MAKAM RECOGNITION USING EXTENDED PITCH DISTRIBUTION FEATURES AND MULTI-LAYER PERCEPTRONS

Furkan Yesiler
Universitat Pompeu Fabra
furkan.yesiler@gmail.com

Baris Bozkurt
Universitat Pompeu Fabra
baris.bozkurt@upf.edu

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

ABSTRACT

This work focuses on automatic makam recognition task for Turkish makam music (TMM) using pitch distributions that are widely used in mode recognition tasks for various music traditions. Here, we aim to improve the performance of previous works by extending distribution features and performing parameter optimization for the classifier. Most music theory resources specifically highlight two aspects of the TMM makam concept: use of microtonal intervals and an overall melodic direction that refers to the design of melodic contour on the song/musical piece level. Previous studies for makam recognition task already utilize the microtonal aspect via making use of high resolution histograms (using much finer bin width than one 12th of an octave). This work considers extending the distribution feature by including distributions of different portions of a performance to reflect the long-term characteristics referred in theory for melodic contour, more specifically for introduction and finalis. Our design involves a Multi-Layer Perceptron classifier using an input feature vector composed of pitch distributions of the first and the last sections together with the overall distribution, and the mean accuracy of 10 iterations is 0.756. The resources used in this work are shared for facilitating further research in this direction.

1. INTRODUCTION

In recent years, mode recognition has been one of the popular tasks in computational studies focusing on non-western music. There exist similarities as well as particular differences in mode descriptions (makam, mugam, raga, etc.) for different traditions. The following components play an important role in describing a mode: scale(s) (involving specification of interval sizes), typical melodic phrases, emphasis notes and melodic contour descriptors. Studying a mode recognition task requires taking culture specific descriptions for these dimensions into consideration. Here, we focus on makam recognition task within the context of Turkish makam music (TMM) where two aspects are specifically highlighted: design rules for overall melodic contour/shape and microtonal scale descriptors.

One of the frequently used features for mode recognition tasks is the pitch histogram/distribution. Pitch histograms have indeed been used as features in various automatic recognition tasks since early days of Music Information Retrieval (MIR) [1]. The most common method for pitch histogram based mode detection is "template matching" where pitch histogram templates are created for different modes, and the distance of the pitch histogram of a recording to these templates are measured to find the closest match, and assign the mode of the closest template as the mode of the recording. Alternatively, other machine learning algorithms (such as K-Nearest Neighbor (KNN)) are applied using pitch distributions as features.

For Western music studies (considering the tonality detection task), most of the methods in literature rely on template matching based on music theoretical [2], psychological [3] and data-driven models [4]. The psychological model proposed by Krumhansl and Kessler [3] is one of the most influential models, and proposes use of 24 templates for major and minor modes where pitch distributions are in the form of pitch class distributions using a 12 dimensional (equal tempered) pitch space.

Pitch histograms are also used in mode recognition tasks for Carnatic, Hindustani, Dastgah and TMM traditions. Raga recognition methods for Hindustani music are proposed by Chordia and Rae [5] using pitch distributions and Support Vector Machine (SVM) algorithm that demonstrated 0.750 accuracy for 31 ragas, and by Chordia and Senturk [6] comparing twelve-dimensional, fine-grained and kernel-density pitch distributions with different classifiers that demonstrated the accuracy of the best system as 0.915 for 23 raga classes. Dighe et al. [7] used a variety of features including chroma features, and achieved average accuracies in a range from 0.835 to 0.977 for 4 ragas in various experimental settings for Carnatic music. Another study by Dighe et al. [8] using pitch histograms demonstrated 0.943 accuracy for 8 ragas in Carnatic music. For Dastgah music, Abdoli [9] achieved an overall accuracy of 0.850 on 5 dastgahs with computing similarity measures between Interval Type 2 Fuzzy Sets of songs and dastgah prototypes. Heydarian [10] compared the performances of various features including chroma features and pitch histograms on automatic dastgah classification for 5 dastgah classes, and reported a mean accuracy of 0.868 with using spectral features and Manhattan distance. Gedik and Bozkurt [11] considered the makam recognition task for TMM, and applied template matching for 9 makams reporting a mean F-measure of 0.680 Ioannidis et al. [12]

compared template matching and SVM approaches using Harmonic Pitch Class Profiles (HPCP) as features. For a dataset containing 9 makams of TMM, their experiments achieved a mean F-measure of 0.730. The most recent attempt was demonstrated by Karakurt et al. [13], and used a dataset of 20 makams with 50 recordings for each makam to perform a joint tonic and makam recognition with using the KNN algorithm. Their results demonstrated a mean accuracy of 0.718.

