A COMPREHENSIVE COMPUTATIONAL MODEL FOR MUSIC ANALYSIS, APPLIED TO MAQAM ANALYSIS

Olivier Lartillot
University of Jyväskylä
olartillot@gmail.com

Mondher Ayari
University of Strasbourg
mondher.ayari@ircam.fr

ABSTRACT

We introduce a new computational framework for music analysis decomposed into a set of modules. Each module addresses a core aspect of music analysis and offers some innovative breakthrough compared to the state of the art. In order to overcome the limitations of local segmentation, we propose an alternative paradigm based on hierarchical local grouping. New mechanisms for ornamentation reduction based on local grouping enable to build a syntagmatic network for the search for ornamented patterns. We propose an approach for modal analysis based on comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. We show how this could be applied in particular for the analysis of Maqam music. Pattern mining is applied for the search for motives, for mode-related patterns as well as metrical analysis. The integration of the modules into a single framework enables to model interdependencies, which play a major role in music.

1. TRANSCRIPTION FROM AUDIO

The analysis can be carried out on MIDI, score representations or audio files. For audio recordings, a first step of transcription attempts to locate temporal position of notes through a combined detection of significant increase of energy and of stabilised pitch. This results in a “protosymbolic” representation where notes are characterised by temporal location and duration, pitch and dynamics. Pitch quantisation and spelling is carried out with interaction through a combined detection of significant increase of energy and of stabilised pitch. This results in a “proto-symbolic” representation where notes are characterised by temporal location and duration, pitch and dynamics. Pitch quantisation and spelling is carried out with interaction through a combined detection of significant increase of energy and of stabilised pitch.

1 Pierre et al. (2010) research also enters into that linear segmentation paradigm, but the underlying principles are not based on the local texture of music, but instead on stylistic rules.

2. LOCAL GROUPING

There has been significant research around the concept of local segmentation, studying the emergence of structure related to the mere variability in the succession of musical parameters. These research, notably by Tenney & Polansky (1980) or Cambouropoulos (2006), focus on the analysis of monodies, and model this structural phenomenon as a segmentation of the monody, which cuts the temporal span at particular instants, resulting into a linear succession of segments. We previously showed that in these approaches the heuristics for segmentation is based on a mixture of several constraints related to what happens both before and after each candidate segmentation point (Lartillot et al., 2013). We presented instead a simpler approach focused only on what happens before each candidate segmentation: this enables to reveal a more complete set of segmentation points, indicate more precisely the temporal locations of the segmentation points, and could also reveal a segmentation hierarchy at multiple structural levels.

We introduce a new formulation of our proposed approach that reveals a much clearer structural description and that can be explained with simple principles. The approaches focuses on grouping instead on segmentation. In other words, what needs to be characterised are not the segments between notes, but instead the groups of notes that are progressively constructed in a hierarchical framework. The approach is applied uniquely to the temporal domain (to the characterisation of the monody as a succession of inter-onset intervals, or IOIs), and does not apply therefore to the pitch domain (the succession of inter-pitch intervals). This is because pitch-based grouping is more related to streaming, i.e., to the construction, from a given monody that features pitch gaps, of internal monodic lines.

This new model for segmentation along the IOI description can be explained as follows:

- A local group $G$ is characterised by a maximal IOI parameter $I$: the IOIs between the successive notes are all lower or equal to that parameter.

- Local groups form strict hierarchies: if one local group $G_1$ has a maximal IOI $I_1$ that is higher than the parameter $I_2$ of another local group $G_2$, and if both groups coincide in the monody, then necessarily $G_2$ is strictly included into $G_1$.

- Local grouping is computed through a single chronological pass of the monody. For a given note $n_k$, we consider the groups $\{G_i\}$ containing the previous note $n_{k-1}$ from larger groups (higher maximal IOI) to smaller groups (lower maximal IOI). Each group $G_i$ is compared with the new IOI $I_{k-1,k}$ between $n_{k-1}$ and $n_k$.

