A Method for Music Symbols Extraction based on Musical Rules

Ana Rebelo
FEUP and INESC Porto
Portugal
arebelo@inescporto.pt

Filipe Paszkiewicz
FEUP
Portugal
filipe.asp@gmail.com

Carlos Guedes
FEUP and INESC Porto
Portugal
carlosguedes@mac.com

Andre R. S. Marcal
FCUP and CICGE and CMUP
Portugal
andre.marcal@fc.up.pt

Jaime S. Cardoso
FEUP and INESC Porto
Portugal
jaime.cardoso@inescporto.pt

Abstract

Optical Music Recognition (OMR) systems are an important tool for the automatic recognition of digitized music scores. However, handwritten musical scores are especially problematic for an automatic recognition. They have irregularities that go from heterogeneous illumination to variability in symbols shape and complexity inherent to music structure. These issues cause serious difficulties when one wants a robust OMR system facilitating search, retrieval and analysis operations. To transform the paper-based music scores and manuscripts into a machine-readable symbolic format several consistent algorithms are needed. In this paper a method for music symbols extraction in handwritten and printed scores is presented. This technique tries to incorporate musical rules as prior knowledge in the segmentation process in order to overcome the state of the art results.

1 Introduction

Music is an *art* form whose *medium* is sound. The word derives from Greek $\mu\varnothing\upsilon\sigma\iota\kappa\dot{\eta}$ ($\tau\acute{\epsilon}\chi\nu\eta$), "(art) of the Muses"[1].

Music notation, the visual manifestation of the interrelated properties of musical sound – pitch, intensity, time, timbre and pace – is a combination and prolonged effort of hundreds of musicians. They all hoped to express by written symbols the essence of their musical ideas. The final notation is a kind of alphabet, shaped by a general consensus of opinion to serve as a general expressive technique.

Since the beginning of the Eras where scholars found bones with lateral holes pierced (flutes) to the 21th century, where Internet contributed to the increase of the expansion of music with online communities like YouTube and MySpace, music has been an important part of the culture of all societies. Nowadays, an enormous number of handwritten music documents still exist only in the form of original or photocopied manuscripts, without being published in a digital format readable by a computer software. Digitalization is an essential tool to preserve this culture heritage. However, in order for composers and other users to be able to interpret the digital music characters, Optical Music Recognition (OMR) algorithms are required.

After an image preprocessing stage (application of techniques such as enhancement, binarization, noise removal, among others), a typical OMR framework has three principal modules: (1) recognition of musical symbols from a music sheet; (2) reconstruction of the musical information to build a logical description of musical notation; and (3) construction of a musical notation model for its representation as a symbolic description of the musical sheet. It is in the first module that staff lines are detected, and usually removed, and the symbols primitives are extracted to perform the recognition. These symbols are merged to form

---

[1] http://en.wikipedia.org/wiki/Music

musical symbols in the second module of musical notation reconstruction. In order to overcome possible classification errors, context information is introduced here, using graphical and syntactic rules. In the last module of the construction of the final representation, the system creates a graphical music-publishing file, such as MIDI or MusicXML, with the information previously stored.

The focus of this paper is on the first module of an OMR system. A method developed for music symbols extraction based in musical rules is presented. The proposed process aims to surmount the current solutions for the recognition of handwritten music scores. The experimental work was conducted over a large data set of scores, consisting both of real handwritten music scores and synthetic scores to which known deformations were applied.

1.1   Related work

The extraction of the music symbols is an operation that usually follows the staff line detection and removal. The operations to perform this task have been studied by the research community (e.g. [11, 12]). Nonetheless, only some works have been produced in the field of handwritten scores (e.g. [4, 8]).

Generally, the adopted approach decomposes the music sheet hierarchically: first the score is analyzed and split by staffs and then the elementary graphic symbols – note heads, rests, dots, etc – are extracted (e.g. [8]). Common procedures make the primitives segmentation step along with the classification operation [11, 12]. Carter [2] uses a Line Adjacency Graph (LAG) to extract symbols, whereas Reed [9] uses LAG only to detect lines and curves. Accidentals, rests and clefs are detected by a character profile method, which is a function that measures the perpendicular distance of the contour of the object from a reference axis, and note heads are recognized by template matching.

