journal of interdisciplinary music studies
season 2011, volume 5, issue 1, art. #11050105, pp. 85-100

Cultural impact in listeners' structural understanding of a Tunisian traditional modal improvisation, studied with the help of computational models

Olivier Lartillot[1] and Mondher Ayari[2]

[1] Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä
[2] Department of Music, University of Strasbourg, France

Background in cognitive and computational research in music segmentation. Theoretical models have been developed with the view to describe how listeners segment music into small chunks. Methodical tests developed in experimental psychology enable us to validate the multiple factors involved in such processes and to estimate the weight of underlying parameters. Computational modelling offers a way to test and develop those models in a more intensive and extensive setting. In all these contexts, the impact of cultural knowledge on segmentation was not studied so far.
Background in intercultural music cognition. Patterns activate learnt schemata, which affect, in turn, the dynamic process of segmentation through the activation of expectations. There is hence a complex interaction between bottom-up analysis of input data and top-down influence of cultural knowledge.
Aims. This study aims to shed light on the complex interdependencies between cognitive mechanisms and cultural background in listeners' structural understanding of music, with the help of an extended computational model.
Main contribution. Tunisian and European musicians analysed the segmentation structure of a traditional Tunisian modal improvisation (*Istikhbâr*) performed on the Nay flute by the late Tunisian master Mohamed Saâda. They signalled segmentation points while listening to a recording of the improvisation and verbally indicated the heuristics guiding their decisions at the same time. Listeners' segmentation decisions based on similar heuristics were clustered across participants.
A computer implementation of low-level heuristics of local discontinuity and parallelism showed that strong segmentation points predicted by the algorithm were generally associated with consensual segmentation points proposed by listeners from both cultures.
The impact of cultural knowledge on the segmentation behaviour was studied by modelling knowledge of Arabic modal structure. The model based on perceptual rules pointed out the most pronounced discontinuities that were consensually detected by most listeners. The integration of cultural knowledge revealed subtler articulation points in the discourse, while predicting more precisely at the same time the heuristics responsible for each of those points.
Implications. The cultural knowledge that has been modelled is based on a set of general mechanisms, such as scales, sets of notes – whatever their specific actualisation in a given culture – and numeric "activation" values associated with each different candidate concept. Those general mechanisms can be used to describe culture specific building blocks that could be reused for the description of the musical knowledge of other cultures as well.

Keywords: cognition, culture, computational model, segmentation, modes, maqam, patterns, improvisation, heuristics.

•*Correspondence:* Olivier Lartillot, PO Box 35(M), 40014 Jyväskylä, Finland;
tel: 358 40 567 7939, fax: 358 14 260 1331, e-mail: olartillot@gmail.com
• *Received:* 22 October 2010; *Revised*: 22 June 2011; *Accepted:* 20 July 2011
• *Available online:* 30 July 2011
• doi: 10.4407/jims.2011.07.005

Introduction

Psychological and cognitive research has offered new perspectives on music understanding including the perception of musical structure and segmentation (for instance, Lerdahl and Jackendoff, 1983). Particular questions raised relate to the relative contribution of culture and nature to music understanding in general and temporal apprehension of music in particular (Imberty, 1981). Ayari's (2008) study on intercultural perception advances the idea that patterns detected in real time in music activate learnt schemata, leading to the development of top-down expectations. In other words, the organization by listeners of music would be based on a complex interaction between on the one hand low-level perceptual processes related to sensory processing of information – founded on Gestalt grouping principles in particular – and on the other hand, culture specific knowledge – including collective memory, specific knowledge, social praxis, knowledge of particular musical styles. We propose to observe the complex interdependencies between cognitive mechanisms and cultural background through the prism of computational modelling, in order to describe the listeners' ability to segment and process music in real time. Our research focuses particularly on the question of segmentation and identification of musical structures.

Several studies have proposed principles of segmentation in music (Tenney and Polansky, 1980; Narmour, 1990; Lerdahl and Jackendoff, 1983). The perceptual validity of such cognitive models can be evaluated with the help of experimental psychology, through a systematic comparison with actual listeners' responses (Deliège, 1987; Clarke and Krumhansl, 1990). On top of that, computer science enables a detailed formalization of models as well as a systematic test of their theoretical implications (Frankland and Cohen, 2004; Temperley, 2001; Cambouropoulos, 2006; Bod, 2001; Pearce, Müllensiefen and Wiggins, 2010). As a productive articulation of these two scientific domains, psychological validation of computer models (Bod, 2001; Melucci and Orio, 2002; Thom, Spevak and Höthker, 2002; de Nooijer et al., 2008; Bruderer, McKinney and Kohlrausch, 2009; Pearce, Müllensiefen and Wiggins, 2010) enables to study in detail the validity of the different components of complex models. We adopt the same paradigm. Additionally, we take cultural influences into account, through a cross-cultural articulation of the psychological experiments.

This study focuses on segmentation of traditional Tunisian *maqām* music. The experiment presented in this paper uses a two-minute long *Istikhbâr* (a traditional instrumental improvisation), performed by the late Tunisian *Nay* master Mohamed Saâda, who developed the fundamental elements of the *Mhayyer Sîkâ* D *maqām* mode. Throughout the piece, starting even before the actual performance of the soloist musician, a pedal tone is constantly played, indicating the main pitch D.

