AUTOMATIC CHORD-SCALE RECOGNITION USING HARMONIC PITCH CLASS PROFILES

Emir Demirel
Queen Mary University of London
e.demirel@qmul.ac.uk

Baris Bozkurt
Izmir Democracy University
barisbozkurt0@gmail.com

Xavier Serra
Universitat Pompeu Fabra
xavier.serra@upf.edu

ABSTRACT

This study focuses on the application of different computational methods to carry out a "modal harmonic analysis" for Jazz improvisation performances by modeling the concept of *chord-scales*. The Chord-Scale Theory is a theoretical concept that explains the relationship between the harmonic context of a musical piece and possible scale types to be used for improvisation. This work proposes different computational approaches for the recognition of the chord-scale type in an improvised phrase given the harmonic context. We have curated a dataset to evaluate different chord-scale recognition approaches proposed in this study, where the dataset consists of around 40 minutes of improvised monophonic Jazz solo performances. The dataset is made publicly available and shared on *freesound.org*. To achieve the task of chord-scale type recognition, we propose one rule-based, one probabilistic and one supervised learning method. All proposed methods use Harmonic Pitch Class Profile (HPCP) features for classification. We observed an increase in the classification score when learned chord-scale models are filtered with predefined scale templates indicating that incorporating prior domain knowledge to learned models is beneficial. This study has its novelty in presenting a first computational analysis on chord-scales in the context of Jazz improvisation.

1. INTRODUCTION

In this work, we perform an automatic analysis of the tonal harmonic context in monophonic improvisation performances specifically in the context of Jazz tradition. As proposed in [1], the cultural context in music should be considered when conducting computational musicological research. Here, we employ *chord-scales* as the unit for computational analysis of Jazz improvisation. This paper proposes a new approach for the retrieval of chord-scale information from Jazz improvisation performances. In our study, various methods are compared for the classification of chord-scale types based on Harmonic Pitch Class Profiles (HPCP) extracted from audio signals. Due to the lack of a dataset with monophonic improvised performances that has temporal labels for the chord-scale that is played, we have curated a new dataset, which is called *The Chord-scale Dataset*. The performance of the proposed methods were tested on this dataset alongside with a baseline score. This work proposes a novel approach for the analysis of the harmonic content in Jazz improvisation, targeting chord-scales as their use in a specific Jazz style is an essential skill [2].

In musical context, a scale is a step-wise arrangement of pitch classes contained within an octave [3]. A Jazz player is expected to draw on everything he or she knows concerning these scales and their relationships with chords [4]. Clearly, the choice of chord-scales to be played within a certain harmonic context is very subjective and dependent on performers artistic decisions. For practical reasons, a limitation on the list of chord-scales to detect or estimate is considered in this work. The readers are encouraged to refer to the Scale Syllabus by Jamey Aebersold [5] for an extended set of chord-scales used in Jazz improvisation.

Intuitively, audio features that represent pitch or pitch class information of the musical content would be ideal for the classification of chord-scales. One approach could be performing chord-scale classification based on (musical) note tracks. However, this approach would require a highly accurate transcription of the improvised content to achieve a robust classifier and predicting accurate results. To skip the automatic transcription step, we employ Harmonic Pitch Class Profiles (HPCP) as features for classification and the estimation of the chord-scale type.

Three methods are proposed for the classification stage. First, a set of predefined binary chord-scale templates are used in a pattern matching strategy based on likelihoods. The second approach is an automatic classification pipeline using Support Vector Machines (SVM). Finally, we apply a similar pattern matching strategy with that of the first method, but replacing predefined templates with chord-scale models learned from labeled data using Gaussian Mixture Models (GMM). Alongside the performance of the proposed chord-scale recognition methods, the choice of different statistical features for classification is tested during experiments.

This paper is structured as follows. First, a big picture of this work is introduced. Then relevant domain knowledge and a review of the previous attemps for musical scale estimation are provided. Then, we introduce the Chord-Scale dataset. Fourth and fifth sections explain the feature preprocessing stages. Later the automatic classification of chord-scale types are explained. The experiments comparing the performances of the proposed classification algorithms and varying statistical features are given in Section

7. Then we conclude with discussions regarding the results of the experiments and possible future directions of this study.

2. BACKGROUND INFORMATION

Due to the high degree of multidisciplinarity of the research area of Music Information Retrieval (MIR), it is essential to have a comprehensive outlook that comprises both *music theoretical* (or domain) knowledge and the advanced computational methods for *music modeling* when doing research for solving an MIR task. Our study for chord-scale recognition proposes combining methodological approaches from both musical and machine learning domains. In this section, we examine the musical theoretical concepts that motivated this work and computational methods to analyze certain musical concepts that are related to chord-scales.

