Understanding Collective Gestural Improvisations; a Computational Approach

by

Hugo Solís García

Submitted in partial fulfilment of the requirements for
the degree of Diploma of Advanced Studies.
Doctorate in Computer Science and Digital Communication
Department of Technology

Tutor: Xavier Serra
Cotutor: Sergi Jordà

Universitat Pompeu Fabra

Barcelona, 30th August 2006

This research work was supported through the award of a scholarship by the *Fondo Nacional para la Cultura y las Artes*, through the *Programa de Apoyo para Estudios en el Extranjero* 2005 and 2006. (EE204-231 and EE105-044)

En apoyo al desarrollo de mis estudios se recibió un apoyo económico del Fondo Nacional para la Cultura y las Artes, a través del Programa de Apoyo para Estudios en el Extranjero 2005 y 2006. (EE204-231 and EE105-044)

“Music is not about sound,
it’s about people.”
John Zorn

Abstract

Gestural Improvisation is a type of free improvisation that aims to create short and long-term relationships in the acoustic parameters of music, such as pitch and intensity. The gestural improvisation style is closely related to the concept of *real-time composition.* This work describes a set of experiments based on *signal processing* methods and *machine learning* techniques that are applied to gestural improvisation music. These methods are commonly applied to other styles of music and employed in music information retrieval and computational musicology, among other fields. As a case-study, a set of twenty-one multi-channel recordings of music performed by the five-member gestural improvisation group *EnsAmble Crumble* is analyzed.

The theoretical framework presented includes a state-of-the-art scenario of computer improvisatory systems as part of an overview of musical concepts related to improvisation and structure. A section explaining the tools and technologies employed for analysis is given as a reference and a guide. The analyzed data is described in detail and a brief description of the author's previous experience in the development of gestural improvisation and improvisatory systems is also given. Finally, the results and conclusions of the analysis are presented with examples from the study: the behavior of one instrument during the collective performance, the global interaction of one instrument across all of the recordings, and the interaction of all the members across all of the improvisations.

In conclusion, basic rules of interaction were found that can be employed in the generation of computational models of musical improvisation. Such models may be integrated into *virtual musical partners* with a better *musical intelligence* than the systems currently available. A set of appendixes includes three articles that complement the work. The appendix contains the description of the *Moz-Art-Global-Art* project.

Resumen

La *improvisación gestual* es un tipo de improvisación libre que busca provocar la creación de las relaciones de corto y largo plazo en los parámetros acústicos de la música tales como altura e intensidad. El estilo que he decidido nombrar improvisación gestual está fuertemente relacionado con el concepto de *composición en tiempo real*.

Este trabajo describe un conjunto de experimentos basados en métodos y técnicas de procesamiento de señal y *machine learning* (aprendizaje mecánico) aplicados a improvisaciones gestuales. Estas metodologías son comúnmente aplicadas a otros estilos musicales y regularmente empleados en *music information retrieval* (recuperación de información musical) y musicología computacional entre otros campos. Como estudio de caso, se toma como referencia el conjunto de 21 grabaciones en multicanal realizadas por un grupo de improvisación gestual de cinco miembros (EnsAmble Crumble).

El marco teórico incluye una descripción del panorama actual en sistemas de improvisación por computadora y una revisión de conceptos musicales relacionados con la improvisación y la estructura musical. Como guía y referencia, otra sección explica las herramientas y tecnologías empleadas para su estudio. Los datos analizados, así como las características del ensamble son explicados en detalle. Se incluye una descripción breve de sistemas de improvisación del autor y de algunas de sus aportaciones musicales.

En la sección final se presentan los resultados y conclusiones de varios análisis que incluyen: las características de ejecución de un miembro del ensamble durante una improvisación particular, las características de ejecución de un miembro dentro del conjunto total de improvisaciones y las características de interacción de todos los miembros dentro del conjunto total de improvisaciones.

Como conclusión se ofrecen reglas básicas de interacción que fueron encontradas durante los análisis y que pueden ser empleadas para la creación de modelos computacionales de improvisación musical. Estos modelos podrían ser integrados a sistemas de interacción con una mejor *inteligencia musical* que los sistemas existentes. Se presenta también una serie de apéndices que incluyen tres artículos y la descripción del proyecto *Moz-Art-Global-Art*.

Acknowledgments

First, I want to thanks to the EnsAmble Crumble for being the musical project that little by little was shaped and built during these two years. After all, this work is about its music. Thanks to Alfonso, Maarten, Gabriela, Sylvain, Amaury, Mauricio, and Edgar for have been such a nice people to work with. The route was nice and uncertain, as improvisation has to be. Live inside and outside the ensemble with you has been great.

Thanks to Sergi, Gunter, Maarten and Marcos; the reacTable* team. It was a pleasure full of learning to be in the 330 room. Thanks also to Sergi Jordà for his interest and support about me to come here. I am in deep gratitude with all of you.

Special thanks to Xavier Serra for guiding the group with such clarity and vision. Thanks also for giving me the opportunity of teaching at the UPF. Thanks to all the members of the MTG for the constant support and help. The MTG is an extraordinary place to work at because of its members.

Thanks to the Essentia team from creating such a nice framework. Thanks to Esteban, Bram, Nicola, Bee Suan, Sebastian, Emilia, Rafael, Konstantin, Perfe, Koppy, José Pedro, Pedro Cano, Lars and many others for be always happy and ready to help and solve questions of any kind.

Thanks to Andrés, Gabriel, and all the people from or involved in the Phonos foundation. Thanks to Jose for all the support in the studio.

Thanks to Thomas Noll and Hendrik Purwins. Thanks for your classes inside and outside the classroom. Thanks to Tristan, your support, comments, and constant help have been always an essential guide. Thanks to Victor Adán for your help. Thanks to both of you for your inspiring thesis.

As always, thanks to my parents and relatives; Isa, Jiu, Jazmín, Mauricio, Julieta, Emiliano, Victoria and María: "Infancia es Destino". Thanks also to the friends from Mexico, you are always present. Again, thanks to the members of the EnsAmble Crumble, and also, to Justo, Catarina, Rosana, Alea, Emilio, Carla, Suad, Kristin, Noemi, Rosamérica among all the people with which we share and live the beauty of Spain and Barcelona.

Thanks to Gabriela. Live with you is never boring nor quite, nor predictable. It is a nice improvisation with a hidden structure. The dúo Juum is on the road.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"My loathings are simple: stupidity, oppression, crime, cruelty, soft music."
Vladimir Nabokov

1.1 Preamble

Gestural improvisation is a musical style that creates its discourse based on the use of not-quantized musical parameters. It is not based on harmonic or rhythmic relationships, and there is not any kind of established structure, grid, or patterns to follow. There is not a discretization in any of the musical dimensions, thus there are constant fluctuations on tempo. The use of glissandi, noisy sounds, and extended techniques in the acoustic instruments is by nature encouraged. Though, everything may seem initially chaotic and unorganized, there is usually a strong search of short-term and long-term structures[1].

Could gestural improvisation be studied with scientific tools? If so, are the conclusions that may be obtained useful? The act of creating free music has, in most of the cases, an association with freedom, spontaneity, and non-deterministic decision making on real time. However, if we look closer to the phenomenon we may conclude not only that gestural improvisation is not as free as we originally think, but also that there are several scientific methodologies and tools that could help us to understand the underlying mechanisms and properties this style.

Signal processing is nowadays a robust field that encompass a broad class of functions to analyze and manipulate data such as audio. Digital Signal Processing (DSP) deals with the digital version of analog signals and offers a vast amount of methods and techniques to treat the signal. The spread out of computers and digital devices has created a culture where almost all the music is recorded, played, processed, or at least preserved in digital format. Because of this, there is a current path of research that intends to extract and understand the musical meaning from the digital audio signal. Gestural improvisation may be a good candidate to be studied with the available DSP techniques.

However, most of the effort in the field of computational musicology has been directed to the study of pop and classical music where almost all the musical parameters are in certain extent discretized. Pitch is usually constraint to scales, and notes have a steady frequency. Harmony is derived from the use of well defined models[24]. Rhythm is created most of the time by a stable and unique pulse, and metric is usually of basic subdivisions[25]. Musical structure is assumed to be created by repetitions and variations of sections, for example verses and choruses[53]. Timbre is usually narrowed to what it could be produced by playing a well-known instrument in a natural way[45].

[1] the term *gestural improvisation* is explained in detail on 2.2.

In addition, the tools that signal processing offers are not enough to get a basic idea of the structure, patterns, and behavior derived from the signal. Thus, other fields should support our research: *Machine Learning* and *Data Mining*. These two fields of Computer Science are two relatively new field that integrate a vast collection of methodologies, techniques, and algorithms that help to get a better understanding of raw data. In our case, these fields helped to derived meaningful conclusions from the collective improvisations that were used as case-study.

However, music is not only a signal. Music has a social and cultural component that may even override any type of conclusion derived only from the analysis of the acoustic data[76]. Because of this, if we intend to study the interaction among musicians while performing free improvisations, it is important to implement not only the audio analyze but also use some of the techniques for modeling social networks and human interaction.

1.2 Description and Contributions

Gestural improvisation is a particular type of free improvisation. However, it searches for short and long-term structures in various musical parameters such as pitch and dynamics. It may be thought as a non-gridy *real-time composition* where musicians are allays aware about the main structure of the work. This work describes a set of numerical analysis of this particular style of musical performance.

As a case-study we analyze the musical production of the EnsAmble Crumble, a five-members group of *gestural improvisation* who recorded more than eight hours of music on multi-channel during a two-months period. The instruments of the ensemble are two electric guitars, synthesizer, violin, and viola.

The first step of the research was to translate the data from the audio files into a perceptual meaningful information suitable for analysis. A set of signal processing techniques based on perceptual models was employed to extract two generic descriptors that may be used for studying the interaction between musicians —loudness and spectral centroid.

An initial exploration for analyzing the signal employs Phase Space Reconstruction, Principal Component Analysis, and Dynamic Time Warping as methods for searching structures and pattern in the signal. In the second step of the research, well-known decision trees and association rules algorithms to detect *archetypical* patterns of interactions between music players were explored.

The results show that, even in music with such a big level of indeterminacy and instability, there are some patterns of interaction among the musicians. Taking into account such results in order to create a model of collective gestural improvisation could help for developing better computer-human improvisatory musical systems that could follow human musical behaviors.

1.3 Overview

This document is divided in six chapters and five appendixes. Each chapter covers a particular area of our research. The appendixes include articles written during the period of the research. If you are only skimming the text, this initial chapter may help to give you the main idea.

In this chapter, *Introduction* a general description of the research -including the contributions- is presented with the goal of establishing the frame and context. It also exposes the motivations and previous works that lead to the current research. In chapter 2, *Musical Background* different aspects of gestural improvisation are examined from a musical perspective. Definitions, examples, and related works are given. In chapter 3, *Numerical Background* all the mathematical, statistical, and computational topics employed during the study are described. The chapter 4, *Gestural Music and Improvisation* has to be seen as an introduction and/or explanation of the dataset that we used for our study. The musical justification came in the form of a description of both my previous works and the labor of the EnsAmble Crumble. Chapter 5, *Analysis and Understanding* describe the experiments and result using the dataset described on chapter 4, applying the tools of chapter 3, all within the musical context of chapter 2. Finally, chapter 6, *Conclusions and Future Work* presents an interpretation of the results of chapter 5, what was obtained, and which are the contributions. It also described what should be done in future work.

Appendix A, presents the paper *The acoustic piano as a conductor of collective musical improvisations.* On it, a suggested method for using an acoustic piano for guiding and shaping collective improvisations is given. This paper was submitted but not accepted to NIME 2006. Appendix B, *The Moz-art-global-art* project, describes a collaborative software for audiovisual improvisations recently developed. Its characteristics, properties, and current stage are exposed. Appendix C, *Graphical Representation of Music* is a final work for the *Introduction to Image Processing* course where basic elements of audio-visual relationships are presented. The Appendix D, *Verbalization of an Improvisation* presents a transcription in Spanish of a verbalized improvisation. You can confront this heuristic experiment developed in 1999 against the current work. Finally appendix E, presents the abstract from the article *Review of Music Cognition Models.* This article was developed by a team of researchers from MTG and I had a partial contribution on it.

1.4 Motivation

1.4.1 A virtual partner

Almost seven years ago, I had the first desire to design a computer musical partner which I could improvise with. This idea has been the main vector of almost all the technical and artistic projects I have realized since then. Two important questions I should have formulated in that moment: "why did I wanted to create a *virtual partner* if there are so many extraordinary live musicians to play with" and "what would be the aesthetic implications of such desire".

The idea of building an *intelligent* and *musical* system to improvise with has become much more harder than I initially could imagine when I had a naive idea about computers, programming, and improvisation. During all these years, I have been emerged in the field of computer music and I have seen and read about many systems for computer improvisation. At the same time, I have improve my technical skills and I have acquired more musical experience which has refined my aesthetic ideas.

The interesting (should I say funny) situation is that after all this time, I feel and hear now farther away from a good result than I was five years ago when I faced all problems with random procedures. I have become -as probably most of the computer music community- extremely skeptic about what computer can offer to *Music Creation* (in capitals) if there are not music ideas behind the technology.

During the natural going back and forward between musical ideas, music technical constraints, technical ideas, and technical constrains creators produce systems, pieces of software, and works that have embedded their own experience. To put in in George Lewis's words when he wrote about his *Voyager* piece of software:

> Musical computer programs, like any texts, are not "objective"

> or "universal," but instead represent the particular ideas of their creators. As notions about the nature and function of music become embedded into the structure of software-based musical systems and compositions, interactions with these systems tend to revel characteristics of the community of thought and culture that produced them. Thus, it would be useful here to examine the implications of the experience of programming and performing with *Voyager* as a kind of computer music-making embodying African-American cultural practice.[42]

On my turn, it was clear that in all my previous works I was omitting a transcendental fact. Besides a naive experiment presented in the appendix D, I had never tried to formalize or model the way in which I improvised. It is here where computer science, signal processing, and music cognition research play all together essential roles. It is impossible to make a computer to improvise if we do not previously formalize the improvisation techniques; and it is impossible to formalize the improvisation techniques if we do not study what we, as human improvisers, do. This work is all about studying how humans improvise. My "virtual partner" is therefore a future work for the moment.

1.4.2 Structure

Over the years, I have become more and more concerned and worried about the notion of structure in music. A piece of music is a path over time where the fluctuations and sudden changes of directions of the musical parameters are in constant evolution. The structure of a piece may be thought as the stratification of such trajectories.

As a listener, my challenge is trying to retain as much as possible of these paths with the better resolution and clarity in order to create a map of the piece.

I am only self satisfied when the work I am listening to is over and I manage to create a mental image of the entire work. However, this level of attention is rare and almost never reachable. In addition, most of the repertoire of the last years has purposed extreme subtle and/or complex structures.

As an improviser, the situation is not different. I am only satisfy when I create an improvisation where the musical ideas that integrates the work have many layers of relationship over different windows of time. In other words, a good improvisation should have consistency in its development. Again, this level of *real-time composition* is extremely hard to reach and only in few times I am satisfied with the result.

The current work is a path of exploration that aims for finding and implementing computer techniques that can help us to get a better understanding of the musical structure. This research belongs to a long-term exploration. The roots can be tracked in several work I developed in the past.

1.4.3 Previous work

Between 1996 and 2001, I did the professional studies of piano performance in Mexico City. These studies were complemented with studies in music composition. From the very beginning, I got deep interest in the areas of improvisation and contemporary music. After a short period, I started to improvise with electronic devices mainly commercial synthesizers and computer software. A course of Java music programming, and a curse of Electronics at New York University started my interest on programming and building my own music tools as a way to get a closed loop of the aesthetic process. As final project at the university, I developed *GAB* [68]; a hardware & software device that interacts on real time with a pianist while improvising. The desire to get a better understanding in the field of computer music and performance lead me to develop the *Improvisatory Music and Painting Interface* (IMPI) [69] at the MIT Media Lab. This software allows a painter to guide an ensemble of improvisers while creating dynamic graphics.

Even thought GAB and IMPI are useful and original systems that fulfill their tasks, I realized -after getting a deeper knowledge in the field of computer music- that a stronger knowledge of Computer Science was required in order to build outstanding systems. While getting involved in the research at the Music Technology Group of the Pompeu Fabra University, I realized the musical usability of some fields such as Machine Learning, Data Mining, and Music Information Retrieval. Applying such knowledge to my own artistic work was a natural path to take for my research.

Parallel to the study of the mentioned areas, the desired to maintain my activity as musician lead to the creation of an improvisatory ensemble. The EnsAmble Crumble was created on March 2005 and started to play once a week. On May 2005 the EnsAmble Crumble started to record its improvisations with regularity. At the beginning of 2006, when my path of research became established, the ensemble started to be recorded with discipline twice a week with multichannel tracks.

1.4.4 Conclusion of the Chapter

Gestural collective improvisation is a complex discipline. In one hand, it implies to search unusual gestures and sounds, and in several cases —as it is my case— to create short-term and long-term structures on real time. However, it is common that the relationships between musical elements and the interaction among musicians become dimmed, at least in certain amount, if we compare them with the type of interaction that may be obtained with notated music. On the other hand, some of the decisions that musicians take while improvising, which create an impact on the general musical discourse, are sometimes unnoticed by the musician or are so embedded in his or her musical language that are hard to described. Could Signal Processing, Machine learning, and Data mining's tools and procedures help to get musical conclusions of collective gestural improvisations? Which kind of observations and generalizations we can obtain? Could this information lead to find archetypes in this music? Could we learn something musically meaningful from such information? Could we infer a model of gestural improvisation? Such type or questions are the main motivation of this work.

Chapter 2

Musical Background

"My music is best understood by children and animals."
Igor Stravinsky

2.1 Improvisation

Musical improvisation is one of the most antique artistic expressions of the human being. Before notation was created humans had the necessity to express themselves in the acoustic domain, and improvisation was one of the techniques employed. In fact, "there is scarcely a single field in music that has remained unaffected by improvisation, scarcely a single musical technique or form of composition that did not originate in improvisatory practice or was not essentially influenced by it. The hole history of the development of music is accompanied by manifestation for the drive to improvise."[5]

The ethnomusicologist Bruno Nettl defines improvisation as

> The creation of a musical work, or the final form of a musical work, as it is being performed. It may involve the work's immediate composition by its performers, or the elaboration or adjustment of an existing framework, or anything in between. To some extent, every performance involves elements of improvisation, although its degree varies according to period and place, and to some extent, every improvisation rests on a series of conventions or implicit rules. The term *extemporization* is used more or less interchangeably with *improvisation*. By its very nature -in that improvisation is essentially evanescent- it is one of the subjects least amenable to historical research."[49] [50]

Improvising is learned with different methods according to the culture and musical style. It could be learned empirically by hearing, copying, and repeating material; or could be thought through in a carefully planned and systematized procedure. "Iranian musicians are told that memorization of the radif, a repertory of 250–300 short pieces, will automatically teach them the techniques of improvisation. Jazz musicians have a variety of learning techniques, including the notation and memorizing of outstanding solos".[49]

All kinds of improvisations are framed by the specific musical and cultural context that surrounds them, no matter if we talk about the improvisation of ornaments in the sixteenth century or the current experimental improvisations using laptops. There are also societies that do not have the notion of improvisation as it is defined in Western music because these societies do not separate composition from improvisation.

Improvisation has a connotation of spontaneity and liberty. Contrary to written music, improvisation offers performers the option of making their own decisions. The level of liberty is framed by the context and by how the improvisatory technique is used. Improvisatory techniques can be employed in different steps of the creative process and can influence the production of music in several ways. It could be, for example, the technique used to compose a score, or it could be the method used during a performance.

Even though it could be thought that improvisation is a field with no constraints, in reality in order to be able to improvise, one must master the language and techniques of the music being produced. In certain ways, improvisation is nothing but composition in real time. In some styles, the boundary between composition and improvisation is blurry especially because both fields take elements form each other. This situation creates one of the main paradoxes of the improvisatory language. For some creators, improvisation should emphasize its compositional quality, for others it should accent its own nature of real-time exploration. Nevertheless, it is accepted that collective gestural improvisation has its own musical form.

In Western music, the notion of musical form has been an important concern throughout history. Each historic period and musical style has created and expanded musical forms, and gestural improvisation has its own. The musical form in improvisatory materials is, in most cases, more diffuse and less structured than the form of compositional works. In addition, it takes longer to develop transitions and variations of the material. Usually, elements and ideas are much more spread out in time. The main reason for this is the difficulty of comprehending the material without a score, and the difficulty of synchronizing events in the case of collective gestural improvisations.

Even if the gestural collective improvisatory style should not necessarily use the syntax of composed music, there are aspects of the latter that could acoustically enrich the expression in the improvisatory domain. Common elements in notated music such as synchronized events and sudden changes by the entire ensemble are hard to obtain with big ensembles. Several techniques have been created in order to import these elements into the improvisatory domain. The simpler one is to create hierarchic ensembles where one members guide the rest. Others, more complex, involves the creation of body language symbols, improvisatory scores, or implementations of computer systems.

2.2 Gestural Improvisation

2.2.1 Notion of Gesture

In the musical human-human and human-computer interaction contexts there is not a unique definition of *gesture* and the term has several definitions and connotations. From the many different definitions "one can reasonably state that there cannot exist one single (*simple!*) definition of the term *gesture*".[11]. For our purposes we may adopted the following definition: a gesture is "a movement of the body that expresses or emphasizes an idea, sentiment, or attitude" [29]. No matter the media, the style, or the technique, a gesture is always a unit that, within the context, acts as a transmitter of expressive information. In the audiovisual domain, a gesture refers to an element or collection of acoustic and visual material that can be recognized as a single unit of expression with an identifiable meaning. In F. Nicolas words, "I call gesture a particular kind of musical moment, which is generally clearly identifiable in my works because it contains a form and a cut-out that are obvious both on reading the score and on listening."[51][11]

2.2.2 Gestural Improvisation versus Free Improvisation

The term *Gestural Improvisation* has not formally described ever but it is associated with visual events or motion activities[38]. It clearly implies the notion described by F. Nicolas. On the other hand, the term *Free improvisation* clearly identify a "musical style" but not the acoustical characteristics. Free improvisation is a generic term that defines a manner of sound production. Contrary, Gestural improvisation defines an acoustical result. It should aims for the productions of musical moments with clear shapes that all together generates an organic form. In this sense, Gestural Improvisation could be associated to the concept of *real-time composition*.

2.2.3 Collective Performance

Collective performance is an experience with particular characteristics. The activity of improvising with other musicians requires the same attention as the performance of written music. Attention must be given not only to the material that each member of the ensemble is producing but also to the final result.

According to the way the members of an ensemble interact, several kinds of ensembles can be designed. Each kind of organization offers advantages and disadvantages over the other configurations, and each one has characteristics that help to emphasize different elements. Some configurations offer more freedom and possibilities for interaction between the members, while others are better at creating well-organized materials.

Six basic models of collective performance can be distinguised: Centralized ensembles with one leader and unidirectional transmission of information between the leader and the rest of the ensemble (fig. 2.1 A); Centralized ensembles with one leader and bidirectional transmission of information between the leader and the rest of the ensemble (fig. 2.1 B); Hierarchic ensembles with group leaders and unidirectional transmission of information between the leaders and the rest of the sub-ensembles (fig. 2.1 C). Hierarchic ensembles with group leaders and bidirectional transmission of information between the leaders and the rest of the sub-ensembles (fig. 2.1 D); Centralized ensembles with one leader

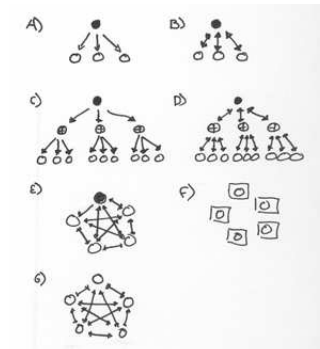


Figure 2.1: Different types of ensembles.

and communication between the leader and the rest of the ensemble, as well as communication between the members (fig. 2.1 E); Non-centralized ensembles with no communication among members (fig. 2.1 F); and Non-centralized ensembles with communication among the members ((fig. 2.1 G).

2.2.4 Notation and Improvisation

The creation of scores that allow certain ranges of variation have been present during the entire history of music notation. The *basso continuo* (fig 2.2)[81] is, for example, one technique where the musicians improvise and complete certain aspects of a score that only presents the harmonic skeleton of the piece.