In (Lartillot and Ayari, 2009), we showed how a computational model mainly focused on perceptual rules, and applied to this same piece, pointed out the most pronounced discontinuities that were consensually detected by most listeners. In this paper, we study further the impact of cultural knowledge on the segmentation behaviour with the help of a computational modelling of Arabic modality. We will show that the integration of cultural knowledge reveals subtler articulation points in the discourse, while predicting more precisely at the same time the heuristics responsible for each of those points.

Method

Listening test

Participants. All participants are musicians, and grouped into three sets of twenty persons:
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.
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. Some of the European musicians had some general theoretical knowledge about improvised modal music (Indian, Turkish, and Arabic music).

Protocol. The individual listening strategies followed by those expert musicians from various cultures were explored with the help of an original experimental protocol, where segmentation decisions were recorded while participants listened continuously to the piece, without the visual support of a score representation. More precisely, the protocol consisted of a series of successive steps:
A first complete listening of the improvisation, where participants were invited to identify the musical material and to recognize the main mode and the modulations. This was followed by three segmentation tasks in an experimental setting, where listeners indicated their segmentation decisions in real-time. They indicated segmentation points by pressing a specified key on a MIDI keyboard, while giving verbal descriptions at the same time.

- In the first task, listeners were asked to segment the piece into phrases that were as musically coherent as possible
- In the second task, listeners were asked to segment the previous phrases into smaller musical ideas and to specify their related musical functions
- In the third – more oriented – task, listeners were asked to segment the improvisation in terms of modal variations: they had to locate transitions between *ajnas*[1] (plural of *jins*) throughout the modal development

The listening test was followed by a melodic reduction task (not discussed further in this paper) and an interview.

Pre-processing. Responses associated with similar musical events were initially temporally scattered on a time interval of 2 to 3 seconds on average, due to the variable delay between what the listeners perceived and their real-time segmentation decisions. We sought to better understand the participants' responses during the segmentation tasks and to observe how participants progressively organise the dynamic structure of the improvisation. In that aim, reactions corresponding to a same perceived musical event were clustered and associated with that musical event, as mentioned by the listeners during the open discussion.[2] In this way, all segmentation decisions verbally described by the participants were taken into consideration. The resulting clusters were precisely repositioned in the score at the corresponding segmentation time given by a referential analysis, carried out by an expert musicologist not constrained by real-time limitations.[3] As an illustration, Figure 1 presents the Tunisian participants' responses after clustering. The graph shows that major sections, indicated by vertical lines, and relevant anchor points within this improvisation were perceived by a large number of participants.

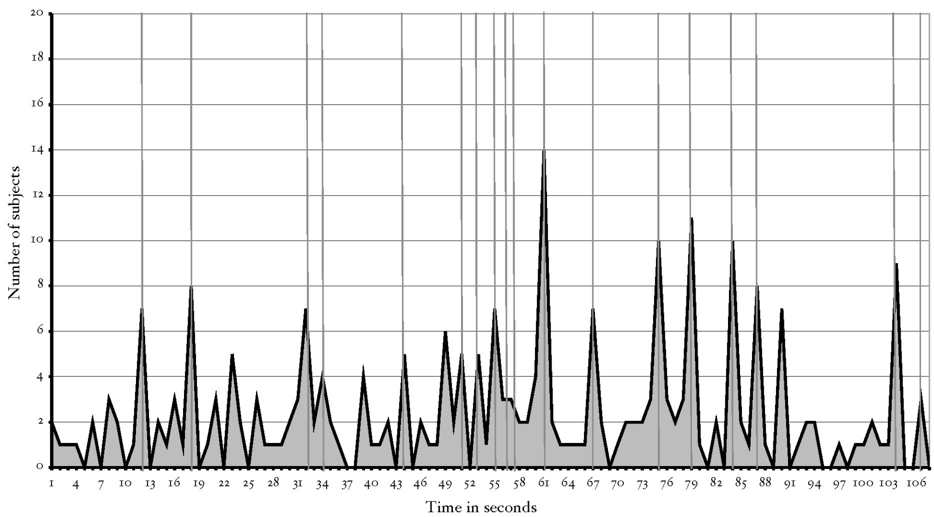

Figure 1. Results of the modal segmentation task performed by the group of Tunisian participants. The curve represents the number of listeners (indicated on the Y-axis) that have indicated segmentation points occurring during the corresponding time interval of one second in the improvisation (positioned on the X-axis). Vertical lines correspond to section divisions given by the musicologist's analysis.

Cognitive modelling

A cognitive model of music segmentation was developed and implemented into computer algorithms. The model is based on a series of heuristics, ordered from low-level acoustic features to high-level cultural knowledge:

- Discontinuities between auditory attributes
- Parallelism, i.e., repeated patterns
- Event stability within functional hierarchy
- Patterns specific to modes
- Transition between modes or subscales
- Formal, stylistic schemas

In the following, each heuristic is described in more details, and its corresponding computer implementation is briefly discussed.

Low-level representation of music. Before discussing the heuristics, we should first of all specify the low-level representation of music considered as input to the structural processes. We decided to start the analysis from a score-like representation. The lower-level processes of extraction of note events from the audio flux will be studied in future research.

