



SCORIA

an international, collaborative BSc/MSc thesis project to create an open-source music scorification pipeline!



Institute of
Computational
Perception


PragueMusicComputingGroup



**Charles
University**

JKU

JOHANNES KEPLER
UNIVERSITY LINZ

WHAT IS “SCORIFICATION”?

When you play on a MIDI keyboard, there are signals being produced about which key you pressed at what time. Recording these signals produces a MIDI file, a symbolic music recording. These files are often used for playback, storage, or music production, but if you want to share the piece with someone else such that they can play it as well, passing just this information is not enough for the other musician to be able to read it. The written standard for communicating your music is not MIDI but music notation, where the musical material is quantized and augmented with several other attributes to make it accessible to musicians versed in music notation.

If a MIDI recording is not enough, how do you get to music notation? You could of course tediously type out all notes in a notation editor. Can notation editors not just open MIDI files? Well yes, they can, but if you ever tried to open a MIDI file you recorded in a music notation editor, you know that they generally struggle to create a readable score. You're more likely to get a visual like this:



a confusing representation of a performance when the actual score looks like this:



This is where music scorification comes in: scorification is the process to get from a symbolic MIDI recording to a readable score, that is, to estimate the quantization of onsets and offsets into musical time and to estimate meter, key, voice, and several more attributes. Scorification is a hard technical problem for which no satisfactory solutions exist yet.

WHAT IS “SCORIA”?:

SCORIA is a scorification project. We aim to collaboratively create a complete pipeline to automatically and beautifully typeset scores from recorded MIDI data. This is a huge undertaking encompassing many music information retrieval tasks. To turn it manageable, we break this goal down into challenging subtasks, each addressed in its own thesis. A thesis topic for each of the following subtasks is available, broken down into time related tasks (MIDI normalization, on the left) and score attribute related tasks (MIDI Scorification):

Task Group	MIDI Normalization:	MIDI Scorification:
API	<p>File I/O: python / partitura</p> <p>Module Input: partitura performance note array / numpy structured array (time unit: seconds)</p> <p>Module Output: partitura score note array / numpy structured array (time unit: quarters)</p>	<p>File I/O: python / partitura</p> <p>Module Input: partitura score note array / numpy structured array (pitch and time information only)</p> <p>Module Output: partitura score note array / numpy structured array (with added columns encoding the estimated attributes)</p>
Individual Tasks	<p>Free Performance Quantization (FP)</p> <p>“Performance to Click” Quantization (PC)</p> <p>MuseScore MIDI Normalization (MS)</p>	<p>Meter / Barline / Tie Estimation (MBT)</p> <p>Key / Pitch Spelling Estimation (KPS)</p> <p>Beaming / Grouping Estimation (BG)</p> <p>Clef / Staff Estimation (CS)</p> <p>Stemming / Voice Estimation (SV)</p> <p>Ornamentation Classification (OC)</p>

ORGANIZATION:

October 24: We are currently looking for students to join SCORIA for their thesis and address each task!

The goal of each thesis is to:

- write a comprehensive overview of both traditional and machine learning based approaches to the topic
- (re)implement or adapt one state-of-the-art algorithm to work within the API specification of SCORIA
- finish at latest by September 2025, when we'll assemble all parts and release the scorification tool

We offer you:

- interesting and challenging music AI tasks
- supervision by experts in the field
- collaborative work with peers across Czechia and Austria
- lasting contributions to an open source project beyond your own thesis
- a specified API for the pipeline
- a utility codebase for symbolic music processing python and data which gets you started on the task right away

We need you to:

- have a strong background in AI/ML/music computing
- be interested in music scores and their intricacies, and you must be able to read music
- be able and willing to complete a thesis within the time allotted.
- be interested in scorification!

If you're interested in a task or thesis or just need more information, please reach out to:

Project Website: <https://cpjku.github.io/scoria>

Institute of Computational Perception / Johannes Kepler University:

coordination: Silvan Peter, silvan.peter@jku.at

Institute website: <https://www.jku.at/en/institute-of-computational-perception/>

Prague Music Computing Group / Charles University:

coordination: Jan Hajič jr., hajicj@ufal.mff.cuni.cz

Institute website: <https://ufal.mff.cuni.cz/prague-music-computing-group>

