

# Usul and Makam driven automatic melodic segmentation for Turkish music

Bariş Bozkurt<sup>1\*</sup>, M. Kemal Karaosmanoğlu<sup>2</sup>, Bilge Karaçalı<sup>3</sup> and Erdem Ünal<sup>4</sup>

<sup>1</sup>Bahçeşehir University, Turkey; <sup>2</sup>Istanbul Technical University Turkish Music State Conservatory, Turkey; <sup>3</sup>Izmir Institute of Technology, Turkey; <sup>4</sup>Information Technologies Institute, Turkey

(Received 30 August 2013; accepted 6 May 2014)

## Abstract

Automatic melodic segmentation is a topic studied extensively, aiming at developing systems that perform grouping of musical events. Here, we consider the problem of automatic segmentation via supervised learning from a dataset containing segmentation labels of an expert. We present a statistical classification-based segmentation system developed specifically for Turkish makam music. The proposed system uses two novel features, a makam-based and an usul-based feature, together with features commonly used in literature. The makam-based feature is defined as the probability of a note to appear at the phrase boundary, computed from the distributions of boundaries with respect to the piece's makam pitches. Likewise, the usul-based feature is computed from the distributions of boundaries with respect to beats in the rhythmic cycle, usul of the piece. Several experimental setups using different feature groups are designed to test the contribution of the proposed features on three datasets. The results show that the new features carry complementary information to existing features in the literature within the Turkish makam music segmentation context and that the inclusion of new features resulted in statistically significant performance improvement.

**Keywords:** maqam, makam, melodic analysis, melodic segmentation, melodic grouping

## 1. Introduction

Automatic melodic segmentation is an extensively studied topic since it is one of the important initial steps of many systems developed for automatic melodic analysis. Melodic segmentation was first introduced and studied within the

domain of music perception and cognition. Computer-based methods to automatically perform such segmentation finds use in the fields of computational musicology and Music Information Retrieval (MIR). MIR systems targeting indexing and search in large databases are examples of such systems. When the segment boundaries are not available, exhaustive search methods are used that are computationally expensive and inefficient. An automatic melodic segmentation tool is potentially useful for studies on higher-level organization (structure) of the musical surface. In this work, we target the development of an automatic segmentation system for Turkish makam music to be used in a computational study of the makam repertoire at the phrase level.

Our specific objective in this work is to develop a statistical classifier that can learn from a manually segmented database of phrases for Turkish makam music and to automatically segment larger amounts of data using the developed classifier. Such an algorithm is aimed to be a part of a battery of automatic melodic analysis tools dedicated to musicological research on makam-specific melodic progressions. Collecting statistics about phrases in pieces of a specific makam, pieces of a specific composer, pieces in a specific period, etc. are some examples of automatic melodic analysis operations we would like to perform. Another objective is the development of an automatic *seyir* analysis tool. *Seyir* refers to the melodic progression of a piece or improvisation. The melodic progression descriptions presented in theory books include both long-time and short-time features. The long-time characteristic is categorized to belong to one of the three classes: ascending, ascending–descending (or alternatively ‘*seyir* in the mid-register’) or descending, a classification for the overall direction of the melodic contour. *Seyir* descriptions also include more local information (short-time characteristic) such as a representation of the ordered sequence of emphasized

Correspondence: Dr Bariş Bozkurt, Electrical and Electronics Eng. Dept., Bahçeşehir University, Istanbul, Turkey. E-mail: [baris.bozkurt@bahcesehir.edu.tr](mailto:baris.bozkurt@bahcesehir.edu.tr)

tones and *çeşniler*<sup>1</sup>. Once such tools are available, a computational study of the *seyir* characteristics of a specific makam can be carried out on a large repertoire.

### 1.1 The ‘makam’ concept and melodic segmentation

Powers and Wiering (n.d.) consider the makam concept as a mode in the sense of a particularized scale or a generalized tune, in the middle of a continuum between scale and tune on its two ends. The popular makam theory books used in music circles today (e.g. Özkan, 2006) include 1–2 page descriptions for each makam specifying a scale, a list of emphasized tones and the tonic, the common *çeşniler* and specification of the overall melodic progression (*seyir*) category. According to Stubbs (1994, p. 1) makam is ‘a practical melody theory, grouping melodies by families or categories that are distinguished by the use of careful microtonal inflections<sup>2</sup> of certain tones according to custom, together with idealized notions of melodic contour’. While a large part of makam literature<sup>3</sup> emphasizes the importance of ‘melody’, computational methods or tools dedicated to the analysis of a large amount of scores to help researchers study the makam concept in relation to melodies are considered in just a few studies<sup>4</sup>.

In the literature on Turkish makam music, analysis of a piece involves the melodic segmentation on scores followed

<sup>1</sup>*Çeşniler* (plural of *çeşni*): A descriptive, functional melodic motive using a specific tri/tetra/penta-chord or a sub-division of it. While, in the Arel theory for Turkish makam music (Arel, 1968) (which is the dominating theory in today’s music circles), the scale of a makam is constructed by stringing together tri/tetra/penta-chords, the chords used in constructing melodies overlap (as the notes of a D Dorian melodic phrase overlaps the notes of a E Phrygian melodic phrase). Davis (n.d.) states that the concept of chords ‘originated in medieval Arab treatise, but disappeared between sixteenth and twentieth centuries. They were reintroduced into Turkish music theory by Rauf Yekta and Sadettin Arel in the early twentieth century, and to Arab music by Yekta’s student, the Syrian Shaykh ali Al-darwish...’. Some authors (e.g. Özek, 2011, p. 6) consider tri/tetra/penta-chords as equivalent of *çeşni*. We think this is misleading since *çeşni* refers to a function (taste), which uses the chord material.

<sup>2</sup>Microtonality is an important aspect of the pitch space for various makam music traditions such as the Arabic maqam music or Turkish makam music. The representation of microtonal pitches (that would conform the practice) and to what degree various makam traditions share the same microtonal pitches is part of controversial discussions some of which can be found in Zannos (1990), Ghrab (2005), and Yarman (2007b).

<sup>3</sup>A review of historical texts for Turkish makam music can be found in Öztürk (2011) where the author discusses that basically two different conceptualizations dominate the literature; ‘a scale-centred approach’ and ‘a melody-centred approach’. The author favours the latter arguing that the scale-centred approach is the result of modernization to ‘resemble the West’.

<sup>4</sup>In Bozkurt, Ayangil and Holzapfel (2014) the authors reviewed in detail computational studies in this domain as well as the basic makam concepts linked to these computational approaches. We avoid duplication of such material here and refer interested readers to that review paper.

by the labelling of the phrases with (tri-tetra-penta) chords<sup>5</sup> used<sup>6</sup> and a further study of the functions within the makam. One of the important challenges for developing automatic tools that can perform similar operations on large repertoires and gather statistical data, is automatic melodic segmentation. In the Turkish makam musicology literature, we do not find any explicitly formalized methodology to perform melodic analysis or melodic segmentation by humans. Due to this lack, we find it advantageous to build an adaptive system that can be tuned to a specific analysis/segmentation style of an expert (given a dataset of phrases manually segmented by the expert). It should be emphasized here that melodic segmentation carried out in this study is performed directly on scores (by the experts), not by listening to audio recordings (which is the common test reference in automatic melodic segmentation literature discussed below).

Lartillot and Ayari’s paper (2009) is the first computational study that examines the link between makam concepts and melodic segments (within Tunisian maqam music). With the aim of studying segmentation methodologies, they consider the perceptual segmentation of an improvisation. In addition to comparing human segmentation with automatic segmentation, the authors also collected data about how labellers made their choices. In an earlier work, Ayari (2005) discusses the following culture specific segmentation strategies (in addition to tracking discontinuities in auditory features and parallelism): ‘a musical phrase tends to end on a scale degree that is hierarchically superior (the tonic in Western tonality, the pivot note or the root of an Arabic Maqam)’ and ‘a recognition of a genre specific motivic formula triggers a transition between genres, producing a segmentation at a relatively high level of the hierarchy’. In our discussions with Turkish makam music experts, we observed that many musicians hold the same intuition while we could not find any explicit reference in the literature. In their computational model, Lartillot (2007) encodes the specific melodic-rhythmic patterns (by defining rules) related to the genre as an additional feature to Gestalt principles of proximity and similarity (which form the basis of most of the automatic melodic segmentation algorithms), and the parallelism criterion (for which they also propose a new approach). Recently, Lartillot, Yazıcı and Mungan (2013) studied the perceptual boundary detection problem for Turkish makam music where they discussed the limitations of presuppositions in the literature (more specifically that of Tenney and Polansky (1980) and Cambouropoulos (2006)). They proposed a new approach that treats the dimensions merged by weighted summations in these previous studies independently and obtained better performance. They also conclude that the results suggest a probable interaction between low-level perceptual processes and culture-specific

<sup>5</sup>Please refer to the previous footnote on *çeşni* for information on chords.

