AUTOMATIC MAKAM RECOGNITION USING CHROMA FEATURES

Emir Demirel
Music Technology Group
Universitate Pompeu Fabra
emir.demirel@upf.edu

Baris Bozkurt
Music Technology Group
Universitate Pompeu Fabra
baris.bozkurt@upf.edu

Xavier Serra
Music Technology Group
Universitate Pompeu Fabra
xavier.serra@upf.edu

ABSTRACT

This work focuses on the automatic makam recognition task for Turkish Makam Music using chroma features. Chroma features are widely used for music identification and tonal recognition tasks such as key estimation or chord recognition. Most of prior work on makam recognition largely rely on use of pitch distributions. Due to the imperfection of automatic pitch extraction for non-monophonic audio, use of chroma features is an alternative that has been showed to be effective in a previous study and we follow the same approach. Our work does not propose a new architecture but rather considers parameter optimization of chroma based recognition for makams. In our tests we use an open-content dataset and perform comparisons with previous studies. As a result of parameter optimization a better performance is achieved. All resources are shared for ensuring reproducibility of the presented results.

1. INTRODUCTION

This study is a continuation of automatic makam recognition studies carried in the CompMusic project (Serra, 2017) and targets improving the performance of chroma feature based automatic makam recognition.

The term '*makam*' mainly refers to a modality system in middle of a continuum defined by a particularized scale and generalized tune on its two poles (Powers & Wiering, 2001). Here we specifically consider the modality system of the Turkish makam music tradition where the following descriptors are considered to be most essential: scale description (involving micro-tonal intervals), overall melodic progression (*seyir*) describing a path from one emphasis note to another until the '*karar*' is reached, preference of specific tri-tetra-penta chords used to form melodies, typical phrases and dynamic range for the melodic contour. For an in-depth review of basic concepts of Turkish makam music and previous computational studies the readers are referred to (Bozkurt, et al., 2014).

Automatic makam recognition can be carried on symbolic or audio data. In (Ünal, et al., 2014), the authors use an n-gram approach for makam detection on symbolic data and report very high accuracies. Makam recognition from audio is a much difficult task due to various characteristics such as heterophony, high variability in interpretations by musicians. The most common used approach in literature (for detection from audio) is the use of pitch distributions (extracted from audio recordings) with a template-matching (or nearest neighbor) strategy (Gedik & Bozkurt, 2010; Karakurt et al., 2016). Pitch histograms have indeed been used as a feature in various automatic recognition tasks since early days of Music Information Retrieval (MIR) (Tzanetakis, et al., 2003). Karakurt, et al. (2016) also presents application of this approach on two other music traditions: Hindustani and Carnatic music (with accuracies 0.92 for 30 ragas and 0.73 for 40 ragas respectively).

Chroma features are frequently used for many tonality related MIR tasks such as chord recognition, tonality detection, audio classification (Dighe et al., 2013; Müller & Ewert, 2011; Jiang et al., 2011) and is a good alternative to pitch distribution features for non-monophonic audio. Chroma based makam recognition has been previously considered by Ioannidis et al. (2011), where the authors follow two distinct approaches for automatic makam classification. First, they apply a makam template-matching method where the templates are constructed from annotated data. Secondly, automatic classification is performed using support vector machines. In our study, we follow a similar approach to the second approach in the aforementioned paper since it shows considerably better performance than template matching approach. We focus on parameter factorization for improving the performance via use of larger window size which reduces noise in features, testing various dimensions for the chroma representation and hyperparameter optimization for the automatic classification stage. We conduct our experiments on the Ottoman-Turkish Makam Music Dataset (Karakurt, et al., 2016), which is the most comprehensive dataset available for computational research on Turkish Makam music. The proposed method is compared with all past approaches using the same set of makams. The performance of our methodology on the same nine makams show that our approach outperforms the prior work of Ioannidis et al. (2011), by more than %10 in overall accuracy and achieves slightly better accuracy scores compared to the state of the art over 20 makams (Karakurt, et al., 2016) which uses pitch histograms as feature.

