

Big Data Optical Music Recognition with Multi Images and Multi Recognisers

Kia Ng
ICSRiM - University of Leeds,
School of Computing and School of Music,
Leeds LS2 9JT, UK
kia.ng@icsrim.org.uk

Alex McLean
ICSRiM - University of Leeds,
School of Music,
Leeds LS2 9JT, UK
alex.mclean@icsrim.org.uk

Alan Marsden
Lancaster Institute for the
Contemporary Arts,
Lancaster University, LA1 4YW, UK
a.marsden@lancaster.ac.uk

In this paper we describe work in progress towards Multi-OMR, an approach to Optical Music Recognition (OMR) which aims to significantly improve the accuracy of musical score digitisation. There are a large number of scores available in public databases, as well as a range of different commercial and open source OMR tools. Using these resources, we are exploring a Big Data approach to harnessing datasets by aligning and combining the results of multiple versions of the same score, processed with multiple technologies. It is anticipated that this approach will yield high quality results, opening up large datasets to researchers in the field of digital musicology.

Optical music recognition. Big data. Music corpus analysis. Music information retrieval.

1. INTRODUCTION

Vast quantities of digital images of musical scores are now available from services such as the International Music Scores Library Project (IMSLP, <http://imslp.org/>). This database alone contains just under a quarter of a million images of scores encompassing over seventy thousand pieces of music. Most of this information is opaque to computational analysis because the scores are scanned and stored as raster images, with none of the symbolic musical information represented. To access the musical information, the score needs to be “read”, converted to musical note data in a symbolic representation. Until the scores have been processed in this way they cannot be used for the purpose of Music Information Retrieval (MIR), or other technological approaches in digital and computational musicology.

Although software to read digital images of words, turning them into encoded text (Optical Character Recognition, OCR), is now commonplace, and underlies such search and analysis services as Google Books, the equivalent task for music (Optical Music Recognition, OMR) has proven to be very challenging (Ng 2004, Bellini 2007, Jones *et al.* 2008). The accuracy of the results is often low, and some researchers have reported that the time taken to find and correct errors is so great as to make it just as quick to enter the data by hand. In

other cases, considerable time is required to train OMR software to get good results from a particular source. Much of the reason for this lack of accuracy is due to the complexity of scores in comparison with text; musical scores have a great deal more vertical structure, and as scores generally include some text, OCR can in a sense be considered a subset of OMR.

As databases of musical scores grow and are consolidated, multiple images of a score of a piece of music are increasingly available. Using more than one image increases the information available for the OMR task. Our project explores two approaches to improving OMR through the use of such sources. The first post-processing approach develops software to take the outputs of OMR software on different sources for the same piece of music, and combines them into a single, more accurate, representation of the piece. The second approach utilises multiple OMR software to improve the accuracy of recognition, building on earlier related work (Byrd & Schindele 2006, Knopke & Byrd 2007).

This paper presents the research programme of this project, which aims to turn the supposed problem of large quantities of inaccessible data into an advantage, making use of the quantity of information available to improve the accuracy of OMR.

2. RESEARCH QUESTIONS

Very large numbers of digital images of music scores are now available in on-line repositories, beyond the scope of manual analysis methods. If musicology is to benefit from this wealth of information, techniques are required for “reading” large numbers of digital image files in order to extract the musical information they contain, at a sufficiently high level of accuracy, for later corpus based analysis. We approach this opportunity via the following core research question;

Can Optical Music Recognition (OMR, the equivalent of OCR for text) achieve a sufficiently high level of accuracy sufficient for building large databases of musical information by making use of the redundant information in multiple sources of digital-image information for single pieces of music?

We call this application of OMR to multiple sources of information *multi-OMR*, and break our central question down into the following components:

- i. What is the best way to align the information from multiple outputs of OMR software (e.g., the output from more than one software program on a single digital image, or the output from a single program when applied to different images of the same piece, or the output when applied to an image of a full score and when applied to images of the individual instrumental parts, etc.) so that information from one can be used to correct errors or fill in gaps in another?
- ii. What level of accuracy in output, with respect to different kinds of information (e.g., pitches, durations, dynamics), can be expected from such multi-OMR?
- iii. What is the best way to adapt the techniques embodied in current OMR software to profit from multiple sources of information (e.g., in determining whether a particular note has an open or filled note-head)?
- iv. What level of accuracy can be achieved from such adapted OMR techniques?

