Chapter for "Visual Perception of Music Notation: On-line and Off-line Recognition"

An Off-line Optical Music Sheet Recognition

P. Bellini, I. Bruno, P. Nesi
Dept. of Systems and Informatics, University of Florence, Florence, Italy
Fax: +39-055-4796425, tel: +39-055-4796363
Corresponding author: nesi@dsi.unifi.it
See also for the research group: http://www.dsi.unifi.it/~nesi/

1 INTRODUCTION

Systems for music score recognition are traditionally called *OMR (Optical Music Recognition)*. This term is tightly linked to *OCR (Optical Character Recognition)* that defines systems for automatic reading of textual documents. Strictly speaking, *OCR* refers to systems that are based on the segmentation and recognition of single characters. Typically, OCR techniques can not be used in music score recognition since music notation presents a two dimensional structure. In a staff the horizontal position denotes different duration for notes and the vertical position defines the height of the note (Roth, 1994). Several symbols are placed along these two directions.

The OMR is a quite complex problem since several composite symbols are typically arranged around the note heads. Despite to the availability of several commercial OMRs: MIDISCAN, PIANOSCAN, NOTESCAN in Nightingale, SightReader in FINALE, PhotoScore in Sibelius, etc., none of these is completely satisfactory in terms of precision and reliability and this justifies the research works around the building of reliable *OMR systems and tool*s.

The OMR systems can be mainly classified on the basis of the granulation chosen to recognise symbols of the music score. There are two main approaches to define basic symbols. The basic symbols can be considered: (i) the connected components remaining after staff lines removal (chord, beam with notes, etc.), or (ii) the elementary graphic symbols such as note heads, rests, hooks, dots, that can be composed to build music notation (Ross, 1970), (Blostein and Baird, 1992), (Bellini, Fioravanti and Nesi, 1999). With the first approach the symbols can be easily isolated from the music sheet (segmented), however, the number of different symbols is very high. The second approach has to cope with huge number of different symbols obtained for composition of the basic symbols. This leads to an explosion of complexity for the recognition tool. Probably a compromise could be the right measure of complexity and capabilities.

The architecture of an OMR system and the definition of basic symbols to be recognised are tightly related to the methods considered for symbol extraction/segmentation and recognition. Generally, the OMR process can be regarded as divided in four main phases: (i) the segmentation, to detect and extract basic symbols from the music sheet image, (ii) the recognition of basic symbols from the segmented image of the music sheet, (iii) the reconstruction of music information, to build the logic description of music notation, and finally (iv) the building of the music notation model for representing the music notation as symbolic description of the initial music sheet.

2 "OPTICAL MUSIC RECOGNITION" (OMR)

OMR is the acronym used to indicate automatic music recognition and reader systems. An OMR system is generically defined as the software that recognises music notation and produces a symbolic representation of music. A robust OMR system can provide a convenient and timesaving input method to transform paper-based music scores into a machine representation for widely

available music software, in the same way, as optical character recognition (OCR) is useful for text processing applications. Converting music sheet into a machine-readable format allows developing applications for an automatic accompaniment, transposing or extracting parts for individual instruments, performing an automated musicological analysis of the music, converting and representing music in different formats (MIDI, Braille Music, etc.).
It is useful to compare the score reading to text recognition. This can suggest strategies for tackling the score reading problem, highlighting problems specific to score reading, and better understanding the capabilities and limitations of current character recognition technology.
Typically, OCR techniques cannot be used in music score recognition since music notation presents a two dimensional structure: in a staff the horizontal position denotes different note duration and the vertical position defines the note height. Several symbols are placed and superimposed along these two directions.
In order to optically recognise text, lines of text are identified by searching for the long horizontal space between lines. Each line is processed in sequence. In each line of text, single characters are handled one at a time, without considering their connection (only at higher level some correction can be applied on the basis of the dictionary). Simple regions labelling algorithms can extract characters for individual handling. Similarly shaped characters are often mistaken as i.e. a 'l' (letter L) for a '1' (number one) even in different contexts, like it may occur in the middle part of narrative text paragraphs. Thus, it is imperative to proof-read the results of any text reading because errors are quite common. In a music page, the first steps used by character recognition programs are not applicable. Isolating glyphs on a page of music is difficult because they overlap and have different sizes. Staff lines are overlapped with almost every other music notation symbols.
The final representation of music is more complex than text where the relevant information is the sequence of characters and the places where the paragraphs break. Even though font information would be desirable, commercial optical character readers supply only limited font information – not closely enough to parse successfully a dictionary definition, for instance. In terms of data structures, this information is typically represented as a sorted sequence of characters with font and size properties attached.
On the front page of a newspaper the relative size and position of text provides information which supplements the meaning of words, gives information about the importance and highlights the relationship with surrounding pieces of information.
A beam groups the set of notes *attached* to it by means of their stems. A key signature is recognised as a *cluster* of accidentals not *beside* note heads. An accidental modifies the note on the right, a dot the note to the left. A slur modifies the performance of notes it marks. The point of view taken in character recognition is inadequate for reading music. In order to address the problem of music recognition different recognition primitives have to be considered as a basis for the architecture of a recognition systems which can handle the diversity of visual relations of music scores.
Automatic music recognition systems can be classified and studied under many points of view since in the literature different approaches have been used.
The identification of the general steps of an ideal music recognition process is not easy, however it is possible to report a list of most relevant steps for the OMR process:

- Digitalisation of music score/sheet to obtain an image;
- Image processing (e.g., filtering) on the acquired image;
- Identification and/or removal of staff lines from the image;
- Identification and/or segmentation of elementary music symbols (basic symbols) allowing to build music notation;
- Reconstruction and classification of music symbols;
- Post graphic elaboration for classifying music symbols;
- Generation of the symbolic representation into a symbolic format for music notation.

The elaboration and filtering of the image is always considered as a unneeded operation, on the contrary the identification and the possible removal of staff is held as a mandatory step by many authors. The graphic analysis (such as basic symbol identification, segmentation, composition and classification) is the core of OMR systems and it is the mostly studied in the literature. Several techniques and solutions have been proposed and many of them are strictly connected with image processing and pattern recognition techniques and methods used in the Optical Character Recognition area. Among reconstruction techniques, the solutions based on the syntactic and semantic knowledge are playing a fundamental role in the phase of post graphic elaboration to help in classifying music symbols. Concerning the generation of the symbolic representation into the chosen format, this is presently an open problem since there is not a standard language able to describe completely the music notation.

2.1 ON-LINE AND OFF-LINE SYSTEMS

Before presenting a selection of the main works in the literature, it is interesting to observe the OMR problem from different points of view, the first step is to distinguish the OMR systems into two main categories: on-line and off-line systems.

In the on-line cases, the system analyses the music sheet directly and provides a result instantaneously: such system could be combined with a robot or a music electronic device connected to a camera to play the music in real time. In this case, the need of generating the music information coding in real time is the most important requirement for such systems. This implies to recognise from each sub-image the music score locally in order to avoid using correction methods based on global information. As the final result, these systems can not consider all aspects of music as, for instance, the interpretation and music effects, since the resolution and the precision is typically low. Another case of on-line system is offered by a new way of writing music with the computer. The user interface technologies and the gesture recognition studies nowadays allow developing new pen based input systems (Anstice et al., 1996) (Ng, Bell and Cockburn, 1998) that aid a user to use a pen in the traditional way. They consist of a tablet, which is normally a touch device on LCD screen (of any size), and an electronic pen or stylus which is used to write on the tablet. The goal of such system is to minimise the input time for data entry into a computer but at the same time they have to deal with the issues derived from the difficult of recognising the human writing.

In an off-line system, the music sheet is digitised by means of a scanner and is stored as an image, later the image is elaborated, analysed and the music information is then converted to a proper symbolic notation code. Such systems have not strong temporal bounds in term of time to spend for the performance in the output production, but only the requirement of quality in the recognition with a low error percentage. This allows spending more computational resources in the global analysis and in the refinement of the identified information.

2.2 SYSTEMS WITH STAFF LINES IDENTIFICATION

The staff lines play an important role in music notation. They mark vertical co-ordinates for the music symbols highness and provide the horizontal direction for the temporal co-ordinates. In the music image analysis, staff lines provide also a dimensional reference, a quality index of the digitalisation, skew and rotation degree of the scanned image. Some authors consider the staff like an obstacle during the recognition of music symbols, for this reason some solutions for staff removal have been studied. As an alternative solution, not considering this problem as an initial step and to start directly with the music symbols analysis means to see the staff as a graphic symbol like the others and a reference for the position of figures. In any case, the position of the staff lines is relevant for the identification of position of music notation symbols.

2.3 USE OF THE MUSIC KNOWLEDGE

To better understand the evolution of OMR systems, it is useful to know how the music knowledge can help the recognition process. The two-dimensional nature of the music notation introduces an important consideration: the music symbols cannot be analysed singularly since the relationships with other symbols may change its nature and interpretation. Relationships among music notation symbols can be expressed by a description based on two-dimensional grammars. In this sense, Fujinaga (1997) gave an important contribution: he holds that the music notation can be formalised by means of a *context-free* and LL(k) grammar because it represents the way of reading used by musicians (top-down parser). A syntactic approach is not complete, leaving out any consideration on the context; for this reason it is necessary to introduce semantic information in the grammar. In this way, the knowledge of the context, together with the syntactic rules helps in making corrections and in improving the segmentation and the objects labelling phases. Some solutions present a recognition process entirely based on graphic identification and graphic classification techniques.