Currently, some styles of jazz have adopted the use of scores that must be completed by the performer. In this style, it is common that the score

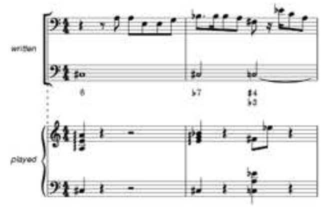



Figure 2.2: Kollman: A Second Practical Guide Thorough-bass (1807).

or Standard show only the basic melody and the harmonic structure. From these elements, the player can build the entire material. In this case, notation helps to preserve both the organization of the ensemble and the structure of the piece. The level of complexity in improvisatory symbols has increased over time. Several contemporary improvisatory scores help to create pieces of complex structure in real time. As will be shown shortly.

2.2.5 Improvisation and Style

In modern Western music, improvisatory techniques for performance have had different levels of acceptance by the musicians depending on the musical style and kind of ensemble. Some improvisatory techniques are commonly used in popular styles such as blues, rock, and jazz. There are also ensembles and styles that are entirely based on improvising such as the performers of electronic improvisations. On the other hand, it is rare to see traditional ensembles such as orchestras, string quartets and brass quintets performing improvisatory concerts.

Some referential points in the history of music improvisation are the nu-

(a) La lontanaza nostalgica utopica futura by Luigi Nono.

(b) Walter Thomson and the Soundpainting orchestra.

(c) Karlheinz Stockhausen, No. 14 Plus Minus detail.

Figure 2.3: Images of Representative Works and Technologies.

merous techniques that involve creation in real time, such as ornamentation, *basso continuo*, and cadenzas. However, 20th century was the boom period for non-standard techniques, nomenclatures, and improvisatory techniques. An incredible number of composers have done explorations in the field, among others: John Cage, Karlheinz Stockhausen (fig. 2.3(c)), Luigi Nono (fig. 2.3(a)) [52], La Monte Young, Terry Riley, Vinko Globokar, and John Zorn.

In *Orchestral improvisation*[1] the conductor stands in front of the entire or-

[1]Here the concept of Orchestra is limited to the classical orchestra. Not to the jazz or Latin orchestral ensemble.

chestra and using traditional techniques for conducting, he or she leads the sonority of the orchestra. Around fifty to seventy musicians are usually involved in this kind of ensemble. Nevertheless, the conductor could control the loudness, the activity of events, and the sections that should or should not play, among other parameters. The level of previous agreement about the meaning of conductor's gestures varies. this type of experience is not practiced often not only because the many logistic difficulties of organizing a big ensemble, but also because orchestral musicians are usually not used to improvise. However, there are several exceptions that arouse interesting examples. Larry Austin, Hans Werner Henze, and Steven Mackey, involved having star improvisers as conductors. Other examples that reimagine the orchestra along improvisational and communitarian lines includes Cornelius Cardew's "Scratch Orchestra," Pauline Oliveros's "Sonic Meditations," and Lawrence "Butch" Morris's "Conductions," in which an improvising conductor functions literally as a centralized conduit of musical current linking other improvisors.[43] Sparkler[33][44] by Tod Machover is a work where, sometimes during the piece, the conductor guides the improvisation of the entire orchestra. In parallel, a computer processes the music produced by the musicians. The works that Walter Thomson realizes with his orchestra deserve special attention. "*Soundpainting* was initially developed as a method of conducting and shaping improvisation during a performance" [71]. Over many years, Thomson developed a vocabulary of hundreds of gestures. "Each player reads the conductor's hands for the style of the next sound, movement or utterance" and each movement has a different musical meaning. Some of them are simple, as, for example, "play loud" and some have a more detailed meaning, such as "long, loud, and high tone" (fig 2.3(b)).

The *Experimental and Free Improvisatory Ensembles* have their roots in the merging of jazz and the *avant-garde* music of the twentieth century. They are usually small, involving merely three to ten musicians. These groups performed entire concerts without any kind of musical constrains. Non-traditional instruments or custom sound devices are usually employed. When technology became popularized, these ensembles started to included computers and electronic devices in order to generate the sound.

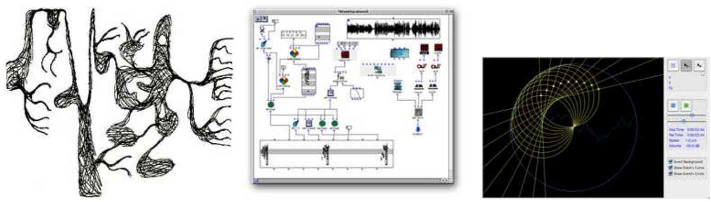
2.3 Computer performance

Computers are powerful instruments that can enrich some aspects of art. They have been used in many artistic contexts since they were created. In the field of computer-assisted composition, there has been a constant improvement of ideas and implementations.

Currently, the variety of approaches and techniques computers offer in the musical field is immense. Their possible uses can range from emulating techniques from the analog world, such as create a random collections of pitches, to the implementation of complex methods of artificial intelligence in order to, for example, automatically complete a musical phrase in the style of the input source.

Nowadays, the ability to create high quality digital audio, and the amount of techniques for synthesizing sounds, has made it so that most of the work in computer music is being focused on the production of digital audio. The situation is exciting because new sonorities are being created; however, the work with acoustic material should be also explored with the same intensity.

During the beginning of computer sciences there were some programs for the composition of music with graphical representations. Some of the important historic references are: The GraphicI used by Max Mathews and L. Rosler and the *Unite Polyagogique Informatique de CEMAMu* designed by Iannis Xenakis (fig. 2.4(a))[13][84]. Nowadays, the majority of them have graphical interfaces and each one has its own visual paradigm. Pure Data[56] by Miller Puckette; Wire[10] by Phil Burk, AudioMunch[7] are some of them. Open Music[54]



(a) Mycenes Alpha by Xenakis.

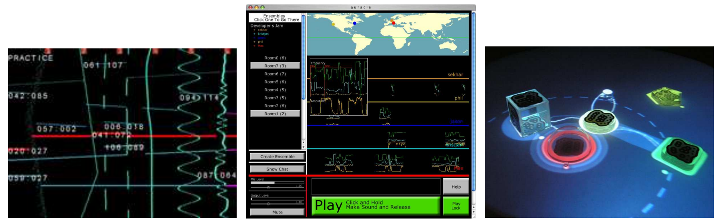
(b) Open Music software.

(c) Example of a work with the IanniX system.

Figure 2.4: Programs and historical references.

by Gérard Assayag and Carlos Agon is particularly interesting in our context because it allows working with notated scores (fig. 2.4(b)).

There are also, a variety of programs that use graphical interfaces in order to create electronic music. In these systems, the graphic is not intended to be part of the final result but just a mechanism for representing the material that it wanted to be sonified. Two of these programs are Metasynth[75] by Eric Wenger, and Iannix[39] designed by La Kitchen (fig. 2.4(c)). On the other hand, few programs are intended for creating music for acoustic ensembles. Hyperscore [22] by Mary Farbood and Egon Pasztor is a piece of software that uses sketches of the user to generate music.



(a) FMOL system by Sergi Jordà.

(b) Auracle's screenshot.

(c) ReacTable's screenshot.

Figure 2.5: Programs for collective computer improvisation.

2.3.1 Computer Assisted Improvisation

The development of algorithms for leading improvisations is older than digital technology. However, computers are ideal devices for setting up systems in which the algorithms could be much more elaborate. Complexity in those algorithms does not necessarily mean that the aesthetic result is more interesting, but experimentation with intricate relationships is much easier, and the expressive possibilities can be extended. Graphical interfaces and hardware devices for the gestural input of information are also elements that have helped in the development of the field. For our analytical purpose, we could divide computer systems into two different branches: one that tries to emphasize the principle of spontaneity and liberty of improvisation, and one that tries to highlight the organization of the material.

The possibilities for transmitting information over different locations, and the simplification of data manipulations using graphical representations has contributed to using computers as an excellent method for helping in collective composition and improvisation. Some computer systems that have been developed for collective computer improvisation are FMOL by Sergi Jordà[35][36] (fig. 2.5(a)); Auracle[58] by Max Neuhaus and others (fig. 2.5(b)); Webdrum by Phil Burk[9].

The *Cognitive Model of Improvisational Action* and its implementation in the *Emonic Environment*[48] is an original work that aims for a new paradigm in human computer interaction. Nemirovsky argues that when we use computers "to interact with media, our experience is that of direct control, and the goal of our interactions is either artifact-production (the editor paradigm) or passive exploration (the browser paradigm)."[48] He proposes "an alternative: a model of media interaction based on the ideas of non-idiomatic improvisation that encourages active exploration of media and its structures. [He] argues that in order to facilitate this kind of exploration, (1) computational tools must actively participate in the creative process and (2) the interaction framework must allow structural exploration of media. This leads to our main claim: improvisation should be considered a valid and appropriate paradigm for media interaction."[48]

The *Voyager*[42] system developed by George Lewis is one of the most well-known systems for computer-interaction systems for music improvisation. The first version was programmed between 1986 and 1988 on Forth. He "conceive a performance of *Voyager* as multiple parallel streams of music generation, emanating from both the computers and the humans -a non-hierarchical, improvisational, subject-subject model of discourse, rather than a stimulus/response setup."[42] The system is technically simple in term of the analysis that it develops for the generation of the output, however the musical quality is obtained not only by the well-tune of the variables but most important, because the input that has been employed has been always top-level musicians. In the same way, the system is well-known not only because it is good, but fundamentally because the player who use it and create it is a great musicians. On Saltz[62] words:

> All that should matter to us are the visual appearance and the acoustic properties of Lewis's performance. But the reality does matter; indeed, the quality of the music plays only a minor role in the fascination this work holds. Lewis's performance with *Voyager* can be most captivating when *Voyager*'s output is the least appealing, and we sense Lewis's attempts to urge the system into more satisfying musical territory.

The *Continuator*[55] has won certain reputation in the community as a system that responds to the input of the musicians with original material within the style of the input. It makes emphasis in the link between generate stylistically consistent material at the same time that that it works as an interactive system. It is a MIDI system that based on Markov models build operational representations of musical styles in real-time context. The system has been build having in mind the benefits and draws -such as lack of long-term information- of representing the state of Markov chains in music. It has four collaborative music modes: single autarcy, master/slave, cumulative, and sharing.

The system builds up a database of patterns detected in the input sequences and uses an indexing scheme which represents all the subsequences -complete and efficient- by implementing a complete variable-order Markov model. It builds a prefix tree by a simple, linear analysis of each input sequence. Later, each new sequence is parsed from right to left and new prefixes encountered are systematically added to the tree. Each node of the tree is labeled by a reduction function which could be the pitch. To each tree node a list of continuations encountered in the corpus is attached denoting the index of the input sequence to avoid duplication. The generation of continuations is produced then with variable-order Markov chains.

```
Input {A B}
Continuation_List({A B}) = {3, 7}  = {C, B}
Random draw, suppose B.
Input {A B B}
Continuation_List({A B B}) = {8}  = {C}
Only one = C.
Input {A B B C}
Continuation_List({A B B C}) = There are none but
                longest possible subsequence = {B C} = 4
Input{A B B C D} no continuation, a node chosen at random. etc...
```

Reduction functions are applied to Pitch, Pitch and Duration, and Pitch Region = Pitch / region_size. The system uses hierarchical graphs as an alternative to Hidden Markov Models for some configurations. For the rhythmic variables four methods for control are employed: natural rhythm, linear rhythm, input rhythm, and fixed metrical structure. The system generates variations by the modification of the basic Markovian generation process by adjunction of a fitness function in real time. It also bias the Markov generation for harmony change with the last 8 notes among others information.

The *Improvisatory Music and Painting Interface*(IMPI) described on chapter 4 and the *Moz-art-global-art* described on appendix B are two systems developed by the author. They should be mentioned because they are practical implementations of the author's ideas, aesthetic, and techniques.

2.4 Psychoacoustic and Philosophical Models

Several psychoacoustic models have been developed in order to create better representations of the way human perceive audio and sounds. All these models take into account psychoacoustic properties. There are several psychoacoustic models that simulate many of the process that occurs during the different steps in which the music perception can be divided. A deeper analysis of the field is out of the scope of this work. However, appendix E is a detailed review of facts and models about music cognition with a emphasis on their computational implementations.

Here, we will concentrate only on some of the few philosophical models of improvisation that have been proposed. On [17] Derrida establish that any *object* can be taken out of its context an re-appropriated given the improviser's desires. For Derrida, the meaning emerges from the marriage of the intention of the writer and reader's perception. The context, thus, is integral to understanding what we see, hear, and read. Therefore, a text can be opened to multiple meanings and interpretations. "Meaning itself is not a fixed construct but rather a fluid divergent process. Deconstructing a 'text' involves showing inherent oppositional or dialectic contradictions in the text's internal logic."[48]

> Following Derrida in his search for a post-structuralist theory of meaning, Deleuze and Guattari introduced the concept of the rhizome[16], a construct that allows for multiple, non-hierarchical entry and exit points in data representation and interpretation. Such a construct is reminiscent of our idea of improvisation, which is defined as a process with no clear structural form.[48]
>
> Among the few cognitive theories of improvisation that have been proposed, Pressing's (in [66]) is particularly interesting. For him, improvisation is association-based and consists of sequential choice

> of elements which either continue or interrupt some aspect of the immediately preceding context. This step-by-step (and sometimes literally note-by-note) explanation of improvisation is in sharp contrast to the work of Sagi and Vitanyi (in the same volume) which attempts to escape the cognitivist approach and focus instead on global features of structure and style. Another, more phenomenological approach that is indirectly applicable to the question of improvisational action is the description of the problem of cognition proposed by Varela[74].[48]

2.5 Structure and Musical Form

Structure and *form* both denote the idea of the internal organization of the musical materials and the relationship among them in a piece of music. Structure is then a model that maintains certain proportions and it works as a framework or skeleton inside a piece of music. In the area of music composition structure and form offer a solution for thinking the internal organization of a piece. In the areas of music theory and music analysis they can be used to explain and track the thoughts of the composer. Finally, in the domain of pedagogy both notions may help to guide a learning process.

It is common in art theory, to refer to the concept of form. Music theory has been deeply concerned about the evolution of musical form throughout history, and it uses the concept to explain several elements of musical phenomenon. The form can be seen as a general map in which it is possible to track the development of elements and ideas. As Whittall suggests:

> Form might be defined simply as what forms have in common, reflecting the fact that an organizing impulse is at the heart of any

> compositional enterprise, from the most modest to the most ambitious. Yet the act, and art, of composition is not synonymous with the selection and activation of formal templates, and composers oblige writers on music to confront the infinite flexibility of the relation between 'form' as a generic category (such as ternary, canon, sonata) and the musical work as the unique result of the deployment of particular materials and processes. Practice particularizes, just as theory generalizes, and discussion of musical form has been especially vulnerable to the tensions which arise between these very different ways of thinking.[77]

On the practice, the form of a musical work is so dependent to the work itself, that the notion of musical form is sometimes questionable. A.B. Marx argued, for example, that "there are as many forms as works of art".[46] This ontological problem is, in fact, a serious problem when we discuss music with open structures as is the case of gestural improvisation. After all, for its own fact of existence, a piece of music has a shape. The aesthetic polemic about form and shape has been present during history and many artist and theorist have strongly questioned their real meaning. However, shape is not structure, at least, for our purposes of analysis. In addition, different kinds of materials are traditionally associated with certain types of forms[2]. On Salzer's words:

> Salzer distinguishes between 'structure' as revealed in Schenkerian voice-leading and harmonic analysis, 'form' as the organization and division of that structure into definite sections, and the relation of those sections to each other, and 'design' as the organization of the compositional surface, in terms of its thematic and rhythmic material.[63]

[2] In Varese's words "New materials require new structures."

Since the 18th century, theorists and composers have reiterated the need for wholeness, symmetry and proportion. However, the musical evolution has opened the door to different kind of tune systems, harmonic systems, and sound quality textures. The compositional techniques are nowadays open to a broad field of possibilities without aesthetic prejudges. The only "parameter" that coexist in all type of musics from any historical period is the relationship of the parts that integrates the piece. Hanslick formulated, now known as *formalist aesthetics*, the principle that "the beautiful is not contingent upon nor in need of any subject introduced from without, but . . . consists wholly of sounds artistically combined".[26]

In other words, the musical form has been the media to maintain the coherence of everyday more complex musical materials. The duration of a piece of music could became gradually longer because composers have managed to create complex musical forms. The Sonata, for example, was a form that for many years contained and molded the composers's imagination. Composers should find the way to fit their ideas inside the rigor of the sonata's form. Though, what happens with the form and the structure in gestural improvisations?

> Carl Dahlhaus was willing to preserve the distinction between a concept of form signifying 'musical coherence on a large scale' and *musique informelle* (exemplified by the radical music of the 1950s and 60s) whose purpose was 'to draw undivided attention to the isolated detail, to the individual musical moment'. Since 'the symptom of extreme musique informelle is the heterogeneous nature of the details from which a musical shape is constructed' and 'disconnected matter stands side by side in sharp contrast', the distinction between 'formed' and 'unformed' music is clear, and aesthetic judgments can derive from this distinction, whether one is regarded as good and the other bad, or both are believed to have equal potential for successful

or unsuccessful use.[77]

The notion of *form* has been associated not only with the basic idea of organization but more deeply with the notion of *organicism*[77]. Therefore, there are constant biological and physical analogies for describing musical events. Schoenberg, said for example “form means that a piece is organized... that it consists of elements functioning like those of a living organism”[65]. In the next chapter we will try to formalize some of these concepts.

2.6 Conclusion of the Chapter

On this chapter we purposed a definition for *gestural improvisation* and said that it is an specific style of free improvisation where structure plays an substantial role. Concepts such as structure, musical form, and improvisation are also studied. We have also seen that there is a long tradition not only on computer composition systems but more specific on computer improvisatory music systems. From the entire panorama a few set of examples were explored and analyzed as reference.

Chapter 3

Numerical Background

"Music is sound's cognitive apologist."
Stephen Smoliar

3.1 Signal Processing and Audio Descriptors

As described in the Introduction, the use of signal processing techniques in the sound and music domains is and old and nowadays robust field. A digitalized musical file may be subject to several digital techniques in order to transform, analyze, or compress the content among others goals. There are however, several limitations that still the boundaries of what is or not feasible. As an example we may mention the not yet solved problem of automatic transcription of polyphonic music.[40]

On this thesis we use several techniques that are commonly employed in sound and music analysis and we will adequate them to our context and requirements. Each of these techniques is by itself an exhaustive field of research. We will limit our scope to employ only the tools that have shown to offer good results in similar works. The rest of this section describe briefly each of the techniques that are employed for the analysis of our improvisations.

3.1.1 Audio Descriptors

Any data that represents some characteristic or property of an acoustic signal may be consider an *audio descriptor*. However the term has been narrowed to numerical values that gives information about the characteristics of an acoustic signal. Ideally, these descriptors should give relevant information by producing some perceptual property of the signal in digital format. Usually, audio descriptors are divided in *Low-level* audio descriptors (LLD) as specified for instance in the MPEG7 standardization format[47], and *high-level* audio descriptors which are more complex process obtained by specific combinations and manipulations of LLDs. LLDs can be organized in various categories including *temporal descriptors* computed from the waveform and its envelope, *energy descriptors* referring to various energy measurements of the signal, *spectral descriptors* computed from the STFT, *harmonic descriptors* computed from the sinusoidal harmonic modeling, and *perceptual descriptors* computed using a model of the human hearing process.[32][27]

Loudness

One important limitation of the spectrogram is the manner in which the intensity of the signal is displayed generally in Decibels SPL. It provides an objective measure of the intensity but it does not properly described the subjective impression of loudness. For this reason, several methods that take into account the sensitivity of the ear to various sound levels of the frequency components contained in the sound have been proposed and implemented. In [72] several algorithms for obtaining the perceptual loudness of the sound are presented. For practical reasons we used the DIN 45631/ISO532B Loudness Model that is based on Zwicker's ported from BASIC to Matlab. This implementation uses a filterbank of one-third-octave filters for the spectral decomposition of the signal. The equation for the specific loudness associated to each Bark band N' (in Sone/Bark) of the dB SPL sound level L_G in a one-third-octave band is

$$N'(z) = E(z)^{0.23} \tag{3.1}$$

The total loudness is the sum of the individual loudness can be expressed therefore as

$$N = \sum_{1}^{nb\_band} N'(z) \tag{3.2}$$

With these descriptor we could obtained a range of values that may be discretized and mapped to a traditional musical scale of intensity from *ppp* to *fff*[1].

Spectral Centroid

The Spectral Centroid is a low level audio descriptor that has certain relationship with the brightness and pitch region of a sound[2]. It measures the average

[1] Or as some radical composers from *pppppppp* to *ffffffff*

[2] During the research the *Brigthness* object developed by Tristan Jehan for the MaxMSP environment was ported for PD. It may be an alternative for real-time analysis of the spectral

frequency, weighted by amplitude, of a spectrum.The centroid of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes.

$$c_i = \frac{\sum f_i * a_i}{\sum a_i} \tag{3.3}$$

Where i is each frame of a sound, a is the amplitude, and f is the frequency band.

We decided to use this particular LLD as a practical solution to get a general notion about the pitch region where the instrument is performing during a specific moment. Other LLD could be employed, however the spectral centroid has shown that it is perceptually relevant. As a proof of its usability we can see that it has been adopted in the MPEG-7 protocol.

Morphological Descriptors

An original approach, not applied on the current stage of the research, but of high interest in our context are the descriptors purposed by Julien Ricard on [60]. A battery of so called *Morphological Descriptors* are described and implemented based on the *typo-morphologies* first defined by the French composer and Musicologist Pierre Schaeffer on his *Traité des objets musicaux*[64].

On his pioneer treatise, Schaeffer identifies four listening modes, *ecouter*, *comprendre*, *entendre*, and *ouir*. Most important, in order to validate his aesthetic that aims for breaking the link between sound object and source of production, he purpose a set of seven categories in three criteria, *form*, *matter*, and

centroid.

variation. *Form* is defined as a property that could be observed if we could freeze the sound; *matter* is related to the time evolution; and *variation* is concerned with the change of *form* and *matter*. In the matter criteria he defined *mass*, *harmonic timbre*, and *grain*; in the form criteria *dynamics* and *pace* (amplitude or frequency modulation); and in the variation criteria *melodic profile* and *mass profile*.

Even though, Schaeffer's thoughts may be consider naive nowadays, there is not only a historical relevance, but also a value in the principles that he defined and still may be explored in research and music creation.

3.1.2 Autocorrelation and self-similarity matrix

"Autocorrelation is a mathematical tool used frequently in signal processing for analyzing functions or series of values, such as time domain signals. It is the cross-correlation of a signal with itself. Autocorrelation is useful for finding repeating patterns in a signal, such as determining the presence of a periodic signal which has been buried under noise, or identifying the fundamental frequency of a signal which does not actually contain that frequency component, but implies it with many harmonic frequencies."[78]

The autocorrelation A of a signal $y[n]$ is calculated as follows:

$$A[\eta] = \frac{\sum_{n=\eta}^{N} y[n]y[n-\eta]}{(\sum_{n=0}^{N-\eta} y[n]^2)^{1/2}(\sum_{n=\eta}^{N} y[n]^2)^{1/2}} \tag{3.4}$$

The self similarity matrix is a representation of the autocorrelation function. In our case, a simplified version is used because only one-dimensional data is compared. The matrix is displayed so that it helps to detected repetitions or imitation segments by finding parallel lines.[23] Both, the x and the y axes are the index of the vector so each point is the union of all the points. Half of the representation is repeated with the diagonal as the symmetry vertex. The diagonal has the bigger value of selfsimilarly because it is the selfcorrelation.

3.2 Statistics

Statistics in conjunction with Computer Science have developed together a robust filed with a deep impact in the analysis of data and information. Several of the techniques employed in this field have a direct application in the music domain. Particular interest has been put in the work of the composer Victor Adan.

