XIV cim

colloquium on musical informatics

computer music: past and future

zattra, mastropietro, de mezzo,
orcalli, meacci, giomi, schwoon,
bernardini, vidolin, gesfin,
teruggi, meneghini, trevisi,
cervelli, di blasio, volpe,
camurri, mazzarino, rodà,
scantamburlo, timmers, seno,
canazza, de poli, mion,
de goetzen, monopoli,
scagliola, ottaviani, brazil,
fernstrom, meudic, miguel,
o'shea, griffith, tarabella,
di scipio, valle, lombardo,
cafagna, vicinanza,
rath, avanzini, rocchesso,
borin, fontana, serafin, young,
evangelista, cavaliere,
de tintis, bresin, nienhuys,
abbatista, sica,
desainte-catherine, brousse,
maicangi, zanon, mariorenzi,
barontini, mayr, novati,
coco, nieuwenhuizen, zaffiri,
ferrand, de felice, wiggins

proceedings

RICERCA
PRODUZIONE
DIDATTICA MUSICALE

TEMPO REALE 8-10.V.2003
FIRENZE - LIMONAIA DI VILLA STROZZI / STUDIO C - RAI

XIV CIM
Colloquium on Musical Informatics

Firenze 8/9/10 May 2003
Studio C/RAI Sede Regionale per la Toscana
Limonaia di Villa Strozzi

Proceedings

Edited by
Nicola Bernardini
Francesco Giomi
Nicola Giosmin

Tempo Reale
AIMI - Associazione Italiana di Musica Informatica

Μια νέα
λειτουργία του κατηγορητικού
εστιατορίου

Σύντομη εισαγωγή στη λειτουργία του κατηγορητικού εστιατορίου

Ονοματεπώνυμο
Ιδρυματική
διεύθυνση
Email

Ημερομηνία
Δημοσίευσης

Foreword/Presentation

While Contemporary Music production still seems to have difficulties to accept new technologies and production modes, Computer Music seems to be now a fully consolidated domain whose effects and benefits can be assessed in a large number of fields, ranging from contemporary or industrial music production to automotive applications, mobile telecommunications, or ecological acoustics.

The Colloquio di Informatica Musicale (CIM), held in Firenze from May 8 to May 10 2003, has now reached its fourteen edition. The CIM has become an event of international relevance since several years now and this edition is keeping up with the previous ones, as these proceedings demonstrate.

While consolidation needs work and dedication, the passage of time brings with itself high prices. Since the last edition of the CIM, two much-beloved and respected pioneers of Electro-acoustic and Computer Music, Teresa Rampazzi and Pietro Grossi, have left us - mixing our sorrow to the mandatory obligation of perpetuating the memory of their work (which coincides with the origins of Computer Music in Italy). This Colloquium and these proceedings are dedicated to the memory of Teresa Rampazzi and Pietro Grossi.

Therefore, a focus on the historical aspects which are starting to emerge in the Italian and international scene of Computer Music (and the analyses of their continuities and discontinuities) seemed to be the most obvious choice as a central theme. Of course, looking back at history does not make sense if it is not directly connected with our future, if it does not provide the direction for future evolution. This is why the central theme of the XIV CIM is in fact: Past and Future. We are convinced that Teresa Rampazzi and Pietro Grossi would have approved the choice.

The organization of an event like CIM takes quite a lot of energy and resources: I wish to thank the Organizing Committee, the Scientific Committee and the Music Committee for their invaluable contribution to the organization and success of this Colloquium.

Nicola Bernardini

Firenze, May 2003

Scientific Committee
Daniel Arfib (LMA-CNRS - Marseille, France)
Stefano Bassanese (Conservatory of Couneo, Italy)
Lelio Camilleri (Conservatory of Bologna, Italy)
Antonio Camurri (DIST - Genova, Italy)
Chris Chafe (CCRMA, Stanford, USA)
Roger Dannenberg (Carnegie Mellon University, USA)
Giovanni De Poli (University of Padova, Italy)
Giuseppe Di Giugno (Federazione CEMAT, Italy)
Godøy Rolf Inge (Oslo University, Norway)
Shuji Hashimoto (Waseda University, Japan)
Douglas Keislar (Computer Music Journal - MIT Press, Berkeley, USA)
Colby Leider (Princeton University, Princeton, USA)
Marc Leman (University of Ghent, Belgium)
Paolo Nesi (University of Firenze, Italy)
Angelo Orcalli (University of Gorizia, Italy)
Yann Orlarey (GRAME, Lyon, France)
Davide Rocchesso (University of Verona, Italy)
Xavier Rodet (IRCAM, Paris, France)
Robert Rowe (New York University, USA)
Eleanor Selfridge-Field (Stanford University, USA)
Lorenzo Seno (CRM - Roma, Italy)
Xavier Serra (Pompeu Fabra University, Spain)
Martin Supper (Berlin University of the Arts, Germany)
Leonello Tarabella (CNUCE-CNR, Pisa, Italy)

Musical Committee
Luigi Ceccarelli (Edison Studio, Rome)
Agostino Di Scipio (Conservatory of Music, Naples)
Francesco Giomi (Tempo Reale, Florence)
Michele Tadini (Agon, Milan)
Alvise Vidolin (Conservatory of Music, Venice)

Organizing Committee
Nicola Bernardini
Lelio Camilleri
Francesco Giomi
Nicola Giosmin

XIV CIM 2003
Computer Music: Past and Future
Firenze 8-9-10 May 2003
Proceedings

Special session: History of electro-acoustic music

Mastropietro A. A Contribution to a (Pre)History of Computer Music Research in Rome: from Evangelisti and Guaccero to Centro Ricerche Musicali p. 1
Mayr A. Pietro Grossi's musical utopia electro-acoustic music in Florence in the sixties and the seventies p. 5
Novati M. The Studio di Fonologia at RAI in Milan and its current archive p. 8
Zaffiri E. Electronic music in the structuralist current at Turin in the sixties p. 10
Zattra L. Teresa Rampazzi: pioneer of italian electronic music p. 11

Restoring, archives and musical productions

Bernardini N., Vidolin A. Medea by Adriano Guarnieri: a report on extreme live-electronics p. 17
De Mezzo G., Orcalli A. Contemporary music archives: towards a virtual music museum p. 22
Geslin J., Teruggi D. Sound transformations: past and future p. 28
Giomi F., Meacci D., Schwoon K. Sound and architecture: an electronic music installation at the new auditorium in Rome p. 33

Analysis

Barontini P., Trevisi S. A computational approach to the analysis of Incontri di fasce sonore by Franco Evangelisti p. 37
Cervelli M. C. Plus Minus: an algorithmic analysis and a musical realization p. 42
Meneghini M. Stria, by J. Chowning: analysis of the compositional process p. 45

Education

Di Blasio S. Educational, Musical and Multimedial Software Archives p. 51

Expressivity and gesture

Camurri A., Mazzarino B., Volpe G. Analysis of expressive gestures in human movement: the EyesWeb expressive gesture processing library p. 54
Camurri A., Timmers R., Volpe G. The expressive functioning of two acoustic cues in three performances of a Skriabin Etude p. 59
Canazza S., DePoli G., Mion L., Rodà A., Vidolin A., Zanon P. Expressive Classifiers at CSC: an Overview of the Main Research Streams p. 64
Goetzen A. de Expressiveness analysis of virtual sound movements and its musical applications p. 69
Monopoli P., Scagliola F. Towards multilevel gestural control p. 75
Rodà A., Scantamburlo M. An XML representation for the expressive performance of a musical score p. 80
Zanon P., Widmer G. Recognition of Famous Pianists Using Machine Learning Algorithms: First Experimental Results p. 84

Psychoacoustic

Brazil E., Fernstrom M., Ottaviani L. Psychoacoustic experiments for validating Sound Objects in a 2-D space using the Sonic Browser p. 90
Ferrand M., Wiggins G. Memory and Melodic Density: a Model for Melody segmentation p. 95
Meudic B. Musical similarity in a poliphonic contest: a model outside time p. 109

Synthesis, signal processing, physical modeling

Avanzini R., Borin G., Bernardini N., Fontana F., Ottaviani L., Rath M., Rocchesso D. An introductory catalog of computer-synthesised contact sounds, in real-time p. 103
Fontana F., Bresin R. Physic-based sound synthesis control: crushing, walking and running by crumpling sounds p. 109
Cavaliere S., Evangelista G. Analysis and Synthesis of Sounds by Means of the Pitch-Synchronous MDCT p. 115
Coco R. PVLAB: an audio processing program based on Phase Vocoder p. 120
DeTintis R., Malcangi M. A framework for speech synchronized animation OF embodied virtual agents p. 124
Di Scipio A. Sound is the interface. Sketches of a Constructivistic Ecosystemic View of Interactive Signal Processing p. 128
Griffith N. J. L., O'Shea D. J. T. Issues for synthsising musical instruments using signal and physical synthesis models p. 132
Lombardo A., Valle A. A Two-Level Method to Control Granular Synthesis p. 136
Serafin S., Young D. Investigation of the playability of virtual bowed strings p. 141
Tarabella L. The pCM framework for realtime sound and music generation p. 145
Cafagna V., Vicinanza D. Sounds obtained via elliptic functions theory p. 150

Specific Applications

Abbattista F., De Felice F., Scagliola F. Gen-Orchestra: A musical blind watch maker p. 154
Brousse N., Desainte-Catherine M. Towards a Specification of Musical Interactive Pieces p. 159
Malcangi M., Nivuori A. Beat and Rhythm tracking of audio musical signal for sincronization of a virtual puppet p. 163
Nienhuys H., Nieuwenhuizen J. LilyPond a free extensible music engraving system p. 167

Mathematical Models

Mariorenzi L., Seno L. A mathematical model for the holophone, a high directivity acoustical radiator p. 173

Miscellaneous Poster Section

Cavaliere S., Sica G. VM-Zone: a tool for interactive didactical experiences in Music p. 177

Concert Notes p. 181

A Contribution to a (Pre)History of Computer Music Research in Rome: from Evangelisti and Guaccero to Centro Ricerche Musicali

Alessandro Mastropietro
PhD in "History and Analysis of Musical Cultures" - "La Sapienza" University, Rome
ale_mastropietro@hotmail.com

0. Introduction. The vicissitudes of computer music research and production date back, in the Rome area, to the second half of the Seventies. In 1974 the Frosinone Conservatory, whose director was at that time Daniele Paris (an outstanding figure in Roman neo-avant-garde music, both as conductor and as a moving force), instituted one of the first chairs of electronic music in Italy. The professor at that time was (and still is) Giorgio Nottoli, formerly a pupil of Walter Branchi at Pesaro Conservatory and a prominent personality with both musical and scientific competence (particularly in the field of electronic engineering).

Nottoli's activity as researcher, musical hardware producer and composer, as well as teacher of several pupils with similar profile, was pursued in a context which was neither neutral nor unresponsive to this research. A receptive ambience was created, as the years went by, by the Roman musical neo-avant-garde, and especially by Guaccero and Evangelisti who were its most representative figures, also from the cultural and aesthetic standpoint, and were able to consider and formulate, in an interdisciplinary way, questions concerning the relations between musical and scientific research as well as to the combined results of the two disciplines.

The depth and richness of this cultural humus is testified by the variety and longevity of the structures which were created by those trained in this school. In the particular field of computer-music research, besides that carried out in industries, the CRM (Centro Ricerche Musicali) is still active in Rome; this centre was founded at the end of the 80's thanks to the initiative of members of SIM (Società di Informatica Musicale), in its turn founded at the beginning of the same decade. Parallel to these two and dedicated essentially (but not only) to research, the activity of other Roman institutions has continued in the specific field of computer music diffusion. "Musica Verticale" has been active for the last 25 years in the electronic music field, together with other Roman institutions (both "generic" and dedicated to contemporary music). The creative path of such figures and institutions will be discussed in terms of digital/musical tools, works and philosophy starting from its historical and aesthetic background. This will be done by pinpointing questions, themes and operative courses up to the more recent and incisive developments in computer-music research and production carried out by CRM until today in the Rome area.

1. Prehistory: ante-digital. In the musical theorizing and praxis of Franco Evangelisti (1924-1980) and Domenico Guaccero (1927-1984), it is possible to point out some relevant points concerning the matters discussed above:

1) Consciousness of the complexity of aesthetic experience and message, that arises both from "internal" (the opus itself, the musical language) and from "external" relations (social and cultural context, means and technologies of production...). The concept that aesthetic experience is a combination of interrelated "orders" was the basis for the new homonymous review (*Ordini*, Orders), whose editorial committee consisted of Guaccero, Evangelisti, Egisto Macchi, Daniele Paris and Antonino Titone. Only the first number of *Ordini* was published (in 1959) - the second, although ready, never saw the light of day owing to the death of the publisher. Among the "orders" investigated was that of musical technique and technology dealt with in the editorial by Adorno (*Musica e tecnica oggi*) and in the articles of Evangelisti (*Verso una composizione elettronica*, pp. 48-53) and Aldo Masullo (*La "struttura" nell'evoluzione dei linguaggi scientifici*, pp. 54-75).

2) The compositional confrontation with electronic technologies and in general with

scientific knowledge. In the case of Evangelisti, this confrontation was, in the course of his artistic development, as precocious as profound: *Incontri di fasce sonore* (1956-7), his third recognized work, testifies a structure and an experimental attitude which are unthinkable without the contribution of scientific knowledge. Furthermore, all his subsequent theorizing (from his interest in Viesi's *Armosonia* to the drafting of his posthumous book *Dal silenzio a un nuovo mondo sonoro*) shows a striving to the complete integration, right from the basic elements, of musical and scientific experience. As for Guaccero, he only started to use electronic tools at the beginning of the Sixties (*Iter inverso* for 16 instruments and *Improvvisazione* for harpsichord, both composed in 1962), and at the same time he wrote his first essays for the periodical *Collage* on the utilization of musical electronics (*Una conclusione provvisoria*, in no. 1 and *Materiali per una verifica sociologica*, in no. 3/4). 3) These last two works, together with Evangelisti's *Spazio a 5* (1959-60) show an extremely precocious interest, well ahead of his times, in real-time electronics. In *Spazio a 5*, an ensemble of voices is amplified and processed electronically in analog, the sole medium that could then make perceivable the timbrical details of different vocal emissions. In Guaccero's *Iter inverso*, a magnetic tape is not used but only amplified violin and cello, plus a synthesizer, while in *Improvvisazione* one of the sections is played and recorded and the section is then played again simultaneously with the recording. At that time, this was a utopian rather than a pioneer resolution, since it drove a typical deferred-time technology beyond its possibilities and prefigured real-time electronics as the solution for complex interactions.

Guaccero's approach to real-time electronics continued with his interest in analog synthesizers (Synket, Moog, VCS3), organizing workshops and experimental performances with these apparatuses first at Studio R7 and then at Centro di Sperimentazione Musicale (which was founded in Rome at the end of 1971 and had its own Group of live electronic improvisation). Possibly Guaccero saw in the VCS3, with its interconnection logic matrix and its capability of real-time operation, not only a chance to restore a gestural expressiveness to performances of electronic music, but also to use electronics for the reciprocal multiplying of live sound signals. In comparison with the prototype for combining live and electronic resources in musical composition (Maderna's *Musica su due dimensioni*), Guaccero's works are based on a multiplicative (non-linear) instead of an additive way of thinking which is more suitable for the control and generation of complex relationships.

2. SIM (Società di Informatica Musicale). In 1980 a group of people with musical as well as scientific competence founded SIM (Società di Informatica Musicale) in Rome: the founder members were Giorgio Nottoli, Massimo Lindoro Del Duca, Francesco Galante, Michelangelo Lupone and Nicola Sani. The last three, all composers, had been Nottoli's pupils in the Frosinone Conservatory; Del Duca, mathematician, physician and guitarist, was to be the first Italian author of an educational book on digital music.

The SIM emerged at once as an organization open to ideas and productive inputs from musical instrument industries as well as from pure research; consequently, it has produced a wide range of musical technologies, both on commission for industries and private or public parties as well as for compositional research projects within the group. Among these, immediately stands out that concerning real-time and the possibility of working with a digital computer in that field. During its lifetime, SIM produced two machines for real-time digital sound-synthesis: the Soft Machine, designed by Giorgio Nottoli, and Fly10, designed by Michelangelo Lupone. They were the first Italian digital systems (besides those of Di Giugno) that were able to work in real-time using as host a PC. Fly10 (conceived in 1983 for APPLE environment and adapted in 1985 for IBM PC) in particular was based on 4 cards with TMS32010 16-bit processors, capable of working in parallel on sound-synthesis processes up to a maximum of 20 simultaneous algorithms.

The parameters of the signal-processing algorithms could be modified by means of a graphic interface and were controlled by a score editor as well as by a double keyboard, programmable in all its 122 keys. For this system, in 1984 Lupone composed *Mira*, later re-elaborated in his more ample *Ciclo Astrale*.

3. CRM (Centro Ricerche Musicali). In 1988 Michelangelo Lupone and Laura Bianchini (who had earlier joined SIM together with Nicola Bernardini) left the SIM and founded CRM (Centro Ricerche Musicali - Centre for Music Research). According to Lupone, the reason for this decision lay in the tendency, not of the SIM itself but of the market, to demand always more computer products which were limited to a conservative musical language, totally unrelated to the musical philosophy that could be developed through utilization of the new technologies.

Lupone and Bianchini developed the Fly30 System, together with Antonio Pellecchia, with the specific aim of re-conceiving the traditional algorithms for digital computation by running them on a floating-point DSP (TMS320c30), an ideal apparatus for digital filtering. Realized in 1990, Fly30 is an object-oriented system, capable of furnishing high precision calculations for real-time analysis, synthesis and sound processing. The system's graphics editor, permitting the interconnection of the programme modules, provided it with a considerable binding and flexibility capacities and it can be considered an extreme evolution of the module connection by matrix present in the VCS3.

In addition to musical production, Fly30 has been used for research and experimentation in psychoacoustic and organological applications, for example: in the simulation of physical models, in the design of virtual spaces (Flyspace), and in a European research project on noise coordinated by Centro Ricerche Fiat. The musical productions realized at CRM are for the most part the work of Michelangelo Lupone and Laura Bianchini and aim at utilizing above all the real-time abilities of the system. Consequently they are often works with live instrumental and vocal performers, as well as designs of a "teatro dell'ascolto" – that is, a theatre in which the sound itself determines, or re-invents, the dramaturgy and the acoustic space. It can be given either without any live gestures or with the performers' gesture-movements; in the latter case. the "teatro dell'ascolto" concept can be extended to other domains (visual, choreographic, vocal) that are traditional in music theatre. Works of this kind include: *Controfiato* and *Contropasso* by Lupone, performances based respectively on the breathing and steps of dancers, the sound of which is electronically processed (choreographies by Massimo Moricone); the "radio-scenes" *In un grattacielo* (text by Enrico Palandri, music by Lupone) and *Immobile e doppio* (text by Susanna Tamaro, music by Bianchini), conceived as "teatro dell'ascolto" but also staged as musical multimedia theatre in Frankfurt am Mein in 1996; ballets *Il Lorenzaccio* and *La Ronde* by Matteo D'Amico (both for Teatro Comunale, Florence, 1994 and 1995); electronic excerpts from Sehn-Sucht by Alessandro Sbordoni; the theatrical action *Come rosse foglie di luna* (texts by Guido Barbieri and Sandro Cappelletto, music by Emanuele Pappalardo, Laura Bianchini, Luigi Ceccarelli, Michelangelo Lupone).

Other productions by authors not part of the CRM staff include, among others: Lucia Ronchetti's *Quaderno gotico* (1991) and *L'Ape apatica* (2000), Teo Usuelli's *Sinite* (1992), Dieter Schnebel's *Studien* (2nd version, 1993) and *Woerte, Tones, Scritte* (1997), Nicola Sani's *In stiller ewiger Klarheit* (1995), Michele Dall'Ongaro's *1995-Post Scriptum* (1996), Maria Cristina De Amicis' *IST-La nota d'arresto* (1997), Guido Baggiani's *Ka-hal* (1999).

A "teatro dell'ascolto", but also a theatre of borderline creativity, has been the leitmotif of the "Musica Scienza" events, a series of initiatives originating in 1993. By means of concerts and conferences (often in the form of performances), "Musica Scienza" has tried to focus special attention on the branch of contemporary philosophy that develops a comprehensive re-thinking of the various ways for approaching Science and the Arts: "Complex Thought". The titles of the ten editions of the festival held to date are indicative of the themes dealt with: *Invenzione e ricerca musicale* (1993); *Caso Caos Necessità. Il pensiero complesso e la migrazione dei concetti* (1994); *La trama delle complessità* (1995); *Ascoltare lo spazio* (1996); *Polifonie multietniche. Musica e tecnologie per una cultura estesa* (1997); *Rumori. Ordine e disordine, linguaggio musicale e innovazione tecnologica* (1998); *Parola versus Suono. Comunicazione, tecnologia ed espressione della musica contemporanea* (1999); *Musica Infinita. Sculture di suoni, immagini, parole* (2000); *Il*

sogno di una macchina. L'interazione uomo-macchina nella performance musicale (2001); *Musica a tre dimensioni* (2002).

In addition to the festival, CRM has organized other events as well as several Courses of Advanced Studies in Computer Music, cycles of thematic lectures which are held by leading experts in computer reearch applied to Music and Art. Various electronic stringed instruments have been presented during performances and sound installations (another research field chosen by CRM), mostly works offering new ways of listening, particularly those exploiting the quality and form of resonant material (Planephones, Infinite...), those which model the acoustic spaces (Sound Pipes), and those which permit "sculpturing the wavefront" (Holophones). All these installations, subsequently developed in various forms, were designed by Lupone with the aim of implying perceptive space as a musical parameter, which can actually be predicted while composing and controlled in its interactions with other sound parameters.

In the courses, as well as in all CRM research activities, a fundamental contribution has been that of the scientific staff, which at present consists of physicist Lorenzo Seno (coordinator) and engineers Marco Giordano and Marco Palumbi. Seno contributes to the équipe a physical-mathematical approach of high philosophical profile and is available for a confrontation with the musical staff on the basis of "complex thought", of chaos theories and therefore of interaction. A further heritage of the practice of Guaccero and Evangelisti, composers who were anything but solitary in their research work, can be observed in the close teamwork in which every individual contributes with his own specific competence, but is prepared for discussion and comparison with the others.

BIBLIOGRAPHY

AA.VV., *Il complesso di Elettra*, Roma, 1996, Cemat.

AA.VV., (C. Boschi edt.), *Musica e scienza. Il margine sottile*, Roma, 1991, Ismez.

MICHELANGELO LUPONE, ENRICO PALANDRI, LAURA BIANCHINI, SUSANNA TAMARO, *Il progetto CRM per una drammaturgia dell' ascolto radiofonico e del suo spazio virtuale*, in *Atti del XI Colloquio di Informatica Musicale*, Bologna, 1995, pp. 135-138.

LORENZO SENO, MARCO PALUMBI, *Metal string. Physical modelling of bowed string - A new model and algorithm*, in *Proceedings of XII Colloqium on musical informatics*, Gorizia, 1998, Aimi / Università di Udine, pp. 47-50.

ALESSANDRO MASTROPIETRO, *Physical Model and Instrumental Sound. An Analysis of Lupone' s Corda di metallo"*, in *Proceedings of XII Colloqium on musical informatics*, Gorizia, 1998, Aimi / Università di Udine, pp. 244-247.

ALESSANDRO MASTROPIETRO, *Ritratto di musicista da sperimentatore: Michelangelo Lupone*, in *Sonus*, fasc. 18 (a. X, nn. 1-2-3), pp. 98-111.

LAURA BIANCHINI, *Designing a virtual theatrical listening space*, in *Proceedings of ICMC 2000*, Berlin 2000, ICMA, pagg. 406-409.

Figura 1. Olofoni: sound projectors. "Con Luigi Nono" – Cemat –Sonora, Roma, Palazzo delle Esposizioni 2000 (photo Massimo Carroccia)

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

PIETRO GROSSI'S MUSICAL UTOPIA ELECTRO-ACOUSTIC MUSIC IN FLORENCE IN THE SIXTIES AND SEVENTIES

Albert Mayr

timedesign@technet.it

ABSTRACT

This paper attempts to outline a critical evaluation of Pietro Grossi's early electro-acoustic works, from the pieces created in his private studio up to the first experiments at CNUCE-CNR in Pisa. I shall also try to recapture the socio-cultural and aesthetic atmosphere of the sixties and early seventies which, I believe, is crucial for the understanding of Grossi's endeavour toward a drastically renewed musical thought and practice. While the most radical aspects of his widespread concerns have not been accepted, let alone taken up consistently by the musical world, his pioneering work deserves to be better known and discussed in depth as it offers, next to very special kinds of aesthetic enjoyment, important stimuli for reflection on what music is - or ought to be.

1. SOME BRIEF HISTORICAL DATA

Pietro Grossi (Venice 1917 - Florence 2002) succeeded in combining the carrier as the first cellist in the orchestra of the Maggio Musicale Fiorentino with that of composer. Up to the late fifties he had written moderately modern pieces for orchestra and various chamber groups. Several of his compositions met with considerable success on the side of colleagues, critics and public. [1]

Then, suddenly, a drastic change took place in his way of thinking about music and of composing it. This manifested itself in an extreme reduction of material and in formal developments that were based exclusively on the successions and groupings derived from the combinatorial analysis of a limited set of elements.

In the early sixties Grossi discovered electronic music; after having spent some time in the Studio di Fonologia Musicale at RAI in Milan he decided to set up his own private studio. That's how in 1963 the S 2F M (Studio di Fonologia Musicale di Firenze) came into existence. Initially its location was at Grossi's home, two years later it was moved to the Conservatory of Florence where Grossi held the first course in electronic music in an Italian Conservatory. In these same years he found out that computers could produce sounds and immediately set out to explore this new opportunity. After various attempts at using the computers of institutions in Florence he succeeded in getting General Electric (located near Milan) interested in his proposal and began his first experiments. In 1968 he organized, in the context of the well-known festival Maggio Musicale Fiorentino, the first international conference of electronic music centres.

In 1969 the CNUCE (Centro Nazionale Universitario di Calcolo Elettronico) decided to open a Musicological Division and Grossi was appointed as its director; a position he held until the early eighties when he started working with the system at the IROE (Istituto di Ricerca sulla Onde Elettromagnetiche) in Florence. In the following years his interest shifted more and more towards graphic work that he could carry out on the personal computer at home and to which he devoted himself until his death.

[1] For a detailed description of Grossi's artistic curriculum and the list of his works see [1] and [2].

2. THE S 2F M

.....when you get right down to it, a composer is simply someone who tells other people what to do. I find this an unattractive way of getting things done. [2]

As should be remembered, in the early sixties electronic music was a very elitist affair. Professional and even semi-professional studio equipment was very expensive, not easy to find and could only be afforded by institutions such as universities or broadcasting corporations. Access to those hieratic places (beyond the occasional short visit) was rather difficult if you did not belong to one of the contemporary music 'Churches' that were influential at the time. This led several composers - who were not among the chosen few - to do 'their own thing', i.e. assemble a private studio. In doing so they usually replaced financial resources with the ability of rummaging through surplus stores of electronic equipment and the co-operation of adventurous and sympathetic technicians.

Beside these - and other - practical aspects there was also another reason for wanting to have one's own studio. The XXth century has often been called the century of ideologies. If this was true for the socio-political arena, it was also true for the European art world. One of the periods in which the artistic climate was heavily influenced by conflicting ideologies were the first decades after World War II, and thus the newcomer, electronic music, could not escape that climate. This also had to do, of course, with the strong tendency toward ideological theorizing that had characterized the European musical tradition since Antiquity. [3] While in the USA the practitioners of the so-called "tape music" adopted a rather pragmatic approach to the new means, in Europe the situation was quite different: for some years at least each of the 'big' centres professed its specific aesthetic credo.

Now, if you as an electro-acoustic composer happened to have your very own brand of musical aesthetics which had little or nothing to do with that of the influential Churches, you were certainly

[2] [3] p. ix.

[3] See, for instance, Carl Dalhaus' statement: "Musiktheorie ist immer schon Dogmatik gewesen", in C. Dahlhaus, *Hermann Helmholtz und der Wissenschaftscharakter der Musiktheorie* in [4] pp. 49-58.

better off if you had your private equipment, albeit little, with which you could pursue your creative and theoretical goals to your heart's content.

This was the case with the three small private studios in Italy that saw the light in rapid succession in the early-mid-sixties: Florence (S 2F M) founded by Pietro Grossi, Turin (SMET) founded by Enore Zaffiri, and Padua (NPS) founded by Teresa Rampazzi. The aesthetic approach in these three studios, although certainly not identical, had much in common. And in common they had a critical attitude toward mainstream electronic music - in Italy represented by the RAI studio in Milan. Inevitably they thus shared also a certain marginalization.

As Grossi began experimenting with his analog equipment (a dozen sine-wave oscillators, a white-noise generator, and two filters - these were the machines initially) his attitude toward sound and music underwent another change. While his last instrumental pieces had been - with all their distinct asceticism - self-contained compositions, now he more and more lost interest in the compositional process as it is conventionally understood in recent Western culture. Creating pieces, with all it involved on the level of invention, formal procedures, and so on, seemed now less urgent to him than a patient, systematic and, of course, even more ascetic exploration of the new sound world. One of his tenets was that even the most humble sonic event deserved such an exploration (or "research" as he was wont of calling it) which now was possible to a much higher degree of accuracy than ever before.

For *Battimenti*, for instance, he chose an acoustic phenomenon, the beats, which, in spite of its attractiveness, is usually discarded by composers, actually regarded as an unwanted result of poor intonation on the part of the performers and left aside as a mere object for psycho-acoustic studies. Grossi instead, with the stubborness his friends and students knew so well - next to his extreme kindness -, devised a systematic plan for a catalogue of beats resulting from the combination of 2, 3, 4, ...10 sine waves, differing by 1 Hz from each other, and started working.

One has also to remember how troublesome such an operation was with the tools available at that time, i.e. capricious oscillators that often had no intention of keeping the frequency that had been patiently set with the aid of a frequency counter. Thus, both because of the laboriousness of this and similar projects and for reasons which are explained below, we collaborators of the initial period (Riccardo Andreoni, Jon Phetteplace and myself) and, later on, the students in the course were recruited to co-operate directly in the realizations; a procedure which was also adopted in the studios in Turin and Padua.

In all the works of that period the micro- and macro-structures were rigorously determined by procedures derived from combinatorial analysis. In *Om* (from *Offerta Musicale*, Bach's *Musical Offering*) - which Grossi liked to call his musical farewell to the well-tempered system - the notes of the celebrated theme were grouped in clusters according to a permutational ordering and the formal articulation of the clusters themselves followed a pattern of the same kind. Even in the humoristic *Three Sketches*, where he exceptionally employed concrete materials, these were subjected to a strict permutational pattern.

Strongly linked to Grossi's attitude toward the sound material and the use of it, were his ideas regarding the procedures by which music is produced, distributed and listened to. Grossi was convinced that the electro-acoustic means as such had pushed into obsolescence the traditional notions of composer and of (self-contained) composition, since any sound work, once it had been recorded on tape, could be easily transformed, dis-assembled and re-arranged, and thus become a new piece or a number of new pieces, which, in turn, could undergo the same procedures, ad infinitum.

He thought that musical composition had to mutate into an enormous, incessant work in progress to be carried on world-wide wherever there was even a small electro-acoustic studio. This also would have meant, of course, doing away with the traditional individual gratifications for the musicians involved in it in terms of glory (and monetary reward). In the first years of activity of the S 2F M he regularly sent the materials created there (for instance sine wave bands with various frequency ratios) to the other studios around the world "to be used for various compositional purposes", as he wrote in the accompanying text.

Electro-acoustic music, that had freed composers from being subject to the good will or, more often, the caprices of instrumentalists, singers and conductors, should also, he felt, be a field where personal ambition and greed gave way to universal co-operation. Obviously these ideas were misunderstood and ridiculed. Grossi who knew the workings of the established music machinery from the inside, having held for thirty years a position of great responsibility in one of Italy's leading orchestras, grossly underestimated the inertia of that machinery and overestimated the readiness of his fellow composers to follow him. Practically only we studio 'insiders', took up Grossi's suggestion of using for our own pieces the materials that had been created collectively or by some other member of the group.

In line with his approach toward the procedures of production Grossi also favored an extreme flexibility with regard to the distribution and presentation of his pieces; to my knowledge he was the first one to set up what now is called sound installations; or he would intersperse fragments of his and our works - austerily marked only S 2F M - between the instrumental pieces performed in the concerts of "Vita Musica Contemporanea" (a festival organized by Grossi).

While, as we have seen, the official musical world usually showed little inclination toward Grossi's work, positive reinforcement, so to speak, came from the visual arts. In all of the three 'alternative' Italian studios there was, almost from the beginning, an exchange of ideas and collaboration with visual artists sharing similar aesthetic ideals. In Florence it was artists such as Auro Lecci, Maurizio Nannucci, Paolo Masi, whose work was close to what at the time was called "arte programmata" - i.e. art based on algorithmic procedures - and who were grouped around the art critic Lara Vinca Masini. On many occasions visual and sonic works belonging to that tendency were presented together, for instance in the exhibition "Ipotesi linguistiche intersoggettive" [4] which in 1967 was shown in several Italian cities and, in the section "musica programmata", included works by Grossi, Zaffiri, NPS, Lecci,

[4] See [5].

Mayr, Nannucci.

3. PISA

If Grossi certainly was an unusual figure in the analog electronic music scene, he was in many ways unique among the practitioners of digital sound. His attitude toward sound materials acquired a new, almost perturbing radicality.

In the early experiments at General Electric he had adopted a system that made the computer generate directly audible oscillations. In very simple terms it worked like this: the machine was made to perform an operation that would last, say, 1/1000 of a second at a certain voltage, then to repeat it at another voltage; when all this then was repeated many times you would obtain an audio frequency of 500 Hz. Of course the repertoire of available frequencies had certain limits - although those of the well-tempered system were all present - and furthermore you could only obtain one wave form, the square wave and a constant intensity. But Grossi did not mind. He continued with this system also at the CNUCE in Pisa for many years, until the arrival of the TAUMUS system which allowed a certain variety of timbres and intensities. During the 'square wave period' he would often say, with even a slight trace of pride, that he had decided to devote his efforts only to the parameters of pitch and duration and if listeners got annoyed by the never ever changing timbre, too bad for them. If I may add a personal anecdote here: being among those who were not so fond of working only with square waves I wrote a little program which modified the symmetry between the two portions of the wave, resulting in a limited, but workable timbrical variation (the so-called duty cycle, or pulse width modulation). But Grossi was not interested at all; he seemed to consider it an undue concession to sensory gratification.

What fascinated him most in the digital world were not the rich and - finally - controllable sounds, but automation, i.e. the possibility of obtaining complex sonic structures that could be easily modified. Right from the beginning he directed his efforts toward real-time operations, at a time when in other computer music centres long waits for sonic results were part of the daily routine of composers.

He became even more outspoken and relativistic in his work-in-progress aesthetics: because of the easiness with which one could continuously obtain new results, any sonic event created at a given moment was to be considered as essentially ephimeral, to be replaced by new events. If you happened to particularly like one of them, you were of course free to save it or put it on tape, but, he maintained, this was of little relevance since the aesthetic substance resided in the process, not the results.

A concept he created and was very fond of was "Artificial Phantasy"; by that he meant that the computer, thanks to its speed, was able to come up with events and structures the human mind would not be able to think of. In fact, he often included random procedures in the operations he set up and eagerly waited to be surprised by what the machine would produce. To him, being a composer did not mean so much "telling other people - or a machine, in his case - what to do", but creating opportunities for making music happen, opportunities for himself and for others.

In contrast to his belief in the supremacy of processes over results he was convinced that computers could and should successfully replace human instrumentalists in the performance of pieces of the traditional, notated repertoire. And so he untiringly transcribed an enormous body of works, ranging from Bach to Stockhausen.

4. CONCLUSIONS

Let me conclude with a personal remark. As for some time now I have looked at the contemporary music scene more from the outside than from the inside, my view may be partial and biased. It appears, however, that in the last decades contemporary music has retired in safer, less perilous waters and lost the adventurous drive that was present up to the seventies of the last century. One may regret this, or one may not. But regardless of personal preferences I believe that Pietro Grossi's work and example is well worth remembering and discussing. One may not be attracted by the sonic fabric of some of his works - although there are many from which one derives not only intellectual, but also sensory pleasure; one may, also, be doubtful regarding his firm trust in the power of machines in bringing forth a new, unlimited creativity.

But every discipline benefits from rethinking - at least occasionally - its substantial aims, possibilities and limits and to this undertaking Grossi's lesson gives an essential contribution.

5. REFERENCES

[1] Giomi, F. and Ligabue, M., L'istante zero - Conversazioni e riflessioni con Pietro Grossi, SISMEL-Edizioni del Galluzzo, Firenze, 1999.

[2] Jacob, M., Pietro Grossi: un percorso nel Novecento, (Tesi di laurea) Universitá di Firenze - Facoltá di scienze della formazione, 1997.

[3] Cage J., A Year from Monday. London: Marion Boyars, 1975 (repr.), p. ix.

[4] Zaminer, F.(ed.), "Uber Musiktheorie", Arno Volk Verlag, Köln, 1970.

[5] AAVV, Ipotesi linguistiche intersoggettive (catalogue), Firenze: Centro Proposte, n.d.

THE STUDIO DI FONOLOGIA AT RAI IN MILAN AND ITS CURRENT ARCHIVE

Maria Maddalena Novati

novati@rai.it

"Luciano Berio and me had the possibility to found in Milan a Electronic Music Studio. The most important experience for us, until now, was the encounter between technicians and us, the musicians. Technicians came towards us with such an interest and comprehension, that they created our own wishes"

In these flattering words, written by Maderna in 1956 [1] is summed up the spirit, not only spirit of collaboration but also of absolute friendship, that from the very beginning characterised the work at the Studio di Fonologia Musicale at RAI in Milan. The technicians were: Marino Zuccheri (who realises with Berio, Maderna and Nono the most important electronic pieces during the twenty years from the fifties to the seventies), Alfredo Lietti (the engineer who designed the complete set of devices, cables, sometimes reconverted to musical aims laboratory devices or discarded devices) Giovan Battista Merighi and Lucio Cavallin.

In the *venetian republic, in the fifth floor of Corso Sempione*, the official language of the Studio is venetian, as Marino Zuccheri often says, referring to the technicians (his colleagues) and to the composers, except for Berio (*the ligurian contamination...*). [2]

The official opening date of the Studio is in June 1955, but from several months already, Berio and Maderna are preparing electroacoustic tapes using concrete and synthetic materials. Berio has the starting-point form America (excited by the *tape music* by Ussachevsky and Luening) and from France (through his friendship with Schaeffer and the Club d'Essai), Maderna instead brings the contributions from the Kohln Studio through his friendship with Stockhausen and Meyer-Eppler and from his frequentations of the Darmstadt summer courses.

At the very beginning the available devices are simply some tape recorders, record-player with which it was possible to change the record speed, some filters, one oscillator and the Martenot Waves. The real change occurred the following year by the acquisition of nine oscillators and with the voice of Cathy Berberian regarded as the *tenth oscillator*: Berio composed with her, among the others *Thema (Omaggio a Joyce)* and *Visage*.

Doc. Lietti, in an informative relation about the studio (probably in 1956) list proudly, among the others *the nine oscillator panel RC model with Wien bridge [...]* controlled by a cathode-ray tube comparator, *a white noise generator, [...] a panel with octave filters, [...] an analizing quartz filter with continue frequency variation, [...] a Toc generator composed by a sawtooth generator Thyratron [...], an amplitude modulator, [...] a ring modulator [...]*, and among the recording devices, *a couple of magnetic tapes equipped with a device for the variation of the recording duration which keeps the frequency of the sounds.* [3] Some of these devices are today stored at RAI in Turin.

At the beginning the broadcast directors do not care about the developing of new compositive strategies in the Studio, or about the individuation of innovative paths in order to write pure electronic music: the broadcast directors only care to develop a new musical *software* for new sound tracks, more modern and exciting. Electronic timbres seem to increase the emotion in the radio dramas. In an internal relation by the head of the Drama Unit [4] the most important dramas in which *the special musical effects of the Fonology Institute (sic!) are listed: they contributed to create a typical atmosphere and to strength in a good way the artistic level.* About *Uomo e Superuomo"* [5] it is also underlined that *also in this drama, electronic music created a particular atmosphere, a shawian hereafter, trying to translate the inner needs of the author: Shaw himself speaks about a special music of ghostlike violins that should introduce the action and comment it during its developing.* So, during the day the Studio works for *the normal exercise* [6]. and during the night, to build up the most important estate of electronic music of the twentieth century.

But soon, RAI directors noticed that the Studio can become the forge for the experiments for the Prix Italia: [7] by definition, the proper place to judge the most important radio and television (from 1957) programs. Rota wins in 1959 with *La notte di un nevrastenico*, Castiglioni in 1961 with *Attraverso lo specchio*, Paccagnini in 1964 with *Il dio di Oro*, Maderna in 1972 with *Ages* (with the direction of Pressburger), Berio in 1975 with *Diario Immaginario*. History will give the reason to the other work that undeservedly did not win or that were not admitted, such as: *Don Perlimplin* by Maderna (1961) or *Omaggio a Joyce. Documenti sulla qualit onomatopeica del linguaggio poetico (1958).*

Pure electronic music, functional music, experimental music, light electronic music, [8] effects, fonologic research and etnomusicology: these were the principal scopes that were developed inside the Studio.

When I began to reorder all the tapes in the archive (in 1996), only 387 analogic spools (65 four tracks, and the remaining one

[1] Handwritten page preserved at the RAI archive in Milan. Cfr. [1] pg. 273.

[2] Cfr. [1] pg. 179 and following.

[3] Relation without date preserved at the Fonology Archive with four attached schemes dated 1955/1956. Cfr. the article of Lietti in [2] and reprinted in [3].

[4] Relation preserved in the Fonology Archive and dated 23/04/1956.

[5] *Uomo e Superuomo* by G. B. Shaw, music by B. Maderna, conducted by A. Brissoni. For a complete list of drama music composed at the Studio by Berio and Maderna updated to 2001 see [4]

[6] This is the definition of Berio in his article in [2]

[7] For more information about the works presented to the Prix Italia and about the motivation for their exclusion see [5].

[8] In this scope it is important to mention the work of Mario Migliardi.

or two tracks) [9]. were still there. Many of them were not collectioned: I completed the inventory by assigning the marker FON, in order to avoid sovrapposition between my catalogation and the original one by Berio, Maderna and Zuccheri.

Since then, thanks to research both in the paper archive and through the consulting of the Radiocorriere and of the central Rome's archive, were added 123 copies on DAT of works produced at the Studio, or just refined there, or comparison copies, in order to better identify the works or the parts of the works founded in Milan.

I founded drama such as *Morte di Wallenstein*, *Salud*, *Santa Giovanna* with Berio's music, or *Mani*, *Aspetto Matilde*, *Laure persecute* with Maderna's music. I catalogued the new tapes with a neutral mark, different from the original numeration of the Studio, using the **Z** letter, while the authentic catalogation used **A** for rehearsal materials and instrumental music, **E** for electronic music, **Q** for four tracks tapes, **R** for effects and materials for radio dramas.

The inventory that I began in 1996 is stored now in a computer database which contains 745 cards with photographs of the contents of all the archives of the Studio and all the contents of the spools were digitalised.

1. REFERENCES

[1] Rizzardi, V., De Benedictis, A. I, Nuova Musica alla Radio, Cidim-Eri-Rai-Amic, Treviso, 2000.

[2] AAVV, Elettronica V/n.3, Edizioni Radio Italiana, 1956.

[3] AAVV, La musica e l'elettronica n. 2/3, RAI/Eri, 1998.

[4] Novati, M., The archive of the "Studio di Fonologia di Milano della RAI", Journal of New Music Research, vol 30 n. 4, 2001.

[5] De Benedictis, A. I., Prix Italia 1949-1972: musique la radio ou radiomusique?" in Musique et Dramaturgie, esthtique de la raprsentation d'avant-garde", Laurant Feneyrou, Paris, CDMC/SACEM, 2001.

[9] For the precise count and classification of the tapes see [1] and [4]

ELECTRONIC MUSIC IN THE STRUCTURALIST CURRENT AT TURIN IN THE SIXTIES

Enore Zaffiri

enore.z@libero.it

In the early 1960s a structuralist current influenced by Gropius' Bauhaus and by the Dutch De Still began to emerge in Turin. In this atmosphere I started to take into consideration the approach to new sound possibilities offered by electronic apparatus. Faced with the dilemma of how to utilize a sound material divorced from any link with tradition, I found myself confronted with two paths: either that of acting on the basis of pure instinct, or that of organizing the material according to a structuralist principle. I opted for the latter course, postponing a freer expression of creativity to the time when I would have acquired further experience and knowledge.

In the meantime I founded with others who were working in the field of aesthetics (Sandro de Alexandris and Arrigo Lora Totino) the Studio of Aesthetic Information, located in Corso Vittorio Emanuele, 32. The objective of the Center was that of conducting interdisciplinary activities among the various sectors of aesthetics and disseminating the results achieved. It was in this context that I came into contact with numerous artists from various disciplines, among them Pietro Grossi who not only instilled me with enthusiasm but also gave me precious suggestions. I had formulated a basic structure for organizing sonic space which could also be applied for organizing visual space. In this way I had established an interdisciplinary principle not dependent on superficial combinations, but on a common structural basis.

The scheme consisted of a plane geometry figure with various internal structural paths through which I could extract the numerical data for organizing sound and visual parameters.

In the meantime I put the Studio at the disposal of a group of young people who were interested in electronic music, promoting a course which was subsequently (in 1968) to be introduced as teaching matter in the *G. Verdi* Conservatory of Turin, after the Pietro Grossi Conservatory in Florence.

Over thirty years later, after various other experiments conducted for the purpose of exploring the possibilities offered me by technology in the creative field, I returned to the geometrical structure of the '60s to organize images and sounds processed by computer, in a context that I have called *painting of the year 2000*, that is, a harmonious combination of sounds and images that develop and transmute together in time, and thus dynamically, rather than in a condition of static immobility like that of a traditional painting.

TERESA RAMPAZZI: PIONEER OF ITALIAN ELECTRONIC MUSIC

Laura Zattra

Dottorato in Scienze della Musica
Università di Trento
laura_zattra@yahoo.it

ABSTRACT

Teresa Rampazzi (1914-2001), pianist and composer, is one of the pioneers of electronic music in Italy and the first Italian woman to produce and promote it.

She started her career as a pianist; in the 50s she attended the Darmstadt's *Ferienkurse*, she played in the Bartók Trio and was a member of the *Circolo Pozzetto*. She was deeply convinced of the necessity to develop Avant-Garde Music to prepare people for the *Neue MusiK* and new electronic paradigm.

In 1965, Teresa created the N.P.S. (Nuove Proposte Sonore) Group, in collaboration with Ennio Chiggio and they started to produce experiments with analogue equipment. After some disagreement, she continued her activity with young engineers and musicians.

From 1972 to 1979, she taught electronic music at the Paduan Conservatory and began to learn and produce computer music at the CSC (Centro di Sonologia Computazionale), obtaining numerous prizes. In 1984 she retired to Bassano (VI), where she continued her musical activity.

1. INTRODUCTION

Two key words characterize contemporary culture: overcoming and flexibility, that is the desire of rapidly overcoming each result with another, better and more functional product, and that flexibility needed to react to this velocity. Teresa Rampazzi, piano performer, composer and pioneer of computer music in Italy, is a good example of these two rules. In 1969 she wrote: "not only are generations forced to assimilate each transformation every ten year, but everyone must change ideas and attitude once or twice during his whole life" [1]. Well, T. Rampazzi's career shows numerous esthetical and musical turns which demonstrate, a posteriori, a clear artistic will. Before dedicating herself to electro-acoustic music, she became an assiduous promoter of Avant-Garde music: she played, for the first time in Italy, the Schönberg Suite op. 25 for piano; in her lounge, she often met Bruno Maderna, Heinz Klaus Metzger, Sylvano Bussotti, etc. When she was 50 years old, she took up analogue music composition. Finally, aged 60 and with her usual great enthusiasm, she began to study and produce computer music.

Nevertheless, T. Rampazzi is especially known for her electro-acoustic activity. She first encountered the new electronic paradigm exactly at its birth, "listening to its cries in the large Marienhöhe concert hall, in Darmstadt" [2, p.122]. Thanks to this discovery, electronic music landed in Padova. Thanks to T. Rampazzi, some composers and engineers got together and founded the N.P.S. (Gruppo Nuove Proposte Sonore) and the CSC (Centro di Sonologia Computazionale). Knowing her activity means to analyze Padova's lively but sometimes contradictory musical life, over the last 50 years.

1.1. Teresa Rampazzi performing activity

T. Rossi (this was her maiden name) was born in Vicenza in 1914. She had been interested in music since her childhood so, "as a typical good family daughter, I played the piano" [3, p.32]. She started studying music with a local teacher (Tonolli), but her father soon sent her to the Milan conservatory where she studied with Arrigo Pedrollo. Here she got to know Bruno Maderna and began to receive friends in her living room, people who would become very important in the contemporary music scene: Franco Donatoni, René Laibowitz, Severino Gazzelloni and Maderna.

In 1952 and 1954, Teresa Rampazzi attended the *Internationale Ferienkurse für Neue Musik* in Darmstadt and listened to electronic experiments made by Eimert. She had understood that that was the only way to completely reject tonal music. But in *Veneto*, the people were not ready to know electronic experimental music or Avant-Garde music. That's why in 1956 Teresa began to play with the Bartók Trio (Elio Peruzzi: clarinet, Edda Pitton: violin, Teresa Rampazzi: piano), and decided to promote the Avant-Garde music by Anton Webern and Alban Berg. In a city where Mahler's music was not known yet, this activity was not so easy. It was therefore crucial to prepare and open the public's mind to the *Neue Musik* and to electronic music.

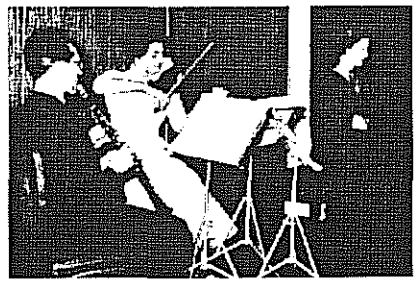


Fig.1. The Bartók Trio in 1959 (from the left: E.lio Peruzzi, Edda Pitton, Teresa Rampazzi).

During this period, Teresa became a member of another group named *Circolo Pozzetto*. The cultural and political circumstances of this merit a little

digression. Its founder, Ettore Luccini, was an intellectual engaged in the Communist Federation of Padova. From 1956 to 1960 he drew together a large number of intellectuals (visual artists, poets, musicians, teachers, etc.) who met in periodical assemblies and talked about the cultural situation. He organized expositions, concerts and conferences in order to let the citizens know the most recent intellectual and artistic trends [4]. The *Pozzetto* "would have been the first positive answer to the isolation of many intellectual communists and to their own isolation towards the Party: a place where it would be possible to meet people, to exchange ideas, to reduce and overcome mental severity" [5]. Nevertheless, the group encountered serious difficulties caused by Luccini's role in the political party. From the beginning he was suspected of promoting free thinking instead of encouraging thought from a political point of view. The Communist Federation had vigorously opposed this activity and eventually, this forced him to leave the city council's Cultural Commission and the Provincial Federation. That's why Luccini ended the *Circolo Pozzetto*.

Considering this background, Teresa Rampazzi's participation did not come from political motivations. It originates from the coincidence of being introduced to the artist Sylvano Bussotti, a member of the group. This could be important to avoid some consequences derived from a political interpretation. However, some years later Teresa would be strictly involved in political sense, during students' and workers' demonstrations. This was part of her passionate character. In any case, from 1956 to 1960 she took part in *Circolo Pozzetto* to promote Avant-Garde and electronic music with auditions and conferences. During these conferences, Teresa talked about *Neue Musik* (Karlheinz Stockhausen, Henri Pousseur, etc.) and about the music composed by her friends F. Donatoni, Niccolò Castiglioni and Maderna. Moreover, in 1959 during auditions dedicated to the music realized between the two world wars, she played with the Bartók Trio and performed music by Bartók, Hindemith and Berg. But her most noteworthy concert is the one set in 1959 with John Cage, H.K. Metzger and S. Bussotti. The four performers 'played' with whatever could resonate, with the piano but also with sticks and screws.

1.2. The Informal Revolution

Combined with her fascination with electronic sounds, Cage's Informal music made Teresa see the possibility to leave behind tonal music and the traditional form 'start-development-end'. At this moment, all preambles for changing her musical activity were complete. She sold her piano in curious circumstances: a legend tells that she demolished it during a performance with Cage. Perhaps the truth is that her informal music techniques worried her friends and audience in such a way that this myth developed. In fact, musicologist Ugo Duse once said: "in Padova there were these crazy people who kicked the piano, slammed the keyboard shut, plucked the strings; this phenomenon, which later spread across Italy, began softly in Padova too [...]. I consider Teresa Rampazzi to be the only person in Padova brave enough to be interested in it" [6]. According to another legend, Teresa's husband bought her piano from her in order to save it from destruction. In any case, it is true that she did not want to see her piano anymore.

In 1964 she met the visual artist Ennio Chiggio, member of the famous *Gruppo Enne*. This was a 'laboratory' whose components produced unpersonalized artistic products. Chiggio himself was fascinated by new sounds of some pieces created at *Centro di Fonologia della RAI di Milano*. Since he worked for an electronic firm, he easily obtained a semiprofessional recorder, a low-frequency generator, and a mixer. Teresa contacted him to borrow a mixer and a tape player for an audition. Chiggio writes: "that meeting was fatal, Teresa was very extraverted and enthusiastic and I showed her my equipment. She talked to me about the Bartók Trio, about Darmstadt, Maderna, Cage and we started to meet very often" [7].

2. THE ELECTRONIC MUSIC

2.1. The Analogue Music: N.P.S. Group

Initially, Teresa and Chiggio made some unofficial experiments. They listen to the *RAI-Radio Televisione Italiana*'s third musical program from which they recorded many pieces, from medieval to contemporary music (a lot of these tapes may be found at Padova University – Music Department). As medieval music was almost entirely vocal, it was a good model for studying how to use simple sinusoid sounds.

Their first experimental work consisted in a sound collage which would have functioned as musical background for an exposition by the *Gruppo Enne* at the Biennale di Venezia. It was a 30 minute tape, played very slowly (4.75 cm/s) to make even longer! On this occasion, Teresa and Chiggio decided to found a group inspired by *Gruppo Enne*. On 20th May 1965, they founded the N.P.S. (Nuove Proposte Sonore) [8]. The choice of the name (New Sound Proposals) shows a firm decision to eliminate any artistic aspiration: this was pure research and each result, named *oggetto sonoro* (sound object), would have been anonymous. These *oggetti sonori* were reminiscent of the Schaefferian masterpiece *Traité des objets musicaux* [9], but they were intended to be its evolution. The first N.P.S. members were Teresa, Chiggio, Memo Alfonsi (a young engineer interested in music) and Serenella Marega (Teresa's friend).

The group's inner organization reflected this ideal collectivism: all instruments were common property, even if individual members had bought them. Amongst the musical equipment they used was the following: a low frequency EICO generator, a radio which produced long-wave frequencies (useful for simulating coloured noise), a tape recorder, a mixer, a two track recorder [7]. In order to reverb sound, they put a loudspeaker at the top of the stairwell and a tape recorder at the bottom of it! "In any case, we could not

avoid the recording of a slammed door, during the so-called composition *Ricerca 4*" [2, p.123].

The manifesto was extremely rigid: "the instrument has no possibilities anymore, it has been raped, destroyed, it can no longer communicate..." [8]. This reflected the deep desire to conduct authentic research without thinking of a potential audience.

In 1965 two types of sound research started: *Ipotesi* (*Hypothesis*) and *Ricerche* (*Research*). Both categories were works from about 3' to 7' minutes, signed by Rampazzi, Chiggio, Alfonsi, Marega and Gianni Meiners. *Ipotesi 1* and *Ipotesi 2* consisted in a opposition of square waves streams and sinusoidal glissandi streams based on the Fletcher studies. In *Ricerche*, N.P.S. members studied the effect of close frequencies which diverge in time reaching the two extremes of the spectrum range. One year later they realized some sound objects – *Operativi* (*Operational*) – with coloured noise, square waves and beatings. *Funzioni* is a research of the glissando's effect. In 1967 N.P.S. produced the works *Ritmi* (rhythmical study) and *Moduli* (impulse and its attack) and, in 1968, *Interferenze* (*Interferences*), *Dinamiche* (*Dynamics*) and *Masse* (*Masses*) [7], [8].

Nevertheless, in 1967 members' ideas started to diverge. Teresa's musical instinct led her to aspire to greater artistic freedom, which was in contrast with Chiggio's motivation. For this reason, Chiggio preferred to leave the group. He once said: "Teresa was like that, she loved and hated to excess, she had enormous enthusiasm and was very stubborn. At that moment we could not reach a compromise so we did not meet for years, even if towards the end our friendship and mutual admiration permitted us to overcome all misunderstandings" (personal communication). In Chiggio's interpretation, "initially, N.P.S. Group had all the ingenuousness and force which characterize Avant-Garde movements. It aimed at "new" proposals, forgetting that from the Middle Ages, many musical trends had employed this term. Its manifesto annihilates every expressive quality of the traditional instruments, but in the end the group would deny all that" [7].

Fig. 2. N.P.S. members during the first period (1965-1967): from the left Chiggio and Alfonsi, behind Rampazzi, Marega, Meiners [7].

During the second period of the N.P.S., Teresa opened her laboratory to young engineers and musicians and started a new epoch dedicating herself to teaching. In November 1968 she began to give free musical instruction. The students, who became new N.P.S. members, were Giovanni De Poli (born in 1946, engineer), Patrizia Gracis (1947, philosopher), Luciano Menini (1948, engineer), Serena Vivi (1945, mathematician) and, from December, Alvise Vidolin (1949, engineer). In a mutual exchange, they spoke about their technology knowledge, whereas Teresa offered her humanistic and musical experience.

At this moment musical equipment was arranged thus: six oscillators with manual control, six oscillators for frequency modulation, a white noise generator, an octave filter, a filter with changeable band, an amplitude modulator, a note switch, a reverb, a 10-channel mixer, an audio signal switchboard, four tapes, a stereo amplifier and a frequency meter [10].

Pieces created by N.P.S. members were now signed. This happened particularly for Teresa's compositions. *Freq.mod 2* (or *Fremod* 2), for example, is based on "fast glissandi in high frequencies, thick exploding glissandi, short explosions of filtered sounds, bands of frequency modulation as a chorus" [8]. It's interesting to notice that even the terminology is experimental and tries to classify sounds objects not definable with traditional terms. In July and September 1970, Teresa Rampazzi was invited by the Washington Catholic University of America and later by the Warsaw Experimental Studio were she exhibited the piece *Insiemi*. In January 1972 the Festival International de Musique Electroacoustique in Paris performed her music during a concert.

The initial attention on the term *sound object* changed now in *musical object*. This meant that the electronic instruments' analysis aimed at the 'musical' synthesis of the different electronic techniques. Moreover, "in our group cooperation continues to be crucial, but we pay attention to the individual proposals" [8]. In fact, from 1970 N.P.S. members did not produce research products anymore, but single works which combined all analytical results with more freedom. They also realized some soundtracks for films and documentaries: *La città*, 1971 (Studio Bignardi); *Vademecum*, 1971 (Max Garnier); *La città feticcio* and *La salute in fabbrica*, 1972 (Giuseppe Ferrara), *Endoscopia*, 1972 (Domenico Oselladore) [8].

In 1972, N.P.S. members bought a Synthi A Ems, which permitted better musical results. The Computer music trend started to seem attractive and considered necessary to develop Teresa's musical ideas.

In October 1972 the conservatory of Padova instituted a new electronic music course and assigned it to Teresa (this was the third course in Italy). She brought with her the whole musical equipment and, consequentially, she ended N.P.S. Group. However, she continued to produce works with the synthesizer: *La cattedrale* (1973) based on the set theory, *Glassrequiem* (1973), *Breath* (1974) and *Canti per Checca* (1975) realized with the voice of her daughter Francesca.

YEAR	TITLE	AUTHORS	Duration
1965	*Ipotesi 1*	A C Ma Me R	4'
	Ipotesi 2	A C Ma Me R	5'
	Ricerca 1	A C Ma Me R	6'
	Ricerca 2	A C Ma R	5'30"
	Ricerca 3	R	7'30"
	Ricerca 4	A C Ma R	5'5"
1966	*Operativo 1*	Ma R	3'
	Operativo 2	Ma R	3'25"
	Operativo 3	R	3'55"
	Funzione 1	Ma R	4'10"
	Funzione 3	Ma R	1'30"
	Funzione 4 (2 tracks)	A Ma R	2'30"
	Funzione 5	Ma R	2'
	Funzione 5a	Ma Me R	1'
1967	*Ritmo 1*	Ma R	3'
	Ritmo 2	Ma R	3'
	Ritmo 3	A Ma Me R	2'40"
	Modulo 1	Ma R	3'
	Modulo 2	Ma R	3'
	Modulo 3	Ma R	2'20"
	Modulo 4	R	3'40"
	Modulo 5	Mk R	2'30"
1968	*Interferenze 1*	Ma R	3'30"
	Interferenze 2	Mk R	4'10"
	Dinamica 1	Ma Mk R	3'
	Masse 1	Mk R	3'
	Masse 2	R	2'30"
1969	*Freq.Mod 1*	Ma R	4'15"
	Freq.Mod 2	R	6'50"
1970	*Imp & Rith.*	R	4'
	Environ	R	7'
	Insiemi	G R	7'20"
1971	*Eco 1*	DP G Men Vid	3'50"
	Filtro 1	DP G Men Vid	6'80"
	Taras su 3 dimensioni	DP Men R Vid	10'50"
1972	*Immagini per Diana Babylon*	R	2h
	Computer 1800	R	8'20"
	Hardlag	DP Me Vid	5'15"

Fig. 3. Table of N.P.S. works [7], [8].
[Legend: A=Alfonsi, C=Chiggio, Ma=Marega, Me=Meiners, R=Rampazzi, Mk=Mazurek, DP=De Poli, G=Gracis, Men=Menini, V=Vivi, Vid=Vidolin]

2.1. Teresa and the 'big monster'

Musical Informatics was rapidly spreading across the world, but its unfriendly language was even more evident to musicians. Furthermore, in Italy the academic society (e.g. engineering faculties) and the conservatories were so definitely separated that composers had no possibilities to learn new digital technologies. For this reason, in 1965 Pietro Grossi did all he could for creating the first electronic music course in the Florence conservatory, followed by other courses including the one in Padova.

Moreover, the positive activity of the N.P.S. Group and the presence, in the Engineering Department, of Alvise Vidolin, Giovanni De Poli and their Professor Giovanni Battista Debiasi, showed that computer music could be seriously considered as a new research trend. In the beginning of the '70s, De Poli, Vidolin and Graziano Tisato, monitored by G.B. Debiasi, began an important investigation in sound synthesis [11]. They worked at the CCA (*Centro di Calcolo di Ateneo*), the university's administrative building, whose computers were not used after 2 PM and could be exploited for research activity. The composer James Dashow worked with them; he brought from the USA the software Music 360 and Music4BF, and named the group *Computer Music Group*. In 1979 Debiasi and the university chancellor institutionalized this association and gave it the name CSC (Centro di Sonologia Computazionale).

In its constitutive statute, CSC intended to develop research, musical production and teaching. Teresa Rampazzi's participation must be considered from this two latter points of view. During the middle '70s, as lecturer she took part in electronic music seminars set in Vicenza (*Seminari di Villa Cordellina – Montecchio Maggiore – Vicenza*), organized by the composer Wolfango Dalla Vecchia (conservatory's director) in collaboration with G.B. Debiasi and with the *Computer Music Group* members. Students came from Italy, France, Germany, England, Romania, Greece, Australia, Canada, USA and South America.

However, Teresa Rampazzi suffered for the lack of collaboration between the conservatory and the *Computer Music Group*. For this reason, in 1974 she convinced the director W. Dalla Vecchia to establish a formal contract, which allowed the students access to the University's machines.

In 1974 Teresa was 60 years old. Her enthusiasm in learning new digital techniques was remarkable. Regarding this, she said: "I thought the computer could be finally a serious person whom you could speak with. I began to study but, unexpectedly, I discovered a great length not in calculating (the computer was incredibly quick) but in writing all the instructions required for a mutual comprehension" [2, p.125]. Teresa Rampazzi's difficulties with computer were also caused by her eyesight problems. For this reason she called it 'the great monster'. But her troubles were not so different from other composers who learned computer language and worked with perforated listings. In 1976 G. Tisato created for her the language ICMS (Interactive Computer Music System). It was a real time software which did editing and mixing and was connected with a video. Teresa selected her 'windows' with a light pen. Her first piece, which is also the first piece realized by *Computer Music Group* together with a piece by James Dashow, was exactly titled *With the light pen* (1976). This piece obtained a special mention at the International Electroacoustic Music Competition in Bourges (France).

Right through this period, she taught to her students analogue techniques during her lessons at conservatory, whereas at CCA she collaborated with them on the same level in producing computer music. She realized other pieces: in 1978 *Computer dances* (Special mention, Bourges, 1978) which is "based on 8 sections in which the signals gradually overlap in a constantly increasing number according to the shortening of the signals" [12a]. It had been realized on the IBM S/7 (16bit) of the CCA, with the ICMS. During its realization, Teresa wrote (it's noticeable her ability and candor to describe her music and compositional processes): "the great monster who keeps me nailed here to my place has given me a time limit and not a moment can be lost. Great reflectors have been

placed around me to enable me to work all night. My eyes are burning; at times I feel I cannot see how everything should be; the entire immense vault covered with figures and a strange type of flowers. Nothing however will be still. I have predisposed things so that everything moves and dances as elegant a manner as possible" [12b].

In 1979 she realized *Fluxus*, (Disk LP EDI-PAN PRC S 20-16, Rome, 1984), based on a fragment by Heraclitus; in 1980 *Atmen noch* which won the Second Price at VIII *Concours Internationale de Musique Electroaoustique de* Bourges (1980, first price not awarded). Among the other pieces we mention *Requiem per Ananda* (1982), which combine some phonemes taken from the Requiem Mass by Louis da Victoria (1548-1611) with digitals signals.

In 1984 her husband died and this experience sharpened even more her refusal towards the past. She sold her home, gave all her musical property to conservatory and musical institutions (now these materials – disks, books, scores, tapes, sketches – may be found at Padova's Music Department) and decided to retire. She first went to Assisi for some years and later she settled in Bassano (Vicenza) where she lived until December 2001. Here she created a little home studio where, helped by the technician Tonino Delfino, she continued to compose (especially with the Yamaha DX7) ad listen to music, faithful to her desire to put music before her personal achievement.

YEAR	TITLE	Traces	Duration
1973	*La cattedrale*	Mono	15'15"
1974	*Breath*	4 tracks	18'34"
	Glassrequiem (Omaggio a Mozart)	Stereo	9'10"
1975	*Canti per Checca*	4 tracks	9'45"
1976	*With the light pen*	Stereo	8'30"
	Melismi (Stockholm – Padova)	4 tracks	7'50"
1977	*Timbri 1597-1977 (Omaggio a Giovanni Gabrieli)*	4 tracks	14'30"
1978	*Computer dances*	4 tracks	10'30"
1979	*Fluxus*	Stereo	10'40"
1980	*Atmen noch*	4 tracks	15'10"
1981	*Metamorfosi*	4 tracks	8'30"
	Danza seconda	4 tracks	8'
1982	*Geometrie in moto*	4 tracks	11'40"
	Requiem per Ananda	4 tracks	8'15"
1983	*Spettri*	4 tracks	9'46"
1984	*Eka'*	4 tracks	19'50"
1987	*Parole di Qoelét*	Stereo	30'
	...Quasi un Haiku...	Stereo	?
1988	*Forse fantasmi*	Stereo	?
'90s	*Incantamento di Silo*	Stereo	?
	Polifonie di Novembre	Stereo	?

Fig. 4. Teresa Rampazzi's works from 1973 to 2001.

3. CONCLUSIONS

Throughout her career, Teresa Rampazzi played an important role in the promotion of contemporary and electronic music. Her cultural and artistic independence permitted her to be sensitove to any intellectual stimulus and to change her ideas without contradiction. She studied in depth every new esthetical choice with obstinacy and determination, forcing herself to forget completely each precedent experience.

Finally, in order to understand her activity as a musician, it is important to underline her dedication to teaching and her perspective as a woman. Nevertheless, she was not a feminist and she did not recriminate sexual differences in the artistic world. She once said: "just as there are many women who have a demanding profession, the same is true in music [...]. Unwittingly I risked compromising my musical interests when I got married. But they were much too important for me" [3, p.66]. As a woman, she sometimes seemed an eccentric composer, but she refused to be considered a woman composer whose gender dominated her music. If someone asked her if there was a feminine way to make music, she answered: "absolutely not. There is neither male nor female music. There are pieces composed by men which seem to be composed by a woman and vice versa, if by 'feminine' you think of something sweet, elegant, delicate. But a woman can be as vigorous as a man, or even more!" [3, p.72].

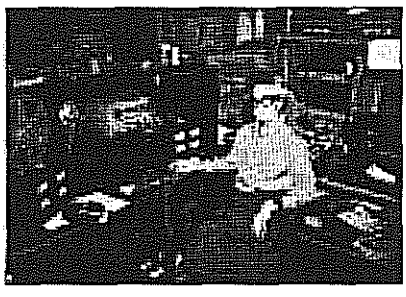


Fig. 3. T. Rampazzi in her home studio in Bassano, Vicenza (late '90s).

4. REFERENCES

[1] Rampazzi, T., "Tempo e ritmo", *Filmspecial*, July 1969.

[2] Rampazzi, T., "Piccolo discorso con Michela", *Autobiografia della musica contemporanea* (Michela Mollia ed.), pp. 122-126, Cosenza, Lerici, 1979.

[3] Galanti, L., *L'altra metà del rigo. La donna e la composizione femminile oggi in Italia*, Imola, Grafiche Galeati, 1983.

[4] *La stagione del Pozzetto. 1956-1960. Documentazione e dibattiti da un avvenimento culturale in Padova*, unique issue, 1979.

[5] *Il Pozzetto. Un orizzonte aperto. Ettore Luccini e la sua lotta contro l'isolamento politico e culturale della sinistra*, Padova, Editore Programma, 1992.

[6] Di Capua, G., *Teresa Rampazzi. Fino all'ultimo suono*, Radiotre, 3/10/17 marzo 1993 (prima puntata).

[7] Chiggio, E., *Oggetto sonoro. Lectures. Musica elettronica – Fonologia n. 7*, Edizioni multimediali del Barbagianni, p.1, March 2002.

[8] *NPS 65-72. Sette anni di attività del gruppo nuove proposte sonore nello studio di fonologia musicale di Padova*, conservatorio 'C. Pollini', Padova, unpublished.

[9] Schaeffer, P., *Traité des objets musicaux*, Paris, Seuil, 1966.

[10] Vidolin, A., Contatti elettronici. La linea veneta nella musica della nuova avanguardia, in *Venezia Arti 1989/3*, Bollettino del Dipartimento di Storia e Critica delle Arti dell'Università di Venezia, 1989.

[11] Durante, S., Zattra, L. (eds.), *Vent'anni di musica elettronica all'Università di Padova. Il Centro di sonologia computazionale*, Palermo, CIMS, 2002.

[12] Rampazzi, T., "*Computer dances* 1978. Technical description [a] and Metaphoric description [b] (in Engligh), unpublished.

MEDEA BY ADRIANO GUARNIERI: A REPORT ON EXTREME LIVE ELECTRONICS

Nicola Bernardini[†], *Alvise Vidolin*[‡]

[†]Centro Tempo Reale, Firenze
[‡]CSC-DEI, Università di Padova
[†]nicb@centrotemporeale.it, [‡]vidolin@dei.unipd.it

ABSTRACT

The present paper describes *Medea*, a full-scale musical work by Italian composer Adriano Guarnieri whose performance has required a considerable amount of human and technical resources. It is to be intended as a performance report to witness the current state-of-the-art in real-world Live Electronics music endeavors.

1. INTRODUCTION

Medea is a large musical work by Adriano Guarnieri for soli, orchestra, choir and Live Electronics which has been described by its author as a *Video-Opera*. This term should not be intended exclusively as a specification of visual requirements or visual technologies implied by this work. Rather, it is the author's indication of the underlying musical vision which includes metaphorical references to video mechanisms such as zooming, edited sequences, etc.

2. WORK METHOD

The collaboration between Adriano Guarnieri and the Centro Tempo Reale has spanned well over ten years. During this period of time, we have established with the composer a work method that has allowed us to define the role of Live Electronics (and electronics in general) in his music with a precision that can be compared to that used for vocal and instrumental part scoring. Usually, this work starts well before the layout of the instrumental part with an approximate definition of the Live Electronics performance environment of the piece[1] related to the role and function of electronics desired by the composer for a given compositional work.

Once this approximate definition is agreed upon, Adriano Guarnieri starts writing the music keeping margin notes on each page about specific electronic performance requirements. When the instrumental score is ready, some meetings are organized to write another score which details the complete Live Electronics setup and each and every action which will have to be performed by the Live Electronics players.

A brief example of the operational data which may be included in this second "score" may include:

- the accurate position of soloists and instrumental players on the stage or in the hall according to the prescriptions of the composer and to the features of the concert venue
- the loudspeaker disposition surrounding the public (the disposition designed for *Medea* is described in Fig.1 as an example; this disposition must then be mapped into the real-world context of each venue; the actual location of in the context of the PalaFenice in Venice is outlined in Fig.2).
- the paths that sounds will have to follow with speeds and accelerations that will be related to music itself
- the sound processing that will carry some instrumental sounds in the register of others or else in other completely abstract dimensions

[1] for a definition of *Live Electronics performance environment* cf. [1].

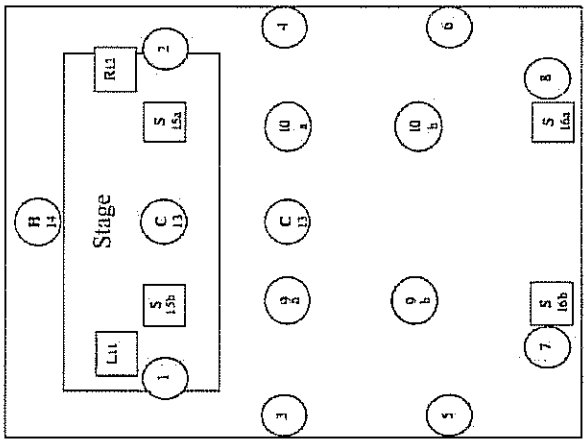


Figure 1: *Medea* - Loudspeaker disposition

Almost inevitably, a number of metaphoric abstract terms (such as "celluloid" movement, "thin metal moans" etc.) are created by the composer during the definition of the Live Electronics score to express the intended results. Most of these terms will be mentioned (in quotes) in this paper because their interpretation proved to be very significant in working out the appropriate related processing.

3. *MEDEA*'S LIVE ELECTRONICS

Live Electronics are a very important musical component of *Medea*. As such, they are described here with the greatest detail allowed by the length of the paper. A description of the roles and functions intended by the composer precedes the technical outline, because it is important to understand the *a priori* musical requirements before delving into implementation details.

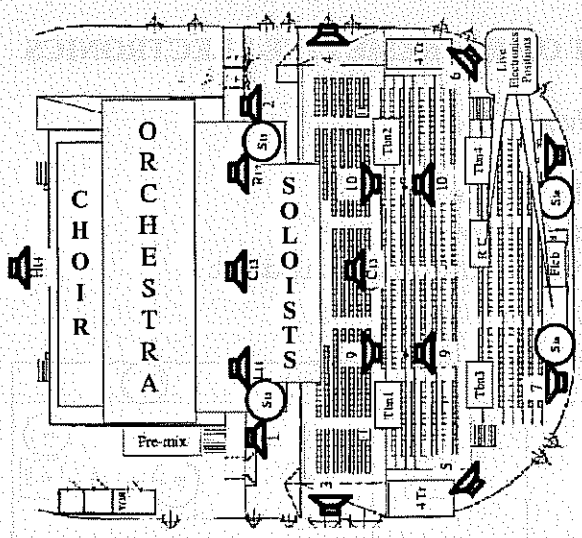


Figure 2: *Medea* - PalaFenice disposition

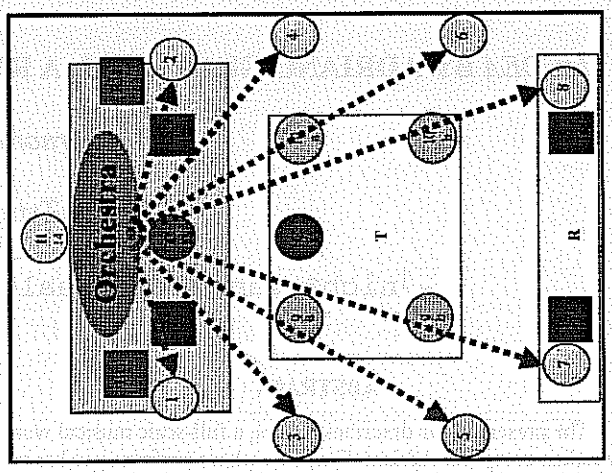


Figure 3: *Medea* - Transparent sound reinforcement

3.1. Roles and functions of Live Electronics

Following ideas developed in previous works (*Orfeo Cantando... Tolse*, *Quare Tristis*, *Pensieri Canuti* and *Passione secondo Matteo*, cf.[2, 3]) the composer has used Live Electronics in *Medea* to enhance the perception of extremely thick textures by displaying them over space and frequency ranges. However, in *Medea* the investigation on sound movement in space does a step forward: the composer proposes the concept of *video opera* not in a visual but rather in a musical sense, where each musical part can be focused upon in turn while being part of a whole ever changing dynamic complex. In *Medea*, the sound processing acquires features and functions which may be termed as "visual": microphones, processing and spatialization become a sort of "sound cameras" which allow global sound views as well as singling out of foreground elements. Each musical page is a scene and the strength of musical writing combined with the design of Live Electronics are able to offer the music contained therein as a whole but also, and simultaneously, as a complex where all details may be picked out individually. Therefore, in *Medea* there are essentially two sound reinforcement modes: "transparent" reinforcement which simulates the acoustical response of an architectural space while boosting its sonic response by calculating appropriate time delays for each speaker (cf.Fig.3) , and sound spatialization which positions natural acoustic sources in a space location which differs from their actual real position. In the latter case it is not only possible to simulate precise locations but also far/near plans and the movement of sound in space along a variety of paths with different speeds.

3.2. Setup

Fig.4 describes the Live Electronics setup of *Medea*.
This schema refers to:

a. a total of 68 microphones that pick up the orchestra, choir and soloists

b. a premix stage for the choir, trumpet section, orchestra and percussion

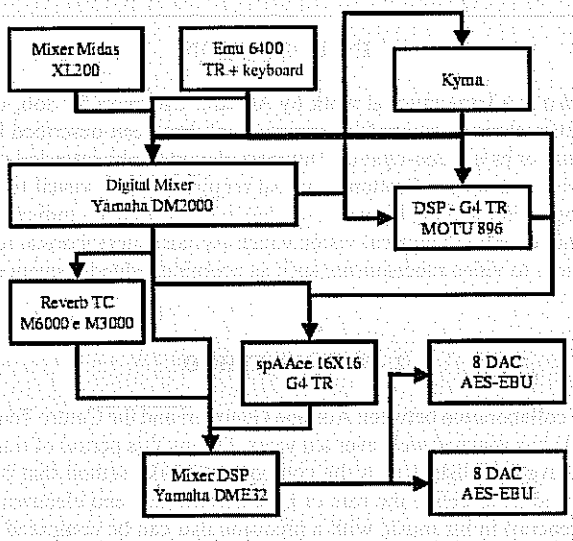


Figure 4: *Medea* - PalaFenice disposition

c. 1 Yamaha DM2000 Mixer (I/O boards: 3 ADAT and 4 AES-EBU; live): devoted to Live Electronics and control

d. 1 Yamaha DME32 Mixer (32 in, 24 out) devoted to global audio I/O management

e. a total of 28 loudspeakers subdivided in 14 audio paths; these audio paths are:

- 8 surrounding groups
- 1 behind the stage
- 1 central
- 2 stereo on the stage
- 2 above the audience

plus two sub-woofer paths driving a total of 4 sub-woofers (2 in front, 2 on the back the audience)

f. a remote control location consisting of: 2 Yamaha 01V, 2 JLCooper MIDI fader boards

g. 2 TC-Electronics reverb units (global reverb, local reverb)

h. 2 G4 machines (one devoted to sound spatialization, the other devoted to digital signal processing)

i. 1 Emu 6400 sampler + mini-keyboard and MIDI faders devoted to sampled choirs and their dynamic control

j. 1 Kyma Capybara 320 Digital Signal Processing workstation devoted to digital signal processing

The G4 machines are both featuring a Max/MSP environment, one devoted to sound spatialization through the *spAAce* program designed by Alvise Vidolin and Andrea Belladonna (cf.[4, 5]) and the other with *ad hoc* digital signal processing programs.

