Segmenting Arabic modal improvisation: Comparing listeners’ responses with computer predictions

Olivier Lartillot
Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland
Olivier.Lartillot@campus.jyu.fi - http://users.jyu.fi/~lartillo/

Mondher Ayari
Department of Music, Université Marc Bloch-Strasbourg 2, France
mondher.ayari@umb.ustrasbg.fr - http://machiavel.ustrasbg.fr/musique/mondher_ayari.html

Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)

Background in cognitive ethnomusicology. Arabic music theory can benefit from new perspectives on music understanding offered by psychological and cognitive research. In particular the perception of musical structure may be explained with the help of Gestalt theory, or other theories of music segmentation and hierarchy (for instance, Lerdahl and Jackendoff, 1983). Particular questions relate to the articulation between culture and nature in music understanding and to the importance of temporal apprehension of music (Imberty, 1981). These questions can be investigated by studying how listeners from different cultures segment improvised music, guided in particular by modal transitions, modal resolutions (Bigand, 1993), and how they progressively construct in parallel a global music meaning. One particular subject of investigation deals with the cognitive processes enabling the recognition of segmentation marks and temporal integration (Ayari, 2003).

Background in computational music analysis. In the last century musicology has formulated the need for methodological development of motivic analysis (for instance, Reti, 1951). The systematic attempts undertaken in particular by linguistics and semiology (Ruwet, 1987; Nattiez, 1990) have been hampered by the underlying complexity of possible strategies and structures. Systematic descriptions of segmentation strategies offered by psychological and cognitive studies might be utilized as a means to guide and make explicit discovery processes. In addition, computational models enable exhaustive analyses of sizable musical pieces, but still struggle to control the combinatorial explosion of structures and to offer musically relevant analyses. It seems necessary in particular to unveil the different factors responsible for those divergences and to progressively build a complex system offering results congruent to listeners’ expectations. In this respect, listening experiments significantly contribute to a progressively improvement of the model.

Aims. Understanding segmentation strategies by comparing listeners’ reactions with segmentations estimated using diverse computational models.

Method. 40 European musicians (20 jazz and 20 non jazz musicians) and 20 expert Tunisian listeners heard a traditional Istikhbâr improvisation played at the Nay flute by Mohamed Saâda. After getting acquainted with the piece, listeners were asked to indicate in real time the perceived segmentations, first on a global level, a second time on a more detailed level, and a third time focusing on modal transitions. In parallel, the same piece has been transcribed and analyzed by computer based on the two heuristics of local segmentation (Tenney & Polansky, 1980; Cambouropoulos, 2006) and parallelism (Lartillot & Toiviainen, 2007). Perceived and estimated segmentations are then compared.

Results. The listeners’ segmentations exhibit fluctuations, due to the variability of listening strategies but also to the real-time configuration of the reaction measurement protocol. This second factor, since undesirable, has been reduced by grouping reactions sharing similar rationalizations (determined after thorough interviews of the subjects) and by positioning the clustered segmentation according to an expert analysis of the transcription. The comparison of the resulting segmentations with the computational analyses validates the models that explain in a more refined way the listening processes.

Conclusions. Most of the segmentations proposed by Tunisian listeners can be partly explained by the presence of local segmentations along time and pitch dimensions. Strong local boundaries consistently induce listeners’ segmentation decisions; weaker boundaries, on the contrary, cannot indicate the perception of segmentation unless they are correlated with other factors such as parallelism. The segmentation by Tunisian listeners of the Istikhbâr can be entirely reconstructed using a methodological articulation between local boundaries and parallelism. On the contrary, the segmentations inferred by European musicians offer less consistency.

