Automatic Segmentation of Musical Flows: a Rational Agents Approach

Ernesto Trajano*
Departamento de Sistemas e Computação - CCT/UFPB
Laboratório de Inteligência Artificial - LabIA
Cx. Postal 10.090, 58.109-970 Campina Grande - PB - Brazil
etrajano@dsc.ufpb.br

Didier Guigue
Departamento de Música - CCHLA/UFPB
Joao Pessoa - PB - Brazil
dguigue@openline.com.br

Edilson Ferneda
Departamento de Sistemas e Computação - CCT/UFPB
Laboratório de Inteligência Artificial - LabIA
Cx. Postal 10.090, 58.109-970 Campina Grande - PB - Brazil
edilson@dsc.ufpb.br

Abstract

Segmentation is one of the main tasks within the musical analysis. We can say that segmentation is the partition of a musical flow into homogeneous segments, accordingly to certain criteria. These criteria are particular to each analytical model. This paper discusses some questions related to a musical analysis methodological model based on the concept of sonic object. In this model, a sonic object may be defined as the combination and interaction of multiple musical components and a gap in the continuity of at least one of these components may identify, in theory, a new structural/logical unit (a new sonic object). This paper is also concerned with the automatization of this process. We propose the use of the concept of rational agent for the segmentation of musical flows. Although the research is still in an early stage, we present some insights concernig the use of this concept in the segmentation of musical flows.

*Master Degree Student in Computer Science (UFPB - Campina Grande), supported by CAPES

1 Introduction

Many works in the artificial intelligence field turn evident, explicitly or implicitly, the importance of computacional environments for supporting researchers in the organization of experimental data and in the generation, avaliation, refutation and revision of theoretical knowledge. Our research program with the GIA (Grupo de Inteligência Artificial da Universidade Federal da Paraíba, Brazil) is concerned with the use of the Automatic Learning applied to discovery-aiding systems. We work on the anatomy of a rational agent, making evident the ensemble of data, heuristics and mecanisms which makes possible that the rational agent completes simple tasks in a very specific domain: geometry [Fer92]. Given the restrictions in the modelling of such an agent, his concepts and the resulting knowledge may be erroneous. As the domain of geometry is very well formalized, the human agent is able to determine counter-examples for the rational agent. Once the rational agent must not produce exact knowledge, but an arguable one, the wrong conclusions and concepts can be corrected through a revision of that agent's knowledge.

We are looking for the conception of an environment able to deal with some problems of the computer-aided musical analysis field. In particular, our work is about the conception of a rational agent that would be able to deal with the automatic segmentation of musical flows[1].

2 Musical Analysis and Segmentation

Musical analysis is a research field in the domain of musicology that seeks to understand and explain the structure of a musical piece. For the analist, a musical piece is defined as "a complex of interdependent elements articulated among each other" [Mee94]. It is the articulation of these interdependent elements, or logical units, that determines the structure, that is, the form of a musical piece. "To deal with musical form is to deal with musical analysis itself" [Mee94].

One of the first tasks when beginning an analysis of a piece is to segment it into several logical units. Mesnage defines segmentation as being "a part of the musical analysis process that consists in underlining in a piece the segments used as meaningfull entities in analysis" [Mes94]. The proceedures of segmentation can be classified in two: the ones of extrinsic base, that is, the proceedures based on concepts borrowed from out of the piece; and the ones of intrinsecal basis, that is, proceedures based, as much as possible, on elements and data found in the piece itself [Mee94].

At first sight, the segmentation process may seem to be simple. But the questions within this process are not at all trivals, even for experienced analysts. Finding fragments within a musical piece does not mean that the fragments are structurally important: the analyst must make his own decision, comparing the content and structure of the fragments. According to the music, the method, and the scope

[1]This paper is part of a Research Project coordinated by Didier Guigue and supported by the CNPQ, in which Ernesto Trajano is developing a Master degree dissertation in computer science, advised by Edilson Ferneda and Didier Guigue.

