Audio Engineering Society

Conference Paper

Presented at the Conference on
Semantic Audio
2017 June 22 – 24, Erlangen, Germany

A Dataset and Method for Electric Guitar Solo Detection in Rock Music

Kumar Ashis Pati[1] and Alexander Lerch[1]

[1] *Center for Music Technology, Georgia Institute of Technology*

Correspondence should be addressed to Kumar Ashis Pati (ashis.pati@gatech.edu)

ABSTRACT

This paper explores the problem of automatically detecting electric guitar solos in rock music. A baseline study using standard spectral and temporal audio features in conjunction with an SVM classifier is carried out. To improve detection rates, custom features based on predominant pitch and structural segmentation of songs are designed and investigated. The evaluation of different feature combinations suggests that the combination of all features followed by a post-processing step results in the best accuracy. A macro-accuracy of 78.6% with a solo detection precision of 63.3% is observed for the best feature combination. This publication is accompanied by release of an annotated dataset of electric guitar solos to encourage future research in this area.

1 Introduction

Music scene analysis and the development of algorithms which allow computers to interpret a piece of music in ways similar to humans, has been gaining more and more importance over the last decade. Research in this area finds applications in automatic annotation and efficient browsing of large musical databases, instrument identification, music transcription, and music performance analysis. The automatic extraction of, for instance, the instruments being played or rhythmic and stylistic properties can have considerable impact on tasks such as music browsing, recommendation, and music production.

Among the many important constituents of a musical scene, this paper focuses on instrumental solos. More specifically, we restrict the scope to electric guitar solos because they form an important feature of several musical genres such as rock, blues, and country music. The terms 'electric guitar solos' and 'guitar solos' will be used interchangeably for the remainder of this paper. Goertzel argues that the popularity of rock music in general has been transformed by the evolution of guitar solos [1]. This is evident from the observation that many songs pertaining to these genres contain a guitar solo in some form or the other. Consequently, many rock music listeners are interested in music containing elaborate and intricate guitar solos. An algorithm that automatically detects and labels solos can thus aid music browsing and search by allowing guitar solo-enthusiasts to identify songs of interest more easily and enable music recommendation engines to showcase such solos effectively by automatically creating targeted previews. Solo detection can, for example, also guide video directors of live music broadcasts in their decisions. In addition, automated detection of guitar solos also allows researchers interested in analyzing

musical and aural characteristics of guitar solos to locate them within a large corpora of rock music, thereby acting as a useful preprocessing step for tasks such as music performance or guitar playing style analysis.

Although guitar solos are seemingly easily identified by most listeners, the definition of guitar solos is not necessarily straight-forward. For example, the distinction between a guitar solo and guitar riffs or licks, both improvisational techniques similar to solos but slightly different in terms of utilization and structure, needs to be formally defined. Furthermore, the automatic identification of guitar solos is complicated by the variety of guitar "sounds", differing significantly from song to song due to usage of effects and different combinations of guitar types and amplifiers. For example, some songs use a "clean" guitar sound whereas others use distorted sounds. A majority of guitarists use effects such as chorus and delays. Hence, guitar solo identification by using only sound and timbre characteristics is a challenging problem. For this reason, the goal of this study is to investigate various audio descriptors from different domains (i.e., temporal, spectral, predominant pitch, and structural segmentation) for the detection and segmentation of electric guitar solos. The descriptors are compared with a Support Vector Machine (SVM) classifier.

We first present a brief overview of the related work in this area in Section 2. The new, publicly available dataset created for this task is described in Section 3. Section 4 presents the overall approach, followed by the experimental design outlined in Section 5. The results and analysis are presented in Section 6. Lastly, Section 7 explores possible avenues for future work and concludes the paper.

2 Related Work

To the best knowledge of the authors, there is no prior work on the specific problem of guitar solo detection within a polyphonic mixture of various instruments. There has been, however, research in related areas such as vocal activity and instrument solo detection.