Success in multi-OMR would go only part of the way towards the sophisticated automatic “reading” of the wealth of score-images available on-line, but we believe it is a step which will produce rapid gains. The current state of OMR technology, briefly stated, is that it is a useful tool, but one which requires a considerable amount of human input, whether in training for use with a particular source, or in correcting errors, to produce useful results. The size of databases, which it is practical to

generate through the use of OMR, is therefore limited. Even a modest increase in the accuracy of OMR will allow considerable increase in this limit.

3. BIG DATA IN DIGITAL MUSICOLOGY

For international corporations such as Google, Facebook, and Twitter, Big Data brings particular challenges, related to databases which are not only spread across the world in huge data centres, but are continually changing, and relate to large portions of the population of countries. For humanities researchers, including musicologists, Big Data is rather smaller, but nonetheless represents a real and growing problem. In particular, datasets of a large and growing scale, present new challenges to researchers, requiring them to develop and adapt technology, to support and extend their exploration and understanding of the data.

The on-line Petrucci Music Library (also called IMSLP, the International Music Score Library Project) currently gives access to digital images of the scores or parts of over 70,000 pieces of music. There are also other easily available sources of digital images of music scores, such as the Neue Mozart Ausgabe (the standard scholarly edition of the complete works of Mozart). This has transformed music higher education, where it is now standard practice for students to access scores on-line rather than on paper. There are two impediments to using IMSLP and similar on-line resources for large-scale musicological research, however.

Firstly, the items are images, and extracting the musical information into a form by which it can be processed by computer requires a process similar to Optical Character Recognition, by which words are extracted from scanned text. There has been research on Optical Music Recognition for more than two decades, and several software programs are available, both proprietary and free/open-source. However, there are still many challenges to improve the robustness of the techniques to a level that is makes it practical to generate a large and accurate digital corpus from sources such as IMSLP.

One web service already using IMSLP data is Peachnote (Viro 2011; <http://www.peachnote.com>), a site which allows scores to be searched by pitch sequences. The database behind this was derived from applying OMR software to scores in IMSLP, but the presence of errors often shows through in the results. While Peachnote is nonetheless a highly useful and rewarding service, more accurate large-scale OMR would open the field up to a far wider range of musicological applications.

IMSLP, in common with other on-line collections of material uploaded by members of the community, is large but messy. Some items are scans of manuscripts, some of printed music, and some are Portable Document Format (PDF) files created with music typesetting software. Some are complete scans of full scores, some are separated into sections, some are scans of instrumental parts, and a few order the pages for printing in booklet style rather than consecutively.

To make use of this material, therefore, either requires careful and expert intervention at the stage of selection of materials in order to determine the precise nature of the data, or some automatic means of identifying what is in a particular item, in the sense of distinguishing parts from scores, and breaks between subsections. For example, some items in IMSLP contain a scan of the full score followed by scans of the individual parts in a single file.

We leave coping with many of the challenges of working with IMSLP and similar resources to another project, while turning the challenge of multiple sources of information often available into an advantage. In doing so we recognise scores that have been printed in multiple editions, and uploaded by multiple enthusiasts, may well be those of greater musicological interest for many cases and purposes. While we do not expect to be able to satisfactorily encode all the source data (e.g. full orchestral scores or complex late-nineteenth-century piano music are still likely to produce unusable results), even a small improvement on current OMR performance would make a very large number of pieces accessible to computational research.

4. TECHNOLOGICAL APPROACH

In the following we will outline our technological approach and design, for multi-OMR. While we will also investigate combining output from proprietary software, our discussion here will focus on the strand of our research engaging free/open source tools and frameworks, a number of which are available. MuseScore (<http://musescore.org/>) is a cross-platform desktop application for editing and rendering notated music, with an advanced plugin framework. Lilypond (<http://lilypond.org/>) is high quality engraving or rendering tool. Music21 (<http://web.mit.edu/music21/>) is a programmers' toolset for computer-aided musicology, provided as a library for the popular Python programming language. Music21 interfaces well with both MuseScore and Lilypond.

The primary free/open source OMR software which we will use is *Audiveris* (<https://audiveris.kenai.com/>), which is an application written in the *Java* language, with a user interface as well as API. We will also investigate use of *Gamera* (<http://gamera.informatik.hsnr.de>), which is a general-purpose framework for document analysis, including some OMR specific functions.

