Tonal-based retrieval of Arabic and Middle-East music by automatic makam description

Leonidas Ioannidis	Emilia Gómez	Perfecto Herrera
Music Technology Group,	*Music Technology Group,*	*Music Technology Group,*
Universitat Pompeu Fabra	*Universitat Pompeu Fabra*	*Universitat Pompeu Fabra*
Iwannidis_8@yahoo.gr	emilia.gomez@upf.edu	perfecto.herrera@upf.edu

Abstract

The automatic description of music from traditions that do not follow the Western notation and theory needs specifically designed tools. We investigate here the makams, which are scales in the modal music of Arabic and Middle East regions. We evaluate two approaches for classifying musical pieces from the 'makam world', according to their scale, by using chroma features extracted from polyphonic music signals. The first method compares the extracted features with a set of makam templates, while the second one uses trained classifiers. Both approaches provided good results (F-measure=0.69 and 0.73 respectively) on a collection of 302 pieces from 9 makam families. Furthermore, error analyses showed that certain confusions were musically coherent and that these techniques could complement each other in this particular context.

1. Introduction

Automatic analysis of musical signals is one of the fundamental activities of Music Information Retrieval (MIR) research, which tries to help musicians and listeners retrieve and access music from large databases in a musically meaningful way. In a musically significant way means that the user is able to retrieve musical pieces by using high-level music semantics such as structure, genre, or instrumentation. One of such semantic concept is the so-called "scale" of a piece, and its automatic determination has been long studied for Western music with very satisfying results [6]. By Western music we consider all music that falls into the broad use of Western music theory and the use of equal-tempered musical scales. On the other hand, we can find music from various parts of the world that make little, or not at all, use of a Western theoretical system. Most of this music is orally transmitted and is restricted to a given cultural group.

MIR, as a young discipline, has mainly focused in Western music culture [1], although a side-branch of this multidisciplinary field, called computational ethnomusicology [4], has recently emerged with the purpose of exploring and studying ethnical music cultures with the tools and techniques developed by MIR researchers. Some of these studies are devoted to infer theoretical systems and representations that did not exist before. It should be noted that in many cases we need to adapt standard MIR techniques to conduct computational ethnomusicology.

The present study is related to makam scales, which are tonal scales/modes used in various traditions such as Arab, Turkish and Persian music. Calling the makam just a tonal scale is a simplification, since it is a fused musical concept consisting on various elements. In general, it is considered to be a compositional tool that defines, besides the intervallic distances of the scale degrees, the tonic, the weight of the other degrees in the scale and the melodic sequence (not in an absolute way but rather in an ascending/descending behavior). Finally it is a common belief among musicians and theorists that makam implies mood or, in a 'Western' sense, genre. Various theoretical models of makam music have been proposed since the 9th century, and some music has been annotated accordingly [10]. Lately, the Arel-Ezgi-Uzdilek System (AEU or simply Arel), which has also adopted the use of the western staff, has been elevated as the 'official theory of Turkish music', and is used today in conservatories [8].

Some efforts have been devoted to the estimation of makam from music material. Gedik and Bozkur have proposed an approach for estimating the makam from monophonic signals by using fundamental frequency envelopes to construct pitch histograms [1,2,3]. The obtained pitch histograms can then be used to build templates, which represent the interval distribution of the different makam types. By measuring a distance from any histogram extracted from a musical excerpt, the model can then estimate the makam. This technique is called template-matching. The mentioned work includes a tonic detection algorithm based on the theory and properties of that music. There is also another proposal that makes use of the note sequence (*seyir*) to estimate the makam [9].This approach obtains a high accuracy using decision trees on the notes sequence, but it is not using any audio information only midi transcriptions of that music.

