"Efficient Dynamic Computer Simulations of Robotic Mechanisms"
Journal of Dynamical System, Measurments and Control, Vol. 104,
pp 205 - 211, 1982.
DC Motors, Speed Control Servo Systems
An Engineering Handbook
prepared by Electro - Craft Corporation, USA

MIDISCAN

the program for reading and processing musical notation

WLADYSLAW HOMENDA(*)(**), ALFREDO CRISTOBAL(**), MERCEDES ARELLANO(**)

Department of Electronics and Telecommunication, Applied Physics Division
Centro de Investigacion Cientifica y de Education Superior de Ensenada
Km. 107 Carretera Tijuana/Ensenada, Ensenada, B.C., Mexico 22860

ABSTRACT

The paper presents problems related to automated recognition of printed music notation. Music notation recognition is a challenging problem in both fields: pattern recognition and knowledge representation. Music notation symbols, though well characterized by their features, are arranged in elaborated way in real music notation, which makes recognition task very difficult and still open for new ideas. On the other hand, the aim of the system, i.e. application of acquired printed music into further processing requires special representation of music data. Due to complexity of music nature and music notation, music representation is one of the key issue in music notation recognition and music processing. The problems of pattern recognition and knowledge representation in context of music processing are discussed in this paper. MIDISCAN, the computer system for music notation recognition and music processing, is presented.

Keywords: music notation recognition, knowledge representation, music representation, MIDI format.

1. INTRODUCTION

There are many ways in which computers have been involved in the world of music. One is the score edition, where musicians can develop a score using an editor to produce a digital output file. This strategy might be a good idea, but it is a non-traditional way to develop a score. Another computer application is music processing that can help musicians in music creation process: automatic composition of music, analysis of musical style and so on.

(*) ***On leave from:*** Institute of Mathemaics, Warsaw University of Technology, Warsaw, Poland

(**) ***Correspondence to:*** Wladyslaw Homenda, Mercedes Arellano, Alfredo Cristobal,
- CICESE Research Center / Applied Physics, P.O.Box 434944 San Diego, CA 92143-4944, USA
  e-mails to: {homenda,arellano,asalas}@cicese.mx
- Wladyslaw Homenda, Instytut Matematyki, Politechnika Warszawska, pl. Politechniki 1, 00-661 Warszawa, Poland
  e-mails to: homenda@plwatu21.bitnet

In this paper we are going to deal with automatic recognition of printed music and its conversion to some digital forms. Here, musicians can continue developing the scores as they used to, and introduce these printed scores into a computer to convert them to a digital output file.

Over the last years, efforts have been focused on automatic musical notation recognition. Several research centers and universities have developed many programs to recognize printed music, c.f. (Computing 1994, Fujinaga, 1988, Itagaki, et al., 1992, Kato et al., 1992). Although some results of these works are quite good, improvement is needed. The obstacles that developers of these systems have had to face seem to have increased and most of them are too complex to be solved by classical methods in comparably short execution time of the system. To show this fact, a couple of the most important problems are listed next:

1. Problems related with the nature of music represented in printed form: here composers feel free to invent new symbols or contractions of them and this freedom has increased the number of musical dialects around the world, most of which are not standardized.
2. Problems related with recognition technology: in this area it is common to find symbols with different sizes and shapes, symbols that overlap one over the other, distortions of the image, etc.

These difficulties have not permitted us to make a universal printed music recognizer system. So, this area is still open to ideas that can improve the methods already applied or create new ones to have a better recognition process.

Some problems mentioned above were solved by MIDISCAN. MIDISCAN is a semi-automatic printed music recognition system which accepts scanned scores and produces a playable file (MIDI file). Such file can be played on synthesizers, computers, and other equipment that can accept MIDI files. Unfortunately, the conversion is not so direct, because of that MIDISCAN represents the recognized scores in an intermediate format called MNOD, cf. (Homenda, 1995).