<sup>6</sup>Chord labels/segments and phrase labels/segments may or may not overlap. For an example please refer to Figure 1, Section 2.

knowledge for perceptual grouping at higher levels for boundary decision.

Concerning the link between phrase boundaries and hierarchies of makam scale degrees, it is known that phrase boundaries are highly correlated with the rhythmic cycle (usul) hierarchies<sup>7</sup>. In the oral tradition of Turkish makam music, the learning/memorization of pieces involves first learning and internalizing the usul pattern and then further memorizing the melodies in relation to the usul pattern (Behar, 1998). While we did not find any theoretical study discussing the link between the phrase boundaries and hierarchies in scale degrees and rhythmic cycles, in this work we target employing this phenomenon using a data-driven approach.

## 1.2 The automatic segmentation literature

In several studies (Bruderer, 2008; Neve & Orio, 2005; Nooijer, Wiering, Volk, & Tabachneck-Schijf, 2008; Pearce, Müllensiefen & Wiggins, 2010), the automatic melodic segmentation literature is reviewed in detail. As suggested by Pearce et al. (2010) melodic segmentation studies can be classified into three main groups: ‘Music-theoretic Approaches’, ‘Psychological Studies’ and ‘Computational Models’. To avoid repeating the detailed surveys in the above-mentioned papers, we will limit ourselves here to studies closely linked with our objective and the methodology that falls into the ‘Computational models’ category.

The Gestalt principles were first considered as an automatic segmentation methodology by Tenney and Polansky (1980) where two primary factors are considered for grouping and segregation: proximity in time and similarity (with respect to some parameter like pitch). Those events that are relatively contiguous in time and/or similar in parameters form groups. Detecting group boundaries are performed by detecting relatively greater separations of events in time and/or dissimilarities in parameters. Following this work, various computational studies implemented discontinuity detection in auditory features by computing deviations in IOIs (inter-onset intervals), OOs (offset-to-onset intervals), and/or pitch intervals and/or a weighted sum of these features to estimate local segment boundary probabilities (Cambouropoulos, 2001; Frankland & Cohen, 2004; Temperley, 2001). Cambouropoulos (2006) states that musical parallelism has also been considered as an important factor for musical segmentation but has been used in very few studies (later in Lartillot and Ayari (2009)) due to the difficulty in formalization. Cambouropoulos (2006) attempts to provide a computational model using parallelism in order to determine the segment boundaries, but without providing a complete melodic segmentation procedure. Lartillot (2007) presents a review of the motivic pattern extraction algorithms that could be beneficial for implementing parallelism. Frankland and Cohen (2004) use a method based on a quantification of local grouping rules of GTTM

(Generative Theory of Tonal Music) (Lerdahl & Jackendoff, 1983).

These algorithms have been comparatively evaluated in various studies. Nooijer et al. (2008) have performed a comparative evaluation of five selected state-of-the-art techniques on a dataset obtained from 40 labellers on 10 melodies. As an outcome of both their own test results and claims of previous comparative works (by other authors), they concluded that human output is closest to three of the algorithms: Grouper (Temperley, 2001), Information Dynamics (Pearce & Wiggins, 2006) and LBDM (Local Boundary Detection Model) (Cambouropoulos, 2001) (and their performances were similar) where the two other algorithms were Temporal Gestalt Units (Tenney & Polansky, 1980) and Melodic Similarity Model (Ahlbäck, 2004). Lartillot and Ayari (2009) also presented a comparison of the Temporal Gestalt Units and LBDM. Pearce et al. (2010) compares the above-mentioned algorithms with their own proposal, IDyOM (Information Dynamics Of Music).

IDyOM falls into the data-driven approaches, which estimates segment probabilities via supervised learning. This approach uses the *information content* defined in MacKay (2003) as a feature (within a multiple viewpoint representation (Conklin, 2001) including pitch and duration features) and a combination of several text compression and statistical language modelling (mainly n-grams) methods. In addition to comparing the mentioned algorithm, Pearce et al. (2010) tested a hybrid approach that combines features of different algorithms and concluded that the hybrid approach has better performance than any of the components. Previously, Bod (2001) used a data-driven approach where the probabilities of phrase boundaries had been computed from pitch-class, interval and duration distributions from a database of segment boundaries in the Essen folk song collection.

## 1.3 Proposed methodology

In this work, we follow the data-driven approach mainly for two reasons. First, building a statistical approach that would learn from a segmentation dataset is advantageous for allowing the musicologist to choose a specific dataset and obtain a tool to produce more data segmented in a similar fashion automatically. Second, we would like to study the contribution of usul- and makam-based features in automatic segmentation quantitatively. The drawback of the approach is that the system functions as a black box and contribution in the understanding of melodic segmentation is limited to listing features that are relevant to the task. In order to achieve this, we used the decision functions (or strength curves)<sup>8</sup> of four

<sup>7</sup>Private communication with makam music experts: Reha Sağbaşı, Ruhi Ayangil and Cem Çırak.

<sup>8</sup>Decision functions/strength curves refer to the sequence of estimated degree/strength values for each event (note for our problem) that is further processed (often by comparing to a threshold) to make a decision about the state (if the note is at the phrase boundary or not) for each event. For example the sequence of LBDM values for a piece is considered as a decision function/strength curve, which can be compared to a threshold to decide the onsets that correspond to the phrase boundaries.

of the above mentioned state-of-the-art techniques (LBDM (Cambouropoulos, 2001), Grouper (Temperley, 2001), Segment probabilities (Bod, 2001) and TGU (Tenney & Polansky, 1980) as features and performed a comparative analysis. The output of the TGU and Grouper implementations provide binary decisions about the notes (i.e. each note is labelled to be the first note of a phrase or not). For these two techniques, the decision functions created are composed of zeros (for non-boundary notes) and ones (for boundary notes). It should be specified that we do not aim for a thorough comparative evaluation of these systems since the context here (segmentation carried directly on scores) is different than the context in which they were designed. Instead, we studied their potential in the segmentation of Turkish makam music pieces. In addition to these features, an usul-based feature and a makam-based feature, proposed in this manuscript, were computed and their contribution to the automatic segmentation accuracy was studied. As explained in detail in Section 3, the usul-based feature was obtained from the distributions of segment boundaries with respect to the specific rhythmic cycle (usul), whereas the makam-based feature was obtained from the distribution of boundaries with respect to the scale degrees/pitches for the specific makam.

The contributions of this study can be listed as follows. The link between the metrical structure and grouping has been previously discussed by Lerdahl and Jackendoff (1983) and taken into consideration by Temperley (2001). However, it is the contribution of this study to examine it for usul structures (i.e. rhythmic modes in a specific culture) in a data-driven approach. We also elucidate its complementary contribution in melodic segmentation relative to the other features. The link between segmental boundaries and makam pitches has been discussed by Ayari (2005) and used in Lartillot and Ayari (2009) via certain rules. Here, we have used a statistical approach to compute a makam-based feature and also measured its relative contribution to the segmentation performance/accuracy. Overall, to the best of our knowledge, this is the first work on automatic segmentation carried out directly on scores in the context of Turkish makam music, tested on a large corpus (containing a total of 21,610 manually segmented phrases). Here we should emphasize that our main goal is not to perform an automatic ‘perceptual’ segmentation that could be generalized or compared to some average human segmentation. Rather, we target a system that learns from a manually created database of segmentations on scores and then segments larger amounts of data in a similar manner. Hence, our data and testing procedure differs considerably from the existing literature on automatic melodic segmentation. We provide test results for several state-of-the-art methods in this alternative context.

The organization of the manuscript is as follows: in the next section, we present the dataset used in this study. In Section 3, we present the feature extraction methodology and in Section 4, the supervised learning technique that forms the basis of the proposed automatic segmentation algorithm.