3 ISSUES IN THE OMR SYSTEMS

The study of automatic recognition of music sheets began in the late sixties when hardware aspects, such as CPU performance, memory capacity and dimension of storage devices were strongly limited. Nowadays, fast processors, high-density hard disk, scanner capable to acquire images at high resolution (more than 300 dpi) are of common usage. Thanks to these devices, the automatic recognition of music sheets has been shifted in pure algorithmic problem of image processing, information modelling and artificial intelligence.
In the following, some aspects related to the OMR technologies are shortly discussed in order to give at the reader a global view of the problems of this area.

3.1 GRAPHIC QUALITY OF DIGITISED MUSIC SCORE

Issues related to the graphic quality of the digitised music score involve the visual representation, the object recognition and music modelling. Aspects concerning the visual representation are mainly print faults and quality of the paper, which can provoke:

- Staves rotation, thus lines are skewed as to the page margin,
- Staves bending, thus lines are not straight (this problem can be found in the original, and can be also caused by using manual scanner or photocopying machine),
- Staff lines thickness variation,
- Mistaken position of music symbols (a note covering both a space and a line).

Aspects concerning information loss can provoke:

- Staff lines are interrupted.
- Shape is not completely drawn or filled (a quaver note having white spots in the notehead).

Aspects concerning unnecessary information, such as:

- Spots, which could be misunderstood as being any music, shape (dot, accent, etc…)

Even though, such issues do not seem so important to the human eye, these kinds of irregularity could seriously affect recognition software. Besides, these flaws are more evident when dealing with ancient music scores, which often occurs in the research field, due to copyright grounds.

3.2 GRAPHIC OBJECT RECOGNITION

Most common problems are:

- Changes of dimension of the same music notation symbol (for instance, the stem height and the beam width in ancient music score; or the difference between a grace and a common note; or

clef changing within a measure) represent a problem for recognition approaches focussed on fixed dimensions of symbols.

- Connectivity and overlapping: for instance slurs could generate problems for the overlapping with other symbols such as stems; whereas if they touch a note or overlap a dynamic indication, slurs generate visual shapes which are not associated with any graphical music symbol.

3.3 COMPLEXITY OF THE MUSIC PIECE

Two aspects concerning the music piece complexity can be focussed on: one is related to the distinction between a single part and a main score; the other deals with 'music architecture'.
As to the former, for most instruments the single part is made of a melodic line written on a single staff except for the piano, the harpsichord and the harp having a double staff and the organ having three staves. On the other hand, the main score is structured so as to offer a simultaneous view of all single instrument parts, thus being characterised by a system of staves. To manage a single part having a staff (or at least three) is relatively easier than to deal with a system of staves. A system of staves bring in a problem related to the method to be used in the staff identification process and in the treatment of the topological information (symbol positions, staff numbering, staff domain, etc).
As to the latter, a distinction needs to be drawn between monophonic and polyphonic music: a monophonic piece has a single voice (layer), whereas the polyphonic one has more than a layer/voice. This distinction affects the semantic recognition and consequently the reconstruction step of music content. In both cases, another important issue is the density of music symbols. In fact, in the case of high density, it is hard to obtain a quite good segmentation since the major difficulty is to isolate fragmented symbols that could be easily unrecognised or misunderstood.

3.4 MANUSCRIPT AND PRINTED MUSIC

To deal with manuscripts is different than to deal with printed music. A part from the already described issues, hand-written music recognition brings into further difficulties, such as:

- Notation varies from writer to writer.
- Simple (but important) and even sometimes brutal changes in notation occur in the same score.
- Staff lines are mostly not the same height, and not always are straight.
- Symbols are written with different sizes, shapes and intensities. The relative size between different components of a musical symbol can vary.
- More symbols are superimposed in hand-written music than in printed music.
- Different symbols can appear connected to each other, and the same musical symbol can appear in separated components.
- Paper degradation requires specialised image cleaning algorithms.

These aspects impacts heavily on the choice of methods to be used in the recognition task and it represents the main differences in dealing with the printed and hand-written music. The technology used for printed music recognition is not completely reusable for the hand written one and generally such systems are designed in a different manner.

3.5 EVALUATION OF AN OMR SYSTEM

The lack of a standard terminology and a methodology does not allow an easy and correct evaluation of results produced by an OMR system. Generally the evaluation is based on indexes used for OCR system. For instance, the error rate, pointing out the rate among recognised and total symbols, is not a valid evaluation index, because it is difficult to define (a) when a symbol has been really recognised, (b) its features, (c) relationships with other symbols, and (d) the relevance of the context. Another aspect is the set of music symbols used to represent the music. It is difficult to fix a complete set of symbols, their number depends on either the language or the chosen coding format. To conduct an objective evaluation of a recognition system it is necessary also to build a

database of test cases. In the character or face recognition field, there are many ground truth databases that enable recognition results to be evaluated automatically and objectively (for more details see http://www.nist.gov/srd/nistsd12.htm). At the present time (Miyao, 2000), no standard database for music score recognition is available. If a new recognition algorithm were proposed, it could not be compared with the other algorithms since the results would have to be traditionally evaluated with different scores and different methods. For this reason, it is indispensable to build a master music score database that can be used to objectively and automatically evaluate the music score recognition system. Since the goal of most music score recognition systems is to make re-printable or playable data, in the first case the shape and position information of each symbol is required, whereas in the second information that includes note pitch, duration, and timing is required. Consequently, the ground truth data must be defined in order to evaluate:

– the result of symbol extraction: evaluation of whether or not each musical symbol can be correctly extracted.
– the result of music interpretation: evaluation of whether or not the meaning of each piece of music can be interpreted correctly and whether or not the music can be correctly converted to playable data.

The following three data representation models have been proposed in order to construct a suitable databases for objectively and automatically evaluating the results of a music score recognition system:

1) Primitive symbol representation including type, size, and position data of each symbol element.
2) Musical symbol representation denoted by a combination of primitive symbols.
3) Hierarchical score representation including music interpretation.

The results based upon the extraction of symbols can be evaluated by the data in systems one and two above, and the results of the music interpretation can be evaluated by the data in system three above. In the overall recognition system, the above data is also useful for making standard patterns and setting various parameters.

3.6 SYMBOLIC REPRESENTATION OF MUSIC NOTATION

One of the fundamental aspects in an OMR system is the problem related to the generation of the symbolic representation in a coding format, to allow music information modelling and saving. This problem is shared by music editor applications. In the literature, many music notation-coding formats are available (Selfridge-Field, 1997), while none of these have been largely accepted as a standard formats: Enigma format of Finale, SCORE, CMN (Common Music Notation), Musedata, SMX (Standard Music eXpression). Others have been proposed as interchange formats: NIFF (Notation Interchange File Format), MusicXML by Recordare, while others as Multimedia music formats: SMDL (Standard Music Description Language), WEDELMUSIC (an XML based music language).

4 RELATED WORKS

In this section, some of the most important OMR solutions (Blostein, 1992 and Selfridge-Field, 1994) and more recent works are discussed. Their presentation follows a chronological order and main aspects and innovative ideas are provided without treating technical details of the implementation. The name of the subsection is that of the main proponent of the solution.

4.1 PRERAU

In the 1970, Prerau (1970) introduces the concept of music image segmentation to detect primitive elements of music notation. He uses fragmentation and assembling methods to identify the staff lines, to isolate fragments of notation and rebuild afterwards music symbols. The process, developed by Prerau, can be described as follows:

1. Staves are scanned to detect parts of music symbols (fragments) lying inside and outside the staff lines. The extraction of detected fragments allows removing staves.
2. Fragments are recombined to build complete symbols. Overlapping rules drive the assembling phase: two symbol fragments that are separated by a staff line are connected if they have horizontal overlap.
3. Dimensions of bounding box are measured for each symbol. Prerau holds that highness and wideness of a symbol are sufficient features for the identification of symbols.
4. Symbols are classified by comparing dimensions with a table where topologic features for each symbol are collected considering as most music symbols typologies as possible.

4.2 CARTER

The contribution provided since 1988 by Carter (1989) is very important. Apart from considering specific questions as the identification of staff, he devoted his efforts to the co-ordination of existing researches with the purpose of defining some reference standards. His more important contribution regards the image segmentation process; it is based on a method that uses the Line Adjacency Graph (LAG). First the music image is vertically analysed and then horizontally to search single vertical paths of black pixels, called segments. Graph nodes correspond to unions of adjacent and vertically overlapped segments (sections) while arcs define the overlapping of different segments in an adjacent column (junctions). The graph so built is analysed to detect the staff lines and symbols lying on it. Using this technique provides interesting results, allowing:

- Identification of empty areas of staff (areas without music symbols); in this way it is possible to detect and label the section containing single music symbols or groups of symbols (beamed notes).
- Identification of staff, even if the image is rotated with an angle up to ten degrees.
- Identification of staff, even if lines are slight bent or broken and if the thickness of lines is variable.

The main phases of OMR systems developed by Carter consist in:

1. Application of the method based on LAG to identify the empty areas of staff lines and to isolate the symbols, groups of overlapped or connected symbols.
2. Classification of the objects coming from the segmentation. Symbols are classified according to bounding box size, the number and organisation of their constituent sections. In case of overlapped or superimposed symbols, specific algorithm are proposed so that for objects that are not linked with the staff.

4.3 FUJINAGA

Fujinaga (1988) proposes in 1988 a new approach to the OMR problem based on an intensive use of projection method, in collaboration with Alphonce and Pennycook. He holds that staff removal is not necessary and it needs only to identify its position. In this sense, the projection of the image along the Y-axis is very efficient. The identification of the symbols and their features is conducted by means of the information provided by a sequence of projections first roughly and then more detailed. He associates the music syntax knowledge with the analysis of projections both in the identification and in the classification of symbols.

Fujinaga provided also an important theoretic contribution: the music notation can be formalised by means of a *context-free* and LL(k) grammar. A pure syntactic approach, where the context is not considered, has many limitations; to solve this aspect he suggests introducing semantic information in the grammar. In the following, main phases of the system are briefly described:

1. The staff is identified by analysing the Y-projection of the image. Groups of five peaks, giving a graphic consistent contribution, mark the staff presence; this allows dividing the image in horizontal segments that include the staff (only for monophonic music).

