OPTICAL MUSIC RECOGNITION BASED ON A FUZZY MODELING OF SYMBOL CLASSES AND MUSIC WRITING RULES

Florence Rossant, Isabelle Bloch

OPTICAL MUSIC RECOGNITION BASED ON A FUZZY MODELING
OF SYMBOL CLASSES AND MUSIC WRITING RULES

Florence ROSSANT, Isabelle BLOCH

ISEP, 21 rue d'Assas, 75006 PARIS, Florence.Rossant@isep.fr
ENST, 46, rue Barrault, 75013 PARIS, Isabelle.Bloch@enst.fr

ABSTRACT

We propose an OMR method based on fuzzy modeling of the information extracted from the scanned score and of musical rules. The aim is to disambiguate the recognition hypotheses output by the individual symbol analysis process. Fuzzy modeling allows to account for imprecision in symbol detection, for typewriting variations, and for flexibility of rules. Tests conducted on a hundred of music sheets result in a global recognition rate of 98.55%, and show good performances compared to SmartScore.

1. INTRODUCTION

Optical Music Recognition aims at reading automatically scanned scores. Although it has been an active research field since the 80's, there are still a lot of unsolved problems. Difficulties result from segmentation problems, printing imprecision, variability of typewriting [1]. The need of structural approaches is now recognized. Research has been conducted in this direction at the symbol level [2] or at a higher level and involves mostly local graphical musical rules [3,4]. Due to the limited work at the higher level, we concentrate on the introduction of local or non-local music writing rules, either graphical or syntactic. Another contribution of this paper is to solve ambiguities by accounting for sources of imprecision and uncertainty. Existing approaches are based on statistical methods [5], or graph-rewriting rules [4]. Here we rely on the fuzzy sets and possibility framework, since it is well adapted to the integration of flexible rules [6] and for dealing with spatial imprecision [7].

We first present an overview of the proposed recognition method (Section 2). The next two sections (3, 4) expose the individual symbol analysis step and the fuzzy modeling. This paper focuses on the fuzzy modeling of symbol classes and graphical rules, that is the main contribution with respect to our previous work. We still refer to [8] for the syntactic part dealing with metric and tonality rules. The global decision algorithm is summarized in Section 5, and we conclude with experimental results in Section 6.

2. SYSTEM OVERVIEW

The inputs of the program are the binary image of a monophonic music sheet scanned at a resolution of 300 dpi and some global information (clef, tonality, metric).

The processing flow (Fig. 1) is divided into three main parts. The individual analysis process performs the segmentation of the objects and the correlation with symbol models stored in a reference base [9]. It provides for each pattern s some recognition hypotheses. Then, the fuzzy modeling part provides for each classification hypothesis a possibility degree of membership to the class. It also introduces a fuzzy representation of the common music writing rules by expressing graphical and syntactic compatibility degrees between the symbols. Finally, the decision process evaluates bar per bar all the hypotheses combinations and chooses the most consistent one.
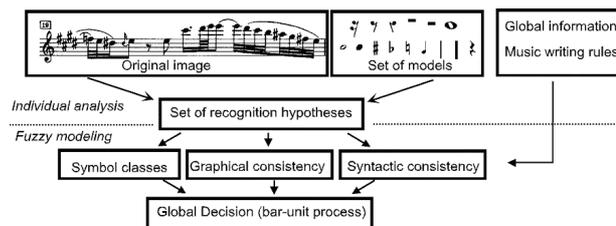

Figure 1: The global processing flow

3. INDIVIDUAL SYMBOL ANALYSIS

Individual symbol analysis is mainly based on pattern matching. We compute in a small search area around each segmented object s the normalized correlation scores $C_s^k(x, y)$ between s and the models M^k of the reference base. Let us denote by $C^k(s)$ the highest value obtained at the (x_k, y_k) coordinates:

$$C^k(s) = C_s^k(x_k, y_k) = \max_{(x, y)} C_s^k(x, y) \qquad (1)$$

These coordinates represent the position of the center of the musical symbol in the hypothesis of the class k.