Finally, in Section 5, test results are presented and discussed. Section 6 concludes the manuscript.

## 2. The dataset

This work is part of a broader research in the domain of melodic analysis of Turkish makam music with the objectives of studying makam-specific melodic progressions and their variation in time, composer-specific choices in melodic organization and mapping symbolic level phrases to the audio data. To provide data for such studies, we have selected pieces of the Turkish makam repertoire: (i) composed in the most commonly used makams (Çevikoğlu, 2007), (ii) with a uniform distribution in time, from the seventeenth century to today, and (iii) from well-known composers such as İsmail Dede Efendi, Hacı Arif Bey and Sadettin Kaynak. Overall, a set of 450 pieces were collected consisting of 30 pieces for each of the 15 distinct makams by rewriting the pieces using Mus2 microtonal notation software (<http://www.mus2.com.tr/>) in the Arel (1968) notation and further converting to the machine readable text format of SymbTr (Karaosmanoğlu, 2012)<sup>9</sup>.

Three experts were asked to mark the phrase boundaries on printed scores, as they would do it for makam melodic analysis. The experts had studied Turkish makam music from different masters, attended different schools, and did not know each other in person. There was no time pressure, as the experts could do the segmentation anytime within a three-month period, and could use their instruments if needed. They did not have access to the segmentation of others for comparison. Their segmentations were manually exported to the machine-readable format using a custom-designed interface.

The experts could not complete the whole task due to time constraints. We finally compiled three sets of data from these experts containing 373, 133 and 128 pieces and 14,965, 2823, 4426 phrases, respectively. The distribution of the data with respect to makams and usuls are given in Table 1. Our efforts to get the data segmented by experts continue. This data will be offered for public access at the end of the process. In Figure 1, we present an example as labelled by all of the experts.

In Figure 1, both the melodic phrase boundaries and the *çeşni* labels are marked. While one of the experts (the second one) preferred to mark tri/tetra/penta-chord labels throughout the piece, other experts have seldom marked such labels indicating only *çeşniler* that use pitches out of the main scale of the makam. In this work, we only consider the phrase boundaries marked by vertical lines as shown in the figure. *Çeşni* labels

<sup>9</sup>Karaosmanoğlu (2012) explained in detail the SymbTr format as well as how this data can be mapped to other formats such as the MIDI representation. In this work, Karaosmanoğlu’s mapping is used to obtain fractional MIDI numbers to represent the microtonal pitches (e.g. in Figure 3, Section 3) of Turkish makam music as defined in the Arel (1968) theory that is mainly based on Pythagorean tuning (with 24 tones in an octave).

Table 1. Number of pieces segmented by the experts with respect to makams and usuls.

Makams	Expert 1	Expert2	Expert3	Usuls	Expert 1	Expert 2	Expert 3
Acemaşiran	12	—	11	Sofyan (4/4)	28	13	7
Beyati	8	—	7	Düyek (8/8)	46	14	15
Buselik	9	—	9	Aksak (9/8)	86	38	37
Hicaz	42	24	16	Semai (3/4)	18	3	7
Hicazkar	17	10	8	Yürüksemai (6/8)	27	10	7
Hüseyini	31	20	10	Aksaksemai (10/8)	2	6	1
Hüzzam	36	21	13	Senginsemai (6/4)	17	3	6
Krd.hicazkar	26	—	16	Türkaksığı (5/8)	17	8	5
Mahur	25	—	22	Nims Sofyan (2/4)	12	5	—
Muhayyer	19	—	16	Curcuna (10/8)	40	4	14
Nihavent	3	—	—	Müsemmen (8/8)	8	—	—
Rast	43	4	—	Devr-i hindi (7/8)	10	5	—
Saba	2	13	—	Ağıraksak (9/4)	25	11	11
Segah	27	20	—	Devr-i kebir (28/4)	7	3	—
Uşşak	26	21	—	Muhammes (32/4)	6	—	4
<b>Total</b>	<b>373</b>	<b>133</b>	<b>128</b>	Ağırdüyek (8/4)	3	—	—
				Hafif (32/4)	3	—	—
Grand Tot.			<b>634</b>	<b>Total</b>	<b>373</b>	<b>133</b>	<b>128</b>

**Gönlümü Dıçar Eden Bu Hale Hep**  
**Hicazkâr Şarkı**

Usul: Devr-i Hindi  
♩ = 84 ⇒ 1 Dk 58 Sn

Beste: Şevki Bey (1860 - 25/6/1891)  
Güfte: ?

Expert 1 (cc): ————  
Expert 2 (eo): - - - - -  
Expert 3 (kb): ······

Fig. 1. Example excerpt of a score from the database, which includes melodic phrase boundaries (vertical lines) labelled by three experts.

and their link to the melodic phrase boundaries will be considered in a future work, and hence, are not discussed here. In the example, expert 2 has marked six boundaries where experts 1 and 3 have marked 17 and 14 boundaries respectively. In Figure 2, the phrase length distributions (in number of notes and in seconds) for the three datasets are given. The tempo information (which is a main multiplier factor in the length in seconds) for the pieces have been manually set by trial and error to achieve an average speed of synthesis that resembles performances by various masters. As clearly observed on the figures, each expert has a different choice in terms of average phrase lengths.

The data from different experts contained some overlap. Comparing the pieces segmented, data from expert 2 and expert 3 have only five pieces in common while all pieces segmented by experts 2 and 3 are also segmented by expert 1. Hence, we can measure the degree of match using the standard measures, precision, recall and F-measure (formally defined in Section 5). We present these results in Table 2. In general, the segmentations by experts 1 and 3 were more in agreement compared to those by expert 2, since expert 2 preferred longer segments.

### 3. Feature computation

#### 3.1 Makam- and usul-based features for melodic segmentation

In this paper, we propose two new features for automatic melodic segmentation: a makam-based and an usul-based feature. These features are computed using a simple learning approach. In the learning phase, distributions of the phrase boundaries in the dataset are collected with respect to pitches for each makam and with respect to note onsets<sup>10</sup> for each usul (some example distributions are shown in Figure 3<sup>11</sup>). In the automatic segmentation phase, these distributions are used to estimate the makam-based and usul-based features.

For a given piece to be segmented, the makam-based feature is computed/assigned for each note, as the value of the phrase boundary distribution for the makam of the piece, for that

actual pitch. In other words, it is the probability of a makam pitch to appear at the phrase boundary.

The usul-based feature is computed/assigned for each note onset in a similar way considering only the time span of the note within the usul, as the highest value of the phrase boundary distribution for the usul of the piece within the duration of the note (its relative time span in the usul cycle). In other words, the probability of each instant within the time span of a note is computed and the maximum value is assigned to the note. As a result, the usul-based feature is the same for all notes in a piece that starts at the same starting instant in the usul and have the same duration. Here, the distributions are computed for each 16th note onset of the usul cycle. In Figure 3, we present sample distributions from all the datasets, for a piece in makam Hicaz and usul Aksak.

The first observation is that the distributions obtained from different datasets are very similar, indicating that all experts made similar choices with respect to these parameters. It is interesting to observe that this consistency holds to a very large extent for the three datasets for all makams and usuls. In the written Turkish makam music literature, we do not find any discussion/specification on phrase boundaries that can provide a theoretical explanation for this observation. The guidelines in the literature mainly specify which *çeyni* is used to emphasize which pitch, but nothing regarding the phrase boundaries. In particular, the emphasis does not necessarily need to be at the end of the phrase. While these findings are not surprising within the general domain of Turkish makam music, this is the first time such a relationship between the phrase boundaries and makam pitches and usul are observed on actual data and the theoretical background needs to be explored further. We consider this line of investigation as part of our future work.

3225, 738 and 1157 phrases were found in usul Aksak and 2165, 656 and 639 phrases were found in makam Hicaz in the three datasets. The distribution of the phrases with respect to beats in the usul cycle (Figure 3(b)) shows that the highest number of phrase boundaries appeared at the first beat of the cycle. In the usul distributions for this odd-beat usul Aksak (9/8), the indices 1, 3, 4 and 7 are the most common positions for the phrase boundaries. It is interesting to note that this is in agreement with the symbolic description of the usul<sup>12</sup> (as shown in Figure 4) corresponding to the first ‘Düm’, ‘Te’, ‘Ke’, and the first ‘Tek’.