In this study, the improvisation has been manually transcribed and encoded in MIDI format, as illustrated in Figure 3. It was possible to use this simplistic representation in this particular study for two reasons. First, the scales used in the improvisation (Figure 2) do not contain any microtonal elements: the chromatic scale, where pitch values are expressed in semi-tones, offers a neutral representation that does not presuppose any implicit scale. Second, due to the absence of evident metrical structure in this improvisation style, durations can be simply expressed in seconds.

Discontinuities between auditory attributes. Local segmentation is founded on relatively contrasting 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, in our context, a pitch leap or a contrastive change in the series of rhythmic values – tends to imply segmentation. This conforms with the Gestalt theory principles of similarity and proximity (Lerdahl and Jackendoff, 1983).

The computational model employed in this study is the Local Boundary Detection Model (LBDM) (Cambouropoulos, 2006). The LBDM mathematically predicts discontinuities perceived between successive notes, based on two rules: a Proximity rule, related to actual pitch and time interval sizes, and a Change rule based on variability between successive intervals. As a result, a discontinuity value is assigned to each interval between successive notes. Segmentations are then predicted and located at note intervals corresponding to relatively high discontinuity values. In our experiment, we use the *Matlab* implementation of LBDM integrated in the *MIDI Toolbox* (Eerola and Toiviainen, 2004).

Parallelism, repeated patterns. Particular schemes, such as sequences of pitches, rhythmic values, etc., are perceived as whole entities, usually called patterns, if they are repeated, developed throughout the piece, with or without variations. This corresponds to the principle of parallelism (Lerdahl and Jackendoff, 1983). In this study, pattern endings are taken into consideration as criteria for segmentation.

The *Istikhbâr* improvisation has been analysed using Lartillot's (2005, 2007) model, which extracts an exhaustive list of repeated patterns in the series of pitch and time intervals. The model was implemented in Common Lisp and integrated into the *OpenMusic* environment (Assayag et al., 1999). A new version in *Matlab* is under development.

Pattern specific to mode. *Mhayyer Sîkâ*, as any *maqām* mode, is associated with a characteristic melodic motif that mainly indicates end of phrases. We hypothesize therefore that the termination of this archetypical motif – underlined in the score in Figure 4 with an extra vertical mark at the right end of each occurrence – contributes to listeners' segmentation.

The detection of this predefined motif in the transcription, being a quite straightforward task in the context of this study, has been performed manually. In further research this heuristic will be automated as well.

Modelling of mode structure. The impact of cultural knowledge on the segmentation behaviour is studied with the modelling of a new set of rules that take into account the modal structure of the improvisation. *Mhayyer Sîkâ*, again as any *maqām* mode, is made up of the juxtaposition of *ajnas* (plural of *jins*), as shown in Figure 2. A *jins* is defined as a group of 3 to 5 successive notes such that one (or two) of those notes is considered as pivotal, i.e., melodic lines tend to rest on such notes. We hypothesise that a transition from one *jins* to another is perceived by both European and Arab listeners as a discontinuity, although due to culture-specific knowledge the feeling of segmentation is stronger for Arab listeners. In Western music, this might be somewhat related to transition from one degree to another, or to modulation from one key to another.

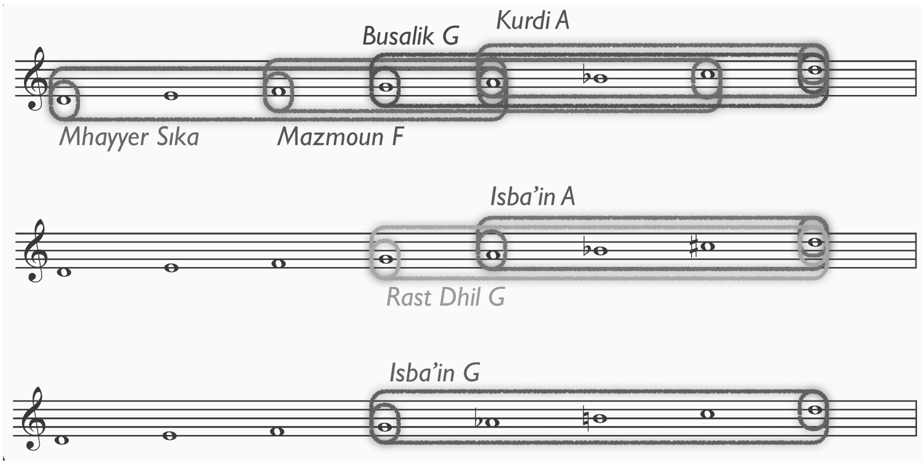

Figure 2. Structure of *Mhayyer Sîkâ* D, a Tunisian *maqām* mode. The *ajnas* constituting the scales are: *Mhayyer Sîkâ* D (main *jins*), *Kurdi* A, *Bûsalik* G, *Mazmoum* F, *Isba'în* A, *Râst Dhîl* G, and *Isba'în* G. Pivotal notes are circled.