To sum up, the work presented in this paper provides a bottom-up demonstration of a chroma-based supervised *mode* recognition architecture, and an evaluation method on an open-content dataset for future research on the topic.

2. DATASET

The *Ottoman-Turkish makam recognition dataset* (Karakurt et al.,2016) is the most comprehensive dataset for computational research on makam music, that is open content and is available for researchers. The entire set for the analysis in this study is composed of 997 audio tracks within the OTMM, which are distributed over 20 makams (Table 1). The tonic frequency of each track (available in the dataset) has been obtained and annotated by extracting pitch at the approximate mid-point of the last note in the performance (which has been annotated manually).

Makam Type	#_of_ Tracks	Makam Type	#_of_ Tracks	Makam Type	#_of_ Tracks
Acemaşiran	50	Huzzam	50	Rast	50
Acemkürdi	49	Karcigar	50	Saba	50
Bestenigar	50	Kurdilihicazkar	50	Segah	50
Beyati	49	Mahur	50	Sultaniyegah	50
Hicaz	50	Muhayyer	50	Suzinak	50
Hicazkar	50	Neva	50	Ussak	50
Huseyni	49	Nihavent	50	**Total**	**997**

Table 1. OTMM – Makam Set / number of tracks

Initially, experiments were performed on the entire OTMM dataset. Even though there exist hundreds of variations of makam types, the set of makams in OTMM is representative of this music tradition. In the previous works of Gedik & Bozkurt (2010) and Ioannidis et al. (2011), experiments contained data from only 9 commonly used makams, which are *Hicaz, Huseyni, Huzzam, Kurdilihicazkar, Nihavent, Rast, Saba, Segah, Ussak*. For the second stage of the experiments, the experimental procedures are performed on this makam set (449 tracks) to observe the effects of parameter factorization on HPCP (Harmonic Pitch Class Profiles) features and performance of supervised learning classifiers, in comparison with the work of the aforementioned study.

3. SYSTEM ARCHITECTURE

Our system uses chroma features for automatic classification. The main advantage of using chroma features for this task is that it discards the need of automatic melody extraction of polyphonic audio, which introduces many complexities.

There are several ways to extract chroma features. Most commonly used techniques include either applying spectral analysis on audio frames and quantizing the frame spectrum into frequency bins (Fujishima, 1999), or employing suitable filter banks for the pitch classes (Müller & Ewert, 2011). In our methodology, we use Harmonic Pitch Class Profiles (HPCP), extracted in a similar fashion as explained in the study of Gómez (2006). Figure 1 shows the general structure of the proposed system. The choices of parameters for each step are explained in detail in this section.

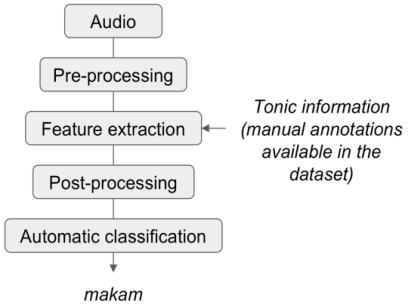

Figure 1. Schematic representation of the system architecture

3.1. Audio Signal Processing

3.1.1. Preprocessing

At the preprocessing stage, DC offset removal is applied on the audio signal using Infinite Impulse Response (IIR) filters. Then, to account for human perception non-linearity, the audio signal is filtered with inverted approximation of equal loudness curves. During our experiments, we have observed that application of these preprocessing steps show some improvement in the robustness of chroma features against transient noise.

3.1.2. Feature Extraction

After filtering out, spectral analysis based on Fourier transform is performed on a frame-based analysis strategy. The frame sizes are chosen as 200ms and hop size as 100ms, which outputs 10 frames per second. As mentioned in Jiang, et al. (2011), larger size windows are preferred over smaller window sizes for mid-level musical information recognition task like chord recognition. Also a smaller window sized window tends to capture more transient noise on the audio signal as opposed to using a larger size window. Besides obtaining chroma features more robust to noise, larger hop size also reduces the computation cost.

