INCORPORATING FEATURES OF DISTRIBUTION AND PROGRESSION FOR AUTOMATIC MAKAM CLASSIFICATION

Erdem Ünal
Tübitak-Bilgem,
Istanbul, Turkey
erdem.unal@tubitak.gov.tr

Barış Bozkurt
Bahçeşehir University,
Istanbul, Turkey
baris.bozkurt@bahcesehir.edu.tr

M. Kemal Karaosmanoğlu
Yıldız Teknik University,
Istanbul, Turkey
kkara@yildiz.edu.tr

ABSTRACT

Automatic classification of makams from symbolic data is a rarely studied topic. In this paper, first a review of an n-gram based approach is presented using various representations of the symbolic data. While a high degree of precision can be obtained, confusion happens mainly for makams using (almost) the same scale and pitch hierarchy but differ in overall melodic progression, *seyir*. To further improve the system, first n-gram based classification is tested for various sections of the piece to take into account a feature of the seyir that melodic progression starts in a certain region of the scale. In a second test, a hierarchical classification structure is designed which uses n-grams and seyir features in different levels to further improve the system.

1. INTRODUCTION

Automatic classification of makams is to some extend a similar research topic as key or mode finding for Western music. While this technology finds use in various information retrieval applications, such a study also provides us an insight about the makam concept. As a concept in oral tradition, makam is often defined with loose verbal descriptions in makam music theory. Attempts in defining measurable features for classifying makams can potentially improve our understanding of makam music.

Computational studies on makam music can be very broadly classified into two categories based on the type of data being processed: symbolic or audio. While some works such as [1, 2, 3, 4] propose systems for makam recognition from audio data, works on symbolic data appear to be much more limited, probably due to lack of machine readable data.

The first study on n-grams for makam recognition was presented in [5] (which is a shorter version of [6]) by Alpkoçak and Gedik. In a recent work[7], Unal, Bozkurt and Karaosmanoğlu proposed a new n-gram perplexity based system and studied the effect of representing the scores in 12TET (Tone Equal Temperament) which was used in [5] and [6] and the Arel system [8] for performance comparison. It is observed that, using a large dataset, and challenging makam couple sets, the system using Arel representation outperformed the system using the 12-TET representation by %3.7 percent on average.

In that work, a recall performance of %88.2 was achieved. As expected, the most confused makams were reported to be the ones that use the same set of pitches, the same set of tetrachord - pentachord formulation and the same tonic.

This study is an extension of [7] to improve the system by including new features of overall melodic direction in the classifier.

The plan of the manuscript is as follows. In the second section, we summarize the system presented in [7]. Later in the third section, we present experimental results on dividing the score into sections and performing classification. In the fourth section, we present a hierarchical classifier where a first level n-gram based classifier is followed by a classifier that uses melodic progression features.

2. THE TESTS USING THE PERPLEXITY BASED SYSTEM

3.1 N-gram models

N-grams are widely used in computational linguistics, probability, communication theory and computational biology as well as music information retrieval [9, 10]. N-grams predict X_i based on $X_{i-(n-1)}$....., X_{i-1}. In theory this is the information calculated by $P(X_i|X_{i-(n-1)},...,X_{i-1})$. Given sequences of a certain set, one can statistically model this set by statistically counting the sequences that belong to it. The main hypothesis to be tested here is that, the short-time melodic contour and the frequency of makam specific notes are selective features for defining makams. This is why n-gram models are selected for training makam models. Given a notation sequence, using perplexity, the system will define how well the input sequence can be generated by the makam models in the database. The makam model that has the maximum similarity score is selected as the output of the system.

In practice, it is necessary to smooth the probability distributions by assigning non-zero probabilities to unseen words or n-grams. Written-Bell smoothing technique available in the SRILM toolkit (http://www.speech.sri.com/projects/srilm/) is used in our experiments.