Implications. The different levels of segmentations as inferred by the expert listeners can thus be explained in term of local discontinuities and parallelism. However, a more precise classification of the segmentation levels requires a taking into account of other factors such as modality in particular. Our collaboration aims at better articulating these aspects into one complex modeling in order to develop the appreciation of listeners’ segmentation strategies. The resulting model would enable detailed and systematic analyses of extensive musical pieces with the help of computer automations, in order to reveal the profuse creativity of compositional and improvisational acts.
**Motivations**

This paper presents a collaborative project articulating ethnomusicology, experimental psychology, music cognition and computer science. The objective of the study is to determine the cognitive factors explaining the understanding of Arabic modal improvisation based on one particular musical example: a Nay flute *Istikhbâr* improvisation by the Tunisian master Mohamed Saâda.

The study will be focused on one particular aspect of musical understanding: namely, *segmentation*, i.e., the spontaneous decomposition, by listeners, of the musical flows into segments.

The segmentation heuristics can be classified according to their dependency on preexisting knowledge and the abstraction of this knowledge, as well as the time span involved. Low-level features are routed in acoustic properties of the musical surface (pitch, timbre, intensity, duration) whereas high-level features rely on learned schemas. One of the main goals of this study is thus to understand the complex interaction between processing based on sensory data and processing based on learned schemas.

Schemas, in this particular study, will be restricted to motivic structures, i.e., to series of successive notes. Arabic modes are usually characterized not only by the use of particular scales, but also by the association with each scale of distinctive melodic phrases (Ayari, 2003). The recognition of these learned phrases and scales is supposed to play an important role in the understanding by expert listeners of the structure of modal improvisations.

Our study however conjectures the possibility of arriving at a similar structural understanding by means of a computational formalization that does not presuppose any preexisting schema, but on the contrary tries to induce these motivic patterns from a direct observation of the regularities in the improvisation itself.

In order to clearly establish the interrelations between the different factors explaining segmentation strategies, various heuristics are formalized and systematically applied on the musical material using computational modeling. The detailed analysis of the possible mappings between the listeners’ response and the models’ prediction enables a comprehensive explanation of the factors underlying the listeners’ understanding of modal improvisation.

**Listening Experiment**

**Choosing the piece**

The experiment is based on a type of Tunisian music in which Mohamed Saâda performs in a style perfectly in keeping with the prevailing grammar of traditional modal improvisations. We presumed that having a traditional improvisational style at our disposal would not create any problems in terms of perception and understanding for an acculturated listener. We hope that in this way we can understand how a listener who has no specific knowledge of the improvised musical culture and style perceives the structure of the piece in question.

For this experiment, we chose an *Istikhbâr* (a traditional instrumental improvisational style) played on Nay flute by Mohamed Saâda, explicitly developing the fundamental elements of the *Mhayyer Sikâ* D modal structure (see figure 1). The transcription of the improvisation, which will be used for the computational analyses, is shown in figures 5 and 6. The improvised melodies of this *Istikhbâr* are made up of melodic motifs that represent modal schemas specific to *Tba’Mhayyer Sikâ*.

![Figure 1](image-url)
Choosing the musicians

Two groups of listeners from two different cultures, European and Arab, listened to the *Istikhbâr*:

- Forty professional European musicians took part in this experiment. Twenty of them play classical, contemporary, rock and electronic music. The others are professional, trained jazz musicians who regularly perform improvised music. Improvisation is a cross-cultural aspect of musical technique: the development of musical material in jazz improvisations shows analogies with Arabic music improvisation techniques. Especially in the jazz model, where improvisation is often based on repetitions and transformations of melodic motifs and/or rhythms, playing with modes, melodic development on degrees that are considered as temporal markers: starting points, recapitulations, and melodic resolutions. Some of the European musicians had some general theoretical knowledge about improvised modal music (Indian, Turkish, and Arabic music).

- In the second part of this experiment, twenty expert Tunisian listeners from Sousse High Institute of Music listened to the piece. The musicians (instrumentalists, singers, and composers) are teachers and students (in second and third cycles of musical training) and play both popular and academic Tunisian music as well as general Arabic music on a regular basis.