This work continues previous research in automatic makam estimation. The first goal of this study is to extend to this problem to polyphonic music and apply chroma features (Harmonic Pitch Class Profiles or HPCP), which were found to be useful alternatives to fundamental estimation in the analysis of polyphonic music signals containing Western classical music [6]. Here we will analyze the limitations of chroma features regarding octave equivalence, given that a high-dimensional pitch space may be needed for our particular problem [2]. Our second goal is to compare some class modeling techniques, used as tools of the trade in MIR, with the template-matching algorithm, where chroma features will form the feature space. This approach has been successfully applied to Western musical key estimation [7].

2. Methodology

2.1. Makam types

Makams are said to be possible in more than 300 different variants, although the actual practice restricts them to approximately 30. In our study only 9 scales are being studied: *hicaz, huseyni, rast, nihavend, kurdili hicazkar, saba, segah, huzzam* and *ussak*, which are claimed to cover the 50% of today's practice [2]. Some of them share many similarities and show just a few dissimilarities in terms of intervallic distances like *ussak* and *nihavend*, see Figure 1, which increase the complexity of its automatic discrimination. Each makam is built by concatenating smaller units called tetrachords and pentachords, made of four and five tones respectively. Figure 1 shows the *hicaz*, *ussak* and *nihavend* makams as they are annotated by the Arel theory. This theory makes use of the Western staff and a larger variety of accidentals to cover the plethora of quarter notes. It should be mentioned that the makam scales do not have the same structure if they are extended beyond the octave, but may modulate to other tetra/pentachords.

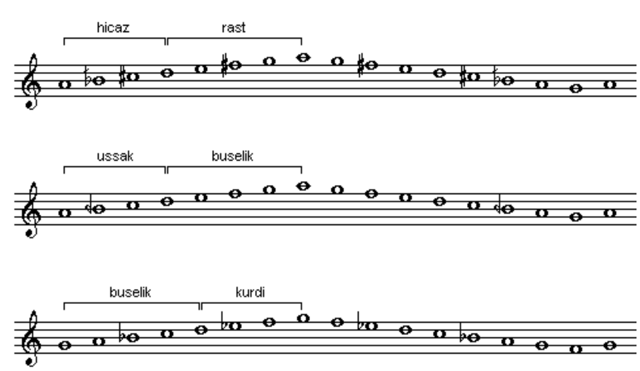

Figure 1. Hicaz, Ussak and Nihavend scales.

2.2. Music collection

For this work a library of 302 polyphonic music pieces was collected from the selected 9 makam families. Songs from various regions were selected (Turkey, Greece, Egypt and Iran). Nevertheless, the big majority comes from Turkey, where it is a common practice to annotate the makam in the title of the piece. That made easy to determine the ground truth information for each piece. All the songs used in our study were extracted from commercial CDs.

We use the AEU theory when needed, as it is claimed to be capable of representing the plethora of intervals in Turkish makam music and also being suitable to Arab makams [8]. We tried to keep an equal distribution among the makams and to include pieces from diverse morphological manifestations of that music, such as taksim (improvisational performance), religious chants, sirto, longa or semai, we address to [11] for further information on the musical forms of Ottoman music. The plurality of styles and origins had the intention to make a non-biased model to a specific region or morphological style. The beginning, the end and a part from the middle of the piece were extracted from all the analyzed pieces. In all cases the excerpts were 60 seconds long and the same experiments were held for all three excerpts. We report results only from the end of files as it yielded better accuracy levels.

3. System architecture

3.1. Descriptors

The feature space includes chroma (HPCP) vectors extracted from the audio signal. We used the approach proposed in [5] with a window size of 46.5 ms. For this study we set the interval resolution to 159 bins per octave. This resolution was chosen as a single octave is divided into 53 equally spaced intervals according the AEU theory and each one being a Holdrian comma. So a 1/3 comma resolution is considered to provide tuning robustness and good resolution [1]. Chroma features were computed and normalized on a frame basis and a global vector was obtained by averaging frame values within the considered excerpt. We considered the frequency range between 40 – 5000 Hz.

