Probabilistic Multilabel Graphical Modelling of Motif Transformations in Symbolic Music

Ron Taieb* Barak Sober† Yoel Greenberg‡

Abstract

This paper presents a statistical framework to quantify motivic transformations in symbolic music, supporting robust computational analysis of musical style. We leverage the Beethoven Piano Sonatas Dataset (BPSD), which provides MusicXML encodings enriched with functional harmonic labels (BPS-FH) and motivic annotations (BPS-Motif). As part of our data handling, we aligned MIDI-based motivic annotations with the MusicXML data, creating a unified symbolic corpus.

We define coarse transformation families (e.g., contour-preserving, rhythm-preserving, harmony-preserving, salient leap-preserving) as binary labels for each instance of a motif. Each sonata is segmented into sentences using harmonic and structural cues, while ensuring that motifs are kept whole and not divided across boundaries.

Within each sentence, transformations are evaluated per motif group. The first motif in the sentence serves as the reference point for labeling its group, chosen for cognitive and musicological reasons: initial statements are usually perceptually salient, sometimes reflect continuity with motifs in the previous sentence due to their position, and are naturally heard as prominent simply by virtue of being presented first—especially in classical styles. All other motifs in the group are labeled relative to this initial motif.

To explore how transformation types relate and co-occur across musical contexts, we leverage probabilistic and graph-based methods that characterize interdependencies among motivic variations. Rather than focusing on prediction, our goal is descriptive: to support musicological interpretation of transformation patterns and expose structure in compositional strategies. Aggregating across sentences, we identify stylistic tendencies that recur across sonatas and may inform future generative modeling.

While existing methods often prioritize classification or automatic generation, our approach aims to model the organization of motivic transformations in a way that is both interpretable and musically grounded.

*Department of Statistics and Data Science, Hebrew University of Jerusalem. ron.taieb@mail.huji.ac.il

†Department of Statistics and Data Science; Center for Digital Humanities, Hebrew University of Jerusalem. barak.sober@mail.huji.ac.il

‡Department of Musicology, Hebrew University of Jerusalem. yoel.greenberg@mail.huji.ac.il

References

[1] Chen, Tsung-Ping, and Li Su. 2018. "Functional Harmony Recognition of Symbolic Music Data with Multi-Task Recurrent Neural Networks." In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, 90–97. Paris, France.

[2] Hsiao, Yo-Wei, Tzu-Yun Hung, Tsung-Ping Chen, and Li Su. 2021. "BPS-Motif: A Dataset for Repeated Pattern Discovery of Polyphonic Symbolic Music." In *Proceedings of the 22nd International Society for Music Information Retrieval Conference (ISMIR)*, 281–288. Milan, Italy.

[3] Karystinaios, Emmanouil, and Gerhard Widmer. 2022. "Cadence Detection in Symbolic Classical Music Using Graph Neural Networks." In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*, 917–924. Bengaluru, India. arXiv preprint arXiv:2208.14819.

[4] Lafferty, John, Andrew McCallum, and Fernando Pereira. 2001. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, 282–289. Williamstown, MA. `https://doi.org/10.48550/arXiv.cs/0108006`.

[5] Yang, Jaewon, and Jure Leskovec. 2013. "Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach." In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM)*, 587–596. Rome, Italy. `https://doi.org/10.1145/2433396.2433471`.

[6] Yang, Ruihan, Dingsu Wang, Ziyu Wang, Tianyao Chen, Junyan Jiang, and Gus Xia. 2019. "Deep Music Analogy via Latent Representation Disentanglement." In *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*, 596–603. Delft, Netherlands.

[7] Zeitler, Johannes, Christof Weiß, Vlora Arifi-Müller, and Meinard Müller. 2024. "BPSD: A Coherent Multi-Version Dataset for Analyzing the First Movements of Beethoven's Piano Sonatas." *Transactions of the International Society for Music Information Retrieval (TISMIR)* 7 (1): 195–212. `https://doi.org/10.5334/tismir.196`.