

EVALUATING OMR ON THE EARLY MUSIC ONLINE COLLECTION

Laurent Pugin

Swiss RISM Office
laurent.pugin@rism-ch.org

Tim Crawford

Goldsmiths College, University of London
t.crawford@gold.ac.uk

ABSTRACT

The Early Music Online (EMO) collection consists of about 300 printed music books of the sixteenth century held at the British Library. They were recently digitized from microfilms and made available online. In total, about 35,000 pages were digitized. This paper presents an optical music recognition (OMR) evaluation on the EMO collection. Firstly, the content of the collection is reviewed, looking at the type of music notation and the type of printing technique. Secondly, for the books for which it is possible (260 books), an OMR evaluation performed using the Aruspix OMR software application is presented. For each book, one randomly selected page of music was processed and the recognition rate was computed using a corrected transcription of the page. This evaluation shows very promising results for large-scale OMR on the EMO or similar collections. The paper also highlights critical points that should be taken into account in such an enterprise.

1. INTRODUCTION

Over the last decade, libraries around the world have been digitizing their collections extensively, making available online not only books but also music scores. For books, digitization projects most often include an optical character recognition (OCR) step that renders the content of the images searchable in a similar way, for example, to Google Books. Enabling full-text access has radically changed the search paradigm for the user; we no longer search in the same way with the complete content of books as we used to with only the restricted metadata available in a library catalogue. When music scores are digitized, however, they usually remain searchable only through textual metadata, since no transcription of the content is made available during the digitization process. This is mostly due to the fact that optical music recognition (OMR), the equivalent of OCR for music, is a very challenging task [11]. Furthermore, performing OMR on a large scale is particularly complicated because of the heterogeneous nature of music notation and of different types of music scores. It makes it difficult to set up a so-

lution that works acceptably well in all cases. However, preliminary attempts to perform OMR on a large scale on the IMSLP Petrucci Music Library have shown interesting results even if at this stage they are still lacking a clear and systematic evaluation [13].

With historical documents in general, performing OCR remains a challenge. In particular, some steps of the OCR process remain critical, in particular the preprocessing of documents that are often highly degraded, and the layout analysis. More recently, large research projects have focused on OCR for historical documents, such as the IMPACT project, with the aim of developing specific tools for the task [2]. For historical music documents, the leading OMR project is SIMSSA, currently with a focus on the handwritten repertoire in neumes and square notation (before 1500) used for experimenting new methods and testing innovative online tools [7].

This paper focuses on printed music from the 16th century, a period of enormous expansion in the availability of music for domestic use. It was made possible by the development of a typographic technique adapted from Gutenberg's invention of typesetting. This remained by far the most widely-used method for printing music until the beginning of the 18th century when it was superseded by engraving. Typographic music printing was invented by Ottaviano Petrucci in 1501 and simplified by Pierre Attaignant in 1527. The former introduced a technique using two or three passes through the press, while Attaignant developed a single-impression technique that made the process commercially sustainable, with the consequence that the number of printed scores increased dramatically from then on. Printed music scores of the time represent a valuable and rich source of material for musicologists, musicians and historians alike.

We present in this paper an evaluation of OMR on the Early Music Online collection, a set of about 300 books held at the British Library and recently digitized from microfilms [12]. The evaluation is twofold. First we evaluated which of the books in such a collection could be processed using the Aruspix OMR software application developed specifically for typeset music prints.¹ In particular, we were interested to see what percentage of a library collection like this can be processed with existing technology. For this, we looked at the type of music notation, the type of format and the type of printing technique. With this in hand, we were able to define a set of books

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2013 International Society for Music Information Retrieval

¹ <<http://www.aruspix.net>>

to be used for the second part of our evaluation, which was to estimate what recognition rate we can obtain in an OMR workflow without human correction of the data.

In the next section, we introduce the EMO data set. The experiments for evaluating the OMR recognition rates and the results are presented in the following sections. Remarks on future work conclude the paper.