In this research, we aim at automatic makam classification using pitch distributions as features and Multi-Layer Perceptron (MLP) algorithm as the classifier. We consider two strategies for improving performance of automatic classification: extending the feature vector with pitch distributions computed from specific portions of each recording and parameter optimization for the MLP model. The first strategy roots from the culture-specific overall melodic contour description for TMM that distinguishes makams using the same scale by their melodic contour. We have tested our system on the Ottoman-Turkish Makam Recognition Dataset published by Karakurt et al. [13] and achieved a mean accuracy of 0.756.

2. METHODOLOGY

In this section, we introduce our methodology which follows the general supervised classification approach in the MIR tasks. The initial step, feature extraction, consists of extracting the predominant melody, estimating the tonic frequency and creating pitch distributions for each recording. The obtained pitch distributions are then used as features, and an MLP model is trained. To test our system, the hyper-parameters of the MLP model are optimized with cross-validation, and the obtained model is evaluated on the test set.

2.1 Predominant Melody Extraction

For extracting the predominant melody in heterophonic recordings, Atli et al. [14] has proposed a tuned version of the algorithm by Salamon and Gomez [15] for TMM recordings (with specific settings on windowing and dynamic range) including a post-processing step to correct octave errors and remove noisy estimates in low-energy regions. In our experiments, the pitch series extracted by using the algorithm of Atli et al. [14] and shared by Karakurt et al. [13] is used.

2.2 Tonic Frequency Estimation

One key feature defining the pitch space of TMM recordings is the use of several possible concert pitch standards (named as the *ahenk* system) instead of a single standard (such as A4=440Hz). While all notations are written in a single key, in interpretation, key transposition is applied. The most commonly used ahenk today is Bolahenk which specifies *'neva perdesi'* (D on staff notation) as about 440 Hz. The key transposition standards define frequency ranges instead of fixed well-defined frequency values. A table of frequency ranges for each ahenk is presented by Erguner [16]. For this reason, a tonic frequency estimation step is essential for aligning pitch distributions in the frequency plane. Gedik and Bozkurt [11] handled the tonic frequency estimation by using pitch histograms with the assumption that the makam is known. This technique relies on sliding the pitch histogram of each recording on the pre-defined makam pitch histogram template, finding the closest matching location, and by assigning the peak that matches the tonic peak of the template as the tonic of the recording, the tonic frequency is estimated. Atli et al. [17] proposed to use a specific property of TMM, all the performances ending in the tonic frequency. A post-filter is used to reduce the existing noise at the end of the recordings; thus, a reliable estimation of the tonic frequency is achieved via estimating the frequency of the last note. The dataset published by Senturk [18] contains manual annotations for the last note from which tonic frequencies are extracted. We use this ground truth tonic data in our experiments.

2.3 Computing Pitch Distributions

In our study, pitch series and tonic frequencies data published by Karakurt et al. [13] are used to represent the pitch series information in intervals (in cents) with respect to the tonic frequency. The size of the distribution should conform the pitch space divisions specified in the music culture being studied. In TMM, the most accepted theory is Arel-Ezgi-Uzdilek (Arel) [19]. Arel theory uses the 53-TET system that divides an octave into 53 equal tempered notes. In our study, the 53-TET system is utilized to create the pitch distributions.

Pitch class distributions are proven useful for mode recognition tasks in which the scales are the main characteristics of modes. In TMM, however, the set of notes are not the only identifier of makams. Along with the emphasis notes, the direction of the melody is one of the main characteristics to separate makams that share the same scale. To utilize this characteristic, the feature vector is extended with 53-bin pitch distributions of the first and the last sections of the recordings. Since makams that share the same scale would present different distributions of notes in the beginning and in the end of the song, extending our feature vector as mentioned above would facilitate classifying makams using the same scale. To find the best performing proportion to use for our task, we test various section sizes from 5% to 50%.