2. MUSIC NOTATION RECOGNITION - PROBLEM STATEMENT

Music notation can be interpreted as a language allowing to document musical information in a legible, archival form. Recognition of music notation can of course be modeled as mappings between the printed notation and the information it represents, cf. Figure 1, cf. (Blostain & Baird, 1992). This general formulation of the task of music notation recognition does not reflects conceptional and technical problems developers of recognition systems are faced.

First of all, music notation does not have a universal definition. Although attempts to codify printing standard for music notation have been undertaken, c.f. (Ross, 1970, Stone 1980), in practice composers and publishers feel free to adopt different rules and invent new ones. Though most of scores keep the standard, they still can vary in details. Moreover, music notation, as a subject of human creative activity, constantly develops and probably will be unrestricted by any codified set of rules. Thus, it may not be possible to built a universal recognition system accepting all dialects of printed music notation. Furthermore, the nature and structure of music, even that printed one, is much more complicated than the structure of a text, so representation of music is comparably much more difficult than, for example, representation of printed text. As the result, comparing music notation recognition with text recognition, there are several applicable computer systems for automated text recognition with considerably high rate of recognition (which can be calculated as ratio of recognized characters to all characters in the text). As to music notation recognition, commercial systems are still very rare despite the fact that several research systems of music notation recognition have already been developed, c.f. (Blostain & Baird, 1992, Fujinaga, 1988, Itagaki, et al., 1992, Kato et al., 1992).

On the other hand, it is not quite clear, what *music information* - the goal of recognition task means, cf. Figure 1. In fact, meaning of this term may depend on the application of recognition task. For some applications, e.g. from musician point of view, a complete solution to the music recognition problem is the specification of: which notes are present, what order they are played in, their time values or durations, and volume, tempo, and interpretation. Thus, in this case the scanning units are generally quite small, for example, a clef, one or more key signature, accidentals, a time signature, note heads in a chord, and following music symbols are not included in scanning units: staff lines, beams, slurs ties brackets, text, and crescendo or

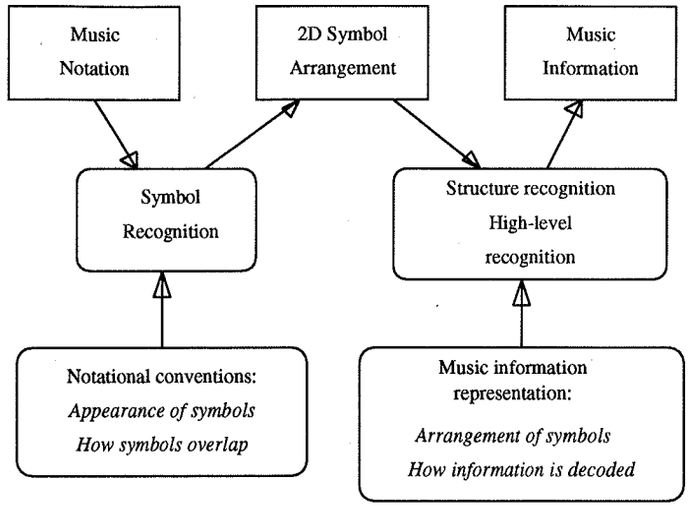

Figure 1. Music notation recognition

decrescendo signs. However, this set of information is not sufficient for producing parts from a score. As it was stated in (Blostain & Baird, 1992), these digital files can be used for many purposes:

1. To adapt existing works to other instrumentations.
2. To convert existing scores to Braille to aid blind musicians.
3. To read works in old editions and produce a new printing.
4. To store scores in digital formats and construct a kind of music library which one can see and hear the music with multimedia equipment, or even change the instruments and features of the score.
5. To print newly written music automatically.
6. To analyze musical structures and styles.

Finally, it should be stated, that designing and developing universal recognition system seems to be currently beyond the horizon. A reasonable strategy for developing recognition system should be based on the application of such a system. This assumption would be useful in restriction both, the class of acceptable notation dialects as well as the format of music information acquired. For example, if recognition of old notations and producing a new printings is considered as a goal of recognition task, it will be effective in fixing the set of rules defining both, old and new notations.

