LESSONS LEARNED FROM A PROJECT TO ENCODE MENSURAL MUSIC ON A LARGE SCALE WITH OPTICAL MUSIC RECOGNITION

David Rizo[1,2] **Jorge Calvo-Zaragoza**[1] **Patricia García-Iasci**[1] **Teresa Delgado-Sánchez**[3]

[1] Pattern Recognition and Artificial Intelligence Group, University of Alicante, Spain
[2] Instituto Superior de Enseñanzas Artísticas de la Comunidad Valenciana, Spain
[3] Biblioteca Nacional de España, Spain

{drizo, jorge.calvo, pgarcia.iasci}@ua.es mariateresa.delgado@bne.es

ABSTRACT

This paper discusses the transcription of a collection of musical works using Optical Music Recognition (OMR) technologies during the implementation of the Spanish PolifonIA project. The project employs a research-oriented OMR application that leverages modern Artificial Intelligence (AI) technology to encode musical works from images into structured formats. The paper outlines the transcription workflow in several phases: selection, preparation, action, and resolution, emphasizing the efficiency of using AI to reduce manual transcription efforts. The tool facilitated various tasks such as document analysis, management of parts, and automatic content recognition, although manual corrections were still indispensable for ensuring accuracy, especially for complex musical notations and layouts. Our study also highlights the iterative process of model training and corrections that gradually improved transcription speed and accuracy. Furthermore, the paper delves into challenges like managing non-musical elements and the limitations of current OMR technologies with early musical notations. Our findings suggest that while automated tools significantly accelerate the transcription process, they require continuous refinement and human oversight to handle diverse and complex musical documents effectively.

1. INTRODUCTION

In recent years, many institutions have digitized their collections to preserve them and make them available online for broader public access. Digital images, however, merely contain a grid of pixels and lack inherent musical meaning; thus, they do not lend themselves to the myriad possibilities offered by music information retrieval and digital musicology approaches, ranging from plain-text content searches to more sophisticated analytical purposes. To leverage these technologies, the music depicted in the images must be encoded in a structured format, such as MEI [1] or MusicXML [2], among others.

Over the past few years, Optical Music Recognition (OMR) technologies have been employed to facilitate the encoding of music scores into structured digital formats [3]. Alfaro-Contreras et al. [4] demonstrated that the most effective method for obtaining digitally encoded scores is through the use of OMR technology. Their research indicates that the accuracy of OMR in recognizing musical notations varies depending on the type of document, the quality of the source material, and the complexity of the notation.

Despite its advances, OMR technology seldom produces flawless results, and the extent of necessary post-editing is determined by the intended use of the digitized content. For instance, some initiatives, such as F-Tempo [1], utilize OMR outputs—even when they contain errors—for conducting search operations. However, when a polished transcription is required, manual corrections become indispensable. This was the case considered in the digitization of a vast array of files for the KernScores database. [2]

The limitations of OMR technology are not solely determined by its recognition accuracy. To date, no OMR system is capable of comprehensively processing the entire spectrum of symbols found in all kinds of musical notations. The complexity of analyzing orchestral scores, with their varied layouts and the inclusion of *ossias*, or managing compositions where different parts are noted on separate sheets, further complicates the scenario. Consequently, in many practical applications, the encoding is ultimately carried out by human transcribers using computerized notation software like MuseScore [3] or Sibelius. [4] In specific projects such as Didone [5], about 4 000 18th-century Italian Opera arias are manually transcribed in Finale [5] before being converted into MusicXML. This methodology was similarly employed to achieve the encoding of modern versions of Renaissance compositions from the "Josquin Research Project". [6]

Furthermore, several OMR solutions exist for tran-

[1] f-tempo.org (accessed April 8th, 2024).
[2] kern.ccarh.org/ (accessed April 8th, 2024).
[3] musescore.org (accessed April 8th, 2024).
[4] www.avid.com/sibelius (accessed April 8th, 2024).
[5] www.finalemusic.com (accessed April 8th, 2024).
[6] josquin.stanford.edu (accessed April 8th, 2024).

scribing Common Western Modern Notation (CWMN), with Audiveris[7] standing out as the sole open-source option alongside several proprietary alternatives, including SmartScore,[8] PhotoScore,[9] and PlayScore 2.[10] The performance of these varies significantly based on the sheet music's complexity and clarity. An evaluation of their efficiency in recognizing content from music theory books is detailed in the work of Moss et al. [6], highlighting the challenges they face in complex situations.