2. THE EMO COLLECTION

In 2011, the British Library together with Royal Holloway, University of London and the UK group of the Repertoire International des Sources Musicales (RISM) conducted a digitization project of sources held at the British Library. In total, 324 volumes of 16th-century printed music, mostly individual voice part-books of vocal music, were re-catalogued, digitized and made available online together with very rich metadata that includes information about the title pages, the dedicatees, the printer and the printing place, and more. It comprises approximately 35,000 pages for a total of about 10,000 pieces of music. The books chosen for this digitization project were selected from the RISM B/I and B/II Series covering printed anthologies. Anthologies (collections of music by multiple composers) were selected as a priority because their content is not as well known as that of individual prints (RISM Series A/I). The books were digitized from the microfilm collection of the British Library and are available under the JISC Collections Open Education User Licence version 1.0.

2.1 Notation types

Most of the books in the EMO collection are in mensural music notation on five line staves, the most common notation of the time. The collection also contains twenty-seven books of lute tablature (e.g., K.1.c.11)¹ and about five books of organ tablature (e.g., K.8.h.22). Six books are interesting cases with a mix of mensural notation and lute tablature (e.g., D.250 and K.2.h.12). There is also one book that contains a very rare case of a mixture of five-line mensural notation and four-line square notation, printed by Etienne Gueynard in Lyon in 1528 (K.9.a.23).

2.2 Book formats

The large majority of the books of the EMO collection are part-books, a standard format of the time. In part-books, each voice or instrumental part is printed in a separate gathering. Sometimes, two or more voices are printed in the same book, for example when a book in four voices (i.e., four part-books) also contains pieces with more voices. In these cases, the different voices appear as in a choir book for the singers or musicians to be able to sing or play together by sharing the part-book. The EMO collection also contains proper choir books

(e.g., K.9.a.11) where all voices are printed in a single book. Finally, the EMO collection contains ten table books mostly printed by Jacques Moderne in Lyon (e.g., K.10.a.7). The table books are similar to choir books with the main difference that the voice-parts are not all printed in the same direction. They were intended to be laid on a table with musicians sitting on different sides of the book. All the table books in EMO contain parts printed upside-down, but books with parts in three or four directions also exist.

2.3 Printing techniques

About 300 books in the EMO collection were printed using the single impression typographic technique introduced by Attaignant. His innovation was to use type-faces where the staff lines and the notes were printed at the same time; this became the most widely-used technique for printed music for nearly two centuries. The EMO collection also contains 13 books printed with the multiple impression technique, mostly printed by Petrucci (e.g., K.1.d.12), one of the very few printers who used this technique. There are also seven books printed from engraved plates by Simone Verovio (K.8.b.17) who was one of the first music engravers, and five books in woodcut mostly by Andrea Antico (e.g., K.8.b.7).



Figure 1. A page printed in multiple impressions by Petrucci left out of the experiments. The processing of pages printed with this technique is problematic when ornate letters are superimposed on the music staves.

3. EXPERIMENTS

3.1 Data set for the OMR evaluation

By looking at the different characteristics in terms of notation type, book format and printing technique, we were able to determine which were the books that could be processed successfully with the OMR tool that was to be used for the evaluation. We selected all the books in mensural notation printed with the single impression technique.

We left out the books in multiple impressions, even though they could also have been processed, albeit certainly yielding less accurate results. As shown on Figure 1, ornate letters are sometimes superimposed on the music staves, which makes small areas of the page badly recognized by the OMR software application used for our evaluation. We also left out lute tablatures and books printed from engraved plates that are not appropriate for the OMR software application. Finally, we also left out

¹ The reference for the books is their signature at the British Library. Images can be accessed from the RISM UK website (www.rism.org.uk).

the table books, the books with a mixture of lute and mensural notation and also the book with square notation.

