PhD proposal: Multi-scale carriers, residual modulation and musical structure

This PhD project aims to develop an informational framework for modeling and multi-scale segmentation of musical data. It is based on the hypothesis that musical structure can be described as one or several multi-scale information carriers, modulated by local variations, contrasts, and structural residuals.

Segmentation is then reformulated as a blind demodulation problem: estimating the carriers that economically explain the musical material, then analyzing residual modulations as elements of variation, transition, or rupture.

The project is situated within the continuity of MIR (Music Information Retrieval) research on automatic music structure analysis.

Approaches based on novelty and self-similarity matrices have shown that musical boundaries can be detected from local changes in the internal geometry of the signal [1]. Surveys in the field have then organized music structure analysis around principles such as repetition, novelty, and homogeneity [2], [3].

Recent work, such as that of Marmoret et al., extends this tradition through barwise segmentation methods based on self-similarity matrices [4].

The present project does not depart from these approaches; rather, it seeks to reformulate them within a framework where a boundary is not only a local discontinuity, but the sign of a reconfiguration of the structural carrier.

This hypothesis also has strong musicological foundations.

- The Schenkerian tradition provides a precedent for conceiving the musical surface as the elaboration of a deeper carrier structure, although this model remains tied to Western tonal art music [5].
- Lerdahl and Jackendoff's "Generative Theory of Tonal Music" (GTTM) contributes the notion of grouping, that is, the hierarchical organization of musical events into nested groups [6]. The project may be understood as a relational extension of this idea: elements grouped around the same carrier need not be contiguous in the temporal stream, but may be linked by similarity, transformation, metrical homology, contrast, or informational dependency.
- Meyer's and Narmour's theories add a cognitive framework based on expectation, realization, and deviation [7], [8].
- Finally, Bimbot, Deruty, Sargent and Vincent's "System & Contrast" model provides a particularly direct precedent by already distinguishing between a carrier system and contrastive elements [9].

From the computational point of view, the project relies on Guichaoua's models of compression and complexity, which link the inference of musical structure to criteria of economy, entropy, and information theory [10].

It also builds on the work of Louboutin and Bimbot on polytopic graphs of latent relations, including their application to chord sequences and hypercubic supports [11], [12]. Louboutin's thesis extends this approach by proposing a multi-scale and multi-dimensional modeling of musical structure through polytopic graphs [13].

These works legitimate the idea that a musical segment can be represented not only as a temporal sequence, but as a relational configuration carried by a combinatorial support.

Methodologically, the PhD project aims to combine matrix transformations of incidence, contrast, and coboundary. Möbius-type transforms make it possible to extract proper or differential contributions; Hadamard-type transforms measure structured contrasts; coboundaries generalize these approaches by localizing relational tensions on combinatorial supports such as chains, grids, hypercubes, systems of faces, or polytopic graphs.

These operators will provide an explicit and interpretable framework for separating carrier and modulation.

The approach may also be related to methods of factorization and blind component separation. Lee and Seung's NMF provides a precedent for decomposition into interpretable parts [14], while Comon's ICA is a foundational reference for the blind separation of latent components [15].

In the present project, however, the goal is not to separate independent sources, but to estimate multi-scale structural regularities and their deviations. Blind demodulation is therefore understood here as a paradigm for the joint estimation of a latent carrier and its modulations.

The expected advances are threefold.

- First, a theoretical advance: formalizing the notion of multi-scale carrier as an informational, relational, and combinatorial object.
- Second, an algorithmic advance: building explicit methods, then learnable ones through deep learning, to estimate carriers and residuals in order to characterize musical segments.
- Finally, a musicological advance: producing interpretable segmentations that describe not only where a boundary appears, but what is maintained, transformed, modulated, or broken within the musical structure.

The specification and evaluation of the developed methods will be carried out in collaboration with partners of the ANR MusiScale project, which aims to analyze the multi-scale structure of musical data, from raw audio signals to scores, and to develop methods accounting for relations between similar elements at different scales [16].

The evaluation protocol, test data, and qualitative criteria will be defined within this collaborative framework, from both MIR and musicological perspectives.

A long version of this PhD proposal is available on request from frederic.bimbot@irisa.fr

References

[1] Foote, J. (2000). Automatic audio segmentation using a measure of audio novelty. Proceedings of ICME 2000.

[2] Paulus, J., Müller, M., & Klapuri, A. (2010). Audio-based Music Structure Analysis. ISMIR 2010.

[3] Nieto, O., Mysore, G. J., Wang, C.-i., Smith, J. B. L., Schlüter, J., Grill, T., & McFee, B. (2020). Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications. Transactions of the International Society for Music Information Retrieval, 3(1), 246–263.

[4] Marmoret, A., Cohen, J. E., & Bimbot, F. (2023). Barwise Music Structure Analysis with the Correlation Block-Matching Segmentation Algorithm. Transactions of the International Society for Music Information Retrieval, 6(1), 167–185.

[5] Schenker, H. (1935/1979). Der freie Satz / Free Composition. First published in 1935; English translation edited by Ernst Oster, 1979.

[6] Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. MIT Press.

[7] Meyer, L. B. (1956). Emotion and Meaning in Music. University of Chicago Press.

[8] Narmour, E. (1990, 1992). The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model; The Analysis and Cognition of Melodic Complexity. University of Chicago Press.

[9] Bimbot, F., Deruty, E., Sargent, G., & Vincent, E. (2016). System & Contrast: A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces. Music Perception, 33(5), 631–661.

[10] Guichaoua, C. (2017). Modèles de compression et critères de complexité pour la description et l'inférence de structure musicale. PhD thesis, Université Rennes 1.

[11] Louboutin, C., & Bimbot, F. (2017). Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments. MCM 2017.

[12] Louboutin, C., & Bimbot, F. (2017). Modeling the Multiscale Structure of Chord Sequences Using Polytopic Graphs. ISMIR 2017.

[13] Louboutin, C. (2019). Multi-scale and multi-dimensional modelling of music structure using polytopic graphs. PhD thesis, Université Rennes 1.

[14] Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by Non-negative Matrix Factorization. Nature, 401, 788–791.

[15] Comon, P. (1994). Independent Component Analysis, A New Concept? Signal Processing, 36(3), 287–314.

[16] ANR MusiScale. Structure musicale multi-échelles. Project ANR-25-CE38-7645.