For early notations, the choices are much more limited. The SIMSSA project [7] considered two software tools—Gamut and Aruspix [8]—for automatic information extraction from images, although these tools are no longer actively supported. Additionally, the project developed an OMR meta-workflow named Rodan, enabling users to assemble custom processing systems from a library of image processing and machine learning modules [9]. While Rodan is not tailored to any particular musical notation, its components are predominantly focused on plainchant. Recently, a web-based OMR application named MuRET has been introduced as a research-oriented tool designed to facilitate the scientific study of the complete OMR workflow across various scenarios and notations [10]. This includes analyzing the real impact of improvements in automatic recognition models and their integration for practical purposes in the work of transcribers.

In this paper, we outline the entire process undertaken in the context of the Spanish PolifonIA project, for which MuRET has been utilized and refined to transcribe the entire collection of white Mensural notation held by the National Library of Spain (BNE) from scratch. We will detail all stages of the process, aiming to provide useful takeaways for other similar projects and transcription tools based on OMR. This includes discussing both manual and automated stages, the steps that may benefit from advancements in OMR techniques, those that still require human intervention, and which processes need to be streamlined due to their significant impact on workflow performance.

To illustrate the aforementioned aspects, figures will detail how, by the end of the project, more than 60 books containing around 12,000 images—some consisting of several pages—were encoded in just 18 person-months. Additionally, the figures will showcase how an iterative approach of transcription, correction, and AI-model training gradually accelerated the whole process.

The remainder of the paper is organized as follows. First, the data that has been transcribed is briefly introduced in Section 2. The following Section 3 describes the whole workflow used for obtaining a final digital score from a set of images in the source collection. This workflow will be analyzed from a quantitative point of view in Section 4, and then discussed from a qualitative perspective in Section 5. Finally, Section 6 concludes the work and discusses possible ideas for future research.

2. DATA

Although the workflow and evaluation described in subsequent sections are somewhat generic, this section provides details of the digitized collection to contextualize its significance.

The collection considered for the project totals 63 works, almost entirely in print editions dating from 1533 to 1811, and mostly written in white Mensural notation. The genres of these works are varied, comprising mainly vocal polyphonic pieces, although there is a presence of instrumental, dramatic, and even treatises. Their functions are predominantly religious, with some presence of profane songs. Their formal structure is linked to this, highlighting the complexity of formats in religious works ranging from Passion Cycle and Missae to the simpler forms of chansons or motets, among others. In polyphonic works, the parts are usually written in separate books.

Regarding printers, the collection features works from the Italian School such as: Scoto, Gardano or Vicenti from the Venetian; Dorico and Robbleti from the Roman; and Carlino and Beltrano from the Neapolitan. Le Roy and Ballard are prominent in the Paris School, along with the Flemish School's Phalesius, Bellere, and Susato. Spanish publishers include Ibarra, Doblado, and Martinez Dávila in Madrid editions.

3. TRANSCRIPTION WORKFLOW

The transcription workflow can be broken down into several sequential phases.

The first stage involves the selection and compilation of works to be transcribed, either in PDF format or as a set of individual images. These are properly ordered through their file names following a lexicographic criterion for avoiding the need for time-consuming manual reordering within the tool.

For the sake of time and organizational management, the works are classified into different collections according to similarities in notation and/or publisher. This allows works sharing similar visual aspects to utilize the same machine learning models without adjustments between them. Considered features include notation type (plain chant, mensural, transitional scores, modern notation), engraving method (handwritten or typeset—where the copyist or printer is noted for sharing typography and layout styles), contents (treatises, instrumental and vocal music including lyrics), and the presence of elements such as basso continuo.