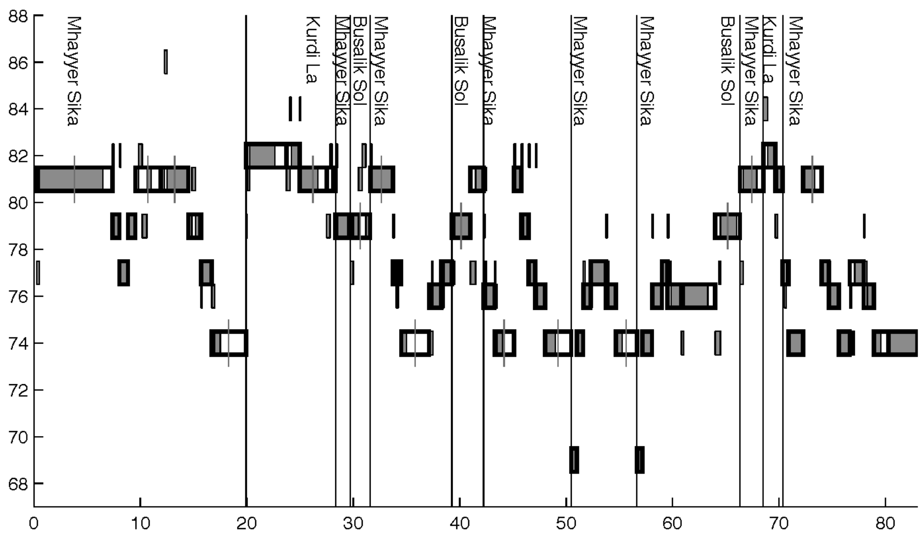

Figure 3. Output of the computational modal analysis. Time in seconds in X-axis, pitch in MIDI-chroma in Y-axis. Actual notes are shown with grey rectangles. Long notes (ornamented or not) are highlighted with bold rectangles. Pivotal notes are indicated with a short vertical line at the middle of the rectangle. Detected jins are shown on the top at the corresponding time position. Candidate segmentation points are indicated with long vertical lines.

This description of Arabic modes has been implemented in the form of a set of general rules, with the purpose of expressing this cultural knowledge in terms of general mechanisms that could be applied, with some variations, to the study of other cultures as well. These rules are detailed in the remaining of this section:

Ornamentation filtering. First of all, a distinction is made between short notes that mostly play the role of ornamentation and longer notes that constitute important steps in melodic phrasing. The simplest method consists in filtering out notes whose temporal distance with subsequent ones (or inter-onset-interval) is shorter than a given constant threshold (in our analysis, 500 ms). A more refined heuristic has been designed that takes into consideration the gross contour profile: an ascending line, for instance, made of a succession of tones with short durations, followed by a descending interval, highlights the highest pitch, climax of the ascending line, even though its actual temporal duration might be short.

In the transcription of the piece in Figure 4, important notes, remaining once ornamentation has been filtered out, are circled. When two neighbouring important notes relate to the same pitch height, they are fused into one single note. This represents a process of reduction of the melodic line linking these two notes into one single pitch. This pre-processed representation is then used as input to the following rules described below. In the figure, such reduction of series of notes into one single pitch is represented by circles encompassing a series of notes: the important note, in this case, is the first and last one, considered as one single event.

A significant amount of work will be required in the future in order to enrich the modelling of mechanisms of ornamentation/reduction. We will also need to take into consideration techniques to study their cognitive justifications and their cultural specificity.

Pivotal note detection. Once ornamental notes are filtered out, what remains are notes that play a role in the modal structure. Amongst those notes, a further distinction is made in order to highlight notes of particularly long durations that play a role of melodic punctuation, and whose pitch values correspond to pivotal points in the modal structure. An easy way of defining important notes is based on a simple constant threshold related to the note duration (or more precisely inter-onset interval): notes whose duration exceeds that threshold (in our analysis, 1.5 second) will be considered as possible candidates for the detection of pivotal points in the modal structure.

Each *jins* is modelled as a concept associated with a value, that represents the degree of likelihood or activation, and allows a comparison between *ajnas* and the selection of the most probable one. This score is represented as a value on a numerical scale referenced by a threshold value: a score above this threshold indicates that the *jins* is considered as a plausible candidate, whereas score below the threshold negates the significance of that particular *jins* for the given musical context.

Each successive note in the improvisation implies an update of the score associated to each *jins*. Four general rules have been defined for the determination and update of scores related to the *jins* candidates:

***Jins* reinforcement.** When the pitch value of a note currently played belongs to a particular *jins*, the score of this *jins* is slightly increased. If the score is below the threshold, the score increases anyway, but remains below the threshold.

Pivotal activation. When a long note currently played corresponds to a pivotal note of a particular *jins*, the score of this *jins* (if inactive) is significantly increased, exceeding the detection threshold, thus confirming the given *jins* as a possible candidate for the current context.

***Ajnas* competition.** When several *ajnas* are activated, the *jins* with highest score is selected as the currently prominent *jins*.

***Jins* deactivation.** When the pitch value of a note currently played does not belong to a particular *jins*, the score of this *jins* is set back to the minimum. When the pitch value of a *long* note currently played does not correspond to a pivotal note of the *jins*, the score is simply decreased.

These rules specify how scores are progressively assigned to each *jins* note after note. We also proposed a few simple rules designed to infer segmentation points from the following *jins* rules: When the previously selected *jins* is not the most dominant activated *jins* anymore:

- If a new *jins* is confirmed, the new modal transition is confirmed, leading to a firm segmentation point, indicated by a '!' punctuation in Figure 5.

- If no *jins* is confirmed, we reach a point of indecision, leading to a possible segmentation at that point, indicated by a '?' punctuation point in Figure 5.