All this software mentioned thus far advertises compatibility with the *MusicXML* standard, however due to the emerging nature of this standard, and differing levels of support, we anticipate having to automate some post-processing to translate scores between this software. Another standard in musicology is the *HumDrum*. We intend to investigate extending our corpus by translating such scores into *MusicXML*, using the *HumDrum* toolkit.

To take best advantage of the above systems, the software we develop are written in the *Java* and *Python* languages. Where possible we use the *Jython* implementation of *Python*, which runs within the *Java Virtual Machine*, and therefore has excellent interoperability with the *Java* language itself.

The core issue for multi-OMR is how to combine the outputs for multiple scans of scores, which may be from different editions, which might contain differing errors or other adjustments. From one perspective, ground truth is an unachievable ideal; scores may be ambiguous, and even expert opinion may differ on the “correct” interpretation. Nonetheless, consensus on correct versus incorrect interpretations of scores is generally achievable, and real controversy is rare.

Due to the open source nature of *Audiveris*, we have access to the confidence level associated with every symbol on a score, as output from its artificial neural network (ANN) evaluator. Therefore once we have aligned the output of two or more scores (building upon the work of Knopke & Bird (2007)), we may employ weighted algorithms, which resolve any points of variance.

5. DEVELOPMENT

To support our initial development phase, we will compile two datasets. The first corpus will consist of Mozart string quartet scores, taken from a number of sources; IMSLP, NMA-Online, MuseData, Kern Scores, and Mutopia. This should provide us with a challenging range of inputs for each quartet, which will be fed through the OMR tools mentioned above.

In addition, we will generate a set of scores from existing digitised MusicXML, using a range of scorer renderers (e.g. Lilypond, Musescore). This will allow us to carry out initial tests, and establish base lines by using OMR to complete “round trips” back to MusicXML. Although the generated scores will in some sense be ground truth data, we anticipate that OMR accuracy will not be uniform across renderings produced by different software. In addition, we will explore manipulating the scores to add “salt and pepper” noise and other common sources of error, to evaluate the confidence ratings of the output.

On carrying out OMR on these scores, we will establish a baseline measure of confidence in the results, to use in aligning and merging the outputs of the different scores and scanners. Multiple measures will be taken, comparing different aspects of the score (pitches, durations, dynamics), to assess relative strengths and weaknesses. We will combine these measures with the note-level confidence measures given by OMR software, where available.

6. CONCLUSION

We have described an ongoing Humanities Big Data project, which aims to improve the results of Optical Music Recognition by combining the results both of multiple sources and multiple OMR scanners. The paper presented our core research questions, the technology available, and broadly described the development process, which we are now undertaking. With this approach we will establish evaluative methods to compare approaches to Multi-OMR, allowing us to explore methods of score alignment and weighting of results to combine OMR results, to find improved accuracy through combined. Our hope is that the result is more robust large scale processing of musical scores, taking advantage of multiple recognition engines and multiple sources of the musical scores now available.

7. ACKNOWLEDGEMENT

This work is conducted with funding support from the Arts and Humanities Research Council (AHRC; AH/L009870/1)

8. REFERENCES

- Bellini, P., Bruno, I., and Nesi, P. (2007). Assessing optical music recognition tools. *Computer Music Journal*, 31(1), 68-93.
- Byrd, D., and Schindele, M. (2006) Prospects for improving optical music recognition with multiple recognizers. In Proceedings of the International Conference on Music Information Retrieval, Victoria, Canada, pp. 41–6
- Jones, G., Ong, B., Bruno, I., and Ng, K. (2008). Optical Music Imaging: Music Document Digitisation, Recognition, Evaluation, and Restoration. In K. Ng, & P. Nesi (Eds.) *Interactive Multimedia Music Technologies* (pp. 50-79). Hershey, PA: Information Science Reference. doi:10.4018/978-1-59904-150-6.ch004.
- Knopke, I., and Byrd, D. (2007) Towards mucisdiff: a foundation for improved optical music recognition using multiple recognizers. In Proceedings of the International Conference on Music Information Retrieval, Vienna, Austria, pp. 122–126).
- Ng, K. (2004). Optical music analysis for printed music score and handwritten music manuscript. *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, 108-127. Hershey, PA: IGI Global. doi:10.4018/978-1-59140-298-5.
- Viro, V. (2011). Peachnote: Music score search and analysis platform. In Proceedings of the 12th International Society for Music Information Retrieval Conference, (pp. 359-362). Miami (Florida), USA.