A hierarchical approach to makam classification of Turkish makam music, using symbolic data

Erdem Ünal[1], Barış Bozkurt[2], M. Kemal Karaosmanoğlu[3]

[1]*TÜBİTAK-BİLGEM, BilişimTeknolojileri Enstitüsü, Kocaeli, Türkiye*

[2]*Electrical and Electronics Engineering Dep., BahçeşehirUniversity, İstanbul, Türkiye*

[3]*Yıldız Technical University, Art and Design Faculty, İstanbul, Türkiye*

[1]erdem.unal@tubitak.gov.tr, [2]baris.bozkurt@bahcesehir.edu.tr, [3]kkara@yildiz.edu.tr

A hierarchical approach to makam classification in Turkish makam music, using symbolic data

A method for hierarchical classification of makams from symbolic data is presented. A makam generally implies a miscellany of rules for melodic composition using a given scale. Therefore, makam detection is to some level similar to the key detection problem. The proposed algorithm classifies makams by applying music theoretical knowledge and statistical evidence in a hierarchical manner. The makams using similar scales are first grouped together, and then identified in detail later. The first level of the hierarchical decision is based on statistical information provided by the n-gram likelihood of the symbolic sequences. A cross entropy based metric, perplexity, is used to calculate similarity between makam models and the input music piece. Later, using statistical features related to the content of the piece, such as the tonic note, the average pitch level for local excerpts and the overall pitch progression, a more detailed identification of the makam is achieved. Different length n-grams and representation paradigms are used including, the Arel theory, the 12 tone equal tempered representation, and interval contour. Results show that, the hierarchical approach is better, compared to a straightforward n-gram classification, for the makams that have similar pitch space, such as *Hüseyni - Muhayyer* and *Rast - Mahur*. Using the proposed methodology, the system's recall rate increases from 88.7% to 90.9% where there exists still some confusion between the makams *Uşşak* and *Beyati*.

Keywords: music information retrieval, computational musicology, makam music, makam detection, hierarchical classification of makams, n-grams

1. Introduction

Music Information Retrieval is gaining worldwide attention since music data is one of the most popular data available on the internet. Some researchers focus on solving problems that have commercial potential such as indexing, classification or retrieval using well known signal processing and mathematical techniques for easier access to digitally available music collections. However, most of the studies address music based

on the 12 tone equal tempered tuning system and very few studies address music traditions, such as Makam Music that is not based on this tuning system.

Even though it is not widely studied, the makam concept originates from a very large geographical region from Balkans to Kazakhstan, Iran, and North Africa. A makam generally implies a miscellany of rules for melodic composition using a given scale that exhibits diverse characteristics from one geographic region to another. While there is some similarity to the concept of key, there is an important distinction that specific rules for melodic progression are applied such as some degrees of the scale being emphasized in an orderly manner.

Recently, some researchers started focusing on makam music, using tools and techniques coming from signal processing, statistics, and pattern recognition. One of the basic motivations is to better understand makam music, its concepts, origins and its relation to a specific culture. In addition, there exists a huge amount of music material available on the internet. However, it is not organized or categorized, thus very hard to access. In order to make available music collections more accessible, the materials need to be analyzed, categorized and organized with respect to their musical content. This can only be ensured if the automatic analysis of the available materials is reliably achieved. One of the most basic problems to start with, for the analysis of makam music, is the automatic classification of makams. Makam classification further finds applications in other analysis problems such as form analysis and tuning analysis. For form analysis, makam transitions provide important clues about section boundaries. For tuning analysis, makam information is useful since some microtonal intervals are makam specific.

It should be emphasized here that there is an important lack of perceptual studies for makam music. We don't have access to any study which reports how well human

beings can distinguish different makams and which cues they use. For the moment, our study relies on makam tags provided on the commercially published scores used by many master musicians. It is among our future goals to conduct such perceptual studies.

Existing studies related to makam analysis and classification focused on audio data [1], [9], [13], [16] and propose signal processing algorithms with limited efficiency. A few makam music studies considered using symbolic data, scores, by applying n-gram analysis and pitch class histogram comparisons [2], [22]. One of the main obstacles for this type of research has been the lack of large repositories of machine readable music scores. For example [2] has important shortcomings in terms of the size and the representation of data (12-TET representation is used). Recently, we made available a collection of makam music scores, named SymbTr [17], which might be the most comprehensive machine readable collection available of Traditional Makam Music in Turkey (TMMT). SymbTr is a compilation from reliable sources of 1,700 TMMT works into text, PDF and MIDI formats featuring distinct examples in 155 diverse makams, 100 usuls and 48 forms. In [22], [23], we proposed a framework for makam classification, where, an n-gram based modeling technique and a perplexity based similarity metric is used. Our experiments were conducted on a subset of the data presented in [17], using 13 makams and 857 songs and we were able to achieve around 88% makam classification accuracy(recall) on average.

While exploiting the potential of symbolic sequential modeling techniques, namely the n-grams, our approach uses musical knowledge and statistics, derived from data and theoretical resources. In this study, incorporating both the statistical information and theoretical knowledge using a hierarchical strategy, a complete method for robust makam classification is proposed. Testing various different experimental set-ups, such as different length n-grams, different representation schemes, and different

theoretical makam related features, the system design offers to analyze, and classify the makams from note level symbolic data.

The main contribution of our current work is the development of a hierarchical framework and use of new features for makams using similar scales. Complexity analysis in [22], [23] showed that, makams using similar scales have high levels of confusion, while more distinct makams are easier to classify. In order to solve the problem for challenging makam sets, a more sophisticated classification paradigm needs to be employed. This work is an extension to [22] and [23], where a hierarchical classification setup is implemented using new musical features in order to increase the makam detection accuracy for challenging makam sets.

The leave-one-out setup is adopted for validation experiments. In the leave-one-out setup, for each test instance, one music piece is selected and left out from the test corpus. The remaining pieces are used for modeling. The left out piece is then rated against previously built makam models. Given a similarity metric, perplexity in the first stage, and total likelihood in the second stage, the decision is made for the left out piece as a single makam class.

In the first stage, makams using similar scales are modeled collectively (*Beyati - Uşşak, Hüseyni - Muhayyer*, *Rast - Mahur*) as a single class, and the n-gram framework is applied directly. After the initial decision is made, knowledge based statistical rules related to melodic progression are used for the final decision. A Gaussian Mixture Model and the Maximum Likelihood Rule approach with a modified Machalanobis distance measurement is used to make the final decision. GMM parameters are trained from a training set, for each of the challenging makam sets, using numerical features for the initial melodic movement and the overall melodic progression. While the initial melodic movement describes how high (in pitch) the music pieces tend to start in the
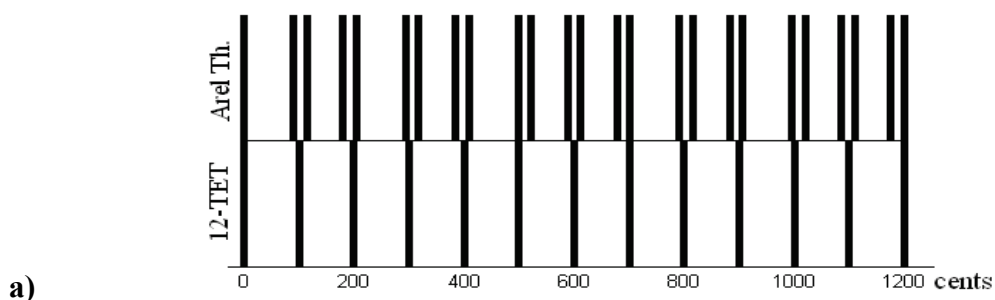
beginning, the overall melodic progression shows how much the melody tends to ascend or descend.

