OMR-ASSISTED TRANSCRIPTION: A CASE STUDY WITH EARLY PRINTS

María Alfaro-Contreras, David Rizo, Jose M. Iñesta, Jorge Calvo-Zaragoza
Department of Software and Computing Systems, University of Alicante, Spain
{malfaro,drizo,inesta,jcalvo}@dlsi.ua.es

ABSTRACT

Most of the musical heritage is only available as physical documents, given that the engraving process was carried out by handwriting or typesetting until the end of the 20th century. Their mere availability as scanned images does not enable tasks such as indexing or editing unless they are transcribed into a structured digital format. Given the cost and time required for manual transcription, Optical Music Recognition (OMR) presents itself as a promising alternative. Quite often, OMR systems show acceptable but not perfect performance, which eventually leaves them out of the transcription process. On the assumption that OMR systems might always make some errors, it is essential that the user corrects the output. This paper contributes to a better understanding of how music transcription is improved by the assistance of OMR systems that include the end-user in the recognition process. For that, we have measured the transcription time of a printed early music work under two scenarios: a manual one and a state-of-the-art OMR-assisted one, with several alternatives each. Our results demonstrate that using OMR remarkably reduces users' effort, even when its performance is far optimal, compared to the fully manual option.

1. INTRODUCTION

Music is a language used and understood all over the world, hence being one of the cornerstones of cultural heritage. In order to represent this art visually, so that it can be transmitted and later interpreted as conceived by the composer, many notation systems have developed and evolved over time. Their engraving in the so-called music score was mostly done by handwriting or typesetting processes until the end of the 20th century, thus there exist millions of music documents only available as physical documents [1].

The ongoing massive digitization of those works by means of scanners is not sufficient for these sources to become truly accessible. For that, they must be transcribed in a structured digital format such as MusicXML [2] or MEI [3], among others, which enables computational tasks such as indexing, large-scale analysis, or editing [4]. This process is often done manually, which leads to an unfeasible challenge at a large scale. Thus, the development of systems capable of automatically performing the transcription process is of substantial importance.

Optical Music Recognition (OMR) is the field of research that studies how to computationally read music notation in scanned documents and store them in a digital structured format [5]. The success of OMR would enhance the value of all the existing musical heritage in digital libraries and facilitate the retrieval of data for musicological research. However, while OMR has been an active research area for decades [6, 7] and current state-of-the-art systems have shown promising results [5], it is not always considered as a real alternative to the manual transcription process. Moreover, there is a lack of technology transfer, i.e., instead of creating a synergistic environment, the digital humanities show a certain reticence towards automatic technologies, as if both were not committed to the same ultimate goal: to study, make accessible, and preserve the existing historical sources of music worldwide [8].

We believe that the mistrust lies in the unattainable goal to which OMR systems have been subjected: a perfectly accurate transcription. Given the vast range of different situations present in real-life recognition scenarios, a perfectly accurate OMR system is a utopia. Therefore, the automatic transcription challenge must be understood as a technology-assisted one, since human-machine interaction is necessary if there is no tolerance for errors in the transcription [9].

In this paper, we study the extent to which OMR systems provide real assistance during a transcription process. For that, we study and compare the transcription process under two scenarios: a manual-working methodology and an OMR-assisted one. We chose a corpus that allows us to compare both workflows with several available open-source tools. We aim at illustrating how OMR systems facilitate the transcription process by measuring the average procedure time per page, to estimate the workdays needed to transcribe a complete music work written in mensural notation.

The rest of the paper is organized as follows: Section 2 overviews some related proposals to this topic; Section 3 thoroughly describes the experimental setup; Section 4 reports the obtained results and their main outcomes; and finally, Section 5 concludes the present work.

2. BACKGROUND

The digital preservation of musical heritage necessarily involves the encoding in a suitable, symbolic, and computer-readable format of the musical content described in the original or scanned, as appropriate, manuscripts. In the majority of cases, this transcription process follows a manual workflow; that is, the corresponding encoding is manually written directly to the computer (either by typing or by mouse-driven actions).

Music notation contains a logical structure and more information than just a series of glyphs positioned over a staff, which makes the transcription process an inherently complex task regardless of the tool used for it. All of this entails a great deal of work that is not feasible at large-scale levels. The use of automatic technologies, in particular OMR systems, would greatly facilitate the task at issue. However, despite a large amount of existing literature on OMR research [5], hardly any system has been developed that goes beyond research tools.