For the computation of the chromagram on the frame level, spectral peaks are detected from the local maxima in the frame spectra (Serra & Smith, 1990). The spectral peaks to be detected are limited within the frequency range 100 – 5000 Hz. The spectral peaks are then mapped to the finite number of frequency bins (12, 24, 36,..) with Equation 1 (Gomez, 2006), where n denotes the HPCP bin of to be considered and a_i and f_i denote the linear magnitude and frequency values of the peak i .

$$HPCP(n) = \sum_{i=1}^{nPeaks} w(n, f_i).a_i \qquad (1)$$

$$HPCP: Harmonic\ Pitch\ Class\ Profile$$
$$nPeaks: number\ of\ spectral\ peaks$$

The frequency mapping process require two important considerations to obtain features that are representative of

the musical signal. Initially, a reference frequency must be set for frequency mapping of the frame spectrum. This reference frequency can also be referred as the first bin of the HPCP vector. Moreover, the number of equally separated bins within one octave (or in other words the size of HPCP vectors) needs to be taken into account as a parameter for music traditions that exploit microtonal intervals to construct melodies. One of the main goals of this work is to shed light on the effect of varying sizes of HPCP vectors for the analysis of non-Western music traditions.

3.1.3. Reference Frequency

In chromagram computation, the center frequencies of each bin in HPCP vectors are determined with respect to a reference frequency. The general approach is to estimate the reference frequency by computing spectral peaks with respect to the standardized tuning frequency of 440Hz. Here, we use the manually annotated tonic frequencies of the tracks available in the data set. Frequency mapping of spectral peaks is performed directly with respect to the tonic as the reference frequency.

3.1.4. Normalization:

As explained in detail in (Müller & Ewert, 2011), normalization on a frame basis is necessary at the post-processing step in order to discard the effects of dynamic variations. In our study, we employ l^1 -norm (Equation 2) which corresponds to normalizing elements of the chroma vector x with respect to the sum of all elements of the vector. By doing so, we obtain the chroma histogram representation of each track in the dataset.

$$\|x\|_1 := \left(\sum_{i=1}^{N} |x(i)| \right) \quad (2)$$

3.2 Classification

3.2.1. Feature Set

The initial set of features are the bins of N-bin normalized and averaged HPCP histograms which are computed as the global mean chroma for each track. The normalized global HPCP mean histograms of a musical performance can also be referred as the normal distribution of averaged pitch-classes of a track. It is expected that this distribution would give an insight about the harmonic structure of the piece. In addition to the averaged HPCP vectors, we also include variance related information in our feature set, by simply computing the standard deviation of bins of HPCP vectors separately and globally for each track. As depicted in Figure 2, the standard deviation histogram shows a similar trend with the mean HPCP histogram. This implies that standard deviation also contains some information related to the makam scale and its emphasized tones, which can be used for automatic classification. In our experiments including standard deviation in the feature set, we have observed a considerable increase in the accuracy of automatic classification, which is explained in Section 4.

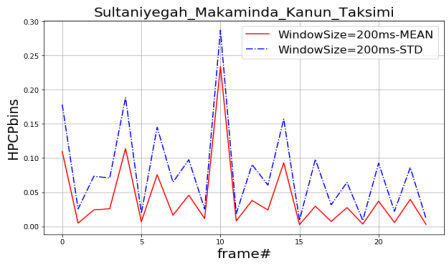

Figure 2. 24-bin - Mean vs. Standard Deviation HPCP histograms

An essential aspect of the makam concept is the fixed 'tonal-spatial' (or tonal-temporal) organization referred as *seyir*. As part of that aspect, the melodic organization and emphasis in the opening of a piece or improvisation plays a crucial role in forming a makam as defined in the theory. To account for more accurate makam estimation, characteristics of the melodic progression also needs to be taken account. It has been previously shown that the first part of the overall melodic contour (i.e. beginning of the performance) carries some discriminative characteristics in Turkish makam music (Bozkurt, 2012; Bayraktarkatal & Öztürk, 2015). To incorporate that, alongside with the global features set, statistical features are also computed locally from certain portions of the beginning of tracks. To determine a more suitable portion of the song, varying percentages of the full track are tested.