The organization of the manuscript is as follows: First, we introduce the concept of makam, its intervalic structure and the collection of scores used in the study. In the next section we explain the n-gram framework and give details on the experimental setup. Then we present the main contributions of this article, which are the hierarchical methodology and the new musical feature set. Finally we give and discuss the results and present our future work plan.

2. Background on makam music

One of most distinctive characteristics of makam music is the use of microtonal intervals. In makam music, the octave is divided into more than 12 unequally spaced intervals, thus it does not use equal temperament. This requires a representation scheme, namely an accidental system, that can handle microtonal pitch intervals.

The tuning system of TMMT is a topic under constant discussion by musicologists and music theorists. The most commonly used system, named the Arel - Ezgi (AE) notation (Arel theory [4]), suggests 24 main notes in an octave. Figure 1 shows the accidentals used and the interval sizes on a full octave compared to those of the (12-TET) equally tempered system.
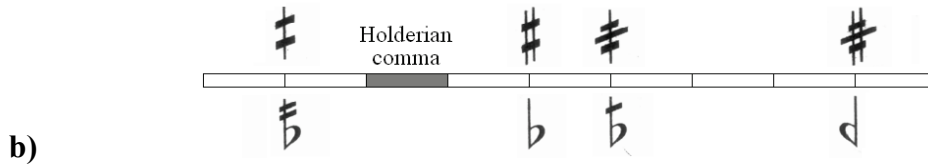

a)

b)

Figure 1. a) 24 pitches of the Arel theory with respect to the 12-TET tones (shown in one octave), b) the accidentals (showing interval sizes within a whole tone) used to represent these pitches on the staff

Since Arel theory is based on a Pythagorean tuning, almost all of the pitches are closely located to a 12-TET tone. It is a well-known problem of the Arel theory that it cannot represent the practice, especially the pitches close to a quarter tone which are common to many makams [8]. Dividing a full tone interval into 9 Holdrian commas (obtained by equal division of an octave into 53 partitions, namely 53-TET), the Arel - Ezgi theoretical notation defines the sharps and flats for 1, 4 (equal tempered sharp), 5 (equal tempered flat) and 8 comma length intervals. The names of the pitches, their representation on the staff, corresponding fingering on ney and sounds (as blown from a ney) can be accessed from http://www.ozanyarman.com/files/neysounds.htm. In a specific makam music piece, the usage of these microtonal accidentals depends on the musical key, more specifically the makam of the piece, as shown in Figure 2.

Figure 2. Scales used for makams *Hüseyni - Muhayyer* (left), and makams *Beyati - Uşşak* (right)

As indicated in Figure 2, in Arel theory, scales are divided into chords. The tonic, *karar*, is the first note of the first tetrachord (A in Figure 2) and the preceding note (G) is the leading tone, *yeden*. The note A (Dügah), is the *karar* of many makams such as *Hüseyni, Muhayyer, Uşşak* and *Beyati*. Usually, an open string in string

instruments is tuned to A or G (Rast, which is another common *karar* note) to be able to provide the pedal tone when needed. In almost all cases, the *karar* is the final note of the piece or improvisation. The *karar* note, i.e. the finalis, appears to be one of the features that discriminate makams into subgroups. The tetrachord - pentachord junction note (E(Hüseyni) for *Hüseyni* and *Muhayyer* and D(Neva) for *Uşşak* and *Beyati*) is very often named as *güçlü*, considered to have the function of a dominant by some authors such as Arel (this is considered to be a "westernized view" by some musicologists [6]). *Güçlü* appears in pitch histograms with high occurrences and is potentially useful for discriminating again subgroups such as *Hüseyni - Muhayyer* and *Uşşak-Beyati*.

Arel theory presents the makam concept very close to the key concept and hence tuning and scale is very central. There is an important deficiency (since very little information is provided) in presenting the *seyir*, i.e. the melodic progression which is considered to be very crucial for describing makams by many scholars [6]. These deficiencies are to some level overcome (with the cost of reducing accessibility of the text) in publications such as [18] (one of the most commonly used resources for education today in Turkey) which is based on Arel theory but includes many detailed descriptions about melodic progression rules for each makam. Arel's formulation with central focus to the scale is considered as a "westernized view" by some authors [6] pointing out that the historical texts of makam music explain the makam concept by descriptions of melodic progression rules and specific functions of the notes instead of emphasis on the scale. Melodic progression rules are explained as road maps in these texts and are often learned by the musicians by studying the repertoire than reading-memorizing these texts. The cognitive processes involved in learning and applying *seyir* is unknown to us today, it is an open topic of research. While in today's music circles three main categories are use to describe the melodic progression (ascending,

descending and mid-range progression), an efficient way of presenting melodic progression rules is also an open topic attracting international researchers [12]. Typically, the progression of the makam is defined for the *zemin (basis)* part where the makam of the piece is "presented" at the beginning and does not necessarily hold for the *meyan* part ("development") where transitions to other makams occur. An ascending progression is considered to start by emphasizing the tonic and then develop towards higher pitches. Descending progression starts by emphasizing the higher tonic and descends towards the tonic. A mid-range progression starts by emphasizing the *güçlü* and descends towards the tonic. An example of mid-range progression is shown in Figure 3 presenting the melograph of *Hüseyni* taksim performed by Fahrettin Çimenli. The solid lines are drawn by the authors to indicate the overall progression.

Figure 3. Melograph of *Hüseyni* taksim by Fahrettin Çimenli. Pitches are indicated as relative to the tonic frequency.

In Figure 3, the progression starts by emphasizing the *güçlü* around 700 cents (the fifth). Then the overall melodic direction is towards the tonic (0 cents) while in the *meyan* part (60-90 secs.), the higher (pitch) part of the scale is used. This example can be considered as a typical mid-range progression. The interested readers are referred to two well-written texts, [12] and [19] which discuss *seyir* in detail.

Melodic progression plays two important roles in our study. The first is the assumption that such rules result in certain distributions of pitches and pitch sequences and hence statistical information about pitches and intervals used in melodies can be used for makam detection. The second assumption is that some information about melodic progression can be captured using techniques such as n-grams which use the statistics of ordered sequences of specific length. Since n-grams capture rather local melodic movements, in our hierarchical approach, we also propose the use of a new feature to describe the overall progression for improving automatic classification performance. The new feature used is discussed in Section 6.

The pitch class histograms and n-grams are used efficiently in mode or key finding algorithms, melodic analysis and cover song identification tasks for Western music studies [11], [21], [24]. While makam detection can be considered as similar to key finding or mode finding problem in Western Music, there appear to be important differences between the concepts of makam, key, and mode [3]. However, assuming that, pitch class based features are also distinctive for makam classification, related algorithms are inherited from the work on Western music. One of the basic goals of this study is to examine the efficiency of n-gram based automatic makam recognition for similar makams and improve the efficiency by hierarchical modeling and use of new features for melodic progression.