2. Symbols are identified by using the X-projection. Values greater than the background noise caused by the staff lines show the presence of music symbols. As a support for the detection of symbols, some syntactic rules mainly based on the position of notation are used.
3. The symbol classification is performed by calculating both the X and the Y projections for each symbol. Projections are used to extract classification features such as width, height, area, and number of peaks in the X-projection. Together with these pieces of information, Fujinaga suggests to consider also the first and the second derivative of projection profiles.

4.4 ROTH

Roth (1994) introduced a totally graphic approach to remove both horizontal and vertical lines and solve the problem of objects touching each other and broken symbols caused by staff lines removal. The system is based on the following steps:

1. Manual correction of the image rotation
2. Statistical analysis of vertical paths. The average length of vertical paths for black and white pixels is used to find and fix the thickness of staff lines and the white space between two lines.
3. Staff lines removal. The staff lines are identified by searching groups of five peaks in the Y-projection profile. Afterwards, the staff is removed by erasing lines having thickness less than the calculated value in step 2.
4. Vertical lines detection and removal. The identification of all vertical lines (bar lines, stems, vertical lines in accidental shapes, etc…) is obtained by using X-projection profile and morphologic operations.
5. Objects labelling. Objects are marked on the basis of graphic features such as: dimensions, number of pixels and centre of mass.
6. Symbol recognition. Symbols are recognised by using the information provided by the labelling step and graphic rules.
7. Information saving: the result of recognition is stored in a proprietary format.

The first version of the system provided modest results and worked with a limited set of symbols. An improvement of the system has been reached introducing a feedback based on the semantic control. The feedback has been used both to make correction on results and to establish which zones are not correctly recognised. This last case allows repeating the graphic analysis with more careful methods and it allows changing threshold levels that control the process.

4.5 COUASNON AND CAMILLERAP

Couasnon and Camillerap (1995) in their research affirm that the music knowledge is very important in the recognition process. In addition, they hold that the information coming from knowledge has not to be used only as a verifying instrument to correct errors, on the other hand it has to be used also to control the whole process.
They formalise the music knowledge by defining a grammar. The grammar describes syntactic rules and represents both the musical and graphical context. The immediate advantage is the possibility to separate the elaboration part of the system from the definition of music rules. In this way if the document shows a different structure only the grammar needs to be re-defined, while using the same parser. In this approach, the segmentation and labelling phases of basic symbols are controlled by the set-up grammar: basic symbols are not labelled during the extraction phase since the labelling phase depends on the verification of syntactic rule consistency.
The proposed system can be divided in the following main phases:

1. Segmentation and labelling of primitives. These operations are driven by the music knowledge in order to control the correspondence of detected graphic elements with syntactic rules.

2. Reconstruction and classification of music symbols. This phase is immediate since the consistency of position and the sequence of primitives have been already tested.
3. Verification and correction of results and control of notes values and figures alignment.

Using the music knowledge to control the whole process allows the evolution of the system and then the recognition of more complex score (scaling up), since it is always possible to define new grammars formalising music notation with different complexity. The system provided good results with polyphonic music score (two voices).

4.6 BAINBRIDGE

Bainbridge (1991, 1996, 1997) focuses his study on the problem of automatic comprehension of music with particular attention to human cognitive process. The wealth of music notations, their evolution, the presence of different set of symbols and the personalisation by musicians and publishers highlight the dynamic nature of the OMR problem.

In the natural process of the music comprehension two main phases can be detected: the recognition of graphic shapes and the application of music knowledge to deduce the meaning. Thus, an optimal OMR system should be constituted by a Drawing Package, where the elementary graphic shapes defining the music symbology are described, together with a Specially Designed Music Language, providing a structure to describe the abstract knowledge of music notation. Both modules should have properties of flexibility and dynamism so as to be easy to fit new music notations or new interpretations of symbols. For this reason, Bainbridge introduces the Primela language that allows defining acceptable configurations of primitive music symbols and to couch the music semantic. A Primela description is written for each type of primitive music shape to be recognised with the pattern recognition routine most appropriate for the task selected. In the Primela language elements such as spatial relationships among primitive symbols, spatial restrictions, graphical rules and Boolean expression are used. Once the primitive objects on the page have been identified, they must be assembled into the larger objects they are a part of. For example, note-heads and stems combine to form notes, and a sequence of sharps in a particular context can form a key signature. The CANTOR system was designed for this reason and to be adaptable so that such alternative forms could also be recognised.

The assembly phase in CANTOR is implemented using Definite Clause Grammars (DCG's). DCG's are similar to BNF, but use a slightly different notation, which make them amenable to implementation in a Prolog environment.

4.7 NG & BOYLE

In 1994, Ng and Boyle (1992, 1994, 1996, 2002) develop the *Automatic Music Score Recogniser (AMSR)* at the University of Leeds (Great Britain). The system works on Unix platforms. The input of the system is an image in bitmap format that is visualised on the screen and the output is coded in Standard Midi File. The used approach is based on the reverse process in writing music: a composer normally writes at first the note head and then the stem or the beam, as last he adds slur and tie. The system selects first the horizontal and thick elements such as slur and subsequently beams and stems. In this way composite and complex music symbols are decomposed in primitive symbols. The system is divided into different sub-systems as follows:

1. The pre-processing sub-system consists in a list of automated processes including:
   - Thresholding, to convert the input grey image into a binary image;
   - Deskewing, the image is rotated whether any correction of the skew, usually introduced during the digitisation, is needed. This step is necessary because the orientation of the input image affects any further process where the projection method is used.
   - Staff lines detection and definition of a constant value obtained by adding the average highness of a staff's black line with the distance between two lines. The constant value so

calculated is used as fundamental unit in further processes and as general normalisation factor.

2. In the sub-segmentation process, the composite music symbols are divided into lower-level graphical primitives such as note heads, vertical and horizontal lines, curves and ellipse. From the initial segmentation, each block of connected features is passed into the classification module. If the object is not classified confidently, and it is too large as a feature, it is passed into the sub-segmentation module to be broken into two or more objects.
3. In the classification process, primitives are classified using a k-Nearest-Neighbour (kNN) classifier based on simple features such as the aspect ratio and normalised width and height, being graphical features of the primitives relatively simple. After recognition, sub-segmented primitives are reconstructed and contextual information is used to resolve any ambiguities. To enhance the recognition process, basic musical syntax and high level musical analysis techniques are employed for instance: automatic tonality detection, harmonic and rhythmic analysis.

The default output format is set to *Exp*Midi (Expressive Midi) (Cooper, Ng and Boyle, 1997), which is compatible with the standard Midi File format, and it is capable of storing expressive symbols such as accents, phrase markings and others.

The pre-processing task for the recognition of printed music has been re-used for hand-written manuscript analysis in (Ng, 2002). The sub-segmentation approach relies on the vertical orientation of the musical features and performs unsatisfactorily for hand-written music due to the characteristically slant and curve line segments. For hand-written manuscript, the sub-segmentation module adopts a mathematical morphology approach (Suen and Wang, 1994), using skeletonisation and junction points to guide the decomposition of composite features, and disassembles them into lower-level graphical primitives, such as vertical and horizontal lines, curves and ellipses.

4.8 ADAPTIVE OPTICAL MUSIC RECOGNITION.

In 1996, Fujinaga (1997, 2001) started to work to the *Adaptive Optical Music Recognition* system. This version enables learning new music symbols and hand-written notation. The adaptive feature allows different copies of the system to evolve in different way, since each of them is influenced by the experience of different users. The system consists in a database of symbols and three inter-dependent processes: a recogniser detects, extracts and classifies music symbols using a k-Nearest-Neighbour classifier. The recognition is based on an incremental learning of examples and it classifies unknown symbols finding out the most similar object which the system has in its database. An editor for music notation allows the user to make correction and a learning process improves the performance and the accuracy of any further recognition sessions, updating continuously the database and optimising the classification strategies.

The whole system is divided into the following main sections:

1. The staff lines removal is performed without removing music symbols. To determine the height of lines and white spaces between two lines, the *vertical run-lengths* is used. All black vertical lines, having a thickness twice as great as the staff line and less than the height of the white space, are erased. To correct the skew of the page, parts of the image are shifted either on the top or the bottom. Objects such as slur and dynamic indications are removed if they have a height similar to the black line thickness and a minimal bounding box comparable with the bounding box containing the line. Staff lines are selected by means of a projection along the vertical axis; the presence of more than one staff and the left and right margin of the page are established by projecting the page along the horizontal axis.
2. The text location and removal tasks are performed by using heuristics. An OCR allows recognising the word; letters as dynamic symbols are not removed since they are not inside a word. This task fails for instance when letters are connected each other or touch the staff lines; a note touching the staff could be recognised as a letter.

3. In the segmentation task, the image is decomposed in disjoined image segments; each segment contains an object to be classified. For each symbol, measurable features (height, width, area, etc…) are extracted and the *feature vector* is calculated by means of *genetic algorithms*.
4. The classification is performed using a k-Nearest-Neighbour. The classifier decides the class the symbol belongs to on the account of its feature vector. The k-NN does not require the knowledge a priori of the symbol distribution in the space of features; this property allows learning new class of symbols.
5. Reconstruction of the score is performed by using a Prolog based grammar.

This system was chosen as the basis for of the Levy OMR system, called Gamera, and expanded with an Optical Music Interpretation system (Droettboom and Fujinaga, 2002).