Figure 5: *Medea* - Live Electronics setup (1)

Fig.5 and 6 show the Live Electronics setup on the sound engineering location, while Fig.7 shows the setup of the remote mixing

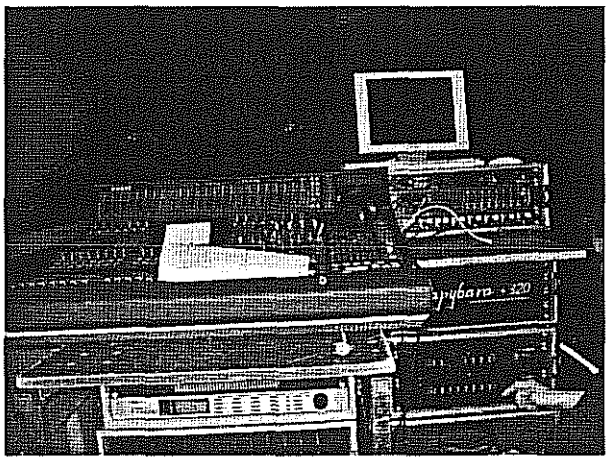


Figure 6: *Medea* - Live Electronics setup (1)

location.

3.3. Sound Spatialization

Besides the transparent sound reinforcement applied to most of the orchestral body of *Medea* (described in Sec.3.1) and some static

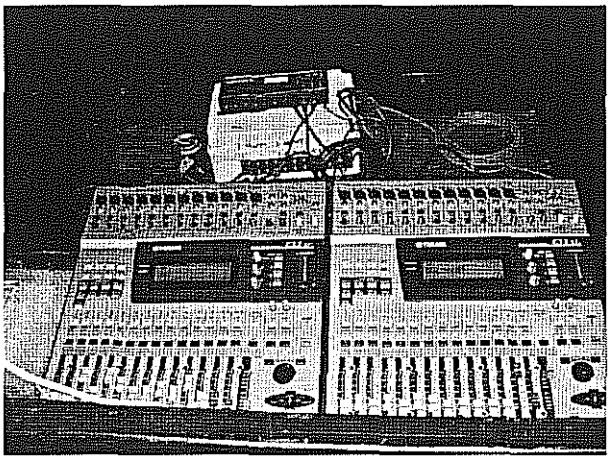


Figure 7: *Medea* - Remote mixing location

sound reinforcement applied to the solo singers, sound spatialization processing plays an important role throughout the work. There are several kind of spatialization patterns applied constantly to some solo instrument or instrumental families, as if specific sound movement patterns were to acquire the function of *leit-motives* and were an integral part of a given instrumental expression.

Different spatial movements are applied to:

the solo contrabass flute: in the first Act, the contrabass flute plays on stage and its sound is simply reinforced; however, in the second and third Act the contrabass flute is physically located behind the public, and the sound is moved back and forth with several different movement modes

trumpets: two trumpet sections are located in the hall to the sides of the public; their sounds are moved across and above the public with positions that are dynamically proportional to the amplitude of the emitted sound (cf.Fig8)

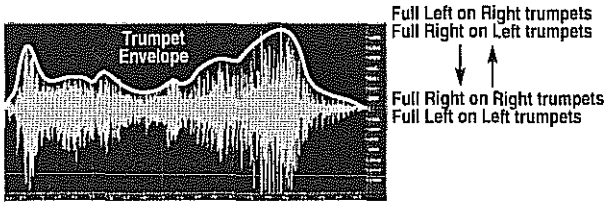


Figure 8: *Medea* - Trumpet Spatialization

trombones: four solo trombones are located among the public; their sounds follow different paths at different times; a new form of "expressive spatialization" has been experimented with them (described in another paper published in the same proceedings, cf.[6])[2]

[2]This experiment is the result of a creative collaboration between the Centro Tempo Reale and the Centro di Sonologia Computazionale (CSC) of the Padova University which has kindly allowed us to use some first-hand experiments produced in the context of the innovative EC-funded project MEGA (IST-1999-20410) on multi-modal expressive interfaces (cf.{\smallhttp://www.megaproject.org}).

choir: the sound of the choir is treated along several different sound spatialization processes:

- transparent sound reinforcement (cf. Sec.3.1)
- surrounding sound reinforcement
- front-only sound movement (termed "celluloid" movement)
- simultaneous clockwise and counter-clockwise circular movements of the different choral voices
- far-near movements

percussion: percussion instruments are treated in different ways:

- timpani follow very rapid and sharp random movements above the public (termed "rain" spatialization)
- cables (which are custom-built instruments made out of multiple-core cables pulled over sharp metal plates and picked up with contact microphones) feature the same spatial processing used for trumpets (cf.Fig.8) but in the front/rear direction
- "celluloid" movement
- Bass drums and tom-toms feature clockwise and counter-clockwise circular movements

3.4. Sound Processing

Other than sound spatialization, timbral processing is a fundamental musical aspect of *Medea*. Given the difficult performance environment (a large hall with a huge stage, many loudspeakers, many microphones and pick-ups, etc.), sound processing had to satisfy strict requirements in term of:

- musical relevance
- stability
- reliability

The palette of timbral processing presented here is just a subset to show how simple, stable and reliable transformations can end up being very significant musical tools. Combined all together, these transformations confer to *Medea* a peculiar soundscape in which instrumental and electronic sounds blend together in a coherent *unicum*.

Fig.9 describes a transformation which allows the soprano and mezzo soloists to produce a high double pedal based on two notes that are sung melodically (a high D - *Re*, and a high B - *Si*).

"Flexatone" flutes are instead described in Fig.10. The result intended by the composer was a subtle morphing between the percussion instrument and the flute. The patch described is one out of several solutions studied in the past. The same patch is also applied to clarinets cascading it out of a one-octave-below transposition of the original sound.

Fig.11 describes the patch to produce "metal" French horns. As their name implies, the metallic (i.e. high-frequency, inharmonic) sound is produced through the patch described.

Something slightly different has been designed to obtain the same effect with trumpets, in order to have a deeper control over amplitudes (cf. Fig.12).

In some passages, Guarnieri had asked for some "thin metal moans", always performed by the trumpet section. Fig.13, together

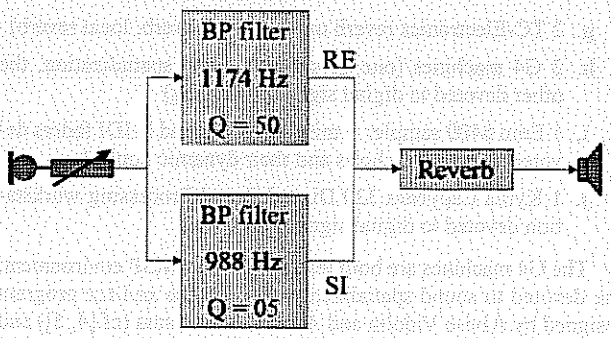


Figure 9: *Medea* - selective reverberation patch

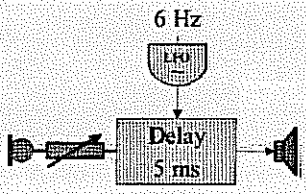


Figure 10: *Medea* - "Flexatone" flute patch

with a peculiar instrumental writing, was the solution found to this end.

In other passages, another requirement was to blend the 4 solo trombones with the choir, as if they were just another voice of the latter. The patch described in Fig.14 was used to add enveloped vibrato to "vocalize" trombones.

4. ACKNOWLEDGEMENTS

The whole production of *Medea* has involved a number of people ranging in the hundreds and its cost has gone well above several hundred thousand Euros. In such a context, the complexity of technological issues is completely overwhelmed by the complexity of the coordination the production itself. This is one of the cases in which some key people really make the difference, whether they occupy key positions or not. This is why the authors would like to warmly acknowledge here the people at BH Audio Services, without whose help the quality of sound (and life during production) would have never been reached, our assistants Nicola Buso and Francesco Canavese who endured hours and hours of rehearsals, Professor Giovanni De Poli (director of CSC, Padova) whose friendship and assistance was so helpful in term of human and technological resources, his graduate student Amalia De Götzen who was responsible for the experimental work on

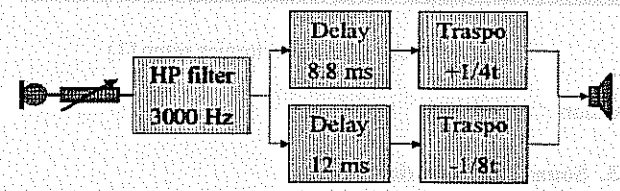


Figure 11: *Medea* - "Metal" Horns patch

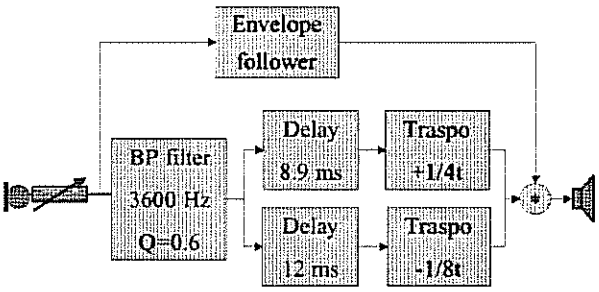


Figure 12: *Medea* - "Metal" Trumpets patch

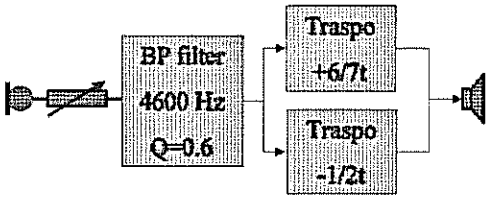


Figure 13: *Medea* - "Thin Metal" Trumpets patch

trombones, Pierangelo Conte (assistant production manager for the Teatro La Fenice) who has solved countless problems of all sorts during rehearsals and performance, Anna Meo (organization director, Centro Tempo Reale) who managed to get us to the Palafenice and back without too many losses, conductors Pietro Borgonovo and Guillaume Tournaire who have patiently sat through many Live Electronics pre-production and production stages, and the technicians of the Teatro La Fenice which have recorded on 24 tracks the complete production filling up a considerable quantity of audio DVDs. Last but not least, our warmest acknowledgement goes to composer Adriano Guarnieri, without whose vision, creativity and work *Medea* would have never seen the light.

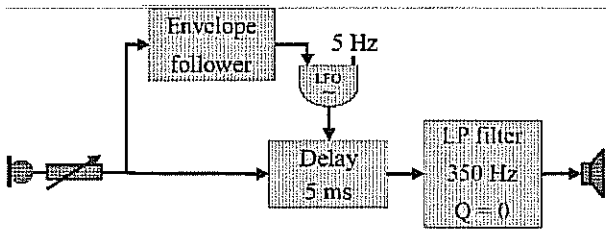


Figure 14: *Medea* - "Vocal" Trombones patch

5. REFERENCES

[1] A. Vidolin, *Ambienti esecutivi*, Musica Verticale - Galzerano Editore, Salerno, 1987.

[2] N. Bernardini and A. Vidolin, "Recording *orfeo cantando... tolse* by Adriano Guarnieri: Sound motion and space parameters on a stereo cd," in *Proceedings of the XII Colloquium in Musical Informatics*, Gorizia, sep 1998, pp. 262–265.

[3] N. Bernardini and A. Vidolin, "The making of *passione secondo matteo* by Adriano Guarnieri: An Outline of Symphonic Live-Electronics," in *Proceedings of XIII Colloquium in Musical Informatics*, L'Aquila, 2000.

[4] Andrea Belladonna and Alvise Vidolin, *Applicazione MAX per la simulazione di sorgenti sonore in movimento con dispositivi musicali a basso costo*, pp. 351–358, Milano, 1993.

[5] A. Belladonna and A. Vidolin, "*spAAce*: un programma di spazializzazione per il live electronics," in *Proc. Second Int. Conf. Acoustics and Musical Research*, Milano, 1995, pp. 113–118.

[6] Amalia De Götzen, "Expressiveness analysis of virtual sound movements and its musical applications," in *Proceedings of XIV Colloquium in Musical Informatics*, Firenze, 2003.

CREATING A DIGITAL ARCHIVE OF ANALOGUE RECORDINGS: TECHNOLOGICAL ASPECTS AND MUSICOLOGICAL IMPLICATIONS

G. De Mezzo and A. Orcalli

Laboratorio MIRAGE – Università di Udine – sede di Gorizia – via A. Diaz 5 – 34170 Gorizia
mirage@cego.uniud.it

ABSTRACT

In the context of the problems related to the archiving and restoration of audio materials, this article presents the techniques and historico-philological criteria developed in the field of preservation, cataloguing, and valorization of contemporary music sound documents by the MIRAGE laboratory of the DAMS course at the University of Udine. The methodological approach adopted is exemplified through the illustration of how the project for the preservation, digitization and cataloguing of the audio archive of the Historical Archives of Contemporary Arts of the Venice Biennale (ASAC – la Biennale di Venezia) was carried out. Among the collections handled by the MIRAGE laboratory, the ASAC collection aptly exemplifies the variety and complexity of the problems inherent in the preservation of audio recordings.

1. INTRODUCTION

The opening up of archives and libraries to a large telecoms community, which has been made available through their integration into the Internet, represents a fundamental impulse for cultural and didactic development. Guaranteeing an easy and ample dissemination of some of the fundamental moments of *the musical culture of our times is an act of democracy* which cannot be renounced and which must be assured to future generations, even through the creation of new instruments for the acquisition, preservation and transmission of information-instruments which must be culturally conceived of in such a way as to avoid subordination to market strategies. This is a crucial point, which is nowadays the core of reflection of the international archive community. If, on one hand, scholars and the general public have begun paying greater attention to the recordings of artistic events, on the other, the systematic preservation and consultation of these documents is complicated by their diversified nature and amount: the data contained in the recordings offers a multitude of information on their artistic and cultural life which cannot be included in traditional bibliographical archiving, which is oriented toward the carriers rather than toward the information they contain.

The preservative re-recording and cataloguing of audio document collections cannot leave out a consideration of the history of the institute or collection in which *they are held. In fact, knowing the documentary* choices and the history and characteristics of the institute that owns them helps define the strategy to adopt during the preservative interventions.

It is well-known that the recording of an event can never be a neutral operation, since the timbre quality and the plastic value of the recorded sound, which are of great importance in contemporary music, are already determined by the choice of the number and arrangement of the microphones used during the recording. In particular, in cases of a non traditional stationing of the orchestra players or of pieces based on improvisation, the positioning of the microphone according to purely documental and presumably «neutral» criteria can be a naïve solution, which in practice *sets serious limits to the identification of the* piece. Moreover, the more sophisticated the interventions of the recording *tonmeister* are, the greater will be the possibility that interpretative elements and manipulations are added to the recording of the event. Thus, musicological and historico-critical competence becomes essential for the individuation and correct cataloguing of the information contained in audio documents.

The commingling of a technical and scientific formation with historico-philological knowledge also becomes essential for preservative re-recording operations, which do not coincide completely with pure digitization, as it is, unfortunately, often thought. In fact, at stake are matters related to the influence that new audio technologies have on piece preservation criteria, on the cultural policy choices made by the institutes, and on the sensitivity of an ever-growing audience who wants an increasingly more direct access to the information.

New perspectives for the treatment of sound materials are opened up by the development of experimentation in the field of multimedia integrated systems (catalographic data/audio signals and/or signals for the automatic management of hybrid archives) and by the evolution of audio technology (24 bit analogue/digital converters, 192 kHz samplers, 17 GB optical media). In this constantly evolving technological scenario, in order to preserve sound documents in a philologically *correct way, it becomes essential to rely, during the re-*recording procedures, on operational protocols aimed at avoiding the overlapping of modernized phonic aspects which alter the original sound content. The delicacy of the process is immediately clear if we think about cases of carrier degradation which run the risk of disintegrating the original. The criteria for the preservation of documents mustn't either be influenced by the market-induced tendency to use a compressed

format (i.e. mp3, WMA, mp3PRO, AAC etc.), thanks to which it is also possible to listen to archive musical pieces and to retrieve them in remote. The low quality of compressed sound, especially if considered in relation to the phonic richness of much contemporary music, imposes the rigorous avoidance of any mixture between the acquisition of documents for conservative aims (preservative copies) and the archiving for common use (access copies). These are, of course, topics which open up new frontiers for on-line research on music documents and on the study of the alterations of auditory sensitivity produced by the new technologies.

Within the project for the global reorganization of its archives, the ASAC of the Venice la Biennale has entrusted the MIRAGE laboratory of the DAMS course at the University of Udine and the CSC of the University of Padua with the planning and creation of the ASAC audio archive.

The project for its preservation and digitization assigned to MIRAGE represents a first step toward an online access to the most prestigious Italian archive of contemporary music. The ASAC patrimony, which is made up of a series of analogue tape sound recordings produced in the span of 40 years, documents the various cultural and artistic initiatives accomplished by this Venetian institute. The enormous amount of these mostly unpublished sound documents, which have an extraordinary musicological and documentary importance, urgently had to be saved from the natural degradation of their analogue carriers.

The ASAC contains material of great interest for the historical reconstruction of the Venetian institute's congresses and administrative life, as well as documents relevant to the history of Italian and international contemporary music. For example, recordings of concerts by the Orchestra della Fenice conducted by F. Donatoni, B. Maderna, Z. Pesko, and K. Stockhausen, by the Ensemble 'die reihe' conducted by F. Cerha, and by the London Philharmonia Orchestra conducted by Giuseppe Sinopoli can be found there. Concerts by Maurizio Pollini and Sviatoslav Richter, a piano recital by S. Strawinsky (music by I. Strawinsky) and performances by the Arditti Quartet have also been recorded there. Furthermore it holds recordings of G. Manzoni's *Modulor*, and Stockhausen's *Hymnen* (conducted technically and artistically by the composer himself), as well as tapes on avant-guard music congresses, such as the 1961 International Congress on Experimental Music, which was attended by the most prestigious centres for music research (i.e. Paris, Köln, Columbia University – New York, Fonologia di Milano, Utrecht, Philips - Eindhoven, Tokyo and Warsaw) and hosted interventions and music by Berio, Pousseur, Nono, Schaeffer and Ussachevsky. There is also good evidence of the experimentation of the Laboratorio di Informatica Musicale (LIMB), coordinated by A. Vidolin at the time.

2. CARRYING OUT OF THE PROJECT

2.1. Preservative Re-recording

The methodology of preservative re-recording is inspired by the ethical criterion of the *historically-faithful reproduction of the document*, which wishes to preserve the sound content of the original recording exactly as it has come down to us.

The indications of the international archive community have been respected in the making of the preservative copy: 1) the re-recording is transferred from the original carrier; 2) if necessary, the carrier is cleaned and restored so as to repair any climactic degradations which may compromise the quality of the signal; 3) re-recording equipment is chosen among the current professional equipment available in order not to introduce further distortions due to the equipment of the time; 4) intentional alterations are compensated for through a correct equalization of the re-recording system and the decoding of any possible intentional noise reduction interventions; 5) unintentional alterations, such as defects introduced by misalignment in the azimuth angle of the recording head, are compensated for.

The project for the digitization of the ASAC has been developed at the MIRAGE laboratory according to the re-recording protocol coherent with this approach. The operational phases of the re-recording are: 1) analysis of the document; 2) optimal signal retrieval from analogue carriers; 3) creation of the preservative copy.

2.2. Analysis of the Document

During this phase the state of the document's preservation is evaluated and the physical characteristics of the carrier and its format are individuated, also on the basis of historical research carried out on the technologies in use at the time of the recording. This knowledge is an indispensable premise for a historically faithful reproduction of the document.

2.2.1. Analysis of the Carrier and Restorative Interventions

The bulk of recordings from the ASAC Audio Archive, which our laboratory has been entrusted with for the re-recording operations and the creation of preservative copies, consists of 90 reels containing analogue recordings on magnetic tape (1/4''), for a total of 100 hours. The reels were selected on the basis of their state of preservation and their musicological and documentary value. The recordings date back to the following years: 1961, 1969-72, 1974-75, 1979, 1982-83.

The magnetic tapes have undergone all the ordinary preservation state checks: winding check, spooling, splice check; where missing, leader tapes have been added at the head and tail (respectively red and green). The cinching, popped strands and pack slip have been documented through digital photographs, which have been enclosed in the preservative copy. If we set aside carriers subject to hydrolysis, the examination of the ASAC documents has not revealed cases of great

losses of magnetic particles, even in the case of the older tapes dating back to 1961.

In order to single out the tapes subject to hydrolysis (sticky shed syndrome), the brands of the tapes and the year of recording have been compared with the data contained in the literature. The degraded tapes have been treated with the desiccation procedure based on oven baking at a thermostatic temperature of 50°C, the heating time is 30^m to avoid thermal stresses, for 1 or more days. These parameters have been chosen in light of the objective of maintaining the information content of the carrier integral and of avoiding the risk of a decay of high frequencies caused by baking at more than 55°C. The measures taken in the course of the thermal treatment performed confirm the amount of water loss following the desiccation procedure (up to 2 grams per reel during the first day of baking).

2.2.2. Analysis of the Recording

The information on the format of the carrier has been inferred from the direct analysis of the tape and then compared with the technical data contained on the case, which was often wrong or missing. The data inferred from the history of audio technology is a source of knowledge which cannot be ignored when defining methods and procedures for the survey of the formats and replay parameters (speed, track format, playback equalization, etc.) adopted during the original recording. Moreover, knowing them allows us to solve specific problems caused by the technical defects of the equipment used for the creation of the document.

Speed: most of the concerts were recorded at 19 cm/s; the congresses and remaining concerts at 9.5 cm/s. Changes in speed have been found within the same tape. Since the information contained on the cases was often unreliable, the speed was additionally checked through the perceptive listening of sample segments taken from the entire recording.

Tracks: the team from the MIRAGE laboratory has created an instrument for the individuation of the number of tracks, which analyses the magnetic structure of the tapes in a non destructive manner by relying upon a thin height-adjustable magnetic head which is connected to a signal monitoring system. This instrument allows for the individuation of the dimension and number of recording tracks with the precision required to retrace the type of heads used for magnetic recording.

Azimuth: signal analyses have been carried out on audio document segments in order to single out and compensate for misalignments in the azimuth angle of the original recording system.

Equalization: in order to determine the correct equalization, we have adopted a methodology based on the following criteria:

a) historico-technical survey of the recording systems of the time;

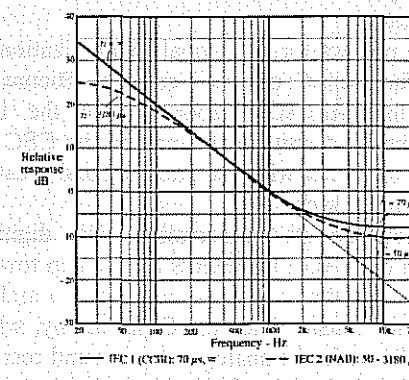
b) analysis of the functionality of the tape-recorders used by la Biennale and still in use at the ASAC;

c) gathering of the witness provided by the sound technicians of the ASAC;

d) frequency analysis of the recorded sound with various equalization curves;

e) perceptive listening of equalized segments with various curves.

The 19 cm/s speed case emblematically illustrates the methodology elaborated by the MIRAGE team. There are, in fact, two different types of equalization, which have been official since the 1960s: NAB (1965; transition time constants: 50-3180 µs); CCIR or IEC 1 (1968; 70 µs). The most striking difference regards the low-frequency attenuation present in the NAB standard but not in the CCIR; this solution aims at attenuating low frequency background noise and in particular at reducing the presence of hums.



European radio broadcasters have adopted the CCIR, while American broadcasters follow the NAB standard. The implementation of the equalization curves in semi-professional recorders on the part of the manufacturers varies from one model to the other and according to the needs of the customer. For example, the A 77 model built by REVOX at the end of the 60s comes with the NAB curve only in recording, while in playback it is also provided with the other equalization curve, so as to guarantee its compatibility to recordings made on preceding models. REVOX's B 77 model can be supplied with either of the curves. More recent professional systems come with both types of equalization.

The technicians of the ASAC audio laboratory carried out the recordings of the tapes examined in our project with REVOX A 77 and B 77 tape-recorders, which were respectively bought in 1969 and 1981 as stated in the ASAC equipment inventory. The two tape-recorders retrieved from the equipment preserved by the ASAC have been tested for functionality. The results confirm the NAB equalization for the A77 model, while the label on the back side of the B 77 and the linearity check carried out on the tapes witness the implementation of the CCIR standard.

The information inferred from the historico-technological background and from the history of the Venetian laboratory has been compared with the indications resulting from the direct survey of the sound content: signal analysis and comparative perceptive listening (see points d, e). The tests have been carried out on a tape-recorder (Studer A812) which offers the possibility of passing from one type of equalization to the other, thereby guaranteeing identical conditions during the testing.

Signal analyses carried out on the same tape segment with the two types of equalization confirm notable

differences in low-frequency. The presence of hums is important for the purposes of our research: the study of its frequency content (decaying, etc.) is therefore another useful element for the determining of equalization.
The same signal samples have been evaluated through perceptive listening based on the following aspects: presence of broadband noise, hiss, high-frequency emphasis, low-frequency emphasis, overall sonority balance, naturalness of the instrumental and vocal timbre.

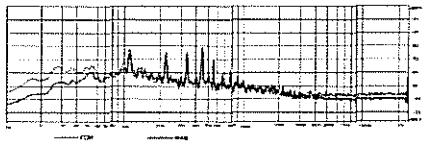


Figure 2. exemplifies the analysis (FFT) of a segment taken from the beginning of the recording of Mahler's *Symphony n. 9* conducted by G. Sinopoli and performed at the 1983 Festival. We can observe that the NAB equalization alters the normal low-frequency noise levels (hums). The effects of these differences can be appreciated during the perceptive listening. The orchestral sound balance is greatly altered by the NAB owing to low and high frequency reduction: the loss of brilliance is followed by the disappearance of the intensity of cello and double bass *pizzicato*, which are necessary to sustain the melody they express at the opening of the symphony's first movement 'Andante comodo'. Instead, with the CCIR equalization, the greater naturalness of the string timbre gives sound the weight necessary to make the orchestral passage fluid. Since the ASAC laboratory owned a CCIR recorder in 1981, the analyses and perceptive listening confirm the hypothesis that the recording was indeed carried out with the REVOX B77 and that the correct equalization is thus the CCIR.
The survey carried out on the tapes produced by the ASAC audio laboratory has given homogeneous results for the comparison between points a-b and d-e of the above-stated criteria.
Intentional alterations: in order to single out the presence of noise-reduction systems (dbx, Dolby), a methodology similar to equalization has been adopted. In the recordings examined, no noise reduction system was found.
Signal: the ASAC collection preserves monophonic and stereophonic recordings. The analysis of the stereophonic signal (correlation and dynamic difference among the channels) has excluded the presence of cases of voluntary signal splitting (pseudo-stereophony).

2.3. Optimal signal retrieval from analogue carriers
On the basis of the information gathered in the first phase, the playback analogue equipment is chosen and the standards for signal digitization are defined.
2.3.1. Transfer technologies
In the methodology adopted by the MIRAGE laboratory, the playback equipment is chosen among the current professional technology available, so as to avoid introducing further distortions and to collect more information than that offered by the equipment of the time. The technico-functional analysis carried out on the ASAC tape-recorders confirms the importance of this choice and reveals their inadequacy: low signal-to-noise-ratio and the unreliability of the tape transport system in guaranteeing the physical integrity of the original document. Besides assuring standards apt for archiving needs, the equipment used by the MIRAGE laboratory is compatible with the historical formats of the documents.
2.3.2. Information Transfer
The transfer from the old to the new format has been carried out without subjective alterations or 'improvements' such as de-noising, etc., in line with the observation that the unintended and undesirable artefacts (noise, clicks, distortions) are also part of the sound document, even if they have been subsequently added to the original signal by mishandling or poor storage. Both have to be preserved with the utmost accuracy.
2.3.3. A/D Conversion
The technological conversion from the analogue to the digital domain is a delicate aspect of the re-recording procedure. Since original carriers may contain secondary information (i. e. bias frequency) which falls outside the frequency range of the primary information (signal) and which may assist in correcting inaccuracies in the original recording, the transfer must be carried out to the highest standard possible. The choice must also assure the compatibility with future standards (192 kHz). In agreement with the directors of the Venetian institute, the audio signal has been acquired at 48 kHz/24 bit and temporarily saved on Hard Disk as an AIFF format file.
2.3.4. Re-recording System
The system has been set up on the basis of the following needs: the saving of the digital data in an audio workstation apt for the operational creation of the preservative copy; the creation of a DTRS digital backup copy; and the alignment level between analogue and digital equipment: 0 VU Meter (+4 dBu) = -14 dBFS. Great care has been dedicated to checking the signal flux and to guaranteeing sound monitoring before leaving the workstation. In fact, every sound document presents original technical aspects. It is precisely because of this «instability» inherent in the document that it is impossible to carry out automatic re-recordings with the simultaneous use of several systems.
2.3.5. Signal Quality Notes
Types of distortions found:
-local noise: clicks, pops, signal dropout due to joints (rare);
-global noise: hums, homogeneous and irrelevant background noise;
-distortions produced during the sound recording phase: electrical noises (clicks, ripples), microphone distortions, blows on the microphone, induction noise.
2.3.6. Descriptive File
The re-recording file contains all the elements gathered in the first two operational phases. The technical

information contained on the cases and the information inferred from the direct examination of the document have been transcribed in different fields. The file consists of five parts:
1) general information on the document;
2) information on the state of preservation of the document;
3) restoration work on the carrier;
4) information on the format and replay parameters;
5) information on the re-recording;
The essential information goes into the general file that comes with the preservative copy.

2.4. Creation of the Preservative Copy

The recordings document: a) Contemporary Music Festival concerts (1969-1983), which couple the performance of historical 20th century pieces with pieces by contemporary authors, usually performed for the very first time, and with concerts dedicated to both traditional and experimental music; b) debate conferences held in occasion of the festivals; and c) avant-guard music concerts (1970). The specificity of the ASAC la Biennale recordings thus lies in the great space they reserve to a contemporary music repertoire, which only few other institutes do (i.e., the Donaueschinger Festival or the Darmstadt Ferienkursen).

2.4.1. General Criteria for the Individuation and Subdivision of the recorded events

The individuation of various pieces of the original document, which belong to such a scarcely-known repertoire, has required documentary research on the sources of contemporary music. The often incomplete documentation contained in the original carrier and the ephemeral and, at times, extemporaneous nature of the event (during Festival concerts there are often variations to the programme as regards to the performing order of the pieces, last minute exclusions and substitutions of pieces) have imposed a systematic control of the content. Given the archive's conspicuous dimension, the possibility that the magnetic tape was wound on the reel of another event or the exchange of unlabelled reels from their cases cannot be excluded among the causes of note-content difference. The individuation task is complicated by the scarcity and difficult retrieval of direct sources (scores or sound recordings) of contemporary repertoires. In some cases, it was possible to individuate a piece thanks to the information inferred from the notes contained in the programme or from publishing house catalogues (instrumentation, length, etc.).

2.4.2. Content Individuation

The individuation of the content is based on the following criteria:
a) information gathered from the case of the original document;
b) information inferred from other documents: the *Annuario* of ASAC events and the programmes of the events and festivals;
c) musicological analysis: comparison with other sound sources and/or scores, information inferred from the programmes, publishing house catalogues, and online resources, etc.

2.4.3. Material Subdivision Criteria

The operations for the subdivision of the material have respected the continuity of the recording of the events and the formal unity of the musical pieces in relation to the capacity of the new carriers. The time intervals corresponding to pauses void of informative content have been eliminated. In the instrumental repertoire the dimensions of the new carriers have allowed for the insertion of the entire piece in most cases. Extremely long pieces (i.e. Mahler's *Symphony n. 9*) have been archived on several carriers by exploiting their division into various movements, thereby safeguarding the continuity of the performance. For example, in the case of the recording of Stockhausen's *Hymnen*, performed at the 1971 Festival in a version for electronic and concrete music with soloists and orchestra, the first two of the four Regions in which the piece consists were performed and recorded without interruptions for a total duration of about 51m and thus beyond the capacity of the new carrier. The third Region, which lasts about 39m, was delimited by a recording pause, while the fourth Region was recorded on a new tape. In the CD version published by Stockhausen-Verlag, the first two Regions are presented without pause, while in the version for electronic and concrete sounds published in 1969 by Deutsche-Grammophon LP the first and second Region were subdivided and presented on the two sides of the disc. The solution adopted by this well-known publishing house has been taken as an example for the subdivision of the Venice performance.

2.4.4. Cataloguing

In conformity with the filing solutions proposed by the ASAC archive directors of the Venice Biennale, two types of files have been created.
The general descriptive file contains the preservation metadata about the recording: the characteristics of the original carrier (its format and state of preservation), the description of the content found on the case (title of the document), the origin and collocation, notes on the title-content correspondence, replay equipment and parameters, the digital resolution, format and all equipment used, and the laboratory involved in the process. The metadata will be a key component in the preservation and management of any digital collection and must be designed to support future preservation strategies.
The track file is entirely dedicated to the musicological description of every single piece (author, title of the work, performers, date and place of performance, length, and sound channels).

2.4.5. Archive Carrier

Although the R-DAT and CD-R (audio) are broad diffusion digital recording system, neither of these systems has a proven record of archival stability. In particular, the audio CD-R is not compatible with the principles that inspire preservative re-recording for low-resolution and data security (redundancy). Among the digital archiving systems (CD-Rom, DVD-R, DLT, AIT-3), a CD-Rom which can accommodate AIFF files

up to about 40 minutes long (stereo) at a 48 kHz/24 bit resolution has been chosen.

2.4.6. Creation of the archive copy

Bearing in mind the principle according to which the audio heritage is made up of primary information – the signal – and secondary or ancillary information – such as handwritten notes, etc. – we have tried to maintain a close continuity with the original carrier and the associated material when creating the CD-Rom copy. For this reason, the ASAC collocation of the original document has been maintained in the identification code of the CD-Rom, with the addition of an indication of the number N of new carriers into which the content of the reel has been divided (i.e., ASAC collocation: A 5600024; new collocation A 5600024 CD n/N). Moreover, the general index-card containing preservation metadata about stored recording, photographs of the carriers and cases, archiving modalities and notes have also been inserted in every CD. A track index-card containing musicological information on the piece has been attached to every audio file. A document containing details on the AIFF format has also been inserted.

3. CONCLUSIONS

The security of sound documents, the restoration and re-recording of their audio content from analogue to digital carriers, the cataloguing of the technical and artistic contents of the preservative copy, and finally the creation of rapid and satisfactory information retrieval and processing systems require an interdisciplinary coordination of numerous abilities: from archiving, music history and the history of musical technology to audio signal processing and carrier chemistry. In the specific field of audio restoration, the formation of specialized personnel is even more necessary, since the removal of impulsive noise and the reduction of broad-band noise is nowadays carried out with noise reduction algorithms which, however refined they may be, inevitably interfere with the quality of the signal.

The influence of the development of means of audio re-recording on taste and aesthetic sensibility affects the entire realm of art and can also be seen in the debate on preservation of audio documents. Although it has its own specificity, audio restoration fits into the more general question which is common to all preservative interventions in the fields of art and which can be summarized in the interrogative already poised by art historians: whether restoration constitutes a moment of pure preservation of the works handed down to us or whether it should rather be oriented toward their adaptation to the new uses and tastes of an ever-growing public, to different cultural policies and to new technological scenarios. The methodological approach adopted by the MIRAGE university laboratory aims precisely at laying solid foundations for a higher professional formation, which will be adjusted to the values already achieved by the Italian 'school' in many other areas of documentation and restoration.

REFERENCES

[1] Bradley, K., «Anomalies in the Treatment of Hydrolysed Tapes: Including Non-Chemical Methods of Determining the Decay of Signals», Technology and Our Audio-Visual Heritage, editor George Boston, pp. 70-83, 1999.

[2] Calas, M.F. and Fontaine, I.M., La conservation des documents sonores, CNRS Editions, Paris, 1998.

[3] Camras, M., Magnetic Recording Handbook, van Nostrand Reinhold, New York, 1988.

[4] Canazza, S., De Mezzo, G., Michelini, G. and Orcalli, A., «Preservation and philological Restoration of Audio Documents by Bruno Maderna», Proc. XIII CIM Colloquium on Musical Informatics, L'Aquila, pp.127-130, 2000.

[5] Canazza, S., De Mezzo, G. and Orcalli, A., «Conservazione e restauro dei documenti sonori al Laboratorio MIRAGE», Suoni in corso – Percezione ed espressione dell'uomo tecnologico, a cura di de Incontrera. C., MittelFest editor, Cividale del Friuli, pp. 347-358, 2002.

[6] Conti, A., Sul restauro, Einaudi, Torino, 1988.

[7] Dorigo, W., «La Biennale di Venezia. Archivio storico delle arti contemporanee. Dall'automazione una risposta adeguata alle necessità di informazione», Informatica e Documentazione, anno 3, n° 2, pp. 116-123, 1976.

[8] Dorigo, W., Caproni, M., Piantoni, M., Zamattio, G., Agosti, M. and Talpo, L., L'automazione dell'Archivio storico delle arti contemporanee, ASAC/Contributi, La Biennale di Venezia, 1979.

[9] Giuliani, R., «Le fonti sonore e audiovisive e la storiografia contemporanea», Rivista Italiana di Musicologia, vol. XXXV, NN. 1-2, pp. 540-584, 2000.

[10] Guide to the Basic Technical Equipment Required by Audio, Film and Television Archives, edited by Boston, G., Paris, 1991.

[11] IASA-TC 03, The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, Version 2, 2001.

[12] Orcalli, A., «On the Methodologies of Audio Restoration», J. of New Music Research, vol. 30, n° 4, pp. 307-322, 2001.

[13] Schüller, D., «Informazioni audio e video. Dalla preservazione dei supporti fisici alla preservazione delle informazioni», Gregory T. and Morelli M., L' eclisse delle memorie, Laterza, Roma-Bari, pp. 21-32, 1994.

[14] Schüller, D., «The Ethics of Preservation, Restoration, and Re-Issuses of Historical Sound Recordings», J. of Audio Eng. Soc., vol. 39, n° 12, pp. 1014-1016, 1991.

[15] Vidolin, A., «La conservazione e il restauro dei beni musicali elettronici«, Le fonti musicali in Italia. Studi e ricerche, 6, pp. 151-168, 1992.

SOUND TRANSFORMATIONS: PAST AND FUTURE.

Daniel Teruggi, Yann Geslin

Groupe de Recherches Musicales
Institut National de l'Audiovisuel - Paris
www.ina.fr/grm

ABSTRACT

History has often opposed *Musique Concrète* and *Electronic Music* as an initial antagonism that was subsequently resolved with the appearance of *Electroacoustic Music*. *Musique Concrète*, an intuitive sound object oriented compositional approach, very quickly found its trends in the media editing technologies (tape splicing in the Studios) and with recorded sound manipulations. Since the first adventures in 1948, continuous evolution has been pursued by hundreds of composers and experimenters, which have renewed the techniques and esthetical approaches. Two concepts have emerged, that are essential to the understanding of our history, our present and our future: *Acousmatics* and *Sound Processing*. Much has been said to explain the first and much to justify the second as an esthetical approach to sound. It seems important to explain why these trends are not only still active but how they have influenced the perception and the conception of Music.

1 Introduction

1.1 Some historical facts

In the beginning of the experiences concerning *Musique Concrète*, the initial intuition on which Pierre Schaeffer worked was that isolated recorded sounds, which were called *sound objects*, could be combined with the same easiness as instrumental sounds did. This would ensure making music quite a similar activity to traditional composition, but using different and continuously renewed sets of sounds [1].

This new conception of composition seemed to have its own rules that emerged very quickly after the first experiences, the strongest fact was that some sounds were completely inappropriate for musical use, while some others were very efficient and easily combined and manipulated. This evidence led Schaeffer to investigate what was there in a sound that could guarantee its "musicality" and to also investigate how our perception worked in order to accept certain events or situations as being musical and how it rejected other sounds, mainly those which had a very strong anecdotal meaning. This investigation gave birth to musical research, which dealt with several technological and psychological disciplines as well as musicology.

In this original new approach, composers were dealing with recorded sounds, which were not referenced in our perception as musical sounds. There was no visual aid that could help identify the sound' s origin. It depended in the capacity of the listener to associate an effective image to a sound, so if sounds were anecdotal, the listening tendency would be to listen to a story instead of listening to music.

This complicated the listening task, since the origin of the sound as well as its semantics had to be determined through listening. Our everyday experience makes us always perceive a sound as a double information source in which we identify the cause or origin of the sound, and we follow the action of the sound through time. In other words, we perceive what the sound is and what it is doing or meaning. So music may come either from already identified sounds like instrumental sounds (which are only used for music, thus reserved for this use), either by sets of sounds not easily associated with known sources so our perception will not loose too much time in source recognition. In the first case, we pay little attention to the origin, since we know the sound; in the later, our perception does a double action of source identification and sense interpretation.

Another mayor trend was that in order for a group of sounds to be considered as a unit and to be perceived together as being musical, there had to exist some kind of homogeneity. Homogeneity was achieved through the use of related sounds (same spectral organisation, similar envelopes, common spectral regions) or by introducing different sounds in a common environment as a reverberation chamber, the reverberation would then function as glue that would permit an efficient assembly of antagonistic sounds. In fact the old rules of acoustics and instrumentation still were efficient when they came to bring sounds together in order to create sense.

1.2 First conclusions

After the first experiences (*Études de bruits*, in 1948) [2] in which a very large variety of sounds were used, Schaeffer' s works, with the addition of Pierre Henry, used a limited set of sounds in order to enhance the musical coherence; i.e. the *Symphonie pour un homme seul*, 1951, mainly uses voice and prepared piano sounds. From there on, the tendency was to use few

starting sounds that were manipulated through adapted tools in order to obtain related sounds which were variations of the initial sources. Investigations were strongly oriented so as to understand the principles that assured coherence and possible combination of sounds with a musical intention. In fact, the main question was: what are the conditions that permit or assure that a combination of sounds is perceived as a structure and not as a set of isolated items? Further on, the question would be: what makes a combination of sounds be musical.

In the mean time, a new set of concepts and words was slowly being developed in order to permit a common vocabulary to describe sounds from this new point of view, between science and musical practice. This concept definition was mainly done through sound experience; this is listening and understanding the functioning of perception. Through a period of 15 years this word led to a major book; the *Traité des Objets Musicaux* [3], published in 1966, in which Schaeffer exposes his major ideas on music and perception and proposes an efficient and complete set of terms for a sound classification based on their *typology* (what kind of sound it is, to which category they belong) and their *morphology* (what is their evolution through time).

The first musical conclusions were that it was efficient to work with sounds that had a low anecdotal reference. Since modifying sounds was the principal method to obtain large homogeneous sets of sounds, it was interesting to work with complex sounds, this is, sounds with rich spectra and complex internal activity that would easily tolerate modifications without the source being recognised. Some implicit rules began to circulate during the first years of *Musique Concrète*. Sounds should not be anecdotal, and complex sounds seemed well adapted for composition. A very small set of sounds should be used, the ideal would be to use one only sound as an origin because you could then follow, study and understand the mechanisms of perception and the proceedings of the composers. These implicit rules or recommendations, would later be abolished through musical experience and through works that proved that music could be done and made good, even if these rules were not followed (one of the major examples is the work *Hétérozygote* composed by Luc Ferrari in 1964 which freely uses anecdotal sounds) [4].

2 Trends and techniques

The main trends that initiated the «concrète« adventure were: no visual references to sounds, and using sound as a material that could be modified in order to fit to musical intentions.

In order to attain both, tools were necessary; and tools appeared at the very beginning of the *Musique Concrète*. Either existing tools, as microphones, turntables and tape-recorders as well as adapted and completely new tools for exploring, manipulating and modifying sounds, as the *Phonogène* or the *Morphophone* [5]. These were the instruments that permitted composers to explore sound and look for unexpected sound results that would always keep a family kind of relation to an original reference sound. During the first years, it was very usual to name the musical works "studies", often composed from a unique sound or exploring the behaviour of a sound.

The use of sounds didn't exclude synthesis or any other kind of sound; the surrounding world was the largest virtual orchestra a composer could find, in which he could pick any sound and make it become music. The limitations were in the method and not in the sources. *Musique Concrète* was mainly experimental; decisions were made after listening and not through pre-conceived schemes; these could exist but needed validation through listening (*Musique Concrète* ideas were radically opposed to the serial approach to composition).

This ensemble of implicit and sometimes explicit rules, set the framework for future developments. Focus was given on sounds and their intrinsic capacity to provide musical material through their processing. As said before, after a first period in which certain sounds were favoured, composers addressed the general domain of sound. It was finally up to them to find the compositional solutions that would permit the good integration of any sound in an esthetical context. However, this open universe of possible sounds was used with great caution, works were never chaotic in their sources (as sometimes caricatured) and the great tendency was (and still is) to have a reduced typology of sounds and a large number of morphological variations. Even in the case of a large set of very different sounds, the homogeneity was obtained through morphology; different sounds doing similar things or having similar profiles.

2.1 What is a sound

This essential question to all listening experiences is the definition of the initial object we deal with. Traditionally, music was made with instruments, which produce sounds that are rather limited in their duration and seem to have a beginning and an end. They can be represented through notes on a score. The sound of music is the combination of those unitary items either in an organised sequence or in a timbral combination; thus, there is a strong tendency to associate in our imagination notes with objects [6].

The same approach was used with *Musique Concrète*: rather short, formally clearly designed objects were used in the first years, and their abstract manipulation (kind of scores that describe their combination) or their definition, were very object oriented. Sound objects replaced notes and composing was conceived as their combination. This conception of sounds in music was also quickly modified, since the sound production was not depending on physical actions on matter, sounds could be extremely long and

homogeneous or continuously changing through time. So the concept of *sound object* lost importance as a method and the limits as to which sounds were best adapted for music, slowly fainted through time and experience.

Sound manipulation in the first 20 years was done mainly through mechanical devices (modified tape-recorders) which permitted a large variety of actions on sound, but with noticeable deterioration of its quality when it was processed several times [7]. The process in itself introduced a loss of quality regarding the original sound and works became very interesting musically, through numerous modifications, but rather opaque acoustically. This was a technology depending problem whose solution came through digital processing.

2.2 Sound richness

A new concept emerged during this period (mainly between 1948 and 1968) and it was the *richness* of sounds. There was no clear definition as to what rich means, the closest would be to say that a sound is rich when it continuously attracts our perception, that is, when it is capable of renewing our listening interest. Richness is somewhere between the redundant simplicity of a sinus sound and the extreme complexity of white noise. Many recorded sounds are close to white noise in their spectra but describing complex variation patterns. But there is more than spectral complexity to the concept. Its origins may be found in Schaeffer's idea that sounds should be *alive*. This liveliness described some ideal sound, which would not be to static or to active, rather constant but with continuous small variations, he called these sounds *convenient objects* [3]. His models in fact were the acoustical instruments that produce homogeneous but always slightly changing sounds. He opposed this definition of liveliness to the static and predictable sounds that in his days were produced by computers (he made a very critic intervention against computers during the International Conference on Music and Computers that was held in Stockholm in 1970) [8].

François Bayle gave a new view to the sound material by specifying that sounds used in acousmatic music have an "outstanding" or salient character that attracts our perception and create a strong impression in our memory [9].

So sounds may be rich, alive and produce strong impressions in our memory. They are rich because they propose complex but not extremely disorganised spectral patterns. They are alive because their behaviour recalls our everyday experience of listening. They produce strong impressions because they behave chaotically through abrupt changes due to physical constraints but within a limited frame of variation.

Richness is a combination of these three characters and it is very often found in the recording of acoustical phenomena. Recording techniques as well as adapted processing tools permit to transmit this richness to processed sounds.

3 Sound processing

3.1 The evolution of processing technologies

The main technical problem during the first twenty years of *musique concrète* was that recording was meant to capture complex variations of spectra and dynamics, but when processed through successive mechanical means, the complexity patterns disappeared to make them become dull and redundant. The quest for rich and interesting sounds led to the improvement of recording techniques; the physical distance at which the sound was recorded determined the amount of complexity or richness that would be included (a good example of this approach can be found in Pierre Henry's *Well Tempered Microphone*). The kind of microphone or the combination of several ones permitted to obtain completely different results.

At the end of the sixties, synthesis was introduced as a technique to produce rich sounds. Synthesis understood not as a process to build realistic sounds but as a method to develop very complex patterns through additive processes and intermodulations. In fact, the objective was to simulate the behaviour of acoustic sounds through synthetic devices. In order to obtain this a large set of interconnected oscillators generated a complex signal that was then controlled as a whole. Acoustic sounds as well as synthetic sounds were used and have produced remarkable works as *De Natura Sonorum* by Bernard Parmegiani, *L'expérience acoustique* by François Bayle or *Signal sur bruit* by Guy Reibel [10]

At the end of the seventies, sound processing through digital devices initiated its long and still lasting evolution. The first approach was to reproduce already existing analog techniques in the digital domain; but little by little new possibilities appeared that introduced completely new concepts and permitted a better organised and systematic work on sounds. Processing actions could be easily reproduced, improved and applied to different sounds. Even if the sound quality was not as good and realistic as that obtained today, it was far better and efficient that any existing technology and was quickly adopted by composers [11].

From non real-time mainframe systems to real-time processing laptops, a continuous evolution in tools and concepts has brought composition through sound processing to its highest summits. Sounds are recorded in relation to kind of processing the composer will realise, which may imply recording several times the same sound in order to enhance certain characteristics.

3.2 The essence of sound processing

In this long and technological changing evolution, there has remained an essential concept: there is a potential amount of information that can be extracted from a sound, which can be effectively used to make music. Music done this way may be either solely conceived for listening attitudes with no visual device as it happens in its acousmatic situation, either associated with images or instruments. Any kind of sound may be used; it is the composer's free choice to create the listening frame in which his music will develop. In the particular case of acousmatic music, this is largely influenced by phenomenological thought, such as how our brain analyses and creates sense to incoming sounds information and organises its complexity into sense [12].

Processing sound is an esthetical choice based on perception processes. Through similarity, it produces homogeneity, which is a very effectual framework for sense construction. One of its essential characteristics is that it transmits morphological trends to other sounds; psychologically, this ensures the transmission of spectral or morphological characteristics to derived sounds. From a compositional point of view it is a very effective way to organise sound material in order to develop evolutions and modulations in music [13].

3.3 Methodology

In opposition to the initial experiences in *Musique Concrète*, in which isolated sounds were put together in order to construct, with a certain difficulty, musical sense, there exists today a methodology issued from practice that has been widely adopted. Even if differences can be found between composers, through studying the methods and listening to results a common way of proceeding can be summarised:

1) *Capturing, mainly through recording, the main sounds that will be used in the process.* This may range from recording, to sampling existing recordings or using synthetic sources. It normally needs a previous conceptual preparation stage in which the general lines of the project are designed.
2) *Processing the sounds.* This is done in order to obtain a large set of related sounds to initial sources. A large array of tools exist which permit to develop very complex sound processing actions.
3) *Listening and editing*, this process, previous to organising the sounds together, permits a good knowledge of the sounds and the selection of the more interesting fragments. Sounds have to be listened a large number of times in order to create abstract images in our mind.
4) *Organising process or mixing.* Here is where the final music is put together, sounds are here often reprocessed and adapted in order to adjust them with other sounds and to develop the intentions of the composer. During this last process, new sounds may be required, which leads sometimes to recording, processing and editing new material.

This is a summarised description of the possible steps to obtain a musical work, the method in itself is largely inspired in the way time dependent media are made (radio and cinema), and it is a consequence of the fact that sound combinations have to be verified through experience [10].

4 Towards the future

Technology and concepts are today largely available in the field. Composing can be done by anyone interested in sound manipulation, and technology is therefore accessible to everybody. The major problem remains always as to *what* to compose and *how*. There are numerous schools of thought and ways of proceeding which may range from general concepts as: *acousmatic music*, *live-electronics*, *multimedia music*, etc. but within each current one may find several approaches based on conceptual definitions.

The field is open to any intention, from popular to serious; and there are large audiences, which follow and identify with different currents. A certain number of techniques are well established and effective tools exist in order to assure an efficient application. Nevertheless there are still issues in which evolution and innovation are possible and necessary.

Processing precision is one; even if the quality of the obtained sounds is remarkable, the old inconvenient through which the tools damage the sound richness is still there. In order to improve results new algorithms are always needed as well as a improved sound definition.

Formalisation is another issue. After 55 years of continuous innovation and improvement much has still to be done to understand the initial questions of what makes sense and what makes music. Semiotics and Phenomenology are powerful understanding tools but no systematic work applied on electroacoustic music has been produced in these domains.

Musical analysis is one of the keys for a major comprehension of music. The electroacoustic domain has remained isolated in its understanding; musicology has had difficulties to study non-written music; and only now appeared efficient tools to assist the analyst [14]. However, work remains still to be done.

Finally, there is still a strong resistance from the musical world concerning electroacoustics and particularly on sound processing. It seems quite an awkward way of thinking nowadays to prescribe or proscribe methods or sound approaches; but we are currently collated with this kind of thinking. Electroacoustics and its more sound-processing form, Acousmatics, are not exclusive domains. These domains are highly embedded in experimentation and open to all sounds, methods and thoughts in order to create, quoting Pierre Schaeffer, "the most general kind of music".

5 Conclusion

Processing sound has frequently been regarded as a rudimentary method of music composition. Either because originally thought as to deal only with noises, either through its experimental approach to composition; it has often puzzled musicians and even brought severe critics as to the «professionalism» of composers. Even if there are deep differences with instrumental composition, the final objective is the same: to construct an organised set of sounds which produces sense for our perception, and music for our mind. Audiences have always been strongly attracted by electroacoustic music, and today concerts bring a large continuously renewed public eager to penetrate to music mainly through listening. And above all, to include the total sound medium, is an expansion of musical thought.

6 References

[1] Schaeffer, P. 1952. «A la recherche d'une musique concrète». Seuil, Paris 1952, 1998.

[2] Schaeffer, P. 1999 "L'œuvre musicale", Ina-GRM CD 1006/7/8, Paris.

[3] Schaeffer, P. 1966. »Traité des objets musicaux, essai interdiscplines». Seuil, Paris.

[4] Luc Ferrari, "Hétérozygote"

[5] Poullin, J. 1954. «L'apport des techniques d'enregistrement dans la fabrication de matières et formes musicales nouvelles. Applications à la musique concrète». In L'Onde Électrique, 34 (324). Société des Radioélectriciens, Éditions Chiron, Paris, pp. 282-291 – see also : «The application of recording techniques to the production of new musical materials and forms. Applications to «musique concrète». National Research Council of Canada - Technical translation TT-646 (D.A. Sinclair), 1957, Otawa.

[6] Schaeffer, P. 1967. »Solfège de l'objet sonore», Text 174pp. + 3CD (english, français, español). Seuil, Paris 1967, Ina, Paris 1998.

[7] Molles, A. ca 1960 «Les Musiques Expérimentales». Trad. Daniel Charles. Cercle d'Art contemporain, Suisse.

[8] Schaeffer, P. 1971, "La musique et les ordinateurs" La revue musicale, Paris.

[9] Bayle, F. 1993, "Musique acousmatique, propositions... ...positions". Buchet/Chastel, Paris.

[10] Teruggi, D. 1994. «The Technical Developments in INA-GRM and their Influences on Musical Composition». Neue Musiktechnologie II. Mainz: Schott 1996, pp. 42-48.

[11] Geslin, Y. 2002. «Digital Sound and music transformation environments: A twenty-year experiment at the Groupe de Recherches Musicales». In Journal of New Music Research, special issue : Musical Implications of Figital Audio Effects vol. 31-2. pp. 99-107. Swets & Zeitlinger, Netherland. ISSN 0929-8215.

[12] Teruggi, D. 1999, "L'interactivité dans le processus de création musicale", in "Interfaces homme-machine et création musicale", Hermes, Paris.

[13] Mailliard, B. 1986. «A la recherche du studio musical». Recherche Musicale au GRM, La Revue Musicale, Paris, pp. 51-63.

[14] Vinet, H., Koechlin, O. 1991. «The Acousmographe, a Macintosh software for the graphical representation of sounds.» ICMC 1991 Proceedings, Montreal.

Sound and architecture: an electronic music installation at the new auditorium in Rome

Francesco Giomi, Damiano Meacci and Kilian Schwoon

Centro Tempo Reale
Firenze
[fg, dm, kilian]@centrotemporeale.it

ABSTRACT

The paper describes the design and the realization of a big sound installation produced by Tempo Reale for the inauguration of Renzo Piano's new auditorium in Rome. Such work required a close collaboration between musicians and architects. The authors describe the general principles of the work, continuing with the criteria adopted to interpret electronic music in space. The general features of the technical system are also discussed.

1. INTRODUCTION

At the beginning of 2002, the Santa Cecilia National Academy asked Tempo Reale to design and realize a big sound installation – conceived by Luciano Berio and dedicated to electronic music – for several locations inside Renzo Piano's new auditorium in Rome.

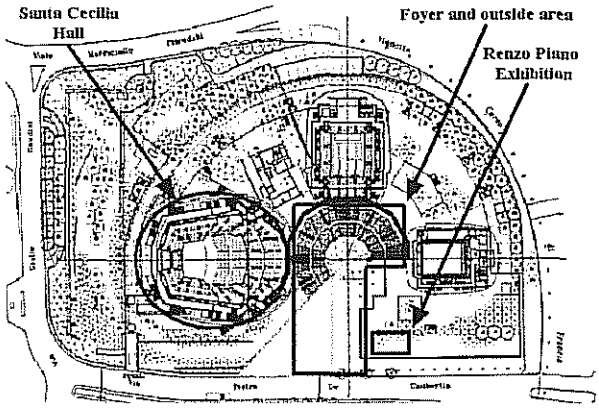


Fig. 1: Map of the auditorium complex; areas involved in the installation are shown in bold

The installation was realized during the two days of inauguration of the auditorium, which took place in April 2002. It involved four different areas, inside and outside the buildings (see fig. 1): the big semicircular foyer, the outside area around the cavea, the Renzo Piano exhibition area and the big "Santa Cecilia" concert hall, still under construction at that time and inaugurated separately in December 2002.

One of the most important aims of the installation was the "discovery" – through music – of the different original spaces of the new architectural structure. This was made possible by the collaboration of the musicians with the group of architects of the Renzo Piano Building Workshop who were in charge of the visual and structural design of auditorium. This collaboration was particularly important because all buildings were still under construction during the period of design and studio preparation of the installation, and it was not possible to have a realistic idea of the architectural scene. This was problematic for example for the placing of the loudspeakers. Moreover, many choices about the visual design of the sound installation had to fit several aesthetical criteria in relationship with shapes, objects and materials already present in the chosen spaces.

Even though there were four areas involved, the installation was musically divided into two parts. The first included foyer, outside (cavea) and exhibition area; in these cases the musical material was derived from a series of electroacoustic pieces by adding a spatial interpretation to their stereophonic characteristics. The general spatialization system developed by Tempo Reale already included a series of complex algorithms but it was extended for this event in order to allow a closer link with the specific architecture. Such algorithms realized movements specifically suggested by the shapes of the involved spaces and by the particular configuration of mounted loudspeakers.

The music used in these areas included excerpts from pieces by François Bayle, John Chowning, Pietro Grossi, Mauricio Kagel, György Ligeti, Bruno Maderna, Bernard Parmegiani, Henri

Pousseur, Steve Reich, Jean Claude Risset, Denis Smalley, Karlheinz Stockhausen and Daniel Teruggi.
The second, autonomous part of the installation was located in the Santa Cecilia concert hall (the biggest hall of the auditorium) and included fragments from Berio's pieces, performed with an elliptical configuration of loudspeakers, specifically located to underline the geometry and the huge dimensions of the space. This part of the installation was not open to general audience but could be experienced through guided tours.
Several people were involved in the realization of the project besides the authors: Lelio Camilleri and Paolo Pachini collaborated on the musical part and Francesco Canavese on the technical aspects. Dino and Massimo Carli of BH-Audio service and by Daniele Tebaldi and Ralf Zuleeg of d&b audiotechnik, the audio equipment company that conceptually collaborated for the installation design, also made relevant contribution.

2. THE SONIC ARCHITECTURE

An important guideline set out by Luciano Berio for the organization of the sonic architecture was the idea of a polyphonic global sensation - that attention should not be drawn away by single outstanding events and their trajectories in space, but should focus on the interaction between the movements of different sound layers.

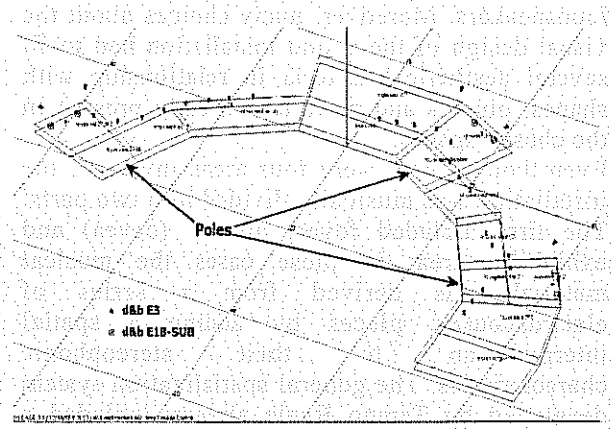


Fig. 2: Positions of loudspeakers in the foyer

The various spatialization algorithms used to characterize these layers were essentially based on indeterministic approaches. For example, continuous movements with irregular changes of direction, or discontinuous movements with random positions and random durations for rests and transitions were used. In all these cases the indeterminacy occurs within certain *constraints* (time ranges for the choice of durations, weights for the probability of certain loudspeaker constellations, etc.)
Many of the algorithms had already been developed at Tempo Reale for performances with a central listening field and a circular spatialization system around it (for example for the installation *Geografia* for the Italian pavilion of the world Expo 2000 in Hannover, with a dome-shaped architectural structure). In fact, the installation in the Santa Cecilia hall with the elliptical configuration of loudspeakers represented such a "classic" situation.
Developing a spatialization concept for the foyer and its corridors was a particular challenge, as there were no areas that could be regarded as central listening positions. Therefore, a multiplicity of listening perspectives had to be considered. In order to organize these spaces, a hierarchy was established between three main zones (defined as "poles" and situated in the larger spaces leading to the cloakrooms) and secondary zones in the adjacent corridor areas. Physically they were differentiated by full-range diffusion systems (for details see below) as poles and chains of small speakers ("rays") reaching into the corridors. The configurations were slightly asymmetrical (as shown in fig. 2), following the architectural structure. In the poles, the acoustical image created by the spatialization was intended to remain close to the original properties of the compositions, whereas following the rays into the corridors, the listener should perceive a more and more fragmentary, but always musically meaningful, acoustical image. To achieve this, a new series of algorithms was developed, adding to the existing system the concept of "expansion" and "contraction" of ranges of speakers. Thus, instead of the neat perception of activation/deactivation of single loudspeakers, there was more the idea of sonic centers continuously moving along the rays of loudspeakers.
In the arrangement of the compositions for this particular "orchestra of loudspeakers" there was the problem of obtaining a polyphony of sound layers from pieces that were available only in mixed-down stereo (or even mono) versions. In general, two types of segmentation were used: one in the time domain (assigning successive events to different spatialization engines) and the other in the frequency domain (assigning different parts of the spectrum, carefully extracted by filtering, to the various engines). Studio experimentation led to the conclusion that in both cases, floating transitions between these spatialized segments were most suitable for the overall polyphonic impression, sometimes by even simply superimposing two distinct spatializations of the same material.

In the Renzo Piano exhibition and in the outside area the use of the spatialization was less "geometrical". Instead of chains of loudspeakers there were only scattered distributions, where different listening zones were predominantly covered by single loudspeakers. This led to a different segmentation of the original material and to a different choice of the musical excerpts and of their temporal organization. In the exhibition area the installation was mostly based on slow passages among the sound layers and on slow transitions between loudspeakers; whereas, on the outside, the musical excerpts used were quite short, with surprising spatialization, giving just a "taste" of what happened in the interior of the architecture.

3. THE TECHNOLOGICAL REALIZATION

Both the audio setups for the Santa Cecilia hall and for the foyer/outside area were based on the combination of two computers and a digital mixer: one computer with a ProTools system and another with a Max/MSP environment (with MOTU 2408 interface), interacting with a mixer from the Yamaha 02R series. The first computer was basically used as a multi-track system to organize the formal structure of compositions, to extract different layers and to assign them to the various spatialization engines (implemented in Max/MSP) of the second computer or (statically) to certain loudspeakers of the foyer and outside areas. The ProTools station was also used to control Max/MSP. Using MIDI communication, it was possible to specify the types and the parameters of the spatialization movements; there were up to four spatialization engines in parallel, routing each input line to a maximum of eight loudspeakers.
For the Santa Cecilia hall, with its eight-point ellipse of loudspeakers (d&b F1220), the described system was already sufficient. The foyer and the outside area were much more complex, especially because of the high number of speakers involved. For the foyer a total of 42 speakers were used, divided into three zones. In each zone there were four speakers (d&b E3) with two sub-woofers (d&b E18-SUB) for the stereo diffusion of the pieces (the above-mentioned poles) and eight speakers (d&b E3) for the spatialization (the rays). In the outside area eight speakers (d&b C6) were placed in four different locations, each pair very far from the others. In these cases, the large distances caused several problems as far as cabling was concerned: therefore, a wireless system was used to connect some of them.

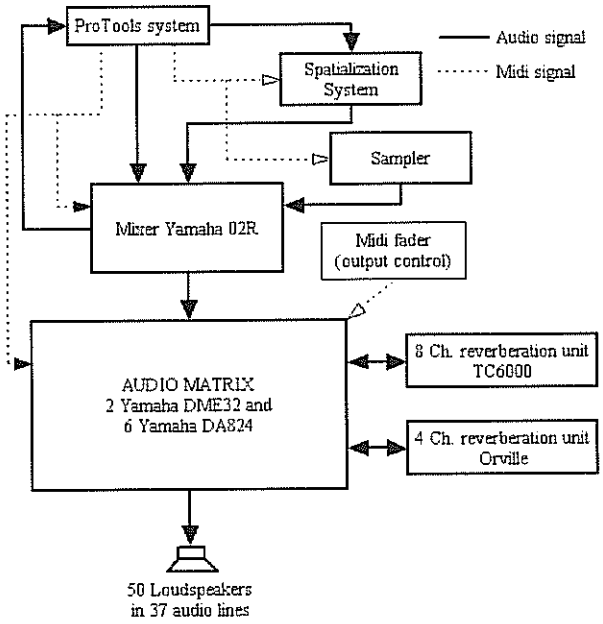


Fig. 3: Technical diagram of the system setup (foyer and outside area)

The management of the huge amount of independent audio lines (37) was the most challenging technical aspect of the installation. An independent control of the equalization of all lines was absolutely needed in order to adjust the sound to particular parts of the architecture that were not specifically conceived for music. Furthermore, there was the need to control the loudness levels separately, as during the days of the inauguration there were other performances, so that sometimes certain areas had to be closed and successively re-opened. Lastly, the routing to two different reverberation units had to be controllable. The first one was used for the rays of speakers, that were hanging from ceiling of the foyer (as they were quite close to the audience, a small reverberation was useful to slightly soften the sound movements), whereas the second one was sometimes used in the external signal lines for special audio effects.
In order to fulfil these technical requirements, two digital mixing engines (Yamaha DME32 with six D/A converters) were connected to the main 02R mixer (eight output channels of the spatialization, stereo output, outside signals), providing a sufficient number of output lines with their relative equalizations. By using the two DME32 mixing engines, MIDI messages from the ProTools stations could be utilized to dynamically change their internal routing and configurations, while the level of each area could be manually adjusted with a separate MIDI controller.

Fig. 4: Detail of one of the rays in the foyer

The setup was completed with a hardware sampler used to reproduce certain events (gong-like samples) that were musically separating the various excerpts of the compositions. It was again triggered by the ProTools station via MIDI messages. The last technical aspect to mention concerns the exhibition area, which was quite simple to solve. In fact, a ProTools system (with a Digi001 card) was routed directly to ten loudspeakers (d&b E3) through a Yamaha 03D digital mixer.

4. CONCLUSIONS

After the planning phase of the work at the Tempo Reale studios, the operative steps of the project took one week of work on site, interacting with the real environment of the architecture (in continuous transformation because of the final rush of the building works) and with the acoustic rendering of spaces. Such a situation caused a series of reconfigurations both from the musical and technical point of view.
The idea of adapting electronic music to a new context was also interesting in the framework of these two days, celebrating the opening of a space for all kinds of music. Symphony orchestras, jazz ensembles, military bands, chamber music groups and vocalists performed mainly in the concert halls, with some extra events in other places of the auditorium. But the mobility of electronic sound offered a starting point for dynamic investigations and discoveries of the various parts of this extraordinary contemporary architecture, emphasizing how this kind of music is challenging the traditional boundaries of concert halls. In this way the installation became a special tribute to that repertoire.

5. ACKNOWLEDGMENTS

The authors would like to thank the following people who contributed to the practical realization of the installation: Susanna Scarabicchi and Massimo Alvisi of the Renzo Piano Building Workshop, Francesca Via of the supervision staff of building works, Fabio Fassone and Vincenzo Cavaliere of the local organization staff, Piergiorgio Cavallari as production assistant of Tempo Reale and the staff of BH-audio as technical reference.

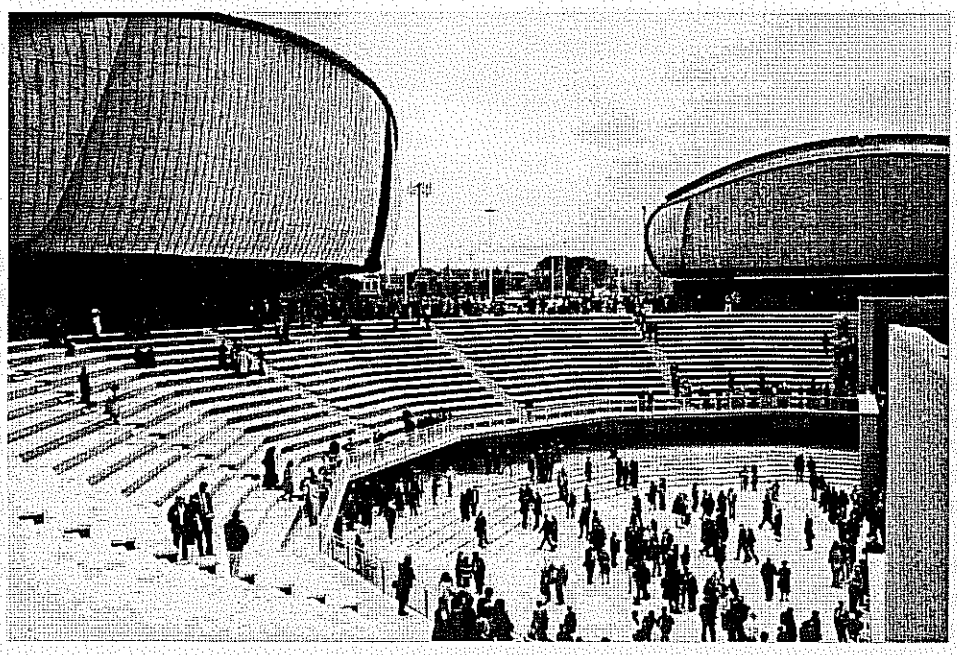


Fig. 5: The auditorium cavea

A COMPUTATIONAL APPROACH TO THE ANALYSIS OF "INCONTRI DI FASCE SONORE" BY FRANCO EVANGELISTI

Patrizio Barontini, Stefano Trevisi

Conservatorio di Musica "A.Boito" di Parma
patriziobar@tiscali.it
stefano.trevisi@tin.it

ABSTRACT

The following paper is an attempt to analyze *Incontri di fasce sonore* by Franco Evangelisti: the analytical method employed is based on a perceptive approach, in which strategies' choices and analysis tools arise from listening.

There are two processes, analysis and synthesis, which are two parts of a single cognitive process: in fact, the need of making heterogeneous data converge in a single expressive unity involves a continuous passage between analysis and synthesis in a fruitful feedback. In that way it is possible to identify the *links*, and therefore to bring out more information about the system than it would be done considering the simple sum of *elements*.

The piece has been analyzed from the perspective of sound objects, valued by means of a computational approach, which allows to observe the different compositional strategies.

1. INTRODUCTION

Incontri di fasce sonore by Franco Evangelisti is one of the most important electroacoustic works produced at the Westdeutscher Rundfunk Studie of Koln in the Fifties. It has been composed in 1956 and it was performed one year later during a program of the Koln's Radio.

There are some evident references to *Studie II* by Stockhausen as far as the construction of basic sound materials is concerned; it appears clearly by a reading both of the score's technical notes and of the piece construction graphical scheme. Even though the bases of the two pieces are similar, they seem radically different to a listening, particularly for sound objects articulation and their formal development.

2. THE REALIZATIVE SCORE

The introduction of the score includes the algorithm used to divide the frequency range between 87 and 11.950 Hz in 91 pitches; the frequency range is defined by avoiding integer or constant ratios (i.e. 5:1 of *Studie II*), and referring to variable ratios of frequency, which are defined by a geometrical progression. The 15 *groups* of pitches - each one formed by 21 *elements* (each element is defined by 7 different sine waves) - are also shown: each *element* is assembled by selecting frequencies using a simple numeric algorithm.

The score is divided into two diagrams which are defined in the time domain. Time is expressed by tape's length (centimetres). A double staff of 21 lines describes sound objects time displacement; each line corresponds to one of the 21 *elements*. Each element is located by a thick line, whose length is proportional to duration, and is identified by a capital letter (*group* of belonging) and by a number (tape's length expressed in centimetres). Different kinds of dotted line define sound transformation processes (reverberation, ring modulation and pitch transposition). The author uses a double staff to avoid a graphic overlapping of lines which could be hard to read, especially where there's a high density of sound materials. In the lower part of the score there are amplitude envelopes of simple sound objects or more complex objects, which are formed by some synchronous elements processed using the same algorithm. The amplitude range is expressed with a range from 0 to –40dB.

This realizative score is not useful for an analytical perceptive approach, because the reading does not enable to recognize quickly the different kinds of musical articulation. Moreover, it is hard to evaluate sound objects' time arrangement because tape's length is indicated in an extremely defined way.

3. THE ANALYTICAL APPROACH

It has been employed a perceptive approach, in which strategies' choices and analysis tools arise from listening.

The two levels of analysis/synthesis are presented separately, but actually they are two parts of a single cognitive process, because the need of leading heterogeneous data to a convergence in a single expressive unity involves a continuous passage between analysis and synthesis in a fruitful feedback. The synthetic step allows a progressive improvement of the analytical tools; this aspect in its turn provides new different interpretative perspectives in a synthetic level, making new emerging qualities stand out. In a complex system, such as a composition is, the *analytical* step (in the strict sense) could only discriminate the *elements*: on the contrary, considering *interactions* which are developed among *elements* is peculiar to a *synthetic* step. In that way it is possible to

identify the *links*, and therefore to bring out more information about the system than it would be done considering the simple sum of *elements*.

4. THE ANALYTICAL PROCESS

From the first listenings it was possible to steer the research to the identification of sound objects and of their transformations. A simple listening diagram has been drawn in order to outline the main directions in the articulation of perceived sound objects.

Four macro-sections have been defined, basing on the morphological distribution and the formal development of the objects.

- *Section 1*, from 00:00 to 01:04: all the object typologies used in the whole work are present; high density of sound materials.
- *Section 2*, from 01:04 to 01:59: focus on linear sound objects; rarefaction of sound materials, arranged in order to highlight a nearly melodic development.
- *Section 3*, from 01:59 to 02:49: iterative and impulsive sound objects; high density of sound materials.
- *Section 4*, from 02:49 to 03:20: high variability and low density of sound objects.

It is possible to make a first classification of sound objects after a further widening of the listening diagram. Morphological categories are defined by contrast, starting from simple perceptive impressions: for example, the differentiation between *punctual* and *linear* objects characterizes the difference between sections 1 and 2; the *continuous/discontinuous* (in particular iterative) contrast allows to isolate the third section; the exclusive presence of directional objects defines the last section. The *directional/non-directional* category discriminates objects pointing in a direction from objects characterized by a equilibrium in their inner structure. The barycentre indicates where the inner direction points, considering the object divided in four segments.

Obj	Start	End	Punctual Linear	Cont Disc	Direct Non-direct	Bary-centre
1	0.4	0.6	Punctual	Cont	Non-dir	
2	0.6	0.8	Punctual	Cont	Non-dir	
3	0.8	1.1	Linear	Cont	Dir	4
4	1.1	1.6	Punctual	Cont	Non-dir	
5	1.6	7.4	Linear	Cont	Dir	1+4

Table 1: example of morphological classification of sound objects at the beginning of the piece.

The categories which describe the specific *typology* of the objects are defined in an empirical way. Timbre is one of the most important criterion, and it permits the identification, for example, of objects with metallic timbre, with a harmonic or non-harmonic spectral typology.

The sonogram is a further analytic instrument, since sometimes it allows more detailing in the morphological discrimination of single sound objects. Such tool could seem unnecessary to analyze a work which consists of easily identifiable sound typologies and whose compositional processes are clear. Nevertheless in this work the sonogram turns out to be an analytical tool which is able, on one hand to point out the nature of some objects with a complex spectrum, and on the other to separate objects inside a thick overlapping of multiple figures.

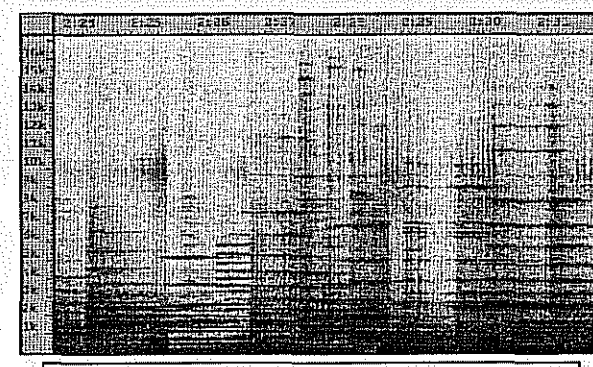


Figure 1: example of a sonogram with overlapping of multiple objects (Section 3).

Finally, the sonogram provides an interaction between visual interpretation and auditory perception of sounds: a visual approach to the work could result complex if we use a realizative score, because a reading of diagrams could easily distort the perception.

5. THE SYNTETIC PROCESS

After the classification of sound objects using the above-mentioned categories, some figures with similar distinguishing marks arranged in the whole space of the composition appeared from the observation in particular of their barycentre. By employing the sonogram, classification of sound objects can be re-calibrated more synthetically in eight categories. These are basically defined by the attack and decay shape of each object:

1) **Pt** – punctual object – attack generated by a cut
2) **Pa** – punctual object – fast attack
3) **Lt12** – linear object - both attack and decay generated by a cut
4) **Lt1f2** – linear object - attack generated by a cut and slow decay
5) **Lat2** - linear object - fast attack and decay generated by a cut
6) **Laf2** - linear object - fast attack and slow decay
7) **Lf1t2** – linear object - slow attack and decay generated by a cut
8) **Lf12** – linear object with slow attack and slow decay.

The classification of sound objects is the starting point for the next step of the analysis, which follows a statistical/distributive approach. The distribution in the time domain of the eight categories is shown in Figure 2.

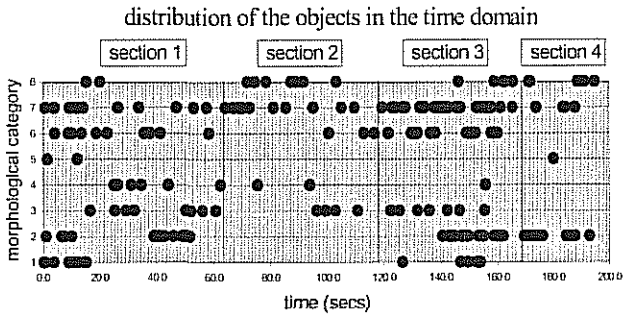


Figure 2: time distribution of sound objects classified in eight morphological categories: a red point refers to "iterative object".

After a first look to the graph, some observations can be done:

- punctual material is generally introduced in complex clusters (lines 1 and 2)
- linear objects are often characterized by an envelope with a slow attack and a cut decay (line 7)
- there are no punctual objects in Section 2
- all the categories can be found in Section 1, while in Section 4 there is the lowest variability of objects.