3. MIDISCAN - RECOGNITION PROGRAM

3.1. Overview of the program

The transformation of data given as printed music into playable MIDI format is the main idea of MIDISCAN software. This transformation is intended to be as far automated as possible. Unfortunately, at present stage of development of both fields: methods of recognition of printed music notation and representation of music data, it is impossible to built fully automated system which could recognize music notation and create playable music data correctly performed with the electronic instrument. Errors appearing in recognition process, even a few of them, cause that correction of recognized music notation is necessary

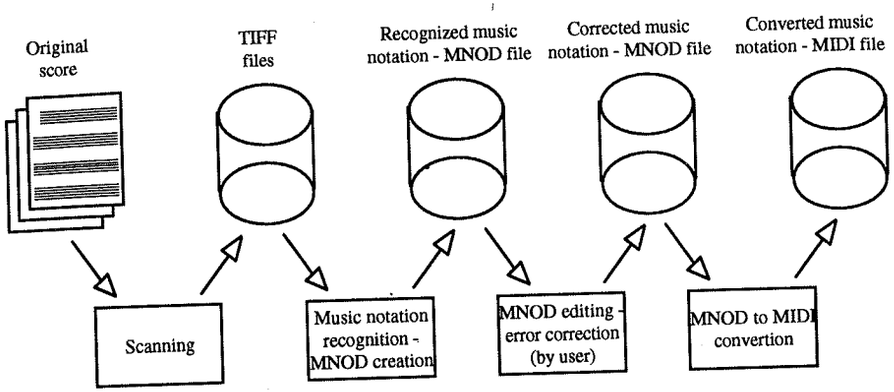

Figure 1.

before it is performed with the instrument. Thus, a correction of acquired music data is necessary. The correction may be done on output, playable data (e.g. MIDI format) or in the middle of the road: before acquired data are converted to playable format.

The first option, correction of a final MIDI format, is regarded as to be less convenient than correction of recognized data before MIDI conversion. The reason is quite clear, for example, key signature correction needs only a few operations before MIDI conversion while pitches of many notes should be corrected if wrong key signature was assumed in MIDI conversion. But this assumption requires the music notation to be represented in a format allowing for editing, correction and, then, MIDI conversion. This assumption implied the necessity of acquired data to be represented in the special intermediate format called MNOD (Music Notation Object Description).

The idea of MIDISCAN is outlined in Figure 1.

Once the score is scanned and stored on a hard disk as a set of TIFF files, MIDISCAN can start its task, i.e. recognition of music notation, writing acquired data as a MNOD file format, editing MNOD format and presenting it to the user for corrections (if necessary), conversion of MNOD file format to MIDI file format.

3.2. Score definition.

Two types of music notation may be processed by MIDISCAN: ensemble and part scores. MIDISCAN does not detect the type of a processed score, it must be defined by user. In case of ensemble score, i.e. score with all voices linked together into systems, user only chooses the score type and defines the sequence of pages of the score (i.e. pages of the original notation scanned and stored in the form of TIFF files). When part score is processed, i.e. score with voices separated from each other, the number of voices and the number of pages for every voice must be described before the sequence of TIFF files is defined. The sequence of pages must keep the order of respective pages of the score.

3.3. System location.

Automated stave location is performed for the whole score before recognition process is started. Simultaneously, the structure of the score is detected, i.e. the way systems are located. The number of staves in every system is detected for the whole score of ensemble type and for every part of the score of part type. Let us recall that the term "system" is used here in the meaning of:

- all staves performed simultaneously and
- joined together in the score or part of the score.

The score for voice and piano with three staves in the system is an example of ensemble type of score. If voice part is missing in the first and second system, the score has two irregular systems with two staves and the other systems are regular with three staves.

