Low Fry* scream, and *Layered* scream. The class labeled 'layered' contains combinations of 2 or more other classes simultaneously (e.g., Mid Fry screams and Sing, or both High and Low Fry screams).

These songs were manually annotated using *Sonic Visualiser* so that the maximum time difference between the start or end of a vocal event and the annotation is less than 0.5 s. The start and end points of the vocal event were localized visually based on the spectrogram of the audio file and validated aurally.

An important consideration is that the categorization of some screams is subjective, and two individuals may assign class labels differently. For example, a 'low-sounding' Mid Fry scream could be perceived as a 'high-sounding' Low Fry scream, and vice versa. As the main criteria for labeling the screams, the vowel characteristics of the sound were used. Typically, a Low Fry scream will have dark vowel characteristics (*/o/* or */u/*), a Mid Fry scream will have vowel characteristics similar to */a/*, and a High Fry scream will have characteristics around */e/* or */i/*. The labels were, thus, assigned based on how the scream sounded with respect to the perceived vowel characteristics; for instance, a scream with prominent low frequencies and vowel characteristics of */u/* or */o/* was labeled as a Low Fry scream.

4. BENCHMARK METHODS

A block diagram of the systems created as a benchmark for future work is shown in Fig. 5, and is described in detail in the following.

4.1 Pre-processing

The audio files were passed through the Spleeter source separation algorithm [19] to separate the vocals from the other components and then divided into overlapping blocks of length 2 s with a 1 s hop size. Each 2 s block is one observation to be classified. All audio files were resampled to a sample rate of 44100 Hz, normalized and downmixed to mono.

4.2 Input Representation

The baseline set of features consists of low level temporal and spectral features that are commonplace in Music Information Retrieval tasks. These features are: 13 MFCCs and Delta MFCCs, RMS, ZCR, Spectral Centroid, Contrast, Flatness and Roll-off (for a feature definition see [20]). These features were extracted using the Librosa python library [21], with a window size of 2048 samples and a hop size of 1024 samples. In addition, VGGish features [22] and the Log-Mel Spectrogram were extracted.

We divide these features into the following feature sets:

1. Feature Set 1: 13 MFCCs, Delta MFCCs, RMS, ZCR, Spectral Centroid, Contrast, Flatness and Roll-off
2. Feature Set 2: VGGish Features
3. Feature Set 3: 13 MFCCs and Delta MFCCs only
4. Feature Set 4: RMS, ZCR, Spectral Centroid, Contrast, Flatness and Roll-off
5. Feature Set 5: Log Mel Spectrogram

4.3 Feature Aggregation

All features in Feature Set 1 were aggregated by taking the mean and standard deviation across each audio block (with duration 2 s). The features in Feature Set 1, 2, 3, and 4 were all z-score normalized across the entire training set to return a feature vector with 0 mean and unit standard deviation. The mel spectrogram input was converted to log scale before use.

4.4 Classifiers

Two multi-class classifiers were used to classify each audio block based on the feature vector. The different classifiers used are a Support Vector Machine (SVM) and a Convolutional Neural Network (CNN). The CNN consists of 3 convolutional layers with dimensions 256, 512, and 1024, each followed by max pooling, respectively, 3 dense layers with dimensions 256, 64, and 16, and an output layer.

5. EXPERIMENTS

The system was tested for two different sets of labels: a 3 class problem (sing, scream, no vocal), as well as a 6 class problem (containing all the 5 labels from the dataset as well as no vocal).