3. The data set

Although widely criticized by musicians, The Arel - Ezgi Theory Notation is the most commonly used notation system today. While a large amount of scanned images of scores are digitally available in this format on the Internet, machine readable data has not been available. In order to fill this gap, we recently published and made publicly available a collection of music scores in machine readable format of TMMT containing
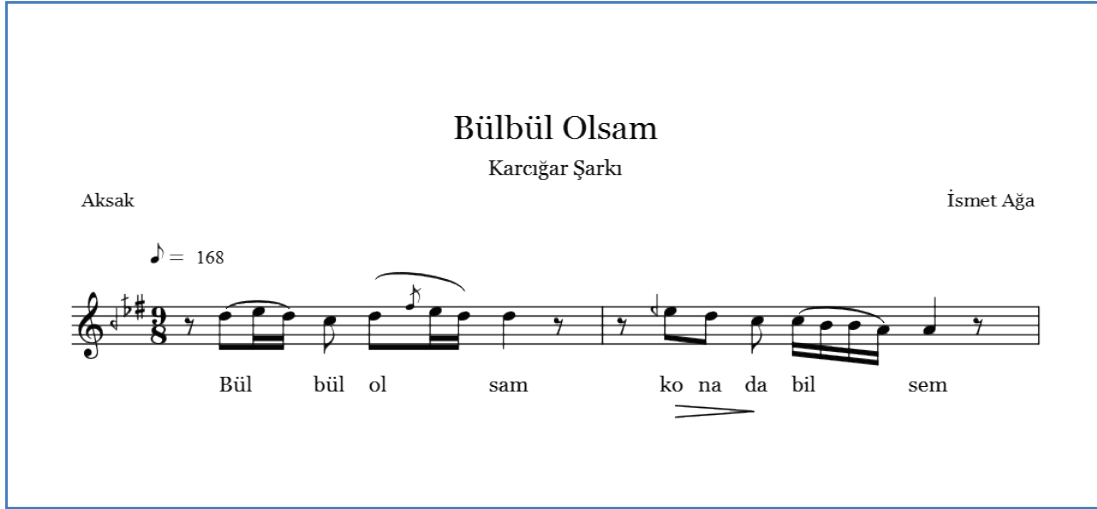
1700 pieces in 155 makams, called the SymbTr [17].

Figure 4. Score notation of a music piece in makam Karcığar

Code	Note 53	Comma 53	Note AE	Comma AE	Num.	Denom.	ms	LNS	VelOn	Syll.
9			-1	-1	1	8	179	100		
9	Re5	D5	327	327	1	8	179	99	96	Bül
9	Mi5b4	E5b4	332	332	1	16	89	99	96	
9	Re5	D5	327	327	1	16	89	95	96	
9	Do5	C5	318	318	1	8	179	95	96	bül
9	Re5	D5	327	327	1	8	179	99	96	ol
8	Fa5#4	F5#4	344	344	1	8	22	99	96	
9	Mi5b4	E5b4	332	332	1	16	89	99	96	
9	Re5	D5	327	327	1	16	89	95	96	
9	Re5	D5	327	327	1	2	715	47	96	sam
9	Mi5b1	E5b1	335	335	1	8	179	95	96	ko
9	Re5	D5	327	327	1	8	179	95	96	na
9	Do5	C5	318	318	1	8	179	95	96	da
...										

Table 1. *SymbTr* representation of the score in Figure 4

The main source of data in SymbTr is TRT (Turkish Radio and Television) and other trustworthy archives, where almost all of the scores have been written using AE notation. An example with score notation and the SymbTr format is given in Figure 4,

and Table 1 respectively. While a detailed explanation of the SymbTr format can be found in [17], we present a short summary here.

"Code" signifies a normal note (#9) or ornamentation. The most commonly used ornamentation codes are as follows: #7 for tremolos, #8 for acciaccatura, #12 for trills, and #23 for mordent.

"Note53" and "Comma53" include the pitch information in a 53-TET resolution and "NoteAE" and "CommaAE" include the pitches specified as in the Arel theory. The latter is obtained by direct conversion of staff notation into text where the former is a corrected version of the pitch by master musicians. Here, "Note" stands for "name of the pitch" and "comma" stands for the interval of the pitch with respect to C_1. "Num.", "Denom." and "ms" columns specify the duration. "Syll." contains the lyrics. "VelOn" is used to specify velocity dynamics. "LNS" (Legato / Normal / Staccato) indicates how tied or detached the notes are. This text representation facilitates viewing the content of the data without need of any specific program but just a text editor and easy data access (compared to for example microtonal MIDI files).

3.1 The selection

Since our basic goal is to validate the usability of distinctive features defined to analyze and classify makams, we selected a subset of this collection with respect to three criteria, namely: commonness, similarity and having sufficient number of examples. For a classification study, it is beneficial to include similar classes and study the effects of such similarities in the classification performance. We have included *Uşşak - Beyati*, *Muhayyer - Hüseyni* makam couples in our set, which are stated to be differing only in melodic progression [18]. These couples share the same set of pitches, the same tonic and *güçlü*/dominant (which is considered to be the boundary of tetrachord - pentachord division of the octave) as shown in Figure 2. Moreover, the two scales used in these

four makams have only one pitch different (F#) which is often replaced with F on descending lines resulting in a very similar group of 4 distinct makams. This is reflected in observations in previous classification work on audio data; *Hüseyni* is also confused with *Uşşak* [13]. Therefore, the set includes challenging examples of classes. The subset corpus used in this study can be seen in Table 2.

Makam name	Total # of Songs	Total # of Notes
Beyati	42	17347
Hicaz	117	39301
Hicazkar	49	14775
Hüseyni	71	23787
Hüzzam	65	23581
Kürdilihicazkar	51	18332
Mahur	54	18039
Muhayyer	50	16774
Nihavent	86	28724
Rast	88	29103
Saba	45	16486
Segah	74	21744
Uşşak	85	26379
TOTAL	877	294372

Table 2. Selected makams and note statistics

4. N-gram methodology and perplexity as a similarity metric

N-gram based statistical modeling and classification is widely used in computational linguistics, computational biology, as well as in Music Information Retrieval [10], [11] and [21]. Given a sequence A with length N, an n-gram model of A is constructed by counting the occurrences of n length subsequences inside A. The model helps predicting the kth value of this sequence, based on k-1 statistics.

Traditional makam music scores include only the main melody and no harmony is involved. Therefore, the main feature that defines the makam is the main melody. In addition, emphasized notes play an important role, hence the number of occurrence is potentially useful as a feature for makam detection (as shown in [13] on audio files).

We can observe this phenomenon in pitch class distributions easily: for makam *Hüseyni*, the first emphasized note and one of the most frequently played notes is hüseyni (E4); for makam *Muhayyer*, the first emphasized note and one of the most frequently played notes is muhayyer (A5); for makam *Neva*, the first emphasized note and one of the most frequently played notes is neva (D4), etc. N-grams provide statistics of ordered sequences (short melodic lines) and hence can be considered as extension of pitch class histograms which have been successfully used in similar problems as in key detection. This is the main reason of our choice of n-grams for makam detection.

In this work, n-gram models are extracted from melodic sequences that belong to the same makams. By doing that, the model will collect information on the frequency of occurrence of common melodic phrases, motifs and note transitions that belong to the same makam. Later, this statistical information can be used to evaluate the similarity of an input sequence for classification purposes.

When performing classification using n-grams, it is desirable that the occurrence of each possible n-length subsequence is assigned a non-zero probabilistic value. Unless this is taken care of, the 'zero frequency problem' will halt meaningful classification when confronted with n-length subsequences that are not seen before. In natural language processing, different smoothing techniques are introduced [15] to solve this problem. Written-Bell smoothing is used in this work, provided by the SRILM toolkit, where this and all other n-gram related computation and experiments were performed [20].

4.1 Perplexity

Perplexity is a form of likelihood estimation that is widely used in NLP (Natural Language Processing). It estimates, how likely a sample or a sample set is generated by a statistical model, thus can be used to evaluate probabilistic models against each other

for classification purposes. Given a proposed probability model q (in our case: a makam model), evaluating q by estimating how well it predicts a separate test sequence or set $x_1, x_2, ...,x_N$ (in our case: a microtonal note sequence) drawn from p, can be performed by using the perplexity of the model q, defined by:

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

(1)

where x_i is drawn from p, and evaluated by q. The exponent in the formula is usually calculated by the cross entropy formula shown in 2.