2.4 Training and Optimizing a Multi-Layer Perceptron

With created pitch distributions, a dataset of 1000 instances and 159 attributes is obtained, and the values for each instance are normalized with respect to maximum value. For evaluation of our model, stratified subsets are created in a random way: a training subset with 900 instances (45 instances per makam) and a test subset with 100 instances (5 instances per makam). 10-fold cross validation is performed with the training subset (i.e. the train set is further divided into train and validation sets). To select the best performing model, a grid search is performed to find the optimum combination of the hidden layer size and alpha coefficient. Hidden layer size defines the number of hidden layers and the number of nodes in each hidden layer. Alpha coefficient is the weight parameter of the regularization term using L2 norm. The hidden layer sizes we considered are one hidden layer with 70, 90, 110, 130, 150, 170, 190 nodes while the alpha coefficients are 0.1, 0.01, 0.001. The momentum coefficient for stochastic gradient descent algorithm is set to 0.5 and the learning rate is 0.001. 10-fold cross validation with all the hyper-parameter combinations gives the best performing combination, and an MLP model using the selected hyper-parameters is created. To estimate the accuracy, the MLP model is finally trained with the entire training subset and the obtained model is used to predict the classes of instances in the test subset. To get a mean accuracy value, this process is repeated 10 times with different randomized and stratified training and test subsets.

3. EXPERIMENTS AND RESULTS

In this section, we introduce the dataset used in the experiments, and present the results for various experimental settings. Our data is balanced in terms of the number of instances for each class. To compare the achieved results, we use the mean accuracy.

3.1 Dataset

For our experiments, the Ottoman-Turkish Makam Recognition Dataset published by Karakurt et al. [13] is used. It contains 50 recordings each from 20 most performed makams in Turkish Music. The recordings are taken from commercial albums with metadata present in Music Brainz, and their makams are annotated by using the information from each relevant album. In case of missing makam information, other sources such as musical scores of the pieces were used. Tonic frequencies and the pitch values of extracted predominant melodies for each recording are also included in the dataset. The recordings are taken from commercial albums available in CD format including live and studio recordings, monophonic and heterophonic recordings. The quality of the recordings vary in terms of noise.

3.2 Analyzing Scale Relationships of Makams Using Pitch Histograms

Prior to the automatic recognition tests, to be able to interpret the results, we first consider forming hierarchical clustering (using the specific implementation in scipy [20]) based on histogram distances, and studying groupings formed automatically. We created pitch histogram templates via averaging tonic aligned pitch histograms of 50 recordings per makam, and calculated the distances between them using Canberra distance [21]. Figure 1 illustrates that makams using the same or similar scales are grouped closer such as “Rast”-“Mahur”, “Huseyni”-“Beyati”-“Ussak”-“Muhayyer”. These makams are indeed among the most confused ones in makam recognition studies. Most of the TMM theory resources present these makams as using very similar scales, and mainly differing in overall melodic contour characteristics.

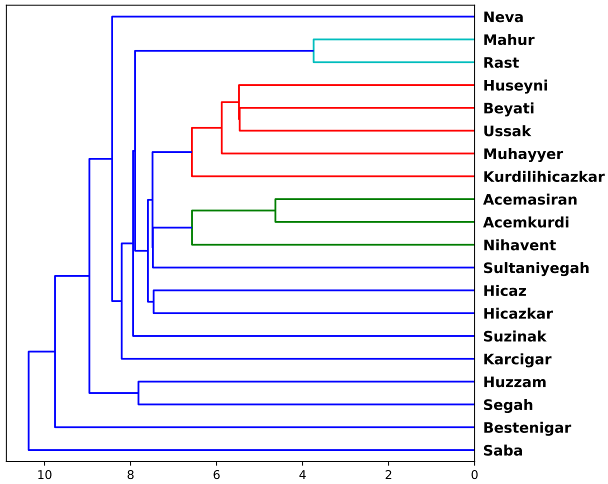

Figure 1. Hierarchical clustering of makams using pitch histograms

3.3 Automatic Classification using the Entire Recordings

To investigate the performance of using MLP models with pitch distributions, the first experiments consider using the pitch distribution of the entire record, without the first and the last sections. As the result of 10 iterations, the achieved aggregated confusion matrix is presented in Figure 2. The mean accuracy of 10 iterations is 0.726.