4.9 LUTH

The work conducted by Luth (2002) is focussed on the recognition of hand-written music manuscripts. It is based on image processing algorithm like edge detection, skeletonisation, run-length. The automated identification begins with the digitising of the manuscript and continues with the following main steps:

1. Image Processing – The aim of this step is the extraction of a binary image with minimal disturbances and maximal relevant structures. The proposed approach for this task is to conduct an adaptive thresholding in small regions by automatic estimation of optimal threshold value for each region. In this way a good quality of the conformity between the binary and the original notation graphics is achieved.
2. Image Analysis – The binary image is decomposed into 5 layers and each layer corresponds to special structures: (i) horizontal lines (staff), (ii) vertical lines (bar, stems), (iii) small circular filled structures (note heads), (iv) line structures (clefs, note flags or accidentals) and (v) other structures (textual information).
3. Object Recognition – It is based on an a-priori knowledge and structuring rules of abstract models (staff lines, note stem, note heads) to assign contextual meanings to the extracted structures. This task also generates a special Notation Graph. It involves a syntactic level of handwriting for the description of relationships among structures in terms of geometry, metrical and features.
4. Identification – A tree of features is built and collects two different forms of structural descriptions: the structure of the abstract object features and the appearance of the features as primitives in the image. Once object features are given in the form of structural descriptions, a graph matching algorithm determines: which image primitives belong to the same object feature, the identity of structure and the correct object features to each image primitive.

5 THE O^{3}MR

In this section, results of the Object Oriented Optical Music Recognition System (O^{3}MR) are described. This system is under development at the Department of System and Informatics of the University of Florence (Italy) and it is inserted in the context of the IMUTUS (Interactive Music Tuition System) and WEDELMUSIC (Web Delivery of Music Scores, www.wedelmusic.org) IST projects. In particular, aspects and adopted solution related to the off-line printed music sheet segmentation problem are shown in this section (Bellini, Bruno and Nesi, 2001).
The general architecture of the O^{3}MR is based on four main components (see Fig. 1):

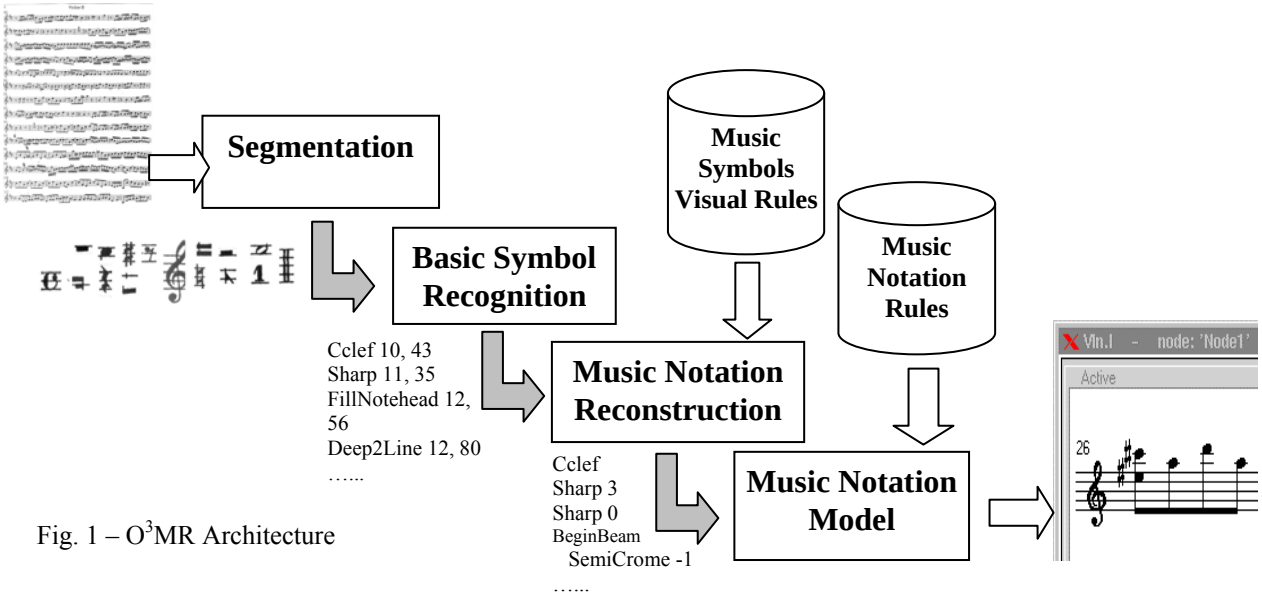


Fig. 1 – O^{3}MR Architecture

- **Segmentation –** the music sheet is processed with the aim of extracting basic symbols and their positions. A basic symbol is an elementary symbol that can be used for building the music notation symbols. For example: the filled note head; the deep lines representing beams, rests, sharps, flats; the empty note head; the accidentals; the thin lines used for drawing the staff, stem, slur, tie, wedges, etc. This means that each basic symbol can be used for building more music notation symbols. The exact identification is performed in the third block of O^{3}MR architecture.
- **Basic Symbol Recognition** – the module performs the recognition of basic symbols by using a neural network. It takes in input image segments of the basic symbols. On the basis of the set of basic symbols a feed–forward neural network has been set and trained to perform the recognition. The output of this module is mainly symbolic. For each recognised basic symbol, the image segment co-ordinates and the confidence value of recognition are produced. When bar lines are recognised, the corresponding image segment is further elaborated in order to estimate the position of staff lines.
- **Music Notation Reconstruction –** The recognised basic symbols are mapped into the elementary components of music notation symbols. For example, a deep line may be a part of a beam as well as of a rest, etc. The decision criteria is based on the recognition context: the position of the basic symbols with respect to the position of staff lines, the confidence level of the first phase of recognition, etc. In order to make simple the identification of elementary symbols, on the basis of their possible relationships, the **Visual Rules of the Music Symbols** have been formalised and used during the recognition. According to this process, for each basic symbol a set of probable elementary symbols are assigned. These elementary notation symbols estimate the probability to be basic symbols on the basis of the context. This module may

request some additional evaluations when the decision cannot be taken with the current knowledge – for example when two Music Notation Symbols are similarly probable.

- **Music Notation Model –** once the basic notation symbols are identified they are composed on the basis of a set of **Music Notation Rules**. Thus, the music model is reconstructed by refining the recognition performed in the previous phase.

5.1 SEGMENTATION

The segmentation is the most critical phase of an OMR process. It has to guarantee that *the segmentation of basic symbols is independent on the music score style, size and on the music processed, from simple (sequence of single notes) to complex (chords in beams with grace notes and several alterations, markers, slurs, expression, ornaments, etc.).*

The first problem addressed in music score segmentation is the management of staff lines that touch the elementary symbols. The removal of overlapping lines requires a complex process of reconstruction of involved symbols, with corresponding loss of information. As a consequence some authors preferred to recognise symbols without removing the portion of lines crossing theme. For these purposes, the use of projection profiles is very common.

The second problem to be considered is that music notation may present very complex constructs and several styles. Music notation symbols are various and can be combined in different manner to realise several and complex configurations, sometimes using not well defined formatting rules (Ross, 1970). This aspect impacts on the complexity of the segmentation problem. A method to cope with the complexity is to regard the music notation as a set of basic symbols whose combination allows producing the entire set of music notation symbols and their combination.

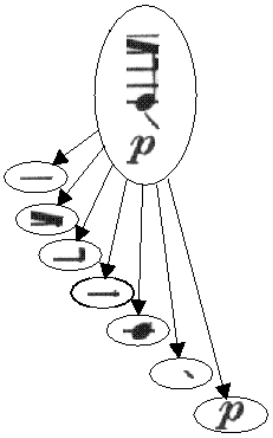


Fig. 2 – Basic symbols decomposition

An image of a music score page grabbed with a scanner is the starting point of the segmentation process. The music sheet image is analysed and recursively split into smaller blocks by defining a set of horizontal and vertical cut lines that allow isolating/extracting basic symbols (see Fig.2).

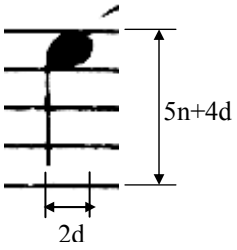


Fig. 3 – Tuning

Staff lines are graphic symbols that are independent on the music content. They give important information about music sheet features, since thickness of staff lines, n, and the distance between staff lines, d, are useful to tune the segmentation process. In fact, they allow defining thresholds, tolerance values for measurements and segmentation of basic symbols of the music score. With the thickness of lines and the distance between two lines, it is possible to describe the graphic features of a note head or to estimate the height of a single staff, as shown in Fig.3.

For these reasons, staff lines are not removed since they are used in the segmentation process and avoid the introduction of elaboration phases to fix symbols partially cancelled by the staff lines removal. The knowledge of staff lines position allows detecting the right pitch of notes in the reconstruction phase.

In more details, the procedure is based on the three elaboration levels as following:

- **Level 0**: the music sheet is segmented to extract sub images including the single music staff. In addition, a set of image score parameters are estimated for tuning the next processing phases.
- **Level 1**: the image segment of each staff is processed to extract image segments that include music symbols and have a width close to that of note heads. This level is performed into three steps: (i) extraction of beams (e.g., group of beamed notes) and isolated symbols (e.g., clefs,

rest, bar line); (ii) detection and labelling of note heads; (iii) detection of other music symbols or parts of them.

- **Level 2**: music symbols, detected at level 1, are decomposed in basic symbols. In this phase, two decomposition methods are used: for image segments containing note heads and for those in which they are missing. In this last case, the image segment may contain other symbols.

5.1.1 Level 0

The main stages of Level 0 are the (i) tuning of the segmentation process by the identification of a set of graphic parameters, (ii) detection of image segments in which staffs are included. According to the idea of a hierarchical structure, the image is decomposed in a set of sub images, each of which includes a staff.