Part type score: only scores with constant number of staves in systems for every part are accepted. Up to 3 staves per system are permitted (e.g. organ part of the score consists of 3 staves, other voices consist of no more staves). A score for string quartet with separated voices is an example of part type score with four parts and one stave in part system for every part.

3.4. Stave location.

Stave location algorithms are based on horizontal projections. Theoretically, for non-distorted and non-skewed images, methods based on projections should be fully effective: high pass filtering gives clear image of a stave as five equidistant picks with the height equal to the length of the stave.

Unfortunately, in real images staff lines are distorted too much to give so clear projections, especially when projection is done for page width or even for wide region. Scanned image of a sheet of music is often skewed, staff line thickness differs for different lines and different parts of stave, staff lines are not equidistant and are often curved, especially in both endings of the stave, for ensemble type of score staves may have different sizes, etc. These problems cause that projections done in wide region are useless for stave location.

On the other hand, projections in narrow region are distorted by notation objects such as ledger lines, dynamic 'hairpin' markings, slurs, note heads, etc. Thus, simple filtering does not give information sufficient for stave location. In MIDISCAN program, horizontal projections in several narrow regions are analyzed for obtaining vertical placement of the staves. Once vertical placement of the stave is located, both endings of it are detected. Projections in relatively narrow regions are used in stave endings location task. Iterative process based on classical bisection algorithm is employed in this process - the process starts in the middle of the stave and then goes in the directions of both endings of the stave.

An advantage of applied methods is that distortions such as non-equidistant staff lines, varying thickness of staff lines, skewed images (skew angle up to 10-15 degrees) and stave curvature, do not influence the process of stave location as well as the process of notation recognition in an observable way.

It is worth mentioning that stave location process is not fully automatic. In some cases, especially for low quality images or for very dense notations, automatic stave location is mistaken and must be corrected manually. Fortunately, the program is able to detect problems it has and only if automated correction of given problem is not possible, location process is suspended unless user fixes the problem.

3.5. Recognition.

Music notation is built around staves. The position and size of symbols are restricted and determined by the stave. So, location and identification of staves must be the first stage of recognition process. Having staves and systems located, the program starts fully automated recognition of music. Recognition is done for every stave, and then, after notation is recognized and analyzed for given stave, the acquired music data are filled into MNOD format.

Recognition strategy can be seen as three step process:

- object location,
- feature extraction
- classification.

The first step - object location - is aimed at preparing a list of located symbols of music notation. Bounding boxes embodying symbols to be recognized are defined for located symbols. The process of object location is based on projection analysis. First, the projection of whole region of given stave on OX axis is processed. The location process is mainly based on the analysis of a derivative the projection, c.f. (Fujinaga, 1988,). The derivative analysis gives the first approximation of object location. Then, for every rough located object, a projection on OY axis is analyzed to obtain vertical location of the object and to improve its horizontal location. The most important difficulties are related to objects which cannot be separated by horizontal and vertical projections. Also wide objects as slurs, dynamic 'hairpin' signs, etc. are hardly located.

The next two steps of recognition process are based on the list of located objects. both steps: feature extraction and classification overlap each other and it is not possible to separate them. Feature extraction starts from extracting the simplest and most obvious features as height and width of the bounding box containing given

object. Examining of such simple features allows for classification only in a few cases. In most cases additional features must be extracted and context analysis must be done. The extraction of features is based on filtering of projections in the bounding box, analysis of chosen columns and rows of pixels, etc. Several classification methods are applied for final classification of object including context analysis, decision trees, and syntactical methods.

Only limited set of music notation objects can be processed in MIDISCAN. This set includes notes, chords, rests, accidentals, clefs, bar lines, ties, key signatures, time signatures, change of key and time signature. Rhythmic grouping can also be inserted into acquired music data, though they are not recognized. Other symbols are going to be recognized and represented in future versions of the program.