One of the first studies for solo detection was carried out by Peterschmitt et al. for classical music [2]. The authors proposed that the spectrum of solo sections is less complex than non-solo sections due to the dominant presence of only one instrument in the solo. They used the 'pitch-mismatch error' (PM error) as described in [3] and observed that this PM error is low for solo sections of a classical piece and high for ensemble sections. The algorithm performed well for harmonic solo instruments like saxophone and violin but did not give reliable results for instruments such as piano and guitar which had less sparse and less harmonic spectra. However, as the size of their dataset was small, the authors did not provide any quantitative performance data. Smit and Ellis investigated another pitch-based approach for solo voice detection in professional folk and classical music [4]. They estimated the dominant period in a signal and used an array of filters to cancel out the corresponding fundamental frequency and its harmonics. Then, the drop in signal energy resulting from this cancellation was used as input to a classifier. The underlying assumption was that the drop in energy would be higher for solo sections compared to non-solo sections because in a solo section the harmonic content of the solo instrument would dominate the mix. The algorithm performed significantly better than a system using only MFCCs (Mel Frequency Cepstral Coefficients) as feature and achieved a precision and recall of around 70%. The small dataset used, however, limits how conclusive these results can be seen.

More recently, there have been a few studies on instrumental solo detection using machine learning approaches. Fuhrmann et al. used feature selection to identify a set of five best features (pitch instantaneous confidence, spectral dissonance, spectral flux, spectral flatness and spectral crest) to identify solos in classical music [5]. The overall accuracy of their algorithm tested on a dataset consisting of 240 audio segments (40 segments for each instrument, each segment was 30 s long consisting of 15 s solo and 15 s of ensemble music) was 77%. Another study using a similar approach was carried out by Mauch et al. [6] on pop songs. In addition to using established audio features such as MFCCs, they also used custom-designed features based on predominant pitch fluctuation, MFCCs of re-synthesized predominant voice and normalized amplitude of harmonic partials. They reported a best-case accuracy of 89.8% with a precision and recall of 61.1% & 51.9% on their dataset comprising of 112 full-length pop songs. Another related study, identifying different modes of guitar playing (chords, solo, bass), was carried out by Foulon et al [7]. The authors used a combination of standard audio features and fundamental pitch based features. Their dataset contained 84 files (nearly 1.5 h) created using guitar sounds of 4

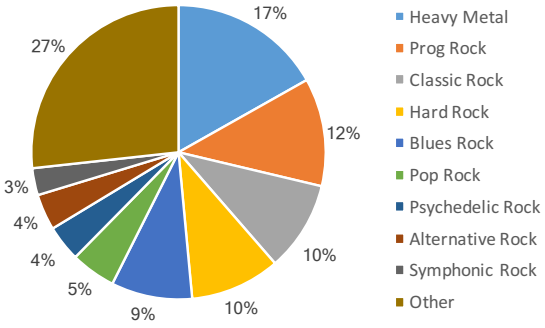

Fig. 1: Distribution of dataset songs with respect to sub-genres

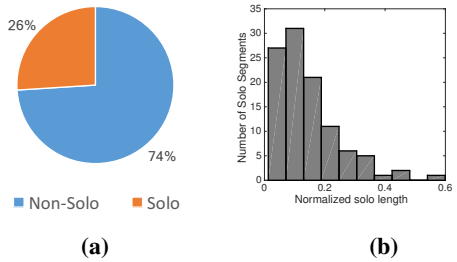

Fig. 2: Dataset Characteristics: (a) % of Solo v/s Non-Solo, (b) Distribution of solo segment length as a % of the song length

different guitar-types. Although, the average f-measure of their algorithm is around 95%, they used sound from a single guitar only and thus, the performance of the algorithm on actual songs involving other instruments is unknown.

3 Dataset & Annotations

Since there exists no publicly available dataset with guitar solo annotations, a new annotated dataset was created for the purpose of this study. A total of 60 full-length songs were chosen; 37 songs are taken from the list "100 Best Guitar Solos of All Time" [1] and 23 additional songs were taken from the authors' personal music collection. The dataset consists of songs from several sub-genres such as Progressive Rock, Classic Rock, Heavy Metal, Hard Rock, and Blues Rock as shown in Fig. 1. The songs were manually annotated with the location of solos in terms of start time, duration, and discographical information[2]. We name this the GSD (Guitar Solo Detection) dataset and have made it publicly available[3].

The dataset consists of 355 minutes of audio out of which nearly 75% constitutes the non-solo part and 25% constitutes the guitar solo. The median length of a guitar solo segment is around 35 seconds (with a minimum segment of 6 seconds and a maximum segment of 2 minutes). The distribution of the guitar solo segments as percentage of the song length is shown in Fig. 2b.