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

(2)

In our experimental setup, a microtonal note sequence of a music piece is evaluated against makam models. Using perplexity, a similarity metric is calculated that shows how close this music piece is to the makam models available in our repertoire.
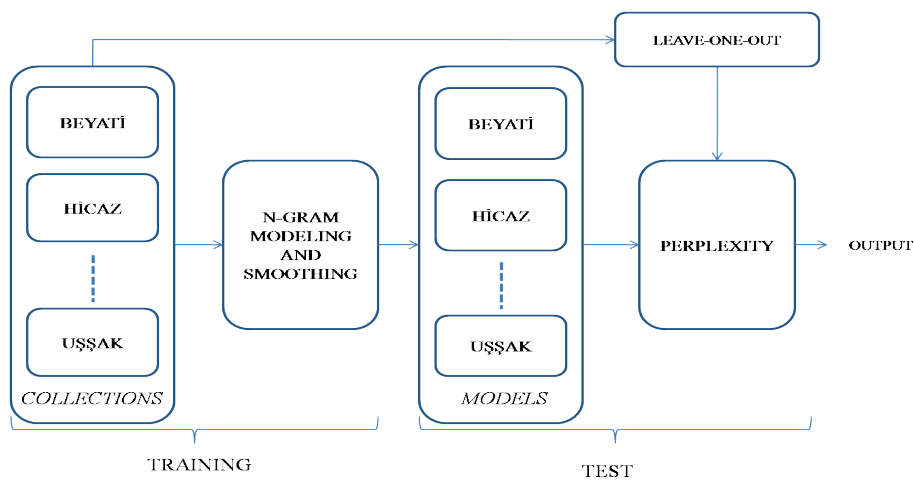
5. N-gram Experiments

Figure 5.The leave-one-out experimental setup: A single piece is left out from the collection to be used as an input for a test instance. The rest of the pieces are used to

build n-gram based makam models.

The experimental setup for the base line system is organized on top of the "leave-one-out" (LOO) strategy. First, a random music piece is selected and pulled out from the collection to be used as an input to the system. The remaining pieces are then used for modeling. The pieces belonging to the same makam are first merged into a single collection. Later, the n-gram statistics are calculated for each of the makam collection. After the appropriate smoothing is performed, the makam models become ready for classification and evaluation.

After the modeling is completed, the left out piece is fed into the system as an input and evaluated against each available makam model in the collection using perplexity. The makam with the lowest perplexity score is returned as the output/decision of the system. The basic experimental setup can be seen in Figure 5.

5.1 Preliminary results

The results of the first set of experiments are going to be explained in two subsections. First, we would like to use a fixed symbolic representation, which is the microtonal AE (Arel-Ezgi notation), and see the effect of increasing of n-grams in the classification performance. And then, selecting a fixed length of n-gram, we would like to see the effect of different representation techniques, namely the microtonal AE, 12-TET notation, and interval contour, over the classification performance.

5.1.1 The effect of increasing n-gram length

The matching performance of the entire system, which is the accuracy (Recall), will be given as a proportion of number of successful classifications over the total test trials. For Total Average, each sample has equal weight, while Weighted Average takes into

account the number of samples in each makam collection as weights. Following in Table 3, output of the system is shown as a confusion matrix.

2-gram RESULTS:														
beyati	hicaz	hczkr	hsyni	Huzzm	krdhz	mahur	muhyr	nhvnt	rast	saba	segah	uşşak		Recall
20	0	0	5	0	0	0	1	0	1	0	0	15	**beyati**	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	4	**rast**	92,2
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	**uşşak**	68,2
55.6	99,2	100	70,3	98,4	96,2	100	60,7	100	91	97,8	94,7	65,9	Prc.	

TOTAL AVERAGE: 87,9
TOTAL WEIGHTED ACC is: 86.7

Table 3. 2-gram results, highlighting the confusion of the makams using the same scale

	n=1	n=2	n=3	Best-N
Beyati	54,8	47,6	47,6	1
Hicaz	100	100	100	1,2,3
Hicazkar	98	98	98	1,2,3
Hüseyni	53,5	63,4	66,2	3
Hüzzam	98.4	95,4	96,9	1
Kürdilihicakar	96,9	100	100	2,3
Mahur	94,4	94,4	94,4	1,2,3
Muhayyer	88	74	64	1
Nihavent	100	98,8	100	1,3
Rast	71,6	92	92	2,3
Saba	97,8	97,8	97,8	1,2,3
Segah	94,6	97.3	95,9	2
Uşşak	75,3	68,2	65,9	1
Tot-Avg	86.5	86,7	86,1	2
W-Avg	86.9	87.9	87.5	2

Table 4**.**The effect of increasing n-length.

Excluding the critical makams that has the potential of high levels of confusion, the system's general performance is promising. Increasing the length of the n-grams also

helped improve the recall performance, from 86.5 to 87.9 (n=1 to n=2). Table 4 shows the effect of increasing the n-gram order for each makam, also indicating the optimum n-length, as increasing the n-length beyond n=3 does not improve our results.

Considering the makam couples *Uşşak - Beyati* and *Hüseyni-Muhayyer*, where we expect the highest confusion, the results are not surprising that the recall rate is considerably low. For *Rast* and *Hüseyni*, the increasing level of n-gram did actually help to decrease the confusion, where recall is increased from 71.6 to 92 (from n=1 to n=3) and 53.5 to 66.2 (from n=1 to n=3), respectively. Another important observation might be that, the increasing of n-gram length did have a negative effect on the recall performance for *Uşşak*, *Beyati* and *Muhayyer*. The result might be an indication that increasing the n-gram order does not help for these makams because instead of carrying distinctive information, the short-time melodic phrases are common for these makams and do not help in classification.

5.1.2 Different symbolic representations

The main representation used in this study is the NoteAE column of SymbTr format. In order to compare the effect of microtonal representation, we compared the classification results of our system with an experiment using the 12-TET representation. In [2] and [14], Gedik claimed that a representation simplified (rounded) to 12-TET still might be useful in makam detection and would be easier to use.

Moreover, we would like to expand our study on symbolic data into practice on audio files. That's why we also re-ran our experiments using the delta/interval contour representation, since microtonal notation is hardly extracted from pure audio while the contour information might be easier to achieve. Thus, we performed experiments using 3 different representation techniques: a) The Arel theory, b) 12-TET and finally c) Delta in Holdrian commas.

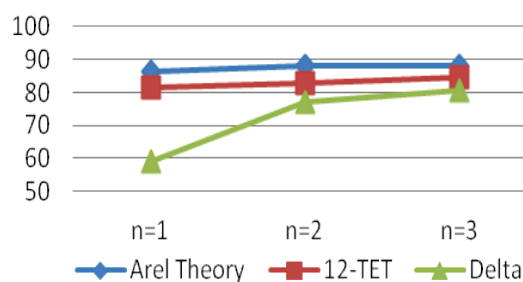| Recall	n=1	n=2	n=3
Arel Theory	86.9	87.9	87.5
12-TET	81.7	82.8	84.5
Delta (in H. commas)	58.9	77.3	80.6

Table 5. Overall performance comparison for varying n-gram length and different representations

As seen from Table 5, the results for the Arel representation outperforms the 12-TET representation by 3.7%, suggesting that, the microtonal information actually has a positive influence in makam classification. One can also see that, the interval contour representation (Delta) can only become useful with increasing length of *n*, however, losing the exact note level information decreases the classification accuracy. Note that, in a real world classification scenario, the system can only use the interval contour information more reliably rather than exact real notes, since exact note level transcription is a challenging task. However, this is a concern for later studies focusing on audio data.