Tuning Process -- The tuning process is performed to estimate the music sheet parameters from the scanned image: (i) the thickness, *n,* of staff lines, and (ii) the distance *d* between staff. To estimate these values, the score image is processed column by column counting sequences of black and white pixels, the most frequent occurrence values fix the number of pixels for the thickness of staff line (*n*) and the space between two lines (*d*). In order to manage variability and noise on the values of these parameters, two intervals have been adopted as follows:

- n_1 and n_2: the minimum and the maximum values for the staff line thickness, respectively
- d_1 and d_2: the minimum and the maximum values for the distance between two staff lines, respectively

The above parameters are used in the next steps of the segmentation process

Staves Detection and Isolation -- The staff detection is the first real segmentation phase of the O^{3}MR system. The goal is to identify a rectangular area in which the staff is located in order to process that image segment to extract the contained basic music symbols. The algorithm for detecting the staffs is based on the recognition of the staff line profile. The profile, obtained by applying the Y-projection to a portion of staff lines image, presents a regular pattern in terms of structure whereas other projections have a variable pattern. In Fig. 4 the staff lines are zoomed and reported in black on the left. In order to distinguish the projection of lines from the other graphic elements, a transformation of profiles has been introduced.

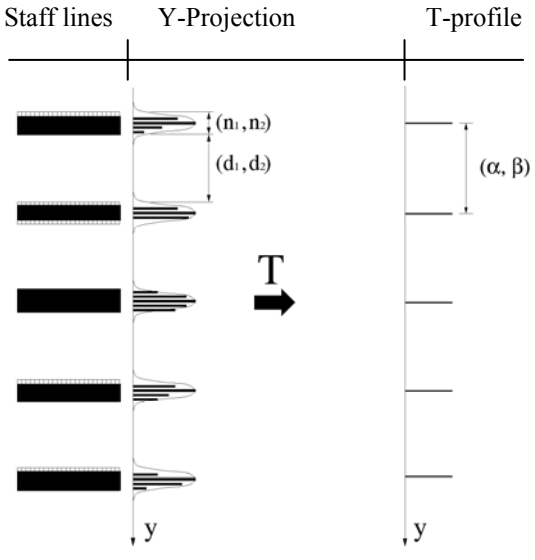


Fig.4 – T-transformation of staff lines profile

The transformation, T, works on the Y-projection of a vertical image segment. The Y-projection is constituted by a set of groups/peaks, each of which is associated with a staff line. When the width is comparable with values defined by [n_1, n_2], then the position of the mean position for the peak is estimated, otherwise it is not considered. The position of the mean values defines the lines in the T-Profile of Fig. 4, and allows characterising the staff in the identification phase. The analysis of the distance between lines in T-profile is used to understand if the profile is due to the presence of a staff.

The distance between lines of the T-domain is strictly related to the values of the above mentioned parameters. In fact, given the [n_1, n_2] range for the thickness of staff lines and the [d_1, d_2] range for the distance between two lines, the distance between mean values expressed in term of tolerance range [α, β] is defined as:

[6.1]
$$\begin{cases} \alpha = d_1 \\ \beta = d_2 + 2(n_2-(n_2-n_1)/2)\ . \end{cases}$$

The staff detection algorithm looks for the "five equidistant" lines structure. This is performed by analysing thin slices of the image sheet. Each slice has a vertical size equal to the whole image score size and width equal to a few pixels (*dx*). The slices processed performing T-transformation. On each slice, a window probe with height $I = 5n_2+4d_2$ looks for the five lines patter. The probe is applied on a sub-section of the slice and analyses it from top to bottom. The probe looks for the starting co-ordinate of the five staff lines.
In order to cope with eventual staff deformations, the above value of *I* has been increased of 20%. The staff detection by means of the probe is performed in two phases: (i) discovering and (ii) centring the staff. In the discovering phase, the probe window is used to detect the staff lines and store the starting co-ordinates of segment in which the staffs are present. In the centring phase, a couple of staff co-ordinates (y_{sup}, y_{inf}) are obtained. These are defined in order to fix the cut lines for extracting the staffs contained in the score image. If *n* is the number of staffs, then each new couple of co-ordinates has been evaluated as:

[6.2]
$$\begin{cases} \hat{y}_{\text{sup}}^{(1)} = 0 \\ \hat{y}_{\text{inf}}^{(1)} = \dfrac{y_{\text{sup}}^{(2)} + y_{\text{inf}}^{(1)}}{2} + \varepsilon \end{cases} \qquad \begin{cases} \hat{y}_{\text{sup}}^{(n)} = \dfrac{y_{\text{sup}}^{(n)} + y_{\text{inf}}^{(n-1)}}{2} - \varepsilon \\ \hat{y}_{\text{inf}}^{(n)} = Y\max \end{cases}$$

Where: $i = 2, \ldots n-1, \varepsilon \geq 0$ (a tolerance value to increase robustness). The introduction of ε tolerance allows getting adjacent or partially overlapped image segments containing the staff. In the above definition, the co-ordinates of the first and last staff are excluded.
An example of staff detection is provided by Figs 5.a and 5.b to better clarify the described method. The considered image has been acquired at 300 dpi with 256 grey levels. The tuning parameters for this image are n = 4 and d = 17, and they determine the following ranges for the staff lines description:

- [3,5] pixels used for the thickness of black lines
- [16,17] pixels for the distance between two black lines
- [16, 25] pixels for the distance among black lines to be considered after the T transform
- I = (25+68)+20% = 112 pixels for the window probe

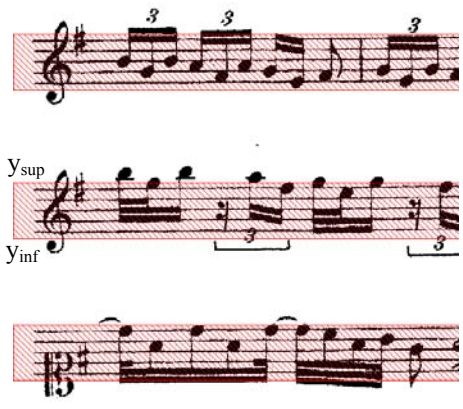


Fig. 5.a – Staff detection

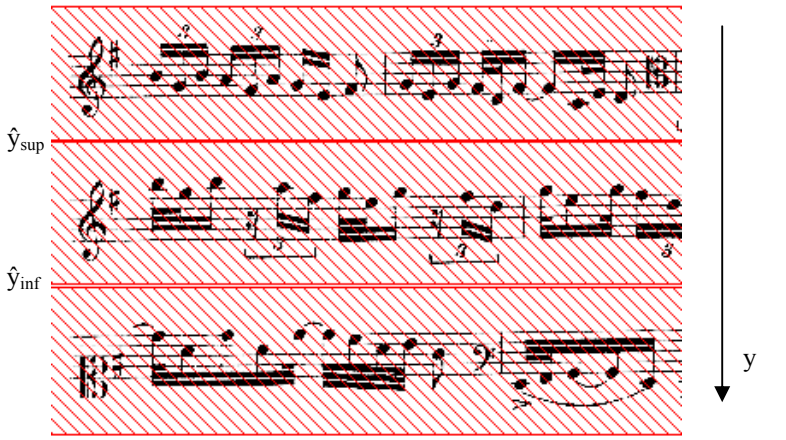


Fig. 5.b – Staff area extension after the application of equations 6.2 with $\varepsilon = 0$

5.1.2 Level 1

Level 1 works on the image segments produced from Level 0 and containing one staff. The aim of this level is to extract the vertical image segments containing music symbols and to produce the lower level sub-image segments in three phases.

Groups & isolated symbols detection -- In this phase, groups of figures and single symbols (e.g., clefs, rest, bar line) are detected. To this end, the staff detection algorithm is applied to produce the value of the binary function, *F*. The detection process has been realised by considering a running image window. This has the height of the image segment coming from level 0, and width, dx_1, equal to 3-4 pixels. The analysis of the image segment is performed by moving the running window from left to right of one pixel a time. The result of staff detection sets the values of binary function, *F* (see Fig. 6). The 0 value is associated with the presence of an empty staff and 1 otherwise. In this phase, the stand-alone music symbols (clef, bar line, time signature, whole notes, etc.) are detected. Whereas, non-stand alone music symbols and all groups of figures have to be processed in the next phase in order to proceed at their decomposition in smaller image segments.

The identification of empty staff allows processing of the corresponding image segment in order to estimate the co-ordinates of staff lines. This information is used in Level 2.