On [2] the composer Victor Adan applied several techniques traditionally employed in the signal processing domain to the area of music analysis and composition. By treating midi files as data signals and applying techniques such as State Space Reconstruction and Modeling, Fourier and Short Fourier Transformations, Wavelets Transformations and Principal Component Analysis the composer got several interesting musical results. He used the model for Interpolation, Abstracting Dynamics from States, Decompositions, and most interesting, for a method that he called *Musical Chimaeras* where he managed to merge two pieces and generate a new one. The generated material is perceptually identifiable as a mix of the sources. The experiments described on [2] were the starting point for our exploration.

3.2.1 Phase Space Embedding and Phase Space Reconstruction[3]

The Phase Space Embedding (PSE) is a technique that can in certain signals show hidden properties of the data that with other techniques remains invisible. For example, some signals that shows nosy behaviors when Fourier transformations are applied to them revels clearer patterns when they are embedded in higher dimensions.

Observed Chaos, Representing Flow, or Dynamical Systems

> A deterministic dynamical system is one where all future states are known with absolute certainty given a state at an instant in time. If the future state of a system depends only on the present state, independently of the time at which the state is found, then the system is said to be autonomous. Thus, the dynamics of an autonomous system are defined only in terms of the states themselves and not in terms of time.[2]
>
> When we are analyzing complex system, it could be the case that we observe only one or at best a few of the dynamical variables which govern the behavior of it. In addition, we often have a sampling rate for our observations. We never sample anything continuously. However this is not a restriction because in some way, all analysis of physical systems take place in discrete time. If we sample a scalar signal $s(t)$ at time intervals τ_s starting at some time τ_0, then our data is actually of the form $s(n) = s(\tau_0 + n\tau_s)$ and the evolution we

[3]Phase Space is also known as State Space.

observe takes us from $s(k)$ to $s(k+1)$.[1] Therefore, we can represent continuous flows

$$\frac{d\boldsymbol{F}(x)}{dt} = \boldsymbol{F}(\boldsymbol{x}(t)) \tag{3.5}$$

as finitely sampled evolutions

$$\boldsymbol{x}(t_0 + (n+1)\tau_s) \approx \boldsymbol{x}(t_0 + n\tau_s) + \tau_s \boldsymbol{F}(\boldsymbol{x}(t_0 + n\tau_s)) \tag{3.6}$$

So the observations take

$$\begin{gathered} s(t_0 + k\tau_s) \rightarrow s(t_0 + (k+1)\tau_s), \\ s(k) \rightarrow s(k+1) \end{gathered} \tag{3.7}$$

Phase Space Embedding

A state space is the set of all possible states or configurations available to a system. For example, a six sided die can be in one of six possible states, where each state corresponds to a face of the die. Here the state space is the set of all six faces. As a musical example consider a piano keyboard with 88 keys. All possible combinations of chords (composed from 1 to 88 keys) in this keyboard constitute the state space and each chord is a state. State spaces can be represented geometrically as multidimensional spaces where each point correspond to one and only one state of the system.[2]

The dimensions of the state space depends on the dimensions of the system, however it is possible and in many times useful to embed the state space into a representation of a higher number of dimensions. Sometimes, the new representation allows to bold -or even discover- information not clearly present in the original representation. In our context, the phase space embedding will help to generate, visualize, and compare the development in the structure of the music studied.

> The theorem called the embedding theorem attributed to Takens and Mañé states that it is possible to go from scalar observations $s(k) = s(t_0 + k\tau_s)$ to multivariate phase spaces. Imagine that we have a dynamical system $\boldsymbol{x}(n) \rightarrow \boldsymbol{F}(\boldsymbol{x}(n)) = \boldsymbol{x}(n+1)$ where $\boldsymbol{x}(t)$ phase space is multidimensional. The theorem tells us that if we are able to observe a single scalar quantity $h(\bullet)$, of some vector function of the dynamical variables $\boldsymbol{g}(\boldsymbol{x}(n))$, then the geometric structure of the multivariate dynamics can be *unfolded* from this set of scalar measurements $h(\boldsymbol{g}(x(n)))$ in a space made out of new vectors with components consisting of $h(\bullet)$ applied to powers of $\boldsymbol{g}(x(n))$. These vectors
>
> $$\boldsymbol{y}(n) = [h(\boldsymbol{x}(n)), h(\boldsymbol{g}^{T_1}(\boldsymbol{x}(n))), h(\boldsymbol{g}^{T_2}(\boldsymbol{x}(n))), ..., h(\boldsymbol{g}^{T_{d-1}}(\boldsymbol{x}(n)))] \tag{3.8}$$
>
> define motion in a d-dimensional Euclidean space. With quite general conditions of smoothness on the functions $h(\bullet)$ and $\boldsymbol{g}(\boldsymbol{x})$, it is shown that if d is large enough, then many important properties of the unknown multivariate signal $\boldsymbol{x}(n)$ at the source of the observed *chaos* are reproduced without ambiguity in the new space of vectors

> $\boldsymbol{y}(n)$. In particular, it is shown that the sequential order of the points $\boldsymbol{y}(n) \rightarrow \boldsymbol{y}(n+1)$, namely, the evolution i time, follows that of the unknown dynamics $\boldsymbol{x}(n) \rightarrow \boldsymbol{x}(n+1)$.[1]

Phase Space Reconstruction[4]

So far, we have seen some important elements, first that a state space can be represented in spaces of higher dimensions for purposes of observation, and second that the dynamics of an autonomous system are defined only in terms of the states and not in terms of time. The last concept to understand is that of reconstructing a phase space of certain system from which we have access only to some of its dimensions. How can we do this?

> To implement the general theorem we can use for the general scalar function $(\bullet)$ the actual observed scalar variable $s(n)$
>
> $$h(\boldsymbol{x}(n)) = s(n) \tag{3.9}$$
>
> and for the general function $\boldsymbol{g}(x)$ we may choose the operation (*method of delays*) which takes some initial vector $\boldsymbol{x}$ to that vector one time delay τ_s later so the T_k^{th} power of $\boldsymbol{g}(\boldsymbol{x})$ is
>
> $$(g)^{T_k}(\boldsymbol{x}(n)) = \boldsymbol{x}(n+T_k) = \boldsymbol{x}(t_0 + (n+T_k)\tau_s) \tag{3.10}$$
>
> then the components of $\boldsymbol{y}(n)$ take the form

[4] Do not get confused with the notion of *embedding* a state space and *representing* a state space in a higher dimension.

$$y(n) = [s(n), s(n + T_1), s(n + T_2), ..., s(n + T_{d-1})] \tag{3.11}$$

If we make the further useful choice $T_k = kT$, that is, time lags which are integer multiples of a common lag T, then the data vectors $\boldsymbol{y}(n)$ are

$$y(n) = [s(n), s(n + T), s(n + 2T), ..., s(n + Td - 1)] \tag{3.12}$$

composed simply of time lags of the observation at time $n\tau_s$

These $\boldsymbol{y}(n)$ replace the scalar data measurements $s(n)$ with data vectors in a Euclidean d-dimensional space in which the invariant aspects of the sequence of points $\boldsymbol{x}(n)$ are captured with no loss of information about the properties of the original symbol.[1]

Calculating the appropriate embedding dimension and the tag delay is not straight forward. This is the central issue of reconstructing the state space and several techniques have been proposed. A simple way of estimating a good embedding dimension is to use the dimension where the correlation sum is very small in proportion to the total number of points. The correlation sum counts the number of distances smaller than ϵ between all pairs of points $(\boldsymbol{x}_i, \boldsymbol{x}_j)$[2]. Another method that can be used for estimating the embedding dimension is the *false nearest neighbors* method [37]. The method consist of comparing the distance between each point and its nearest neighbor in a m dimensional space against the distance between the same points in a $m + 1$ dimensional space. If the distance between the pair of points grows beyond a certain threshold this means that such points are false nearest neighbors that are not close each other

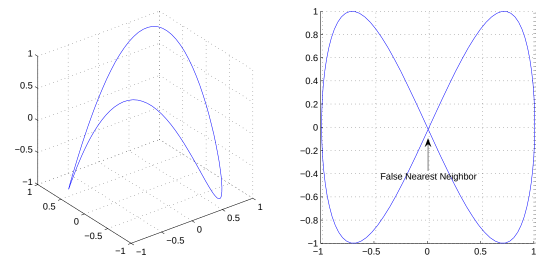


Figure 3.1: False Nearest Neighbor.

in the bigger dimensional space. In figure 3.1 it is possible to observe that the data in a two dimensional space generates two points at the same place while the same data in a tree dimensional spaces does not.

Cao [12] proposed another method of false nearest neighbors. The threshold is eliminated by considering only the average of all the changes in the distance between all points as the dimension increases, and then taking the ratio of these averages. The mean is defined as

$$E[m] = \frac{1}{N} \sum_{alln} \frac{\|\boldsymbol{x}_n^{(m+1)} - \boldsymbol{x}_{(1)n}^{(m+1)}\|}{\|\boldsymbol{x}_n^{(m)} - \boldsymbol{x}_{(1)n}^{(m)}\|} \tag{3.13}$$

and $E1(m) = E[m+1]/E[m]$ as the ratio between the averages of consecutive dimensions. The ratio function $E1(m)$ describes a curve that grows slower as m increases, asymptotically reaching 1. The value of m where the growth of the curve is very small is the best embedding dimension.[2]

Finally, calculating the time lag τ, which in theory is irrelevant since the reconstructed space is topologically equivalent turns out important for model estimation and prediction. A measure of mutual information between a series and itself delayed by τ provides a reasonable solution [1].

Let $p_{s_n}(i)$ be the probability that the series $s[n]$ takes the value i at the time n, and $p_{s_n,s_{n-\tau}}(i,j)$ be the probability that the series $s[n]$ takes the values i at time n and j at time $n-\tau$ for all n. The mutual information is then[2]:

$$I(s_n; s_{n-\tau}) = \sum_i \sum_j p_{s_n,s_{n-\tau}}(i,j) \log_2 \frac{p_{s_n,s_{n-\tau}}(i,j)}{p_{s_n}(i)^2} \tag{3.14}$$

3.2.2 Principal Component Analysis

The main purpose of this technique is that of decorrelating a set of correlated variables and simplifying a dataset.

It is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset

> that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the 'most important' aspects of the data. But this is not necessarily the case, depending on the application.[80]

The Principal Component Analysis is of popular use in the field of signal processing[5]. The *covariance matrix* is a nice way to organize the covariance of multidimensional information. Just to remember, the *variance* is a measure of the spread of data in a data set of one dimension. It is almost identical to the *standard deviation*

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)} \tag{3.15}$$

thus, the covariance is a slight modification of the variance but measured between two dimension

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \tag{3.16}$$

The *covariance matrix* is then a matrix of the covariance of all the dimensions of the data against each other.

$$C^{m*n} = (c_{i,j}, c_{i,j} = cov(Dim_i, Dim_j)) \tag{3.17}$$

[5] [67] is a good tutorial for PCA.

where C^{m*n} is a matrix with n rows and n columns, and Dim_x is the xth dimension.

Eigenvectors is a special case of matrix multiplication where a matrix is multiply by a vector and the resulting vector is an integer multiple of the original vector. Such vectors can be though -and represent- an arrow pointing from the origin to the (x, y) value of the vector. The square matrix (in our case the covariance matrix) can be thought of as a transformation matrix. Thus, imagine the multiplication, as a transformation of the vector from it's original position. Each eigenvector is associated with an *eigenvalue*. They indicate the amount by which the original vector was scaled after the multiplication. There are as many eigenvectors as dimensions and all of them are perpendicular (*orthogonal*).

The steps for producing the PCA (as described on [67]) are therefore

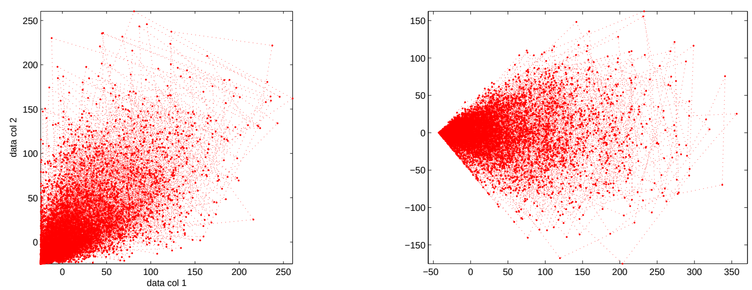
1. Subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. This produces a data set whose mean is zero.

2. Calculate the covariance matrix.

3. Calculate the eigenvectors and eigenvalues of the covariance matrix.

4. Order the data dimensions by their corresponding eigenvalue from highest to lowest. This gives the components in order of significance. Here is where you have to decide if you ignore the components of less significance -losing more and more information while discarding more and more components-. On the other hand, if all the important information is the initial components you may end up with a compress version of the data but using less dimensions.

5. Form a *feature vector*. This is a matrix of the eigenvectors in the columns.



(a) Original Data. (b) Data after PCA.

Figure 3.2: Principal Component Analysis transformation of the energy of the Vienna I improvisation of Keith Jarrett.

6. The final step is to derive the new data set by multiplying the feature vector transposed and the mean-adjusted data also transposed.

3.2.3 Dynamic Time Warping

Dynamic Programming

In Computer Sciences, Dynamic Programming (DP) is a method, - a set of algorithms- that reduce the time for solving tasks that can be broken down into subproblems which are reused several times. The requirement is that the sub-problems can be used to find the optimal solutions of the overall problem. Thus, an optimization is obtained using recursion. The three-step process to solve a problem with optimal substructure (according to [79]) is then:

1. Break the problem into smaller subproblems.
2. Solve these problems optimally using this three-step process recursively.
3. Use these optimal solutions to construct an optimal solution for the original problem.

There are two basic approaches for Dynamic Programming routines[79]:

1. Top-down approach: The problem is broken into subproblems, and these subproblems are solved and the solutions remembered, in case they need to be solved again. This is recursion and *memoization* combined together.
2. Bottom-up approach: All subproblems that might be needed are solved in advance and then used to build up solutions to larger problems. This approach is slightly better in stack space and number of function calls, but it is sometimes not intuitive to figure out all the subproblems needed for solving given problem.

There are several algorithms that uses DP, the one we are interested on is *Dynamic Time Warping*.

Dynamic Time Warping

Dynamic time warping (DTW) is an algorithm for measuring similarity between two sequences which may vary in time or speed. It is typically used for text processing, and alignment and similarity of biological sequences, such as proteins or DNA sequences. A well known application has been automatic speech recognition, to cope with different speaking speeds. In general, it is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions, i.e. the sequences are "warped" non-linearly to match each other. DTW has been superseded in several fields by Hidden Markov Models (HMMs), a related method which uses training data to learn optimal parameters settings.

On [19] Simon Dixon gives a clear description of the technique, he explains that

> DTW aligns time series $U = u_1, ..., u_m$ and $V = v_1, ..., v_n$ by finding a minimum cost path $W = w_1, ..., w_l$, where each W_k is an ordered pair (i_k, j_k), such that $(i, j) \in W$ means that the points u_i and v_j are aligned. The alignment is assessed with respect to a local cost function $d_{U,V}(i, j)$, usually represented as an $m*n$ matrix, which assigns a match cost for aligning each pair (u_i, v_j). The cost is 0 for a perfect match, and is otherwise positive. The path cost $D(W)$ is the sum of the local match cost along the path:
>
> $$D(W) = \sum_{K=1}^{l} d_{U,V}(i_k, j_k) \tag{3.18}$$
>
> Several constraints are placed on W, namely that the path is bounded by the ends of both sequences, and it is monotonic and continuous. Other local path constraints are also common, which alter the monotonicity and continuity constraints. Additionally, global

path constraints are often used which constraints the path to lie withing a fixed distance of the diagonal. By limiting the slope of the path, these constraints prevent erratic solutions and reduce the search space.

The minimum cost path can be calculated in quadratic time by dynamic programming, using recursion:

$$D(i,j) = d(i,j) + \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} \tag{3.19}$$

3.3 Machine Learning and Data Mining

Machine Learning (ML) is a generic name for a set of several statistical procedures -some of them extremely complex and some very simple- for obtaining meaningful knowledge from data sets[6]. In practice, a computer is always involved in the analysis due the huge amount of data and mathematical operations that are involved. In fact, the jump between statistics and ML is precisely, that the last one grew and evolved tided to Computer Science[7].

Data Mining (DM) is a subtle variation of the Machine Learning field. The core concepts, procedures, and techniques are pretty much the same in both areas. The slight difference, is that DM is more focused on the analysis of *raw* data of *extensive* size while ML may be also interested in the understanding and/or modeling of systems. However, as already mentioned, the difference for

[6]The practical application of these techniques is broad. They are constantly apply in financial, medicine and biology, pattern recognitions, humanities, and of course in art.

[7][3] is a good introduction to the field.

our particular interests is so subtle, that we will speak about them as the same field.

With ML we can build numerical models that can explain our data, make predictions, classify our content, recognize patterns, extract knowledge, or compress the amount of data among others. Formally, the model may be *predictive* if the goal is to make predictions in the future, or *descriptive* to gain knowledge from data.[82]

3.3.1 Sensing and Modeling Network Interaction

An inspiring work to me and another starting point for the research described here is the work of Tanzeem Khalid Choudhury[14]. On her PhD thesis she developed a computational framework for learning the interaction structure and dynamics for people interaction automatically.

> Knowledge of how groups of people interact is important in many disciplines, e.g. organizational behavior, social network analysis, knowledge management and ubiquitous computing. Existing studies of social network interactions have either been restricted to online communities, where unambiguous measurements about how people interact can be obtained (available from chat and email logs), or have been forced to rely on questionnaires, surveys or diaries to get data on face-to-face interactions between people. The aim of this thesis is to automatically model face-to-face interactions within a community. The first challenge was to collect rich and unbiased sensor data of natural interactions. The "sociometer", a specially designed wearable sensor package, was built to address this problem by unobtrusively measuring face-to-face interactions between people. Using the sociometers, 1518 hours of wearable sensor data from 23 indi-

> viduals was collected over a two-week period (66 hours per person). This thesis develops a computational framework for learning the interaction structure and dynamics automatically from the sociometer data. Low-level sensor data are transformed into measures that can be used to learn socially relevant aspects of people's interactions - e.g. identifying when people are talking and whom they are talking to. The network structure is learned from the patterns of communication among people. The dynamics of a person's interactions, and how one person's dynamics affects the other's style of interaction are also modeled. Finally, a person's style of interaction is related to the person's role within the network. The algorithms are evaluated by comparing the output against hand-labeled and survey data.[14]

3.3.2 Decision trees and Classification rules

On of the most rudimentary techniques of ML are the *Decision trees*. In this technique the output from the algorithm is a table of decisions. This type of algorithms look up for the appropriate conditions that best define the behavior of the raw data. If, for example, our data set shows that *all* the time that a particular instrument plays loud, the rest of the members stop playing, then this condition is set as a condition with a high value. On the other hand, if this happens only 10% of the times this condition is less influential or even discarded. The problem is to decide which attributes to leave out without affecting the final decision and also how to create the rules.

> Decision trees is a "divide-and-conquer" approach to solve a learning problem. Nodes in a decision tree involve testing a particular attribute. Leaf nodes give a classification that applies to all instances that reach the leaf, or a set of classifications, or a prob-

> ability distribution over all possible classifications. To classify an unknown instance, it is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf.[82]

Classification rules are a popular alternative to decision trees. In this case, a set of tests in the form of logical operators are evaluated in order to generate a conclusion. The conlcusion is a class or classes that are applied to instances covered by the rule. In the practice, conflicts arise when several rules with different conclusions can be applied and usually a process of pruning is required. *Association rules* are a type of classification rules that can predict not only the class, but also any attribute which allows the prediction of combinations of attributes.

3.4 Tools

The essentia project

The *essentia* framework is a python[8] framework for analysis and classification of acoustic sources. It is the evolution of audioclass[9].

> Audioclass is a is a Eureka project that addresses the needs of post-production audio studios to access huge collections of audio samples in a precise, effective and efficient way. Indeed, sound producers create the sound that goes along the image in cinema and

[8] www.python.org

[9] Audioclass was developed by the Music Technology Group, Barcelona: http://www.iua.upf.es/mtg; the Tape Gallery, London: http://www.sound-effects-library.com; and the DUY Research, Barcelona: http://www.duy.com/

> video productions, as well as spots and documentaries. Some sounds are recorded for the occasion. Many occasions, how ever, require the engineer to have access to massive libraries of music and sound-effects. Audioclass explores means to search and explore the collection the appropriate sounds enhancing the search engines, currently based on classic text-retrieval techniques, with technologies that actually listen to sound. The technologies, normally experimented in academic environments are to be used in real industrial production systems. Of the three major facets of audio in post-production: music, speech and sound effects, the project concentrates on Music and Sound Effects.[4]

Essentia includes a set of low-level audio descriptors, and a set of high-level audio descriptors. Some of these descriptors can be found in other systems and are standardized procedures. Others descriptors are experimental descriptors developed by the essentia team. Most of the extraction of audio descriptors for the experiments and research described on this paper was employing this framework.

3.4.1 Weka

All the machine learning procedures applied in this work were using the *Weka* workbench[10]. The Weka workbench is a collection of state-of-the.art machine learning algorithms and data preprocessing tools and includes all the common algorithms for standard ML and DM problems: regression, classification, clustering, association rule mining, and attribute selection[11].

[10] http://www.cs.waikato.ac.nz/ml/weka/.
[11] [82] is a good source to deep into weka.

3.5 Conclusion of the Chapter

On this chapter, we saw that there are several tools from the signal processing and the machine learning fields useful for our analytical purposes. Loudness and spectral centroid were taken as two basic audio descriptors exploitable for our research. Phase Space Reconstruction, Principal Component Analysis, Dynamic Time Warping, Decision Trees, and Association Rules are few of the vast techniques for analyzing our chosen audio descriptors. The Essentia and the Weka projects are two useful and practical frameworks to work with.

Chapter 4

The EnsAmble Crumble

"Notation is to improvisation as the portrait to the living model."
Ferruccio Busoni

The EnsAmble Crumble is an ensemble of gestural improvisation created at the beginning of 2005. Its music is employed as a case-study for the current research. In this chapter, the musical characteristics of the ensemble are presented. They are preceded by a section of previous related works.

4.1 Previous work

In this section, a short description of previous works that have led to the current research is presented. Some of them are purely musical, others are technical, and some are an imbrication of art and technology.

Figure 4.1: NICROM presentation and Four paintings by Mauricio Zárate created during a concert.

4.1.1 NICROM

In 2001, while I was a last-year student at the National School of Music at Mexico City, I founded and played in a trio for improvising electro-acoustic music with action-painting. We were two musicians and one painter. Rodrigo Garibay played wind instruments, I played synthesizers and effects processors, and Mauricio Zárate experimented with watercolor and oil.

The concerts consisted of a one hour-long gestural improvisation session with a projection on stage of the material that the painter was creating in real time (fig. 4.1(a)) [28]. The relationship between the audio and the visual material was subjective and based solely on the human perception of the members. It was, in this way, a human feedback between both mediums: The painter listened to the acoustic material and used it to continue the painting; the musicians looked at the painting and used it as a base from which to generate music.

The kind of graphic language that the painter created was abstract and a lot of emphasis was put in the use of color (fig. 4.1(b)) [28]. The physical gesture used in the creation of a stroke was an important criteria in relating the visual element with the music. This relationship got lost as soon as the stroke was finished but the final painting preserved at least in a subjective way the acoustic intention.

The music that was produce was, in general, non-tonal material with a strong emphasis on the timbrical quality of the music. Many parts were also noisy with irregular and inconstant rhythmic patterns. The clarinet and the saxophone were digitally processed in conjunction with the electronic material created with the synthesizers and the computer.

Because there were no agreements before the performance and all the decisions were made during the concerts by a subjective interpretation of the material, the level and type of interactions between the members varied freely over time. There were not assigned roles. At any moment one member could be leading the improvisation, and later on be guided by other members. Sometimes there was no clear leader, creating in this way a segment consisting of three independent lines.

During a concert, between 15 and 20 paintings were created and about 10 musical ideas developed. Some moments of the performance used to be solid and coherent but there were also segments where there was not any recognizable idea. Even though the transitions between segments were smooth and fluid, the general shape of the improvisation was on many occasions uninteresting.

It was this lack of control over the big structures that made me think of the idea of creating a system that could help in creating better-structured improvisation.

Figure 4.2: Interface and controller of the GAB system.