2.1 The Chord-Scale Theory

The chord-scale theory is a method of mapping a list of scales to a list of chords and the theory aims to explain the inter-relationships between the chords and scale in the context of Jazz improvisation. For last few decades, jazz musicians use the approach of chord-scales when improvising over chord progressions. The essence of the chord-scale approach is if a chord is diatonic to a certain scale, than that scale can be used as a resource for creating melodic lines.

In traditional approach for tonal harmony, chords are built in three tones or triads. In Jazz tradition though, seventh degrees and additional color tones are commonly used, describing a chord or a chord progression with all its potential tonal possibilities. Chords are vertical structures of notes where the chord tones are separated by leaps while scales form a step-wise arrangement of notes [2]. Playing consecutive steps is much easier for playing fast and accurately which makes the chord-scale theory a preferable approach for improvisation among jazz musicians.

The list of chord-scales involved in this study for analysis are determined according to the content of Gary Burton's on-line course of Jazz Improvisation. In addition to the scales involved in the course, we consider other mother scales included in [6] that are commonly used in Jazz performances. The complete list of chord-scale mapping to be considered in this study can be seen from Table 1.

2.2 Chord-Scale Type Recognition

The task of automatic chord-scale recognition (or detection) from musical audio is relatively a new field of research in Sound and Music Computing (SMC) and has not been investigated as deeply as other MIR tasks like *automatic music transcription*, *chord recognition*, *musical similarity*, *style identification*, etc. However the retrieval of chord-scale related information has a potential to be beneficial for many MIR tasks and applications including the aforementioned titles.

Even though there is no well developed literature specifically focusing on the automatic chord-scale recognition task, there are relevant research that study related musicological concepts to chord-scales. In [7], Weiss and Habryka proposes an algorithm for visualizing the tonal characteristics of classical audio recording through the concept of scales. In the second approach presented in the paper, the authors apply maximum likelihood estimation on audio frames using chroma features to estimate the scale-type.

Conceptually, chords and chord-scales are theoretically and practically related concepts in musical practice. The same may hold true for the recognition of both musical concepts as separate Music Information Retrieval (MIR) tasks. Both chords and chord-scales are musical harmonic structures that are defined in terms of pitch classes. This relevance of these concepts inspired our study to employ some of the common probabilistic approaches used for chord recognition. In [8], the authors study the effects of employing binary chord templates and probabilistic chord models, where they do not report a significant performance improvement of using learned chord models as opposed to predefined binary templates. In [9], the authors use chromas extracted in various ways with the goal of establishing timbre invariant feature vectors. For chord recognition, Gaussian Mixture Models are used as the pattern matching strategy.

3. DATASET

We have curated a new dataset for the context of the study presented in this paper. *The Chord-Scale Dataset* consists of 39 monophonic improvisation performances in 12 chord-scale types (Table 1) commonly used in Jazz improvisation. The recording device used during the data collection sessions is Zoom H-6. The total duration of the dataset is around 40 minutes and the recordings belong to either tenor saxophone or trumpet performances. Throughout each recording the tonic and the tonal harmonic context is kept constant. The musical phrases (or *motif*s) in each recording are annotated with their start and end times. In order to maintain a balanced dataset in terms of number of phrases per chord-scale type, we have included 4 additional tracks from *Jamey Aebersold Jazz, Volume 26: The "Scale Syllabus"* [5]. In total, there are 43 tracks and 236 data points, each representing a musical phrase. Hence, the chord scale recognition in our study is performed on the phrase level

The dataset is shared publicly on *freesound.org* [1]. The phrase onset and offset annotations can be found in the same repository with the code to generate and reproduce the experiments [2]. The annotation format in this dataset is similiar to the chord progression annotation format proposed in [10]. For reproducibility purposes, we have made the frame-level HPCP features extracted from the recordings available in the *github* repository. Moreover, this dataset also includes instrument labels for each recording.

[1] https://freesound.org/people/emirdemirel/packs/24075/
Pack ID: 24075

[2] https://github.com/emirdemirel/Chord-ScaleDetection

4. FEATURE EXTRACTION

pitch class Profiles (PCP) [11] or chroma features [12] [13] have been a popular tonal representation of musical signals since their introduction around two decades ago. In our system, we have used Harmonic Pitch Class Profile (HPCP) features for the recognition of the chord-scale type from a musical phrase. HPCP features are a specific type of chroma features which are extracted via weighted mapping of harmonic peaks in the frame spectrum onto 12 pitch classes [14].