- If the new IOI is equal to or smaller than the maximal IOI of the given group $G_i$, the group is extended with this new note $n_k$. In order to tolerate slight slow down, even IOIs that are a little higher than the maximal IOI are accepted. More precisely,
the condition for group extension is the following:

$$\log \frac{I_{k-1,k}}{I_{i_0}} < \delta$$  \hspace{1cm} (1)

where $\delta$ is fixed to .3 in our current tests.

- If the new IOI is larger than the maximal IOI of the given group $G_{i_0}$, i.e. if

$$\log \frac{I_{k-1,k}}{I_i} > \delta$$  \hspace{1cm} (2)

the group is closed: it is followed by a longer silence so will not be extended any more. All the other groups $\{G_i\}_{i > i_0}$, with even lower maximal IOIs, are closed as well.

- Hence only the larger groups with higher maximal IOIs, $\{G_i\}_{i < i_0}$, have been extended. If the next IOI $I_{k-1,k}$ is even smaller than the maximal IOI $I_{i_0-1}$ of that smallest group, i.e. if:

$$\log \frac{I_{k-1,k}}{I_{i_0-1}} < \epsilon$$  \hspace{1cm} (3)

where $\epsilon$ is fixed to -.4 in our current tests, then a new group $G_i$ is created, whose maximal IOI is set to the current IOI, i.e. $I_i = I_{k-1,k}$.

This approach generates a hierarchical structuration of the monody that is very intuitive to understand, as shown in the example in Figures 2 to 5.

Each local group starts immediately at the onset of its first note. Concerning the temporal location of the closure of the group – that can be called group offset –, we can apply the very same heuristics we introduced previously in the context of local segmentation (Lartillot et al., 2013): In order to detect that a given group $G_i$ is closed, we need to wait at least the temporal interval $I_i$ after the last note’s onset in order to check whether a new note appear during that temporal span or not. If no new note appear, the group is closed. Thus the group offset can be assigned to that moment at $I_i$ after the last note’s onset.

3. ORNAMENTS REDUCTION

Previous computational attempts to model processes related to melodic reduction in music primarily (Gilbert & Conklin, 2007; Marsden, 2010) formalize general aspects without detailing concrete conditions for reduction. We present a set of rules founding the detection of ornaments, based on local grouping.

3.1 Local group’s head

By definition, a local group terminates with a note that is followed by a duration (before the next note) that is significantly longer than the IOIs within the group. As such, the local group can be perceived as a phrase that terminates with a concluding note that has a more structural importance. This hypothesis might not be always valid, in particular in the presence of particular accentuations at particular

Figure 3: Analysis of the first stave of the improvisation displayed in figure 1 using the same convention as in figure 2.

Figure 4: Analysis of the second stave of the improvisation displayed in figure 1 using the same convention as in figure 2.

Figure 5: Analysis of the third stave of the improvisation displayed in figure 1 using the same convention as in figure 2.
Figure 1: Beginning of a traditional Tunisian modal improvisation *Istikbâr* played by flute master Mohamed Saâda on the *Mhayer Sikâ* maqâm.

Figure 2: In black: piano roll representation of the beginning of a traditional Tunisian modal improvisation (*Istikbâr*). In blue: local groupings. In red: Local groups’ heads. In grey: passing note.
notes within the group. But in more general case, it seems to offer some general interest. Following this observation, we propose to formalise this hierarchy of notes in local groups by associating with each local group a main note, or “head”, to follow Lerdahl & Jackendoff (1983)’s Time-Span Reduction terminology, which would in the simple case be the last note of the group, as circled in red in Figures 2 to 5. The other notes would be considered as “subordinate events”, or – why not – as the “tail” of the group.

When subordinate events in a local group have same pitch than the final note of the group, they all form a single note – a “meta-note”, as we called in Lartillot & Ayari (2012) –, which will become the actual head of the group. The subordinate events of the groups can be considered as forming an ornamentation – such as a cambiata or a trill – of the group’s head. They are highlighted by red rectangles in Figures 2 to 5.

