

AN INTEGRATED FRAMEWORK FOR TRANSCRIPTION, MODAL AND MOTIVIC ANALYSES OF MAQAM IMPROVISATION

Olivier Lartillot

Swiss Center for Affective Sciences, University of Geneva
olartillot@gmail.com

Mondher Ayari

University of Strasbourg & Ircam-CNRS
ayari.mondher@gmail.com

ABSTRACT

The *CréMusCult* project is dedicated to the study of oral/aural creativity in Mediterranean traditional cultures, and especially in Maqam music. Through a dialogue between anthropological survey, musical analysis and cognitive modeling, one main objective is to bring to light the psychological processes and interactive levels of cognitive processing underlying the perception of modal structures in Maqam improvisations.

One current axis of research in this project is dedicated to the design of a comprehensive modeling of the analysis of maqam music founded on a complex interaction between progressive bottom-up processes of transcription, modal analysis and motivic analysis and the impact of top-down influence of higher-level information on lower-level inferences.

Another ongoing work attempts at formalizing the syntagmatic role of melodic ornamentation as a Retentional Syntagmatic Network (RSN) that models the connectivity between temporally closed notes. We propose a specification of those syntagmatic connections based on modal context. A computational implementation allows an automation of motivic analysis that takes into account melodic transformations. The ethnomusicological impact of this model is under consideration.

The model was first designed specifically for the analysis of a particular Tunisian Maqam, with the view to progressively generalize to other maqamat and to other types of maqam/makam music.

1. INTRODUCTION

This study is illustrated with a particular example of Tba' (traditional Tunisian mode), using a two-minute long Istikhbâr (a traditional instrumental improvisation), performed by the late Tunisian Nay flute master Mohamed Saâda, who developed the fundamental elements of the Tba' Mhayyer Sikâ D. This example is challenging for several reasons: in particular, the vibrato of the flute does not allow a straightforward detection of note onsets; the

Copyright: © 2012 Olivier Lartillot et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

underlying modal structure has rarely been studied in a computational framework; the absence of a clear metrical pulsation complicate the rhythmic transcription¹.

The long-term aim of the project is to develop a computational model that is not focused on one single piece, or one particular style of modal music, such as this Tunisian traditional Istikhbâr improvisation, but that is generalized to the study of a large range of music, Arabo-Andalusian maqam but also Turkish makam for instance.

2. BOTTOM-UP ANALYSIS

The aim of music transcription is to extract elementary musical events (such as notes) from the raw audio signal, and to characterize these events with respect to their temporal locations and durations in the signal, their pitch heights, dynamics, but also to organize these notes into streams related to particular musical instruments and registers in particular, to integrate the notes in an underlying metrical structure, to indicate salient motivic configurations, etc.

Computational techniques to detect these events are based on three main strategies:

- A first strategy consists in detecting saliencies in the temporal evolution of the energy of the signal. This method does not work when single notes already feature significant temporal modulation in energy (such as vibrato) or when series of notes are threaded into global gestures where the transition between notes is not articulated in terms of dynamics.
- An alternative consists in observing more in details the spectral evolution, and in particular in detecting significant dissimilarities between successive frames with respect to their general spectral distributions. Yet still global comparisons frame by frame cannot generally discriminate properly between spectral discontinuities that are intrinsic to the dynamic of a single note and those that would relate to transition between notes.
- Another alternative consists in analyzing the temporal evolution of the note pitch heights and to infer, from this continuous representation, periods of

¹ The emergence of local pulsation in non-metric music is an important question that we plan to study extensively in forthcoming works.

stability in pitch height corresponding to notes. This method is particularly suited to instrument featuring vibrato, such as flute.

This section details our proposed method that follows this third pitch-based strategy.

