

Let's Formalize Music Notation for OMR

Organization

- **Jorge Calvo Zaragoza**, University of Alicante, Spain
- **Jan Hajič Jr.**, Charles University, Czech Republic
- **Alexander Pacha**, TU Wien, Austria

Duration: full day

Logistical requirements: tables, projector, power plugs

Technical requirements: Internet connection

Workshop description

Optical music recognition (OMR) has undisputedly made significant progress in recent years, but it is still difficult to leverage this progress in the various application scenarios. Among the obstacles are a dearth of interoperability between individual OMR contributions and systems: it is fine that system A, for instance, detects notation symbols better than system B, but system A cannot easily be swapped into another OMR pipeline because it either uses a slightly different set of objects, or at least represents the output differently.

Despite a long history of efforts, there is also prominent lack of accepted OMR standards — not just evaluation metrics. While there are multiple ongoing initiatives to improve the way how music notation is engraved (SMuFL) and encoded (MEI), OMR has specific needs, such as a clear distinction between graphics and semantics or representing syntactically incorrect scores. While this situation has been recently improved with the introduction of extensive datasets¹ that have elaborate specifications, the dataset builders have resolved the issues involved in designing a data model for OMR differently.

We believe that the underlying problem is, that there is no agreed-upon computer standard for describing music notation (a set, a sequence, a graph, a formal grammar, etc.). Until one has defined such a data structure, one cannot define, e.g. distance functions acting over pairs of such data structures.

We intend the workshop to be an open forum to discuss how music notation can be formalized in computers (and, consequently, how OMR can be evaluated within this formalism). Participants should bring along some basic knowledge of OMR research, be comfortable with music notation and its terminology, and ideally be aware of common computer data structures. The workshop will start with a presentation about the topic and the open problems, followed by

¹ <https://apacha.github.io/OMR-Datasets/>

discussion groups. Attendees are expected to actively participate in these discussions and contribute to the decisions to be taken. In the second half of the workshop, we plan to bring the outcomes of the discussion into practice and work on specifications and actual code.

The Music Encoding Conference represents an ideal opportunity to get this kind of work done since a significant portion of both the OMR community and the MEI community will be present. On the one hand, there is sustained interest in OMR from the music library community because it reduces the cost of digitizing written music. On the other hand, OMR formalization should not be developed (only) by people on the technical side, but by all stakeholders. We, therefore, invite all kind of people that have an interest in the development of the OMR field to attend this workshop and contribute with new ideas.