As mentioned in section 1, each makam can have different weights in the degrees of the scale, apart from the tonic itself. For instance in figure 1 we observe that *ussak* is constructed from a *ussak* tetrachord followed by a *buselik* pentachord, and the *nihavend* makam is constructed by a *buselik* pentachord followed by a *kurdi* tetrachord. In terms of intervallic distances they are almost identical, except for a comma value in the second degree of the *ussak* scale. But it should be expected that in the case of *ussak* there are more occurrences in the sub-dominant than in the dominant. To capture this feature, the peaks of each HPCP vector were detected and its ranking was computed in addition to the usual raw energy values.

3.2. Template-matching approach

This model is inspired on the one presented in [3] and it is summarized in Figure 2. In order to build makam templates, we average chroma (HPCP) vectors for all the excerpts of a given makam, as shown in Figure 2 (left). Before averaging, each chroma vector is shifted with respect to its tonic using the method proposed in [1]. In this method, the makam annotation is used to compute its corresponding theoretical makam template. The theoretical template is built with Gaussians centered in the intervals dictated by AEU. Then, the chroma vector is ring-shifted (159 bin per octave) and the tonic bin is found as the position providing the maximum cross-correlation with the theoretical template. We finally average all the shifted chroma vectors from the same makam. In theory each makam scale has a characteristic tonic note, but in practice this doesn't always happen; for example, the makam *rast* has as tonic the middle *D*, but due to modernization of this music we can have the *rast* scale transposed to other tonic notes depending the instrumentation or the singing voice. Also, old traditional recordings do not use 440Hz as tuning frequency. Thus, we cannot establish the tonic note from the makam type annotation and the previous procedure is needed.

For estimating the makam of a new excerpt (Figure 2, right) we compute the distance between its chroma vector (HPCP) and each makam template, and the smaller one defines the estimated makam. This distance measure is computed for different ring-shifted versions of the HPCP vector. After performing some initial experiments, we compared two different methods.

The first one shifted the HPCP vector with respect to the tonic position. This tonic position was computed as explained above, and then used available ground truth information. The second method shifted the HPCP vector with respect to its three highest peaks, so that no ground truth information is needed. By shifting the vector towards all 159 bins, the method gave extremely low accuracy levels. For a music excerpt where the makam information is not known, we then assumed that the tonic should always have a high value of occurrences. After performing some empirical experiments, we decided to select the three highest peaks as potential tonic candidates. Both methods used Euclidean distance as the distance measure.

3.3. Class-modelling approach

The same features were used to train a classifier to discriminate between the 9 different makam categories. The experiments were done in Weka[1]. We performed two different experiments. In the first one the feature set was formed by the average HPCP vector, while in the second one we shifted it with respect to its tonic, as explained in Section 3.2. Although various classification techniques were tested we finally report only the results obtained with Support Vector Machines (SVM), as they yielded the best performance.

At first we trained the classifier using the same feature space as in the template-matching model to have a direct comparison of the two models, thus by providing only the energy values from the 159 bins vector. It is usual in this research trend to apply numerical transforms to the feature data to achieve a better training of the classifier. We applied the ranking transform of the HPCP vector, which ranks the bins according to their number of occurrences, and then the log transform, which consists on computing the log of each value. The first showed a very slight boost in the overall results while the log transform added an overall 2.6% improvement. As the feature space was more than the half of the number of instances, we applied an attribute selection procedure, using the *CFS* function, which shortened the number of attributes to 46. This also provided a final boost in the classification accuracy. The results are summarized in section 4. A 10 fold cross-validation was used in order to ensure enough generalization power to the model and do not get stuck to an over-fitted model.

4. Results

4.1. Case study

We illustrate with some examples the chroma features extracted from the selected musical excerpts. Figure 4 shows an averaged template constructed from all *nihavend* excerpts, built with the method described in Section 2, and a HPCP vector of one single *rast* excerpt. The dashed lines indicate the distances from the tonic of each degree in the makam, as dictated by AEU. We observe that they agree with the peaks of the computed templates.

