

Human-Computer Interaction for Optical Music Recognition tasks

Jorge Calvo-Zaragoza

*Dept. of Software and Computing Systems
University of Alicante, Spain
jcalvo@dlsi.ua.es*

Jose Oncina

*Dept. of Software and Computing Systems
University of Alicante, Spain
oncina@dlsi.ua.es*

Abstract

The need to digitise music scores has led to the development of Optical Music Recognition (OMR) tools. Unfortunately, the performance of these systems is still far from providing acceptable results. This situation forces the user to be involved in the process due to the need of correcting the mistakes made during recognition. However, this correction is performed over the output of the system, so these interventions are not exploited to improve the performance of the recognition. This work sets the scenario in which human and machine interact to accurately complete the OMR task with the least possible effort for the user.

1. Introduction

Music is one of the main components of cultural heritage. Over the centuries, musical scores have been stored and preserved in cathedrals, libraries and museums to ensure its maintenance. However, this measure has deprived the access to these scores. Digitizing music scores allows greater dissemination and integrity of this culture, so since decades many effort has been devoted to the development of tools for this purpose.

Nowadays, edition tools that allow actions based on 'mouse and click' to locate musical symbols in empty scores are available. Although this mechanism can be very accurate, their use is very time consuming. Moreover, digital instruments (such as MIDI keyboard) from which the musical information can be directly transferred to the PC while playing the score can also be found. However, this mechanism can not be completely accurate and capture all the nuances of the score. Furthermore, this method requires the user to be able to play the piece perfectly, which is not a trivial matter.

The emergence of Optical Music Recognition (OMR) [1] systems represented the more comfortable alternative to the user. These systems take a scanned

image of a score and try to extract its musical information in order to export it to a digital format (such as MIDI, MusicXML or PDF). OMR raised in 1966 [7], related to fields like Image Processing, Document Analysis and Pattern Recognition. A good example of a common framework to perform OMR processes can be found in [8].

Unfortunately, despite several research efforts, these systems are far from achieving good accuracy rates, especially for handwritten scores [9]. The scores are analysed by OMR tools and its output has to be corrected using a conventional score editor. Hence, the user has to be inevitably involved in the transcription process. However, this involvement is unexploited since it could be used for improving the performance of the OMR system itself. An scenario in which the user and the machine interact to achieve, with the last effort for the user, the perfect transcription of a score can be approached.

Previous studies have shown how to take advantage of interactive frameworks for pattern recognition tasks [11, 12]. Following this line, we present in this paper an overview about human-computer interaction for OMR tasks.

The remainder of this work is as follows: Section 2 presents the field of study, its foundations and its main scenarios to exploit. Section 3 enumerates some of the first areas to explore in the near future. Finally, Section 4 concludes the paper.

2. Human-Computer Interaction for Optical Music Recognition

As occurs in other fields of pattern recognition, the user is the most reliable source to validate an OMR process. The productivity that can be taken from human assistance in OMR tasks is an issue that is yet to be explored. Until now, research has focused on improving the accuracy and speed of the algorithms involved in the process. It is needed a deep analysis about the main weaknesses of these algorithms and how human

participation can improve its performance or be used to produce new algorithms. In this case, there are two critical processes pertaining to OMR system performance: segmentation and classification.

The segmentation stage is aimed to detect and isolate the existing musical symbols in the score. This process is quite complex and requires a comprehensive procedure. It usually involves the following steps:

1. Preprocessing: correction of rotation, binarisation, scaling.
2. Staff lines detection and removal
3. Symbol isolation

The preprocessing stage is focused on providing robustness to the system. If posterior stages always have as input an image with the staff lines aligned with respect to the horizontal axis, with equal relative sizes and where the only possible values for a pixel are background or foreground, the systems tend to generalise more easily. Each of these steps can be addressed in different ways and in the literature each author chooses those techniques that are considered more appropriate.

The staff lines detection and removal is one of the most critical aspect of the process since both the detection and the classification of musical symbol relies on its accuracy. Much research has been conducted concerning this step (a good comparative study can be found in [2]). Although this stage can be approached in many ways, it finally becomes a trade-off between keeping information and reducing noise. Aggressive approaches greatly reduce the noise but can eliminate relevant information. Moreover, less harmful processes end up producing a high amount of noisy areas.