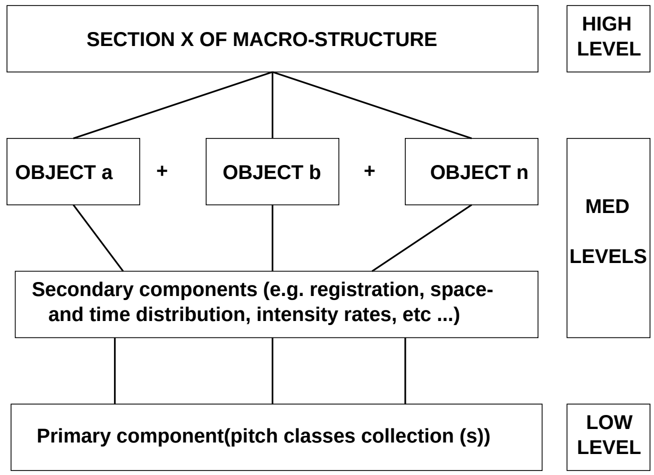

Figure 1: sonic objects and their position in the structural levels.

of the analysis, he may pay mostly attention to the low-level material, such as pitch-classes, and the way they are bring together to build melodic or harmonic structures. Or he can also take in account the so-called secundary parameters [Mey89], such as densities, textures, dynamics and others.

3 Automatic Segmentation of Musical Flows

Our approach is based on the concept of sonic object [Gui97b, Gui97a]. In this model, a sonic object is defined as the combination and interaction of multiple statistical components — grossly the same kind of components Meyer call secundary parameters [Mey89, Gui97a]. The sonic objects function as a medium-level structure, as opposed to lower-level (pitch or pitch-classes structures), and upper-level (macro-structure). Figure 1 depicts the structural levels.

The way these medium-level objects are linked is an important vector of form in 20th century music. Thus our approach may be described as object oriented. It seeks to clarify:

1. the inter-relations between the medium level of sonic objects, the lower level of pitch classes, and the upper level of macro-structure;

2. the transformational dynamics between consecutive objects, which configure a crucial aspect of the pieces formal *kinesis* and allows the form to be inferred from the succession of more or less contrasted sonic objects.

The first step of the model is to reconfigure the musical flow into a discrete sequence of logical units - the sonic objects. We assume that a gap in the structural continuity of at least one of that statistical components implies, in theory, a break in the underlying sonic continuity and, consequently, may identify a new structural/logical unit (a new sonic object) [Gui97b]. Thus, segmentation means finding the continuity gaps for each component.

Nonetheless, these components are not all equivalent, nor in their capacity to provoke gaps in the sonic continuum and, thus, orient the segmentation, nor linear interchangeable. They act on more or less global levels of the musical surface. For instance, gaps may appear as: silences, fermati, . . . ; change of the basic pulsation, global tempo or meter; global sonic changes (i.e., instrumental group in orchestral music; *pedal* change in piano music, etc.); interruptions in phrasing and articulation ; breaks in the homogeneity of some relevant components such as intensities, registers, densities, etc.

The algorithmic structure we are working on as a software program is based on the following assumptions and steps:

1. The computer is given a set of musical (e.g., Midi), graphical or sonic (audio) parameters in which structural gaps may appear, from the most generic to the most specific one. This is an example:

    (a) Macro ruptures:

        i. sound/silence parameter, the most generic, is identified in a score through silence or suspension marks and other graphical or textual indications of interruption of the sound;

        ii. macro-formal switches (changing of *tempo*, of metrical unit, etc.). etc.

    (b) Global ruptures:

        i. global change of sound: change in instrumental structure, distribution or density (commonly referred as orchestration parameters), global changes in individual sound production (action on the pedals of the piano — activation or interruption —, *sordinas* on brass instruments, "registers" for the organ, *pizzicato* and other modalities of sound production on bowed instruments, and so on);

        ii. end (or beginning) of a global articulation (generally indicated by legato slurs, or the like);

        iii. ruptures in global intensities;

        iv. ruptures in global registration.