At most three recognition hypotheses are selected. A recognition hypothesis assigns the pattern s to a class k, if the associated correlation score is greater than the minimum threshold t_m (0.3). We add also the possibility that there is no symbol ('-' in Table 1) when the highest correlation score obtained for class k is less than the decision threshold $t_d(k)$. These thresholds allow to take into account that some classes are more sensitive to typewriting variations (small value for $t_d(k)$) or have a higher probability of false detection (large value).

Fig. 2 shows two bars, (a) and (b), and the recognition hypotheses superimposed on the original images. Table 1 indicates some of the associated correlation scores.
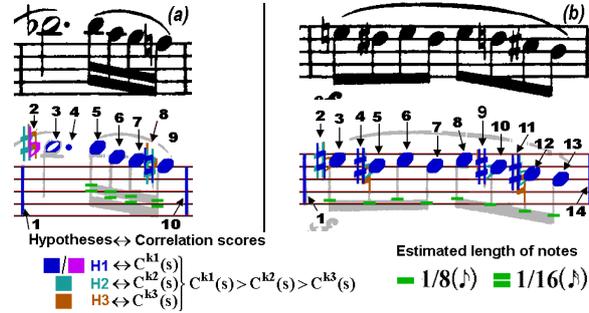

Figure 2: Original images and recognition hypotheses

(a)	2	3	8		(b)	2	9	11
H1	- or ♭ 0.66	⬭ 0.60	♮ 0.58		H1	♮ 0.62	♯ 0.58	♯ 0.59
H2	♯ 0.43		♯ 0.48		H2	♯ 0.52	♭ 0.58	♭ 0.49
H3	♮ 0.43		♭ 0.39		H3	♭ 0.50	♮ 0.55	♮ 0.45

Table 1: Some hypotheses and correlation scores

We can see the limits of the individual analysis: the correlation scores may be very ambiguous, and the highest does not always correspond to the right hypothesis.

4. FUZZY MODEL

Ambiguities result from segmentation defects and variations in typewriting. This section aims at modeling the information extracted from the score during the first analysis and the knowledge about music writing. Fuzzy sets and possibility theory offer a good framework for this aim. It allows us to deal with uncertainty concerning the shape and the location of the symbols, and to express musical rules that have different degrees of flexibility.

4.1. Symbol classes

The correlation scores provide similarities between each analyzed symbol and models. Therefore we define the

degree of possibility $\pi_k(s)$ that s belongs to class k as an increasing function of $C^k(s)$:

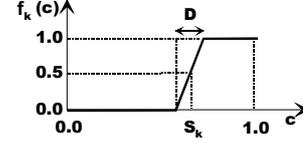$$\pi_k(s) = f_k\left(C^k(s)\right) \qquad (2)$$

Figure 3: Possibility distribution of class k

The shape of the possibility distribution for class k (Fig. 3) is defined by two parameters: D (always 0.3), and S_k that is learnt from the individual analysis results:

$$S_k = \frac{t_d(k) + n(k)m(k)}{n(k)+1} \qquad (3)$$

where $n(k)$ is the number of objects having a correlation score with model M^k larger than the threshold value $t_d(k)$, and $m(k)$ is the average value of the scores. Because of typewriting variations, the reference model M^k may not match closely the objects of class k in the processed score. In this case, $m(k)$ is close to $t_d(k)$, and S_k also close to $t_d(k)$; $m(k)$ takes larger values otherwise, so that the possibility distribution is shifted to the right.

The shape of the distribution π_k does not need to be estimated very precisely [6]. It has experimentally proved to be robust. The most important is that it is not a binary function (there is no crisp threshold) and that it is increasing: the higher is the correlation score, the higher is the degree of possibility.

Table 2 shows the results obtained for the two bars. Compared to Table 1, the possibility degrees present less ambiguity. It should be noticed that the classification rank may change: see for example object 9 in bar (b), for which the possibility of a flat is now eliminated.