Configuration	acc	bal-acc	f1
Feature Set 1 + SVM	82.20	82.10	82.18
Feature Set 2 + SVM	82.06	82.23	82.10
Feature Set 3 + SVM	77.12	76.95	77.21
Feature Set 4 + SVM	79.55	79.40	79.60
Feature Set 5 + CNN	**87.33**	**87.58**	**87.42**

Table 1. Results for the 3-class problem in Exp. 1 (values shown in %)

5.1 Experiment 1: 3-Class Problem

All scream classes are combined into a single class, resulting in the target set of classes *Sing*, *Scream*, and *No Vocal*. The following configuration were evaluated:

1. Feature Set 1 + SVM
2. Feature Set 2 + SVM
3. Feature Set 3 + SVM
4. Feature Set 4 + SVM
5. Feature Set 5 + CNN

5.2 Experiment 2: 6-Class Problem

As opposed to Experiment 1, Experiment 2 treats each scream class separately, resulting in the target set of classes *Sing*, *Low Fry*, *Mid Fry*, *High Fry*, *Layered*, and *No Vocal*. This experiment investigates the two best-performing SVM configurations and the CNN configuration from Exp. 1:

1. Feature Set 1 + SVM
2. Feature Set 2 + SVM
3. Feature Set 5 + CNN

5.3 Performance Metrics

The performance metrics used in this study are:

1. Accuracy: *acc*
2. Macro-Accuracy: *bal-acc*
3. Balanced F1 Score: *f1*

These metrics were computed with the sklearn python library [23].

6. BENCHMARK RESULTS

The results of both the 3-class and 6-class classification problem are presented below, followed by a discussion of the results. The results for a 3 class implementation, with blocks being classified into sing, scream and no vocal are compared to a 6 class implementation, where the audio block was classified into Sing, Low Fry scream, Mid Fry scream, High Fry scream, Layered screams and No Vocal.

6.1 Experiment 1: 3-Class Results

The results for each experiment are shown in Table 1 and the class-wise recall of each combination are shown in Fig. 6.

Figure 7 shows the t-SNE plot of Feature Set 1, and we can see a distinction between the 3 different classes, although some overlap between the classes *Sing* and *Scream*. We can make the following observations. First, combined Feature Set 1 outperforms Feature Sets 3 and 4 with a gap of roughly 5%. This is expected as these sets are subsets

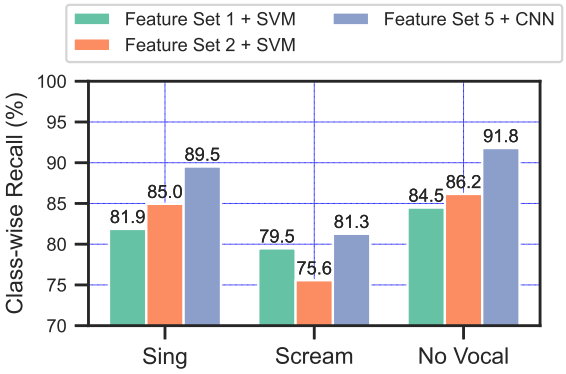

Figure 6. Class-wise recall for the 3-class problem.

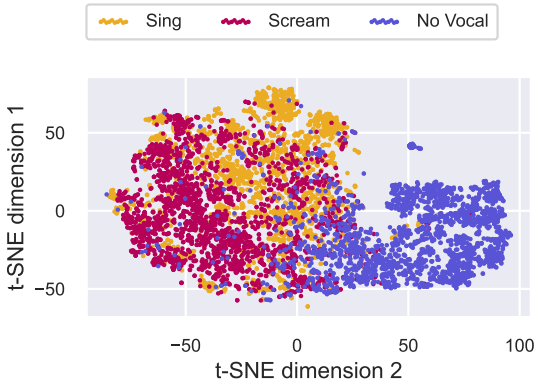

Figure 7. t-SNE projections of the feature space (Feature Set 1).

of Feature Set 1. Second, the combined Feature Set 1 and the VGGish Feature Set 2 show the best performance and perform similarly with recall above 82%. This means that the VGGish features, trained on a different task, contain a similar, semantically meaningful, information for classification as the combination of common baseline features. To a degree it is surprising that Feature Set 2 does not clearly outperform the traditional feature set as VGGish features have been shown to be powerful in music tasks such as musical instrument classification [24]. Third, the results show that the CNN with spectrogram input is able to detect the presence of screams with 87.6% balanced accuracy, which is notably higher accuracy than any SVM-based approach. It seems that the CNN is able to utilize the information in the spectrogram and is able to detect spectral patterns efficiently.