Experiment Protocol

After listening a first time in order to get acquainted with the piece, musicians were asked to judge the level of familiarity with the musical style, to describe the modal behavior of the *Istikhbâr*, to guess the simple melodic schema founding the whole original *Istikhbâr* mode, and to reduce the totality of the improvisation in order to sketch the schema of its melodic development. These tasks are not considered in the presented study, but are used for subsequent research.

The part of the experiment that is central to this study is related to the determination of segmentation strategies. Three levels of segmentations are considered. In order to structure the sound material they hear in real-time, the listeners are free to choose and change at any time the segmentation criteria and the hierarchical levels they consider appropriate while listening to the first and second segmentation.

1. Listeners were therefore asked to segment the *Istikhbâr* into small pieces that are as musically coherent as possible and to indicate the places where the piece was segmented as well as, during a second listening, the instant of transition between successive musical ideas. Listeners press a key on a keyboard while describing the melodic sequence in order to record their reactions. A program is used to synchronize the audio sequences, the timer display, and the recording of the verbal response corresponding to each segmentation point.

2. The second segmentation consists of segmenting the previous sequences into small musical ideas while specifying the musical function of each one.

3. A third segmentation task is oriented toward the detection of modal transitions between the various *Iqd* genres during the development of the modal melodies. This task is not studied in this paper, as no computational model has been modeled for this particular question.

Besides, interviews with our expert listeners were carried out in order to interpret the listeners’ reactions and understand the difficulties they might have experienced.

Post-processing of the observation data

Figure 2 shows the responses of the Tunisian experts to the first segmentation task. The subjects did not all use the same criteria when segmenting the piece: some perceived broad hierarchical forms in the improvisation that they could easily describe, whereas others turned their attention hurriedly to a lower level of segmentation.

In order to better understand the subjects’ responses during this first segmentation task and to see how they organized (or, more
precisely, began to organize) the dynamic structure of the improvisation, data post-processing was performed using data clustering. The goal of this post-processing operation is to show the major trends in the listeners’ divisions and to evaluate the number of reactions made at precise moments during the improvisation.

The reactions of the subjects occasionally appeared before but more often after the articulation points, as determined by a musicological analysis of the transcription, indicated by the vertical lines in figure 2. Certain Tunisian subjects were able to recognize the playing style of the performer by identifying the dynamism in the development of melodies, which allowed them to anticipate the continuation and the conclusion of the musical discourse. On the other hand, a large number of subjects reacted late because they waited for the beginning of the following melodic movement before finally signaling the end of the sequences. They were thus in a state of uncertainty about what was coming in the music, about the resolution or the completion on a more or less stable note in the hierarchy of degrees. Using a real-time perceptive analysis, the temporal difference between the responses on the time axis is therefore of a moderate size (from 2 to 7 seconds), related to organizational complexity of the melodic movements.

For the reasons stated above, it seems important to regroup reactions that have been associated, in the listeners’ verbalization, with the same musical objects, and the resulting clusters have been reallocated to the temporal endings of those mentioned objects. The results graphic (figure 3) presents the whole Tunisian subjects’ responses, some of which have been grouped together along the time axis with respect to their significations. This graph makes it possible to look at listeners’ perception of the major forms during the first segmentation task. The major sections and the most relevant anchor points that characterize this improvisation were perceived (with a variable reaction time of 2 to 3 s on average) by a large number of subjects.

Similarly, the responses to the second segmentation task by the Tunisian subjects are shown in figure 4. Here again, some responses are regrouped and repositioned according to the musical descriptions they relate to. This analysis makes it possible to look at the organization of the internal structure of the improvisation as described by the 20 Tunisian musicians. The most relevant articulation points in the improvisation and the subsections that organize its structure were perceived at the same time by a variable number of subjects (3 to 17 subjects).

The positions of the segmentations, after clustering, and the number of listeners’ accounts related to each of these segmentation points will be used for the final mapping with the model predictions, as shown in figures 5 and 6.