This diagram confirms the principles of perceptive subdivision which defines the four macro-sections. The same diagram describes in particular the distribution of discontinuous-iterative objects, which coagulate in Section 3. The presence of iterative elements in the previous sections implies a further research about their spectral morphology: its purpose is to verify the transformation by defining the iterative object which characterizes the whole Section 3.

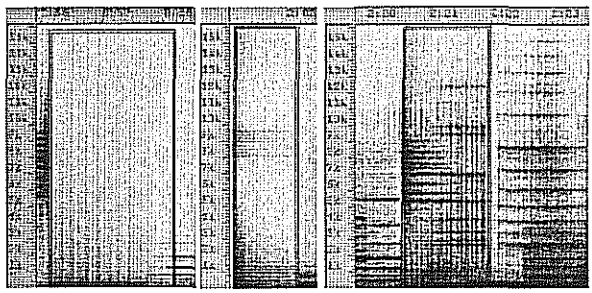


Figure 3: sonograms' examples of the iterative object's transformation: proto-iterative object (Section 1), fast iterative object (Section 3), slow iterative object (Section 3).

Duration of sound objects is described in Figure 4.

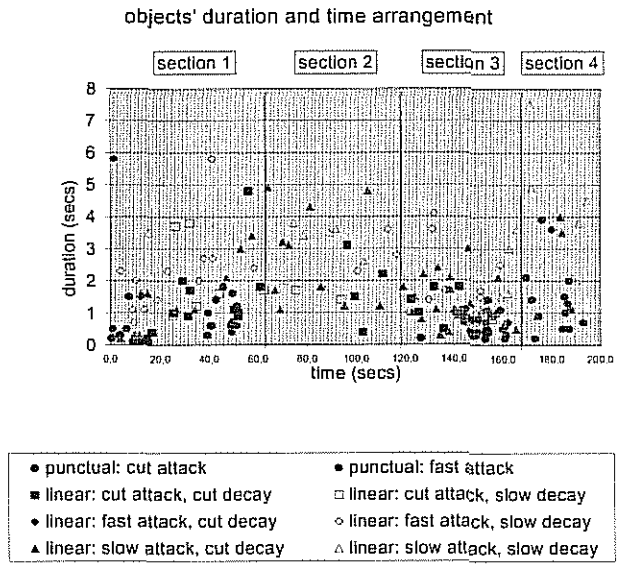


Figure 4: objects' duration and time arrangement.

It is possible to deduce some information about duration's ratio among the eight morphological categories:

- there are no objects with duration longer than 8 seconds; most objects' duration is shorter than 3 seconds
- punctual objects have a maximum duration of 2 seconds, except in Section 4 because of reverberation
- in Section 4 linear objects with slow attack and slow decay (Cf12 category) have a duration's increase, in comparison with the previous sections
- Section 1 and 4 have the largest variability of durations.

The graph shows a different distribution of duration which follows the partition in four sections: in fact the first and the third section display durations which are shorter than two seconds, whereas the second section presents the longer average durations.

In section 2 long lengths corresponds to linear objects: this fact defines the identification of a basic sound object's typology, the *texture*. Such typology was not defined *beforehand* (following for example the score or more directly the title), but it is the result of an interpretation of morphological data, which is getting more and more refined. The *texture* object is further classified in three typological sub-categories, basing on different kinds of timbre:

- linear object – bell
- linear object – scraped metal
- linear object – cluster on a vibraphone.

In the same way other two typologies, which can have the function of signal-figures to the listener, are defined:

- linear discontinuous (iterative) sound
- punctual sound with a wooden instrument's timbre.

The last object takes shape in the beginning as a very high tone punctual figure.

The distribution in time domain of the typological categories described above is shown in Figure 5.

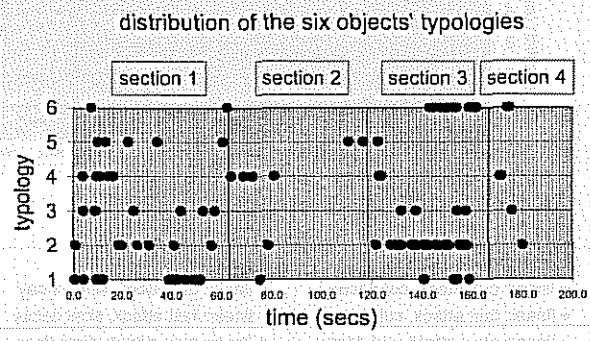


Figure 5: distribution of the six objects' typologies:
1) punctual object, very high tone
2) linear iterative object
3) linear object, bell
4) linear object, scraped metal
5) linear object, cluster on a vibraphone
6) punctual object, wooden percussion with reverb

In the last step we can recognize timbral-morphological features which represent perceptive points of reference: a morphological-distributive approach turns into a compositional one which makes perceptive paths clear by means of a *network* of signal-figures. The table described above is completed by the description of perceptive links among the objects, organized in the six typological categories.

Taking gestures - which belong mainly to a real-timbre context (percussion) - into account as perceptive references, suggested a new interpretation of the objects by assigning to the most of them a direct reference to an instrumental gesture. For example: cymbal scraped with a metal mallet, gong, blowing in a metal pipe, *crotali* with the bow, vibraphone with the bow, drum, and so on.

6. CONCLUSIONS

The progressive refinement derived by the mutual integration of analytic and synthetic process permits a focalization on what seemed to be a peculiarity of this work at a first listening: a low degree of surrogation (in Smalley's meaning) with regard to the source-bonding criterion. The composer, even though he starts from electronic materials which are similar to those that WDR composers used, he seems to take a different expressive direction, moving to a more figurative language both in the timbre (mimesis with real acoustic instruments) and in the objects' articulation.

7. REFERENCES

[1] "Cologne-WDR Early Electronic Music"(Compact Disc 9106), BVHAAST, 1992.

[2] Delalande F., "Le condotte musicali", CLUEB, Bologna, 1997.

[3] Evangelisti F., "Studio elettronico – Incontri di fasce sonore", U.E. 12863, 1956-57.

[4] Giomi F., Ligabue M., "Evangelisti's composition Incontri di fasce sonore at W.D.R.: aesthetic-cognitive analysis in theory and practice", in: Journal of New Music Research, vol.27 (1-2 1998). p.120-145.

[5] Smalley D., "La spettromorfologia", Musica/Realtà, n°50, Luglio 1996, p.121-137.

PLUS MINUS: AN ALGORITHMIC ANALYSIS AND A MUSICAL REALIZATION

Maria Clara Cervelli
Conservatorio di Musica A. Casella – L'Aquila
Università degli Studi di L'Aquila, Facoltà di Ingegneria

mclaracervelli@hotmail.com

This work is formed by two parts.The first part is made by analysis of the score and proposes flow diagrams that can drive the realization choises. Last part describes a particular realization of the work that is finalized to the individuation of the modernity elements that are contained in the score.

1. Introduction

Plus Minus, work by Stockhausen of 1963, constitues an example of the theorization of the musical composition's rules which control the evolution of most common sonorous parameters for a musician (i.e. pitch, intensity, timbre, duration).Originally created as an exercise for the first class of "Cologne New Music Courses", it is a work that provides, in an abstract and symbolic way, the criteria and the processes to follow for a composition development.

The work constitues of 7 pages containing symbols, 7 pages containing notes and an introduction containing the necessary guidelines for the realization of the composition.

Each page constitues of 7 main groups to be utilized as central sounds, and 6 sound collections (subsidiary notes) which must be utilized as 'ornaments' of the aforementioned groups or main 'types'. Every page is based on a fundamental pitch "Zentraltone", while the remaining sound are derived starting from the fundamental by following the Fibonacci series expressed in semitones or the series' multiples.

Every page of symbols constitues of 53 squares, each square representing an event to be realized.

2. Criteria and processes

Due to its sensibility to the initial choices (e.g. materials, pages, symbols' assignment), the work allows for an infinite number of compositions as long as the criteria and the vertical and orizontal organizational processes are accounted for.

In order to realize a composition, a page of notes must be matched to a page of symbols. The grouping of the 7 couples chosen, performed one after the other, without any interruption, constitues a 'layer'. A specific version may contain one or more layers laid upon each other, up to a maximum of 7. For the realization, an additional task to be performed is the decodification of the symbols related to each of the specific events. As a matter of fact, each event is characterized by a series of information contained in a symbolic manner within its square. Each symbol has a precise functionality and by analizing such functionality it is possible to classify such symbols in 3 main sub-groups: information charaterizing the structure and internal articulation (*event microstructure*); information regarding the possible event modification during the composition's evolution (*horizontal macrostructure*); blending or contrasting elements with events which are simultaneous or close temporally in other layers (*vertical macrostructure*).

2.1 Microstructure

The characteristics related to the event's internal structure are of 4 different types (see figure 1):

- the relative position of the central sound (event's specific reference) and accessories (vibrations, attacks, central portions, and decays of the central sound).This determines the event's type, establishing which group of notes to utilize in order to create the event.
- the timbre's type to assign to the event (soft sound, hard sound, soft noise, hard noise, soft sound-noise, hard sound-noise)
- the subsidiary notes to be introduced in the event specified position and based upon the provided rhythmic characterization

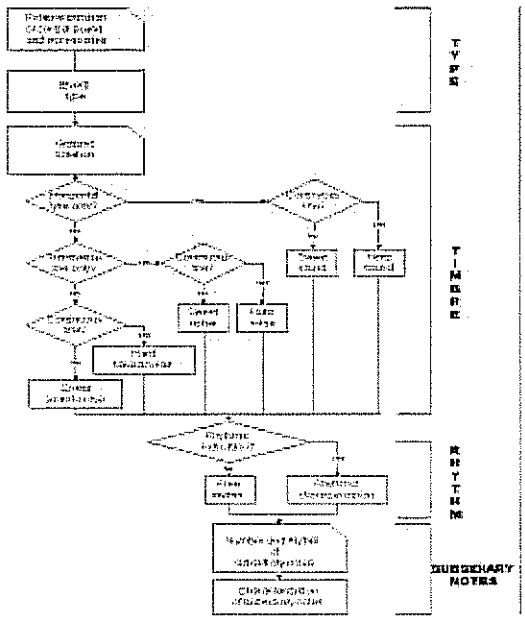


Figure 1 - Sequence of operations to be performed in order to determine the event's microstructural characteristics

2.2 Horizontal macrostructure

The symbols which establish the relationships between different subsequent events of a layer (horizontal macrostructure) can be classified in 3 chategories (see figure2) :

- either pause or lenght duration, which determine the separation of subsequent events or their total or partial blending.
- 'crescendo' or 'diminuendo' signs that regulate the intensity, pitch or duration relationship between subsequent events

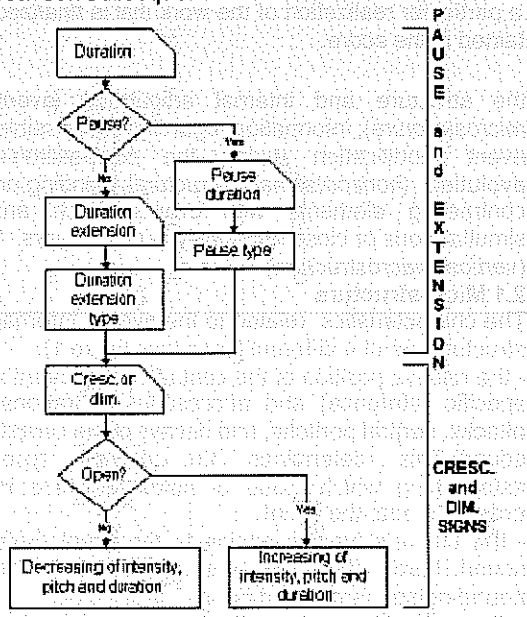


Figure 2 - Interpretation of the elements related to the horizontal macrostructure

A quite peculiar role in the general structure of the work is represented by the 'flags'"; they allow, in fact, to substantially modify the structure of an event, leading to its proliferation (i.e., increase of the number of parts that constitute it), to a substantial renovation (the number of total parts reaches the value of +13), to a negation (the number of parts becomes negative), or to a suppression of the event itself (the number of total parts reaches the value of –13).

The relevance of such symbol's role is proven by the fact that even the title itself of the work is given by the type of operation (i.e., addition or subtraction of parts) that is performed starting from such indication.

The "flags" can be seen either as a micro-structural element (which acts on the intimate nature of the event itself), or as a mean of horizontal connection, since they (better than any other elements) can give information on the layer's evolution in its entirety, influencing the relative weight of each of the blocks and by substantially contributing to the determination of the character of each horizontal layer.

2.3 Vertical macrostructure

A last set of symbols gives specific information on the methods of vertical connection between the various layers (see figure 3):

- temporal connection (synchronous between events of different layers or sequence of these events)
- frequency connection (repetition of common pitches, or suppression of these pitches)
- intensity connection

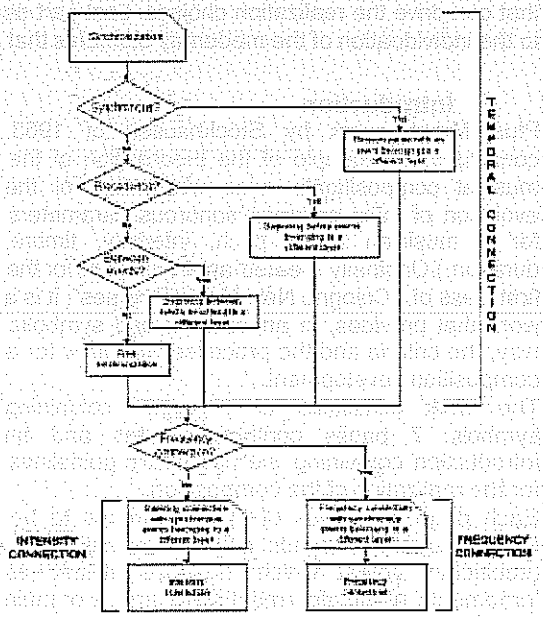


Figure 3 - Interpretation of the elements related to the vertical macrostructure

2.4 Free events

Each of the pages also contains a series of empty squares (with some square parenthesis inside), which definition is 'free events'. The creation of such events (i.e., the process of filling the empty spaces) is obtained by following the indications provided by the horizontal arrows, present at the top of the page.

In such spaces, it is possible to insert events originating from the previous page, from the following page, or events that are completely free. Hence, the presence of such events fosters even more the creation of a solid horizontal structure of the layer. This structure, consequently, enables to insert cross-references relative to what had occurred in the past, anticipations of the successive evolutions, or elements of surprise, which would allow renovating the attention or would produce certain coherence between the layers.

Each page of symbols , therefore, represents a complex object belonging to a small system constituted by seven units having the same informational weight.

Each of these objects represent an internal story linked to the micro-variations of timbre, of rhythm and of structure for each of the events that constitute it, linked to the events' dialectic, to the connection between the objects more closely related to it, and to a rich network of relations between other objects, which above or below it, tell the evolution of other systems, constituted by individual units and characteristics (see figure 4).

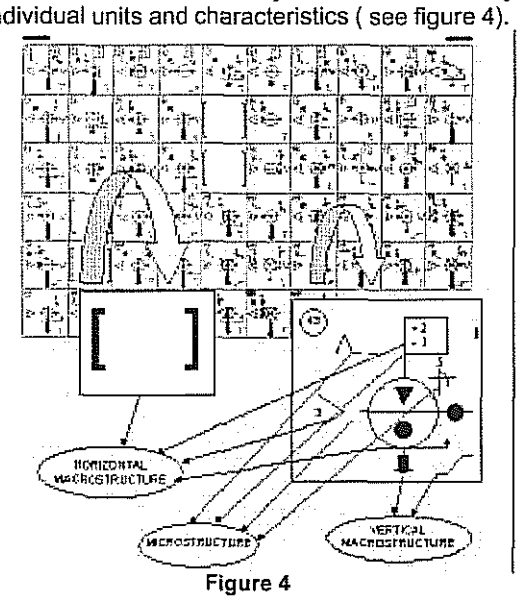


Figure 4

3.Plus Minus today: a realizing experience

The logic and the structure that can be identified in Plus Minus is certainly an offspring of Stockhausen's personal development, and of a series of reflections that were born from the musical events of that period. The work is a clear attempt of rationalization and synthesis of all the principles that had driven the musical production during the preceding decades.

At this point, it is inevitable the following question: does Plus Minus still have a real significance if seen with the eyes of the contemporaneousness? A version of Plus Minus was created to attempt to give an answer to such question.

The version is constituted of 3 layers overlayed.

The realization has been developed in three sequential steps:

- Determination of the pages' order
- Determination of the materials
- Writing of the score, and realization of the digital support

3.1 Determination of the pages' order

The first step taken was the numbering of both the 7 pages of notes and the 7 pages of symbols. Such numbering has been performed by starting from the casual position that the pages were at that specific moment; this way an initial chance was maintained while a specific order was successively imposed to this randomness.

Once the pages have been ordered, they have been distributed among the 3 layers by following a graphic method and a formal method. The distribution has been done so that by placing the 3 sequences one upon the other, two squares would be obtained which would contain on the diagonals, either the number '3' (number of layers created) or the number '7' (number which governs all the work's structure)(see figure 5). An additional choice made was the one of assigning the same final page (the one identified by the number '1') to all the three layers.

4	5	3	7	6	2	1
5	3	4	2	7	6	1
3	4	5	6	2	7	1

Figure 5

To enable an easier distinction between layers, always pages with a different "Zentraltone" had been assigned, with the exception of the last page. At this point, in fact, each of the horizontal bands has affirmed clearly its personality, and it can manifest it even by circling around the same dominant sound.

For the pages of symbols, a different approach has been used.

The analysis quickly has shown that not all the pages could be chosen as initial pages, since some of these pages present an exchange of events (i.e., free events) with the preceding page. The presence in some of these pages of exchanges with the following pages also did not allow for a random choice of the final pages.

Once the possible three initial and final pages had been identified for each different layer, the number of exchanges present in each page was studies.

For each different layer in the first part of the realization, pages that included only few exchanges were considered so that all the materials could be exposed, could be elaborated and only subsequently (during the second part), the game of cross-references and anticipations, and the introduction of surprise elements, could be intensified.

Page	Free events tot	To Prec.	To Foll.	From Prec.	From Foll.	Free events
1	7	1	1	2	3	2
2	2	1	-	-	-	2
3	2	1	1	-	-	2
4	3	-	-	-	1	2
5	5	1	1	1	2	2
6	4	-	-	1	1	3
7	2	-	-	-	-	2

Figure 6 - Exchanges present in each page. The number of events which are completely free has been determined by the following method: Free Events=Tot. Free Events – (From Preceding + From Following)

3.2 Choice of Materials

At a second stage, the materials to assign to each of the layers were chosen. Each layer has been assigned to an acoustic instrument and to a part on digital support.

The acoustic instruments used are: piano, violin and a percussion set (vibraphone, 2 tom and bamboo chimes).

A significant and relatively conditional choice (especially for the piano and the violin) has been the one of imposing that the instrument players would create all the requested typologies of timbre by utilizing exclusively the execution's possibilities of their instruments, in conjunction with an unchanged electronic elaboration based on the non-linear Distortion.

This has led to a reflection on the potentials of each of the instruments, and even more to a reflection on the concepts of sound and noise, and on which would the intensities be, the modalities of attack, the rhythmical characteristics to attribute to each of the typologies of timbre. As a matter of fact, for example, to realize on the piano an event that could be identified with the attribute "sweet sound", involved a choice that included all the parameters aforementioned.

3.3 Writing of the Score and Realization of the digital support

The last step is the writing of the score. Even in this step, a significant choice is made for the formal evolution.

The use of the "flags" has been treated in a particular way.

For each layer, three events to which impose a development that would include initially the negation, and in a following step the complete elimination, have been chosen.

For each of these, the "negative tape" has been created on a digital support, and the diffusion of such part has been assigned to an installation, to be placed on the stage with the performers, constituted of an intertwining structure of 4 tubular pipes on which 4 speakers have been installed with an angle of irradiation of 120.

The installation offers itself as a fourth interpreter, which collects in itself all the elements of negation coming from the layers, and proposes itself in a dialectic relationship with each of the instruments, while at the same time creates an additional link (besides the ones already present in the score) between the horizontal paths.

The tape has been built by starting from the actual material of the instruments, as the "picture's negative" by negating all the parameters: timbre, intensity, pitch and duration. The negation of pitch and intensity has been created with a pitch's movement, and a decrease in the amplitude of the signal coherently to the negative number reached by the event; the negation of the duration has been created by assigning to each event an envelope of the amplitude, which would constitute a presence of the number of significant signal portions, equal to the negative number reached by the event; finally the timbre has been negated with a progressive destruction of the characteristics peculiar to each fragment, leading each of them towards noise.

For the remaining events, on the contrary, the "flags" have been utilized in order to keep unaltered (as much as possible) the original structure, and the most consistent work has been the one of a study of the micro-variations.

The most stimulating part of the work has been the research of elements that could arise to the listener's attention the personality of each of the "types", hence creating a series of objects that would result coherent among themselves, but at the same time, distinct and easily identifiable.

The realization of Plus Minus has therefore proposed a series of developments and reflections; based on the experience made, an answer to the question asked can be answered in the following manner.

The work is certainly an offspring of its times, but it still presents elements of timeliness; the involvement with such a work can lead to an intimate reflection of the sonorous materials, on the potential that such materials can have, and on the transformations that can be impressed to them; above all, though, this involvement can help to reflect on the personal will of expression and on the personal path of formation.

By creating Plus Minus, one is more or less rigidly forced to follow an externally imposed form, obliged to follow a path that is not imposed by internal expressive needs, but that is already traced ahead of time; the neglect or acceptance of such path forces nonetheless to perform a critical and well-thought analysis of oneself paths and personal needs.

Aknowledgements

Special thanks to Emiliano Bamia, Lucia Marucci and Paola Salvatore who performed the described version of Plus Minus and to Michelangelo Lupone for the precious advices during the score realization.

References

[1] K.Stockhausen *Plus Minus*
Universal Edition, 1963

[2] K.Stockhausen *Kontakte*
Universal Edition, 1960

[3] K.Stockhausen *Solo*
Universal Edition, 1965

[4] K.Stockhausen *Intervista sul genio musicale*
Saggi Tascabili Laterza, 1985

[5] K.Stockhausen *Entretiens avec J.Cott*
J.CLattes, 1974

[6] R.Maconie *The works of K.Stockhausen*
Marion Boyars, 1976

[7] P.Boulez *Note di apprendistato*
Einaudi, 1968

[8] P.Boulez *Pensare la musica oggi*
Einaudi, 1968

[9] F.K.Prieberg *Musica ex Machina*
Einaudi, 1963

[10] F.Galante, N.Sani *Musica espansa*
Ricordi-Lim, 2000

STRIA, BY JOHN CHOWNING: ANALYSIS OF THE COMPOSITIONAL PROCESS

Matteo Meneghini
CSC-DEI Università di Padova
menego@dei.unipd.it

ABSTRACT

Stria is a piece fully generated by the computer, with the creation of all the parameters needed to play each sound, starting by a certain number of input sessions, elaborated by algorithms. This paper contains the partial synthesis of the analysis of this piece; the analysis was conducted starting from the original algorithms, the listening of the piece and a few literature documentation. Direct communication with J. Chowning was important too.

1. INTRODUCTION

Stria was composed by John Chowning, at the Stanford University (USA), in 1977, while working at the Center for Computer Research in Music and Acoustics (CCRMA). After he discovered that frequency modulation could be efficiently applied to the synthesis of sound (1967-1971), he composed several works using the results of his research: together with *Stria* we remember *Turenas* (1972) and the later PhonŒ *or Phoné* (1981). Each of those compositions is intended to give value to a specific technique he had worked on: in *Turenas* he used his studies about spatialization to define the travel of sounds in a quadraphonic space, in *Phoné* he dealt with spectral fusion between sounds, using an algorithm applying frequency modulation to the synthesis of sung vocal tones. In *Stria*, as we will see, he used computer synthesis to interrelate the small-scale sound design to the whole composition's structure.

Before going into the details of the structure of this piece, it is important to get the basic knowledge about the numeric construction which is at the base of this work: the golden mean.

2. GENERAL PROPERTIES

In this section we will analyse the general properties which characterize *Stria*, starting with some basic definitions about the golden mean, and continuing with the description of the pitch space and spectrum division, of the instrument played and of the temporal structure of the piece.

2.1. The Golden Mean

Considering the geometric and architectonic origin of the golden mean (or golden section), we start considering a segment of length z=1, and look for its part x such as the ratio between z and x is equal to the ratio between x and the remaining part of the segment itself. To do this, we can consider the equality

$$\frac{1}{x} = \frac{x}{1-x} \quad (1)$$

Solving this equation for x, we find that

$$x = \frac{1}{2}\left(-1+\sqrt{5}\right) = 0.618 \quad (2)$$

We can then extend this result to the continuous proportion

$$\frac{1-x}{x} = \frac{x}{1} = \frac{1}{1/x} = \ldots \quad (3)$$

which numerically corresponds to

$$\frac{0.382}{0.618} = \frac{0.618}{1} = \frac{1}{1.618} = \ldots \quad (4)$$

In ancient times, the important ratio we have obtained this way was considered a rule of physical perfection. It is easily recognizable in many human works (eg. in architecture) and in nature too.

In music, the golden section represents (in a good approximation) a minor sixth, in western notation. As a matter of fact, an eight-semitones space is defined by the ratio

$$\sqrt[12]{2^8} \cong 1.6$$

which is near to the golden mean.

Another important property to remember is connected to the Fibonacci succession: each of its terms, starting by 0,1,2, is obtained with the sum of the two immediately preceding terms. In particular, it can be proved that the ratio between two consecutive terms of this succession quickly tends to the golden section. From this, we can easily say that the powers of G=1.618 are ordered in accordance with the Fibonacci succession, i.e. that the equation

$$G^n = G^{n-1} + G^{n-2} \quad (5)$$

is true.

These properties are very important in reading *Stria*, and must be remembered in the following analysis.

2.2 Pitch Space And Spectrum

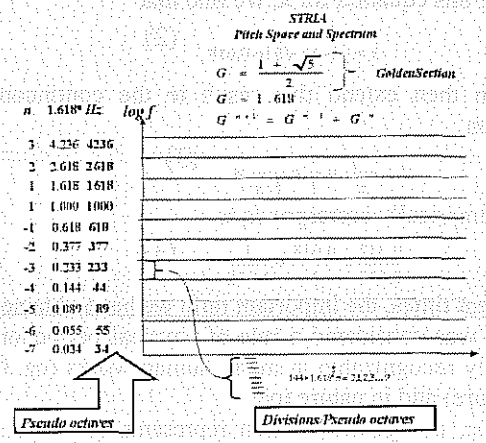
After a long series of experiments on FM synthesis, Chowning tried to discover an inharmonic ratio to re-define the concept of octave: he needed a ratio which generated FM synthesis components some of which are exactly powers of that ratio. After many tests (executed before programming) and fascinated by the sound of FM ratios $c:m=G^n:G^m$ where n and m are integer powers, he found that the golden section really had all the properties he was looking for (1974, Berlin). He re-defined the concept of octave (usually based on the ratio 1:2), using the ratio 1:G=1:1.618. Each pseudo-octave generated was then equally divided into 9 tones, by the factor

$$G^{\frac{k}{9}} \quad (6)$$

An eighteen tones division was also available, obtaining a sort of semitones. The pseudo-octaves used in *Stria* are generated around the central frequency f=1000 Hz, and the fundamentals of each octave are expressed by

$$G^{-3}f, G^{-2}f, G^{-1}f, f, Gf, G^{2}f \ldots$$

A good representation of the pitch space is given by the following figure (due to Chowning himself).



In this figure we can see the fundamental notes of each pseudo-octave, the division of each octave in 9 tones, and the properties of the golden mean. The fundamentals of each octave are connected to powers of G: the advantage due to the use of the golden mean is that this relationship is also linear.

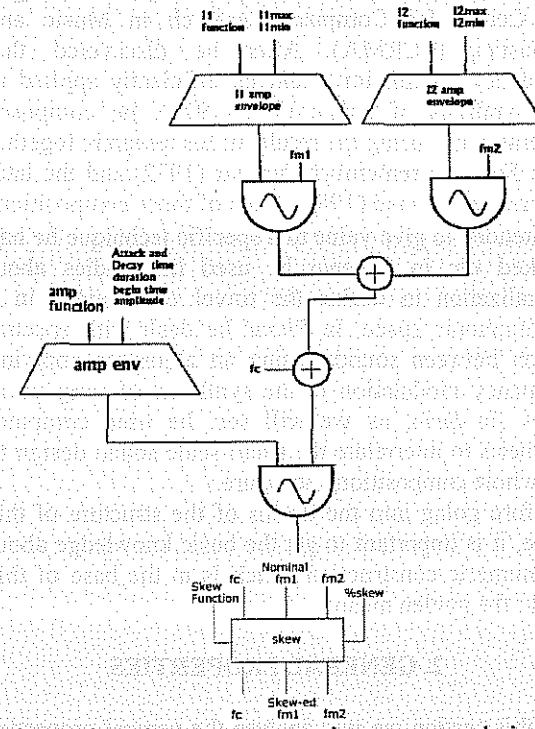
As we have seen, G is the limit of the ratio between two successive terms of a Fibonacci succession: then we can derive that by adding two following powers of G, we obtain another power of G, by the table

Power of G	Linear combination a+bG
$0.056=G^{-6}$	13-8G
$0.090=G^{-5}$	5G-8
$0.146=G^{-4}$	5-3G
$0.236=G^{-3}$	2G-3
$0.382=G^{-2}$	2-G
$0.618=G^{-1}$	G-1
$1=G^{0}$	1
$1.618=G^{1}$	G
$2.618=G^{2}$	1+G
$4.236=G^{3}$	1+2G

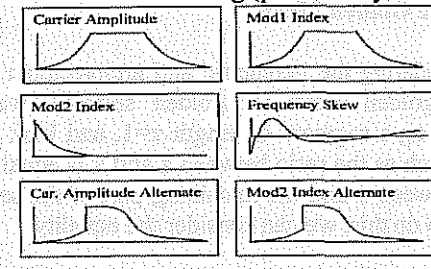
The spectral components obtained with the sum of powers of G, can be expressed by linear combinations a+bG. Chowning decided to use this property defining a carrier to modulator ratio for the FM synthesis equal to G: in this way the components generated were sums or differences between powers of G, which were also in a linear relationship a+bG with G. With this efficient mechanism, the whole pitch space was ordered in way that there was no component in discordance with the golden ratio. In *Stria* Chowning used eight pseudo-octaves, three above and five below the central frequency (f=1000 Hz): all these pseudo-octaves are used in the composition.

2.3 The Instrument

Using this efficient division of the audio spectrum, Chowning generated all the sounds with a unique instrument: starting with the input parameters, the algorithms generated 30 parameters for each instrument. The whole piece can be intended as played by a 26 instruments orchestra: each instrument generates one sound every time it is called to play; the sound played is different at every call, having different parameters used in the call itself. The basic scheme used for the instrument is the FM modulator, with double modulator: all the oscillators used were sine functions, and the modulators were summed to the carrier frequency. The sound was then shaped in time: Chowning applied envelope generators to the amplitudes of all the oscillators, thus varying the amplitude of the signals, and changing the spectral content of the sound in time. A light deviation (called skew) was added to the frequencies of both the carrier and the modulators in proportional way, to obtain a major liveliness and reality. The two modulators allowed Chowning to increase the spectral density without using large indexes. Large modulation indexes would have reduced the contribution of the carrier in the modulated sound, by reducing the zero-order Bessel function, which Chowning didn't want. The instrument can be represented by the figure



where the skew generator is represented in an equivalent way, considering in input the nominal values of the frequencies, and the parameters defining the skew; in output there are the skew-ed values of the frequencies, which will be applied to the sin oscillators. The amplitude envelopes used for the oscillators were the following (provided by Chowning)



There were two possibilities for the amplitude envelopes: the normal and the alternative one. In the normal case, the sound started as modulated by the second modulator, continued as a double modulator FM, became a FM sound modulated only by the first modulator, and finished as only the carrier. The alternative case was used at the climax of the piece, to produce a sort of *ssshBoom* effect (as Chowning called it): this effect was due to a rough variation of the second modulating index (generating a very rich spectrum), accompanied by a step variation of the carrier amplitude.

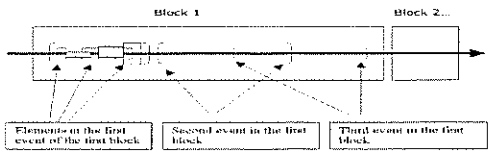
The skew function was defined by a small deviation at the attack of the sound, that becomes even smaller in a short time. It is difficult to hear this small deviation on a single sound, but it is easily recognizable in the superposition of many musical elements, as a sort of beating between components. The maximum amplitude of the skew function is determined by the frequency of the tone to play, using a low-pass function, in order to obtain a sharp deviation at low frequencies, and a small deviation at high frequencies. The human hear is more sensitive in frequency variations at high frequencies, than in low frequencies. This justifies the trend of the function defining the skew percent in frequency.

For each instrument is defined a set of three spatialization parameters: the reverb to apply to the sound, the apparent angle of the source and the apparent distance of the source. The last parameter is defined in accordance with the theory explained by Chowning in "*The Simulation of Moving Sound Sources*" (1971, [1]) and in "*Perceptual Fusion and Auditory Perspective*" (1990, [2]). In both these papers we find that the perception of the distance is based on the ratio between the direct sound intensity (varying with the distance) and the reverberated sound one (fixed with the distance).

Anyway, the spatial control used in *Stria* (no Doppler effect was used) was not intended to be precise: the sounds resulting using these parameters were similar to the ones that would be obtained in a reverberant cathedral, amorphous, big and undefined.

2.4 The Temporal Structure

Stria can be considered composed by a microstructure and a macrostructure. The whole composition is divided in blocks, a sort of input sessions, each of which is saved into three different files. Each block is composed of few events: each of these is defined by a set of input data, and is composed by a great number of elements (single sounds, played by single instruments), generated by the algorithms starting from the input values. The following picture represents the linear temporal composition of the piece:



The arrow represents the temporal evolution of *Stria*, the big rectangles are the blocks, and between brackets we can see the events; in the first event are represented also the elements composing it (by small rectangles). We can consider the elements as the atoms of which *Stria* is constructed. Each element is a single sound, and the succession and the superposition of these sounds is the whole piece.

This is the macrostructure of *Stria*: each element is defined by some microstructure parameters, defining its characteristics, as will be clear in a following section; some of these parameters are generated using the golden mean. While generating of the events the algorithms created the single sounds (elements) by the definition of their parameters:

Chowning used recursion to generate further sounds (child elements) superimposed to the original ones (parent elements); this will be discussed in a successive section. The whole piece is 17 minutes long: in the first part of *Stria* the intensity increases and after 10 minutes (number which approximately stands in the golden ratio with the total length), there is a climax, followed by a quasi-silence moment, after which the intensity grows again. The organization of the sound events in the time-frequency space is opposite to the traditional one: usually, in fact, the low pitch events are longer than the high pitch ones; in *Stria*, instead, the longer events have higher pitch, and vice versa.

The pitch of the sounds in *Stria* decreases towards the climax, and increases after this moment. Also the attack and decay time of each sound is determined by its frequency (e.g. an high pitch sound will have a slow attack).

3. THE PROGRAM

The program is written in SAIL (Stanford Artificial Intelligence Language, [3]), a language created in Stanford similar to ALGOL and PASCAL. It consists of a few procedures, called by a main program; the events are generated by the procedure EVENT2, receiving data from input, and calling other service routines to generate the frequencies (INHARM), the times (PROPORTION), the spatialization parameters (AZIM) and to write the output files (WRITE). Each call of the program defines one block, composed by few events: each of these events is defined by input parameters; the output of the program consists of three score files, one of which containing 30 parameters for each instrument to play.

3.1 The Frequency Generation

Frequency generation is managed by the procedure INHARM, called by EVENT2, when generating each element in the event. Each element created in the current event has a different value of the frequency of the note played, in accordance with the variation of a parameter (*num*) at every call of INHARM. It is the variation of this parameter that creates the melodic line of the event. In EVENT2 the frequency of each sound to play is generated by the expression

$$f = fff \cdot freq \cdot k \quad (7)$$

where *freq* is the base frequency of the event, i.e. the frequency of the fundamental note of the pseudo-octave on which the whole event is constructed (equal for every sound in the event); *fff* is the coefficient (scale frequency) that determines the note played by the current element in the event, by the product *fff*freq*, varying by element to element in the event (at every call of INHARM); *k* is a coefficient used to calculate the frequency to play on each oscillator by the product *fff*freq*k*. For example the carrier frequency is calculated from the note *fff*freq* by the expression

$$c = fff \cdot freq \cdot f_c \quad (8)$$

fff and *k* are multiples of G, and *freq* is given by an expression like G^j*1000. *fff* varies from element to element, at every call of the INHARM procedure.

For each event is defined a frequency space variable, by the expression

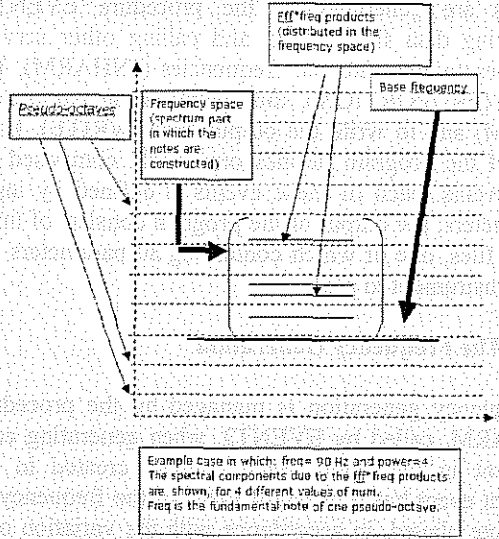
$$space = ratio^{power} \quad (9)$$

where ratio is equal to G=1.618, and power is integer (positive or negative). This variable represents the frequency space to be occupied of the event, i.e. the dimension of the spectral space occupied by the event. Power is the number of pseudo-octaves used in the current event (above or below *freq*, in accordance with its sign). Each element will play a note in one of these pseudo-octaves, that will be divided in 9 or 18 tones.

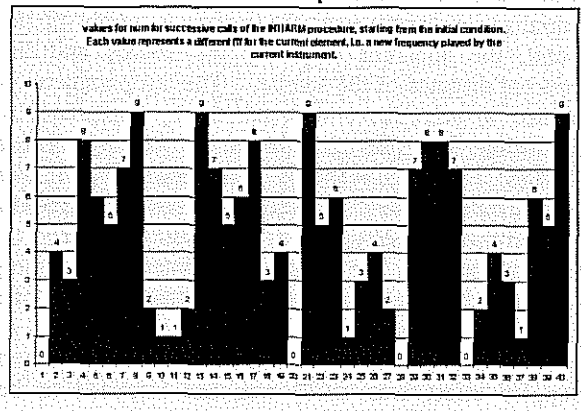
This division is done by the variable *fff*, which defines the note to play, and is calculated by the expression

$$fff = (ratio^{power})^{num\,/\,divx} = ratio^{\frac{power \cdot num}{divx}} \quad (10)$$

where *num* is different for every element generated, and is calculated by the service routine INHARM, called by EVENT2; *divx* is a variable which is equal to the number of divisions chosen for the frequency space (9 or 18). It is important to note that for |*power*|=1 Chowning didn't use the 18-notes division, as a way to avoid noisy spectra.



This figure shows the relationships between the parameters used to define the frequencies of each element in a particular case. Now we will see how *num* is generated, i.e. how the melodic line is created in *Stria*. Num is constructed by a table of 10 values, that creates a succession of number used to calculate *fff*. This succession, in the case of a 9-notes division, is periodic with repetition period 40 (if not re-initialized), meaning that the melodic line would be repeated only every 40 elements; in *Stria* there is no event with more than 40 (parent) elements; this means that the melodic line isn't repetitive (for the child element the base frequency is different from the parent's one, and no repetitive melodic line is possible, even though continuing to read the 40-period succession). The generation of the values of num in the subsequent events can continue to follow the succession (creating the periodicity) or re-initialize it from the initial values. This choice (done by input) is useful to increase the melodic variety in the piece. In the case of an 18-notes division of the frequency space occupied by the event, the table is read in a different way, generating a succession with period 20, of values of num comprised in the range 0..18; *divx* is equal to 18, in this case. In this way Chowning allowed some events to generate elements playing a kind of semitones. It is important to note that the recursive sounds (child elements) can be constructed only on a 9-division space, while the parent sounds can have *divx* equal to either 9 or 18.



The last figure represents a particular case of a melodic line (of values of num) generated starting by the initial conditions of INHARM for 40 successive elements in an event, for a 9-notes division of the frequency space. From the values of *fff* and *freq* for the current element, procedure EVENT2 calculates the carrier frequency and the second modulator by the formulas

$$c = fff \cdot freq \cdot f_c \quad m_2 = fff \cdot freq \cdot f_{m2} \quad (11)$$

where f_c and f_{m2} are the frequency coefficient already explained. The determination of the first modulator frequency is quite different, and for this oscillator there is the possibility to maintain the same first lower (or upper) side frequency constant for all the elements in the event, besides the traditional way (which would create different components for all the elements in the event). In this case the formula used (for a constant lower side) is

$$f_{m1} = fff \cdot freq \cdot [-ratio^{\frac{9-num \cdot power-1}{9}} \cdot (f_c - f_{m1}) + f_c] \quad (12)$$

and remembering (10) the first lower side frequency is

$$f_{ml} = (f_c - f_{ml}) \cdot freq \cdot ratio^{8/9} \quad (13)$$

which is constant with num, i.e. for each element in the event.

3.2 Time Generation

Another important topic is time generation: each instrument has time parameters, as begin time, duration and attack and decay time. For the determination of the first two parameters, the procedure EVENT2 (generating the events) calls the service routine PROPORTION, for two reasons: to calculate a global weight factor (*sc_prop*) and to calculate the temporal weight of each element (*prop*) in respect to the total attack duration of the event. From this last parameter, knowing the attack duration of the event (*at_dur*), i.e. the part of the event in which the elements can begin to play, the begin time of the next element is calculated by the current begin time by the formula

$$nextbeg' = nextbeg + prop \cdot at\_dur \quad (14)$$

The total attack duration of the event is then partitioned between all the (parent) elements in the event. In each event the elements are numbered by a counter, *cnt*: the first instrument that begins to play is the number 1...

The duration (*el_dur*) of each element is determined by considering the remaining time from the beginning of the instrument play to the end of the event, weighted by a factor directly related to the number of the element generated, in accord with the expression

$$el\_dur = (beg + dur - el\_beg) \cdot \left(\frac{cnt}{elements}\right)^{2-ext} \quad (15)$$

where beg and dur are respectively the begin time and the duration of the event, *elements* is the number of parent elements in the event and *el_beg* is the begin time of the current element. *Ext* represents a weighting factor for an exponential interpolation, and is comprised in the range $0.8<ext<1.5$ in *Stria*. An important situation happens when there is no overlapping between two successive elements: in this case Chowning imposed the overlapping condition on the elements, making longer the element with no overlap, by the formula

$$el\_dur = (nextbeg - el\_beg) \cdot 1.25 \quad (16)$$

where *nextbeg* represents the begin time of the next instrument, and *el_beg* the current one.

The attack time of each element in the event is determined by an exponential interpolation between two values: the attack time of the first element, and the attack time of the last element in the event. To make this interpolation a parameter *interp* is used, to calculate the position of the element in the event: this parameter is obtained by

$$\text{int}erp = \frac{el\_beg - beg}{at\_dur} \quad (17)$$

which represents the distance of the beginning of the element from the beginning of the event, normalized to *at_dur* (it is comprised between 0 and 1 for all the elements in the event); by this parameter it is easy to calculate the attack time of the current element by exponential interpolation between the initial and final values (INITATT and ENDATT) with the formula

$$attack\_time = el\_dur \cdot INITATT \cdot \left(\frac{ENDATT}{INITATT}\right)^{interp} \quad (18)$$

The same for the decay time. These parameters can be determined also depending by the frequency (by an input choice) in way to obtain short attacks for low pitches and long attacks for high pitches; the inverse for the decay time.

3.3 Spatialization

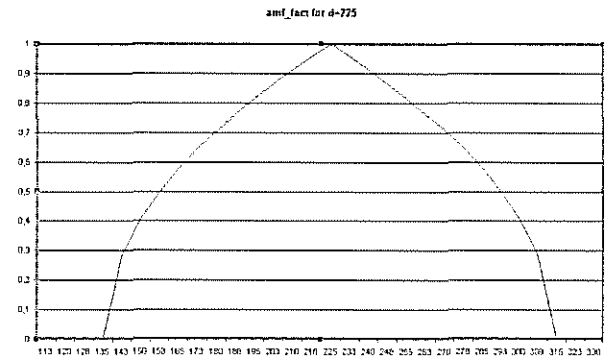
Spatialization is given by three factors: the reverb percent of all the sounds in the event, the apparent angle of the source and the apparent distance of the source of the sound.

The reverb is constant for every parent element in the event, while the apparent angle of the source is given by the quadraphonic diffusion of the sound, by calculating the four amplitude factors, for the four speakers. For the speaker positioned at the angle *d* (equal to 45, 135, 225 or 315) that has to emit a sound apparently diffusing from the angle deg, the amplitude coefficient is calculated by the formulas

$$amp\_fact = \sqrt{\frac{\text{deg} - d + 90}{90}} \quad \text{for d-90<deg<d} \quad (19)$$

$$amp\_fact = \sqrt{\frac{d + 90 - \text{deg}}{90}} \quad \text{for d<deg<d+90} \quad (20)$$

and for the speaker positioned at d=225 degrees, the amplitude diagram is



Each event has many elements, numbered by the variable *cnt*: the reference angle *el_deg* of each parent element is calculated by the formula

$$el\_\text{deg} = ev\_\text{deg} + 360^{\left(\frac{cnt}{elements-1}\right)^{0.5}} \quad (21)$$

where *ev_deg* is the reference angle of the current event: this means that the elements in the event rotate around the listener, more than 360 degrees per event. Another rotation is introduced by event to event: at the end of the generation of an event (a parent or a recursive one), in fact, there is a rotation of –90 degrees: this means that a slow rotation on the events is superimposed to the one due to the elements, in the opposite direction. As we will see in the following section, the child events generally spins around the listener faster than the parent ones, generating dynamism.

The apparent distance from the source is given by varying the ratio between the intensity of the direct sound component and of the reverberated one: the second one is kept fixed in the whole generation of parent elements, while the first one can vary by element to element, in exponential way with the formula

$$dis = (cnt \cdot DIS\_SCALE)^{0.8} \quad (22)$$

where *DIS_SCALE* is the distance scale of the event, and *dis* is the distance parameter of the current element, numbered by *cnt*, obtaining an apparent movement of the source (varying its position exponentially).

3.4 Recursion

In the generation of an event a recursion may appear: after the creation of each element, procedure EVENT2 checks if two conditions are verified:

- The number of recursions done in the event is less than the maximum one imposed by input (usually one recursion per event)
- The value of the weighting factor prop for the current element is a particular one (condition which is verified on average one time every five elements)

If both these conditions are true the current element is the parent of a child event, and procedure EVENT2 calls itself with other parameters, to construct the new event, which has different characteristics from the parent one.

The child event has duration and attack duration shorter than the parent's ones, because these parameters are scaled by *prop*<1 in the recursive calls: this means that the recursive events are shorter than the original ones; in addition, being shorter the attack duration of the child event, its elements will be closer than the original ones, increasing the dynamism (shorter event with closer elements).

The number of child elements is kept minor or equal to 9 (while the number of parent ones can be major), to avoid the possibility of instrument overflow. The base frequency for the new event is given by the expression *fff*freq*, i.e. it is equal to the frequency of the note played by the parent element which generated the recursion; the space variable is chosen by imposing |power|=1, and this means that the recursive event occupies only one pseudo-octave, above or below its base frequency (above if the original value of power was <1, and vice versa). In the case in which the new base frequency is greater than 1618 Hz, power is kept equal to -1, in way to avoid frequency divergence.

The frequency space occupied by the child event is divided in 9 notes (no 18-tones division), in order to avoid too rich spectra. The child event begins in the moment in which the parent one begins, and has attack and decay times calculated by the values by input and independent by the frequency. The reverb percentage for the childs is 1.2 times the parent's one, resulting in a major reverb percent in a minor duration. The rotation around the listener is generally faster than the parent's one, depending on a smaller number of elements in a shorter event.

The generation of the parent elements continues after the creation of the child ones, starting from the moment in the score in which it was interrupted.

At this point it is easy to understand the meaning of recursion in *Stria*: in the generation of an event, one element can create a recursion, which generate another event, shorter, starting from the moment in which the parent element begins.

This new event is shorter and has closer elements, spinning around the listener faster than the original ones, has a frequency space occupation one pseudo-octave wide and a base frequency equal to the note played by the parent element; the reverb percentage is bigger, for a shorter time, to increment the acoustic weight of this event. This, in other words, means that the recursion generates an explosion of the sounds in correspondence of the recursive call, with many close and short sounds rotating fast around the listener, creating a big dynamism.

4. CONCLUSIONS

After this analysis it easy to understand that the most meaningful aspect in Stria is the formalized process which controls both the global and the low level parameters in the composition of the sounds. The recursion is used to add importance, dynamism and speed to the sounds, thus generating an acoustic explosion.

5. REFERENCES

[1] Chowning, J, "The Simulation of Moving Sound Sources," *J. Audio Eng. Soc.*, 19(1), 1-6, 1971.

[2] Chowning, J., "Perceptual Fusion and Auditory Perspective." in P. Cook (ed.), *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics.* Cambridge, MA: MIT Press, 1999.

[3] Smith N. W., *SAIL TUTORIAL*, 1976, http://pdp-10.trailing-edge.com/decuslib20-01/01/decus/20-0002/sail.tut.html

[4] Chowning, J., Bristow, D., *FM - Theory & Applications*, Tokyo: Yamaha Music Foundation, 1988

[5] Dodge, C., Jerse, T. A, *Computer music: synthesis, composition and performance*, N.Y: Schirmer, 1985

[6] Snijders C. J, *De gulden Snede*, De Driehoek, 1969

[7] Original algorithms in SAIL, and output files

[8] Personal communication with John Chowning. Special thanks to him for his careful reading this paper.

E.M.M.S.A. - Educational, Musical and Multimedial Software Archives

Stefania Di Blasio

Centro Studi Musica & Arte (CSMA) - Florence
emmsa@musicarte.it - www.musicarte.it/emmsa

ABSTRACT

The article describes a new service that was established in Florence about a year ago, the EMMSA – Educational, Musical and Multimedial Software Archives. It is a research centre concerning educational technology and specialized in music teaching. In this centre there are educational software and multimedia CD-rom, specialized magazines, articles and books concerning the use of educational technology in Education and in Music Learning and also materials produced by schools carrying out projects in which teachers and children have used computers and educational software.

1. What is EMMSA ?

EMMSA is an educational, musical and multimedial software library. Besides a wide range of software there are articles, magazines and books concerning the use of educational information technology, text-books about educational psychology and music teaching, a collection of projects developed by Schools, Educational Institutes and Associations. All digital and paper documents present in the centre are classified in a database that facilities the search of materials to consult (articles and books) or to test on the computers (software, web site and CD-rom). The database is arranged according to category, subject, age, school level, allowing you to find the software or pubblication which is most suitable for your needs. EMMSA equipment consists of: 2 Macintosh, 1 PC Window, printer, Midi keyboard, CD writer, sound amplification and Internet connection. Emmsa computers can be used to consult software on the spot and the presence of two system computers (Mac – Pc) allows the MacOs and Window software to be tested.

Set up in February 2002, the EMMSA project is an Integrated Regional Plan promoted by the Right to Study ideals in collaboration with the Florence Town Council Educational Authority of Florence. The project designer is Stefania Di Blasio from the Centro Studi Musica & Arte of Florence.

2. EMMSA aims

Everybody talks a lot about educational technology and about its use in Education; most Schools are fairly well equipped with computers both in terms of quantity and quality, but very often teachers are unfamiliar with specialized software for carrying out interesting experiences with their students. Knowledge concerning educational technology and relevant software is rather limited also among high school, conservatory and university students. Filling this gap isn't at all easy because it is difficult to find information and material, above all it's impossible to have access to the software except by purchase. EMMSA wants to overcame this difficulty providing a software reference library and an advice point. So Emmsa's most important aim is to make the most useful educational musical software available to teachers, educators, students and families

There are a lot of very interesting projects concerning the use of Technology in Music Education from early childhood to the highest level in the Universities, so EMMSA has a remarkable collection of useful links where details and materials can be found.

Moreover EMMSA wants to give help and encouragement to teachers and operators in carrying out experimental musical projects based on new experiences of active learning assisted by computer technology An other EMMSA goal is to update educational methodologies through appropriate and creative use of computer technology in Education.

In Music Education and Teaching EMMSA wants: to increase the role of Music Composition and Creative Activities through Computer Technology and Computer Music; to foster development of perceptive abilities and musical memory from early childhood at school or at home through computers, appropriate educational software and multimedia cd-Rom.

3. Who is EMMSA aimed at ?

EMMSA is aimed at educators of every level and order including music teachers at elementary, middle and high school, university and conservatory level, cultural and social workers and therapists, students, parents and families. Teachers and Educators can apply to EMMSA to consult catalogued materials or to test and familiarize themselves with software and Cd-rom. They can also get a free advice for carrying out projects in their Institutes. EMMSA experts give assistance in setting up the project, identifying the appropriate hardware and software and in finding them. Students can look for and test particular software typology (music notation, sequencer, ear-training) or they can do research into special aspects of educational technology. Families and children generally approach EMMSA to discover and to use very interesting educational music software, intelligent games and multimedia Cd-rom that offer not only entertainment and fun moments but also opportunities for learning and knowledge especially for children that don't have appropriate home computers.

4. How to get access to EMMSA

You can get access to EMMSA in different ways:

fully access to our service can be had on Tuesday afternoons and Friday mornings in the Emmsa workshop. Via Internet on the EMMSA web site it is possible to find information, access the online general catalogue and particular records or book appointments and seek help for experimental projects by e-mail.
EMMSA Services are:
Consultation and research service in the general catalogue
Computer laboratory for experimenting software and multimedial cd-Rom selected
Tutoring while projects are underway: programming, experimentation and verification
EMMSA gathers material (cd rom, papers, web site) produced by Schools, Educational Institutes, Music Schools
Consultation of the published material concerning the use of computer technology in teaching and in learning.
Web consultation and Internet navigation

5. Why Technology in Education ?

The latest theories concerning learning and the thesis of contemporary pedagogists make reference to educational technology and to the changes they involve in applied methods and in teachers' role. EMMSA intends to promote technology use in Education, especially in Music Education suggesting musical computer activities integrated into school curricola.
Computer technology forces us to revise our thoughts concerning teaching in order to embrace the complexity of a learning process in which there are as many cognitive styles at work as there are ways of gaining knowledge. In a computer assisted pedagogic pathway, traditional and multimedial teaching collaborate in order to develop consistency in integration between the different dimensions - cognitive, emotional and creative - within the knowledge gaining experience.

6. Why should Technology be used in Music Education?

Because Educational software help to plan out *Constructionist* music lessons where teachers and students collaborate actively in the carrying out of research experiences through play, simulation and experimentation. Computers allow children to manipulate sounds and to create musical structures even if they aren't able to play a musical instrument. Musical technology is efficient in developing musical skills and in improving the child's multisensorial dimension. So the presence of computer in music education may have different goals: to broaden children's perception capabilities and musical memory; to help children to create sound structures with the same immediacy with which they draw or paint; to carry out multimodal experiences alongside reflection on what they are doing, seeing and hearing.

7. How Technology is used?

EMMSA experts propose two different ways of using computers:
- Children work in groups with a computer linked to a projector. The children take it in turns to use the computer, the others participate actively by watching the image projected on the screen and carrying out related activities. The whole group interacts. This method allows a high level of interaction in the group making use of the contribution given by individuals to common activities that are thus shared.
- In pairs or individually in the computer laboratory when there are particular requirements. This method can be used at a lates moment, when a remarkable level of individualization is required in the students' proposals and personal production.

8. The software in the EMMSA catalogue

The software available for consultation in EMMSA office is costantly increasing through the acquisition of demo, freeware or shareware copy from the web. Periodically when the very limited funds are available we order complete version. We can list:

- Creative and educational software for pre-school and elementary school children.
- Educational games.
- Interactive tales and living books.
- Music notation program and sequencer for the creation of arrangements and pieces.
- Self-learning programs for children and young people.
- Musical Software aimed at children for composing, varying, dismantling and recomposing, improvising.
- Musical software for improving hearing perception, concentration and memory; for getting to know musical code and childhood repertory; for learning through play (sound puzzles, memory…).

9. Problems

It isn't so easy to find software designed to stimulate the child's sensorial and cognitive abilities.
Musical educational software is often designed only as a training activity which dedicates little space to imagination. After searching through hundreds of titles for the most intuitive, creative and immediate, the teacher or operator must experiment different ways of integrating technology with formative activities such as movement, singing games, listening, playing...

10. Projects for the future

In order to obtain documentation and to publicize experiences already carried out EMMSA is in contact with large number of Schools and Educational Institutes. Anybody wishing to send material can contact Stefania Di Blasio: diblasio@musicarte.it – emmsa@musicarte.it
EMMSA is updating its web site where articles, demos, useful links and other material will be found. EMMSA is seeking to adhere to an international network of organizations and projects through the

national Institute of multimedial documentation for Education as INDIRE of Florence.

11. REFERENCES

[1] Bamberger, J. "Developing musical intuitions: A project-based introduction to making and understanding music". New York: Oxford University Press, 2000.

[2] Berz, W. L., & Bowman, J. "Applications of research in music technology". Reston, VA: Music Educators National Conference, 1994.

[3] Brown, A. "Music, media and making: humanizing digital media in music education". Internation Journal of Music Education, 33, 10-17, 1999.

[4] Clements, D. "Teaching creativity with computers". Educational Psychology Review, 7(2), 141-161, 1995.

[5] Dillon, A., & Gabbard, R. "Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style". *Review of Educational Research, 68*(3), 322-349, 1998.

[6] Fletcher-Flinn, C., & R., G. "The efficacy of computer assisted instruction (CAI): A meta-analysis". Jounral of Educational Computing Research, *12*(3), 219-242, 1995

[7] Folkestad, G., Hargreaves, D., & Lindström, B. "Compositional strategies in computer-based music-making". *British Journal of Music Education, 15*(1), 83-97, 1998.

[8] Gardner, H. "The unschooled mind: How children think and how schools should teach". New York: Basic Books, 1991.

[9] Hickey, M. "The computer as a tool in creative music making". *Research Studies in Music Education, 8* (July), 56-70, 1997.

[10] Kassner, K. "One computer can deliver whole-class instruction". *Music Educators Journal, 86*(6), 34-40, 2000.

[11] Liu, M. "The effect of hypermeida authoring on elementary school students' creative thinking". *Journal of Educational Computing and Research, 19*(1), 27-51, 1998.

[12] Lord, C. H. "Harnessing technology to open the mind: Beyond drill and practice for aural skills". *Journal of Music Theory Pedagogy, 7*, 105-117, 1993.

[13] MacGregor, R. C. "Learning theories and the design of music compositional software for the young learner". *International Journal of Music Education, 20*, 18-26, 1992.

[14] Papert, S. "The children's machine: Rethinking school in the age of the computer". New York: Basic Books, 1993.

[15] Reese, S., & Hickey, M. "Internet-based music composition and music teacher education". *Journal of Music Teacher Education*, 25-32, 1999.

[16] Taylor, J., & Deal, J. "Integrating technology into the K-12 music curriculum: A pilot survey of music teachers". In S. Lipscomb (Ed.), *Sixth International Technological Conference on Directions in Music Learning* (pp. 23-27). San Antonio, TX: IMR Press, 1999.

[17] Webster, P. R. "Creative thinking, technology, and music education". *Design for Arts in Education, 91*(5), 35-41, 1990.

[18] Williams, D., & Webster, "Experiencing music technology " (2nd ed.). New York: Schirmer Books, 1999.

ANALYSIS OF EXPRESSIVE GESTURES IN HUMAN MOVEMENT: THE EYESWEB EXPRESSIVE GESTURE PROCESSING LIBRARY

Antonio Camurri, Barbara Mazzarino, Gualtiero Volpe

InfoMus Lab (Laboratorio di Informatica Musicale)
DIST – University of Genova
Viale Causa 13, I-16145 Genova, Italy
http://infomus.dist.unige.it
{toni, bunny, volpe}@infomus.dist.unige.it

ABSTRACT

This paper presents some results of a research work concerning algorithms and computational models for real-time analysis of expressive gestures in human full-body movement. The work has been carried out at the DIST - InfoMus Lab in the framework of the EU IST Project MEGA (Multisensory Expressive Gesture Applications, www.megaproject.org).

The MEGA project is centered on the modeling and communication of expressive and emotional content in non-verbal interaction by multi-sensory interfaces in shared interactive Mixed Reality environments. It focuses on music performance and full-body movements as first class conveyors of expressive and emotional content. Analysis of expressiveness in human gestures can contribute to new paradigms for the design of interactive systems.

As a main concrete result of our research work, we present the EyesWeb Espressive Gesture Processing Library, a collection of software modules for the EyesWeb open software platform (distributed for free at www.eyesweb.org).

1. INTRODUCTION

Our research is focused on the design of interactive systems mainly for theatre and performing art applications, explicitly considering and enabling the communication of expressive, emotional content. Such a research work includes (i) the analysis and classification of expressive gestures in music (audio) and movement (video), (ii) the real-time generation of audio and visual content depending on the output of the analysis, (iii) a study of the interaction mechanisms (mapping strategies) enabling the results of the analysis to be employed (transformed) in automatically generation of audio and visual material.

In this paper we focuses on the first aspect, and in particular we address algorithms and computational models for the extraction of a collection of expressive features from human movement. Since the particular interest in interactive systems for performing art, and since it can be considered as the artistic expression of human movement, dance has been chosen as a particular test-bed for our research.

After introducing the conceptual framework underlying the research work, models and algorithms will be presented with reference to a concrete output of the research process: the EyesWeb Gesture Processing Library, a collection of software modules for the EyesWeb open software platform (distributed for free at www.eyesweb.org)

2. AUTOMATED EXTRACTION OF MOVEMENT CUES: CONCEPTUAL FRAMEWORK

In the design of a multimedia interactive system where expressive information is extracted and used, depending on the required kind of mapping strategy, several expressive cues can be needed from the analysis side, situated at different levels of complexity and carrying different kinds and amounts of information. For example, the generation of a particular output (e.g., a sound, a colored light) can directly depend on low-level motion features (e.g., position of a dancer on the stage, speed of the detected motion), or can be the result of the application of a number of decision rules considering the context, the history of the performance, the information about the classified expressive intention of a dancer (e.g., in term of basic emotions: joy, grief, fear, anger).

In order to extract and provide such a variety of possible expressive cues, a layered approach [1] has been adopted to model expressive gestures, from low-level physical measures (e.g., in the case of human movement, position, speed, acceleration of body parts) toward descriptors of overall (motion) features (e.g., fluency, directness, impulsiveness).

If, from the one hand, such high level descriptors are grounded within the consolidated tradition of biomechanics, on the other hand, they are inspired to studies by psychologists (e.g., [2]) and researchers on human movement coming from the fields of performing arts and humanities, e.g., Rudolf Laban and his Theory of Effort [3][4].

The layered model is also motivated by an integrated multimodal representation of different channels of information (visual, acoustic, etc). That is, a similar approach can be used to analyze audio input, to individuate expressive gestures in audio and to find a

common representation of music and movement gestures.

Further, such a layered approach is suitable for both modeling (that is, generating) movement (for example of avatars, virtual characters, robots in Mixed Reality scenarios) and for recognizing (analyzing) movement qualities.

Figure 1 sketches the layered model: for each layer inputs and outputs are displayed, as well as the kind of employed techniques.

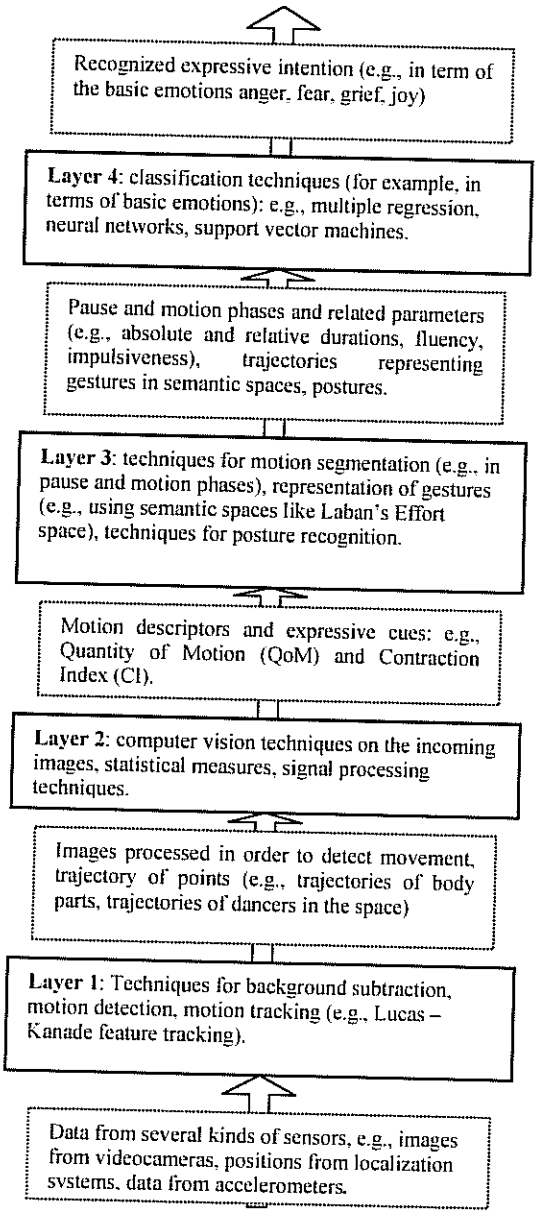


Figure 1: the conceptual framework

Layer 1 and 2 are responsible for the processing of the incoming video frames and for the extraction of a collection of movement cues.

Layer 1 mainly employs consolidated computer vision techniques such as, for example, techniques for background subtraction and motion detection, for real-time analysis and recognition of human motion and activity (see for example the temporal templates technique for representation and recognition of human movement described in [5]), for motion tracking (e.g., Lucas-Kanade feature tracking [6], tracking of colored blobs).

Layer 2 performs cue extraction mainly by means of statistical and signal processing techniques.

Layer 3 deals with gesture segmentation and representation. A possible representation consists in considering gestures as trajectories in semantic, expressive spaces. Such a representation can have a cross-modal valence, since both movement and music gestures can be expressed as a trajectory in properly defined spaces.

Layer 4 collects inputs from Layers 2 and 3 and tries to classify dance fragments. A possible classification is in term of the four basic emotions anger, fear, grief and joy [7]. Another classification can be in term of Laban's basic efforts (e.g., pushing, gliding...). For example, [8] describes the use of neural networks for recognition of the Laban's Effort qualities (direct/indirect, quick/sustained...).

Many techniques are available for this task: statistical methods like multiple regression, neural networks (e.g., classical back-propagation networks, Kohonen networks), support vector machines, methods based on fuzzy sets, decision trees.

3. THE EYESWEB EXPRESSIVE GESTURE PROCESSING LIBRARY

The *EyesWeb Expressive Gesture Processing Library* is the main concrete output of our research work. It includes a collection of blocks (software modules) and patches (interconnections of blocks) contained into three main sub-libraries:

- *The EyesWeb Motion Analysis Library*: a collection of modules for real-time motion tracking and extraction of movement cues from human full-body motion. It mainly covers Layer 1 and 2 in the conceptual framework.
- *The EyesWeb Space Analysis Library*: a collection of modules for analysis of occupation of 2D (real as well as virtual) spaces. If from the one hand this sub-library can be used to extract low-level (Layer 1 and 2) motion cues (e.g., how much time a dancer occupied a given position on the stage), on the other hand it can also be used to carry out analyses in semantic spaces (Layer 3 and 4).
- *The EyesWeb Trajectory Analysis Library*: a collection of modules for extraction of features from trajectories in 2D (real as well as virtual) spaces. Again, this sub-library can be used for analyses situated both at Layer 1 and 2 (trajectories in physical spaces) and at Layer 3 and 4 (trajectories in semantic, expressive spaces).

3.1. The EyesWeb Motion Analysis Library

The EyesWeb Motion Analysis Library applies computer vision, statistical, and signal processing techniques to extract expressive cues from human full-body movement. It mainly covers Layer 1 and 2 in the conceptual framework.

A first task consists in individuating and tracking motion in the incoming images. Firstly, background subtraction is used to segment the body silhouette. Algorithms based on searching for body centroids and on optical flow based techniques (e.g., the Lucas and Kanade tracking algorithm [6]) are available.

Starting from silhouettes and tracking information a collection of expressive parameters is extracted. Three of them are described in the following.

- *Quantity of Motion* (QoM), i.e., the amount of detected movement. It is based on the Silhouette Motion Images. A Silhouette Motion Image (SMI) is an image carrying information about variations of the silhouette shape and position in the last few frames. SMIs are inspired to motion-energy images (MEI) and motion-history images (MHI) [5][9]. They differ from MEIs in the fact that the silhouette in the last (more recent) frame is removed from the output image: in such a way only motion is considered while the current posture is skipped. QoM is computed as the area (i.e., number of pixels) of a SMI. It can be considered as an overall measure of the amount of detected motion, involving velocity and force.

- *Silhouette shape/orientation of body parts*. It is based on an analogy between the image moments and mechanical moments: in this perspective, the three central moments of second order build the components of the inertial tensor of the rotation of the silhouette around its center of gravity: this allows to compute the axes (corresponding to the main inertial axes of the silhouette) of an ellipse that can be considered as an approximation of the silhouette: orientation of the axes is related to the orientation of the body [10].

Figure 2 shows the ellipse calculated on a reference dance fragment.

Figure 2: silhouette shape and orientation.

By applying the extraction of the ellipse to different body parts, other information can be obtained. For example, by considering the main axis of the ellipses associated to the head and to the torso of the dancer, it can be possible to obtain an estimate of the directional changes in face and torso, a cue that psychologists consider important for communicating expressive intention (see for example [11]).

- *Contraction Index*, a measure, ranging from 0 to 1, of how the dancer's body uses the space surrounding it. It is related to Laban's "personal space". It can be calculated in two different ways: (i) considering as contraction index the eccentricity of the ellipse obtained as described above, (ii) using a technique related to the bounding region, i.e., the minimum rectangle surrounding the dancer's body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette. Intuitively, if the limbs are fully stretched and not lying along the body, this component of the CI will be low, while, if the limbs are kept tightly nearby the body, it will be high (near to 1).

The EyesWeb Motion Analysis Library also includes blocks and patches to extract measures related to the temporal dynamics of movement. A main issue is the segmentation of movement in pause and motion phases. A motion phase can be associated to a dance phrase and considered as a gesture. A pause phase can be associated to a posture and considered as a gesture as well. From this point of view, segmentation of movement is strictly related to segmentation and representation of gestures and therefore considered as part of Layer 3. For example, in a related work ([7]) the QoM measure has been used to perform the segmentation between pause and motion phases. In fact, QoM is related to the overall amount of motion and its evolution in time can be seen as a sequence of bell-shaped curves (*motion bells*). In order to segment motion, a list of these motion bells has been extracted and their features (e.g., peak value and duration) computed. Then, an empirical threshold has been defined: the dancer is considered to be moving if the area of the motion image (i.e., the QoM) is greater than 2.5% of the total area of the silhouette.

Several movement cues can be measured after segmenting motion in motion and pause phases: for example, blocks are available for calculating durations of pause and motion phases and inter-onset intervals as the time interval between the beginning of two subsequent motion phases. Furthermore, descriptive statistics of values of extracted cues can be computed on motion phases: for example, it is possible to calculate the sample mean and variance of the QoM during a motion phase.

3.2. The EyesWeb Space Analysis Library

The EyesWeb Space Analysis Library is based on a model considering a collection of discrete potential functions defined on a 2D space [12]. The space is divided into active cells forming a grid. A point moving in the space is considered and tracked. Three main kind of potential functions are considered: (i) potential functions *not* depending on the current position of the tracked point, (ii) potential functions depending on the current position of the tracked point,

(iii) potential functions depending on the definition of regions inside the space.

Objects and subjects in the space can be modeled by time-varying potentials. For example, a point moving in a 2D space (corresponding to a stage) can be associated to a dancer. Objects (such as fixed scenery or lights) can be modeled with potential functions independent from the position of the tracked object: notice that "independent from the position of the tracked object" does not mean time-invariant. The trajectory of a dancer with respect to such a potential function can be studied in order to identify relationships between movement and scenery. The dancer himself can be modeled as a bell-shaped potential moving around the space by using the second kind of potential functions. Interactions between potentials can be used to model interactions between (real or virtual) objects and subjects in the space.

Regions in the space can also be defined. For example, it is possible that some regions exist on a stage in which the presence of movement is more meaningful than in other regions. A certain number of "meaningful" regions (i.e., regions on which a particular focus is placed) can be defined and cues can be measured on them (e.g., how much time a dancer occupied a given region).

This metaphor can be applied both to real spaces (e.g., scenery and actors on a stage, the dancer's General Space as described in [4]) and to virtual, semantic, expressive spaces (e.g., a space of parameters where gestures are represented as trajectories): for example, if, from the one hand, the tracked point is a dancer on a stage, a measure of the time duration along which the dancer was in the scope of a given light can be obtained; on the other hand, if the tracked point represents a position in a semantic, expressive space where regions corresponds to basic emotions, the time duration along which a given emotion has been recognized can also be obtained.

The EyesWeb Space Analysis Library implements the models and includes blocks allowing the definition of interacting discrete potentials on 2D spaces, the definition of regions, the extraction of cues (such as, for example, the occupation rates of regions in the space). For example, Figure 3 shows the occupation rates calculated on a rectangular space divided into 25 cells. The intensity (saturation) of the color for each cell is directly proportional to the occupation rate of the cell. The trajectory of the tracked point is also displayed.

3.3. The EyesWeb Trajectory Analysis Library

The EyesWeb Trajectory Analysis Library contains a collection of blocks and patches for extraction of features from trajectories in 2D (real or virtual) spaces. It complements the EyesWeb Space Analysis Library and it can be used in conjunction with the EyesWeb Motion Analysis Library.

Blocks can deal with lot of trajectories at the same time, for example the trajectories of the body joints (e.g., head, hands, feet) or the trajectories of the points tracked using the Lucas-Kanade feature tracker available in the Motion Analysis sub-library.

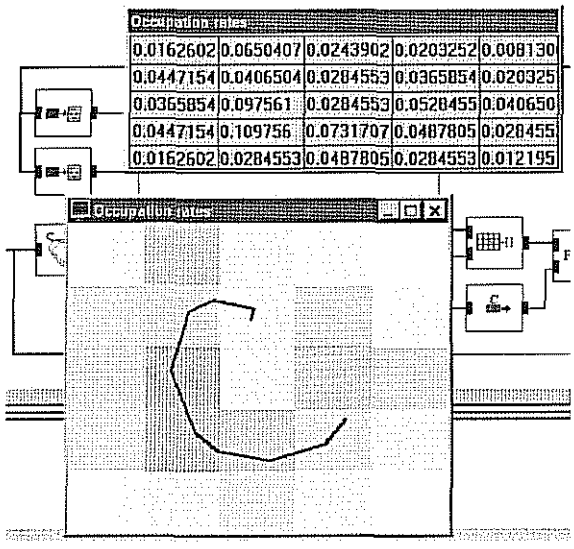


Figure 3: occupation rates

Features that can be extracted include geometric and kinematics measures.

Examples of geometric features are the length of a trajectory, its direction and its *Directness Index*. The Directness Index (DI) is a measure of how much a trajectory is direct or flexible. In the Laban's Theory of Effort it is related to the Space dimension. In the actual implementation the DI is computed as the ratio between the length of the straight line connecting the first and last point of a given trajectory and the sum of the lengths of each segment constituting the given trajectory. Therefore, the more it is near to one, the more direct is the trajectory (i.e., the trajectory is "near" to the straight line).

Figure 4 shows for example a trajectory (in red) and its computed direction (the green segment, whose length is proportional to the overall displacement).

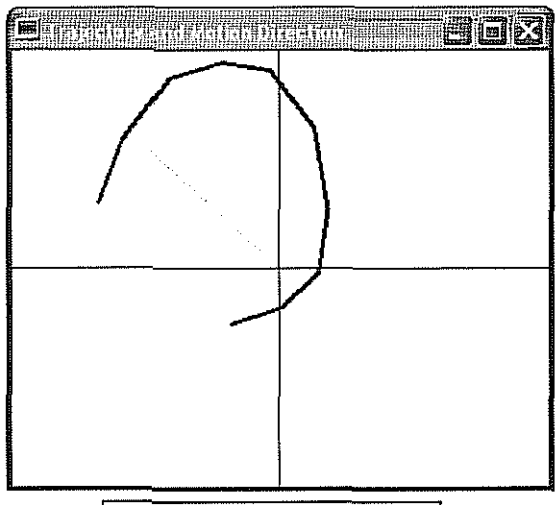


Figure 4: motion direction

The available kinematics measures are velocity, acceleration, and curvature. Their instantaneous values are calculated on each input trajectory. Numeric derivatives can be computed using both the symmetric and the asymmetric backward methods (the user can select the one he prefers). Acceleration is available both in the usual x and y components and in the normal-tangent components.

Descriptive statistic measures can also be computed:

(i) *Along time*: for example, average and peak values calculated either on running windows or on all the samples between two subsequent commands (e.g., the average velocity of the hand of a dancer during a given motion phase)

(ii) *Among trajectories*: for example, average velocity of groups of trajectories available at the same time (e.g., the average instantaneous velocity of all the tracked points located on the arm of a dancer).

As in the case of the EyesWeb Space Analysis Library, trajectories can be real trajectories coming from tracking algorithms in the real world (e.g., the trajectory of the head of a dancer tracked using a tracker included in the EyesWeb Motion Analysis Library) or trajectories in virtual, semantic spaces (e.g., a trajectory representing a gesture in a semantic, expressive space).

The extracted measures can be used as input for clustering algorithms in order to group trajectories having similar features. In the real space this approach can be used to identify points moving in a similar way (e.g., points associated to the same limb in the case of the Lucas-Kanade feature tracker). In a semantic space, it could allow grouping similar gestures, e.g., gestures communicating the same expressive intention.

The EyesWeb Expressive Gesture Processing Library has been employed in a number of artistic events in the framework of the MEGA project (list and description of performances available at the project website `www.megaproject.org`). It consists of a distinct and separate add-on with respect to the EyesWeb software platform and includes the research and development work carried out during the last year.

Novel blocks for the EyesWeb Gesture Processing Library are currently under development, including for example refined motion tracking (Layer 1), extraction of new cues (Layer 2), machine learning techniques for high-level gesture analysis (Layers 3 and 4).

4. ACKNOWLEDGMENTS

We thank Matteo Ricchetti and Riccardo Trocca for discussions and their concrete contributes to this research project. We also thank the other members of the EyesWeb staff.

This work has been partially supported by the EU – IST Project MEGA (Multisensory Expressive Gesture Applications) and by the National CNR Project CNRG0024AF "Metodi di analisi dell'espressività nel movimento umano per applicazioni in Virtual Environment".

5. REFERENCES

[1] Camurri, A., De Poli G., Leman M. "MEGASE - A Multisensory Expressive Gesture Applications System Environment for Artistic Performances", Proc. Intl. Conf. CAST01, GMD, St Augustin-Bonn, pp.59-62, 2001.

[2] Wallbott, H.G., "The measurement of Human Expressions", in Walbunga von Rallfer-Engel, "Aspects of communications", pp. 203-228, 1980.

[3] Laban, R., Lawrence F.C., "Effort", Macdonald & Evans Ltd. London, 1947.

[4] Laban, R., "Modern Educational Dance" Macdonald & Evans Ltd. London, 1963.

[5] Bobick, A.F., Davis J., "The Recognition of Human Movement Using Temporal Templates", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3): 257-267, 2001.

[6] Lucas B., Kanade T., "An iterative image registration technique with an application to stereo vision" in Proceedings of the International Joint Conference on Artificial Intelligence, 1981.

[7] Camurri A., Lagerlöf I., Volpe G., "Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques", International Journal of Human-Computer Studies, Elsevier Science, in press.