One may argue that monophonic musical signals can be conveniently transcribed and analyzed in symbolic domain for more precision and explainability. Such automatic transcription can be achieved in two steps: First, pitch-tracking would be applied on the improvised performance. Then, the pitch track would be transcribed into musical note level. The automatic music transcription procedure introduces much more complex level of computation. Thus, we propose to use Harmonic Pitch Class Profile features for the simplicity in conducting a subbranch of automatic harmonic analysis that our study focuses on. Given the chord-scale types included in this study being octave invariant sets of pitch classes, meaning that the relative octave differences between pitches that belong to the same pitch class does not influence the chord-scale type estimation proposed in this study. According to this consideration, using HPCP features does not introduce inefficacy for achieving the chord-scale recognition task.

4.1 Preprocessing

First, the audio signal is filtered with inverted approximation of equal loudness curves in order to account for the non-linear perception of the spectra in the human auditory system [15]. The sampled and filtered audio signal is divided into series of analysis frames of size N_{frame} and hop size of N_{hop}. Then, each analysis frame $x(n + l \cdot N_{hop})$ is multiplied with a "Hanning" window function $w(n)$, to obtain the windowed audio signal $x_w(n)$, where $n = 0, 1, ..., N_{frame} - 1$ and l indicates the number of the frame that is analyzed.

4.2 Harmonic Pitch Class Profiles

Chromagrams or Pitch Class Profiles (PCPs) are widely used in many applications that aim to extract mid-level musical information from the audio signal, like automatic chord recognition key extraction [16], cover song identification [17] and such, since they were introduced in [11].

In principle, Harmonic Pitch Class Profiles [14] are modified versions of Pitch Class Distributions (PCDs) which are introduced in [11]. In principle, the HPCP extraction procedure applies a weighted mapping on spectral peaks detected on frame-based spectra to a finite number of pitch classes. In our context, we use 12 pitch classes as there are 12 pitch classes defined within one octave in equally the tempered tuning system.

In order to maintain tonic invariant HPCP features, the frame level HPCP vectors are constructed with respect to the tonic frequency of the improvised phrase, so that the reference frequency f_{ref} (*HPCP bin # 0*) corresponds to the tonic frequency. The tonic frequency is computed using the following formula:

$$f_{ref} = f_{tuning} \, .2^{\frac{\delta k \ .100}{1200}} \tag{1}$$

where f_{tuning} is the global tuning frequency of the performance, δk is the distance in semitones between the keys of the tuning frequency K_{tuning} and the target key K.

4.3 Post-Processing

The goal of the post-processing steps explained in this section is to prepare the feature data for the classification stage. In order to establish dynamic invariance in the feature vectors, each of the frame-based HPCP vectors are normalized with respect to a suitable norm. In our methodology, we employ $unitSum$ $(l - 1)$ norm. By applying $unitSum$ norm, we obtain relative weights of pitch classes in the frame-based HPCP vectors.

The pitch class mapping of spectral harmonic peaks causes artifacts on frame-level HPCP vectors. In order to minimize these artifacts, we only consider HPCP bin with the maximum value in the frame-level HPCP vectors and set the rest of the bins to zero (Equation 2). The resulting feature vectors are denoted as $HPCP'(i)$ This process is valid for our case since the analysis audio files are monophonic / one-instrument performances and it is expected to have only one dominant pitch class in the feature frames.

$$HPCP'(i) = \begin{cases} HPCP(i), & \text{if } i = \max_i HPCP(i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

We apply further processing on the frame-based chroma features in order to reduce the influence of noise and artifacts in the summarized features for classification. The logic of our noise removal method is as below:

$$\textbf{if} \quad argmax(HPCP'(n)) \neq argmax(HPCP'(n \pm 1) : \\ HPCP'(n) = 0 \\ k = 0, 1, 2, ..., 11 \tag{3}$$

The recordings in the data set are monophonic and the post-processed feature vectors have only one non-zero component that is the pitch class with maximum energy in the unprocessed vectors. The non-maximum pitch classes are discarded on frame level targeting to obtain an overall pitch class profile distribution that has less influence from the harmonic artifacts and noise.

5. FEATURE SELECTION

The chord-scale recognition approaches proposed in this paper are performed on phrase-wise summarized HPCP features. Frame-level features are summarized into 2 statistical aspects: the mean and the standard deviation. For each pitch class (or bins in HPCP vectors), the mean and standard deviations are calculated over the manually annotated phrase segments. Then, the summarized features are normalized on *l-2* norm.

