Music recognition, encoding, and transcription (MuRET) online tool demonstration

David Rizo[1,2], Jorge Calvo-Zaragoza[1], Juan C. Martínez-Sevilla[1], Adrián Roselló[1], and Eliseo Fuentes-Martínez[1] ⋆

[1] Universidad de Alicante
[2] Instituto Superior de Enseñanzas Artísticas de la Comunidad Valenciana (ISEA.CV)
drizo@dlsi.ua.es

Abstract. Most of the musical heritage is only available as physical documents. Their mere availability as scanned images does not enable tasks such as indexing or editing unless they are transcribed into a structured digital format. Many transcription processes have been traditionally performed following a fully manual workflow. At most, it has received some technological support in particular stages, like optical music recognition (OMR), or transcription to modern notation with music edition applications. A new online tool named MuRET has been recently developed, which covers all transcription phases, from the manuscript image to an digital score. MuRET is designed as a machine-learning based research tool, allowing different processing approaches to be used, and producing both the expected transcribed contents in standard encodings and data for research activities. The objective of the demonstration is to showcase it for an efficient transcription process and provide guidelines on how to get the most out of it.

1 Description of the demonstration

MuRET is a research oriented optical music recognition tool (OMR) based on a series of machine learning techniques, mainly deep neural networks, that has been recently ported to be an online application [4] from the original desktop application proposal.

The demonstration will focus on showing all the possibilities that MuRET [4] offers and the process required to convert a series of input images into a digital score in MEI format. MuRET being a research tool, a discussion will be held on possible extensions of interest to the community. Specifically, the demo will consist of the following items:

1. Collection handling (Fig. 1a) in order to organize large corpora.
2. Section and images management (Fig. 1b), used to group, correctly ordering the images to be transcribed, and setup the correct nature of the document (parts based, incipits book, etc.).

⋆ This work has been supported by the Spanish Ministerio de Ciencia e Innovación through project MultiScore (No. PID2020-118447RA-I00), supported by UE FEDER funds.

3. Document analysis (Fig. 2) to show how the system has detected the different regions, namely, staves, title, and lyrics, and how user can edit them.

4. Part linking (Fig. 3) in order to let the system identify which instrument belongs each image or crop of the image.

5. Region contents recognition (Fig. 4) where the machine learning models identify the sequence of symbols contained in each region using different approaches depending on the content type, lyrics or music, or the granularity and interaction strategy. In this step, the recognized symbols are just graphical representations (denoted as *agnostic* representation in [2]) without musical meaning.

6. Music encoding of individual staves (Fig. 5) to obtain an actual music encoding of the agnostic representation obtained in the previous step. We will introduce the extension of the formats **kern and **mens used to accommodate layout information besides the musical content itself. The possibility of transliterate early notations into modern ones will be also explored.

7. Scoring up and exporting (Fig. 6) as the final step in a transcription project, where user can obtain a whole score from the different spread parts, and export a MEI file, either as a whole MEI file or divided into parts including facsimile information to be used by other tools such as MP Editor [3].

8. Offline model training and uploading to allow the user to use his/her already tagged collections to create new fine-tuned models.

9. User action logs analysis from interaction data to obtain the actual transcription times and study the real improvement of new models and approaches.

(a) User collections

(b) Two sections of a document of a complete opera shown at left column

Fig. 1: Work organization

At the end of the tutorial, attendees should understand the operation of MuRET and how systems based on machine learning can be interactively improved. Also, we hope that attendees will perceive how the use of MuRET, even without being able to guarantee absolute accuracy, significantly decreases the temporal cost of transcription compared to a completely manual process [1].

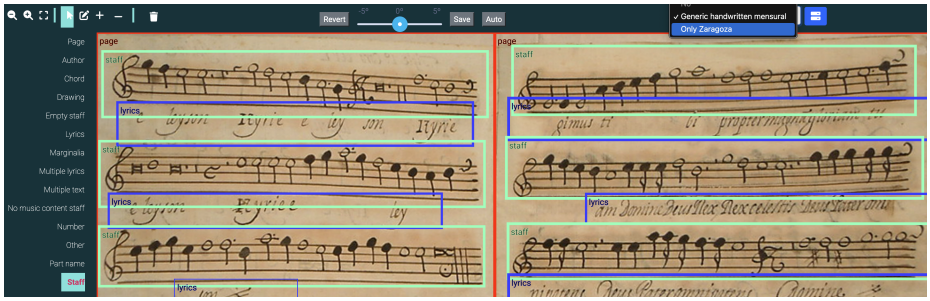
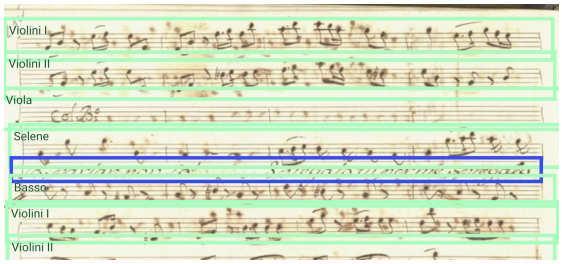
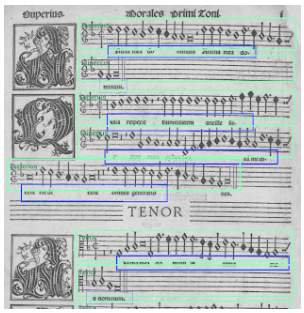


Fig. 2: Document analysis screen excerpt. In this example, only the staves and lyrics regions are segmented. The snapshot shows controls to rotate, manually or automatically, the image, and two possible classifiers to perform the operation automatically. The current catalog of region types shown at the left of the image can be easily modified.