(a)	2	3	8		(b)	2	9	11
H1	- or ♭ 0.17	⬭ 0.44	♮ 0.30		H1	♮ 0.43	♯ 0.40	♯ 0.43
H2	♯ 0.00		♯ 0.07		H2	♯ 0.20	♭ 0.00	♭ 0.00
H3	♮ 0.00		♭ 0.00		H3	♭ 0.00	♮ 0.20	♮ 0.00

Table 2: Some hypotheses and possibility degrees

4.2. Graphical consistency

Musical writing rules impose some compatibility of position between two successive objects. Due to possible imprecision in the score printing and in the segmentation, as well as to symbol density variations, these rules cannot be used in a crisp way, and are rather a matter of degree.

4.2.1. Compatibility between accidental and note

An accidental should be placed before a note and at the same height. The possibility degree that the object s_n is an

accidental of class k_n, and that a following nearby object s_m (m>n) is a note of class k_m, is a function of the compatibility degree $C_p(s_n^{k_n}, s_m^{k_m})$ between both symbols:

$$C_p\left(s_n^{k_n}, s_m^{k_m}\right) = \begin{cases} \alpha_l f_l(\Delta l) + \alpha_h f_h(\Delta h) \text{ if } f_l(\Delta l) > 0 \text{ and } f_h(\Delta h) > 0 \\ 0 \qquad\qquad\qquad\qquad\qquad otherwise \end{cases} \quad (4)$$

where f_l, and f_h are two functions defining the admissible values for Δl and Δh, the differences in horizontal and vertical position between s_n and s_m:
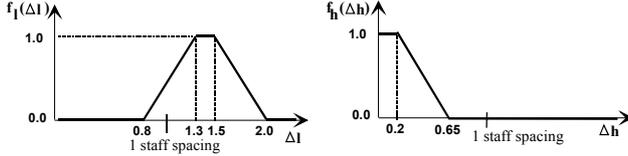

Figure 4: Accidental / note graphical compatibility

This combination is a compromise between two criteria, excluding the cases where at least one is not satisfied at all. The chosen coefficients $\alpha_l=0.2$ and $\alpha_h=0.8$ express their relative importance. Using a degree between 0 and 1 instead of a crisp threshold on each criterion allows us not to discard completely an accidental which is not exactly at the theoretically expected position. For object 2 in bar (a), we obtain a compatibility degree with the next note of 1.0 in the hypothesis of a flat, and of 0.77 in the hypothesis of a sharp. These results rank correctly the flat before sharp and strengthen the right hypothesis H1.

4.2.2. Compatibility between notes and augmentation dot

During the first analysis process, augmentation dots are searched for in an area next to the note head [9]. The typical location is represented in light gray in Fig. 5a. But smaller or larger horizontal distances Δl may occur. That is why the searching area has been extended to the black rectangle. A new compatibility degree has also been defined as a function of Δl (Fig. 5b). It expresses that the typical distance is in the range of 1 to 2 staff spacing and that smaller or larger distances become less and less possible.
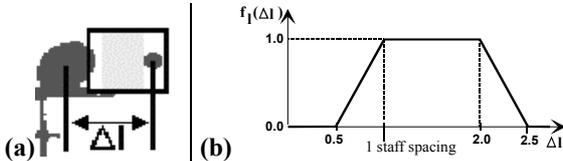

Figure 5: (a) Search area for an augmentation dot, (b) Note / augmentation dot graphical compatibility

Some confusion may occur if a staccato dot above a note head is also in a searching area for an augmentation dot (see Fig. 7b). Therefore we define a compatibility degree between the hypothesis that s_n is an augmentation dot and the hypothesis that a following object s_m is a note, as a function of the horizontal and vertical differences between them (Fig. 6, Eq. (5)):

$$C_p\left(s_n^{k_n}, s_m^{k_m}\right) = Max\left[f_l(\Delta l), f_h(\Delta h)\right] \quad (5)$$