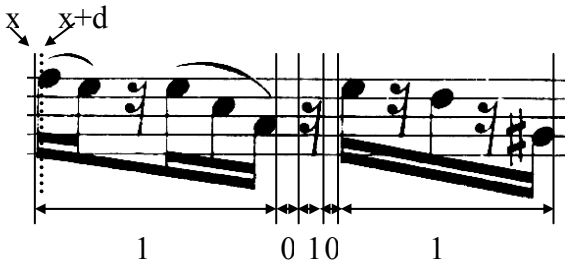


Fig. 6 – Symbol segment detection, function F

Note head detection & labelling -- The goal of this phase is to slice the complex image segments produced by the previous phase and marked as F=1. In the slices produced by this phase one or more single note heads have to be present along the y-axes. The proposed approach is based on searching the presence of single note head. In western notation, note heads may present at least a diameter equal to the distance between two staff lines. To consider the note head width equal to $2d_1$ is a reasonable approximation. For these reasons, image segments coming from the previous phase are processed on the basis of their width. In this case, only image segments larger than $2d_1$ are considered. In the X-projection, we have: (i) spikes due to note stems and vertical symbols; (ii) offsets due to horizontal symbols like staff lines, beams, slurs, crescendo, etc. (iii) smoothed dense profile due to note head. In order to extract the note heads the dense profile contribution has to be extracted from the X-projection. This means to eliminate the other two contributions. To this end, a thin running window is considered on the image segment containing the staff with symbols (see Fig. 7). The running window scans the image with a step equal to 1 pixel. For each step/pixel the Y-projection is calculated. In the projection, the presence of a note head can be detected on the basis of its width, H, which is in the range $[2n_2, 2n_2+d_1]$. Since the objective of the process is to detect the location of note heads, only the maximum value of H along the projection of the running window is considered (Hy6). This value is reported in the final X-projection shown in Figs. 7 and 8a. The note heads produce higher peaks since they are deeper than beam lines. On the other hand, when note heads are missing the maximum value produced from the beam is reported in the X-projection of max Y-projection. The evident difference in the value of the two cases is amplified by the adoption of a running window that brings to consider several overlapped thin windows of the same note head. The final step consists of isolating the note heads. This is performed by using a double threshold mechanism to obtain the results shown in Fig.8b. The first threshold is defined as the compromise between the dimension of the note head and that of the beams. The second filtering is performed on the width of the remaining peaks that are considered as due to the presence of note

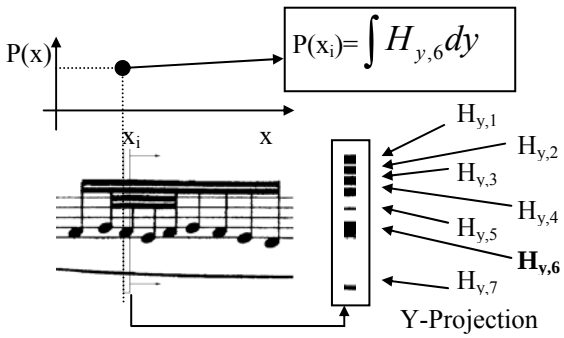


Fig. 7 – Building of X-Projection

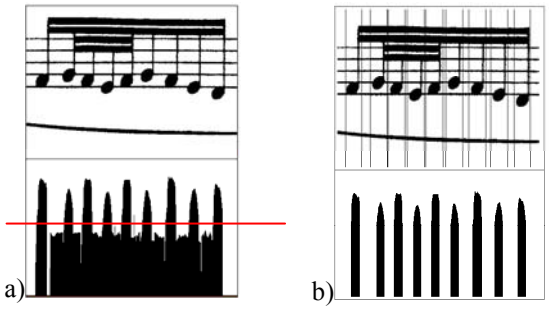


Fig. 8 – a) X-projection before thresholds application; b) X-projection after thresholds application and ready for extracting image segments with note heads.

only if their width is larger than $d_1/2$. For each peak of the X-projection the mean position of the peak along the X-projection is estimated. If C is the x co-ordinate of the mean value, the couple of points is defined as following: $(C-d_1, C+d_1)$. Each couple of points defines the width of the image segment that includes the note. In particular, the width is equivalent to that of the note head (see Fig. 8a).

In this process, each image segment containing a note head is labelled. This information is used by Level 2. The above process for labelling segments that contains note heads is also performed on the image segments marked with F=1. Please note that with the presented approach also chords having notes on the same side are managed (see next section). Image segments that do not contain note heads are non-sliced by the described process.

Other Symbols Detection -- Image segments that have to be processed typically include groups of symbols very close to each other. In most cases, these symbols are separated by small spaces generating local minima in the X-projection profile, such as in Fig.8. Thus, detecting these points means to allow slicing the image segment and thus to extract the basic symbols. To this end, an iterative method has been developed. As a first step, a low pass filter is applied to smooth the profile. The smoothed profile is analysed in order to find the position of the minima. Its position is used to divide the image segment in two new sub-images. The same procedure is applied at the two sub-images when their width is greater or equal than $d_1/2$. The process stops when the maximum width of the processed image segments is lower than $(5/2)\ d_1$. The last operation consists in sorting the points for vertical image segment in order to define image segments.

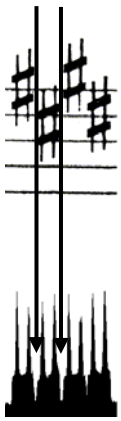


Fig.8 – Isolation points and minimums of X-projection

In presence of a "constant" profile the segmentation process produces image segments with a width comparable to that of note heads, $d_1/2$. Image segments having a width lower than the staff line thickness are not considered in the next segmentation level. This process is capable to cope with key signatures, and accidentals, since it allows decomposing the image segments non-decomposed in the previous step.

5.1.3 Level 2

Level 2 is the last phase of the segmentation process. In this phase the images coming from the previous Level 1 are decomposed into a set of basic symbols. This phase covers an important role since the recognition and reconstruction are strictly connected to it. The produced image segments must include graphic details: (i) belonging to the set of defined basic symbols in repeatable manner, (ii) additional information needed for their recognition. The first aspect impacts on the reliability of the recognition phase, while the second on the application of the rules of the reconstruction phase.

The previous segmentation phase produced different results depending on the presence or not of note head. In this process, two different segmentation methods are applied to the received image segments: (i) including, and (ii) non-including note heads. This distinction is possible by using the information given by labels defined in the note head detection phase. Both segmentation methods are based on the Y/X-projection and produce couples of vertical co-ordinates to extract basic symbols. The result of this level is provided in the Fig. 9.

Images with note heads -- In image segments containing a note head, this can be connected to other basic symbols. This implies that a specific process for their division is needed. Ornaments (turn, mordent, trill, etc.) and horizontal (slurs, crescendo, etc.) symbols can be more easy identified since they are not connected or strongly close to the note head. The graphic components of the note are mainly: (i) note head, and (ii) beams or hooks. In the Y-projection of notes, the stem contributes adding an offset to the profile of the projection linking the profile of the note head with beams and hooks. In the case of a quarter and whole note, it is only an offset. In both cases, the contribution of stem is used as the lower value for the threshold that is applied for the extraction of basic symbols. The proposed approach is based on the result obtained in (Marinai and Nesi, 1999) and it consists of the following steps:

1. **Staff lines removal from Y-projection.** According to the processing of the last segment containing an empty staff the position of the staff lines is known (see Level 1). The contribution of the staff lines is removed by masking their contribution lines having a width of n_2.
2. **High-pass filtering.** This filtering phase allows eliminating the low frequency components and removing the offset due to the stem.
3. **Extraction points computation**: it is based on the computation of extraction points by means of a threshold mechanism (Marinai and Nesi, 1999). When two successive segments are closer than n_2 they are fused in unique segment.

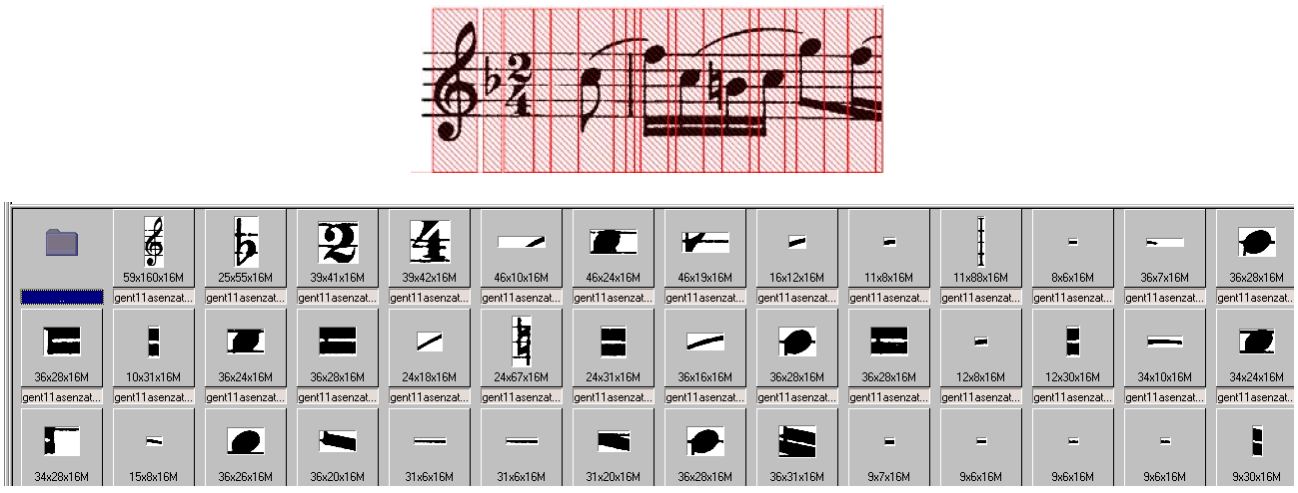


Fig. 9 – Decomposition in basic symbols: output of level 2

5.2 RECOGNITION OF BASIC SYMBOLS

To perform the recognition and classification of basic symbols (see Fig. 9 and 10) an MLP-Backpropagation neural network is used (Rumelhart and McClelland, 1986). It takes in input the simple normalised image segments (8 x 16 pixels) of the basic symbols and provides the class and the confidence of classification (goodness percentage) as output. More than 20.000 basic symbols have been collected and distributed into a database constituted by 54 classes of elementary shapes. This set allows considering classical music scores. On the basis of set of basic symbols a feed–forward neural network has been set and trained to perform the recognition.