In Figure 1, the summarized feature histograms are shown for both of *mean* and *standard deviation* features.

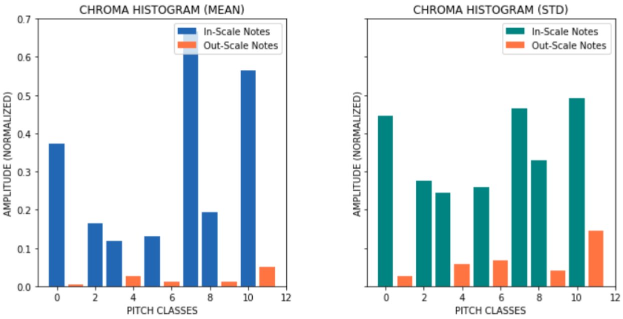


Figure 1: pitch class distributions of statistically summarized features. **Mean** features (*left*) and **std** features (*right*). Sample : 'Improvisation on Minor Scale - Trumpet - Phrase # 2'

It can be seen from the plots in Figure 1 that the histograms have non-identical but similar distributions over pitch classes, which indicates both features contain relevant information. Since these histograms represent statistical summarization of tonal features, it can be maintained that as both statistical features capture similar tonal aspects from the acoustic signal.

6. CLASSIFICATION

In this study, we propose 3 distinct chord-scale recognition algorithms, The first approach applies pattern matching using Maximum Likeliest Estimation (MLE) based on predefined binary chord-scale templates. In the second approach, we use Support Vector Machines (SVM) for classification of chord-scale types. Finally, we propose to replace binary templates in the first approach with chord-scale models that are learned using Gaussian Mixture Models (GMM).

6.1 Binary Template Matching

The first chord-scale recognition method we have developed follows a Binary-Template Matching (BTM) strategy, which is inspired by the scale matching method proposed in [7]. The method proposed is essentially a *maximum likelihood estimation* procedure where the likelihoods are computed as the product of pitch classes in the chroma vectors. We propose to compute the likelihoods as the sum of the pitch classes, due to the methodological and contextual differences explained in Section 4.3.

The chord-scale type likelihoods S are obtained by computing the inner products of the statistically summarized chroma vectors $HPCP$ with each of the chord-scale templates T_s:

$$S(s) = \sum_{i=0}^{11} HPCP(i) \cdot T_s(i) \tag{4}$$

Table 1: Scale Dictionary

Scale Types s	**Binary Templates** T_s
Ionian (Major)	(1 0 1 0 1 1 0 1 0 1 0 1)
Dorian	(1 0 1 1 0 1 0 1 0 1 1 0)
Phrygian	(1 1 0 1 0 1 0 1 1 0 1 0)
Lydian	(1 0 1 0 1 0 1 1 0 1 0 1)
Mixolydian	(1 0 1 0 1 1 0 1 0 1 1 0)
Aeolian (Natural Minor)	(1 0 1 1 0 1 0 1 1 0 1 0)
Locrian	(1 1 0 1 0 1 1 0 1 0 1 0)
Melodic Minor	(1 0 1 1 0 1 0 1 0 1 0 1)
Lydian b7	(1 0 1 0 1 0 1 1 0 1 1 0)
Harmonic Minor	(1 0 1 1 0 1 0 1 1 0 0 1)
Altered (Super Locrian)	(1 1 0 1 1 0 1 0 1 0 1 0)
Whole Tone	(1 0 1 0 1 0 1 0 1 0 1 0)
Half-Whole Step Diminished	(1 1 0 1 1 0 1 1 0 1 1 0)

Finally, the chord-scale type s with the maximum likelihood would be determined as the estimated or detected scale type.

$$S' = \max_{s} S(s) \tag{5}$$

where S' denotes the final estimated chord-scale type.

6.2 Support Vector Machines

Support Vector Machine (SVM), first introduced in [18], is a classification and regression tool that uses a hypothesis space with linear functions in a high dimensional feature space, trained using a learning algorithm from optimization theory that implements a learning bias from statistical learning theory [19].