Some of the most popular commercial OMR systems are PhotoScore,[1] SmartScore,[2] and PlayScore.[3] With high recognition accuracy for printed scores, they constitute a great alternative to users who want to scan and play musical content. However, they are restricted to only one type of music notation, namely Common Western Music Notation (CWMN), and are proprietary solutions, i.e., it is necessary to pay a license or a subscription to get full access.

On the side of non-commercial systems, we find the following systems:

- Aruspix, a cross-platform software for OMR of early music prints, mainly those printed during the 16th and 17th centuries [10]. It transforms the music content of each page into an editable digital music format, allowing the correction of recognition errors by the user, used as feedback to dynamically improve its performance. For a more direct correction process, Aruspix possesses superimposition and collation features.

- Audiveris, an OMR system devised to extract the musical content from printed or handwritten score sheets in order to edit them further in music edition applications [11]. It only supports CWMN.

- Rodan, a web-based customizable OMR system [12]. The user inputs an image and creates the most appropriate workflow to process it. Once the corresponding preferred adjustments are selected, the same processing can be applied to all similar images. This is a double-edged sword because it puts the manuscript at the center of the workflow at the expense of having minimum knowledge of the technologies that can constitute an OMR system, something that is not required for librarians or musicologists. This might pose significant risks to the effective and timely achievements of the project objectives.

- "MUsic Recognition, Encoding, and Transcription" (MuRET), a web-based application that divides the transcription process into different steps [13]. It is a technology-centered research tool, which allows the use of different transcription approaches ranging from manual to OMR-assisted ones, producing in those cases the transcribed contents in standard encodings. MuRET allows a simultaneous graphical comparison between the original and the encoded score, favoring a quick detection of errors.

Despite the various OMR systems developed, their inclusion in the transcription process of digital libraries is far from being widely common. They are relegated to a pure research application spectrum due to the mistrust caused by their imperfect behavior, therefore making human supervision necessary.

This paper attempts to provide insights into the use of OMR systems in the transcription process of a music work to see whether, despite not being perfectly accurate, the time spent on error correction compensates for the time saved, compared to a fully manual transcription paradigm.

3. METHODOLOGY

The main focus of this work is to study to what extent an OMR approach can be useful for digital libraries in the music score transcription process. It is important to emphasize that we do not intend to benchmark OMR tools. Therefore, we design a methodology aimed at estimating the effort, in terms of user time, saved by performing a transcription process assisted by OMR technology, instead of a fully manual one.

In the following sections, we thoroughly detail our methodology. First, we describe and justify the chosen music collection; then, we illustrate the different transcription pipelines considered, as well as the tools and encoding languages involved; and finally, we explain the metrics considered that allow comparison between the previously described transcription modalities.

3.1 Corpus

We must first select a suitable music work that allows meaningful comparisons. The test case considered in this work is the *Magnificat omnitonum cum quatuor vocibus* by Cristóbal de Morales,[4] hereafter referred to as *Magnificat* corpus. It is a collection of one hundred and twenty-six typeset pages corresponding to a Spanish choir book of the 16th century written in white mensural notation. Figure 1 shows a short example of this set.

We have chosen printed mensural notation for several reasons. First, several open-source tools with varied approaches are available, which allows us to compare different transcription paradigms. Second, the graphical com-

[1] https://www.neuratron.com/photoscore.htm
[2] https://www.musitek.com/smartscore-pro.html
[3] https://www.playscore.co/
[4] RISM A/6 M 3597. http://bdh.bne.es/bnesearch/detalle/bdh0000100234

Figure 1. Staff from the *Magnificat* corpus.

plexity of the symbols is quite regular, which lets us draw conclusions using a low number of transcribed pages. Given so, we expect that the obtained conclusions can be extrapolated to other (similar) corpora. Lastly, the choice of a printed typeset work rather than a handwritten one is due to the desire of avoiding possible transcription ambiguities that may be caused by the specific writing style of the author of the work, which could affect our study.

3.2 Transcription pipelines

The transcription of a music score is understood as the process of fully transcribing the document into a structured digital format with the ultimate goal of keeping the same musical information that could be retrieved from the physical score itself [5]. In this work, we study two main transcription workflows, each one also considering different alternatives:

(i) A manual transcription, where the musical content is directly typeset in a chosen standard format, MEI [3], PAEC [14, 15], or Humdrum [16].

