Segmentation of Tunisian Modal Improvisation: Comparing Listeners' Responses with Computational Predictions

Olivier Lartillot[a]; Mondher Ayari[b]

[a] University of Jyväskylä, Finland [b] Université Marc Bloch-Strasbourg, France

PLEASE SCROLL DOWN FOR ARTICLE

Segmentation of Tunisian Modal Improvisation: Comparing Listeners' Responses with Computational Predictions

Olivier Lartillot[1] and Mondher Ayari[2]

[1]University of Jyväskylä, Finland; [2]Université Marc Bloch-Strasbourg, France

Abstract

This study is aimed at an exploration of segmentation strategies through a comparison between listeners' reactions and predictions estimated using computational models. In a listening experiment, Tunisian subjects of various degrees of expertise were asked to listen to a traditional Tunisian improvisation and to indicate in real time the perceived segmentation, first on a global level and, then, on a more detailed level. In parallel, the same piece has been transcribed and analysed by computer, based on heuristics of local discontinuity and parallelism. A detailed analysis of the possible mapping between listeners' responses and models' predictions suggests an explanation of the factors underlying the listeners' understanding of modal improvisation. Most of the segmentation positions proposed by the subjects can be partly explained by the presence of local discontinuities along time and pitch dimensions. Strong local discontinuities relate to listeners' segmentation decisions; weaker discontinuities, on the contrary, cannot explain the perception of segmentation unless they are combined with other factors such as parallelism. Segmentation by Tunisian listeners can be approximated using a model combining local discontinuity and parallelism.

1. Introduction

This paper presents an interdisciplinary project involving musicology, experimental psychology, music cognition and computer science. The objective of the study is to explore the cognitive factors underlying the structural understanding of improvised modal music using one particular musical example: a Nay flute *Istikhbâr* improvisation performed by the Tunisian master Mohamed Saâda.

The study will focus on one particular aspect of musical understanding: namely, segmentation, i.e. the spontaneous decomposition, by listeners, of the musical flow into segments. The general aim of this study is to try to clarify segmentation strategies by comparing listeners' reactions with segmentation estimated using computational models based on local segmentation factors and parallelism.

The segmentation heuristics can be classified according to their dependency on pre-existing knowledge, the abstraction of this knowledge, and the temporal extent of underlying retentive processes. Low-level features are embodied in acoustic properties of the musical surface (pitch, timbre, intensity, duration) whereas high-level features rely on learned schemata. The long-term objective of our study is to understand the complex interaction between processing based on sensory data and processing based on learned knowledge.

In the study presented in this article, schemata will not be considered in their broadest generality, but will be restricted to motifs, i.e. schemata simply defined by a succession of notes. In order to establish clearly the interrelations between the different factors explaining segmentation strategies, various heuristics are formalized and systematically applied to the musical material by means of computational modelling.

Correspondence: Olivier Lartillot, Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, P.O. Box 35, Jyväskylä 40014, Finland. E-mail: olivier.lartillot@jyu.fi

2. Background

2.1 Background in cognitive musicology

Arabic music theory can benefit from new perspectives on music understanding offered by psychological and cognitive research (Ayari & McAdams, 2003). In particular the perception of musical structure may be explained with the help of Gestalt theory, or other theories of music segmentation and hierarchy (e.g. Lerdahl & Jackendoff, 1983).

Particular questions relate to the interaction 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 musical meaning. One particular subject of investigation deals with the cognitive processes enabling the recognition of segmentation points and temporal integration (Ayari, 2005).

One of the hypotheses underlying this interdisciplinary project is that patterns are able to activate learnt schemata which themselves may affect, in return, the dynamic process of segmentation. Therefore, there is a complex interaction between the analysis of the input data and the influence of cultural knowledge in the way a given musical sequence is organized. Hence, we must observe this process from different perspectives including music analysis, cognitive ethnomusicology and computer modelling of musical pattern identification, in order to describe the real-time potential of music understanding.

2.2 Background in computational music analysis

During the last century, musicology has formulated the need for methodological explanation of motivic analysis (Reti, 1951). Systematic attempts undertaken in particular by linguistics and semiotics (Ruwet, 1987; Nattiez, 1990) have been hampered by the underlying complexity of possible strategies and structures. Formalized 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 (Cambouropoulos, 1998, 2001, 2006; Rolland, 1999; Conklin & Anagnostopoulou, 2001; Meredith et al., 2002; Bod, 2002; Lartillot, 2005; Ahlbäck, 2007) enable exhaustive analyses of sizeable musical pieces, but still struggle to control the combinatory explosion of structures and to offer musically relevant analyses. It seems necessary in particular to unveil the different factors responsible for such divergence and to build progressively a complex system offering results congruent to listeners' expectations (Lartillot, 2007). In this respect, listening experiments significantly contribute to a progressive improvement of the model.

2.3 Segmentation heuristics

Ayari (2005) proposes a non-exhaustive list of heuristics ruling the perception of segmentation when listening to formulae-based improvised music, and which can be organized into three different classes.

(a) Factors depending minimally on acculturation, such as:
- (a1) Discontinuities between auditory attributes: any significant departure, for a given musical parameter, from a domain of values with which a given stream of notes complies—departure such as a pitch leap, a change of timbre, a modification of sound level, of temporal shape of individual events, of rhythmic values, etc.—tends to imply segmentation, in line with the *Gestalt* theory principles of similarity and proximity (Lerdahl & Jackendoff, 1983; Deliège, 1987).
- (a2) Extraction and memorization of repeated patterns: particular schemes, such as sequences of pitches, rhythmic values, etc., are perceived as whole entities, usually called patterns, if they are repeated several times, developed throughout the piece, with or without variations. This corresponds to the principle of parallelism (Lerdahl & Jackendoff, 1983). A large part of musical structures highlighted by music analysis, such as motifs and metrical structures, emerge from the detected repetition of particular patterns. The status of motivic patterns, in particular, is related to their temporal spans: local texture of cells, thematic signature of motifs, stable respiration of phrases, structural punctuation of whole-part repetitions.