Considering Figure 3(a), we observe that for makam Hicaz, the *karar*<sup>13</sup> pitch (A4, MIDI No: 69.06) appears as the most frequent phrase boundary (ending note). The second most

<sup>10</sup>A note onset corresponds to the starting instant of a note. Where applied, the specific mention of an onset instead of a note for a boundary aims to refer to an instant rather than a note that spans a time interval.

<sup>11</sup>The *aksak* usul shown in Figure 4 has a length of nine, which results from summing up the notated durations, assuming long notes to be of length 2, and short notes to be of length 1. As Turkish makam music usually uses a Western staff notation, this usul is notated with a 9/8 time signature. While we refer to a specific usul pattern here (sound examples for the Aksak pattern can be found on <http://www.freesound.org/search/?q=aksak>), ‘aksak’ is also used to refer to the category of asymmetric usul cycles. An usul can be encountered in a wide range of tempi. Certain tempo classes are denoted by the denominator of the time signature, which is referred to as *meretebe* in Turkish language. Using the example of *Aksak*, its slow version is called as *Ağıraksak* (lit. trans. *slow aksak*), and shown as a 9/4.

<sup>12</sup>Turkish makam music rhythmic modes/cycles are notated using syllables that indicate accents. The upper line is executed using the right hand and the lower line is executed using the left hand. The syllable ‘Düm’ is usually related to a stronger accent; however, no generally valid way to differentiate the strength of the accents exists. This or similar forms to notate usul are reported to be used since nineteenth century in Turkish music (Marcus, 2001).

<sup>13</sup>*Karar* (*Perdesi*): Final resolution pitch/note for a makam.

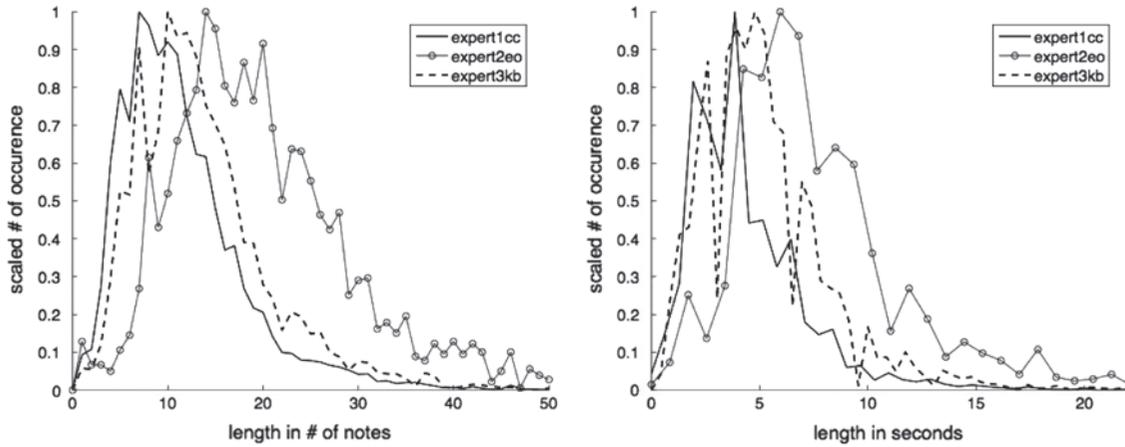


Fig. 2. Phrase length distributions of the three datasets in number of notes and in seconds.

Table 2. Comparison of segmentations by three different experts.

	Precision	Recall	F-Measure
Expert 1 versus 2 (133 common pieces)	0.4618	0.8046	0.5715
Expert 1 versus 3 (128 common pieces)	0.7424	0.8448	0.7827
Expert 2 versus 3 (5 common pieces)	0.8604	0.3712	0.5096

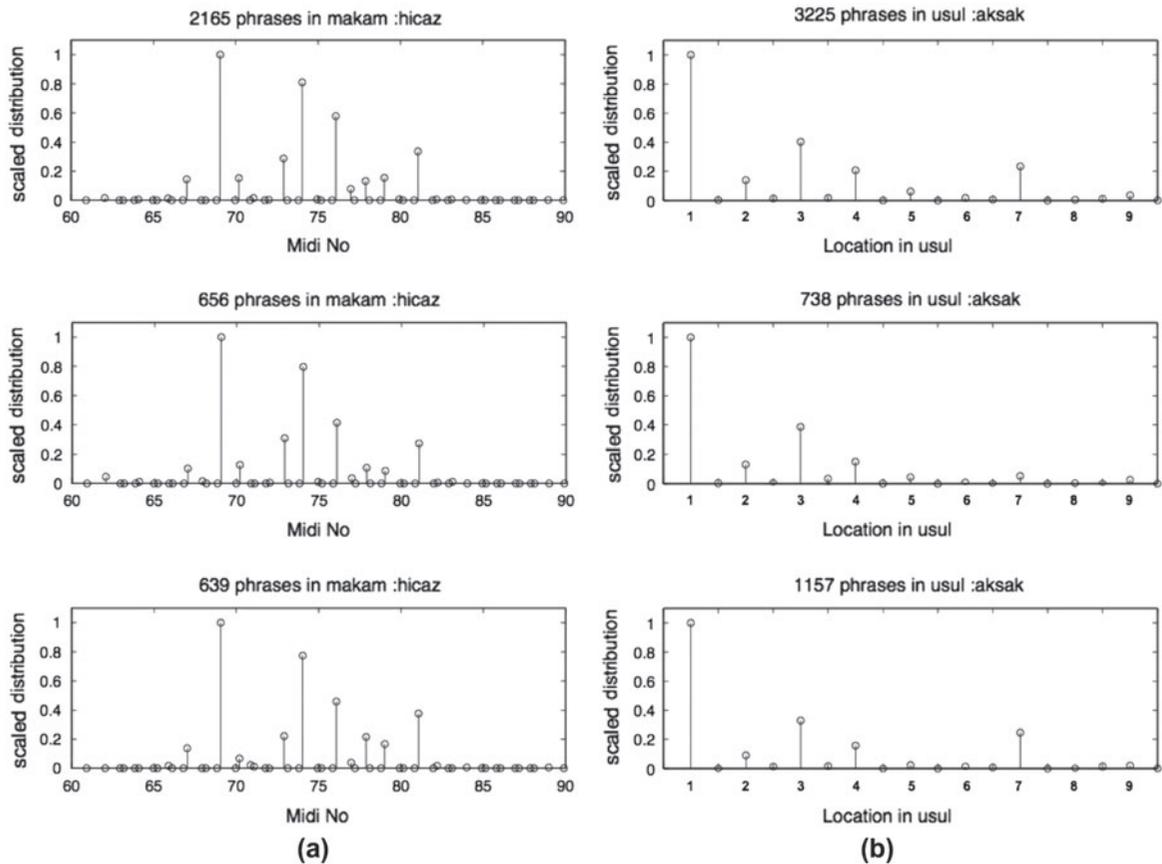


Fig. 3. Distributions computed with respect to (a) pitches of makam Hicaz and (b) 16th note onsets in the usul cycle Aksak (9/8),<sup>11</sup> from the three datasets.

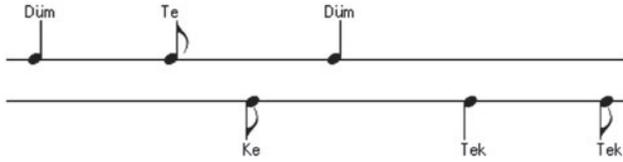


Fig. 4. Symbolic description of the usul Aksak (9/8).

frequent phrase boundary note is the *güçlü*<sup>14</sup> (D5, MIDI No: 74.04). Similar observations hold for the rest of the makams in terms of hierarchies of scale degrees and frequent phrase ending notes. This supports the claims of Ayari (2005) that the hierarchy of makam pitches and phrase boundaries are related. Another example is provided in Figure 5 for makam Hüseyinî and usul Ağıraksak (9/4)<sup>15</sup>. Again, the distributions from the different datasets are very similar for both the makam and the usul. In the makam distributions, the *karar* (A4, MIDI No: 69.06) is the most prominent pitch, followed by the *güçlü* (E5, MIDI No: 76.07) for the first and third datasets and pitch D5 (MIDI No: 74.04) for the second dataset. Considering the usul distributions, the most prominent component appears in the sixth 8th note onset (location 3.5). For this usul (which is considered to be ‘syncopated-like’), this emphasis is specifically stated in the oral culture. Additionally, at many instances in compositions using this usul, rests are observed just before this instant in the usul cycle.