2.1 Perplexity

Perplexity is a metric that is widely used for comparing probability distributions. The perplexity can be stated as the perplexity of the distribution over its possible values of x. Given a proposed probability model q (in our case: a makam model), evaluating q by asking how well it predicts a separate test sequence or set x_1, x_2, ...,x_N (in our case: a microtonal note sequence) also drawn from p, can be performed by using the perplexity of the model q, defined by:

$$2^{\sum_{i=1}^{N} \frac{1}{N} \log_2 q(x_i)} \quad (1)$$

For the test events, we can see that better models will assign better probability scores thus a lower perplexity score which means it has a better potential to compress that data set. The exponent is the cross entropy per definition:

$$H(p,q) = -\sum_x p(x) \log_2 q(x) \quad (2)$$

The cross entropy thus the perplexity is the similarity measure for the test instance and the makam models in the search space. For each of the makam model defined, the system calculates the similarity metric to evaluate which makam is the most similar to the input sequence given.

2.2 Data

The symbolic data used is a subset derived from the largest symbolic database of TMMT we recently announced [11]. The makam selection is based on two criteria: commonness and similarity. On purpose, makam couples such as *Hüseyni* - *Muhayyer* and *Beyati* - *Uşşak* have been included in the set. The scale for makam *Hüseyni* and makam *Muhayyer* (top figure) and makam *Uşşak* and makam *Beyati* (bottom figure) are the same as presented in Figure 1. From makam music theory, we know that pitch hierarchy, melodic direction, typical phrases and typical makam transitions appear to be the discriminating features for makams having the same set of pitches and tonic.

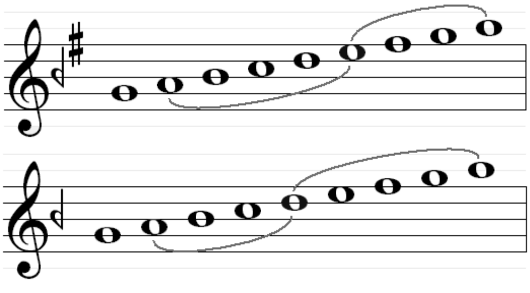

Figure 1. Scale used for makam *Hüseyni* and makam *Muhayyer* (top), makam *Beyati* and makam *Uşşak*

Due to the availability at the time of the experiments, this study uses the following subset:

Makam name	*Total # of Songs*	*Total # of Notes*
Beyati	42	17 347
Hicaz	117	39 301
Hicazkar	49	14 775
Hüseyni	71	23 787
Hüzzam	65	23 581
Kürdilihicazkar	51	18 332
Mahur	54	18 039
Muhayyer	50	16 774
Nihavent	86	28 724
Rast	88	29 103
Saba	45	16 486
Segah	74	21 744
Uşşak	85	26 379
TOTAL	**877**	**294 372**

Table 1. Makam coverage and note statistics for the test database

An additional filtering has been applied to the data compared to the data used in [7]. It has been observed on some examples in [7] that interludes, consisting of repeated short melodic segments of some pieces do not obey the melodic progression rules defined for the specific makam. Personal communication with masters on this issue resulted in the decision that the interludes can be filtered out. Therefore, the data has been preprocessed and all interludes of pieces with lyrics are filtered out.

2.3 Experimental setup

For testing, the leave-one-out technique is used. For each of the test trials, one song from the database was chosen as the input. The rest of the pieces are used for modeling the makam classes.

3. USING DIFFERENT SECTIONS OF THE PIECE FOR CLASSIFICATION

In makam theory, *seyir*, the overall melodic progression is described as a road map or an ordered sequence of emphasized notes in a piece or improvisation. For makams using the same scale and tonic, this progression is the main discriminating feature.

In order to observe the general melodic progression of the selected makams in our dataset, we down-sampled the melodic contours of each piece so that they have the same length (of 20 points) and plotted these as points in Figure 2. The solid line shown in the figures are obtained by averaging all melodic contours. Figure 2 presents the obtained average melodic progression for makam *Muhayyer* and *Hüseyni*. The highest differences of the two progressions are observed during the first quarter. Similar observations are made for other very close makam couples. For this reason, an experiment is designed to study n-gram based classification success using only the first quarter of the piece.