(b) Event stability in relation to a functional hierarchy between attributes: 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 *Maqâm*) and perceived as more stable. Ending on a less stable note creates a feeling of incompletion or of musical tension. As illustrated by the results of the study carried out by Bigand (1993), listeners tend to perceive a segmentation point when a melodic or harmonic sequence resolves into a stable note. Similar effects arise when expert listeners segment Arabic *Maqâmât* (Ayari, 2005). Due to the dependence of this type of indicator on cultural knowledge, the results are expected to vary according to listeners' cultural origins.

(c) Factors highly dependent on listeners' knowledge, related to mode and performance styles as well as sociocultural aspects:
- (c1) Modulation between modes, tonalities: *Maqâm*, on which this study is centred, is a musical style essentially determined by its modal
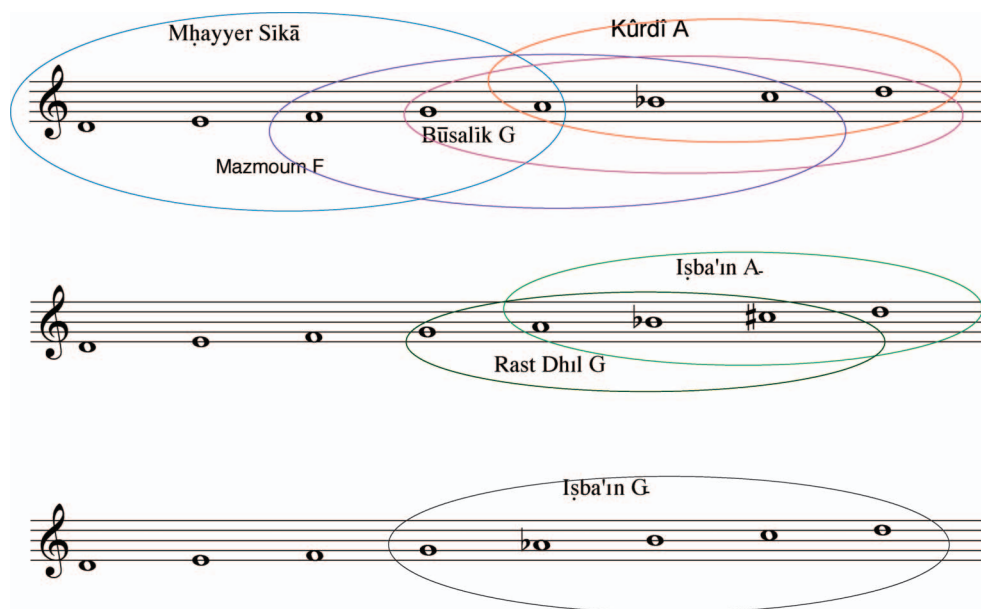
Fig. 1. The *Mhayyer Sîkâ* mode is composed of several *'Iqd-* genres, each genre partially covering a particular scale and gravitating towards a pivot note: *Mhayyer Sîkâ* genre (scale D E F G A and pivot note D), *Bûsalik*, *Râst Dhîl* and *Isba'în* (all with same pivot note G), *Mazmoum* (pivot note F), *Kûrdî* and *Isba'în* (both with pivot note A) (Lartillot & Ayari, 2006).

structure (Figure 1), which provides a scheme guiding the elaboration of improvisations in this style. This modal structure is understood by expert musicians and listeners as implicitly made up of the juxtaposition of musical *'Iqd*-genres, which are characterized not only by scales (i.e. particular set of pitches), but also by emblematic melodico-rhythmic *motivic formulae*. Figure 2 shows for instance a formula-pattern central to the *Mhayyer Sikâ 'Iqd*-genre that will be studied in our experiment. A recognition of a genre-specific motivic formula triggers a transition between genres, producing a segmentation at a relatively high level of the hierarchy. An analogue (but not identical) to this phenomenon in western music would be a change in tonality or a harmonic modulation.

- (c2) Formal and stylistic schemata: the high-level schemata discussed in this last category are linked to the musical process and to melodic introduction, exposition, development, recapitulation, and closing gestures specific to a given musical structure such as a *Taqsîm*, a sonata, a minuet, etc. These large sections, and parts of these sections, are often characterized by prominent harmonic or modal modulations that only listeners from that culture are able to follow. When listening to a piece of music within a given cultural context, expert listeners tend to partly follow a strategy based on the structural organization of the piece. The cognitive

Fig. 2. Melodico-rhythmic pattern emblematic of the *Mhayyer Sikâ 'Iqd*-genre.

schemata that are activated prompt a more focused observation, which anticipates what should happen next. These expectations can be more or less precise depending on the style of the piece, the performer, and the culture in question.

3. Listening experiment

3.1 Objectives and hypotheses

The objective of the experiment is to broaden our understanding of listeners' segmentation strategies through a comparison of listeners' segmentation decisions—collected via listening tests—with segmentation prediction generated by cognitively-grounded computational models.

Due to the complexity of the problem and the limited scope of the computational models, in this experiment we test whether, and to which degree, listeners' segmentation can be explained by heuristics based uniquely on local discontinuities and pattern repetitions. Although this reduction in the scope of the model probably implies a simplification of the study and the suppression of subtle

but important aspects, we hypothesize that the proposed reductionist focus on fundamental cognitive mechanisms could reveal interesting conclusions, as discussed at the end of the article.