3.2.2. Supervised Learning

Automatic classification of the songs according to makam classes are performed using supervised learning. Different combinations of the statistical features set defined above were used to train a support vector machine with radial basis function kernel. The evaluation of each test feature subset are given in Section 4.

The training of support vectors is done with radial basis function kernels. For each test on feature subsets, the hyperparameters of the support vector classifier has to be optimized for each iteration. For training support vector machines with RDF kernel, there are two hyperparameters that need to be tuned to achieve good performance: penalty parameter (regularization constant) C and the kernel coefficient γ. We apply grid search method for hyperparameter optimization which is an exhaustive searching process over a set of defined parameters. For the regularization constant, iterations are done on the following set: C={0.001,0.01,0.1,1,10,100,1000 } for the penalty parameter and γ={0.001,0.01,0.1,1} for gamma. The grid search parameters are validated using 10-fold cross validation on the train set. The parameter combination that gives the best average cross-validated accuracy is used to build the overall model. Then, we test our classifier model, which is trained on the training set, and make pre-

dictions on the test set. To ensure that there is no over-fitting in the results and achieve a high generalization power, the experiments are repeated 10 times over randomized and stratified test & train data splits. In the results, we report the average accuracy and F measure of these experiments. Furthermore, to maintain reproducibility, the random seed is fixed and documented in our shared code. The overall result of predictions with the best performing feature set are shown via confusion matrix in Section 4.

4. RESULTS

The effects of HPCP parameterization on automatic classification with SVMs are tested using stratified 10-fold cross validation. The hyperparameters of the classifier are tuned on the training set using grid search and makam estimations are performed on the testing set which is randomly selected but stratified %10 of the whole set. In order to obtain statistically less biased results, the evaluation pipeline is performed on ten different and randomized train/test splits.

4.1. Experiments on 20 makams:

Automatic classification is performed over 997 songs in 20 makams. In our study we test the performance of using 12, 24, 36 and 48 bins in the HPCP vectors. In addition to the scale of vector size, the effect of different combinations of statistical features in the feature set are tested. Finally, F-measures and accuracy scores are reported. Since the dataset is balanced, weighted macro scores over the dataset are appropriate measures for evaluation.

Table 2 presents resulting F scores of cases with varying number of bins and feature sets which represent the global statistics of HPCP vectors. It is seen that standard deviations of HPCP vectors improve the classifier's performance. This improvement is more significant as the number of bins increase.

F-Measures	**12 - bins**	**24 - bins**	**36 - bins**	**48 - bins**
Mean	0.64	0.64	0.65	0.66
Stdev.	0.65	0.7	0.69	0.7
Mean+Std	0.65	0.7	0.7	0.7

Table 2. F Measure of varying number of bins and feature set combinations (Full track)

As explained in Section 3.2, the makam of the song is generally introduced with emphasis at the beginning of the track. In addition to the above experiments, we also provide a comparison of the global chroma features and chroma features obtained locally from the beginnings of the songs. To determine a good estimation of size of such a region for the beginning of the songs, an iteration over varying portions of the track (from %5 to %40) is performed. (Figure 3) In the figure only the combination of mean and standard deviation features are shown since they outperform the only-mean HPCP features. This analysis has a potential for revealing some future directions for automatic makam recognition task, including a further structural analysis for this tradition.