When, on the contrary, the current long note corresponds to the main pivotal note of the selected *jins*, the modal development reaches a state of stability. This can be considered as a possible important punctuation of the phrase, leading to a candidate segmentation point indicated by a '.' punctuation point in Figure 5. These rules have been implemented in a *Matlab* script. The output of the algorithm after analysing the beginning of the studied improvisation is given in Table 1 in the Result section. An example of graphical output returned by the algorithm is shown in Figure 3. A score representation of the same results is shown in Figure 4 and 5.

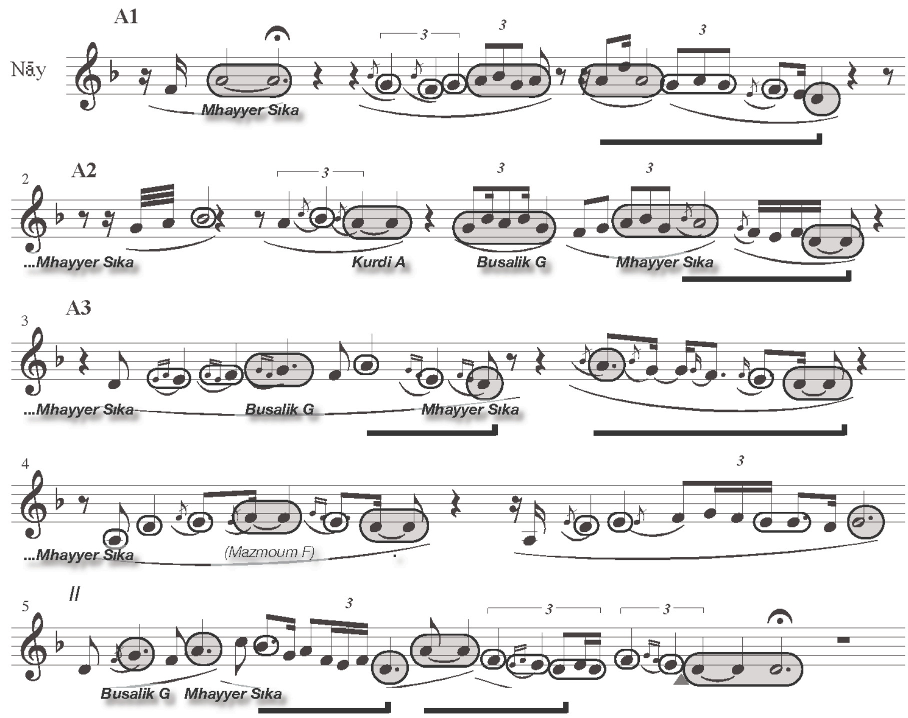

Figure 4. Score representation of the computational modal analysis of the first part of Mohamed Saâda's *maqam*. The terminations of the archetypical *Mhayyer Sîkâ* motif are indicated by bold lines under the staves showing one vertical mark at their right ends. The succession of most likely *ajnas* is indicated below the staves. Important notes (as opposed to ornaments) are circled, and pivotal notes are highlighted with grey ovals that encompass the whole underlying ornamentation.

Formal and stylistic schemas. This last heuristic considered in our list is related to high-level structural configurations that we have not studied yet in our research project, but that we plan to consider and model in future works.

Results

Figure 5 shows the segmentation of the first part of the improvisation, both by the participants and by the computational implementation of the models. The participants responses are displayed above the staves using downward triangles of three colours: black for the first broad (top-level) segmentation, white for the second more detailed (low-level) segmentation, and grey for segmentation based on modes. Above the triangles is indicated the number of participants who segmented at that particular location, for each class of listeners: Tunisians (t), European jazzpersons (j) and non-jazzpersons (n). As mentioned, due to the real-time setting of the experiment, segmentation points have been relocated during a post-processing phase, based on the listeners' own justification of their segmentation choices.