[8] Zhao, L., "Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures", Ph.D Dissertation, University of Pennsylvania, 2001

[9] Bradsky G., Davis J., "Motion segmentation and pose recognition with motion history gradients", Machine Vision and Applications 13:174-184, 2002.

[10] Kilian J., "Simple Image Analysis By Moments" OpenCV library documentation, 2001.

[11] Boone, R. T., Cunningham, J. G., "Children's decoding of emotion in expressive body movement: The development of cue attunement" Developmental Psychology, 34, 1007-1016, 1998.

[12] Camurri A., Mazzarino B., Trocca R., Volpe G. "Real-Time Analysis of Expressive Cues in Human Movement." Proc. Intl. Conf. CAST01, GMD, St Augustin-Bonn, pp. 63-68, 2001.

THE EXPRESSIVE FUNCTIONING OF TEMPO AND DYNAMICS IN THREE PERFORMANCES OF A SKRIABIN ETUDE

Renee Timmers, Antonio Camurri, & Gualtiero Volpe

Infomus, DIST, University of Genova
renee74@xs4all.nl, music@dist.unige.it,
volpe@infomus.dist.unige.it

ABSTRACT

The way a performer phrases music is important for the way the music is perceived by listeners. Phrasing influences the structural interpretation of the music by listeners as well as his or her emotional engagement with the music; at least, those are the hypotheses that this study likes to confirm. An experiment was run in which twelve listeners heard three performances of an Etude of Skriabin. They pressed a button to indicate the phrase-boundaries of the music and moved a slider to indicate their emotional engagement with it. Two main expressive parameters for a pianist were measured, tempo and key-velocity, and related to the responses of the listeners. It was found that tempo correlated with the indication of phrase boundaries, while key-velocity was especially related to the listeners' emotional engagement. Further examination showed that both aspects can actually be related to phrasing, though tempo to local phrasing and dynamics to global phrasing. The pianist used dynamics to express the music's overall form and to modulate the tension of the music throughout the entire piece. The listeners responded emotionally to this.

1. INTRODUCTION

Three findings from previous research are especially relevant for this study. The first is that expressive variations in, for example, tempo and dynamics are often related to the musician's structural interpretation of the music [1], [2]. The second is that these variations are also often related to the expression of a certain emotion by the performer [3], [4]. And the third is that these aspects of a performance influence the structural or emotional interpretation of the music by listeners depending on which interpretation is asked for [2], [3], [4].

This study aims to continue and reconcile these findings and hypothesizes that it is the musician's phrasing that influences simultaneously the listener's structural and emotional interpretation. The way a musician phrases the music influences the attention of the listener and so modulates to what extent the listener is emotionally engaged. It is especially when a listener gets the idea of a larger phrasing, a transgression of the local fixation, that he is emotionally moved. As shown in previous literature [5], [6], the pianist especially uses two expressive variables, tempo and dynamics, to express the phrase-structure of the music. Accordingly, he also uses these to move the listener.

2. METHOD

2.1 Musical Performances

An explorative experiment was run to investigate the hypotheses. A professional pianist was asked to perform an emotionally engaging piece of his choice at a concert that was organised for the experiment's purpose. He performed the piece first without public in a normal manner (to be referred to as p1) and an exaggerated manner (to be referred to as p2) and then performed the piece with public in a normal, concert manner (to be referred to as p3). He performed on a Yamaha Disklavier, which made it possible to register MIDI information of the performance. In addition, audio recordings were made and presented to the participants of the experiment.

The pianist chose to perform Etude Op. 8 no. 11 by Alexander Skriabin, which is a slow and lyrical piece (Andante cantabile) in a late Romantic style that has a considerable amount of modulations. According to the pianist, the piece can be played with a lot of freedom. Theoretically, the piece has a simple A B A with coda structure (A A' B A'' A''' C to be more precise), but the pianist interpreted the line of the music differently: The first main target of the music is a release of tension halfway the B section. Everything preceding this target point is a preparation for this tension release. The A section is anyway preparatory; it leads towards the start of the B section, which is the real beginning of the piece. After this release of tension, the music builds up towards the dramatic return of the theme of the A section. This prepares for the second possible point of tension release halfway the coda at a general pause. The release is however not continued and the piece ends most sad.

2.2 Participants

Twelve people participated in the experiment among them were four musicians. The participants varied greatly in musical experience. Some of them never had had music lessons and hardly listened to classical music, while others basically performed classical music already their entire live.

2.3 Procedure

The participants sat behind a desk with a slider and a joystick before them. They heard the three performances of the Skriabin Etude twice in random order. The first time they heard the music, they indicated the phrase boundaries in the music by pressing the button of the joystick. The second time they heard the music, they indicated to what extent they were emotionally engaged with it by moving a MIDI-slider up and down. The whole procedure was explained to them by a written instruction and a practice trial.

3. RESULTS

3.1 Performance Data

The key-velocity and onset-times of notes were extracted from the MIDI files. From this, the average key-velocity for each quarter note, which roughly corresponds to the dynamics of the performance, was calculated as well as inter-onset-intervals (IOI's) between successive quarter notes, which is a measure for local duration. The quarter note was taken as unit, because it gives both sufficiently detailed information about the performances and sufficient consistency between listener response data for which synchronisation is an issue (see below).

The resulting profiles of quarter note key-velocity and quarter note IOI are plotted in Figure 1, top panels. Separate graphs are plotted for p1, p2 and p3. Vertical dotted lines indicate section boundaries. Bar numbers are given at the bottom. The profiles were highly similar for the three performances: they all started in a slow tempo and with soft dynamics, had considerable crescendi and accelerandi in the A section, a diminuendo and crescendo in the B section accompanied by first a highly variable tempo and thereafter an accelerando, a fast and loud return of the A section with limited variation in tempo and dynamics, a soft and slower repeat of the theme, and a coda that fades away in dynamics and tempo (see Figure 1 top panels).

In addition to this global pattern, the IOI-profile shows the characteristic peaks of phrase-final lengthenings. It shows this at a fairly high density and large magnitude. There is no large-scale hierarchy in the phrase-final lengthenings that has larger legnthenings at major boundaries and short lengthenings at minor boundaries. Instead rubato is quite steep throughout the piece, except in the forte return of the A section (A''). The key-velocity profile shows drops in velocity at most phrase-boundaries, though these are balanced by strong crescendi in most sections.

This is the basic pattern of all three performances. Differences between them are that performance 2 is clearly an exaggerated version of the other performances: all variations in tempo and dynamics are larger. Performance 1 is the more modest version of the three, while performance 3 is in between. Performance 1 and 3 further contain a crescendo and decresendo in the coda that performance 2 lacks.

3.2 Listeners' Data

The indication of phrase-boundaries was measured at a sampling rate of 10 Hz. The measure was 0 when the participant did not press the button and 1 if the he or she pressed the button to indicate a phrase boundary. The data was filtered to be 1 only at the onset-time of a phrase-boundary indication. For the rest of the time, the measure was put to 0.

For each quarter note in the performance, the number of people who indicated a phrase-boundary was calculated by summing the number of boundary indications per quarter note over participants. The resulting "segmentation measure" was expressed as a multiple of chance-level. Chance-level was defined as the number of boundary indications per quarter note if the total of boundary indications would have been equally distributed over all quarter notes of the piece. From this recalculation, the quarter-note level turned out to be the most statistically reliable unit for this measure (better than the 8th-note or half-bar level for example). It should be noted that some kind of data-reduction was necessary, since the participants would never react exactly at the same time and could have been indicating the same boundary, even if onset-times of the key-press differed 1 or 2 seconds.

The indication of emotional engagement was also measured at a sampling rate of 10 Hz using a MIDI-slider that had a range from 0 to 127. The average level of the MIDI-slider ("emotion measure") per quarter note was calculated for each participant separately and averaged over participants.

The first result from these measures is that participants disagreed quite strongly, though the averages per performance are statistically reliable (see bottom panels of Figure1). On average the indication of section boundaries was 1.6 times chance level, which is significantly above chance, but not very much. The indication of emotional engagement was diverse between subjects to such extent that no significant differences can be found between the three performances: the standard deviation of the difference between the emotion measure of two performances is larger than the average difference itself. The average profile within pieces is statistically reliable: the standard deviation per quarter note is relatively small with respect to the average per quarter note (on average 1:2.7, 1:2.2 and 1:2.4 for performance 1, 2 and 3 respectively).

The second and third results are the profiles themselves: The segmentation measure shows most boundary indications at the start and in the B section. It shows a considerable drop in the number of boundary indications in the return of the theme. Highest peaks are between repeats of the theme in the first and second A section and at two places that the pianist mentioned explicitly: a general pause halfway the coda and a point of release of tension halfway the B section (see downwards pointing arrows in Figure 1).

The emotion measure increases towards the B section. It decreases and increases within the B section, reaches a maximum at the return of the theme, and decreases at the repeat of the theme and in the coda. For performances 1 and 3 it shows a local revival in the coda.

3.3 Relation Performance and Listeners' Data

The quarter note IOI and key-velocity measures were correlated with the emotion and segmentation measures per quarter note. This was done directly and with a time-delay of one, two and three quarter notes of the performance data with respect to the listeners' data. The best correlations were obtained between the performance data and the segmentation data if the performance data was not delayed or delayed for only one quarter note. For the emotion data, however, the best correlations were achieved if the performance data was delayed by three quarter notes. The optimal correlations are shown in Table 1.

Generally, the emotion measure was highly correlated with key-velocity and much less negatively correlated with duration. The segmentation measure was on the other hand more highly correlated with duration than (negatively) with key-velocity. A direct correlation between the segmentation measure and the emotion measure at the quarter note level was not significant. The number of boundary indications per section did however highly correlate negatively with the average indication of emotional engagement per section.

Table 1: Correlations at the quarter note level (column 2 and 3) and at the level of sections (last column).

Performance 1

	IOI	Velocity	Segm. M
Velocity	-,50		
Segm. M	,50	-,34	
Emotion M	-,40	,74	-,90

Performance 2

	IOI	Velocity	Segm. M
Velocity	-,54		
Segm. M	,55	-,34	
Emotion M	-,49	,77	-,88

Performance 3

	IOI	Velocity	Segm. M
Velocity	-,49		
Segm. M	,47	-,34	
Emotion M	-,43	,75	-,80

It should not be taken as a surprise that the correlations with the segmentation measure are relatively lower than with the emotion measure, since the segmentation measure still contains the between participant variability, while the emotion measure is a pure average. In addition, if we use a multiple regression analyses that takes key-velocity and IOI as independent variables and segmentation measure or emotion measure as independent variables, the contributions of velocity is reduced to non-significant for the segmentation measure, while the contribution of IOI also cancels out for the emotion measure. Duration becomes the main predictor of segmentation and key-velocity of emotion.

3.4 Relation Phrasing and Emotion

The correlations presented in Table 1 and the multiple regression analyses suggest that, for these performances, tempo was especially a cue for phrase-boundaries, while dynamics was especially a strong cue for the intensity of emotion. This may in turn suggest that phrasing did not play an important role for the emotional engagement of listeners. Still there seems a relation between phrasing and emotion, since the correlation between the number of phrase boundary indications per section and the average emotional engagement per section was very high. How should we interpret these results?

From the descriptions of the performance data, it became clear that phrase-final lengthenings occurred at a high density, which means that they indicated rather local phrase-boundaries, and that there was limited hierarchy within these lengthenings. It became also clear that the dynamics and the tempo profiles showed an overall pattern of increase and decrease that was especially apparent in the key-velocity profile, but masked by the large phrase-final lengthenings in the profile of the quarter note IOI. The interpretation is that the pianist used both dynamics and tempo to phrase the music, though tempo was especially used to locally phrase music, while dynamics was more saliently used to express the larger form of the music. The listeners responded emotionally to this expression of the larger form.

And the overall form is as described by the pianist: the first part is an introduction and builds up to the B section, which he considered as the real beginning of the piece. This beginning is again a preparation for the first target of the piece: the release of tension at the middle of the B section. Hereafter tension builds up towards the dramatic return of the theme, which leads via a repeat of the theme in contrasting dynamics to the second important target of the theme: the second possible release of tension at the general pause. After the general pause, the release is not given and all hope is lost. The piece ends most sad. The pianist most skilfully expresses this interpretation in the patterning of dynamics (see arrows in the key-velocity panel of Figure 1). The resulting phrasing is over the entire piece with subdivisions at measures 22 and 36. The return of the theme is the culminating point of the piece where after tension can release. According to the pianist, this tension cannot however be fully resolved.

4. SUMMARY

This study has investigated the expressive functioning of two acoustic cues, tempo and dynamics, in three performances of an etude of Skriabin and has shown that both tempo and dynamics were used to express the phrasing of the music, though tempo expressed local phrasing and dynamics the larger form. The local phrasing was especially reflected in the segmentation indications of the listeners. The global phrasing was reflected in the ratings of emotional engagement of the listeners. This followed the pattern outlined by the dynamics and reached a maximum at the turning point of the piece and local minima at section divisions.

5. DISCUSSION

This paper has investigated the relation between experience of emotional intensity and expressive features of a performance. It did not consider other factors that could have contributed to the experience such as melodic movement, harmony or rhythm. In a follow up study such a comparison is planned. For now it suffices to see the extent to which these two expressive relate to emotional experience. We see promising evidence in this study that the way of phrasing influences the structural interpretation of the music by listeners, but moreover also the emotional experience of it. A relation between tempo and phrasing and dynamics and emotion has also been shown in previous studies and this study provides a confirmation of these findings [5], [6], [7], [8]. However, this study has also given insight into the simultaneouty of these relations and it has demonstrated how skilfully the pianist modulated these. This skilfullnes is probably crucial for the emotional experience. An aspect that should be investigated further in future research. Just a crescendo or *diminuendo without* proper preparation would not have given the strongly emotional responses that this study got. Although the participants varied strongly in responses, they did mention to be highly moved and were made sure to give responses according to their feelings.

6. REFERENCES

[1] Palmer, C. (1997). Music Performance, Annual Review of Psychology, 48, 115-138.

[2] Sloboda, J. A. (1983). The communication of musical metre in piano performance. Quarterly Journal of Experimental Psychology, 35 (A), 377-396.

[3] Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. Psychology of Music, 24 (1), 68-91.

[4] Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. JEP: Human Perception and Performance, 26, 1797-813.

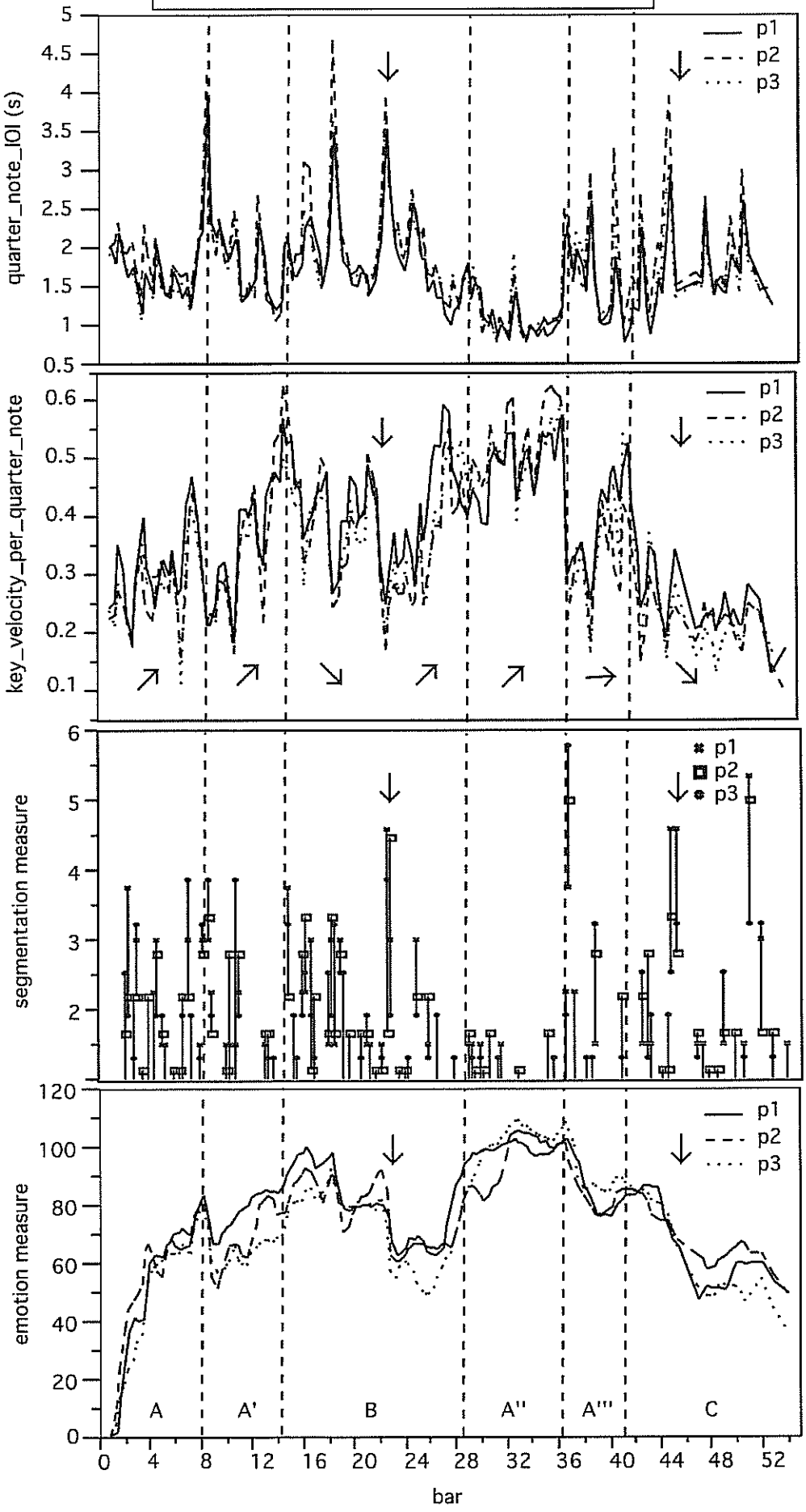
[5] Todd, N. P. (1992). The dynamics of dynamics: a model of musical expression. Journal of the Acoustical Society of America, 91 (6), 3540-3550.

[6] Clarke, E., & Windsor, L. W. (2000). Real and *Simulated Expression*: A listening study. Music Perception, 17, 277-314.

[7] Madsen, C. K. (1996). Empirical investigations of the aesthetic response to music: Musicians and non-musicians. In B. Pennycook & E. Costa-Giomi (eds), *Proc. of the Fourth Int. Conf. of* Music Perception and Cognition, pp. 103-110. Montreal, McGill University.

[8] Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K.282-Segmentation, tension and musical ideas. *Music Perception*, 13, 401-32.

Figure 1: Graphs of performance and listeners' data



EXPRESSIVE CLASSIFIERS AT CSC: AN OVERVIEW OF THE MAIN RESEARCH STREAMS

Sergio Canazza, Giovanni De Poli, Luca Mion, Antonio Rodà, Alvise Vidolin, Patrick Zanon

CSC - DEI, University of Padua
Via Gradenigo 6/a, 35131 Padova, Italy
{canazza,depoli,randy,vidolin,patrick}@dei.unipd.it,ar@csc.unipd.it
http://www.dei.unipd.it/ricerca/csc/

ABSTRACT

Music can be a communication mean between performer and listener. Several studies demonstrated how different expressive intentions can be conveyed by a musical performance and correctly recognized by the listener. Some models for the synthesis can be found in the literature. In this paper we describe three automatic expressive analysis methods based on studies made at CSC during the last year. A brief overview of their implementation is presented. Finally, some results of the validation are sketched.

Keywords: Machine Recognition of Music, Music Analysis, Psychoacoustics, Perception, Cognition, Real-time Systems, and Studio Report.

1. INTRODUCTION

Playing music is a complex task that requires to professional pianists high motor control skills and fine cognitive abilities [1]. The former allows the performer to act on the instrument with high precision and the latter is used as a feedback to tune and correct the movements during the piece. Together these two aspects allow him to communicate his interpretation of the piece. On the other side, listeners use their cognitive capacity to understand what the performer is communicating with his performance. It has proven that music can be played in different ways in order to communicate the structural interpretation of the piece, tensions [2] and expressive content [3]. There is a general agreement among performers and audience on this kind of communication [4]. Starting form these observations, several models to synthesize expressive performances has been developed [5], [6]. These works allowed some studies on the analysis side using the "analysis through the synthesis" approach [7] in which the synthesis models were used to understand the expressive content of a musical performance.

The expressive analysis of musical performances can be used for the realization of the Automatic Content Processing (ACP) that is essential in the today's rapidly evolving panorama of the Internet exchange of multimedia. In fact, Musical Information Retrieval (MIR) is an active research field since the techniques developed for the indexing of the textual information are inappropriate for the multimedia. ACP can also be used according to the MPEG 7 standard for the content description of multimedia products. Several projects where developed with this aim. For example, CUIDADO [8] is aiming to provide a sound palette engine and a music retrieval system as well; they are flexible and they can adapt their retrieval capabilities according to the users' preferences. Other projects were developed outside Europe, for example the Machine listening project of Hashimoto [9], [10], in which a computer is continuously listening to pieces and training itself to recognize some authors or styles; moreover, a computer is also searching in internet for new timbre sounds and classifying them using machine learning techniques. Research is currently developing on the automated structural analysis by Dannenberg [11]: a computer was instructed to recognize the sections' structure of audio data. Machine learning techniques are used to recognize the performer's style [12] at ÖFAI, and by Dannenberg at al. to classify the musical styles [13]. Friberg and Bresin [14] work on the audio cues extraction and classification for the expressive content. Their work is the most similar according to our approach; the main difference deals with the kind of data used: MIDI data in our case, instead of Audio data.

In this paper we will deal with the automatic expressive analysis tool of musical performances developed at CSC in the last year. These methods give an insight on this complex task and can provide some ideas on future works.

The paper is organized as follow: in the next section a brief description of the method used to derive the analysis algorithms is presented with a brief description of each of them. Then a more detailed description of the implementation of the algorithm is presented. The third section is devoted to some validation.

2. METHODS

At CSC several experiments were conducted, both on well-known pieces, and on improvisations. The data were recorded into MIDI files, then several perceptual and acoustical analysis were carried out in order to extract models.

In the former approach, experiments on known pieces yielded a synthesis model [5] that is able to synthesize a performance conveying an expressive intention by transforming a *neutral* one (i.e. a literal human performance of the score without any expressive intention or stylistic choice), both with reference to the score. The model uses the results of several perceptual tests and acoustic analysis [15] and acts on the score by adding some micro deviations in the acoustical parameters (tempo, legato and loudness) that a performer usually intro-duces while playing. This is made by a transformation of the values already present in the neutral performance by means of two sets of coefficients named *K-coefficients* and *M-coefficients*: a K-coefficient changes the average values of an acoustic quantity (for example, the tempo), and the respective M-coefficient is used to scale the deviations of the actual values of the same parameter from the average. In this way, each expressive intention can be represented by a set of 6 parameters. The

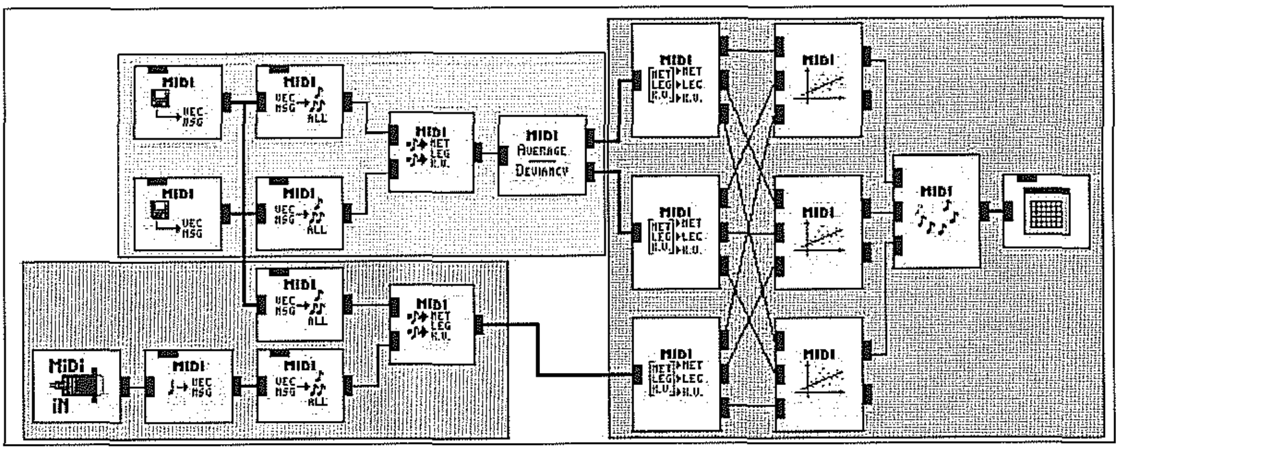


Figure 1: The three parts composing the Expressive Analyzer: the Neutral Analyzer (upper left rectangle), the Real Time Analyzer (lower left rectangle) and the Expressiveness Recognizer (right rectangle).

model was implemented in an EyesWeb block called "Expressive Sequencer" (ES). The ES model was then reversed in order to extract the expressive content from a given performance. This approach requires the knowledge of the score: the analysis process is based on the comparison between the values of Key Velocity, Tempo and Legato of each note of a pre-recorded neutral performance with the values of the same quantities of an expressive performance, played in real time.

To overcome to the limitation imposed by the knowledge of the score, a simple patch developed with EyesWeb were developed, using some results of the previous study [15] on the main relevant sonological parameters. The patch tries to label in real time the expressive content of an expressive improvisation using simple statistical analysis of the relevant perceptual sonologic parameters (notes per second, loudness and legato). This work suggested us to use more sophisticated statistical techniques tested on a new set of performances in order to analyze real time improvisations. A set of improvisations were then recorded and analyzed to infer a Bayesian network able to give as output the probability that the input performance is played according an expressive intention.

3. THE IMPLEMENTATION

3.1. The Expressive Analyzer

We made the implementation of the analysis model using EyesWeb, a product of the Music and Informatics Lab of the University of Genoa [16]. It turned out to be a graphical environment planned for the development of projects based on the analysis and processing of multimedia data streams and the creation of audio/video interactive applications. The graphical interface is quite similar to other software-environment like PD or Max, in which the user can build up an application using a library of blocks that can be connected together into a patch. In our work we developed a number of new blocks that implement dedicated functions and finally we connected them in a patch for the real time analysis of a performance.

Figure 1 shows the patch that realizes analysis using the score knowledge (Expressive Analyzer). The system can be divided into three different parts: the *Neutral Analyzer*, the *Real Time Analyzer* and the *Expressiveness Recognizer*.

The Neutral Analyzer sub-patch is the set of modules located in the upper left rectangle of the patch in figure 1. It is dedicated to the computation of the profiles of the sonological parameters (tempo, articulation and intensity) related to the neutral performance: the two leftmost blocks read the data from two MIDI files (the score and the neutral performance respectively) and a filtering of the non-interesting MIDI event is applied; then the two data streams reach a module called *ComputeLaws* (2 ins 1 out). In this module, the tempo, the legato and the key velocity of the neutral performance are computed note-by-note. The block aligns the input streams using two internal buffers, in which the MIDI events can be cumulated until two MIDI events with the same note number are found on both the input. Then, the output is sent to the last module of this sub-patch, the so-called *ComputeAverage* (1 in 2 outs), which calculates the average for each of the three sonologic parameters over a sliding window of a fixed amount of time. The two outputs of this module are the *expressive profiles*: the former data stream contains the average values of the sonologic parameters; the latter contains the values obtained from the difference between the input and the average.

The Real Time Analyzer sub-patch is the set of modules that lies below the Neutral Analyzer in figure 1. This block achieves the same computations to those accomplished by the previous sub-patch. In this case however, the computations of the expressive quantities are relative to the expressive performance that is received as a real time stream of MIDI events. Also in this case the ComputeLaws block take care of the correct synchronization between the pre-recorded score and the live performance. No average is computed in this case.

The final sub-patch, on the right of the picture, is the Expressiveness Recognizer. It performs two separate tasks: the computation of the K and M-coefficients and the expressive labelling. The K and M-coefficients are computed for each of the three expressive quantities considered, using a multiple linear regression. Then, they are compared with some predefined sets that are specific for each expressive intention. The closest set is selected on the base of a weighted Euclidean distance.

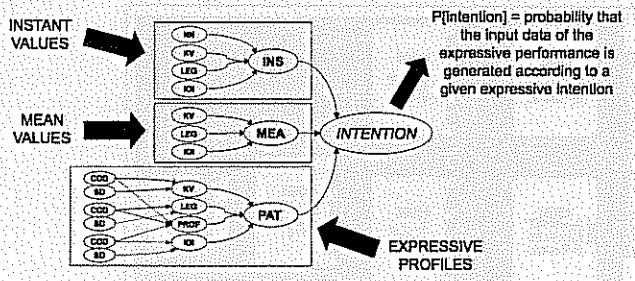


Figure 2: The Bayesian network structure. The three main categories of cues used are evidenced in the dotted rectangles: the instant values, the mean values, and the pattern. Marginal nodes (on the left) are inferred firstly. Each node conditions its descendant nodes until node intention, which value is the probability that input performance is played with a given expressive intention.

3.2. CSC Descriptor of Content (CDC)

The patch called also CDC realizes the second analysis approach. It is based on the statistical analysis of the main relevant perceptual sonologic parameters. This method overcomes to the requirement of the score knowledge that is here not necessary. The patch uses as input an expressive performance codified into a MIDI format, and starts calculating the intensity, the value of legato, and the number of notes per second. The values are then compared with suitable thresholds and the results of the comparison are finally combined in a case based style to select the guessed label describing the expressive content of the input improvisation. The values of the thresholds were tuned by hand. This simple but quite robust work suggested us to use more sophisticated statistical techniques in order to analyze real time improvisations.

3.3. Bayesian Analyzer

The general structure of the patches realizing the Bayesian networks is shown in figure 2: it is our third analysis approach. It is based on the statistical analysis of the relevant perceptual sonologic parameters carried out by a suitable Bayesian network. As in the previous method, the score knowledge is not necessary.

The analyses we made were: a factor analysis of the results of perceptual tests, a mean values' analysis, a factor analysis on expressive profiles and a note-by-note analysis.

The analyses evidenced three categories of possible cues to be used: *instant values*, *mean values*, and *patterns*. Thus, the network was accordingly organized in three main parts, as depicted in figure 2: in the first one, the instantaneous values of the sonologic parameters (NN, KV, LEG and IOI) are evaluated as the note-by-note analysis suggested; in the second part, the mean values (KV, LEG and IOI) are taken into account as ANOVA and the first factor analysis indicated; finally, the third part is devoted to evaluate if in the input data there is some relevant pattern that can be used to recognize the intention (factor analysis of the profiles). Three of the four nodes (KV, LEG and IOI) compare the codifications and standard deviations of tempo, intensity and articulation with values given by analysis on performances, while node PROF yields the comparison among the parameters' profiles. The three analyses are collected together to provide as output the probability that the performance in the input is played with the given expressive intention. Thus graph is built using 21 nodes as shown in figure 2. This analysis methodology needs a network for each expressive intention. After the definition of the topology, we specified the size and type of each node assuming that all nodes are discrete and binary, and the arcs specify the independence assumptions that must hold between the random variables. The task consists in computing the probability that the node intention is true, supposed as known the whole set of roots' values. Several criteria were used to estimate these probabilities. According to perceptual test and factor analysis, the probabilities of the nodes should be proportional to listeners' choices in perceptual test, which has been carried out [17] on several piano improvisation inspired by a set of adjectives. Its results have been related to rhythmic events evaluated on a factor analysis and on a mean values' analysis. Thus, the whole set of hypothesis on adjectives' character allows the construction of a model based on a Bayesian network, to relate all the discovered relationships and give a top down approach to intentions' recognition. In fact, as strong is perceived the relation between a parameter value and an expressive intention, as high is the probability that flows down to set true the node intention.

Intention	Pianist A	Pianist B	Pianist C	Pianist D
Light	52%	45%	75%	29%
Hard	54%	43%	33%	45%
Heavy	69%	78%	27%	86%
Soft	53%	61%	68%	89%

Table 1: Percentage of correct classified notes for each piece and performer. The performances with less than 50% of correct classification are evidenced.

4. VALIDATION

4.1. Expressive Analyzer

The model has been tested by means of four professional pianists. They were asked to play the same musical piece according to five different expressive intentions, even considering a first neutral (without expression).

The musical piece chosen for the test was an excerpt from W. A. Mozart's sonata K.545.

The score of the piece was previously codified into a MIDI file. Then each of the four musicians performed and recorded into another MIDI file his neutral version of the piece (according to his personal idea of neutral). Finally, each one of the four pianists performed the piece according to four different expressive intentions: hard, soft, heavy and light. After the tuning process, the system was tested on the same data used for the tuning and the results are presented in table 1. The table shows the percentage of correct classified notes for each piece and performer. The performances with less than 50% of correct classification are evidenced with a grey background.

The results were in general quite good, even if there is a certain degree of misunderstanding for some pianists (C and D) and for some expressive intentions (Light and Hard).

For the pianists B and C, a more detailed insight is given in figure 3 and figure 4. In the first one, it is possible to see that the recognition of the Heavy performance is very good, even if there is a number of notes that were classified as Hard. In the second figure, the confusion was made among Light and Soft performances. Also in this case, the correct recognition percentage is quite high.

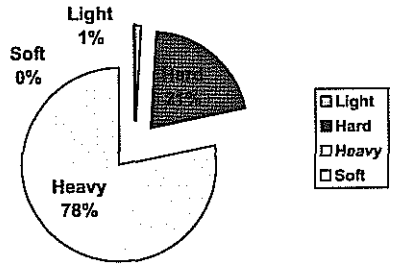


Figure 3: Recognition graph for the heavy performance played by the pianist B. 78% of the event is correctly recognized.

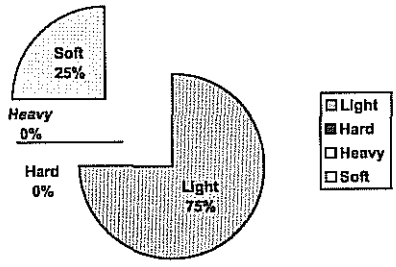


Figure 4: Recognition graph for the light performance played by the pianist C. 75% of the event is correctly recognized.

4.2. Bayesian Analyzer

The networks were trained with a set of 8 different kind of improvisations: 1. (SLA) slashing; 2. (HEA) heavy; 3. (HOP) hopping; 4. (VAC) vacuous; 5. (BOL) bold; 6. (HOL) hollow; 7. (FLU) fluid; 8. (TEN) tender. Thus, we performed 64 simulations (8 networks subjected to 8 performances), in which we computed the mean output probability; the table 2 shows the results.

The computation has been carried out by a Matlab Toolbox for Bayesian Networks [18], which yields the probability of node intention. Most of the networks give higher probability (gray cells) when they are subjected to the data used for their training. Hollow and Fluid networks give the highest value of probability to the corresponding performances; Vacuous and Tender networks confuse the input data, giving higher probability to other intentions (circled cells). Anyway, Tender network recognize the correspondent performance as second (circled beside). We noticed that intentions confused in perceptual test (as mean and factor analysis revealed) caused difficulties in assigning decision rules in their networks; thus, those intentions are recognized with difficulty.

5. CONCLUSIONS

An overview of the main automatic expressive analysis method was presented. The methods present a satisfactory behavior if used with a suitable tuning of the parameters. However, the validation methodology has to be improved by separating the training examples and the testing ones. Some generalities in the models can be underlined, since the second model was simply obtained from the studies that yielded the first. Further improvement are in the agenda: the use of Hidden Markov Models for the recognition of the timing patterns, and the analysis of the asynchronicity between different melodic lines.

Input Data

Networks	SLA	HEA	HOP	VAC	BOL	HOL	FLU	TEN
SLA	0,70	0,46	0,64	0,49	0,54	0,35	0,52	0,48
HEA	0,36	0,77	0,21	0,25	0,34	0,13	0,21	0,28
HOP	0,68	0,76	0,77	0,48	0,67	0,50	0,72	0,67
VAC	0,34	0,20	0,27	0,41	0,21	0,48	0,40	0,54
BOL	0,63	0,52	0,58	0,08	0,69	0,39	0,61	0,31
HOL	0,35	0,48	0,25	0,69	0,36	0,89	0,29	0,25
FLU	0,59	0,39	0,58	0,42	0,57	0,49	0,83	0,59
TEN	0,44	0,49	0,37	0,26	0,40	0,29	0,56	0,52

Table 2: Mean probabilities given by Bayesian networks using the eight performances as input: the cell (x, y) contains the mean P value given by network x subjected to performance y; the values on the diagonal represents the mean values given by networks subjected to the performances they were trained with.

All the material can be freely downloaded and tested from the CSC's web site [19].

6. ACKNOWLEDGEMENTS

This research was supported by the EC project IST-1999-20410 MEGA.

7. REFERENCES

[1] Palmer, C., "Music Performance", Annual Review Psychology, 48: 115-38, 1997.

[2] Krumhansl, C., "A Perceptual Analysis of Mozart's Piano Sonata K. 282: Segmentation, Tension and Musical Ideas", Music Perception, 13(3):401-432, 1996.

[3] Gabrielsson, A., "Music Performance. The psychology of music." In D. Deutsch (ed.) The psychology of Music, 2nd. ed. New York: Academic Press, 1997.

[4] Canazza, S., De Poli, G., Rodà, A., Vidolin, A., "An abstract control space for communication of sensory expressive intentions in music performance", Journal of the New Music Research, 2002 (accepted for publication).

[5] Canazza, S., De Poli, G., Drioli, C., Rodà, A. and Vidolin, A., "Audio Morphing Different Expressive Intentions for Multimedia Systems" IEEE Multimedia, 7(3): 79-83, 2000.

[6] Friberg, A., Frydn, L., Bodin, L. and Sundberg, J., "Performance Rules for Computer-Controlled Contemporary Keyboard Music." Computer Music Journal, 15(2): 49-55, 1991.

[7] Canazza, S., De Poli, G., Rodà, A., Soleni, G. Zanon, P. "Real time analysis of expressive contents in piano performances", Proc. International Computer Music Conference, Göteborg, pp. 414-418, 2002.

[8] Vinet, H., Herrera, P., Pachet, F. "The CUIDADO Project: New Applications based on Audio and Music Content Description", Proc. International Computer Music Conference, Göteborg, pp. 450-454, 2002.

[9] Suzuki, K., Taki, Y., Konagaya, H., Hartono, P., Hashimoto, S., "Machine Listening for Autonomous Musical Performance Style", Proc. International Computer Music Conference, Göteborg, pp. 61-64, 2002.

[10] Hartono, P., Suzuki, K., Qi, H., H., Hashimoto, S., "Subjective Preference Oriented Global Sound Database", Proc. International Computer Music Conference, Göteborg, pp. 446-449, 2002.

[11] Dannemberg, R., "Listening to 'Naima': An Automated Structural Analysis of Music from recorded Audio", Proc. International Computer Music Conference, Göteborg, pp. 28-34, 2002.

[12] Widmer, G., "Using AI and Machine Learning to Study Expressive Music Performance: Project Survey and First Report." AI Communications, 14(3), 149-162, 2001.

[13] Dannenberg, R., Thom, B., Watson, D., "A Machine Learning Approach to Musical Style Recognition", in Proc. International Computer Music Conference, San Francisco, pp. 344-347, 1997.

[14] Friberg, A., Schoonderwaldt, E., Juslin, P., Bresin, R. "Automatic Real-Time Extraction of Musical Expression", Proc. International Computer Music Conference, Göteborg, pp. 365-367, 2002.

[15] De Poli, G., Rodà, A. and Vidolin, A., "Note-by-note Analysis of the Influence of Expressive Intentions and Musical Structure in Violin Performance", Journal of New Music Research, 27(3): 293-321, 1998.

[16] Camurri A., Coletta P., Peri M., Ricchetti M., Ricci A., Trocca R., Volpe G. (2000). "A real-time platform for interactive performance", Proc. ICMC-2000, Berlin, pp. 374-379.

[17] Bonini, F., Rodà, A. "Expressive content analysis of musical gesture: an experiment on piano improvisation", Workshop on Current Research Directions in Computer Music, Barcelona, 2001.

[18] Murphy, K., 2001. "The Bayes Net Toolbox for Matlab", in Computing Science and Staistics, vol. 33.

[19] CSC's home page: `http://www.dei.unipd.it/ricerca/csc/`

EXPRESSIVENESS ANALYSIS OF VIRTUAL SOUND MOVEMENTS AND ITS MUSICAL APPLICATIONS

Amalia de Götzen

CSC-DEI, Università di Padova
corvo@dei.unipd.it

ABSTRACT

This paper describes a work which investigates sound motion expressiveness. This work is divided in three parts: the design of a perceptive test, the statistical analysis of the collected data and then a musical application. The main idea is to find a way to use sound motion like a musical parameter to convey a specific expressive content related to performance gestures. The purposes of this work can be stated as follow:

- To show how multi-modality can be conveyed by acoustic means by changing the position of sound objects into a virtual space;
- To open a new multi-modal channel to convey expressiveness;
- To establish an explicit connection between sound movement and expressiveness;

These purposes are related to the idea that much contemporary music can take advantage of expressive spatialization. In particular, this framework has been used in the opera *Medea* by Italian composer Adriano Guarnieri which was premiered at the PalaFenice in Venezia in October 2002.

1. INTRODUCTION

The need for this kind of investigation comes from some contemporary music works in which sound motion is a musical parameter like intensity, timbre, pitch etc.. While the connection between music and emotion of these afore mentioned parameters has been deeply investigated spatialization still remains a new research path. In fact, in scoring *Medea*, the composer Adriano Guarnieri stresses a desired expressive tuning between performer and sound movement gestures in order to obtain a stronger musical message. This work has been conducted within the MEGA (Multisensory Expressive Gesture Application, cf. `http://www.megaproject.org`) project which concentrates on expressive and emotional content modeling and communication in non-verbal interaction through the use of multi-sensory interfaces in shared interactive Mixed Reality environments.

1.1. Gesture

Generally speaking, gesture is often referred to dancer movements and sometimes to specific body expressions. However, gesture can be considered also a structure with definite semantics defined into an abstract space. In this way a musical phrase can be considered a gesture which can express an emotion using only musical parameters, where the music is the abstract space. This connection between music and body movement is explicit in dance: a specific choreography can be used to better express a given musical work in a ballet and vice-versa; the emotional states carried by the music are the same ones expressed by the body movement. Moreover, there are many psychological studies about music and emotion that maintain that music can represent the dynamic properties of emotions like speed, strength and intensity variations. In particular, Gurney asserts that music can express these emotions through the association of affinities between musical characteristics and body movements which can show emotions [1]. Furthermore, Imberty [2] underlines that there are some kinetic tension and release schemes which are typical for both body and emotion so that movements and emotional states are a coherent set and gesture is a communication channel. Sound movement in the space appears to be a very good choice to explore this connection between gestures and emotions. Fig.1 describes the many steps of this kind of work: we will concentrate on the last three in order to show how spatialization can be used in musical performance.

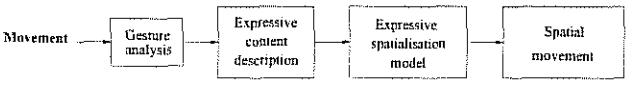


Figure 1: *Connection between two kinds of movements*

2. A PERCEPTIVE TEST

In order to evaluate whether sound movement can be considered a musical parameter and whether it has some expressive content for the listener a psychoacoustic test was performed. Its design was aimed at establishing a perceptive paradigm to construct a model that can be used in a musical context. This preliminary study focused on three parameters which are considered to be the basic components of sound movement: speed, legato/staccato and path. Variations of these parameters were designed to obtain eight stimuli/gestures as shown in Tab.1. These variables in these parameters were designed to have only two opposed values in each examples:

- speed: 0 = 200ms, 1 = 2000ms (where this sustain time represents the sound permanence on every loudspeaker)
- path: 0 = continuous circular, 1 = discontinuous random
- legato/staccato: 0 = 0.2 overlap, 1 = 2 overlap (where the overlap shows the degree of crossed overlap between the loudspeakers amplitude envelopes: 0.2 means staccato, while 2 means legato)

	Path	Legato/Staccato	Speed
A	0	0	0
B	1	1	1
C	0	1	0
D	1	0	0
E	0	0	1
F	1	0	1
G	0	1	1
H	1	1	0

Table 1: Audio examples used in the test

With these kinds of movements, 8 audio examples were produced featuring white noise sound, while another 8 examples featured an harmonic continuous sound (a looped trombone sample) in order to evaluate if timbre has an influence on movement perception too. Listeners where placed in the center of a circular 6 loudspeakers setup. 20 subjects (10 musicians, 10 non-musicians) were selected and a dimensional approach was used: the subjects were asked to mark in a bi-dimensional space an increasing order alphanumeric character at each stimulus. The bi-dimensional space was organized to establish valence on the vertical axis and arousal on the horizontal one. This is an approach of representing emotional states, drawn from the psychological tradition, which is alternative to the categorical one. The clearest common element of emotional states is that the person is materially influenced by feelings that possess a valence: they are centrally concerned with positive or negative evaluations of people or things or events; furthermore, emotional states involve a disposition to (re)act in certain ways. Arousal states are simply rated in terms of the person's disposition to take some action rather than no action at all [3].

In Fig. 2 we can see the valence-arousal space and the Russel circumplex model used to evaluate the emotions.

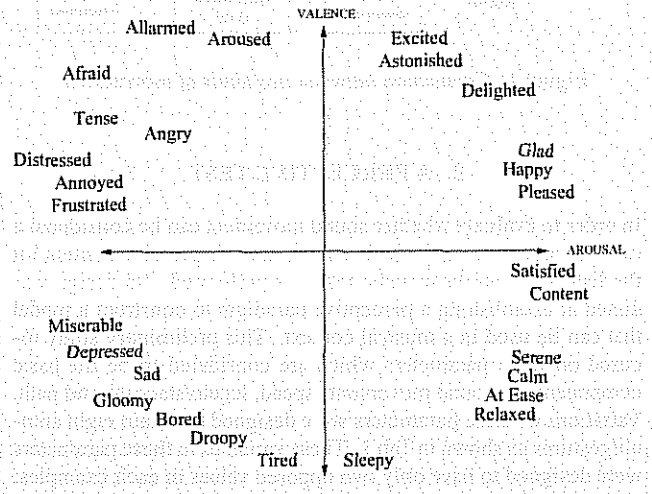


Figure 2: *The Russel circumplex model in the Valence/Arousal space*

2.1. Data analysis

The data collected has been analyzed statistically in order to obtain the desired information. The following analyses were performed:

- *Cluster analysis*
- ANOVA test
- Correlation
- T-Student Test

A preliminary cluster analysis was performed on the data. Cluster analysis is a method to assemble by similarity data which are similar and to obtain a dendrogram which is an easy way to plot these similarity characteristics. A complete agglomeration with Euclidean distance method was used in this case. We tried to verify for each timbre if a different behavior could be observed between musicians and non musicians. Fig. 3 and in Fig. 4 do not outline two radically separated groups, even though in the case of the harmonic sound the two groups are a slightly more distinct. After the

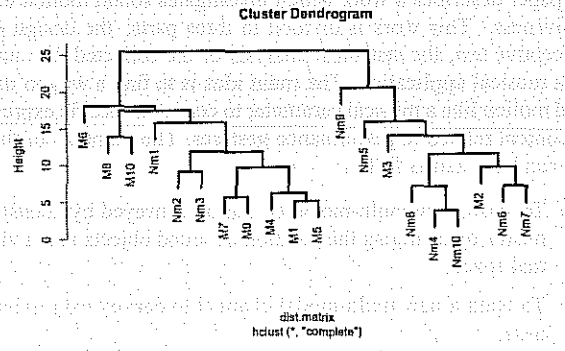


Figure 3: *Trombone data dendrogram*

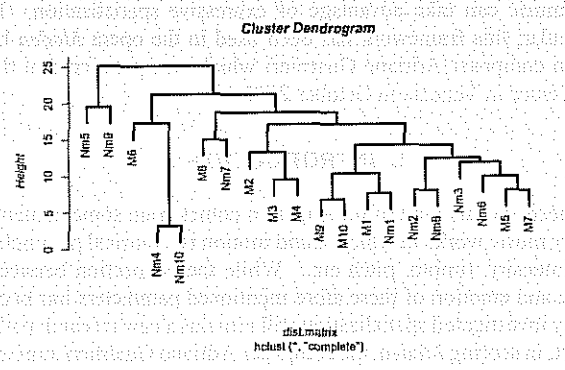


Figure 4: *White noise data dendrogram*

cluster analysis a one factor ANOVA test for each dimension was performed in order to evaluate if the mean value of the collected data was significant. The results of the ANOVA test are synthesized in Tab.2 where the significance threshold is below $p=0.05$. We can observe that the harmonic sound mean values are very significant for the arousal axis while they tend to be less significant for the valence axis even though they are always under the threshold. On the other hand, white noise mean values are not significant for the valence axis.

The arousal axis seems to be an important discriminant in distinguishing movements and finding some other law between our three parameters and the movements. The results obtained in this study were compared to the Russel model. In this space several

	Trombone	White noise
	P value	P value
Musicians - Ar.	$p < 0.001$	$p < 0.001$
Musicians - Val.	$p < 0.02$	$p < 0.05$
Not musicians - Ar.	$p < 0.001$	$p < 0.02$
Not musicians - Val.	$p < 0.005$	$p > 0.05$
Total - Ar.	$p < 0.001$	$p < 0.001$
Total - Val.	$p < 0.002$	$P < 0.005$

Table 2: P-value in the ANOVA test

studies have found the expressive content of many musical parameters like rhythm, melody and harmony [4] [5]. The similarity between the the results obtained in these studies and those emerging from the test described in this paper is indeed a striking one: Fig.5 shows the mean values for each stimulus distributed among harmonic sound and white noise. The collected data and its statistical

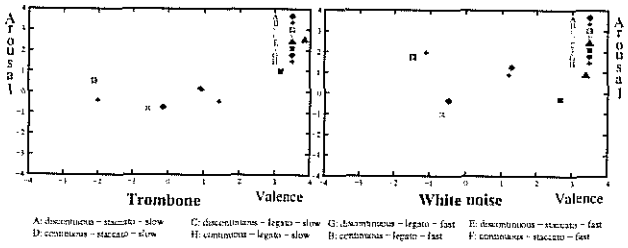


Figure 5: *Trombone mean values*

analysis allow some conclusions to be drawn:

- The values obtained from each timbre identify a similar space location for the same movements; there are some movements that seem to be further correlated for both tests; these are: C-H, A-D, B-G, E-F. These movements are correlated by featuring the same speed and the same legato/staccato characteristics, while the path is not overly important; the the T-test further outlines this evaluation and the correlation test shows a better correlation of musician answers;
- Legato/staccato and speed are stronger parameters than the path, even though the Russel model features a specific emotion related to path;
- Timbre influences subject answers as shown in Fig. 5: it was observed that very often the valence component was also dependent on the pleasantness of the sound heard; for example: in the white noise case, someone referred to it as a reminiscence of sea wave, while others were just plainly annoyed by it; consequently, the standard deviation over this collected data set was found to be bigger than that related to harmonic sound.
- Speed is strongly related with the arousal axis and it is also the most coherent parameter in subjective appreciation in the examples with the two different timbres.
- Different kinds of path (continuous or discontinuous) appear to be a more subtle parameter that can lead to the distinction among emotions which belong to the same area, while speed and legato/staccato appear to be stronger parameters in direct relation to emotions. In fact, these latter parameters have been widely studied in psychoacoustics and they are strongly connotated from the expressive point of view [4]: a smooth amplitude envelope, for example, is associated to emotions like sadness, fear or tenderness, while a steep slope envelope is related to happiness or anger like a staccato articulation is.

Fig.6 is an attempt to show the same relations using the Russel model inserting the harmonic sound mean values in the valence/arousal space. A relation between movements and perceived emotions can

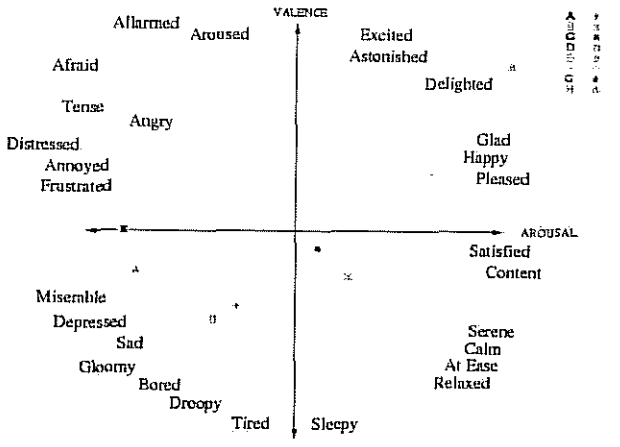


Figure 6: *Harmonic sound data in the Russel model*

be observed there. Tab.3 shows the design of a model to use this data in musical applications.

	Speed	Legato	Path
frustrated, distressed	-	+	discontinuous
depressed unhappy	-	+	continuous
tired, asleep	-	-	
serene, relaxed	+	+	continuous
happy, glad	+	-	continuous
excited, delighted	+	-	discontinuous

Table 3: Emotions associated to each movement

3. A MUSIC APPLICATION: *MEDEA*, OPERA-VIDEO BY ADRIANO GUARNIERI

Medea is an innovative work which stems out of an in-depth research in the multimedia domain; the importance and the amount of research dedicated to sound movement in space is quite considerable in this work. Guarnieri insisted explicitly on wanting to achieve an expressive matching between instrumental and sound movement gestures so that both gestures would reinforce each other producing a more powerful and complete message. Sound spatialization is a new way of conveying expressive content, and in this perspective it can be used to open a new expressive channel that can interact with other audio transformations and visual analyses, broadening the spectrum of multi-modality capabilities. *Medea* has offered the possibility to experiment with spatialization controls driven by the very same instrumental gesture (coming in this case from one of the solo trombone players) while allowing the connection between advanced technological research and

a real-world complex musical event. To be specific, the four solo trombone players are located in the hall among the public and follow very distinct individual movements. It is interesting to notice that some of the pages of the score suggest specific gestures to the trombone players (standing up, stirring up, etc.) That is precisely when spatialization is driven by their gestures. A webcam picks up the movements of the players and digital image analysis performed on this signal allows to extract some parameters that control the speed (and therefore part of the expressive features) of spatialization.

3.1. The space in the score

The score of *Medea* has been written with a precise reference to sound spatialization as a fundamental feature of the opera. This appears clearly when considering the distribution of musicians in the hall and the considerable number of tasks assigned to live-electronics. The musicians in the hall are to be considered as a *sonic body that lives among the public* to create a sort of gravitational center for the trumpets that are located to both sides of the audience: the presence of trombones with their gestural posture becomes a *central expressive feature. The importance* of space as a musical parameter is evident when considering the articulated legenda related to space at the beginning of the score. In that legenda, there are four different sound *reinforcement modes:*

- **AT**: Transparent reinforcement; delays are used to keep a natural perspective of sound positioning itemize
- **Celluloide**: random movement among the 4 stereo front speakers and the front central cluster
- **Pioggia**(Rain): fast random movement on some specific speakers above the public
- **Olophonic**: movement *simulation operated by controlling* volumes through low-pass filtering

Each instrument has its own spatialization modes: the score marks very precisely each transformation and movement. In keeping the trombone perspective, we can define 11 different modalities:

- Tb1= 3-9; Tb2= 4-10; Tb3= 5-9; Tb4= 6-10: *trombone n.1* is statically located on speakers n.3 and 9; trombone n.2 is statically located on speakers n.4 and 10; trombone n.3 is statically located on speakers n.5 and 9; trombone n.4 is statically located on speakers n.6 and 10; Thus, each trombone is reinforced simultaneously by a loudspeaker belonging the circle (1-2-3-4-5-6-7-8) and by 2 of the 4 loudspeakers suspended over the audience (9a-9b-10a-10b).
- 4 circular *movements.* 2 clockwise, 2 counter-clockwise, with 4 (slow) different timings; in this case, the loudspeakers involved are those that compose the circle surrounding the audience; *the clockwise path will then be produced* by the sequence 1-2-4-6-8-7-5-3 while the counter-clockwise path will be produced by the sequence 1-3-5-7-8-6-4-2. The starting point of each circle is not a *fixed one: it will change* according to the situation.
- Each trombone has a random space movement whose timing features are controlled by EW (speed CTL 0/127; legato/staccato and overlap): sound spatialization is thus related *to the trombone* player gestures through an EyesWeb software application
- AT: transparent reinforcement
- Localized random movement (timing and path): 9-3; 10-4; 9-5; 10-6, excluding the area where the trombone is actually playing; as an example: trombone 1 will move randomly among the 10-4, 9-5 and 9-6 areas.
- Static Tb1=1-3; Tb2=2-4; Tb3=5-7; Tb4=6-8: each trombone is located statically on the indicated loudspeakers
- Slow movement: 4 lines, clockwise + counter-clockwise: circular movements as above.
- Circular movement 1-3 clockwise; 2-4 counter-clockwise: trombones 1 and 3 will move clockwise, while trombones 2 and 4 will move counter-clockwise.
- Tb1=1; Tb2=2; Tb3=7; Tb4=8: *the trombones* are fixed statically in the indicated positions.
- From Tb1=1; Tb2= 2; Tb3= 7; Tb4= 8 to AT: the trombones move from a fixed indicated position to transparent sound reinforcement.
- Far away towards Tb1=3; Tb2=4; Tb3=5; Tb6=6: the trombones are virtually moved out of the hall along the indicated positions

In Fig.7 we can see the loudspeakers disposition and also the position of each trombone.

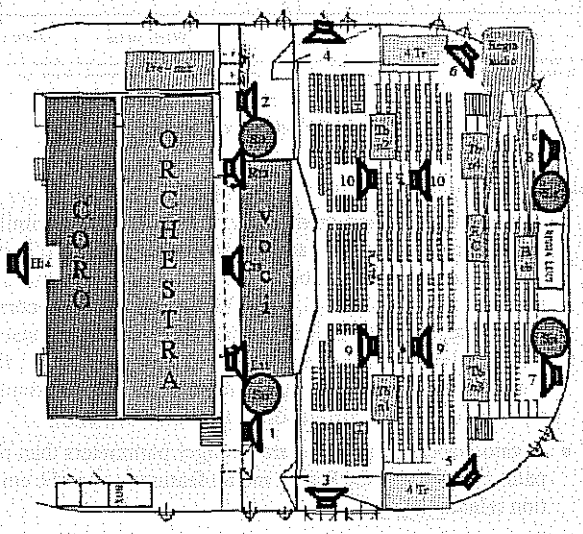


Figure 7: *Loudspeakers and instruments setup (courtesy of Alvise Vidolin)*

3.2. Hardware setup

The opera's live-electronics device is fairly complex as it amounts to:

- Microphones (for a total of 59): many different types of microphones have been used; in particular, trombones have been picked up with clip microphones hanged *onto the trombone* bell
- 1 Midas KL200 Mixer (48 input, 16 out; choir premix): dedicated to choir, trumpet, orchestra and percussion premix.

- 1 Yamaha DM2000 Mixer (I/O boards: 3 ADAT and 4 AES-EBU; live): devoted to live-electronics and control
- 1 Yamaha DME32 Mixer (32 in, 24 out): global audio I/O management
- 31 loudspeakers subdivided in 14 audio paths: 8 surrounding groups, 1 behind the stage, 1 central, 2 stereo on the stage, 2 above the audience
- 2 Yamaha 01V remote control
- 2 JLCooper Midi fader
- 1 M6000 TC-electronics multi-effect: local reverb
- 1 M3000 TC-electronics multi-effect: global reverb
- G4 TR for spAAce - Hammerfall RME PCI - MIDI I/O (TR - serial): computer devoted to sound spatialization
- G4 TR DSP - MOTU 896 AV (8 ADAT + 8 Analog) MIDI (TR - serial): digital signal processing of trombones, choir, trumpets and solos.
- Emu 6400 + keyboard: note and dynamic control of sampled choirs
- Kyma AV + PC control (AV): management and control of some transformations

In Fig.8 we can see the connection between all these devices, while Fig.9 describes all MIDI channels

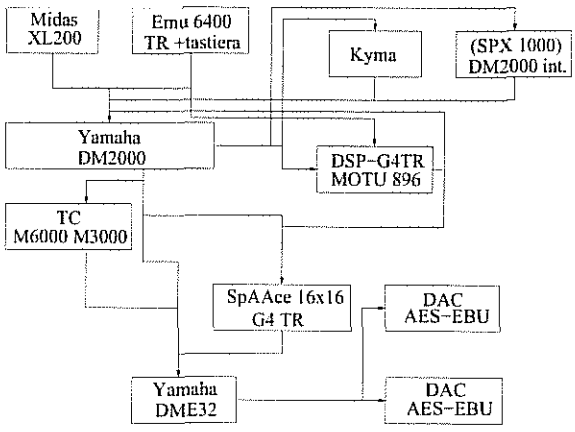


Figure 8: *Hardware setup (courtesy of Alvise Vidolin)*

3.3. Software setup

In this complex structure the actual sound spatialization is achieved through the "G4 spAAce" machine showed in the illustration and the spAAce software, cf.Fig.8, written by Alvise Vidolin and Andrea Belladonna in the Max/MSP environment on the Apple Macintosh platform [6]. This system is able to simulate moving sound sources featuring a great variety of configurations and of movements in all directions. The system is based on Inter-aural Amplitude Differences (IAD) which may be controlled via MIDI messages thus guaranteeing the availability of remote controls and the possibility of managing a system with a reduced set of controls located far away from the core mixing consoles. The movement control of sound sources can be managed by two operational modes: "Realtime" and "Playlist". The Realtime mode features gestural

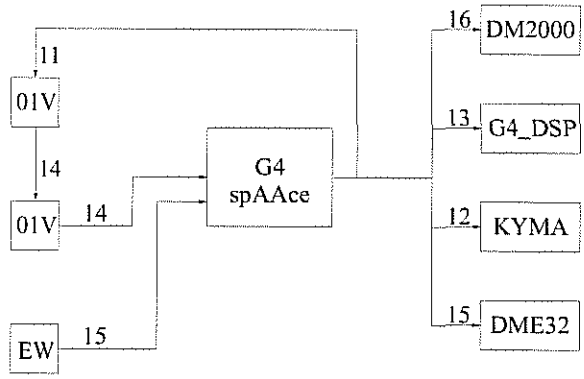


Figure 9: *MIDI channels (courtesy of Alvise Vidolin)*

controls for the majority of space management parameters; the Playlist mode allows the definition of a list of space movement events that can be activated either automatically as a sequence in time or manually by an operator in stepwise fashion. In the case of Guarnieri's *Medea* the manually-operated Playlist mode has been used: all movements and transformations were executed following the score and the conductor and all sound movements were fixed beforehand with the exception of those coming from the trombones. For these latter sounds, movements derived from the gestures of trombone player 3 have been captured by a webcam and digitally processed by the EyesWeb program to provide the speed parameter of each movement using a gesture-speed mapping. The patch used is displayed in Fig.10: its functionality derives from a translation of the image bitmap in terms of speed: very intense instrumental gestural activity ("stirring up") leads to a large bitmap variation and therefore to high speed, while reduced gestural activity corresponds to a moderate speed of movements. The original project was more articulated and complex: the idea was to control not only speed but also legato/staccato properties through the instrumental gesture. Red LED lights were attached to the clip microphone placed over the trombone bell, and the speed calculation was performed detecting the LED position in space. The purpose was to discriminate four fundamental types of movement whose micro-variations would then be derived by continuous interpolation among them. In Fig.11 we can see the trombone player during the première performance.

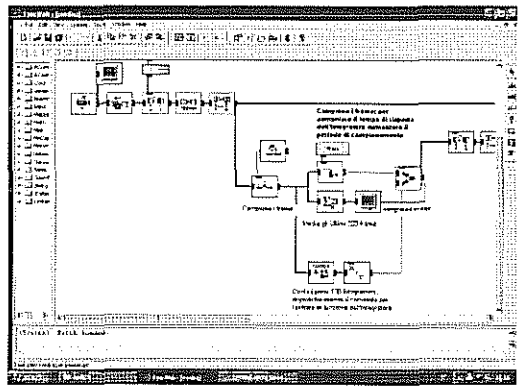


Figure 10: *EyesWeb patch*

Figure 11: *Trombone 3 during the première performance of Adriano Guarnieri's* Medea

4. CONCLUSIONS AND FUTURE PERSPECTIVES

Space has become a completely legitimate musical parameter, and future generations of composers will certainly take it into account. The space dimension is becoming also a common parameter and it may be used as an hyper-realistic operation or else in an expressive dimension, according to the composer aesthetic choices. In this work it is clearly shown that it is possible to directly translate instrumental gesture into spatial movement even though such a translation is fairly complicated in its conception, both from the technological and from the musical point of view. One thing to be kept in mind is that a physical gesture is generally a quick movement in relatively restricted spaces. When we project these movements into space, movements of several meters will correspond to movements that span some centimeters: small gestural speeds will become considerable sound speeds. Thus, there is a substantial scale difference between sound in space movements, which should be much slower than the gesture that is performed by the player. A "transducer" must be found then to map the physical gesture to the sound gesture when some kind of direct or inverse coherence is sought, because the two processes move over two different space and time scales. However the expressive component of sound movement is a research domain which promises to reach many good results also on the musical side. Space is a parameter which is much more complex than simple movement:

1. The correlation between movement speed, its variability, etc. and the agogic mechanisms which are already well known in music (legato/staccato, fluid/rigid, etc.)
2. The expressive meaning of proximity/distance (with its relative speed of change, etc.) along with an efficient concert-grade implementation
3. The coupling/decoupling of movements with the agogic parameters of a musical phrase (for example, what will happen when a crescendo is played by a sound source that is simultaneously being moved far away from the listeners? The crescendo will be perceived just the same because its perception is connected to timbre as much as to amplitude, but what will happen to its expressivity? will it be modified? etc.)

These and other aspects will require a considerable amount of further investigation.

5. ACKNOWLEDGEMENT

The author wishes to thank all the brave souls that accepted to patiently sit through the test sessions, Prof.Giovanni De Poli for allowing the work to be carried out at the CSC, Antonio Rodà and Sergio Canazza for their constant help and useful suggestions, all the staff at BH Audio for the pleasant moments spent during hectic production times, Alvise Vidolin and Nicola Bernardini for allowing her to participate to the making of a masterwork, and last but not least Adriano Guarnieri, without whose music none of all this would exist.

6. REFERENCES

[1] Budd, M., Music and the emotions: the philosophical theories, Routledge ed., London, 1992.

[2] Imberty, M., Suoni Emozioni e Significati, Editrice CLUEB Bologna, Bologna, 1986.

[3] Cowie, R., Douglas-Cowie, E., Tsuputsoulis, G., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G., Emotion Recognition in Human-Computer Interaction, IEEE Signal Processing Magazine, January 2001.

[4] Juslin, Sloboda, Psychological perspectives on music and emotion, Music and Emotion - Theory and research, Oxford University Press, Oxford, 2001

[5] Gabrielsson and Lindstrom, The influence of musical structure on emotional expression, Music and Emotion - Theory and research, Oxford University Press, Oxford, 2001.

[6] Belladonna, Andrea and Vidolin, Alvise, spAAce: un programma di spazializzazione per il Live Electronics, pagg. 113-118, Proc. Second Int. Conf. Acoustics and Musical Research, Milano, 1995.

TOWARDS MULTILEVEL GESTURAL CONTROL

Francesco Scagliola, Pino Monopoli

Conservatorio di Musica "Niccolò Piccinni"
Bari

fscagliola@libero.it, pinomonopoli@tiscali.it

ABSTRACT

In this paper we will discuss the possibility, granted by computer music, to shape all controllable sound parameters by superimposing some control function at several structural level.

This apparently simple, 'technical' feature leads to some important compositional considerations, which will be described in detail.

1. INTRODUCTION

The use of computer to make music allows a very high level of accuracy in the control of several sonic characteristics: some sound parameters, which in instrumental music are strictly connected to the performing practice and are only partially and approximately noted (for example loudness level), or even not noted at all (vibrato), in computer music can be precisely controlled, becoming elements of compositional pertinence.

We can control the sound parameters in a static or in a dynamic way, according to two different approaches in formal construction. In the first case we have a «sequential form», while in the second a «morphic form», in Trevor Wishart's meaning. In the sequential approach the form is generated by juxtaposing fixed values, while in the morphic one the form is generated as dynamic articulation of the continuum, where the sound is shaped with a motion between different fixed values:

> «The crescendo is an example of the most elementary morphic form, a linear motion from one state (in this case, of loudness) to another. In a world of stable loudness fields (not taking into account the subtleties of loudness articulation in the performance practice) this was a startling development. But traditional notation gives no means to add detail to this simple state-interpolation.» ([1], p. 108)

On the one hand the computer allows to run in a more precise and deeper way this type of transition; on the other hand it permits to operate in an analogue way on different sound parameters:

> «The combined powers of the computer to record sound and to perform numerical analysis of the data to any required degree of precision immediately removes any technical difficulties here. For those willing to deal with precise numerical representations of sonic reality, a whole new field of study opens up.» ([1], p.109)

Not only: by superimposing different gestures to different sound parameters it is possible to create a kind of new counterpoint, in which the several (controllable) parameters of the sound take their own gestural autonomy.

This gestural approach is moreover extremely connected non only with the compositional aspect, but also with a properly perceptional-analytic side. This is the approach which is the basis for the spectromorphology theorized by Dennis Smalley, who, for his part, states precisely that «spectromorphology is not a compositional theory or method, but a descriptive tool based on aural perception» ([2], p. 107).

In Smalley view, besides, as in electroacustic and computer music there are no coded hierarchical levels of formal structuration, the gesture itself, considered at different temporal scales, guides the perceptional process:

> «In my spectromorphological approach, the concepts of gesture [...] may be applied to smaller or larger time-spans which may be at lower or higher levels of structure. Finding the ' right' levels of temporal dimensions to apply the attributes of these concepts must remain the perceiver' s decision.» ([2], p.114)

2. MULTILEVEL GESTURES

We saw the possible use of multidimensional gestures suggested by Wishart, that is the use of different gestures applied to the 'multidimensional space' of sound parameters.

In this paper we will discuss the possible superimposition of different gestures on a single parameter, at several structural levels. In this way we

can realize a multilevel gestural control over any controllable parameters. Better: over the parameters we decide to control in our composition.

Taking in account the aforesaid example of the orchestral crescendo, we can think of a multilevel one: sound loudness can be modified first by a control function, then by a second one, then by a third, and so on. The global result of these superimpositions will be the product of the used functions.

Using this kind of process, we can use several linear control functions that, applied at several levels, will generate a total nonlinear control function.

An example: let us consider three linear control functions, which can be applied to whatever parameter: the first function arises from 0 to 1 in 4 seconds; the second in the same time moves from .8 to .6; the third for 2 seconds stays at .6, then goes down to .4.

These are the functions:

$$f_1(x) = \frac{x}{4} \tag{1}$$

$$f_2(x) = \frac{16 - x}{20} \tag{2}$$

$$f_3(x) = \begin{cases} \frac{6}{10} & \text{if } x < 2 \\ \frac{8-x}{10} & \text{if } x \geq 2 \end{cases} \tag{3}$$

with their correspondent plots:

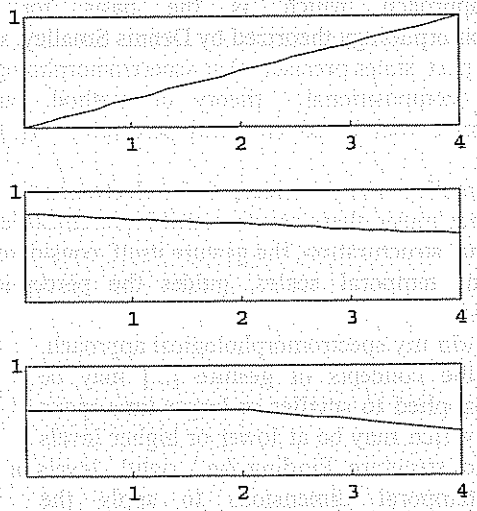


Figure 1: plot for functions (1), (2) and (3), respectively.

The result of the superimposition of the three gestures, that is the product of the three function, is the following:

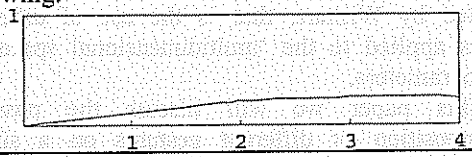


Figure 2: product of functions (1), (2) and (3)

with, eventually, its rescaled version:

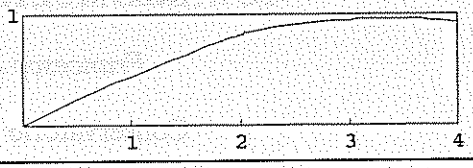


Figure 3: rescaled version of plot in figure 2

3. STRUCTURE AND FORM

The superimposition of several levels of linear functions is a very important tool – from a compositional point of view – since these functions can be applied over different time lengths. In this way the gestures become a fundamental mechanism for integrating and differentiating the sound material, in the sense of Dufourt thought regarding the (new) possibilities – and the new problems – that the use of electronics for music broughts in:

> «The new fact is that thought controls the creation of operative rules. The new problem, on the contrary, is the sense to be given to the organization that those rules make possible, that is unifying and specifying, differentiating and juxtaposing. [...] we know that it is necessary to work on systems of relations, state principles of functional solidarity, try to simplify the technical processes, and open the way to organization likeness.» ([3], pp.28-29; translation is ours)

The central idea of our work is the capability of creating a complete formal structuration by superimposing several linear gestures on different time scales: we search a formal and structural unity starting from a certain number of different sound fragments and using a gestural process for controlling a sonic parameter.

It is a matter of investigating a possible coincidence between structure and form; we intend with structure the *application* of processes to a sonic material, while with form the *perception* of processes applied to a sonic material (see [4]). So, in our approach, we seek a correspondence between compositional and perceptual-analytic process.

So, we start from fragments: the initial assumption is that these fragments are marked by a certain staticity of one or more controllable parameters: for example loudness, or pitch, or cut frequency of the filters, or again an internal rhythmic feature, and so on. The gesture, applied over several level but on a parameter at a time, will formally and structurally 'orientate' the parameter itself.

A first level of gestural application is the single fragments: a different gesture is superimposed on each fragment, so that every fragment assumes its own

'gestural autonomy' with regard to the parameter we are controlling. We have to note that, since this phase, we can differentiate originally similar sonic materials or create perceptual associations between different materials. For example, a common gesture applied over two distinct fragments leads to a certain similitude between them; on the contrary, a fragment can be duplicated and shaped by two different gestures: in this case we will hear the common origin of the two fragments, but with a difference in the motion of the sonic parameter.

A second level of gestural application permits to group the fragments, superimposing a unique control function over two or more fragments. This fragments can be coincident, partially coincident or one after the other in time: for example we can superimpose a gesture from 0" to 10" over three fragments, the first which starts at 0" and has a duration of 4", the second which starts at 3" with a length of 2", the third which starts at 5" and plays for 5".

The grouping generates unity, but several groups with different gestures can generate also contrast, if the gestures are quite different.

Furthermore, we can iterate this process grouping the original fragments in different ways, using new gestures, superimposing the gestures one to another, and establishing new sonic relations.

Besides, we can introduce a temporal segmentation of the piece, subdividing it in several subsequent sections, each of them with its own gestural behaviour. In this case, all the fragments that lie in a section will be oriented by the gesture which characterize it.

Also this process can be iterated, creating several time segmentations and multiplying the control functions.

4. A SMALL EXAMPLE

The mechanism described before can be sketched in as follows: let us consider 10 sonic fragments, each of which having its own gestural autonomy. In a first compositional step, they are composed by superimposing and juxtaposing, in time, one to another (Figure 4).

In a second compositional phase, we create several group using the starting fragments by superimposing, over each group, a unique gesture. Naturally, different gestures for each group are used. (Figure 5) This gestural superimposition will modify the original fragment gestures, as showed in Figure 6.

Furthermore, the entire piece is subdivided into five time subsequent sections; each of them is characterized by a particular gesture, which can be in continuity or in contrast with its adjacent sections (Figure 7). This new level of gestural superimposition will modify again the gestures of the fragments, which now are oriented in a higher formal and structural (piece-oriented, in one sense) way (Figure 8).

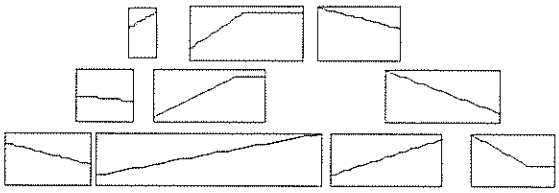


Figure 4: composition of gestural-oriented fragments

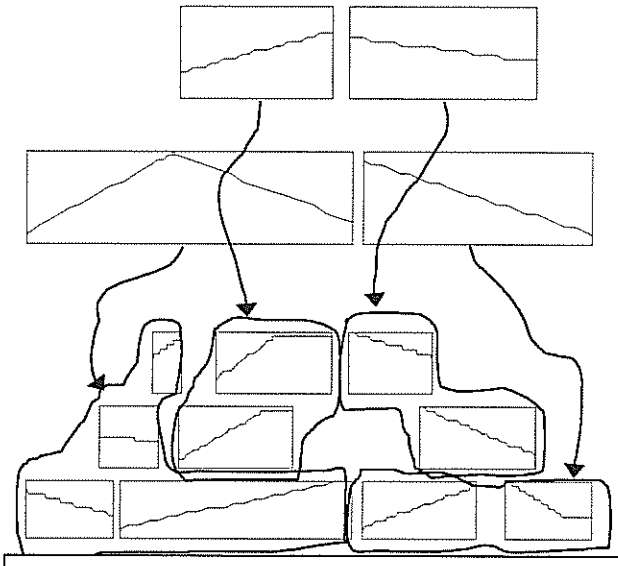


Figure 5: grouping of fragments by superimposition of different gestures

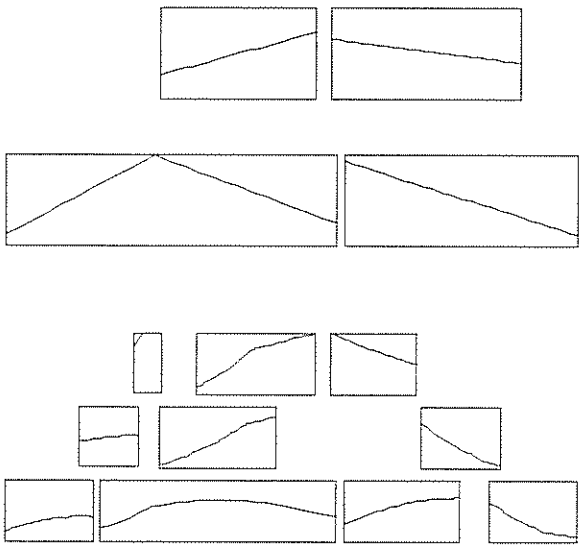


Figure 6: effect of 'gestural grouping' on fragment gestures

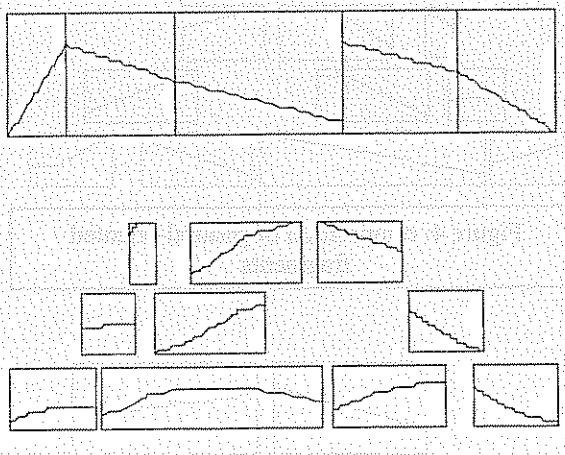


Figure 7: articulation in five sections by gestural superimposition

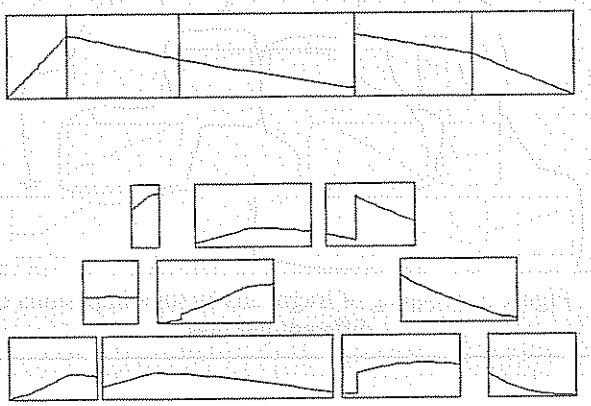


Figure 8: effect of superimposition of 'section gestures' on the fragments

5. FURTHER DEVELOPMENTS

Two relevant aspects will be investigated in a future work in order to better study the superimposition of gestures at several formal-structural levels.