Meta-note can form is any hierarchical level: For instance, in Figure 3, the meta-note is formed on the smallest local group, but in Figure 4 the meta-note around time 13 second in the improvisation is formed on a intermediary hierarchical level.

So far, these subordinate events consist of the notes forming the local group. But a richer understanding of the structural configuration is that the subordinate events of a local group G consist actually of the smaller local groups that belong to that group. And this hierarchical structuration continues recursively for the smaller local groups. Thus when searching for the subordinate events that have same pitch than the last note of group G, we don’t need to remember all the notes in the group, but only to check the pitch of the heads of the local groups one level down in the hierarchy.

### 3.2 Passing note

Within a local group, all notes do not have the same importance. A monotonous and uniform conjunct melodic motion is a series of notes such that:

- inter-pitch intervals between success notes are all of 1 or 2 semi-tones and in the same direction (up or down),
- inter-onset intervals between successive notes are very similar,
- no note is particularly accentuated.

In such configuration, the intermediary notes form passing notes: these subordinate elements play mainly a role of filling the interval gap between the starting and ending points of the line. For that reason, these intermediary notes are generally not perceived as note that play a more global role outside that particular melodic line. Intermediary notes are shown in grey in Figures 2 to 5.

This can have an impact in different modules of the integrated analysis framework. For instance, in the previous paragraph dealing with local group’s head (section 3.1), a note within a local group that has same pitch than the last note of the group cannot be included in the group’s head if it is a passing note.

Passing notes can exist in melodic lines of any length, so for a given note $n_k$, its “passingness” can be defined based simply on the previous note $n_{k-1}$ and next note $n_{k+1}$, according to the following conditions:

- the inter-pitch intervals between $n_{k-1}$ and $n_k$ and between $n_k$ and $n_{k+1}$ are of 1 or 2 semi-tones and in the same direction (up or down),
- their inter-onset intervals are very similar:

$$\left| \log \frac{I_{k-1,k}}{I_{k,k+1}} \right| < \beta \quad (4)$$

where $\beta$ is fixed to .1 in our current tests.

### 3.3 Syntagmatic network

Melodic “reduction” is commonly understood as a process of eliminating the ornamentation (here, the subordinate elements) of the monodic surface in order to keep the deeper structure (on various hierarchical levels) made of the more important notes. Our conception of melodic reduction, however, does not impose such reduction of information, but on the contrary, integrates the deeper structure information with the monodic surface. More precisely, the monodic surface is formalised as a chain of connections between successive notes, i.e., a chain of syntagmatic connections, or a syntagmatic chain, following Saussure’s terminology. The deeper structure can be represented by adding new syntagmatic connections between successive elements in the deeper hierarchical levels. We obtain hence a syntagmatic network presenting a set of possible alternative syntagmatic chains. We are currently formalising heuristics ruling this construction of syntagmatic chains. They could include for instance:

- The head of any local group is syntagmatically connected to the note $n_i$ preceding the group as well as to the head of any local group closed by $n_i$.
- The head of any local group is syntagmatically connected to the note $n_j$ succeeding the group as well as to the head of any local group started by $n_j$.
- In a series of passing note, there is a direct syntagmatic connection between the notes just before and after that series.
- A syntagmatic chain of notes $\{n_{k-1}, n_{k-2}, \ldots \}$ of same pitch form a single meta-note (Lartillot & Ayari, 2012) whose onset is set at the first note $n_{k-1}$. This meta-note can be syntagmatically connected to any note succeeding any note $n_{k-2}, n_{k-3}, \ldots$ constituting the meta-note.

This concept of syntagmatic network follows Lartillot & Ayari (2012), but we propose here a much simpler network with connections justified by stronger perceptual heuristics, based on local grouping and ornamentation reduction.
The interest of this network is that motivic pattern can be searched for on any path, which enables to detect pattern repetition with or without ornamentation.

Example of syntagmatic networks is shown in yellow in Figures 3.

