

## DEFACING MUSIC SCORES FOR IMPROVED RECOGNITION

Scott Sheridan <sup>1</sup>, Susan E. George <sup>1</sup>

<sup>1</sup>School of Computer and Information Science  
University of South Australia

**ABSTRACT:** The area of Optical Music Recognition (OMR) has long been plagued by an inability to provide a definitive method for locating and identifying musical objects superimposed on musical stave lines. The first step in the process of recognising musical symbols in OMR has previously been to either remove the stave lines, or ignore them. Removing stave lines leads to many problems of fragmented and deformed musical symbols, or in the case of ignoring them, a lowered chance of recognition. Most OMR systems attempt to correct these deficiencies later on in the process through many varied approaches including bounding box analysis, k-nearest-neighbour (k-NN) and neural network (ANN) classification schemes. All of these have a level of success, but none have provided nearly the desired level of accuracy.

This paper aims to show that this removal of the stave lines before symbol recognition is not the only first step and may not be the best. Instead of removing stave lines, more should be added! This process is called 'defacing' since it adds stave lines to the score at a 1/2 stave line width, and actually overwrites the score - apparently complicating the recognition procedure. However, the addition of signal to the image means that subsequent symbol recognition is 'normalised' and a musical symbol will look the same whether it was above, below or on a stave line. As a result of this, a classification system trained with double stave lines should provide a higher level of accuracy than the traditional approaches of removing/ignoring the stave lines.

### INTRODUCTION

Automated Music Recognition is an area of computer science devoted to translating musical scores into meta-data that can be read, understood and easily modified by a computer. This technology can be used for easy editing of a score for a composer, translating a music score from a piano piece to a full orchestra piece at the press of a button, or something as simple as playing a piece of music from the score without having any musical knowledge, or an audio track of the piece. It can also be used for data mining; for example a composer may want to see whether a particular melody of her devising has been used before in another piece of music. Another useful task the meta-data can be used for is to be able to present the score in any of the different styles of music that there are and have been throughout history. These tasks are just a few simple examples of the usefulness of having a meta-data representation of a piece of music, and obviously the ability to automatically generate this data from a musical score is invaluable.

The main tasks of an OMR system when attempting to interpret an image of a musical score are as follows:

- (i) Identify the stave lines,
- (ii) Locate the musical objects on the score,
- (iii) Identify the symbols that the objects represent, and
- (iv) Understand what the position of the symbols relative to the stave lines and each other means.

This paper will briefly describe stages (i) to (iv), expanding how step (ii) and (iii) can be assisted by the novel method of 'image defacing'. Image defacing can be understood as 'semantically normalising' the musical symbols within an image, to facilitate 'syntactic recognition' of the object. Without the added stave lines the semantics of a musical object intrudes upon both locating the object (step (ii)) and recognising the object (step (iii)). For example, a note symbol placed *between* stave

lines has a different musical pitch, and hence meaning, to the same note symbol placed *on* a line, In trying to find and recognise the basic shape of a note symbol, whether it occurs on a line or between lines is irrelevant. Without 'image defacing' the placing of the musical symbol on the staff (ultimately so important for meaning) intrudes upon the locating and recognition process. Approaches have previously either removed staff lines (often fragmenting symbols in the process) or worked around them (having problems identifying the foreground symbol from the background line). By adding staff lines, and apparently defacing the image, we propose that symbol recognition can actually be enhanced. Every note symbol will look the same, regardless of its final pitch and semantic meaning, making the recognition task easier. The later processing that builds up the musical meaning (eg step (iv)) will receive more accurate data about any note symbols in the image, and can better interpret the recognised symbols with respect to the original lines to capture the musical meaning. Thus, in this approach to OMR, the defacing of the music score is applied before the location of the musical objects, but its usefulness is most obvious and relevant during the identification stage.

#### IDENTIFYING STAVE LINES

Staff line identification is usually the first step in an OMR system, as the data that they provide is essential for understanding the pitch of notes. Identifying staff lines first provides the additional benefit of providing a measure of the width between staff lines, which is useful for determining the scale of the musical score. For example, note heads are almost always 1 staff line widths high, so finding this height before trying to identify notes is very important.

The most common method for finding staff lines (and the method that is used in this project) is horizontal projection.

Horizontal projection takes a sheet of music, and counts the number of black pixels along each row of the image. This method shows 5 distinct peaks in the number of black pixels for each staff, identifying the positions of each staff line of that staff. However, to as one of the pioneers of the OMR field noted, the 5 staff lines are not exactly parallel, horizontal, equidistant, of constant thickness, or even straight [Pre70] To help cope with this, scores are broken into smaller portions and the staff lines from each portion's beginning and end are joined after their identification.