In our SVM classification model, we use statistically summarized feature vectors as the feature data for the classifier. The SVM classifier holds Radial Basis Function (RBF) kernels which provide more flexible mapping of feature spaces. The penalty parameter C and the kernel coefficient γ of the classifiers are optimized using Grid Search / Cross Validation as in [20]. The following parameter grids are iterated over to choose the best pair of hyperparameters:

$$C \ : \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$$
$$\gamma \ : \{0.001, 0.01, 0.1, 1\}$$

6.3 Gaussian Mixture Models

Finally, we propose a probabilistic approach that resembles the binary-template based likelihood strategy in Section 6.1 but applies pattern matching on learned chord-scale models from labeled data. In this approach, each chord-scale model is constructed using Gaussian Mixture Models (GMM) defined in terms of mean μ and a covariance matrix. A GMM is a weighted sum of multivariate Gaussian distributions [21], which is defined as :

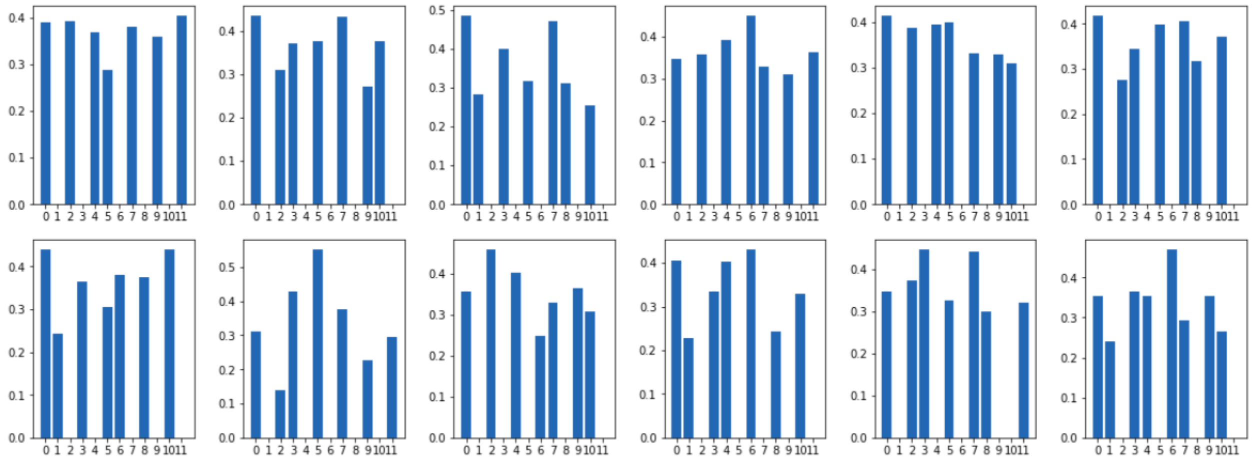


Figure 2: Chord-Scale Models learned using GMM, $x\ axis$ = *HPCP index*, $y\ axis$ = weights assigned by GMMs
Top row: *major, dorian, phrygian, lydian, mixolydian, minor*
Bottom row: *locrian, melodic minor, lydian b7, altered, harmonic minor, half-whole step diminished*

$$p(\mathbf{x}) = \sum_{k=1}^{K} c_k \frac{1}{\sqrt{|2\pi \sum_k|}} exp\big(-\frac{1}{2}(\mathbf{x}-\mu_k)^T \sum_{k}^{-1}(\mathbf{x}-\mu_k)\big) \quad (6)$$

where K is the number of mixture components in the Gaussian distribution. In the case of K = 1, the distribution employs the maximum likeliest estimate of parameters. In our approach, we construct a GMM with one component (K = 1) for each chord-scale type Then the estimation of the likeliest chord-scale in Equation 4 is applied by replacing the binary templates T_s by μ.

$$S(s) = \sum_{i=0}^{11} HPCP(i) \cdot \mu_s(i) \quad (7)$$

Finally, the chord-scale type with the maximum likelihood is determined (as in Equation 5) as the chord-scale estimate.

7. EXPERIMENTS

The experiments are conducted to test the choice of statistical summarization of features and the performance of the proposed algorithms for chord-scale recognition. The method introduced in [7] is also tested on the *Chord-Scale Dataset* as comparison with the proposed classification algorithms during evaluation.

The pattern matching algorithms (BTM and GMM) employ MLE for classification. MLE is applied on the mean vectors for each chord-scale model. Note that the out-scale pitch classes in GMM models have non-zero values. Since the MLE procedure in this study does not give penalty to out-scale pitch classes, these features potentially decrease the classification performance. In order to overcome this and obtain a higher performance, we apply element-wise multiplication on the binary-templates and the learned models, which filters out the out-scale pitch classes (setting them to zero) and eliminating the effect of these features on MLE-based classification. The resulting chord-scale models are shown in Figure 2. The performances of both the filtered and unfiltered GMM models are tested during the experiments.