4. MODAL ANALYSIS

4.1 Scale identification

Previous methods in tonal and modal analysis traditionally compute global pitch statistics (of either whole pieces or on successive arbitrary time frames) that are compared to mode or key templates (Krumhansl, 1990; Gomez, 2006; Gedik & Bozkurt, 2010). The main limitations are that these arbitrary time frames often encompass complex modulations or spurious note events that are foreign to the main mode, and that the templates force a single stereotypical representation of each mode.

In contrast, we are conceiving a new paradigm for modal analysis carried out for each successive note in the protosymbolic representation, and based on a comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. A scale can also be identified while recognizing only subset of it, by taking also into account pivotal notes in the scale and longer and main notes in the local context.

Below is a more precise description of the approach, currently under development, specialised here on the analysis of Arabic Maqam music.

- Each Maqam mode is defined by:
  - A scale: a series of pitches, indicated relatively to the origin of the scale.
  - A juxtaposition of ajnas (plural of jins), as shown in Figure 6. 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. The pitch scale of that jins is also expressed relatively to the pivot, to which is associated the value 0. One of the ajnas is considered as the main jins of the Maqam mode: it is typically the one that starts at the tonic of the scale.
  - For each successive note \( n_k \) being analysed, we keep track of any combination of set of pitches \( \{ S_i \} \) that have been recently played. For each set of pitches \( S_i \), one pitch is selected as pivot, it is the pitch of the note with longest duration, among the notes associated with that set of pitch. The pivot is fixed as the pitch origin, with value 0, and the other pitches of the scale are expressed with respect to the pivot.
  - The combinatory of possible subset of pitches can be reduced by taking into account the local grouping structure discussed in section 2. For a given note \( n_k \) of pitch \( p_k \), a pitch that is not expressed by head of previous local groups, but that appears only as subordinate event(s) of one or several groups \( G_i \) will not form a pitch subset with \( p_k \) unless the subset also includes the pitch expressed by the head of the group \( G_i \). This represents the hierarchical structure of pitch scale, where the less important pitches, appearing as ornaments, can be considered in the scale only if the whole ornament including the group head is included.
  - For each set of active pitches \( S_i \), we can measure the degree of fit \( SJ_{i,j} \) with each different possible jins \( J_j \), which is defined as the number of pitches in \( S_i \) that can be associated with a pitch in the jins scale, divided by the total number of pitches in \( J_j \). Hence this score \( SJ_{i,j} \) indicates the proportion of the jins scale covered by \( S_i \). Note that scales related to \( S_i \) and \( J_j \) are both expressed relatively to their pivot, which is in both case equal to 0. The alignment of pivots means that the jins \( J_j \) compared to \( S_i \) is at a specific transposition.
  - For each possible transposition of each Maqam mode, and for each constituting jins \( J_j \), we can find the set of active pitches \( S_i \) that fits best, i.e. the one with highest score \( SJ_{i,j} \). For each Maqam mode, and for each possible transposition, we obtain hence a table of scores that shows how much its constituting ajnas are covered throughout the piece.
  - The development of a Maqam mode can be considered as a succession of ajnas. But it appears that an important part of ornamentation transgresses somewhat the border fixed by the ajnas. These ornamentation could be understood as transitory ajnas that appear locally and are superposed to the longer-term logic of the ajnas developed in a larger scale.
  - Again, the local grouping structure can be used to infer the temporal scope of the different ajnas discovered. Particular notes that are subordinate events of local groups can infer particular ajnas that do not exceed the scope of the local group.
  - We are conceiving methods for choosing the actual Maqam mode, among the different possible candidates. For a new Maqam to be detected at a given note \( n_k \), its main jins needs to be activated for that note.

4.1.1 Example

Here there ideas are illustrated through the analysis of the beginning of an improvisation based on Mhuyyer Sîkâ maqam (Lartillot & Ayari, 2011). To make the analysis more challenging, we ignore the pedal note D played from the beginning of the improvisation to the end, emphasising the tonic of the mode.