A structural method based on the construction of graphs for each symbol is proposed by Randriamahefa [7]. In [11] a fuzzy model based on a robust symbol detection and template matching is developed. The author proposes a segmentation process with two main steps: individual analysis of musical symbols and fuzzy model. In the first step, the vertical segments are detected by a region growing method and template matching; then beams are detected by a region growing algorithm and a modified Hough Transform; the remaining symbols are extracted again by template matching. From this first step three recognition hypotheses result; the fuzzy model is then used, as a second step, to make a consistent decision.

Other techniques for extracting musical symbols include rule-based systems to represent the musical information, a collection of processing modules that communicate by a common working memory [10] and pixel tracking with template matching [12].

A mathematical morphological approach for the music symbols segmentation is proposed in [5]. The authors apply a skeletonization technique with an edge detection algorithm and a stroke direction operation to segment the music score. In [4] the authors start by using median filters with a vertical structuring element to detect vertical lines. Then a morphological opening using an elliptical structuring element is applied to detect noteheads. The bar lines are detected considering its height and the absence of noteheads in its extremities. Clefs symbols are extracted using Zernike moments and Zoning, which codifies shapes based in statistical distribution of points. In [8] the segmentation of the objects is based on a hierarchical decomposition of a music image. A music sheet is first analyzed and split by staffs. Subsequently, the connected components are identified.

Notwithstanding the several techniques already existent in the literature, the results in the handwritten symbol segmentation are still far from ideal. Algorithms that use music rules in the extraction of the music symbols are an advantage for the OMR systems. This prior knowledge tries to deal with the inconsistencies of the writing music leading to better results.

2   Musical symbols detection

The musical symbols detection is a stage on an OMR system where operations to localize and to isolate musical objects are applied. The proposed algorithms must be robust to several problems imposed by music symbols. The complexity of this phase, caused by printing and digitalization, as well as the paper degradation over time, is directly related, not only with the distortions inherent in staff lines, but also with broken, connected and overlapping symbols (see Figure 1), differences in sizes and shapes (see Figure 2), noise, and zones of high density of symbols (see Figure 3).

(a) Broken minim symbols.   (b) Connected notes in a chord.   (c) A slur that overlapped a beam.
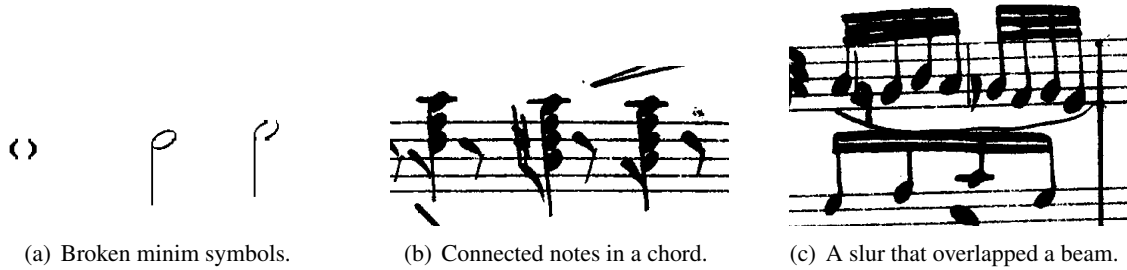
Figure 1 : *Broken, connected and overlapping symbols.*

(a)   (b)   (c)

(d)   (e)   (f)

Figure 2 : *Variability in sizes and shapes between handwritten and printed scores.*

(a) Music scores with noise.   (b) An high density of symbols.
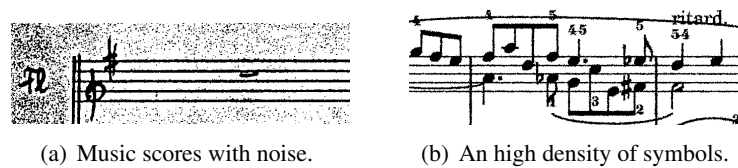
Figure 3 : *Noise and zones of high density of symbols.*

2.1   Segmentation process architecture

The music symbols can be split into four different types: (1) the symbols that are characterized by a vertical segment (stem) and an oval note head: crotchet (e.g. ♩), notes with flags (e.g. ♪) and minim (e.g. ♩); (2) the symbols that link the notes: beams (e.g. ♫); (3) the remaining symbols connected to staff lines: clefs, rests

(e.g. ♪), accidentals (e.g. ♭, ♯, ♮) and time signature (e.g. ₵); (4) the symbols above and under staff lines: ties, slurs (e.g. ⌒) and accents (e.g. >).