(ii) An OMR-assisted transcription, where the system performs the score transcription directly from the corresponding image, and the user corrects possible errors afterward.

Our objective is to establish a difference, quantified as the procedure time, between both paradigms. It must be noted that the transcripts will be verified at all times since it is established that, with or without the presence of OMR technology, errors might occur.

In the following sections, we will elaborate on the two transcription paradigms considered.

3.2.1 Manual transcription modality

In this modality, the manual transcription process commonly used in digital libraries is done with the computer by directly typing the encoding format or choosing graphical music symbols from a toolbar. To evaluate this paradigm, we consider the following tools:

- Verovio Humdrum Viewer (VHV) [5], an online digital music editor and interactive notation rendering interface for Humdrum files [17]. Mensural music notation can be encoded in Humdrum using the **mens exclusive interpretation [18].

- Oxygen XML Editor, a paid [6] and multi-platform editor of widespread use in the Music Encoding Initiative (MEI) community. MEI is an eXtensible Markup Language (XML) based format that enables the transcription of a wide range of music notations. We use Verovio [19] [7] to render the XML files edited in Oxygen XML Editor.

- The Computerized Mensural Music Editing (CMME) is a "what you see is what you get" score notation tool that allows inputting mensural notation visually [20]. For this work, it could be considered as an equivalent to MuseScore [8], Finale [9], or Sibelius [10] but for mensural notation.

- The MuRET web application [11], using the graphical annotation modality for manual transcription. When considering a manual paradigm, MuRET allows two scenarios: the typesetting encoding of the music score in one of the standard formats or its manual annotation at the graphical symbol level. The former is the same process as the one performed with VHV or Oxygen XML Editor, whereas the latter is another approach to the transcription modality proposed by the CMME. The graphical annotation of MuRET consists of creating corresponding bounding boxes for each of the symbols in the score and classifying them with graphical labels, selected from a catalog. The user focuses on finding the shape and vertical position in the staff that match those of the to-be-transcribed symbol, without the need of knowing its musical meaning. Finally, the tool converts the sequence of graphical symbols into the final encoding that the user must check and correct.

3.2.2 OMR-assisted transcription modality

This modality considers an OMR system to automatically transcribe the musical content of a music score. Assuming that OMR never guarantees a perfect transcription, the objective of this paradigm is not improving accuracy but reducing the effort, measured as time, that users invest in aiding the machine to attain such perfect results. In this work, we evaluate this transcription methodology under two scenarios:

- A pre-built scenario, that is, we use an OMR system previously built for working with corpora of similar characteristics. For that, we consider Aruspix, as it has been created to recognized typeset music scores of the 16th and 17th centuries, hence being appropriate to transcribe the test case considered in this work—a typeset body of work of the 16th century.

- A from-scratch scenario, that is, we train a new OMR model for this work. To develop this approach, training pairs, consisting of problem images together with their corresponding transcript solutions, are required. This implies that we need to first manually transcribe some pages of the corpus in order

[5] https://verovio.humdrum.org

[6] For the experiment, we use a trial version of 30 days.

[7] https://www.verovio.org/index.xhtml

[8] https://musescore.org

[9] https://www.finalemusic.com

[10] https://avid.com/sibelius

[11] https://muret.dlsi.ua.es/muret

to train the model to be able to use it for the transcription of the remaining pages. Therefore, we will study how many pages are necessary to train a model that yields results that can be considered effective. In this scenario, we use the OMR-assisted workflow of MuRET, as it allows the retraining of its own OMR model. Thus, the training pages are transcribed with the aforementioned manual annotation from MuRET.

3.3 Evaluation procedure

To avoid biases in the results, associated both with the lack of knowledge of musical notation and the use of the tools considered, a person with a good level of computer literacy and sufficient knowledge of music to understand the test case has been chosen to carry out the proposed experimentation. The task performer is familiar with the tools and coding languages considered, and so the transcription scenarios are not challenging by themselves.

We compare the different transcription modalities by measuring the time to execute the task at hand, i.e., the time it takes to transcribe a complete score will be timed, including that needed for revision and correction of possible errors processes. This time will be referred to as *procedure time*. This will allow us to obtain the average transcription time per page for the different scenarios, which we will use to estimate the total time needed, measured in 8-hour workdays, to transcribe the entire work with each of the tools.