2.1 Autocorrelation and spectrogram combined method

We propose a method for pitch extraction where two strategies are carried out in parallel. The first strategy based on autocorrelation function focuses on the fundamental component of harmonic sounds, and can track multiple harmonic sources at the same time [8]. The audio signal is decomposed using a two-channels filterbank, one for low frequencies below 1000 Hz, and one for high frequencies over 1000 Hz. On the high-frequency channel is performed an envelope extraction using a half-wave rectification and the same low-pass filter used for the low-frequency channel. The periodicity corresponding to note pitch heights is estimated through the computation of an autocorrelation function using a 46.4 ms-long sliding Hanning window moving every 10 ms. Side-border distortion intrinsic to autocorrelation function is neutralized by dividing the autocorrelation with the autocorrelation of its window [6]. A magnitude compression of the amplitude decreases the width of the peaks in the autocorrelation curve, suitable for multi-pitch extraction. After summing back the two channels, the sub-harmonics implicitly included in the autocorrelation function are filtered out from the halfwave-rectified output by subtracting time-scaled versions of the output. A peak picking frame by frame of this representation results in a pitch curve showing the temporal evolution of the fundamental components of the successive notes played by the musical instruments. One drawback of this method is that the frequency is not clearly stabilized on each note, showing fluctuation.

The second strategy for pitch extraction is simply based on the computation of a spectrogram using the same frame configuration as for the first method. In this representation, the curve of the fundamental component is indicated with better accuracy and less fluctuation, but harmonics are shown as well, so the fundamental curve cannot be tracked robustly. The advantages of the two methods are combined by multiplying point by point the two matrix representations, so that the fundamental curve is clearly shown and the harmonics are filtered out [7].

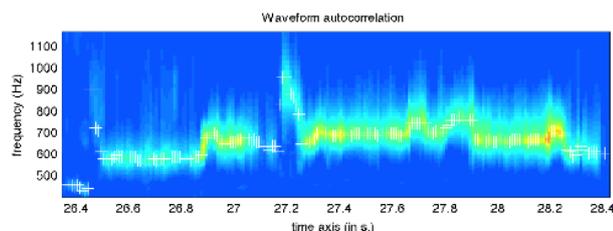


Figure 1a. Autocorrelation function of each successive frame (each column) in an excerpt of the improvisation.

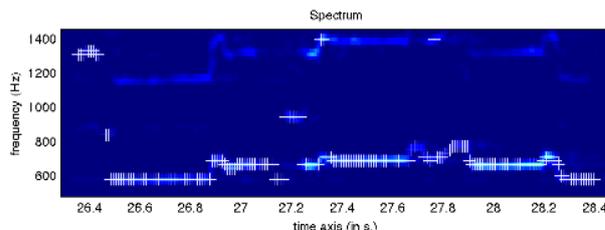


Figure 1c. Spectrogram computed for the same excerpt.

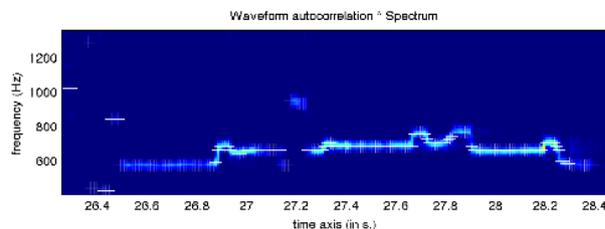


Figure 1e. Multiplication of the autocorrelation functions (Figure 1a) and the spectrogram (Figure 1c).

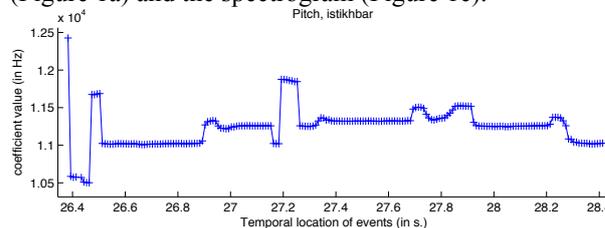


Figure 1f. Resulting pitch curve obtained from the combined method shown in Figure 1e.

2.2 Pitch curve

Global maxima are extracted from the combined pitch curve for each successive frame. In the particular example dealing with nay flute, the frequency region is set within the frequency region 400 Hz – 1500 Hz.

Peaks that do not exceed 3% of the highest autocorrelation value across all frames are discarded: the corresponding frames do not contain any pitch information, and will be considered as silent frames. The actual frequency position of the peaks is obtained through quadratic interpolation.