Crotchet is usually the symbol that appears with higher frequency in a score. Hence, it is natural to start the segmentation process for stem detection in order to detect, after that, the note heads. Besides, almost all musical rules are dependent on this symbol. For instance, beams only exist connecting crotchet. Similarly dependency can be established for accidentals: they only exist before each note head and at same height. In Table 1 the reader can find the possible graphical and musical rules for the music notation in study.

Symbols	Description
Clefs:	The largest of all symbols, beginning below the staff and extending above it, and they are always placed in the beginning of a staff.
Accidentals:	These symbols exist before each notehead and at same height (placed on the same staff line or space).
Beams:	The connection between eighth notes.
Accents:	They are placed above beams, bellow and above noteheads.
Quarter, 8th's, 16th's, 32th's, …, and Minim:	These symbols are characterized by three components: note head (the round part), stem (the vertical line attached to the note head) and flag (the line that comes off the top or bottom of the note stem and always placed to the right of the stem).
Rests:	The written position of a rest between two bar lines is determined by its location in the meter.
Ties and Slurs:	A tie must be on the same line or the same space; a slur is placed above or below the note head.
Dots:	These symbols are always placed to the right of note heads and in the center of a space; with note heads with flags they are located in the space above and beyond the tail; if the note heads are placed on lines the dots are in the center of the space above; with chords where the lower notehead is on a line its dots must go in the space beneath it.

Table 1 : *Music notation.*

We can also have a set of accidentals symbols, appearing after the clef sign, composing the key signature. Besides all these music symbols, a musical score must have a sign to define the meter of the music. Time signatures establish the number of beats in each bar line and which note value constitutes one beat. This type of sign is sometimes represented by two figures where the numerator occupies the two top spaces and the denominator occupies the two bottom spaces, or simply it is represented by a C used to indicate $\frac{4}{4}$ or a ₵ used to indicate $\frac{2}{2}$. The time signature is placed after the clef symbol and any key signature.

A music sheet is first analyzed and split by staffs, as yielded by the staff lines removal step. This choice aims the disintegration of a complex problem towards various simple problems. The processing continues with the segmentation process proposed in this paper and illustrated in Figure 4. The extraction of all music symbols is based in contextual information and music writing rules. For each staff (set of staff lines) a clef detection process (step I) is applied. If this symbol exists, the processes to detect key (step II) and time signatures (step III) is performed. If a clef symbol does not exist (for instance, the current music sheet can be the second sheet of that music work) the algorithm searches for stored information about the last analyzed music sheet. After step III (Time Signature Detection) a procedure to find stems is made (step IV). If these objects are present in the staff then the algorithm searches for the existence of closed note heads (step V) –
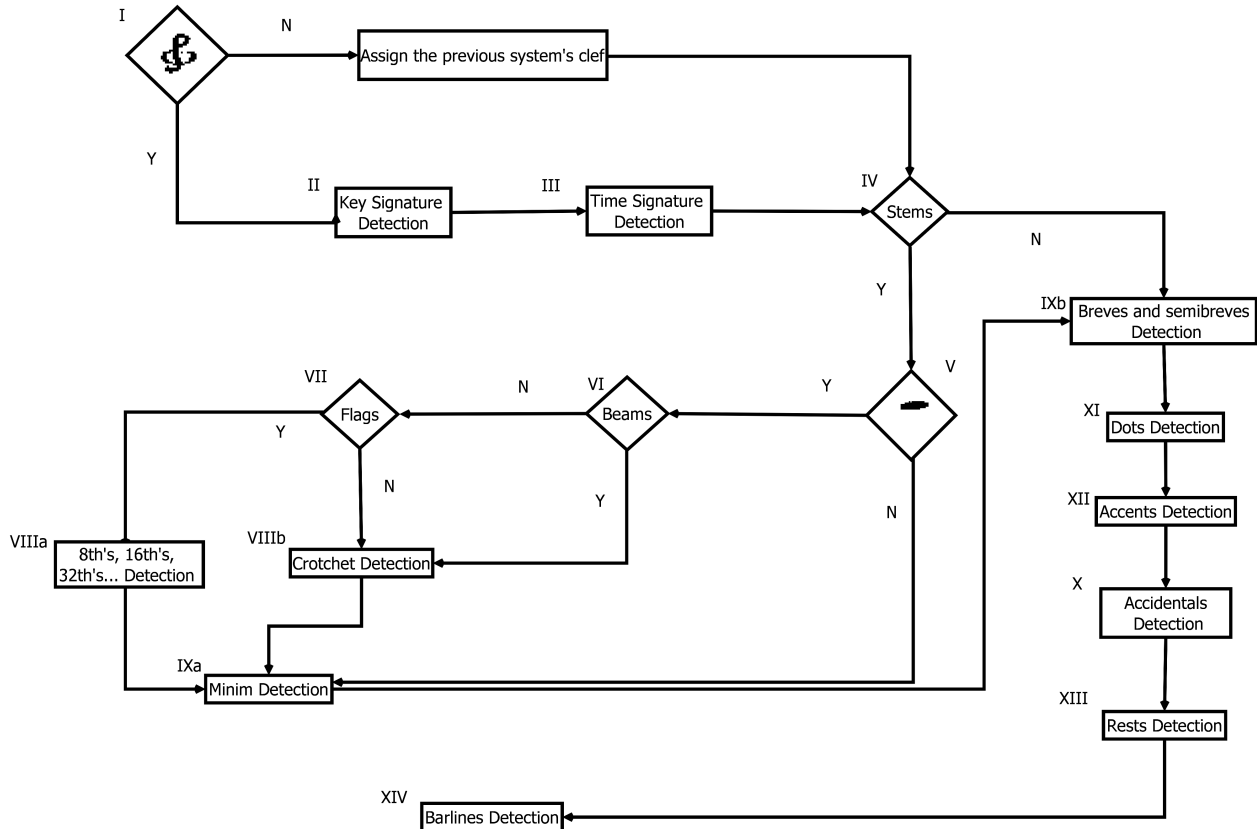
Figure 4 : *Diagram of the musical symbols extraction algorithm.*