4. RESULTS

In this section, we present and discuss the results obtained in our experiments, in terms of the average procedure time per page, in the following order: first, those obtained for the manual paradigm; and second, those concerning the OMR-assisted one. Afterward, we compare both transcription paradigms when estimating the workdays needed to transcribe the full *Magnificat* corpus in each of them.

4.1 Manual paradigm

We first introduce the results obtained in a manual transcription process for each of the tools considered in this scenario, as described in Section 3.2.1. Table 1 shows the average procedure time per page and its standard deviation.

An inspection of the results in Table 1 reveals that the tool used in a manual transcription influences the procedure time. On the one hand, we observe that the procedure time of Oxygen XML Editor is higher than that of VHV. This is rather expected, since the encoding vocabulary size is much larger when considering an XML standard format like MEI, as in the case of Oxygen, instead of the compact Humdrum syntax, as in the case of VHV, which codifies music symbols with less than 5 characters. Furthermore, when considering graphical interfaces for the manual transcription, MuRET depicts a higher procedure time than CMME. In the latter, the graphical label corresponding to the musical symbol to be transcribed is directly dragged to

Table 1. Average time and its standard deviation for the transcription of one page of the *Magnificat* corpus for each of the tools considered in a manual transcription paradigm. Oxygen stands for Oxygen XML Editor.

Manual transcription paradigm

	VHV	**Oxygen**	**CMME**	**MuRET**
Average time per page	$27'\,06''$	$56'\,14''$	$33'\,37''$	$49'\,19''$
Standard deviation	$3'\,52''$	$5'\,49''$	$3'\,07''$	$11'\,27''$

the desired staff position, while in MuRET a corresponding bounding box must first be created and then graphically labeled in terms of shape and vertical position in the staff, respectively, which slows down the correct labeling of the music symbol.

We consider it necessary to comment on the importance of a good design of the tool from a user experience point of view [21], and specifically the ease and speed of error correction offered by each of the tools used in the manual paradigm, as it has been also shown in other similar tasks [22]. To begin with, we find that the music sheet is rendered as it is being transcribed in VHV, which facilitates the comparison process because a division of the computer screen allows having both the original score and the transcribed version in the same viewing plane. This facilitates the detection of errors, which are easily corrected thanks to the reduced character syntax of the Humdrum encoding used in VHV. The situation changes in the case of Oxygen XML Editor because it lacks an instantaneous rendering. The result cannot be visualized until the transcription is finished and for that, an external tool, like Verovio, is required. This makes error correction a slow and tedious process. [12] On the other hand, the error correction in the CMME is similar to that of VHV, with the difference that in the former we correct graphical symbols instead of text characters. Finally, we must mention that in MuRET, the source image is present along with the transcription rendering during the whole process, facilitating the user's ability to detect and correct errors.

Note that we are not drawing conclusions from the standard deviation values because we believe that they are mainly caused by the variations in the number of symbols per page and not by the operation of the different alternatives. We nonetheless provide them for the sake of completing the results.

4.2 OMR-assisted paradigm

In order to gain insights into the OMR-assisted transcription paradigm, two scenarios are evaluated: (i) a pre-built one, where we employ an OMR system built for working with corpora of similar characteristics to the one we want to transcribe (Aruspix), and (ii) a from-scratch one, where

[12] The interactive MEI encoding tool MEISE [23] has been discarded as it failed to render mensural notation.

an OMR system is built anew by means of MuRET.

We measure the average procedure time per page in both scenarios. In the from-scratch one, as the goal is to know the number of pages that must take part in the training set to achieve a model with an acceptable recognition accuracy, we will measure the average procedure time by increasing the training set by one page each time, until reaching 10 pages. After that, the number of pages will increase by 5, as the number of pages becomes less relevant. Note that in this case, the procedure time only refers to the model recognition and the revision of the recognized manuscript times.

Table 2 and Table 3 show the average times and their standard deviations obtained in the two OMR-assisted scenarios considered. As in the previous paradigm, we will not take into account the standard deviation during the analysis because they are mainly caused by variations in the number of symbols per page.