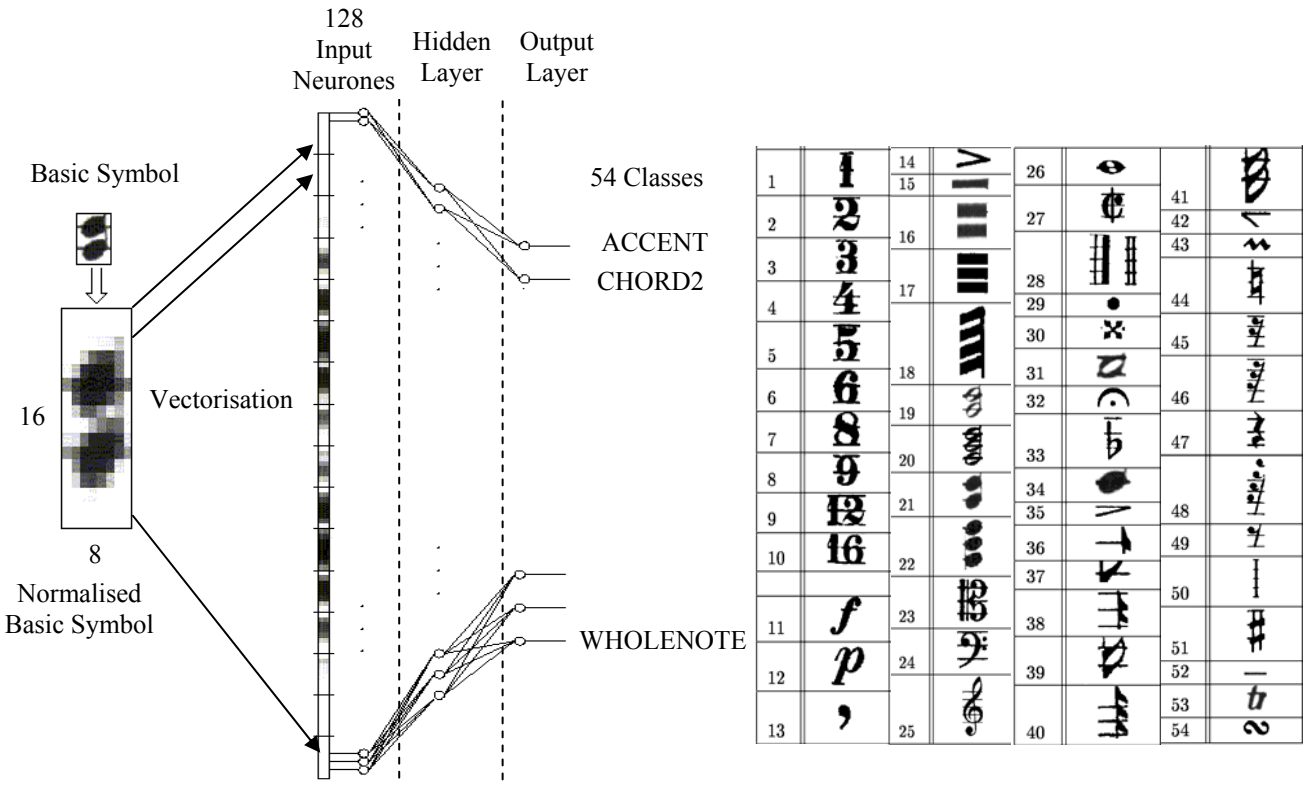
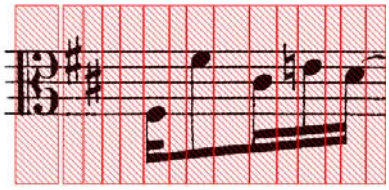


Fig. 10 – The MLP Classifier Structure and the Database of basic symbols

The MLP network structure consists in 128 input, a hidden layer of 256 neurones and 54 output neurones; the neuronal activation function is sigmoidal with output values in the [0,1] range. The normalised image is vectorised and provided as input (see Fig. 10). The reached percentage range of classified symbols oscillates between 80-90%, this oscillation depends mainly by the different kinds of font and by the likeness among different basic symbols (i.e., the augmentation dots and black note heads, beam and part of slur, etc.).
For this reason, a confusion table has been defined in order to consider the mistakes introduced by the network. By means of the table and the confidence value associated to the classification, corrections can be done considering the position that the confused symbol occupies in the staff. The error recovery is performed in the reconstruction phase.
The output of this module is mainly a symbolic description. For each recognised basic symbol, the image segment co-ordinates, dimensions of bounding box and the confidence value of recognition are produced:

Class	X	Y	Width	Height	Confidence
CLEFTREBLE	55	48	57	153	0.99

Moreover, when bar lines are recognised, the corresponding image segment is further elaborated in order to estimate the position of staff lines. This information is extracted and communicated to the successive module to be considered as reference lines (STAFF5 line).
In presence of groups of note such as beamed notes, the segmentation task provides the start and the end point of groups, this information is useful since it reduces the context analysis and provides a way to control the reliability of results.
In the Fig. 11, an excerpt of the symbolic description and the related image are shown.



Output

```
STAFF5 48 57 75 94 113 133
CLEFALTO 49 54 54 84 0.31
STAFF5 107 57 76 95 114 134
SHARP 108 37 25 63 0.99
STAFF5 131 57 76 95 114 134
SHARP 134 65 23 61 0.90
STAFF5 207 57 76 95 114 133
BEGIN
FNOTEHEAD 208 123 32 23 0.99
BIGLINE1 208 165 32 14 0.55
BIGLINE1 208 180 32 15 0.81
BIGLINE1 240 178 21 14 0.97
FNOTEHEAD 261 56 34 23 0.99
BIGLINE1 261 174 34 16 0.50
THINLINE 295 170 44 16 0.36
FNOTEHEAD 339 83 34 25 0.98
BIGLINE1 339 152 34 14 0.80
BIGLINE1 339 168 34 14 0.61
NATURAL 373 46 26 61 0.50
BIGLINE1 373 150 26 14 0.51
BIGLINE1 373 166 26 14 0.66
FNOTEHEAD 399 64 34 25 0.99
BIGLINE1 399 145 34 15 0.68
BIGLINE1 399 163 34 15 0.46
BIGLINE1 433 144 19 13 0.95
BIGLINE1 433 160 19 15 0.76
FNOTEHEAD 452 75 31 23 0.99
THINLINE 452 143 31 14 0.97
THINLINE 452 159 31 14 0.97
END
.....
```

Fig. 11 – Classification and symbolic description

5.3 MUSIC NOTATION RECONSTRUCTION

In this section, the music reconstruction module foreseen in the O^{3}MR will be briefly described. Referring to the Fig. 12, the module is constituted by three main components:

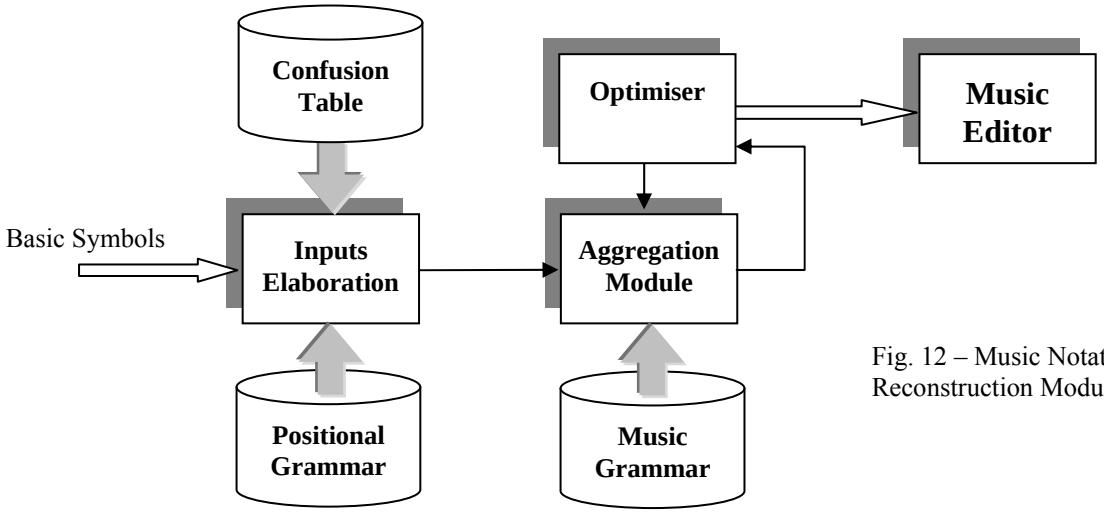


Fig. 12 – Music Notation Reconstruction Module

Inputs Elaboration – The Basic Symbols recognised by the previous step are assembled in strips according to the image segments produced during the segmentation. In this phase, by means of the confusion table of the MLP Neural Network and a set of rules based on the probability that a symbol can occupy that position in the score (Positional Grammar), it is possible to perform the recognition mistakes recovery. At the same time, if a symbol can assume different meanings related to its position (i.e., bass clef or baritone clef), by using the position rules it is possible to establish the right assumption.

Aggregation Module – Symbols are aggregated an order to rebuild music notation symbols. A rule based Music Grammar describes the relationships among music symbols. Rules are strictly connected to the way of writing music and take in account how basic symbols have to be configured in order to obtain music symbols and how each music symbol is affected by adjacent symbols. The aggregation is conducted in two steps: (i) vertical aggregation and (ii) horizontal aggregation. In the first step, if for instance the strip contains a black note head and a beam, the aggregated music symbol is a one-beamed eighth note. In the second step, if the aggregated music symbol is adjacent to an accidental symbol and this symbol stays on the left of the note, then the two symbols are fused together defining an eighth note with accidental and beamed. In this way it is possible to consider the music context where the music symbol is in. During this phase, single measure is rebuilt considering the duration of assembled music symbols and indications of start measure and end measure given by the recognition of bar lines.

Optimiser – The input of the optimiser is a music measure assembled by the Aggregation Module. If the measure is already the optimal measure then is ready to be converted in to the symbolic notation model, else it needs to be elaborated again by the Aggregation Module. The goal of this feedback is to identify the strip that generated the incorrect music symbol. The Aggregation Module re-considers and modifies the strip. The new strip and the grammar rules will produce a new music symbol that will be considered again in the evaluation of measure correctness. The goal is to determine the "best" strips in order to minimise a cost functional that takes into account the duration of measure and the number of recognised basic symbols involved in the aggregation step.