Recognition confidence depends on many features of a printed score: font, printing quality, image quality, notation density, etc. The obvious, general rule may be formulated that the higher quality of printed music and scanned image, the higher the rate of recognition.

Recognition efficiency of MIDISCAN program may be estimated as 95% for good quality of image and 80-85% for low quality of image. Precise calculation of recognition rate is strictly related to applied calculation method (c.f. note in chapter 1.), but the scope of the paper does not allow to discuss extensively this problem.

4. MUSIC REPRESENTATION

Making an analogy between computer processing of a printed text and a printed music, it is worth underlining that, as to text processing, design and implementation of widely accepted data representation is considerably easy. Rich Text Format (RTF) format is an examples of such a representation.

Music data representation is far more difficult and, up to now, there is no universal representation widely used and commonly accepted. Music data formats used in computer systems are intended more for particular task rather than for common use. Even if a particular format is widely spread and commonly applied, it is used for special tasks rather than for any purpose. For example, "MIDI (Music Instrument Digital Interface) data format was established as a hardware and software specification which would make it possible to exchange information between different musical instruments or other devices such as sequencers, computers, lighting controllers, mixers, etc. This ability to transmit and receive data was originally conceived for live performances, although subsequent developments have had enormous impact in recording studios, audio and video production, and composition environments" (c.f. (MIDI, 1990)). Nevertheless, MIDI format, as performance oriented, is not an universal one. E.g., it is very difficult or even impossible to represent in MIDI format graphical features related to music notation. On the other hand, format used by notation programs are notation oriented and cannot be easily used as universal format of music representation.

Because recognition confidence in MIDISCAN is not satisfactory enough for direct conversion of recognized music notation into MIDI format, it is necessary to edit acquired data, correct it and then convert into MIDI format. These tasks need acquired data to be stored in some form suitable for both editing and conversion. For this particular aim, a special format was developed. This format is called MNOD format (Music Notation Object Description). It plays the role of an intermediate format between printed music notation and playable MIDI format.

It was assumed that recognized music notation would be edited for correction in the form compatible with original score. This assumption allows for simultaneous displaying of original score and acquired data. It makes editing and checking correctness of acquired data as easy as comparing two notations which should be identical. All differences can be easily corrected, even if user is not familiar with music and music notation.

MNOD format applied in MIDISCAN for data representation meets all these requirements. Its main features give the possibility of data interpretation form both perspectives: notation oriented and performance oriented, that makes music data easily accessible for both purposes: editing and MIDI conversion. Unfortunately, MNOD format applied in MIDISCAN does not represent all commonly used notational symbols. Only symbols edited in MNOD editor (see section 3.5. for the list of the symbols) are represented in the format.

MNOD format is structured hierarchically. The levels in this hierarchy reflects data accessibility for above purposes. MNOD format may be regarded as two different structures permeated each other.

The notation oriented structure may be seen in the following levels:

- score,
- page,
- stave on page,
- objects of music notation

while performance oriented structured may be outlined as below:

- score,
- part of score (applicable to the scores of part type),
- system,
- stave in system,
- vertical events,
- objects of performed music.

This comparison gives only general view on differences between those attempts. Extended discussion on this topic is out of the scope of this paper.

It is worth mentioning that levels of both structures differ in their meaning even if they are called similarly. E.g., notation structure reflects sequential organization of staves on page while performance structure organizes staves according to the systems of the score, regardless of their order on the page. Similarly, objects of music notation are seen differently in both structures. For example, in notation structure, notes must have such features as their position on the page, while their relative position to each other is unimportant. On the other hand, relative placement of notes is significant for performance structure, while their position on the page is of less importance.

The approach to music representation applied in MIDISCAN is flexible and easy to control: displaying data in graphical form on the screen, converting music data to MIDI format and independent music data processing.

