Scriptoria

A Crowd-powered Music Transcription System

Samiotis, Ioannis Petros; Lofi, Christoph ; Alaka, Shaad; Liem, Cynthia C. S.; Bozzon, Alessandro

DOI
10.1145/3487553.3524252

Publication date
2022

Document Version
Final published version

Published in
WWW 2022 - Companion Proceedings of the Web Conference 2022

Scriptoria: A Crowd-powered Music Transcription System

Ioannis Petros Samiotis
Delft University of Technology
Delft, Netherlands
i.p.samiotis@tudelft.nl

Christoph Lofi
Delft University of Technology
Delft, Netherlands
c.lofi@tudelft.nl

Shaad Alaka
Delft University of Technology
Delft, Netherlands
s.alaka@student.tudelft.nl

Cynthia C. S. Liem
Delft University of Technology
Delft, Netherlands
c.c.s.liem@tudelft.nl

Alessandro Bozzon
Delft University of Technology
Delft, Netherlands
a.bozzon@tudelft.nl

ABSTRACT

In this demo we present Scriptoria, an online crowdsourcing system to tackle the complex transcription process of classical orchestral scores. The system's requirements are based on experts' feedback from classical orchestra members. The architecture enables an end-to-end transcription process (from PDF to MEI) using a scalable microtask design. Reliability, stability, task and UI design were also evaluated and improved through Focus Group Discussions. Finally, we gathered valuable comments on the transcription process itself alongside future additions that could greatly enhance current practices in their field.

CCS CONCEPTS

• **Information systems → Crowdsourcing**; *Digital libraries and archives*; • **Applied computing → Sound and music computing**; • **Human-centered computing → Usability testing**.

KEYWORDS

crowdsourcing, music transcription, focus group discussions, iterative design, digital archives

ACM Reference Format:
Ioannis Petros Samiotis, Christoph Lofi, Shaad Alaka, Cynthia C. S. Liem, and Alessandro Bozzon. 2022. Scriptoria: A Crowd-powered Music Transcription System. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3487553.3524252

1 INTRODUCTION

Music transcription is a challenging topic in computer vision. Despite the latest improvements in Optical Music Recognition (OMR), professional orchestras still rely heavily on manual transcription of orchestral music scores due to their length and complexity (e.g. multiple parallel instruments with varying notations, bad quality of source material, hand-written annotation, etc.). Improving the transcription process of classical music scores would be a significant contribution to ongoing efforts to preserve this type of valuable cultural inheritance.

Inspired by recent works in microtask crowdsourcing [1, 8] and *nichesourcing* [6], we conducted requirement analysis with experts and designed Scriptoria, a crowd-powered music transcription system. The system consists of multiple modules which process incoming PDF files of scanned scores, process them and segment them into smaller parts. Each segment is then annotated through a crowdsourcing pipeline of incremental transcription. The results are then aggregated and published in an online repository.

We evaluated our system with Focus Group Discussions with members of Dutch youth orchestras in two iterations. Between these iterations, we made improvements to our system based on the feedback we received. In this paper, we finally present some of the most valuable insights gathered through our discussions with the participants.

Our work has parallels to studies such as [2], the *Allegro* system, and [3], which focused on user input and task design. However, both studies focus on single-user transcription, while our workflows allow many contributors to participate for scalability [7].

2 REQUIREMENTS AND DESIGN

We present Scriptoria in two parts: (a) the back-end architecture of the transcription pipeline and processing modules; and (b) the task interfaces that users interact with. Both the back-end and front-end of this crowdsourcing system are hosted on the Dutch national e-infrastructure with the support of SURF Cooperative. The source code is published on GitHub[1,2].

2.1 System Architecture

Our back-end accommodates a crowd-assisted OMR pipeline. It processes PDF input data (image processing and segmentation), generates crowdsourcing tasks for the non-automated parts, and finally aggregates results to build an MEI version of the original orchestral music score (see Figure 1).