The next step involves uploading the works to MuRET. This tool employs OMR models that drastically reduce the image sizes to heights of 256 pixels. Although it utilizes IIIF servers that manage image resizing, for transcription purposes, it is not necessary to import high-quality images, but rather those with sufficient resolution to be readable on the user's device. To avoid wasting server processing time and space, a prior down-sampling of images is advisable.

In the next stage, we refine the content imported into the tool. Most of the imported image sets contain covers, empty pages, and indexes that, while not containing strictly

[7] github.com/Audiveris (accessed April 8th, 2024).

[8] www.musitek.com (accessed April 8th, 2024).

[9] www.neuratron.com/photoscore.htm (accessed April 8th, 2024).

[10] www.playscore.co (accessed April 8th, 2024).

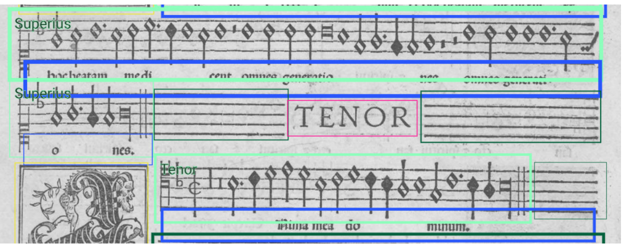


Figure 1. Example of document analysis and part assignment. Box colors represent different region types.

music information, are useful for extracting metadata and should be discarded but not removed, as they can guide the transcription process. Although some processes exist to automate this detection of pages with actual music content [11], MuRET does not include this desirable feature. Once the images are loaded and filtered, longer works to be transcribed must be divided into sections, such as the different parts of a mass (Kyrie, Agnus, etc.) or the movements of a concerto.

The final block aims to perform the actual transcription of the works. It consists of four main operations that will be detailed below: analyzing the document layout and dividing it into regions of interest, associating each staff with a part or instrument, recognizing the music contained in each staff and its encoding, and, finally, using all that information, scoring up all the parts to form a final digital score.

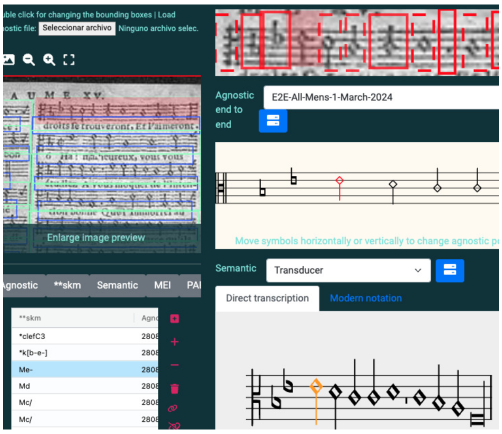
The document analysis and staff-level recognition of music symbols are performed using deep learning technologies [10]. Generally, we follow the same scheme for handling new works to be transcribed. First, models trained with previous collections are applied, mistakes are corrected, and then iteratively, new models are built, either specific for the collection if it is very different from previous documents, or following the proposal in [12], general for all transcribed collections. When faced with a new manuscript, the strategy is to first evaluate with the latest general model. If this does not perform well—which is evaluated subjectively by the user—we proceed to label, with or without the help of the OMR output, about twenty pages of the new work, then build specific OMR models and, in addition, enrich the general model for future works.

3.1 Document analysis

Upon arranging the images, the initial action in transcribing a manuscript, termed *document analysis*, involves dividing each image into distinct elements. This process detects various region types within the images, such as staves, lyrics, part names, among others, as illustrated in Figure 1. Typically, an image encompasses only a single page. However, scans of entire books are also common, resulting in images that depict multiple pages simultaneously, akin to the example shown in the figure.