[1] http://guitar.about.com/od/guitaristsatoz/tp/100-Greatest-Guitar-Solos.htm, last accessed on 27th Jan 2017

[2] obtained from https://www.discogs.com, last accessed on 6th Apr 2017

[3] https://github.com/ashispati/GuitarSoloDetection/tree/master/Dataset

An important consideration for annotating the audio is the definition of a guitar solo. Looking back at previous work in this area, Peterschmitt et al. define a solo as a "section of a piece where an instrument is in foreground compared to the other sections of the piece" [2]. We extend this definition to guitar solos and define a guitar solo as a section in a song where the guitar is in the foreground compared to other instruments used in the song. However, this definition fails to differentiate solo from either riff or lick, which are both sometimes in the foreground during a song. According to Kernfeld, a riff refers to a short melodic sequence which may be repeated to accompany an underlying harmonic structure [8]. A lick, on the other hand, is defined as a short melodic motif or phrase used for improvisations [9]. In addition to these definitions, it is commonly observed that many guitar riffs rely on the use of power chords to provide a "hook" line for the song and repeat over measures. Licks, on the other hand, are much shorter in length (usually less than 1-2 measures) and usually follow vocal or other instrumental phrases.

Considering the above, we define a section of a song as a guitar solo if

- the guitar is in the foreground compared to other instruments or vocals,
- the guitar plays improvised melodic phrases which do not generally repeat over measures or in other structural segments, and
- the solo is longer than 2 measures.

4 Approach

The approach used in this paper is inspired by the work of Mauch et al. [6] but focuses on the specific case of guitar solos using a larger set of features. Fig. 3

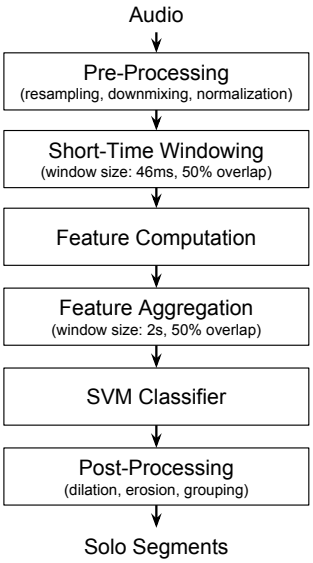

Fig. 3: Block Diagram of the presented approach

displays a flow-chart of the overall system. In the Pre-Processing block, the input audio is resampled to 44100 Hz, downmixed to a single channel, and normalized. The Short-Time Windowing step splits the audio into overlapping blocks (block size: 46 ms, overlap: 50%). The subsequent blocks are explained in more detail in the subsections below.

4.1 Feature Computation

Three sets of features are computed: baseline features, pitch-based features, and structural features.

4.1.1 Baseline features

This set of features consists of low level spectral and temporal features which have been chosen due to their widespread use in several Music Information Retrieval (MIR) tasks. The features used are: spectral centroid, maximum amplitude per frame, zero crossing rate, spectral crest factor, spectral flux and the 2nd–13th MFCCs. This results in a 17-dimensional feature vector per block of the input audio. More detailed information on the computation of these features can be found in [10].

4.1.2 Predominant pitch-based features

A guitar solo has the guitar in the foreground of the mix and will therefore tend to have a strong predominant pitch component. Hence, features based on measures of predominant pitch should help in detecting solo segments. The predominant pitch of a lead guitar will be distinct from that of another instrument like a bass guitar. In addition, a measure of the confidence with which this predominant pitch is detected might be able to distinguish between solo and non-solo parts. For computing the predominant fundamental frequency in Hz, we use the Melodia pitch extraction algorithm [11] as it has been shown to work well with polyphonic music signals. The implementation of the Melodia algorithm as provided in the Essentia library [12] is used. The predominant frequency is then converted into the MIDI pitch domain using the relation

$$m = 69 + 12\log_2\left(\frac{f}{440}\right),$$

where f is the predominant frequency in Hz and m is the resulting MIDI pitch. The Melodia algorithm also provides "pitch confidence", a measure of the confidence with which the predominant pitch is detected. The value of this pitch confidence measure is zero for non-voiced segments and positive for voiced regions. We considered both these measures, i.e., the fundamental pitch value and the pitch confidence as features, resulting in a 2-dimensional feature vector per block of the input audio.

4.1.3 Structural segmentation-based features

As pointed out in our discussion of the definition of a guitar solo, a solo does not usually repeat in a song. Considering the structural constituents of a typical song, a solo would generally occur during a segment of a song which is not repeated. Hence, a feature which incorporates the structural information of a song may be helpful towards distinguishing between solos and riffs. Under this assumption, two features based on the result of structural analysis of a song are designed to aid in solo detection. The structural segmentation is carried out using *msaf* — a python-based framework provided by Nieto et al. [13]. The convex non-negative matrix factorization (cnmf) based algorithm is used for the boundary detection and the Laplacian segmentation is used for labeling the segments. For more information about the algorithms, the reader may refer to the *msaf*

documentation[4]. Based on the time-stamp of the block under consideration, the segment label for the block is determined. The first feature is the number of repetitions of this segment label in the song and the second feature is the duration of this segment normalized by the song length.