Core system requirements for our prototype were to: (1) design the system in a modular and distributed fashion and (b) store in the system all the data resulting from processes throughout each of the steps of our music transcription pipeline, to make them easily accessible by all the system's modules. We set the first requirement to enable scalability and support easier maintainability. Each of the

[1]https://github.com/cakefm/crowd_task_manager
[2]https://github.com/cakefm/scriptoria

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04.
https://doi.org/10.1145/3487553.3524252

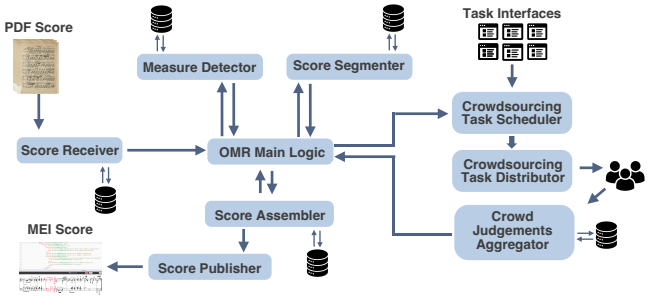


Figure 1: Schema of Crowd Task Manager's modules.

modules in the prototype represents a step on the transcription pipeline and serves a specific functionality. This helps to easily replace parts of the pipeline with more sophisticated ones, without breaking the overall operability of the system. We implemented a central module which holds the logic steps of the transcription pipeline which sends messages that dictate which of the modules should be activated and when. Each module inside the pipeline imports data from our local database and stores data to it to make them available to the other modules.

When a PDF score is sent to the back-end, first rasterisation is performed followed by some standard image pre-processing. First, the contrast of the page is maximized, after which the page is binarized. Following this, any rotations in the page that might have occurred due to scanning of the original score are rectified. Following these steps, a top-down approach is followed in analyzing the page structure. First, systems will be separated, which are subsequently segmented into vertical blocks, which are then segmented into measures. Each segment is stored in a MongoDB database, alongside their identifiers and they become available to the front-end through an API.

Consulting an expert from the Royal Concertgebouw Orchestra[3] (RCO) of Amsterdam, we identified the most important elements of an orchestral music score, alongside the minimum-viable-product requirements for a final transcribed score. The music notations we focus in our transcription pipeline where: (a) clefs, (b) time signatures, (c) key signatures, (d) rhythmic information of notes, (e) pitch information of notes. We then broke down the transcription pipeline into consecutive tasks that can be conveyed in a microtask crowdsourcing fashion, focusing on those individual important music elements. The tasks we designed were:

- **Clef recognition**: indicate whether one or multiple clefs are visible in a segment, and if so, which;
- **Time signature recognition**: indicate whether a time signature indication is visible in a segment, and if so, which;
- **Key recognition**: indicate whether a key signature indication is visible in a segment, and if so, which;
- **Rhythm transcription**: transcribe the rhythm of the musical content in a segment;
- **Pitch transcription**: transcribe the pitches of the musical content in a segment.

[3]https://www.concertgebouworkest.nl/en

Each task has specific inputs (segments of the given score) and outputs (music notations), designed to be performed easily and efficiently by the users. As contributors work on the tasks, in the back-end, a score is built up in the open Music Encoding Initiative (MEI) format, and each contribution is stored as a commit on GitHub.

To allow for coherent completion, we implemented a scheduling algorithm which follows a hierarchy of importance for MEI elements. For each segment, the clef, key and time signatures are essential, as they could alter all subsequent music elements (notes/rests), which depend on them. These crowdsourcing tasks co-exist with automated methods such as measure detection, image segmentation and XML-tree aggregation, creating a hybrid system where human-machine collaboration achieves the shared goal of generating MEI orchestral pieces from PDF input.

2.2 Task Design and Interfaces

A dedicated front-end server was developed, to allow dynamic rendering of UI elements and dynamic route matching for the different types of tasks. This is based on a NodeJS server which hosts all the necessary components such as interfaces, UI elements and dedicated task type components, while handling communications with the back-end through Axios. The front-end can access each score segment through the back-end's API and renders their images dynamically on the browser. The user input is translated to MEI headers and communicated back to the back-end.