Figure 2. Tunisian subjects responses for the first segmentation task. The X-axis represents the time of the piece in seconds and the Y-axis the number of subjects. The red circles group together reactions that have, according to the descriptions provided by the Tunisian listeners, similar musical “percepts”, for example: “the end of the melodic movement”, “the end of the section,” etc. The vertical lines indicate the general organization of the improvisation, as established by a musicological analysis. The resulting subparts A1, A2, etc. are indicated in the transcriptions in Figure 5 and 6.
Intercultural comparison

One aim of the experimental study was to test the impact of cultural knowledge and musical and cultural background on the perception and comprehension of a traditional Tunisian improvisation. Indeed the underlying modal structure was not known to the European subjects, although the development of musical ideas might be understood without any modal context.

The results indicate that both groups (jazz musicians and non-jazz musicians) had difficulty anticipating the progression of the melodic development. The different modulations were perceived by many listeners, particularly jazz musicians. Yet, as they could not rely on learned schemas specific to the given style, it was relatively difficult for them to be able to anticipate the musical gestures. These musicians were often surprised by the arrival of musical events such as modal transitions, thematic recapitulations, partial resolutions, etc.

One of the problems experienced by the two European groups of subjects concerned the real-time selection of criteria for segmentation, causing a lack of coherence between the reactions of the European subjects. Certain subjects had difficulty in catching the general structure, but not in perceiving the underlying musical ideas, while others, on the contrary, could more easily segment at a more general level than at a specific level.

Another difficulty was due to the absence of a regular pulse that disturbed some of the non-jazz musicians’ listening, as well as the fact that the phrases were not of the same magnitude.

For these reasons, this study will be focused on the understanding of the segmentation strategies related to Tunisians listeners only.
Computational Modeling

As explained in the introduction, segmentations processed can be broadly decomposed into two categories: those that relate to sensory appreciation of the surface level, and those that pertain to high-level structures depending on cultural knowledge. This section details modeling strategies for these two types of segmentation.

Modeling local segmentations

Local segmentations are segmentations of the temporal flux of notes at instants where relatively contrastive discontinuities along one or several musical dimensions (usually temporal distances and pitch intervals between notes) can be observed. This rule follows the Gestalt theory rules of proximity and similarity. Two rule-based models have been tried on the improvisation:

- Tenney and Polansky’s (T&P) model (1980) detects two levels of segmentation: the higher level decomposes the discourses into segments whereas the lower level decomposes each segment into “clangs”. The segmentation is based on a detection of large values within a sequence of distance between successive notes.

- Cambouropoulos’ Local Boundary Detection Model (LBDM) (2001) introduces a more complete description of local segmentation based on two rules: a Change rule and a Proximity rule, that calculate boundary strength values for each interval of a melodic surface according to the strength of local discontinuities. The resulting segmentation points are valued from 0 to 1, allowing therefore a more refined hierarchical ordering of the segments than in T&P model.

Both models have been tested using the implementations available in the MIDItoolbox (Eerola and Toiviainen, 2004).

Modeling parallelism

Another important factor of segmentation is related to the identification of repetitions, or “parallelism” (Lerdahl and Jackendoff, 1983): sequences of notes, or durations, etc., are perceived as whole entities, usually called patterns, if they are repeated several times. The pattern structure of the improvisation has been analyzed in details using a computer-based motivic pattern extraction. Several algorithms has been proposed in the literature (Lartillot, 2007). In this study, we chose to use the algorithm developed by the main author of this paper (Lartillot and Toiviainen, 2007), as it offers the following advantages:

- Closed pattern mining techniques adaptively reduce the structural redundancy and avoid combinatorial explosion, without any loss of information.

- Heuristics have been added for the taking into consideration of melodic variations such as ornamentation, which are very common in Arabic modal improvisations. With the addition of these heuristics, pattern repetition can be detected even when ornamentations erode the motivic equivalences on the surface level.