Finally, symbol isolation is performed by searching the remaining meaningful objects in the score. The main problem is that some of the musical symbols are broken by the earlier stages (especially because of staff lines detection and removal).

Once single pieces of the score have been isolated, an hypothesis about the type of each one is emitted in the classification stage. The high variability of handwritten music symbols (see Fig. 1) is the main difficulty to overcome. In addition to this, all errors committed in previous stages (incorrect binarisation, partially extracted symbols, remains of staff lines, etc.) are carried, which complicates even more this recognition.

Based on what has been explained in these lines, there are several scenarios in which the user could attend the machine. Depending on how it is done and the objective pursued, we can divide this assistance into four categories: error detection, online learning, active learning and supervision. It should be emphasized that



Figure 1. Four handwritten eighth notes: example of variability in music notation.

the purpose of these scenarios is to reduce user effort. Since human participation is mandatory -OMR systems are not sufficiently accurate and a score with errors is not acceptable- the real goal is to make the effort required to get a perfect transcription be less than it would be if conventional OMR systems were applied and the output posteriorly edited.

Next subsections describe these four scenarios. It should be noted that these scenarios are not mutually exclusive but they can be combined depending on the suitability of each one at each particular time.

2.1. Error Detection

In the error detection scenario, the user simply marks the places where the system has made a mistake. This information can help to improve both the segmentation and the classification stages.

For segmentation assistance, the user can mark a place where there is a symbol and the system has not found anything; or just the opposite, the user can indicate an area in which the machine creates a symbol and there is actually nothing. In this way, the system can know better the current sheet and extract more reliable features of the staff.

Regarding the classification, marking a symbol which has been misclassified is a way of getting useful domain knowledge. These errors can be propagated to all those symbols which have been classified in a similar way and modify the model learned to be adapted to this change. Moreover, if a definite order is selected in the corrections (for example, left-to-right), each correction implicitly validates previous hypothesis, which could be very helpful for improving the posterior guesses.

2.2. Online Learning

Online learning is a machine learning model where the true label of a certain hypothesis is discovered and used for improving the performance of the learning algorithm.

In an online learning scenario for OMR, the user does not only mark where there is a classification error

but he specifies the correct hypothesis. In this case, the learned classification model gets even more useful information, being able to modify the model to suit the corrections given by the user. As it occurred in the previous scenario, each correction can also validate the previous hypotheses.

This scenario implies a greater support for the system, but it also requires a more active user involvement.

2.3. Active Learning

Active learning refers to a machine learning approach in which the algorithm can query an oracle to know the true label of a certain sample. Further details about active learning can be found in [10].

At any given time, the OMR process may require the user expert assistance. If this information is correctly analysed, it could be beneficial to both the segmentation and classification stages.

To have a more accurate segmentation process, the machine could ask the user to label some areas of the image. For instance, he could indicate if a particular piece of the image is empty of musical symbols so that the algorithm has a seed to learn about interesting (in any way) areas. This may serve to have a greater knowledge of the score, which could be exploited to understand the features of the staff (beneficial for detecting staff lines) or the sheet (beneficial for binarisation).

Moreover, once the symbols have been isolated, the machine could ask the user to label some of the detected symbols. Thus, the algorithm obtains domain information and it might be able to make future hypotheses more accurately.

2.4. Supervision

In this scenario, the user would act as a reviewer of the involved procedures. After each step, the system shows the output and the user must validate or reject the result. Therefore, the process only progresses when every step has been accepted by the user. The main advantage of this framework is that mistakes are not carried through the different steps, so the final result will be free of errors.

This validation allows avoiding general selection of magic numbers in the segmentation stage. After each rejection, the involved procedures could tune some of its parameters to fit the current score.

In the classification stage, each hypothesis about a symbol could be presented to the user. User verdict not only helps to finally achieve a transcription free of errors, but the machine could learn from each rejection, as it happens in the error detection scenario. Furthermore,

each validation provides a completely reliable prefix, which can be used to improve the accuracy of the next hypotheses. This scenario is especially interesting because an action requires the minimum user effort. Nevertheless, the process could be extended for too long.