    (c) change in the configuration of other components: relative density, distribution modes (the way the sounds are spatially or temporally distributed), directionality, articulation marks, pulsation patterns[2].

2. The user must be allowed to enter a *precision rate* for the ongoing task, according to the scope and the expected level of globality of his/her analysis. Thus, the user, globally choosing how deep and low the program will have to go down into the input file, determines the amount of data to be evaluated and returned. Obviously, this precision rate is to be considered by the program as it runs the next steps.

[2]This parameters and criteria are thoroughsly described in [Gui97a, Gui97b].

3. Each piece (or a section of it, depending on the scope of the analysis) generates its own hierarchy of relevant parameters for structural breaks criteria. That is to say, nor all the parameters may apply in all music, nor their hierarchical position remains the same. Some standard deviation and other statistic-related algorithms may help the computer to decide, before some weighting algorithm is applied to each end-selected parameter.

4. At the same time, the scanning of the file must allow the computer to define, for each relevant parameter, the minimal intervalic[3] value from which it may be considered to have a structural gap. Obviously, an interval of zero will give a truly microscopic analysis where each new value corresponds to a structural gap, which is very far from an useful information. This minimal gap value is also highly determined by the *precision rate* previously entered by the user.

5. After running on, the program outputs two main groups of data:

   (a) the resulting segmentation of the piece, in the form of a linear sequence of logical units. This output can be, for instance, a collection of small Midi or audio files.

   (b) a quantified gap weight of each unit, obtained through the synthesis of the evaluation of the observed gaps in the value lists obtained from the selected parameters. This gap weight will help the user to profile the diachronic formal structure of the piece.

The main questions in this implementation arise from making the computer decide and/or know the following:

1. For a given input - a music properly coded - and assumed that the computer is given previously a knowledge base, what are the relevant components? That is, how to segment a musical piece only from its representation and a knowledge base?

2. How to hierarquically classify these components, assumed that some of them are more active as continuity-breaking vectors than others?

3. To what extent the target of the analysis may infer some of these previous choices? In other words: how could work the user's interface to allow him/her to orient the level or kind of segmentation the computer is going to make?

4 Intelligent Agents

An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors [RN95]. Intelligent

[3]Interval here does not have the usual meanning used for musicians, that is, the distance between two notes. Here, interval means the distance between two numerical values of a list.

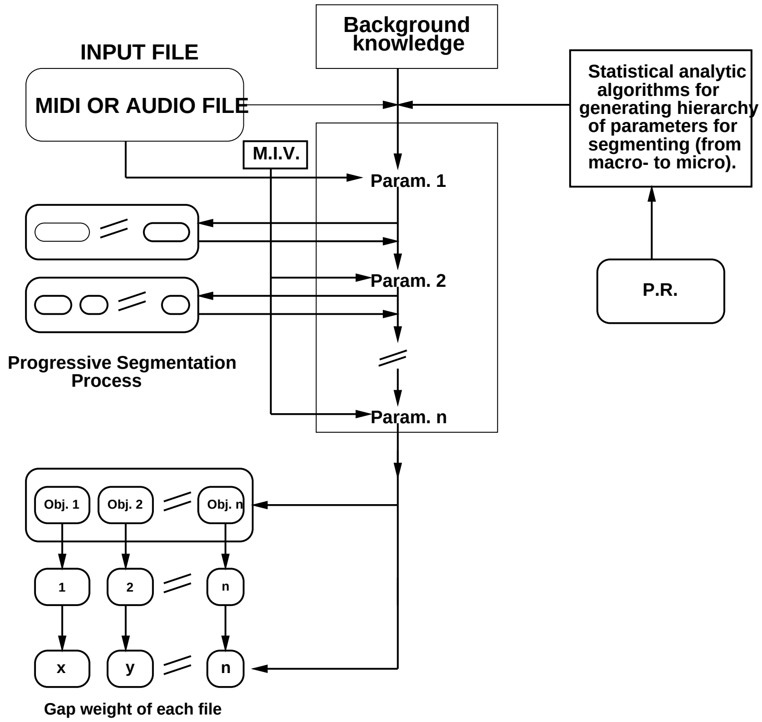

Figure 2: Description of the segmentation process. Legend: rounded boxes: input and output files; squared boxes: the kernel of the program; P.R.: user defined precision rate; M.I.V.: calculating the Minimal Intervalic Value.