3.2 Stimulus

The experiment stimulus is a traditional Tunisian *Istikhbâr* improvisation played on a traditional flute, called Nay, by Mohamed Saâda, who was a conductor, composer, and professor at Tunis University. We assume that expert listeners will have no difficulty segmenting an improvisation that follows a traditional style. The *Istikhbâr* improvisation explicitly develops the fundamental elements of the *Mhayyer Sîkâ* D mode (Figure 1). The transcription of the improvisation, which will be used for the computational analyses, is shown in Figures 7 (a) and (b), Section 4.4. The improvised melodies of this *Istikhbâr* are made up of melodic motifs that represent modal schemata specific to *Tba' Mhayyer Sîkâ*. The organization of the musical structure adheres to the scheme traditionally used; the originality of this transcribed improvisation is expressed in the particular way the melodic movements are articulated one with each other.

3.3 Listening test

Twenty expert Tunisian listeners from the High Institute of Music of Sousse, Tunisia, participated in the experiment. These musicians (instrumentalists, singers, and composers) are teachers and students, and play both traditional and modern Tunisian music, as well as Arabic music in general, Western music, jazz, etc.

In order to study the listeners' diverse processes of segmentation and recognition, a series of tasks were designed. The part of the experiment that is central to the present study relates to the determination of segmentation strategies.[1] Three levels of segmentation are considered. In order to structure the musical material in real-time, subjects were free to choose and change at any time the segmentation criteria and the hierarchical levels they considered appropriate during the first and second segmentation phases.

- Listeners were first asked to segment the *Istikhbâr* into phrases that were as musically coherent as possible and to indicate the places where the piece was segmented, as were also indicated, during a second listening, the instants of transition between successive phrases. Listeners pressed a key on a keyboard while verbally describing the improvisation. A program was used to synchronize the audio sequences, the timer display, and the recording of

the verbal response corresponding to each segmentation point.
- Listeners were then asked to segment the phrases into motifs while specifying their related musical function within the overall structure of the piece.
- A third segmentation task was oriented towards the detection of modal transitions between the various *'Iqd* genres. This task is not studied further in this paper, as no computational model has been developed for this particular topic.

Following this segmentation task, interviews with the listeners were carried out, in order to interpret the listeners' reactions, to understand the difficulties experienced by listeners during the identification, reduction, and segmentation stages, and finally to contrast different listening strategies contributing to the real-time perception of a musical work.

3.4 Analysis of the empirical data

3.4.1 First segmentation task

Figure 3 shows the responses of the Tunisian listeners to the first segmentation task. For example, subject No. 11 segmented the piece into two parts (denoted A and B and shown respectively in Figures 7 (a) and (b), Section 4.4), while subject No. 3 segmented it into 19 sections. This indicates that the Tunisian listeners did not all use the same criteria when segmenting the piece (Ayari, 2008).

In order to understand better the subjects' responses during this first segmentation task and to observe how subjects progressively organize the dynamic structure of the improvisation, responses associated with similar percepts were clustered. At the same time, a reference analysis was carried out by an expert musicologist,[2] who, contrary to the subjects of the experiments, studied the transcription beforehand and was offered the possibility to listen to the improvisation as many times as needed. We hypothesize here that both musicologist and subjects use the same set of segmentation heuristics mentioned earlier. On the other hand, the reference segmentation given by the musicologist, since not constrained by real-time limitations, is considered as more robust and more accurate temporally. Indeed, listeners' segmentation occasionally appeared before but more often after the referential articulation points indicated in the musicologist's analysis of the transcription. Some Tunisian subjects were able to follow the articulation of the performer's style, to identify the dynamism in the development of melodies and therefore 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

---

[1]For this reason, the other tasks of identification and reduction (cf. Ayari, 2008) will not be discussed here.

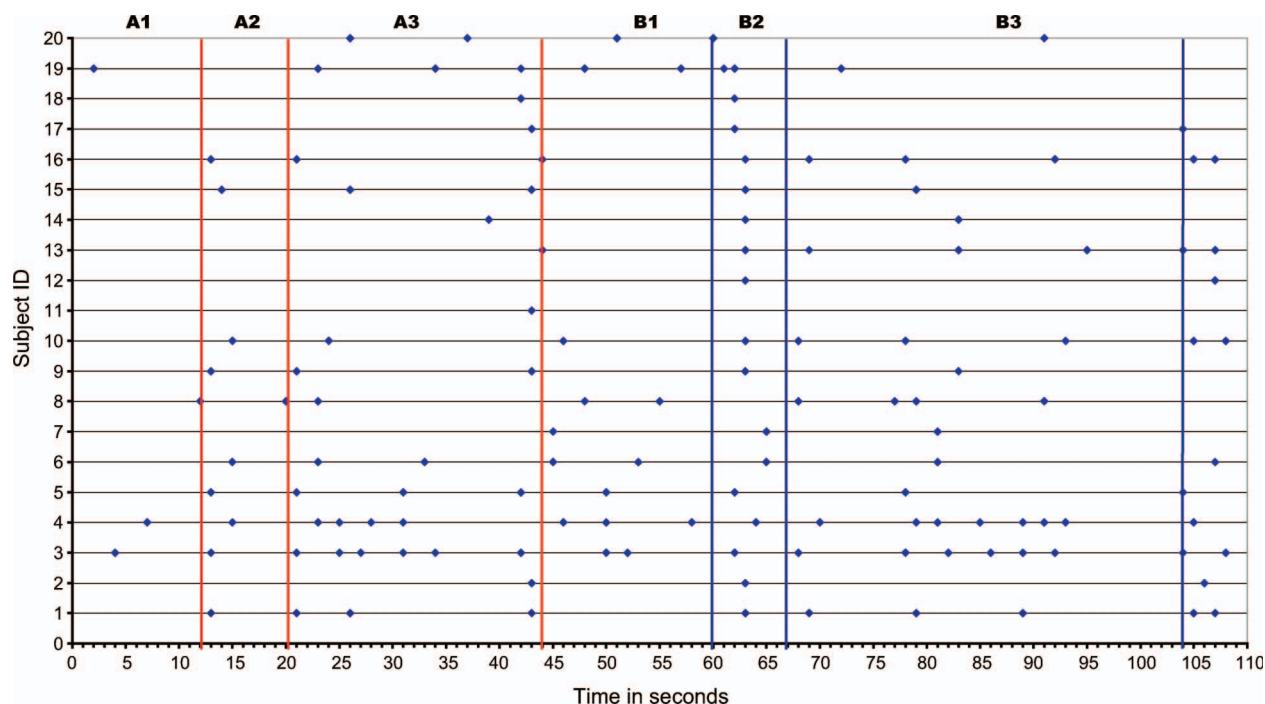[2]Namely, the second author of this article.