crotchet –, otherwise the algorithm goes to breves and semibreves detection (step IXb). For each closed note heads a beams detection (step VI) process is applied. The closed note heads with beams cannot have flags. For the others a flag detection (step VII) step is needed. If the algorithm does not detect any closed note heads then it goes directly to step IXa (Minim Detection). The remaining stages are accidentals detection (step X), dots detection (step XI), accents detection (step XII), rests detection (step XIII) and bar lines detection (step XIV).

The multiple steps of the extraction process use the connected components technique with a possible threshold for the distance between objects. This freedom degree for the neighbors pixels is due to the existence of broken symbols, as already mention in this paper. For clefs and rests detection, matrices with 10 rows and 5 columns are created, containing for each cell the number of black pixels of each symbol. These will be masks that will be used to calculate correlation degrees. This procedure starts with an image from the music score with a certain height and width; while the correlation is below a threshold the width of the image is increased. The stop criterion, if no match symbol is found, is the maximum number of iterations. To detect key and time signatures, the algorithm searches for connected components that are placed a certain number of pixels away from the previous detected symbol. For the key signature, distances between each accidentals are also considered. Another important issue that is taken into account is the maximum number of accidentals allowed per key. For the note heads detection, the connected components chosen were those with a width and height that do not differ by more than 40%. These values were chosen with the aim to preserve the geometric features of the notes heads. If the procedure is detecting closed note heads then the algorithm rejects the objects selected with white pixels inside. For every symbol, musical rules were considered before the application of the connected components technique. For instance, flags are symbols that are always placed to the right of the stem. As opposed to these signs, accidentals exist before each

notehead and at same height. In this manner, the connect components process is only applied in these areas. It is worth noting that the threshold values used were all obtained experimentally.

3   Evaluation metrics and results

The data set adopted to test the proposed architecture for the music symbols extraction consists of both handwritten (see Figure 5(a)) and synthetic scores (see Figure 5(b)). In the latter, 18 printed scores available from [3] written on the standard notation were considered. Several deformations were applied to this database: rotation, curvature and typeset emulation; see [3] for more details. In total, 882 images were generated. The real scores consists on a set of 23 handwritten scores from 6 different composers. These type of scores were binarized using two different procedures: Otsu's and BLIST's methods [6]. Note that the synthetic scores are already in binary format.

(a) Handwritten scores.   (b) Synthetic scores.