Table 2. Average time and its standard deviation for the transcription of one page of the *Magnificat* corpus for the pre-built OMR-assisted transcription paradigm, in which Aruspix is the OMR system used. Standard deviations are not considered relevant as they are mostly linked to the variability of the number of symbols of each page and scarcely to the transcription tool.

	Average time per page	Standard deviation
Pre-built OMR-assisted paradigm	8′ 39″	3′ 10″

It should be noted that Aruspix has been considered as a static model, i.e., the dynamical improvement feature has not been used as each page has been recognized independently. We want to establish comparisons between two possible OMR-assisted transcription methodologies.

We now proceed to compare the two OMR-assisted scenarios. The pre-built scenario allows us to evaluate how OMR systems facilitate the transcription process without the need to transcribe some test pages to first train the model. Looking at Table 2, we observe that it takes less than 9 minutes to correctly transcribe a typeset musical score written in mensural notation. This indicates that we are leveraging previous efforts and existing labeled data. Moreover, the transcription process is smooth thanks to the superimposition feature that allows for straightforward comparison with the recognized results.

Oppositely, the second situation allows us to study how many test pages have to be manually transcribed to train an OMR system before it can be used in the transcription process. According to the results in Table 3, the average procedure time decreases as the number of training pages increases. This drop is very steep at the beginning, especially when considering relatively small training sets of three pages or less. A model trained with one page gives a mean transcription time of approximately 49 minutes, whereas one trained with 3 pages gives an average procedure time of more or less 12 minutes. Moreover, training sets of 6 pages, or more, estimate lower transcription times than the pre-built OMR-assisted scenario.

4.3 Manual vs. OMR-assisted

To establish a comparison between both transcription paradigms, manual and OMR-assisted, we estimate the workdays needed for the complete transcription of the *Magnificat* corpus in both of them, by multiplying the number of pages with the average procedure time per page obtained in Section 4.1 and Section 4.2, respectively.

It is important to note that, as seen in the previous sections, in both the manual and the pre-built OMR-assisted paradigms, the average procedure times are constant as they contemplate static scenarios where both the user effort and the tool's performance can be estimated as averages. Hence, the total estimated transcription time will also be a constant value. However, in the from-scratch OMR-assisted scenario, the average transcription time of a page changes as a function of the number of training pages, as they influence the accuracy of the model used. Therefore, the time spent in the transcription of the complete corpus will vary as a function of the training pages, and to estimate it, the time spent in manually transcribing those training pages must be taken into account. We have to point out that the model's training time is not considered because it can be done in background [13] .

Figure 2 reports the workdays needed to transcribe the *Magnificat* corpus completely (126 pages) for both transcription paradigms.

By examining Figure 2, we draw several conclusions. The most important one is that the pre-built OMR-assisted transcription paradigm yields much shorter times, almost up to 5 workdays less, compared to the best case of the manual transcription paradigm. This result could be intuited when comparing the average procedure times per page obtained in the previous sections. This demonstrates that OMR systems can be embraced as a really helpful alternative, even when it does not completely automatize the process.

Additionally, the results show that by labeling only 3 pages to train an OMR model, the workdays needed for the total transcription of the corpus are less than in the manual paradigm for any of the tools considered there. This means that even if a model has to be built for a specific corpus, the OMR model-assisted transcription saves the user time, compensating for the time spent correcting errors. However, this trend is not constant; in our case, after 10 pages of training, the workdays for the from-scratch OMR-assisted scenario begin to increase. This is because the recognition accuracy of the model does not increase enough to reduce the average procedure time per page by a large amount, resulting in the time spent on the manual transcription not being compensated. In other words, a slight reduction in the OMR-assisted transcription time, of the order of a minute per page, does not make up for the manual transcription

[13] It takes approximately 7 minutes to train a state-of-the-art OMR model with a training set of 30 pages with a low-profile GPU unit.

Table 3. Average times and their standard deviations for the transcription of one page of the *Magnificat* corpus for the from-scratch OMR-assisted transcription paradigm. This scenario uses MuRET as it allows to retrain its OMR model. Times are given as a function of the number of training pages, as they directly affect the accuracy of the model.