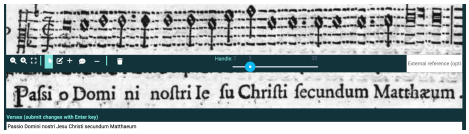


(a) Parts in orchestral score

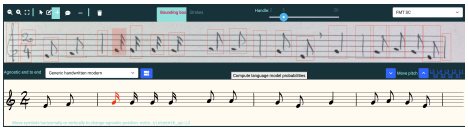


(b) Parts in choir book

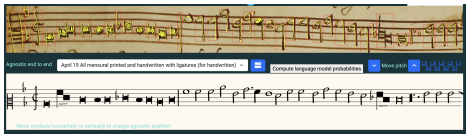
Fig. 3: Different parts and arrangements. All regions must be attributed to a part.



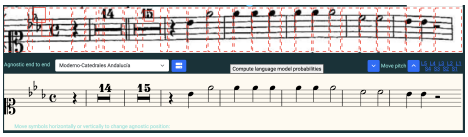
(a) Lyrics. See text transcribed at the bottom of the snapshot.



(c) Bounding boxes of agnostic symbol

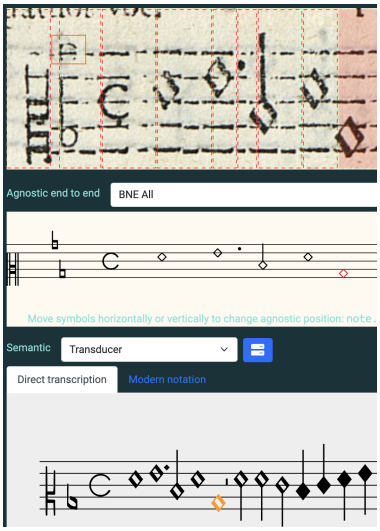


(b) Strokes

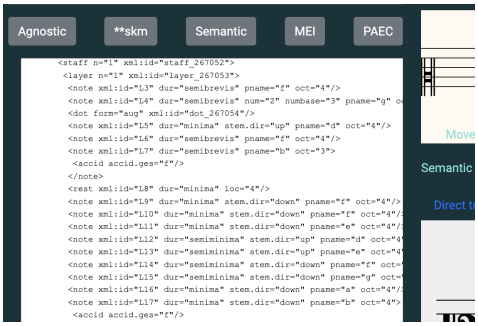


(d) Staff-level end-to-end

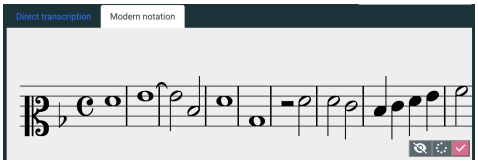
Fig. 4: Transcription of regions.



(a) Agnostic to semantic

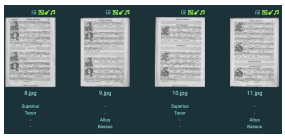


(b) MEI conversion of semantic content

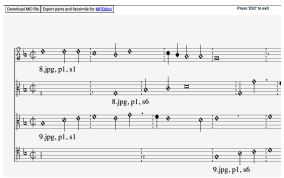


(c) Modern transcription

Fig. 5: Semantic contents recognized from the image.



(a) Image selection



(b) Score preview

(c) Parts based MEI including facsimile (see more details in [3]).

Fig. 6: Previsualizing and exporting

References

1. M. Alfaro-Contreras, D. Rizo, J.M. Iñesta, and J. Calvo-Zaragoza. OMR-assisted transcription: a case study with early prints. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, pages 35–41, Online, November 2021. ISMIR.
2. J. Calvo-Zaragoza and D. Rizo. Camera-PrIMuS: Neural end-to-end optical music recognition on realistic monophonic scores. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, 2018*, pages 248–255, 2018.
3. K. Desmond, L. Pugin, J. Regimbal, D. Rizo, C. Sapp, and M. E. Thomae. Encoding polyphony from medieval manuscripts notated in mensural notation. In *Music Encoding Conference Proceedings 2021*, page 197–219. Humanities Commons, May 2022.
4. D. Rizo, J. Calvo-Zaragoza, J.C. Martínez-Sevilla, A. Roselló, and E. Fuentes-Martínez. Design of a music recognition, encoding, and transcription online tool. In *16th International Symposium on Computer Music Multidisciplinary Research, Tokyo*, (accepted) 2023.