3. Future developments

As a starting point, the first developments must entail the basics of the interaction between the human and the system. Specifically, we think that there are three aspects that must be considered at first:

1. Development of an optimised OMR system for reducing human interactions. It has been demonstrated that in sequential pattern recognition tasks, giving the most likely output is not optimal for reducing human corrections [6]. This condition has not been taken into account in any current OMR system yet.
2. Development of a system for online recognition of handwritten musical notation. When the user has to correct a misclassification of a symbol, he can search the correct label in a list of musical symbols. However, it is more natural and comfortable to draw the correct symbol over the score itself. To this end, it is necessary to develop a system that recognise musical symbols from the strokes drawn by an user. Although some works has already been done about this issue [5, 3, 4], it is still unexplored how offline and online classification can be more profitably combined.
3. Analysis of the exploitation of user assistance in the segmentation stage. The process of segmentation is a key stage to achieve a good transcription. So far it has not been analysed how this process can be improved using the assistance of the user. Several procedures are involved in this stage (see Section 2), so it would be appropriate to locate the most interesting aspects to focus the human efforts on what is most profitable. Since the staff lines detection and removal step is one of the keys in the performance of an OMR system, developing a new algorithm that takes advantage of user feedback in any of the presented scenarios is advisable.

It should be noted that each of these items represents a new and independent line of research, given the magnitude of the involved field. Once these ways have been explored and exploited, it would be interesting to see how they can be combined in the future to produce an efficient human-computer OMR system.

4. Conclusions

This work aims to establish the starting point for research in Human-Computer Interaction for Optical Music Recognition tasks. First, it has been explained the need to develop such kind of systems. Due to the inaccuracy of state-of-art OMR systems, the user has to correct the mistakes so it is worth using this unavoidable effort to improve the score recognition process.

Focused on the process of OMR, four scenarios were presented according to the nature of the user intervention: error detection, where the user only marks where an error has been made; online learning, where the user corrects the errors made by the system; active learning, where the user is asked to provide the system with the label of a specific sample; and supervision, where the user sequentially accepts or rejects the steps involved in the process.

Regarding the future research, it have been listed the three open lines that should be explored at first: development of an OMR system optimized to reduce human interactions, development of a system for online recognition of handwritten music notation and analysis of the exploitation of user assistance in the segmentation stage.

In the future it is intended to achieve a system that optimises the user effort to obtain the perfect transcription of a music score.

References

- [1] D. Bainbridge and T. Bell. The Challenge of Optical Music Recognition. *Language Resources and Evaluation*, 35:95–121, 2001.
- [2] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga. A comparative study of staff removal algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(5):753–766, 2008.
- [3] S. E. George. Online pen-based recognition of music notation with artificial neural networks. *Comput. Music J.*, 27(2):70–79, June 2003.
- [4] K. C. Lee, S. Phon-Amnuaisuk, and C.-Y. Ting. Handwritten music notation recognition using hmm – a non-gestural approach. In *Information Retrieval Knowledge Management, (CAMP), 2010 International Conference on*, pages 255–259, 2010.
- [5] H. Miyao and M. Maruyama. An online handwritten music score recognition system. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 461–464, 2004.
- [6] J. Oncina. Optimum algorithm to minimize human interactions in sequential computer assisted pattern recognition. *Pattern Recognition Letters*, 30(5):558–563, 2009.
- [7] D. Pruslin. *Automatic recognition of sheet music*. Sc.d. dissertation, Massachusetts Institute of Technology, 1966.
- [8] A. Rebelo, G. Capela, and J. S. Cardoso. Optical recognition of music symbols - a comparative study. *IJDAR*, 13(1):19–31, 2010.
- [9] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. Marcal, C. Guedes, and J. Cardoso. Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, pages 1–18, 2012.
- [10] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [11] A. H. Toselli, E. Vidal, and F. Casacuberta. *Multimodal Interactive Pattern Recognition and Applications*. Springer, 2011.
- [12] E. Vidal, L. Rodriguez, F. Casacuberta, and I. Garcia-Varea. Computer assisted pattern recognition. In *Proceedings of the 4th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, June 2007.