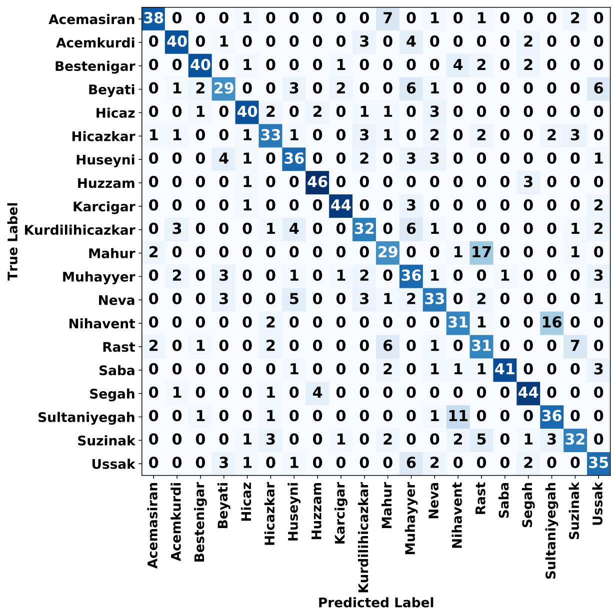

Figure 2. Confusion matrix using overall distributions

In terms of the mean accuracy, we observe that the most accurately classified makams are “Huzzam” with 0.920, “Karcigar” and “Segah” with 0.880 mean accuracy values. The most confused makam pairs are “Rast”-“Mahur” and “Nihavent”-“Sultaniyegah”.

3.4 Analyzing the Characteristics of the First and the Last Sections of the Recordings

Here, we present two examples for demonstration of differences observed in histograms of different parts of a piece. In Figure 3, we present pitch histograms of two makam pairs, “Huseyni”-“Muhayyer” (left) and “Beyati”-“Muhayyer” (right) computed via averaging histograms of all recordings in a given makam. The distribution of the note annotated by Peak 1 is considerably higher for “Huseyni” than “Muhayyer” in the first 10% of the recordings; however, in the last 10%, such relationship is not observed. For Peak 2, the figure demonstrates that the ratio of distributions are significantly different in the first 10% of the recordings compared to the last 10%. Similar trend is observed for the distributions of “Beyati” and “Muhayyer” for Peak 3, 4 and 5. These observations are in-line with TMM theory which explicitly specifies emphasis of these specific notes in the introduction of a composition or improvisation.

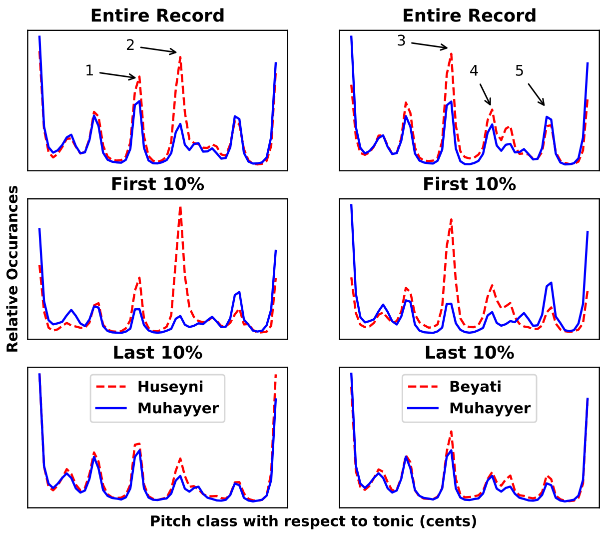

Figure 3. Pitch histograms of the entire recordings and the first and the last sections

3.5 Automatic Classification with Including the First and the Last Sections of the Recordings

For analyzing the effects of extending the feature set with the first and last sections of each recording, the second set of experiments consider the extended feature sets. Pitch distributions of the first and the last sections of each recording are added as extra features. In Figure 4, the mean accuracy results of using different sizes of the first and the last sections of the recording to extend the feature sets are presented. The best performing method is using the first and the last 30% of the recording in addition to the entire record, although using 10% demonstrates slightly lower results. The highest mean accuracy of the 10 iterations for using 30%-size sections is 0.756.