4.1.2 GAB

GAB is an electronic system developed in 2001 as final project for the bachelors degree at the National School of Music in Mexico City after a scholar residency at New York University. It allows the reinterpretation of musical material in real time by a pianist improviser [68]. It was developed with JMSL [18] and it was extensively used during the Nicrom performances.

GAB was designed to be used by a pianist during the performance of musical improvisations. The system consists of a computer program written in Java that performs all the mathematical calculations required to produce the reinterpreted material, and a small controller enabling the pianist to modify the variables of the system (figure 4.2). GAB attempts to combine the design and use of musical instruments with improvisation as a creative means of expression.

GAB is composed of a software program and a hardware device. The software is a Java program that reads the input data in MIDI format; analyzes this information; creates new material according to some algorithms; and finally outputs the new material on several MIDI channels. The algorithms create the new material based on a set of modifiable values that are established in real time with the controller.

The controller is a device designed to manipulate and control the software of the system. It uses a *BasicStamp*[20] microcontroller in order to read the values of the different potentiometers and convert them into a digital signal formatted according to the MIDI specification[1]. It contains eight knobs and eight 1/4 inch plug-ins to connect variable controllers such as dynamic pedals, pressure sensors, light sensors, etc. It sends 16 signals containing continuous MIDI messages in a fixed channel.

The objective of the system was to create a tool that enriches the pallet of possibilities that a pianist performer could have when an improvisation is developed. The intention was that the acoustical result would be completely dependent of the material of the input material. In this way, the system would expand the particular language that the musician would be using. The intention was not to create a random generator with autonomous processes, but more an accompanying device. GAB can be thought as an ensemble of virtual pianists that can create variations of the original material.

4.1.3 Improvisatory Music and Painting Interface

The *Improvisatory Music and Painting Interface* (IMPI) is an audiovisual system developed as final project for the Masters degree at the MIT Media Laboratory. It was created with two main goals first, to assist ensembles of acoustic

[1] Nowadays, I would not recommend the use of the BasicStamp. There are better alternative such as using directly the *Pic* microcontroller or using devices with strong community feedback as *Arduino*.

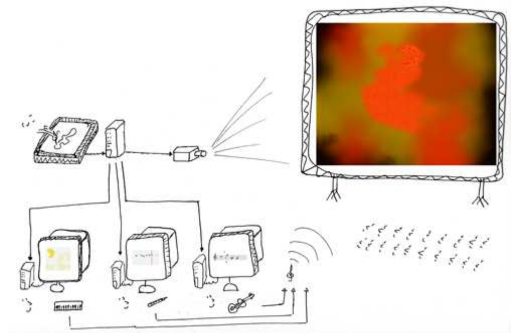


Figure 4.3: Improvisatory Music and Painting Interface.

musicians in the creation of organic improvisations where elements of the compositional technique, such as synchronization between instruments and abrupt changes of the material, could be created. Second, to create a dynamic graphical expression that, in conjunction with the music, produces a coherent and malleable audiovisual work with perceptual meaning between all the elements.

In the versions that have been developed, a "conductor" is in charge of shaping the improvisations and creating the visual material by drawing with a specific syntax on a digital tablet. With the objective of recognizing the gesture of the stroke, several elements such as speed, pressure, location, and duration are analyzed. In the versions developed, the musical elements that the "conductor" can control are the level of participation and activity of each member of the ensemble, the general contour of the improvisation, the levels of intensity, and the pitch regions of the music. The visual elements that can be controlled by

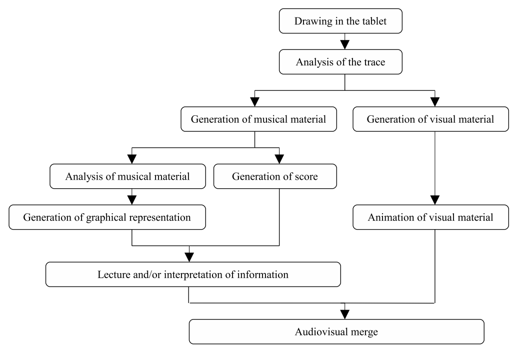


Figure 4.4: Diagram of IMPI's functions.

the conductor are the shape, the location, and the color of the objects. The rest of the functions are automated using preestablished functions.

The musical information generated by the conductor is distributed and sent to the ensemble by a computer network. Ideally, each musician should have a computer to read from, substituting for the traditional paper score. On the screen, each musician receives the material that the conductor is creating for him or her. The information is represented with several degrees of abstraction, and musicians are free to choose which representation to follow. Standard Music

Notation (SMN) is one of the possible representations, and should be sight-read while it is scrolling through the window. Dynamic shapes that represent musical content is another type of symbology. In this case, the player has the potential for greater freedom; however, the shape of the improvisation is preserved.

The visual information is also generated by the drawings of the conductor. The gestures used in the drawings produce dynamic images that should be projected on the stage with the music. In the implemented versions, the traces created by the conductor are directly mapped to the shape of the object created in the main visual output. Dynamic filters and transformations are then applied to the image to generate blurring and merging among the shapes.

The kind of music that the system should ideally generate is non-tonal, non-rhythmic, and pattern oriented, which means that strong emphasis should be put on the control of timbral qualities and global textures. Because the interface is for acoustic instruments and there is particular interest in the timbre element, the instruments should use all their resources. Extended techniques are ideal resources to represent the intention. In order to create the sensation of continuum transformations in the pitch domain, intervals smaller than the semitone should be employed. If the instrument allows glissandi, they should also be used.

In order to create the sensation of movement over space in the musical domain, IMPI should use fundamental transformations in pitch and intensity. In order to create the sensation of continuum transformation in the time domain, the synchronicity of events and the use of regular pulses or clear rhythmic patterns should be carefully employed. Politempis, irregular rhythmic patterns, and avoidance of repetitions should also be considered. IMPI has its own paradigm in the way it creates the translation of these ideas into acoustic and visual materials.

4.2 The EnsAmble Crumble

The EnsAmble Crumble[2] is an electro-acoustic improvisatory ensemble formed at the beginning of 2005. Most of the members belongs to the Music Technology Group of the UPF and most of us have interdisciplinary studies in music and sound technology. The ensemble was created as an opportunity to develop musical ideas in the form of free electro-acoustic improvisations. Three concerts have been performed in different venues of Barcelona.

The initial desire was that each member could merge his or her instrument with digital technologies. However after some months the group took an acoustic rout and only the author kept working with electronic devices. During the year and a half that the group has been playing two members have leaved the ensemble and two new members have been integrated in different moments. Nevertheless, all the recordings that were selected for analysis have been done with the same members.

The work procedures of the ensemble are traditional following common methods of rehearsal. The members join together once or twice a week and play for about two or three hours. All the rehearsals take place in the IUA-Phonos audio studio. About tree long improvisations are performed on each session. After each improvisation, an informal discussion take place where each member gives his or her own impression of the result.

One of the biggest issues that the ensemble has faced is the tendency of merging sections of music that do not have relationship between each other. Several times, the members think that the improvisations lost direction, get stoke, and became redundant. For this reason, in some rehearsals specific exercises have been developed with the intention to solve such lack of direction. Most of these exercises impose certain constrain to the improvisation such as

[2]http://www.iua.upf.es/ hsolis/EnsAmbleCrumble/

“everybody improvise in *piano*”, ”avoid playing if more than two players are improvising in that moment”, “improvise by responding only to the impulse of only and only one other player”.

As explained on chapter 2, there are several types of improvisations. The EnsAmble Crumble has a structurized-improvisation language which means that it aims for create and developed musical ideas that have exposition, development, realization, and conclusion. To reach this goal, another type of subjective experiments have been realized. Metaphors and sonification of histories have been realized such as improvise for two minutes representing how a glass falls from a table to the floor and breaks in multiple fragments; improvise representing the inner movement of a river and how the river moves floating objects in the surface.

4.2.1 Current Members

In order to get a better idea about the ensemble, a brief description of the instruments and the current members is presented. In addition, on table 4.1 and table 4.2 information about the musical studies, experience, and expertise is presented.

The *Guitar A* in the data-set is an electric guitar played by Maarten Grachten (The Netherlands, 1976). The instrument was never processed electronically and most of the time played in a “semi-standard” manner. However, sometimes extended techniques that enrich the sound pallet were observable. Glissandi and moving of tuning heads were frequent. Scratching of the winding strings with the nail and external objects were sometimes employed.Maarten Grachten started playing guitar at age ten, and took both classical and electric guitar lessons for several years. From 1989 to 2001 he has played guitar and drums in various rock-experimental groups, often with a focus on improvisation.

The *Guitar B* is also an electric guitar played by Sylvain Le Groux (France,

member	INS	E-ENS	E-EXP
guitar A	Guitar 19, drums 5, bass 5.	Expressive-progressive, rock-metal-jazz: 16 years total in 8 different groups.	Some improvisation in rock ensembles.
guitar B	Guitar 14, flute 12.	Jazz band (guitar): 3 years, classical orchestra (flute): 5 years.	Some.
synth.	Piano 15.	Classic Rock Jazz and Contemporary ensembles.	2 years in experimental groups.
violin	Violin 18.	Orchestra and camera 15 years, pop-folk 3 years.	1 year in an experimental group.
viola	Viola 15.	Orchestra and camera 10 years.	No.

Table 4.1: Information about the players. INS: Instruments that plays and years of practice. E-ENS: Experience in musical ensembles. E-EXP: Experience in experimental ensembles.

1977). This instrument was played with few effects from the amplifier and was played most of the time with techniques that produced noisy and fragmented sounds. The vibrato bar was extensively employed and also the use of metal objects to pluck the strings. Sylvain Le Groux studied classical flute and Music Theory at the National Conservatory of Nantes. He studied basic harmony and counterpoint at Paris Conservatory and Electroacoustic music at Studio Delta P (La Rochelle, fr.).

The *Synthesizer*[3] is a kurzweil K2000 controlled with a M-Audio Oxone keyboard played by Hugo Solís (Mexico City, 1976). The sound-bank during all the improvisations was limited to ten sounds, however the timbre of each sound could be drastically altered with the knobs of the controller. All the sounds share a noisy quality and some could produce semi-rhythmic patterns that were moderately employed. None of the sounds had a long enveloped that could generate dense textures without a physical effort of the player. Hugo Solís García studied piano performance, composition and music technology at the UNAM in Mexico City. He has played piano and electronics throughout Mexico and has collaborated in many interdisciplinary projects in conjunction with dancers, painters, filmmakers and radio-artists.

The *Violin* was played by Alfonso Pérez (Spain, 1977). The instrument was played most of the the time in a standard manner. However, glissandi and natural harmonics were sometimes employed . The instrument was amplified with a variable gain which allows to pick up soft sounds. In some moments, the instrument's body was even used as percussive resonator with the fingers. Alfonso Pérez studied Violin at the conservatory of Palencia and Computer Science at the University of Valladolid (Spain). His background in music is mainly with classical and folk music. He has played in several orchestras and music ensembles in Spain and Austria.

[3]Called also synth, keyboard, or electronics indistinctly over the document.

	IN-L	SO-L	HA-L	CO-L	MT-L	CP-L	RMS	WMS	PP
gtrA	5	1	basic	no	no	no	30	basic	no
gtrB	12	8	basic	2	advance	basic	90	basic	no
synth	10	4	advance	2	advance	advance	60	advance	no
vl	7	5	basic	no	medium	no	80	medium	no
vla	10	5	medium	no	medium	medium	80	medium	no

Table 4.2: Information about the players B. IN-L: Instrument lessons in years. SO-L: Solfege lessons in years. HA-L: Harmony lessons. CO-L: Composition lessons in years. MT-L: Music Theory lessons. CP-L: Counterpoint lessons. RMS: Reading Music Scores 1-100. WMS: Writing Music Scores. PP: Perfect Pitch.

The *viola* was played by Gabriela Villa (Mexico City, 1976). The instrument was played with extended techniques, natural and artificial harmonics were extensively employed. *Scordaturas*, *sul tasto*, *sul ponticello*, and *Col legno* were also used. The instrument was also amplified. Gabriela Villa has collaborated in many interdisciplinary projects and has been member of several orchestras from México. At the Conservatory of the Roses in Morelia México, she studied viola and music analysis. She did her basic musical studies at the Universidad Nacional Autónoma de México in Mexico City.

4.2.2 Music Data Base

A brief description and statistical information of the music used in the experiments are presented in order to give a better understanding of the type and kind of material.

The database is build of twenty one improvisations that the ensAmble Crumble recorded in studio from March to May of 2006. The improvisations where recorded on multichannel, one mono channel at 44.100 16 bits for each instrument. All the improvisations were recorded with the same acoustic and technical conditions. The mixer, microphones, amplifiers, and instruments were always the same to preserve the same sonority. The electronic instrument employed a reduced amount of sounds for not altering the sound context. After all the improvisations were recorded. All the channels of all the recordings were initially normalized using a Root Mean Square normalization (RMS) to $-6dB$ in order to have a common value of reference.

The root-mean-square energy r an audio signal x of length n is defined as

$$r = \frac{\sqrt{\sum_{i=0}^{n-1} \|n\|^2}}{n} \tag{4.1}$$

In addition, in most of our experiments, the data has been smoothed by applying a low-pass filter.

$$y_t = c * x_t + (1 - c) * y_{t-1} \tag{4.2}$$

Where c is a value between 0 and 1 that set the amount of smoothness and t is time expressed in our case as a frame. With this technique we loose a little bite of temporal resolution but in our context is not a problem because we want to analyze long-term structures.

As can be seen in Table 4.3 between two and three improvisations per week were recorded. The length of the pieces goes from four minutes to forty six minutes. The mean of the length is twenty four minutes and the standard deviation about thirteen. The synth and the viola are present in all the improvisations and the guitar A in all minus one. Seven improvisations contains the entire ensemble and sixteen has four or more than four players[4].

[4]The sequential number of the improvisations is the identification tag for the pieces along the document.

Impro	Instruments	Duration	Date
01	gtrA, gtrB, synth, vla	30'09	04-03-06
02	gtrA, gtrB, synth, vla	10'12	04-03-06
03	gtrA, gtrB, synth, vla	09'27	04-03-06
04	gtrA, gtrB, synth, vl, vla	37'13	04-07-06
05	gtrA, gtrB, synth, vl, vla	13'18	04-07-06
06	gtrA, gtrB, synth, vl, vla	32'29	04-10-06
07	gtrA, gtrB, synth, vl, vla	18'07	04-10-06
08	gtrA, synth, vl, vla	45'01	04-20-06
09	gtrA, gtrB, synth, vl, vla	25'31	04-20-06
10	gtrA, gtrB, synth, vl, vla	17'51	04-26-06
11	gtrA, gtrB, synth, vl, vla	27'41	04-26-06
12	gtrA, synth, vl, vla	45'55	05-08-06
13	gtrA, synth, vla	27'30	05-11-06
14	gtrA, synth, vla	24'35	05-11-06
15	gtrA, synth, vla	08'11	05-11-06
16	gtrA, synth, vl, vla	39'21	05-19-06
17	gtrA, synth, vl, vla	09'03	05-19-06
18	gtrA, synth, vl, vla	35'07	05-25-06
19	synth, vl, vla	07'40	05-25-06
20	gtrA, synth, vl, vla	46'57	06-01-06
21	gtrA, synth, vla	4'14	06-01-06

Table 4.3: Information about the improvisations.

Chapter 5

Analysis and Understanding

"I like each one of Picasso's paintings by itself,
but I also like all the Picasso's work conceived as a
long route from which I know each one of the periods.
The famous metaphysical questions,
Where do we came from? and
Where are we going?, have in the art a
a clear and concrete sense,
and are not absent of answer."
Milan Kundera

In this chapter the experiments developed during the research are presented. First, the behaviors of one player in one performance are studied. Not only because musically is simpler to study single fragments but also as an introduction to the basic techniques. Later, the overall behavior of one performance alone the twenty one improvisations is studied in order to detect global conducts of musicians. Finally, the behavior of the musicians is studied inside the entire ensemble context.

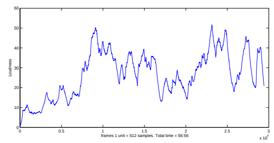
5.1 Analysis of the audio signal

5.1.1 One player, one performance

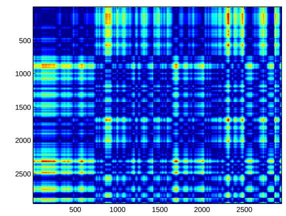
Loudness and Spectral Centroid

As a first practical example of how to implement the low-level techniques, described in the Background chapter, and obtain musical conclusions the improvisation that Ramón Lopez and Daniel Humair performed in February 8th at the Metronom Gallery at Barcelona is analyzed. I decided to use this material because this one hour long improvisation with two drums sets uses a musical language that is far away from the tradition concept of "percussion music" and it is a challenge to computer listening. The pulse of the music is used to create timbre changes and the flow of the music was obtained by continuous transitions of pitch-range, intensity, and timbre.

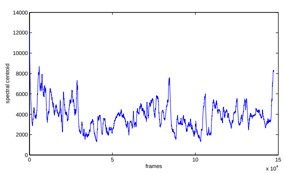
In the figure 5.1 it is possible to observe the evolution of the loudness and the spectral centroid during the improvisation. In both cases, the smooth was set to 0.0002 which is a good compromise between resolution and smoothness for this example. Both curves are presented with their self similarity matrices. For our initial analysis we observed that the loudness curve gives a better information than the spectral centroid because in the former the general curve increases among time and decrease at the end.



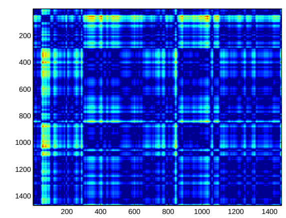
(a) Loudness Curve.



(b) Loudness SSM.



(c) Centroid Curve.



(d) Centroid SSM.

Figure 5.1: Evolution of loudness and spectral centroid of a one-hour improvisation.

Energy and Structure

On figure 5.1.1 there is a set of graphics all related with the energy development during the piano improvisation that Keith Jarrett gave on 1991 at the Vienna State Opera[1][30].

The energy E of a signal $y[n]$ for ith frame of the signal is calculated as follows:

$$E[i] = \sqrt{\sum_k y_i[k]^2} \tag{5.1}$$

On this experiment we embed the energy data into a state space. Figure 5.2(a) plots the energy over time, the duration of the performance was 42'05. Figure 5.2(b) shows us the curve that helps us to choose the best lag for our embedding. Remember from chapter 3 that the best lag is that where is the knee of the curve. In our case ten is a good value. Figure 5.2(c) shows a plot of possible dimensions from which we should choose the lowest dimension with an abrupt change in value. In our case, the best dimension is three which let us visualize the result without any reduction. Once that the lag and dimension are estimated we can produce the state space, figure 5.2(d) shows such state space embedding in a 3D plot. For visualization purposes only one of each 100 points is plotted. As it is observable, there is not a special curve or structure in the plot which indicates that the experiment do not gives valuable information[2]. Even

[1]From this improvisation, Keith Jarret says "I have courted the fire for a very long time, and many sparks have flown in the past, but the music on this recording speaks, finally, the language of the flame itself"[30].

[2]This may be arguable because this shape could be seen as a 3D static representation of the structure that could help in certain cases to compare musical pieces.

though, there are not important conclusion, we still presenting on figure 5.2(e) the same data after applying a principal component analysis that let us see the rotation of the structure in the axes where most of the data is aligned.

5.1.2 Overall performance of one player

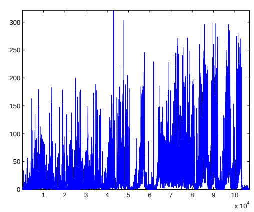
Another thread of exploration is to study the behavior of the musicians along the entire set of improvisations. Are there any well-defined patterns and/or constants in the performance of the players? Does the musician tends to react in the same way or create the same type of structures during each session?

We have seen in table 4.3 that the length of the improvisations is in average of twenty four minutes. Inside each long improvisation several ideas were produced. For these reason, the improvisations were segmented with an automatic process that search for pauses longer than ten seconds in the performance. Figure 5.3 shows the places of segmentation for improvisation one. In this case, six sections of about four minutes were founded.

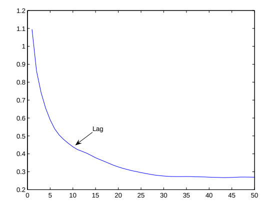
Table 5.1 shows the segments of each one of the twenty one improvisations obtained from the viola instrument longer than one minute and divided from at least five seconds.

75 segments (sections) were obtained from the segmentation. Each section was correlated with four *archetypical* shapes that represent types of evolution: *crecendo*, *decrecendo*, *stable piano*, and *stable forte* (figure 5.4). Table 5.2 presents the five sections of improvisation 1. The second column indicates the type of shape that was chosen after taking the maximum value of the four correlations. The confidence values are not conclusive.

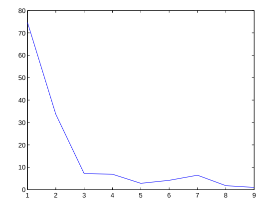
Because the confidence of the data is not significant. A second experiment was planned. From the four *archetypical* shapes a collection of four hundred similar shapes (one hundred for each shape) were created by multiplying the original shape with a vector of random numbers with a normal distribution of



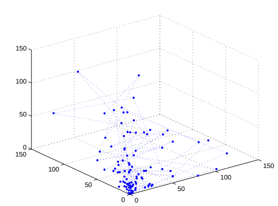
(a) Energy.



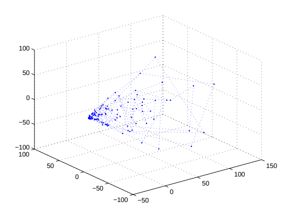
(b) Lag estimation.



(c) Dimension estimation.



(d) State space embedding.



(e) SSR after PCA.

Figure 5.2: Analysis of the energy of the improvisation *Viena I* by Keith Jarrett.

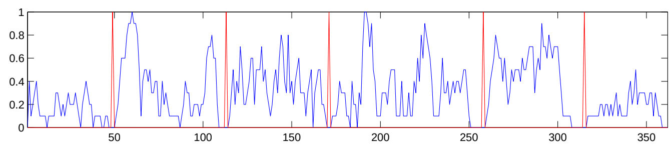


Figure 5.3: Viola's loudness of the entire improvisation 1 and the six segmented sections of about minutes each.

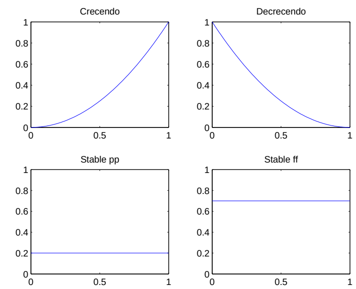


Figure 5.4: Four *archetypical* music shapes.

Improvisation	segments
1	305 290 405 255 225
2	65 190
3	80 295
4	395 130 65 220 725 640
5	545
6	160 305 90 310 65 170
7	395 145 315
8	405 470
9	810
10	110
11	480 80
12	530 90 80 425 115 195 290 70
13	75 95 225
14	385 75
15	270 95 75
16	200 80 320 90 145 150 275 295 105 265
17	135
18	160 240 85 320 215 615 130 95
19	100
20	790 150 250 425 130 175 315
21	255

Table 5.1: Number and duration in seconds of the segments of the 21 improvisations for the viola player.

Section	shape	confidence	crecendo	decrecendo	stable piano	stable forte
section A	crecendo	0.3145	13.3061	8.6994	8.7100	11.5900
section B	stable forte	0.2788	8.6366	8.0892	6.2400	8.8800
section C	stable forte	0.2953	12.7988	15.5188	9.4800	15.8400
section D	stable piano	0.3239	7.2260	7.3940	9.6900	5.6100
section E	stable forte	0.4018	7.8859	10.1950	0.8000	12.6800

Table 5.2: Shape of the five sections of improvisation 1 for the viola player.

one and a standard deviation of 0.5. With this technique the general shape was preserved but different instances created.