Our target contributors were semi-experts (youth/student members of classical orchestras), so we factored their expertise during the task design process. As described in Section 2.1, we designed a separate task type for each of the five music notations. The detection tasks for **clef**, **time signature** and **key signature**, presented the user with the original segment of the score (an image of a given measure) and they had to indicate the existence of the given music notation, while identifying its characteristics (e.g. if a clef exists, select its type). For the **rhythm** and **pitch** detection tasks, the user was presented with the image of the original segment to the left, so they can immediately compare their choices on the rendered MEI snippet to the right (see Figure 2).

Due to our contributors' expertise, a certain high level of input was expected. Their expertise, combined with majority voting aggregation and tree aligning algorithms, would ensure high quality of output, therefore rendering possible verification tasks inessential.

3 FOCUS GROUPS AND ITERATIVE DESIGN

We conducted focus group discussions with semi-experts and young professionals members of multiple classical youth to evaluate our workflow and task designs. During the discussions, we investigated current methods they employ to transcribe orchestral music scores, but also encouraged them to explore the requirements and workflows of a future, more feature-rich version.

Interviews with experts played a crucial role to the design of our transcription system. The initial design of our prototype was based on feedback and requirements received by a professional expert in the RCO and from the youth orchestra Krashna Musika[4]. For the focus group discussions, we reached out to several youth

[4]https://www.krashna.nl/en/

orchestras in the Netherlands, who were enthusiastic to participate. Following an iterative design methodology, we split the participants into two groups; the feedback received from the first group was used to update the designs in our transcription system and the second group was presented with the final version.

3.1 Recruitment

Due to the COVID-19 crisis, all studies were conducted through online videoconferencing. For both rounds of studies, a similar protocol was followed. First study was conducted in 5 sessions, with 30 participants in total. The second study took place 4 months later and it was conducted in 4 sessions, with 33 participants in total. The participants of both studies were members of Dutch youth orchestras, namely: Collegium Musicum[5], Quadrivium[6], NJO[7], Sweelinck[8], Nijmeegs Studentenorkest CMC[9], Amsterdams Studenten Orkest[10], S. M. G. 'Sempre Crescendo'[11] and Almeers Youth Symphony Orchestra[12].

3.2 Focus Group Structure

First, informed consent was asked of the participants and a musical background survey based on the Goldsmiths Musical Sophistication Index (Gold-MSI) [5] was conducted. After a round of introductions, the researchers discussed with the participants on current transcription practices and motivations behind the proposed workflow. Subsequently, the participants were invited to engage with the transcription system for an hour. During this hour, they would gradually complete the different tasks and go through the different task stages. The participants were asked to work individually, and only ask for help/clarification in case they really ran into technical problems; the researchers remained on the call for answering such questions. Finally, participants were invited to fill in a Post-Study System Usability Questionnaire (PSSUQ) [4] survey, and the researchers moderated a task-by-task discussion in which participants were encouraged to share qualitative feedback on their experiences and opinions on possible improvements.

All study sessions consider the first pages of Ludwig van Beethoven's Sextet in E-flat major, op. 71, where for the scanned score we used a PDF from the IMSLP[13]. The amount of transcription work was adjusted, based on the number of participants in each session.

4 EVALUATION AND IMPROVEMENTS

4.1 First study

As expected, the musical background survey indicated that many of the players had extensive musical instrument training, representing a considerable diversity of instrument experience. While this could indicate different expertise on specific music notations (such as types of clefs), nevertheless, the UI and transcription tasks of those notations were designed to be primarily based on visual recognition of similar artefacts.

In all sessions, participants managed to fully complete all tasks. We found that the rhythm and pitch transcription tasks are much more time consuming compared to the clef, key signature and time signature tasks. The qualitative feedback by the participants, indicated that the UI design of these tasks could still be further improved.