- The improvisation starts with a short note with pitch F.
• The next note (pitch A) is very long. We obtain the subset (F,A) with pivot on A. Evidently, a huge number of possible ajnas from various scale at various transposition can be associated to this minimalistic description, so no particular modal context is inferred for the moment.

• The third note (pitch Bb) is very short. Subset (A,Bb) with pivot on A is inferred. The combination of the subset (F,A) and (A,Bb) leads to an identification of the Mhayyer Sikâ scale with (F,A) belonging to the main jins while (A,Bb) belongs to the Kurdi A jins.

• The next note (pitch G) is longer than the previous one, closing a 2-note local group whose pitch subset (G,Bb) belongs to the Busalik G jins. But in a larger scale, the subset (F,G,) belongs to the main jins.

• The next two notes (pitch Bb and F) form a local group related to the Mazmoun F jins. In larger scale, F develops further the main jins.

• The analysis continues similarly.

4.2 Modal pattern identification

But beyond scales, ajnas and pivotal notes, Maqam modes are also characterised by particular melodic lines. We conceived an innovative method (Lartillot, 2005) for listing sequential patterns and tracking their possible cyclical repetition. We are currently integrating the detection of these modal patterns through an interaction between local grouping, ornamentation reduction and the motivic analysis module.

For instance, the main characteristic melodic line associated with Mhayyer Sikâ maqam is the pitch series A F E D. The modal pattern can be explicitly found along particular paths of the syntagmatic network constructed based on the method presented in section 3.3.

5. FURTHER WORKS

Adequate rhetorical description of notes requires the inference of the underlying pulsation, and more generally of multiple pulsation levels forming a metrical structure. Current beat tracking methods are based on global description of periodicity (such as autocorrelation function) (Dixon, 2007), which, here also, fails to grasp particular idiosyncrasies of music, such as ornaments or sudden changes of tempo. Besides, periodicities can also be expressed as successive repetitions of sequential patterns. This method is applied among others to detect pulsation and construct the metrical structure. The metrical structure cannot always be inferred from a mere search for periodicity in the signal, but may sometimes require the recognition of learned rhythmic patterns.

Our aforementioned method for sequential pattern mining is used to detect not only metrical periodicities, but also any type of sequential repetitions such as melodic themes and motifs (Lartillot, 2005). Ornaments reduction enables to detect varied repetitions; modal and metrical analyses add further musical representations (such as diatonic scale and rhythmic values) along which the sequential pattern mining is carried out as well.

Following Lerdahl & Jackendoff (1983), grouping structures are founded not only on local discontinuity (cf. local grouping module), but also on higher-level aspects, such as mode, metre and motifs. We are extending our previous works (Lartillot & Ayari, 2009) focused initially on linear segmentation in order to integrate multi-level grouping. Contrary to the purely hierarchical representation in Lerdahl & Jackendoff (1983), we propose a model that allows overlapping and multiple segmentation alternatives. One main application of this grouping structure construction is the detection of musical forms. This would be modeled as a mapping between the global grouping structure constructed on a given piece with form patterns that constitute the predefined cultural knowledge.

6. CONCLUSION

One main objective of the CrèMusCult project\(^2\) was to study the impact of cultural knowledge in listeners’ understanding of music and in our proposed cognitive modeling of music analysis. Cultural knowledge is implemented as specification of list of modes, metrical structures, motifs, forms, but could also take the form of particular parametrization such as in the local grouping module.

The complete framework presented in this paper forms a package, called MusMinr, of MiningSuite, a new Matlab environment for audio and music analysis. A standalone graphical user interface, also called CrèMusCult, offers to musicologists the possibility of easily visualize the complete analysis of musical pieces of their choices.

\(^2\) Funded by the French research agency (ANR) during the years 2011-2013.
7. ACKNOWLEDGMENTS
This study has benefited from stimulating discussions with Mathieu Giraud, Petri Toiviainen and Funda Yazici.

8. REFERENCES