The collection was used as the training set for feeding the J48 classifier. The results for the classification and the cross-validations were:

```
=== Error on training data ===
Correctly Classified Instances         403               99.7525 %
Incorrectly Classified Instances         1                0.2475 %
Kappa statistic                          0.9967
Mean absolute error                      0.002
Root mean squared error                  0.0315
Relative absolute error                  0.5281 %
Root relative squared error              7.2667 %
Total Number of Instances              404

=== Confusion Matrix ===
   a   b   c   d   <-- classified as
 101   0   0   0 |   a = CRECENDO
   0 101   0   0 |   b = DECRECENDO
```

```
   0   1 100   0 |   c = STABLE_PP
   0   0   0 101 |   d = STABLE_FF

=== Stratified cross-validation ===
Correctly Classified Instances         394               97.5248 %
Incorrectly Classified Instances        10                2.4752 %
Kappa statistic                          0.967
Mean absolute error                      0.014
Root mean squared error                  0.1118
Relative absolute error                  3.7215 %
Root relative squared error             25.8171 %
Total Number of Instances              404

=== Confusion Matrix ===
   a   b   c   d   <-- classified as
 101   0   0   0 |   a = CRECENDO
   0 101   0   0 |   b = DECRECENDO
   1   2  96   2 |   c = STABLE_PP
   1   0   4  96 |   d = STABLE_FF
```

The obtained model was later used on the real data. The algorithm classified 73 of the 75 sections as *crecendo* with a confidence of 1.0. The remaining two sections were classified as *decrecendo* with 1.0 and 0.8 of confidence. The high but wrong values of one experiment, and the low values of the other are not promising. It is naive to shape structures of one minute or more with basic shapes.

5.2 Understanding the Musical Content

Finding long-term relationships among players during an entire improvisation is not an trivial task. The rest of the chapter, describes some experiments developed with the goal of finding patterns of interaction among the players.

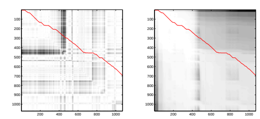
5.2.1 One player inside the ensemble

As it was seen in the chapter 3, the *Dynamic Time Warping* (DTW) is a method that can be used to get the optimal path between two sequences. Therefore, DTP can be used to get not only a measure about how similar two sequences are, but also, to see how far from a source sequences, a target sequence is in certain period of time.

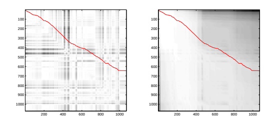
On figure 5.5 a series of five correlation matrices is displayed, each one with its correspondence accumulated distance matrix. The source of the data was taken from the improvisation number 10. In the x axis the loudness of the isolated instrument is displayed. On the y axes the mean of the loudness of the rest of the ensemble is displayed. Thus, we create a correlation matrix where differences of loudness between the particular instrument and the rest of the ensemble are clearly observable. For example, in figure 5.5(a) it is clear that the keyboard played with a different intensity than the rest of the ensemble at the first quarter of the improvisation. At the end of the performance shown in figure 5.5(e) the same behavior is also observable.

Over each matrix, there is a red line with a decreasing slop. It shows the optimal path obtained by the Dynamic Time Warping, all of them are similar but none of them create a complete diagonal. A DTW with constrains in the slope is therefore required.

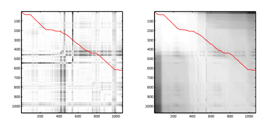
Another visualization technique that helps us to see the interaction between musicians is presented on figure 5.6. Here the first sixteen minutes of the im-



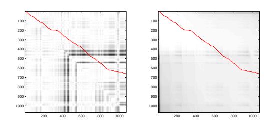
(a) Keyboard vs the rest of the ensemble.



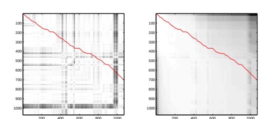
(b) Guitar A vs the rest of the ensemble.



(c) Guitar B vs the rest of the ensemble.



(d) Violin vs the rest of the ensemble.



(e) Viola vs the rest of the ensemble.

Figure 5.5: Correlation matrices between one player against the rest of the ensemble and optimal path using DTW.

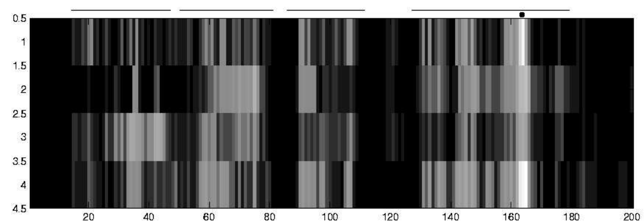


Figure 5.6: Grid of loudness interaction between the guitar A and the rest of the ensemble. First sixteen minutes of improvisation 4.

provisation 4 is presented. Each horizontal line correspond to the interaction between the guitar A against the other four members in the following order guitar B, synthesizer, violin, and viola. It is possible to see a high correlation between all the members in unit 160 (each unit equals five seconds) indicated by a small circle above the graphic. In addition, it is possible to see that the sixteen minutes could be dived into four sections of constant interaction dived with black regions. The sections are indicated by horizontal lines above the graphic.

With the same principle, figure 5.7 shows a network map based on the loudness of each player extracted from all the improvisations. x-axis and y-axis represents the five musicians in the traditional order guitar A, guitar B, synthesizer, violin, and viola. The color values indicates, for each musician, the amount of segments of five seconds where they have a similar level of loudness with each of the other subjects. From the image, it is observable that the viola was the musicians that played more time with the same loudness with the rest of the ensemble. The interaction between viola and keyboard was more permanent

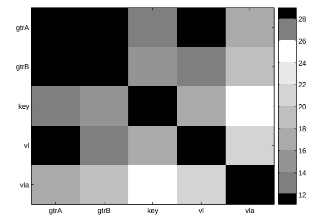


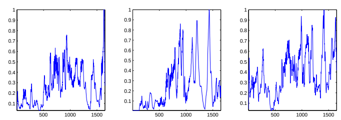
Figure 5.7: Network map based on loudness information of improvisation 4.

during the entire performance. On the other hand, it seems that guitar A, and guitar B, played more often in moments when the rest of the ensemble was not playing.

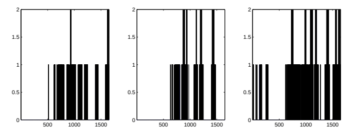
5.2.2 Do I play, you play, or we play?

On figures 5.8 and 5.9 the results of an initial experiment are presented. Figure 5.8(a) shows the loudness of guitar A, synthesizer, and viola during the improvisation 13 which has a length of 27'30 minutes. The data is discretized in three regions, *Not Significant* if the loudness is low, *Accompaniment* if the loudness is in the middle range, and *Soloist* if the loudness exceeds two thirds of the range. Soloist should be understand as *important* because it could be possible the have all the members in the soloist range. Figure 5.8(b) shows the discretized data.

The discretized data is then employed to build three decision-tree learners using the J4.8 algorithm one for each instrument of the ensemble with a cross-validation of 10. Figure 5.9 shows such trees. In the case of the guitar's tree



(a) Loudness Curves.

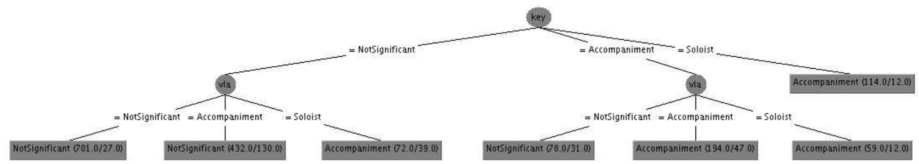


(b) Loudness States.

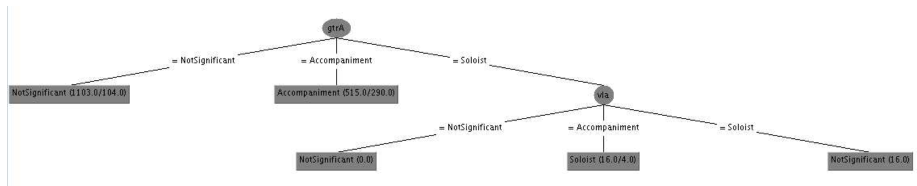
Figure 5.8: Loudness Curves and Loudness States from guitar, keyboard, and viola.

there were 1352 correctly classified instances (81.9394 %) and 298 incorrectly classified instances (18.0606 %). In the case of the keyboard tree there were 1252 correctly classified instances (75.8788 %) and 398 incorrectly classified instances (24.1212 %). Finally, in the viola's tree there were 1083 correctly classified instances (65.6364 %) and 567 incorrectly classified instances (34.3636 %). In this case, the guitar would be the best referential point to defined which other instrument may or may not play. However, we still need to know why, and when.

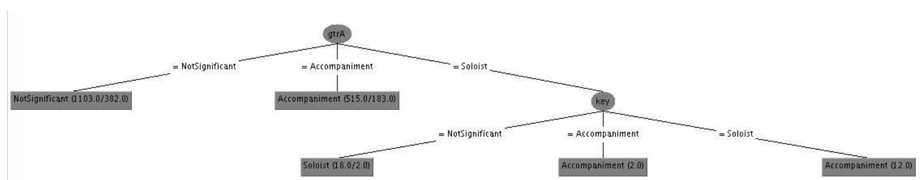
If the previous experiment gives an idea about relationships between musicians, the following experiments are refined explorations of the same idea with the aim of searching particular patterns of interaction. For these experiments, the data of all the twenty one improvisations was normalized, discretized, and segmented in fragments of two seconds giving a total of 15462 frames. Each frame containing ten parameters —loudness and pitch region for the five instruments. The possible values for the loudness and the pitch region are presented on table 5.3. Because the goal of the experiments was to model the behavior of



(a) Decision tree with the Guitar as root.



(b) Decision tree with the Keyboard as root.



(c) Decision tree with the Viola as root.

Figure 5.9: Simple Decision tree using the J4.8 algorithm.

Parameter	Possible Values
Loudness	? NP PPPP PPP PP P MP MF F FF FFF FFFF
Pitch range	? NR LLL LL L MLL ML M MH MHH H HH HHH

Table 5.3: Possible values for the discretized data.

Gtr A		Gtr B		Synth		Violin		viola	
Loudness	Pitch	Loudness	Pitch	Loudness	Pitch	Loudness	Pitch	Loudness	Pitch
NP	NR	?	?	PPPP	HHH	FFFF	LLL	MF	M
PP	LL	PPPP	LL	NP	NR	?	?	PPPP	ML
PP	LL	PPPP	L	NP	NR	PPPP	H	PPPP	MH
PPPP	L	PPPP	ML	PPPP	M	PPP	ML	PPP	ML
...	...	...	...	...	...	...	...	...	...

Table 5.4: Values of four frames of discretized data.

the entire ensemble, the data of the channels with an absent player was treated as *unknown* (?). *NP* meaning not playing; *NR* meaning not register because not playing; *PPPP PPP PP P MP MF F FF FFF FFFF* are a eleven-step musical analogy to music dynamics; and *LLL LL L MLL ML M MH MHH H HH HHH* segments the pitch-region range into three parts, low, medium, and high each with sub-regions. Table 5.4 shows an example of the data format where four frames are shown.

With this representation of our data, several experiments were developed that can be divided into two categories: *static models*, those where time is not consider; and *time aware models*, those where the moment in time of a particular interaction is taken into account. As experiments of the first model several association-rule learners were tested. Initially, the *Apriori* algorithm was ran over the entire data. As expected, the rules with a confidence value bigger than 0.9 were obvious, for example that there is not assigned pitch register if the musician is not playing. Therefore, we decided to search for more subtle rules. The next list shows the rules that the *PredictiveApriori* algorithm found with an accuracy bigger than 0.9 using only the pitch-range data. The interpretation of the results is disused on chapter 6.

```
 1. gtrA_C=LL gtrB_C=NR key_C=LLL vl_C=LL 13 ==> vla_C=L 13    acc:(0.99288)
 2. gtrA_C=LL gtrB_C=NR key_C=LLL vla_C=L 13 ==> vl_C=LL 13    acc:(0.99288)
 3. gtrB_C=L key_C=MLL vla_C=L 10 ==> gtrA_C=LL 10    acc:(0.99103)
 4. gtrB_C=NR key_C=LLL vl_C=LL 30 ==> vla_C=L 29    acc:(0.98978)
 5. gtrA_C=L key_C=ML vla_C=LL 7 ==> vl_C=L 7    acc:(0.98522)
 6. gtrA_C=ML gtrB_C=LL key_C=LL vl_C=ML 7 ==> vla_C=ML 7    acc:(0.98522)
 7. gtrA_C=H key_C=LLL vla_C=MHH 6 ==> gtrB_C=L 6    acc:(0.98067)
 8. gtrB_C=LLL key_C=L vla_C=L 6 ==> vl_C=L 6    acc:(0.98067)
 9. gtrB_C=L vl_C=ML vla_C=L 6 ==> key_C=NR 6    acc:(0.98067)
10. gtrB_C=M key_C=L vla_C=MLL 6 ==> vl_C=L 6    acc:(0.98067)
11. gtrA_C=LLL gtrB_C=L key_C=LL vl_C=L 6 ==> vla_C=MLL 6    acc:(0.98067)
12. gtrA_C=LL gtrB_C=LL key_C=LL vl_C=M 6 ==> vla_C=M 6    acc:(0.98067)
13. gtrA_C=LL gtrB_C=LL vl_C=M vla_C=M 6 ==> key_C=LL 6    acc:(0.98067)
14. gtrA_C=L gtrB_C=L key_C=ML vla_C=ML 6 ==> vl_C=L 6    acc:(0.98067)
15. gtrB_C=LL key_C=LL vl_C=M vla_C=M 6 ==> gtrA_C=LL 6    acc:(0.98067)
16. gtrA_C=L vla_C=HHH 5 ==> gtrB_C=L 5    acc:(0.9729)
17. gtrA_C=NR gtrB_C=LLL key_C=LLL 5 ==> vl_C=L 5    acc:(0.9729)
18. gtrA_C=LLL key_C=MHH vl_C=LL 5 ==> vla_C=L 5    acc:(0.9729)
19. gtrA_C=H vl_C=MH vla_C=NR 5 ==> key_C=NR 5    acc:(0.9729)
20. gtrA_C=HH gtrB_C=NR vla_C=HH 5 ==> key_C=NR 5    acc:(0.9729)
21. gtrB_C=NR key_C=MH vl_C=LL 5 ==> gtrA_C=LL 5    acc:(0.9729)
```

```
22. gtrB_C=NR key_C=HHH vla_C=NR 5 ==> gtrA_C=NR 5    acc:(0.9729)
23. gtrB_C=M key_C=NR vl_C=NR 5 ==> gtrA_C=NR vla_C=NR 5    acc:(0.9729)
24. key_C=MLL vl_C=LL vla_C=MLL 5 ==> gtrA_C=LL 5    acc:(0.9729)
25. gtrA_C=LLL gtrB_C=NR key_C=LLL vla_C=MLL 5 ==> vl_C=L 5    acc:(0.9729)
26. gtrA_C=LLL gtrB_C=LL key_C=MLL vl_C=L 5 ==> vla_C=L 5    acc:(0.9729)
27. gtrA_C=M vl_C=HHH 4 ==> key_C=NR 4    acc:(0.95875)
28. gtrA_C=NR gtrB_C=H vl_C=MHH 4 ==> vla_C=MH 4    acc:(0.95875)
29. gtrA_C=NR gtrB_C=HH vl_C=L 4 ==> vla_C=MLL 4    acc:(0.95875)
30. gtrA_C=NR key_C=L vl_C=H 4 ==> vla_C=L 4    acc:(0.95875)
31. gtrA_C=NR key_C=ML vl_C=HHH 4 ==> vla_C=NR 4    acc:(0.95875)
32. gtrA_C=NR vl_C=H vla_C=M 4 ==> gtrB_C=NR 4    acc:(0.95875)
33. gtrA_C=NR vl_C=HH vla_C=MH 4 ==> key_C=NR 4    acc:(0.95875)
34. gtrA_C=H vl_C=HHH 3 ==> key_C=NR 3    acc:(0.93098)
35. gtrA_C=HH vl_C=HH 3 ==> key_C=NR 3    acc:(0.93098)
36. gtrB_C=LLL vl_C=HH 3 ==> key_C=NR 3    acc:(0.93098)
37. gtrB_C=NR key_C=NR vl_C=NR vla_C=NR 63 ==> gtrA_C=NR 58    acc:(0.90791)
```

The results with the same data-set using the *tertius* algorithm.

```
Tertius
=======
 1. /* 0.198896 0.038539 */ vla_C = NR ==> gtrA_C = NR or key_C = NR
 2. /* 0.197989 0.084837 */ gtrA_C = NR ==> key_C = NR or vla_C = NR
 3. /* 0.195520 0.029227 */ vla_C = L ==> gtrA_C = LLL or gtrB_C = LLL or vl_C = LL
 4. /* 0.195478 0.031102 */ vla_C = L ==> gtrA_C = LLL or gtrB_C = HHH or vl_C = LL
 5. /* 0.193516 0.031038 */ vla_C = L ==> gtrA_C = LLL or gtrB_C = HH or vl_C = LL
 6. /* 0.193284 0.034336 */ vla_C = L ==> gtrB_C = HHH or vl_C = LL
 7. /* 0.193189 0.032202 */ vla_C = L ==> gtrB_C = LLL or vl_C = LL
 8. /* 0.191739 0.034077 */ vla_C = L ==> gtrB_C = HHH or key_C = HHH or vl_C = LL
```

```
 9. /* 0.191627 0.031749 */ vla_C = L ==> gtrB_C = HHH or key_C = LLL or vl_C = LL
10. /* 0.191549 0.030973 */ vla_C = L ==> gtrA_C = LLL or gtrB_C = H or vl_C = LL
```

5.3 Conclusion of the Chapter

Our data set can be represented in a suitable way to be analyzed by our studied tools. The data was studied with three different approach: the behavior of one musician during one improvisation, the overall behavior of one musician inside the entire data set, and finally the interactive patterns between musicians during the twenty one improvisations. Each approach offered answers to different questions that are described in detailed in the next chapter.

Chapter 6

Conclusions and Future Work

"Music is a personal force that resists processing, packing *or* understanding.*"*
Brian Whitman

6.1 Conclusions

In the previous chapters we set a musical framework to describe and bound our musical interest. Later, we explained the numerical tools we would used for our research. After that, we gave a description of our data and the characteristics of the ensemble that produce it. Finally, we applied the selected tools on the data and we got several results. What does these results show to us?

If I would have to give one and only one conclusion, I would agree with Brian Whitman; music resists processing, packing, and understanding[1]. The chosen audio descriptors, the representation of our data set, and the statistical tools among others decisions were the product of a trade off between pragmatism, knowledge, efficiency, and simplification of a complex problem.

6.1.1 Numerical Conclusions

From one player, one performance

On section 5.1.1 the loudness and the spectral centroid of a one-hour improvisation was plotted as a time evolution and also in the form of a similarity matrix. The loudness curve could be seen as a plot of the musical intensity that in certain cases may revel long-term structures. If there would be repeating patterns, they would be shown with more clarity in the similarity matrix. In our experiment there are not clear repetitions that could be bold with the similarity matrix. The problem of finding structures should be address with other techniques.

On section 5.1.1 the energy of the Keith Jarret's *Vienna* piano improvisation was embedded in a three-dimensional space using the State Space Reconstruction technique. This technique, that may help to revel structures and pattern in some signals that show only noise when are transform with Fourier procedures, did not show significant information and produced a chaotic shape. The structure of this shape is however unique and could be seen as a static representation of the temporal vector. The PCA over the data could have helped if we would have required to project the signal or reduce the dimensionality of the result. In our case, the SPR gave already a three-dimensional space. The used of the SSR method should be deeply explored and tried with shorter fragments of audio.

[1] As George E. P. Box said "All models are wrong, but some are useful."

From one player inside the ensemble

Contrary to the results of the previous section and mainly because in this case we are not analyzing the signal directly but using such signal as representation of the interaction between players, here we obtained clearer results. Section 5.2.1 presents a set of experiments to detect the behavior of one player during one improvisation. From the correlating matrices between musicians and the dynamic time warping we see that musicians of the ensemble tends to maintain the general values of the rest of the ensemble. Only on few and punctual moments a player tend to be “outside” the rest of the ensemble. The grid of figure 5.6 let us see not only the sections of the improvisations, but also the amount of interaction between players. On similar way, figure 5.7 let us see that there are almost null interaction between some members that we may identify as *complementary*, while there are other members with frequent interaction identify as *united*.

From overall performance of one player

While segmenting the long improvisations into pieces —knowing that we tended to change musical material several times during one improvisation— was a good idea, trying to extract what we called *archetypes* from the signal did not show good results neither with the correlation nor with the decision tree classifier. The confidence of both experiments was too low to rely on the results.

From understanding the musical content

The most interesting conclusions came from the exploration of the interactivity of all the members during all the improvisations. After discretized and normalized all the corpus of our improvisations, the data was treated with several algorithms. The initial three tree decisions tress shows that there are certain rules of interaction. It is naive to think that such rules could model an entire improvisation, but they explain some situations in conjunction with other tools.

The set of rules from the *PredictiveApriori* algorithm indicates also certain statistical behaviors. Do they mean something musical? The first fact that we may observe is the correlation of the pitch range between all the instruments. This behavior was verbally recognized by the members during the discussions and it seem that we never manage to break such overlap of pitch register. Also, it seem that keyboard would stop to play if the rest of the ensemble is playing soft as shows rule 9 of the PredictiveApriori results.

It is clear that this conclusions, derived from the basic experiments explained in the document are only an initial step in the forthcoming research. With the obtained results, it is not possible to extract any type of *grounded true* but such results could help and support other studies.

Finally, as Kopiez suggest, "up until now there has been a significant lack of methods for the study of entire performances with time-frames of more than a few minutes"[41]. This work is a contribution to the field.

6.1.2 Musical Conclusions

As a musicians and composer I will always resist the idea of numerically modeling the process of creating music[2]. In the same way, I also have serious doubts about modeling the processes that happen when people improvise together. We may explain in detail the behaviors, recreate the actions, and even predict the comportment of an improvisatory ensemble but we would still far away from explaining the sensible reasons of our actions. Nevertheless, the exploration has to be done, only a methodological study can help to enrich and increase our perceptual resolution as creators and receivers of art[3]. All kind of resources

[2]In that way I share the idea that art is a purified decantation of our personal history.

[3]As Socrates said "Know yourself".

should be employed for studying our musical behaviors and numbers is one of them.

6.2 Future Work

6.2.1 Analysis

As it is observable on chapter 5 the number of experiments and the techniques employed are limited. The experience and the work developed up to know are a good starting point for a deeper exploration, however in order to obtain a better scope of our problem it is important to develop more experiments and apply a wider number of techniques.

Being worried about the global structure, and the main form of material, we have leave out of study short-term relationships and interactions. It is hard to get a clear idea of the general structure without studying the details. It is important to develop experiments for studying the interactions between musicians in short-time windows. A *buttom-top* analysis is missing and it is encouraged.

In this context, analysis and experiments should be addressed for studying the relationship between micro and macro structures. Several musical and physical theories have been purposed in this area[61] that explains and/or justify a connection between short-term structures such as wave forms in sounds, and bigger structures such as rhythmic patterns, or musical structures. For example, the Mexican composer Julio Estrada has suggested a compositional relationship between different size structures[21].

Several procedures that are employed for the analysis of more stable music should be employed in fragments of few seconds. The employed of such techniques could be the base for the *buttom-top* analysis. Among these techniques the music structural analysis of Bee Suan[53], the characterization of music complexity of Sebastian Streich[70], the tonal description of music audio signals of Emilia Gómez[24], and the computational rhythm description of Fabian Gouyon[25] should be necessary studied and applied. A work that should be studied because of its clarity and consistency if we decide to explore to pitch class variable of our data is the work of Hendrik Purwins[57]. The analysis of long structures by Kopiez[41] should be studied with detail due to the similarities between our explorations.

A missing element so far that is essential is human classification, evaluation, and validation. Using real musicians to hand-label interesting moments and interactions and classifying and grading fragments according to its musical interest have to be done. The obtained data could be use on several experiments. In this way, formal questionnaires after each performance would have being of value.

6.2.2 Understanding

As mention in the conclusions, most of the techniques used on this research still in a naive stage if our intention is the extraction of subtle patterns of interaction between musicians. There are not only a vast numbers of basic techniques, but also complex procedures that still have to be explored. To mention, *Multilayer Perceptrons*, *Hidden Markov Models*, *Bayesian networks*, *Supported Vector Machines*, and *Correspondence Analysis*. In addition, the use of the State Space Reconstruction was only grasp.