The distributions depicted in Figures 3 and 5 form the basic makam-based and usul-based information from which two features are computed.

### 3.2 Existing decision functions/strength curves used as features

As stated in Section 1, most of the automatic segmentation algorithms use measures of inter-onset intervals (IOI) and pitch changes as the basic constituents of decision functions/strength curves computed as a phrase boundary probability. Then a threshold is generally applied to obtain boundary decisions. In order to study the contribution of the proposed features on the segmentation performance within the supervised learning framework, our basic system computes a feature vector (of five features) for each onset in the symbolic data, containing (i) the IOI, (ii) the ratio of the current IOI and the next IOI, (iii) the absolute pitch interval (API) at the onset, (iv) the usul-based feature and (v) the makam-based feature.

In addition, the decision functions of the following state-of-the-art systems were collected to form a set of additional features for comparison: LBDM (Cambouropoulos, 2001), TGUs

(Temporal Gestalt units) (Tenney & Polansky, 1980), Segment probabilities (Bod, 2001) as implemented in MIDI Toolbox (Eerola & Toiviainen, 2004) with the default parameters ( $W_{API} = 0.25$ ,  $W_{IOI} = 0.50$  and  $W_{OOI} = 0.25$  for LBDM) (IOI having double the weight compared to the weight of the pitch interval for TGU), and Grouper (Temperley, 2001) using the implementation by Sleator and Temperley (<ftp://ftp.cs.cmu.edu/usr/ftp/usr/sleator/melisma2003>).

Since the detailed explanations for these approaches are available in several publications, we will present here very brief descriptions of the algorithms.

TGU involves a two-step processing of a decision function. The decision function is computed as a weighted sum of IOI and pitch interval for each note. In the first step, local maxima are found by detecting samples (of the decision function) for which the sample value is higher than both the previous and the next samples’ values. The segments bounded by these maxima are named as clangs. Then a second-order decision is applied in the same manner, this time on clangs (instead of notes) to identify the segments.

LBDM computes a decision function as a weighted sum of three functions; that of IOI, pitch interval and rest duration. Each function is computed by multiplying the current value by the sum of the degree of changes from previous and next samples. The degree of change is computed as a ratio between the difference between the two values and sum of the two values. LBDM provides a continuous decision function instead of binary decisions for boundaries.

The Grouper algorithm uses a dynamic programming approach for processing the features (both IOIs and OOs (offset-to-onset intervals)) to predict the phrase boundaries using Phrase Structure Preference Rules (PSPRs). These rules specify higher preferences for large IOIs and OOs, an average number of notes for segments and segmentations at parallel points in the metrical structure.

Bod (2001) uses a data-driven approach as an alternative to TGUs. The method employs grammar-learning techniques to compute the probabilities of phrase boundaries from pitch-class, interval and duration distributions from the Essen Folk Song Collection (Schaffrath, 1994).

In order to be able to use the Midi Toolbox, the SymbTr format was directly converted to the note matrix representation of the Midi Toolbox. Minor modifications are made in order to cope with the microtonal midi information represented with fractional numbers in our data. The results of these four algorithms were merged with the five features described earlier to obtain a setup with nine features in total.

To study the contribution of our proposed (makam-based and usul-based) features, we also carried out tests for an alternative setup with seven features, leaving out these two features from the original set of nine features. In addition, the individual contributions of the makam and usul features to the segmentation task were evaluated by carrying out segmentations using these two features individually.

To summarize, the following setups are designed and tested on each dataset:

<sup>14</sup>*Güçlü* is the note that is mainly emphasized in the interlude, which often (but not always) coincides with the first note of the second chord making up the scale.

<sup>15</sup>Notation and sound example for this usul can be found on: <http://www.freesound.org/people/barisbozkurt/sounds/140545/>

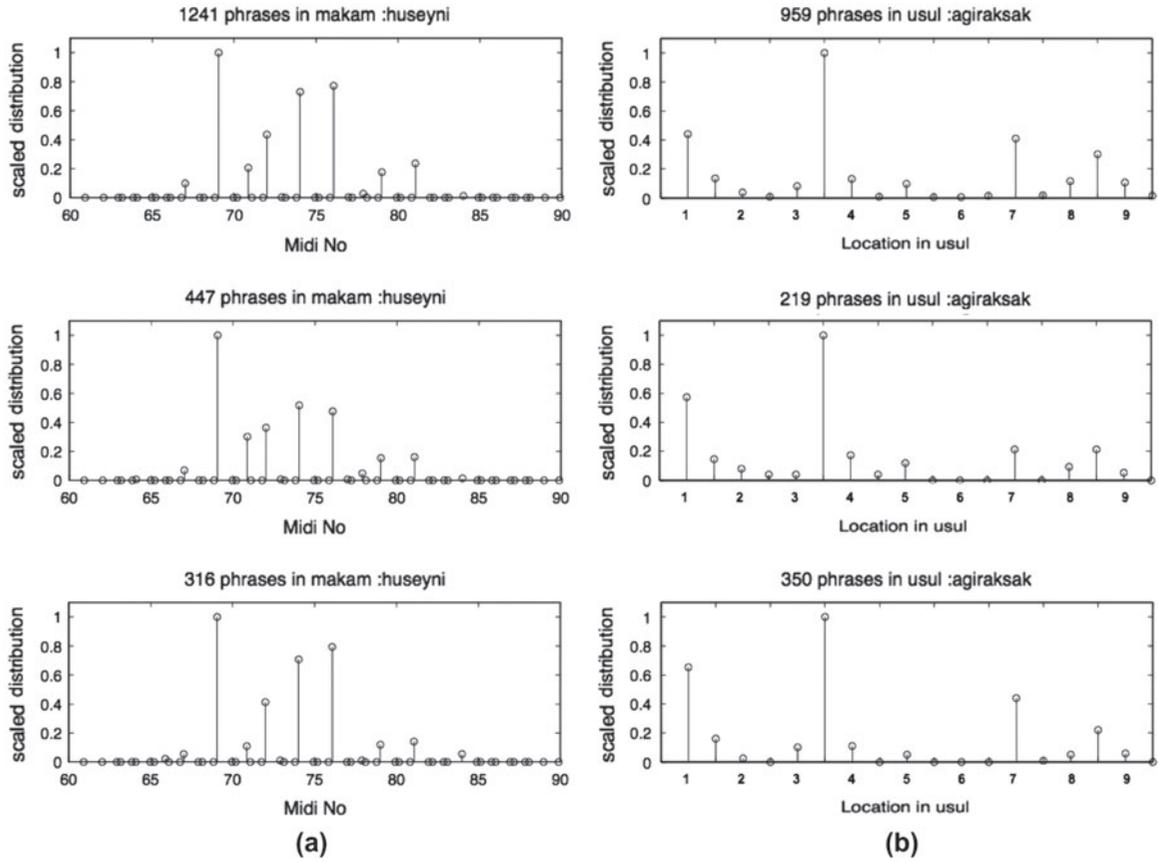


Fig. 5. Distributions computed with respect to (a) notes of makam Hüseyini and (b) each 8th note onset in the usul cycle Ağıraksak (9/4),<sup>15</sup> from the three datasets.

- (1) Our basic setup that uses IOIs, ratio of consecutive IOIs, pitch change, usul-based and makam-based features (setup using five features)
- (2) An extended setup where LBDM, TGU, Segment Probabilities and Grouper are appended to these five features (setup using nine features)
- (3) The extended setup except the two proposed (makam-based and usul-based) features (setup using seven features) to study the contribution of these two features
- (4) Makam- and usul-based features considered separately by themselves
- (5) Each of LBDM, TGU, Segment Probabilities and Grouper considered separately

#### 4. The supervised learning method used for automatic segmentation

Binary pattern classification is the quintessential pattern recognition problem studied at great length in the machine learning literature, with various algorithms such as the nearest neighbor rule (Cover & Hart, 1967; Fukunaga & Hostetler, 1975), support vector machines (Cortes & Vapnik, 1995; Vapnik, 1998), artificial neural networks (Haykin, 2008), and discriminant

functions (McLachlan, 2004). These methods can be trained to construct decision rules to separate the pattern classes of interest based on a training set of data containing samples representing the two classes.