From-scratch OMR-assisted paradigm														
Training pages	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**	**20**	**25**	**30**
Avg. time per page	48′ 02″	32′53″	11′ 56″	9′ 48″	8′ 54″	6′ 46″	5′ 25″	5′ 25″	4′ 58″	4′ 47″	3′ 52″	3′ 33″	3′ 31″	3′ 15″
Standard deviation	2′ 49″	9′ 29″	3′ 40″	3′ 41″	3′ 18″	3′ 09″	2′ 51″	2′ 41″	2′ 45″	2′ 48″	2′ 44″	2′ 42″	2′ 49″	2′ 44″

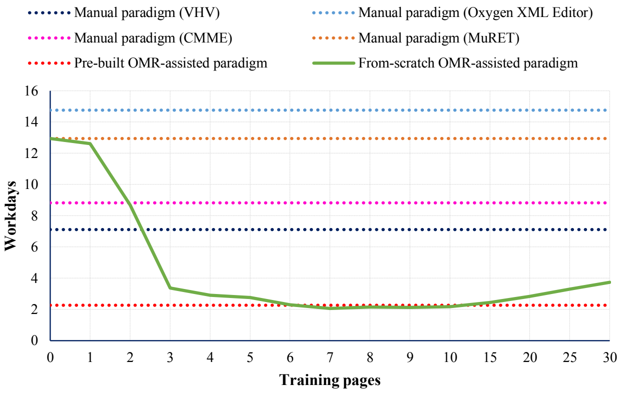

Figure 2. Workdays (assuming 8 hours per day) needed to transcribe the 126 typeset pages of the *Magnificat* corpus. In the manual and pre-built OMR-assisted paradigms, the curves are constant as they simulate static scenarios, where the user effort can be estimated as an average. However, in the from-scratch OMR-assisted case, the tendency varies depending on the number of pages used to train the OMR system, since this affects the accuracy of such a model and therefore, the subsequent error correction effort made by the user. It should be noted that in this scenario there is an inflection point after which the effort made in the manual transcription of the training pages does not compensate for the improvement of the model.

time, of the order of fifty minutes per page, needed to train such a system.

As the last point to mention, we expect that if more complex compositions are introduced, such as cross-staff notation, the difficulty of correction with the corresponding tools increase. In addition, if the OMR systems behave as expected, then we could say that the trend of the curves will maintain and the conclusions drawn can be extrapolated to other corpora. In any case, the methodology presented would still be valid.

5. CONCLUSIONS

The transcription of existing musical heritage, available only as physical documents, is a necessary activity for their preservation, access, and dissemination. Optical Music Recognition (OMR) was born with the goal of facilitating the manual transcription process by its automation. However, after decades of research and promising results, its inclusion in the transcription workflow as a user tool has not materialized. Acceptable but not accurate results have relegated OMR systems to just representing a scientific challenge to solve.

For all the stated above, we posed the following question in this work: to what extent do OMR systems facilitate the transcription process? To answer the question, we set up an experiment that allows us to draw meaningful comparisons between two transcription methodologies: a fully manual one and an OMR-assisted one. For that, we transcribe the content of a printed early music work written in white mensural notation under both paradigms and compare the procedure time. Additionally, we evaluate the OMR-assisted paradigm from two points of view: (i) a pre-built one, where an OMR system built to recognize works of similar characteristics to those of the test case, to take advantage of previous efforts and see if the error correction compensates such savings; and (ii) a from-scratch one, where we train an OMR system from start, to see if the manual transcription of the training pages is rewarded later with the automatic transcription.

The obtained results estimate that in both cases of the OMR-assisted paradigm, the user time is less than that of the best case of the manual paradigm, indicating that posterior correction of errors in the automatic transcription is more than offset by the time saved when compared with manual transcription. The correct usability of the system to correct errors is an important issue that needs to be taken into account, as the usefulness of any transcription tool affects the process itself. Thanks to the simultaneous rendering of the transcript, the detection, and change of possible errors is a smooth process in the OMR-assisted paradigm.

Despite the first impression that OMR systems fail, they can be considered as a useful transcription tool, as they reduce the cost of the most valuable non-renewable resource, time. As future research, we aim at introducing more complex corpora, such as those involving handwriting and/or polyphonic pieces, as well as more task completers, to go beyond a single case study and provide more general insights into the analyzed question.

6. ACKNOWLEDGMENTS

First author's research is funded by the Spanish Ministerio de Universidades through grant FPU19/04957.

7. REFERENCES

[1] L. Pugin, "The Challenge of Data in Digital Musicology," *Frontiers in Digital Humanities*, vol. 2, p. 4, Aug. 2015.