6.2 Experiment 2: 6-Class Results

The results of the 6-class problem are given in Table 2. We can observe that the performance is considerably lower for the 6-class problem with the two top-performing feature sets from Exp. 1. The VGGish features in Feature Set 2 seem to slightly outperform the low-level Feature Set 1. The CNN did not perform as well as the combination of

Configuration	acc	bal-acc	f1
Feature Set 1 + SVM	44.24	41.92	38.03
Feature Set 2 + SVM	**45.53**	**45.91**	**40.13**
Feature Set 5 + CNN	42.89	40.87	38.79

Table 2. Results for the 6-class problem in Exp. 2 (values shown in %)

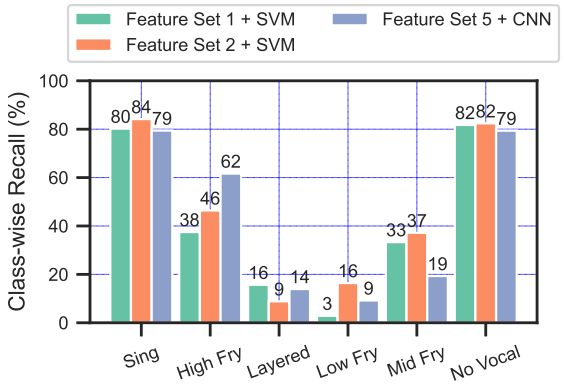

Figure 8. Class-wise recall for the 6-class problem.

VGGish features and SVM; the results of the CNN appear to be biased towards High Fry screams (see below).

Looking at the class-wise recall in Fig. 8, we observe that the systems could still identify the sung vocal and absence of vocals with high accuracy in the same range as the 3 class results shown above, however, they could not accurately distinguish between the different types of screams. We also see that the recall of the High Fry scream in the CNN is significantly higher than the other experiments, which is due to the classifier predicting most screams to be High Fry screams.

Investigating the confusion matrix in Fig. 9 gives us more details of the problem with the screams. We can see that several classes are being predicted incorrectly. Blocks labeled 'Layered' were often predicted as other classes, especially 'Sing' and 'High Fry' this could be because the layered class contains combinations of different classes, including the 'Sing' vocals. We also see that 'Low Fry' screams are often predicted as 'Mid Fry' due to the high degree of overlap between these classes in the feature space.

7. CONCLUSION

We introduced a new annotated dataset to aid and encourage further research in vocal detection in heavy metal music. Both the dataset and code have been made publicly available. While targeting scream detection, the dataset is also suitable for related tasks such as Vocal Activity Detection.

We presented a set of benchmark experiments on the automatic detection and classification of vocals in heavy metal music with the presented dataset. In these experiments, various temporal, spectral, and cepstral, and VGGIsh features were evaluated and compared with a CNN with log-mel spectrogram input.

In conclusion, with the dataset presented in this paper, we

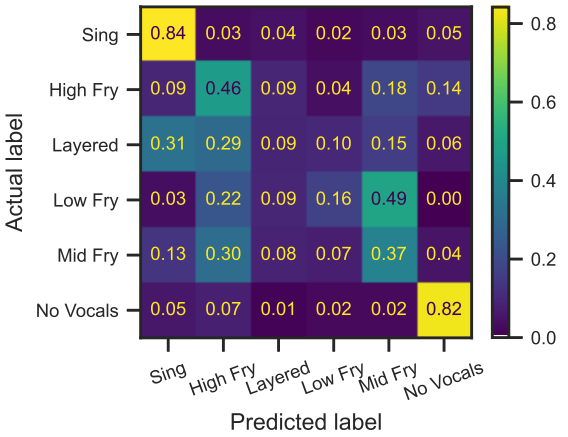

Figure 9. Confusion matrix for the 6-class problem (SVM).

were able to detect the presence of vocal events and classify them into sung vocal and screamed vocal with good accuracy. However, the same cannot be said for classifying the screams into the different types, as the different scream classes overlap within the feature space and cannot be separated easily. Thus, the dataset provides a new challenging task that can currently not be solved with satisfying results.