According to (5), the compatibility degree is high as soon as one of both criteria is well satisfied (Fig.7).
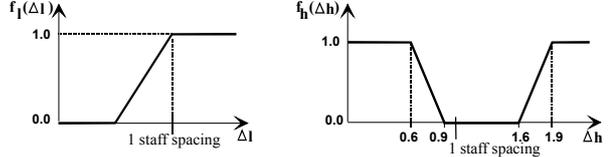

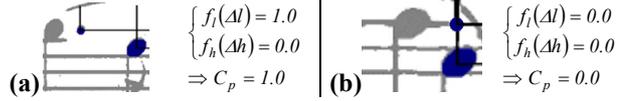**Figure 6: Dot / note graphical compatibility**

Figure 7: Dot / note graphical compatibility (examples)

4.2.3. Compatibility between any two symbols

For any other pair of nearby symbols, we just express that they should not overlap. For that, we estimate the positions of the bounding box of each symbol (Fig. 8a), knowing the typical sizes of the two considered symbol classes. Let us denote by Δl the difference between the right border of s_n and the left border of s_m, m>n. The admissible values for Δl are defined by function f (Fig. 8b). It allows negative values for Δl in order to deal with imprecision on the symbol size and position, and with high symbol density. In the vertical direction, we must consider different configurations. For example, if s_n is above s_m, we compute the difference between the bottom border of s_n and the upper border of s_m. The admissible values for Δh are again defined by f. The global compatibility is expressed as the maximum of both criteria: just one has to be satisfied.

$$C_p\left(s_n^{k_n}, s_m^{k_m}\right) = Max\left[f(\Delta l), f(\Delta h)\right] \quad (6)$$
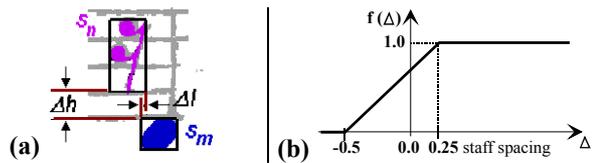

Figure 8: Graphical compatibility for any two symbols

4.2.4. Fusion

We have computed compatibility degrees between all pair of nearby symbols, according to their classes. In case of music sheets with high symbol density, there may be more than one nearby symbol before or after s_n. Equation (7) defines the global graphical compatibility degree for object s_n classified in class k_n:

$$C_p\left(s_n^{k_n}\right) = \left[\min_{j<n} C_p\left(s_n^{k_n}, s_j^{k_j}\right)\right] \cdot \left[\min_{l>n} C_p\left(s_n^{k_n}, s_l^{k_l}\right)\right] \quad (7)$$

It is the product of two terms representing respectively the compatibility with the previous nearby objects, and with the next nearby objects. The use of the product t-norm, instead of the minimum for instance, makes this rule more precise.

5. FUSION AND GLOBAL DECISION

The global decision is made bar per bar. We consider all possible combinations of symbol hypotheses. The ones that include at least one zero graphical compatibility degree are impossible and thus immediately discarded. For the others, we compute the average of all possibility degrees (2), graphical compatibility degrees (7) and syntactic compatibility degrees [8]. The resulting coefficient conveys the symbol class consistency for the configuration. It is then multiplied by another possibility degree checking duration consistency [8], so that we get just one final possibility degree merging all criteria. The decision algorithm chooses finally the configuration that results in the highest final possibility degree, with priority to the ones that achieve the strict metric constraint (number of beats per bar).

6. RESULTS AND CONCLUSION

Tests of the proposed method have been conducted on a large database (100 music sheets, about 42500 symbols) coming from various composers and publishers. The average recognition rate is now 98.55% . It has been slightly increased with respect to the results in [8], although we have introduced more difficult scores, thanks to the improvements realized in the graphical modeling part. The latter has proved to be very efficient especially in case of high symbol density.