6. Using additional features for makam classification

For further improvement, especially in discriminating similar-scale makams, the use of additional features is considered in a recent study [7]. The following list of features is observed to be useful in describing or teaching in instructional (theory and practice)

books:

- Scale, intervals and intonation of specific notes in the scale

- Melodic range

- Overall melodic progression, *seyir*

- Typical phrases, typical transitions or flavors[1]

- Tetrachords, pentachords constituting the scale and notes defined as tonic, dominant, leading tone.

In [7], quantitative features are defined for the features listed above. Then each quantitative feature is studied by observing their distributions computed from the makam collections. Following the observations (which are not included here due to space limitations), quantitative features that are potentially discriminative are tested within our makam detection algorithm. Here, we discuss the features that lead to improvement in the algorithm.

As a result of such study, a quantitative feature for the overall melodic progression, *seyir*, is considered to be useful for improving n-gram based classification. *Seyir*, is often considered as the most important characteristic that differentiates makam from key in Western music. It is usually considered as a roadmap of emphasized notes. An example is the makam *Rast* as explained in [5]: "The melodic progression begins with the *Rast* flavor on *rast* (G) due to the makam's ascending character. Following the half cadence played on the dominant neva (D), suspended cadences are played with the *Segah* flavour on *segah* and the *Dügah* flavor on *dügah* (A). The extended section is

---

[1] Trichords, tetrachords or pentachords used to create a melodic phrase emphasizing a certain degree. Interested readers are invited to refer to [12] for a detailed analysis of "flavours" in improvisations.

presented and the final cadence is played with the *Rast* with *acem* (F) flavor (should be note) on the tonic *rast* (G)". This description should be considered as a rough guide, not an exact point by point analysis.

Our main focus here is to use the melodic progression information for improving automatic classification of makams that use the same musical scale and tonic (*Hüseyni - Muhayyer*, *Beyati - Uşşak*). To our knowledge, *seyir* has not been computationally studied before. A detailed analysis of *seyir* would necessitate labeled data in sequence of flavors and emphasized notes and detailed musicological analysis of pieces to guide such work. For the moment, we lack both labeled data and an automatic means to extract such information. Therefore, as an initial effort to compute *seyir*, our rather simple method is limited to automatically extracting a feature and investigating its usefulness within an automatic makam detection test. Computational analysis of *seyir* is among our future goals and we are currently collecting manual labels from musicians on scores to start such detailed study.

To be able to define a quantitative feature for *seyir*, we used a down-sampling strategy of the melodic progression of the music pieces into a fixed length index (20 points in this study), and observe them graphically. Figure 6 shows the down-sampled melodic progression graphs for one of the challenging makam couples in our data set, *Muhayyer* and *Hüseyni*.
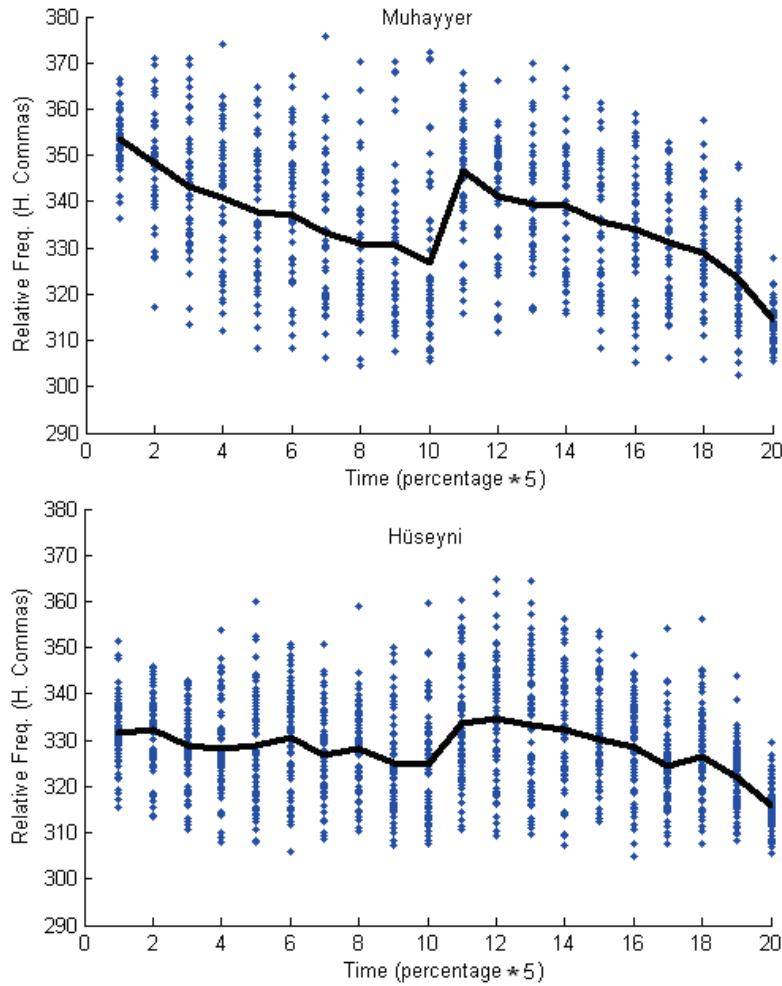
Figure 6. Melodic progression of makam *Muhayyer* and *Hüseyni*.

The *'x* axes' (time axis) in the graphs represent the index numbers and the '*y* axes' represent the pitch in Holdrian commas with respect to C_1 at the particular index. Dots are the samples and the solid black lines are the average progression for the entire makam collection. One can see from the graphs that, the beginning part of the average progressions might be more diagnostic for the makam classification task where the most melodic movement differentiation was observed. This observation is in-line with the findings of [6]. Two approaches are considered for improving automatic makam detection. First, in order to test the effect of this difference on makam classification within the n-gram based learning and classification approach, 3 tests were performed:

(1) models constructed from the full length music pieces, tests performed with the entire piece,

(2) models constructed from the full length music pieces, tests performed only using the first quarter of the piece,

(3) models constructed from the first quarter of the music pieces, tests performed only using the first quarter of the input melody.

Second approach is to define a quantitative feature and use this feature in a hierarchical framework as explained in the next section.

	N=1	N=2	N=3
Full vs full	86.9	87.9	87.5
1st quarter vs full	87.7	88.7	87.7
1st quarter vs 1st quarter	83.7	86	85.7

Table 6. The results for the experiments *1,2* and *3*

The comparison of all results with respects to tests (1), (2), and (3) with the effect of increasing n-gram order can be seen in Table 6. 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 accuracy of the makam detection can be slightly (%0.9) improved if testing is performed only from 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 lack of data required for modeling however, this effect can only be analyzed with more data in later studies.

7. Hierarchical classification using melodic features

Until this point, all the makams are considered as independent classes and individual nodes. However, as the results suggest, even though increasing the length of the n-

grams and partial modeling and testing strategy helps, the makams that have common musical scale are confused with each other at a certain level. In order to further improve our results, we would like to incorporate global melodic features alongside the local note movements (n-grams) in our experiments.