In this work, we have used Fisher's linear discriminant analysis (Duda, Hart and Stork, 2001) to identify the phrase boundaries based on the set of features collected for each note described in the previous section. This particular choice of learning method was based on the computational expense in carrying out more sophisticated machine learning algorithms on large datasets such as the ones used here, as well as the vulnerabilities of these methods to undercut the less populated pattern class when a significant difference in the number of class representative samples are available in the training sets: the largest dataset in our study contained 168,777 notes, of which only 14,965 represented phrase boundaries.

Mathematically, given a dataset  $\{(x_i, y_i)\}$  of feature vectors  $x_i$  and their class labels  $y_i \in \{0, 1\}$ , with a label of 1 indicating a phrase boundary, for  $i = 1, 2, \dots, \ell$ , Fisher's linear discriminant is defined by the function

$$f(x) = x^T (\Sigma_0 + \Sigma_1)^{-1} (\mu_1 - \mu_0),$$

where  $\mu_0$  and  $\mu_1$  represent the sample means of the feature vectors in the two classes, and  $\Sigma_0$  and  $\Sigma_1$  their respective

covariance matrices. A decision for a note with feature vector  $x$  is formed by computing the sample means and covariance matrices from the feature vectors in the training dataset, and thresholding the discriminant function defined above by a critical value  $f_c$ , so that the note is labelled as a phrase boundary if  $f(x) \geq f_c$ . The critical value is to be set to satisfy a trade-off between false detections and missed phrase boundaries on the training data.

Note that in our implementation that followed a Leave-One-Out (LOO)-like procedure (explained in the next section), the decisions for the notes in a piece were formed by computing the sample statistics on the feature vector dataset that omitted all the notes of the piece in consideration, to ensure that the knowledge of the other notes in the same piece did not influence the decision on the note in question.

## 5. Tests and results

There exists no generally accepted method for comparative evaluation of melodic segmentation systems (Nooijer et al., 2008). We agree with Nooijer et al., 2008 and Pearce et al. 2010 that the validation should be chosen (among various approaches) that fits the target application and the nature of the data. Here, we use one of the most common measures, the F-measure defined as

$$F = 2 \frac{R^{precision} R^{recall}}{R^{precision} + R^{recall}}$$

with the precision and recall rates calculated by

$$R^{precision} = \frac{\sum_i 1('x_i \text{ was labelled as a phase boundary and } y_i = 1')}{\sum_i 1('x_i \text{ was labelled as a phrase boundary}')}$$

and

$$R^{recall} = \frac{\sum_i 1('x_i \text{ was labelled as a phrase boundary and } y_i = 1')}{\sum_i 1(y_i = 1)}$$

respectively. Note that occasionally in the literature, the terms sensitivity or the true positive rate are used to represent the fraction of segment boundaries accurately recognized as such, synonymous to the recall rate. Finally, the specificity rate was also calculated in a similar manner, as

$$R^{specificity} = \frac{\sum_i 1('x_i \text{ was not labelled as a phrase boundary and } y_i = 0')}{\sum_i 1(y_i = 0)},$$

to represent the fraction of notes correctly recognized not to be on a phrase boundary. Recall that the note with a feature vector  $x_i$  was labelled as a phrase boundary if  $f(x_i) \geq f_c$ , for some decision threshold  $f_c$ . In addition, two-by-two confusion matrices were calculated for each segmentation experiment in order to facilitate the calculation of these performance evaluation measures. Specifically, the first row of a confusion matrix recounted the number of notes that were not marked as a phrase boundary by the expert (for which  $y_i = 0$ ) correctly identified as such in the first column, and those that are

mistaken for a phrase boundary in the second column. The second row counted the notes that were marked to be on a phrase boundary by the expert ( $y_i = 1$ ) but missed by the automated segmentation procedure in the first column, and those that were correctly identified in the second column. A good segmentation, then, is marked by a greater number of notes accumulated along the diagonal where the true nature coincides with the predictions.

In order to avoid any overlap between the data used for training and testing, we used a LOO (leave-one-out)-like procedure in the segmentation experiments. The leave-one-out cross-validation procedure is a well-known technique in statistical pattern classification, where classifiers are trained on a dataset that does not include the specific pattern to be classified, eliminating the possibility of any knowledge on the pattern in question of affecting the results. In accordance with this principle, in the training phase, the piece to be segmented was taken out of the dataset and the distributions were computed from the rest of the data. Indeed, the distributions obtained by taking out different pieces were almost identical since the data size is sufficiently large. In the testing phase, automatic segmentation trained in this manner was applied only to this piece and compared with manual segmentations. These two steps were repeated separately for all pieces and the results were collected. In all tests, the critical decision threshold  $f_c$  was determined automatically, by maximizing the F-measure on the labelling obtained on the training dataset that omitted the notes of the piece in question, in full agreement with the LOO-like cross-validation procedure described above.

### 5.1 Test results for the basic setup (5 features)

In Tables 3 and 4 we present the test results considering each note as a data point where 1 corresponds to the phrase boundary class.

### 5.2 Test results for the extended setup (9 features)

In Tables 5 and 6 we present the results for the extended setup.

### 5.3 Test results for the setup without the new features (7 features)

In Tables 7 and 8 we present the results for the setup without the new features.

### 5.4 Test results for the makam and usul features alone

In order to identify how much of the predictive power of the configurations above depend on the makam and usul features, we have carried out the same analysis on these features alone. The resulting evaluation measures are presented in Table 9.

Table 3. Confusion matrices for the basic setup with five features.

	First dataset			Second dataset			Third dataset		
	0	1	Total	0	1	Total	0	1	Total
0	148016	5796	153812	55428	1646	57074	52385	2318	54703
1	6400	8565	14965	1518	1305	2823	1742	2684	4426
Total	154416	14361	168777	56946	2951	59897	54127	5002	59129

Table 4. Evaluation measures for the basic setup with five features.

	First dataset	Second dataset	Third dataset
Recall	0.5723	0.4623	0.6064
Specificity	0.9623	0.9712	0.9576
Precision	0.5964	0.4422	0.5366
F-measure	0.5841	0.4520	0.5694

Table 5. Confusion matrices for the extended setup with nine features.

	First dataset			Second dataset			Third dataset		
	0	1	Total	0	1	Total	0	1	Total
0	148421	5391	153812	55356	1718	57074	53035	1668	54703
1	6020	8945	14965	1339	1484	2823	1811	2615	4426
Total	154441	14336	168777	56695	3202	59897	54846	4283	59129

Table 6. Evaluation measures for the extended setup with nine features.

	First dataset	Second dataset	Third dataset
Recall	0.5977	0.5257	0.5908
Specificity	0.9650	0.9699	0.9695
Precision	0.6240	0.4635	0.6106
F-measure	0.6106	0.4926	0.6005

Table 7. Confusion matrices for the setup with seven features.

	First dataset			Second dataset			Third dataset		
	0	1	Total	0	1	Total	0	1	Total
0	148018	5794	153812	55503	1571	57074	52703	2000	54703
1	6644	8321	14965	1625	1198	2823	1984	2442	4426
Total	154662	14115	168777	57128	2769	59897	54687	4442	59129

Table 8. Evaluation measures for the setup with seven features.

	First dataset	Second dataset	Third dataset
Recall	0.5560	0.4244	0.5517
Specificity	0.9623	0.9725	0.9634
Precision	0.5895	0.4326	0.5498
F-measure	0.5723	0.4285	0.5507

Table 9. Evaluation measures for the makam and usul features alone (DS: dataset).

	Makam-based feature			Usul-based feature		
	DS 1	DS 2	DS 3	DS 1	DS 2	DS 3
Recall	0.5668	0.3216	0.5999	0.5997	0.6458	0.6139
Specificity	0.7114	0.8700	0.7096	0.9226	0.9067	0.9162
Precision	0.1604	0.1091	0.1432	0.4299	0.2551	0.3721
F-measure	0.2501	0.1629	0.2312	0.5008	0.3657	0.4633

Table 10. Evaluation measures for the individual features (DS: dataset).