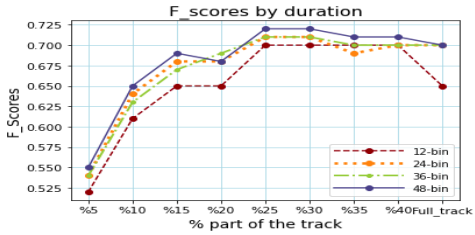

Figure 3: F_scores of classification with local statistical features (Mean + St.dev.)

The results in Table 2 and Figure 3 indicate that higher resolution in the HPCP vector increases the classification scores. Features with vector size of 48 shows slightly better performance than the rest, hence this resolution is set for the following experimental steps. Moreover, the classifier has a better performance when trained with the local chroma histograms instead of global. Regarding statistical features, further investigation is performed to highlight possible directions to improve classification performance. Table 3 shows the evaluation of the combinations of local and global statistical features. In this step, the local features are obtained from the first %30 of the whole track.

Feature_Set (HPCP)	**12-bins**		**48-bins**	
	F Measure	Accuracy	F Measure	Accuracy
Mean(full)	0.64	0.65	0.66	0.67
Std(full)	0.65	0.65	0.67	0.72
Mean(Full)+Std(Full)	0.65	0.66	0.67	0.7
Mean(Local)	0.65	0.65	0.67	0.67
Std(Local)	0.66	0.67	0.73	0.74
Mean(Local)+Std(Local)	0.7	0.71	0.72	0.72
Mean(Full)+Std(Local)	0.71	0.72	0.74	0.74
Mean(Local)+Std(Full)	0.71	0.72	0.74	0.75
Std(Local)+Std(Full)	0.72	0.73	**0.76**	**0.77**

Table 3. Evaluation scores of classifier models with varying feature set combinations, 12-bin vs. 48-bin (Analysis on 20 makams)

With the best performing feature set combination and HPCP parameters, our system is able to score **77%** overall accuracy. These results are comparably better that the current state of the art methodology on the task (Karakurt et al., 2016), where their best performing parameters result in 71.8% accuracy.

In Figure 4, we present the confusion matrix for our system using 48-bin HPCP vectors (mean HPCPs of the

first 30% together with standard deviation of HPCPs of the whole track). Here, the confusion matrix includes the classification instances in the tests over all of the ten randomized sets. We discuss our observations on the confusion matrices in Section 5.

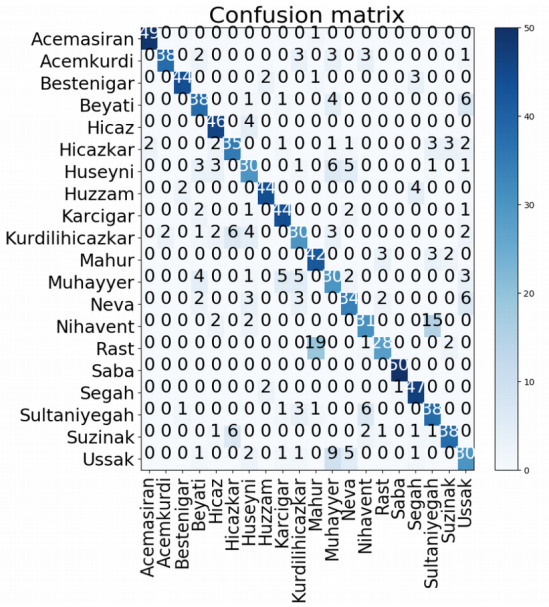

Figure 4. Confusion Matrix of 20 makams analysis

4.2 Experiments on 9 makams:

To have a clear comparison with prior work, we have performed the same experimental procedure over 449 songs in 9 makams. The makam set for this stage of experiments is chosen in consideration with previous research on the topic. (Ioannidis, et al., 2011) In their study, Ioannidis, et al. (2011) use 159-bin HPCP vectors as the feature set which is constructed in parallel with the theoretical knowledge. The experiments in this paper consider bin variations within [12,24,36,48].