6.2.3 A model of gestural improvisation

The collection of observations and conclusions obtained during this research in addition with those obtained from future experiments have the intention to help in building a model of gestural improvisation that could described or partially described basic mechanics and *archetypes* that may happens when human improvised. It is naive to think that it is possible to came up with *The Model* that encompass the improvisatory activity but it is feasible to create a framework of what a particular musician or group of musicians do when play in the same *style*. Two strategies should be explored

- *A static model* that reacts and interacts to the input with a set of fixed rules. The output, however, should produce a result within the style and constrains of the live musicians. There are several models of this type.

- *A dynamic model* that should pass over a set of iterative interactions with the musicians before it could produce the desire result. We may see this as a *rehearsal-concert* model. After all, most musicians get together and practice together for many hours before giving a concert. If, as we mention several times over this work, structure is an important element of music then the model should have a set of such structures before be able to produce new ones. Therefore, the necessity to feet the model with finalized improvisations.

6.2.4 Future projects and implementations

The Virtual Improviser

As mention in the motivations section of the introduction. The final goal of this research is the creation of a virtual musician using a model of gestural improvisation that could play with live musicians. The idea is grasped in the appendix A however, all the knowledge obtained from the experiments developed so far and also from future experiments have to be taken into account. The work still in the first stage of progress. The knowledge has to be model, the model implemented, and finally the model used and tuned. The system which could play with us with some "musical intelligence" is still a long-term future work not only for me but for the entire community devoted to the task.

Moz-art development

The Moz-art-global-art project will be used and performed on November. The software should be not only improve for then, but most important the results of this research may be employed for implementing human-computer interaction procedures that may increase the experience. For example, the creation of a virtual player that could react with some pseudo-logic rules according to the player or players input may be an interesting experiment. In counterpart, the interaction between humans may be recorder in order to see the behavior and mechanisms that governs the dialog. Future analysis and study of such interaction could lead to new conclusions.

Appendix A

The acoustic piano as a conductor of collective musical improvisations

The acoustic piano as a conductor of collective musical improvisations *

An experimental union of the *IMPI*: Improvisatory Music and Painting Interface and the *LIAS*: Learning Improvisatory Audio System

Hugo Solís García
Music Technology Group
Pompeu Fabra University
Ocata 1, 08003
Barcelona, Spain
+34 93 542 21 04
hsolis@iua.upf.es

ABSTRACT

This paper describes the design and implementation of a collective improvisatory system controlled by an acoustic piano. The system works by analyzing on real time the audio output of an acoustic piano using standard audio descriptors ported to PD. This data is mapped to musical parameters that are used to generate diferent visual representations on computer screens using Java. The rest of the ensemble uses this visual information in order to extend the improvisation. Three representations are available: standar music notation —using JMSL—, sketch notation, and dynamic drawings. The core of the system has been derived from two projects developed by the author: the Improvisatory Music and Painting Interface *IMPI* and the Learning Improvisatory Audio System *LIAS*.

Keywords

Collective improvisation, extended piano, visual representation of music, JMSL

1. INTRODUCTION

Shaping collective free improvisations in order to obtain solid and succinct works with surprising and synchronized events is not an easy task. In Western music, the notion of musical form has been an important concern throughout history. Each historic period and musical style has created and expanded musical forms, and free improvisation has its own. The musical form in improvisatory materials is, in most cases, more diffuse and less structured than the form of compositional works. In addition, it takes longer to develop transitions and variations of the material. Usually, elements and ideas are much more spread out in time. The main reason for this is the difficulty of comprehending the material without a score, and the difficulty of synchronizing events in the case of collective improvisations.

Even if the free collective improvisatory style should not necessarily use the syntax of composed music, there are aspects of the latter that could acoustically enrich the expression in the improvisatory domain. Common elements in notated music such as synchronized events and sudden changes by the entire ensemble are hard to obtain with big ensembles. Several techniques have been created in order to import these elements into the improvisatory domain. The simplest one is to create hierarchic ensembles where one member guides the rest. Others, involves the creation of body language symbols, improvisatory scores, or implementations of computer systems. This experiment is a proposal of this last technique.

This document is organized in the following way. We begin by giving a contextual framework about the extensibility of acoustic instruments, the computer as improviser, and the possibilities of extending digitally the acoustic piano. We then discuss separately the two core technologies that were merged for this exploration. In the central part of the document, we describe how by using both systems we can shape musical improvisations by analyzing the audio from an acoustic piano and mapping the description of such data into digital scores. These scores are presented on computer screens and updated on real time. This representation is used by the rest of the ensemble. The detailed technologies, concepts, and specifications of the exploration are presented pointing out the contributions, problems, and future work.

1.1 The Constant Dilemma: Transforming, Extending, Hacking, or Creating new musical interfaces

The creation of musical interfaces has been a historical activity directly related with the development of music. The production of such musical interfaces has been done with different approaches. Just as a schematic classification, we

*This project has the support from the PAEE-FONCA

can propose a basic organization of such approaches [8]

- Transforming procedures. In this case, a gradual transformation of an existing device is observed. These transformations implie small alterations of some part of the instrument.

- Extending procedures. In this case, the transformation implied the addition of new parts or technologies. In recent years, these extensions have included digital devices.

- Hacking procedures. Alternatively, some inventors have recycled or have modified the nature of the instrument, creating therefore an entire new instrument.

- Creation procedures. Finally, a less common but more radical approach has been the creation from scratch of the entire interface. On many cases, the creation of a new instrument suggest also new performance techniques.

1.2 The improviser computer, an ever-new old challenge

Improvising music is an essential act of the music phenomena. This technique has been present in multiple situations with several connotations and techniques in all musical periods and all cultures even thought "by its very nature — in that improvisation is essentially evanescent — it is one of the subjects least amenable to historical research."[14]

With the introduction of electronic technologies that allow the generation in real time of materials impossible to be created by one person the paradigms of real-time and performance were modified. Therefore, the field of musical improvisation suffered severe transformations and the opportunities for exploration were multiplied. Different approaches were taken for many developers of improvisatory music systems. The improvisation with electronic devices derived into a deeper and more complex concept: the idea of the computer improviser.

Several works in the area should be described. However, only some of them will be enumerated as a general overview of the field. The work of George Lewis who developed his software *Voyager* and uses it with his trombone [10], The Robert Rowe's system *Cypher* who pushed to field into high levels of music analysis [19]; the *Continuator* of Francois Pachet which has interesting results in spite of its simplicity [16]; and the works of Danenberg, Al Biles, Sergi Jordà, among many others.

1.3 The acoustic piano is dead! Long live to the acoustic piano!

For some people the acoustic piano is a frozen instrument loosing territory besides electronic pianos and synthesizers. For others, even if the core physical modifications of the instrument stopped around 1860 [17], the piano is still open to improvement either by changes on the instrument itself or by extending and hacking the interface. Being one of the most technologically developed acoustic instruments, the piano is also one of the most well-known musical instrument. Its capabilities of integration into almost all the music styles have done that a huge amount of music is still composed and performed on it.

As any other instrument, it has had a long evolution over several hundred of years: an evolution as a musical device on its own —a technological evolution—; an evolution inside a musical context —musicological evolution—; and because of the last one, an evolution in its performance techniques. As a live instrument, the piano, its music, and its performance are exposed and influenced by the present time. In the software domain an outstanding example that must be mentioned is the piece *Jeux Deux*. In this collaboration between Tod Machover and Mark Downie "the music, for pianist Michael Chertock and Yamaha Disklavier, aligns, manipulates and propels visuals prepared from analyses of the performer's hand movements, creating a real-time illustration of the musical score and its interaction with the physicality of performance." [4] [3]

As observed, the acoustic piano is source of new aesthetic contributions and digital media has much to offer to the instrument. On this project, the author searches for the integration of both domains by combining his experience on both areas and applying them in the area of improvised music as will be explained later.

2. THE ORIGINAL TECHNOLOGIES

2.1 Improvisatory Music and Painting Interface

The Improvisatory Music and Painting Interface (IMPI) [21] system is a computer program for the creation of audiovisual improvisations performed in real time by ensembles of acoustic musicians. It was developed by the author between 2002 and 2004 at the MIT Media Lab. The coordination of these improvisations is obtained using a graphical language. This language is employed by one "conductor" in order to generate musical scores and abstract visual animations in real time. Doodling on a digital tablet following the syntax of the language allows both the creation of musical material with different levels of improvisatory participation from the ensemble and also the manipulation of the projected graphics in coordination with the music. The generated musical information is displayed in several formats on multiple computer screens that members of the ensemble play from. The digital graphics are also projected on a screen to be seen by the audience. One of the main goals of the system is the translation of planned compositional elements (such as precise structure and synchronization between instruments) into the improvisatory domain. The graphics that IMPI generates are organic, fluid, vivid, dynamic, and unified with the music.

2.1.1 IMPI on live application

At the end of March 2004, the Media Laboratory hosted the visit of the composer and improviser John Zorn. During the visit a colloquium and workshop with the Performance with Experimental Musical Instruments team were realized. The experience was guided by Zorn who contributed constant comments and suggestions to help to improve the improvisations. The second version of IMPI was tested with the entire ensemble of six musicians and myself as the conductor. Each of us had a computer so that the musicians could read the scores and the author could generate them by drawing with the mouse and the digital tablet. Several improvisations were performed using IMPI. It worked more as a shaper of form than as a generator of musical content. It was interesting to see that during the performance, the improvisatory language was preserved at the same time that

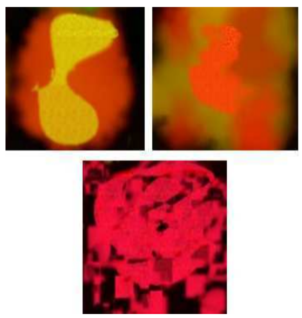


Figure 1: Three snapshots from the graphic output of IMPI.

some moments of silence and solo sections were easily created.

At the end of the improvisations, a discussion took place where John Zorn and all the musicians gave their impressions. Zorn said that given the audiovisual systems for live music applications he was familiar with, IMPI posed one of the most interesting musical paradigms due to the multi-representation of the musical content. The musicians of the ensemble described in detail their impressions. The general agreement was that the graphics reduced their attention to the performance of other musicians, which is something fairly common when musicians work with scores that they do not know in detail. The discussion helped to point out that some of the mappings were in fact, interesting and potentially expressive, and some parts of the design of the musicians's and the conductor's interfaces should be modified. It also showed that the system requires careful study by the conductor and musicians in order to generate coherent results, such as consistent interactions between musicians and stronger relations between audio and visuals.

2.1.2 *JSML within IMPI*

The Java Music Specification Language (JMSL) [2] was created by Nick Didkovsky and Phil Burk as an evolutionary successor of the Hierarchical Music Specification Language (HMSL) created by Phil Burk, Larry Polansky, and David Rosenboom. JMSL is a "Java-based development tool for experiments in algorithmic composition, live performance, and intelligent instrument design" [1]. JMSL is an excellent package for the creation of algorithmic music. Its concepts of hierarchical organizations, musical tasks, and parallel and sequential collections, are entirely natural to the way music is organized no matter what kind of structures and elements create the musical language. In addition, JMSL is well integrated with JSyn. It "uses native methods written in C to provide real-time audio synthesis for Java programmers. JSyn can be used to generate sound effects, audio environments, or music. JSyn is based on the traditional model of unit generators which can be connected together to form complex sounds" [1]. This means that if, in the future, IMPI grows to integrate the production and manipulation of electronic and/or electro-acoustic music, the extension will be easy to integrate.

The current version of JMSL offers the ability to render scores in real time. The first version of IMPI uses its own score rendering, as will be seen in the section Versions and Evolution; during the process of improving the system, the option of using the JMSL's score package seemed much more logical. The score package of JMSL is a Java package totally integrated with the rest of JMSL, and allows the creation and transformation of scores written with Standard Music Notation. The method of creating, modifying, or transforming the score is entirely open to the choice of the programmer and can be done entirely in real time. In addition, the JMSL's scores can be saved as musicXML scores which means that the final file can be opened into music editors such as Finale, Turandot, and Igor Engraver, among others.

2.2 LIAS

LIAS:Learning Improvisatory Audio System is a training audio-to-audio mechanism with the final goal of extending musical improvisations of acoustic instruments. The system is currently under development and implies a long-term research. As a final goal, the system should contribute not only to the extension of the timbre but most important to generate musically coherent relationships between the musician or musicians and the machine by applying signal processing and standard machine learning techniques. This extension will generate new digital content on real-time in response to the musical material performed by the player or the players. The generated material should be meaningful and coherent with respect to the input. In order to obtain these results, a musical model has to be implemented and electroacoustic music already written has to be analyzed. The work is oriented but not limited to generate audio content. Other media such as dynamic images and control of hardware devices may be explored in the future. The entire work should create a new aesthetic proposal and the research should be a theoretical contribution to the field. The work is focused on improvisatory expressions, therefore experiments and research should be conducted in the field. As validation and proof of concept, several practical experiments will be developed. The work described on this document is the first one of the series.

2.2.1 *Applying state-of-the-art knowledge of audio description and machine learning techniques for artistic purposes*

Many people works on audio description and machine learning techniques applied to music. However, most of this research is oriented to solve commercial and industrial problems such as creating audio recommendation systems, audio identifiers for radio stations, music querying by humming, among many others.[1] Therefore, most of the work is done in pop-music that has the majority of consumers. Few people applied the knowledge to study other type of sound expressions and, when that is the case, the work has a traditional musicological approach. In consequence, there are few artistic explorations in the use of such cutting-edge technologies. There are some exceptions, and several musicians are implementing new techniques to their artistic production, however these contributions are mainly in the field of electronic music.

[1]Most of the research at the Music Technology Group is on such direction.

One of the main contributions of this work is therefore implementing the current knowledge in the fields of audio description and machine learning into the domain of improvisatory music. These tools are the technical methods that will be employed in order to create an artistic proposal that try to push the current state of improvisatory music systems by using recent knowledge in the technical domain.

2.2.2 *Analyzing the acoustic piano with DSP techniques*

With the signal processing techniques available nowadays, it is possible to obtain on real-time a huge amount of information of an acoustic signal represented as digital data. Combining some of these descriptors it is possible to obtain more musically meaningful information such as on set detection, beat and tempo, pitch and harmony, etc. [6] Even complex problems such as obtaining polyphonic pitch detection on real time can be extracted in certain cases [20].

Many of the available systems for extending acoustic instruments use special hardware devices for extracting the information of the piano. This project is, on the other hand, a system that search for the extension using only the available signal processing procedures which seams to be a more natural procedure. Several detailed information that can be extracted by mechanical devices such as MIDI keyboards will be lost if we uses only audio detection. However, the hypothesis is that the lack of short-term resolution may be substituted by the use of adaptable mappings [13], and also by a better understanding of the high-level musical structure.

2.2.3 *Modeling structure*

Several works have been developed in the recent years in order to extract patterns from digital data. The results are used then to recognize low-level and high level musical structures. Most of this work has been developed in the context of popular music where usually there is a tonal harmony, a steady rhythm, and recognizable sections in the form of verses and choruses.[15] In order to obtain certain musical understanding of the patterns, several models have been proposed. Most of them generate hierarchical trees in one way or another. The most adopted model is probably the model of F. Lerdahl and R. Jackendoff [9] which has the disadvantage of working only with a particular style of music where the sections and phrases are well delimited. Another alternative that could be studied is the mathematical model of Guerino Mazzola which seems to be a suitable option for modeling patterns more diluted. [12] Thus, another branch of the research is to study and implement the model or models that best fulfill our requirement of having a tool that permits to keep tracking of the material performed on the instrument and obtain information about the overall shape, the leading, and the direction of the musical ideas. This final model should be able to work dynamically and update constantly on real time.

2.2.4 *Implementing a learning algorithm*

Rehearsal is essential for a good concert! Either is a soloist improvisation or a group performance, the process of learning and tuning details before the actual concert is fundamental in the live music paradigm. As mentioned before, many music systems do interact with live musicians. However, few of them are conceived as systems where actual learning algorithms are core elements of the implementations. The author proposes, on the other hand, a system that requires rehearsal. Thus, the musician or musicians should interact several times with the system before the final concert.

Based on the model described in the last section and using standard machine learning techniques [22] the final system will analyze the activity of the player or players. It will also try to obtain as much musical knowledge as possible: type of musical materials employed; type of transformations and developments of such materials; type of global shape of the performance; type of interactions between musicians, etc. Once the system has obtained some information about the music, the system would be ready to become part of the ensemble and live musicians would be able to interact and improvise with it. Ideally, the system should generate motives and musical phrases that could be musically meaningful for the performance. The type of sounds that the system should generate must be in accordance to the material, and tentatively they should be generated based on the acoustic material of the acoustic instruments.

3. THE ACOUSTIC PIANO AS A CONTROLLER FOR COLLECTIVE MUSICAL IMPROVISATIONS

On this experiment we use our improvisatory model of *IMPI* as an initial proof-of-concept of our musical goals of *LIAS*. By employing some additional technologies and implementing a basic generative algorithm [18] we developed an experiment where the acoustic piano lead a collective musical improvisation. The kind of musical style this model is intended for is non-tonal, non-rhythmic, and texture-oriented, which means that strong emphasis is put on the control of timbral qualities and continuum transitions.

The intention of the experiment is to shape the improvisation by using a leading instrument. Collective improvisation is an experience with particular characteristics. The activity of improvising with other musicians requires the same attention as the performance of written music. Attention must be given not only to the material that each member of the ensemble is producing but also to the final result. According to the way the members of an ensemble interact, several kinds of ensembles can be designed. Each kind of organization offers advantages and disadvantages over the other configurations, and each one has characteristics that help to emphasize different elements. Some configurations offer more freedom and possibilities for interaction between the members, while others are better at creating well-organized materials. [21] The basic models are:

- Centralized ensembles with one leader and unidirectional transmission of information between the leader and the rest of the ensemble.
- Centralized ensembles with one leader and bidirectional transmission of information between the leader and the rest of the ensemble.
- Hierarchic ensembles with group leaders and unidirectional transmission of information between the leaders and the rest of the sub-ensembles.
- Hierarchic ensembles with group leaders and bidirectional transmission of information between the leaders

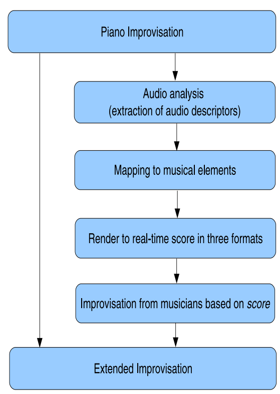


Figure 2: Chart of the information flow in the system.

and the rest of the sub-ensembles.

- Centralized ensembles with one leader and communication between the leader and the rest of the ensemble, as well as communication between the members.

- Non-centralized ensembles with no communication among members.

- Non-centralized ensembles with communication among the members.

3.1 Audio analysis and generative algorithm

The audio model of our experiment is based on the work developed by Tristan Jehan, Tod Machover, and Mike Favio on the orchestral interactive piece *Sparkler* of Machover [11]. On that score, the `analyzer~` MaxMSP object developed by Jehan [7] is used for study the audio content of an entire orchestra on real time. On the other hand, our model uses this object to extract meaningful information from the acoustic piano [2]. The perceptual features that the object can extract are: pitch, loudness, brightness, noisiness, onsets, and Bark scale decomposition. Only some of this features are employed and some are used as "intuitive" guides as is the case of the pitch detector that is intended for monophonic signals. As in the case of *Sparkler* other features are derived from the raw data and the "Activity" is estimated from loudness dynamics over a large period of time.

In a similar way than *Sparkler* "the generative algorithm [18] does not rely on a particular beat or sequence of notes but is very flexible" [7]. In this way, the material that is generated from the audio of the piano and render in several ways (notated score, sketching guide, and drawing gesture) on the computer screens has *textural* and *stochastic* characteristics.

[2] An implementation of the object for PD is under development by the author

3.2 Score representation

Several mappings have been tried and because each one offers some benefits and some disadvantages the solution proposed is to let the musicians to choose from a pallet of possibilities. In this way, the conducted musicians could have a range of freedom. One of the solutions was to create different kinds of musical representations. Each representation would give to the musician a different level of independence, ranging from the determinism of the Standard Music Notation, to letting the musicians create their own interpretations of graphical shapes. In addition, the possibility of creating scores for musicians that do not know how to read music seemed to be a nice option for inviting all kind of musicians to interact with the system.

- The Score representation. The top window of the interface shows the material that is sent by the painter in the Standard Music Notation style. This window would correspond to the window of the first version of IMPI; however, it was not fully implemented when the I decided to work on the next version. The musicians should know how to read scores, and he or she can not alter the information received.

- The Sketch representation. Below the score window there is another representation of the musical material. It was designed to preserve the musical intention and structure of the improvisation, but musicians are not required to know how to read music and they have a bigger frame for developing personal ideas. This representation of the music is, in many ways, comparable to the piano rolls; however, it also preserves similarities to the scroll window. Musical notes are represented with rectangles. The length of the rectangle represents the length in time, and its darkness represents the intensity of the note. Almost white rectangles represent soft notes and black rectangles represent louder notes. Notes are constantly scrolling from right to left, as occurs in the score window. The position of the rectangle in the Y-axis represents the pitch of the note. The lower part of the window indicates the lowest register of the instrument, and the high section of the window indicates the highest register of the instrument. Information runs from right to left.

- The Drawing representation. Below the sketch window, between other two square windows, there is a white square area where the musicians receive only the graphical information that was intended to be directed to them. Musicians do not receive the final animation, but only the objects that were sent to the sub-ensembles they belong to. Musicians should use this graphical information in order to create a subjective acoustic representation of the visuals [3]. Of all the representations, this is the one that offers almost complete independence to the musicians. Even with all the liberty that this option offers, it is expected that at least a slight intention of the structure defined by the conductor is still preserved.

[3] An interesting reference on audiovisual music is [5]

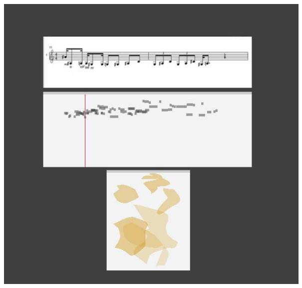


Figure 3: Screen for the musicians where the three representations of the material are presented.

Because musicians have the option of altering or modifying the information they receive or even generate a personal line of music, a natural interaction occurs. Since in this version musicians are not necessarily constrained to what is presented in the notated score, they can interact with the same conditions that traditional free collective improvisation offers. Musicians can choose among the different musical representations according to their wishes and abilities, and they can voluntary switch among representations.

4. CONCLUSIONS AND FUTURE WORK

As can be observed, *LIAS* is a long-term research with strong emphasis in theoretical aspects. However, it requires constant experimentation and several proof-of-concept implementations. The model described on this document is the first experiment with the goal of setting the basic concepts. This experiment has necesary to be extended and tested on real musical contexts. However, there are, in this moment, several important elements that can be pointed. The poly-representation of music as a medium for shaping collective improvisations is the main contribution.

There are still, several features and explorations that must be done with the system. First of all, the mapping between the audio and the musical representation have to be extended. Also, other generative algorithms must be test and implemented. Finally and most important, implementing "learning" and "intelligent" procedures have to be developed.

5. ACKNOWLEDGMENTS

The author wants to thanks the Fondo Nacional para la Cultura y las Artes (FONCA) throw the Programa de Apoyo para Estudios en el Extrenjero 2005 for the grant which make this work possible. He also wants to thanks the people from the Interactive Systems Group at MTG which are always willing to help, specially to Sergi Jordà. Thanks to Tod Machover and to the Opera of the Future group at the MIT Media Lab where IMPI was developed. Thanks to Nick Didkovsky for JMSL. Thanks to Tristan Jehan for his constant help. Thanks to Alfonso, Amaury, Maarten, and Sylvain for the comments.

6. REFERENCES

[1] N. Didkovsky. http://www.algomusic.com.

[2] N. Didkovsky. Java music specification language, v103 update. In *ICMC '04: Proceedings of the Internationa Computer Music Conference*, 2004.

[3] M. Downie. http://www.openendedgroup.com.

[4] M. Downie. *Choreographing the Digital Agent: Live Performance Graphics for Dance Theater*. PhD thesis, Massachusetts Institute of Technology, 2005.

[5] B. Evans. Foundations of a visual music. *Computer Music Journal*, 29(4), 2005.