The implementations of machine learning algorithms used in this study are obtained from **scikit-learn**, which is a *Python* module that integrates a broad range of state-of-the-art machine learning algorithms for medium scaled problems [22].

8. RESULTS

As explained in Section 6.2, the hyperparameters of the SVM classifier are optimized using Grid Search. First, the dataset is split into two balanced subsets: train and test sets. 10-fold cross validation is applied on the train set for optimizing the hyperparameters of the SVM classifier. More explicitly, the train set is split in 10 folds and one fold is chosen as the validation set at each iteration. The validated hyperparameter pair is then used to evaluate the performance of the classifier on the test set. In order to increase the generalization power of this procedure, we classification scores are obtained 10 times on pseudo randomized development and test splits. Then, the average performance scores of all iterations are reported as the final SVM classifier score. The other classifiers included in our experiments do not hold hyperparameters for optimization. For the GMM based methods, the chord-scale models are trained on the same development set as in the SVM classification procedure and tested on the test split. Similarly, 10 iterations on the pseudo randomized splits are performed and the average is reported. Since the BTM method does not require training, the classifier is directly tested on the same test sets.

Table 2 shows the classification performances of the algorithms presented in this paper alongside the comparison of $HPCP.mean$ and $HPCP.std$ features. The results are provided in terms of accuracy scores (%). $HPCP.std$ features outperforms $HPCP.mean$ for almost all methods which implies *standard deviation* is a more suitable choice of statistical summarization of features for this task.

For MLE based classification, additive method performs better than multiplicative likelihood estimation method. This may be due to how the features are processed, as mentioned in Section 4.3. The unfiltered learned GMM models perform worse than predefined binary template matching (BTM) algorithm. More so, multiplicative MLE performs poorly (around 20%). It appears that GMM models learn parameters that assign more weight on the out-scale pitch classes in the overall pitch class distribution of chord-scale models. These weights on out-scale pitch classes cause an expected decrease in accuracy score for MLE based classification as the method relies on the weights assigned to each pitch class by the mixture model.

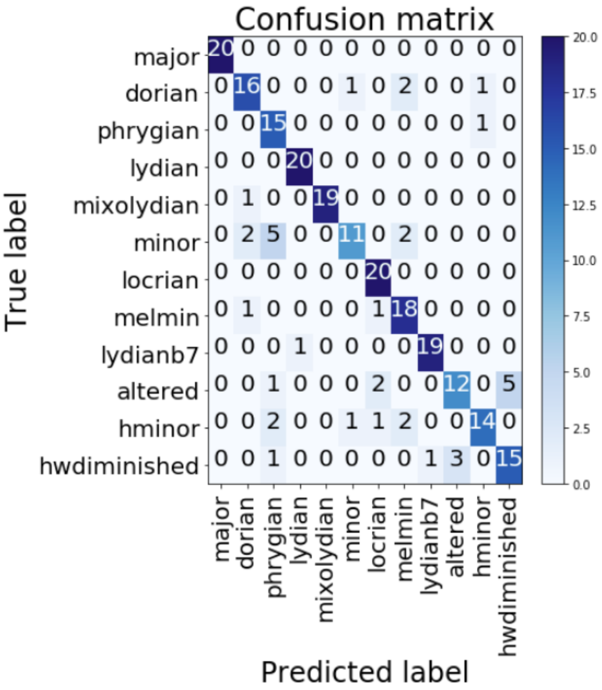
Table 2: Results - Accuracy Scores

Method	HPCP.mean (%)	HPCP.std (%)
BTM-MLE (Mult.)	72.03	68.64
BTM-MLE (Add.)	79.23	79.66
SVM	76.25	80.83
GMM-MLE (Mult.)	19.92	21.18
GMM-MLE (Add.)	69.92	79.23
GMM-MLE-filt (Mult.)	71.19	71.61
GMM-MLE-filt (Add.)	76.69	**84.32**

The accuracy scores increase evidently when MLE is applied on GMMs that are filtered using the predefined binary chord-scale templates. These filtered GMMs show a performance increase between $6-60\%$ depending on the chord-scale recognition method and the feature set used for classification. Leveraging domain knowledge appears to be effective for our task. Overall, *Additive MLE on filtered GMMs using* $HPCP.std$ *features* perform the best among all the proposed algorithms. For comparison with the baseline, aforementioned algorithm outperforms the method proposed in [7] applied in this context by around 15%.