First of all, the different 'weight' to be attributed to each level, that is the values that the control functions can assume. As an intuitive approach, at a higher level of structural organization must correspond a higher difference between the minimum and the maximum values of the function. According to this empirical intuition, in our plots different weights have been attributed: gestures at lower level (fragments) vary from 0 and 1, gestures at grouping level from 0 and 1.5 and gestures at section level from 0 and 2. But we think that a more precise definition of these difference can be studied.

Furthermore, we will investigate the 'changes of direction' in the control functions, at the several level. That is: is it possible to determine a certain formal variety operating on the coincidence or, on the contrary, on the dephasing of the control mutations?

We think these are two fundamental steps in the aforesaid search of a certain correspondence between form and structure.

6. REFERENCES

[1] Wishart T., 1994, Audible Design. A plain and easy introduction to practical sound composition, Orpheus the Pantomime Ltd, 1994

[2] Smalley D., Spectromorphology: explaining sound-shapes, in «Organised Sound», Volume 2, Number 2, August 1997, pp.107-126, Cambridge University Press, 1997

[3] Dufourt H., Musique, poivoir, écriture, Bourgois, Paris, 1991

[4] Sciarrino S., Le figure della musica da Beethoven ad oggi, Ricordi, Milan, 1998

7. APPENDIX: A SIMPLE IMPLEMENTATION IN CSOUND

```
; multi.orc

sr          =          44100
kr          =          4410
ksmps       =          10
nchnls      =          1

            instr 1

; instrument 1 produces three gestures as global variables

gk_level01  linseg .8,p3,2
            ; hi level gesture
gk_level02  linseg .8,p3/2,.2,p3/2,1.5
            ; mid level gesture
gk_level03  linseg .8,p3/3,1,p3/3,.2,p3/3,1q
            ; low level gesture

            endin


            instr 2

; instrument 2 creates a simple sine-waveform at a frequency of 6000 Hz
; which is multiplied by the product of the three gestures in instrument 1

iamp        =          5000
ifreq       =          6000

asig   oscili  iamp,
gk_level01*gk_level02*gk_level03 * ifreq, 1
            out        asig

            endin
```

```
; multi.sco

f1 0 4096 10 1

; this line activates the gestures (instrument 1)
i1 0 20

; this line activates the oscillator
i2 0 20
```

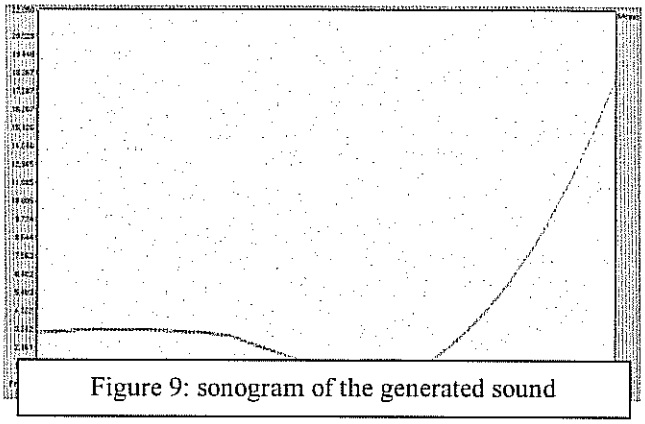


Figure 9: sonogram of the generated sound

AN XML REPRESENTATION FOR THE EXPRESSIVE PERFORMANCE OF A MUSICAL SCORE

Antonio Rodà, Massimo Scantamburlo

CSC – DEI, University of Padua
ar@csc.unipd.it
http://www.dei.unipd.it/ricerca/csc/

ABSTRACT

At the time being, XML seems to be the most promising format to represent music contents at a symbolic level. In this paper, we consider the problem of representing the music information related to the expressive performance of a musical score. We developed an XML-based environment that will allow us to test novel solutions and put in evidence the various problems concerning this subject.

1. INTRODUCTION

The coming of the Internet era is deeply changing the way of thinking to music contents. The new technologies allow a novel approach to music, that go beyond the simple listening activity, and include a lot of interactive features, such as finding and selection of contents, customization, deep integration with other media, hyper-navigation, etc.. Furthermore, this approach gives rise to new business models, that are destined to replace the traditional way of marketing, but require to solve a lot of problems concerning security, author rights, and payment transitions. One of the main topic in this field, is the design of a suitable format to represent all the information needed to face all these tasks, as the traditional standard (i.e. MIDI, WAV, MPEG1-layer3, etc.) dont comply enough the new requirements. The format would have to meet two different needs: i) to be adequately general, in order to represent different aspects of the music content, such as graphic, symbolic, and physic world ii) to be shared by as large as possible community of researchers, industries and users.

One of the more promising format, that can comply these requests is the eXtensible Markup Language (XML) [1]. As its more known brother HTML, XML is derived from SGML: they are all programming languages based on markup, that is we can encapsulate part of a document in a section and attach some additional information to each kind of section. The main result we can achieve through this action is a flexible structure of the data contained in a document.

XML is yet widely used and supported by a lot of software packages that require to represent data in a structured form. Nevertheless, the employ of XML in the music field need to define which data have to be represented, and how they have to be structured. In the last years, many researcher groups proposed an XML scheme suitable to music representation, but, up to now, none of them was accepted like a standard by all the community. The main problem is that a music content can be considered from different points of view (e.g. audio, symbolic, graphic), and can be addressed to different kind of users (e.g. musician, musicologist, teachers, common people) with different aims. Therefore, depending on the specific task, the attention need to be focused on one musical aspect, or a subset of all the possible aspects, as it is very difficult to satisfy all the requirements at the same time. Some proposals are focused on the representation of audio signal (e.g., [2]); other on the symbolic representation of music (see [3] for a review of the main proposals). There are also some interesting attempts to define an XML-based framework ([3], [4]) for the structured representation of several musical contents such us musical sheet (or an equivalent symbolic representation), audio recordings, video recordings, etc.

As regards the symbolic representation, one of the less investigated aspect is the coding of information related to the music performance, i.e. all the information that are needed in order to allow a computer to play/perform a musical sheet in a non-mechanical (human-like) way (see [16] for an alternative non XML-based approach to this problem). This capability would be very useful in a lot of contests: the fruition of musical libraries, when an audio recordings is not available; the educational field, in which computer generated performance can be a useful tool to study interpretative models [5]; the development of new interaction paradigm, based on the communication of emotions, that can be conveyed by means of music [6].

In the last years, several computational models for the automatic human-like performance of music were developed, using different strategies: analysis-by-synthesis method, analysis-by-measurement method, machine learning techniques (see [7], [8] and [9] for a review). Almost all these models, however, require extra information that normally are not coded in a musical sheet or in a MIDI file, such as a structural or an harmonic analysis of the score. This fact imply the need for developing a format to represent these information in a suitable and flexible way.

In this paper we present an XML scheme
an XML-based framework for the automatic expressive performance of musical score. This is not a proposal for a generic XML music representation, as we considered only one aspect (what concerns expressiveness) of the problem. This work, instead,

aims to demonstrate how and with which advantages and problems XML can be used to represent music information oriented to automatic performance. In order to face in a complete manner this issue and be able to test the reliability of our representation, we developed some tools that allow to easily produce a correct XML representation starting from a score written in Finale, and to rendering an expressive audio performance of the score starting from the XML representation.

2. EXPRESSIVENESS MODEL

We based our XML representation on an existing expressiveness model developed by the CSC of the University of Padua (see [10] for a detailed description). This model, focused on western tonal music, presents some interesting characteristics, that are very useful for the development of an automatic player:

a) it assumes that there isn't a unique "right" performance of a score, but the same score can be played with different expressive styles (called expressive intention); in this sense, the model makes a distinction between a *neutral performance* (i.e., a human performance of the score without any expressive intention or stylistic choice) and an *expressive performance*.

b) it is based on the hierarchical structure of the musical piece, i.e. the subdivision of the musical language in periods, phrases, and words.

c) it employs information that are expressed in an abstract way (closer to the musical language than to the physical one), i.e. using the parameters *intensity* instead of *key-velocity*, or an *accelerando* instead of a duration expressed in *ms* or in *tick*, etc.

The item (a) implies that the system can change its performance following the user's expressive intentions, or depending on the user's actions: this capability makes possible a music fruition with different degree of interaction (see [11] for the description of a possible scenario).

Many researches put in evidence that the musical structure, item (b), is one key aspect of musical language (at least for western music), and a lot of interpretative model uses this information to generate the performance (e.g., [12], [13], [14]). Moreover, a musical structure description is very important for content retrieval applications, as the music segmentation plays the same role of word segmentation in textual retrieval engine.

Finally, the item (c) allows to face some common problems that arise when a computer plays a symbolic representation like MIDI. In this standard, in fact, the duration and the intensity of each note is exactly specified, by means of *ticks* and *key-velocity*: a MIDI file, therefore, is more similar to a (symbolic) "frozen" recordings than a musical score to be interpreted. As a consequence of it, if the instrument that plays the file is changed (e.g. in client-server applications, or when the file is shared with other users), the audio results are often not so god, because some parameters need to be changed to comply the requirements of the new instrument. Moreover, MIDI standard or equivalent representations don't codify a lot of information regarding timbre and articulation aspects such as spectrum, amplitude envelopes, vibrato, etc.

This is a list of the parameters taken into account by the expressiveness model, as reported in [10]:

- *time*: starting moment of the event in tick, as specified in the score.
- *duration*: note duration in tick, or number of sub-events forming the event, as specified in the score.
- *pitch*: e.g. a4 or e5.
- *channel*: number of the channel or track the events refers to.
- *elasticity*: degree of expressive elasticity of the event. It indicates when it is possible to work on the parameters of that group of notes (intensity, metronome, timbre characters) to reach a certain expressive intention without distorting the piece itself.
- *dynamics*: intensity curve that describes the profile of intensity deviations in the event.
- *metronome*: metronome curve that describes the profile of the metronome deviations in the event.
- *expression*: symbol of the adjective to be applied to the phrase + intention degree.
- *attack time*: duration of the attack expressed in a perceptual scale.
- *legato*: legato-staccato degree expressed in a perceptual scale.
- *intensity*: perceptual loudness.
- *brightness*: perceptual measure of the high frequency spectral components.
- *vibrato*: rate and extent vibrato expressed in a perceptual scale.
- *portamento*: glissando speed degree in a perceptual scale.

The first 4 parameters represent the music information as noted in the score (metric position, duration, pitch, track). The *dynamics* and *metronome* parameters describe dynamics and timing deviations in term of profiles, that reproduce different kinds of *crescendo-decrescendo* and *accelerando-rallentando* patterns.

The last six parameters are expressed in a perceptual scale, in which the unit represents the difference between two perceptual levels (e.g. the difference between f and ff for loudness). Because of their (abstract) definition, these parameters are independent from the particular musical instrument that will play the score. Finally, the parameter *expression* allow to specify a label representing the user's *expressive intention*: up to now, the model supports a limited set of sensorial adjectives, but this list can be easily extended.

3. XML REPRESENTATION

To develop an XML representation of the symbolic music information, it is necessary to rigorously define

the structure of such information, i.e. to design a Document Type Definition (DTD).

The first choice made was regarding what kind of information must be provided as tags (or elements) and what as attributes. Data regarding expressiveness are provided as elements since we focus our attention on this aspect. We also want to respect the logical structure of our score so we have events (phrase, note, chord and voice two) presented as tags. On the other hand information such as pitch or duration of a note, key velocity and others MIDI-like parameters are given as attributes of the events. Information about metronome, time-division, base-time and key-signature must appear at the beginning of the document in the section named <HEADING>. Figure 1 shows a fragment of the DTD: we see how the event PHRASE is defined. A phrase can be made by sub phrases, notes, chords and voice two's in any order and quantity. We define also the attributes that represents MIDI parameters and tags regarding expressive curves.

```
<!-- event PHRASE -->
<!-- a PHRASE can be made with more phrases and/or
chords and/or voice two's in every combination -->
<!-- ELASTICITY is required -->
<!ELEMENT PHRASE ((ELASTICITY | EXPRESSION? |
METRONOME? | DYNAMICS?)+, (PHRASE* | NOTE* |
VOICE_TWO* | CHORD*)+)>
<!ELEMENT EXPRESSION (#PCDATA)>
<!ELEMENT ELASTICITY (#PCDATA)>
<!-- curves -->
<!ELEMENT DYNAMICS (#PCDATA)>
<!-- METRONOME has been already declared -->
<!-- MIDI parameters are required and appear as attributes
-->
<!ATTLIST PHRASE
    level CDATA ""
    duration CDATA ""
    channel CDATA ""
    time CDATA "">
```

Figure 1: the phrase event definition in the DTD.

Beside the existence of a lot of tools for the parsing, the editing and the writing of XML files, the most evident advantage of this representation is its readable layout: in this way we have a clear vision on phrase division, duration and hierarchy, on what notes belong to a chord and so on. The self explicative labels for tags and attribute (e.g. <PHRASE> or <PORTAMETO> make immediately accessible the data and the music structure (see Figure 2).

4. SYSTEM ARCHITECTURE

To test the reliability of the XML representation, we developed some tools for the creation, editing and performance of the XML file (see Figure 3). Finale can be used to write the score and to describe the musical structure of the score. The latter task is performed using the smart shape tool, so that each slur individuate a segment; the slurs can be hierarchically structured as shown in Figure 4, so to individuate phrases, sub-phrases and words.

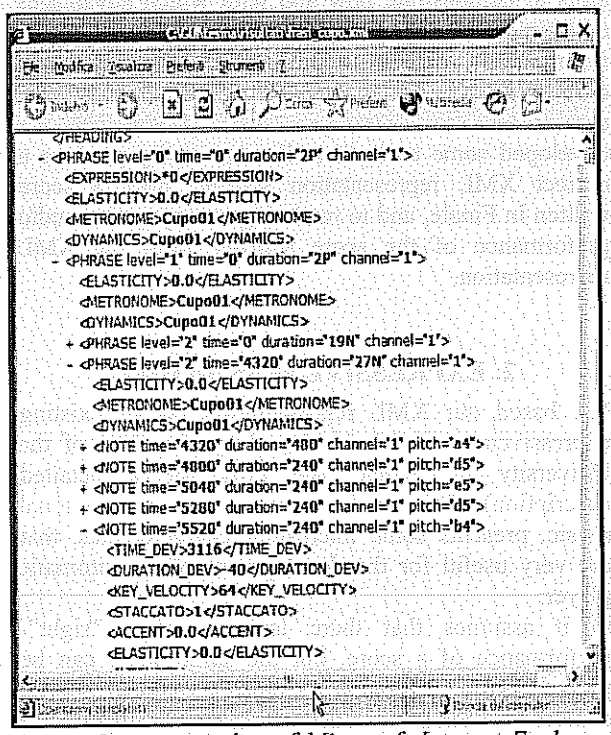


Figure 2: a screenshot of Microsoft Internet Explorer 6.0 used as XML viewer.

Moreover, using the text tool, it is possible to add some textual labels, that will be interpreted as the user's expressive intentions.

The resulting ETF file, generated by Finale, is then processed by the XML Creator tool. This tool parses the ETF file and translates the information following the DTD described in the previous section. In addition, when the routine finds a textual label that is supported by the expressiveness model (see section 2), it automatically calculates the musical parameters and the dynamics and timing curves in order to render the desired expressive intention. As asserted in section 3, XML file contains information at an abstract level, in order to allow an efficient rendering with many different synthesizer and synthesis techniques. However, it is necessary another elaboration stage (called Virtual Performer), in order to drive the particular instrument. Ideally, it would be necessary a Virtual Performer for each instrument (or each instrument family) that will be used, like it occurs for the device drivers. At the time being, we developed a Virtual Performer to play an FM instrument, implemented in CSound language. The output of the Virtual Performer, therefore, is a SCO file for CSound, that finally performs the audio rendering of the score.

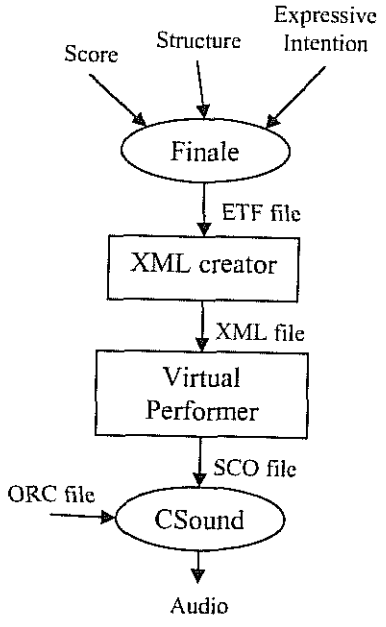


Figure 3: system architecture

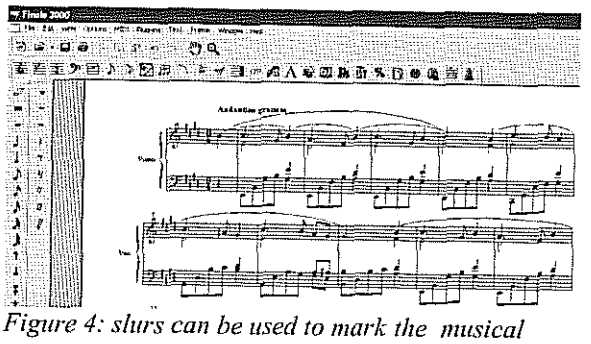


Figure 4: slurs can be used to mark the musical structure of the score.

5. CONCLUSIONS

Starting from an existent expressiveness model, we developed an XML format to represent all the information needed to an automatic expressive performance of a musical score. The representation is characterized by the use of abstract musical parameters and a structure description of the score. Moreover, we developed a framework for the creation and rendering of the XML files, that will allow us to validate the reliability of such representation. Some other important issues to be faced are the extension to a really polyphonic content (at this moment polyphony is restricted to chord or voice-two structures), and the enlargement of the expressive intention set supported by the model. Finally, an effort will be done to verify if the XML representation presented in this paper, could be made compatible with other XML-based format (e.g. MusicXML [15]), in order to move us toward a unique efficient and general XML standard for musical applications.

6. ACKNOWLEDGEMENTS

This research was supported by the EC project IST 1999-20410 MEGA (www.megaproject.org).

7. REFERENCES

[1] World Wide Web Consortium XML web site: http://www.w3.org/XML/.

[2] Gazon, D. and Andriarin, X. "XML as a means of control for audio processing, synthesis and analysis". Proc. of MOSART, Workshop on current research directions in Computer Music, Barcellona, November 15-17, 2001, pp. 85-91.

[3] Haus, G., Longari, M. "Towards a Symbolic/Time-Based Music language based on XML". Proc. of MAX 2002 International Conference, Milan, September 19-20, 2002

[4] Bellini, P. Nesi, P. "WEDELMUSIC Format: An XML Music Notation Format for Emerging Applications", Proc. of WEDELMUSIC2001 IEEE Conference, Florence, Italy, November 2001, pp. 79-86.

[5] Battel, G.U., Fimbianti, R. "Analisys of expressive intentions in pianistic performances", Proc. of International Workshop on Kansei, Genova October 1997, pp. 128-133.

[6] Juslin, P. N., Friberg, A., & Bresin, R. „Toward a computational model of expression in performance: The GERM model." Musicae Scientiae special issue 2001-2002, pp 63-122.

[7] Gabrielsson, A. "Music Performance". The psychology of music, In D. Deutsch (Ed.) The psychology of Music, 2nd. ed. New York: Academic, 1997.

[8] Palmer, C. "Music performance". Annual Review of Psychology, 48, 1997. pp. 115-138.

[9] Zanon, P., De Poli, G. "Estimation of parameters in rule systems for expressive rendering in musical performance". Computer Music Journal, 2002 (in press).

[10] Canazza S., De Poli G., Di Sanzo G., Vidolin A. "A model to add expressiveness to automatic musical performance". Proc. of International Computer Music Conference. Ann Arbour. pp. 163-169.

[11] Canazza S. De Poli G., Drioli C., Rodà A., Vidolin A. "Audio morphing different expressive intentions for Multimedia Systems". IEEE Multimedia, July-Sept., Vol. 7, N° 3, pp. 79-83.

[12] Todd, N. P. McAngus . The kinematics of musical expression. Journal of the Acoustical Society of America, 97, pp 1940-1949, 1995.

[13] Friberg, A. (1991). "Generative Rules for Music Performance: A Formal Description of a Rule System", Computer Music Journal, 15 (2), pp.56-71.

[14] De Poli, G., Rodà, A., & Vidolin, A. "Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance". Journal of New Music Research, vol. 27, no. 3, pp. 293-321, 1998, Special Issue.

[15] MusicXML Software page. http://www.musicxml.org/software.html.

[16] Kuuskankare, M. and Laurson, M. "ENP, Musical Notation Library based on Common Lisp and CLOS." In Proc. of ICMC'01, Havana, Cuba, Sept. 2001.

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

RECOGNITION OF FAMOUS PIANISTS USING MACHINE LEARNING ALGORITHMS: FIRST EXPERIMENTAL RESULTS

Patrick Zanon

CSC – DEI, University of Padua
Via Gradenigo 6/a, 35131 Padova, Italy
patrick@dei.unipd.it
http://www.dei.unipd.it/~patrick

Gerhard Widmer

University of Vienna and ÖFAI
Schottengasse 3, A-1010 Vienna, Austria
gerhard@ai.univie.ac.at
http://www.oefai.at/~gerhard

ABSTRACT

The paper addresses the question whether a machine can learn to identify famous performers (pianists) based on their style of playing. A preliminary study is presented where different machine learning algorithms are applied to performance data derived from Mozart sonata recordings by several famous pianists. It is shown that the algorithms learn to recognize pianists at a level better than chance, and that some pianists seem easier to recognize than others. The study identifies a number of limitations of the current approach (regarding both data and learning algorithms) and points to a variety of fruitful directions for further research.

1. INTRODUCTION

The work presented here is part of a large investigation into the use of novel computational methods for studying basic principles of expressive music performance [9]. One of the questions we study is whether and to what extent aspects of *individual artistic style* can be quantified. And one of the possible approaches to this question is to investigate whether machines can learn to distinguish and recognize different performers based on their style of playing.

Previous research has shown that this seems indeed possible, to a certain extent [7, 8]. In a study with 22 different pianists (teachers and students of the University of Music in Vienna) a new machine learning algorithm achieved a surprising level of recognition accuracy. However, that study was limited in many respects, particularly with regard to the data that were available (recordings by 22 different pianists, but only two pieces). On the other hand, the performance measurements were extremely precise, because the recordings had been made on a Bösendorfer SE290 computer-controlled piano.

The present paper describes first steps towards generalizing this research. We extend the study towards the analysis of famous world-class pianists, and work with larger collections of recordings. At the same time, that raises a data problem, because performances of famous artists are only available as audio recordings, and it is impossible to extract the same kind of exact performance information (e.g., details of timing and articulation) from these recordings. In the experiments to be reported below, performance information will be available only at an extremely crude level (essentially, only high-level tempo and loudness changes over time), and the question is whether performer identification is still possible at this level.

The paper is organized as follows: section 2 describes in detail the experimental methodology followed, including a description of the data, a characterization of the performance features extracted from the recordings, and a list of the machine learning algorithms tested. Section 3 presents preliminary experimental results that show that the learning algorithms can at least learn to identify performers at a level better than chance. The results also point to a number of problems and prompt us to identify several promising directions for further research. The next steps to be performed in this project are then detailed in section 4.

2. METHODOLOGY

2.1. The performance data

For the experiments, commercial recordings of piano sonatas by W.A. Mozart by six different concert pianists were collected, and a sizeable number of pieces were selected for performance measuring and analysis. The pieces are listed in Table 1, and the pianists in Table 2.

ID	Sonata	Movement	Key	Time sig.
kv279_1	K.279	1st mvt.	C major	4/4
kv279_2	K.279	2nd mvt.	C major	3/4
kv279_3	K.279	3rd mvt.	C major	2/4
kv280_1	K.280	1st mvt.	F major	3/4
kv280_2	K.280	2nd mvt.	F major	6/8
kv280_3	K.280	3rd mvt.	F major	3/8
kv281_1	K.281	1st mvt.	Bb major	2/4
kv282_1	K.282	1st mvt.	Eb major	4/4
kv282_2	K.282	2nd mvt.	Eb major	3/4
kv282_3	K.282	3rd mvt.	Eb major	2/4
kv330_3	K.330	3rd mvt.	C major	2/4
kv332_2	K.332	2nd mvt.	F major	4/4

Table 1: Movements of Mozart piano sonatas selected for analysis.

From the audio recordings, rough measurements characterizing the performances were obtained. More precisely, tempo and general loudness were measured at the level of the beats, by using an interactive beat tracking program [3] to find the beat in the audio signal and computing beat-level tempo changes from the varying inter-beat intervals. Overall loudness of the signal at these time points was extracted from the audio signal and is taken as a very crude representation of the dynamics applied by the pianists. No more detailed information (e.g., about articulation, individual voices, or timing details below the level of the beat) is available.

These sequences of measurements can be represented as two sets of performance curves — one representing beat-level tempo,

ID	Name	Recording
DB	Daniel Barenboim	EMI Classics CDZ 7 67295 2, 1984
RB	Roland Batik	Gramola 98701-705, 1990
GG	Glenn Gould	Sony Classical SM4K 52627, 1967
MP	Maria João Pires	DGG 431 761-2, 1991
AS	András Schiff	ADD (Decca) 443 720-2, 1980
MU	Mitsuko Uchida	Philips Classics 464 856-2, 1987

Table 2: Pianists and recordings.

the other beat-level loudness changes — or in an integrated two-dimensional way, as trajectories over time in a 2D tempo-loudness space [6]. We have developed a graphical animation tool called the *Performance Worm* [4] that displays such performance trajectories in synchrony with the music. A part of a performance as visualized by the Worm is shown in Figure 1. Note that the display is interpolated and smoothed. For the machine learning experiments reported below, only the actually measured points were used; no interpolation or smoothing was performed.

Thus, the raw data for our experiments is tempo and overall loudness values measured at specific time points in a performance (either every beat according to the time signature or, where beat tracking was performed at lower levels, at subdivisions of the beat). For each measured time point, the following is stored: t_i (absolute time in seconds), B_i (calculated tempo in bpm), L_i (loudness level measured in sone [11]), and some bit-coded flags that represent hierarchical structural information. More precisely, *ftbb* indicates whether the current time point coincides with a beat track point (trivially true for all), a beat, or the beginning of a bar; *fs*1234 indicates four levels of phrase structure (this information was added manually by a musicologist).

The raw data so obtained had to be refined in order to be homogeneous and usable in the learning process. In fact, data usually comes from different sources, and could have been produced by different persons, sometimes with different strategies. This introduce some noise that can affect the output of the learners. Thus, some time had to be spent in cleaning of the data in order to have the most homogeneous information representation as possible.

In our case, some of the pieces were tracked at the level of the beat (as defined by the time signature), some at the half beat level. Moreover, some pieces start in different ways, according to the performers' decisions: for example, some of them start with an upbeat, and some others with two. We decided to skip all the non-common information, by sampling the pieces with higher tracking resolution, and by discarding all the non-common up beats.

Moreover, most of the players repeated some sections, while others didn't (i.e., Gould). Also in this case, we decided to discard all the non-common sections, so that the learners would work with comparable data for all the performers, and not with over-represented sections that could affect their output.

2.2. Instances and features

Each measured time point, along with its context, is used as a *training example* for the learning algorithms. In other words, an example or *instance* for the learners is a subsegment of a tempo-loudness trajectory (see Figure 1), centered around a specific time point. Altogether, this procedure results in some 23.000 instances for all the six pianists.

The instances are represented by a set of *features* that are extracted from the raw trajectories. The features were calculated over a sliding window w_i of two bars (the context size). Thus, if in the original data there are n instances, then there will also be n windows and n sets of features. Of course, at the beginning and at the end, some of the features were calculated over a narrower window than two bars.

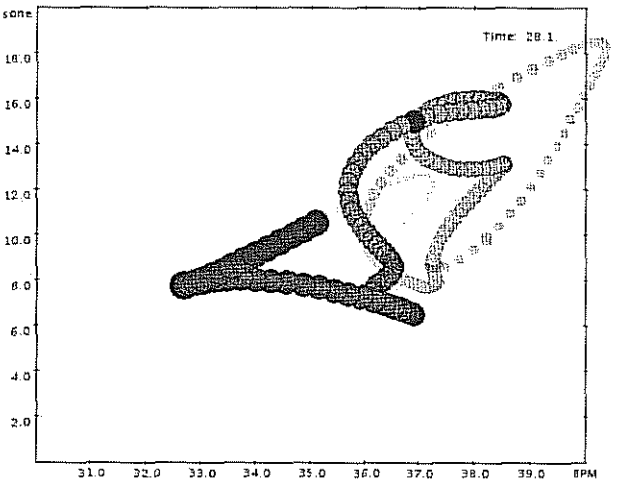


Figure 1: Snapshot of the *Performance Worm* at work: First four bars of Daniel Barenboim's performance of Mozart's F major sonata K.332, 2nd movement. Horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in *sone* [11]. Movement to the upper right indicates a speeding up (*accelerando*) and loudness increase (*crescendo*) etc. etc. The darkest points represents the current instant, while instants further in the past appear fainter.

Notice that the number of measured points included in a window w_i may be different between pieces, since there are different tempo indications and different tempo tracking resolutions. For example, in the piece kv280_3 there is a low sampling rate (1 beat per bar), while in kv282_1 there are 8 tracked points per bar.

The sliding window method produces some redundancy in the data, since the windows were overlapping; this should be an advantage in learning. Some of the features were calculated using a window which extended beyond the boundary between two sections. This is no problem in most cases, but in some cases the two sections may be non-contiguous in the original data. We chose to allow this discontinuity in the data, because the number of affected instances is negligible. Thus, it was not necessary to 'instruct' the code to recognize section boundaries.

Caution had to be taken with some of the extracted performance information; In particular, the features derived from loudness had to be filtered in some way, because they can trivially reveal some of the performers. For example, Gould's CD recordings are older (1967) than the others (1980-1991), resulting in a significantly lower recording level. That would permit the learners to detect this famous performer simply by loudness difference. Thus, a normalization in the data was carried out in order to mask this information: all the relevant loudness-derived features were normalized using the actual window average $\mu_L(w_i)$.

For each instance of the original raw data, the following features were computed both for tempo and loudness: the average value within the window $\mu(w_i)$, the standard deviation $\sigma(w_i)$, the

Operation	Tempo	Loudness	Others
None	B_i	$L_i^{(1)}$	$ftbb_i$, $fs1234_i$
Average and Median	$\mu_B(w_i)$, $\nu_B(w_i)$	$\mu_L(w_i)^{(1)}$, $\nu_L(w_i)^{(1)}$	–
Standard Deviation	$\sigma_B(w_i)$	$\sigma_L(w_i)^{(1)}$	–
Min, Max and Range	$m_B(w_i)$, $M_B(w_i)$, $R_B(w_i)$	$m_L(w_i)^{(1)}$, $M_L(w_i)^{(1)}$, $R_L(w_i)^{(1)}$	–
Normalization	$b_i(w_i)$, $\sigma_b(w_i)$, $m_b(w_i)$, $M_b(w_i)$, $R_b(w_i)$	$l_i(w_i)$, $\sigma_l(w_i)$, $m_l(w_i)$, $M_l(w_i)$, $R_l(w_i)$	–
Correlation	$\Sigma_{tB}(w_i)$	$\Sigma_{tL}(w_i)$	$\Sigma_{BL}(w_i)$
Directness	$\Delta_{tB}(w_i)$	$\Delta_{tL}(w_i)$	$\Delta_{BL}(w_i)$
Derivative	$M\delta_B$, $\mu\delta_B$	$M\delta_L$, $\mu\delta_L$	–

Table 3: Complete set of features extracted from the data for each instance. $^{(1)}$indicates that the corresponding feature must not be used by the learner (if it were, it would trivially reveal some of the performers, on the basis of the CD recording level).

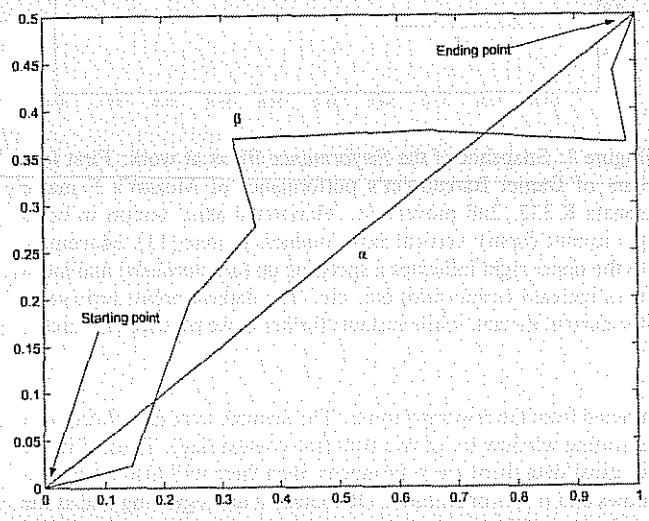


Figure 2: The definition of the directness index in an X-Y space.

median $\nu(w_i)$, the minimum $m(w_i)$, the maximum $M(w_i)$, and the range $R(w_i) = M(w_i) - m(w_i)$. For each of these features, the corresponding normalized ones were also calculated by division by the mean. The normalized features are indicated with lower-case letters of the tempo/loudness subscripts. For example, if $\sigma_B(w_i)$ is the tempo standard deviation, then $\sigma_b(w_i) = \sigma_B(w_i)/\mu_B(w_i)$ is the corresponding normalized version. Additional features were added that represent correlations: $\Sigma_{tB}(w_i)$ is the correlation between the time and the tempo, $\Sigma_{tL}(w_i)$ is the correlation between the time and the loudness, and $\Sigma_{BL}(w_i)$ is the correlation between the tempo and the loudness. *Directness of movement* is a feature that captures aspects of the curvature of a trajectory segment: it measures the ratio between the length of a direct movement from A to B (α) and the length of the actual trajectory between the same points (β) in a bi-dimensional space (see Figure 2). These features were calculated in different spaces: in the time-tempo space ($\Delta_{tB}(w_i)$), in the time-loudness space ($\Delta_{tL}(w_i)$) and in the tempo-loudness space ($\Delta_{BL}(w_i)$). Finally, some derivatives were calculated for tempo and loudness: the maximum of the absolute value of the derivative ($M\delta(w_i)$) and the average of the absolute derivative ($\mu\delta(w_i)$).

Table 3 summarizes all the features. The loudness-derived features that cannot be used in the learning process are also shown.

ID	WEKA Name
01	trees.j48.J48
02	bayes.NaiveBayes -K
03	*lazy.kstar.KStar*
04	meta.ClassificationViaRegression -W LinearRegression
05	misc.VFI
06	trees.DecisionStump
07	rules.ConjunctiveRule
08	rules.Ridor

Table 4: Learning algorithms used in the experiments.

2.3. The learning algorithms

For our first experiments, we selected a representative set of standard machine learning algorithms with different representations and *biases* (see Table 4). All of these are available in the Waikato Environment for Knowledge Analysis (WEKA)[1] [10], and the names given in Table 4 are the names (including parameters) by which they are called in WEKA.

The following learners were selected: J48 is a state-of-the art decision tree learner; NaiveBayes is a probability-based algorithm that applies Bayes' rule for prediction and assumes independence between the individual features; KStar is a Nearest-Neighbor classification algorithm; *ClassificationViaRegression* is a 'meta-learner' that induces linear discriminant functions for the individual classes and combines these into an n-class classifier by voting; VFI ('Voting Feature Intervals') is an extremely simple classifier that lets each feature vote for the class in isolation; DecisionStump learns extremely simple one-level decision trees; ConjunctiveRule is a similarly simple algorithm that learns one conjunctive decision rule per class; and Ridor is a algorithm for directly learning classification rules from examples and generating exceptions for the default rule with the least (weighted) error rate.

All of these algorithms read the same data format and produce predictive models for discrete classification problems.

2.4. Testing methodology

Two kinds of aspects of the learned models are of interest: *qualitative* — do the models capture relevant and interpretable aspects of performance style? can we learn anything new about performance from them? — and *quantitative* ones — how well can the classifiers identify performers in new recordings? what is the recognition rate that can be achieved?

[1]The Java source code of WEKA is publicly available at `www.cs.waikato.ac.nz`

	Classifiers									
Piece	**01 [%]**	**02 [%]**	**03 [%]**	**04 [%]**	**05 [%]**	**06 [%]**	**07 [%]**	**08 [%]**	**Average [%]**	**DEF [%]**
kv279_1	30.52	34.24	23.75	32.65	16.81	20.90	20.48	29.39	26.09	16.68
kv279_2	17.91	15.97	25.15	33.96	14.10	21.27	23.21	28.81	22.55	17.24
kv279_3	32.19	27.05	26.89	26.52	18.42	17.36	16.94	24.46	23.73	16.73
kv280_1	19.30	23.87	19.85	32.61	16.83	18.84	20.27	18.96	21.32	16.71
kv280_2	20.07	30.69	19.98	25.34	21.75	20.68	22.08	24.87	23.18	16.77
kv280_3	16.68	21.42	18.44	19.67	17.12	18.00	16.68	17.12	18.14	16.68
kv281_1	17.15	13.25	15.12	24.23	15.51	17.92	21.86	21.90	18.37	16.73
kv282_1	25.81	21.93	29.63	30.57	18.34	23.57	23.40	29.22	25.31	16.93
kv282_2	21.50	28.10	24.70	34.75	17.42	24.36	24.15	25.41	25.05	16.62
kv282_3	33.50	33.66	26.45	26.95	16.54	18.76	18.10	32.27	25.78	16.71
kv330_3	19.31	15.35	17.60	28.05	16.28	23.85	21.51	18.28	20.03	16.72
kv332_2	22.88	27.69	31.97	33.54	9.98	27.80	16.67	29.94	25.06	16.67
Average	23.07	24.44	23.29	29.07	16.59	21.11	20.45	25.05	–	16.76
Weighted Average	22.85	24.42	22.88	29.41	16.66	21.17	20.68	24.76	–	16.75

Table 5: Preliminary results: classification accuracy; DEF refers to the *default (baseline) accuracy*, i.e., the accuracy one would achieve by always predicting the class that is most frequent in the training data. 'Weighted average' is the mean classification accuracy when weighted by the relative size (number of instances) of the different test pieces.

In the first experiments, we focused on quantitative issues, as these are easier to measure, in particular when we learn many different models with different learning algorithms. The first question to be studied was to what extent it is possible for machine learning algorithms to learn to identify performers in new recordings, and which are the most promising learning algorithms. This was tested via *cross-validation* at the level of pieces (sonata movements): each of the algorithms was trained on all of the sonata movements except one, the learned classifiers were then tested on the remaining movement, and the percentages of correct predictions were recorded. This process was repeated in a circular fashion, so that each sonata movement served as test piece exactly once for each classifier. The important thing about this procedure is that the predictivity of learned classifiers is always tested on independent data that was not used in the training phase, so that we get a realistic estimate of a classifier's expected performance.

3. RESULTS

The results of these cross-validation experiments are summarized in Table 5, which lists the *classification accuracies* achieved by the individual classifiers on each of the test pieces (after having been trained on the other pieces). Classification accuracy is defined as the percentage of instances in the test piece that were assigned the correct class by the classifier. This is to be compared to the so-called *default* or *baseline accuracy*, which is the success rate one would achieve by 'intelligent guessing', i.e., by always predicting the class that is most frequent in the training data.

At first sight, the results look disappointing. The accuracies achieved by the individual classifiers are far from the optimum of 100%; they range from 9.98% (classifier 05 on test piece kv332_2) to 34.75% (classifier 04 on kv282_2) on individual pieces, and from 16.66% to 29.41% on average. This is put into perspective by noting that 6-way classification is a difficult problem: the default accuracy would be around 16.67% (see the last column in Table 5). All the classifiers predict significantly above this baseline on average, except for classifier 05, which obviously fails to learn anything sensible. So clearly, there is significant information in the performance data that can contribute to identifying the performer, even though the data in its current form is very abstract and incomplete.

Looking at the results piece by piece reveals that performer identification may be easier in some pieces than in others. The average performance over all classifiers on a piece-by-piece basis (see penultimate column in Table 5) is above the baseline in every case, ranging from 17.89 (kv280_3) to 25.05 (kv279_1). If we remove classifier 05 from the table, the average accuracy achieved by the remaining classifiers is even higher, ranging from 18.29 (kv280_3) to 27.42 (kv279_1).

A closer look shows that not all classifiers perform well or poorly on the same pieces (see, e.g., test piece kv279_2, where classifier 02 achieves its third-poorest result, while classifier 04 achieves its second-best). That indicates that it may be fruitful to join classifiers into so-called *ensembles* [2] which combine their predictions, e.g., by voting on the class of a new test case.[2] It is known from systematic research in machine learning that classifier combination is particularly promising if the errors of the individual classifiers are highly uncorrelated [1, 8]. Experiments with ensembles of classifiers are currently on our agenda.

In order to analyze which pianists are easier or more difficult for the classifiers to recognize, Figure 3 shows a so-called *recall-precision diagram*. Recall and precision are concepts from the field of information retrieval; they give an indication of the trade-off between being able to recognize (or retrieve) as many instances of a given target class as possible (true positives), and erroneously classifying other instances as belonging to the target (false positives). More precisely, assume there are P true instances of the target class C and N instances of other classes $\overline{C}_i$, and that a classifier classifies tp instances correctly as C, fp incorrectly as C (while they really belong to one of the $\overline{C}_i$); then *recall* is defined as tp/P and *precision* as $tp/(tp + fp)$.

As Figure 3 shows, the results for different pianists occupy different regions in recall-precision space. Glenn Gould (GG) seems to the most easily recognizable pianist,[3] relatively speaking: most of the learners manage to correctly recognize between 35 and more than 60 % of the examples related to Gould, with a precision that

[2] Another indication for the promise of ensemble learning is the fact that classifier 04, which is in itself a kind of simple ensemble learner, performed significantly better than any of the other learners in our experiment.

[3] Not surprisingly, some might say ...

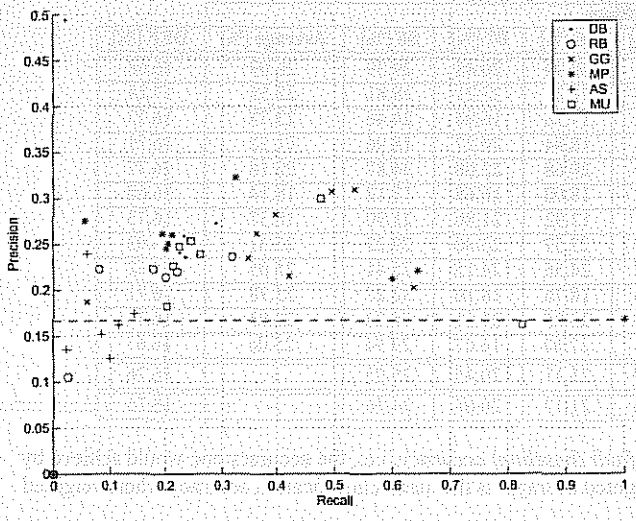


Figure 3: Recall-precision diagram showing recognition results for different pianists and all classifiers. The pianists are identified by different point types. The horizonal line at $y = .1667$ indicates the *default precision* that would result if we always predicted the artist we are interested in (which would give the optimal recall of 1.0) — note that one of the classifiers actually did this with Maria João Pires (MP).

is well above the baseline precision (between .2 and .315). Also Maria João Pires (MP) and Daniel Barenboim (DB) seem to have recognizable characteristics. On the other hand, all the classifiers have problems recognizing András Schiff. This should, however, not be misconstrued to mean that Schiff has a less individualistic style. All we can say at the moment (especially given the very preliminary phase of our data analysis) is that the current set of learning algorithms finds it easier to find partial discriminant descriptions of Gould's style, given the current set of features. More detailed experiments will be needed to learn something meaningful about aspects of style.

Again, the generally quite low recall and precision values may seem disappointing, but it must be kept in mind that these values derive from an n-way classification setting. The learners tried to learn a model that distinguishes between all the pianists simultaneously, rather than focusing on one particular pianist and trying to distinguish him or her from the others. We expect to get much better results if the problem is turned into n two-class discrimination tasks.

All the accuracy figures reported above refer to the classification of individual test examples (which are not entire pieces, but specific time points in a performance, with a window around them). That is, what is counted is how many of the individual instances each of the classifiers assigned to the correct pianist. In a more natural classification scenario one might be interested in the classification of an entire performance: who was the pianist behind a given recording? The simplest way to do that would be to assign the piece to the pianist who collects the most predictions over the set of instances that make up the piece. We do not have these figures ready at this moment (that will require some rewriting of the experiment scripts), but will have them available at the conference.

In summary, the first experimental results do indicate that it may be possible for a machine to recognize famous artists from their style, at least to some extent. The machine learning algorithms achieved recognition rates significantly above the baseline, which indicates that they can pick up some relevant information from the recordings. At first sight, the absolute accuracy figures look rather poor. However, our current training data (performance measurements) are extremely crude and incomplete (only beat-level tempo and beat-level overall loudness; no information about the loudness of individual voices, about articulation, about the timing of individual voices, etc.). Also, n-class identification task are notoriously difficult, and we expect to get much better results by converting the original problem into a large number of 2-way discrimination problems (one pianist against all others, or pairwise discrimination), and by employing more complex learning schemes. Some plans along these lines are listed below.

4. NEXT STEPS

This paper has presented the first preliminary experiments that study whether a machine can learn to recognize famous performers (concert pianists) based on aspects of their style of playing. Only very high-level and incomplete performance information was available for the experiments, and the results were mildly positive. Many improvements are possible, and the following next steps are currently on our research agenda:

- classification of entire pieces instead of individual instances (for instance, by voting over the instances);
- experiments with combinations of several different classifiers (*ensemble learning methods* [2]);
- reformulation of the n-class problem as n two-class problems (distinguishing one pianist from all the others) and appropriate combination of the resulting classifiers to solve the original n-class identification problem; that might also lead to learned models that tell us something about aspects of individual style;
- *round-robin* learning [5], i.e., breaking up the n-class problem into $\frac{n(n-1)}{2}$ two-class discrimination problems, one for each unique pair of artists; it has been shown that in this way classification accuracy may be significantly improved (and, surprisingly, also the efficiency of learning in terms of run-time);
- identification of those features that contribute to the discernibility of the individual pianists; by reducing the set of features to those that are truly relevant, additional improvements in classification accuracy may also be achieved;
- refinement of the performance measurements; we are working on methods to extract more detailed information regarding timing and dynamics from audio recordings.

We are confident that the current results can still be considerably improved, and that the models induced by the learning algorithms in the two-class discrimination tasks will provide some interesting insights into some of the things that make up the recognizable style of a great artist.

5. ACKNOWLEDGMENTS

This research was supported by an ERASMUS scholarship to the first author, the EC project HPRN-CT-2000-00115 MOSART, and

the project Y99-INF (START Prize by the Austrian Federal Government). The Austrian Research Institute for Artificial Intelligence (ÖFAI) acknowledges basic financial support from the Austrian Federal Ministry for Education, Science, and Culture.

6. REFERENCES

[1] Bauer, E. and Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning* 36:105–169.

[2] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In J. Kittler and F. Roli (Ed.), *First International Workshop on Multiple Classifier Systems*. New York: Springer Verlag.

[3] Dixon, S., "Automatic Extraction of Tempo and Beat from Expressive Performances", Journal of the New Music Research, 30(1):39-58, 2001.

[4] Dixon, S., Goebl, W., and Widmer, G., "The Performance Worm: Real Time Visualization of Expression based on Langner's Tempo-Loudness Animation", Proc. International Computer Music Conference (ICMC 2002), Göteborg, Sweden, pp. 361-364, 2002.

[5] Fürnkranz, J. (2002). Round Robin Classification. *Journal of Machine Learning Research* 2:721–747.

[6] Langner, J. and Goebl, W. (2002). Representing Expressive Performance in Tempo-Loudness Space. *Proceedings of the ESCOM Conference on Musical Creativity*, Liège, Belgium.

[7] Stamatatos, E. (2002). Quantifying the Differences between Music Performers: Score vs. Norm. *Proceedings of the International Computer Music Conference (ICMC'2002)*, Göteborg, Sweden.

[8] Stamatatos, E. and Widmer, G. (2002). Music Performer Recognition Using an Ensemble of Simple Classifiers. *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI'2002)*, Lyon, France.

[9] Widmer, G. (2001). Using AI and Machine Learning to Study Expressive Music Performance: Project Survey and First Report. *AI Communications* 14(3), 149–162.

[10] Witten, I.H. & Frank, E. (1999). *Data Mining*. San Francisco, CA: Morgan Kaufmann.

[11] Zwicker, E. and Fastl, H. (2001). *Psychoacoustics. Facts and Models*. Springer Series in Information Sciences, Vol.22. Berlin: Springer Verlag.

PSYCHOACOUSTIC EXPERIMENTS FOR VALIDATING SOUND OBJECTS IN A 2-D SPACE USING THE SONIC BROWSER

Laura Ottaviani

Dipartimento di Informatica
Università degli Studi di Verona
Strada Le Grazie, 15 - 37134 Verona, Italy
ottaviani@sci.univr.it

Eoin Brazil and Mikael Fernström

Interaction Design Centre
Dept. of Computer Science and Information Systems
University of Limerick, Ireland
Eoin.Brazil@ul.ie
Mikael.Fernstrom@ul.ie

ABSTRACT

The Sonic Browser is a software for navigating among sounds, in a bi-dimensional space, using the hearing system. Its two main purposes are its use for conducting psychophysical experiments, and its application for managing catalogues of huge sound collections.

In this paper, we focus on the former scenario, presenting our recent psychophysical experiments using the Sonic Browser for validating the sound models developed by the Sounding Object project, looking at the quality of the synthesized sounds and the relationship between the physical parameters involved in the sound synthesis.

1. INTRODUCTION

The Sonic Browser is a software, born at the Interaction Design Centre in the University of Limerick in 1996 and in continuous improvement since then [1, 2], which allows the user to navigate a bi-dimensional multimedia space primarily through listening.

This tool represents a useful application for the Sounding Object (SOb) project[1]. The SOb project has pioneered several recent attempt to synthesize sounds with physical-based sound models, trying to reduce their degree of generality, while maintaining their information effectiveness. This is accomplished by the process of cartoonification, which exaggerates certain sound features, discarting other aspects. In this way, the information is clearer and more effective, and the computational load for the physical models is reduced, allowing them to be controlled interactively by means of gestural or graphical interfaces [3].

The Sonic Browser main purpose is managing huge catalogues of sound collections, a difficult problem in the sound designer community and among Foley artists. The development of sound models, whose parameters control the sound source characteristics, has lead to the use of the Sonic Browser for conducting psychophysical experiments, in order to test and validate the sounds produced by these sound models. [4].

In this paper, we focus on this latter application by comparing Sound Objects to real sound recordings and investigating the perceptual scaling of the physical parameters that control the sound models, which yelds data that can be used to inform and calibrate the features of these sound models.

The Sonic Browser can be useful in this analysis for three main reasons. First, it allows the users to browse the sound space, providing fast and direct access to the sounds and making the task more natural to them. Second, it lets the users move the sounds according to a bi-dimensional evaluation scale and, therefore, create their own perceptual space. Finally, there is the main feature of the Sonic Browser: the *aura*, the circular area, in the tool interface, surrounding the cursor, which can be resized or even turned off. The user listens simultaneously to all the sounds included in the aura. This feature of the tool, which represents the heart of the Sonic Browser and what differentiates it from other tools [5], despite not being so important in the validation scenario as it is for the cataloguing scenario, it has demonstrated to be very useful to users, who applied it for comparing stimuli and previous estimations.

In this paper, we will introduce the experiments, showing the results[2].

2. THE EXPERIMENTS

The experiment aims to begin to understand how the synthesized sounds produced by our models are scaled in comparison with physical dimensions, by means of the Sonic Browser. In this experiment we focused on two dimensions: perceived height of the object drop and perceived size of dropped objects, that specified the scales of the 2-D plot in the application. Our exploration was not limited only to the scaling task, but also encompassed the perceived realism of the event. Therefore, we divided the experiment in two phases, one concerned with the scaling task per se and the other one focused on the realism judgement. Moreover, as we wanted to compare the sound rendering of two different approaches in the sound modelling, the stimuli set included, besides recorded events, Sound Objects from both of these modelling approaches.

The experiment was preceded by a pilot experiment, which used only one modelling approach. The pilot probe allowed for a first observation of the type of results and it highlighted which sounds were most suitable to focus on in the main experiment.

Two techniques were involved for collecting experimental data. First, data logging was collected by the application for the object positioning in the 2-D space. Second, the user was asked to comment aloud on the thinking process, as it is established by the *Thinking Aloud Protocol*.

The Thinking-Aloud Protocol is widely used in usability testing and it represents a way for the experimenter to have a "look"

[1] http://www.soundobject.org

[2] Further details and more data representations are available at http://www.soundobject.org

at the participants' thought processes [6]. In this approach, the users are asked to talk during the test, expressing all their thoughts, movement and decisions, trying to think-aloud, without paying much attention to the coherency of the sentences, "as if alone in the room".

Employing this protocol, we were able to collect not only the data concerning the stimuli positions in the 2-D space of the Sonic Browser, but also the comments of the users during the experiment, which expressed the reasons, for instance, of a particular judgement or their appreciation of the Sound Objects realism. The tests were all recorded by a video-camera.

In the next subsections we will present both the pilot and the main experiment, introducing procedures and results of the test.

2.1. Participants

The pilot probe and the main experiment involved respectively 4 and 5 volunteers, all students or workers at the Computer Science Department of the University of Limerick. All of them referred to have a musical training, in average 5 years for the pilot probe, while 8 years for the main experiments, in the following ranges respectively: 2–10 and 6–10 years. No participant reported to have hearing problems, while two of the participants to the main experiment required glasses for reading.

2.2. Stimuli

The stimuli sets of both the experiments included recorded sounds and Sound Objects and consisted of 18 sounds, but in a different proportion: 9 recorded and 9 synthesized for the pilot probe, while 6 real and 12 synthetic for the main experiment.

The recorded sounds were produced by 3 steel balls, weighing 6, 12 and 24 g, and falling on a wooden board of 1500 x 500 x 20 mm from a height of 10, 20 and 40 cm, respectively, by positioning the microphone at 3 different distances: 20 - 40 - 80 cm, respectively. Recordings were done with a MKH20 Sennheiser microphone, and a 16 bit sound card sampling at 44.1 kHz rate.

These stimuli were used in previous experiments conducted by the SOb project on the perception of impact sounds [7]. In this study, Burro found the relationship between the physical quantities of weight, distance and height and the relative perceptual quantities. He argued that manipulating one of the physical parameters affects more than one of the perceptual quantities.

In the pilot probe, we decided to keep the height of the dropped balls constant (h=20 cm), while in the main experiment we kept constant distance (d = 80 cm), while changing the height.

All the synthesized sounds in the pilot probe and 6 in the main experiment were designed with the PD-modules modelling impact interactions of two modal resonators [8], simplified returning only one mode, and they used either glass or wood as the material property. On the contrary, the remaining 6 synthesized stimuli of the main experiment were designed with the complete model of the impact interactions and the dropping event, as well. In this latter case, we preferred to keep the material constant, since we noticed some difficulties during the pilot probe for the users to evaluate and compare the dimensions of events involving different material. We decided on wood as the material, even if it is not clear if the wood is the material of the impactor or of the surface. In fact, even if the real sounds come from steel balls, they were referred to by the participants as wood. This perception arouse from the bigger influence of the surface material in certain cases.

2.3. Procedure

The two experiments were conducted in the isolation room of the recording studio in the Computer Science Department at UL. The stimuli were presented by stereo headphones to the users through the Sonic Browser. The experiments were conducted applying the Thinking-Aloud Protocol and the participants sessions were all recorded on video-tapes. A more detailed analysis of the results of verbal protocol was conducted [4] and is available online[3].

After the perception estimation task, the participants, during the second phase of the tests, were asked to tag the sounds that they thought were unrealistic.

At the end of each session, a seven point Likert scale questionnaire with six sets of semantic differentials was filled out by each participant who were asked to express their responses to the interfaces and to the tasks, from 0 to 6, where 0 is "poor" and 6 is "excellent". In the main experiment, three questions, asking about the learnability, interpretation of the application and the difficulty in replaying the last sound, were added to the questionnaire after the pilot probe.

The users estimated the data positions in the bi-dimensional scales without a comparison stimulus or a reference scale. Despite being pre-defined, i.e. being limited to the screen, the ranges of perceptual evaluations were relative to each user. The perceptual space boundaries were considered by all the users relative to their maximum value, as they reported at the end of the task. In fact, we noticed an initial difficulty by the participants of referring to the screen space. On the contrary, they showed a preference of defining their own boundaries. In order to be able to compare the results of each participant, we decided to normalize the data coordinates, which identify the locations in the 2-D space, between 0 and 1.

2.4. Results and Observations

2.4.1. The pilot probe

From a global observation of the collected data, we can notice that the participants estimated correctly the height from the real sounds, h=20 cm for all of them, since most of the real sounds, barring five outliers, were positioned by the users in the central area of the evaluation space. On the other hand, the size estimation varies to a degree between users. This could be influenced by either the distance and/or the conditions in which the real sounds were recorded.

Looking at the individual perceptual scaling and tagging information sorted by users, reported in fig. 1, we notice that two participants in particular (users 2 and 3) made an obvious distinction between real and synthetic sounds.

It is interesting to observe the single stimuli, looking at fig. 2 which represent, through a box plot, the individual perceptual scaling of height and size respectively sorted by stimuli. For our purposes, we will focus on the synthesized sounds.

From these plots, we can see different perceptual scaling by the users along the two dimensions. In fact, we can find some sounds that were judged coherently by most of the users, at least along one dimension, while others (sound10, sound16 and sound17) whose scaling is spead across the evaluation space, showing a difficult by the participant to estimate them. It is interesting to notice that sound17 was tagged as unrealistic by all the participants to the probe. Therefore, the data spread could be due to the lack of

[3] `www.soundingobject.org`

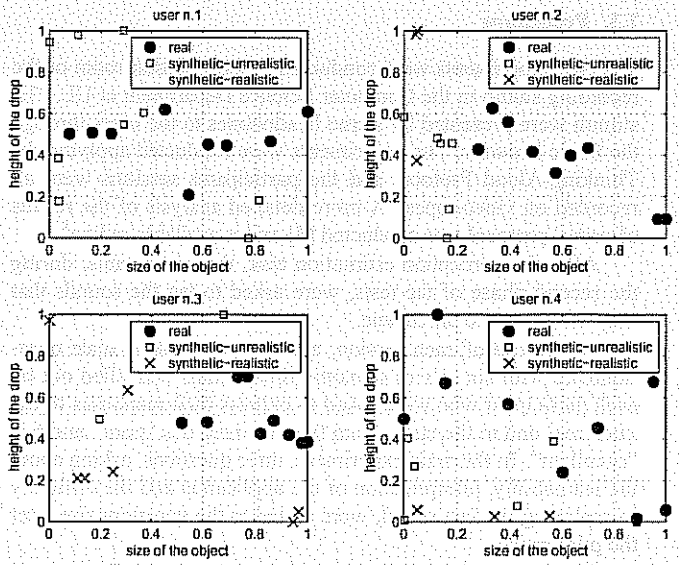


Figure 1: *Pilot probe: representation of the individual perceptual scaling and tagging information sorted by users.*

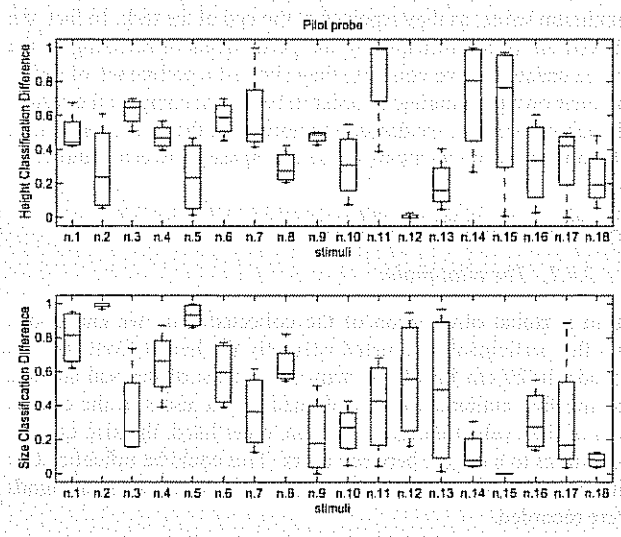


Figure 2: *Pilot probe: representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli.*

realism provided by the sound. On the other hand, sound18 was judged uniformly by all the users, and especially for the size dimension. Finally, the other five stimuli of the synthesized set were all judged uniformly in one dimensions.

In fig. 3, the results of the questionnaire, with cumulative participant responses displayed per question, can be seen, with 0 representing a negative result to the question and 6 a positive one.

We can see that while the task was found to be non trivial (question 1), the users rated the ease of use of the application above average (question 2). The participants judged the sounds to be realistic (question 3) and of high quality (question 4). In the application, there is a slight delay of up to 0.3 of a second when playing

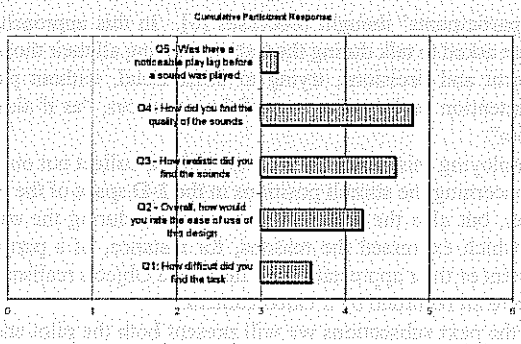


Figure 3: *Results of the questionnaire for the pilot probe.*

an audio file. The users found it to be acceptable but noticeable (question 5).

2.4.2. The main experiment

In fig. 4, we report the data from the main experiment sorted by users. As for the pilot experiment, we can see the classification by sound groups. Moreover, we notice that two of the participants (user 1 and user 2) only performed minor judgements on size of the real sounds. They referred, in fact, that they perceived other parameters changing, such as distance and material. This complex influence of the three parameters has already been discussed by Burro [7].

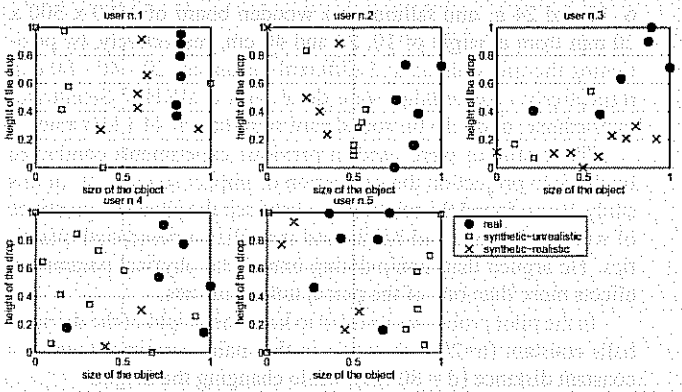


Figure 4: *Representation of the individual perceptual scaling and tagging information sorted by users.*

It is interesting to observe the single stimuli, as we did for the pilot probe, looking at fig. 5. We again focused on the synthesized sounds.

We can observe that there is more uniformity in perceptual scaling in two dimensions, than in the pilot experiment. For instance, four stimuli were judged uniformly by the participants, where sound7 and sound11 could be considered to show a strong uniformity in both the dimensions, while sound10 and sound14 a slight uniformity in either dimension. Moreover, four of the stimuli were perceived uniformly in one dimension, and, finally, three

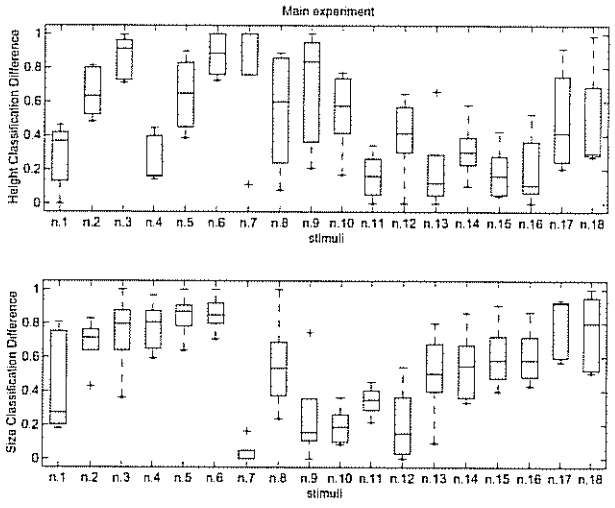


Figure 5: *Representation, by a box plot, of the perceptual scaling of the height and size sorted by stimuli.*

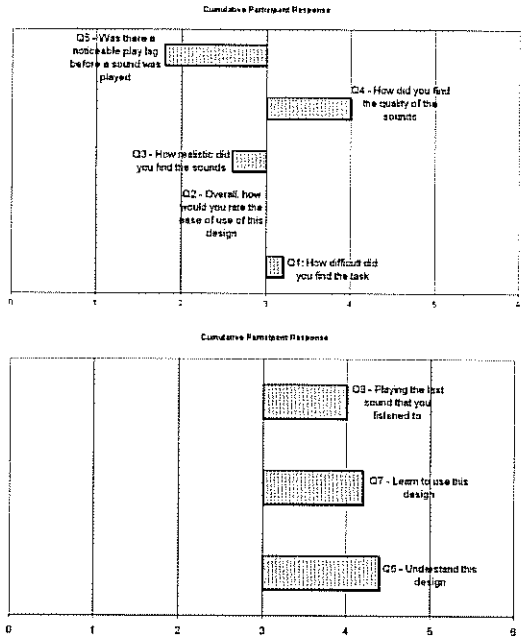


Figure 6: *Results of the questionnaire for the main experiment: the same questions presented in the pilot probe and additional ones.*

stimuli (sound8, sound16 and sound18) were dispersed in the perceptual estimation space and one, sound17, highly dispersed.

Contrary to the results of the tagging task in the pilot probe, there is no stimuli, in this experiment, that was tagged by all the participants. The maximum user consensus regarding unrealistic stimuli was achieved by 3 users. The real sounds were perceived again as realistic, duplicating the results of our pilot study.

In fig. 6, as we did for the pilot probe, we report the results of the questionnaire with cumulative participant responses displayed per question, highlighting the three additional questions.

We can see that the task was found to be non trivial (question 1) and the ease of use of the application was rated only on average (question 2). The participants judged the sounds to be of high quality (question 4), but not so realistic (question 3). This can be attributed to the inclusion of two different types of sound objects containing either one or two modes as well as the lack of room acoustics within the sound object sounds and the presence of a "buzz tail" at the end of the two mode sound object sounds. The slight delay when playing an audio file resulted in this experiment to be very noticeable to the users (question 5). This fact is evident by the verbal protocol, as well, since many participants were irritated by it. As what the added questions concern, the application was judged to be easy to understand (question 6) and to learn (question 7) and, also, to play back a sound (question 8), despite the slight delay.

3. CONCLUSIONS

Examining the results we can state that sound convey information about dimensions even if they have only one mode. Apart from one case in the pilot probe, the unrealistic perception of sounds did not affect the participant's perception of the sound's dimensions. This illustrates that the "realism" of a sound does not affect the amount of information extracted by a participant. Moreover, unrealistic synthetic sounds can be recognized as unrealistic events but their high-level parameters can still be extracted and evaluated, as it is stated by the technique of sound cartoonification. This issue was already discussed by Rath [9], focusing on sound objects in combination with cartoonification.

Our experimental analysis and our user debriefing suggest three aspects which should be further investigated. First, the inclusion of "room acoustics" and the necessary elements of reverberation within a sound object to allow for a more natural sounding event. Second, to further investigate the material and surface perception, because it was pointed out frequently by the participants. Third, to study the distractors found within the sounds used in the experiments, concerning the relationship between speed as well as temporal pattern of the bouncing and the realism of the sound, and the "metallic zips" occurring at the end of each of the two mode sound objects. These distractors illustrate the need for a further refinement of the perceptual parameters within the sound models to prevent user confusion when judging a sound's properties.

4. ACKNOWLEDGEMENTS

This research has been supported by the European Commision under contract IST-2000-25287 ("SOb - the Sounding Object"). `http://www.soundobject.org`. We would like to express our thanks to all the postgraduate students and workers at the University of Limerick who participated to these experiments. We say thank you to the members of the Interaction Design Centre at UL for their useful comments and accommodating spirits.

5. REFERENCES

[1] E. Brazil and M. Fernström, "Let your ears do the browsing - The Sonic Browser", *The Irish Scientist*, 2001

[2] E. Brazil, M. Fernström, G. Tzanetakis and P. Cook, "Enhancing Sonic Browsing Using Audio Information Retrieval", *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, July 2-5, 2002

[3] D. Rocchesso, R. Bresin, M. Fernström, "Sounding Objects", IEEE Multimedia, 2003. In press.

[4] E. Brazil, L. Ottaviani and M. Fernström, "Experiments in a 2-D space using the Sonic Browser for validation and cataloguing of Sound Objects". In *Perceptual Assessment of Models*, pages 40-76. SOb project, 2002. Available at `http://www.soundobject.org`.

[5] G.P. Scavone, S. Lakatos and C.R. Harbke, "The Sonic mapper: an interactive program for obtaining similarity ratings with auditory stimuli", *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, July 2-5, 2002.

[6] M.T. Boren and J. Ramey, "Thinking aloud: Reconciling theory and prac-tice", *IEEE Transactions on Professional Communication*, 43(3):261-278, September 2000.

[7] R. Burro. "Interaction effects in bouncing ball". In *Physically-based Sound Spaces and Psychophysical Scales*, pages 23-35. SOb project, 2002. Available at `http://www.soundobject.org`.

[8] M. Rath, "Sound design around a real-time impact model". In *Models and algorithms for sounding objects*, pages 32-45. SOb project, 2002. Available at `http://www.soundobject.org`.

[9] M. Rath, "Some audio cartoons of fluid movements". In *Models and algorithms for sounding objects*, pages 57-59. SOb project, 2002. Available at `http://www.soundobject.org`.

MEMORY AND MELODIC DENSITY: A MODEL FOR MELODY SEGMENTATION

Miguel Ferrand, Peter Nelson — University of Edinburgh, Scotland, UK, {mferrand,pwn}@music.ed.ac.uk

Geraint Wiggins — City University, London, UK, geraint@soi.city.ac.uk

ABSTRACT

We present a memory-based model for melodic segmentation based on the notion of melodic density. The model emphasises the role of short-term memory and time in music listening, by modelling the effects of recency in the perception of boundaries. We describe the model in detail and compare it with Cambouropoulos' Local Boundary Detection Model for a series of melody examples. First results indicate that this new model is more conservative, as it generates fewer total boundaries but preserves most boundaries that coincide with the limits of recurring patterns.

1. INTRODUCTION

It is known that listeners identify segmentation boundaries when abstracting musical contents. The ability to partition a melody in several segments provides a structural description of the piece of music. Thus, segmentation can be seen as a pre-processing stage for other tasks such as pattern discovery or music search.

Pattern finding algorithms, in particular, are known to be computationally expensive, and therefore can benefit from a reduction of the initial search space. A low-level segmentation can provide an efficiency gain by pre-processing a melodic sequence, and generating an initial set of boundaries which may be used as markers for pattern search [1]. One such method is The Local Boundary Detection Model (LBDM) [2], a segmentation model that identifies discontinuities in a melodic surface based on Gestalt principles of perception. The LBDM is an essential reference amongst segmentation algorithms, mostly due to its simplicity and generality [3, 2]. As the author emphasises, the LBDM is not a complete model of grouping in itself, as it relies on complementary models (i.e. pattern similarity) to select the most relevant boundaries. Although in that context this may not be considered a weakness of the model, excessive boundary generation may become a disadvantage if we intend to use the LBDM in isolation, and when segmentation is to be used as a reliable data reduction technique.

The LBDM has a fairly short memory as it considers at most 4 consecutive events at a time. As a consequence, there is limited interaction between neighboring boundaries and sometimes small "oscillations" can be identified as salient boundaries. This type of limitation has also been referred to by Lerdahl & Jackendoff in their Generative Theory of Tonal Music [4].

Research on auditory perception and memory has underlined the influence of time in the perception of differences and in the establishment of temporal relations in sequential processes. Studies have shown that listeners retain auditory information for some time, even after the end of stimulation [5]. This means that several past (although relatively recent) stimuli may draw the listener's attention, and may be retained as the actual most recent and prominent stimuli. Some researches have suggested that listeners perceive a musical surface by focusing on successive zones, that can be viewed as a "sliding window" along the musical piece [6]. The size of this window (determined by short-term memory restrictions) should limit the amount of musical material that can be looked back on when processing a melodic sequence. Within this time window, recency effects are likely to apply, as documented in [7, 8].

2. THE LBDM

The LBDM calculates a boundary profile for a melody, using Gestalt-based identity–change and proximity–difference rules, applied to several parameters describing a melody. The refined version of this algorithm [2] takes as input a melodic sequence converted into several independent parametric interval profiles $P_k = [x_1, x_2, ...x_n]$ where $k \in \{pitch, ioi, rest\}, x_i \geq 0$ and $i \in \{1, 2, ...n\}$. A *Change* rule assigns boundaries to intervals with strength proportional to the degree of change between neighboring consecutive interval pairs. Then a *Proximity* rule scales the previous boundaries proportionally to the size of the intervals.

The strength of the boundaries at each interval x_i is given by the following,

$$s_i = x_i \times (r_{i-1,i} + r_{i,i+1}) \tag{1}$$

where

$$r_{i,i+1} = \begin{cases} \frac{|x_i - x_{i+1}|}{x_i + x_{i+1}} & x_i + x_{i+1} \neq 0 \wedge x_i, x_{i+1} \geq 0 \\ 0 & x_i = x_{i+1} = 0 \end{cases}$$

For each parameter k a sequence s_k is calculated, then all sequences are normalised and combined in a weighted sum to give the overall boundary strength profile. The suggested weights for the 3 different parameters are $w_{pitch} = w_{rest} = 0.25$ and $w_{ioi} = 0.5$ (see [9] for an overview on the behavior of the LBDM with different parameter tunings). The local peaks in the resulting boundary profile indicate local boundaries in the melodic sequence. A threshold must be defined a priori, above which, a peak is identified as a boundary. For additional details on the implementation of the LBDM the reader is referred to [2].

3. MELODIC DENSITY SEGMENTATION MODEL

We now describe a new model for melodic segmentation which identifies segmentation boundaries as perceived changes in melodic

Table 1: *Order* and *recency* of pitch intervals for a sequence of events. Intervals are in semitones.

e_{i-3}	e_{i-2}	e_{i-1}	e_i	*event*
53	50	50	48	*pitch*
				$order(n)$
	3	0	2	1
		3	2	2
			5	3
...	2	1	0	$recency(m)$

density. We will designate this model as Melodic Density Segmentation Model (MDSM). In contrast with the LBDM, that measures the accumulated boundary strength and identifies local maxima, *the MDSM calculates the accumulated melodic cohesion between* pitch intervals, and then identifies local minima (i.e. points of low melodic density) as local boundaries. This new segmentation method also incorporates a short-term memory window and models the effects of recency with an attenuation function.

Before a formal description of the model is presented, some of its characteristics and underlying assumptions must be explained.

It is conjectured that pitch intervals may be formed (and perceived) between *all* notes occurring over an interval of time (short term memory window) and not just between consecutive notes. In Table 1 a short sequence of 4 midi notes is depicted together with the pitch distances between all pairs of events. The order of an interval determines the distance between the present and previous event considered. Thus, an interval of order k with respect to a given event e_i is denoted by (e_{i-k}, e_i). For example, from table 1 intervals (e_{i-1}, e_i) and (e_{i-2}, e_{i-1}) have order 1, intervals (e_{i-2}, e_i) and (e_{i-3}, e_{i-1}) have order 2, etc...

Recency effects apply in two different ways. The higher the order of an interval, the greater the temporal separation between the events, and therefore the weaker the perceived link between the two. On the other hand, more recently formed intervals have a stronger contribution to the melodic cohesion of the sequence than earlier formed ones. The recency of an interval with respect to an event e_i is given by the time that separates e_i and the latest event of the two that constitute the interval. These two factors are combined to determine the overall contribution of each interval at any given moment in time. In Table 1, recency is indicated in the bottom row. Increasing values of recency express less recent intervals. Let's consider here for simplicity, that all events in the previous example have equidistant on-set times and equal duration. Then intervals (e_{i-2}, e_i) and (e_{i-2}, e_{i-1}) will have equivalent contribution, since the former is an interval of order 2 (meaning that events are separated by 2 duration units) but with recency 0, and the latter has order 1 but recency 1 (meaning that the interval is separated from the reference event e_i by 1 duration unit).

The melodic cohesion of an interval is defined here to be proportional to the frequency of occurrence of that interval in the interval framework associated with the melody being analysed. Later, we will discuss in more detail how these interval frequencies are obtained.

A short–term memory window determines the span of recent events that can form intervals. The size (duration) of this window is fixed. The tempo of the piece will determine the number of recent events that can be recalled and influence the perception of a boundary.

We can now formalise the notion of melodic density (MD) as the weighted sum of the contributions of all intervals occurring over a period of time determined by the memory window. So given a sequence of N events $(e_1, ..., e_N)$ representing a melodic sequence the melodic density d_i at event i, is defined as:

$$d_i = \sum_{m=0}^{t_i - t_{i-m} < M} \; \sum_{n=1}^{t_i - t_{i-m-n} < M} f(r_i(m,n)) \cdot a_i(m,n) \qquad (2)$$

where $f(r)$ is a function that returns the frequency of an interval, and $f(r) \in [0,1], r_i \in 0, 1, ...12$, and $r_i(m,n) = |p_{i-m} - p_{i-m-n}|$ denotes a pitch interval in semitones, where p_k denotes the MIDI pitch of event e_k, and

$$a_i(m,n) = (1 - \frac{t_i - t_{i-m-n}}{M})^2 \qquad (3)$$

is the attenuation function, where t_i denotes the onset time of event e_i, and M is the duration of the memory window (in seconds). It is worth noting that a Gestalt-based principle of proximity is encapsulated in the attenuation function, as this will return values closer to 1 for recent and low-order intervals, and values closer to 0 for remote and high-order intervals.

Finally, boundaries are indicated by local minima in the melodic density profile obtained from Equation 2.

4. EXPERIMENTS AND RESULTS

To assess the behavior of the model we used both the LBDM and the MDSM on a set of melody examples. For each of the examples we also obtained a pattern boundary profile, which indicates the location of recurrent patterns within the melodic sequence (see [1] for details).

The interval frequencies given by function f were obtained from the combined frequencies of intervals that occurr in major and minor scales. This major-minor framework is described by Camboroupoulos in his General Pitch Interval Representation (GPIR) [1]. The memory window M was set to 4 seconds.

Table 2 summarises the boundary counts for each melody, including pattern boundaries and the segment boundaries generated by both the LDBM and the MDSM. A boundary is marked correct if its location coincides with a pattern boundary, with a tolerance of +/-1 event. A threshold of 70% was adopted to filter only the most prominent peaks from the boundary profiles.

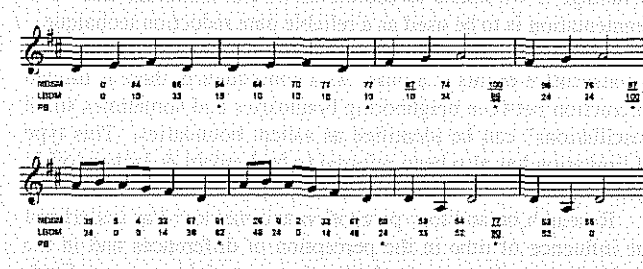


Figure 1: Normalised *MDSM* and *LBDM* boundary profiles for melody number 2 (Frere Jacques). Underlined values indicate selected peaks. Pattern Boundaries(PB) are indicated in the bottom row

Table 2: Results obtained for 7 melodies, showing the total no. of pattern boundaries (PB), and for both the LBDM and MDSM: total no. of pattern boundaries found ($_{fnd}$), no. of pattern boundaries not found ($_{notfnd}$) and no. expurious boundaries found ($_{ex}$)

		LBDM			MDSM		
Melody	PB	fnd	$notfnd$	ex	fnd	$notfnd$	ex
1. L. Row	5	5	0	0	5	0	0
2. Frere J.	7	3	4	0	5	2	0
3. Twinkle	5	5	0	2	4	1	1
4. Y.Doodle	5	5	2	3	5	0	2
5. L'H.Arme	9	8	1	0	9	0	0
6. Mozt.Gm	6	6	0	14	6	0	3
7. Beet.9th	9	9	0	0	9	0	0
Total	46	39	7	19	43	3	6

Table 3: F-measure for the LBDM and MDSM

Model	P	R	F
LBDM	0.85	0.67	0.75
MDSM	0.93	0.88	0.91

In total the LBDM generated 58 boundaries against only 49 by the MDSM. From the analysis of Table 2 it may be observed that both models find approximately the same number of pattern boundaries, but the MDSM is more conservative, generating only 6 excessive boundaries, against the 19 of the LBDM. In the melodies where excessive boundaries where found, the MDSM always register a lower count. However It must be noted that melody number 6 alone (theme of Mozart's Symphony in Gm) is responsible for the majority of the excessive boundaries generated by the LBDM. For a numerical comparison between the performance of both models the F-measure [10] was used. The F-measure is given by the weighted harmonic mean of $Precision(P)$ and $Recall(R)$.

$$F_{measure} = 2 \times \frac{P \times R}{P + R} \tag{4}$$

where

$$P = \frac{PB_{fnd}}{PB_{fnd} + PB_{notfnd}}, R = \frac{PB_{fnd}}{PB_{fnd} + PB_{excess_{fnd}}} \tag{5}$$

In table 3 we can see that although the MDSM only has a slightly higher $Precision$, it has a significantly higher $Recall$ resulting in a higher value of F.