	LBDM			TGU			Segment Prob.			Grouper		
	DS 1	DS 2	DS 3	DS 1	DS 2	DS 3	DS 1	DS 2	DS 3	DS 1	DS 2	DS 3
Recall	0.553	0.451	0.492	0.408	0.428	0.427	0.361	0.315	0.366	0.489	0.501	0.516
Specificity	0.950	0.959	0.964	0.946	0.927	0.943	0.865	0.900	0.867	0.939	0.921	0.935
Precision	0.516	0.355	0.526	0.423	0.226	0.375	0.206	0.135	0.182	0.439	0.239	0.392
F-measure	0.534	0.397	0.508	0.415	0.296	0.399	0.262	0.188	0.243	0.462	0.324	0.445

### 5.5 Test results for individual features from the literature

We present the evaluation measures obtained by the features derived from the literature in Table 10.

In a separate test, three components of LBDM, namely API, IOI and OOI, are used as features (without computing a weighted sum) and tests were repeated with these three features using the same classifier to train weights of LBDM. Interestingly, this system performed very close to LBDM with fixed default weights, and hence the results with the default weights are included in Table 10.

### 5.6 Statistical significance analysis on the F-measures

We have carried out statistical comparison tests to see if the hybrid system using nine features achieves meaningful improvements upon the existing methods. To this end, we have first compared the F-measures of the four methods from the literature, evaluated above with the average F-measures of the hybrid system across the three datasets using a paired *t*-test. The results showed that the improvement in the F-measure by the hybrid system was highly significant: The average improvement was 0.0882 on LBDM (P-value 0.0044), 0.1979 on TGU (P-value < 0.001), 0.3369 on Segment Probabilities (P-value 0.0023) and 0.1576 on Grouper (P-value 0.0014). The small P-values attest to the statistical significance of the observed amelioration in the average F-measures.

As a final test for the performance achieved by the hybrid system using nine features, we have compared the F-measures obtained for every piece in the three datasets by the hybrid system to those obtained by the four methods from the literature. The results of a paired *t*-test analysis on the piece-wise F-measures provided in Table 11 below provided a decisive performance validation for the hybrid system using nine features. The average piece-wise F-measures obtained by the hybrid system were 0.5999, 0.4728 and 0.5991 for the

three datasets, respectively. The improvements on the four alternative methods were statistically significant, with P-values ranging between  $10^{-6}$  to  $10^{-113}$ .

### 5.7 Discussions

Before starting to discuss the test results, it is important to emphasize here once more that the above presented results can only be interpreted as some measure for the potential of using decision functions of specific algorithms/systems on datasets of segmentation applied directly on scores (without listening to an interpretation) in a supervised learning framework.

We start with the observation that the basic setup using five features (Table 4) (IOI, ratio of consecutive IOIs and pitch change, usul-based and makam-based features) has a better performance than the other methods when used individually (Table 10). Among individual systems from the literature, LBDM has the highest potential followed by Grouper. These results support the results of previous comparative work such as Nooijer et al. (2008).

Comparing the performance evaluation measures in Tables 3–10, we observe that the highest performing setup is the one using all nine features. This result is in-line with the study of Pearce et al. (2010) where they tested a similar hybrid system (combining features of several approaches) and arrived at a similar conclusion. For this specific task of automatic segmentation via supervised learning on the dataset of segmented scores by experts, the following F-measures were obtained: 0.6106 (first dataset), 0.4926 (second dataset) and 0.6005 (third dataset) (Table 6). The disadvantage of using a large collection of features is that the system becomes bulky. While such systems serve as effective tools for automatizing the segmentation task, the insight brought to understanding the nature of segmentation by using such tools is minor.

Table 11. Average improvement achieved by the hybrid system using nine features on the piece-wise F-measures against the methods from the literature.

Alternative method	Dataset 1		Dataset 2		Dataset 3	
	Improvement	P-value	Improvement	P-value	Improvement	P-value
LBDM	0.0649	<0.001	0.0843	<0.001	0.0770	<0.001
TGU	0.1883	<0.001	0.1750	<0.001	0.1990	<0.001
SegmentProb.	0.3451	<0.001	0.2970	<0.001	0.3634	<0.001
Grouper	0.1341	<0.001	0.1486	<0.001	0.1454	<0.001

The contribution of the proposed makam-based and usul-based features can be studied by comparing the test results of the setup using seven features obtained by excluding these features from the original set of nine features. We observe that the new features have brought an improvement of 0.0383–0.0641 (or 6.69–14.96%) in F-measures. A complementary test evaluating the usefulness of these features revealed that while these two features alone can be used to achieve an average level of accuracy in recognizing the segment boundaries (Table 9), the performance levels achieved by collaborative configurations were markedly better (Tables 4 and 6). It is interesting to note that the usul-based feature used alone leads to a higher performance level (Table 9) than TGU, Segment Probabilities and Grouper (Table 10).

While the makam-based and usul-based features are tested within a statistical learning framework, these features could also be used as components of decision functions. For example these features could be merged with other features such as IOI and API using a weighted sum to form a new decision function. Such an approach mainly differs from the statistical learning approach by using fixed weights instead of weights estimated/learned from the data to achieve an optimum performance.

Compared to LBDM, TGU, Segment Probabilities and Grouper, our approach is computationally heavier and more complex to implement (like any other data-driven method). In addition, a database of manually segmented melodic phrases is required for training. The advantage is the adaptability of the system to a specific database.

Among the datasets included in this study, the second dataset appears to be the most difficult one to learn since the expert preferred to group the events into longer chunks. Comparing the segmentations of the second expert with the first expert (on 133 common pieces, Table 2), we observe that the Recall rate is high. This suggests that the segmentations of expert 2 are scarcer but located in relatively similar positions considering the beats in the usul cycle and the degrees of the makam scales.

## 6. Conclusion

In this work, we considered the task of automatic segmentation as an expert would perform for melodic analysis directly on scores of Turkish makam music. The most common melodic analysis performed for research and education of Turkish makam music today includes the steps of

segmenting, assigning segments to specific (tetra-penta) chords or *çeşniler* to study the context change within pieces or compositional preferences (e.g. Eroy, 2010; Gönül, 2010; Kılınçarslan, 2006). To the best of our knowledge, melodic phrase segmentation (which does not necessarily coincide with *çeşni* segmentation) has not been discussed explicitly in any text.

We have tested four of the state-of-the-art methods, namely LBDM, TGUs, Segment Probabilities and Grouper, in this specific context even though they were not designed for such a task. Our choice of methods was driven by the results of the previous comparative studies and the availability of the resources and/or implementations for academic use. The systems were tested on large datasets collected from three experts.

Since the aim in this study was to perform automatic segmentation on an existing large dataset as segmentation provided in a smaller dataset, we followed a data-driven approach. To this end, Fisher's linear discriminant analysis was used to derive the decision criterion to separate the phrase boundaries from the remaining notes. We introduced two new domain-specific features for the task; a makam-based and an usul-based feature. The test results show that the two features carry complimentary information to existing features in the literature. Their inclusion led to statistically significant improvement in automatic segmentation. During computation of these features, high degrees of consistencies were observed between the three datasets on phrase boundary distributions with respect to onsets in usul cycles and pitches in makam scales. We believe that the specific regularity of Turkish makam music phrases in these dimensions is the source of the approach's relative success.

In this study, we mainly aimed at developing segmentation tools that would facilitate musicological research for understanding the makam concept and its practice by analysing large repertoires. One such example is the SymbTr database (Karaosmanoğlu, 2012), which contains 1700 Turkish makam music scores and does not include any melodic segmentation information. We conclude that the extended system using all nine features has the potential to be used towards that end.

Here, the tests were only carried out on Turkish makam music examples. We expect that similar links between melodic segment boundaries and tonal/modal hierarchies and metrical hierarchies are present in many other tonal/modal music traditions. Therefore the approach presented here can potentially be useful for other musical repertoires. This, however, needs to be explored on data from the specific culture of interest.

## 6.1 Future work

As shown by Cambouropoulos and Tsougras (2004) it is known that the beginning and ending points of ‘significant’ repeating musical patterns influence the segmentation of a musical surface (referred to as the musical parallelism). For Tunisian music, Ayari (2005) points out the importance of ‘genre specific motivic formula’ for segmentation. The approach detailed in this paper lacks such a dimension. Currently, we are working on automatic detection of makam specific motives on the data used in this work. The incorporation of makam specific motives and repeating musical patterns into the automatic melodic segmentation will be part of our future work. We plan to use the approach proposed by Lartillot and Ayari (2009) for merging motive detection with the existing system.