Feature Set(HPCP)	**12-bins**		**48-bins**	
	F_Measure	Accuracy	F_Measure	Accuracy
Mean(Full)	0.76	0.77	0.8	0.8
Std(Full)	0.81	0.82	0.82	0.82
Mean(Full)+Std(Full)	0.81	0.81	0.85	0.85
Mean(Local)	0.77	0.77	0.82	0.82
Std(Local)	0.8	0.81	0.86	0.86
Mean(Local)+Std(Local)	0.82	0.82	0.86	0.87
Mean(Full)+Std(Local)	0.82	0.83	**0.89**	**0.89**
Mean(Local)+Std(Full)	0.84	0.84	0.87	0.87
Std(Local)+Std(Full)	0.86	0.86	**0.89**	**0.89**

Table 4. Evaluation scores of classifier models with varying feature set combinations (Analysis on 9 makams)

In order to provide a concise comparison, only the results of the best performing local features of 12-bin and 48-bin feature vectors, in combination with global statistics are shown. (Table 4) The classification scores of proposed methodology shows a robust performance with the classification accuracy of **%89**. This result outperforms the prior works of Ioannidis et al. (2011) where their best performing approach scores an F-measure of %73. Additionally, Table 4 shows the scores for the case where the HPCP vector resolution is 12-bins per octave, which is the standard resolution in MIR. Finally, confusion matrix of the best resulting model for the test with 9 makams is illustrated in Figure 5.

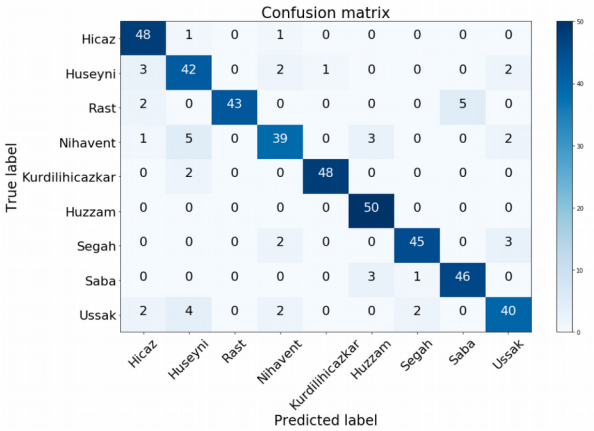

Figure 5. Confusion Matrix of 9 makams analysis

5. DISCUSSIONS

The research presented in this paper explores the significance of parameter selection for extracting chroma features and proposes the use of different statistical features for automatic makam classification. The outperforming results of second stage of the experiments (analysis over 9 makams) is highly due to parameter factorization. Larger size for windows serves better than a smaller size one for automatic modality detection tasks. This observation agrees with the previous work done in Western music traditions (Müller & Ewert, 2011; Jiang, al., 2011). Besides the window sizes, higher HPCP vector resolution has a positive effect on the classification performance. Another important aspect that may have an impact on the performance increase is the hyperparameter optimization for the automatic classifier.

Our study shows that adding various statistical features to the feature set shows a significant improvement for automatic classification as well as the other factorization explained above. Tests over the number of bins reveal that further research is necessary to study the size of chroma vectors when analyzing music from non-Western traditions. The results show better performance for sizes greater than 12. In Figure 3, Table 3 & Table 4, it is observed that the features obtained from the beginning of the tracks result in improvement of classification accuracy. Even though this does not contradict with the the-

ory, further research is necessary on the structure of songs in makam music tradition.