4.2. Automatic classification

Standard evaluation measures (Precision, Recall and F-measure) were computed for both approaches and all makam classes. They are summarized in a comparative chart in table 1.

Table 1. Overall Precision, Recall and F-measure for the different approaches.

	Template-Matching			Class Modeling		
	P	R	F	P	R	F
HPCP	0.50	0.45	0.45	0.298	0.31	0.29
Tonic-shifted	0.69	0.69	0.69	0.727	0.73	0.73

[1] http://www.cs.waikato.ac.nz/ml/weka/

We observe that the best overall performance, F-measure equal to 0.73, is obtained with the class modeling method using tonic-shifted HPCP vectors. These results are similar to the ones obtained in previous studies [3] reported an F-measure of 0.68 for the same makam classes and similar tonic detection procedure). We also confirm that the use of the tonic detection method improves to a great extent the results for both approaches.

To have a direct comparison with related work in Western music we can consider the results obtained in MIREX 2005 [12], an international competition for comparing key detection algorithms. The best working algorithm correctly estimated the key in around 86% of the tested pieces (24 possible keys). This accuracy is obtained for MIDI synthesized recordings of classical music. Nevertheless, an analogous problem to makam estimation for Western music would be scale estimation (or mode). We can argue that state-of-the-art accuracy for mode estimation is around 86% for real recordings from different musical genres (reported in our previous work [5]). Considering that we are dealing here with 9 different makam types instead of 2 possible modes (major and minor), we can consider that the accuracy levels obtained here are comparable to previous work in Western material.

Regarding the results for the different makams, we observed that some makams were confused between each other whereas some of them were easily discernible. This is expected, e.g., as *hicaz* and *rast* use very distinct intervals, so by measuring the distance from each template vector is enough for distinguish them.

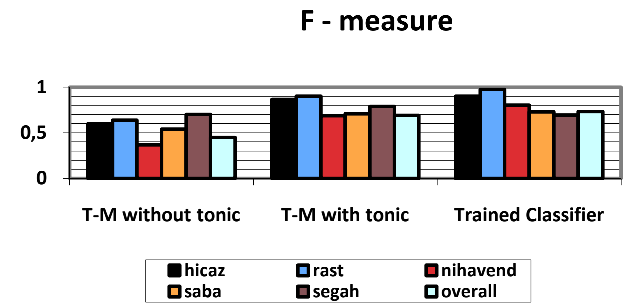

Figure 2. F-**measure for different makams and approaches.**

On the other hand, makams such as *nihavend*, *ussak* and *huseyni* share similarities in terms of intervals, which yield frequent confusions. The error analysis showed that there is large confusion in the template-matching approach between *huseyni* and *ussak* makams and low performance for the *nihavend* scale, as shown in Figure 3. We observe that some of these problems were overcome in the trained model, such as a good classification of *nihavend,* which had a poor performance in the template matching model.

We also performed an experiment to test the model for different time regions of the songs. The presence of the tones of a makam scale is expected to be more likely in the beginning and in the end of a music performance. The compositor/performer will start by presenting the genre of the piece, and the same is expected for closing the performance. Table 2 provides the results for the Template-Matching (using HPCP) for the different excerpts.