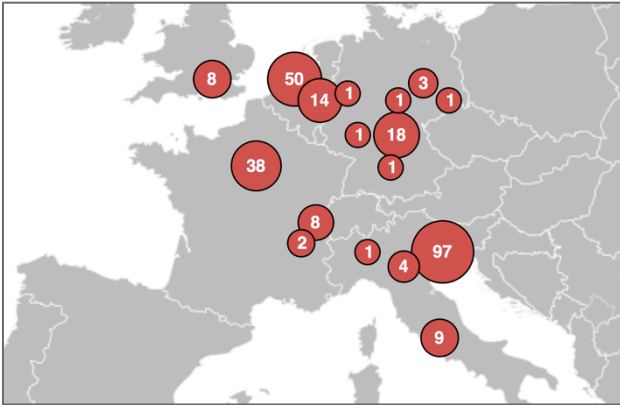


Figure 2. The 260 books selected for the OMR evaluation were printed in 17 different cities, reflecting the importance of the music printing centres of the time.

We ended up with a total of 260 books representing 80% of the EMO collection. These 260 books were printed by no less than 48 different printers coming from 17 cities (plus two from unknown locations). Interestingly, the provenance of the books reflects quite precisely the music printing production of the different cities at that time (see Figure 2); in particular, more than a third of the books were printed in Venice [3].

The diversity of the data set can also be visualized by looking at the different font shapes. In Table 1, we present the 18 different G-clef shapes found in the pages we processed for our evaluation.

	Amadino		Antico		Ballard
	Donangeli		Ang. Gardano		Gerlach
	Hucher		Moderne		Petrejus
	Phalèse		Scotto		Short
	Susato		Vissenaken		Waelrant
	[unknown]		[unknown]		[unknown]

Table 1. The font and woodcut shapes in the books used for the evaluation are highly variable as illustrated by the 18 different G-clef shapes appearing in the data set.

3.2 Procedure

The experiments were performed using the original gray-scale images in uncompressed TIFF format at 400dpi, the resolution used by the British Library during the EMO digitization process. This resolution, however, does not provide accurate information about the original document size because of all the different parameters that were involved in the making of the microfilms and the scanning. As an indication, looking at the staff height in pixels, we noticed in the EMO images that the staff height goes from 60 up to 300 pixels, with an average value of 122 pixels and a median value of 108 pixels. Of course, this is only an indication since the actual staff sizes on the sources themselves vary.

In the EMO images, each image always contains two pages. Since the OMR tool used for our experiments works on single page images, the images had to be split. The images were split automatically by selecting a little more than half the image width for each page (55% of the width). Because the books are always centered in the EMO images and because the tool performs a border removal, this simple method worked well in all cases. For each of the 260 books of our data set, one page of music was randomly selected. For the books where the random selection process did not retrieve a page containing music (9 books), the first subsequent page of music was selected by hand.

The recognition process in the tool used for the experiments includes a binarization step that converts the gray-scale image into a black and white one. This step has been shown to be critical on degraded music documents [8]. The OMR tool offers two options for the binarization, one for documents in good condition (B1), and another based on an algorithm specifically designed for highly degraded documents with marked bleed-through (B2). The B1 algorithm is the Brink & Pendock 1996 algorithm, and B2 its modified three-class version, Pugin 2007, the two best-performing algorithms described in [4].

The choice of the binarization option is left to the user and is not determined by the tool. In our experiments, we allowed the evaluator to choose in advance the algorithm that seemed the most appropriate for each book based on a visual evaluation of its degradation. B1 was applied to 159 books, and B2 to 101. In order to evaluate how appropriate the choice of the evaluator was, we also processed all the pages for which B2 was selected with the B1 options and compared the results.

All the pages were processed without human interaction. That is, without any processing indication or correction other than the binarization method selection. In that sense, the results provide a clear indication of what would be obtained with this technology on a large-scale OMR process involving minimal human intervention for preparing the process and no data correction at all.

4. RESULTS

All but 2 of the 260 pages were successfully recognized. Only two pages produced unusable results because the staff detection step in the OMR process failed. The two pages were taken from books printed by Giorgio Marescotti (C.219.a) and Antonio Gardano (K.3.1.7).