Figure 5 : *Examples of real and synthetic scores used in this work.*

Three metrics were considered: the accuracy rate, the average precision and the recall. They are given by

$$accuracy = \frac{\#tp + \#tn}{\#tp + \#fp + \#fn + \#tn}, \quad precision = \frac{\#tp}{\#tp + \#fp}, \quad recall = \frac{\#tp}{\#tp + \#fn}$$

where tp are the true positives, tn are the true negatives, fn are the false negatives and fp are the false positives. A false negative happens when the algorithm fails to detect a music symbol present in the score; and a false positive is when the algorithm falsely identifies a musical symbol which is not one. These percentages are computed using the symbols position reference obtained manually and the symbols position obtained by the segmentation algorithm. The performance of the procedure can be seen in Table 2.

	Binarization Method	Precision	Recall	Accuracy
Handwritten scores	Otsu	88.88%	85.28%	74.40%
	BLIST	94.25%	78.72%	73.11%
Printed scores		98.73%	81.15%	79.97%

Table 2 : *Results of the music symbols extraction algorithm using both printed and handwritten scores.*

The encouraging results obtained for handwritten scores motivate further investigations in our work. The BLIST's method gives the highest precision, whereas Otsu's method gives the lowest error and the highest recall. The performance obtained by the most commonly used commercial OMR system is approximately 90% [1]. However, the estimation of this value and the type of printed music scores used is not clear. With the proposed algorithm the precision was 98.73%.

4   Conclusion

A method for music symbols extraction in handwritten and printed scores was presented. The results achieved can be improved through an error detection in the symbols extraction phase. These errors can be symbols missing, symbols confusion and wrong symbols added. The idea is to check the number of figures between bar lines in order to coincide with the time signature. We also suggest to incorporate degrees of confidence in each extracted or classified symbol. The inclusion of prior knowledge of syntactic and semantic musical rules in the OMR recognition process can also lead to better results in the future.

References

[1] P. Bellini, I. Bruno, and P. Nesi. Assessing optical music recognition tools. *Computer Music Journal*, 31:68–93, March 2007.

[2] N. P. Carter. Automatic recognition of printed music in the context of electronic publishing, 1989. In Dorothea Blostein and Henry S. Baird, *A Critical Survey of Music Image Analysis*, in *Structured Document Image Analysis*, Baird, Bunke, and Yamamoto, pages 405–434, Springer-Verlag, Heidelberg, 1992.

[3] C. Dalitz, M. Droettboom, B. Czerwinski, and I. Fujigana. A comparative study of staff removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:753–766, 2008.

[4] A. Forns, J. Llads, and G. Snchez. Primitive segmentation in old handwritten music scores. In Wenyin Liu and Josep Llads, editors, *Graphics Recognition. Ten Years Review and Future Perspectives*, volume 3926 of *Lecture Notes in Computer Science*, pages 279–290. Springer Berlin / Heidelberg, 2006.

[5] K. Ng, D. Cooper, E. Stefani, R. Boyle, and N. Bailey. Embracing the composer: Optical recognition of handwritten manuscripts. In *Proceedings of the International Computer Music Conference*, pages 500–503, 1999.

[6] T. Pinto, A. Rebelo, G. Giraldi, and J. S. Cardoso. Music score binarization based on domain knowledge. In *Proceedings of The 5th Iberian Conference on Pattern Recognition and Image Analysis*, 2011.

[7] R. Randriamahefa, J. P. Cocquerez, C. Fluhr, F. Pepin, and S. Philipp. Printed music recognition. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, pages 898–901, 1993.

[8] A. Rebelo, G. Capela, and J. S. Cardoso. Optical recognition of music symbols: A comparative study. *International Journal on Document Analysis and Recognition*, 13:19–31, March 2010.

[9] K. T. Reed and J. R. Parker. Automatic computer recognition of printed music. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 3, pages 803–807, Washington, DC, USA, August 1996. IEEE Computer Society.

[10] J. W. Roach and J. E. Tatem. Using domain knowledge in low-level visual processing to interpret handwritten music: an experiment, 1988. In Dorothea Blostein and Henry S. Baird, *A Critical Survey*

of Music Image Analysis, in *Structured Document Image Analysis*, Baird, Bunke, and Yamamoto, pages 405–434, Springer-Verlag, Heidelberg, 1992.

[11] F. Rossant and I. Bloch. Robust and adaptive OMR system including fuzzy modeling, fusion of musical rules, and possible error detection. *EURASIP Journal on Advances in Signal Processing*, 2007:160–160, January 2007.

[12] F. Toyama, K. Shoji, and J. Miyamichi. Symbol recognition of printed piano scores with touching symbols. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 480–483, Washington, DC, USA, 2006. IEEE Computer Society.