SVMs have comparable performance with filtered GMM-MLE. Note that using $HPCP.std$ features perform better than using the $HPCP.mean$ feature, supporting our claim that std features are a better fit as features for the classification tasks included in this study.

In Figure 3, the confusion matrix of the best performing algorithm and the feature set is provided. The chord-scale type with highest error rate is the *minor* and *altered* chord-scale types. The misclassified instances for minor scale are classified as either *phrygian*, which differs from minor scale by only one pitch-class. Most misclassified instances for *altered* scale are estimated as *half-whole diminished* scale. Even though these scales have different functionalities in tonal harmony, they are similar to each other in terms of diatonic pitch classes or scale degrees, which seem to be the major source error in classification. To overcome this problem, tonal harmonic constratints may be further applied to increase the classification performance. In general, this algorithm shows a reasonable performance and makes musically reasonable mistakes.



	major	dorian	phrygian	lydian	mixolydian	minor	locrian	melmin	lydianb7	altered	hminor	hwdiminished
major	20	0	0	0	0	0	0	0	0	0	0	0
dorian	0	16	0	0	0	1	0	2	0	0	1	0
phrygian	0	0	15	0	0	0	0	0	0	0	1	0
lydian	0	0	0	20	0	0	0	0	0	0	0	0
mixolydian	0	1	0	0	19	0	0	0	0	0	0	0
minor	0	2	5	0	0	11	0	2	0	0	0	0
locrian	0	0	0	0	0	0	20	0	0	0	0	0
melmin	0	1	0	0	0	0	1	18	0	0	0	0
lydianb7	0	0	0	1	0	0	0	0	19	0	0	0
altered	0	0	1	0	0	0	2	0	0	12	0	5
hminor	0	0	2	0	0	1	1	2	0	0	14	0
hwdiminished	0	0	1	0	0	0	0	0	1	3	0	15

Figure 3: Confusion Matrix for the best performing method (GMM-MLE (filtered) with HPCP.std features)

9. CONCLUSION

We have tackled a relatively new domain of analysis in Music Information Retrieval research, which focuses on the retrieval of chord-scale information from improvised Jazz solos. The automatic recognition of chord-scale would be of use for various musical applications including style recognition or transfer, automatic harmonic analysis, performer identification and educational purposes. Our study applies a comparative study between a rule-based and two conventional machine learning approaches. The proposed methods were also used in a task specific manner for other related MIR tasks like chord recognition [13] [8], key detection [23] [24] [16]. The results of our study show a similar trend with the prior study on these related tasks. GMMs alone do not perform any better than the predefined binary template (BT) matching method. However, filtering the GMM scale models with these predefined templates is shown to be beneficial for the proposed chord-scale recognition pipeline. This result reveals the advantage of incorporating prior domain knowledge with learned models. Moreover, summarization of frame-based HPCP features over phrase-wise audio segments for chord-scale classification using their standard deviations perform more robustly compared to summarizing in terms of the mean of frame-based features, which agrees with the prior study in [20]. We have conducted the experiments on the Chord-Scale dataset, which is introduced in this work and shared publicly for reproducible and open science. The code to reproduce the experiment results and scale annotations can be found in the github repository. There are numerous possible future directions that the study presented here can take. Chord-scale recognition can be performed on MIDI notes (features) transcribed with an advanced AMT algorithm. For skipping the automatic transcription step, there are several chroma extraction methods that provide features that are more robust to noise or the variance of the instrument type [13] [25]. The application of more advanced machine learning methods like neural networks require big-scale datasets. Hence, new strategies for obtaining more data in the context of Jazz improvisation need to be explored.

Acknowledgments

The authors acknowledge Toprak Barut and Hikmet Altunbaslier for all their help in data collection. This work is supported by ERC funded TECSOME project. The author E.D received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowsa-Curie grant agreement No. 765068.

10. REFERENCES

[1] X. Serra, "The computational study of a musical culture through its digital traces," in *Acta Musicologica*, vol. 89, no. 1. International Musicological Society, 2017, pp. 24–44.

[2] B. Nettles and R. Graf, *The Chord Scale Theory and Jazz Harmony*, 1997.

[3] W. A. Fraser, "Jazzology: A study of the tradition in which jazz musicians learn to improvise (afro-american)," in *9th Conference on Interdisciplinary Musicology, 2014*, 1983. [Online]. Available: https://repository.upenn.edu/dissertations/AAI8406667

[4] D. Baker, *Jazz Improvisation (Revised): A Comprehensive Method for All Musicians*. Alfred music, 2005.