7.1 Future Work

There is anecdotal evidence online and within the heavy metal community for additional categories of vocal effects such as 'guttural vocals' and 'pig squeals'. Pending further investigation into this, we plan the extension of the dataset with additional audio files as well as extending the annotations to include these additional subsets of extreme vocal effects.

At present, the dataset has limited samples containing clean vocals sung over distorted instrumental sections, as most of the sections containing clean vocals in the songs used were also softer in nature. The dataset also has fewer samples of Low and High Fry screams (this is representative of their use in modern metal), and can be expanded upon by including further examples of these vocals.

8. REFERENCES

[1] O. Nieto, "Unsupervised clustering of extreme vocal effects," in *Proceedings of the 10th International Conference Advances in Quantitative Laryngology*, 2013, p. 115.

[2] N. J. Purcell, *Death Metal Music: The Passion and Politics of a Subculture*. McFarland, 2003.

[3] O. Nieto, "Voice transformations for extreme vocal effects," Master's thesis, Pompeu Fabra University, 2008.

[4] E. Smialek, P. Depalle, and D. Brackett, "Musical aspects of vowel formants in the extreme metal voice," *International Conference on Digital Audio Effects Conference*, pp. 1–8, 2012.

[5] C. T. Ishi, K.-I. Sakakibara, H. Ishiguro, and N. Hagita, "A method for automatic detection of vocal fry," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 47–56, 2007.

[6] E. Smialek, P. Depalle, and D. Brackett, "A spectrographic analysis of vocal techniques in extreme metal for musicological analysis," in *Proceedings of the International Computer Music Conference*, 2012.

[7] K.-I. Sakakibara, L. Fuks, H. Imagawa, N. Tayama, and others, "Growl voice in ethnic and pop styles," in *Proceedings of the International Symposium on Musical Acoustics*, 2004.

[8] W. Huang, T. K. Chiew, H. Li, T. S. Kok, and J. Biswas, "Scream detection for home applications," in *Proceedings of the Conference on Industrial Electronics and Applications*. IEEE, 2010, pp. 2115–2120.

[9] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri, and N. Ellouze, "Improved one-class svm classifier for sounds classification," in *Conference on Advanced Video and Signal Based Surveillance*, 2007, pp. 117–122.

[10] P. Laffitte, D. Sodoyer, C. Tatkeu, and L. Girin, "Deep neural networks for automatic detection of screams and shouted speech in subway trains," in *International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 6460–6464.

[11] M. K. Nandwana, A. Ziaei, and J. H. Hansen, "Robust unsupervised detection of human screams in noisy acoustic environments," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2015, pp. 161–165.

[12] J. Pohjalainen, P. Alku, and T. Kinnunen, "Shout detection in noise," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2011, pp. 4968–4971.

[13] N. Hayasaka, A. Kawamura, and N. Sasaoka, "Noise-robust scream detection using band-limited spectral entropy," *AEU - International Journal of Electronics and Communications*, vol. 76, pp. 117–124, 2017.

[14] C. Eckers, D. Hütz, M. Kob, P. Murphy, D. Houben, and B. Lehnert, "Voice production in death metal singers," *Proceedings of the International Conference on Acoustics/35th German Annual Conference on Acoustics*, pp. 1747–1750, 2009.