Fig. 3. Responses of each Tunisian subject to the first segmentation task. The *X*-axis represents the temporal progression of the piece in seconds and the *Y*-axis the subject ID number. Dots correspond to segmentation points indicated by each subject while listening to the piece in real-time. Vertical lines delimit the major sections of the improvisation according to the musicologist's analysis.

following melodic movement before signalling the end of the previous sequences, and were thus in a state of uncertainty about what was coming next in the musical discourse. In the collected data, the temporal difference between the responses on the time axis is of a moderate amount (from 2 to 7 s) and seems to vary according to the organizational complexity of the melodic movements.[3]

For the reasons stated above, it seemed important to cluster the reactions corresponding to the same perceived musical event that the Tunisian listeners mention themselves during the interview following their segmentation. In this way, all the segmentation decisions described verbally by the subjects are taken into consideration. The resulting clusters are precisely repositioned on the score at the corresponding segmentation time given by the referential analysis. Figure 4 presents the Tunisian subjects' responses, some of which have been grouped together on the time axis whenever they share similar structural functions, as verbalized by the subjects. The graph indicates that the major sections, indicated by vertical lines, 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.

---

[3]More details concerning the clustering of listeners' reactions are given in the section dealing with the analysis of the segmentation stages in Ayari (2008).

3.4.2 Second segmentation task

Similarly, the responses to the second segmentation task by the Tunisian subjects are shown in Figure 5.

Figure 6 shows general trends in the Tunisian subjects' second segmentation, indicating the number of reactions corresponding to similar musical percepts at precise moments during the improvisation. Here again, responses are regrouped and repositioned according to the musical descriptions they relate to. This analysis makes it possible to observe the organization of the internal structure of the improvisation as described by the 20 Tunisian subjects. 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 temporal positions of the segmentation points, after clustering, and the number of subjects reporting each of these segmentation points will be used for the final mapping with the model predictions, as shown in Figures 7 (a) and (b), Section 4.4.