The importance of staff line identification is extended in this project, as it is used for the placement of extra staff lines. The extra staff lines are placed at  $\frac{1}{2}$  staff width horizontally on the score, extending above and below the original 5 staves for the height of the score for use during the identification stage of the OMR process.

#### LOCATING MUSICAL OBJECTS

Musical object location has the conflicting goals of locating musical objects without regard for staff lines, and relating the location of the object to the staff lines for the pitch of the object.

Musical objects are 'superimposed' upon staff lines, meaning they look the same regardless of where they are on a score. However the data that OMR systems must deal with is 1 dimensional, so identifying the superimposed musical objects while ignoring the background of staff lines, and still remembering the staff lines' location relative to that object is the problem that must be dealt with in this step.

Historically, and still the most prevalent, first step in locating musical objects has been to remove the staff lines from the image, thus isolating the remaining shapes for identification. ([BB92], [BC97], [BB03]). However this method has many drawbacks to counter the advantages. Even with the best methods developed so far, the staff line removal results in fragmented musical objects, where areas of musical objects that coincide with the staff lines are removed by the staff removal algorithm. For example, Figure 1 illustrates a musical symbol (bass clef) where staff lines have been removed and the symbol fragmented



Figure 1, Original Base clef, Base clef with stave lines totally removed and Base clef with stave lines removed by Clarke et al's algorithm. [CBT88]

The removal of the stave lines is also useful for presenting the musical object to a recognition agent without the stave lines to interfere with the recognition of the object. This is because the recognition software is trained on the pure symbol and has to treat any remnants of the stave lines, or fragmentation resulting from the stave line removal as noise, which has led to a tremendous effort in improving stave line removal algorithms. The project that this paper is based on contends that the removal of the stave lines is the wrong first step. Defacing the score before locating the musical objects

One of the reasons that the stave lines were removed in the dawn of OMR was to help the primitive software and hardware locate the musical objects using bounding boxes, which required white space to determine where an object started and ended. This reason is no longer an issue, as more modern methods, even simple ones such as horizontal and vertical projection (the technique of counting how many black pixels in a row or column), can locate a musical object without having to remove the stave lines.

This brings into question the whole process of removing the stave lines. One of the aims of the project is to prove that not only is removing the stave lines unnecessary, but that even *adding* stave lines doesn't decrease accuracy and if used correctly can even improve accuracy during the identification stage.

#### IDENTIFYING MUSICAL OBJECTS

Identifying the musical objects is the self explanatory step of discerning what the located musical objects actually are. It is here where the most variety comes into OMR. There have been many different methods proposed for this step, such as bounding box analysis, horizontal and vertical projection, template matching, contour tracing, k-Nearest-Neighbour and Artificial Neural Networks. The identification of musical objects is made difficult by the sheer number of ways that the note heads, stems, rests, holds, flags and all of the innumerable other musical symbols can be combined into a single musical object. It is also made more difficult by the differing styles of notation, and differing rules governing the sizes and appearance of the musical symbols in a diverse selection of music scores. As a simple example, quarter notes have 3 commonly used flag appearances, depending on the style that was used when typesetting the score. These difficulties are only compounded by the fact that these symbols can be joined by beams, and by bad typesetting.

Let us not forget that this all takes place on a 2 dimensional plane (defined by time horizontally and pitch vertically), making this task much different from Optical Character Recognition, which only needs to consider the time progression, and need not worry about superimposed symbols, nor the effect a symbol has on its neighbours.

This stage has been dominated by techniques to minimise or repair the damage that was inflicted upon the score during the location stage by the removal of the stave lines and then present the musical objects as close as possible to the object would be if it had never been superimposed on the stave lines ([BB92], [BC97], [BB03]). This is the reasoning that is being challenged. The musical objects *have* been superimposed upon a stave, so if this is taken into account when training a recognition agent, it no longer has to deal with the extra noise that is created during removal of the stave line.

## DEFACING THE MUSIC SCORE TO IMPROVE RECOGNITION

The process of defacing the music score involves placing horizontal lines that are exactly halfway between the existing staff lines and extend to the top and bottom of the score at  $\frac{1}{2}$  staff line width for the height of the score. The technique eliminates the need to remove the staff lines, and may improve the recognition rates in a system that has been trained to recognise the objects on a defaced score.

Defacing the music score is a perhaps anti-intuitive step in an OMR system, however the potential benefits can be demonstrated by showing the 6 possible appearances a whole note could take on a score compared to the appearances of these same notes once the score has been defaced.

