HISPAMUS: Handwritten Spanish Music Heritage Preservation by Automatic Transcription

José M. Iñesta, Pedro J. Ponce de León, David Rizo, José Oncina, Luisa Micó,
Juan Ramón Rico-Juan, Carlos Pérez-Sancho, Antonio Pertusa
Software and Computing Systems, University of Alicante, Spain
{inesta,pierre,drizo}@dlsi.ua.es

***Abstract*—The HISPAMUS proposal aims at enhancing the Hispanic music heritage from the 15th to the 19th centuries, by exploiting the digital resources of these collections. In addition, thousands of oral tradition melodies that were compiled by folklorists in the 1950s decade are digitized just as images, currently without the possibility of content-based search or study. It is necessary to develop services and tools for the benefit of archives, libraries, scholars, computer scientists and general public. HISPAMUS tries to provide smart access to archival manuscripts of music scores, allowing its reuse and exploitation. In order to reach this ambitious goal, our group can provide cutting-edge technology in the fields of Machine Learning, Pattern Recognition, and Optical Music Recognition.**

***Index Terms*—Heritage, Notation transcription, Encoding, Optical music recognition, Handwritten, Mensural notation.**

I. Introduction

Musical notation has evolved over the centuries into various writing systems. In particular, Spanish white mensural notation was the dominant code for writing music in Spain (and in Latin America) between the 16th and 18th centuries, producing large collections of handwritten documents yet to be made accessible to the public.

All this cultural heritage has been, in part, hidden from the public eye because its custody is restricted to certain actors (the Church, private collections, etc.) and its interpretation by contemporary musicians is far from their correct understanding. Accordingly, musicologists work for the study and appreciation of such compositions, from the cultural and economic point of view, is limited by the tools within their reach. In all these cases, just the scanned or photographed images of the scores are available, and not their transcription into symbolic music formats. How do we make them digitally available to the public for its use and valorization?

The HISPAMUS project aims at generating the software tools needed to convert manuscript score images to a modern digital format, like MEI/MusicXML, ready to render modern notation scores, allowing the public to enjoy and search into the digital contents of these works, either as a musicologist or as a musician.

The main objectives of the project can be summarized as follows:

- Foster a significant technology progress beyond the current state of the art in digital music content production and management by means of optical music recognition and machine learning technologies.
- To make it possible to access both early and traditional music works in new ways, for study, analysis, performance, etc., within the framework of new technologies and web services.

The OMR techniques available today are not ready for being applied to these kind of handwritten documents. Thus, one of the main goals of this project is to process this kind of documents, covering the whole workflow from the digitized images to the production of the digital symbolic format.

One important feature will be the possibility to render also a translated version to a modern notation score, readable by a contemporary musician, and browse the digital contents of these works, either as a musicologist or as a musician.

From the technical point of view, the processing stages (see Fig. 1) will be designed using machine learning (ML) techniques. Most available systems are based on heuristic rule methods, but one of the problems of OMR is the existence of too many repertoire-dependent context rules [1]. Therefore, it is hard to extend them to recognize early music notations. By applying ML technologies we can build models based on *training pairs*: sets of data examples presented together with a ground-truth label. This opens up the possibility of adapting the system, in principle, to any music notation, if labeled data is available for the system to learn.

II. Methodology

The software to be developed through the HISPAMUS project is required to recognize the contents of printed or manuscript scores from different notations, the encoding into all current standards, and the assisted transcription into a edited version, suitable for preparing a critical edition. In addition, as this system is conceived as a research tool, it has to be equipped with features to measure the efficiency and effectiveness of the different tasks and approaches.

The scheme in Fig. 1 describes the main workflow in the project. The system is fed with digitized images of the manuscripts. The user has to organize the images according to the books. The OMR processes include a layout analysis for separating music from texts, and identifying the different regions in the pages.

In all cases, our OMR operations generate two kinds of sequence symbol encodings. One is named *agnostic*, which recognize symbols by their shape and position, without analyzing the musical meaning of the symbol in the score. The

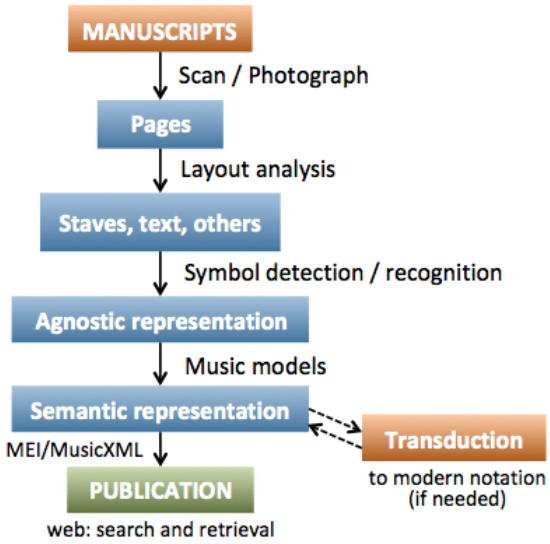

Fig. 1. Main workflow in HISPAMUS.