4. Computational modelling

In this study, as a first approach, among the five different segmentation heuristics described previously, only three of them will be taken into consideration in the proposed modelling, namely:
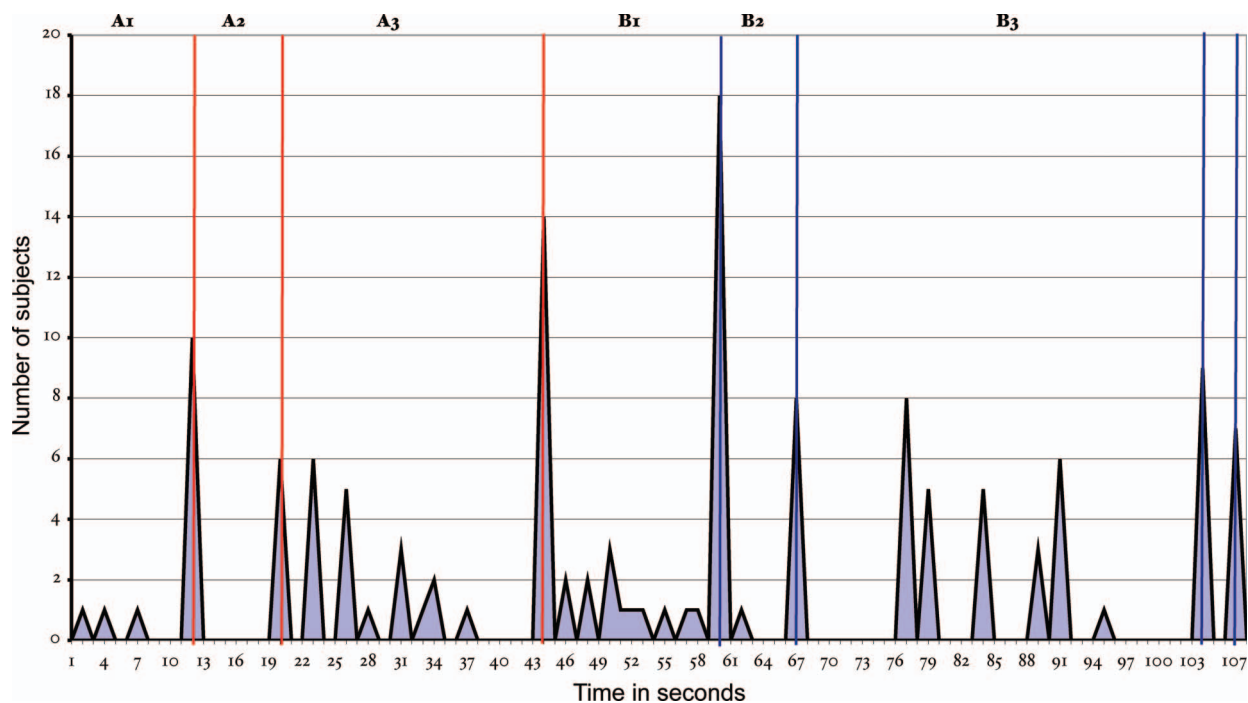
Fig. 4. Global overview of the Tunisian subjects responses for the first segmentation task. The *X*-axis represents the temporal progression of the piece in seconds and the *Y*-axis the number of subjects. Reactions have been clustered into underlying musical percepts, according to the descriptions provided by the Tunisian listeners, such as: 'end of melodic movement', 'section ending', etc. Vertical lines delimit the major sections of the improvisation according to the musicologist's analysis.
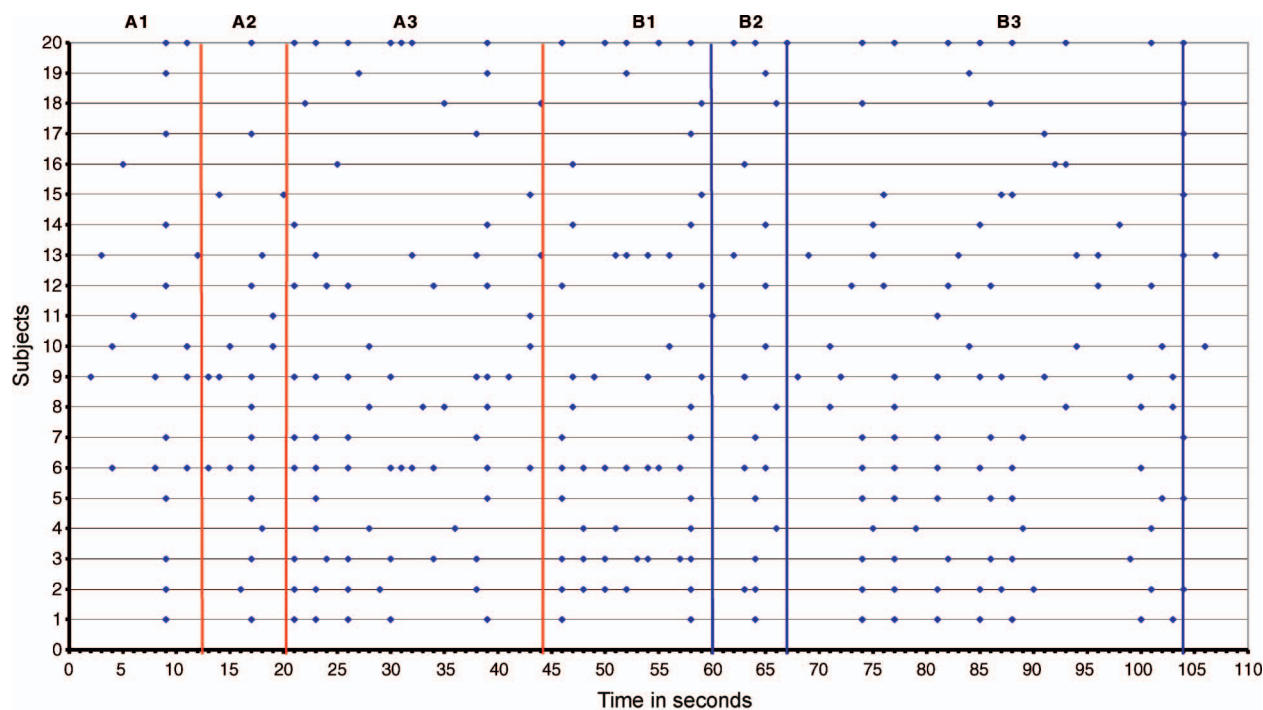

Fig. 5. Reactions of each Tunisian subject during the second segmentation. See Figure 3 for a description of the axes.

● Discontinuities between auditory attributes (heuristic a1): this relates to sensory appreciation of the surface level.

● Parallelism (heuristic a2): also stemming from the surface level, but requiring memorization processes.
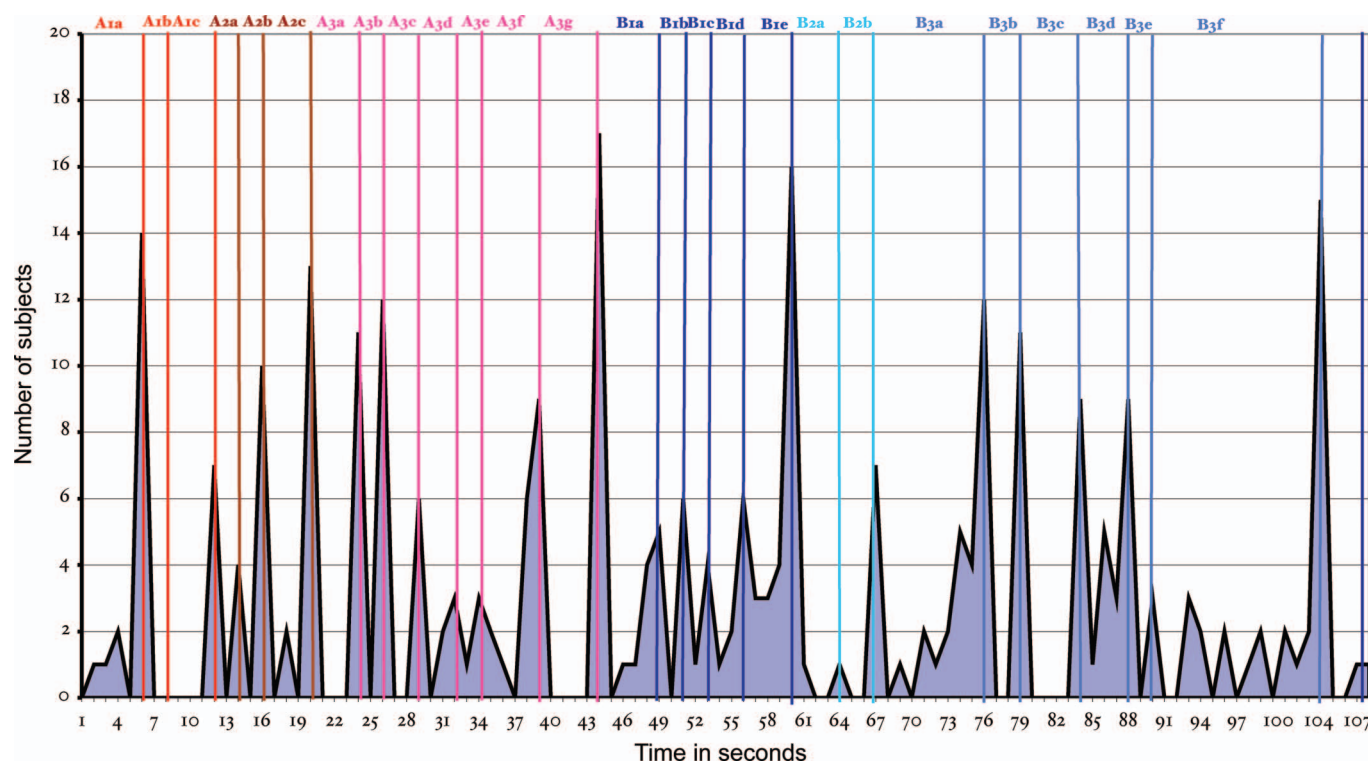
Fig. 6. Global overview of the Tunisian subjects response during the second segmentation, after data clustering. See Figure 4 for a description of the axes.