3.2 Part management

The majority of materials requiring processing are polyphonic, composed of multiple voices or instruments. These materials come in various formats, such as compositions with parts spread over several pages, or choir-books that display two voices on a single page (see Figure 1), among others. Occasionally, the document intended for transcription is dedicated to music theory, as seen in music treatises [6], predominantly featuring textual content with occasional musical illustrations. Currently, the assignation of parts is performed manually.



Agnostic end to end: E2E-All-Mens-1-March-2024

Semantic: Transducer

**skm	Agn
*clefC3	2808
*k[b-e-]	2808
Me-	2808
Md	2808
Mc/	2808
Mc/	2808

Figure 2. Agnostic representation and its semantic conversion in MuRET.

3.3 Region-wise content recognition

After distinguishing and assigning the various staves to their respective parts, it becomes essential to extract the musical elements located within each staff.

The approach applied divides the recognition of the musical content in a staff in two steps (see Figure 2). First, it extracts what is referred to as *agnostic representation* [13], i.e., tokens that have not yet been assigned a specific musical meaning, as well as their absolute vertical positions on the staff, regardless of the clef used. Then, these are automatically transduced into a meaningful `**mens` encoding [14], that can be manually post-edited.

After the automatic recognition, the eventual mistakes must be corrected. We found four different kinds of errors, with different impacts on the time required to be corrected. The easiest mistake is that of the vertical position of a recognized symbol (1), that is amended just with a mouse or keyboard action. A symbol whose type is wrongly detected (2) requires a slight higher effort, as it takes some seconds to find the expected symbol among all the possibilities. The removal of a symbol (3) is a very quick operation, while adding an undetected symbol (4) requires drawing a box over the manuscript image.

Note that for those difficult manuscripts for which all automatic models generate too many errors, as that shown in Figure 2, it might be preferable to manually add all agnostic symbols as described above.

3.4 Scoring up and exporting

As above mentioned, most of the works transcribed in the project are organized into separate parts or choral books, where different voices or instruments are scattered across

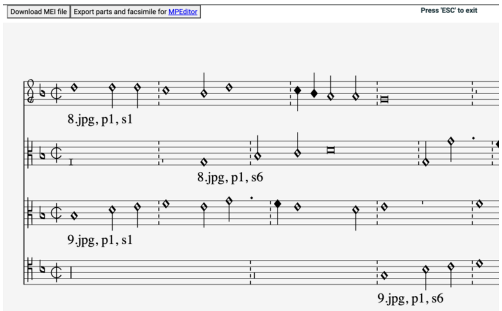


Figure 3. Example of alignment in MuRET. Some textual data is included below the staves to indicate reference points of the corresponding source image. Dashed bar lines are used to help detecting alignment errors.

different pages. Having already identified to which instrument each staff belongs (Sect. 3.2), this operation is simply accomplished by concatenating all the staves of the same part.

However, in Mensural notation, a preliminary step is required to correctly align the voices. In this notation, some notes may have different durations depending on the context, despite their appearance. The contextual resolution of durations, resulting in changes called *perfection* and *alteration*, can be carried out in MuRET either automatically, by applying the rules established in [15], or manually, by editing the `**mens` code.

In any case, mistakes such as missing symbols, incorrect duration elements, or invalid perfection assignments can only be detected by visually inspecting the aligned score (see Figure 3).

The final step of the process is exporting the transcription into an interchange or storage format. In the particular case of MuRET, the MEI standard is considered, offering two possible export formats: a parts-based MEI format that includes graphical information in the facsimile element or the arranged score MEI file.

4. QUANTITATIVE EVALUATION

MuRET records all operations performed by the user, saving the timestamp of each action and the element on which it is performed.

In this evaluation we address three questions. The suitability of using a transcription tool such as MuRET in a real-world scenario, the relative importance in OMR operations compared to the other tasks, and the ability of machine learning approaches to improve their accuracy as training datasets are iteratively expanded.