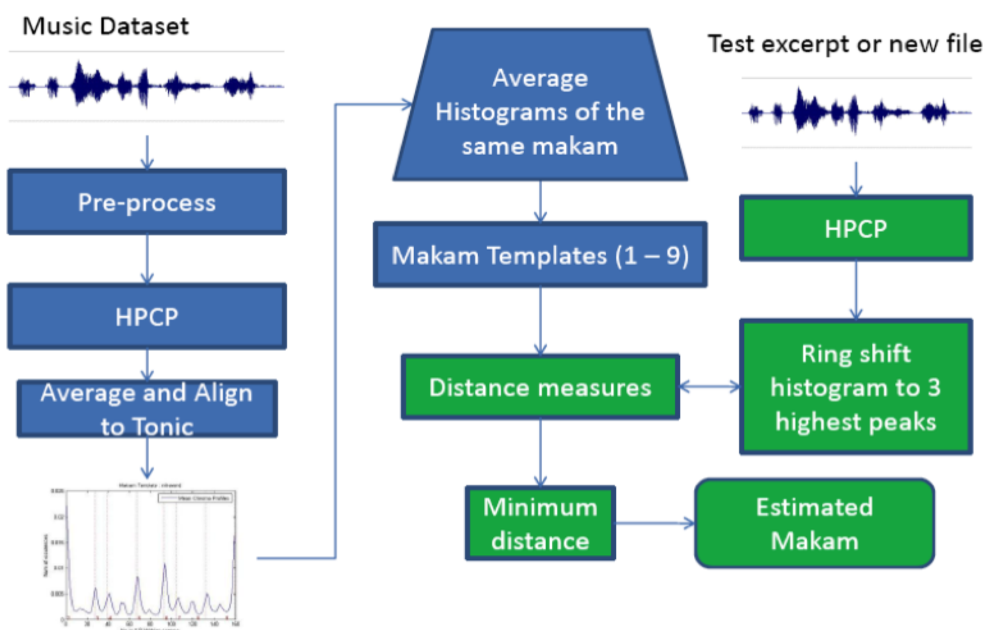

Figure 3. General schema of the template-matching model.

Similar behavior was found for the other approaches. The above assumption was verified as the best performance was found in the beginning and the end of the pieces. Because of these results, the reported results correspond to the analysis of the last portion of the file.

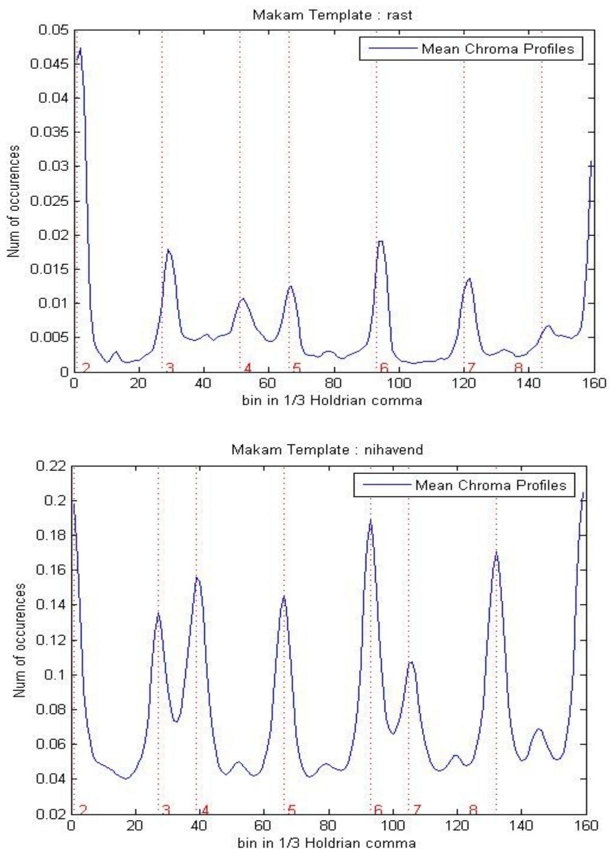

Figure 4. Rast (top) and Nihavend (bottom) chroma vectors.

Table 3 provides the confusion matrix for the best performing method (classifier with tonic-shifted HPCP). The highest confusion still remains for the same three classes as mentioned above, which verifies that the errors are 'musically' meaningful. The highest confusions in certain classes are marked in bold.