In Figure 5, we present the aggregated confusion matrix of 10 iterations with using 30% for the section size. “Segah”, “Huzzam”, “Acemasiran” and “Hicaz” are the top 4 most accurately classified makams with the mean accuracy values 0.920, 0.880, 0.820 and 0.820, respectively. The lowest accuracy results are found for “Beyati”, “Kurdilihicazkar” and “Rast” (they are confused with makams using the same or similar scales).

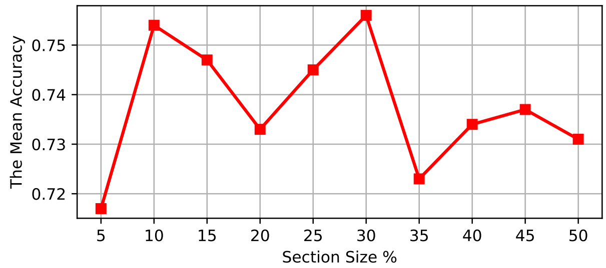

Figure 4. The mean accuracy values for different first and last section sizes

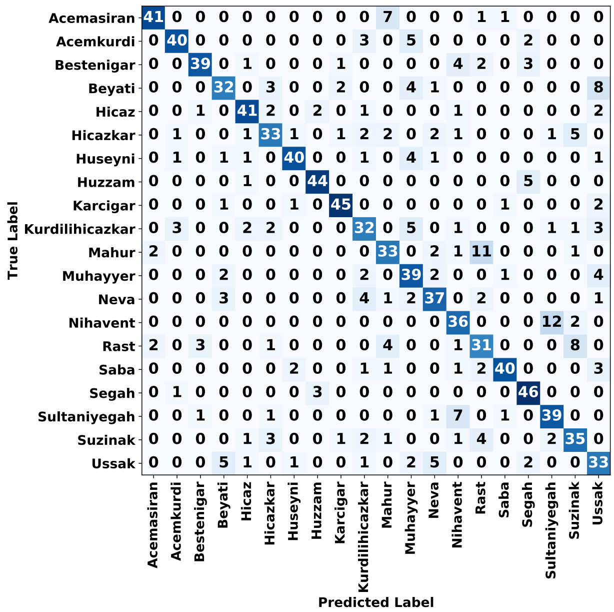

Figure 5. Confusion matrix using the extended feature vector with the first and the last section distributions

In terms of the mean accuracy, we observe that extending our feature set improves the performance of the classifier. For “Nihavent”, “Huseyni”, “Mahur” and “Neva” makams, the mean accuracy increases 10%, 8%, 8% and 8%, respectively. The extended feature set has a positive or neutral effect on the mean accuracy, except for 4 cases: for “Huzzam” and “Ussak” 6%, for “Bestenigar” and “Saba” 2% decrease in the mean accuracy is observed.

4. DISCUSSION AND CONCLUSION

In this study, we have considered the automatic makam recognition task for TMM using an MLP classifier with pitch distribution features. The initial set of experiments using the pitch distributions of the entire recordings resulted in a mean accuracy of 0.726. The misclassified makam pairs such as “Rast”-“Mahur”, “Nihavent”-

"Sultaniyegah" are similar in terms of scales; hence, it is an expected result considering TMM theory. The parameter optimization with cross validation and using an MLP model as a classifier can be accounted for the slight improvement in the mean accuracy over the state of the art.

We have further considered extending the feature vector by including pitch distributions of the first and the last sections of the recordings. Testing with values ranging from 5% to 50%, we observed that using 30% as the size of the sections resulted in the highest mean accuracy. The classification performances for a number of makams that share the same or similar scales with others such as "Nihavent"-"Sultaniyegah" and "Beyati"-"Huseyni"-"Muhayyer" are improved by this extension of the feature vector, and the highest achieved mean accuracy of the proposed method is 0.756. To facilitate further research on this field, the source code used to produce each step of the experiment is shared. [1]

5. REFERENCES

[1] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch histograms in audio and symbolic music information retrieval," in *Journal of New Music Research*, vol. 32, no. 2, 2003, pp. 143–152.