6 FUTURE TRENDS AND CONCLUSIONS

The Optical Music Recognition problem is a wide research field and presents nowadays many open problems. Several music recognition systems have been implemented but it is difficult to compare them because of different set of data used and to judge how much each system is tuned to particular examples for which results are reported. This problem is due to the lack of a standard evaluation metrics. To cover this lack, one of the tasks of the MUSICNETWORK IST project (http://www.interactivemusicnetwork.org) is to define the guidelines for evaluating the capabilities and the performance of Optical Music Recognition Systems. The main goal of the task is to set up different music scores with different levels of complexity as test cases and a set of rules and to use in the validation of OMR systems.

In music reading, as in other applications, it is difficult to scale up a small working prototype into a large and complete system. A system based on syntactic methods with a grammar could be easily manageable if it uses for example 30 production rules, but if this number is elevated to hundreds of production rules may be the system would become difficult to manage. Another example is found in passing from a system that manages monophonic music to an another one managing polyphonic music. In order to realise that, one of the possible requirement may be to redesign the whole system.

The high diversity of scores due to the high possibility to combine the music notation impacts in deep on the project and develops of an OMR system and on the possibility to create an architecture that can be scale up.

Finally, another important open problem is how to solve the problem of image noise for example: when a primitive is expected to be connected for extraction purposes but it is actually broken into more fragments of image or two objects touch each other crating a unique unknown symbol.

7 REFERENCES

Anstice J., Bell T., Cockburn A. and Setchell M. (1996). The Design of a Pen-Based Musical Input System. OzCHI'96: The Sixth Australian Conference on Computer-Human Interaction. Hamilton, New Zealand. 24-27 November, 1996. pages 260-267. IEEE Press.

Bainbridge, D. (1991). Preliminary experiments in musical score recognition. Department of Computer Science. The Kings Buildings, Mayfield Road, Edinburgh, GB, University of Edinburgh.

Bainbridge, D. (1996). Optical music recognition: A generalised approach. Second New Zealand Computer Science Graduate Conference.

Bainbridge, D. (1996), An Extensible Optical Music Recognition System, in the Australasian Computer Science Conference (Melbourne), pp. 308-317.

Bainbridge, D. (1997). Extensible optical music recognition. Christchurch, New Zealand, University of Canterbury.

Bellini, P., Fioravanti, F., & Nesi, P. (1999) Managing Music in Orchestras, IEEE Computer, Sept., 1999.

Bellini, P., I. Bruno, I., Nesi, P. (2001), "Optical Music Sheet Segmentation", Proceedings of the 1st International Conference of Web Delivering of Music. Florence: IEEE press.

Blostein, D., & Baird, H. S.. (1992). A critical survey of music image analysis. In Structured Document Image Analysis, ed. H. S. Baird, H. Bunke and K. Yamamoto, 405-34. Berlin: Springer-Verlag

Carter, N. P. (1989), Automatic Recognition of Printed Music in Context Electronic Publishing, Ph.D thesis, University of Surrey, February.

Carter, N. P. and R. A. Bacon (1990). Automatic recognition of music notation. Proceedings of the International Association for Pattern Recognition Workshop on Syntactic and Structural Pattern Recognition: 482.

Carter, N. P. (1992). A new edition of Walton's Façade using automatic score recognition. Proceedings of International Workshop on Structural and Syntactic Pattern Recognition: 352-62.

Carter, N. P. (1992). Segmentation and preliminary recognition of madrigals notated in white mensural notation. Machine Vision and Applications 5(3): 223-30.

Carter, N. P. (1993). A generalized approach to automatic recognition of music scores, Department of Music, Stanford University.

Carter, N. P. (1994). Conversion of the Haydn symphonies into electronic form using automatic score recognition: a pilot study. Proceedings of SPIE 2181: 279-90.

Carter, N. P. (1994). Music score recognition: Problems and prospects. Computing in Musicology 9: 152-8.

Choi, J. (1991). Optical recognition of the printed musical score. M.S. Thesis, Electrical Engineering and Computer Science, University of Illinois at Chicago.

Cooper, D., Ng, K. C. and Boyle, R. D. (1997). MIDI Extensions for Musical Notation: Expressive MIDI. In E. Selfridge-Field (Ed.), Beyond MIDI - The Handbook of Musical Codes. (pp. 402-447). London, UK:The MIT Press.

Coüasnon, B., and J. Camillerapp. (1995). A way to separate knowledge from program in structured document analysis: Application to optical music recognition. International Conference on Document Analysis and Recognition: 1092-7.

Droettboom, M., I. Fujinaga and K. MacMillan (2002). Optical music interpretation. Proceedings of the Statistical, Structural and Syntactic Pattern Recognition Conference.

Droettboom, M., K. MacMillan, I. Fujinaga, G. S. Choudhury, T. DiLauro, M. Patton and T. Anderson (2002). Using Gamera framework for the recognition of cultural heritage materials. Proceedings of the Joint Conference on Digital Libraries.

Fujinaga, I. (1988). Optical music recognition using projections. *M.A*: Thesis.

Fujinaga, I., (1997) Adaptive Optical Music Recognition, Ph.D Dissertion. McGill University, Montreal, CA.

Fujinaga, I. (2001). Adaptive optical music recognition. 16th Inthernational Congress of the International Musicological Society (1997). Oxford, Oxford University Press

Luth, N. (2002), Automatic Identification of Music Notation, Proceedings of the 2nd International Conference of Web Delivering of Music. Darmstadt (Germany): IEEE press.

Marinai, S. & Nesi, P. (1999), Projection Based Segmentation of Musical Sheet Proc. of the 5th Intern. Conference on Document Analysis and Recognition, ICDAR'99, IEEE press, IAPR (International Association on Pattern Recognition), Bangalore, India, pp.515-518, 20-22 September 1999.

Miyao, H. and R. M. Haralick (2000). Format of ground truth data used in the evaluation of the results of an optical music recognition system. IAPR Workshop on Document Analysis Systems.

Ng E., Bell T., Cockburn A. (1998). Improvements to a Pen-Based Musical Input System. *OzCHI'98: The Australian Conference on Computer-Human Interaction. Adelaide, Australia. 29 November to 4 December, 1998.* pages 239--252. IEEE Press.

Ng, K. C. and R. D. Boyle (1992). Segmentation of music primitives. BMVC92. Proceedings of the British Machine Vision Conference: 472-80.

Ng, K. C., and R. D. Boyle. (1994). Reconstruction of music scores from primitive Sub-segmentation: School of Computer Studies, University of Leeds.

Ng, K. C., and R. D. Boyle. (1996). Recognition and reconstruction of primitives in music scores. Image and Vision Computing 14 (1): 39-46.

Ng, K. C. (2002). Optical music analysis: A reverse engineering approach. EVA 2002 Florence, Italy

Ng, K. C. (2002). Music manuscript tracing. Graphics Recognition: Algorithms and Applications, Springer-Verlag. 2390: 330-342.

Ng, K. C. (2002). Document imaging for music manuscript. 6th World Multiconference on Systemics, Cybernetics and Informatics, Florida, USA.

Ng, K. C. (2002). Optical music analysis: A reverse engineering approach. EVA 2002 Florence, Italy.

Prerau, D. S. (1970). Computer pattern recognition of standard engraved music notation. Ph.D. Dissertation, Massachusetts Institute of Technology.

Ross, T. (1970), The Art of Music Engraving and Processing, Hansen Books, Miami.

Roth, M.(1994). An Approach to Recognition of Printed Music, tech. rep., Swiss Federal Institute of Technology, ETH Zurich, Switzerland.

Rumelhart, D. E., and J. L. McClelland (1986), Parallel Distributed Processing: Exploration in the Microstrure of Cognition, Vol. 1 Cambridge University Press.

Selfridge-Field, E. (1994), "Optical Recognition of Musical Notation: A Survey of Current Work," Computing in Musicology, Vol. 9, 1993-4, pp. 109-145.

Selfridge-Field, E. (Ed.). (1997). Beyond MIDI - The Handbook of Musical Codes. London, UK:The MIT Press.

Suen, CY and Wang, PSP (1994), Thinning Methodologies for Pattern Recognition, Series in Machine Perception and Artificial Intelligence, Vol 8, World Scientific, 1994.

Biographies

Pierfrancesco Bellini is a contract Professor at the University of Florence, Department of Systems and Informatics (Italy). His research interests include object-oriented technology, real-time systems, formal languages and computer music. Bellini received a PhD in electronic and informatics engineering from the University of Florence (Italy), and worked on MOODS, WEDELMUSIC, IMUTUS, MUSICNETWORK projects of the European Commission.

Ivan Bruno is a PhD candidate in software engineering and telecommunication at the University of Florence (Italy). His research interests include optical music recognition, audio processing, computer music, object-oriented technologies and software engineering. He worked on WEDELMUSIC, VISICON, IMUTUS, MUSICNETWORK projects of the European Commission.

Paolo Nesi is a full professor at the University of Florence (Italy), Department of Systems and Informatics. His research interests include object-oriented technology, real-time systems, quality, system assessment, testing, formal languages, physical models, computer music, and parallel architectures. Nesi received a PhD in electronic and informatics engineering from the University of Padova (Italy). He has been the general Chair of WEDELMUSIC conference and of several other internal conferences. He is the co-ordinator of the following Research and Development multipartner projects: MOODS (Music Object Oriented Distributed System, http://www.dsi.unifi.it/~moods/), WEDELMUSIC (WEB Delivering of Music Score, www.wedelmusic.org), and MUSICNETWORK (The Interactive Music Network, www.interactivemusicnetwork.org). Contact Nesi at nesi@dsi.unifi.it, or at nesi@ingfi1.ing.unifi.it.