Table 2. Evaluation results for the template-matching model using HPCP and the different excerpts of the pieces.

	Precision	**Recall**	**F-**
Begin	0.50	0.40	0.37
Middle	0.39	0.39	0.36
End	**0.50**	**0.45**	**0.45**

5. Conclusions

We have presented a method for classifying 9 makam types from polyphonic music signals by means of chroma features. We obtained a F-measure of 0.73 evaluated on a music collection of 302 pieces. The comparison of template-matching and class-modeling methods revealed some differences. On one hand, the trained classifier yielded better figures of precision and recall than the template-matching method, but the template-matching model can bring better results in the absence of a tonic detection procedure. As part of our future research we include: i) the automatic computation of the tonic information for each new excerpt where ground truth knowledge will not be required; ii) detection of the note sequence (*seyir*), an important makam discriminating feature; and iii) exploring the combination of template-matching and class-modeling, as their error diversity could yield improvement is the overall performance of a hybrid system.

Table 3. Confusion matrix for the best-performing method (class modelling with tonic-shifted HPCP).

makam	**hic**	**hus**	**ras**	**nih**	**kurd**	**sab**	**seg**	**huz**	**uss**
hic	45	1	0	0	3	0	0	0	1
hus	1	**13**	0	2	2	2	0	0	**8**
ras	0	0	38	1	0	0	0	0	0
nih	0	1	0	31	0	1	0	0	3
kurd	1	2	0	0	15	2	0	0	2
sab	0	2	0	4	1	27	0	0	4
seg	1	0	0	0	0	0	**16**	**8**	0
huz	1	0	0	0	0	0	**5**	**28**	0
uss	2	**8**	0	3	4	2	0	0	**11**

6. References

[1] Bozkurt B., "An Automatic Pitch Analysis Method for Turkish Maqam Music", Journal of New Music Research, Vol. 37, issue 1, 1-13, 2008.

[2] Gedik Ali C., Bozkurt B., "Pitch-frequency histogram-based music information retrieval for Turkish music", Signal Processing Volume 90, Issue 4, 1049-1063, 2010.

[3] Gedik Ali C., Bozkurt B., "Automatic Classification of Turkish Traditional Art Music Recordings by Arel Theory", Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08), Greece, 2008.

[4] Tzanetakis G., Ajay Kapur, W. Andrew Schloss, Matthew Wright, "Computational Ethnomusicology", Journal of interdisciplinary Music Studies, Vol. 1, issue 2, fall 2002

[5] Gómez E., “Tonal Description of Music Audio Signals”, PhD Thesis, Department of Information and Communication Technologies, Univeristat Pompeu Fabra, 2006.

[6] Ozgur I., “An Algorithm for Audio Key Finding”, Music Information Retrieval Evaluation Exchange (MIREX), 2005.

[7] Gómez E., Herrera P., “Estimating the tonality of polyphonic audio files cognitive versus machine learning modeling strategies”, International Society for Music Information Retrieval Conference (ISMIR), 2004

[8] Ozan Y., “79-tone Tuning & Theory for Turkish Maqam Music as a Solution to the Non-Conformance between Current Model and Practice”, Istanbul Technical University: Institute of Social Sciences, 2007.

[9] Alpkoçak, A.and Gedik,A.C. 2006, “Classification of Turkish Songs according to makams by using n grams”, Proceedings of the 15. Turkish Symposium on Artificial Intelligence and Neural Networks,TAINN, Mugla

[10] H. S. Powers, et al. "Mode." Grove Music Online. Oxford Music Online. 21 Apr. 2010, http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg5

[11] C. Tanrikorur, “The Ottoman Music”, 10 Jun 2010, http://www.turkmusikisi.com/osmanli_musikisi/the_ottoman_music.ht

[12] Audio Key Finding Results, MIREX 2005 http://www.music-ir.org/mirex/wiki/2005:Audio_Key_Finding_Results