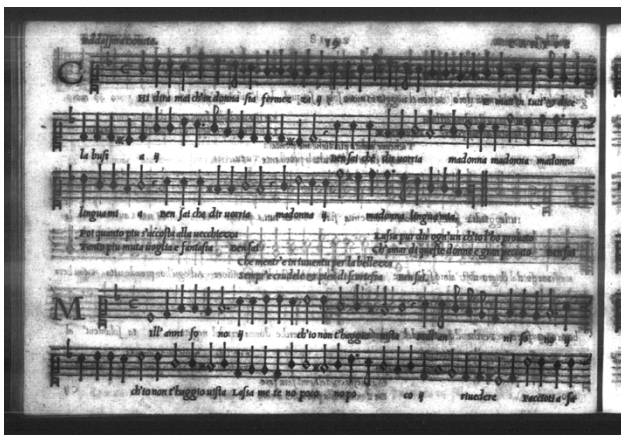


Figure 3. For two pages in the data set the recognition process failed. With the first, both the quality of the typeface used for printing and the quality of the microfilm were poor. With the second, the bleed-through was too strong and the microfilm quality poor.

Both failures can be explained by the poor quality of the documents as shown in Figure 3 for the prints by Marescotti and Gardano. In the first case, the quality of the print itself was very bad because it had been printed with a deteriorated typeface that produced extremely irregular staff lines. The quality of the microfilm and of the image was also problematic. In the second case, the book was highly degraded, with very marked bleed-through. Nothing in the layout of the documents indicates that the recognition would have been problematic, had they been in better condition.

For all the pages that were processed successfully (258), the OMR transcription was fully corrected by hand in order to be able to compute recall and precision following the evaluation procedure previously used for the tool [10]. This method uses a standard best alignment approach for evaluation of the number of substitutions, deletions and insertions. For all symbols, exact matching is required, e.g., for a music note, a match occurs only when both pitch and duration are the same. It is also important

to mention that the misrecognition of a clef does not affect the evaluation of the subsequent pitches, even though technically pitches would be incorrect – it is also not rare to see cases where the clef is actually wrong in the source itself.

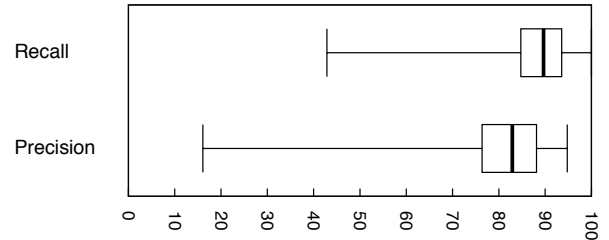


Figure 4. Aside from a few pages that produced poor results, the quartiles show that three quarters of the pages produced a recall rate between 85% and 100% and a precision rate between 76% and 95%.

The median recall rate is 90% and the median precision rate is 83%. Some pages did produce bad results. However, it can be seen that this is the case for only a few of them. Although the lowest recall rate is just 43%, the lower recall rate quartile is 85%, as shown on Figure 4. We noticed that the lowest recall rate was obtained on a fairly atypical page; it contained only two staves with few music symbols on them and also has added handwritten slurs, which explain the poor results.

Looking at the results by printer did not highlight very significant differences. Figure 5 shows the quartiles for the five printers for which we had more than 10 books in the data set. Overall, it seems that document degradation has more impact on the results than the printers, although the difference is slight.

Some of the books in the data set were printed using nested typefaces. With this technique, types with not only five lines but also four or three lines are used together with line elements overlapping several types horizontally. The technique had the advantage of requiring a smaller number of type shapes. It was used mostly in the Low Countries, in France and in Germany, by printers such as Robert Ballard and Tielman Susato. Our experiments show that the nested technique used by some printers do not cause major problems. Though this might be an explanation why the results obtained on the prints by Ballard are more variable, the results obtained on prints by Susato (19 books) are the same as that obtained on the prints by Antonio Gardano (30 books) with an average recall rate of about 92% and an average accuracy of 86%.