agents have several characteristics. Among them autonomy is a very important one. An agent is autonomous to the extent that its actions choices depend on its own experience, that is, the autonomy of an agent is close related to the quantity of its built-in knowledge.

There exists a variety of basic agent program designs, depending on the kind of information made explicit and used in the decision process. Factors like efficiency and flexibility are taken into account, but the final design of an agent will depend on its perceptions, actions, goals and environment[4].

4.1 Rational Agents

Rational agents can be viewed as a particular type of intelligent agent. They have a special characteristic: rational agents are able to explain how they solved a given problem. In reality its rationality, its explanation capacity, is communicational: this makes the agent able to interact with other machines or even the human being [Sal97]. To be rational, an agent must respect the following facts:

1. The principle of reason: it generates the causality, since nothing exists without a reason;

[4]There is a classification of the environments and its influences in the agents final design in [RN95], pp. 45-49.

2. The principle of identity: it maintains the permanency (A is the same of A);

3. The principle of contradiction: nothing exists without its contrary.

The rational agent exists for organizing, maintaining and distributing a knowledge, working as an information server. He is conceived to help us to categorize and theorize [Sal97] and, as such, he is built through categorial structures and defined in function of big structural invariants. He has four conceptual frames:

1. A frame defining the form;

2. A frame defining the rules and

3. A frame defining the relationships between forms and rules.

The rational agents theorize through the communications made between each other: they are used to build a semi-empirical theory. These communications generate facts, hypothesis, objects, concepts, lemmas, etc. about a given domain.

5 Agents and Segmentation

The importance of applying the concept of rational agent in the automatic segmentation of musical flows resides mainly in the agents' capacity of producing an arguable knowledge. As the agent theorizes and categorizes a domain, the interaction of this agent with the analyst may lead to a more complete and coherent explanation of the domain. For example, if the agent's result - let's call pre-segmentation - seems to be incoherent with the analysts' experience, this one can ask the agent to review its knowledge and make another pre-segmentation based on this knowledge review.

The importance of this interaction and the resulting agent's knowledge review is indeed pertinent to our context. An analytical methodology, and consequently its general rules for segmentation, assumes that in a given context - a tonal or non-tonal system, for instance - certain things occur. But it is known that art, and music in particular, is not a static and pre-defined context and things may go in a not expected way. A rational agent designed for the segmentation will certainly face situations not represented in its knowledge of the domain, in its theory. One way to solve this problem is to permit that the agent interact with the user and review its theory.

Figure 3 depicts the agents interactions in the system we propose. It introduces a new element: the work group. The work group has an epistemical activity: he formulates rules and prescriptions about a micro-world (in this case a musical piece). His task is to determinate the terms that permit the rational agent to reason and act in this world [Sal97].

Basically, the process of segmentation is done in the following way: the user gives the agent, and eventually the work group, its previous analysis of the piece[5].

[5] This previous analysis is the description of the parameters that seems to be more relevant for the piece.