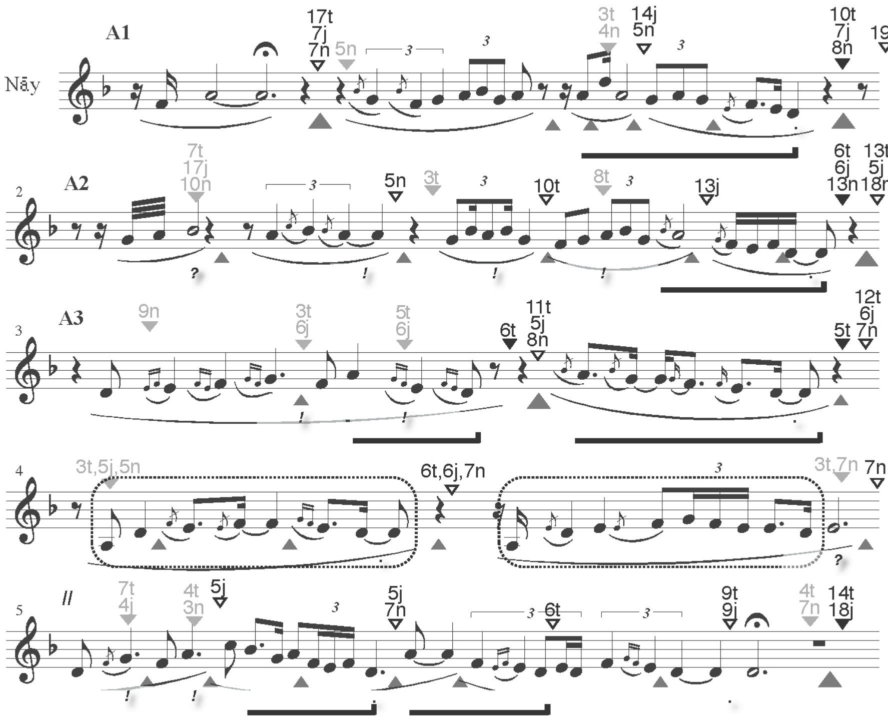

Figure 5. Segmentation of the first part of Mohamed Saâda's *maqam* improvisation by Tunisians (t), European jazzmen (j) and non-jazzmen (n) (over the staves) and by computer (on and under the staves). Local segmentation predicted by the LBDM model is displayed below the staves with upward triangles. The termination of the archetypical *Mhayyer Sîkâ* motif is indicated with a bold line under the staves showing one vertical mark at its right end. Modal segmentation points are indicated by punctuations below the score. See the text for more explanation.

Local segmentation predicted by the LBDM model is displayed below the staves with upward triangles. Strong perceptual discontinuities (large triangles below staves, corresponding to a LBDM value larger than .25) can generally be associated with consensual segmentation points proposed by participants of all cultures. These strong discontinuities coincide with listeners' segmentation into phrases (first task) and musical ideas (second task), though without giving a precise hierarchy between these two levels of representation. Weak perceptual discontinuities (small triangles below staves, corresponding to a LBDM value lower than .25) cannot be easily explained the same way (Lartillot & Ayari, 2008, 2009). Another heuristic for segmentation induction, based on propagation of segmentation expectations, enables to explain interesting segmentation behaviour by listeners, especially in the second part of the improvisation.[4]

Below is described in more detail the progressive modal analysis, note after note, of the beginning of the improvisation. The corresponding quantitative results of the modal analysis are given in Table 1.

Table 1. Modal analysis of the two first lines of the improvisation. Each successive row in the table is related to each successive important note (circled) in the score. For each note the score for each candidate *jins* is shown (except the *ajnas Rast Dhil* G and *Isba'in A*, which were explored during the second part of the improvisation only). Resulting segmentation point possibilities are indicated in rightmost column.

Note	*Mhayyer Sika* D	*Mazmoun* F	*Busalik* G	*Kurdi* A	*Rast Dhil* G	*Isba'in* A	**Segmentation**
0: D0	**2**	0	0	0	0	0	
1: A	**4**	0.5	0.5	2	0.5	0.5	
2: G	**5**	0.6	0.6	0!	0.6	0!	
3: F	**6**	0.7	0!	0	0!	0	
4: G	**7**	0.75	0.5	0	0.5	0	
5: A	**9**	0.8	0.6	2	0.6	0.5	
6: A	**9**	0.8	0.6	2	0.6	0.5	
7: G	**10**	0.85	0.7	0!	0.7	0!	
8: F	**11**	0.9	0!	0	0!	0	
9: D0	**13**	0!	0	0	0	0	
10: Bb	0	0.5	0.5	0.5	0.5	0.5	**? (Indecision)**
11: A	1	0.6	0.6	**2.5**	0.6	0.6	**! (Decision)**
12: G	2	0.7	**2.6**	0!	0.7	0!	**! (Decision)**
13: A	3	0.75	**3.6**	2	0.75	0.5	
14: F	**4**	0.8	0!	0!	0!	0!	**! (Decision)**
15: D	**6**	0!	0	0	0	0	**. (Stability)**

- We propose to consider the pedal note played in the background throughout the improvisation as if it were a long note actually played by the musician, at least at the beginning of the piece, before listeners progressively got less attentive to that static pedal note. In this respect, the long D note (note #0 in Table 1) can be related to the main pivotal note of the *jins Mhayyer Sika* D, which is thereby activated. As *Mhayyer Sika* D is the only *jins* activated at that point, it is considered as the current *jins*.