[5] J. Aebersold and P. A. LONG, *Volume 1 [-] of A new approach to jazz improvisation*. J. Aebersold, 1974.

[6] M. Levine, *The Jazz Theory Book*. O'Reilly Media, Inc., 2011.

[7] C. Weiss and J. Habryka, "Chroma based scale matching for audio tonality analysis," in *Proceedings of 9th Conference on Interdisciplinary Musicology*, 2014, pp. pp.168–173. [Online]. Available: http://publica.fraunhofer.de/dokumente/N-345665.html

[8] T. Cho, R. J. Weiss, and J. P. Bello, "Exploring common variations in state of the art chord recognition systems," in *In proceedings of the Sound and Music Computing Conference*, 2010, pp. 1–8.

[9] N. Jiang, P. Grosche, V. Konz, and M. Müller, "Analyzing chroma feature types for automated chord recognition," in *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio*, 2011.

[10] V. Eremenko, E. Demirel, B. Bozkurt, and X. Serra, "Audio-aligned jazz harmony dataset for automatic chord transcription and corpus-based research," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*, pp. 483–490. [Online]. Available: https://doi.org/10.5281/zenodo.1291834

[11] T. Fujishima, "Real-time chord recognition of musical sound: A system using common lisp music," in *Proceedings of International Computer Music Conference*, 1999, pp. pp.464–467. [Online]. Available: http://hdl.handle.net/2027/spo.bbp2372.1999.446

[12] G. H. Wakefield, "Mathematical representation of joint time-chroma distributions," in *Proceedings ofAdvanced Signal Processing Algorithms, Architectures, and Implementations IX*, vol. 3807, 1999, pp. 637–646.

[13] M. Muller and S. Ewert, "Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features," in *Proceedings of International Society of Music Information Retrieval Conference*, 2011, pp. pp.215 – 220. [Online]. Available: http://doi.org/10.1.1.399.9397

[14] E. Gómez, "Tonal description of music audio signals," Ph.D. dissertation, Universitat Pompeu Fabra, 2006. [Online]. Available: https://doi.org/10.1287/ijoc.1040.0126

[15] D. Bogdanov, N. Wack, E. Gómez Gutiérrez, S. Gulati, P. Herrera Boyer, O. Mayor, G. Roma Trepat, J. Salamon, J. R. Zapata González, and X. Serra, "Essentia: An audio analysis library for music information retrieval," in *In proceedings of International Society of Music Information Retrieval Conference*. [Online]. Available: http://hdl.handle.net/10230/32252

[16] E. Gómez, "Key estimation from polyphonic audio," in *Music Information Retrieval Evaluation Exchange (MIREX05)*, 2005.

[17] J. Serra, E. Gómez, P. Herrera, and X. Serra, "Chroma binary similarity and local alignment applied to cover song identification," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6. IEEE, 2008, pp. 1138–1151. [Online]. Available: https://doi.org.10.1109/TASL.2008.924595

[18] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152. [Online]. Available: http://doi.org/10.1145/130385.130401

[19] V. N. Vapnik, "An overview of statistical learning theory," in *IEEE transactions on Neural Networks*, vol. 10, no. 5. IEEE, 1999, pp. 988–999. [Online]. Available: http://doi.org/10.1109/72.788640

[20] E. Demirel, B. Bozkurt, and X. Serra, "Automatic makam recognition using chroma features," in *Proceedings of Folk Music Analysis Conference*, 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1239435

[21] J. H. Jensen, D. P. Ellis, M. G. Christensen, and S. H. Jensen, “Evaluation of distance measures between gaussian mixture models of mfccs.” in *Proceedings of International Society of Music Information Retrieval Conference*, 2007, pp. 107–108. [Online]. Available: https://www.ee.columbia.edu/~dpwe/pubs/JenECJ07-gmmdist.pdf

[22] 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,” vol. 12, pp. 2825–2830, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1953048.2078195

[23] Ö. Izmirli, “Template based key finding from audio.” in *International Computer Music Conference*, 2005. [Online]. Available: http://www.music.mcgill.ca/~ich/research/misc/papers/cr1293.pdf

[24] G. Peeters, “Chroma-based estimation of musical key from audio-signal analysis,” in *In proceedings of International Society of Music Information Retrieval Conference*, 2006, pp. 115–120.

[25] M. Mauch and S. Dixon, “Approximate note transcription for the improved identification of difficult chords.” in *IProceedings of International Society of Music Information Retrieval Conference*, 2010, pp. 135–140.