In Figure 1 we show the boundary profiles of both models together with the score of melody no. 2 (Frere Jacques). For ease of comparison, the melodic density profile of the MDSM has been inverted [1] and normalised in the range 0-100%. From this example it seems clear that some of the boundaries generated by the LBDM were eliminated due to the 70% selection threshold, although smaller peaks can be found in the vicinity of the pattern boundaries that were missed.. An adjustment of the selection threshold to considerably lower values, will result in a significant increase of the number of peaks that are extracted, and consequently in an increase of the number of spurious boundaries. On the other hand, we would expect that an increase of the selection threshold would increase the selectivity of the model. In Figure 2 we can observe that this is not always the case. Most of the peaks of the LBDM profile have values over 80% or even 90%, thus making the elimination of the excessive boundaries difficult to achieve only by adjusting the selection threshold. The example of Figure 2 highlights also that most of the boundaries "filtered" by the MDSM are not coincident with pattern boundaries.

[1] recall that for the MDSM boundaries are obtained from the lower peaks on the profiles

5. DISCUSSION

The boundary selectivity reported on the MDSM, results partially from the propagation of the intervals over a time window creating a "smoothing" effect. However this effect can be also a drawback of this approach. In some cases, boundaries can be shifted forward or prolonged due to a slower decay of the melodic density function. This is visible in Figure 1 where the boundary peak after the third measure is followed by a significantly slow decay of the MDSM values (specially when compared with the sharp drop on the LDBM profile), until it meets the following peak. This may have an impact on the accuracy of the boundary locations, in particular when matched without tolerance, against pattern boundaries.

Although tempo was kept constant in this study, the MDSM is robust to small changes in tempo. This is mainly due to the discrete nature of the events, combined with a memory window of fixed size. For example, with a tempo of crotchet=60, a memory window of 5 seconds would include 5 crotchets (or the equivalent in duration), and an increase of the tempo to crotchet=72 would be necessary to include an additional crotchet in the calculations. Few studies have addressed the effects of changes in tempo in music perception [11]. Although the present model was designed to account for changes in tempo, a systematic evaluation of these effects has not yet been included. For such analysis we may require that listeners be tested on the effects of changes in tempo to provide data to be compared with the model.

The choice of the attenuation function (a decaying polynomial), is the result of preliminary experiments with the algorithm, where several decaying functions were examined. However, it must be said, the differences were not conclusive. It seems intuitive that, in general, less recent notes have a smaller contribution to the melodic cohesion of a sequence, than more recent ones. However, to the best of our knowledge, there is no theoretical or experimental evidence to support the choice of a specific memory decaying function.

As mentioned previously, interval frequencies were obtained from the combined statistics of interval counts from major and minor scales. Since one of the motivations of this work is to devise a model that can segment melodies without any domain specific knowledge, we propose that these frequencies may be acquired from a music corpus that is representative of the melodies being analysed. This idea is supported by several studies, some of which were carried out outside the western musical culture, that report, for example, the prevalence of small melodic intervals in melodic lines [7, 12]. If indeed the melodic preferences of a particular musical culture are reflected in the musical material, it seems rea-

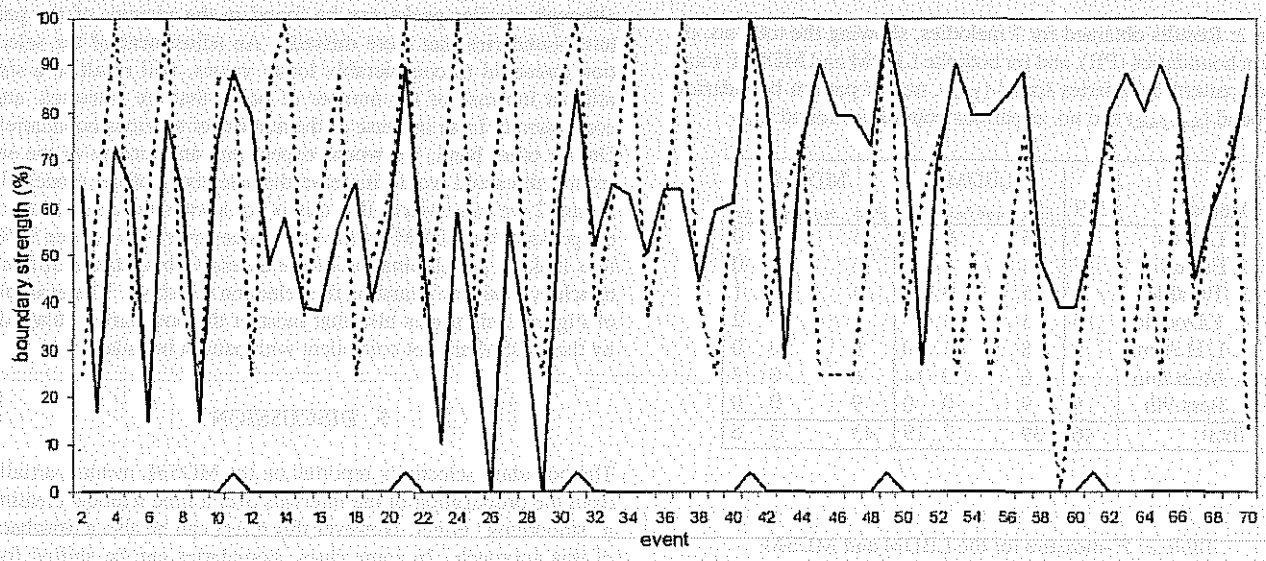


Figure 2: Boundary profiles obtained with LBDM (dotted line) and MDSM (solid line) for melody no. 6 (theme of Mozart's Symphony in Gm). Pattern boundaries are indicated by arrows at the bottom of the chart.

sonable to reverse this process, by using implicit intervalic information to interpret the musical material.

6. CONCLUSIONS

We presented the MDSM, a memory-based melodic segmentation algorithm based on the concept of melodic density. We compared this algorithm with the LBDM, for a set of melody examples. It was shown that in general the MDSM has higher selectivity than the LBDM, generating fewer total boundaries but preserving most boundaries indicated as pattern boundaries. This suggests that the MDSM may be used successfully as a pre-processing method for pattern finding algorithms, providing additional reduction of the search space without the cost of eliminating many candidate pattern boundaries.

The contribution of this new approach lies in the way it incorporates pitch and time, and in particular in the use of tempo as a parameter together with a short-term memory window, thus seeking a more cognitively realistic approach to melodic segmentation.

7. ACKNOWLEDGMENTS

This research was funded by EPSRC research project GR/N08049/01

8. REFERENCES

[1] Emilios Cambouropoulos, *Towards a General Computational Theory of Musical Structure*, Ph.D. thesis, University of Edinburgh, 1998.

[2] Emilios Cambouropoulos, "The local boundary detection model (lbdm) and its application in the study of expressive timing," in *Proceedings of the International Computer Music Conference (ICMC'2001)*, Havana, Cuba, September 2001.

[3] Emilios Cambouropoulos, "Melodic cue abstraction, similarity, and category formation: A formal model," *Music Perception*, vol. 18, no. 3, pp. 347–370, 2001.

[4] Fred Lerdahl and Ray Jackendoff, *A Generative Theory of Tonal Music*, M.I.T. Press, Cambridge (Mass.), 1983.

[5] M. Eysenck and M.T. Keane, *Cognitive Psychology: a Student's Handbook*, Psychology Press, 3rd edition, 1995.

[6] Emmanuel Bigand, "Contributions of music to research on human auditory cognition," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, Stephen McAdams and Emmanuel Bigand, Eds., chapter 8, pp. 231–277. Oxford University Press, 1993.

[7] Carol L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Number 17 in Oxford Psychology Series. Oxford University Press, Department of Psychology, Cornel University, 1990.

[8] Ludger Hofmann-Engl and Richard Parncutt, "Computational modeling of melodic similarity judgments: two experiments on isochronous melodic fragments," http://freespace.virgin.net/ludger.hofmann-engl/similarity.html, 2000.

[9] Belinda Thom, Christian Spevak, and Karin Höthker, "Melodic segmentation: Evaluating the performance of algorithms and musical experts," in *Proceedings of the 2002 International Computer Music Conference (ICMC'02*, Göteborg, Sweden, 2002.

[10] C. J. Van Rijsbergen, *Information Retrieval, 2nd edition*, Butterworths, London, Boston, 1979.

[11] Stephen Handel, "The effect of tempo and tone duration on rhythmic discrimination," *Perception & Psychophysics*, vol. 54, no. 3, pp. 370–382, 1993.

[12] Paul von Hippel, "Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals," *Music Perception*, vol. 17, no. 3, pp. 315–327, 2000.

SIMILARITY BETWEEN MIDI SEQUENCES IN POLYPHONIC CONTEXT : A MODEL OUTSIDE TIME

Benoit Meudic

Musical Representation Team
Ircam, Paris
meudic@ircam.fr

ABSTRACT

In the context of pattern extraction from polyphonic music, we challenge an approach outside time for computing the similarity between two musical sequences that neither modelizes temporal context nor expectancy. If theses notions might play a role in our perception of musical patterns, we propose in a first step to investigate the limits of a system that ignores them. Our approach relies on a new representation of the polyphonic musical sequence that is quantized in equally-spaced beat-segments and on a new definition of the notion of similarity in a polyphonic context. In agreement with ([1], [2]), we think that text-matching methods, or pure mathematical algorithms are not directly convenient for music analysis. We think that the similarity relationships between musical sequences are the result of a cognitive process that implies to evaluate the algorithms in terms of their cognitive relevance. As few experiments have been made on people's cognitive criteria for similarity measuring, we base our criteria on heuristics that were inspired from some musical issues. Three different sets of features have been considered: pitches, pitch contours and rhythm. For each set, a similarity measure is computed. The global similarity value results from the linear combination of the three values. The algorithm was tested on several pieces of music, and interesting results were found. At the same time, new questions were raised on the notion of similarity (this research is part of the European project Cuidado).

1. BACKGROUND

The interest for music similarity has been growing for a few years. Several techniques associated with different musical representation formats have been proposed for performing similarity measurement. Most of the techniques are context-independent: the computation of the similarity value between two musical sequences does not depend on the events that have occured before or between the two sequences. Against the rule, an interesting proposition for an inductive model can be found in [3]. Very few approaches propose to analyse polyphonic data. When they do, polyphony is not considered as a specific musical issue: most of the papers aim to transform polyphony in a monophonic approximation, while others consider polyphony as a part of a more general multidimensional mathematical issue [4]. In the last case, musical results are not provided and similarity is only based on exact repetition.

Among the following approaches, the issue of polyphony is not tackled, but the proposed strategies are interesting because they illustrate numerous different ways to compute the similarity.

An approach using dynamic programming can be found in [5]. The similarity between two sequences relies on an edit distance that measures the number of basic operations (adding, deleting, moving a note...) that transform one sequence into the other. The difficulty here is to determine both the basic operations and their "cognitive cost".

Another interesting approach [1] uses statistical information about pitches and durations together with a contour representation extracted from scores in order to obtain feature maps that are formed by an unsupervised learning algorithm. Unfortunately, it is implicitly assumed that the similarity between the different melodies is a transitive relation (distances in the two dimensions super-map are euclidean), whereas this is hardly the case in music. Moreover, the temporal succession of the events is not considered in the features (except in the contour) and maybe other features should be taken into account. Also, note that this approach is not context independent as a learning process is required for initialization.

Another approach is described in [2]. The similarity is based on the length of the vector of the differences between two sequences of melotones (pitches representation) or cronota (durations representation). However, this measure has limitations. For instance, intervals between not contiguous notes that would be common to the two compared sequences are not being considered in the measure. Moreover, the cronota sequence is represented by multiples of the 1/16 note, which is not compatible with ternary rhythmic values. Choosing the common denominator of all the durations as a basic unit would imply a far too complex analysis, as most of the very small durations do not play a role in the similarity measure. We will propose another representation format in paragraph 3.

Lastly, [6] considers a set of several features such as pitches (or duration) profiles, intervals and contour as binary features. The similarity between two sequences increases with the number of shared features. This global approach does not allow for local variations. For instance, with this approach, contours can be identical or different, but not similar.

2. QUESTIONS ON SIMILARITY

The interest for music similarity has raised with new commercial applications, such as query by humming,

that have emerged from the growth of Internet. However, the notion of similarity remains very difficult to define. Usually, similarity is not considered differently according to the different objectives for which it is computed. However, we think that it would be a first step in order to make this notion clear.
For instance, query by humming and pattern extraction are very different tasks. In one case, the goal is to match two sequences that should be the same but are different because they have been interpreted from the original score. In the other case, the goal is to match two sequences that are perceived as similar but that have initially been written differently by the composer. If the first matching do not specially requires cognitive considerations (two sequences are similar because they should physically be the same), the second one often needs them: two sequences can be heard as similar while being physically different.
Another issue arises when dealing with polyphonic sequences that happen very often in music. The term polyphony doesn't automatically mean several pitches at the same time, but several voices at the same time. For instance, a "Suite" of J.S Bach for violoncelle can be polyphonic whereas only one event is played at the same time. An answer to the issue of polyphony would be the automatic separation of the different voices, but no algorithm can currently perform that. Thus, we think that polyphony must be taken into account in a similarity measure, which raises new issues that will be tackled later.
Another problem arises from temporal considerations: should a similarity algorithm be independent of the context? This is hardly the case when considering that our culture, education and memory influence our perception of similarity. However, one can still wonder is there exists some universals.

3. AIMS

The method we present in this article is a new model for computing the similarity value between two non-identical polyphonic sequences.
In a first step, we have tried to challenge the notion of context. For instance, we do not introduce knowledge about tonality neither we modelize induction, memorization or learning processes. We know that this draws the limits of the system, and we know that a further step would be to integrate theses notions. But we are still very interested in exploring the limits of a system "outside time". Besides, we will not try to extract *all* but *a set* of significant patterns from a polyphonic music. Thus, it may happen that two musical sequences that are very similar perceptually are not recognised as such by our algorithm. In a first step, our limited goal is that sequences recognised as very similar by the algorithm should indeed be very similar perceptually.

4. THE SIMILARITY MODEL

4.1. Musical representation : the b.s

Performing similarity measures on not quantized music is rather difficult as soon as identical sequences with different tempi will physically appear as different whereas they are cognitively heard as similar. Our model integrates this cognitive aspect by representing each pattern as a quantified sequence of beat segments (b.s). Each b.s is a polyphonic sequence of pitches, onsets and durations in the MIDI format (see figure 1). The beat-tracking algorithm we use is described in [7].

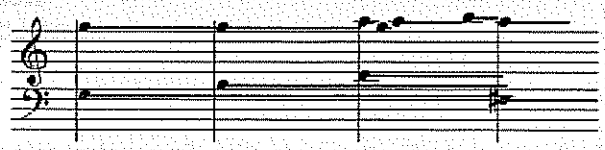


Figure 1: Beginning of the "variations Goldberg" from Bach. The vertical lines delimit the beat segments. Horizontal lines are the durations of each event.

4.2. Similarity in a polyphonic context

We assume that the notion of similarity between two polyphonic musical sequences makes sense. No information is available on the different voices of each sequence. Computing the intersection between the two sequences would appear as an intuitive way to measure what is common between the two sequences. However, the intersection could be empty while the two sequences would be perceived as similar.
Thus, we state that a sequence x is similar to a sequence x' if x is approximatively included in x'. For instance, when listening to music, we try to associate one sequence already heard with the current sequence we are hearing. We do not intersect the two sequences, but we evaluate the similarity between one sequence and a reference one. Thus, in our model, we understand similarity between two sequences x and x' as the distance from x to a certain sequence sub(x') included in x'. Note that this measure is not symetric (see equation 2).

4.3. Introduction to some cognitive aspects of the model

An important cognitive aspect of our model is that a musical sequence of b.s is considered as a whole entity (it may contain an abstract cognitive structure), and not solely as the concatenation of smaller entities. We think that several relations between non-adjacent events emerge from the whole entity. Theses relations play a role in the cognitive processes for recognizing the similarity between two sequences. To integrate this aspect, the similarity value between two sequences will not be computed from the addition of the similarity values between the smaller components:

$$S(x, x') + S(y, y') \neq S(xy, x'y') \quad (1)$$

Where S(x, x') designs the similarity value between sequences x and x', and xy designs the concatenation of sequence x and y.
Another cognitive aspect (see 4.2) is that our similarity measure is not symmetric in a polyphonic context:

$$S(x, x') \neq S(x', x) \quad (2)$$

If x is approximatively included in x', x will be very similar to x'. But x' will not automatically be approximatively included in x.
Last, according to cognitive aspects in [8], a similarity measure is not transitive:

$$S(x, y) + S(y, z) \leq \text{or} \geq S(x,z) \quad (3)$$

Our similarity computation provides a real value between 0 and 1 that state how similar are the sequences (1 is for identical). In our pattern extraction project, we use a similarity matrix for the representation of the results, and we perform clustering, but this will not be presented in this paper.
We will now describe our model for similarity. Because of lack of space, we have chosen to provide a general overview of our algorithm rather than a detailed description of a part of it.
In our model, sequences of b.s are of same length (length is expressed in number of b.s), so that each position of b.s in a given sequence can be matched *with the same position of b.s in another sequence*. We compute three different similarity values by considering three different sets of features: pitches (chords, pitch intervals etc...), pitch contours (contour at the top and at the bottom of the polyphony) and rhythm.

4.4. Similarity measure for pitches

We consider here the chords and the pitch intervals features. A similarity value is computed from two b.s sequences seq1 and seq2 of same length.
The only events falling on the downbeats are considered. This may be arguable, but two reasons have conducted this choice:

- Considering all the polyphonic events would require too much running time.
- The downbeats are often perceived as salient temporal position. Two sequences whose pitches coincide on the downbeat but differ elsewhere are often recognised as very similar (this has been confirmed in our experiments).

Usually, a downbeat event (dwb.event) is a chord, but it can also be a note or a rest.
In order to consider all the possible relations (see figure2) between non-adjacent dwb.event, the global similarity results from the computation of a similarity value between all the pairs of dwb.event in seq1 (dwb.event-seq1(i), dwb.event-seq1(j)) and their corresponding pairs in seq2 (dwb.event-seq2(i), dwb.event-seq2(j)). i and j (> i) are indexes of the considered b.s.
The similarity value for a pair depends on:

- the length of the different combinations of intersections between the four dwb.events considered as chords (harmonic similarity)
- the length of the intersection between all the intervals between dwb.event-seq1(i) and dwb.event-seq1(j) and the intervals between dwb.event-seq2(i) and dwb.event-seq2(j) (melodic similarity).

The similarity values between pitches or intervals are :

- 1 for equal pitches or equal intervals
- 0.5 for transposed chords
- 0 otherwise

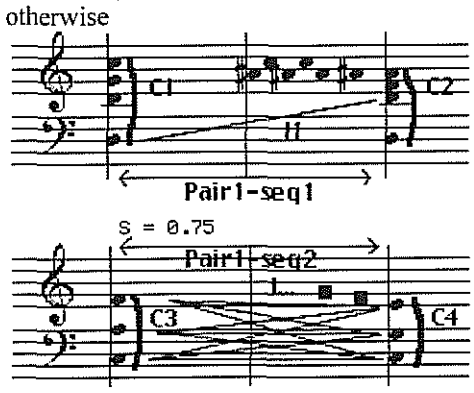


Figure 2 (see part 5 for details on the symbols): Two similar patterns in Sonata8Am-Part2 from Mozart. All vertical and horizontal intervals between dwb.events C1 and C2 of Pair1-seq1 and dwb.events C3 and C4 of Pair1-seq2 are compared.

4.5. Similarity measure for contours

As reported in [1], contour plays an important role in the perception of melodic similarity.
Our model compares the upper and lower contours of two b.s sequences seq1 and seq2 of same length.
As above, the only events falling on a downbeat (dwb.events) are considered. An up (down) contour is *the sequence of the intervals between the upper* (lower) pitches of the dwb.events. Each contour of each sequence is compared with the two contours of the other sequence. Contours are very similar (see figure3) if the intervals from one sequence are similar to the corresponding intervals from the other sequence (two intervals are similar if their difference is less than 5 half tones).

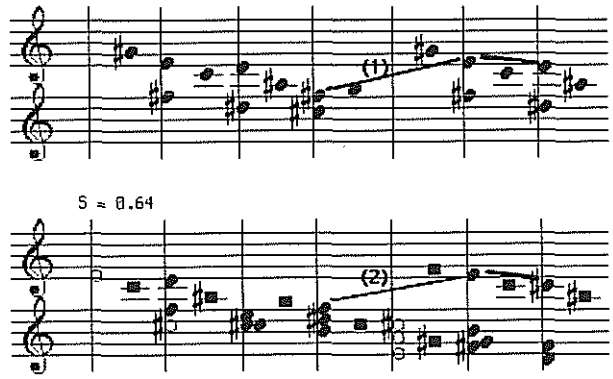


Figure 3 (see part 5 for details on the symbols): Two similar patterns in Pierrot Lunaire from Schoenberg. Lines (1) and (2) show similar contours.

4.6. Similarity measure for rhythm

We believe that rhythm is a component of the cognitive criteria we use in the recognizing of similarity. Our model compares the rhythmic structure of two sequences of b.s seq1 and seq2 of same length.

In a first step, seq1 and seq2 are normalized so that the total duration of the b.s will be the same for seq1 and seq2. Then, for each b.s, onsets (temporal positions) in seq1 are associated to the corresponding onsets in seq2. Two onsets of two b.s form a pair if they share similar temporal positions in the b.s. If an onset of one sequence do not form a pair with an onset of the other sequence, then it is deleted. The similarity between two sequences of b.s is the mean of the similarity between each corresponding b.s. (as seq1 and seq2 have same length, each b.s of seq1 correspond to one b.s of seq2) The similarity between two corresponding b.s is the mean of the similarity between each pair of corresponding onsets (here, relations between non-adjacent events are not considered). Corresponding onsets are already similar because they share the same temporal position in the b.s. The similarity increases with the length of the intersection of the durations of the events corresponding to the onsets of a pair (an approximation value is considered for the intersection).

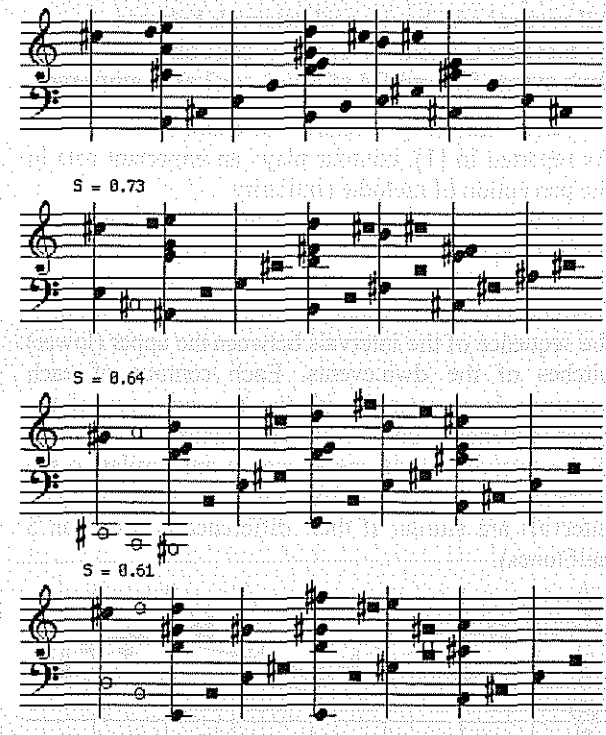


Figure 4 (see part 5 for details on the symbols): Three patterns similar to the first pattern in Sonata AM d664 op121, 1rst Part from Schubert.

4.7. Overall similarity measure

Each of the three above models (pitches, contour and rhythm) computes a similarity value. The three values can then be linearly combined into a global similarity measure. In this case, different weights can be applied to the different measures. In our experiments, we have chosen to give a higher weight to the rhythmic and the pitch based measures as they are more "selective" than the contour measure: the contour is expected to be common to more sequences than the pitch or the rhythmic successions of events.

5. MUSICAL EXEMPLES

Each of the above musical exemples (Figure2, Figure3 and Figure4) shows a reference pattern (the above one) together with other similar patterns. Durations are not represented. In figure4, the similar patterns are sorted according to their decreasing similarity value S. The events that determined the similarity between the patterns (for pitches, contours and rhythm) are represented in black. The square symbols only determined the similarity for rhythm.

6. CONCLUSION

We have presented the general lines for a polyphonic similarity model that integrates some cognitive principles. Interesting musical results have been presented that show that patterns can be extracted without considerations on the temporal context. We think that further investigation should be done in this direction. For instance, one could introduce a dissimilarity measure in order to consider events that contribute to the perceptual differentiation of sequences. Other features, such as statistical features, or the dynamics of the notes, should also be considered.

7. REFERENCES

[1] Petri Toivainen & Tuomas Eerola ``A computational model of melodic similarity based on multiple representations and self-organizing maps", Proceedings of the 7th ICMPC Sydney 2002.

[2] Hofman-Engl, L "Melodic transformations and similarity - a theoretical and empirical approach", Phd Thesis, Keele University, 2002.

[3] Lartillot O, "Generalized Musical Pattern Discovery by Analogy from Local Viewpoints", in Discovery Science, LNCS 2534, Springer-Verlag, 2002

[4] Wiggins, Lemstrom, Meredith "SIA(M)ESE: An algorithm for transposition invariant, polyphonic content-based music retrieval", Proceedings of the 3rd ISMIR, Ircam, 2002

[5] M. Mongeau and D. Sankoff, "Comparison of musical sequences", Computers and the Humanities 24:161-175, 1990.

[6] Cambouropoulos and Widmer, "Melodic clustering : Motivic analysis of Schuman's Träumerei", Proceedings of the 3rd Jim, Bordeaux, 2002

[7] Meudic B, "A causal algorithm for beat tracking", 2nd conference on understanding and creating music, Caserta, 2002

[8] Tversky A, "Features of similarity", journal of Psychological Review, p327-352 1977

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

AN INTRODUCTORY CATALOG OF COMPUTER-SYNTHESIZED CONTACT SOUNDS, IN REAL-TIME

M. Rath, F. Avanzini, N. Bernardini, G. Borin, F. Fontana, L. Ottaviani, D. Rocchesso

Dipartimento di Informatica
Università degli Studi di Verona
rath@sci.univr.it

ABSTRACT

As part of the SOb European project several cartoon models of contact sounds of solid bodies, "hitting", "bouncing", "dropping", "breaking", "rolling", have been developed and implemented as modules (and sub-patches) for free real-time sound software pd [1]. The models are accessed through perceptually meaningful parameters and run with low computational load on standard PC hardware.

The underlying idea of *cartoonification*, its motivation and background in psychoacoustic research are sketched first. The main common sound-core of most models, a physics-based algorithm of impact-interaction with interacting resonators in modal description, is shortly presented. The impact module is embedded in patches of higher-level control to model more complex contact scenarios. The structure, use and potential of the resulting sound objects is described.

While the results are a possible basis for reactive sonic interfaces in Human-Computer-Interaction, they can as well be exploited for musical purposes.

1. CATALOG

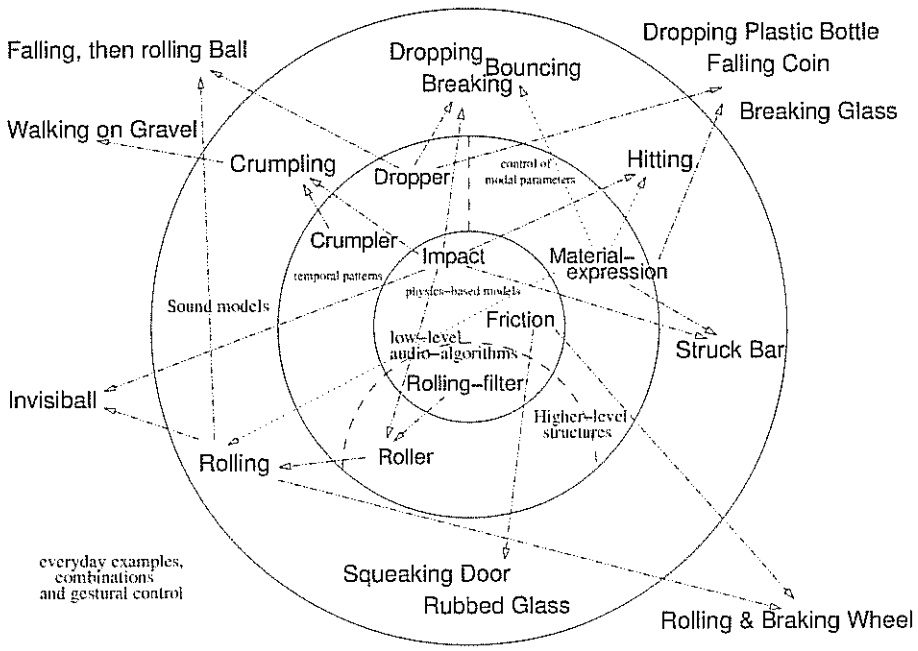


Figure 1: Overview of real-time sound models of contact scenarios and their underlying structures, as developed during the course of the SOb project. The graphical layout in nested (circular) fields reflects the structural hierarchy: Physics- and geometrically-based low-level audio algorithms in the center, completed with surrounding higher-level objects, resulting sound models in the largest circle and finally more concrete and complex example-scenarios, that may also envolve gestural control. Arrows indicate dependencies and are to be read as "contributes"/"used to realize".

[1] http://iem.kug.ac.at/pd/

2. INTRODUCTION

In 1969, Risset published a ground-breaking catalog of computer-synthesized sounds [1], which served the purpose of illustrating the emerging techniques of sound analysis and synthesis. Those examples and studies are still influential for composers and sound scientists, especially those working with signal-processing tools and in the context of musical sounds. The discipline of psychoacoustics has provided, over the years, a solid support to connect signal processing to human perception.

A new stream of studies started in the early eighties from the observation that everyday listening is different from musical listening [2]. Both new psychoacoustic and sound modeling methods and results are needed for this new framework. On the perceptual side, the viewpoint of ecological psychology is very useful [3]. On the modeling side, the physically-based modeling paradigm seems to be the best sound production strategy to address everyday listening in interactive applications.

The EU-funded project "the Sounding Object" (SOb) [2] was launched in 2001 to provide a corpus of knowledge in everyday sound perception, accompanied by suitable new methods and tools for physics-based sound modeling and for high-level control of these models.

The SOb project aims at "sounding objects" that incorporate a (possibly) complex responsive acoustic **behavior**, expressive in the sense of ecological hearing, rather than the (re-)production of fixed isolated signals. Although "real" sounds hereby serve as an orientation, realistic simulation is not necessarily the perfect goal: simplifications which preserve and possibly exaggerate certain acoustic aspects, while losing others considered less important, are often preferred. Besides being more effective in conveying certain information, such "*cartoonifications*" are often cheaper to implement, just like graphical icons are both, more clear and easier to draw than photo-realistic pictures.

Physical modeling naturally relates to synthesis controlled in terms of ecological parameters. The straight approach though, the description of a given physical system through differential equations and their numerical solution, often leads to (possibly highly) realistic simulations that are computationally expensive and lack flexibility and generality. We thus combine closely physics-based models in the above sense with structures that remind of classical techniques of sound synthesis, trying to integrate ecological expressivity, flexibility and computational efficiency.

Contacts of solid bodies form a large class of sound-emitting processes in every-day surroundings. The perception of ecological attributes, like material and size of involved objects and their way of interaction, hitting, sliding, rolling..., from contact sounds is common experience and has been examined by psychoacoustic studies. Our works show that, from the standpoint of cartoonification, many typical forms of contact-interaction can be successfully modeled on the basis of a physically founded but "abstracted", flexible and efficient one-dimensional impact or friction algorithm. Specific characteristics of the macroscopic scenarios which are of high perceptual relevance are modeled explicitly, for instance as macro-temporal distributions of micro-impacts.

This paper is intended to be a explanatory guide to our collection of sound models and examples, as they are available nowadays from the SOb project website as pd plugins and patches. Detailed explanations of the inner structure and the development of all models can be found in [4]. All contact sound models are based on low-level models of basic interactions: impacts and frictions, which are described in sections 3.1 and 3.2. Higher-level models describing phenomena with complex temporal patterns are presented in section 4. Finally, section 5 briefly gives some examples of how the sound models can be associated to everyday objects, thus providing their typical sonic behavior in an interactive, real-time fashion.

[2] `http://www.soundobject.org`

3. THE LOW-LEVEL PHYSICS-BASED MODELS

3.1. Impact

In contrast to several studies of contact sounds of solid bodies that focus on the resonance behavior of interacting objects and widely ignore the transient state of the event, our approach is based on a physical description of impact interaction processes [5]. This physical model involves a degree of simplification and abstraction that implies efficient implementation as well as adaption to a broad range of impact events.

We consider two resonating objects and assume that their interaction depends on the difference x of two (1-dimensional) variables connected to each object. In the standard case of examined movements in one spatial direction, x is the distance variable in that direction (negative distance $\triangleq$ contact). Possible simultaneous interaction along other dimensions are excluded at this stage. This leads to a compact efficient algorithm that strikes the main interaction properties. The impact force f is stated as a nonlinear term in x (and $\dot{x}$):

$$f(x(t),\dot{x}(t)) = \begin{cases} k(-x(t))^{\alpha} + \lambda(-x(t))^{\alpha} \cdot (-\dot{x}(t)), & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (1)$$

Here, k is the elasticity constant, i.e. the hardness of the impact. α, the exponent of the non-linear terms, shapes the dynamic behavior of the interaction (i.e. the influence of initial velocity), while λ controls the dissipation of energy during contact, accounting for friction loss. One inlet of the `impact_...`[3] modules takes a list of `interactor`-parameters containing the aforementioned values (in the same order).

Alternative versions, "`linpact_...`" exist with a simpler, linear, force term,

$$f(x(t),\dot{x}(t)) = \begin{cases} -kx(t) - r\dot{x}(t), & x < 0 \\ 0, & x \geq 0 \end{cases} \quad (2)$$

(and accordingly only two `interactor`-parameters k and r), that trade richness in detail for reduced computational cost.

The two interacting, resonating objects are built under the premises of modal synthesis [6] [4]. This formulation supports particularly well our main design approach for its physical generality and, at the same time, for its intuitive acoustic meaning. One `resonator` is here characterized by the number of its modes (which

[3] The difference between `impact_modalb` and `impact_2modalb` is explained later in the section.

[4] In fact, plugins are realized in a modular structure, that enables the connection of numerous different `resonators` as well as `interactors`. Digital waveguide resonators are in preparation; linearized impact and friction are the additional existing `interactors`. All modules so far share the modal `resonator`-functions.

can be chosen freely as an argument given at module-creation) and for each mode the three modal parameters: mode-frequency, exponential decay-time and level ("weight") of the mode at the point of interaction. Accordingly the `impact_/linpact_...` modules have inputs for lists (of length "number of modes") of frequencies, decay-times and `weighting-factors`. It is often satisfactory and more convenient to use the modules `impact_modal` and `linpact_modal`, where (in contrast to `impact_2modal` and `linpact_2modal`) the first `resonator` is reduced to an inertial (point-)mass[5] and characterized only by one ("mass"-)parameter. This practical and computational simplification parallels the notion that in many practical contact scenarios the vibration of one involved object is hardly or not perceived.

Further, for each resonator, an arbitrary number of "`pick-ups`" can be defined, which are characterized by lists, given at one inlet, of `weighting-factors` (for all modes). The first `pickup` is identical with the interaction-point (and always exists).

Finally, all modules have three audio inlets, for the input of signals representing external forces on both resonators (again at the point of interaction) and an additional `offset`, used mainly for surface profiles in rolling-/sliding-models.

3.2. Friction

For friction modeling, we use a computational structure very similar to the one used for impacts. The pd plugin is called `friction_2modalb`.

The underlying model describes the average behavior of a multitude of micro-contacts made by hypothetical bristles extending from each of two sliding surfaces. When a modal decomposition is adopted for both interacting objects, the equations are

$$\begin{cases} m_{ei}\ddot{x}_{ei} + r_{ei}\dot{x}_{ei} + k_{ei}x_{ei} = f_{ee} - f_f, & (i = 1 \dots N_e) \\ m_{rj}\ddot{x}_{rj} + r_{rj}\dot{x}_{rj} + k_{rj}x_{rj} = f_{re} + f_f, & (j = 1 \dots N_r) \\ v = \sum_{i=1}^{N_e} \dot{x}_{ei} - \sum_{j=1}^{N_r} \dot{x}_{rj}, & \text{(relative velocity)} \\ \dot{z} = f_{NL}(v,z) = v\left[1 - \alpha(z,v)\dfrac{z}{z_{ss}(v)}\right], & \\ f_f = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v, & \text{(friction force)} \end{cases} \tag{3}$$

where the x variables represent the modal displacements, while z is the mean bristle displacement. The terms f_{ee} and f_{re}, as indicated by e in the second subscript, represent external forces. As far as the form of functions α and z_{ss} is concerned, we adopt a couple of previously proposed functions[7].

High-level interactions between the user and the audio objects rely mainly upon three interaction parameters. These are the external forces f_{ee} and f_{re} (see equations (3)) acting on each of the two objects, which are tangential to the sliding direction, and the normal force f_N between the two objects. The remaining parameters belong to a lower level control layer, as they are less likely to be touched by the user and have to be tuned at the sound design level.

Such low-level parameters can be grouped into two subsets, depending on whether they are related to the resonators' internal properties or to the interaction mechanism. Each mode of the two resonating objects is tuned according to its center frequency and decay time. Additionally, a modal gain (which is inversely proportional to the modal mass) can be set for each resonator mode, and controls the extent to which the mode can be excited during the interaction.

[5]This is the special case of a modal resonator with only one resonant mode of frequency 0 and infinite decay time (undamped).

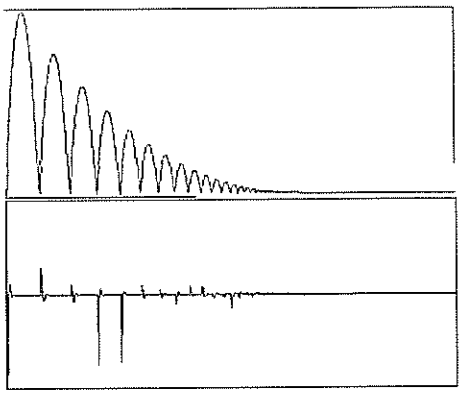


Figure 2: Temporal movement of an inertial mass (above) "bouncing" on a two-mode resonator (at `pickup-point`, below).

A second subset of low-level parameters relates to the interaction force specification. The triple $(\sigma_0, \sigma_1, \sigma_2)$ (see equations (3)) define the bristle stiffness, the bristle internal dissipation, and the viscous friction, and therefore affects the characteristics of signal transients as well as the ease in establishing stick-slip motion. A triple of parameters is used to set the shape of the curve z_{ss}. Specifically, the Coulomb force and the stiction force are related to the normal force through the equations $f_s = \mu_s f_N$ and $f_c = \mu_d f_N$, where μ_s and μ_d are the static and dynamic friction coefficients. Finally, the breakaway displacement z_{ba} is also influenced by the normal force. In order for the function $\alpha(v,z)$ to be well defined, the inequality $z_{ba} < z_{ss}(v) \forall v$ must hold. Since $\min_v z_{ss}(v) = f_c/\sigma_0$, a suitable mapping between f_N and z_{ba} is

$$z_{ba} = c f_c/\sigma_0 = c\mu_d f_N/\sigma_0, \qquad \text{with} \qquad c < 1. \tag{4}$$

4. HIGHER-LEVEL STRUCTURES

4.1. Bouncing, Dropping, Breaking

Short acoustic events like impacts can strongly gain or change in expressive content, when set in an appropriate temporal context. One example is the grouping of impacts in a "bouncing" pattern, as it results from a constant external (gravity-)force term.

The one-dimensionality of the impact algorithm only allows the immediate simulation of symmetrical, basically spherical, bouncing objects; these simulations through an external gravity term are very realistic in detail, "too realistic" from a standpoint of *cartoonification*: The exact (accelerating) tempo of bouncing is coupled to the impact parameters, and simplifications on the elementary sound level necessarily affect the higher level pattern. A strict physical simulation of irregular bouncing objects on the other hand, would be highly complex to control and implement, computational "overkill". Instead, an explicit modeling of typical bouncing-patterns leads to *cartoonifications*, that are efficient to implement and able to express ecological attributes like regularity/irregularity of the bouncing object. The main notions behind the structure and parameters of the "`dropper`" object are shortly sketched in the following.

The first basic principle behind the process is the loss of macro-kinetic energy of the global vertical, horizontal and rotational movement, in friction and microscopic (e.g. acoustic) vibration. These loss-terms in exactness are different for each impact, and can in this detail only be found from an elementary simulation as above. Under the assumption, that the loss of (macro-)energy with each bounce is proportional to the remaining kinetic energy, we receive an exponentially (in the number of reflections) decaying energy term. If we further for spherical bouncing objects concentrate on the vertical movement and ignore the horizontal and rotational terms as independent, the kinetic energy at floor level is proportional to the square roots of collision velocity and the duration of the following bounce. We thus arrive at analogous exponentially decaying terms for impact velocities and temporal intervals in a regular bouncing movement. The implementation of this basic scheme in fact proved to be convincing in comparison with the afore-described implicit simulation (compare figure 2) as well as recordings of bouncing (round) wooden balls. For irregular objects, energy can be transferred between the vertical, horizontal and rotational terms, of which only the vertical velocity (and therefore the maximum height) contributes a simple term to the impact intervals and velocities, while the contribution of the rotational movement is not expressible in a simple form (and that of the horizontal movement is basically zero). Energy transfer thus results in deviations of both, impact-intervals and -velocities, from, but generally bounded by, the (exponentially decaying) values of the regular case. Similarly, also the effective relative masses and the `weighting~factors` of resonant modes are modulated through the rotation (and therefore changing contact points) of an irregular object. Summing up, while generally the exact movement in the non-spheric case can only be simulated through a detailed solution of the underlying differential equations, which would not make sense in our context of real-time interactivity[6], controlled-random patterns of impact parameters can generate expressive cartoonifications. Another important observation are static stages in the bouncing movement also of non-spherical shapes with certain symmetries of regular aspects (e.g. such as disks or cubes). In these cases the transfer of energy between the vertical, horizontal and rotational terms can take place in regular patterns, closely related to those of spherical objects. This phenomenon is exploited in some modeling examples; often however, such movements include rolling aspects, suggesting a potential improvement through integration of rolling models. A very prominent sound example with an initial "random"- and a final regular stage is that of a falling coin.

Following the previous observations, the "`dropper`" generates temporal patterns of impact velocities triggered by a starting message. Control parameters are:

1. The time between the first two reflections, representing the initial falling-height/-velocity, together with
2. the initial impact velocity.
3. The acceleration factor is the quotient of two following maximal "bounce-intervals" and describes the amount of microscopic energy loss/transfer with each reflection, thus the speed of the exponential time sequence.
4. The velocity factor is defined analogously.

Note that these parameters should for a spherical object be equal (see above), while in exactness being varied (in dependence of actual impact velocities) in the general case. In a context of cartoon-based auditory display they can be effectively used in a rather intuitive free fashion.

5. Two parameters specify the range of random deviation below the (exponentially decaying) maxima for temporal intervals resp. impact velocities. The irregularity/sphericity of an object's shape is modeled in this way.
6. A threshold parameter controls, when the accelerating pattern is stopped, and a "terminating bang" is sent, that can e.g. trigger a following stage of the bouncing process.

Warren and Verbrugge [8] study on the perception of breaking- and bouncing-scenarios is a starting point for our related modeling efforts. They showed, that sound artefacts, created through layering of recorded collision sounds, were identified as bouncing or breaking scenarios depending on their homogeneity and the regularity and density of their temporal distribution.

Again the main ideas behind the structure of the breaking-model are shortly sketched: Typical fragments of rupture are of highly irregular form and rather anelastic, and tend to "nod" rather than bounce, i.e. perform a decelerating instead of accelerating movement. It is further on important to keep in mind that emitted fragments mutually collide, and that the number of such mutual collisions rapidly decreases, starting with a massive initial density; those collisions do not describe bouncing patterns at all. Following these examinations the breaking-model was realized by use of the `dropper` with high values of "randomness", and a quickly **decreasing** temporal density, i.e. a time-factor > 1, set "opposite" to the original range for bouncing movements. Supporting Warren and Verbrugge's examination, a short noise impulse added to the attack portion of the pattern underlined the breaking character.

As another insight during the modeling process, several sound attributes showed to be important. Temporally identically grouped impacts seem to be less identifiable as a breaking event, when tuned to a metallic character in their modal settings; this may correspond to the fact that breaking metal objects are rather far from everyday experience. Also, extreme mass relations of "striker" and struck resonator in the impact settings, led to more convincing results. Again, this is in correspondence with typical situations of breakage: a concrete floor has a practically infinite inertia in comparison to a bottle of glass.

4.2. Rolling

Among the various common mechanical interactions between solid objects, "rolling" scenarios form a category that seems to be characteristic also from the auditory viewpoint: Everyday experience tells that the sound produced by a rolling object is often recognizable as such, and in general clearly distinct from sounds of slipping, sliding or scratching interactions, even of the same objects. This may be due to the nature of rolling as the most prominent **continuous** interaction process, where the mutual force on the involved objects is described as an impact without additional perpendicular friction forces.

Consequently, the impact-algorithm has been embedded in a complex higher-level structure to reach an efficient cartoonification, that can express various ecological attributes of rolling-

[6] Also, it seems questionable, how precisely shapes of bouncing objects (except for sphericity) can be recognized acoustically?

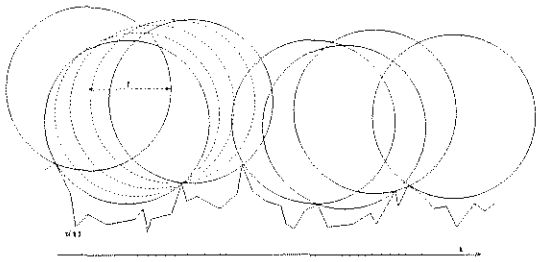


Figure 3: Sketch of the fictional movement of a ball, perfectly following a surface profile s(x). Relative dimensions are highly exaggerated for a clearer view. Note that this is **not** the de-facto movement; this idealization is used to derive the `offset-curve` to be used by the impact-model.

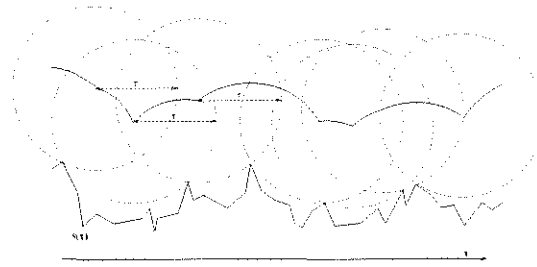


Figure 4: Sketch of the effective `offset-curve`, resulting from the surface $s(x)$.

scenarios: material, size and shape of the involved objects, as well as velocity or acceleration/deceleration (*transformational attributes* [2]). The main observations behind the development and structure of the model are presented shortly:

4.2.1. Rolling interaction with the impact-model as lowest-level building block on a driving `offset-curve`

Rolling-contact between two objects is restricted to distinct points: the supporting surface is **not** fully "traced"/followed, nor is the surface of the rolling object. Figure 3 sketches the idea; the rolling object is here assumed to be locally spherical without "microscopic" surface details. These assumptions are unproblematic, since the micro details of the surface of the rolling object can be simply added to the second surface (to roll on) and the radius of the remaining "smoothed macroscopic" curve could be varied; in conjunction with following notions, even an assumed constant radius however showed to be satisfactory.

The actual movement of the rolling object differs from the idealization of figure 3 due to inertia and elasticity. In fact, it's exactly the consequences of these physical properties, which are described by, and substantiate the use of the impact model(-equations). It is further important to notice that, in contrast to slipping-, sliding- or scratching-actions, the interaction force on the two objects involved in a simple rolling-scenario is approximately perpendicular to the contact surface (the macroscopic mean curve), pointing along the connection line of the momentary point of contact and the "center of the rolling object". This fact is not reflected in the sketches, since here relative dimensions are highly unrealistic, exaggerated for purposes of display). Summing up, the final vertical movement of the center of the ball can be approximated by use of the one-dimensional impact-model with the `offset-curve` shown in figure 4.

In a naive approach, the calculation of contact points is computationally highly demanding: In each point x along the surface curve $s(q)$[7], i.e. for each sampling point in our practical discrete case (at audio rate), the following condition, which describes the momentary point of contact p_x, would need to be solved:

$$f_x(p_x) = max_{q\in[x-r,x+r]} f_x(q) \quad where \quad (5)$$
$$f_x(q) \triangleq s(q) + \sqrt{r^2 - (q-x)^2}, \quad q \in [x-r, x+r]$$

[7]Here we use the unproblematic assumption that the surface curve is presentable as a real function $s : I \longmapsto R$, $I \subseteq R$ an interval.

To facilitate a practical, real-time, implementation, a "smart" algorithm had to be developed, that reduces the number of calculations/comparisons by factors up to 1000. The ideal `offset-curve` is then calculated from the coordinates of the contact points.

4.2.2. Surface profile

The surface-signal which is processed by a "rolling-filter" as above might be derived through scanning/sampling of "real" surfaces. A flexible statistics-based generation though is preferable in our context over the sumptuous, static storage of fixed profiles. One such approach is *fractal noise*, i.e. noise with a $1/f^{\beta}$ power spectrum, the real parameter β reflecting the fractal dimension or roughness. Practical results of modeling following the so-far developed methods however became much more convincing, when the bandwidth of the surface-signal was strongly limited. This does not surprise, when one keeps in mind that typical surfaces of objects involved in rolling scenarios, are generally smoothed to high degree. (In fact, it seems hard to imagine, what e.g. an uncut raw stone rolling on another surface, typically modeled as a fractal, let's say a small scale reproduction of the alps, would sound like?) Smoothing on a large scale, e.g. cutting and arranging pieces of stone for a stone floor, corresponds to high-pass-filtering, while smoothing on a microscopic level, e.g. polishing of stones, can approximately be seen as low-pass-filtering. In connection with this resulting band-pass, the $1/f^{\beta}$ characteristics of the initial noise signal lost in significance. Band-pass filtered white noise thus was chosen as a cheap and efficient solution; it can eventually be enhanced by an additional second-order filter, whose steepness finally represents a "microscopic" degree of roughness as a very coarse approximation of the fractal spectrum.

Of course, the parameters of the impact itself, in particular the elasticity constant k, can/must also be carefully adjusted to surface-, e.g. material properties and strongly contribute to the expressiveness of the model.

4.2.3. Higher-level features

Typical scenarios of rolling tend to show characteristic "macroscopic" acoustic features, that appear to be of high perceptual relevance, especially for velocity-expression. Macro-temporal periodicities result from typical patterns of more or less regular nature as found on many "ground" surfaces (such as joints of stone- or wooden floors, the periodic textures of textiles or the pseudo-periodic furrows in wooden boards). Seemingly even more important, for rolling objects, that are not perfectly spherical (in the sec-

tion relevant for the movement), the velocity of the point of contact on both surfaces and the effective force pressing the rolling object to the ground vary periodically. In order to model such deviations from perfect sphericity, these two parameters must be modulated, our practical experience showing a higher significance of pressing-force; a good choice are obviously sinusoidal or other narrow-band modulation signals (since objects that differ too much from a spherical shape, that are too edgy, don't roll!). Of course all (quasi-)periodic modulations have to reflect the rolling-velocity in their frequency.

Finally it is to be noted that, like in everyday listening, acoustic rolling scenarios are recognized and accepted more easily with "typical dynamics": As an example, consider the sound of a falling marble, that bounces until constant contact to the ground is reached, now the rolling action gets acoustically clear and the average speed slowly decays to zero.

4.3. Crumpling

Like most other sounds presented in this catalog, `crumpling` results from providing the impact model with a control layer. Since `crumpling` does not model physical contacts between solid objects, but rather special time sequences of bends, the use of closed-form formulas expliciting interaction mechanisms can be avoided.

Impacts are triggered following stochastic laws which are derived from the physics of crumpling [9]. Such laws rule the dynamic and temporal statistics of those impacts. By including a notion of *energy* in the crumpling process, we can control the time length and the overall dynamics of individual *events*, each one consisting of a collection of "crumples".

The physics-based approach to crumpling-sound reproduction produces a control layer with physical parameters. The advantage of having such parameters at hand is twofold: first, those physical controls can be interfaced with the impact driving parameters directly; second, the user interface presents a consistent control panel to the user, without the need of intermediate maps layered in between the model and the user interface. By means of this design approach we were able to synthesize sounds of *crushing* cans.

Aiming at yet a higher level of scenarios to be modeled, the "user" can be a top-level control structure, which triggers events according to some rule. Rules governing the temporal evolution of *walking* and *running* exist, which are physics-based [10]. Those rules drive the crumpling model parameters directly, in a way that we have obtained interesting walking and running sounds.

Crushing, walking and running are extensively described in an article submitted for these proceedings .

5. FAMILIAR (SOUNDING) OBJECTS

The expressiveness of the sound models is best recognized, when parameters are set to example-values within the wide ranges, that are connected to scenarios familiar from every-day experience. Such demonstrations often envolve combinations of several models; we have chosen some items, partly accompanied with basic visualizations, from rather simple to complex ones:

- The sole impact-model can be tuned to struck bars of different sizes and materials,
- as the low-level friction-model can realize squeaking doors and rubbed glasses.
- The rolling-model with its strong ecological potential (velocity, direction, size ...) sonifies different interactive "games" with rolling balls.
- Rolling and friction are two states of an interactive wheel–brake construction.
- The dropper-object delivers convincing bouncing balls as well as dropping plastic bottles, metallic coins and breaking glasses.
- Natural is the combination of dropping and subsequently rolling balls.
- Typical scenes of crumpling are crushing cans and the sound of walking on gravel.

6. ACKNOWLEDGMENTS

This work has been supported by the European Commission under contract IST-2000-25287 (project "SOb - the Sounding Object": `www.soundobject.org`).

7. REFERENCES

[1] J.-C. Risset, *An Introductory Catalog of Computer-synthesized Sounds*, Murray Hill, New Jersey: Bell Laboratories, 1969.

[2] W. W. Gaver, *Everyday listening and auditory icons. PhD thesis*, University of California, San Diego, 1988.

[3] W. W. Gaver, "How Do We Hear in the World? Explorations in Ecological Acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, Apr. 1993.

[4] F. Avanzini, F. Fontana, L. Ottaviani, M. Rath, and D. Rocchesso, *Models and Algorithms for Sounding Objects*, 2002, Progress Report of the SOb project, available at `http://www.soundobject.org`.

[5] F. Avanzini and D. Rocchesso, "Modeling Collision Sounds: Non-linear Contact Force," in *Proc. COST-G6 Conf. Digital Audio Effects (DAFX-01)*, Limerick, Dec. 2001, pp. 61–66, Available at `http://www.soundobject.org`.

[6] J. M. Adrien, "The Missing Link: Modal Synthesis," in *Representations of Musical Signals*, G. De Poli, A. Piccialli, and C. Roads, Eds., pp. 269–297. MIT Press, 1991.

[7] F. Avanzini, D. Rocchesso, and S. Serafin, "Modeling interactions between rubbed dry surfaces using an elasto-plastic friction model," in *Proc. Conf. on Digital Audio Effects (DAFx-02), COST-G6*, Hamburg/Germany, 2002.

[8] W. H. Warren and R. R. Verbrugge, "Auditory perception of breaking and bouncing events: a case study in ecological acoustics," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, no. 5, pp. 704–712, 1984.

[9] P. A. Houle and J. P. Sethna, "Acoustic emission from crumpling paper," *Physical Review E*, vol. 54, no. 1, pp. 278–283, 1996.

[10] R. Bresin, A. Friberg, , and S. Dahl, "Toward a new model for sound control," in *Proc. Conf. on Digital Audio Effects (DAFX-01), COST-G6*, Limerick/Ireland, Dec. 2001, pp. 45–49.

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

PHYSICS-BASED SOUND SYNTHESIS AND CONTROL: CRUSHING, WALKING AND RUNNING BY CRUMPLING SOUNDS

Federico Fontana

Dipartimento di Informatica
University of Verona
fontana@sci.univr.it

Roberto Bresin

Speech, Music and Hearing, TMH
Royal Institute of Technology, KTH
roberto@speech.kth.se

ABSTRACT

Three types of ecological events (crushing, walking and running) have been considered. Their acoustic properties have been modeled following the physics-based approach. Starting from an existing physically-based impact model, we have superimposed to it the dynamic and temporal stochastic characteristics governing crushing events. The resulting model has been finally triggered by control rules defining walking and running time patterns.

This bottom-up design strategy was made possible because the sound synthesis and sound control models could be directly connected each other via a common switchboard of driving and control parameters. The existence of a common interface specification for all the models follows from the application of physics-based modeling, and translates in major advantages when those models are implemented as independent, self-contained blocks and procedures connected together in real-time inside a software like *pd*.

1. INTRODUCTION

Physics-based sound synthesis has its roots in some traditional paradigms for the generation of sound for musical purposes. In these paradigms the synthesis model is not specified by requirements on the signal, but follows by a comprehension of the generation process and, consequently, by its reproduction [1].

This approach to sound design, hence, deals with the cause and not with the effect. In other words we can say that sound is obtained by designing (or modeling) a corresponding process, whatever it is and even regardless of its effects. It is not an uncommon result that this process, in the end, comes out to be unsuitable for sound generation. Nevertheless, there are cases in which a given process produces effective and convincing sounds.

Such an approach leads to models whose control is specified *in parameters having a direct relationship with the synthesis process*. On the other hand, any control devoted to manipulate specific sound features must be applied indirectly through a map, that acts as an intermediate layer in between the synthesis engine and the control panel. For this reason, the physics-based approach is generally not so effective in contexts where sound must be controlled using traditional signal-based operations (for instance, spectral manipulations).

In the case of a synthesis process that models the evolution of a physical system, we are in front of a *special* sound synthesizer. Once its suitability to sound generation has been proved, this synthesizer offers peculiar (i.e., physical) control parameters to the user. By means of those parameters the user can directly select quantities such as *force*, *pressure*, *mass*, and *elasticity*. It is clear that a control panel like that is not optimal for all contexts. At the same time it invites the user to explore a wide range of possibilities by real-time direct manipulation of the control parameters.

Physical modeling has been successfully applied to musical instrument modeling [2, 3, 4, 5, 6, 7], mainly because the ease of control of peculiar parameters, such as, in a virtual piano model, the force applied to a key by the player [8]. On the other hand, physical modeling hardly surpasses synthesis techniques based on signal sampling if only the parameter of sound reproduction accuracy is taken into account during the evaluation of a synthesizer. Physical modeling is competitive as long as the level of interactivity between the user and the virtual instrument is considered as a parameter of quality. Though, the exploration of physical models is relatively at an early stage. The horizon at which designers can look at is still heavily bounded by the amount of computational resources they can afford, as long as the real-time constraint comes into play along with reasonable limits in the cost of the final application. Unfortunately, despite a moderate need of memory, physical models demand a large amount of computational power (for this reason, techniques aiming at transforming at least part of those computational power requirements into figures of memory occupation have been proposed [9]). This is not the case of sampling: since a large amount of fast-access memory is available at a low cost, designers can store huge databases of sound samples in the application.

Physics-based (or physically-based) sound synthesis relocates the physical modeling background into a different framework, that is, the synthesis of *ecological* sounds [10]. Those sounds are normally experienced during everyday listening, and contribute to our perception of the world. Sounds accounting for events such as the bouncing of an object falling on the floor, the friction between two objects, a crushing can, the walking of a person, are detected, recognized and possibly localized by our hearing system. These sounds convey *vital* information, via the auditory channel, about the surrounding environment, and their reproduction is very interesting in applications involving audio virtual reality (or audio VR). The more realistic the virtual scenario, the more effective the audio VR system must be. Hence, we must not be surprised if the simulation of an everyday listening environment becomes much more realistic when it conveys also the auditory information about ecological events such as those seen above [11]. In this paper we will deal in particular with *crushing*, *walking* and *running*.

Physics-based sound synthesis has proved its validity in the modeling of elementary ecological sounds, such as impacts and collisions: these sounds can be synthesized importing the nonlinear hammer-string interaction block [8] inside a model that captures the most relevant physical phenomena occurring during a

collision (impact) between two objects [12, 13, 14]. This model yields, at its interface level, a set of controls for recreating realistic collisions between objects made of different materials. The same controls enable the user to interact with the virtual objects directly.

To describe higher-level ecological events, such as crushing or walking [15], we have developed a bottom-up design approach that, starting from the existing impact block, builds up higher-level sound models. In this way we respect the original physical modeling approach, according to which a virtual musical instrument is assembled by putting together some basic, general-purpose building blocks.

In Section 2 and 3 we will describe the way individual impacts can be assembled together producing a single crushing event. As we will see, crushing events are the result of a statistical rather than deterministic sequence of impacts, in a way that we are not asked to figure out formulas expressing closed-form relationships between different types of sounding objects (those relationships need to be calculated, for example, in the study of contact sounds: handling them is usually not easy [14]). We instead will find a way to create consistent (from a psychophysical point of view) collections of "atomic" (impact) sounds starting from the stochastic description of so-called *crumpling sounds*.

In Section 4, we will describe how a set of crushing events can be controlled by *rules* so to produce realistic sequences of walking and running sounds.

2. CRUMPLING SOUND

Crumpling sounds occur whenever our hearing system identifies a source whose emission, for some reason, is interpreted as a superposition of microscopic crumpling events.

Aluminum cans emit a characteristic sound when they are crushed by a human foot that, for example, compresses them along the main axis of the cylinder. This sound is the result of a composition of single crumpling events, each one of those occurring when, after the limit of bending resistance, one piece of the surface forming the cylinder splits into two facets as a consequence of the force applied to the can.

The exact nature of a single crumpling event depends on the local conditions the surface is subjected to when folding occurs between two facets. In particular, the types of vibrations that are produced are influenced by shape, area, and neighborhood of each facet. Also other factors play a role during the generation of the sound, such as volume and shape of the can. The can acts as a volume-varying resonator during the crumpling process.

A precise assessment of all the physical factors determining the sound which is produced by a single crumpling event is beyond the scope of this work. Moreover, there are not many studies available in the literature outlining a consistent physical background for these kinds of problems. On the other hand, it is likely that our hearing system cannot distinguish such factors but the most relevant ones. For this reason we generate individual crumpling sounds using the impact model.

Studies conducted on the acoustic emission from wrapping sheets of paper [16] concluded that crumpling events do not determine *avalanches* [17], so that fractal models in principle cannot be used to synthesize crumpling sounds [18]. Nevertheless, crumpling paper emits sound in the form of a stationary process made of single impulses, whose individual energy E can be described by the following power law:

$$P(E) = E^{\gamma} \quad , \tag{1}$$

where γ has been experimentally determined to be in between -1.3 and -1.6. On the other hand a precise dynamic range of the impulses is not given, although the energy decay of each single impulse has been found out to be exponential.

The temporal patterns defined by the events is an important factor determining the perceptual nature of the crumpling process. A wide class of stationary temporal sequences can be modeled by *Poisson's processes*; each time gap τ between two subsequent events in a temporal process is described by an exponential random variable with *density* $\lambda > 0$ [19]:

$$P(\tau) = \lambda e^{\lambda\tau} \text{ with } \tau \geq 0 \quad . \tag{2}$$

Assuming a time step equal to T, we simply map the time gap over a value kT defined in the discrete-time domain:

$$k = \text{round}(\tau/T) \quad , \tag{3}$$

where round$(\cdot)$ gives the closest integer to its argument value.

The crumpling process consumes energy during its evolution. This energy is provided by the agent that crushes the can. The process terminates when the transfer of energy does not take place any longer, i.e., when a *reference energy*, E_{tot}, has been spent independently by each one of the impulses forming the event s_{tot}:

$$s_{\text{tot}}[nT] = \sum_i E_i s[nT - k_i T] \quad \text{with} \quad E_{\text{tot}} = \sum_i E_i \quad , \tag{4}$$

where $s(nT)$ is a signal having unitary energy, accounting for each single crumpling.

In order to determine the dynamic range, suppose to constrain the individual energy E to assume values in the range $[m, M]$. The probability P that an individual impulse falls in that range is, using the power law expressed by (1):

$$P[m \leq E < M] = \int_m^M E^{\gamma} dE = 1 \quad . \tag{5}$$

This equation allows to calculate an explicit value for m if M is set to be the value corresponding to full-scale, beyond which the signal would clip. In this case we find out the minimum value coping with (5):

$$m = \{M^{\gamma+1} - \gamma - 1\}^{\frac{1}{\gamma+1}} \quad . \tag{6}$$

2.1. Driving the impact model

During a crushing action, the creases on the object's surface become increasingly dense. Hence, vibrations over the facets increase in pitch since they are bounded within areas that become progressively smaller. This hypothesis inspires our model for determining the pitch in a single crumpling sound.

Given a segment having a nominal length D_0, initially marked at the two ends, let us start the following procedure: Each time a new impulse is triggered, a point of this segment is randomly selected and marked. Then, two distances are measured between the position of this mark and its nearest (previously) marked points. The procedure is sketched in Figure 1, and it is repeated until some energy, as expressed by (4), is left to the process. The values L_i

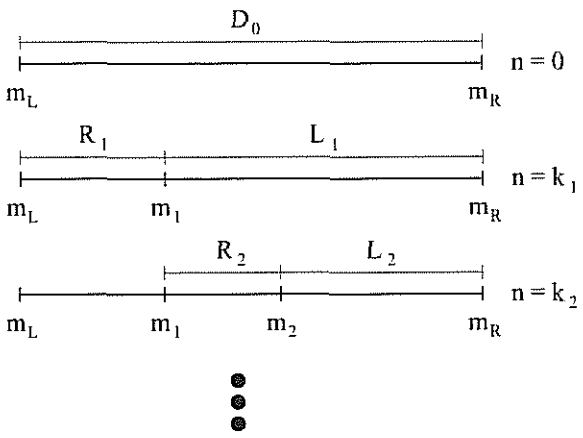


Figure 1: *Sketch of the procedure used to calculate the pitch of the impulses as long as the process evolves.*

and R_i, corresponding to the distances calculated between the new mark m_i (occurring at time step k_i) and the leftward and rightward nearest marks (occurred at previous time steps), respectively, are used as absolute values for the calculation of two driving frequencies, f_L and f_R, and two decay times, τ_L and τ_R, and also as relative weights for sharing the energy E_i between the two impacts, x_{f_L,τ_L} and x_{f_R,τ_R}, forming each crumpling sound:

$$\begin{aligned} E_i s[nT - k_i T] &= E_i \frac{L_i}{L_i + R_i} x_{f_L,\tau_L}[nT - k_i T] \\ &+ E_i \frac{R_i}{L_i + R_i} x_{f_R,\tau_R}[nT - k_i T] \quad , \end{aligned}$$

where the driving frequencies (decay times) are in between two extreme values, f_{MAX} (τ_{MAX}) and f_{MIN} (τ_{MIN}), corresponding to the driving frequencies (decay times) selected for a full and a minimum portion of the segment, respectively:

$$\begin{aligned} f_L &= f_{\text{MAX}} - \frac{L_i}{D_0}(f_{\text{MAX}} - f_{\text{MIN}}) \\ f_R &= f_{\text{MAX}} - \frac{R_i}{D_0}(f_{\text{MAX}} - f_{\text{MIN}}) \quad , \end{aligned}$$

in which the symbols f must be substituted by τ in the case of decay times.

We decided to operate on the so-called "frequency factor" and decay time of the impact model: the former is related to the size of the colliding object; the latter accounts for the object material [14]. We considered both of them to be related to the crumpling facet area: *the smaller the facet, the higher-pitched the fundamental* frequency and the shorter the decay time of the emitted sound.

3. CAN CRUSHING

In the case of the aluminum can, crushing occurs in consequence of some force applied to it. This action is usually performed by an agent having approximately the same size as the can surface, such as the sole of a shoe. As the agent compresses the can, sound emission to the surrounding environment changes since the active emitting surface of the can is shrinking, and some of the creases become open fractures in the surface. We suppose that the internal pressure in the can is maximum in the beginning of the crushing process. This pressure relaxes to the atmospheric value as long as the process evolves, due to pressure leaks from the holes appearing in the surface, and due to the decreasing velocity of the crushing action. These processes, if any[1], have a clear effect on the evolution in time of the spectral energy: high frequencies are gradually spoiled of their spectral content, as it can be easily seen from Figure 2 where the spectrogram of a real can during crushing has been plotted.

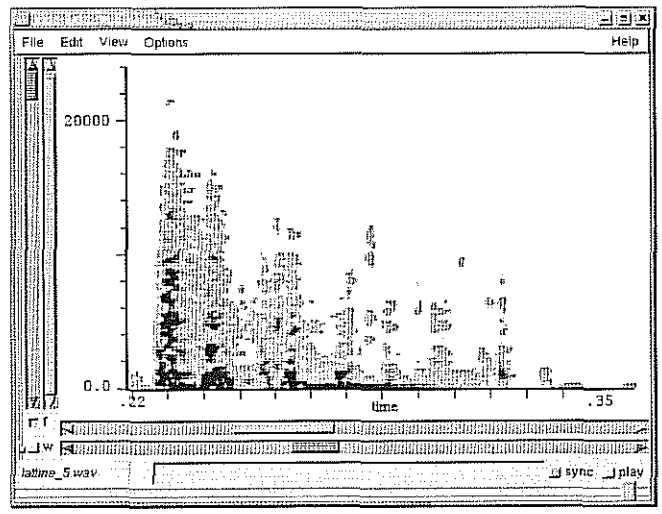


Figure 2: *Spectrogram of the prototype sound of a crushing can.*

The whole process is interpreted in our model as a time-varying resonating effect, realized through the use of a low-selectivity linear filter whose lowpass action over the sound s_{tot} is slid toward the low-frequency as long as the process evolves. Lowpass filtering is performed using a first-order lowpass filter [20]. In the case of crushing cans we adopted the following filter parameters:

- lowest cutoff frequency $\Omega_{\text{MIN}} = 500$ Hz
- highest cutoff frequency $\Omega_{\text{MAX}} = 1400$ Hz.

Using those parameters, the cutoff frequency is slid toward the lowest value as long as energy is spent by the process. More precisely, the cut frequency ω_i at time step k_i is calculated according to the following rule:

$$\omega_i = \Omega_{\text{MIN}} + \frac{E_{\text{tot}} - \sum_{k=1}^{i} E_k}{E_{\text{tot}}} (\Omega_{\text{MAX}} - \Omega_{\text{MIN}}) \quad . \qquad (7)$$

This kind of post-processing contributes to give a smooth, progressively "closing" characteristic to the crumpling sound.

3.1. Parameterization

Several parameter configurations have been tested during the tuning of the model. It has been noticed that some of the parameters have a clear (although informal) *direct* interpretation:

- E_{tot} can be seen as an "image" of the *size*, i.e., the height of the cylinder forming the can. This sounds quite obvious, since E_{tot} governs the time length of the process, and this length can be in turn reconducted to the can size. Sizes

[1] We are still looking for a thorough explanation of what happens during crushing.

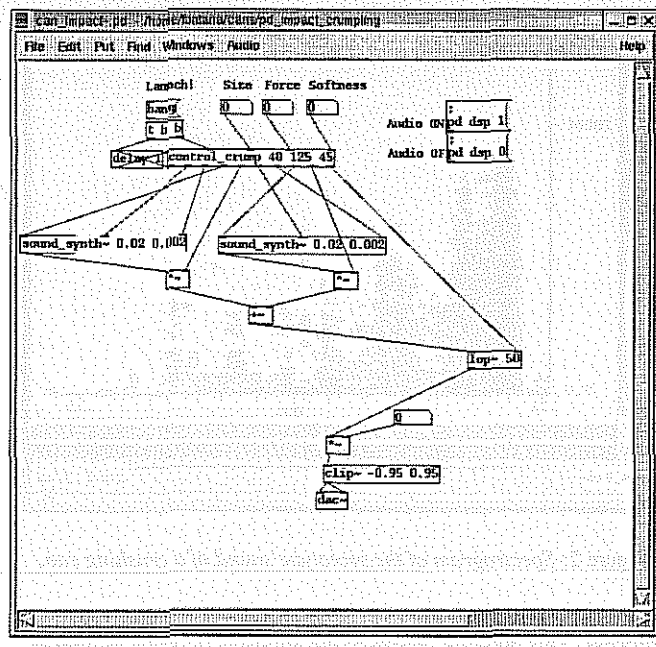


Figure 3: *Screenshot of the pd-module implementing the crushing can model.*

which are compatible with a natural duration of the process correspond to potential energies ranging between 0.001 and 0.1;

- low absolute values of γ result in more regular realizations of the exponential random variable, whereas high absolute values of the exponential statistically produce more peaks in the event dynamics. Hence, γ can be seen as a control of *force* applied to the can. This means that for values around -1.5 the can seems to be heavily crushed, whereas values around -1.15 evoke a softer crushing action. Thus, γ has been set to range between -1.15 and -1.5;
- "soft" alloys forming the can can be bent more easily than stiff alloys: holding the same crushing force, a can made of soft, bendable material should shrink in fewer seconds. For this reason, the parameter p_s governing the frequency of the impulses in the Poisson process can be related to the material stiffness: the higher p_s, the softer the material. *Softness* has been set to range between 0.001 (stiff can) and 0.05 (soft can).

We noticed that variations in the definition and parameterization of the crumpling sound s can lead to major differences in the final sound. On the other hand the statistical laws which we used for generating crushing sounds are general enough for reproducing a wide class of events, including paper crumpling and plastic bottle crushing. In that case s must be properly shaped to accommodate for different kinds of events.

3.2. Implementation as *pd* patch

Crushing cans have been finally implemented as a pd *patch* [21]. The modular implementation of pd allows to have the crumpling sound synthesis model, the higher-lever statistical module, and the time-varying post-processing lowpass filter decoupled inside the same patch (see Figure 3). The only limitation with this implementation is represented by the firing rate used by the statistical module (labeled as `control_crump`) to feed control data to the two sound synthesis modules (labeled as `sound_synth~`) producing the two signals x_{f_L} and x_{f_R}. This limitation comes from the presence of a structure containing the statistical module in loop-back with a `delay` block, which limits the shortest firing rate to 1 ms.

On the other hand the chosen implementation allows a totally independent design of the statistical module and of the sound synthesis module. The statistical module has been realized in C language, as a pd class, whereas the sound synthesis module has been implemented as a sub-patch nesting inside itself several pre-existing building blocks, which come together with pd. This modular/nested approach leaves the patch open to independent changes inside the modules, and qualifies the patch in its turn for use as an individual block inside higher-level patches. For this reason we could straightforwardly integrate crushing sounds in a framework including specific rules for producing walking and running sounds.

4. WALKING AND RUNNING SOUNDS

Most of sound synthesis techniques available today are capable of reproducing excellent sounds. At the same time these techniques do not take into consideration the fact that sounds usually appears in a sequence, such in the case of a music melody or of a sequence of footsteps. For example the synthesis of a sequence of trumpet sounds can give poor results if the musical context and the performance style, such as in *legato* and *staccato* articulation, is not taken into consideration. Therefore, in order to arrange single events in a meaningful way, we need control functions for the specific sound synthesis technique that is used [22].

Physics-based synthesis techniques allow a direct manipulation of parameters connected to physical properties of the sound model. These techniques are therefore particularly suitable to be controlled for the production of organized sound sequences. In particular, a complete model for the production of walking sound has been recently developed by Cook, following a similar approach [15]. One major difference lies in the low-level synthesis model: in that case, each individual impact is generated using a signal-based rather than a physics-based approach. On the other hand, the upper levels are designed starting from considerations on the statistics and the energy of the process, hence having several similarities with the corresponding control layers used in our model.

4.1. Control models

Some recent works reported strong relations between body motion and music performance. This is mainly due because sound production is a result of body movements such as hands, wrists, harms, shoulders and feet in pianists and in percussionists, lips, lungs, and tung in singers and in wind instrument players. Besides these obvious relations, some researchers have found more indirect associations between music performance and body motion. Friberg and Sundberg demonstrated the way the *Final Retard* in performances of Baroque music can be derived from measurements of stopping runners [24].

In another work [26] it was found that *legato* and *staccato* playing in piano performance can be associated to walking and running respectively. During walking there is an overlap time when both feet are on the ground, and in *legato* scale playing there is an overlap time (key overlapped time) when two fingers

are pressing two keys simultaneously. During running there is an time interval when both feet are in the air, and in *staccato* scale playing there is a time interval (key detached time) when all fingers are in the air simultaneously.

These and other results contributed to the design of some performance rules implemented in the Director Musices performance system [27, 28] and related to body motion. In previous experiments it has been demonstrated that *legato* and *staccato* are among the expressive parameters which help in synthesizing performances which are recognized as sad and respectively as happy by listeners [23, 29].

Given the considerations above, it was tempting trying to apply the performance rules, which present similarities with walking and running, to the control of crumpling sounds.

4.2. Controlling footstep sounds

In a first experiment, performance rules where used for the control of the sound of one single footstep extracted from a recorded sequence of walking and running sounds on gravel [26]. Stimuli including both sequences of footsteps controlled with the rules, and sequences of footsteps obtained by looping the same sound were produced. These stimuli were proposed to subjects in a listening test. Subjects could distinguish between walking and running sounds, and classified the stimuli produced with the performance rules as more natural. Still the sound controlled was a static sound. Thus, a natural development is to control a physics-based model implementing footstep sounds.

Informal listening test of the sound produced by the crushing can model, presented in Section 3, classified this sound as that of a footstep on cold snow. This result suggested the possibility of controlling the model using performance rules such those for the control of *Final Retard*, and *legato* and *staccato* articulation.

A pd-module implementing the control of footstep sounds is represented in Figure 4. A sequence of footstep sounds with IOI set by the "Tempo" slider is produced by pressing the `Walk` button. If the `Slow_down` button is pressed, the IOIs become larger, i.e. slower tempo. This effect is implemented using the *Final Retard* rule [24]. Parameters `Size`, `Force`, and `Softness` are set to values which give a realistic sound of footstep on cold snow. In the case of walking/running sounds these three parameters represent the size of the foot, the force applied by the foot, and the softness of the ground.

The spectrogram of walking sounds produced by the module of Figure 4 is shown in Figure 5. The IOI between footsteps increases because the `Slow_down` button was pressed and the "Final Retard" rule was activated.

An outcome of the present work is that analogies between locomotion and music performance, briefly presented in Section 4, combined with the control properties of physics-based sound modeling, open new possibilities for the design of control models for artificial walking sound patterns, and in general for sound control models based on locomotion.