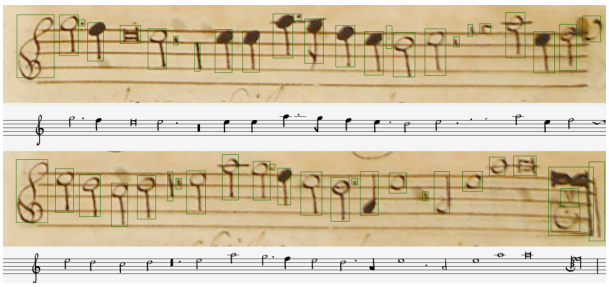

Fig. 2. Snapshot of the system in its current stage of development. Bounding boxes for the symbols and their recognition are displayed.

other encoding is the *semantic* one, where meaningful musical information is encoded into a standard music format.

The system is designed under the assumption that no OMR approach is able to provide a perfect performance. Thus, the user intervention is required. This assessment and correction task can be accomplished at graphical symbol level (agnostic) or by editing the music content obtained directly (semantic).

From our experience, we believe that it is difficult to find a single approach for OMR that can be effective in all situations. Therefore, we implement three types of automatic recognition, analyzing the image at different levels of granularity:

a) User-driven symbol recognition: the first approach is based on the work of Calvo-Zaragoza et al. [2], where the user locates the symbols manually. Typically, the computer mouse can be used to indicate the position of the symbols, but we also allow in MuRET, the software application developed for the project, the use of a digital pen, resulting in a more ergonomic interaction.

b) Holistic staff recognition: in this approach, each staff is processed without previous symbol segmentation. This can be achieved by using continuous models like hidden Markov models [3] or recurrent neural networks [4].

c) Full page recognition: the last approach we consider is a neural model that is applied directly to a full page [5]. In this case, given the image, the model returns a list of predictions, each of which indicates a bounding box and the category of the object therein.

These approaches work at the graphical level (agnostic representation), ignoring the semantics of the music notation itself that can be recovered in a post-processing stage. For example, in the case of notes, the vertical position within the staff, along with the clef and eventual alterations, allows inferring their pitches.

Our intention is also to take advantage of this infrastructure to carry out user-centered studies, which will provide information about which of these approaches is the most effective in each situation. The ultimate goal is to obtain the structured encoding of the music sources with the least possible effort from the user. The effort of the user is closely related to the accuracy of the recognition models, but also with the type of corrections and interactions required.

Music editing: As mentioned above, one of the fundamental objectives of HISPAMUS is to obtain a transcription of the manuscript without any ambiguities, so that both the musicologist and the publisher can generate a critical edition from it. This stage includes two tasks: the optional correction of the recognized graphical symbols from the OMR, and the correct assignment of musical functions to those graphical symbols.

Furthermore, the current goal of MuRET is not to generate final edited preprints, but to produce contents to be sent to on line services or publishers. As mentioned above, the system should export in any interchange format will be edited by these publishers to fulfil their publishing workflow. Thus, a printed image must be generated with the aspect of how approximately the transcription should look like. MuRET exports this kind of output using PDF.

Acknowledgements

HISPAMUS is the project TIN2017-86576-R, supported by the Spanish Ministry, partially funded by the EU.

References

[1] D. Byrd and J. Simonsen, "Towards a standard testbed for optical music recognition: Definitions, metrics, and page images," *Journal of New Music Research*, vol. 44, no. 3, pp. 169–195, 2015.

[2] J. Calvo-Zaragoza, D. Rizo, and J. M. Iñesta, "Two (note) heads are better than one: Pen-based multimodal interaction with music scores," in *17th International Society for Music Information Retrieval Conference*, 2016, pp. 509–514.

[3] J. Calvo-Zaragoza, A. H. Toselli, and E. Vidal, "Handwritten music recognition for mensural notation: Formulation, data and baseline results," in *14th IAPR International Conference on Document Analysis and Recognition*, 2017, pp. 1081–1086.

[4] J. Calvo-Zaragoza and D. Rizo, "End-to-end neural optical music recognition of monophonic scores," *Applied Sciences*, vol. 8, no. 4, pp. 606–629, 2018.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.