Acquiring contextual information from recognized music notation and checking correctness of recognized notation is the aim of independent data processing applied in the program. This processing allows for locating of pickup/close-out measures, analyzing voice lines, verifying bar lines or change of key and time signature consistency, monitoring data integrity. In general, the possibility of independent music data processing considered in wider context creates a lot of research problems related to knowledge representation, that makes music representation interesting from more general point of view.

5. EXPERIMENTAL RESULTS

MIDISCAN was developed and ported on PC 386 and compatible computers in WINDOWS environment. Hardware requirements are similar as for WINDOWS environment. Digitizing resolution of 300 dpi is suggested for acquired binary images of the size A4. It gives TIFF file of the size approximately equal to 1MB. Both orientation of a page: portrait and landscape is accepted. Pages/files of bigger size can be effectively processed on computers with at least 8 MB RAM. Processing time, which is different for different notations, depends on resolution of the image and on notation density.

For example, it took about 75 minutes to recognize first movement of Bach's Brandenburg Concerto no. 6. The experiment was done on PC 386 / 8 MB RAM, 33 MHz clock. The score consisted of 15 pages of A4 format, 18 staves per page, 6 staves per system, scanned at 300 dpi resolution. The printing quality of the score was rather low, music notation density - medium. The recognition efficiency for this score was considerably high. Three staves were incorrectly localized and, what was important, system detected all three errors and signalized them to the user. Estimated recognition rate overheaded 90%. It was calculated as the ratio of missing or mistaken objects or their features, to all objects.

Another experiment done for Beethoven's "Fur Elise" gave similar result in speed processing (i.e. comparable recognition time per stave), but the rate of recognition was higher due to higher printing quality of tested score. All 36 staves of this score were located accurately, score structure was also correctly detected, estimated recognition rate was over 95%.

More experiments done for different scores confirmed that context information implied form music notation, such as pickup and close-out measures and voice lines analysis and location, was correctly and comparably easily acquired.

MNOD to MIDI conversion was very fast and took about one minute for Bach's score and a few seconds for that of Beethoven's.

To compare similar parameters of other recognition systems see (Blostain & Baird, 1992, Fujinaga, 1988, Itagaki, et al., 1992, Kato et al., 1992)..

6. CONCLUSIONS

The paper describes MIDISCAN - a recognition system for printed music notation. Music notation recognition is a challenging problem in both fields: pattern recognition and knowledge representation. Music notation symbols, though simple and well characterized by their features, are arranged in sophisticated and confusing way in real music notation, which makes recognition task highly difficult and still open for new ideas, as for example, fuzzy sets application in skew correction and stave location. On the other hand, the aim of the system: conversion of acquired printed music into playable MIDI format, requires special representation of music data. This representation should be adequate for both: source - notation oriented, and target - performance oriented music data. Regarding further development of this system, the effort should be put on following tasks: improving recognition methods, extending class of recognized objects, improving and extending music representation format.

See also (Aikin, 1994, Computing, 1994, Homere, 1994, Lindstrom, 1994) for reviews of MIDISCAN.

REFERENCES

Aikin J. (1994): Musitek Midiscan: Sheet music recognition software, *Keyboard*, March 1994, pp. 136-140.

Blostein D. & Baird H. S. (1992), A Critical Survey of Music Image Analysis, *Structured Document Analysis, H.S.Baird, H.Bunke, K.Yamamoto (Eds)*, pp. 405-434, Springer-Verlag, 1992

Computing (1994), *Computing in Musicology*, vol. 9, 1993-1994, pp. 109-166.

Fujinaga I. (1988), Optical music recognition using projections, *Master's thesis*, McGill University, Montreal, Canada, September 1988.

Homenda W. (1995), Optical pattern recognition for printed music notation", *Proc. of the SPIE's Intern. Symp. on Optical Pattern Recognition VI*, April 1995, Orlando, Florida.

Homere S. (1994), Musitek Midiscan: Donnez la parole a vos partitions, *Keyboards Magazine*, 1994.