4.2 Feature Aggregation

The output of the feature extraction is a 21-dimensional feature vector per input block of audio (17 baseline features, 2 predominant pitch-based features, and 2 structural segmentation-based features). A feature matrix is created using these feature vectors obtained from all input blocks for a song. The computed features are then aggregated over texture windows (length: 2 s, overlap: 50%). The aggregation is carried out by computing the median and standard deviation of each feature over all the blocks within one texture window. This results in an aggregated feature vector per texture window which is twice the size as the feature vector per input block.

4.3 SVM Classifier

A binary SVM classifier is used for classifying each texture window as solo or non-solo. The implementation of the SVM classifier is taken from the libSVM library [14] and the classifier is used with default parameters and a radial basis function kernel (C was set to 1 and gamma to $1/num\_features$ where $num\_features$ is the number of features used).

4.4 Post-Processing

Guitar solos usually occur as continuous segments. Therefore, a post-processing step is applied to the classifier output to ensure that neighboring time segments are more likely to form a single segment. The SVM classifier output is passed through simple erosion and dilation operations to break disjoint segments and connect adjacent segments. Erosion and dilation are common techniques used for image processing tasks and were first described by [15]. Here, a simple 1-dimensional version is used: first, the dilate operation convolves a mask (an all-ones column vector of length 3) with the classifier output, and second, the erode operation convolves a different mask (all-ones column vector of length 3 normalized by the length of the mask) with the output of the dilate operation. Segments are classified as solo if they are greater than 1. The result is then passed through a grouping algorithm which ensures that only contiguous segments longer than k seconds are classified as solos. The parameter k can be tuned to the task requirements.

[4] http://pythonhosted.org/msaf/index.html, last accessed on 27th Jan 2017

5 Experimental setup

Each of the 60 songs in the dataset is split into texture windows for feature aggregation as outlined above. Each texture window is labeled as solo or non-solo depending on its time-stamp and the ground-truth annotations. Seven experiments were performed based on different feature combinations using the baseline features (B), predominant pitch features (P) and structural segmentation features (S). The different combinations are B, P, S, BP, BS, BSP and BSP_PP where PP stands for post-processing.

The dataset was divided into 10 folds (6 songs each) and 10-fold cross-validation was carried out. The folds were distributed on the basis of songs instead of total number of texture windows. Although this results in a slightly non-uniform data distribution in the folds due to the variability in song length, it ensures that data points from the same song are not used for both training and testing. Sampling ensured a uniform class distribution in the training set to prevent bias in the classifier. Testing was carried out on entire songs. The 10-folds' results were averaged to get the overall performance.

In order to measure the performance of the algorithm, the following metrics are used:

(i) Micro-Accuracy, (m): This refers to the sum of correct classifications (both solo and non-solo) divided by the total number of texture windows considered.
(ii) Macro-Accuracy (M): This refers to the simple average of the % solo accuracy and the non-solo accuracy considered individually.
(iii) Precision (p): This is the ratio of correct solo classifications to the sum of correct solo classifications and incorrect solo classifications.
(iv) Recall (r): This is the ratio of correct solo classifications to the sum of correct solo classifications and incorrect non-solo classifications.
(v) f-Measure (f): This is the harmonic mean of the precision and recall.

Table 1: Overall performance of the different feature combinations (Micro-Accuracy m, Macro-Accuracy M, Precision p, Recall r, f-measure f, and specificity S, R1 is the accuracy of the classifier if it classifies all data-points as non-solo), in BSP_PP case: k = 4s, all metrics are in %

	R1	B	P	S	BP	BS	BSP	BSP_PP
m	74.7	78.4	67.8	63.1	79.5	79.3	80.7	**82.6**
M	50.0	74.3	69.5	57.2	75.8	75.4	76.7	**78.6**
p	-	54.1	43.2	33.9	56.3	56.1	57.5	**63.3**
r	-	67.0	69.8	49.6	68.9	68.3	69.9	**71.8**
f	-	59.8	53.4	40.3	62.0	61.6	63.1	**67.3**
S	-	81.6	69.1	64.9	82.6	82.6	83.6	**85.4**

(vi) Specificity (S): This is the recall of the non-solo class and is the ratio of the number of correct non-solo classifications to the sum of correct non-solo classifications and incorrect solo classifications.