Studying the confusion matrices in Figure 4 and Figure 5, we observe that most of the mis-classifications occur for makams which have very similar or the same scales, like *Mahur* and *Rast*, or *Muhayyer* and *Huseyni*, or *Beyati* and *Ussak,* or *Nihavent* and *Sultanıyegah.* This implies that the classifier learns and extracts musically meaningful information from the chroma features, regarding makams. Moreover, the common confusions in similar scales indicate the necessity of expanding the analysis into more other dimensions, like expanding chroma feature vector into two octaves, since octave equivalency does not hold for certain makams. Further tests are done to observe the effect of smoothing the chroma frames in time, which did not show any improvement in the performance of our system. At the classification stage, we have tried to reduce the dimensionality of our feature space using principal component analysis, which neither showed an increase in the results. Thus detailed discussions related to smoothing and dimensionality reduction are not provided in this study.

6. CONCLUSION

Our approach of automatic makam classification with chroma features sets a baseline for further research on the topic. Moreover, for reproducibility purposes, we share a Jupyter Notebook demonstration of our work[1]. The list of MusicBrainzIDs of the songs in this study can be found in the same repository. The future directions of our research include testing various other chroma features (NLSS Chroma, Chroma Toolbox) on automatic makam classification task and applying structural analysis for segmentation of makams by detecting the harmonic changes in the performance.

8. REFERENCES

Bayraktarkatal, M. E., Öztürk, O. M. (2012). Ezgisel kodların belirlediği bir sistem olarak makam kavramı: Hüseyni makamı'nın incelenmesi. *Porte Akademik*, 3, 24.

Bozkurt B. (2008), An Automatic Pitch Analysis Method for Turkish Maqam Music", *Journal of New Music Research*, 1-13.

Bozkurt B. (2012). Features for analysis of makam music, In P*roceedings of the 2nd CompMusic Workshop*; 12-13, 61-65.

Bozkurt B., R. Ayangil & A. Holzapfel (2014). Computational Analysis of Turkish Makam Music: Review of State-of-the-Art and Challenges. *Journal of New Music Research*, 43(1) pp. 3-23.

Dighe, P., Agrawal, P., Karnick, H., Thota, S. & Raj, B. (2013). Scale independent raga identification using chromagram patterns and swara based features. In *Multimedia and Expo Workshops (ICMEW), IEEE International Conference* on, 1–4.

Gedik, A. C. & Bozkurt, B. (2010). Pitch-frequency histogram based music information retrieval for Turkish music. *Signal Processing*, 90(4), 1049-1063.

Ioannidis, L., Gómez, E. & Herrera, P. (2011). Tonal-based retrieval of Arabic and middle-east music by automatic makam description. In *Proceedings International Workshop on Content-Based Multimedia Indexing*.

Karakurt, A., Sentürk, S. & Serra, X. (2016). MORTY: A Toolbox for Mode Recognition and Tonic Identification. In *Proceedings of the 3rd International Workshop on Digital Libraries for Musicology*.

Fujishima, T. (1999). Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music. *International Computer Music Conference* (ICMC), 464-467

Gómez, E. (2006). Tonal Description of Music Audio Signals. *The Astronomical Journal*, *35*(5), 220.

Müller, M. & Ewert, S. (2011). Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-Based Audio Features. *12th International Society for Music Information Retrieval Conference* (ISMIR).

Jiang, N., Grosche, P., Konz V. & Müller, M. (2011). Analyzing Chroma Feature Types for Automated Chord Recognition In *Proceedings of 42nd AES Conference*

Powers, H. S.& Wiering, F. (2001). et al. Mode. *Grove Music Online, Oxford Music On*line: http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg5S

Serra, X. (2017). The computational study of a musical culture through its digital traces. *Acta Musicologica*; 89(1), 24-44.

Serra, X., and Smith, J. (1990). "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition. *Computer Music Journal* 14 (4), 12-24.

Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2003). Pitch histograms in audio and symbolic music information retrieval. Journal of New Music Research, 32, 43–152.

E. Ünal, B. Bozkurt, and M. K. Karaosmanoglu. (2014). A Hierarchical Approach to Makam Classification of Turkish Makam Music, Using Symbolic Data. *Journal of New Music Research*, 43(1) 132-146

1 *https://github.com/emirdemirel/Supervised_Mode_Recognition*