The frequency axis of the pitch curve is represented in logarithmic domain and the values are expressed in cents, where octave corresponds to 1200 cents, so that 100 cents correspond to the division of the octave into 12 equal intervals, usually called “semi-tones” in music theory. This 12-tone pitch system is the basis of western music, but is also used in certain other traditions as well. The maqam mode considered in this study is based also on this 12-tone pitch system. More general pitch system can be expressed using the same cent-based unit, by expressing intervals using variable number of cents.

2.3 Pitch curve segmentation

Pitch curves are decomposed into gestures delimited by breaks provoked by any silent frame. Each gesture is further decomposed into notes based on pitch gaps. We need to detect changes in pitch despite the presence of frequency fluctuation in each note, due to vibrato, which can sometimes show very large amplitude. We propose a method based on a single chronological scan of the pitch

for the current context. When the pitch value of a note currently played does not belong to a particular *jins*, the score of this *jins* is decreased.

These rules above found the first version of the computational modeling of modal analysis we initially developed [1]. One major limitation of this model is that any note not belonging to the predominant *jins* (the one with highest score), even a small note that could for instance play a role of ornamentation, may provoke a sharp drop of that score. The solution initially proposed was to filter out these short notes in a first step, before the actual modal analysis. Yet automating such filtering of secondary notes arises further difficulties, and it was also found problematic to consider such question independently from modal considerations.

A new model is being developed that answers those limitations. The strategy consists in automatically selecting the notes that contribute to a given *jins* and in discarding the other notes. For each *jins* is hence constructed a dedicated network of notes; in some cases, this network connects notes that are distant from each other in the actual succession of notes of the monody, separated by notes that do not belong to the *jins* but that are considered in this respect as secondary, playing a role of ornamentation. Constraints are added that require within-*jins* notes to be of sufficient duration, with respect to the duration of the shorter out-of-*jins* notes, in order to allow the perception of connection between distant notes.

3.2. Extension of the Model

The computational model presented in the previous section is currently enriched by integrating not only the modelling of individual ajnas, but also of a larger set of maqamat. Similarly to the modelling of ajnas, with each maqam is associated a numerical score that varies throughout the improvisation under analysis. This value represents a degree of likelihood, or activation, and allows a comparison between maqamat and the selection of the most probable one. The score of each maqam is based on two principles: scales and constituting ajnas.

A larger set of maqamat – including their possible transpositions – and their ajnas is progressively considered. In this general case, the detection of maqamat and ajnas cannot rely on absolute pitch values any more, but instead on the observation of the configuration of pitch intervals, in order to infer automatically the actual transposition of each candidate *jins* and of the resulting candidate maqamat.

3.3. Impact on Transcription

Sometimes the short notes that play a role of appoggiaturas or other ornamentations are not associated with a very precise pitch information as a degree on the modal scale. Although a precise scale degree can in many cases be assigned based on the audio analysis, this particular pitch information is not actually considered as such by expert listeners if its actual value contradicts with the implicit modal context. In such case, this pitch information is understood rather as an event with random pitch [2]. Such

filtering of the transcription requires therefore a modal analysis of the transcription.

4. MOTIVIC ANALYSIS

We stress the importance of considering the notion of note succession – or syntagmatic connection – not only between immediately successive notes of the superficial syntagmatic chain, but also between more distant notes. Transcending the hierarchical and reductionist approach developed in Schenkerian analysis, a generalised construction of syntagmatic network, allowed by computational modelling, enables a connectionist vision of syntagmaticity.

4.1. Retentional Syntagmatic Network

We define a *Retentional Syntagmatic Network* (RSN) is a graph whose edges are called syntagmatic connections, connecting couple of notes perceived as successive. Combination of horizontal lines, typical of contrapuntal music in particular, are modeled as syntagmatic paths throughout the RSN.

A syntagmatic connection between two notes of same pitch, and more generally a syntagmatic chain made of notes of same pitch, are also perceived as one single “meta-note”, called *syntagmatic retention*, related to that particular pitch, such that each elementary note is considered as a repeat of the meta-note on a particular temporal position. This corresponds to a basic principles ruling the Schenkerian notion of *pitch prolongation*.

Since successive notes of same pitch are considered as repeats of a single meta-note, any note n of different pitch that comes after such succession does not need to syntagmatically connect to all of them, but can simply be connected to the latest repeat preceding that note n . Similarly, a note does not need to be syntagmatically connected to all subsequent notes of a given pitch, but only to the first one. The actual note to which a given note is syntagmatically connected will be called *syntagmatic anchor*.