This is done by adopting a hierarchical strategy for the classification process. For this, the makams that have the same musical scale are grouped together as shown in Figure 7. First, the local note movements define the first stage of the classification, and then rule based global melodic features are used for the final decision.
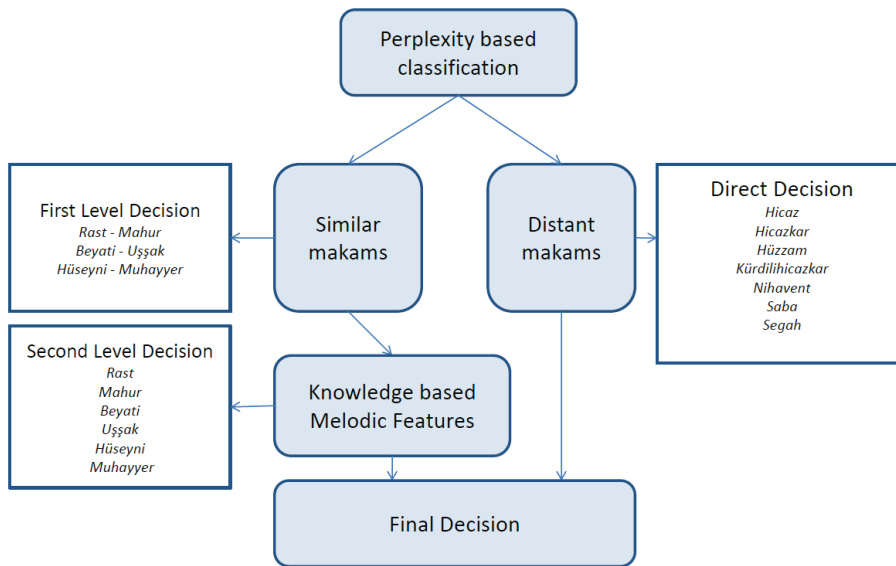

Figure 7. The hierarchical classification approach.

As seen from Figure 7, in the first step, makams with similar scale are grouped together, and the left out makams are considered individually. Considering the confusing makam couples (*Uşşak - Beyati*; *Hüseyni - Muhayyer*; *Rast - Mahur*), and the left out makams (*Hicaz, Hicazkar, Kürdilihicazkar, Hüzzam, Nihavent, Saba and Segah*) as single entities, the system uses perplexity between the first level makam models and the input piece for the first stage. And then, the system uses rule based decision criteria to perform lower level identification.

The decision rules applied in the second level of the hierarchical methodology are only applicable for binary decision between the confusing makams, but not for identifying them directly at the first level. So, when an *Uşşak* input piece is evaluated by the system, it first needs to be defined as a member of the '*Uşşak - Beyati*' and later by using the decision rules, it can only be identified as *Uşşak*. If, the makam is not a member of the confusing makam couples, the decision is made at the end of the first step, meaning that no further analysis is needed.

7.1 Theoretical clues

7.1.1   First stage: Tonic Information

While *Rast* and *Mahur* makams are successfully categorized, there is still potential for improvement. One very important rule about the *Rast - Mahur* makams is that the pieces belonging to this makam couple ends with the note *rast* (same as the makam name) and it is decoded as the symbol G4 in the SymbTr format. It is a general rule for makam music that the last note in the piece is the *karar*/tonic, for these two makams, *rast*. The exception to this rule is very rare. This rule gives us the chance to correct classification errors for pieces that are identified as other confusing makams *Uşşak - Beyati* and *Hüseyni - Muhayyer*. Once a music piece is classified as *Uşşak - Beyati* or *Hüseyni - Muhayyer,* the system first checks if the tonic rule applies. If the ending note is confirmed to be *rast*, then the output of the first classifier is corrected to be *Rast - Mahur*, instead of *Uşşak - Beyati* or *Hüseyni - Muhayyer*.

Applying this rule increased the classification accuracy (recall) for *Rast - Mahur* detection from 82.4% to 97.8% and the general classification accuracy from 88.6% to 90.9% (for n=3). This rule is only applicable at the second stage, since the tonic of some of the other makams (*Nihavent, Hicazkar and Kürdilihicazkar*) is also *rast*.

7.1.2 Second stage: Start Index versus Sum of Deltas

As seen from Figure 2 in Section 2, even though the musical scale of makam *Hüseyni* and *Muhayyer* are the same, there are certain differences in the melodic progression. According to our observations for the melodic progression, the starting pitch position of the pieces might carry makam specific information. This observation is in-line with the findings of musicological studies such as [6]. As we have presented in the introduction, there are three basic types of *seyir;* ascending, descending and mid-range progression. All progressions end on the tonic and the main difference is where the progression starts. The *seyir* describes the progression of the first part of the piece and following this first part, there exist "regions of creativity" where transitions to other makams may take place. In order to quantify this, a feature called the Start Index is defined. Start Index is defined as the average pitch of the first 5% of the melody. Also, overall decay of the melody is represented with a sum of deltas (intervals between consecutive pitches) feature. In the following figures 8, 9 and 10 we provide graphs for the makam couples under investigation, providing a statistical decision boundary for the second stage decision with respect to Start Index and the Sum of Deltas features.

Even though we present the graphs for two independent features (Start Index, Sum of Deltas), the decision boundary was calculated using only the Start Index, since the Sum of Deltas did not improve our decision results. The decision boundary was calculated from the Gaussian Mixture Models (GMM) models of the confusing makam sets using the nearest mean classifier. Since the features can be assumed to be independent, have equal variance, the decision of each instance can be calculated using the Euclidian distance instead of the Machalanobis distance according to the Maximum Likelihood rule.

For *Rast - Mahur* decision, the boundary is at the 330 comma, while the *Hüseyni - Muhayyer* decision is at 345. The two decision boundaries applied for *Rast - Mahur*

and *Hüseyni - Muhayyer* resulted in an 89.8 % recall rate. Similarly a decision boundary is also calculated for *Uşşak - Beyati* (at the 322 comma). However, the recall rate is considerably low, which is around 60.5%. The overall evolution of the recall rate from Stage 1 classifier to Stage 2 classifier can be seen at Table 7 and Table 8 for n=3 where the best performance was achieved.
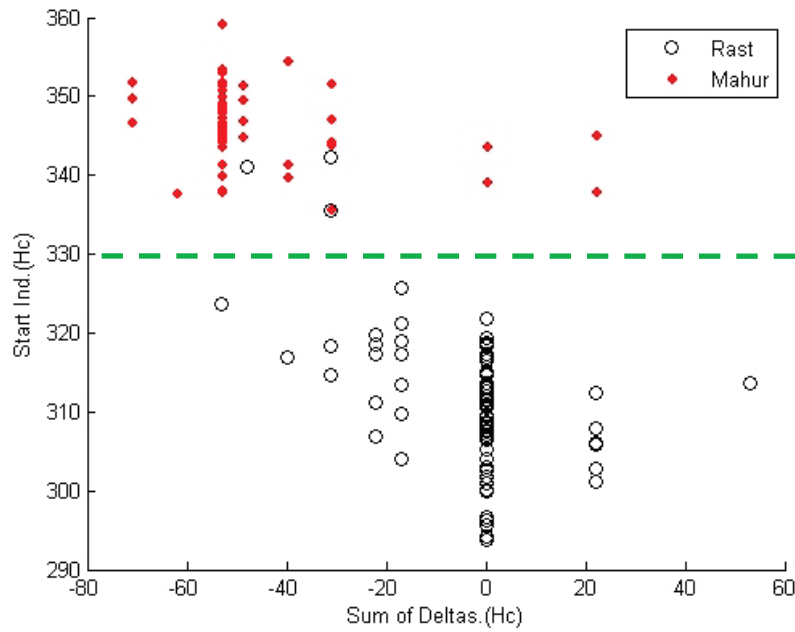