A distinction is being made between macro and micro-accuracy because, as discussed in Sect. 3, the dataset is heavily skewed towards the non-solo class and hence, the results may appear unrealistically good even if the classifier never detects solos. It is also worth noting that all the above metrics are computed on a per song level (using a macro averaging scheme) to ensure that the algorithm performs well for all songs irrespective of the song length.

6 Results & Discussion

The performance metrics obtained from the experiments are shown in Table 1. Compared to the baseline (B case), it can be seen that there is a small improvement in performance when either the predominant pitch based features or the structural segmentation based features are added to the baseline features (BP & BS cases respectively). However, improvement in performance in either case is not substantial. The reason for this might be that the pitch-based features rely on predominant pitch only and contain no information about which instrument is actually contributing to the predominant pitch. The predominant pitch might be influenced by vocals and other instruments in many cases. The segmentation based features, on the other hand, look only at very simple properties associated with the structural segments. This becomes clear when we look at the performance of the pitch-based features and segmentation-based features on their own (P and S case respectively) which are both inferior to the performance of the baseline features. However, the results improve when both the pitch-based features and the segmentation-based features are added to the baseline feature set (BSP case). It appears that the combination of these features contains more information for the classifier to exploit. Post-processing on the BSP case further improves the performance. McNemar tests conducted on the results obtained from different pairs of feature combinations indicate statistical significance with p-value $< 10e{-}8$. It is worth noting that the standard deviation of accuracy values between the different folds is approx. 5%.

Considering that only standard spectral and temporal audio features were used in the B case, the accuracy of almost 78% seems good when compared to the random selection accuracy for a 2-class problem (which would be 50%). However, it should be noted that given the distribution of the classes in the dataset, wherein almost 75% of the data points are in the non-solo category, the detection accuracy for baseline features alone is not impressive. This is clearer when we consider that the R1 micro-accuracy (which is the accuracy of the classifier if it classifies all the data points as non-solo) for this dataset would be 74.7%. Hence, the macro-accuracy is a better indicator of performance in this case. The best macro-accuracy of 78.6% was observed for the BSP_PP case ($k = 4$ s). The classifier makes a lot of errors in classifying solos compared to the non-solos. Higher values of specificity also indicate this bias. This is evident in the cumulative confusion matrix shown in Table 2. Looking at precision, recall and specificity, we observe that there is a substantial improvement in the precision of solo-detection due to post-processing (almost 5%). This indicates significant reduction in the number of false positives in the post-processing step.

Overall, the results are in line with those obtained in previous studies on solo detection [4, 6, 5], but consid-

Table 2: Confusion Matrix (for BSP_PP case, k = 4s)

		Predicted	
		Non-Solo	Solo
Actual	Non-Solo	**13216**	2228
	Solo	1410	**4001**

ering the different sizes of datasets used, the difference in genre of music (classical, pop, rock etc.) and the fact the current work is very specific to guitar solos, a direct comparison cannot be made.

7 Conclusion

This paper presented a preliminary study of features for the automatic detection of electric guitar solos in rock music. In the absence of substantial work in the area, a manually annotated dataset of guitar solos was created to aid and encourage future research. Both the dataset and the code have been made available publicly[5]. The study provides a baseline reference against which the performance of future guitar solo detection methods can be evaluated. Timbre, predominant pitch, and structural segmentation-based features were used which, in combination, led to improvement in detection accuracy. Although predominant pitch based features have been used in previous works on solo detection, the improvement in detection accuracy due to structural segmentation based features points to a promising direction for future research.

There are a lot of possible avenues to explore for future work. The current dataset contains only a limited number of songs and is mostly homogeneous with respect to genre. An expansion of the dataset, possibly with songs from other genres like blues and country music, would allow for more generality and could increase confidence in the results. In addition, songs not containing guitar solos or containing solos of other instruments might also be included in the dataset so as to ascertain the robustness of the algorithm. Another area with respect to the dataset that needs more exploration pertains to 'album-effect' which has been observed in several MIR studies [16, 17]. Albums and to a certain degree artists have their own unique timbre / sound which leads to improvement in classification results for several MIR tasks. The presented dataset contains 6 songs by Dream Theater and 3 songs each by Pink Floyd and Led Zeppelin. It also contains several songs from the same album (3 songs from Images and Woods, 2 songs each from Dark Side of the Moon, Black Clouds & Silver Linings, Led Zeppelin II, and Nightmare). A larger dataset with more variety in songs will also help us ensure that album-effect is minimized. Another aspect worth noting is that neither the performance of the fundamental pitch detection algorithm nor the performance of the structural segmentation algorithm have been evaluated on this specific dataset. The robustness of these algorithms can have significant impact on the subsequent processing steps and the overall result. A possible avenue for future work would be to evaluate the performance of the pitch and segmentation based features when different algorithms are used to extract them.