This enables to significantly reduce the complexity of the RSN: instead of potentially connecting each note with each other note, notes only need to be connected in maximum to one note per pitch, the syntagmatic anchor, usually the latest – or the soon-to-be – played note on that particular pitch. The RSN can therefore be simply represented as a matrix [3].

The definition of the RSN is highly dependent on the specification of the temporal scope of syntagmatic retentions. In other words, once a note has been played, how long will it remain active in memory so that it get connected to the subsequent notes? What can provoke an interruption of the retention? Can it be reactivated afterwards?

One main factor controlling syntagmatic retention is modality: the retention of a pitch remains active as long as the pitch remains congruent within the modal framework that is developing underneath. We propose a formalized model where the saliency of each syntagmatic connection is based on the modal configurations that integrate both notes of the connection, and more precisely

on the saliency of these modal configurations as perceived at both end of the connection (i.e., when each note is played).

4.2. Motivic Pattern Mining

An *ornamentation* of a motif generally consists in the addition of one or several notes -- the ornaments -- that are inserted in between some of the notes of the initial motif, modifying hence the composition of the syntagmatic surface. Yet, the ornamentation is built in such a way that the initial -- hence *reduced* -- motif can still be retrieved as a particular syntagmatic path in the RSN.

The challenge of motivic analysis in the presence of ornamentation is due to the fact that each repetition of a given motif can be ornamented in its own way, differing therefore in their syntagmatic surface. The motivic identity should be detected by retrieving the correct syntagmatic path that corresponds to the reduced motif. Motivic analysis is hence modelled as a search for repeated patterns along all the paths of the syntagmatic network [5].

We proposed a method for comprehensive detection of motivic patterns in strict monodies, based on an exhaustive search for closed patterns, combined with a detection of cyclicity [5]. That method was restricted to the strict monody case, in the sense that all motifs are made of consecutive notes. The closed pattern method relies on a definition of specific/general relationships between motifs. In the strict monody case, a motif is more general than another motif if it is a prefix, or a suffix, or a prefix of suffix, of the other motif. The application of this comprehensive pattern mining framework to the analysis of RSNs requires a generalization of this notion of specific/general relationships that includes the ornamentation/reduction dimension.

Figure 3 shows a theoretical analysis of a transcription of the first part of the Nay flute improvisation. The lines added in the score show occurrences of motivic patterns. Two main patterns are induced, as shown in Figure 4:

- The first line of Figure 4 shows the main pattern that is played in most of the phrases in the improvisation, and based on an oscillations between two states centered respectively around A (added with Bb, and represented in green) and G (with optional F, and represented in red), concluded by a descending line, in black, from A to D. This descending line constitutes the emblematic patterns related to the *Mhayyer Sikâ maqam*, and can be played in various degrees of reduction through a variety of different possible traversals of the black and purple syntagmatic network.
- The second line shows a phrase that is repeated twice in the improvisation – plus another more subtle occurrence – and based on an ascending (blue) line followed by the same paradigmatic descending line aforementioned.

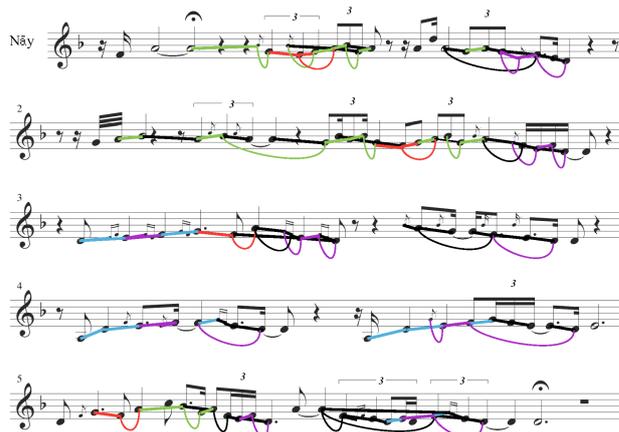


Figure 3. Motivic analysis of the first part of the improvisation. The lines added in the score show occurrences of motivic patterns, described in Figure 4.