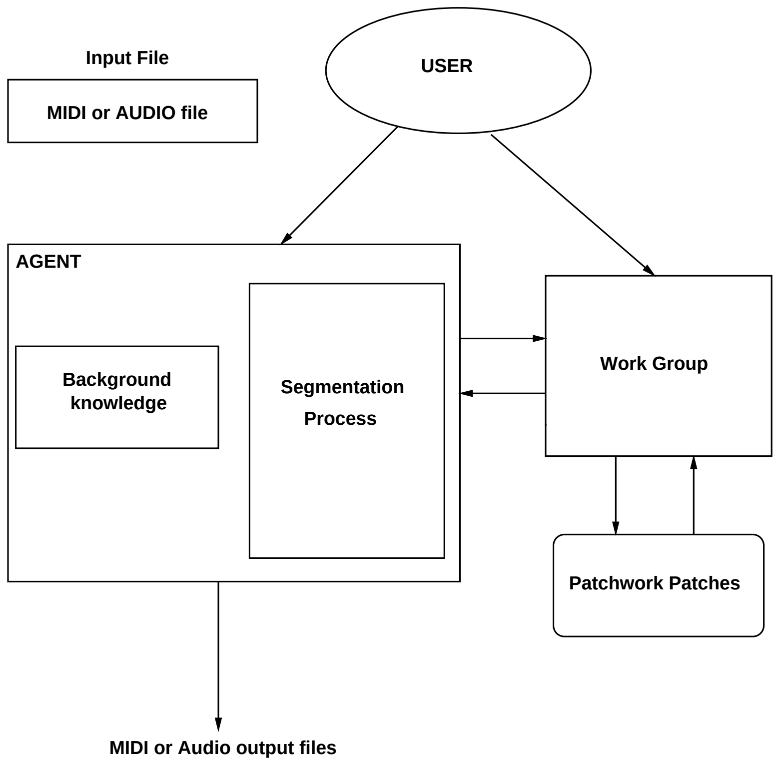

Figure 3: software description.

The agent will interact with the work group trying to solve the problem, accordingly to the opinions of the user. The work group will verify the users opinions and the agent results consulting some statistical algorithms implemented in the Patchwork environment. These algorithms deal with the verification of similarities in many parameters among sonic objects[6]. If the sonic objects resulting from the segmentation present similarities in a pre-determinated number of parameters, the segmentation is considered sound. If the work group does not find similarities among the resulting objects, he suggests the agent to redo the segmentation. The agent must then decide if he will ask the user for more information or if he has sufficient knowledge to try to infer a new order in which the parameters should be considered.

6 Previous conclusions and future work

As we said, this is a work in progress. Nonetheless, we can present some previous conclusions concerning the proposed model:

1. In the musicological side, the segmentation process is already tested, although not automaticaly, and works accordingly to our expectations [Gui97a, Tra98];

[6] For a complete description of these algorithms see [Gui97b].

2. The use of rational agents in the process of automatic segmentation of musical flows is, up to our knowledge, new;

3. The capability of the rational agent to produce arguable knowledge is an important feature for its use in the automatic segmentation process;

4. The multi-agent architecture permits that the agent collaborates with another agents, represented here by the work group and the user itself. Due to the complexity of the segmentation task, collaboration is a desirable feature.

For the completion of this work several things must be done, among them:

1. We must test the design of the agent: although the agent's design (fig. 2) seems to be sound and respecting the various aspects of the musicologal model, we can not state its validity for our purposes. Only an implementation and tests with this agent can verify the validity of this design;

2. Another important feature is the work group. It must be established how many agents will be there, as well as their capabilities. The use of the patches that the work group makes must also be better defined;

3. A study on the computational complexity of this model is also indispensable;

4. The communication model between agent and the work group must be defined, as well as the communication model between agent and user.

References

[Ber87] Wallace Berry. *Structural functions in music*. Dover, New York, 1987.

[Coo94] Nicholas Cook. *A Guide to Musical Analysis*. Oxford University Press, Oxford, 1994.

[Fer92] Edilson Ferneda. *Conception d'un Agent Rationnel et Examen de son Raisonnement en Géométrie*. Tese de doutorado, Universidade de Montpellier II, 1992.

[Gui97a] Didier Guigue. Sonic object: A model for twentieth-century music analysis. *Journal of New Music Research*, Vol. 26(N° 4):346–375, 1997.

[Gui97b] Didier Guigue. *Une étude "pour les Sonorités Opposées": Pour une analyse orientée objets de l'œuvre pour piano de Debussy et de la musique du XXè siècle*. Presses Universitaires du Septentrion, Villeneuve d'Ascq, 1997.