- The successive repetition of a same pattern leads to periodicity (Cambouropoulos, 1998) that, if not carefully controlled by the computational model, may leads to a combinatorial explosion of induced structures, known to be detrimental to the effectiveness of the algorithm. The concept of cyclic pattern (Lartillot, 2007; Lartillot and Toiviainen, 2007) introduced in our model enables a non-destructive control of the phenomenon, and suggests an interesting conjecture regarding the perception of motivic structures. The analysis of the Istikhbâr improvisation will show examples of cyclic patterns.

Inducing segmentation from parallelism

The motivic pattern extraction algorithm, when applied to the Istikhbâr improvisation, highlights on the transcription score the set of notes that are considered as repeated patterns. In order to apply these results to our investigation centered on the question of segmentation, the patterns need to be translated themselves into segmentation points. Studies have suggested that commencements of patterns are more likely to trigger segmentation decisions than endings, as each occurrence of a motive is
more likely prone to vary in its endings and thus make the detection of its closure more difficult (Cambouropoulos, 2006). One main problem here is that in a real-time listening context, the detection of the pattern itself is not instantaneous, as it generally requires the observation of at least several intervals. Therefore, unless the pattern was already expected, or features a very characteristic beginning, a segmentation can hardly be induced at its beginning. For this reason, in general, only pattern endings will be taken into consideration as criteria for segmentation.

In the case of cyclic patterns, where the repetitions immediately follow one each other, the segmentation can be more problematic. According to the most straightforward hypothesis, the segmentations might be expected at the borders, similarly to non-cycle pattern. But due to the concatenation of repetitions, their separation might in some cases not be perceptible any more (unless local boundaries prevail at the border).

In this study, we suggest another heuristics for segmentation induction based on 

**propagation of segmentation expectations:** during the repetition of a pattern, the segmentations found in the previous occurrence(s) are expected to appear during the ongoing occurrence as well, on the same position. The segmentations that are propagated in this way can be local boundaries, or can be induced by other patterns included in the cyclic pattern. One interesting consequence of this paradigm is that, due to the expectation, not only the endings of the included pattern, but also its starting point can be taken as possible candidates for segmentation.

**Explanatory Mapping Between the Observations and the Predictions**

Figures 5 and 6 show the analysis of the complete improvisation both by the subjects and the computational implementation of the models. The subjects’ reactions are displayed over the staves using downward triangles of two colors: green for the first broad segmentation, and yellow for the second more detailed segmentation. For each location is indicated the number of subjects who segmented at that location. As indicated above, due to the real-time context of the experiments, the temporal positions of the listeners’ segmentation present a one-second wide margin of error. As previously explained, the listeners’ justification of their segmentation has been used to relocate more precisely the segmentation points in the score.

**Impact of local segmentations**

The segmentations predicted by the models are displayed below the staves. Upward triangles indicate segmentation points based on local segmentation models. Green triangles are related to LBDM, with the associated strengths indicated below each triangle. Purple triangles result from the T&P model: besides segment decomposition, clang decomposition is also indicated by dimmed triangles, for the two first staff of the improvisation. The clang decomposition is too detailed and too frequent to offer a relevant factor for the explanation of the more sporadic listeners’ segmentations. But in the same time, some important segmentation points proposed by the listeners are related to clang boundaries instead of segment boundaries (such as the first clang boundary in the score); and reversely, some segment boundaries (such as the first one in the score) do not relate to any segmentation proposed by the listeners. Hence the T&P model cannot offer reliable explanations in our experiment.

LBDM segmentation offers better congruence with listeners’ segmentation: all segmentation points with strength equal or superior to .25 are corroborated by some of the listeners. All the other segmentations points proposed by the listeners (except one, mentioned below) can be related to LBDM segments of lower strength, but the selection of these points cannot be explained using the LBDM solely. The other segmentation factors will be used specifically to select, among the low-strength LBDM segmentations, those that can explain listeners’ segmentations.