The first question is evaluated by comparing the performance of the tool with the theoretical hypothesis proposed in [4]. The first two rows of Table 1 show the times reported in [4] for processing 126 typeset pages of a *Magnificat*, either totally manually, or using an OMR. [11] Note that in that work, only the agnostic representation is obtained, and the time required for performing all the other tasks, such as the document analysis, or document preparation is discarded. Automatic processing times are in all cases less than 1 second after loading the models into memory.

The next two rows show the process performed in the current project with the same *Magnificat*. First, the time to perform OMR processes (document analysis and recognition of agnostic symbols in each staff), then the entire transcription process, including all phases of the workflow. The final review of the scoring up has been excluded from these figures because in many cases the time is spent on musicological discussions of the manuscript rather than mechanical issues.

Finally, we have added to the table the worst case of those encountered in the project because it is a very difficult one due to the very low resolution of the images, which would have been extremely tedious to transcribe without the help of OMR (see Fig. 2), and the best case found for which the existing general OMR models have been able to correctly detect almost all symbols, and no part management was required.

The times reported demonstrate the suitability of using an OMR approach, but also the major impact on the whole process of the other, non-directly OMR processes, which cannot be overlooked.

Table 1. Summary of annotation times per page.

Scenario	**Avg. Time/Page**
Magnificat work	
Manual agnostic annotation [4]	$49'19'' \pm 11'27''$
OMR of agnostic representation [4]	$15'23'' \pm 2'44''$
OMR: doc. analysis and agnostic	$22'07'' \pm 20'42''$
Whole transcription process	$29'09'' \pm 23'37''$
Whole project collection	
Worst case (whole transcription)	$52'31'' \pm 23'31''$
Best case (whole transcription)	$4'30'' \pm 1'51''$

Regarding the second question, compare the relative importance of classic OMR operations with other operations such as document preparation or parts management for an entire collection, we show in Table 2 the times of all actions performed in MuRET grouped by all the workflow phases described in Sect. 3. The figures show that as it could be expected, the recognition of the musical symbols in each staff is the most time consuming task, followed by the semantic conversion and the document analysis, and what a priori could seem a slow operation, the manual assignment of parts to the staves, is a very small portion of the total, even lower than the preparation of images and organization into sections prior to the transcription itself.

Finally, to evaluate how incremental training of OMR models leads to better OMR behavior, we report in Fig.4 the number of operations performed on each image throughout the life of the project. Using the date axis is interesting because as the project has progressed, we have had more accurate OMR models because we have been trained on more data.

In the figure, we have used the number of operations instead of times because the time depends on the laptop on

[11] This value is computed from the values of Figure 2 in [4]

Table 2. Summary of time per phases.

Phase	Processing times
Document preparation	830′08″
Document analysis	4.913′36”
Part management	429′37″
Agnostic representation	1.9536′59″
Semantic encoding	10.923′55″

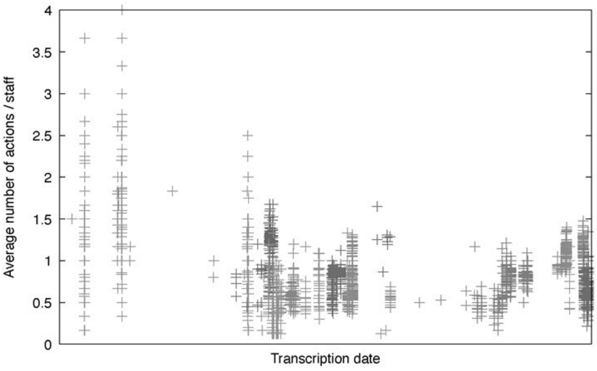


Figure 4. Evolution of transcription operations over time for all images in the project. Each point represents the number of operations required to transcribe a page.

which the operation was performed, since all classification models in MuRET are executed in the browser. Also, depending on the work, the number of staves of each image varies. To solve this, the graph shows the relationship between the number of operations and the number of staves of each image.

The figure shows how the average number of operations over time tends to be lower as the date progresses. We observe that an initial specialization of the OMR engine does help, and after that the user effort is stabilized.