- The first note played by the musician, note #1[5] in Table 1, with pitch A, confirms the prevalence of the *jins Mhayyer Sika* D, where A is a pivotal note. It also suggests *Kurdi* A as a new candidate *jins*, since A is also a pivotal note here. The other *ajnas* taking part in the *Mhayyer Sika* modal structure are weakly activated as well.
- Note #2, G, confirms the *jins Mhayyer Sika* D as the prevalent subscale, but denies *Kurdi* A as a candidate *jins*, as G does not belong to that subscale. A*jnas Mazmoun* F and *Busalik* G are still weakly activated, without reaching their detection threshold, as none of their pivotal notes have been detected yet. (The current note G is played insufficiently long for it to be considered as a pivotal note for *Busalik* G).
- Note #3, F, still confirms the *jins Mhayyer Sika* D, slightly increases the low activation of *Mazmoun* F, and deactivates *Busalik* G.
- Note #4, G, still confirms the *jins Mhayyer Sika* D and slightly increases *Mazmoun* F, *Busalik* G and *Rast Dhil* G.
- The following notes #5 to #8 lead to similar activation patterns as before.
- Note #9, low D, still confirms the *jins Mhayyer Sika* D, and rejects all other *jins* since this low D is only present in the *jins Mhayyer Sika* D.
- Note #10, Bb, does not belong to the *jins Mhayyer Sika* D, which is therefore rejected. No other *ajnas* are sufficiently activated (since their pivotal notes have not been detected in the current context). We therefore reach a state of indeterminateness, leading to a possible segmentation point (indicated by '**?**' in Table 1 and in Figure 5).
- Note #11, A, reactivates *Mhayyer Sika* D as a candidate *jins*, but particularly activates *Kurdi* A, since A is the main pivot of that *jins*. *Kurdi* A is therefore considered as the most prevalent *jins*, leading to a modal decision and to a segmentation point (indicated by '**!**').
- Note #12, G, played with a long duration, strongly activates *Busalik* G, confirms *Mhayyer Sika* D as an alternate candidate, and desactivates *Kurdi* A. A new modal transition is therefore detected, leading to a candidate segmentation point (indicated here also by '**!**').
- Note #13, A, played with a long duration, confirms both *Mhayyer Sika* D and *Busalik* G, and strongly activates *Kurdi* A as well. In the proposed model, *Busalik* G remains the most probable *jins* at this point. This is justified in particular by the fact that this *jins* was the most prevalent in the previous step.
- Note #14, F, confirms the strong activation of the *jins Mhayyer Sika* D and the slight activation of *Mazmoun* F, but rejects all other subscales. The *jins Mhayyer Sika* D takes the lead once again, leading to a new modal segmentation candidate.
- Note #15, low D played with a long duration, strongly confirms the *jins Mhayyer Sika* D and rejects all other options. As we reach the principal

pivotal point of the main *jins* of the modal structure, we reach a stable point leading to a possible segmentation candidate (indicated here by a dot '**.**').

Besides those considerations based on scales, subscales and pivotal notes, modes are also characterised by specific short melodic motifs. The termination of the archetypical *Mhayyer Sîkâ* motif is indicated by bold lines under the staves in Figure 4, showing one vertical mark at their right ends.

Both European and Arab listeners could perceive parts of the composition process developed in the maqam, but the modulation from one *jins* to another was more strongly perceived by expert listeners as provoking a segmentation in the musical grammar: most modulations – even subtler ones – were detected by at least 3 and up to 7 Tunisian participants for each modulation, while the majority of these modulations were not detected by European participants. As we already reported in (Lartillot and Ayari, 2009), expert listeners tend to detect end of phrases or musical ideas at terminations of the archetypical *Mhayyer Sîkâ* motif, and strong local boundaries are mainly associated with ends of phrases. The further integration of modal analysis developed in this new study shows in addition that stabilisation to the mean pivotal point (as in the middle of stave 4) can also signal to listeners the end of a musical idea. It also explains why a strong discontinuity, such as the first one at the beginning of the improvisation, was not considered as an end of phrase: there is no modal stabilisation to the main pivotal point. It seems therefore that the integration of cultural knowledge allows for a clearer understanding of listeners' segmentation judgements and of the impact of their musical expertise. Besides, this comparison enables us to discuss the relevance of the computational predictions, and to guide further improvements of the cognitive modelling.

Discussion

By implementing a multi-component model designed to capture aspects on the cross-section of music analysis, perception and cognition into a computational model, theoretical hypotheses can be tested through a fully systematic procedure. Predictions of the computational model, once compared with concrete musical cases and listeners' judgments, enable us to question the theoretical hypotheses and to suggest ways of improving both the resulting computational algorithms. For instance, our first attempt at comparing a computational model (at that time mainly based on perceptual heuristics without much cultural knowledge) with listeners' reaction to the same piece revealed some weaknesses in the modelling (Lartillot and Ayari, 2009), suggesting the need to integrate further higher-level heuristics in the model. In further work, the resulting computational model, once validated, could be applied to the analysis of more complex pieces of music, and to large databases of music.

This study of segmentation strategies by listeners of various cultures shows that, whereas a cognitive model purely based on perceptual rules may offer some explanation of listeners' behaviours, the integration of cultural knowledge creates a deeper but at the same time clearer interpretation of the ways listeners constructed a structural understanding of the improvisation: The modelling of mode-based segmentation strategies enabled to reveal subtler articulation points in the discourse, while predicting more precisely at the same time the heuristics responsible for each of those points.

It should be noted however that what we called "low-level heuristics", such as those based on local discontinuities along pitch and time dimensions, are not completely independent from cultural background. In particular, the symbolic representation that is used as input for the analyses is already a product of acculturation, inducing particular categorizations of time and pitch dimensions. Besides, we might notice that even a low-level heuristic such as local discontinuity can be largely dependent on culture: in particular, pitch-gap discontinuity does not seem to have a large segmental impact in this style of music, whereas temporal gaps can be better explained by integrating them into the modelling of Arabic mode (as they help define pivotal notes).

The study was focused on one particular improvisation, but the heuristics that have been developed and implemented are meant to describe general characteristics of music perception. We plan to test the complete model on other improvisations with the same modal structure, and to check the validity of the segmentations and structures returned by the algorithms through listening tests. In order to progressively extend the domain of application of the model, the cultural knowledge will be completed with the integration of other *maqam* modes (scales, *jins*, pivotal notes, representative motivic patterns). A new computational challenge here is that the model should include a process that selects the correct mode out of the list of available modes based solely on the transcription of the improvisation. What is more, the taking into account of microtonality will require a generalization of the symbolic representation and the integration of a transcription module. In future work, we plan to establish a general model where such cultural knowledge would be implicitly learned through exposing the computational model to a corpus of music. The extensive experimentation of the complex model on real-world music might offer new insights into the complex interaction between cognitive constraints and cultural knowledge.