- Specific melodic-rhythmic patterns related to modal *'Iqd*-genres (heuristic c1): although pertaining to high-level structures depending on cultural knowledge, we will hypothesize in this experiment that they can be inferred directly from the piece itself via the parallelism heuristic.

This section details modelling strategies for these types of segmentation.

4.1 Modelling local segmentation

As explained in Section 2.3, local segmentation is based on relatively contrastive discontinuities along one or several musical dimensions (usually temporal distances and pitch intervals between notes). This rule follows the *Gestalt* theory principles of proximity and similarity. The computational model employed in this study is Cambouropoulos' (2001) *Local Boundary Detection Model* (LBDM), which is based on two rules, a *Change* rule and a *Proximity* rule, that calculate boundary strength values, for each interval of a melodic surface, indicating the relative importance of each successive local discontinuity. The resulting segmentation points are valued from 0 (absolutely continuous) to 1 (extremely discontinuous). The LBDM has been chosen as it has been shown to perform rather well (de Nooijer et al., 2009). For comparison, an analysis using Tenney and Polansky's

(1980) model was also attempted, without offering any significant corroboration with the experimental data (Lartillot & Ayari, 2008). Both LBDM and Tenney and Polansky's models have been tested using the implementations available in the *MIDItoolbox* (Eerola & Toiviainen, 2004).

4.2 Modelling parallelism

Concerning parallelism, the motivic structure of the improvisation has been analysed in details through computer-based motivic pattern extraction. Several algorithms have been proposed in the literature (see overview in Lartillot, 2005, 2007). In order to reduce the combinatory explosion of results produced at the pattern extraction phase, most algorithms employ filtering heuristics that select a sub-class of the resulting patterns based on global criteria such as pattern length, pattern frequency (within a piece or among different pieces), and so on. The main limitation of this method comes from the lack of selectivity of these global criteria. For instance, by selecting longest patterns, one may discard short motifs that listeners may nevertheless consider as highly relevant. And conversely patterns repeated even only twice may be perceived if these repetitions are sufficiently close in time, such that the first occurrence is still kept in the short term memory—and therefore easily recalled—while the second occurrence is heard.

The model developed by the first author of this paper (Lartillot & Toiviainen, 2007) is primarily aimed at discovering the origins of the combinatorial explosion hampering the efficiency of computational models, and at building as simple a model as possible that may be able to closely mimic the listeners' structural perception. We propose heuristics ensuring a compact representation of the pattern configurations without loss of information, thanks to an adaptive and lossless selection of most specific descriptions:

- *Closed pattern* mining heuristic (Pasquier et al., 1999), which only selects patterns whose *support* (i.e. number of occurrences in the score) is higher than the support of the pattern in which they are *included*—such as suffixes or prefixes, for instance—adaptively reduces the structural redundancy and prevents 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, where secondary notes can be added in the neighbourhood of primary notes, in both time and pitch dimensions. We are developing algorithms that automatically discover, from the raw surface level of musical sequences, musical transformations revealing the sequence of pivotal notes forming the deep structure of these sequences. With the addition of these heuristics, pattern repetition can be detected even when ornamentation erodes the motivic equivalence on the surface level.
- Combinatorial explosion can be caused by another common phenomenon provoked by successive repetitions of a single pattern. As each occurrence is followed by the beginning of a new occurrence, each pattern can be extended by a description that is identical to the first description of the pattern. This extension can be prolonged recursively, leading to a combinatorial explosion of patterns that are not perceived due to their complex intertwining (Cambouropoulos, 1998) and is known to be detrimental to the effectiveness of the algorithm. The concept of *cyclic pattern* (Lartillot, 2007; Lartillot & Toiviainen, 2007) introduced in our model enables a non-destructive control of the phenomenon, and suggests an interesting insight regarding the perception of motivic structures: because the last state of each occurrence of the repeated pattern is synchronized with the first state of the following occurrence, listeners tend to fuse these two states, and to perceive a loop from the last state to the first state, leading to a cyclic pattern. This cycle-based modelling seems to explain a common listening strategy, and resolves the problem of combinatorial redundancy. The analysis of the *Istikhbâr* improvisation includes examples of cyclic patterns, as we will see in Section 4.4.