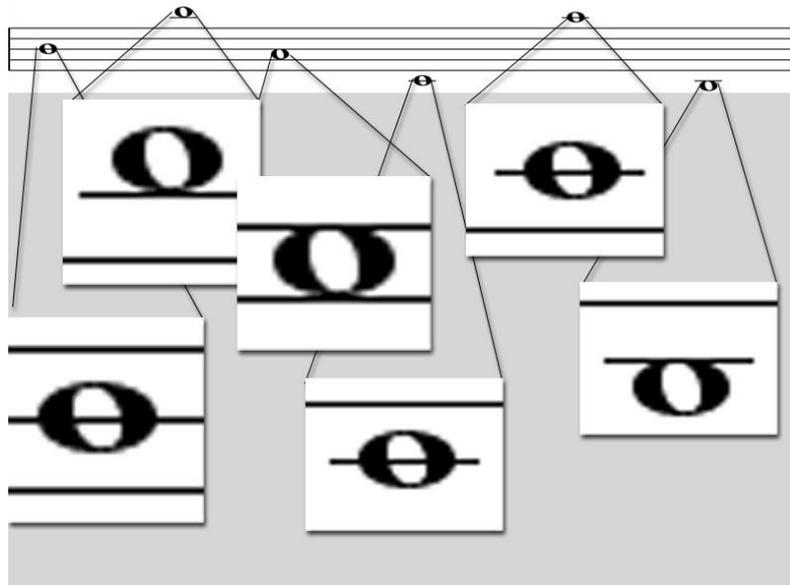


Figure 2: The 6 appearances of a whole note

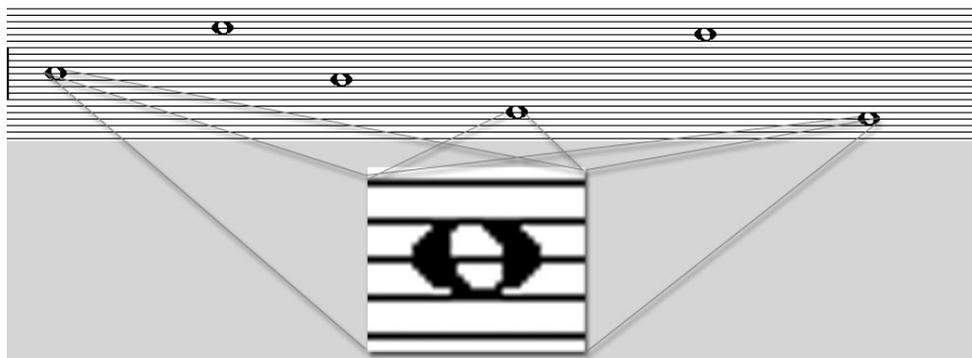


Figure 3: The same 6 notes when the score has been defaced

As can be seen in the above figures, the whole note looks different in the original score, but looks exactly the same, no matter what its position in the defaced score. The need for the staves as a reference for pitch was eliminated when they were identified, as it is easy to remember where the original staves were without the need for them to be visually obvious.

The training of a recognition agent is now simplified. The recognition agent (such as k-NN or ANN) needs only to be trained on the defaced example of a musical object as it is the only way that the object will appear on the score (barring noise, which both of these methods are designed to handle). The alternative, to train the recognition agent on the note without any staff lines, makes the recognition more difficult because the agent must treat the staff lines as noise. This last point is why staff line removal remains prevalent, however with the defacing technique the staff line removal becomes an unnecessary burden.

Potential problems

Defacing the score has some undesirable attributes in certain typesets, where it can partially obscure musical objects such as beams, slurs and hold dots. The answer to this problem is to keep a copy of the score before it was defaced for analysis. All of these objects are easily picked up if they have been missed on the original score. The identification of beams and slurs remains a problem that defacing the score cannot improve, but any solution to a score that has not been defaced can be easily applied to an OMR system that uses defacing by using the original copy for these symbols.

OBTAIN AN UNDERSTANDING OF THE MEANING OF THE MUSICAL OBJECTS

Understanding the meaning of a musical score is where the main benefits from OMR are located. Without understanding of the score, an OMR system becomes just an overly complex photocopier. To understand a score, the system must understand the relationships between separate musical objects and their effect on the objects that follow. This interaction between objects is quite complex, and sometimes the same semantic meaning can be arrived at by different ways. For example, a half note followed by a hold dot means exactly the same thing as a half note tied to a quarter-note. Also the same symbols in different places in relation to an object have differing meanings. Apart from the obvious example of pitch being determined by the note heads' location in regards to the stave lines, a dot above a note has a very different meaning to a dot that follows a note.

The implementation of this step varies as widely as the object recognition, and this step has recently become used more and more as a tool for correcting and/or re-evaluating identified musical objects as a result of the musical understanding gained through this step. This step is not affected by the defacing of the score, either positively or negatively and it remains a valuable way to evaluate object's identifications in all OMR systems.