[2] M. Good and G. Actor, "Using MusicXML for file interchange," in *Proceedings Third International Conference on WEB Delivering of Music*, Leeds, UK, Sep. 2003, p. 153.

[3] A. Hankinson, P. Roland, and I. Fujinaga, "The Music Encoding Initiative as a Document-Encoding Framework," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami (Florida), USA, Oct. 2011, pp. 293–298.

[4] G. Jones, B. Ong, I. Bruno, and K. Ng, "Optical Music Imaging: Music Document Digitisation, Recognition, Evaluation, and Restoration," in *Interactive Multimedia Music Technologies*. IGI Global, 2008, pp. 50–79.

[5] J. Calvo-Zaragoza, J. H. Jr., and A. Pacha, "Understanding Optical Music Recognition," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–35, Sep. 2020.

[6] D. Bainbridge and T. Bell, "The Challenge of Optical Music Recognition," *Computers and the Humanities*, vol. 35, no. 2, pp. 95–121, May 2001.

[7] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. S. Marçal, C. Guedes, and J. Cardoso, "Optical music recognition: state-of-the-art and open issues," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 173–190, Oct. 2012.

[8] M. Urberg, "Pasts and Futures of Digital Humanities in Musicology: Moving Towards a 'Bigger Tent'," *Music Reference Services Quarterly*, vol. 20, no. 3-4, pp. 134–150, Oct. 2017.

[9] L. Chen, E. Stolterman, and C. Raphael, "Human-Interactive Optical Music Recognition," in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, New York City, United States of America, Aug. 2016, pp. 647–653.

[10] L. Pugin, "Aruspix: An Automatic Source-Comparison System," *Computing in musicology: a directory of research*, no. 14, pp. 49–59, 2006.

[11] H. Bitteur, "Audiveris," 2004. [Online]. Available: https://github.com/Audiveris/audiveris

[12] A. N. Hankinson, "Optical Music Recognition Infrastructure for Large-Scale Music Document Analysis," Ph.D. dissertation, McGill University, Montréal, Québec, Canada, Dec. 2015.

[13] D. Rizo, J. Calvo-Zaragoza, and J. M. Iñesta, "MuRET: A music recognition, encoding, and transcription tool," in *Proceedings of the 5th International Conference on Digital Libraries for Musicology (DLfM 2018)*, Paris, France, Sep. 2018, pp. 52–56.

[14] B. S. Brook, "The Simplified 'Plaine and Easie Code System' for Notating Music: A Proposal for International Adoption," *Fontes Artis Musicae*, vol. 12, no. 2-3, pp. 156–160, Jan. 1965.

[15] J. Howard, "Plaine and Easie Code: A Code for Music Bibliography," in *Beyond MIDI: The handbook of musical codes*. JSTOR, 1997, pp. 362–372.

[16] D. Huron, "Humdrum and Kern: selective feature encoding," in *Beyond MIDI: The handbook of musical codes*. MIT Press, 1997, pp. 375–401.

[17] C. S. Sapp, "Verovio Humdrum Viewer," in *Music Encoding Conference*, Tours, France, May 2017.

[18] D. Rizo, N. Pascual-León, and C. S. Sapp, "White Mensural Manual Encoding: from Humdrum to MEI," *Cuadernos de Investigación Musical*, no. 6, pp. 373–393, 2018.

[19] L. Pugin, R. Zitellini, and P. Roland, "Verovio: A library for Engraving MEI Music Notation into SVG," in *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, Taipei, Taiwan, Oct. 2014, pp. 107–112.

[20] T. Dumitrescu, "Computerized Mensural Music Editing (CMME) Project," 1999. [Online]. Available: https://www.cmme.org

[21] J. Tidwell, *Designing Interfaces*. O'Reilly Media, Inc., 2005.

[22] J. D. Valor Miró, J. A. Silvestre-Cerdà, J. Civera, C. Turró, and A. Juan, "Efficiency and usability study of innovative computer-aided transcription strategies for video lecture repositories," *Speech Commun.*, vol. 74, no. C, p. 65–75, Nov. 2015.

[23] N. Beer, M. Hartwig, and K. Herold, "The MEI Score Editor: Use Cases and Future Perspectives," in *Music Encoding Conference Proceedings 2013 and 2014*, 2016, pp. 142–146.