4.3 Inducing segmentation from parallelism

The motivic pattern extraction algorithm, when applied to the *Istikhbâr* improvisation, highlights the set of notes that are considered as repeated patterns (see the transcription score in Figures 7 (a) and (b), Section 4.4).[4] In order to apply these results to our investigation centred on the question of segmentation, the patterns need to be translated into segmentation points. Studies have suggested that beginnings of patterns are more likely to trigger segmentation decisions than endings, since endings are generally more affected by melodic variation, resulting in a more difficult detection of their termination (Cambouropoulos, 2006). One main problem here is that in a real-time listening context, the detection of patterns themselves is not instantaneous, as it generally requires the observation of at least some of the musical components characterizing the pattern. Therefore, unless the pattern is already expected, or features a very characteristic beginning, a segmentation can hardly be induced at its beginning.[5] For this reason, in this study, only pattern endings will be taken into consideration as criteria for segmentation.

In the case of cyclic patterns, where the repetitions immediately follow each other, segmentation can be more problematic. According to the most straightforward hypothesis, segmentation might be expected at borders, similarly to acyclic patterns. But due to the concatenation of repetitions, those borders might not in some cases be perceptible any more (unless local boundaries prevail at those points).

In this study, we suggest another heuristic for segmentation induction based on *propagation of segmentation expectations*: during the repetition of a pattern, segmentation detected in previous occurrence(s) is expected to appear during the ongoing occurrence as well, in the same phase. Segmentation that is propagated in this way relies on local boundaries, or is induced by other patterns included in the cyclic pattern. One interesting consequence of this paradigm is that, due to the implied expectation, not only the endings of the included pattern, but also the starting points can be chosen as possible candidates for segmentation.

---

[4] A first version of the computational analysis was presented in Lartillot and Ayari (2006), but without actual comparison with the listeners' segmentation.

[5] In the special case of patterns that are sufficiently long and salient to signal segmentation of their own, listeners might decide to indicate a segmentation even when the new occurrence has been detected after some delay. These types of patterns, which do not seem to play a major role in the piece of music studied in this article, should nevertheless be considered in future work.

4.4 Explanatory mapping between observations and predictions

Figures 7 (a) and (b) show the analysis of the complete improvisation both by the subjects and the computational implementation of the models. The subjects' responses are displayed above the staves using downward triangles of two colours: black for the first higher-level segmentation, and white for the second more detailed segmentation. For each location the number of subjects who segmented at that particular location is indicated. As mentioned above, due to the real-time context of the experiment, segmentation points have been relocated based on the listeners' own justification of their segmentation choices.

4.4.1 Impact of local segmentation

Local segmentation predicted by the LBDM model is displayed below the staves with upward triangles, and associated strengths are indicated below each triangle. LBDM segmentation offers good congruence with listeners' segmentation: all segmentation points with strength equal or superior to 0.25 are corroborated by some of the listeners. All the other segmentation points proposed by the subjects (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 alone. Other segmentation factors will be used to select, among the low-strength LBDM segmentation points, those that can explain listeners' choices.

4.4.2 Impact of parallelism

Patterns extracted by our algorithm are displayed below the staves with horizontal lines; vertical marks indicate the implied segmentation points, and letters from *a* to *e* denote the corresponding motivic classes, as described below. Several phrase-like patterns, noted I, II and III, are circled on the staves.

The termination of acyclic patterns, indicated by lines under the staves showing only one vertical mark at their right ends, contributed to listeners' segmentation. The most typical example of such patterns is the archetypical *Mhayyer Sîkâ* motif (pattern *a*, cf. also Figure 2). Each termination of this pattern provokes a segmentation decision by listeners, even when local boundary strength is low (under 0.25). An extreme case is shown at the end

Fig. 7. (a) Transcription, analysis and segmentation of Mohamed Saâda's *Istikhbâr* by Tunisian listeners (over the staves) and by computer (under the staves). See Section 4 for an explanation of the annotations. This figure shows the first part of the improvisation. (b) Second part of the analysis of Mohamed Saâda's *Istikhbâr*.

of part A: one occurrence of the motif is followed by a repetition of the last pitch, but with a particularly long duration (a few seconds), indicating an important punctuation of the improvisation. A segmentation purely based on local boundary would not infer any segmentation between the end of the motif and the final punctuation. However many listeners perceived this segmentation, because they were expecting a segmentation immediately at the end of the motif.

Pattern *b* shows the same ending as in the archetypical *Mhayyer Sîkâ* motif, 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 stave 6 in part B features not only an occurrence of the archetypical *Mhayyer Sîkâ* motif, but also a repetition of a longer phrase (phrase *II*), whose first occurrence appeared at the end of part A. Similarly, the ending of stave 8 includes both an occurrence of the *Mhayyer Sîkâ* motif and the successive repetition of another phrase (*III*). Moreover, as the *Mhayyer Sîkâ* motif is an implicit component of phrase *III*, its expected starting point can also support the internal segmentation (indicated by the downward triangle associated with 5 listeners), that was otherwise insufficiently explained by a local segmentation of strength 0.1.

Cyclic patterns are represented in the score by lines with multiple vertical marks. Cyclic pattern *c* can be seen both in stave 2 of part A and in staves 2 and 3 of part B. Following the propagation rule introduced in Section 4.3, the local segmentation easily perceived in the first cycle of the pattern (with a strength of 0.35) and a little less accentuated in the following cycle (with a strength of 0.25) is also perceived by listeners in the following cycles, despite a significant drop in local boundary strength. In part B, the local segmentation of the first phase of each occurrence of this cyclic pattern *c* is also triggered by the termination of another pattern (pattern *f*). The segmentation of the subsequent phases follows the propagation rule as before.

Finally, the short cyclic pattern *e* 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 boundaries.