[6] T. Jehan. *Creating Music by Listening*. PhD thesis, Massachusetts Institute of Technology, 2005.

[7] T. Jehan and B. Schoner. An audio-driven perceptually meaningful timbre synthesizer. In *Proceedings International Computer Music Conference*, pages 381–388, La Habana, Cuba, 2001.

[8] S. Jorda. *Digital Lutherie: Crafting musical computers for new musics' performance and improvisation*. PhD thesis, Universitat Pompeu Fabra, 2005.

[9] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1993.

[10] G. Lewis. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal*, 10:33–39, 2000.

[11] T. Machover. *Sparkler - musical score*. Boosey and Hawkes, New York, 2001.

[12] G. Mazzola. *The Topos of Music; Geometric Logic of Concepts, Theory, and Performance*. Birkhauser, 2002.

[13] D. Merril. Flexigesture: A sensor-rich real-time adaptive gesture and affordance learn. Master's thesis, Massachusetts Institute of Technology, 2004.

[14] B. Nettel. Improvisation. http://www.grovemusic.com.

[15] B. S. Ong. *Towards Automatic Music Structural Analysis: Identifying Characteristic Within-Song Excerpts in Popular Music*. PhD thesis, Universitat Pompeu Fabra, 2005.

[16] F. Pachet. The continuator: Musical interaction with style. In *Proceedings International Computer Music Conference*, Göteborg, Sweden, 2002.

[17] E. M. Ripin and S. Pollens. Pianoforte. http://www.grovemusic.com.

[18] R. Rowe. *Interactive Music Systems*. MIT Press, 1992.

[19] R. Rowe. *Machine Musicianship*. MIT Press, 2001.

[20] M. Ryynanen and A. Klapuri. Polyphonic music transcription using note event modeling. In *Proceedings of Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2005.

[21] H. Solís. Improvisatory music and painting interface. Master's thesis, Massachusetts Institute of Technology, 2004.

[22] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.

Appendix B

The MOZ-ART-GLOBAL-ART project

The *MOZ-ART-GLOBAL-ART project*[1] is a piece of software for collective audiovisual improvisations over the net produced by the composer Mauricio Valdés and the author. It was commissioned by the *Instrumenta 2006* festival which should have taken place in Oaxaca Mexico on June 2006[2]. The conditions of the commission were first, to create a tools that allows musicians far away from the place of the festival to contribute and become active participants during the

[1] http://moz-art-global-art.com/

[2] The festival was moved to November for social and political problems in Oaxaca.

performance and second that the result should have a minimum level of musical quality to be presented during the festival among other concerts of well-known musicians.

B.1 Origin of the system

In order to guaranty a minimum control of the result two different approaches were initially considered: making a program where the piece would be gradually and collaborative produced and the result presented during the concert, and making a program that generates the result on real-time but having several constraints that helps to *model* the output. Another issue that had to be discussed was if the invitation to collaborate in the improvisation should be public or only specific composer contacted. Mauricio and I decided to make a real-time experience because having the net for making an off-line experience may seems contradictory. Later on, the festival's artistic director decided to invite composers only with the idea that trained musicians would be concerned about the aesthetic result besides the novelty of the technical experience.

Knowing that only experimented composers would interact between them with a real-time application lead to some possibilities that had to be considered. Should the material of the concert had to be build on several interactions? or, would it be a one-time event on real time? After several evaluations, and having in mind that it would be hard to organize several rehearsals with several composers all around the world, we decided that the best option would be a piece of software that allows the players to practice off-line alone, rehearsal with others at any time, and obviously perform during the concert.

Considering the time that the players would invest in average before the performance and also in order to guaranty that composer with few experience in the use of digital media could feel comfortable we decide to create a simple

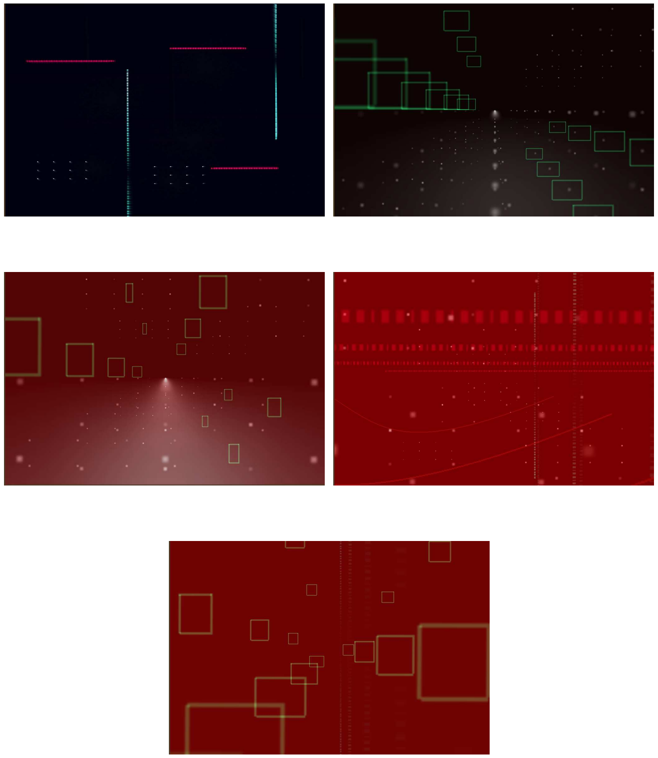


Figure B.1: Moz-art-global-art main window screenshots.

and straight-forward interface. Other factor that was considered was the type of experience -acoustic or audiovisual-. We opted for and audiovisual experience, not only because both of us are interested in this type of experience but also because the graphics could help to give consistency and guidance to the improvisations.

Finally, in order to maintain a simple interface that does not interfer with the audiovisual production and does not distract the audiovisual creation, we decided to minimize the interaction among players and leave the interaction in the domain of the human-decision making. Therefore, we decided that users would be in charge from his or her material only without altering other's work.

B.2 Technology

Once that the essence of the project was defined, the technical aspects had to be selected. The initial proposal was to use MaxMSP [85] or PD [56]. PD has GEM [15] and MaxMSP has Jitter[34] as graphical counterparts and they could be both used. However we wanted to minimize as much as possible all the external factors that could avoid users to run our program and it seemed that installing, running PD with GEM or MAX with Jitter, and interacting over the net was not a straight-forward procedure. Many potential users could get scare about installing and configuring. Another considered alternative was the *Processing* enviroment [59] however making the audio to run could be tricky for some users. Therefore, we decided to build our system from scratch using Java. The Java toolkit offers the *java web start* technology which appeared to use as a good method that guaranties the execution of the application on any platform with a minimum set of installation steps.

> Java Web Start is an application-deployment technology that gives you the power to launch full-featured applications with a sin-

> gle click from your Web browser. You can now download and launch applications, such as a complete spreadsheet program or an Internet chat client, without going through complicated installation procedures.
>
> With Java Web Start, you launch applications simply by clicking on a Web page link. If the application is not present on your computer, Java Web Start automatically downloads all necessary files. It then caches the files on your computer so the application is always ready to be relaunched anytime you want-either from an icon on your desktop or from the browser link. And no matter which method you use to launch the application, the most current version of the application is always presented to you.[31]

For the audio manage we employed the Jsyn [8] and JSML [18] libraries that I have used in several other applications and shows a good performance and reliability. On the other hand, Java imaging still slower than native graphics. Fortunately, several rappers for java have been developed that allows the used of native graphic capabilities. We employed *JOGL*[73] which let us code OpenGL [83] within Java.

The server was written using purely Java and was coded fundamentally to transmit the data produced by one player to the rest of members currently connected (figure B.2). It is therefore a Centralized Network Model.[6] The server has secondary mechanisms such as counting the members connected, sending synchronizations cues, and sending play and stop messages to all the members.

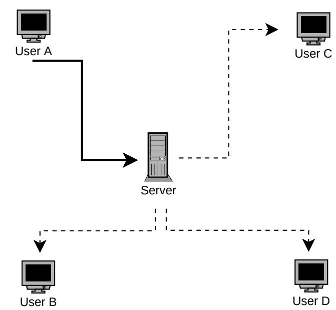


Figure B.2: Interaction configuration.

B.3 Interface

Figure B.3 shows the control panel that users employ for interacting and creating the audiovisual material. At the top left side five control tools are placed. Only four of them are visible on the image. The first tool in left to right order is the play and stop button. Users can start and stop the program if they are working off-line. If they are connected to the Internet, the symbol is frozen and the piece is automatically synchronized with the time and start stop status of the server. The second icon, that is a break horizontal line is the connection indicator. It is a completed line if the connection with the server is established. It is also the button to connect and disconnect from the server. The third icon indicates the number of current people connected. If the user is not connected no number is displayed. The four icon indicates if the user interaction is evaluated and sent to the others or not. The reason of having such indicator is because the human input is controlled in order to avoid the overdrive if many people is playing at

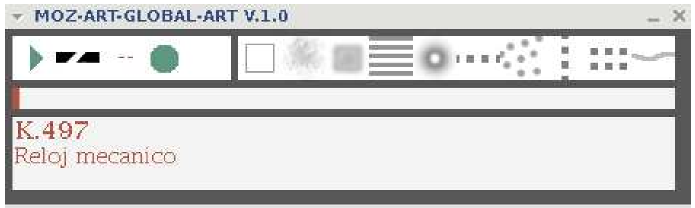


Figure B.3: Control panel.

the same time. The last area display a counter of the current interaction. The input source is the mouse, this icon reflects the amount of time that the current object is been produced.

The top right section of the panel shows the ten objects that the current version include. The icon reflects the visual type of shape that it generates. By putting the mouse over any of them, the icon is highlighted and a brief description of the acoustic characteristic is showed in the bigger bottom section. The long narrow panel is a time-line indicator where the little red rectangle moves across the line from left to right during the piece.

B.4 Interaction

The mouse is the input device. The mechanism is clear. The user should select a particular object, from the available ten, according to his or her desire. Once that the object is picked up and the icon highlighted, the user may instantiate new objects by pressing the mouse in the display window. The visual and acoustical objects's behaviors are different but all of them generates the output based on the following variables:

- *Mouse location.* The x and y coordinates influence the pitch and amplitude of the gadget. Some objects have a linear mapping where the x axis is the pitch and the y axis the intensity, other objects have a more complex mapping.

- *Duration of the impulse.* When the mouse is pressed the object is instantiated. However, it is triggered when the mouse button is released. The duration between the pressing and releasing of the button becomes a scale factor of the real duration of the object. Most of the objects have a factor of ten that showed to be a good trade-off between human activity and density of the piece.

- *Moment of the creation.* The behavior of a particular object is strongly determined by the moment at which it is created. Thus, the same click of the same duration and location have a different result depending of the moment in time of the piece.

Even though, all the objects are produced in the same way, each one has a particular behaviors and set of ranges. The acoustic and visual development is not modifiable by the user. Among all the possibilities, we created objects that move around the canvas, others grows on size, others change of color. The acoustic behavior is also different, some produces rhythmic or pseudo-rhythmic patterns, other produces clouds and clusters, others high-pitch sounds (see Table B.1. The objective was to create a big set of alternatives that composers could choose from in order to produced different types of materials.

Object description	visual behavior	acoustical behavior
Square	Green squares that gradually transform in rectangles.	Rhythmic minimalistic pattern.
Cloud	Purple cloud that grows in size.	Additive synthesis with stable pitch.
Blurry filled square	Blue square that grows in size	Pure sound with high pitch.
Collection of horizontal lines	Blue horizontal lines that randomly appears.	Low to high glissandi of short sounds.
Blurry ring	White ring that grows in size.	Noisy stable sound.
Horizontal squared dots	Red squared dots that simulated horizontal displacement.	Mechanical pulses.
Collection of random squared dots	Random dots of bright colors that generate clouds.	Granular sounds.
Vertical squared dots	Blue squared dots that simulated horizontal displacement.	Mechanical pulses.
Grid of six dots	Grid of six white dots that change in intensity	Pure sounds.
Blended horizontal line	Red horizontal line that gradually blends.	Rough sound.

Table B.1: Acoustic and visual behaviors of the objects.

B.5 Conclusion and Future Work

The moz-art-global-art project is a work on progress. Since the presentation was moved to November the project is currently on hold. It has not been used by real users, however its nature makes it an excellent place to test and implement ideas about the virtual improviser. The results and conclusions obtained for the current research are easily implemented on this project.

Appendix C

Graphical Representation of Music

An Essay about the Graphical Representation of Sound and Music

Hugo Solís

July 29, 2005

DRAFT VERSION NOT FOR PUBLISHING

Abstract

On this paper my personal thoughts about the current state on visual representation of sound is given. The paper starts with an explanation about why the author decide to write about the topic. Then a section where some ideas about the translation among medias is presented as a general overview of the multi-representation of data. After, a brief description about the history of music notation is given as an example of the visual representation of music information. Then the core of the paper is presented. Here, several ideas are explored such as visualization of acoustic data, representation of low and high level structures of sound and music, and the work of Paul Klee and Wassily Kandinsky as examples of artistic visualizations of music. Finally, a conclusion is presented.

1 Introduction

This work is part of the requirements for the *Introduction to Image Processing* course. Students were free to choose the topic of the work. The author decided to write about his own thoughts about an area of his interest that is directly related to his research and his artistic production: The Graphical Representation of Sound and Music. Many of the ideas were first explored on the author's thesis. *Improvisatory Music and Painting Interface* [6] This paper does not pretend to be a hardcore technical paper in the sense that does not present mathematical formulas nor describe a new technique, method, strategy or procedure to represent sound and music in the graphic domain. Is the author thought, that during the last academic year the acquisition of technical knowledge has highly exceed the production of creative work, and more over, the production of "aesthetic ideas" that could support such technical knowledge. Thought, the author wants to take advantage of this opportunity to write about a free topic to freely write his thoughts on the field as a way to organize his ideas and find possible paths to lead his future work. In my personal case, as a creator, only with a constant wondering of the benefits, lacks, constrains, and possibilities of new technologies will be possible to take advantage of them in order to create materials where the technology could be justify. Several articles, works, and examples are studied and commented over this paper. However none of them is analyzed in deep because the intention of the author is not to give a formal case-study but getting a general overview of his understanding about the field. Over the last two year, the author has been interested on the relationships between music and sound, from simple structures to the high level ideas. Several ideas have emerged during this period. In this paper, the author want to focus on one of them: the visual representation of sound and the transformation of ideas between different medias.

2 The representation of information among different medias

Most of the information and ideas that human beens produce can be represented with multiple mediums. In this way, an image can be described with words; a text that describes quantities and relationships among these quantities can be notated as a mathematical formula; a collection of notes can be represented in a score; and so on. Of course, the level of detail and resolution of such informations varies enormously. There are techniques to describe some ideas that are more suitable than others according to the type of information. Even feasible, nobody would make a serious atlas of the world with only text! But the core advantage, or at least the one I want to focus on on this essay, is not the suitability nor the potentiality to describe information in this or that media but the kind of information that becomes uncover with the transformations between different representations. The hypothesis to bold on this text is that the ability to represent the same material with different media is one of the most important elements to produce evolution and analysis of such information. The history of each media is marked by the evolution on its representation. The detail of such representation could be a good measure to know how much the human has been concern about the structure of the material. We can mention, as an example, that music has generate over the years its own methods to be preserved and transmitted in the form of scores. On the other hand dance has not developed an spread

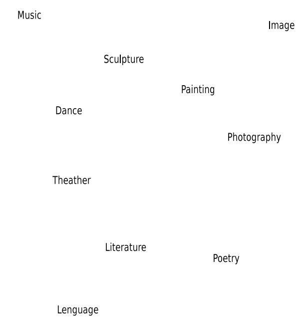


Figure 1: Heuristic map of relationships between human expressions.

well accepted methods to be notated. Can this situation be explained because the number of parameters to control in the last media exceeds the former one? It is a partial answer. The differences between the evolution in the notation of music and dance is also and fundamentally product of the aesthetic requirements of each media.

Let mention several human expressions for comparative purposes: music, dance, painting, poetry, and theater. Let analyze in detail the situation of different human expressions and the possibility to transit among them. The Figure 1 shows an intuitive map space of some human expressions. As you can see the distance between poetry and novel is shorter than the distance between music and photography. It is important to notice that this do not represent interpretations or translations, but the level of integrity that is preserved when we transit from one media to other. Music is then, one of the most portable expressions. When we talk about a painting for example, the painting is the piece of art itself and being in front of the paint is the only way to see the elements that conform the work. A precise literal description would not be part of the art work. The painting and only the painting is the art piece. On the other hand music is first the sound that your ears get during a concert, but also an score that represents the sounds. The music is the art piece but also the score. Finally, another important element that must be consider when we analyze the representation of information is that in some cases, these representations become an independent entity with their own history and evolution. In other words, the representations create their own history. In the case of music for example, the notation of music has change among the time and this change has its own characteristics that can be approach apart from the music itself. There is a music history, but there is also a history of the music-representation.

3 Music notation

Writing is one of the most important inventions in human history. It allows humans to preserve precise ideas for long periods of time without gradually altering the materials such as happens with oral transmission. In addition, writing permits going back and recovering, comparing, and modifying ideas generated at different times. In the musical field, the first attempts to notate music were around the fourth millennium. From then to the present, there has been an extraordinary evolution of the written language that creates, by itself, an entire chapter in music history.

With the evolution of musical ideas in Western music, musical creation became more complex and more idiomatic. Notation contributed to this evolution in at least two ways: First, the possibility of notating music helped to create and manipulate bigger and more complex materials. Second, the necessity of notating complex ideas generated better notations, and better notations made it possible to think of even more complex ideas. The feedback be-

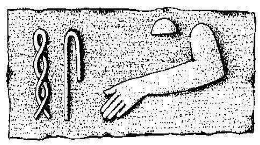


Figure 2: Symbols associated with music in ancient Egypt.

tween the two different representations of the same idea has contributed enormously to the development of Western music. Even if music notation has contributed greatly to the evolution of musical ideas, it is important to consider not only that many styles and music from different parts of the world are not written, but also that the actual notation system is far from a perfect method for writing music. A musical score is a graphical representation of the acoustic ideas of a composer. However, in order to keep the representation simple and easy, scores do not symbolize the musical events with extreme precision. There is an implicit acceptance of this “lossy” compression by the composer and the performers. In the past, the stylistic frame helps performers to reconstruct the acoustic idea, and currently recording helps performers to get the acoustic intentions of the work. Because in most cases scores offer a margin for the interpretation of their content, there is a paradigm in Western musical tradition. For some performers, the score is just a guide that must be use as a medium for transmitting the performers ideas, while for others; the score must be followed with exhaustive precision with the intention of preserving the ideas of the composer with a minimum of alteration. Another interesting situation is the displacement of music notation by new techniques for preserving acoustic material such as analog and digital recording. Since each day there is more music that cannot be notated with traditional techniques, since there is a deep interest in the exploration of new sounds, and since digital technology offers faster methods for creating, preserving, manipulating, and analyzing music, it seems that, at least conceptually, digital representation can be seen as the modern technique for notating music. When computers were introduced in the musical world, the imitation of the possibilities of analog devices was the area that received the most development. In the field of music notation, most of the efforts have been focused on the representation of music as a traditional score instead of taking advantage of the new media and creating new representations. However, several developers have taken the risk of pushing the digital media in order to create novel representations of music, and each day there are more people creating new representations that could present meaningful data in faster, cleaner, and more compressed ways. Some of these works are focused on creating representations that could change dynamically over time. Others are focused on showing information that is not presented in conventional scores.

4 Visual Representation of Sound

On this section some techniques that has been used for representing music and sound are exposed. The visual representations described on this section were chosen by the author based on the originality, efficiency, and quality of each and most of them were recently developed. Music can be analyze with different perspectives and also at different levels of resolution. It is possible to be interested in the internal structure of one piece, it is possible to try to find similarities between a set of pieces, it could be that we were interested in the sounds used on the piece. The intersection between image and sound is an area that has always been present in the mind of humans. Audiovisual stimuli are part of our daily lives and a huge amount of information is obtained from the unconscious analysis of the relationships between the visual and the acoustic elements of an audiovisual stimulus. Usually, both our visual and acoustic channels of perception run in synchrony. Any disassociation or unsolved relationship in the information of both channels creates tension and expectancy. Imagine, for example, witnessing a car crash in the street, and not receiving any corresponding acoustic stimuli; or hearing somebody whispering to your ear with nobody near you. Throughout history a huge number of theories for analyzing musical and visual expressions have been developed. On the other hand, the number of techniques for analyzing audiovisual works is close to null and there is not yet a grammar of audiovisual expression that is comparable to the theories of harmony that exist in the musical domain. However, there has been a few important attempts as will be described in the Historic References section. It is interesting to follow the development of audiovisual thoughts throughout history, such as the concern of merging music and visuals using people in ballet; the introduction of cinema, which opened a new field for the relationship between music and graphics; the use of digital technologies that has popularized and widely spread out the audiovisual expression. The audiovisual expression takes elements from the visual and the musical fields. The entire set of elements and behaviors of both domains are present in this merge. In addition, new kinds of components and relationships have been created. The new elements will be determined by several factors that will be described in the following sections.

4.1 Visualization of acoustic data

The graphication of air pressure over time is an ideal method to represent acoustic events. In some way, the dent of a vinyl recording is a 3D representation of sound. The time in the x axes tide in spiral over the z axes, and the level of air pressure stamped in the y axes. The principle of all recording methods is based on such logical representation. In the case of digital recordings time and pressure is sliced but the principle is preserved. When observing a graphics of the air pressure it is possible to obtain some information of the signal such as general intensity, moments of silence and changes in amplitude. However the graphics is meaningless for other kind of information such as pitch or timbre. Another representation that gives a lot of information is to graphic the spectrogram of the air pressure. Applying the Fourier Transformation into a signal, it is possible to get more information of the source. The FFT gives the level of energy by frequencies. This is used to get information about the pitch and the register among others. Modern techniques of signal processing have been used to measure different aspects of the sound such as brightness, noisiness, and frequency centroid. Thus, the representation of the air pressure is good for some purposes, and the representation of the FFT is good for showing other kind of information. In the recent time, a new and interesting method has been used to represent on the wave some of the information obtained with the FFT by coloring the data.

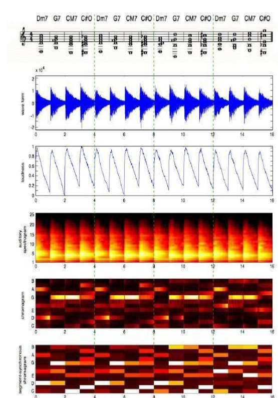


Figure 3: Typical II-V-I progressions played on a piano. No distinction can be made by looking at the waveform, the loudness curve, or the 25-critical band auditory spectrogram since this is only "one" timbre. However the chromagram allows us to differentiate chords and recognize similarities between chords of the same class even when they are inverted. The last pane shows the chromagram computed on a per-segment basis, while the previous one is computed every 6 ms with a 93-ms long window. Our representation is much simpler (only one chroma per segment), and preserves the most important information.[3]

4.2 Visualization of musical information

On the previous section some methods where the key-element is the acoustic information were described. On the other hand, the techniques described on this section are more focused on the representation as musical elements. Over the history there has been a constant interest on representing music events as can be observed in the astonish evolution of nowadays scores. However is clear that scores are an abstraction of musical events where only the frequency and the intensity are notated. The rest of the musical event is abstracted and intentionally oversimplify. Thus, contemporary composers have created new methods where other elements could be also described. Some new notations has been generalized but others belong only to a particular aesthetic or even a particular composer. Since there is not a formalized or unique way to describe sonic information the methods to visualize single music entities such as notes or isolated sounds varies a lot. Since the introduction of noisy and not-tuned elements into music artists started to think about methods to represent also sound.