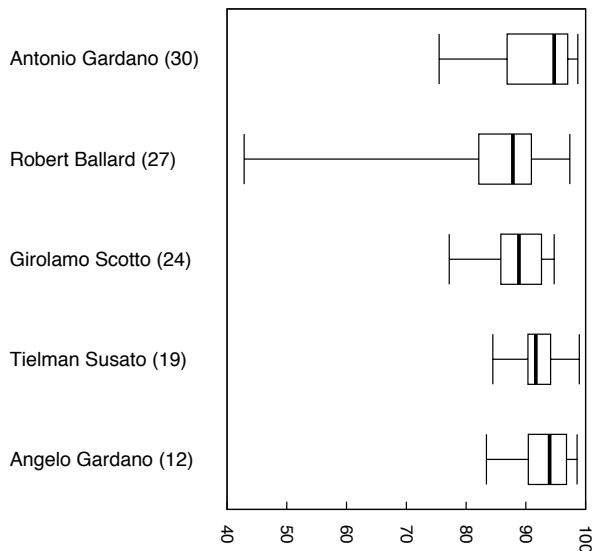


Figure 5. The quartiles for the prints with more than 10 books in the data set shows differences, albeit small ones. The results on the prints of Ballard are more variable and can be lower.

For the binarization algorithm, the results show that the choice made by the human evaluator before the processing of the book was appropriate in most cases. More precisely, the B2 binarization option gave better results than B1 in 80% of the cases where it was chosen by the evaluator. In the 21 cases where this was an inappropriate choice, it produced a recall rate significantly lower (by more than 5%) in one third of the cases. Figure 6 shows a page where B2 did improve the results and a case where it did not. In the first case, the recall rate was 69% with B1 and 91% with B2. In the second case, it was 97% with B1 but only 83% with B2, undoubtedly because the bleed-through is not marked enough.



Figure 6. Two pages for which selecting the B2 binarization option did (above) and did not (below) improve the recognition results.

For the case where algorithm B2 improved the recognition results, we can clearly see looking at the binarized images shown on Figure 7 that B1 was not appropriate, in particular where the ornate letter printed on the other side of the page clearly appears through the paper.



Figure 7. The binarization results with B1 (above) and B2 (below). The B2 option was clearly more appropriate for removing the bleed-through, in particular where the ornate letter is printed on the other side.

Correcting the results for this evaluation highlighted a handful of characteristics that evidently caused recognition errors. Prints with a strophic repertoire, mostly French chansons and lieder, with multi-line lyrics printed under the notes, often caused pre-processing imprecisions that then had an impact on the final recognition process. The problem with multi-line lyrics was that some of them were not properly recognized as text and the software thus attempted to read them as music symbols. Oblique ligatures (appearing on 6 pages) were inconsistently recognized. Other rarities that caused problems were passages with an unusual number of ledger lines (1 page) and the aforementioned handwritten slurs (1 page). In terms of document condition, the typical problems that locally affected the recognition process are: the presence of library stamps (2 pages), stains (4 pages), scratched-out notes (1 page), pages with extremely light inking (2 pages), or curvature on the edge of the page (9 pages). Additionally, in a few cases, because of very high bleed-through (5 pages), the accuracy of the corrected transcription used for the evaluation is questionable due to readability issues.

5. CONCLUSION AND FUTURE WORK

By evaluating OMR on the EMO collection, this paper sets a baseline for large-scale projects in the field. Our evaluation showed that about 80% of the collection can be processed and that we can expect a recall rate of be-

tween 85% and 100% for three quarters of the pages. In comparison, the Bibliothèque Nationale de France retained a minimal recognition rate of 60% for OCR results to be included as full-text [5]. This was based on an evaluation that showed this to be the limit for results to be useful for the user. Although searching text is different from searching music, and the usability of the data needs to be clearly estimated, our evaluation already shows an interesting potential for making the EMO or similar collections searchable. Furthermore, the evaluation does not take into account the improvement that can be obtained with adaptive OMR techniques that have proven to be extremely valuable in such cases [9].

The EMO collection is certainly an excellent example of what we can expect to find in library collections of similar size. Studying it was informative about which types of sources can be problematic. Whilst 20% of the sources had to be left out, there are other state-of-the-art tools that could be used to process some of them, in particular the lute tablature [6]. For others, specific modules would need to be developed, for example for the layout analysis of table books, while books such as the ones printed from engraved plates are certainly more problematic.