5. ONGOING RESEARCH

In the particular case of walking/running sounds, we have observed that the "Force" and "Softness" parameters should vary with "Tempo". Higher tempi should correspond to larger values of the force and lower degree of softness. A next step in our research will be therefore the identification of such a mapping. Studies on

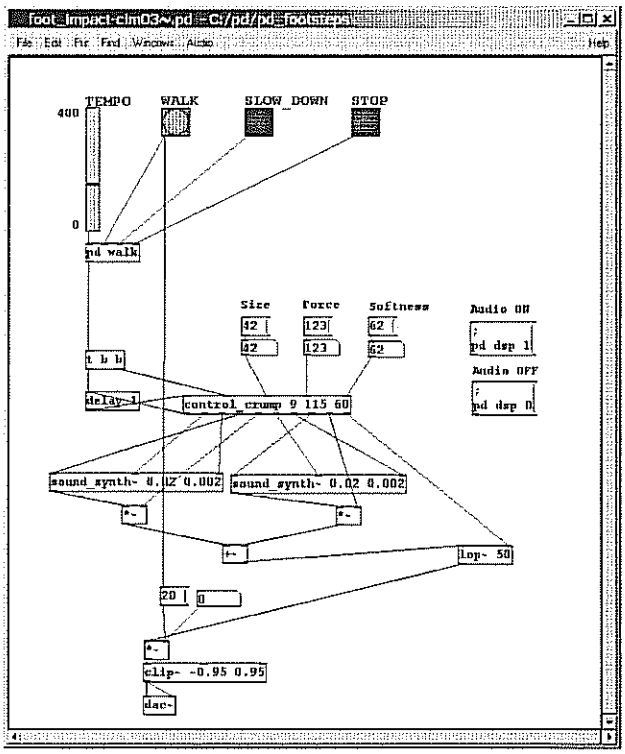


Figure 4: *Screenshot of a pd-module implementing the control of footstep sounds.* `Tempo` *(in bpm),* `Walk`, `Slow_Down`, *and* `Stop` *control the timing of walking steps.* `Size`, `Force` *and* `Softness` *determine the foot size, its force on the ground and its softness. The model controlled is that of the crushing can presented in Figure 3.*

expressive walking could help toward a possible solution. In a recent study [25] ground reaction force by the foot during different gaits was measured. Value and time envelope of this force varied with the different walking technique adopted by subjects. These techniques were also characterized by different tempi.

In general, the proposed rule-based approach for sound control is only a step toward the design of more general control models that respond to physical gestures. Ongoing research in the framework of the Sounding Object project is dealing with the implementation of rule-based control of impact sound models and of friction sound models.

6. ACKNOWLEDGMENTS

This work was supported by the European Commission: SOb - The Sounding Object project, no. IST-2000-25287, `http://www.soundobject.org`.

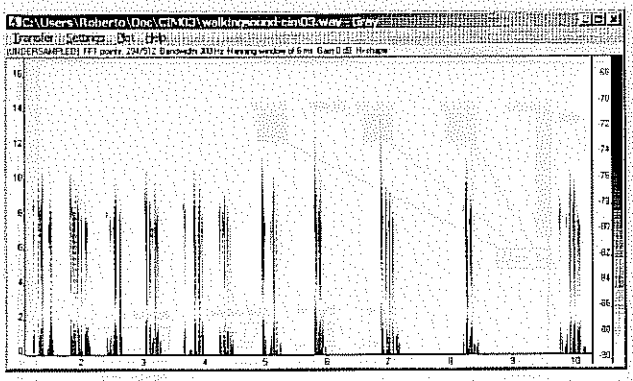


Figure 5: *Spectrogram of a stopping walking sound produced by the pd-patch of Figure 4: the inter-onset-interval (IOI) between footsteps becomes larger.*

7. REFERENCES

[1] G. De Poli, A. Piccialli, and C. Roads, *Representations of Musical Signals*, MIT Press, Cambridge, Mass., 1991.

[2] K. Karplus and A. Strong, "Digital Synthesis of Plucked String and Drum Timbres," *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, 1983.

[3] M. Karjalainen, U. K. Laine, T. I. Laakso, and V. Välimäki, "Transmission-line modeling and real-time synthesis of string and wind instruments," in *Proc. Int. Computer Music Conf.*, Montreal, Canada, 1991, ICMA, pp. 293–296.

[4] G. Borin, G. De Poli, and A. Sarti, "Algorithms and structures for synthesis using physical models," *Computer Music Journal*, vol. 16, no. 4, pp. 30–42, 1992.

[5] J. O. Smith, "Physical modeling synthesis update," *Computer Music Journal*, vol. 20, no. 2, pp. 44–56, 1996.

[6] C. Cadoz, A. Luciani, and J.-L. Florens, "CORDIS-ANIMA: A Modeling and Simulation System for Sound Synthesis - The General Formalism," *Computer Music Journal*, vol. 17, no. 1, pp. 19–29, Spring 1993.

[7] G. De Poli and D. Rocchesso, "Physically-based sound modeling," *Organised Sound*, vol. 3, no. 1, 1998.

[8] G. Borin and G. De Poli, "A hammer-string interaction model for physical model synthesis," in *Proc. XI Colloquium Musical Informatics*, Bologna, Italy, Nov. 1995, AIMI, pp. 89–92.

[9] V. Välimäki, C. Erkut, and M. Laurson, "Sound synthesis of plucked string instruments using a commuted waveguide model," in *Proc. of the 17th International Congress on Acoustics (ICA)*, Rome, Italy, Sept. 2001, vol. 4, Available only on CD.

[10] W. Gaver, "How do we hear in the world? explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, Apr. 1993.

[11] G. Kramer, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison-Wesley, Reading, MA, 1994.

[12] F. Avanzini and D. Rocchesso, "Controlling material properties in physical models of sounding objects," in *Proc. Int. Computer Music Conf.*, La Habana, Cuba, Sept. 2001, pp. 91–94.

[13] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. Conf. on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, Dec. 2001, pp. 61–66.

[14] M. Rath, D. Rocchesso, and F. Avanzini, "Physically based real-time modeling of contact sounds," in *Proc. Int. Computer Music Conf.*, Göteborg, Sept. 2002.

[15] P. R. Cook, *Real Sound Synthesis for Interactive Applications*, A. K. Peters, L.T.D., 2002.

[16] P. A. Houle and J. P. Sethna, "Acoustic emission from crumpling paper," *Physical Review E*, vol. 54, no. 1, pp. 278–283, July 1996.

[17] J. P. Sethna, K. A. Dahmen, and C. R. Myers, "Crackling noise," *Nature*, , no. 410, pp. 242–250, Mar. 2001.

[18] M. R. Schroeder, *Fractal, Chaos, Power Laws: Minutes from an Infinite Paradise*, W.H. Freeman & Company, New York, NY, 1991.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2nd edition, 1984.

[20] S. K. Mitra, *Digital Signal Processing. A computer-Based Approach*, McGraw-Hill, New York, 1998.

[21] M. Puckette, "Pure data," in *Proc. Int. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997, ICMA, pp. 224–227.

[22] R. Dannenberg and I. Derenyi, " Combining Instrument and Performance Models for high-quality Music Synthesis," *Journal of New Music Research*, vol. 27, no. 3, 1998, pp. 211–238.

[23] R. Bresin and A. Friberg, "Emotional coloring of computer controlled music performance," *Computer Music Journal*, vol. 24, no. 4, 2000, pp. 44–62.

[24] A. Friberg and J. Sundberg, "Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners," *J. Acoust. Soc. Amer.*, Vol. 105, No. 3, 1999, pp. 1469–1484.

[25] A. Friberg, J. Sundberg, and L. Frydén, "Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion," *Journal of New Music Research*, Vol. 29, No. 3, 2000, pp. 199-210.

[26] R. Bresin, A. Friberg, and S. Dahl, "Toward a new model for sound control", in *Proceedings of the COST-G6 Conference on Digital Audio Effects - DAFx01*, Limerick, Ireland, 2001, pp. 45-49.

[27] A. Friberg, V. Colombo, L. Frydn, and J. Sundberg, "Generating Musical Performances with Director Musices", *Computer Music Journal*, vol. 24, no. 3, 2000, pp. 23-29.

[28] R. Bresin, A. Friberg, and J. Sundberg, "Director musices: The KTH performance rules system", in *Proceedings of SIGMUS-46*, Kyoto, 2002, pp. 43-48.

[29] P. N. Juslin, A. Friberg, and R. Bresin, "Toward a computational model of expression in performance: The GERM model," *Musicae Scientiae*, Special issue 2001-2002, 2002, pp. 63-122.

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

ANALYSIS AND PROCESSING OF SOUNDS BY MEANS OF THE PITCH-SYNCHRONOUS MDCT

Gianpaolo Evangelista and Sergio Cavaliere

Dipartimento di Scienze Fisiche
Università di Napoli "Federico II"
gianpaolo.evangelista@na.infn.it

ABSTRACT

In this paper we explore the Modified Discrete Cosine Transform (MDCT), based on cosine modulated filter banks, as a tool for the analysis and processing of musical tones. We show that a complexification of the MDCT channels yields a suitable representation of sounds in which extraction of characteristics such as amplitude envelopes and instantaneous frequencies of the partials is efficiently and accurately achieved. As an extension, we introduce a pitch-synchronous version of the MDCT, the PS-MDCT, which can be adapted to the time-varying pitch of instrument sounds. The PS-MDCT is an invertible, orthogonal transform of signals that can be used as a synthesis by analysis technique. The implications are that this representation is useful in several audio effects like pitch shifting and time stretching.

1. INTRODUCTION

In sinusoidal representations [1], in Spectral Modeling Synthesis [2] based on sinusoids plus noise, as well as, in the phase vocoder [3][6][5][4][7], the Short-Time Fourier Transform (STFT) plays a central role for the analysis and synthesis of the instantaneous amplitudes and frequencies of the sound partials. The intuition is that to each individual partial there corresponds a spectral peak, whose position, amplitude and phase can be estimated by means of the discrete Fourier transform. Peak parameter estimation can be updated in time by subdividing the signal into several, possibly overlapping, windowed frames composing the STFT. Tracking of the spectral peaks is necessary in order to cope with real-life signals in which the pitch of the partials is not constant. A pitch-synchronous version of the STFT can also be devised to analyze pseudo-periodic or pseudo-harmonic sounds, e.g., tones with slight frequency deviation as in vibrato. In that case, the frame length and hop size of the STFT can be adapted to the fluctuations of the pitch. Other alternatives include modification, e.g. regularization, of the signal by remapping each period to a standard one or by frequency warping. Misalignment of the signal pitch with the analysis frequency bins causes, in any case, leakage into adjacent bins. Therefore the energy of a partial, even when this is very coherent and narrow band, spreads out to several adjacent bins. Consequently, extraction of characteristics has to take this fact into account in order to obtain meaningful estimates. It is well known, for example, that extraction of the amplitude envelopes of the partials is affected by ringing error unless a proper number of frequency bins are averaged.

In this paper we focus on an alternate signal representation based on the Modified Discrete Cosine Transform (MDCT). The representation can be made pitch-synchronous by conforming each period of the signal to the longest expected period. The signal is then analyzed by means of a cosine modulated filter bank having a constant number of channels equal to the number of samples in the maximum period. A similar representation was introduced by one of the authors as a building block in the Fractal Additive Synthesis method [8][9]. Unlike the pitch-synchronous STFT, in which there exist basis elements resonating on the harmonic frequencies of the partials, in the pitch-synchronous MDCT (PS-MDCT) each harmonic partial is split into a pair (doublet) of sinusoidal terms, each lying on one side of the harmonics and representing a single side-band. A superposition of these two terms provides a broad-band representation of each of the harmonics. The MDCT of a real signal is real valued. However, as we will show, both amplitude and phase of each harmonics can be estimated by introducing a complexification of the doublets. The main advantage of using the MDCT is that no further averaging of frequency bins is required in order to achieve a reasonable estimation of these quantities. This property can be exploited in synthesis parameter extraction, as well as, in special effects, such as pitch shifting and time stretching [16][15].

2. THE MODIFIED DISCRETE COSINE TRANSFORM

The Modified Discrete Cosine Transform (MDCT) is contained in the MPEG-1 audio coding standard and its subsequent versions as a signal representation suitable for perceptually adaptive quantization. The MDCT is based on a sliding window version of the type IV Discrete Cosine Transform (DCT). A signal $s(n)$ is analyzed by means of the orthogonal projections

$$S_p(r) = \langle s, \varphi_{p,r} \rangle = \sum_k s(k)\varphi_{p,r}(k),$$

on the basis set

$$\varphi_{p,r}(k) = \varphi_{p,0}(k - rP), \quad p = 0, ..., P-1,\ r \in \mathbb{Z},$$

where

$$\varphi_{p,0}(k) = \sqrt{\frac{2}{P}} w(k) \cos\left((2k + 1 - P)(2p + 1)\frac{\pi}{4P} \right).$$

This basis set is obtained by translation of a window $w(k)$ modulated by a cosine term. The MDCT can be implemented by means of the critically sampled P-channel filter bank shown in figure 1. In order to ensure invertibility, the length $2P$ window must satisfy the following conditions [13][12][11]:

$$w^2(k) + w^2(k + P) = 1$$
$$w(k) = w(2P - 1 - k).$$

A particularly popular window satisfying the above requirements is the sine window:

$$w(l) = \begin{cases} \sin\left(\frac{\pi}{2P}\left(k+\frac{1}{2}\right)\right), & k = 0, ..., 2P-1 \\ 0 & otherwise \end{cases}. \quad (1)$$

The design of longer windows preserving the orthogonality and invertibility of the transform is also feasible [14].

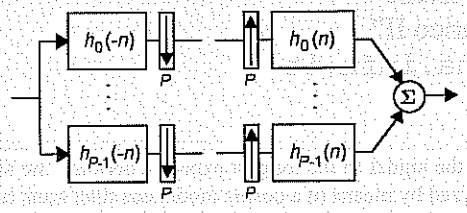


Figure 1: *P-channel critically sampled filter bank to compute the MDCT and its inverse.*

In the frequency domain, each basis element $\varphi_{p,r}(k)$ corresponds to a narrow band centered on the frequency

$$\omega_p = \left(p + \frac{1}{2}\right)\frac{\pi}{P}. \quad (2)$$

Therefore the center frequencies of the basis elements occur at half-integer multiples of a fixed frequency $\frac{\pi}{P}$. These have to be compared with the analysis frequencies of the DFT, which are integer multiples of $\frac{2\pi}{P}$. The situation is depicted in figure 2, where the MDCT analysis bands are shown together with the analysis frequencies of a length-P DFT represented by vertical lines. It should be noticed that the analysis frequencies of the pitch-synchronous STFT are usually integer submultiples of the $\frac{2\pi}{P}p$ frequencies. This is due to the usual choice of an analysis window covering more than one period P. In fact, in the non-overlap case of length P window the choice is restricted to a rectangular window having bad frequency selectivity. As a result the information concerning the harmonics highly overlaps, thus degrading envelope estimation. The choices of a length $2P$ window are wider. However, the "mute" frequencies that are odd integer multiples of $\frac{\pi}{P}$ fall exactly halfway between the harmonics. Therefore the information contained in the corresponding frequency bins is shared by two adjacent harmonics. This frequency overlap impairs correct envelope estimation. Performance is improved by using a length $3P$ or longer window, at the expense of time resolution. However, in that case, averaging of the frequency bins for accurate amplitude and especially phase estimation requires many terms. Moreover, while amplitude estimation can be achieved by means of the square root of the sum of the energies of neighbor bins, combining these bins in order to improve phase estimation is more cumbersome.

Accurate estimation of the instantaneous frequency requires phase differences between adjacent STFT frames [10]. The same method applies to the complexified MDCT coefficients, as shown in the next section.

2.1. Estimation of Instantaneous Amplitude and Frequency

Consider the analysis of a P-periodic signal by means of MDCT. Each harmonics is represented by essentially two terms given by the two sidebands of each line shown in figure 2. More precisely, in order to represent the harmonic at frequency $\frac{2\pi q}{P}$ we need two terms, one at frequency $\frac{2\pi q}{P} - \frac{\pi}{2P}$ and the other one at frequency $\frac{2\pi q}{P} + \frac{\pi}{2P}$. These terms are obtained from (2) by selecting $p = 2q-1$ and $p = 2q$, respectively, corresponding to two sidebands of the harmonics of a P-periodic signal.

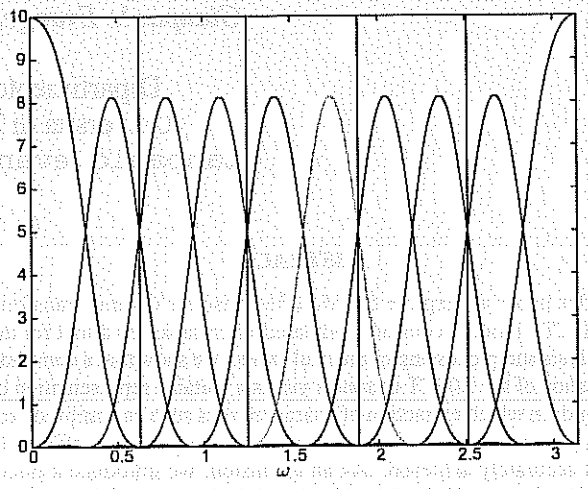


Figure 2: *Analysis bands of the MDCT*

In order to combine the pair of adjacent channels corresponding to each harmonic we form the complex doublet

$$C_q(r) = S_{2q-1}(r) + jS_{2q}(r). \quad (3)$$

In polar form we have

$$C_q(r) = \rho_q(r)e^{j\theta_q(r)}$$

with

$$\rho_q(r) = \sqrt{S^2_{2q-1}(r) + S^2_{2q}(r)}$$

and

$$\theta_q(r) = \arctan\left(\frac{S_{2q}(r)}{S_{2q-1}(r)}\right).$$

The sequences $\rho_q(r)$ and $\theta_q(r)$ respectively are the amplitude and the phase of the doublet. While $\rho_q(r)$ can be directly interpreted as a downsampled version of the amplitude envelope of the q-th harmonics, the phase $\theta_q(r)$ is a relative term. Nevertheless, the derivative of the phase can be interpreted as the instantaneous frequency shift from the harmonic frequency $\frac{2\pi q}{P}$. In order to see this, consider the case of a sinusoidal input signal

$$s(k) = \cos(\omega k)$$

with frequency ω not necessarily multiple integer of $\frac{2\pi}{P}$, i.e., where the period of the discrete-time signal is not integer and differs from the number of channels P. In this case

$$S_p(r) = \langle s, \varphi_{p,r}\rangle = \sqrt{\frac{2}{P}}\,\mathrm{Re}\left\{\sum_k e^{-jk\omega} f_p(k - rP)\right\} =$$

$$= \sqrt{\frac{2}{P}}\,\mathrm{Re}\left\{e^{-jrP\omega}F_p(\omega)\right\},$$

where

$$f_p(k) = w(k)\cos\left(\frac{k(2p+1)\pi}{2P} + \alpha_p\right),$$

with

$$\alpha_p = \frac{(1-P)(2p+1)\pi}{4P}. \tag{4}$$

Consequently, the DFT $F_p(\omega)$ of $f_p(k)$ is

$$F_p(\omega) = \frac{e^{j\alpha_p}}{2} W\left(\omega - \frac{(2p+1)\pi}{2P}\right) + \\ + \frac{e^{-j\alpha_p}}{2} W\left(\omega + \frac{(2p+1)\pi}{2P}\right).$$

Writing ω as $\omega = \frac{2\pi q}{P} + \Delta\omega$, with $|\Delta\omega| \leq \frac{\pi}{P}$, then:

$$S_p(r) = \sqrt{\frac{2}{P}}\,\mathrm{Re}\left\{e^{-jrP\Delta\omega} F_p\left(\frac{2\pi q}{P} + \Delta\omega\right)\right\} \tag{5}$$

Since $W(\omega)$ is essentially nonzero in a neighborhood of $\omega = 0$ then only the elements for $p = 2q-1$ and $p = 2q$ are essentially nonzero with

$$F_{2q-1}\left(\frac{2\pi q}{P} + \Delta\omega\right) \approx \frac{e^{j\alpha_{2q-1}}}{2} W\left(\Delta\omega + \frac{\pi}{2P}\right) \\ F_{2q}\left(\frac{2\pi q}{P} + \Delta\omega\right) \approx \frac{e^{j\alpha_{2q}}}{2} W\left(\Delta\omega - \frac{\pi}{2P}\right). \tag{6}$$

Furthermore, for $\Delta\omega \approx 0$ and using as analysis window the sine window (1) we have:

$$W\left(\Delta\omega + \frac{\pi}{2P}\right) \approx \frac{e^{j\frac{\pi}{4P}}}{2j} e^{-j\left(P-\frac{1}{2}\right)\Delta\omega} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}} \\ W\left(\Delta\omega - \frac{\pi}{2P}\right) \approx -\frac{e^{-j\frac{\pi}{4P}}}{2j} e^{-j\left(P-\frac{1}{2}\right)\Delta\omega} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}}, \tag{7}$$

since the other terms in

$$\frac{\sin\left(P\Delta\omega \pm \pi\right)}{\sin\left(\frac{\Delta\omega}{2} \pm \frac{\pi}{2}\right)}$$

are negligible. Thus, we have:

$$S_{2q-1}(r) \approx -\frac{1}{2\sqrt{2P}} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}} \times \\ \sin\left(rP\Delta\omega - \alpha_{2q-1} + \left(P - \frac{1}{2}\right)\Delta\omega - \frac{\pi}{4P}\right) \tag{8}$$

and

$$S_{2q}(r) \approx \frac{1}{2\sqrt{2P}} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}} \times \\ \sin\left(rP\Delta\omega - \alpha_{2q} + \left(P - \frac{1}{2}\right)\Delta\omega + \frac{\pi}{4P}\right). \tag{9}$$

In order to obtain the phase of the complex doublet we make the following considerations. Since

$$\alpha_{2q-1} = \alpha_{2q} + \frac{\pi}{2} - \frac{\pi}{2P}$$

then (8) becomes:

$$S_{2q-1}(r) \approx \frac{1}{2\sqrt{2P}} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}} \cos\left(rP\Delta\omega - \alpha'_{2q}\right)$$

with

$$\alpha'_{2q} = \alpha_{2q} - \left(P - \frac{1}{2}\right)\Delta\omega - \frac{\pi}{4P}, \tag{10}$$

while, with the same notation,

$$S_{2q}(r) \approx \frac{1}{2\sqrt{2P}} \frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}} \sin\left(rP\Delta\omega - \alpha'_{2q}\right).$$

Thus,

$$\theta_q(r) = \arctan\frac{\sin\left(rP\Delta\omega - \alpha'_{2q}\right)}{\cos\left(rP\Delta\omega - \alpha'_{2q}\right)} = rP\Delta\omega - \alpha'_{2q}. \tag{11}$$

Therefore, the phase of the doublet is linear and its slope is proportional, via the factor P, to the frequency deviation of the partial from the center analysis frequency $\frac{2\pi q}{P}$. The instantaneous frequency deviation $\Delta\omega$ can be extracted from the complexified MDCT data by computing the difference of the unwrapped phase θ_q in adjacent frames, for each partial or MDCT doublet indexed by $q = 1, ..., P-1$:

$$[\Delta\omega]_q = \frac{\theta_q(r+1) - \theta_q(r)}{P}. \tag{12}$$

Finally, the instantaneous (normalized) frequencies of the partials are given by

$$[\Delta\omega]_q + \frac{2\pi q}{P}.$$

For the amplitudes we obtain the following result:

$$\rho_q(r) = \frac{1}{2\sqrt{2P}} \left|\frac{\sin P\Delta\omega}{\sin\frac{\Delta\omega}{2}}\right|.$$

Thus, the amplitude is affected by a known scaling factor deriving from the partial frequency detuning from the center analysis frequency. This result continues to hold under the assumption that the amplitude envelope is approximately constant within any window of length $2P$. This formula allows for a simple means to compensate the extracted envelope for the bias introduced by the $\Delta\omega$ detuning.

Envelopes extracted from an oboe sound by means of the tuned MDCT method are shown in figure 3. The original sound contained a certain amount of vibrato and pitch instability. Nevertheless, the number of channels of the MDCT filter bank was tuned to the average number of samples in a signal period. This result should be compared with the estimation of the envelopes using a STFT with a Hann window of length $2P$ and hop size of P samples, shown in figure 4. It is apparent that the STFT extracted envelopes are not accurate, especially in the second half of the signal, due to pitch instability generating interference terms in adjacent frequency bins.

An example of estimation of the instantaneous frequencies of the partials is shown in figure 5. There, the frequency deviation (in Hz) of the first four harmonics of the oboe sounds, extracted from MDCT data by complexification and frame-to-frame differencing is shown. The vibrato and pitch trend of the signal are clearly visible from the figure.

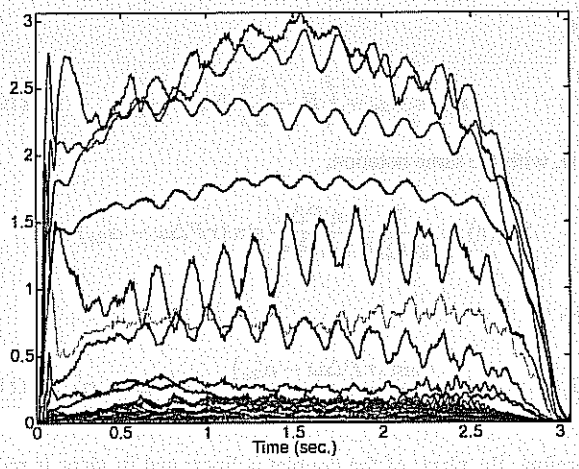


Figure 3: *Tuned MDCT estimation of the envelopes of the partials of an oboe sound.*

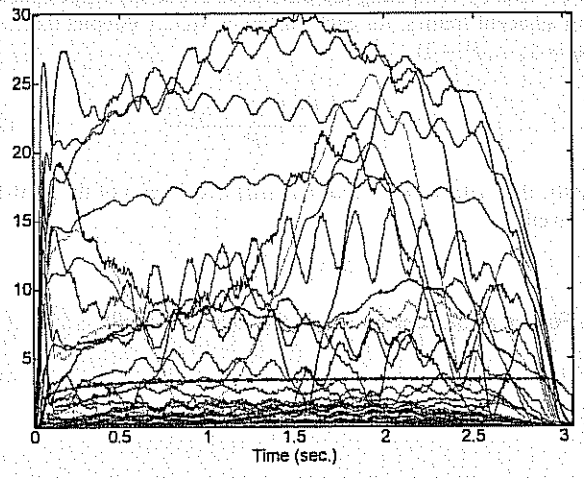


Figure 4: *Tuned STFT estimation of the envelopes of the partials of an oboe sound.*

3. THE PITCH-SYNCHRONOUS MDCT

Glissando or deep vibrato can significantly alter the pitch of sounds in time. In the previous section we showed that accurate estimates of the amplitudes of the partials of pseudo-harmonic tones can be efficiently achieved by means of a tuned MDCT scheme. Due to the analysis bandwidth of $\frac{2\pi}{P}$ around each harmonics, the scheme is robust to small deviations of the pitch. However, for large pitch deviations it is desirable to frequently tune the transform in order to reduce the estimation error. Several possibilities are available for introducing a pitch-adaptive time-varying representation having the same flavor as the MDCT. One method is to design a time-varying perfect reconstruction filter bank allowing for switching the number of channels in time. This turns out to be a difficult problem and at the moment only a partial solution is available where the frequency responses of the filters fail to have satisfactory stop-band rejection properties. Alternately, one can use a pitch-synchronous time-varying frequency warping algorithm [17] in order to stabilize the pitch of the sound prior to MDCT analysis. Although the method is computationally intensive, it gives excellent results. Additionally, an approximate version is available, which is useful for real-time implementations.

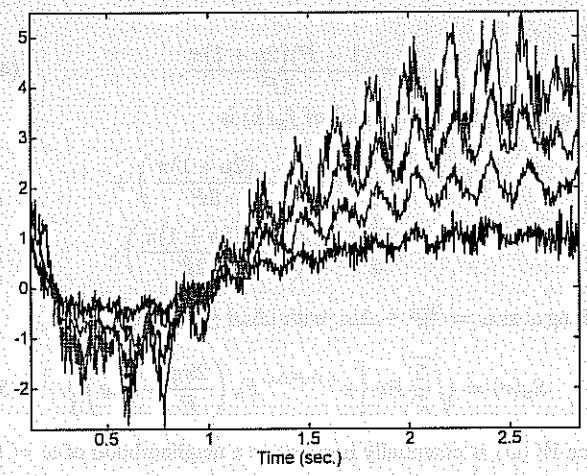


Figure 5: *Instantaneous frequency estimation by means of MDCT: deviation from the center frequencies (in Hz) of the partials of an oboe sound.*

A third way of introducing a PS_MDCT is to use a pitch-synchronous representation of the signal in which each signal period is stored in a variable length array [18]. This method does not require any computation at all except for data indexing or multiplexing. From a signal $s(k)$ we extract a sequence $P(r)$ of period lengths and we form a sequence of vectors $\mathbf{v}(r)$, whose components $v_q(r)$ are given as follows:

$$v_q(r) = s\left(q + \sum_{i=0}^{r-1} P(i)\right) = \sum_k s(k)\delta(k - q - M(r)),$$

where $q = 0, 1, ..., P(r) - 1$, and

$$M(r) = \sum_{i=0}^{r-1} P(i).$$

Each variable-length array can be conformed to a maximum size $P_{\max}$ by suitably extending each signal period. This extension is arbitrary but, in the context of MDCT and wavelet transforms, we found it useful to extend the period as constant by appending to the period samples several repetitions of the last sample in the period itself until we achieve maximum length. In other words, we form the extended period vectors $\widetilde{\mathbf{v}}(r)$ whose components are

$$\widetilde{v}_q(r) = \begin{cases} v_q(r) & if\ 0 \le q < P(r) \\ v_{P(r)-1}(r) & if\ P(r) \le q < P_{\max} \end{cases}.$$

The period-length conformed extended signal is then formed as follows

$$\widetilde{s}(k) = \sum_i \sum_{q=0}^{P_{\max}-1} \widetilde{v}_q(r)\delta(k - q - iP_{\max}).$$

The intuition is that since the MDCT analysis filters are bandpass they act as generalized differentiators. Thus, appending a constant tract to the period will only moderately influence the outputs of the

analysis filter bank. The analysis of the period-conformed signal is then performed by means of a cosine modulated filter bank having a constant number $P_{\max}$ of channels. By means of this extension, each MDCT doublet is tuned to a harmonic partial, independently of pitch variations. Therefore the results of the previous section can be applied in order to obtain instantaneous amplitude and frequency estimates. The only difference is that in order to obtain the instantaneous frequencies one needs to add to the $\Delta\omega$ terms in (12) the time-varying frequencies $\frac{2\pi q}{P(r)}$ instead of the constant ones. Additionally, estimation of the $\Delta\omega$ terms by frame-to-frame differencing has to take into account the time-varying pitch characteristics.

4. CONCLUSIONS

While estimation of characteristics of musical tones such as instantaneous amplitudes and frequencies of the partials has been performed in the past by means of the STFT or vocoder schemes, in this paper we showed that alternate algorithms to perform the same task are available in a suitable MDCT based representations. In many cases the new algorithms are more accurate or efficient than their STFT counterpart. MDCT analysis can be generalized to signals with significantly time-varying pitch, by means of the given pitch-synchronous scheme. The MDCT and its inverse represent alternative building blocks for phase vocoders, with applications to pitch shifting and time stretching.

5. REFERENCES

[1] McAulay, R. J., Quatieri, T., "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, Vol. 34, pp. 744-754, 1986.

[2] Serra, X., Smith, J. O. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, 14(4):12-24, 1990.

[3] Flanagan, J. L., Golden, R. M. "Phase Vocoder," *Bell Syst. Tech. J.*, Vol. 45, pp. 1493-1509, Nov. 1966.

[4] Portnoff, M. R., "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, Vol. 29, pp. 374-387, June 1981.

[5] Dolson, M. "The phase vocoder: A tutorial," *Computer Music Journal*, 10(4):14-27, 1986.

[6] Moorer, J. A. "The Use of the Phase Vocoder in Computer Music Applications," *J. Audio Eng. Soc.*, Vol. 26, pp. 42-45, Jan./Feb. 1976.

[7] Puckette, M. "Phase-locked vocoder," *IEEE ASSP Workshop on Appl. of Signal Proc. to Audio and Acoust.*, pp. 222 -225, La Paltz, New York, Oct 1995.

[8] Polotti, P., Evangelista, G. "Analysis and Synthesis of Pseudo-Periodic 1/f-like Noise by means of Wavelets with Applications to Digital Audio," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 1, pp. 1-14, Hindawi, March 2001.

[9] Polotti, P., Evangelista, G., "Fractal Additive Synthesis by means of Harmonic-Band Wavelets," *Computer Music Journal*, 25(3):22-37, 2001.

[10] Puckette, M. S., Brown, J. C. "Accuracy of frequency estimates using the phase vocoder," *IEEE Trans. on Speech and Audio Proc.*, Vol. 6, No. 2, pp.166 -176, Mar 1998.

[11] Vetterli, M., Le Gall, D. "Perfect reconstruction FIR filter banks: some properties and factorizations," *IEEE Trans. on Acoustic, Speech, and Signal Proc.*, Vol. 37, pp. 1057-1071, July 1989.

[12] Malvar, H. S., "Modulated QMF filter banks with perfect reconstruction," *Electronics Letters*, Vol. 26, No. 13, pp. 906-907, June 1990.

[13] Princen, J. P., Bradley, A. B. "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, Vol. 34, No. 5, pp. 1153-1061, 1986.

[14] Koilpillai, R. D., Vaidyanathan, P. P. "Cosine-modulated FIR filter banks satisfying perfect reconstruction," *IEEE Trans. on Signal Proc.*, Vol. 40, No. 4, pp. 770 -783, Apr 1992.

[15] Laroche, J., Dolson, M. "Improved phase vocoder time-scale modification of audio", *IEEE Trans. on Speech and Audio Proc.*, Vol. 7, No. 3, pp. 323 -332, May 1999.

[16] Portnoff, M. R. "Time-Scale Modification of Speech Based on Short-Time Fourier Analysis," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, Vol. 24, pp. 243-248, June 1976.

[17] Evangelista, G., Cavaliere, S. "Audio Effects Based on Biorthogonal Time-Varying Frequency Warping," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 1, pp. 27-35, Hindawi, March 2001.

[18] Evangelista, G. "Pitch Synchronous Wavelet Representations of Speech and Music Signals," *IEEE Trans. on Signal Proc.*, special issue on Wavelets and Signal Processing, vol. 41, no. 12, pp. 3313-3330, Dec. 1993.

PVLAB: an audio processing program based on Phase Vocoder

Remigio Coco

Classe di Musica Elettronica
Conservatorio "O. Respighi" - Latina
rcoco@jumpy.it

ABSTRACT

The Phase Vocoder is a well-known technique of signal analysis, successfully used for the analysis and resynthesis of audio signals. In this paper, a GUI program for audio signal processing, based on the Phase Vocoder, is presented. The name of the program is PVLAB (Phase Vocoder LABoratory), it is written in C++, and it is available only on Windows platform, for the moment. It can do several types of transformation, grouped into four categories: time transformations, frequency transformations, amplitude transformations and miscellaneous transformations (those transformations that involve more than one domain). It works on mono audio files in WAV format, or analysis files obtained with Csound; stereo files will be supported in future.

1. INTRODUCTION

Phase Vocoder analysis is carried out by means of short-term FFTs; input signal segments can be contiguous or overlapping, giving more or less time resolution. The result of the analysis phase is a number of *frames*, each containing (N/2 + 1) pairs {A,f} (amplitude, frequency), where N is the number of FFT points. PVLAB gives the opportunity to the user to set the number of FFT points, window type (Hamming, Bartlett, etc.) and overlapping amount (number of points between two consecutive frames of analysis – limited to powers of two). In PVLAB, the analysis data are processed further, as we will see, yielding a new set of analysis frames, and then the signal is resynthesized. The resynthesis technique is additive, considering a bank of (N/2 + 1) sinusoidal oscillators updated at each frame with the pairs {A,f} (ref. [1], [2], [3], [4], [5]).

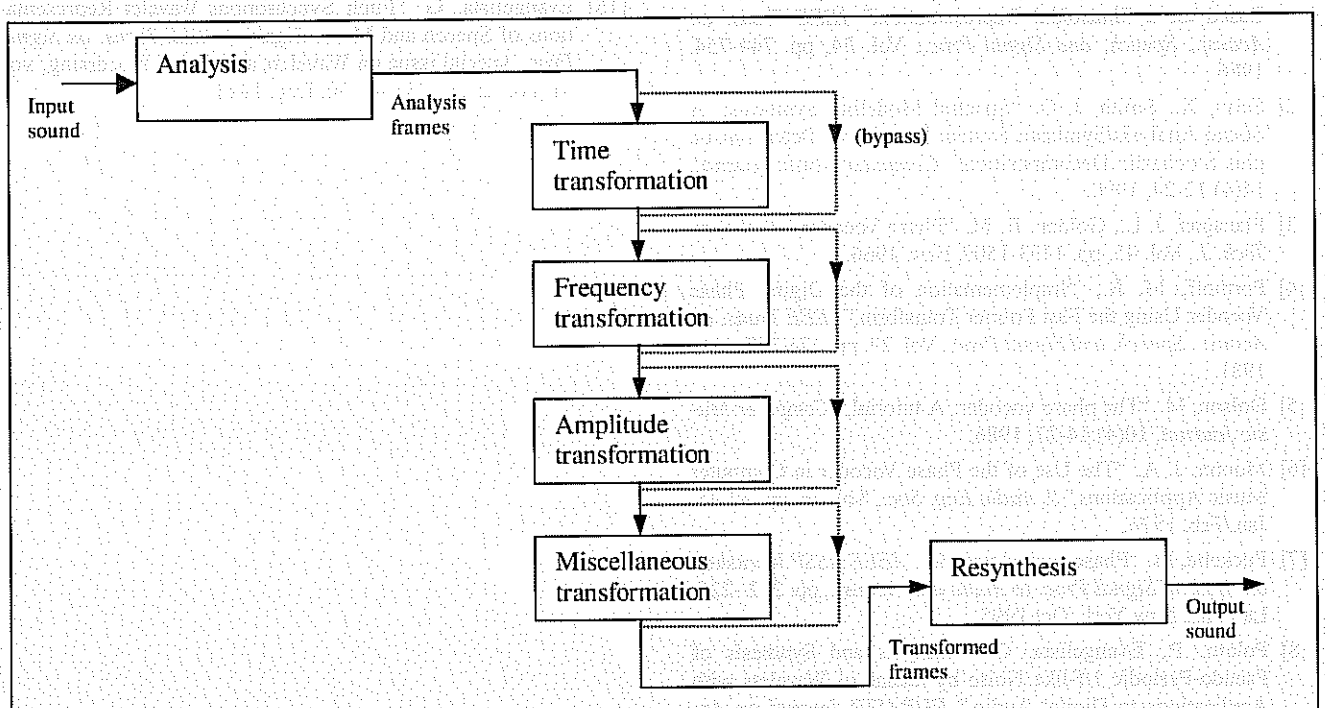


Fig. 1 - Block diagram of signal processing in PVLAB

1.1. Time transformations

In the time domain, there is only one transformation in PVLAB: the modification of the order in which the frames are written to the output. The user can define a *curve {input time vs. output time}*, also defining the output duration. The simplest application of this transformation is time contraction or expansion of the input signal: the user should choose a linear curve y=x and modify the output duration. Interpolation between adjacent frames can be enabled or disabled: if disabled, output frames are identical to input frames, except for the time location.

1.2. Frequency transformations

There are three transformations in PVLAB, belonging to the frequency group: *Remapping*, *Quantize* and *Random add*. *Remapping* is straightforward: the user defines a curve {output frequency vs. input frequency}, and each frequency value contained in the pair {A, f} is modified accordingly. The simplest effect achievable with this technique is pitch shifting: one has simply to define a curve of the form y=ax. PVLAB allow also the user to define two frequency-remapping curves and an interpolation curve, so that the spectral deformation can be varied in time. The *Quantize* transformation is a particular case of remapping: the frequencies of the output can assume only discrete values, calculated with the formula

$$f_n = f_0 \, r^n \quad \text{with } n = \ldots -2,-1,0,1,2,\ldots \qquad (1)$$

where f_0 is the base frequency and r is the quantization ratio.

For example, if base frequency is 27.5 Hz and quantization ratio is 1.059463, the quantization grid corresponds to the frequencies of the temperated scale. For each {A, f} pair, the new frequency value equals the grid value nearest to the old value. Finally, in the *Random add* transformation, a random value is added to the original frequency value: the maximum value added is 4 times the frequency resolution, given by (sampling rate)/(n. of FFT points). The user can set the amount of randomness on a frequency basis, defining a curve {random amount vs. frequency}, or two curves and an interpolation curve.

1.3. Amplitude transformations

The amplitude transformations group contains three items: *Transfer function*, *Threshold*, *Inverse threshold*. The *Transfer function* is the classical curve {amplitude coefficient vs. frequency}, that allow the user to perform very precise filtering. Filtering can also be dynamic, interpolating between two filtering curves. *Threshold* and *Inverse threshold* operations are trivial: the output values of the amplitude are equal to the input values if these are above (or below, respectively) the threshold value, otherwise they are set to zero. The *Threshold* transformation can be used for noise reduction, or to extract the strongest harmonics of an instrumental sound, for example. In PVLAB, the threshold value is specified as a fraction (from 0.0 to 1.0) of the maximum RMS value of the input signal.

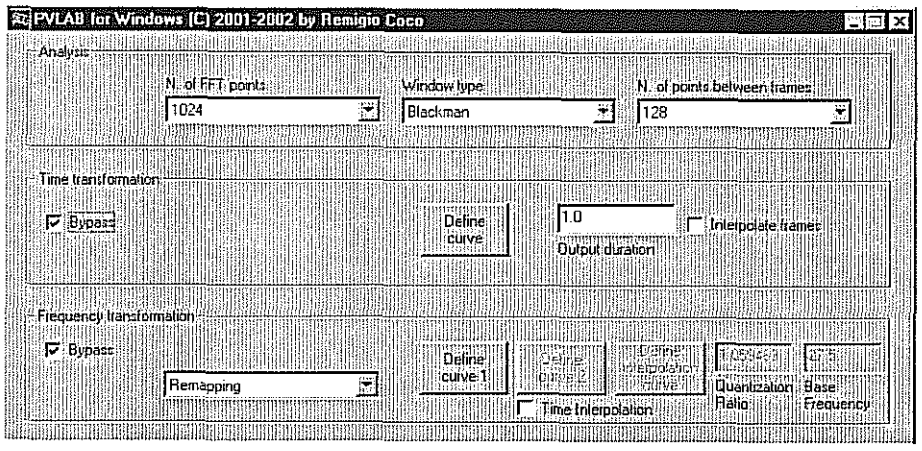


Fig. 2 Screenshot of the upper part of the main window

1.4. Miscellaneous transformations

This group of transformations include the following: *Delay vs. Frequency*, *Stereo-ization*, *Morphing (interpolation)*, *Morphing (cross-synthesis)*, *Morphing (density)*. In the first tranformation, the user can define a curve {delay vs. frequency}, so that each pair {A, f} of the output signal is delayed in comparison with the corresponding input pair; the amount of delay varies with the frequency (hence the name of the transformation). The user has to define the normalized {delay vs. frequency} curve (or two curves and an interpolation curve), and also the maximum delay (in seconds). The effect of this transformation is a sort of "spectral arpeggio", in which the partials of the input signal are not synchronized anymore, but they are presented in a different order. The *Stereo-ization* transformation generates a stereo output signal, based on a curve {Left-Right balance vs. frequency}. The amount of signal passed to each stereo channel varies with the frequency, so that the user could split, for example, high frequencies to the left, and low frequencies to the right. By defining two curves and an interpolation curve, one can also create interesting "moving" effects.

The three following transformations are similar, and are all sound morphing techniques. In all these, the user has to define two curves, {amplitude morphing factor vs. time} and {frequency morphing factor vs. time}; of course, he has to supply two audio or analysis files to the software, let's call them A sound and B sound.

The output signal duration is the maximum duration between A and B; the shorter sound is prolonged, repeating the last frame. In the *Morphing (interpolation)* technique, each output value of amplitude and frequency is calculated by linear interpolation between A values and B values, according to the morphing factor defined by the user (0.0 means only A, 1.0 means only B). In the *Morphing (cross-synthesis)* technique, each output value of frequency is obtained by linear interpolation, exactly as in the previous case. Instead, the output amplitude values are obtained "filtering" the A signal with the spectrum of B signal, and the morphing factor is simply a mixing factor between the original signal A and the filtered signal A#B. The cross-synthesis is achieved normalizing the amplitude values of B, dividing each value by the maximum RMS; then, for each frame and each frequency bin, the amplitude values of A are multiplied for the corresponding normalized value of B. Finally, in the *Morphing (density)* transformation, the morphing factor controls how many values of amplitude and frequency (separately) are taken from A and from B. For example, if amplitude morphing factor is 0.4, each frame is made of 40% amplitude values taken from B, and the remaining 60% from A; distribution of the values inside each frame is casual (i.e. the location of A and B values among the N/2+1 values is randomly chosen).

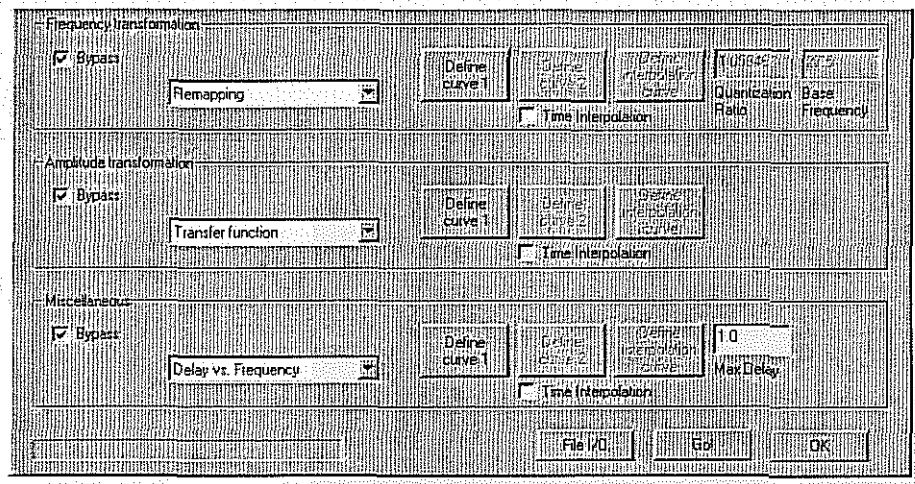


Fig. 3 Screenshot of the lower part of the main window

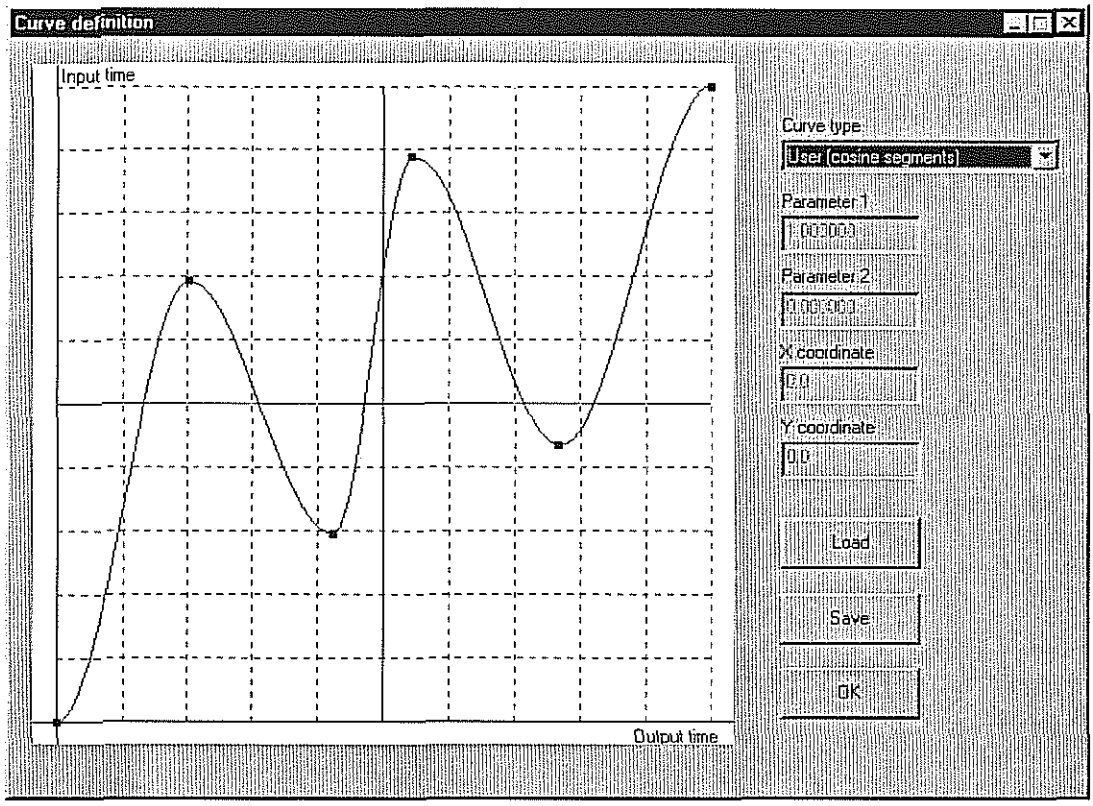


Fig. 4 - Screenshot of the curve definition window

2. CONCLUSIONS

The PVLAB program is useful in a huge range of audio processing needs. Its main features are speed of execution, in comparison with general purpose software like Csound, and the capability to define curves with an intuitive and easy-to-use GUI.

3. REFERENCES

[1] Dodge, C. and Jerse, T.A., "Computer Music. Synthesis, Composition and Performance", Schirmer Books, New York, 1985.

[2] Dolson, M., "The phase vocoder: a tutorial", Computer Music Journal Vol. 10, N. 4: pp. 14-27, 1986.

[3] Wishart, T., "The composition of Vox-5", Computer Music Journal Vol. 12, N. 4: pp. 21-27, 1988.

[4] Dobson, R.,"The operation of phase vocoder. A non-mathematical introduction to the Fast Fourier Transform", Composers' Desktop Project (CDP), York, 1993.

[5] Oppenheim, A.V., and Schafer, R.W., "Digital Signal Processing", Prentice Hall Inc., New Jersey, 1988.

A FRAMEWORK FOR SPEECH SYNCHRONIZED ANIMATION OF EMBODIED VIRTUAL AGENTS

MARIO MALCANGI[1], AND RAFFAELE DE TINTIS[2]

[1] *DICO - Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano*
Via Comelico 39, 20135 Milan, Italy
malcangi@dsi.unimi.it

[2] *DSPengineering, Milan, Italy*
rdt@dspeng.com

ABSTRACT

In this paper we describe a system dedicated to embodied agents speech animation. The described framework derives from the speech signal all the necessary information needed to drive 3-D facial models. Using both digital signal processing and soft computing (fuzzy logic and neural networks) techniques, a very flexible and low-cost solution for the extraction of lip and facial-related information, has been implemented. The developed system is robust to background noise. It also performs speaker and language independent recognition. For this reason neural network training operations are not required.

1. INTRODUCTION

The purposed framework accomplishes two main targets. The former was the implementation and test of a phonemic recognition system dedicated to speech animation [1], the latter was the development of a real-time system dedicated to entertainment and mobile visual communication. These novel applications must be supported by systems requiring very compact init operations. As audio can be acquired in any real-life location under different conditions, they must be robust to background noise occurring at any time. Also, calibration and markers settings operations are highly time consuming. As a result no optic or magnetic sensors can be employed.

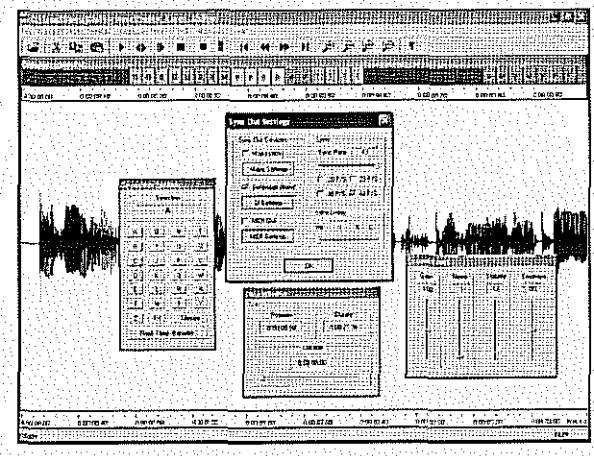


Figure 1: *The LipSync 2.0 main panel*

Real-time speech animation is more effective and supports different accuracy levels also on low cost platforms.

2. SPEECH/SILENCE DECISION vs. USABILITY

In applications targeted for entertainment and visual mobile communication, audio is captured in presence of background noise. For this reason more applicability derives from robustness to diffuse noise. In order to transmit only speech information to the pattern matching subsystem, it is very important to identify the start and end-points of speech utterances. Moreover, it is necessary to model noise representing frames without utterance. Comparing input signal parameters to the noise model, the system must be able to avoid incorrect activations of the speech classification process [2][3].

The model of background noise adopted in our framework is based on the following parameters:

- Q_N = Average noise energy
- Z_N = Average noise zero-crossing rate
- P_N = Average noise spectral power inside the speech frequency range 100-4000 Hz.

These parameters are computed at initialisation time or periodically if the noise is not stationary.

In order to model speech information, the following parameters are computed at run-time:

- $Q_S(t)$ = Time domain signal energy
- $Z_S(t)$ = Zero-crossing rate
- $P_S(t)$ = Spectral power within the range 100-4000 Hz (speech frequency range)
- S(t) = Period of continuous "non-speech" coding collected at current time
- PDim(t) = Period of a valid speech frame
- DP(t) = Ratio between word energy and energy computed at actual frame
- PDist(t) = Period measured between last valid speech frame and previous frame

Speech end-point evaluation is computed by means of fuzzy logic rules evaluating, at each analysis frame, the following input parameters:

- $Input[0] = Q_S(t) / Q_N$
- $Input[1] = Z_S(t) / Z_N$
- $Input[2] = P_S(t) / P_N$

- Input[3] = S(t)
- Input[4] = PDim(t)
- Input[5] = DP(t)
- Input[6] = PDist(t)

3. THE IDENTIFICATION FRAMEWORK

Analysis data represent fuzzy rules input and are computed every 20 ms estimating both time domain and frequency domain information.
At each speech signal frame, the following analysis data represent neural network input:

- Zero-crossing rate
- Total signal energy
- 10 z-transformed LPC coefficients

In the implemented framework, animation data playback can be performed at rates in the range 1-50 frames per second. Low-pass filtering is applied on analysis data in order to reduce speech analysis frame rate at animation frame rate.

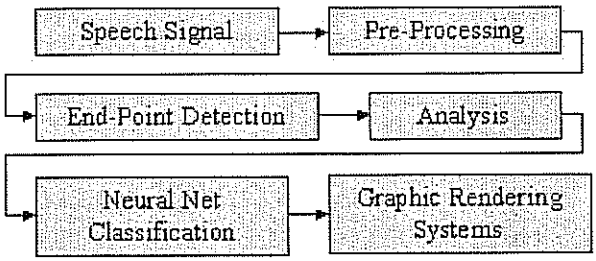


Figure 2: *The automation framework.*

Two different neural networks have been trained: the first targeted for vowel recognition, and the second targeted for consonant recognition.
The use of two small specialized neural networks instead of one proved to occupy less memory and also demonstrated to be better performing.
Response from the voiced/unvoiced decision computed by LPC analysis is used for network selection.
In each analysis frame, the identified phoneme is coded and transmitted to a 2-D/3-D modelling system by MIDI interface (live animation) or by file passing (non real-time mode).
Once the gender of the speaker is specified by the user, the system is able to carry out speaker independent coding on unlimited vocabularies.

4. LIP SYNCHRONIZATION

Speech animation does not require the recognition of all of the different phonemes in a chosen language but only of a fixed set of visemes (a set of visually distinguishable speech posture involving lips, tongue and teeth). In order to accomplish a reduction from phonemes to visemes, it is important to notice that some spoken phonemes have different spectral content mainly due to different tongue positions inside the mouth even if their corresponding lips position is the same or very similar [4].
After reduction, the lips related events recognized by Lip*Sync* are the following:

- Speech signal amplitude (RMS)
- No-speech frames
- Vowels
- Consonants

Vowels are individually recognized as [a], [e], [i/y as in "*youth*"], [o], [u]". Consonants are grouped in 7 classes corresponding to similar lip positions: [(r), (b, m, p), (v, f, th), (d, l, n, t), (s, z), (ch, g, j, sh), (c, k, g)].
The purposed model for speech animation is based on a simplification of the Nitchie lip reading model [5] based on 18 different visemes. Twelve visemes were derivated from the 18 included in the Nitchie original model.
A model for coarticulation was also adopted. We implemented a simplification of the Cohen and Massaro model [6], based on Lofqvist's work on speech gestures [7]. Simplification was necessary in order to support real-time animation.
The Cohen and Massaro model avoids impulsive transitions between phonemes while preserving speech dependent timing characteristics during coarticulation.
At each analysis frame the recognized phoneme is mapped into a set of facial control parameters that will be used during the phoneme segment. The phoneme segment is the time location in the utterance where the phoneme presents his influence. Generally it will be a multiple of the analysis frame. In order to compute actual parameters we must introduce two functions:

Real-Time Dominance:

$$RTD_{sp}(t) = \alpha_{sp} e^{-(\theta sp\ \tau)^{\wedge} c} \quad \tau >= \tau_{SA},\ 0 \text{ otherwise}$$

$$\tau_{SA} = t - t_A$$

RTD_{sp} is the real-time dominance function for parameter P.
α_{sp} is the magnitude of RTD_{sp}.
τ is the time distance from the phoneme activation.
t_A is the phoneme activation time.
θp and C are the control parameters for the exponential function.

The amplitude of the real-time dominance function is controlled by the α_{sp} parameter. The θ_{sp} and C parameters control its shape.

Real-Time Blending

Blending is applied to the dominance functions of each parameter in order to compute final parameter values.
In our model the real-time blending function is defined as follows:

$$RTF_p(t) = \Sigma_n (D_{sp}(t) T_{sp}) / \Sigma_s D_{sp}(t)$$

n = 1,..,N and N is the number of active phoneme segments at frame t. A phoneme segment is active when its dominance function is non zero.

5. FACIAL MODELLING

In normal speech flows, fast high amplitude changes can result in non natural facial movements and not sufficient synchronization quality.
In order to follow amplitude variations we adopted the classic average magnitude function [8]. If N is the window length, the average magnitude function is defined as follows:

$$M(t) = \Sigma_n |s(n)| \qquad n = 0,\ldots,N$$

Energy dynamics used to modulate the amplitude of the lip opening strength, result in more natural movements. Continuous modulation always produces different instances of the same phoneme, in that the mouth shape corresponding to different occurrences of the same phoneme will always be different, similar to what happens in real-life mouth movements.
Speech volume can be used to control different components of facial modelling . In loud intensity frames:

- Facial muscles become stretched
- Eyebrows tend to frown
- Forehead tends to wrinkle
- Nostrils tend to extend

The embodied avatar model should embed speech energy as a variable parameter to control the above aspects so that facial movements can be automated.
Moreover, some emotional aspects can be estimated using speech analysis in order to improve facial modelling.
Anger, for example, presents high average pitch with large fluctuations. Also volume is high on average and presents large fluctuations [4].
In the implemented system, the computed emotional related events are:

- Average Loudness (L_A)
- Loudness Fluctuations Rate, Intensity (L_{FR}, L_{FI})
- Average Pitch (P_A)
- Pitch Fluctuations Rate, Intensity (P_{FR}, P_{FI})
- Pauses Length (S_L)

An estimation scheme for anger and sadness would be as shown in Fig.3:

	L_A	L_{FR}	L_{FI}	P_A	P_{FR}	P_{FI}	S_L
Anger	High	High	High	High	High	High	Low
Sadness	Low	Low	Low	Low	Low	Low	High

Figure 3: *Anger/Sadness Estimation Scheme.*

Both emotions and intonation produce variations of pitch. An accurate pitch estimation however is not necessary and less computational intensive related parameters can be used. Average zero-crossing rate and average zero-crossing rate fluctuations are important variables that can be computed.
Zero-crossing rate is a very rough measure of pitch in wideband signals anyway it is always correlated to pitch in voiced frames. Analysis studies [9] show that the mean short-time average value for 10 ms analysis frames is 14 for voiced phonemes and 49 for unvoiced phonemes.
It is possible to track zero-crossing rate only on voiced frames. In our model, voiced frames are assumed to be only those frames with zero-crossing rate lower than 25 and presenting energy values not lower than half the average loudness. This reduces errors due to zero-crossing rate overlapping between voiced and unvoiced distributions.
Low pass filtering is applied in order to provide smooth zero-crossing rate envelope tracking.

6. CONCLUSIONS

The presented framework features the requirement of robustness needed for novel applications in entertainment and mobile visual communication.
Future work will include more intensive feature recognition for better facial modelling and automation. Also prosodic aspects identification must be considered a key feature.
Fine tuning of system performance can be obtained by reductions of the set of coded visemes for applications requiring low computational cost.

7. REFERENCES

[1] M.Malcangi, R. de Tintis, "LipSync: A *Real-Time System for Virtual Characters Lip-Synching*", in Proc. *XIII Colloquium on Musical Informatics,* L'Aquila, Italy, 2000

[2] M.Malcangi, R. de Tintis, "*Real-Time Facial Modelling For Avatars Animation*", in Proc. *III DSP Application Day*, Milan, Italy, 2002

[3] J.C. Junqua, B. Mak, B. Reaves, "*A robust algorithm for word boundary detection in presence of noise*", IEEE Trans. Speech and Audio Processing, Vol. 2, No. 3, July 1994.

[4] J.A. Markowitz, "*The Data of Speech Recognition*" in" "*Using Speech Recognition*", Prentice Hall, 1996

[5] E.B. Nitchie, "*How to Read Lips For Fun and Profit*", Hawthorne Books, New York, 1979.

[6] M. Cohen and D. Massaro, "*Modeling coarticulation in synthetic visual speech*", In N.M. Thalmann editors, *Models and Techniques in Computer Animation*. Springer-Verlag, Tokyo, 1993.

[7] A. Löfquist, "*Speech as audible gestures*", In W.J. Hardcastle and A. Marchal editors, *Speech Production and Speech Modeling*. Kluwer Academic Publishers, Dordrecht, 1990.

[8] L.R. Rabiner, R.W. Schafer, "*Digital Processing of Speech Signals*", Prentice Hall

Inc., Englewood Cliffs, New Jersey, 1978.

[9] Y. Cao, S. Sridharan, M. Moody, "*Voiced/Unvoiced/Silence Classification of Noisy Speech in Real Time Audio Signal Processing*", 5th Australian Regional Convention, April, 1995, Sydney, (AES Preprint N. 4045)

'SOUND IS THE INTERFACE'
Sketches of a Constructivistic Ecosystemic View of Interactive Signal Processing

Agostino Di Scipio

Scuola di Musica Elettronica, Conservatorio di Napoli, Italy
discipio@tin.it

ABSTRACT

This paper takes a systemic perspective on interactive signal processing, and introduces the author's Audible Eco-Systemic Interface project (AESI). Based on a constructivistic view, it discusses cybernetics principles of a bio-ecological kind (energy exchange, structural closure, organizational openess, system/environment coupling) in the design of signal processing interfaces, with implications for actual musical work (not only live performance situations or studio work, but also sound installations).

Crucial to the subject is an understanding of 'interaction' in terms of a network of interdependencies among system components, and in terms of the system dynamics and its 'structural coupling' to the external environment. The paper illustrates the meaning of these notions as actually implementated in the current release of the AESI (created with Kyma5.2).

1. INTRODUCTION

Talk of 'interactivity' is today common, ubiquitous. And often meaningless, too. The history of the 'interactive arts' and their paradigms [1] is documented in an overwhelming body of literature.

It is not the goal of this paper to overview existing work in the area of 'interactive music systems'. It rather focuses on the paradigm of 'interaction' inherent in most efforts in this research area, and particularly on interactive signal processing interfaces. What kind of 'systems' are musical systems called 'interactive'? I try to answer adopting a particular system-theory view – a *radical constructivistic* view [2], of a kind that emerged from efforts in the cybernetics of living systems [3], including social systems and 'ecosystems' [4]. This requires a reformulation of the notion and function of what is meant by 'interaction'. To illustrate the argument, I describe the design philosophy developed in some recent personal efforts, resulting in the *Audible Eco-Systemic Interface* project, with the aim to implement a kind of 'sonorous ecosystem'.

2. WHAT KIND OF SYSTEMS ARE 'INTERACTIVE MUSIC SYSTEMS'?

Typical interactive music systems can be viewed as dedicated computational tools capable of reacting in some way upon changes they detect in their 'external conditions', namely in the initial input and the run-time control data. Such data are usually set and adjusted by some agency – a performer, or group of performers (could be a composer, too, either working in the studio or experimenting on stage, improvising) – using some control device (mechanical or visual interface).

Therefore, *the agent's operations, as reflected in the control data, implement the system external conditions and all changes therein.*

The agent's operations upon the control data are turned by the computer (according to a variety of digital signal processing techniques and/or program routines operating at a higher, more symbolic level) into sound events, or cause some transformation of input sound material. Upon hearing, the agent adjusts the data by means of the available controls, eventually 'playing' the system almost as if it were a new kind of music instrument (however, the 'instrument' metaphor cannot really be generalized). A variety of known interactive performance situations have been described in available publications on these matters, ranging from the 'solo instrument' set-up, to the 'duo' and larger 'ensembles', where many performers and/or computer systems are interconnected and play together (see [5] for a summary).

When it comes to live digital signal processing interfaces, it becomes clear that the system design itself, and particularly the interactions mediated by the user interface (interdependencies among control variables), become the very matter of composition [6] and can no longer be separated by the internal development of sound. Interestingly, in recent work we can observe that not only the visual interfaces, but also the design of mechanical devices can be crucial to the extent that it directly affects the details in the output sound (a peculiar example is [7]).

Notwithstanding the sheer variety of devices and protocols currently available, all 'interactive music systems' – including developments specific to the internet – share a basic design principle, namely a linear communication flow (figure 1).

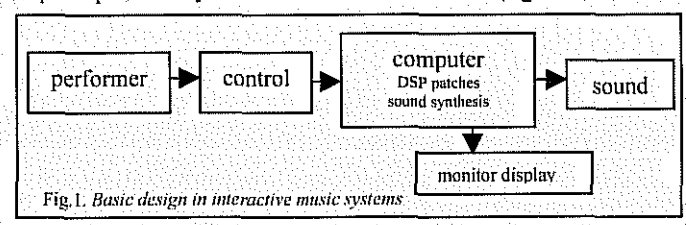


Fig.1. *Basic design in interactive music systems*

Usually, this design assumes an implicit recursive element, namely a loop between the 'sound' and 'performer': the computer output somehow affects the performer's next action (reaction), which in turn will eventually affect the computer system in some way (figure 2).

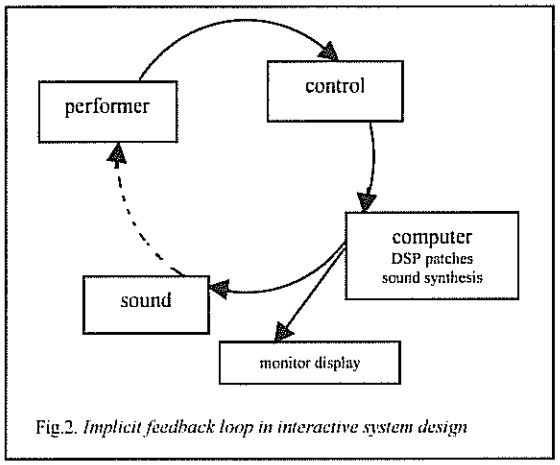


Fig.2. *Implicit feedback loop in interactive system design*

That recursion is a potential source of dyamical behaviour, of noise and unpredictable developments. Yet the basic design remains a linear one, i.e. from the performer's action to the computer reaction. The performer is first the initiator agent of the computer's reaction, and only secondly, and optionally, might become the very *locus* of feedback, injecting some features of nonlinearity into the overall system ('system', then, includes the performer). Most 'live electronics' music can be referred to that design, which also returns in most computer music endeavours, even in very recent discussions [8].

2.1 *Where* and *when* are systems 'interactive?

In the above mentioned approach, 'interaction' means that the computer change its internal state depending on the performer's action, and that the latter is potentially itself oriented in his/her action by the computer output (dashed line in figure 2). There – in the perfomer's mind and ears – lies the only source of unforeseen developments, of dynamical behaviour. Of course one could introduce all sorts of complex transfer functions in the mapping of data from one domain to another (similar to many musical instruments with their complex mapping of gesture into sound) but that is an option which is not in essential in the ontology reflecting 'interaction': *agent acts, system re-acts.*

Put it in another way, in existing 'interactive music systems', and particularly in their DSP components, *the system is not itself able to directly cause any change or adjustment in the 'external conditions' set to its own process*, it does not generate or transform the control data it needs to change its own internal state. The model prompted includes no recursion.

Also, 'interaction' is not usually referred to the mechanisms implemented within the computer: the array of available generative and/or transformative methods usually consists of separate functions, i.e. they are conceived independent of one another, and function accordingly. The agent selects the particular function(s) and process(es) active at any given time, and the latters' output data are simply summed together. As an example, the sudden occurrence of, say, too large a mass of notes or sound grains or other musical atom units would not automatically induce a decrease in amplitude (a perceptually correlate dimension of density): such an adjustment remains a chance left to the performer. No interdependency among processes is usually implemented within the 'interactive' system.[1] 'Interaction' (either between system and external conditions, or between any two processes within computer) is rarely understood and implemented for what it is in real living systems (either human or not): a by-product of implemented lower-level interdependencies among system components. Interfaces usually do not allow to create a communication between DSP processes, but only to independently handle their parameters, in the form of runtime variables.

3. FROM 'INTERACTIVE COMPOSING' TO 'COMPOSING THE INTERACTION'

In a different approach, the aim would be first of all to create a *dynamical system*, possibly exhibiting an adaptive behaviour to the surroundings ('external conditions'). It should be able not only to 'hear' what happens 'out there' (an 'observing' system, capable of tracking down relevant features of the external world, not demanding this from an external agency), but also to become in the end a *self-observing* system [10], ultimately using information on the external conditions to orient its own internal sequence of system states (*self-organization*).

This idea is motivated with a notion of 'interaction' as a means, not an end in itself, i.e. a prerequisite for something like a 'system' to emerge. Therefore, 'interaction' should be the object of design (hence, composition), and more precisely the by-product of carefully planned-out interdependencies among system components.[2] The overall system behaviour (dynamics) should be born of those interactions, in turn born of lower-level interconnections.

This is a move from 'interactive performance', or 'interactive composing' (as in the pioneering work of Joel Chadabe and others) to 'performing the interaction', or 'composing the interaction'. In the latter approach, one designs, implements and maintains a network system whose emergent behaviour in sound one calls music. When a such a system enters a non-destructive relationship to the surrounding environment – the system's *house*, literally: its οικος – it becomes an *eco-system*.

4. THE *AUDIBLE ECO-SYSTEMIC INTERFACE*

The *Audible Eco-Systemic Interface* (AESI) reflects a self-feeding loop design (figure 3). A chain of causes and effects is established without any intervention on a human performer's part (but that doesn't mean that a perfomer cannot enter the loop[3]): the

[1] The observation also applies to 'interactive music systems' capable of 'listening' to, say, an instrumentalist playing thereby making decisions based on features tracked down ('machine listening' [9]). In 'score following', runtime control variables are updated following the successful or unsuccessful matching of an instrumental performance against a stored event list (score): therefore here, too, decisions are 'dynamically' made by means of a predetermined knowledge-base (score representation).

[2] A similar approch is in George Lewis' interactive improvisation programs and musical pieces [11].

[3] Note that 'interaction' here should not be implicitly understood as 'man / machine interaction'. The AESI project leans on 'ambience / machine interaction'. Clearly, the ambience may host one or many performers. A similar annotation should be made for 'interface', as used later in the paper.

computer emits some initial sound, that is heard through the loudspeakers and also fed back into the computer, by microphones scattered around the performance place; the computer analyzes the microphone signals, and the features accordingly extracted are used to generate low rate control signals and drive the system's internal process (transformations or synthesis of sound material); also, the computer matches the microphone signals against the original synthetic or sampled signals, and the difference-signals (the difference values between original and fed-back sound), reflecting the contribution of the room resonances to the sound as actually heard, are used to adapt the computer performance to the room response: the AESI variables are in a constant flux of changes depending on the resonances in the environment, as elicited by the sound event initally emitted.

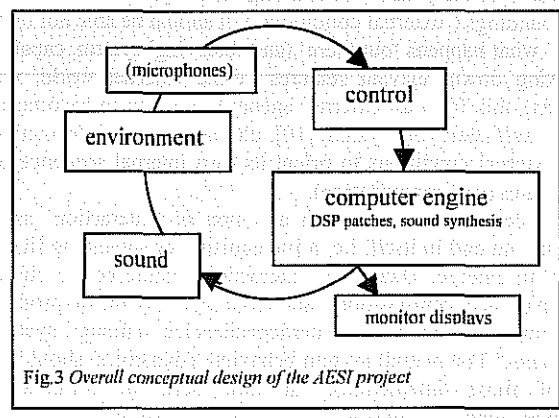


Fig.3 *Overall conceptual design of the AESI project*

The sound (room resonances) affects the DSP control variables, and the latter affects the sound. By means of this recursive coupling, the transformative or generative functions implemented in the computer become *iterated functions*. If they include some nonlinear mapping, they become *nonlinear iterated functions*, a source of dynamical behaviour inherent to the system, can subject to be modelled in mathematical terms [12]. Different from more usual interactive systems, this recursion is structural to the overall design, not optional: it is independent of human performers, although clearly human performers can enter the process cycle and become a part of the overall system (yet he/she would not be the only source of dynamical behaviour).

Observe that, because here the fed back sounds become source signals for the generation control signals, the feedback loop in this design is not in the audio frequency domain, but confined in the low frequency domain (control rate). The precise rate depends on the time scales in the feature-extraction processes, the integration time of observing functions, and other time-related variables in the mapping of control data (time dilation of control signals, time maps of input data, etc). The role of time variables is really crucial in the AESI project, as their side-effects contribute to install an element peculiar to dynamical systems such as ecosystems: namely the fact that dramatic long-term consequences could follow from what currently appears as marginal detail.

4.1 Ecosystemic dynamics

Variables. In the AESI project, system variables are represented by the control signals generated with feature-extraction processes. Sonic features so far taken into account include: amplitude, density of events (rate of onset transients), transient delay between microphones (resulting in perceptual precedence effects), and psychoacoustical variables such as 'brilliance' and other statistical spectral properties. Clearly, pitch and other frequency-related measures could be also considered. Beside, it is good idea to create many control signals out of a single feature-extraction process.[4]

Functions. Considered in the AESI are the following functions (described in their systemic meaning, because the particular math expression can change depending on context): *compensation* (e.g. decreasing amplitude, either that of the input or the output to some process, when the overall density increases); *following* (running after and finally matching some variable, with some delay; e.g. the *hysteresis* effect, implemented as *latency* filter circuits); *redundancy* (supporting a given predominant feature, rather than contrasting it; e.g. automatically adjusting the integration time of variable followers in order to make them identical with, or close to, the rate of repeated events – as used, for example, in dynamical tempo tracking and other adaptive filters); *concurrency* (giving way to an alternative feature competing with the predominant one; e.g. introducing high-frequency spectral energy when low-frequency energy is predominant in the room).

The overall goal of these functions is to implement and regulate the interdependencies among variables, depending on external conditions. By way of that, they make the internal routines and DSP modules to interact in a meaningful way. Finally, by means of those emergent interactions, they indirectly implement the system dynamics.

Depending on the long-term effects, a function contributes to install the system overall orientation. One can consider at least the follwing two

Competing orientation criteria. Include *omeostasis* (the system's tendency to keep and mantain a constant or recurrent behaviour, which determines its 'identitiy' to an external eye or ear); and *omeoresis* (the opposite tendency, to introduce and follow a more varied behaviour, which determine the system's ambiguity to an external eye or ear).

These contrasting criteria are implemented in terms of the above functions, and should always counterbalance one another. A rich system dynamics is typically achieved only as a result of their competition.

4.2 Some observations

In the AESI project, the overall system/environment coupling function is not explicitly formulated: it emerges as a property of the system dynamical self-organization (its ability to change state and to change the interdependencies among system components). The particular self-organizing behaviour in turn depends not only on the mapping of the control signals onto the parameter space (for the given sound synthesis or processing algorithms adopted), and the time variables (time scales at which the input sound stream is analyzed in the feature-extraction process, and time

[4] *The real-time generation of control signals is an art in itself, and opens* questions that are worth of a specific discussion.

maps of control signals). It also heavily depends on the sound material being used to excite the environment resonances, and clearly on the specific room acoustics (material and geometrical characteristics of the system environment).

The AESI runtime process unfolds as it 'learns' about the environment. Its coupling to the latter is indirect: it happens in the medium of sound. In a way that is *not* at all metaphorical, *by way of doing something to the environment* (sending sounds to it) it actually learns about itself as a system and develops its own internal organization.

Finally, note that all exchanges between the AESI and its οικος take place in the medium of sound: *sound is the interface* (all processes and equipement involved, e.g. microphones, are vehicles or transformers of sound signals).[5] As finally perceived by the listeners, sound bears traces of the structural coupling it is born of. One can speak of *audible interfaces*, i.e. interfaces whose process (the mediation between system and environment) *is* or *can* be something actually heard.

5. PRELIMINARY CONCLUSIONS

A thorough technical description of the AESI implementation is out of the scope of the present paper (it would extend to the discussion of the particular DSP methods adopted, the feature-extraction algorithms themselves, etc. – not to mention the type of microphones, their placement, relative distance and orientation, the size and geometry of the room, placement of loudspeakers, etc.). Furthermore, many details have been so far only planned out but not fully implemented yet. The purpose of the present paper was to introduce the perspective and the basic motivations.

In this approach the notion of 'interaction' is reformulated (and implemented) in terms of dynamical interdependencies among system components. The idea that a computer 'reacts' to a performer's gesture is replaced with a 'structural coupling' of system and environment. The system *acts upon* the environment, observes the latter's reactions, and then reacts based on the environment's response. Also, by tracking down its own previous internal states, and previous interactions with the ambience, it developes based on its own history, i.e. 'cognizant' of the past (system's *memory*, long term effects, etc).

Reflecting a radical constructivistic epistemology [15], the *Audible Eco-Systemic Interface* represents a 'structurally closed' but 'organizationally open' system. The 'closure' is meant to preserve the system identity (that's the case with typical 'interactive systems', too). The 'openess' reflects the system's ability to exchange energy (sound) with the environment [13] and to determine its own internal causal sequence of states, thanks to a rich dynamics emerging in the interactions among system components (this is not the case with existing 'interactive music systems').

Addendum

The current AESI implementation was created with SymbolicSound's KYMA5.2 (using the CAPYBARA320 as DSP engine). In a preliminary musical work composed with and for the AESI, titled *Ecosistemico udibile n.1* (2002), short sound impulses were used as the only raw sound material introduced into the feedback loop of the AESI. That was precisely in order to experimentally test the system's *impulse response*. The pulse material had been created with the PULSARGENERATOR program [14] during the author's composer residency at CCMIX (Centre de Creation Musicale Iannis Xenakis), in Paris, April 2002.[6]

References

[1] Dinkla, S., "The History of Interface in Interactive Art". *ISEA Proc.* 1994.

[2] von Glasersfeld, E., "The Roots of Constructivism". Paper presented at the Scientific Reasoning Research Institute, 1999.

[3] Maturana, H. and Varela, F., *Autopoiesis. The realization of the living*. Dordrecht, 1980.

[4] Morin, E., *La méthode. La nature de la nature*, Paris, 1977.

[5] Rowe, R., *Interactive Music Systems*, MIT Press, 1993

[6] Hamman, M., "From Symbol to Semiotic: Representation, Signification and the Composition of Music Interaction," *Journal of New Music Research*, 28(2), 1999.

[7] Smith, T. and Smith, J.O., "Creating Sustained Tones with the Cicada's Rapid Sequential Buckling Mechanism", *NIME Proc.*, 2002

[8] Schnell, N., and Battier, M., "Introducing Composed Instruments. Technical and musicological implications", *NIME Proc.*, 2002.

[9] Rowe, R., *Machine musicianship*, MIT Press, 2001

[10] von Foerster, H., "On self-organizing systems and their environment." In (C.Yovits, ed.) *Self-organizing systems*, New York, 1960. Also: von Foerster, H., *Sistemi che osservano* (edited by M.Ceruti and U.Telfener), 1987.

[11] Lewis, G., "Interacting with the Latter-Day Musical Automaton", *Contemporary Music Review*, 18(3), 1999.

[12] Collet, P., and Eckmann, J.-P., *Iterated Maps on the Interval as Dynamical Systems*, Boston, 1980.

[13] von Bertanlanffy, L., *General System Theory*, NY, 1968.

[14] Roads, C., "Sound Composition with Pulsars", *Journal of the Acoustical Engineering Society*, 49(3), 2001.