Using only the first quarter reduces the data considerable for modeling. For that reason, the following tests are performed:

i) the whole input sequence tested against the models derived from the whole music pieces

ii) the first quarter of the input sequence tested against the models derived from the whole music pieces

iii) the first quarter of the input sequence tested against the models derived from only the first quarter of the music pieces

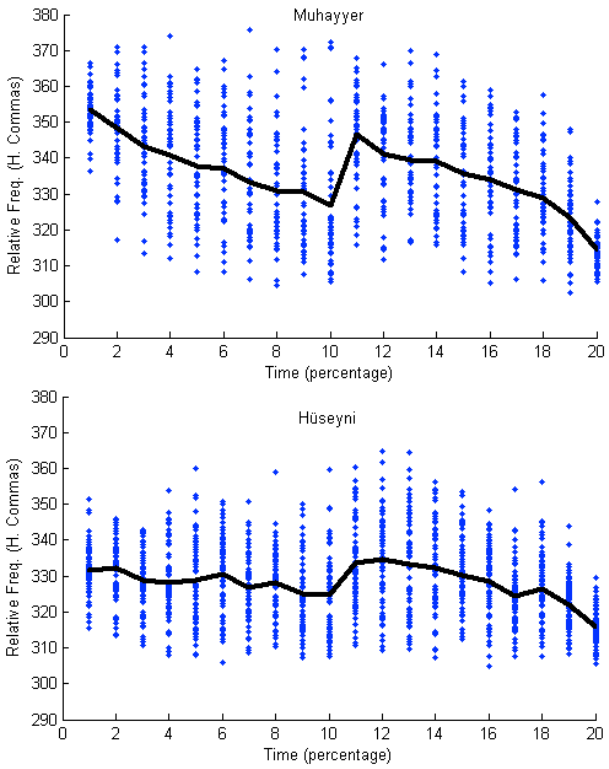

Figure 2. Melodic progression of makam *Muhayyer* and *Hüseyni*.

The tests are repeated for increasing number of n and best result is taken (it appears to be n=2 for all these cases). In the Appendix, we present the confusion matrices for the three tests. The second test, where the models are built from the entire pieces and tested only against the first quarter of the input provides the best result for average makam detection accuracy. We observe that, compared to classification using the entire piece, accuracy of the makam detection can be slightly (%0.9) improved if detection is performed on the first quarter of the piece. Performing both modeling and testing on the first quarter of the piece provides lower accuracy values. We think this is due to the reduction in the size of the remaining data for modeling.

4. HIERARCHICAL CLASSIFICATION

As a result of close observation on the confusion matrices, a hierarchical classification is considered to be worth testing. The first level classification groups makams *Muhayyer, Hüseyni, Uşşak, Beyati* and *Rast* in one group, *Segah* and *Hüzzam* in one group. Then in higher levels other features are used to further perform classification within a group.

As observed in Figure 2, the starting region (the first value of the down-sampled melodic contour) for progression is a potentially discriminating feature. In addition, in [6], authors propose use of sum of deltas of consecutive notes (i.e. summation of all melodic intervals of the piece) showing the total overall progression of the piece as a numerical value in commas. In Figure 3, we present *Hüseyni* and *Muhayyer* data on this two dimensional feature plane. It is obvious that these features (starting point and sum of melodic intervals) are potentially useful for discriminating such close makams.

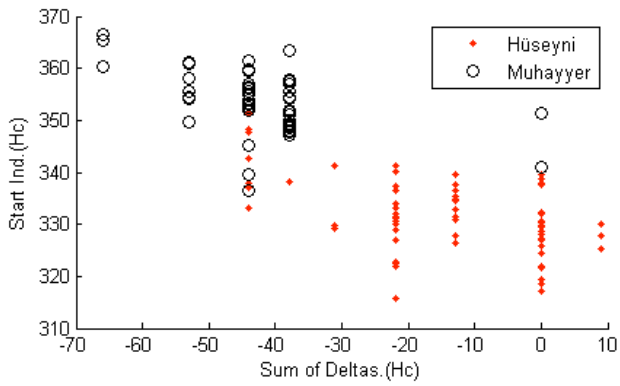

Figure 3. Seyir features: starting region and sum of deltas for makams *Muhayyer* and *Hüseyni*.