[2] H. C. Longuet-Higgins and M. Steedman, "On interpreting bach," in *Mental processes: Studies in cognitive science*. MIT Press, 1987, pp. 221–241.

[3] C. L. Krumhansl and E. Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," in *Psychological Review*, vol. 89, no. 4, 1982, pp. 334–368.

[4] D. Temperley and E. W. Marvin, "Pitch-class distribution and the identification of key," in *Music Perception: An Interdisciplinary Journal*, vol. 25, no. 3, 2008, pp. 193–212.

[5] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions," in *Proc. of the 8th Int. Conf. on Music Information Retrieval (ISMIR 2007)*, Vienna, 2007, pp. 431–436.

[6] P. Chordia and S. Senturk, "Joint recognition of raag and tonic in North Indian Music," in *Computer Music Journal*, vol. 37, no. 3, 2013, pp. 82–98.

[7] P. Dighe, P. Agrawal, H. Karnick, S. Thota, and B. Raj, "Scale independent raga identification using chromagram patterns and swara based features," in *Electronic Proc. of the 2013 IEEE Int. Conf. on Multimedia and Expo Workshops*, San Jose, 2013, pp. 1–4.

[8] P. Dighe, H. Karnick, and B. Raj, "Swara histogram based structural analysis and identification of Indian Classical ragas," in *Proc. of the 14th Int. Society for Music Information Retrieval Conf. (ISMIR 2013)*, Curitiba, 2013, pp. 35–40.

[9] S. Abdoli, "Iranian traditional music dastgah classification," in *Proc. of the 12th Int. Conf. on Music Information Retrieval (ISMIR 2011)*, Miami, 2011, pp. 275–280.

[10] P. Heydarian, "Automatic recognition of Persian musical modes in audio musical signals," Ph.D. dissertation, London Metropolitan University, 2016.

[11] A. C. Gedik and B. Bozkurt, "Pitch-frequency histogram-based music information retrieval for Turkish music," in *Signal Processing*, vol. 90, no. 4, 2010, pp. 1049–1063.

[12] L. Ioannidis, E. Gomez, and P. Herrera, "Tonal-based retrieval of Arabic and Middle-East music by automatic makam description," in *9th Int. Workshop on Content-based Multimedia Indexing*, Madrid, 2011, pp. 31–36.

[13] A. Karakurt, S. Senturk, and X. Serra, "MORTY: A toolbox for mode recognition and tonic identification," in *3rd Int. Digital Libraries for Musicology Workshop*, New York, 2016, pp. 9–16.

[14] H. S. Atli, B. Uyar, S. Senturk, B. Bozkurt, and X. Serra, "Audio feature extraction for exploring Turkish makam music," in *Proc. of 3rd Int. Conf. on Audio Technologies for Music and Media*, Ankara, 2014, pp. 142–153.

[15] J. Salamon and E. Gomez, "Melody extraction from polyphonic music signals using pitch contour characteristics," in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 6, 2012, pp. 1759–1770.

[16] S. Erguner, *Ney, Metod*. Istanbul: Erguner Muzik, 2007.

[17] H. S. Atli, B. Bozkurt, and S. Senturk, "A method for tonic frequency identification of Turkish makam music recordings," in *Proc. of 5th Int. Workshop on Folk Music Analysis*, Paris, 2015, pp. 119–122.

[18] S. Senturk, "Computational analysis of audio recordings and music scores for the description and discovery of Ottoman-Turkish makam music," Ph.D. dissertation, Universitat Pompeu Fabra, 2016.

[19] H. S. Arel, *Turk Musikisi Nazariyati Dersleri*. Istanbul: Husnu Tabiat Matbaasi, 1930/1968.

[20] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–, [Accessed 26- May- 2018]. [Online]. Available: http://www.scipy.org/

[21] G. N. Lance and W. T. Williams, "Computer programs for hierarchical polythetic classification ("similarity analyses")," in *The Computer Journal*, vol. 9, no. 1, 1966, pp. 60–64.

[1] https://github.com/furkanyesiler/pitchdistamr