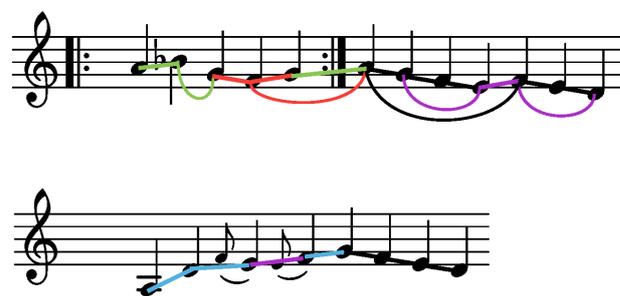


Figure 4. Motivic patterns inferred from the analysis of the improvisation shown in Figure 3.

4.3. Impact of Motivic Analysis on Transcription

Motivic analysis plays a core role in rhythmic analysis, not only for measured music, but also in order to take into account the internal pulsation that develop throughout the unmeasured improvisation. Successive repetition of a same rhythmic and/or melodic pattern are represented with similar rhythmic values. In our case, for instance, motivic repetitions help suggests a regularity of rhythmical sequences such as A C Bb C A / G Bb A Bb G in stave 2, or D / E D E / F E F / G F G at the beginning of stave 3. The motivic analysis enables in particular to track the rhythmical similarities despite any accelerandi (which often happens when such motives are repeated successively).

Another reason why pure bottom-up approaches for music transcription does not always work is due to the existence of particular short parts of the audio signal that cannot be analyzed thoroughly without the guidance of other material developed throughout the music composition or improvisation. For instance, a simple vibrato around one note might sometimes, through a motivic analysis, be understood as a transposed repetition of a recently played motif [2].

5. COMPUTATIONAL FRAMEWORK

The MiningSuite is a new platform for the analysis of music, audio and signal currently developed by Lartillot in the *Matlab* environment [4]. One module of *The Min-*

ingSuite, called *MusiMinr*, enables to load and represent in Matlab symbolic representations of music such as scores. It also integrates an implementation of the algorithm that automatically constructs the syntagmatic network out of the musical representation. Modes can also be specified, in order to enable the modal analysis and the specification of the RSN. Motivic analysis can also be performed automatically.

MusiMinr also integrates a module that performs transcription of audio recordings of pieces of music into score representations. Actually, the whole musical analysis is progressively performed, including the syntagmatic, modal and motivic analyses, in the same time as the transcription itself. In this way, higher-level musical knowledge, such as the expectation of a given modal degree or a motivic continuation, is used to guide the transcription itself.

Acknowledgments

This research is part of a collaborative project called “*Creativity / Music / Culture : Analysis and Modelling of Creativity in Music and its Cultural Impact*” and funded for three years by the French *Agence Nationale de la Recherche* (ANR) under the program “Creation: Processus, Acteurs, Objets, Contextes”.

6. REFERENCES

- [1] O. Lartillot, M. Ayari, "Cultural impact in listeners' structural understanding of a Tunisian traditional modal improvisation, studied with the help of computational models," in *J. Interdisciplinary Music Studies*, 5-1, 2011, pp. 85-100.
- [2] O. Lartillot, “Computational analysis of maqam music: From audio transcription to musicological analysis, everything is tightly intertwined,” in *Proc. Acoustics 2012 Hong Kong*.
- [3] O. Lartillot, M. Ayari, “Prolongational Syntagmatic Network, and its use in modal and motivic analyses of maqam improvisation,” in *Proc. II International Workshop of Folk Music Analysis*, 2012.
- [4] O. Lartillot, “A comprehensive and modular framework for audio content extraction, aimed at research, pedagogy, and digital library management,” in *Proc. 130th Audio Engineering Society Convention*, London, 2011.
- [5] O. Lartillot, “Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering,” *J. New Music Research*, 2005, 34-4, pp. 375-393.
- [6] P. Boersma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” *IFA Proc.*, 1993, 17, pp. 97–110.
- [7] G. Peeters, “Music pitch representation by periodicity measures based on combined temporal and spectral representations,” *Proc. ICASSP*, 2006.
- [8] M. Tolonen, M. Karjalainen, “A computationally efficient multipitch analysis model,” *IEEE Trans. Speech and Audio Proc.* 2000, 8, 708–716.