Figure 8. The Start Index vs the Sum of Deltas for *Rast - Mahur* confusion and the Maximum Likelihood decision boundary
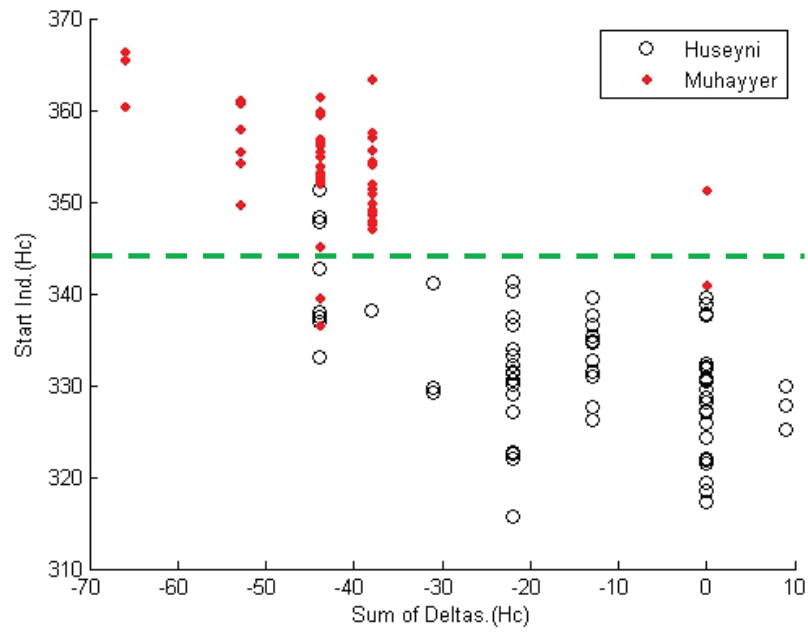
Figure 9. The Start Index vs the Sum of Deltas for *Hüseyni - Muhayyer* confusion and the Maximum Likelihood decision boundary
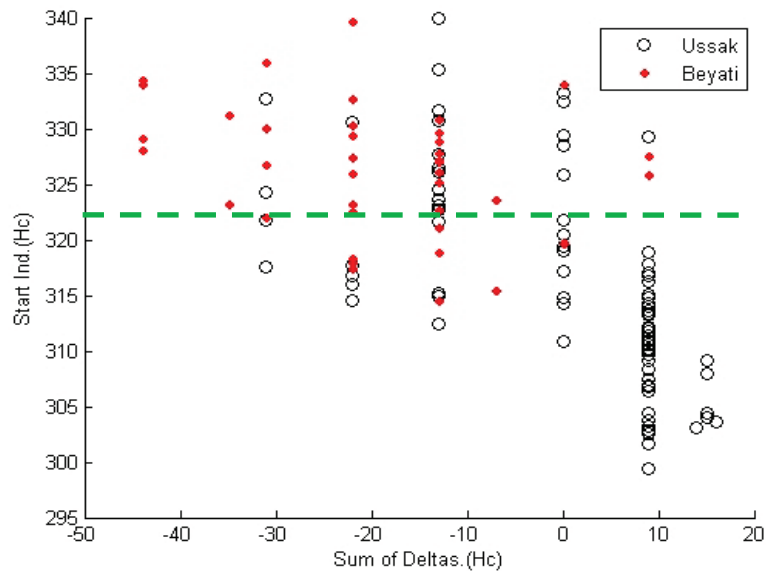

Figure 10. The Start Index vs the Sum of Deltas for *Uşşak - Beyati* confusion and the Maximum Likelihood decision boundary

3gram Results

bytus	hicaz	hczkr	hsymh	huzzm	Krdhz	nhvnt	rstmh	saba	segah		
104	0	0	21	0	0	0	0	0	2	bytus	81.9
0	117	0	0	0	0	0	0	0	0	hicaz	100
0	0	49	0	0	0	0	0	0	0	hicazkar	100
14	1	0	105	0	0	0	1	0	0	hsymh	86.8
0	0	0	0	63	0	0	0	0	2	huzzm	96.9
0	0	0	0	0	51	0	0	0	0	krdhz	100
0	0	0	0	0	0	86	0	0	0	nhvnt	100
0	0	0	1	0	0	0	141	0	0	rstmh	99.3
0	0	0	1	0	0	0	0	44	0	saba	97.8
0	0	0	1	2	0	0	0	0	71	segah	95.9
88.1	99.2	100	81.4	96.9	100	100	99.3	100	94.7		

Total Weighted ACC is : 94.6
Total ACC is : 95.9

beyati	hicaz	hczkr	hsyni	huzzm	Krdhz	mahur	mhyrr	nhvnt	rast	saba	segah	uşşak		
24	0	0	5	0	0	0	0	0	0	0	0	13	beyati	57.1
0	117	0	0	0	0	0	0	0	0	0	0	0	hicaz	100
0	0	49	0	0	0	0	0	0	0	0	0	0	hczkr	100
7	1	0	59	0	0	0	2	0	1	0	0	1	hsyni	83.1
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	54	0	0	0	0	0	0	mahur	100
6	0	0	3	0	0	0	41	0	0	0	0	0	mhyrr	82
0	0	0	0	0	0	0	0	86	0	0	0	0	nhvnt	100
0	0	0	1	0	0	3	0	0	84	0	0	0	rast	95.5
0	0	0	1	0	0	0	0	0	0	44	0	0	saba	97.8
0	0	0	1	0	0	0	0	0	0	0	71	0	segah	95.9
12	0	0	16	0	0	0	0	0	0	0	2	55	uşşak	64.7
49	99.2	100	68.6	96.9	100	94.7	95.3	100	98.8	100	94.7	79.7		

Total Weighted ACC is : 90.9
Total ACC is : 90.2

Table 7. The classification performance for Stage 1 and Stage 2. (bytus = combination of *Beyati* and *Uşşak*, hsymh = combination of *Hüseyni* and *Muhayyer*, rstmh = combination of *Rast* and *Mahur*)

	BaseLine (%)	Hierarchical (%)	Improvement (%)	Significance (p<0.05)
Beyati	47.6	57.1	9,5	+
Hicaz	99.1	100	0.9	+
Hicazkar	95.9	100	4.1	+
Hüseyni	66.2	83.1	16.9	+
Hüzzam	96.9	96.9	0	-
Kürdilihicazkar	100	100	0	-
Mahur	94.4	100	5.6	+
Muhayyer	82	82	0	-
Nihavent	98.8	100	1.2	+
Rast	93.2	95.5	2.3	+
Saba	97.8	97.8	0	-
Segah	95.9	95.9	0	-
Uşşak	70.6	64.7	-5.9	+
TOTAL	88.7	90.9	2.2	+

Table 8. The effect of hierarchical classification with respect to the base line and significance test result using (p<0.05)

As seen from Tables 8 and 9, the overall recall performance of the classifier improved 2.2% (absolute), where the specific recall rates of *Beyati*, *Hicaz*, *Hicazkar*, *Hüseyni,*

Mahur, Nihavent and *Rast* have been improved, while only the recall rate for *Uşşak* makam has decreased. The system's overall recall rate is promising, and we had a chance to support our claims about the feasibility of a hierarchical methodology for makam detection; however, the efficiency of the *Uşşak - Beyati* classification is low. This is not too much of an unexpected result. In literature, we find works (such as [25]) that claim that the discrimination between *Uşşak* and *Beyati* is artificial, based on observations on pieces which are classified as *Uşşak* by one author and *Beyati* by another. An example is the piece "Canım Tezdir Sabredemem" by Tanburi Mustafa Çavuş, which is classified as *Uşşak* by Dr. Suphi Ezgi and *Beyati* in a respected song-book publication.

8. Discussion and Conclusion

In this study, we proposed a hierarchical framework for statistical and rule based classification of makams using symbolic data. First the baseline system that uses n-grams and perplexity in a single classification stage was built. Here, the basic idea was to test the potential of short length (n-length) melodic frames for classification, using different length n-grams.