We are currently working on the development of the hierarchical classifier and the results will be presented during the workshop.

5. CONCLUSIONS

The present study is a continuation of a very recent work on n-gram based makam classification. In the first step, we tested the effect of using only the first quarter of the pieces to classification performance and observed that minor improvement is achieved. As the second step, a hierarchical classifier taking into account some melodic progression features is being developed. Tests will be performed and presented in the workshop.

Acknowledgments

This work was funded in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 267583 (CompMusic) and in part by TÜBİTAK ARDEB grant no:3501-109E196. All staff notation representations used in this paper is taken from Mus2okur software, a digital encyclopedia for Turkish music (http://www.musiki.org/).

6. REFERENCES

[1] S. Abdoli,. "Iranian Traditional Music Dastgah Classification," in Proc. International Society for Music Information Retrieval ISMIR, 2011.

[2] N. Darabi, N. Azimi and H. Nojumi, "Recognition of Dastgah and Makam for Persian Music with Detecting Skeletal Melodic Models," in Proc. 2nd IEEE BENELUX/DSP Valley Signal Processing Symposium, 2006.

[3] A. C. Gedik, and B. Bozkurt, "Pitch-frequency histogram-based music information retrieval for Turkish music," Signal Processing, 2010, 90(4), 1049-1063.

[4] L. Ioannidis, E. Gómez, and P. Herrera, "Tonal-based retrieval of Arabic and Middle-East music by automatic makam description," in Proc. CBMI, 2011.

[5] A. Alpkoçak and A. C. Gedik. "Classification of Turkish songs according to makams by using n grams," in Proc. the 15. Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN), 2006.

[6] A.C. Gedik, C. Işıkhan, A. Alpkoçak, Y. Özer, “Automatic Classification of 10 Turkish Makams”, in Proc. Int. Cong. on Representation in Music & Musical Representation, İstanbul, 2005.

[7] E. Unal, B. Bozkurt, M. K. Karaosmanoğlu, “N-gram based Statistical Makam Detection on Makam Music in Turkey using Symbolic Data,” in Proc. Int. Society for Music Information Retrieval (ISMIR), 2012.

[8] H. S. Arel, "Türk Musikisi Nazariyatı Dersleri, Hazırlayan Onur Akdoğu," Kültür Bakanlığı Yayınları /1347, Ankara, 1991, p.70.

[9] S. Doraisamy, "Polyphonic Music Retrieval: The N - gram Approach," PhD Thesis, University of London, 2004.

[10] S. Downie "Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text," PhD thesis, University of Western Ontario, 1999.

[11] M. K. Karaosmanoğlu, “A Turkish makam music symbolic database for music information retrieval: SymbTr,” in Proc. Int. Society for Music Information Retrieval (ISMIR), 2012.

7. APPENDIX

2-gram RESULTS:														
byati	hicaz	hczkr	hsyni	huzzm	krdhz	mahur	muhyr	nhvnt	rast	saba	segah	ussak		
20	0	0	5	0	0	0	1	0	1	0	0	15	byati	47,6
0	117	0	0	0	0	0	0	0	0	0	0	0	hicaz	100
0	0	48	0	0	1	0	0	0	0	0	0	0	hczkr	98
4	1	0	45	0	0	0	14	0	1	1	0	5	hsyni	63,4
0	0	0	0	62	0	0	0	0	0	0	3	0	huzzm	95,4
0	0	0	0	0	51	0	0	0	0	0	0	0	krdhz	100
0	0	0	0	0	0	51	1	0	2	0	0	0	mahur	94,4
3	0	0	5	0	0	0	37	0	1	0	0	4	muhyr	74
0	0	0	0	0	1	0	0	85	0	0	0	0	nhvnt	98,8
0	0	0	0	0	0	0	1	0	81	0	0	6	rast	92
0	0	0	0	0	0	0	1	0	0	44	0	0	saba	97,8
0	0	0	0	1	0	0	1	0	0	0	72	0	segah	97,3
9	0	0	9	0	0	0	5	0	3	0	1	58	ussak	68,2
55,6	99,2	100	70,3	98,4	96,2	100	60,7	100	91	97,8	94,7	65,9		
TOTAL WEIGHTED ACC is: 87.9				TOTAL MAKAM ACC is : 86.7										