Itagaki T. et al. (1992), Automatic Recognition of Several Types of Music Notation, *Structured Document Analysis, H.S.Baird, H.Bunke, K.Yamamoto (Eds)*, pp. 405-434, Springer-Verlag, 1992

Kato H. & Inokuchi S. (1992), A Recognition System for Printed Piano Music Using Musical Knowledge and Constraints, *Structured Document Analysis, H.S.Baird, H.Bunke, K.Yamamoto (Eds)*, pp. 405-434, Springer-Verlag, 1992

Lindstrom B. (1994), Musitek Midiscan: Optical Music Recognition is finally a reality, *Electronic Musician*, February 1994, pp. 151-154.

MIDI (1990), *MIDI 1.0, Detailed Specification*, Document version 4.1.1, February 1990.

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

Stone K. (1980), *Music Notation in Twentieth Century: A Practical Guidebook*, W.W.Norton & Co.,New York, 1980.

A computer-aided object-oriented analysis

Didier GUIGUE
DEPARTAMENTO DE MÚSICA/CCHLA
UNIVERSIDADE FEDERAL DA PARAÍBA
Home add.:Rua Antônio P. da Rocha, 90
58045 - 380 João Pessoa - PB - BRAZIL
e-mail: didier @ brufpb.bitnet

Most of analytic tools developped during this century are exclusively devoted to pitch structure. It is however an evidence that, at least since Debussy, forming and linking complex abstract sound-objects are too decisive composing features. This paper exposes the basis of a computer-aided information and evaluation method which may bring out significant data for an objective analysis of the formal functions of these sound-objects. The Prelude La Cathédrale engloutie , by Debussy, exemplifies the way this method can contribute to high-level structure understanding.

1. The goal

Most of analytic tools developped during this century are exclusively devoted to the basic low-level components of music writing: pitches (or pitches classes). It is however an evidence that, at least since Debussy, forming and linking complex abstract sound-objects are too decisive composing features. Musicology and music theory generally approach object-oriented formal structures in a empirical way, asserting, without demonstrating, that they depend on the laws structuring pitches, considered as the most important, or, indeed, the *unique* formal dimension. This paper exposes the basis of a computer-aided method of information and evaluation of the components of the musical score which may bring out all the necessary data for an objective analysis of the formal functions of this sound-objects. Restricted — at least for the moment — to the piano literature, it is aided by a set of algorithms realized inside the *Patchwork* environment developped at the IRCAM Institute, Paris [1], as part of a PhD research on the 20th Century piano music.

2. The score reading

The source score has to be previously segmented in a sequence of units, basically defined by an homogeneous content; this homogeneity may be mainly due to the invariance of (most of) its elements, or to continuous teleological changes. The segmentation process is reached threw a continuity break scanning of the score, at any desired level: macro-formal stops (rests, fermatas), interruptions of pedal sustains and/or phrase slurs, continuity breaks in intensities, registers, rhythmical outlines, densities, a.o. These rupture criteria do not depend on the thematical/periodical structure (phrases and other time-groupings), although coincidences may be founded, especially at higher hierarchical levels. Nevertheless, independence is fundamental to check connections between the object-level organization and lower ones. Each unit is thus a sole (written) sound-object [See fig.1 for a segmentation sample of the last bars of Debussy's La Cathédrale engloutie]. Once segmented, the score is stored in several Midi files (one per segment, i.e. one per sound-object) to be exported to Patchwork [see, in the sample *Patchwork* window fig.3, the Midi-import patches and Midi-object storage].

3. The sound-objects evaluation

Each object is analysed by a set of specific algorithmic patches, we name *interpreters*, connected in a hierarchically structured "frequency-modulation" bidimensional network [fig.2], where qualitative evaluations (in round boxes) modulate quantitative ones (square boxes), producing "synthetical" interpreters at a subsequent level.

The "S" dimension is exclusively concerned with achronic aspects of the musical writing: inside the "*SPACE* " sub-group of evaluation patches, the *AMBITUS* single interpreter calculates the range-filling rate of