One specific question relates to the level of generality of the proposed model: can it be applied to other cultures as well or is it too specific to the corpus under study? We proposed to formalise maqam modes as scales, i.e. series of notes and intervals, articulated with a series of subscales and pivotal points (the *ajnas*). Whereas the notion of scale is a very general musical concept that can be applied to other cultures as well, the theory of *ajnas* is quite specific to maqam music. A generalisation of the model to western tonal music, for instance, would require an adaptation of the concept of subscale that would correctly describe the notion of tonal *degree*. On the other hand, the idea of associating a numeric "activation" value with each different candidate concept (each possible mode, each possible subscale within one mode) is a general cognitive strategy that could be directly used for the modelling of other cultures as well.

We are integrating all the components of the computational model into a common framework developed in Matlab, that we will release as a module, called *CréMusCult*, within our new environment called *The MiningSuite* (Lartillot, 2011), which will be freely available for download.

Acknowledgments. We would like to warmly thank Renee Timmers and the anonymous reviewers for valuable help in the improvement of the article.

References

Assayag, G., Rueda, C., Maurson, M., Agon, C., & Delerue, O. (1999). Computer assisted composition at Ircam : PatchWork & OpenMusic. *Computer Music Journal, 23*, 59–72.

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

Bod, R. (2001). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research, 31*, 27–37.

Bruderer, M. J., McKinney, M. F. & Kohlrausch, A. (2009). The perception of structural boundaries in melody lines of Western popular music. *Musicae Scientiae 13*, 273 – 313.

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

Clarke, E. F., & Krumhansl, K. L. (1990). Perceiving musical time. *Music Perception, 7*, 213 – 252.

Deliège, I. (1987). Grouping conditions in listening to music. *Music Perception, 4*, 325–360.

Eerola, T. & Toiviainen, P. (2004). *MIDI Toolbox: MATLAB tools for music research*. University of Jyväskylä: Kopijyvä, Jyväskylä, Finland. Available at http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/.

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*, 499–543.

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*, 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, PA: Information Science Reference.

Lartillot, O. (2011). A comprehensive and modular framework for audio content extraction, aimed at research, pedagogy, and digital library management. *Proceedings of the Audio Engineering Society (AES) Convention*, Paper number 8375.

Lartillot, O., & Ayari, M. (2008). Segmenting Arabic modal improvisation: Comparing listeners' responses with computer predictions. *Proceedings of the Conference on Interdisciplinary Musicology (CIM08)*.

Lartillot, O., & Ayari, M. (2009). Segmentation of Tunisian modal improvisation: Comparing listeners' responses with computational predictions. *Journal of New Music Research, 38*, 149–159.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Melucci M., & Orio, N. (2002). A comparison of manual and automatic melody segmentation. *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 7–14.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model*. Chicago, IL: University of Chicago Press.

De Nooijer, J., Wiering, F., Volk, A., & Tabachneck-Schijf H. J.M. (2008). Cognition-based segmentation for music information retrieval systems. *Proceedings of the Conference on Interdisciplinary Musicology (CIM08)*.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*, 1367–1391.

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.

Thom, B., Spevak, C., & Höthker, K. (2002). Melodic segmentation: Evaluating the performance of algorithms and musical experts. *Proceedings of the International Computer Music Conference (ICMC)*, 65–72.

---

[1] These notions are further discussed in the next section.

[2] Examples of musical events considered by groups of listeners, for the first segmentation task, were: end of phrase, end of melodic movement, end of part, melodic modulation; for the second segmentation task: exposition of a particular degree of the mode, end of exposition, end of musical idea, end of small melodic movement, affirmation of a particular *jins*, transposition of a motif, development, variant, melodic descent.

[3] For further justifications for this methodology, cf. Lartillot and Ayari (2009).

[4] More details in (Lartillot & Ayari, 2009).

[5] In the note enumeration in Table 1, notes playing simple ornament role are not taken into account.

Biographies

Olivier Lartillot is an Academy of Finland Research Fellow at the Finnish Centre of Excellence in Interdisciplinary Music Research, at the University of Jyväskylä. His research in the areas of computer science, music analysis and music cognition are dedicated to the development of a computational framework for music analysis from symbolic and audio domains. He obtained a degree in engineering at Supélec Grande École, France, and a PhD degree in Computer Science at Ircam / University of Paris 6 in 2004. He also obtained a BA in Musicology from the University of Paris-Sorbonne.

Mondher Ayari is Lecturer at the Department of Music of the University of Strasbourg. His research in ethnomusicology and music cognition is dedicated to the history, analysis and perception of Arabic and Oriental improvised music. Music Degree, Tunis National Conservatory, 1990, Master Degree (musicology), Superior Music Institute of Tunis, 1994, Ph.D. (esthetics, science and technology of art), Paris 8, 2000. Author, L'écoute des musiques improvisées : essai de psychologie cognitive (L'Harmattan, 2003). Editor, De la théorie musicale à l'art de l'improvisation : analyse des performances et modélisation musicale (Delatour, 2005)