5. QUALITATIVE EVALUATION OF THE WORKFLOW: OPPORTUNITIES

In this section we analyze the suitability of the involving stages of the transcription workflow to draw good practices for the development of OMR tools. We will discuss which operations we believe could be fully or partially automated to speed up the process and avoid tedious and repetitive work as much as possible.

The preparation of the works to be transcribed and its correct organization have been decisive for the success of the project. Since there is not yet a universal OMR model capable of dealing with any possible entry, the grouping of the pieces according to time period and typographic or calligraphic style, and the arrangement of the transcription following these groups, has been a key factor. In cases where, for some reason, we interleaved a piece out of that order, the performance decreased. While this clustering process was performed manually by computer scientists and musicologists, automating it could help to know in advance which existing OMR model could be applied to a new manuscript. Also, it is interesting to automatically detect whether no model is able to process the manuscript and manual labeling of a number of pages is required to build a specific one.

A factor that we have already mentioned is that of the image resolution and, implicitly, the weight of the files. Although IIIF servers are able to deal with the resizing of images, we have experienced a noticeable speed-up when the uploaded images are of smaller sizes.

Initially in the project, each work was processed following the different steps sequentially image by image. After processing some, instead, another approach was proved to be more convenient: perform all the operations of each phase for all the images of the work in batches. This allowed us to follow up on the work and detect possible errors made or not detected. It is important to note that in cases where we did all the tasks on each page and only reviewed them once, we made more errors. This approach was enhanced by a new feature added to MuRET in the middle of the project: the possibility of automatically tagging all work for later correction, which drastically improved transcription times by saving us OMR processing times for each page and staff (done "offline").

A key aspect with a huge impact on the throughput of the workflow has been the (sometimes questionable) decisions of the MuRET developers in terms of UI/UX. The simplicity on the correction of the agnostic staff level automatic transcription and its automatic conversion into a meaningful semantic encoding in `**mens` format helped to minimize the impact of inaccurate OMR model predictions. A paradigmatic example has been the change in MuRET for the way of processing ligatures. The first OMR models in MuRET were not able to detect different mensural ligatures, but all different ligatures as a common symbol. The conversion of all ligatures to their final `**mens` encoding took longer than automatically encoding and correcting an entire page. During the transcription project, this tool was able to detect all the individual components of the ligature (plicas and note heads). Being quite accurate, when failing, the correction of the individual components took the same time as deleting the whole detection and adding them again. In a later version, this approach was changed by another one where the ligature was converted in a lower number of elements (different notes with or without plicas) with a bit worse OMR performance. However, for the purpose of final correction times, this change was appropriate because from then on, the correction time for errors was equivalent to the correction time for any other element.

Following this line, an aspect that could improve the efficiency of use of the system would be an easiest correction procedure of wrongly detected agnostic symbols. Currently, the user has to locate the symbol into a grouped list of possibilities. Even though this a specific criticism to MuRET, any simple mechanism in any transcription tool for locating the desired element to use, as some keyboard filtering approach, would significantly reduce the correction times.

The separation between agnostic representation and its final semantic encoding has proven to be an efficient way of processing early music. The ease of checking that the graphic symbols are the same as those in the manuscript, regardless of musicological considerations, has greatly accelerated the process, allowing each member of the team to focus on one of the phases, leaving the expert musicologist to deal only with the final transcription. It's worth to mention, that in this process, a specialized language model to detect syntactic mistakes would have improved the efficiency of the process, as we have devoted most of the time to visually inspect the output of the OMR classifiers, even more than correcting wrong symbols.

Although the conversion of the agnostic representation of each staff into a final encoding is performed automatically, we've found cases where it has been necessary to make some adjustments, such as the encoding of implied accidentals. MuRET does not use any *WYSIWYG* approach but asks the user correct directly writing `**mens` format. Having a steep learning curve, the code has proven to be efficient for performing this kind of operations.