We can also notice that some categories that had to be left out are rather highly represented in EMO, quite likely because it contains only anthologies. If we compare this with the RISM A/I Series of individual prints, out of about 4,200 prints inventoried for the 16th century, around 3% are lute tablatures (10% in EMO), 0.6% are prints by Petrucci in single impression (3%) and only three prints by Verovio in engraving are inventoried (2%). Of course, this does not take into account the numerous re-editions to be found in the A/I Series. It does, however, show that the 20% exclusion part we ended up with in EMO is certainly an upper limit if 16th century prints are considered on a larger scale and 5% to 10% is a more likely percentage range on a collection with individual prints.

In terms of workflow, detecting which pages are missing music would be necessary for avoiding preparatory work. Experiments have already been conducted on this [1] and the SISSMA project is currently working on similar issues. Our experiments have also shown that strophic music was not always properly handled in the preprocessing. This confirms it is always worth improving the layout analysis because it is a critical step in the process. For the binarization, having the user select the binarization algorithm for a book is acceptable because it is a quick task. The cases where B2 was a bad choice could certainly have been avoided with better training of the user. However, user-selected binarization is not optimal because it assumes the degradation to be identical throughout the book, which is of course not always the case. Having the most appropriate binarization algorithm selected automatically would be a valuable improvement. Future work will also include reassembling the parts for

reconstructing the score, which will open the door to polyphonic processing of the data.

6. REFERENCES

- [1] D. Bainbridge, and T. Bell: "Identifying music documents in a collection of images," in *Proceeding of the ISMIR 2006 Conference*, pp. 47–52.
- [2] H. Balk, and L. Ploeger: "IMPACT: Working together to address the challenges involving mass digitization of historical printed text," *OCLC Systems & Services*, 25(4), pp. 233–248, 2009.
- [3] J. Bernstein: *Print Culture and Music in Sixteenth-Century Venice*, Oxford, Oxford University Press, 2001.
- [4] J. A. Burgoyne, L. Pugin, G. Eustace, and I. Fujinaga, "A comparative survey of image binarisation algorithms for optical recognition on degraded musical sources," in *Proceedings of the ISMIR 2007 Conference*, pp. 509–12.
- [5] G. Cron: "Corpus." Presentation, Journée IMPACT, BnF, Paris, September 19, 2011.
- [6] C. Dalitz, and C. Pranzas: "German Lute Tablature Recognition," in *Proceedings of the ICDAR 09*, pp. 371–375.
- [7] A. Hankinson, J. A. Burgoyne, G. Vigliensoni, and I. Fujinaga: "Creating a large-scale searchable digital collection from printed music materials," in *Proceedings of the 21st ACM Conference on World Wide Web*, pp. 903–908.
- [8] L. Pugin, J. A. Burgoyne, and I. Fujinaga: "Goal-directed evaluation for the improvement of optical music recognition on Early music prints," in *Proceedings of the ACM-IEEE JCDL 2007*, pp. 303–04.
- [9] L. Pugin, J. A. Burgoyne, and I. Fujinaga, "MAP adaptation to improve optical music recognition of early music documents using hidden Markov models," in *Proceedings of the ISMIR 2007 Conference*, pp. 513–16.
- [10] L. Pugin, J. Hockman, J. A. Burgoyne, and I. Fujinaga, "Gamera versus Aruspix: Two optical music recognition approaches," in *Proceeding of the ISMIR 2008 Conference*, pp. 419–24.
- [11] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. Marcal, C. Guedes, and J. S. Cardoso: "Optical music recognition: state-of-the-art and open issues," *International Journal of Multimedia Information Retrieval*, 1(3), pp. 173–190, 2012.
- [12] S. Rose: "Early Music Online," *Early Music Performer*, 30, pp. 22–25, 2012.
- [13] V. Viro: "Peachnote: Music score search and analysis platform," in *Proceedings of the ISMIR 2011 Conference*, pp. 359–362.