The design of better and more task-representative features is certainly a promising direction as the current features only contain very limited information. Features which capture the essence of the sound of a guitar or which describe the harmonic content of the segments may be considered as possible alternatives. Feature learning also has obvious potential to generate features that represent the task better. Another possible approach to solve this problem would be to investigate it from a source separation perspective. For example, one could think about extending the work of Cono et al. on separating the solo part of a song from the accompaniment [18], and then using a novelty function based approach to segment the solo locations. Finally, considering the recent popularity and success of deep learning based methods in several MIR tasks, applying deep learning to this task might also be of interest.

[5] https://github.com/ashispati/GuitarSoloDetection

References

[1] Goertzel, B., "The Rock Guitar Solo: From Expression to Simulation," *Popular Music & Society*, 15(1), pp. 91–101, 1991, doi:10.1080/03007769108591426.

[2] Peterschmitt, G., Gomez, E., and Herrera, P., "Pitch-based solo location," in *Proceedings of MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, Spain, 2001.

[3] Maher, R. C. and Beauchamp, J. W., "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," *The Journal*

of the Acoustical Society of America, 95(4), pp. 2254–2263, 1994, doi:10.1121/1.408685.

[4] Smit, C. and Ellis, D. P. W., "Solo Voice Detection Via Optimal Cancellation," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 207–210, New York, NY, USA, 2007, doi:10.1109/ASPAA.2007.4393045.

[5] Fuhrmann, F., Herrera, P., and Serra, X., "Detecting solo phrases in music using spectral and pitch-related descriptors," *Journal of New Music Research*, 38(4), pp. 343–356, 2009.

[6] Mauch, M., Fujihara, H., Yoshii, K., and Goto, M., "Timbre and Melody Features for the Recognition of Vocal Activity and Instrumental Solos in Polyphonic Music." in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 233–238, Miami, USA, 2011.

[7] Foulon, R., Roy, P., and Pachet, F., "Automatic classification of guitar playing modes," in *Proceedings of the International Symposium on Computer Music Modeling and Retrieval (ISMIR)*, pp. 58–71, Marseille, France, 2013.

[8] Robinson, J. B., "Riff," in B. Kernfeld, editor, *The New Grove Dictionary of Jazz*, Oxford University Press, 2nd edition, 2002.

[9] Witmer, R., "Lick," in B. Kernfeld, editor, *The New Grove Dictionary of Jazz*, Oxford University Press, 2nd edition, 2002.

[10] Lerch, A., *An introduction to audio content analysis: Applications in signal processing and music informatics*, John Wiley & Sons Inc., Hoboken, New Jersey, 2012.

[11] Salamon, J. and Gomez, E., "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), pp. 1759–1770, 2012, doi:10.1109/TASL.2012.2188515.

[12] Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., and Serra, X., "Essentia: An Audio Analysis Library for Music Information Retrieval." in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 493–498, Curitiba, Brazil, 2013.

[13] Nieto, O. and Bello, J. P., "Systematic exploration of computational music structure research," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 547–553, New York, NY, USA, 2016.

[14] Chang, C.-C. and Lin, C.-J., "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp. 27:1–27:27, 2011, doi:10.1145/1961189.1961199, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[15] Haralick, R. M., Sternberg, S. R., and Zhuang, X., "Image analysis using mathematical morphology," *IEEE transactions on pattern analysis and machine intelligence*, (4), pp. 532–550, 1987.

[16] Kim, Y. E., Williamson, D. S., and Pilli, S., "Towards Quantifying the" Album Effect" in Artist Identification." in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 393–394, Victoria, Canada, 2006.

[17] Pampalk, E., Flexer, A., Widmer, G., et al., "Improvements of Audio-Based Music Similarity and Genre Classificaton." in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, volume 5, pp. 634–637, London, UK, 2005.

[18] Cono, E., Dittmar, C., and Schuller, G., "Efficient implementation of a system for solo and accompaniment separation in polyphonic music," in *Proceeings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 285–289, IEEE, Bucharest, Romania, 2012.