**Impact of parallelism**

Patterns extracted by our algorithm are displayed below the staves with graduated lines; several patterns are also highlighted
directly in the staves using rounded rectangles. Lines showing only one graduation at their right ends represent acyclic patterns whose endings contributed to the listeners’ segmentation. The most typical example of this pattern is the archetypical Mhayyer Sîkâ motive, shown in red in the score. Each termination of this pattern provokes a segmentation decision by listeners, even when local boundary strength is low (under .25) or even inexistent (such as the last occurrence in part A).

In blue lines is represented a related pattern: it shows the same ending than the archetypical Mhayyer Sîkâ motive, but with a different commencement, and might suggest that listeners segmented just before the last D note of part A, i.e. before hearing an actual pause in the discourse.

The ending of staff 6 in part B can be explained by an occurrence of the archetypical Mhayyer Sîkâ motive, but it can be supported as well by the repetition of a longer phrase, highlighted by the dotted rounded rectangle, whose first occurrence appeared at the end of part A. Similarly, the ending of staff 8 can be explained by the Mhayyer Sîkâ motive and by the successive repetition of the phrase shown in previous staff, which terminated by a stronger local segmentation. Moreover, as the Mhayyer Sîkâ motive is included inside the dotted-line phrase, its starting point can also support the internal segmentation (indicated by the yellow triangle associated with 5 listeners), that was otherwise insufficiently explained by a local segmentation of strength .1.

Cyclic patterns are represented in the score by multi-graduated lines. The cyclic pattern shown in brown line can be seen both in staff 2 of part A and in staves 2 and 3 of part B. Following the proposed propagation rule, the local segmentation easily perceived in the first cycle of the pattern (with a strength of .35) and a little less accentuated in the following cycle (with a strength of .25) are also perceived by listeners in the following cycles, although the local boundary strengths are lower than .25. This formalizes a common communication principle intuitively used in music performance: once a particular musical idea has been clearly presented, its subsequent repetitions do not require the same attention both from performers and listeners (and even from the composer, when considering written music), as the structural content of the pattern has been already understood.

In part B, the local segmentation of the first phase of each occurrence of this cyclic pattern is triggered also by the ending of another pattern (dotted graduated line). The segmentation of the subsequent phases follows the propagation rule as before.

Finally, the little cyclic pattern in black at the beginning of part B leads to a segmentation that remains ambiguous, due to the absence of dominant local boundary segmentation. Hence listeners might segment either after A, or after D. This shows the difficulty of inferring segmentation using cyclic patterns in the absence of strong local boundary.

Conclusions

As the analysis has shown, the Tunisian listeners’ segmentation can be globally explained using systematic models entirely formalized and implemented in computer.

One limitation of the protocol of the current experiment is that the exact temporal location of the listeners’ segmentation cannot be set with absolute certainty. Indeed, due to the real-time context of the listening experiment, listeners may sometimes miss certain segmentation or indicate them with some delay. In order to strengthen the results and analyze the segmentation in more details, an update of the experimental protocol may be attempted, that would offer the subjects the possibility to stop and rewind the musical flow at any time and modify slightly the temporal position of the segmentation points.

Acknowledgments. This collaboration has been initiated in a French project financed by CNRS under the ACI “Complex Systems for Human and Social Sciences”, during the years 2003-2005, with Gerard Assayag (Ircam, Paris), Stephen McAdams (McGill University, Montreal) and Petri Toiviainen (University of Jyväskylä).
References


Available at http://www.jyu.fi/music/coe/materials/miditoolbox.

Figure 5. Transcription, analysis and segmentation of Mohamed Saâda’s Istikhbâr by Tunisians listeners (over the staves) and by computer (under the staves). See last section for explanations concerning the annotations. This figure shows the first part of the improvisation.
Figure 6. Second part of the analysis of Mohamed Saâda’s Istikhbâr.