

Task-Oriented Evaluation for Automated Music Simplification

Andrew McLeod, Hanna Lukashevich, Jakob Abeßer

The task of music simplification involves the conversion of a piece of music into a version that is, in some sense, easier or simpler [1]. While some models have been proposed for the task, no standardized evaluation metric or procedure has yet been established. We argue that the reason for this is the wide range of use cases for which music simplification could be applied.

A major class of use cases for music simplification is that of making music easier to play, with different goals: allowing a musician with a lower skill level to play something that still captures the essence of the full piece; helping a musician to learn the full version of a piece by learning progressively more difficult versions [2]; or improving music accessibility by, e.g., reducing hand span for piano or limiting the fret range of a piece for guitar.

A second class involves the abstraction of a piece of music for the purpose of a deeper understanding or analysis (e.g., [3]). Such a simplification might result in a version which could help musicologists in producing a harmonic analysis of a given piece, or a version which allows performers and accompanists to better improvise along with it.

Finally, a third class considers the use of music simplification for the purpose of some downstream MIR task. Examples include query-by-humming [4], where a user might query using a simplified version of the desired melody; or music generation, where (pre-)training a model on abstracted versions might help to avoid picking up genre-specific biases from the training data.

Each of these use cases and goals is distinct, and music simplification models designed for one might perform poorly on another. We argue therefore, that use case and goal specific evaluation procedures are required. We consider how this might look for each with the help of initial results from a music simplification model.

References

- [1] P. Ramoneda, M. Suzuki, A. Maezawa, and X. Serra, “Difficulty-aware score generation for piano sight-reading,” *Expert Systems with Applications*, vol. 321, p. 132088, 2026, ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2026.132088>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417426010018>.
- [2] P. Ramoneda, E. Parada-Cabaleiro, D. Jeong, and X. Serra, *Difficulty-controlled simplification of piano scores with synthetic data for inclusive music education*, 2025. arXiv: 2511.16228 [cs.SD]. [Online]. Available: <https://arxiv.org/abs/2511.16228>.
- [3] P. Ericson, Y. Rammos, and M. Rohrmeier, “Musereduce: A generic framework for hierarchical music analysis,” in *Music Encoding Conference 2022, Halifax, Canada, May 19-22, 2022*, Music Encoding Initiative, 2023, pp. 40–51.
- [4] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima, “A practical query-by-humming system for a large music database,” in *Proceedings of the eighth ACM international conference on Multimedia*, 2000, pp. 333–342.