For the **time signature detection** task, participants suggested to include buttons for commonly occurring time signatures to further minimize the need for textual input. They also indicated that the **key signature detection** task was confusing for some, due to the high complexity of the annotation (key signatures can occur in multiple places, even in a small segments).

For the **rhythm transcription**, the UI was found cumbersome and it was suggested to expand to more buttons with common preset choices. Furthermore, the absence of note beams when transcribing, made the visual comparison between the reference and the entered input more difficult.

Regarding the **pitch transcription**, participants indicated that the task involved elaborate user input and that it would be useful to include shortcuts for common actions and input dragging. Furthermore, at the moment of these studies, the way the default pitch of the rhythm notes was registered during the previous task, was deemed insufficient to clearly visualise the notes in this task.

Execution time is estimated based on commit logs of the Git repository. Through the user evaluation we identified issues with this method for two main reasons: there may be time lag between subsequent commits due to the input volume processed by our system; results that did not alter the MEI snippet, where not committed (e.g. indicating no clef in given segment).

Finally, general feedback focused on the inconsistency of the 'submit' button's look and feel across tasks. Also, task instructions were found either unclear or too 'wordy'. These usability issues were also apparent in the PSSUQ survey results.

4.2 Implemented improvements

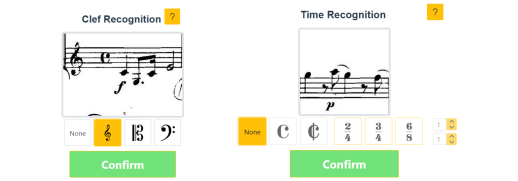
As noted in the previous section, there were some issues with the initial GUI designs that impacted user efficiency and the overall user experience. To rectify this, we implemented multiple changes.

For **time detection**, we added preset buttons with frequently occurring time signatures. For **key signature detection**, contributors can select the type of key signature, and click a button to increment the count, which will show a preview of the key signature. For **rhythm transcription**, a full redesign of the GUI was performed, replacing the slider with expanded preset buttons (see Figure 2). In addition, note navigation/deletion has been replaced by a single undo button. Finally, beam support has been added. For **pitch transcription**, octave adjustment buttons have been added, and notes get initialized in the middle of a staff, depending on the preceding clef. Furthermore, for note navigation, keyboard shortcuts were now implemented.

In terms of general improvements, the 'submit' button was standardized in terms of look, feel and location across tasks. Furthermore, the help text was replaced by a help button, launching a

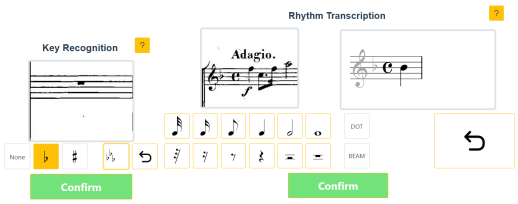
[5] https://www.collegiummusicum.nl/en/
[6] https://www.esmgquadrivium.nl
[7] https://www.njo.nl/english/orchestra/orchestra
[8] https://www.sweelinckorkest.nl
[9] https://www.nijmeegsstudentenorkest.nl
[10] http://www.amsterdamsstudentenorkest.nl/en/
[11] https://www.smgsemprecrescendo.nl
[12] https://www.stichtingajso.nl/english/ajso/
[13] https://imslp.org/wiki/Sextet_in_E-flat_major%2C_Op.71_(Beethoven%2C_Ludwig_van)

floating window, that shows an animation of how the task is supposed to be performed, along with a description. Finally, Verovio, the used score online editor, was no longer loaded for tasks that do not use any MEI preview, which improved loading times for the time, key signature and rhythm transcription tasks. During any interface loading, a loading progress indicator was also included. In the back-end, we also improved system logging, so more refined timing information could be included in our analyses (e.g. registering MEI snippet submissions with no alterations).



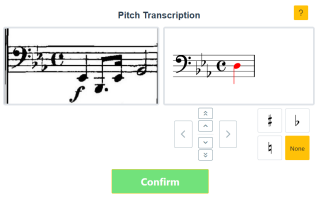
(a) Clef transcription

(b) Time signature transcription



(c) Key signature transcription

(d) Rhythm transcription



(e) Pitch transcription

Figure 2: Improved designs for the transcription tasks.