Table 1. Results for the first test: model derived from the whole, test performed on the whole

2-gram RESULTS:														
byati	hicaz	hczkr	hsyni	huzzm	krdhz	mahur	muhyr	nhvnt	rast	saba	segah	ussak		
20	0	0	5	0	0	0	1	0	0	0	0	16	byati	47,6
0	116	0	1	0	0	0	0	0	0	0	0	0	hicaz	99,1
0	0	47	0	1	1	0	0	0	0	0	0	0	hczkr	95,9
5	1	0	47	0	0	0	12	0	1	1	0	4	hsyni	66,2
0	0	0	0	63	0	0	0	0	0	0	2	0	huzzm	96,9
0	0	0	0	0	51	0	0	0	0	0	0	0	krdhz	100
0	0	0	0	0	0	51	1	0	2	0	0	0	mahur	94,4
2	0	0	5	0	0	0	41	0	0	0	0	2	muhyr	82
0	0	0	0	0	1	0	0	85	0	0	0	0	nhvnt	98,8
0	0	0	0	0	0	0	1	0	82	0	0	5	rast	93,2
0	0	0	0	0	0	0	1	0	0	44	0	0	saba	97,8
0	0	0	0	2	0	0	1	0	0	0	71	0	segah	95,9
8	0	0	8	0	0	0	5	0	3	0	1	60	ussak	70,6
57,1	99,1	100	71,2	95,5	96,2	100	65,1	100	93,2	97,8	95,9	69		
TOTAL WEIGHTED ACC is: 88.7				TOTAL MAKAM ACC is : 87.6										

Table 2. Results for the second test: model derived from the whole, test performed on the first quarter

2-gram RESULTS:														
byati	hicaz	hczkr	hsyni	huzzm	krdhz	mahur	muhyr	nhvnt	rast	saba	segah	ussak		
23	0	0	3	0	0	0	1	0	0	0	0	15	byati	54,8
0	113	0	3	0	0	0	1	0	0	0	0	0	hicaz	96,6
0	0	41	0	5	1	0	0	0	0	0	2	0	hczkr	83,7
5	2	0	54	0	0	0	4	0	2	1	0	3	hsyni	76,1
0	0	1	0	61	0	0	0	0	0	0	3	0	huzzm	93,8
0	0	3	0	1	42	0	0	4	0	0	1	0	krdhz	82,4
0	0	0	1	0	0	51	2	0	0	0	0	0	mahur	94,4
0	0	1	4	0	0	0	45	0	0	0	0	0	muhyr	90
0	0	0	0	0	2	0	0	83	0	0	1	0	nhvnt	96,5
2	0	0	3	1	0	0	1	0	72	0	0	9	rast	81,8
0	0	0	0	0	0	0	0	0	0	44	0	1	saba	97,8
0	0	0	1	2	0	0	0	0	0	0	70	1	segah	94,6
14	0	0	10	0	1	0	0	0	5	0	0	55	ussak	64,7
52,3	98,3	89,1	68,4	87,1	91,3	100	83,3	95,4	91,1	97,8	90,9	65,5		
TOTAL WEIGHTED ACC is: 86				TOTAL MAKAM ACC is : 85.2										

Table 3. Results for the third test: model derived from the first quarter, test performed on the first quarter