Comparison between the results provided by our program (Fig. 9a, 10a) and those output by SmartScore 3.2 Pro Demo [10](Fig. 9b, 10b), one of the most efficient commercial software, shows that the proposed method is able to solve problems for which SmartScore fails.
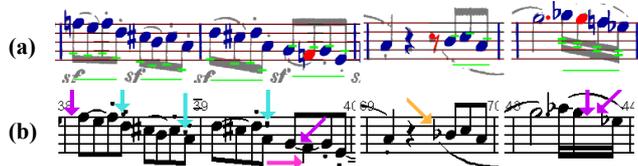

Figure 9. (a) Results with our program: 0 error.
(b) Results with SmartScore: 9 errors

For example, there are some confusions between staccato dots and duration dots with SmartScore (Fig. 9b), while we avoid this problem thanks to our graphical model. We can also point out that symbols touching each other are often not recognized with SmartScore (Fig. 9b), unlike our program that models such configurations. Symbol 9 in

bar (b) is now correctly recognized as a natural, thanks to the fusion with the syntactic rules (Fig. 10a), while it is suppressed by SmartScore. Lastly, the models of symbol classes and syntactic consistency allow here to avoid the confusions made by SmartScore on the eight rest (Fig. 9b) and the half note (Fig. 10b). Examples were extracted from two music sheets. The global symbol recognition rate is 92.0% for SmartScore, 98.7%. for our program. 85.3% of quavers get the correct length with SmartScore, 99.3% with our method.
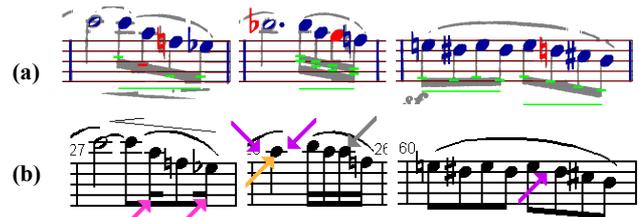

Figure 10. (a) Results with our program: 0 error
(b) Results with SmartScore: 7 errors.

These results show the interest of the proposed method, in particular the fuzzy modeling of symbol classes, that allows to deal with imprecision and variations of symbol shapes, and the fuzzy modeling of the graphical rules, expressed as compatibility degrees between symbols, that allows to deal with spatial imprecision and flexibility.

7. REFERENCES

[1] D. Blostein, H. Baird. "A Critical Survey of Music Image Analysis". *Structured Document Image Analysis*, Springer, Verlag, pp 405-434, 1992.
[2] G. Watkins. "The Use of Fuzzy Graph Grammar for Recognising Noisy Two-Dimensional Images", *NAFIPS Conference* , pp 415-419, 1996.
[3] B. Coüasnon, B. Rétif,. " Using a Grammar for a Reliable Full Score Recognition System", *International Computer Music Conference*, Banff, Canada, pp 187-194, 1995.
[4] H. Fahmy, D. Blostein, "A graph-rewriting paradigm for discrete relaxation: application to sheet-music recognition". *International Journal of Pattern Recognition and Artificial Intelligence*, *Vol. 12, No. 6*, pp 763-799, 1998.
[5] M.V. Stückelberg, C. Pellegrini, M. Hilaro, "An Architecture for Musical Score Recognition using High-Level Domain Knowledge". *4th ICDAR Conference,* vol. 2, pp 813-818, 1997.
[6] D.Dubois, H. Prade. "*Fuzzy Sets and Systems: Theory and Applications*." Academic Press, New-York, 1980.
[7] I. Bloch, H. Maître. "Fusion of Image Information under Imprecision". In B. Bouchon-Meunier, Ed., *Aggregation and Fusion of Imperfect Information, Series Studies in Fuzziness*, Physica Verlag, Springer, pp 189-213,1997.
[8] F. Rossant, I. Bloch, "A fuzzy model for optical recognition of musical scores", *Fuzzy Sets & Systems 141*, pp165-201, 2004.
[9] F. Rossant, "A Global method for music symbol recognition in typeset music sheet," *PRL 23 (10)*, pp. 1129-1141, 2002.
[10] SmartScore 3.2 Pro Demo, http://www.musitek.com/