In that regard, another important feature, without which the correction operations would have been more tedious, has been the proper synchronization of the views of the different representations of the selected transcribed musical symbol: when selecting the agnostic symbol, it was automatically highlighted in the original manuscript preview, and in the final encoding. The absence of this feature in the final MuRET scoring-up process has made the final review and correction time consuming and error-prone.

For dealing with many different works with a large number of images each, it is very important to keep track of the status of the work. MuRET asks the user to record the status of each phase (document analysis, part linking, music transcription) for each image. Although a priori this seems reasonable, we usually forgot to perform this operation, and the simple task of going back individually to mark each image and step as completed has been a time consuming operation. For any transcription tool, it is extremely important to include a project management tool to easily annotate and visualize, either individually for each image or in batch, the progress status of the transcription, including the addition of user comments.

An interesting result of our transcription experience is that some operations do not require any algorithm, but are simply performed with a correct graphical user interface. This has been the case for document analysis labeling of new manuscripts for which no model was good enough to correctly identify the regions of interest. At the beginning of the project, when this situation arose, we had to manually label a number of pages of the manuscript to build a new model that was subsequently improved with new samples. For collections in which the layout of the regions of interest and the parts to which they belong is repeated over several pages, this process does not need any complex machine learning process, but a process of reusing the existing tagging is enough. During the project, MuRET included a tool to copy the document analysis and link parts to other images. This simple tool turned this tedious and repetitive operation of tagging the pages first into only a minor issue.

A notable case occurs in the event that the tool, or a component of a tool, does not support a required feature. For instance, bar-lines crossing a note in late Mensural notations or the rendering of *signum congruentiae* is not supported in Mensural notation by the engraving tool used in MuRET, Verovio [16]. In those cases, our principle has been to store a specific element, such a text, and print them to be visualized, and once they are supported by the tools, replace them.

Finally, when focusing on the transcription of musical content, most tools discard many non-musical elements such as titles, part, instrument or voice names, capital letters miniatures. All this information, if automatically detected, could help to the users to have a better overview of large works to organize the transcription process.

6. CONCLUSIONS AND FUTURE WORK

Most of the OMR community's efforts are focused on achieving high accuracy rates in automated music reading. We have shown in Section 4 that the use of an OMR tool has proven to be an adequate means to transcribe a whole collection of works saving an enormous amount of time and effort for the user. While this approach is valid, it is important not to overlook aspects that are not intrinsically OMR and that can impact even more than the performance of the transcription tool on the effort required to transcribe collections.

In this work, we have shared our experience in transcribing a complete collection of works written in Mensural notation, describing all the steps taken and discussing issues we believe are important to achieve a streamline process, both from the perspective of the OMR tool used and in the preparation of the collections to be transcribed.

This paper has not addressed aspects that would be interesting to explore in the future. Some are related to the functioning of the computer system itself, such as the impact of classification times of automatic systems on the overall process and program response delays, as well as the measurement of the impact of execution errors or a comprehensive study from the perspective of human-computer interaction (HCI) in operations such as editing the staff transcription made or the final scoring up.

Other factors to consider are purely musical, such as the use of musical language models, both melodic and harmonic, for error detection, the impact of using one musical encoding over another, assistance in aligning lyrics with music, the treatment of abbreviations in the lyrics, or the detection of specific properties of the notation type such as the semitonia subintelecta in Mensural notation, the processing of multiple voices in piano-form music, the detection of hidden graphical elements such as the digit '3' in triplets in common western music notation, or finally the specific cases described by Byrd and Simonsen [17].

Regarding the OMR system, it is interesting to compare different strategies at work within a complete transcription system, not just in isolation. For instance, replace the MuRET stages (document analysis, agnostic representation, semantic encoding), for those based on graphical primitives and later semantic encoding reconstruction [18], or the direct obtaining of the final encoding from a complete page [19].