The symbolic data selected from [17] included 13 makams and 877 music pieces. The song set was a challenging one due to makams that have high level of similarity, especially for the ones that use the same scale.

Pieces with long forms (such as *ayin*) were not included in the data set to avoid large regions with makam transitions within a piece. Even short pieces may include makam transitions (especially in the third sections or *meyan* sections) but such transitions are most of the time very short hence is not very problematic considering the system's statistical approach.

The leave-one-out strategy was used. A music piece was randomly selected from the corpus and left out. The remaining pieces were used for modeling n-grams. In order to compensate for the zero frequency problem, the Written-Bell smoothing technique available in the SRILM toolkit was used. After the smoothing is finished, the models became ready for classification. The left out piece was then compared to the available models in the collection using a cross entropy based similarity metric called, perplexity. The model which has the minimum perplexity for the specific input piece was the classification output of the system.

The classification accuracy (recall) of the baseline system was calculated to be 87.9%. For the makams that use the exact same scale (*Uşşak - Beyati, Hüseyni - Muhayyer*), a high level of confusion was observed.

Different representations were used, including the Arel theory notation, 12-TET notation and interval representation in order to see the effect of microtonal representation over the classification accuracy. The results showed that the microtonal representation improves the classification accuracy when compared to 12-TET, boosting the performance from 84.5% to 87.5% recall rate. We also tested our system using the interval contour representation because in real life, contour information is easier to extract when compared to extracting exact microtonal notation. Results showed that, using the comma level contour information, the system works with 80% recall rate showing that the performance of the Arel representation is closely achievable under real conditions.

The complexity analysis of the tests suggests that most of the confusion comes from the makams that use the same musical scale. Even though the increasing length of n-grams and concentrating on distinctive parts of the melody helps, a hierarchical approach is more desirable, since there might be some statistical clues related to the

musical structure, for better classification. First, makams that use the same musical scale are grouped together, to be classified in the first stage of the hierarchical framework, using perplexity. And then, statistical observations are applied as decision rules for the latter stage where the final decision is made.

Three distinctive features were tested for the second stage: the tonic as the final note, Start Index within the first 5 percent of the melody, and the Sum of Deltas.

We then used a down-sampling strategy for detailed analysis of the melodies belonging to the same makam as discussed in Sections 2 and 3. Observations suggest that, for the makams that are using the same musical scale, some rules about the melodic progression might help to separate one from another. The Start Index, the feature we defined for average position of the first 5% of the melody was used. Results showed that, on average, 2.2% improvement was achieved, suggesting that the melodic features related to the global progression of the melody actually help in better classification of makams that have similar musical scale. The base-line using n-grams is already quite efficient (%87.9) and an improvement of 2.2% using a simple to compute *seyir* feature signifies that there is still room for improvement when a detailed computational analysis of *seyir* and some results of makam perception studies are available.

As stated earlier, our final goal is to apply this study to more practical usage such as categorization of audio signals. Future work will focus on adaptation of this study into the audio domain. Since a direct transcription is going to be noisy with insertions and deletions, the biggest challenge is to develop a method robust against transcription errors and similar audio related variations (different keys, instrumentation, tempo, etc...)

10. Acknowledgement

11. References

1. S. Abdoli, "Iranian Traditional Music Dastgah Classification," *in Proceedings of the Int. Society for Music Information Retrieval (ISMIR), Florida*, p. 275-280, 2011

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

3. T. Aoyagi, "Makam Rast: Intervallic Ordering, Pitch Hierarchy, Performance and Perception of a Melodic Mode in Arab Music," *PhD Thesis, University of California*, 2001.

4. H. S. Arel, "Türk Musikisi Nazariyatı Dersleri, Hazırlayan Onur Akdoğu," *KültürBakanlığı Yayınları /1347, Ankara*, p.70. 1991.

5. M. Aydemir, Turkish Music Makam Guide, *Pan Yayıncılık, Istanbul*, 2010.

6. M. E. Bayraktarkatal, O. M. Öztürk, "Ezgisel Kodların Belirlediği Bir Sistem Olarak Makam Kavramı: Hüseyni Makamı'nın İncelenmesi," *Porte Akademik,* 2012, 3(4), p. 24.

7. B. Bozkurt, "Features for analysis of makam music," *in Proceedings of the 2nd CompMusic Workshop*, 2012

8. B. Bozkurt, O. Yarman, M. K. Karaosmanoğlu, C. Akkoç, "Weighing diverse theoretical models on Turkish maqam music against pitch measurements: A Comparison of peaks automatically derived from frequency histograms with

proposed scale tones," *Journal of New Music Research*, Volume 38, Issue 1, p. 45, 2009

9. N. Darabi, N. Azimi and H. Nojumi, "Recognition of Dastgah and Makam for Persian Music with Detecting Skeletal Melodic Models," *in Proceedings of the The second annual IEEE BENELUX/DSP Valley Signal Processing Symposium,* 2006.

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

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

12 E. B. Ederer, "The Theory and Praxis of Makam in Classical Turkish Music 1910-2010," *PhD Thesis, University of California, Santa Barbara, 2011.*

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

14. A.C. Gedik, C. Işıkhan, A. Alpkoçak and Y. Özer, "Automatic Classification of 10 Turkish Makams," *in Proceedings of the International Congress on Representation in Music & Musical Representation*, İstanbul, 2005.

15. H. S. Heaps "Information Retrieval: Computational and Theoretical Aspects," *Academic Press, Inc. Orlando, FL, USA ©1978 ISBN:0123357500*

16. L. Ioannidis, E. Gómez, and P. Herrera. *"*Tonal-based retrieval of Arabic and Middle-East music by automatic makam description*," in Proceedings of the 9th International Workshop on. Content-Based Multimedia Indexing (CBMI), IEEE, 2011. p. 31-36.*

17. M. K. Karaosmanoğlu, "A Turkish makam music symbolic database for music information retrieval: SymbTr," *in Proceedings of the International Society for Music Information Retrieval (ISMIR)*, 2012.

18. İ.H. Özkan, "Türk Musikisi Nazariyatı ve Usulleri - Kudüm Velveleleri," *Ötüken Neşriyat, İstanbul, 1990.*

19. F. W. Stubbs, "The art and science of taksim: an emprical analysis of traditional improvisation from 20th century Istanbul." *PhD Thesis, Wesleyan University*, 1994.

20. A. Stolcke: "Srilm – an Extensible Language Modeling Toolkit," *in Proceedings of the international conference on spoken language processing. 2002. p. 901-904.*

21. E. Unal, E. Chew, P. Georgiou and S. Narayanan, "Perplexity based cover song identification System for short length queries," *in Proceedings of the Int. Society for Music Information Retrieval (ISMIR), Florida*, 2011

22. E. Unal, B. Bozkurt, M. K. Karaosmanoğlu, "N-gram based Statistical Makam Detection on Makam Music in Turkey using Symbolic Data," *in Proceedings of the Int. Society for Music Information Retrieval (ISMIR), 2012.*

23. E. Unal, B. Bozkurt, M. K. Karaosmanoglu, "Incorparating features of distribution and progression for automatic makam classification", *in Proceedings of the* 2nd CompMusic Workshop, 2012

24. E. Unal, Elaine Chew, Panayiotis Georgiou and ShriNarayanan,"Statistical modeling and retrieval of polyphonic music", *in Proceedings of the IEEE Multimedia Signal Processing Workshop*, 2007, *p. 405-409*

25. A. Zeren, "Müzik sorunlarımız üzerine araştırmalar," *Vol. 107*, *Pan Yayıncılık, İstanbul*, 2003.