[Mee94] Nicolas Meeùs. De la forme musicale et de sa segmentation. *Musurgia*, Vol. 1(N°1):7–23, 1994.

[Mes94] Marcel Mesnage. Techniques de segmentation automatique en analyse musicale. *Musurgia*, Vol. 1(Nº 1):39–49, 1994.

[Mey89] Leonard B. Meyer. *Style and Music*. The University of Chicago Press, Chicago, 1989.

[RN95] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood-Clifs(NJ), 1995.

[Rus97] Stuart J. Russell. Rationality and intelligence. *Artificial Intelligence*, Nº 94:57–77, 1997.

[Sal97] Jean Sallantin. *Les Agents Inteligents: Essai sur la Rationalité des Calculs*. Collection informatique. Hermès, 1997.

[Tra98] Ernesto Trajano. A Klavierstück XI de Stockhausen: "uma imensa melodia" de timbres. In *Anais do XI Encontro Nacional da ANPPOM*, pages 286–291, 1998.

[Val95] Leslie G. Valiant. Rationality. In *Proc. of the 8th Annual Conference on Computer Learning Theory*, Santa Cruz (CA), July 5-8 1995.

About the authors

Born in 1975, **Ernesto Trajano** graduated in Piano at the Federal University of Paraíba. He is member of the GMT - Research Group on Music, Musicology and Applied Technology - since 1997. In the moment, he is a Master degree student in Computer Science at the Systems and Computation Department of the Federal University of Paraíba. His main research points are: Computers & Music, Artificial Intelligence, Musicology and Computer-aided Musical Analysis. His Master dissertation is focused on the automatic segmentation of musical flows, using AI concepts. More information, can be found at: *http://www.liaa.ch.ufpb.br/˜gmt*.

Didier Guigue was born in 1954 in France. He graduated in Piano and Bassoon but is also active as a musicologist, composer and performer. Master in Aesthetics, Sciences & Technology of Arts (University of Paris-8, France). Doctor in 20th Century Music & Musicology (IRCAM/*École des Hautes Etudes en Sciences Sociales*, France). He has lived in Brazil since 1982, where he teaches music analysis and computers applied to music at the Federal University of Paraíba. As a researcher and consultant at the CNPQ (National Brazilian Council for Research) and member of the IRCAM Forum, France, he works on computing in 20th Century music analysis. Related papers are published in *La Revue de Musicologie* and *The Journal of New Music Research*, among others. More information, including a bibliography with links to texts and papers, can be found at: *http://www.liaa.ch.ufpb.br/˜gmt*.

He composes instrumental and electro-acoustic music, in a variety of styles which range from experimental computer music to progressive pop, rock, jazz. He also composes for video, film, dance and theater projects, including the award-winning music for the films *A Margem da Luz* and *A Árvore da Miséria*.

Recently he had his works played at the International Computer Music Conference 1997, Thessaloniki, Jazz Festival Brazil/Argentina, New York, XII Bienal de Música Contemporânea Brasileira, Rio de Janeiro, III CEAIT Electronic Music Festival, California Institute of Arts, Valencia, USA, among others. Electroacoustic works are included in CDs published by Organized Sound (Cambridge University Press, 1998, Vol.3 No.1) e Leonardo Music Journal (Special 10th anniversary issue to be currently released). A full CD of his own, entitled *Vox Victimæ*, has been released in Brazil (1999, Eldorado/CPC-UMES, São Paulo). More about this work can be found at: *http://www.paraiwa.org.br/didierguigue*.

Edilson Ferneda: graduated in Computer Science (*Instituto Tecnológico em Aeronáutica*, São José dos Campos, Brazil, 1979), Master in Computer Science (Federal University of Paraíba, Campina Grande, Brazil, 1984) and Doctor in Computer Science (*Université de Montpellier*, France, 1992). Professor at the Systems and Computation Department of the Federal University of Paraíba.