4.3 Elements of Music, Basic Units and High-Level Structures

The smallest unit in the standard approach to music is the note. The elements of a note are intensity, duration, pitch, and timbre. These elements have a direct relationship to measurable acoustic components. The intensity is determined by the amplitude of the wave, the duration by the time of the sound, the pitch by the perceived fundamental frequency, and the timbre by the spectral content. Physical location and the acoustic conditions of the space where the note is produced are two elements that are external to the note, but influence the way the note is perceived. These parameters could remain unchanged during the production of the note, or they could continuously change over time, offering in both cases a huge amount of possibilities and combinations that together defined the characteristics of the acoustic stimuli. Notes are combined over time, creating relationships between them. Notes that are cre-

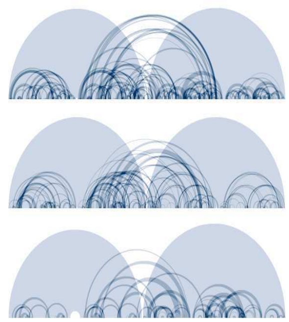


Figure 4: Bach, Three of the Goldberg Variations The images are as closely related as the music and show the AABB form.[7]

ated at the same time generate harmonic relationships and spread over time create rhythms. Harmony and rhythm are two of the most important elements in music. During history, there has been a constant exploration in both domains that has led to complex and intricate structures. During the last century, alternative theories for explaining complex musical processes have been created. These theories search for methods that could coherently explain the relationships between all the musical elements among all levels of structure.

4.4 Elements of Visual Analysis, Basic Units and High-Level Structures

As in music, visual expression is created with minimal units and the relationship between such units. The analog to the note in the visual domain is the shape. Shapes have internal characteristics such as form, color, and texture. Shapes are affected by external factors such as location and light. In his book, *Visual Perception, a psychology of the creative eye* Rudolf Arnheim [1] describes some elements that can be considered the high level structures of the visual domain. For Arnheim, shapes have weight and direction. These two elements generate the balance of the piece. The space is the three dimensional virtual space that is created by the relationships between shapes. Arnheim describes how the shapes share contours, how the figures and the ground interact, and how the levels of depth are created. Arnheim also analyzes the concept of movement in two different contexts: first, as the physical displacement of objects in time, considering speed and direction; and also as an illusion in static works, produced by the simulation of gravitatory effects and the direction of the shapes. Finally, Arnheim analyzes tension as another perceptual element that is associated with the movement and the illusion of movement without motion. Two other historic works that should be studied are the *Padagogisches Skizzenbuch* [5] by Paul Klee, because of its fundamental ideas about the characteristics of the line, and *Punkt und linie zu flache* by Wassily Kandinsky [4], because of its details about composition and visual organization.

4.4.1 The Visual Analysis of Wassily Kandinsky

According to Kandinsky, the point is the basic visual element and it has static characteristics. The point can be conceived as an abstract concept without size, but also as an element with weight, color, limits, and even shape. It is possible to make an analogy between the point and the musical note. In fact, in *Punkt und linie zu flache* there is an interesting analogy between the first notes of the fifth Beethovens symphony and a collection of points of different sizes where the size corresponds to the duration of the note and the location to the pitch. Some other audiovisual relationships are also established; however, the level of precision is not particularly interesting because many of the concepts that are used are only empirical explanations of basic phenomena. The painter establishes that the line is a secondary visual element extended from the point. This extension is the product of applying specific forces to the point. The characteristics of these forces determine the type of line. Thus, tension and direction are two characteristics of the line. For example, if the line is horizontal, it appears simpler, rigid, static, with a blacker tendency. When vertical, it reaches the clearest form of the infinite and warm possibility of movement, and has a whiter tendency. Finally, the diagonal ones keep characteristics of the two others, but are colder than the horizontal lines, and have a red, gray or green tendency. Other characteristics of basic lines are analyzed, for example the expressive quality of the angle formed by the intersection of two lines. In addition, the painter establishes that shapes have a natural relationship with color. The square would tend to feel red, the triangle yellow, and the circle blue. The stroke of the line, and how its thickness changes, strongly defines the expressive qualities of the trace. The repetition of lines and their combinations, as well as the variations of their characteristics create rhythm and tensions. According to Kandinsky, the top of a canvas is more dynamic than the bottom, in the same way that the left is more active than the right. *Punkt und linie zu flache* was written in 1926. It is interesting to observe that many of the ideas that are described in the text are still valid, and intended to be systematic. Some of the concepts have a universal agreement because they are based upon fundamental principles of perception. On the other hand, many concepts were diffuse and lacked of a methodological approach. They did not reach the level of consistency that is required to create a solid theory of visual analysis. There are also concepts that would be much clearer and more precise if they were treated with a

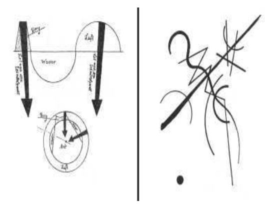


Figure 5: Illustrations from Klee and Kandinsky's visual analysis books.

mathematical or physical approach.

4.4.2 The Visual Analysis of Paul Klee

In 1925, Paul Klee wrote *Padagogisches Skizzenbuch*. According to Klee, lines can be divided into three kinds: active lines, intermediate lines, and passive lines. He also explores structures based on repetitions, their variations, and the positions of certain objects in the plane, as well as structural materials present in nature, such as muscular structures, bone structures, and join structures. The aspects of dimension and bidimensionality are also studied. Like Kandinsky, Klee analyses the case of the horizontal line and the vertical line. However, he emphasizes the aspects of balance and equilibrium, which lead to a scheme about development and direction within a painting. He uses examples from the observation of his surroundings to describe physical phenomena such as acceleration, circular movement, spin rotation, pendulum movement, spiral movement, and free trajectory. Finally, as in the case of Kandinsky, Klee gives an interesting evaluation of color, and describes many physical phenomena with the approach of a practical painter.

4.5 Mapping between Music and Graphics

How one creates expressive and coherent mappings between the visual and the acoustic components in audiovisual systems is the biggest unknown in the field. "There is no objective mapping from sounds to image or vice versa" [2]. Depending on how direct the relationship between both fields is, and how the basic structures of the languages are used in the mapping, the audiovisual relationship has a degree or level. If the mapping uses basic elements and/or easily recognizable perceptual structures, it is said that the relationship is of low level. Mapping, for example, the creation of a small circle in the canvas to the production of each musical note, and the location of the circle to the pitch of the note, indicates a low-level mapping. Low-level mappings are easy to implement and easy to follow as an audience. However, their expressive potential is limited. As soon as the strategy is detected, which usually occurs almost immediately, the interest in the relationship is lost and either the music or the image becomes an ornament. On the other hand, low-level mappings are excellent methods for highlighting a relation between a musical and a graphical element. High-level mappings involve relationships between the music and the graphic material that are complex and not easily recognizable. Instead of linking low-level units such as notes or color, these mappings merge higher structures such as relationships between colors, musical harmony, the contour of the musical phrase, and speed, velocity and trajectory in the movements. The list of possibilities and elements that can be taken into account in the creation of high-level mappings is infinite. The decision of which elements should be considered, and how should they be interrelated, is part of the design and implementation of the system. Well-planned mappings are harder to conceive and implement, but they can offer richness and endlessness discourse. On the other hand, it is very easy to create abstract relationships that lose all perceptual meaning. An entire gamut of mappings, from the low-level mapping to the high-level, is available for exploration during the creation of audiovisual works.

5 conclusions

On this paper several ideas were explore with the goal to find possible paths of explorations. As can be seen the audiovisual field has been seriously explored over the years and several approaches have been taken for such explorations. However, the field still open to explorations in other areas. The most important idea to recover from this essay is that the translation of information among medias is fundamental to have a better understanding of the material.

References

[1] Rudolf Arnheim. *Art and visual perception, a psychology of the creative eye.* University of California Press, 1954.

[2] Enrique Franco, Niall J.L. Griffith, and Mikael Fernstrom. Issues for designing a flexible expressive audiovisual system for realtime performance and composition. In *Preceedings of the 2004 Conference New Interfaces for Musical Expression*, 2004.

[3] Tristan Jehan. *Creating Music by Listening.* PhD thesis, Massachusetts Institute of Technology, 2005.

[4] Wassily Kandinsky. *Punto y línea sobre el plano.* Ediciones Coyoacán, 1994.

[5] Paul Klee. *Bases para la estructuración del arte.* Premia Editora, 1978.

[6] Hugo Solís. Improvisatory music and painting interface. Master's thesis, Massachusetts Institute of Technology, 2004.

[7] Martin Wattenberg. The shape of song. ¡http://www.turbulence.orgWorks/song/¿, June 04.

Appendix D

Verbalization of a piano improvisation (1999)

The following transcription was extracted from a file recorded on 1999. The experiment consisted on verbalized the musical decisions that the player took while the improvisation was on progress[1].

```
Atonal, evitar giros melódicos.
Elemento uno.
Repetir....y ampliar.
Repetir.... en otro registro.
Pausa.
Nuevo elemento agudo.
```

[1]Unfortunately the audio file is lost.

Textura distinta... aumentando... repetir.
Aumentar velocidades.
Repetir grave.
Introducir textura.
Ritmo.
Textura contrapuntística.
Se está cayendo.
Nuevo elemento.
Se está perdiendo hilo conductor.
Regresar al elemento primero...para retomar.
Variarlo.
Choque.
Nuevo elemento otra tonalidad.
Construir acordes.
Ir pensando en un nuevo momento...sin sonoridad, sin pedal.
Cuidado...grave nuevamente.
Repetir, ampliar, ir a registros no trabajados.
Desarrollar primera célula.
Ir abajo.
Poco a poco retomar elemento fuerte.
Nuevo elemento en el registro grave.
Pregunta respuesta. Agudo
Momento con choque cromático.
Otro
Otra vez, desarrollar.
Poco a poco crecendo...
Cuidado con los acordes tonales.
Desarrolla...
Cuidado con elemento no apareciente.

Manos contraste, una sube otra baja.
Momento cromático.
Desarrollo de este elemento.
Momento de reposo.
Repetición de notas una octava.
Momento súbito, preparamos momento súbito. Crecendo.
            Para retomar motivo en otro grado.
Ahora, no hemos trabajado pianos.
Disminuir la textura con un piano y cambiando la sonoridad
            que mantenga el elemento de unión sobre el trabajar
            algo un poco más melódico. Que conserve características.
Y poco a poco hacer pequeñas variaciones en el ostinato.
Introducir notas nuevas en stacato.
No una pausa con una inflexión de notas como hace rato.
Un acorde en stacato.
Vamos a subirnos medio tono todo.
Continuamos con la línea melódica.
Repetimos el cambio de tonalidad.
Cuidado, se está formando algo un poco tonal.
Corregimos retomando la primer tonalidad...ahora el cambio.
Vamos a ir preparando la unión de los dos elementos de distintas tonalidades.
Grados.
Ahora en el primero y en la mano derecha se esta proponiendo...
Ahora en una tonalidad, ahora la segunda y la primera.
Octavas en todas las manos.
Preparamos crecendo, eso.
Estamos en otra tonalidad.
Desarrollamos, la segunda.
Bajamos, bajamos, bajamos, hasta, otra vez.

Quitamos pedal súbito, no funcionó.
Nos quedamos en una sonoridad.
Sin pedal, por octavas.
Estamos utilizando siempre los mismos motivos, ya nos estamos aburriendo.
Esto se está parando, que no se caiga otra vez, ya no quiero retomar el mismo motivo, vamos a intentar poco a poco tomar la segunda tonalidad que hicimos he ir agregando notas.
Silencios... muy bien.
Aquella celulita nos va a servir.
El trabajo rítmico es pobre muy pobre.
Cuidado con esas quintas.
Muy bien.
No tiene nada que ver... con lo que estabamos haciendo, pero funciona... como un contraste general.
Cuidado con ese acorde menor.
Y súbitamente vamos, las notas...primeras.
Cuidado, esto se esta volviendo tonal.
Una variación rítmica.
Variación del ostinato del pedal, hace mucho que no aparecía una grave... nos sirve como pausa, no la retomes, súbete... cromáticamente, irle metiendo pedal y ve preparando el mismo tema...
Las notas que eran similares.
Define un motivo.
Incrementa el tempo.
Transformalo en un trino.
Y ve preparando en final retomando en el grabe el mismo motivo transportado una cuarta aumentada arriba y no pierdas el trino arriba.
Pedales largos.
Se disuelve el trino de la mano arriba y nos queda...

Ralentando cada vez más.
Preparamos la última nota... con la que empezamos.
Que no... Utilizamos agudos? Si
Dejamos esta como la nota final.

Appendix E

Review of Music Cognition Models (Abstract)

Review of Music Cognition Models

Hendrik Purwins[1], Amaury Hazan[1], Ricard Marxer[1]
Perfecto Herrera[1], Maarten Grachten[2], and Hugo Solis[1]

[1]Music Technology Group
Universitat Pompeu Fabra

[2] Institut d'Investigacio en Intel.ligencia Artificial
Spanish Scientific Research Council

June 13, 2006

Abstract

We present a review of facts and models about music cognition, with a special emphasis on their computational implementations. We adopt a multiple-view position, including facts from the different disciplines involved: neuroscience, psychology, cognition, artificial intelligence and musicology. The article summarizes the methodologies that these disciplines use to approach the phenomena of music understanding, the localization of musical processes in the brain, and the flow of cognitive operations involved in turning physical signals into musical symbols, going from the transducers to the memory systems of the brain. Special attention is devoted to musical facets such as melody, rhythm, and tonality. We then focus on formal models developed to emulate, explain and predict phenomena involved in early auditory processing, pitch processing, grouping, source separation, melody, rhythm, and music structure computation. We make distinctions between generic computational architectures that can be streamlined and tuned to deal with specific musical phenomena and, on the other hand, specific models that only have a local predictive power. Criteria on how to evaluate them are presented and discussed.

Bibliography

[1] H. D. Abarbanel. *Analysis of Observed Chaotic Data (Institute for Nonlinear Science)*. Springer, Berlin, 1st edition, 1996.

[2] V. Adán. Hierarchical music structure analysis, modeling and resynthesis: A dynamical systems and signal processing approach. Master's thesis, Massachusetts Institute of Technology, 2005.

[3] E. Alpaydin. *Introduction to Machine Learning*. MIT Press, 2004.

[4] webpage, Accessed on June 2006. http://audioclas.iua.upf.edu/.

[5] D. Bailey. *Improvisation. Its Nature and Practice in Music*. Moorland, 1980.

[6] A. M. Barbosa. *Displaced Soundscapes: Computer-Supported Cooperative Work for Music Applications*. PhD thesis, Universidad Pompeu Fabra, 2006.

[7] R. Bencina. Audiomulch. Audiomulch webpage. [Online; http://www.audiomulch.com].

[8] P. Burk. Jsyn: A real-time synthesis api for java. In *Proceedings of the International Computer Music Conference*, 1998.

[9] P. Burk. Webdrum. SoftSynth webpage, Accedded on 2004. http://www.transjam.com/webdrum/.

[10] P. Burk. Wire. SoftSynth webpage, Accedded on 2006. http://www.softsynth.com/wire/.

[11] C. Cadoz and M. M. Wanderley. *Trends in Gestural Control of Music*, chapter Gesture - Music. Ircam, 2000.

[12] L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, 1997.

[13] webpage, Accessed July 2004. http://www.ccmix.com.

[14] T. K. Choudhury. *Sensing and Modeling Human Networks*. PhD thesis, Massachusetts Institute of Technology, 2004.

[15] M. Danks. Real-time image and video processing in gem. In *In Proceedings of the 1997 International Computer Music Conference*, pages 220–223, 1997.

[16] G. Deleuze and F. Guattari. *A thousand plateaus: capitalism and schizophrenia*. University of Minnesota Press, 1987.

[17] J. Derrida. *Of grammatology*. Johns Hopkins University Press, Baltimore, 1976.

[18] N. Didkovsky. Java music specification language, v103 update. In *ICMC '04: Proceedings of the Internationa Computer Music Conference*, 2004.

[19] S. Dixon. An on-line time warping algorithm for tracking musical performances. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.

[20] S. Edwards. *Programming and Customizing the Basic Stamp*. McGraw-Hill, 2001.

[21] J. Estrada. Focusing on freedom and movement in music: Methods of transcription inside a continuum of rhythm and sound. *Perspectives of New Music*, 40(1), 2002.

[22] M. Farbood. Hyperscore: A new approach to interactive, computer-generated music. Master's thesis, MIT Media Laboratory, 2001.

[23] J. Foote and M. Cooper. Visualizing musical structure and rhythm via self-similarity. In *Proceedings International Computer Music Conference*, La Habana, Cuba, 2001.

[24] E. Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universidad Pompeu Fabra, 2006.

[25] F. Gouyon. *A computational approach to rhythm description — Audio features for the computation of rhythm periodicity functions and their use in tempo induction and music content processing*. PhD thesis, Universidad Pompeu Fabra, 2005.

[26] E. Hanslick. *Vom Musikalisch-Schönen: ein Beitrag zur Revision der Aesthetik der Tonkunst*. Leipzig, 1854. Eng. trans., 1891.

[27] P. Herrera, X. Serra, and G. Peeters. Audio descriptors and descriptor schemes in the context of MPEG-7. *International Computer Music Conference*, 1999.

[28] Hugo solis webpage. Online; accessed 3-August-2006.

[29] M.-W. Inc. Merriam-Webster Online Dictionary. http://www.m-w.com.

[30] K. Jarret. Vienna concert. CD, 1991. CD cover.

[31] [Online; accessed 3-August-2006].

[32] T. Jehan. *Creating Music by Listening*. PhD thesis, Massachusetts Institute of Technology, 2005.

[33] T. Jehan, T. Machover, and M. Fabio. Sparkler: An audio-driven interactive live computer performance for symphony orchestra. In *Proceedings International Computer Music Conference*, Göteborg, Sweden, 2002.

[34] Jitter. Cycling'74 webpage. [Online; http://cycling74.com].

[35] S. Jordà. Sonigraphical instruments: From FMOL to the reacTable*. In *Proceedings of the 3rd Conference on New Interfaces for Musical Expression (NIME 03)*, pages 70–76, 2003.

[36] S. Jorda. *Digital Lutherie: Crafting musical computers for new musics' performance and improvisation*. PhD thesis, Universitat Pompeu Fabra, 2005.

[37] M. Kennel, R. Brown, and H. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, 45:3403–11, 1992.

[38] P. Khooshabeh, E. Smith, and J. Thomspon. Gestural musical improvisation and programming. In *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 333–334, 2005.

[39] webpage, Accessed on 2004. http://www.la-kitchen.fr.

[40] A. P. Klapuri. Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–282, 2004.

[41] R. Kopiez, M. Bangert, W. Goebl, and E. Altenmuller. Tempo and loudness analysis of a continuous 28-hour performance of erik satie's composition "vexations". *Journal of New Music Research*, 32(3):243–258, 2003.

[42] G. Lewis. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal*, 10:33–39, 2000.

[43] G. E. Lewis. Improvisation and the orchestra: A composer reflects. www.americancomposers.org. [Online; accessed 3-August-2006].

[44] T. Machover. *Sparkler - musical score*. Boosey and Hawkes, New York, 2001.

[45] K. D. Martin. *Sound-Source Recognition. A Theory and Computational Model*. PhD thesis, MIT Media Lab, 1999.

[46] A. Marx. *Die Lehre von der musikalischen Komposition, praktisch-theoretisch*. Leipzig, 1837. Eng. trans., 1910.

[47] The official MPEG website. [Online; accessed 3-August-2006].

[48] P. Nemirovsky. *Improvisational Interaction: a Framework for Structural Exploration of Media*. PhD thesis, Massachusetts Institute of Technology, 2006.

[49] B. Nettl. Improvisation. http://www.grovemusic.com.

[50] B. Nettl, editor. *In the Course of Performance: Studies in the World of Musical Improvisation*. University of Chicago Press, 1998.

[51] F. Nicolas. Pur la beauté du geste. Conference presented at Université européenne de la recherche, June 1995.

[52] L. Nono. La lontananza nostalgica utopica futura. score ed. Ricordi, 1988.

[53] B. S. Ong. *Towards Automatic Music Structural Analysis: Identifying Characteristic Within-Song Excerpts in Popular Music*. PhD thesis, Universidad Pompeu Fabra, 2005. Doctoral Pre-Thesis Work.

[54] Ircam webpage, Accessed on 2004. http://forumnet.ircam.fr/.

[55] F. Pachet. The continuator: Musical interaction with style. In *Proceedings International Computer Music Conference*, Göteborg, Sweden, 2002.

[56] M. Puckette. Pure data. In *Proceedings of the International Computer Music Conference*, pages 269–272, 1996.

[57] H. Purwins. *Profiles of Pitch Classes Circularity of Relative Pitch and Key —Experiments, Models, Computational Music Analysis, and Perspectives*. PhD thesis, Elektrotechnik und Informatik der Technischen Universitat Berlin, 2005.

[58] C. Ramakrishnan, J. Freeman, and K. Varnik. The architecture of auracle: a real-time, distributed, collaborative instrument. In *Proceedings of the Conference on New Interfaces for Musical Expression*, Hamamatsu, Japan, 2004.

[59] C. Reas and B. Fry. Processing: a learning environment for creating interactive web graphics. In *In Proceedings of the SIGGRAPH 2003 conference on Web graphics*, 2003.

[60] J. Ricard. *Towards computational morphological description of sound*. PhD thesis, Universidad Pompeu Fabra, 2004. Doctoral Pre-Thesis Work.

[61] C. Roads. *Microsound*. MIT Press, 2002.

[62] D. Z. Saltz. The art of interaction: Interactivity, performativity, and computers. *The Journal of Aesthetics and Art Criticism*, 55(2):117–127, Spring 1997.

[63] F. Salzer. *Structural Hearing*. New York, 1952.

[64] P. Schaeffer. *Traité des objects musicaux*. Edition du Seuil, Paris, 1977.

[65] A. Schoenberg. *Fundamentals of Musical Composition*. Faber and Faber, 1999. [written 1937-1948].

[66] J. Sloboda, editor. *Generative processes in music: the psychology of performance, improvisation, and composition*. Clarendon Press, 1988.

[67] L. I. Smith. A tutorial on principal components analysis, 2002. Online; accessed 3-August-2006.

[68] H. Solís. Gab: Sistema de reinterpretación para pianos. Master's thesis, Escuela Nacional de Música UNAM, November 2001.

[69] H. Solís. Improvisatory music and painting interface. Master's thesis, Massachusetts Institute of Technology, 2004.

[70] S. Streich. *Automatic Characterization of Music Complexity: a multifaceted approach*. PhD thesis, Universidad Pompeu Fabra, 2005. Doctoral Pre-Thesis Work.

[71] W. Thompson. webpage, Accessed July 2004. http://www.wtosp.org.

[72] J. Timoney, T. Lysaght, M. Schoenwiesner, and L. M. Manus. Implementing loudness models in matlab. In *Proceedings of Digital Audio Effects Workshop 2004 (DAFx)*, pages 177–180, 2004.

[73] D. Twilleager, J. Kesselman, A. Goldberg, D. Petersen, J. C. Soto, and C. Melissinos. Java technologies for games. *Comput. Entertain.*, 2(2):18–18, 2004.

[74] F. J. Varela. *The embodied mind: cognitive science and human experience.* MIT Press, Cambridge, Mass., 1991.

[75] E. Wenger. Metasynth. Metasynth webpage. [Online; http://metasynth.com].

[76] B. A. Whitman. *Learning the Meaning of Music.* PhD thesis, Massachusetts Institute of Technology, 2005.

[77] A. Whittall. Form. Grove Music Online ed. L. Macy, Accessed 2 August 2006. http://www.grovemusic.com.

[78] Wikipedia. Autocorrelation — wikipedia, the free encyclopedia, 2006. [Online; accessed 3-August-2006].

[79] Wikipedia. Dynamic time warping — wikipedia, the free encyclopedia, 2006. [Online; accessed 3-August-2006].

[80] Wikipedia. Principal components analysis — wikipedia, the free encyclopedia, 2006. [Online; accessed 3-August-2006].

[81] P. Williams and D. Ledbetter. Continuo. Grove Music Online ed. L. Macy, Accessed 2002. http://www.grovemusic.com.

[82] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.

[83] M. Woo, J. Neider, T. Davis, and D. Shreiner. *Programming guide: the official guide to learning OpenGL*. Addison-Wesley, 1999.

[84] I. Xenaxis. Mycenes alpha. score, 1978.

[85] D. Zicarelli. An extensible real-time signal processing environment for Max. In *Proceedings International Computer Music Conference*, pages 463–466, Ann Arbor, Michigan, 1998.

Colophon

- This work was developed using linux (Ubuntu 5.10).
- This document was produced with $\LaTeX\,2_\varepsilon$.
- Graphics and images were edited using Gimp.