5. Discussion

As the analysis has shown, the Tunisian listeners' segmentation of the *Istikhbâr* improvisation can be explained to a large extent using systematic models formalized and implemented on the computer. Most segmentation points inferred by Tunisian listeners can be partly explained by the presence of local discontinuity

along time and pitch dimensions. Strong local boundaries can be easily related to listeners' segmentation decisions; weaker boundaries, on the contrary, are not sufficient to explain the perception of segmentation unless they are supported by other factors such as parallelism. Most segmentation from Tunisian subjects of the *Istikhbâr* improvisation can be explained in terms of interdependencies between local boundaries and parallelism, that we proposed to formalize using a heuristic of propagation of segmentation expectations. This interaction explains a common communication principle intuitively used in music performance: once a particular musical idea has been clearly presented (such as the first repetitions of motif *c*), 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 introduced and understood.

The different levels of segmentation as inferred by expert Tunisian listeners can thus be explained using a model that was not initially designed for this type of music, but is based on general cognitive mechanisms such as local discontinuities and parallelism. However, a more precise classification of segmentation levels requires taking into account other cultural factors influencing segmentation: functional hierarchy between attributes and the implied event stability, the overall modal structure (as illustrated in Figure 1), featuring both scale and phrase components, and higher-level formal and stylistic schemata. The computational modelling of the complete segmentation process would suggest a systematic assessment of the model based on the congruence with the listeners' responses, offering hence a new conceptual framework to Arabic musicology.

The final objective of the project in the future is thus to determine the influence of cultural knowledge on the perception and understanding of a piece of music, and the role of universal cognitive processes. Our aim is to better articulate these aspects through the design of a complex model integrating both cognitive processes and cultural knowledge, in order to develop the appreciation of segmentation strategies by listeners of various cultural backgrounds (Arabic, European, etc.). We expect that a careful modelling of this cultural knowledge will enable us to differentiate precisely between the response of Tunisian and non-Tunisian musicians.[6] The resulting

---

[6]An extension of this study (Ayari, 2008) shows in particular that segmentation inferred by European musicians offers less consistency, although Jazz musicians seem to be able to grasp major aspects of the phrasal and modal construction of the improvisation. Further experimental data have been collected, in order to test precisely the degree of pertinence of the cognitive schemes used by Tunisian listeners when analysing in real-time modern improvisations based on contemporary styles.

model would enable detailed and systematic analyses of extensive musical pieces with the help of computer automation, in order to reveal the creativity of compositional and improvisational acts.

Acknowledgements

References

Ahlbäck, S. (2007). Melodic similarity as a determinant of melody structure. *Musicae Scientiae, Discussion Forum*, *4A*, 235–280.

Ayari, M. (2005). *De la théorie musicale à l'art de l'improvisation: Analyse des performances et modélisation musicale*. Sampzon: Delatour France.

Ayari, M. (2008). Performance and musical perception analysis. *Intellectica*, (48–49), 51–80.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsim). *Music Perception*, *21*(2), 159–216.

Bigand, E. (1993). Contributions of music to research on human auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 231–273). Oxford: Oxford University Press.

Bod, R. (2002). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, *31*(1), 27–36.

Cambouropoulos, E. (1998). *Towards a general computational theory of musical structure* (Doctoral dissertation). University of Edinburgh, UK. Available at http://users.auth.gr/~emilios

Cambouropoulos E. (2001, September 18–22). The Local Boundary Detection Model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference (ICMC'2001)*, Havana, Cuba.

Cambouropoulos E. (2006). Musical parallelism and melodic segmentation: A computational approach. *Music Perception*, *23*(3), 249–269.

Conklin, D., & Anagnostopoulou, C. (2001, September 18–22). Representation and discovery of multiple viewpoint patterns. In *Proceedings of the International Computer Music Conference (ICMC'2001)*, Havana, Cuba.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, *4*(4), 325–350.

Eerola, T., & Toiviainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. Kopijyvä, Jyväskylä, Finland: University of Jyväskylä. Available from http://www.jyu.fi/music/coe/materials/miditoolbox

Imberty, M. (1981). *Les Écritures du Temps: Sémantique Psychologique de la Musique*. Paris: Dunod.

Lartillot, O. (2005). Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering. *Journal of New Music Research*, *34*(4), 375–393.

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. (2008, July 2–6). Segmenting Arabic modal improvisation: Comparing listeners' responses with computer predictions. In *Proceedings of the Conference on Interdisciplinary Musicology*, Thessaloniki, Greece. Retrieved from http://cim08.web.auth.gr/cim08_papers/Lartillot-Ayari/Lartillot_Ayari.pdf

Lartillot, O., & Toiviainen, P. (2007). Motivic matching strategies for automated pattern extraction. *Musicae Scientiae, Discussion Forum*, *4A*, 281–314.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Meredith, D., Lemström, K., & Wiggins, G. (2002). Algorithms for discovering repeated patterns in multi-dimensional representations of polyphonic music. *Journal of New Music Research*, *31*(4), 321–345.

Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*. Princeton, NJ: Princeton University Press.

Nooijer, J. de, Wiering, F., Volk, A., & Tabachneck-Schijf, H.J.M. (2008, July 2–6). Cognition-based segmentation for music information retrieval systems. In C. Tsougras & R. Parncutt (Eds.), *Proceedings of the Fourth Conference on Interdisciplinary Musicology (CIM08)*. Retrieved from http://cim08.web.auth.gr/cim08_papers/Nooijer-Wiering/Nooijer-Wiering.pdf

Pasquier, N., Bastide, Y., Taouil, R., & Lakhal, L. (1999). Efficient mining of associative rules using closed itemset lattices. *Information Systems*, *24*(1), 25–46.

Reti, R. (1951). *The Thematic Process in Music*. Basingstoke: Macmillan.

Rolland, P.-Y. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, *28*(4), 334–350.

Ruwet, N. (1987). Methods of analysis in musicology. *Music Analysis*, *1*(2), 11–36.

Tenney, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory*, *24*, 205–241.