7. ACKNOWLEDGMENTS

This paper is part of the I+D+i TED2021-130776A-I00 (PolifonIA) project, funded by MCIN/AEI /10.13039/501100011033 and European Union NextGenerationEU/PRTR. The authors would like to thank Antonio Madueño, Adrián Roselló, Juan Carlos Martínez-Sevilla, and Antonio Ríos-Vila for their essential contribution to the project.

8. REFERENCES

[1] P. Roland, "The music encoding initiative (MEI)," in *Proceedings of the First International Conference on Musical Applications Using XML*, jan 2002, pp. 55–59.

[2] M. Good and G. Actor, "Using MusicXML for File Interchange," *Web Delivering of Music, International Conference on*, vol. 0, p. 153, 2003.

[3] J. Calvo-Zaragoza, J. Hajič, and A. Pacha, "Understanding optical music recognition," *ACM Computing Surveys (CSUR)*, vol. 53, no. 4, pp. 1–35, 2020.

[4] M. Alfaro-Contreras, D. Rizo, J. 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*. Online: ISMIR, Nov. 2021, pp. 35–41.

[5] A. Torrente and A. Llorens, "The Musicology Lab: Teamwork and the Musicological Toolbox," in *Music Encoding Conference Proceedings 2021*. Humanities Commons, 2022, pp. 9–20.

[6] F. Moss, N. Nápoles-López, M. Köster, and D. Rizo, "Challenging sources: a new dataset for omr of diverse 19th-century music theory examples," in *Proceedings of the 4th International Workshop on Reading Music Systems (WoRMS 2022)*, November 2022.

[7] I. Fujinaga, A. Hankinson, and J. Cumming, "Introduction to SIMSSA (single interface for music score searching and analysis)," in *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*, ser. DLfM '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1–3.

[8] L. Pugin, J. Hockman, J. Burgoyne, and I. Fujinaga, "Gamera Versus Aruspix: Two optical music recognition approaches," in *9th International Conference on Music Information Retrieval, Drexel University, Philadelphia, USA, September, 2008*, 2008, pp. 419–424.

[9] I. Fujinaga and G. Vigliensoni, "The art of teaching computers: The SIMSSA optical music recognition workflow system," in *27th European Signal Processing Conference, EUSIPCO 2019, A Coruña, Spain, September 2-6, 2019*. IEEE, 2019, pp. 1–5.

[10] D. Rizo, J. Calvo-Zaragoza, J. 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*, November 2023.

[11] A. Pacha and H. Eidenberger, "Towards self-learning optical music recognition," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017, pp. 795–800.

[12] J. C. Martinez-Sevilla, A. Rosello, D. Rizo, and J. Calvo-Zaragoza, "On the performance of optical music recognition in the absence of specific training data," in *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023, Milan, Italy, November 5-9, 2023*, A. Sarti, F. Antonacci, M. Sandler, P. Bestagini, S. Dixon, B. Liang, G. Richard, and J. Pauwels, Eds., 2023, pp. 319–326.

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

[14] D. Rizo, N. Pascual-León, and C. Sapp, "White Mensural Manual Encoding: from Humdrum to MEI," *Cuadernos de Investigación Musical*, 2019.

[15] M. E. Thomae, J. E. Cumming, and I. Fujinaga, "The mensural scoring-up tool," in *Proceedings of the 6th International Conference on Digital Libraries for Musicology*, ser. DLfM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 9–19.

[16] L. Pugin, R. Zitellini, and P. Roland, "Verovio - A library for Engraving MEI Music Notation into SVG." in *International Society for Music Information Retrieval*, jan 2014.

[17] D. Byrd and J. G. Simonsen, "Towards a standard testbed for optical music recognition: Definitions, metrics, and page images," *Journal of New Music Research*, vol. 44, pp. 169–195, 1 2015.

[18] J. Hajič and P. Pecina, "The muscima++ dataset for handwritten optical music recognition," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 39–46.

[19] A. Ríos-Vila, J. M. Iñesta, and J. Calvo-Zaragoza, "End-to-end full-page optical music recognition for mensural notation." in *ISMIR*, 2022, pp. 226–232.