RESULTS

In order to validate that this method has merit, two identical multi-layer perceptrons were trained. One was trained with a data set that had not been normalised, and the other was trained with the normalised version of the same data set.

Table 1. Some examples of training data and their desired output.

Input Data (normalised vs. original)		Desired Output (whole half quarter)		
		(1	0	0)
		(0	1	0)
		(0	0	1)
		(0	0	0)
		(0	0	0)
		(0	0	0)

After training the networks on their respective data sets, they were tested on some sample scores. Potential note heads that were presented to the networks were selected by concentration of black pixels within a specified bounding box. This knowledge-free technique was used to simplify the search algorithm and eliminate the need for music-specific searches which are more vulnerable to noise and could miss valid notes before they can be tested on the networks.

Both networks performed well on the simpler music scores, but when more complex scores were tested, the normalised network outperformed the network trained on non-normalised data by a

substantial margin. The main cause of errors in both cases was not missing note heads, but rather mistaking areas of the score as note heads. By normalising the score, the amount of mistakenly identified note heads was reduced.

Table 2: The results from the two networks on the sample music scores.\*

Sheet #	True Whole Notes Identified	True Half Notes Identified	True Quarter Notes Identified	False Whole Notes	False Half Notes	False Quarter Notes
# 1	6/6	0/0	0/0	0	0	0
normalised	6/6	0/0	0/0	0	0	0
# 2	0/0	8/8	0/0	0	0	0
normalised	0/0	8/8	0/0	0	0	0
# 3	0/0	0/0	8/8	0	0	0
normalised	0/0	0/0	8/8	0	0	0
# 4	1/1	4/4	8/8	0	0	1
normalised	1/1	4/4	8/8	0	0	0
# 5	0/0	1/1	12/12	0	0	2
normalised	0/0	1/1	12/12	0	0	0
# 6	0/0	8/8	2/2	0	1	0
normalised	0/0	8/8	2/2	0	1	0
# 7	2/2	0/0	0/0	0	0	0
normalised	2/2	0/0	0/0	0	0	0
# 8	3/3	3/3	8/8	2	3	3
normalised	3/3	3/3	8/8	0	0	1
# 9	2/2	3/3	12/13	3	4	8
normalised	2/2	3/3	13/13	1	1	0
TOTAL	14/14	29/29	50/51	5	8	14
normalised	14/14	29/29	51/51	1	2	1

\*all 18 music scores can be found at: [www.macs.unisa.edu.au/Sceptre/thesis/scores](http://www.macs.unisa.edu.au/Sceptre/thesis/scores)

## CONCLUSION

OMR is an image processing problem that requires locating and recognising musical symbols in the image that are superimposed upon stave-lines. The stave-lines are integral to the final meaning of the musical symbol, but have been found to intrude upon both the location of musical objects in the image, and the classification of those objects. Existing solutions that attempt to remove stave-lines often fragment and deform the musical symbols in the process and detract from the final recognition achieved. Solutions that attempt to locate and identify symbols in the presence of stave lines often have difficulties in separating the foreground symbol from the background stave – again detracting from final recognition. These difficulties are magnified when the musical symbols are not typeset but rather handwritten. Commercial systems are still *unable* to deal satisfactorily with handwritten music – usually totally failing to even locate written symbols, let alone make a correct interpretation of them in the image.

This paper has investigated the contribution that ‘image defacing’ can make to OMR proposing the ‘normalisation’ of symbols within the image by the addition of stave lines. Improvements in locating and recognising symbols (handwritten or other) that normalisation can achieve will provide the OMR system with better information with which to build up the final interpretation of the music represented within an image.

REFERENCES

- [BB92] Baird, H.S. and Blotstein, D, A critical survey of music image analysis, in *Structured Document Image Analysis*, eds. Baird, H.S., Bunke, H and Yamamoto, K, (1992); 405-434
- [BB03] Bainbridge, D & Bell, T.C. A music notation construction engine for optical music recognition, *Software, Practice and Experience* 33, 173-200, (2003)
- [BC97] Bainbridge D, Carter N. Automatic reading of music notation. *Handbook on Optical Character Recognition and Document Image Analysis*, Bunke S, Wang PSP (eds.). World Scientific: Singapore, (1997); 583–603.
- [CBT88] Clarke, A. T., Brown, B. M. and Thorne, M. P. , Inexpensive optical character recognition of music notation: a new alternative for publishers, *Proceedings of the Computers in Music Research Cont.* , 84-87, 1988
- [Pre70] Prerau, D. S. Computer Pattern Recognition of Standard Engraved Music Notation. *PhD thesis, Massachusetts Institute of Technology*, 1970