This work mainly aimed at developing an automatic segmentation tool using both existing tools and additional culture specific features. Exploring the link between phrase boundaries and makam and metrical hierarchies needs dedicated studies, which could not be performed due to insufficient resources (perceptual data and information from the literature). We are currently looking for collaborations with musicologists to fill this gap. We hope that the database created in this work (which is available on request from the authors) and the results provided therein will be useful to researchers studying these phenomena.

## Funding

This work is supported by the Scientific and Technological Research Council of Turkey, TUBITAK, Grant [112E162].

## References

- Ahlbäck, S. (2004). *Melody beyond notes: A study of melody cognition*. Göteborg: Göteborg University.
- Arel, H.S. (1968). *Türk Musikisi Nazariyatı (The Theory of Turkish Music)*. ITMKD yayınları, no: 2, İstanbul: Hüsnüabiati matbaası.
- Ayari, M. (2005). *De la theorie musicale a l'art de l'improvisation: Analyse des performances et modelisation musicale*. Sampzon: Delatour France.
- Behar, C. (1998). *Aşk olmayınca meşk olmaz: Geleneksel Osmanlı/Türk müziğinde öğretim ve intikal*. İstanbul: Yapı Kredi Yayınları.
- Bod, R. (2001). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, 30(3), 27–37.
- Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational Analysis of Turkish Makam Music: Review of State-of-the-Art and Challenges. *Journal of New Music Research*, 43(1), 3–23.
- Bruderer, M. (2008). *Perception and modeling of segment boundaries in popular music* (PhD dissertation). Technische Universiteit Eindhoven, Eindhoven, The Netherlands.
- Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference* (pp. 17–22). San Francisco: ICMA.
- Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: A Computational Approach. *Music Perception*, 23(3), 249–268.
- Cambouropoulos, E. & Tsougras, C. (2004). Influence of musical similarity on melodic segmentation: Representations and algorithms. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*. Paris, France.
- Çevikoglu, T. (2007). Klasik Türk Müziğinin bugünkü sorunları. In *Proceedings of International Congress of Asian and North African Studies (Icanas 38')*. Ankara, Turkey.
- Conklin, D. (2001). Representation and discovery of multiple viewpoint patterns. In *Proceedings of the International Computer Music Conference* (pp. 479–485). Havana, Cuba.
- Cortes, C., & Vapnik, V.N. (1995). Support vector networks. *Machine Learning*, 20(1–2), 273–297.
- Cover, T.M., & Hart, P.E. (1967). Nearest neighbor pattern classification. *Information Theory*, 13(1), 21–27.
- Davis, R. (n.d.) Mode §5: Middle East and Asia: ‘maqam’, ‘makom’. *Grove Music Online*. Edited by Deane Root. <[oxfordmusiconline.com](http://oxfordmusiconline.com)> (accessed 21.04.2013)
- Duda, R.O., Hart, P.E., & Stork, D.G. (2001). *Pattern classification*. New York: Wiley-Interscience.
- Eerola, T. & Toiviainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. Kopijyva, Jyvaskyla, Finland: University of Jyvaskyla. <http://www.jyu.fi/music/coel/materials/miditoolbox>
- Eroy, O. (2010). *Tekirdağ Bölgesi Çingene Müziklerinde Kullanılan Ezgi Yapılarının İncelenmesi* (MS Thesis). Kırıkkale University, Kırıkkale, Turkey.
- Frankland, B.W., & Cohen, A.J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff’s A Generative Theory of Tonal Music. *Music Perception*, 21(4), 499–543.
- Fukunaga, K., & Hostetler, L.D. (1975). k-nearest-neighbor bayes risk estimation. *IEEE Transactions on Information Theory*, 21(3), 285–293.
- Gedik, A.C., & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, 90(4), 1049–1063.
- Ghrab, A. (2005). The western study of intervals in Arabic Music, from the eighteenth century to the Cairo congress. *World of Music*, 47(3), 55–79.
- Gönül, M. (2010). *Nevres Bey’in Ud Taksimleri Analizi ve Ud Eğitimine Yönelik Araştırmaların Oluşturulması* (PhD Thesis). Selçuk University, Konya, Turkey.
- Haykin, S. (2008). *Neural networks and learning machines* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Karaoşmanoğlu, M.K. (2012). A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of the International Society for Music Information Retrieval (ISMIR)* (pp. 223–228). Portugal.
- Kılınçarslan, H. (2006). *Dede Efendi’nin Hüzzam Mevlevi Ayininin Makam, Usul ve Ezgisel Yönden İncelenmesi* (MS thesis). Selçuk University, Konya, Turkey.

- Lartillot, O. (2007). Motivic pattern extraction in symbolic domain. In J. Shen, J. Shepard, B. Cui & L. Liu (Eds.), *Intelligent Music Information Systems: Tools and Methodologies* (pp. 236–260). Hershey, Pennsylvania: Information Science Reference.
- Lartillot, O., & Ayari, M. (2006). Motivic pattern extraction in music, and application to the study of Tunisian modal music. *South African Computer Journal*, 36, 16–28.
- Lartillot, O., & Ayari, M. (2009). Segmentation of Tunisian modal improvisation: Comparing listeners' responses with computational predictions. *Journal of New Music Research*, 38(2), 117–127.
- Lartillot, O., Yazıcı, Z.F. & Mungan, E. (2013). A pattern-expectation, non-flattening accentuation model, empirically compared with segmentation models on traditional Turkish music. In *Proceedings of the 3rd International Workshop on Folk Music Analysis*. Amsterdam.
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- MacKay, D.J.C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.
- Marcus, S. (2001). Rhythmic modes in middle-eastern music. In V. Danielson & S. Marcus (Eds.), *Garland Encyclopedia of World Music, Volume 6: Middle East* (pp. 89–92). London: Routledge.
- McLachlan, G.J. (2004). *Discriminant analysis and statistical pattern recognition* (Wiley Series in Probability and Statistics). New York: Wiley-Interscience.
- Neve, G., & Orio, N. (2005). A comparison of melodic segmentation techniques for music information retrieval. *Research and advanced technology for digital libraries. 9th European Conference, ECDL 2005, Vienna, Austria, September 18–23, 2005. Proceedings*. (Lecture Notes in Computer Science, Vol. 3652 pp. 49–56). Berlin: Springer.
- Nooijer, J. de, Wiering, F., Volk, A., & Tabachneck-Schijf, H.J.M. (2008). Cognition-based segmentation for music information retrieval systems. In *Proceedings of the Fourth Conference on Interdisciplinary Musicology (CIM08)* (pp. 399–407). Sapporo.
- Özek, E. (2011) *Türk müziğinde çeşni kavramı ve icra teori farklılıklarının bilgisayar ortamında incelenmesi* (PhD thesis). Haliç University, Turkey.
- Özkan, İ. H. (2006). *Türk Musikisi Nazariyatı ve Usulleri Kudüm Velveleleri*. Istanbul: Ötüken Neşriyat.
- Öztürk, O.M. (2011). Turkish modernisation and makam concept: Some determinations on two musical systems. In *Proceedings of the ICTM World Conference*, Canada.
- Pearce, M., Müllensiefen, D., & Wiggins, G.A. (2010). Melodic grouping in music information retrieval: New methods and applications. *Advances in music information retrieval* (Studies in Computational Intelligence, Vol. 274). Berlin: Springer.
- Pearce, M.T., & Wiggins, G.A. (2006). The information dynamics of melodic boundary detection. In *Proceedings of the Ninth International Conference on Music Perception and Cognition* (pp. 860–865). Bologna.
- Powers, H.S. & Wiering, F. (n.d.). Mode §1: The term. Grove Music Online. Edited by Deane Root. Retrieved 21 April, 2013 from <http://oxfordmusiconline.com>
- Schaffrath, H. (1994). The ESAC electronic songbooks. *Computing in Musicology*, 9, 78.
- Stubbs, F.W. (1994). *The art and science of taksim: an empirical analysis of traditional improvisation from 20th century Istanbul*. Middletown, CT: Wesleyan University.
- Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.
- Tenney, J., & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, 24, 205–241.
- Vapnik, V.N. (1998). *Statistical learning theory*. New York: Wiley.
- Yarman, O. (2007b). *79-tone tuning & theory for Turkish maqam music. Social sciences* (PhD thesis). Istanbul Technical University, Turkey.
- Zannos, I. (1990). Intonation in theory and practice of Greek and Turkish music. *Yearbook for Traditional Music*, 22, 42–59.