4.3 Second study

Following the results of the first study, the musical expertise of the participants was equally high and diverse in terms of instrument of choice. Following our improvements based on collected feedback, the participants in all sessions of the second study were very enthusiastic about our system. We identified major improvements in the amount of time spent per task, while also the both the feedback from discussions and the PSSUQ questions was positive.

Our improvements in the UI of **rhythm transcription** and **pitch transcription** tasks seemed to also assist better the users, to complete successfully their tasks.

5 INSIGHTS FOR FUTURE VERSIONS

From our discussions with the RCO experts and the members of youth orchestras, we received invaluable insights on the traditional transcription methods employed by professionals in classical orchestral music. We first-hand witnessed challenges, such as: messy handwritten annotations that obscure the printed music notations, imperfectly scanned pages and damaged scores. Although such challenges exist in all types of printed music scores, the length and complexity of orchestral pieces amplifies them, causing the automatic transcription methods to frequently fail. Professionals and amateurs alike, still lean on manually transcribing scores from the ground up, using dedicated software, online solutions or even in cases, pen and paper.

Participants in our Focus Groups discussed extensively their transcription habits. The majority of the youth orchestras relied on one or two people who transcribed the score for all the rest. Our microtask approach was happily welcomed and the participants indicated that a collaborative, task-based workflow could potentially improve productivity and the social bonding of a group.

Our discussions with the participants brought another valuable insight on the annotation needs of orchestras: almost unanimously the participants pointed towards sharing performance annotations between orchestras. When performing music pieces, each orchestra adds their own interpretation that can often be quite unique and separate from others. The potential of digitizing and sharing those performance annotations on top of the other music notations and sharing them between users of the transcription platform, was deemed to be a key future to its success.

ACKNOWLEDGMENTS

This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. We thank Marcel van Tilburg for his insights in orchestral music scores and Carlo van der Valk for his engineering contributions.

REFERENCES

[1] Alessandro Bozzon, Marco Brambilla, Stefano Ceri, and Andrea Mauri. 2013. Reactive crowdsourcing. In *Proceedings of the 22nd international conference on World Wide Web*. 153–164.

[2] Manuel Burghardt and Sebastian Spanner. 2017. Allegro: User-centered Design of a Tool for the Crowdsourced Transcription of Handwritten Music Scores. In *Proceedings of the 2Nd International Conference on Digital Access to Textual Cultural Heritage* (Göttingen, Germany) *(DATeCH2017)*. ACM, New York, NY, USA, 15–20. https://doi.org/10.1145/3078081.3078101

[3] Liang Chen and Christopher Raphael. 2017. Human-Directed Optical Music Recognition. *Electronic Imaging* 2016, 17 (feb 2017), 1–9. https://doi.org/10.2352/issn.2470-1173.2016.17.drr-053

[4] James R Lewis. 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors Society Annual Meeting*, Vol. 36. Sage Publications Sage CA: Los Angeles, CA, 1259–1260.

[5] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one* (2014).

[6] Jasper Oosterman, Alessandro Bozzon, Geert-Jan Houben, Archana Nottamkandath, Chris Dijkshoorn, Lora Aroyo, Mieke HR Leyssen, and Myriam C Traub. 2014. Crowd vs. experts: nichesourcing for knowledge intensive tasks in cultural heritage. In *Proceedings of the 23rd International Conference on World Wide Web*. 567–568.

[7] Ioannis Petros Samiotis, Christoph Lofi, and Alessandro Bozzon. 2021. Hybrid Annotation Systems for Music Transcription. In *3rd International Workshop on Reading Music Systems*.

[8] Ioannis Petros. Samiotis, Sihang Qiu, Andrea Mauri, Cynthia C. S. Liem, Christoph Lofi, and Alessandro Bozzon. 2020. Microtask crowdsourcing for music score Transcriptions: an experiment with error detection. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*.