ms3: A parser for MuseScore 3 files and data factory for annotated music corpora.

Johannes Hentschel ⓘ [1][*][¶] **and Martin Rohrmeier** ⓘ [1][*]

1 École Polytechnique Fédérale de Lausanne, Switzerland ¶ Corresponding author * These authors contributed equally.

Summary

Digital Musicology is a vibrant and quickly growing discipline that addresses traditional and novel music-related research questions with digital and computational means (Honing, 2006; Huron, 1999; Urberg, 2017). Research questions and methods often overlap with or draw on those from diverse disciplines such as music theory and analysis, composition history, mathematics, cognitive psychology, linguistics, anthropology, or computer science (Volk & Honingh, 2012; Wiggins, 2012). Corpus research, i.e., the computational study of representative collections of texts (in the case of linguistics) or compositions (in musicology), plays a prominent role in this trans-disciplinary quest to "make sense of music" through scientific models (London, 2013).

At the heart of corpus-driven methods into tonal music (roughly speaking, Western music since the late 16th century) lies the score, a symbolic representation of the relations between sound events over time. Building on notational conventions that have emerged over the course of circa 1,000 years, they constitute a well-proven cultural practice of mapping events from the continuous space of physical time and frequency to an idealized, virtual, discrete space of musical time (perceived temporal bins) and pitch (perceived frequency bins), that bears witness of categorization as a prevalent principle of human cognition. Much like texts, scores represent both a means of communicating music to readers and performers (who transduce them into cognitive processes and sound), and a conventionalized way to reduce, bin, quantize, and protract the complex information inherent to physical, analog, or digital audio signals and musical thought (inner hearing). Apart from often being the only surviving trace of music from earlier centuries, scores encode additional semantic information on sounding events that is unavailable in their physical signal, e.g., the intended or ascribed relation of a pitch to its surrounding pitches (paradigmatic relations), and to musical time (syntagmatic relations). Consequently, a score can also be viewed as an abstraction over all possible ways a human can interpret and actuate it, and hence is an indispensable tool for encoding, communicating, and comparing its various interpretations and recordings thereof.

Understanding a score as an abstract relational model of the composition or musical utterance it represents sheds light on its essential role for interrelating heterogeneous data from and for performance research, music analysis, neuroscience, stylometry, music psychology, ethnography, or music information retrieval (Abdallah et al., 2017; Cook, 2005). Whereas text corpora for the long-established discipline of computational linguistics are abundant, the advent of similarly large curated datasets of symbolically encoded (as opposed to scanned) digital scores yet awaits (Shanahan, 2022).

Statement of need

Acknowledgements

References

Abdallah, S., Benetos, E., Gold, N., Hargreaves, S., Weyde, T., & Wolff, D. (2017). The Digital Music Lab: A Big Data Infrastructure for Digital Musicology. *Journal on Computing and Cultural Heritage*, *10*(1), 1–21. https://doi.org/10.1145/2983918

Cook, N. (2005). *Towards the compleat musicologist?*

Honing, H. (2006). On the Growing Role of Observation, Formalization and Experimental Method in Musicology. *Empirical Musicology Review*, *1*(1), 2–6. https://doi.org/10.18061/1811/21901

Huron, D. (1999). *The New Empiricism: Systematic Musicology in a Postmodern Age* (Music and {{Mind}}: {{Foundations}} of {{Cognitive Musicology}} No. 3). The 1999 Ernest Bloch Lecture.

London, J. (2013). Building a Representative Corpus of Classical Music. *Music Perception*, *31*(1), 68–90. https://doi.org/10.1525/mp.2013.31.1.68

Shanahan, D. (2022). What the history of computational musicology can tell us about the future of corpus studies. *Future Directions of Music Cognition Virtual Speaker Series*. https://doi.org/10.17605/OSF.IO/CB26R

Urberg, M. (2017). Pasts and Futures of Digital Humanities in Musicology: Moving Towards a "Bigger Tent." *Music Reference Services Quarterly*, *20*(3-4), 134–150. https://doi.org/10.1080/10588167.2017.1404301

Volk, A., & Honingh, A. (2012). Mathematical and computational approaches to music: Challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music*, *6*(2), 73–81. https://doi.org/10.1080/17459737.2012.704154

Wiggins, G. A. (2012). Music, mind and mathematics: Theory, reality and formality. *Journal of Mathematics and Music*, *6*(2), 111–123. https://doi.org/10.1080/17459737.2012.694710