[15] P. Ribaldini, "Heavy metal vocal technique terminology compendium: A poietic perspective," Master's thesis, University of Helsinki, 2019.

[16] A. Loscos and J. Bonada, "Emulating rough and growl voice in spectral domain," in *Proceedings of the International Conference on Digital Audio Effects*, 2004, pp. 49–52.

[17] M. Guzman, K. Acevedo, F. Leiva, V. Ortiz, N. Hormazabal, and C. Quezada, "Aerodynamic characteristics of growl voice and reinforced falsetto in metal singing," *Journal of Voice*, vol. 33, no. 5, pp. 803–e7, 2019.

[18] K. Kato and A. Ito, "Acoustic features and auditory impressions of death growl and screaming voice," in *9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, 2013, pp. 460–463.

[19] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, "Spleeter: a fast and efficient music source separation tool with pre-trained models," *Journal of Open Source Software*, vol. 5, p. 2154, 2020.

[20] A. Lerch, *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Hoboken: Wiley-IEEE Press, 2012.

[21] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, 2015.

[22] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, and others, "CNN architectures for large-scale audio classification," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2017, pp. 131–135.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[24] S. Gururani, M. Sharma, and A. Lerch, "An Attention Mechanism for Music Instrument Recognition," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. Delft: International Society for Music Information Retrieval (ISMIR), 2019.

9. APPENDIX

The following songs were included in the dataset (Song No. Artist – Song Name):

1. Abbath – Ashes Of The Damned
2. After The Burial – Lost In The Static
3. Amon Amarth – Destroyer of the Universe
4. Amon Amarth – Live For The Kill
5. Amon Amarth – Twilight Of The Thunder God
6. Be'lakor – Venator
7. Behemoth – Ecclesia Diabolica Catholica
8. Behemoth – Bartzabel
9. Behemoth – Blow Your Trumpets Gabriel
10. Born of Osiris – White Nile
11. Cannibal Corpse – High Velocity Impact Spatter
12. Children of Bodom – Under Grass And Clover
13. Children of Bodom – Living Dead Beat
14. Children Of Bodom – Are You Dead Yet
15. Children of Bodom – Sixpounder
16. Children Of Bodom – Everytime I Die
17. Children Of Bodom – In Your Face
18. Dark Tranquillity – Lost to Apathy
19. Dark Tranquillity – Atoma
20. Death – Pull the Plug
21. Death – The Philosopher
22. Decapitated – Kill The Cult
23. Decapitated – Blood Mantra
24. Ensiferum – In My Sword I Trust
25. Enslaved – Caravans To The Outer Worlds
26. Godless – Deathcult
27. Gojira – Stranded
28. Gojira – Silvera
29. Immortal – Northern Chaos Gods
30. In Flames – Cloud Connected
31. Lamb of God – Memento Mori
32. Lamb of God – Laid to Rest
33. Lamb of God – Omerta
34. Lamb of God – Now You've Got Something to Die For
35. Lamb of God – The Faded Line
36. Ne Obliviscaris – Pyrrhic
37. Ne Obliviscaris – And Plague Flowers the Kaleidoscope
38. Nevermore – Born
39. Of Mice & Men – Bones Exposed
40. Of Mice & Men – Obsolete
41. Opeth – Blackwater Park
42. Parkway Drive – Carrion
43. Rings of Saturn – Senseless Massacre
44. Slayer – War Ensemble
45. Slayer – South Of Heaven
46. Slipknot – Psychosocial
47. Suffocation – Clarity Through Deprivation
48. Suicide Silence – No Pity for a Coward
49. Suicide Silence – Disengage
50. Suicide Silence – You Only Live Once
51. Suicide Silence – Slaves To Substance
52. Tesseract – Nocturne
53. Textures – Storm Warning
54. Textures – Old Days Born Anew
55. Thy Art Is Murder – Reign Of Darkness
56. Veil of Maya – Overthrow
57. Wintersun – Time