[15] see Riegler, A. www.univie.ac.at/constructivism/, 2000.

[5] In a similar vein, [11] writes: "There is no built-in hierarchy of human leader/computer follower: no 'veto' buttons, pedals or cues. All communication between the system [he means the computer, which is just a system component, though] and the improvisor takes place sonically. [Such] a performance... is in a very real sense the result of a negotiation..." (p.104). While in Lewis' approach the 'improvisor' is the only 'ambience' set to the computer, the only source of external conditions (noise), in the AESI project the human component is another possible source of information inhabiting the shared ambience.

[6] This composition was premiered in a concert at Keele University, Stoke-on-Trent, UK (October 2002). Kurt Hebel took care of the set-up and supervised the system performance. The italian prèmiere will take place in the 14th CIM concerts, in Florence, with the supervision of Alvise Vidolin.

ISSUES FOR SYNTHESISING MUSICAL INSTRUMENTS USING SIGNAL AND PHYSICAL SYNTHESIS MODELS

Donagh J. T. O'Shea, Niall J. L. Griffith

Department of Computer Science and Information Systems
University of Limerick
Limerick
Ireland
donagh.oshea@ul.ie, niall.griffith@ul.ie

ABSTRACT

This paper discusses the two most widely used general methods of sound synthesis for the creation of instrumental sounds and composition.

1 INTRODUCTION

The use of signal generation to create new musical instruments has contributed significantly to both our understanding of how musical timbres are perceived and to music composition. The development of new musical instruments and the synthetic mimicry of existing instruments have involved three collaborating and complementary aspects. These are the development of a better understanding of a) the aspects of sound that the human auditory system exploits in processing sound and which contribute to our identification of instruments; b) the properties of the physical systems that constitute an instrument and its interactions with its environment including feedback between the player and instrument; c) techniques for spectral synthesis. In the development of more realistic synthesis of existing instruments, psychological models of perceptual descriptors have augmented the critical listening that has been applied to the development of physical and spectral based models. When synthesising an existing instrument the problem is that of mimicking the characteristics of a known producer. However, when the aim is to create a novel instrument this is not the case. The nature of synthesis is such that the space of instrumental timbres is essentially infinite. One problem confronting the instrument designer is how to make an instrument that is consistent over different playing dimensions such as pitch, mode and intensity when there is no physical system present? The different forms of synthesis have characteristics that allow for rather different solutions to this problem.

2 Characteristics of Physical and Spectral Models

Physical models use mathematical approximations of the excitation and resonance patterns in a physical system such as a musical instrument. The input parameters reflect physical parameters appropriate to the system. The output is a digital signal that reflects the output state of the model in response to the input state of the parameters to the model.

Signal models are constructed from the analysis of signal content – often a spectral analysis of the acoustic result of a state of vibration in a system, physical, synthetic or complex. Even when a new and unheard instrument is envisaged, an internal mental model of the projected instrument's productivity is used as a guide to implementation. The input to a signal model must specify the parameters governing the actual spectral information in the signal, according to the synthesis method used (additive, granular, FM, etc.).

2.1 Strengths and Weaknesses of Physical and Spectral Models

Physical models have the advantage, in modelling instruments, that the result of all the available physical parameters contained in the model is output as a function of the model. Variations in dynamics, articulation and range are, once the model has been defined, intrinsic to it. The consistency of the resulting instrument is derived entirely from the model. Physical models also have the advantage that the controller mapping is a one-to-one function, even if this gives rise to a control surface problem, as might be the case with 37 parameters governing a model of the human voice [1], seeing as a performer may not be comfortable with so much internal wiring.

Signal models, on the other hand, must provide explicitly for variations in the instrument's performance, including pitch, mode and dynamics. This impinges not only on the consistency of the instrument as it is played, but also on how it is played. One assumption is that some form of physical interface device initiates and controls production and involves a strategy for mapping between the manipulation of the device and the parameters of the generative signal model. The mapping strategy may be configured in any way including arbitrarily. This may cause control problems if the mapping is in some way counter-intuitive, too complex, or too simple [2]. This control difficulty inevitably interacts with the main conceptual

advantage of using a signal model; that the control space can be extended to allow playability in any direction on any dimension once a production mechanism has been established that «feels like an instrument». In other words, the total feedback from the instrument must provide enough stimulation to encourage further engagement with it and enough stability to prevent frustration arising from inability to repeat actions consistently.
The data on which the model is based is primarily perceptual. In theory this allows any conceivable instrument to be created. However, the complexity of the perceptual system in comparison to a vibrating acoustic system means that arriving at a rich, coherent signal model is a significantly greater challenge, even if the computation at run time is roughly equivalent.
The feedback into the design process has both formal and informal aspects. It is informal in as much as it is based on the designer's iteration through the loop of listening to the output sound and adjusting elements of the instrument according to observations from listening. The formal aspects are based on measures derived from experiments. Such measures are correlated with attributes of the signal, for example, between spectral centroid and brightness. However, there remains the problem of reverse engineering such abstract and non-unique properties in consistent ways.

3 Psychological Measures and their Interaction with Synthesis Parameters

The analysis of acoustic signals is a complex task. Classification cues which may be used to categorize and identify sonic events, such as spectral centroid [3], [4], [5] and cepstral coefficient [6], [7], [8] have the disadvantage that these measures are statistical and effectively unrecoverable. The effect of SPL, frequency, duration, and waveform on perceived loudness, pitch, duration and timbre is complex and interdependent. Ignoring contextual considerations, polyphony/multiple simultaneous events, and non-linearities in this relationship, timbre is still a contentious area, apparently processed by the auditory system in separate streams for time and frequency domains. The complex interaction of these domains in the cognition of even single stream stimuli creates ambiguities that have yet to be fully understood, such as the predominance of attack or steady state cues in the tracking and recognition of sounds [9], listening strategies [10], and the relative importance and number of dimensions contained within the conceptualisation of timbre.

4 The Problem of Composition versus Specification

Musical composition involves an understanding of the psychological/musical effect of how a sound is presented. It is the most sophisticated use of timbre that human beings engage in. The musical paradigm allows the simultaneous comprehension of elements that usually cannot be perceived simultaneously. The unique manner in which music allows an understanding, in musical terms, of the complexity of interaction between genre, culture, personality and emotion has so far defied translation into scientific or linguistic terms, by virtue of its extreme complexity. However the musical sense of this complexity is inherently understandable through concepts such as «expression» [11]. «Expression» is a complex phenomenon. In Western Art Music it is mediated through the score and the performers' interpretation while in other Musics it arises through an analogous shared sense of important features. Socially constructed «meaning» is arrived at through mutually understood metaphors suitable for describing how to achieve the desired results, as well as how to interpret those results. The challenge for computer-mediated music is to emulate this high level causality of intention and result in a meaningful and effective way. A significant aspect of this integration lies in how the instrument is played. This suggests that a systematic investigation into how mappings are constructed in the absence of physical coupling will inform how mappings are to be designed. For controllers to allow consistent navigation of potentially infinite timbre space requires synthesis methods that are controllable by intentional input, and control devices that produce intentional output.

4.1 Bonding in the Mapping Concept

Although devices such as the Hyperbow [12] and the vBow [13], for example, go some way towards addressing the control surface and haptics of physical models, controller mapping for spectral models is arbitrary, and does not easily lend itself to strategies that may become standard, conventional or generally accepted beyond keyboard/MIDI style note/velocity type events, given that the flexibility to re-map these inputs is often the most admired feature of tools which use these types of control. Devices such as these do contribute greatly to our understanding of the underlying physical acoustic systems we often use to make music, and the gestural primitives we associate with «expressive» performance. They do not contribute to the expansion of the sonic palette available to the composer, or to our understanding of how acoustic artefacts are interpreted as «expressive».
The growing understanding of how an instrument's sound, its form and control are all closely related emphasises the depth of the problem of how to control parameters that have no physical corollary.
Many pieces require unique controllers, specifically designed for a single performance, installation or tour. This situation is not conducive to encouraging musicians to spend years of practise mastering the subtleties of expressive potential within a control paradigm. On the other hand, systems like Pro Tools, Kyma, Max/MSP, Reason, Csound and others do provide a paradigm within which virtuosity is both possible and valuable. They also impart a sense of being able to 'play' within the environment in a largely unstructured way. Part of the reason for this is that

these are stable systems that have great scope for re-use in different contexts. They continue to be useful over the course of many years, allowing users to refine their skills. This would suggest that higher level organizational strategies may be more relevant to the production of music beyond traditional techniques and to the utilization of the power of modern computation for the realization new music. Computers were not designed to be musical instruments. Their power lies in their ability to provide frameworks within which complex processes can be facilitated and complex low and intermediary level calculations can be accomplished quickly and effortlessly, allowing the user to concentrate on higher level organization processes.

The computer as a tool requires definitive organization at a low atomistic level and human created algorithmic specification. The understanding of human created methods of interaction with music must be refined to account for strategies of specification. Sound Synthesis has yet to adequately address two more important issues: a) how to derive parameters for synthesis accurately and systematically, and b) how to label the control 'handles' in a way that integrates them with the synthesis parameters, beyond ASDR, LFO, etc.

Resynthesis algorithms have undergone considerable development and improvement in recent years [14], but they have not reached a stage where they can overcome the limitations inherent in Fourier Transform mathematics as applied to real signals, and certainly do not allow the systematic reconstruction of instrumental sounds from frequency or time domain input. Changes over range, intensity and articulation, which come free with physical models, must be explicitly modelled in spectral synthesis.

Some spectral models, such as SMS [15] and the Timbre Engine [16] have been successful by separating noisy and harmonic portions of sounds. The task of linking such processes to a systematic classification system based on psychologically based categorization schemes is a highly complex one.

Some aspects of expression can be idiosyncratic, personal and rooted in culture and genre conventions [17], [18]. The freedom to manipulate timing, articulation and dynamic expressive control can and should be left to the performer, whether real-time or pre-orchestrated. These aspects of musical expression are well established, documented and defined, although little understood in formal terms of psychological affect, mainly because of personality, psychological state and cultural complications [11]. The control of the micro-structure of timbre within these meso-level events, however, is a problem that is very poorly understood, although some strides are beginning to be made in this direction [19]. This is partly due to the brevity of its history – until the arrival of digital computers, there were no tools to look at the micro-level timbral events, except intuition and empiricism. Since the early nineteen seventies, many of the previous assumptions about timbre have been undermined by systematic studies [20], [21], [22].

5 Ideal Principles for Mediating Control of Synthesized Instruments

Arguably one of the most important lessons to be derived from the history of synthetic musical instrument design is the necessity of defining a high level conceptual space for the control of timbre. This control space to be effective has to a) organise and control timbre as a hierarchically embedded element of a musical composition, subordinate to pitch, dynamic, and articulation controls, b) be flexible enough to provide automatic parameter manipulation via metaphoric, visual, or other means of specification, and c) be rooted in perceptual judgements of sound, categories of sounds, and hierarchically nested constraints. If the possibility of specifying the type of sound to the depth of detail desired could be realised this would allow musicians to develop their own systems of specification or naming categories. A learning mechanism that can, for example achieve timbral transposition, and account for dynamically varying feature vectors is a framework within which the power of synthesis could be used efficiently. The need to provide instruments that people can play without painstakingly learning syntax and parameter control as is available in Csound, or Moog type methods persists. Some admirable developments in the field of novel controllers may be beginning to allow people to become computer instrumentalists without being instrument builders.

6 Conclusion

In conclusion, synthetic musical instruments present the instrument designer with a threefold challenge. The idea of a high level conceptual control space stands as a short hand for three aspects of the notion of an 'instrument'. Firstly, the instrument has to be consistent over its range and modes of playing. This is akin to the consistency found in the identity of a natural acoustic instrument. Secondly, the instrument has to be playable. This involves notions of plasticity and flexibility but more inportantly, and related to its consistency it has to engage the player in its productivty. Thirdly, in order to be able to accomodate these it has to support hierarchical constraints and adapt to these. Just as a traditional composer tries to integrate diverse elements of the musical fabric on all levels - phrase, note, timbre, articulation, etc. are interactive and condition each other – the synthetic system must account for these interdependencies.This is equivalent to specifying how in a particular passage the instrument should be played to sound right. It is only when these three qualities are realised that the player will be able to use it to create music that follows its own logic rather than the demands of the instrument's mediating control strategy and build the kind of intra-musical relationships that are the essence of the dynamically changing context that is musical process. In this circumstance we will be able to claim that we have created an instrument that is truly synthetic and truly an instrument.

7 References

[1] Cook, P. R., «SPASM, a Real-time Vocal Tract Physical Model Controller; and Singer, the Companion Software Synthesis System», Computer Music Journal, 17 :1, pp. 30-44, 1992.

[2] Hunt, A., Wanderley, M. M. and Paradis, M., «The importance of parameter mapping in electronic instrument design», Proc. NIME-02, Dublin, Ireland, pp. 149-154, 2002.

[3] Fujinaga, I., «Machine recognition of timbre using steady-state tone of musical instruments», Proc. ICMC98, University of Michigan, Ann-Arbor, 1998.

[4] Beauchamp, J. W. and Lakatos, S., «New spectro-temporal measures of musical instrument sounds used for a study of timbral similarity of rise-time and centroid-normalized musical sounds», Proc. 7th Int. Conf. On Music Perception and Cognition, Sydney, Australia, pp. 592-595, 2002.

[5] Kendall, R. A., «Musical timbre beyond a single note, II: Interactions of pitch chroma and spectral centroid», Proc. 7th Int. Conf. On Music Perception and Cognition, Sydney, Australia, pp. 596-599, 2002.

[6] Eronen, A. and Klapuri, A., «Musical instrument recognition using Cepstral coefficients and temporal features», Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000, Istanbul, Turkey, pp. 753-756, 2000.

[7] Gu, L. and Rose, K., «Split-band Perceptual Harmonic Cepstral Coefficients as Acoustic Features for Speech Recognition», Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2001, Salt Lake City, Utah, pp. 125-128, 2001.

[8] McGrath, D., «Design of a Musical Instrument Classifier System based on Mel Scaled Cepstral Coefficient Supervectors and a Supervised Two-layer Feedforward Neural Network», Proc. 13th AICS, Limerick, Ireland, pp. 215-220, 2002.

[9] Sandell, G.J., "SHARC Timbre Database." Electronic publication: World Wide Web URL http://sparky.ls.luc.edu/sharc/, 1994.

[10] Bregman, Auditory Scene Analysis, MIT Press, Cambridge, Massachusetts, 1990.

[11] Juslin, P. N., Friberg, A. and Bresin, R., «Toward a computational model of expression in music performance: the GERM model», Musicae Scientae, Special Issue 2001-2002, pp 63-122, 2002.

[12] Young, D., «The Hyperbow Controller: Real-time Dynamics Measurement of Violin Performance», Proc. NIME-02, Dublin, Ireland, pp. 65-70, 2002.

[13] Nichols, C., «The vBow: Development of a Virtual Violin Bow Haptic Human-Computer Interface», Proc. NIME-02, Dublin, Ireland, pp. 29-32, 2002.

[14] Ellis, D. P. W., Prediction-driven computational auditory scene analysis, PhD Dissertation, MIT, 1996.

[15] Serra, X., Bonada, J., Herrera, P. and Loureiro, R., «Integrating complementary spectral models in the design of a musical synthesizer», Proc. ICMC97, Thessaloniki, Greece, 1997.

[16] Martenakis, G. and Jensen, K., «The Timbre Engine - Progress Report», Proc. MOSART IHP Network, Barcelona, Spain, 2001.

[17] Gabrielsson, A. and Juslin, P.N., «Emotional Expression in Music Performance: Between the Performer's intention and the Listener's Experience», Psychology of Music, 24, pp. 68-91, 1996.

[18] Matthews, M. V., «Harmony and nonharmonic partials», J. Acoust. Soc. Am., 68(5), 1980.

[19] Roads, C., Microsound, MIT Press, Cambridge, Massachusetts, 2002.

[20] Grey, J. M., «Multidimensional perceptual scaling of musical timbres», J. Acoust. Soc. Am., 61(5), 1977.

[21] Ehresman, D. and Wessel, D., Percepttion of Timbral Analogies, IRCAM Report 13/78, 1978.

[22] von Bismark, G., «Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes», Acustica, Vol.30, pp. 146-159, 1974.

A TWO-LEVEL METHOD TO CONTROL GRANULAR SYNTHESIS

Andrea Valle, Vincenzo Lombardo

MultiLab e Dipartimento di Informatica - Università di Torino
andrea.valle@unito.it, vincenzo@di.unito.it

ABSTRACT

This paper presents a formal system for the algorithmic control of composition based on granular synthesis. The system features two description levels: a low level, that organizes grains into a graph structure, and a high level, that distributes the graphs of the low level in specific locations of a space. A composition is a trajectory in the space, appropriately interpreted to control a number of parameters of physical and musical relevance. The paper is organized as follows: first, we introduce the composition process with granular synthesis and we briefly outline the current approaches to control; second, we describe the formal system in terms of the two levels that compose it; finally, we see how the system can be viewed as a generalization of the note approach and the stochastic approach.

1. THE COMPOSITION WITH GRANULAR SYNTHESIS

Granular synthesis is a general term that encompasses various kinds of synthesis techniques based on a grain representation of sound, i.e. sonic events are built from "elementary sonic elements" of very short duration ([1]; as a general reference, see [2]). Different organization techniques can lead to very different timbral and compositional results (see [1], [2], [3], [4]). So, one of the main questions arising while working with grains is how to move from single grain level (*microstructure*) up to compositional design (macrostructure), possibly passing through note level (ministructure) and rhythm level (mesostructure) (following [5]: 266; see also [2]: 3). We can distinguish two major approaches: the *note* approach and the *stochastic* approach.

In the case of the note approach, the focus is on microstructure as embedded in ministructure: so, ministructure defines the sound objects and granularity defines the timbre of each object (i.e. drum roll, rolled phonemes, flutter-tongue, [6]: 56). Granular synthesis and granulation of existing sound objects are methods to create/transform elements at the "note" level. As in traditional composition, there is a logic gap between sound and structure ([7]). This is the approach implemented in grain-based modules of DSP applications ([8], [1], [4]), and in Csound built-in opcodes ([9]).

More radically, in the stochastic approach, granularity is intended as a compositional feature. Having to work with a pulviscolar matter, composers involved in granular synthesis have often decided to avoid an "instrumental-music approach" to promote textural shaping as a general compositional feature, in order to "unite sound and structure" ([7]: 120). Various stochastic methods and strategies have been used to control grain densities, distribution in frequency spectrum, waveshape in the time course (see the "classic" works by Xenakis, Roads, Truax). In Xenakis ([5]), the sound is thought as an evolving gas structure. The audible field is modelled according to the Fletcher-Munson diagram, which is subdivided in a finite number of cells. Each instant is described through the stochastic activation of certain cells in the diagram (a "screen") and each screen has a fixed temporal duration. The sound/composition is an *aggregatum* of screens collected in a "book" in "lexicographic" order (as in the series of sections of a tomography). In Truax ([10], [7]), massive sound texture is obtained via the juxtaposition of multiple grain streams ("voices", like in polyphony): the parameters of each grain stream are controlled through tendency masks representing variations over time (i.e. timbre selection, frequency range, temporal density, [7], [10]). This approach is well known in the literature as Quasi-Synchronous Granular Synthesis (QSGS, [11], [4], [2]). In Roads [11], grains are scattered probabilistically over frequency/time plain regions ("clouds"). The compositional work relies on controlling cloud global parameters (i.e. start time and duration of the cloud, grain duration, density of grains, etc.). In these three cases, compositional strategies are based on the direct control of the creative process with an empty uniform time/frequency canvas[1]. Not surprisingly, the compositional metaphor in Roads is explicitly related to painting, using different brushes with different (sound) colours ([11]: 143).

The goal of this paper is to provide a new perspective on the composition process with granular synthesis by introducing a formal system based on two description levels. As we see below, the system can be viewed as a generalization of the note and the stochastic approaches.

[1] See also the graphic notation in [12]: 156-57. This is the standard spatial metaphor in different granular synthesis implementations using tendency masks: in a Csound-oriented perspective, see for example the software *GSC4* ([13]) and *Cmask* ([14]).

2. GEOGRAPHY: A TWO-LEVEL SYSTEM FOR GRAIN GENERATION AND CONTROL STRUCTURE

In this section we describe the formal system GeoGraphy, that models the composition process with two components, one for the generation of grain sequences, another for the parametric control of waveforms.

First, we introduce some terminology. A composition is a set of *tracks*; each track is a grain sequence (Figure 1), where the single grains are waveforms that result from granular synthesis and parametric control. The formal system GeoGraphy consists of two components: a graph-based generator of grain sequences (i.e. tracks), and a map-based controller of grain waveform parameters.

The grain generator (level I) is based on directed graphs, actually a multigraph ([15]), as it is possible to have more than one edge between two vertices (Figure 2). A vertex represents a grain; an edge represents the sequencing relation between two grains. Grains can be either sampled waveforms with fixed durations, or waveforms generated by a synthesis process with a duration that is overtly marked on the vertex (all the vertices in Figure 2 represent sampled grains). A label on an edge represents the temporal distance between the onset times of the two grains connected by the edge itself[2]. A grain sequence is a path on the graph, that in case a graph contains loops can also be infinite. The generation of a grain sequence is achieved through the insertion of dynamic elements into the graph, called *graph actants*. A graph actant is initially associated with a vertex (that becomes the origin of a path); then the actant navigates the graph by following the directed edges according to some probability distribution[3]. Multiple independent graph actants can navigate a graph structure at the same time, thus producing more than one grain sequence.

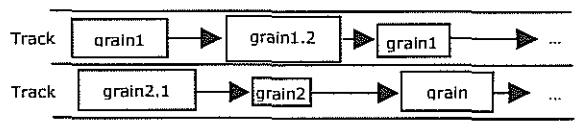


Figure 1: A musical piece

[2] All durations in the formalism can be made dependent on some probability distribution, so to act as a general stochastic grain generator. This feature together with the track (or voice) structure of the musical piece allows GeoGraphy to simulate the expressive power of QSGS.

[3] For those readers that are familiar with Petri nets, a graph actant can be viewed as a *token*. The probabilistic control of the token also reminds to stochastic Petri nets.

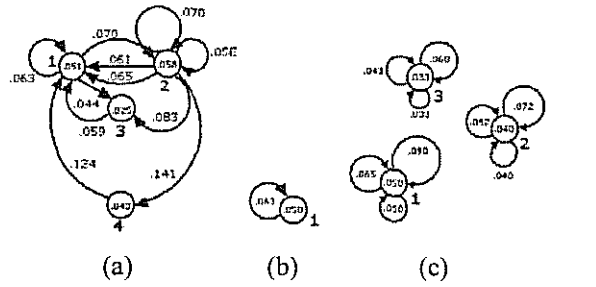


Figure 2: (a) A complex multigraph; (b) A graph of one vertex and one edge; (c) A graph consisting of three disconnected subgraphs.

In Figure 2 there are three examples of graphs. The graph in Figure 2a is a multigraph with several connections (almost completely connected). It also contains loops. One possible result is in Figure 3, where some amplitude control, a typical Gaussian envelope, has been applied to avoid clicking. Starting from vertex 4, the graph actant generates a grain of duration 43 milliseconds (vertex label), then it reaches vertex 1 with a delay of 124 milliseconds (edge label), it loops two times on vertex 1, generating two grains of 51 milliseconds with a delay of 63 milliseconds, then it leaves vertex 1 for vertex 2 and so on. As grain duration and delay of attack time are independent, it is possible to superpose grains (vertex label > edge label, see the last two grains in Figure 3).

In Figure 2b there is a graph with one vertex and one edge that loops on the unique vertex. The grain sequence produced by such a graph is the exact repetition of the grain associated with the vertex; each repetition starts after 63 milliseconds with respect to the beginning of the previous repetition.

In Figure 2c there is a graph consisting of three disconnected subgraphs, each with one vertex and three edges that loop on the vertex itself. If we assume a single actant on each graph, the system generates three simultaneous streams of grains. If we associate each vertex a grain of fixed frequency, we yield a spectrum consisting in three rows (Figure 4), a "stratus" in Roads' typology ([11]: 165, [2]: 104).

In order to control the setting of the parameters associated with the grain waveforms, the idea implemented in the GeoGraphy system (level II) is to position the graphs in a space, and then to control the parameter values by navigating the space (*control space* or *map of graphs* – Figure 5). Once the single sound streams have been defined through the generation of graphs, the composer distributes the graphs onto a map, and then designs a *trajectory* that allows to decide how the several sound streams contribute to the piece. The control of the parameters occurs with reference to the spatial metaphor: parameters value ranges are mapped onto spatial distance, and the nearer is a trajectory to some vertex, the higher is the value of some parameter for the grain waveform represented by that vertex.

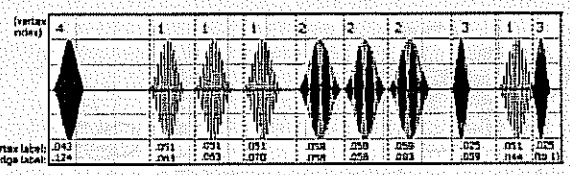


Figure 3: A grain sequence generated by the graph of Figure 2a.

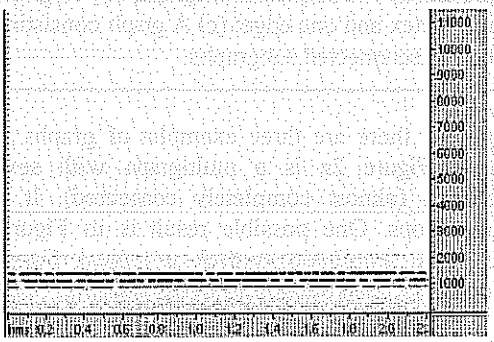


Figure 4: Spectrum of the signal generated by the three subgraphs of Figure 2c ("stratus").

So, the parametric controller of the GeoGraphy system is a spatial *map* of graphs. Theoretically, the space can have any number of dimensions, but we limit our composition model to the Euclidean space (actually, the examples in this paper feature only two dimensions). In order to place the graphs on the map, each vertex is associated with coordinates in the space. A map contains a finite number of graphs (n), which work independently, thus generating a grain sequences, where a is the total number of the graph actants that navigate in all the graphs. As there is at least one graph actant for each graph, there will be a minimum of n sequences ($a \geq n$).

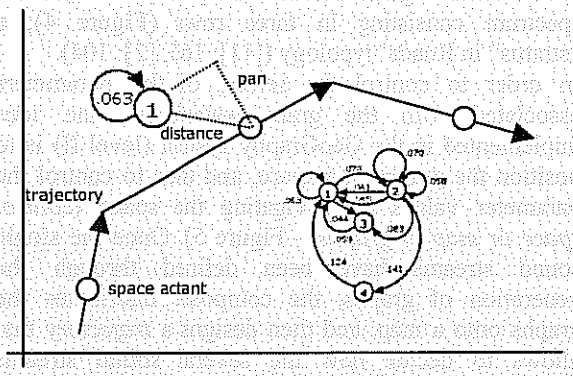


Figure 5: A control space with one space actant and its trajectory, and two graphs. Distance and panning are marked with respect to one vertex

Trajectory design occurs by defining a new actant (the *space actant* – Figure 5) which navigates at constant rate in the space following a *trajectory*. Each vertex emits a grain (determined by the passage of the graph actant); when a grain is generated, the device calculates the distance between the space actant and the vertex. This distance is then used as a general control parameter. If we consider the space actant as a directed human head (Figure 5), the displacement of the vertex from the frontal and the lateral axes is used to control panning, while the Euclidean distance is used to control amplitude. Trajectories can be explicitly defined by the composer or generated algorithmically (e.g., *brown motion*). Different trajectories determines different strategies of exploration of the map space.

The two levels of GeoGraphy are described in detail in ([16]). In the next section we discuss some expressivity issues of the model; in particular we see how it generalizes over the note approach and the stochastic approach.

3. EXPRESSIVITY ISSUES

The space typically considered in granular synthesis techniques is the physical continuum time/frequency. This space is not necessarily musically meaningful: it can be quite complex to define musical meaningful parameters while working with it without great efforts. On the contrary, a general map space as the one we have proposed can embed different topologies of musical features among the ones proposed in literature. This can be done through the application of grouping strategies to vertices in the control space. The two compositional dimensions are density (number of grains for surface unit) and structure organization of vertices (i.e. qualitative relationship over grains). For example, in order to enrich the spectral content of a musical object we can collapse several vertices in a single location. The graph structure can contribute to stress the hierarchical relationship in the group: the star-structured graph of Figure 6a generates a sequence of the form $1n1n\ldots$, thus stressing the importance of vertex 1. The vertices (or group of vertices) can be distributed following timbral spaces discussed in literature ([17]: 76; [18]; [19]:13; [20]: 59; [21]: 48; [22]: 197), but also pitch spaces like the Two-dimensional Melodic and Harmonic Maps proposed by ([23]: 374-78). Harmonic/inharmonic, sparse/dense, low/high, are categories that, together with topology features like centre/periphery, are of typical use.

If one chooses to consider one of the axis of the map space as frequency, the space actant displacement on that axis can be thought as a band pass filtering (if the distance has a threshold, bandwidth is represented exactly by the diameter of the audible circle, Figure 6b). Vertex groupings in Figure 6c follows cloud patterns as described by ([11]: 166; [4]: 182; [2]: 105) (from left to right: cumulus, stratus, glissandi).

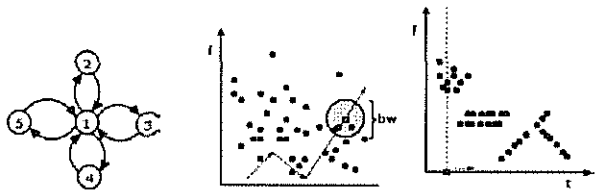


Figure 6: A star graph; (b) Space actant (square) as a bandpass filter (bw: bandwidth) (c) Space actant as a scanning device (audible circle compressed to dashed line) (edges are omitted)

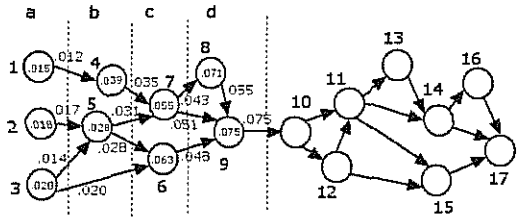


Figure 7: Starting from one of the vertices 1-3 every path has a duration between 143 (vertices 3+6+9) and 221 milliseconds (2+5+7+8+9). The subgraph 1-9 can consequently be considered as a note event leading to the note event, resulting from the subgraph 10-17

Note that the time/frequency domain, and hence the stochastic approaches, can be considered a special case of the map space in which: a) one axis represents frequency; b) the other represents time; c) the trajectory coincides with the time axis; d) the circle surrounding the space actant is compressed to its diameter parallel to the frequency axis; e) there is not distance threshold or, if present, it is as large as the frequency axis.

It must be noted that a map space should be used with caution in simulating a time/frequency space. In fact, the map space is filled not with actual events but only with potential events (vertices), in the sense that it is possible that a grain is emitted at the very moment in which the space actant scanning the timeline trajectory is passing (as the result of the activations of the vertex), but it is not necessary, as it depends on the grain generator (it is like scanning a sky filled with pulsating stars). Also note that density does not depends only on spatial distribution of the grains on the map, but mostly on graph structures.

Also the note approach can be simulated in GeoGraphy, i.e. GeoGraphy allows composers to work with vertices that represents events of greater complexity than grains. Consider the graph in Figure 7. In the example, vertices 1-9 represent a subgraph in which all the possible paths starting from 1-2-3 have a duration in range between 143 and 221 milliseconds, which can be considered as the duration of the whole subgraph as a higher-level note event. The whole graph can be considered as a two note cell, the first resulting from subgraph 1-9 an the second one from subgraph 10-17. The vertices 1-9 are grouped so that four sets of increasing edge and vertex labels emerges (a-d): with opportune tuning, this kind of relation can be used to model a typical percussive noisy attack with subsequent periodicity and decay.

The advantages of the GeoGraphy model for the composer rely mostly upon the symbolic approach, in that each graph structure represents a set of relations between vertices, which can be thought as *objets sonores* ([24]). Sound objects as defined by Schaeffer are symbolic objects encoding sonic properties apt to be used in compositional practice. In this sense, they are considered meaningful-to-the-ear objects, while physical dimensions need a continuous monitoring to be musically relevant. Sound objects have been placed at the analogous of the phonological level in language, as sets of relevant sonic features (*acoulogie* in [24], sonology in Laske's term, see [25] for a general discussion). Geography is intended to be a model to structure sequences of well-fit sound events (*objets convenables*), of which notes in traditional composition, but also microsounds, are special cases.

4. CONCLUSIONS

This paper has presented a formal system for the algorithmic control of composition based on granular synthesis. The system features a grain level, that organizes grains into a graph structure, and a spatial level, that distributes the graphs in specific locations of a space. The composition process is the design of a trajectory in the space, appropriately interpreted to control a number of parameters of physical and musical relevance. We have also seen how the system can be viewed as a generalization of the current approaches to composition with granular synthesis.

3. REFERENCES

[1] De Poli, G., "Granular Representations of Musical Signals: Overview", in De Poli, G. - Piccialli, A. – Roads, C. (eds.), Representations of Musical Signals, The MIT Press, Cambridge (Mass.)-London, 1991.

[2] Roads, C., Microsound, The MIT Press, Cambridge (Mass.)-London, 2001.

[3] Roads, C., "Introduction to Granular Synthesis", Computer Music Journal, vol. 12, n. 2, 1988.

[4] Roads, C., The computer music tutorial, The MIT Press, Cambridge (Mass.)-London, 1996.

[5] Xenakis, I., Formalized Music. Thought and Mathematics in Composition, Bloomington, Indiana, 1971 (rev. ed. Pendragon Press, Stuyvesant (NY), 1991).

[6] Wishart, T., Audible Design. A Plain and Easy Introduction to Practical Sound Composition, Orpheus the Pantomime, 1994.

[7] Truax, B., "Composing with Real-Time Granular Synthesis", Perspectives of New Music, vol. 28, n. 2, 1990.

[8] Jones, D. L. – Parks, T. W., "Generation and Combination of Grains for Music Synthesis", Computer Music Journal, vol. 12, n. 2, 1988.

[9] Lee, A. S.C., "Granular Synthesis in Csound", in Boulanger, R., The Csound Book. Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming, The MIT Press, Cambridge (Mass.)-London, 2000.

[10] Truax, B., "Real-Time Granular Synthesis with a Digital Signal Processor", Computer Music Journal, vol. 12, n. 2, 1988.

[11] Roads, C., "Asynchronous Granular Synthesis", in De Poli, G. - Piccialli, A. - Roads, C. (ed.), Representations of Musical Signals, The MIT Press, Cambridge (Mass.)-London, 1991.

[12] Roads, C., "Granular Synthesis of Sound", in Roads, C.-Strawn, J., Foundations of Computer Music, The MIT Press, Cambridge (Mass.)-London, 1985.

[13] Giordani, E., "GSC4 – Sintesi granulare per Csound", in Bianchini, R. – Cipriani, A., Il Suono Virtuale. Csound per PC e Mac. Teoria e Pratica, ConTempo, Roma, 1998.

[14] Bartetzki, A., "Csound Score Generation and Granular Synthesis with Cmask" (on-line paper: http://www.kgw.tu-berlin.de/~abart/CMaskPaper/cmask-article.html, visited on Nov 27th, 2002).

[15] Diestel, R., Graph Theory, Springer (electronic ed.), New York, 2000.

[16] Valle, A., "Del *Gran Paese*, or: Some brief remarks on space", unpublished paper, 2002.

[17] Plomp, R., Aspects of Tone Sensation. A Psychophysical Study, Academic Press, London,1976.

[18] Wessel, D., "Timbre Space as a Musical Control Structure", Computer Music Journal, vol. 3, n. 2, 1979.

[19] Rasch, R.A. - Plomp, R., "The Perception of Musical Tones", in Deutsch, D., The Psychology of Music, Academic Press, Orlando, 1982.

[20] Slawson, W., Sound Color, University of California Press, Berkeley-Los Angeles-London, 1985.

[21] Risset, J. C. – Wessel, D., "Exploration of Timbre by Analysis and Synthesis", in Deutsch, D., The Psychology of Music, Academic Press, Orlando, 1982.

[22] Lerdahl, F., "Les hiérarchies de timbres", in Barrière, J.B. (ed.), Le timbre. Métaphore pour la composition, Christian Bourgois-IRCAM, Paris, 1991 (French transl. of "Timbral Hierarchies", Contemporary Music Review, vol. 2, n. 1, 1987).

[23] Shepard, R. N., "Structural Representations of Musical Pitch", in Deutsch, D., The Psychology of Music, Academic Press, Orlando, 1982.

[24] Schaeffer, P., Traité des objets musicaux, Seuil, Paris, 1966.

[25] Bel, B., "Symbolic and Sonic Representations of Sound-Objet Structure", in Balaban, M., Ebcioğlu, K., Laske, O. (eds.), Understanding Music with AI: Perspectives on Music Cognition, The AAI Press/The MIT Press, Cambridge (Mass.)-Menlo Park-London, 1992.

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

INVESTIGATION OF THE PLAYABILITY OF VIRTUAL BOWED STRINGS

Stefania Serafin, Diana Young

CCRMA, Department of Music
Stanford University, Stanford, CA
MIT, Media Lab
Cambridge, MA
serafin@ccrma.stanford.edu, diana@media.mit.edu

ABSTRACT

Driving a bowed string physical model using a bow controller, we explore the potentials of simulating the gestures of a violinist using a virtual instrument. After a description of the software and hardware developed, preliminary results and future work are discussed.

1. INTRODUCTION

One of the aspects that made the family of bowed string instruments successfull is the incredible degree of nuances and expressivity that a player can obtain with a bow. In building virtual bowed strings instruments, the same expressivity is a desirable characteristic. The issue of expressivity in synthetic bowed string's research is well known in the computer music field. The importance of controlling a bowed string physical model with input parameters that simulate a physical gesture was first underlined in the work of Chafe [3] and Jaffe and Smith [6]. In this research the combination of the input parameters of a bowed string physical model was used to reproduce different bow strokes such as detaché, legato and spiccato.

Although the goal was to reproduce a particular sound specific to a certain performer's gesture, no real-time input controller was used. To our knowledge, one of the first attempts to control a bowed string physical model using devices other than the traditional mouse and keyboard was proposed in the early 90s by Cadoz and his colleagues at ACROE. Using a device capable of providing haptic feedback the player was able to feel the synthetic bowed string. Recent results of this research are detailed in [4].

Recently, an increasing interest has been shown in controllers for virtual bowed strings developed using physical models.

Preliminary experiments [11] used a Wacom tablet to control a real-time waveguide bowed string model developed in the Max/MPS environment.

The Wacom tablet allows a straightforward mapping of the parameters of the bowed string physical model, since it is able to capture the pressure and two axis position of the pen provided with the tablet itself; these values can be easily mapped to the bow pressure, velocity and bow-bridge distance. Moreover, the tablet is able to detect the horizontal and vertical tilt angle of the pen. This immediately shows an advantage of the Wacom tablet versus more traditional input devices such as a mouse and a keyboard, i.e. the possibility of having the same degrees of freedom as a bow in contact with a string. The tablet, however, shows limitations that are mainly illustrated by the difficulties of using it as a performance instrument. This is due also to the lack of haptic feedback (because of the lack of elasticity of the tablet compared to the bow hair) and the dramatic difference between its ergonomics and that of a traditional violin bow.

This lack of force feedback was compensated for when controlling the same bowed string physical model using the Moose, a device built by Sile O'Modhrain and Brent Gillespie. Experiments show that the playability of the bowed string physical model greatly increases when haptic feedback is provided [8].

In order to introduce both force feedback and ergonomics that are reminiscent of a traditional violin interface, Charles Nichols built the vBow [7], a haptic feedback controller. The goal of the vBow is to be able to introduce a new violin interface that addresses the prior limitations of MIDI violins as well as to provide a controller that can also play other real-time synthesis tools. Extended techniques for bowed string physical models have been explored using the Metasax [2], an extended tenor saxophone designed by Matthew Burtner.

2. ASPECTS OF PLAYABILITY

In virtual musical instruments and musical acoustics, the word *playability* has different definitions. While in this paper we focus on playability of virtual bowed strings, the same issues can be applied also to all other expressive virtual instruments.

According to Jim Woodhouse [15], playability of virtual instruments means that the acoustical analysis of the waveforms produced by the model fall within the region of the multidimensional space given by the parameters of the model. This is the region where *good tone* is obtained. In the case of the bowed string, *good tone* refers to the Helmholtz motion, i.e. the ideal motion of a bowed string that each player is trying to achieve. The Helmholtz motion is given by an alternation of stick-slip-stick-slip, in which the string sticks to the bow hair for the longest part of its period, slipping just once. Experiments show that simulated bowed strings have the same playability as real bowed strings as calculated by Schelleng [10].

Further experiments also show that the playability of virtual bowed strings increases when accurate friction models that account for the thermodynamical properties of rosin are taken into account.

This above mentioned definition of playability is interesting from a musical acoustician's point of view but does not reflect performance issues. In these experiments, in fact, the input parameters that drive the bowed string model corresponding to the right hand of the player are kept constant for each simulation. This is a

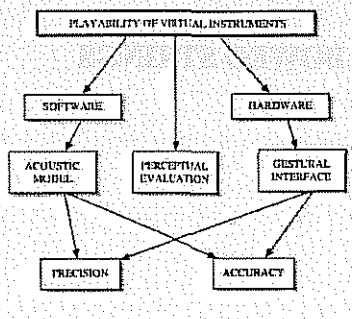


Figure 1: Playability chart of a virtual musical instrument

situation that is clearly not the same as that which occurs in violin performances. As in performance it is the continuous evolution of the input parameters that constitute the nuance that are the characteristics of an expressive performance. In order to address this issue, Askenfelt [1] studied the contribution of bowing parameters in different bow strokes, trying to determine the physical limits of the input parameters in order to achieve a specific stroke. He determined the maximal duration of the pre-Helmholtz attack allowed in order to judge a particular stroke as acceptable.

In interactive performances, other issues related to playability must be considered. Sile O'Modhrain [8], for example, studied the influence of haptic feedback on the playability of virtual bowed strings.

She discovered that haptic feedback greatly increases the playability of virtual instruments. In this context, playability is referred to as the ability of bowed string players to consistently perform different bow strokes that produce violin sounds that are judged as perceptually acceptable by professional bowed string players and also have waveforms that reside within the playability region as defined by [15].

Another important issue in virtual musical instruments' playability refers to the precision and accuracy of both the hardware and the software that comprise the virtual instrument itself.

In the situation where a controller is used to drive a synthetic model, we may say that a controller is precise if it allows the player to control subtle variations of the parameters with great care, and we may say that it is accurate if the data collected by the device may be easily correlated to a real value of gesture. Correspondingly, we may say that the model is precise if it reproduces the nuances in the sonorities that are produced by a real instrument, and it is accurate when the physical input/output data can be matched to measurements performed on real acoustic instruments. An issue of importance in considering both of these aspects is that of latency, which may be adversely affected by limitations of either precision and accuracy. Of course, in musical performance, minimizing latency is a priority.

Obviously, the acceptance level of responsiveness varies according to the instrument played. For example, percussion instruments require a higher responsiveness than woodwind instruments. In general, it is reasonable to assume that transient instruments require a higher level of responsiveness that sustained instruments.

In the case of the bowed string, the issue of responsiveness is crucial when fast bow strokes such as staccato, balzato, martellato are performed.

Another more general definition of playability is the ability of the virtual instrument to be ergonomically playable, that is the player should be able to physically manipulate the interface freely and with ease.

A chart that summarizes all the playability issues mentioned above is shown in figure 1.

In this paper we are interested in exploring all the previous definitions of playability and extending them by driving a violin bowed string physical model using a bow controller.

Rather than building or adopting an alternate controller or working with a haptic device such as in the examples mentioned above, we are interested in exploring the possibility of reproducing traditional bowing techniques using a bow controller that behaves in a manner as closely related to that of a traditional violin bow as possible.

This allows us to validate both the model and the controller by comparing it to the behavior of the traditional instrument.

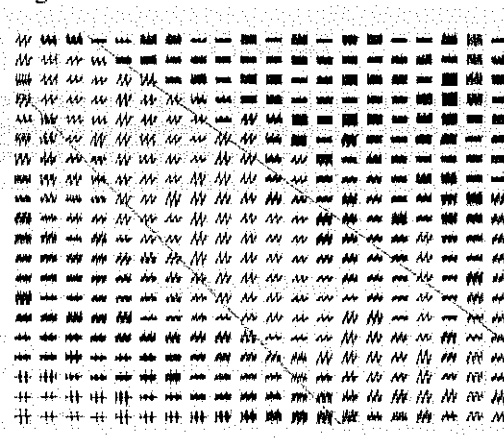


Figure 2: Playability plot obtained by capturing the waveforms after steady-state motion is obtained. Horizontal axis: bow force. Vertical axis: bow position.

3. A BOWED STRING PHYSICAL MODEL

We built a bowed string physical model that combines waveguide synthesis [14] with latest results on bowed string interaction modeling [9].

A schematic representation of the model is shown in figure 3. In this model, the bow excites the string in a finite number of points, which represent the bow width. The frictional interaction between the bow and the string is modeled considering the thermodynamical properties of rosin [13]. The bow width is modeled by discretizing the region of the string in contact with the bow using finite differences and calculating the coupling between the waves propagating along the string and the frictional interaction between the bow and the string at each point. Once the velocity of the string at the contact point has been calculated, the waves propagating along the string are modeled using digital waveguides. More precisely, transversal and torsional waves propagating toward the bridge and the fingerboard are modeled as pairs of one dimensional digital waveguides.

The outgoing velocity at the bridge is filtered through the body's resonances and corresponds to the output waveforms perceived by the listener.

For a detailed description of the physical model see, for example, [12].

The input parameters of the model corresponding to the right hand of the player are bow position relative to the bridge (normal-

ized between 0 and 1, where 0 corresponds to the bridge, 1 corresponds to the nut, 0.5 corresponds to the middle of the string), bow pressure, bow velocity, and amount of bow hair in contact with the string. The model has been implemented as an external object in the Max/MSP environment.

Figure 2 shows the waveforms obtained by running the model with a constant bow velocity of 0.05 m/s, and varying the bow force and bow position between 0 and 5 N and 0.01 and 0.4 respectively. The waveforms are captured after a steady-state motion is achieved. Inside the two straight lines appears the playability region as measured by Schelleng with the same parameters' configuration, Note how the synthetic model and the real instrument are in good agreement concerning the playability region's results.

4. THE BOW CONTROLLER

The bow controller provides three different types of measurement that reflect the nuances of gesture that may be observed in traditional string instrument bowing technique. The controller's sensing system employs commercial MEMS accelerometers to measure three axes of acceleration, electric field sensing to track bow position and bridge distance from the bowing point, and foil strain gauges to detect changes in the downward strain of the bow stick as well as in the orthogonal direction (toward the scroll).

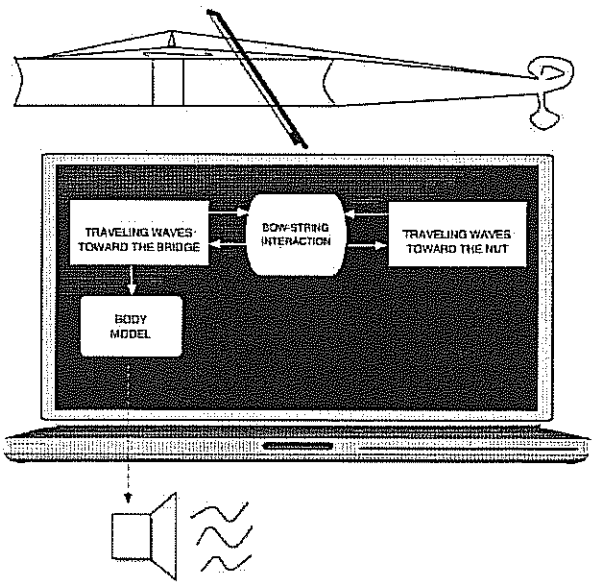


Figure 3: A bowed string instrument and the corresponding simplfied block diagram of its model.

These sensors were chosen so as to offer a player and composer the ability to capture data concerning all of the parameters of bowing that contribute to the interaction between the bow and the string: bow speed and bowing point (from the position tracking), downward force and bow width (both reflected by strain sensors). Additionally, the accelerometers were included as a means of observing with great precision changes in bowing direction as well as the differences in characteristics of various kinds of string attacks.

The placement of the sensors and the accompanying electronics was carefully designed so as to provide the best results in measurement, while also maintaining as completely as possible the balance, weight, and feel of the bow. Also a priority in this implementation is the protection of the hardware itself and robustness of the overall system, as the interface was intended for use in live performance and rigorous laboratory experiments.

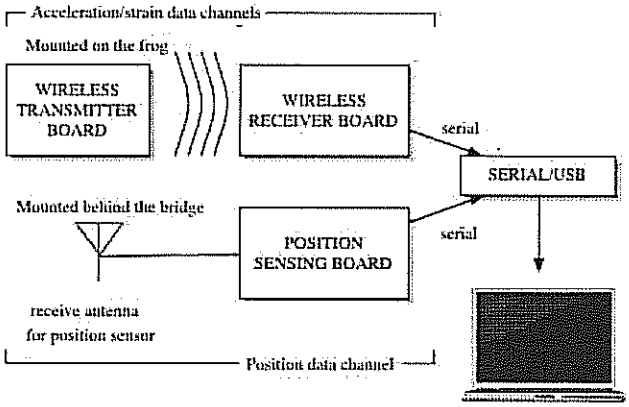


Figure 4: Data flow for the violin controller.

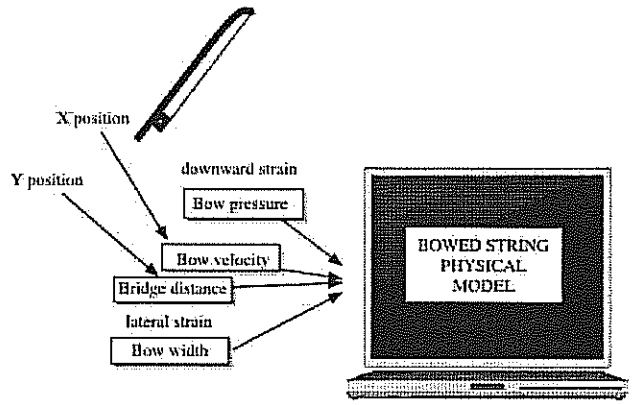


Figure 5: One to one mapping of the bow controller to the bowed string physical model.

The strain sensors are effectively integrated into the composite material of the bow, as they are permanently adhered around the midpoint of the stick. They are protected from wear by a thin layer of flexible tubing, and their connecting wires run down the lower length of the bow to the electronics board mounted on the frog (while still allowing the frog to slide freely in order to adjust the tension of the bow hair). This electronics board houses the accelerometers, the microcontroller, and the wireless transmitter, which sends the acceleration and strain data to a remote receiver board containing a serial port. This board also sends two separate signals to either end of a resistive strip that runs the length of the bow stick, acting as an antenna for the position measurement. The resultant mixed signal is received by an antenna mounted behind the bridge of the violin, and this signal is connected to another electronics board that determines the different amplitudes of the two received signals.

The mechanical layout of the electronics on the bow allows the player to use traditional right hand bowing technique, while keep-

ing the hardware out of harm's way from the players hands and the strings of the violin. The electronics add about 30g of weight to the original carbon bow, but because of the careful distribution of the weigh along the length of the bow the balance point of the bow is still well within the normal range for a traditional violin bow. In addition, the remote electronics boards are small and light, and the bow is wireless and runs on a camera-style battery with a lifetime of over 20 hours, and so the interface is highly portable.

The data from the bow controller is carried by two separate serial buses: one for the acceleration and strain data and one for the position data. In this experiment, a commercial serial/USB converter is used to connect the two serial lines with the input of a Macintosh computer [16].

5. MAPPINGS

In order to build an expressive virtual musical instrument, the capture of the gesture of the performance is as important as the manner in which the mapping of gestural data onto synthesis parameters is done. In the case of physical modeling synthesis, a one-to-one mapping approach of control values to synthesis parameters makes sense due to the fact that the relation between gesture input and sound production is often hard-coded inside the synthesis model [5]. Because both the physical model and the bow controller are developed according to physical input and output parameters, the mapping between the two is straightforward. Downward bow force of the controller is directly mapped into bow force in the physical model. Bow velocity and bow-bridge distance are captured by measuring the horizontal and vertical position of the bow respectively. Moreover, lateral strain sensors are mapped into the amount of bow hair in contact with the bow.

6. PRELIMINARY EXPERIMENTS

The first experiments completed to begin the integration between the bow controller and the violin physical model were encouraging. With the model parameters of bow-bridge distance, bow velocity, and bow width for a fixed frequency held constant, we mapped the downward strain data from the bow controller to the downward force parameter. While applying pressure on the strings of our test violin (without actually drawing the bow across the strings) using the bow controller, we were able to quickly produce sonorities from the model that sounded appropriate for the amount of pressure we applied to the bow.

We also experimented by drawing the bow across damped strings to determine the response for changes in downward force that occur during different bow strokes. Again, the response of the model and the feel of this simple mapping produced a satisfying interaction and was interpreted by us as extremely promising for continued work.

7. DISCUSSION AND FUTURE WORK

The initial experiments for this research project yielded positive results. Players were able to produce bow strokes that felt natural and were judged as acceptable from bowed string players. Furthermore, the mapping between the amount of force and the change in amplitude of the sound seemed intuitive.

In the future, we will continue to map each of the remaining input parameters of the physical model to the appropriate sensor data and aim to iteratively build a convincing link between physical controller and virtual violin. As this work progresses, we will also evaluate the overall system by examining all aspects of playability for its component parts.

8. REFERENCES

[1] A. Askenfelt. Measurement of the bowing parameters in violin playing. *Journal of the Acoustical Society of America*, 86(2):503–516, August 1989.

[2] M. Burtner and S. Serafin. The exbow-metasax: Compositional applications of bowed string physical models using instrument controller subsitution. *Journal of New Music Research*, 31, 2002.

[3] C. Chafe. Simulating performance on a bowed string. *Technical Report*, STAN-M-48, May 1988.

[4] J.-L. Florens and C. Henry. Bowed string synthesis with force feedback gesture interaction. In *Proc. International Computer Music Conference*, pages 115–118. ICMA, 2001.

[5] A. D. Hunt, M. Paradis, and M. M. Wanderley. The importance of parameter mappings in electronic instrument design. In *Proc. NIME*, 2002.

[6] D. Jaffe and J. Smith. Performance expression in commuted waveguide synthesis of bowed strings. In *Proc. International Computer Music Conference*, pages 343–346. ICMA, 1995.

[7] C. Nichols. The vbow: A virtual violin bow controller for mapping gesture to synthesis with haptic feedback. *Organized Sound*. In press.

[8] S. O'Modhrain, S. Serafin, C. Chafe, and J. Smith. Qualitative and quantitative assessment on of a bowed string instrument. In *Proc. International Computer Music Conference*. ICMA, Aug. 2000.

[9] R. Pitteroff. Mechanics of the contact area between a violin bow and a string. part i:reflection and transmission behaviour. part ii: Simulating the bowed string. part iii: Parameter dependance. In *Acustica-Acta Acustica*, pages 543–562, 1998.

[10] J. C. Schelleng. The bowed string and the player. *Journal of the Acoustical Society of America*, 53(1):26–41, Jan. 1973.

[11] S. Serafin, R. Dudas, M. M. Wanderley, and X. Rodet. Gestural control of a real-time model of a bowed string instrument. In *Proc. International Computer Music Conference*. ICMA, Oct. 1999.

[12] S. Serafin, J. O. Smith, III, and J. Woodhouse. An investigation of the impact of torsion waves and friction characteristics on the playability of virtual bowed strings. New York, Oct. 1999. IEEE Press.

[13] J. H. Smith and J. Woodhouse. The tribology of rosin. *J. Mech Phys.Solids*, 48:1633–1681.

[14] J. O. Smith. Physical modeling using digital waveguides. 16(4):74–91, Winter 1992.

[15] J. Woodhouse. On the playability of violins. Part I: Reflection functions. Part II: Minimum bow force and transients. *Acustica*, 78:125–136,137–153, 1993.

[16] D. Young. The hyperbow controller: Real-time dynamics measurement of violin performance. In *Proc. NIME*, 2001.

The pCM framework for realtime sound and music generation

Leonello Tarabella

computerART project of I.S.T.I. "A.Faedo"
Research Area of C.N.R. - via Moruzzi 1, 56124 PISA
l.tarabella@cnuce.cnr.it
www.cnuce.pi.cnr.it/tarabella/cART.html

ABSTRACT

I started to write a very basic library of functions for sound processing and for driving the gesture interfaces realized at cART project of CNR, Pisa. In the long run the library became a very efficient, stable and powerful framework based on pure C programming, that is "pure-C-Music", or pCM.
This programming framework gives the possibility to write a piece of music in terms of synthesis algorithms, score and management of data streaming from external interfaces. The pCM framework falls into the category of the "embedded music language" and has been implemented under the Metrowerks' Code Warrior C-compiler.
As a result a pCM composition consists of a CW project assembled with all the necessary libraries including a DSP.lib able to implement in realtime the typical synthesis and processing elements such as oscillators, envelope shapers, filters, delays, reverbs, etc. The composition itself is a C program which mainly consists of the Orchestra() and Score() functions. Everything here is compiled into machine code and runs at CPU speed.

1. INTRODUCTION

At "cART" attention has been focused in designing and developing original general purpose man-machine interfaces taking into consideration the wireless technologies of infrared beams and of real-time analysis of video captured images [1,2].
The basic idea consists of remote sensing gesture of the human body considered as a natural and powerful expressive "interface" able to get as many as possible information from the movements of naked hands [3] such as the Twin Towers, Handle and the Imaginary Piano. The TwinTowers [4,5] is infra-red light sensing device for real-time control of electro-acoustic music which gives the performer the sensation of "touching the sound"; the Handel system [6] is based on real-time FFT analysis of video captured images for recognizing shape, position and rotation of the hands; in the Imaginary piano, hands movements are recognized and used for controlling sound and musical structures of the piano.
These devices and systems has been previously presented at both technical and artistical levels in many different conferences and contemporary art festivals [7,8,9].
Another application based on the video captured image analysis system is PAGe, Painting by Aerial Gesture, which allows video graphics realtime performances. Video clips of performances realized with these gesture interfaces can be found at the reported web page.

1.1. Interfaces and mapping

Specific targets of the research consist of studying models for mapping gesture to sound. In fact the different kind of gestures such as continuos or sharp movements, threshold trespassing, rotations and shifting, are used for generating sound event and/or for modifying sound/music parameters.
For classic acoustic instruments the relationship between gesture and sound is the result of the physics and mechanical arrangement of the instrument itself. And there exist one and only one relationship.
Using a computer based music equipment, it's not so clear "what" and "where" is the instrument; from gesture interfaces to loudspeakers which actually produce sound, there exist a quite long chain of elements working under control of the computer which performs many tasks simultaneously: management of data streaming from the gesture interfaces [10], generation and processing of sound, linkage between data and synthesis algorithms, distribution of sound on different audio channels, etc.
This means that a music composition must be written in term of a programming language able to describe all the components including the modalities for linking gesture to sound, also said how to "map" gesture to sound.
"Mapping" makes therefore part of the composition [11].
In order to put at work the mapping paradigm and the facilities offered by the gesture interfaces we realized, at first we took into consideration the most popular music languages: MAX/DSP and Csound. Unfortunately, both languages resulted not precisely suited for our purposes mainly because Csound was not so realtime as declared and Max was not so flexible for including video captured images analysis code.

2. pCM

The pCM framework has been implemented first for Macintosh computers using the CodeWarrior C compiler by Metrowerks. As a result a pCM composition consists of a CW project which includes all the necessary libraries including a DSP.lib (Digital Signal Processing library).
This library consists of a number of functions (at the moment more than 50) able to implement in realtime the typical synthesis and processing elements such as oscillators, envelope shapers, filters, delays, reverbs, etc..
The composition itself is a C program consisting of four void functions: Init(), Orchestra(), Score() and End() properly invoked by the main program which controls the whole "machinery":

```
void mainMechanism();
{
 ........;
 Init();
 do
 {
   Orchestra();

   Score();
 }
 while (finish);
 End();
}
```

For pratical reasons of consistency it's a good idea the four functions make part of the same file which also includes declaration of the variables visible by the all functions and expecially from *Orchestra*() and *Score*(). Everything here is written following the C syntax: synthesis algorithms, score and declaration of variables and data structures. Everything here is compiled into machine code and runs at CPU speed.
The *Init()* function includes everything regarding the initialization of envelopes, tables, delays, reverbs, variables and data structures and the loading of samples. Usually it also includes calls for opening Midi, TCP/IP, Audio and/or Video Input channels.
As a counter part the *End*() function is called at the end and is used for closing channels and dispose previously allocated memory.

2.1. Orchestra and instruments

An instrument is defined in the *Orchestra()* function and consists of code for sound synthesis and processing; it is continuously called at audio sampling rate, that is 44100 times per second. An instrument is defined in terms of an ordinary C program - with all the programming facilities such as *for*, *do-while*, and *if-then-else* control structures - which calls functions belonging to the DSP.lib; sound is generated by assigning the results of the whole computation to two predefined "system variables" *outL* and *outR*. The *iN* predefined "system variable" allows to select different instruments inside the same Orchestra: what follows the *iN* variable assignament is related to that instrument.
Look at the following simple example.

```
....
  iN=4;
  sig  = ampli*Env(1)*Oscil(1,freq);
  outL = sig*pan;
  outR = sig*(1.-pan);
.....
```

where *sig* is a local variable, *ampli*, *freq* and *pan* are global variables loaded by Score(), Env(1) and OscSin(1,freq) belong to the DSP.lib and sounds like that:

```
float Oscil (int nOsc,  float freq)
{
  float pos;
  pos = oscPhase[iN][nOsc] + freq;
  if(pos>=tabLen)pos=pos-tabLen;
  if (pos<0) pos=pos+tabLen;
  oscPhase[iN][nOsc] = pos;
  return Tabsen[(long)pos];
}

float Env(int nEnv)
{
 float  vval,vv,pos;
 long   ntabEnv =envNum[iN][nEnv];
 pos  = envPos[iN][nEnv];
 vval=*(envTable[ntabEnv]+pos);
 if((long)pos<envLenght[ntabEnv])
   envPos[iN][nEnv]=pos++;
  return vval;
}
```

All the DSP.lib functions return a value normalized between -1.0 and +1.0 or (0.0 and +1.0 in the case of envelopes); it's up to the programmer/composer to properly rescale the value when necessary.
The system variables *inL* and *inR* make it possible to input audio signal in a sound process algorithm. In the next example

```
.........
outL = Reverb(1,inL);
outR = GetDelay(1); PutDelay(1,inR);
.............
```

the left channel signal is reverberated and the right one is delayed. The delay length is declared at Score() level using `NewDelay(n,dur)` where `n` identifies that particular delay line and `dur` is expressed in seconds.

2.2. The score

The Score() is a C function which prepares parametric values and loads the global variables (*ampli*, *freq* and *pan* in the example) used by the active instrument in Orchestra().
There exist different modalities for writing a score: following the algorithmic composition approach, writing sequences of predefined events, getting values coming from the external gesture interfaces and, finally, combining in different ways these techniques.
Suppose we want to control in real-time a very simple instrument using movements of the mouse by linking the vertical position to frequency and the horizontal position to left-right panning. In the MacOs environment, the mouse position is found invoking the

GetNextEvent(..) tool-box function which returns the x,y position values in the "Event.where.v/h" variable and used as follows:

```
pan    = Event.where.h/1023.;
freq   = Event.where.v+200.0;
```

These two lines make part of the Score() which is automatically and repeatedly called by the main mechanism. Since the mouse spans between 0 and 1023 horizontally and from 0 to 767 vertically, the variable *pan* (ranging between 0. and 1.) and *freq* (ranging between 200. and 967.0) communicate proper values to the instrument for changing frequency and panoramic position. The following example explains the dynamic relationship between the Orchestra and the Score.

3. AN EXAMPLE OF COMPOSITION

A composition consists of five different sections:
1- global variables declaration for communication between the Orchestra() and the Score(); 2- Init(), for opening channels, declaring envelopes, reverbs and delay lines, loading samples, etc..; 3- Orchestra() where sound synthesis and processing algorithms are defined; 4- Score() defined in terms of algorithms and management of data streaming from external gesture controllers; 5- End() for closing channels and disposing envelopes and samples. The following listing reports a real working example.

```
//======== FMouse =======
//  global  variables
float     frq,pan,vert,val;
int       enva, samp;
bool      mousePressed;
Point     position;

int Init ()//-------------------------
{
  float env[]={3, 0.0,0.0, 0.1,1.0,
               4.0,0.0};
  enva  = NewLinEnv(env);
  mousePressed=false;
  OpenAudio();
}//-------------------------------
void End()
{
  DisposeEnv(enva);
  CloseAudio();
}//================================
void Orchestra() // simple FM
{
 iN = 1;
 if (noteon[1]) TrigEnv(1);

val=Env(1)*Oscil(1,frq+Oscil(2,frq*.5));
 outR = val*pan;
 outL = val*(1.-pan);
}


void Score ()
{
  if (Button() && !mousePressed)
  {
   iN=1; noteon[iN] = true;
   pan          = Event.where.h/1023.;
   frq          = Event.where.v+200.0;
   mousePressed= true;
  }
  if (theEvent.what==mouseUp)
       mousePressed=false;
}//--------------------------------
```

Functions in bold belong to the DSP.lib. Orchestra is automatically called at 44.100 Hz sampling rate while Score is called from 50 to 100 times per second depending on the CPU request after the complexity of Orchestra.
Each time the mouse button is depressed (Score has the control) the predefined boolean variable *noteon[iN]* is set to true; then, the Orchestra trigs the envelope defined in Init and sound is produced with pitch related to the mouse vertical position and panoramic position related to the mouse horizontal position.
Actually, the layout of both Score and Orchestra is more complex and consists of a number of " *case N: break;* " blocks which define different situations at micro level of sound processing and at macro level of events control.

4. THE DIGITAL SIGNAL PROCESSING LIB

This is the library of predefined functions which perform the micro level computation for generating and processing sound. What follows is a very small excerpt of the collection of fuctions at the moment developed and upgraded when requested.

```
float  Noise();
float  Oscil (int nOsc, float frq);
void   TrigKarplusStrong(int nOsc);
float  KarplusStrong(int nOsc,float frq);
int    NewLinEnv (float v[]);
void   TrigEnv (int nEnv);
float  Env (int nEnv);
int    LoadSample (char nomesmp[]);
void   TrigSample (int nSmp );
float  Sample(int nSmp );
float  Bandpass(int nFilt, float inp,
                float  freq, float q);
float  Lowpass(int nFilt, float signal,
               float cutfreq);
float  Hipass(int nFilt, float signal,
              float cutfreq);
void   NewDelay(int nDly, float dur);
void   PutDelay(int nDly, float val);
float  GetDelay(int nDly);
void   NewReverb(int nRev);
float  Reverb(int nRev, float sign);
void   OpenAudio();
```

It is possible to here recognize the very well known building blocks for a software synthesizer such as oscillators, envelope generators, filters and so on. How to connect them for carrying out a synthesis algorithm it's a matter of C language syntax. For example, in order to reverb a filtered white-noise signal, the following code may be written:

```
float temp1 = Noise();
float temp2 = Lowpass(1,temp1,alfa);
float sign  = Reverb(1,temp2);
```

```
outL=outR=sign;
```

But it also possible to use a more compact expression following the peculiarity of C-programming:

```
outL=Reverb(1,Lowpass(1,Noise(),alfa));
```

In both cases `alfa` is a global variable used to control the synthesis algorithm with values coming from the Score().

4.1 Other facilities

It's also possible to generate (offline or in realtime under the control of data streaming from gesture interfaces, sequences of events automatically activated by the Scheduler(), a special mechanism which triggers sounds at the right times and change parametric values in the instruments of an Orchestra.
Other facilities supported by the pCM framework are:

-a) CD tracks playing
-b) direct to memory recording
-c) midi message management.
-d) realtime image grabbing.

```
CDtrackSearch(short nTrack);
CDplay ();
CDstop ();
CDvolume(short vol);
```

This is particularly useful to realize a "sound bridge" when passing to a new situation which requires stopping the main mechanism of sound generation.

```
Record("soundFilename.aiff",long lenght)
```

This is called in the Init() and starts the recording of the global sound result onto memory with no loss of quality. The file is saved onto disk in .aiff or .wav format.

```
OpenMidi();
GetMidiData();
SendMidiMessage(int status,
                int data1, int data2);
NoteOn(int chan, int num, int vel);
etc..
```

The `OpenVideo()` function allows to open a video channel and to grab frames coming from a video camera. A "videoCallBack" function is installed when the video channel is open and called as service interrupt routine when a new frame has been digitalized and put in memory: the analysis of images is performed by processing the planar x-y matrix of values corresponding to the pixels of frames. Extracted parameters values are then used in real time in the Score() as part of a pCM project.

The gloabal variable *realTime* reports the current time elapsed after the last call to `ResetTimer()` void function.

5. TOWARD THE OBJECT ORIENTED PARADIGM

Due to the great interest of the students of my course on Computer Music at Pisa University towards pCM, some of them (Diego Colombo, Andrea Buffa and Giuseppe Croccia) particularly skilled and hardworking and who really like programming and "put their hands" inside computers, decided to port the pCM framework onto PC/Windows environment rewriting it following the Object Oriented paradigm.
Starting from the basic approach of the main mechanism and the DSP.lib functionalities of pCM, they started to write the "MusiXBox Library" by highlighting the roles of the Musician, the Musical Instrument and the Audience.
Instruments are constructed as an extention of SoundObjects consisting of:

- synthesis algorithm
- SoundGenerator
- Building blocks such as Delay, Reverb, Reson, KarplusStrong, BandPass, etc.
- Sound Controller
- JackStreamer

The SoundGenerator implements all the functionalities of oscillators, envelopes, filters etc.
SoundControllers make it possible to map values coming from external gesture interfaces onto the synthesis algortihms.
A JackStreamer is an abstraction of the real connection between the elements of a single instruments and between the instruments as it happens in the real world.
An instrument produces a monoaural sound source and can be assembled together with other instruments to be passed to a Musician which executes the task for rendering sound. It also controls the spatial position and moving speed including the doppler effect. In fact, the Renderer is connected to the Musicians and for each of them it can implement different kind of spatial synthesis depending on the available audio equipment: Mono, Stereo, Quadri, DolbyDigital, etc.

The technology of rendering is based on the DirectX9 by Microsoft which make it possible to stay apart from specific drivers and the peculiarities of sound boards and still gets the maximum of what at disposal on a particular workstation.
The basic libraries have been realized in ansi C++; in this way only the main rendering engine must be rewritten for different platforms (Win32, MacOS e Linux).
We plan to put these framework available on the Net properly documented and easy to use and to extent it depending on particular needs of users.

6. CONCLUSION AND AKNOWLEDGMENTS

The pCM framework has been efficiently used for composing and performing [many pieces of music under the control of the gesture tracking systems and devices realized at cART project in Pisa. It has been also included as special topic in the course of Computer Music 1 yearly teach at the Computer Science Facoulty of Pisa University. Special thanks are due to Massimo Magrini who greatly contributed to set up the pCM main mechanism currently in use and the many facilities for data communication and audio processing. Also thanks to Roberto Neri who recently graduated in Electronic Engineering at Pisa University with a thesis regarding the upgrading of the pCM DSP.lib.

7. REFERENCES

[1] Tarabella L., Magrini M., Scapellato G., "Devices for interactive computer music and computer graphics performances" , IEEE First Workshop on Multimedia Signal Processing, Princeton, NJ, IEEE cat.n.97TH8256, Computer Society Press, (1997).

[2] Tarabella L., Bertini G., "Wireless technology in gesture controlled computer generated music", Proceedings of MOSART2001, Barcelona, nov.2001

[3] Tarabella L., Bertini G., "Giving expression to multimedia performances" – ACM Multimedia 2000, Workshop "Bridging the Gap: Bringing Together New Media Artists and Multimedia Technologists" Los Angeles, 2000

[4] Tarabella L., Bertini G., Sabbatini T., "The Twin Towers: a remote sensing device for controlling live-interactive computer music". In Procs of 2nd international workshop on mechatronical computer system for perception and action, SSSUP, Pisa, (1997)

[5] Rowe, R., Machine Musicianship Cambridge: MIT Press. March 2001 ISBN 0-262-18296-8, pagg.343-353

[6] Tarabella L., Magrini M., Scapellato G., "A system for recognizing shape, position and rotation of the hands" in in Proceedings of th Internationl Computer Music Conference '97 pp 288-291, ICMA S.Francisco, (1997)

[7] Tarabella L.: "Five minutes for Joseph Beuys", for electronic viola and Twin Towers – 49a Biennale di Venezia, Platea dell'Umanità, june 2001

[8] Tarabella L.: "Kite: for copper wire and TwinTowers" – Perfomance commissioned by the Guglielmo Marconi Foundation and Assindustria Bologna, for the centenary of the first transatlantic radio-telegraphic transmission. – 26 March 2002, Sala Farnese, Bologna and 25 april 2001, Sasso Marconi.

[9] Tarabella L.: "Suite for M" for TwinTower and Imaginary Piano – NIME2002, Media Lab, Dublin. May 25th, 2002.

[10] Tarabella L., Bertini G., Boschi G. : "A Data Streaming Based Controller For Real-Time Computer Generated Music" Proc. Int'l Symposium on Musical Acoustics ISMA 2001, Perugia, Italy, sept.10-14 2001, Vol. 2 pp. 619-622.

[11] Tarabella L., Bertini G., "The mapping paradigm in gesture controlled live computer music" in Procs of "2nd Conference Understanding and Creating Music", Caserta, November 2002, Seconda Università di Napoli.

SOUNDS OBTAINED VIA ELLIPTIC FUNCTIONS THEORY

Vittorio Cafagna and Domenico Vicinanza***

**DMI - University of Salerno, Italy*, cafagna@unisa.it
***Departiment of Phisics - University of Salerno, Italy*, vicinanza@sa.infn.it

ABSTRACT

Synthesis by means of two-variables functions has been investigated in many different ways (see, e.g., [1], [2]) since the pioneering article by Mitsuhashi ([3]). The usual point of view is to choose a surface, a closed curve on the surface, and then to produce a waveform sampling the function on the curve. Our aim in this paper is to take up this method of synthesis with an attitude towards systematic exploration. Instead of choosing a single surface, or a small collection of surfaces, we decided to consider an entire class of functions and to start an organized investigation of the sounds than can be generated using the above approach. The functions we choose are the elliptic functions. The reasons for the choice are manifold. First of all, elliptic functions are doubly-periodic, so it seems natural to expect an interesting behaviour from a sonic point of view. Second, the theory of elliptic functions is one of the highlights of the 19th century mathematics, mainly the creation of the great mathematicians Gauss, Abel, Jacobi and Weierstrass. It is a beautiful theory, rich in formulas connecting in many different ways different elliptic functions. Elliptic functions are not a mere collection of objects, but a highly structured class. In particular it is noteworthy that any two elliptic functions are related by an algebraic relation (a property which is obviously false for trigonometric functions: think of e^z and e^{e^z}), enabling thus the use of algebra to buid a sort of additive synthesis.

1. ELLIPTIC FUNCTIONS

A function $f : \mathbb{C} \to \mathbb{C} \cup \{\infty\}$ is doubly periodic with periods ω_1 and ω_2 if ω_1 and ω_2 are linearly independent over $\mathbb{R}$ and if $f(z + \omega_1) = f(z) = f(z + \omega_2)$ for all $z \in \mathbb{C}$. An elliptic function is a meromorphic doubly periodic function. For ω_1 and ω_2 fixed, the corresponding elliptic functions form a field. The parallelogram with vertices $0, \omega_1, \omega_2, \omega_1 + \omega_2$ containing the sides adjacent to the origin, but not containing the other two, is called the fundamental parallelogram Π. Any translate $\alpha + \Pi, \alpha \in \mathbb{C}$, is called a period parallelogram. For any period parallelogram $\alpha + \Pi$, any point in $\mathbb{C}$ is congruent modulo the lattice $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ to one and only one point of $\alpha + \Pi$. Therefore one can think of elliptic functions as lattice-periodic functions.

Here follows a list of elementary properties:

1. A non-constant elliptic function has at least one pole in any period-parallelogram.
2. A non-constant elliptic function connot have just one simple pole in a period-parallelogram.
3. The number of zeros of a non-constant elliptic function, on a period-parallelogram, is equal to the number of poles in it (counting multiplicity). The number of poles in a period-parallelogram is called the *order* of the elliptic function.
4. A non-constant elliptic function of order h assumes, in a period-parallelogram, every complex value h times (taking multiplicity into account).

2. THE JACOBI FUNCTIONS $sn\ u$ AND $cn\ u$

Let

$$u = \int_0^x \frac{dt}{\sqrt{1-t^2}\sqrt{1-k^2t^2}} \tag{1}$$

and

$$K = \int_0^1 \frac{dt}{\sqrt{1-t^2}\sqrt{1-k^2t^2}} \tag{2}$$

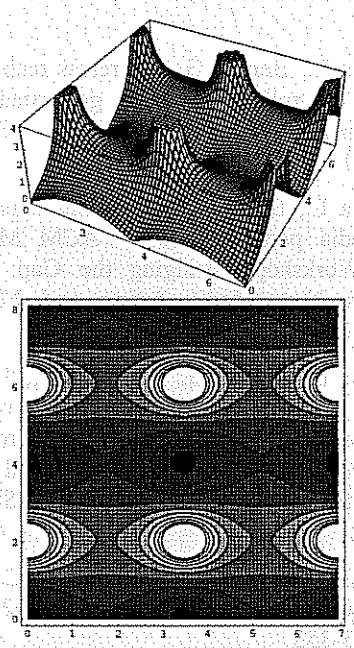


Figure 1: 3D plot and contour plot of the absolute value of Jacobi's $sn\ u$ function.

Suppose that x and K are real and $0 < k < 1, -1 \leq x \leq 1$, and $\sqrt{(1-t^2)}$ and $\sqrt{1-k^2t^2}$ are positive. Equation (1) defines u as an odd function of x, increasing from 0 to K as x increases from 0 to 1. Conversely, the same equation allows us to define x as an odd function of u which increases from 0 to 1 as u increases from 0 to K. This last function is called the Jacobi $sn\ u$ function (fig. 1) ([4]). Therefore we write:

$$u = sn^{-1}\, x \tag{3}$$

and

$$x = sn\, u \tag{4}$$

Following the example offered by trigonometry we can define $\sqrt{1 - sn^2\, u}$ and call it $cn\ u$:

$$cn\ u = \sqrt{1 - sn^2\, u} \tag{5}$$

It is easy to prove that

$$sn^2\, u + cn^2\, u = 1 \tag{6}$$

One can also observe that $sn\ u$ is and odd function of u, while $cn\ u$ is an even function of u. The analogy with sinus and cosinus is evident.

3. WEIERSTRASS'S ELLIPTIC FUNCTION $\wp(Z)$

The Weierstrass elliptic function $\wp(z)$(fig. 2) is defined as follows ([4]) :

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L'} \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right] \tag{7}$$

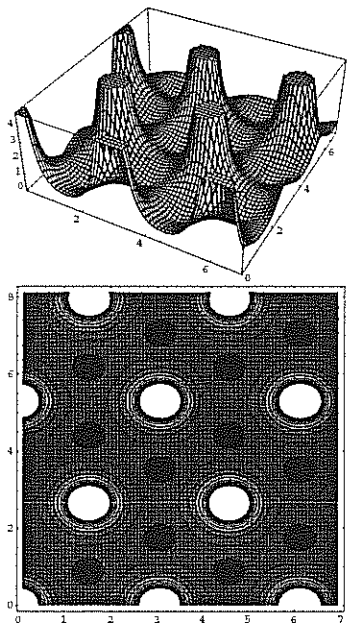


Figure 2: 3D plot and contour plot of the absolute value of Weierstrass's $\wp(z)$ function.

where the sum is taken over the set of all non-zero periods, denoted by L'. This series converges uniformly on compact sets not including the lattice points. The expression of the first derivative of $\wp(z)$ is

$$\wp'(z) = -2 \sum_{\omega \in L'} \frac{1}{(z-\omega)^3} \tag{8}$$

We list the main properties of $\wp$ and $\wp'$:

1. The poles of $\wp(z)$ are given by $z = \omega$.
2. The principal part of $\wp(z)$ at $z = 0$ is $1/z^2$.
3. $\wp(z) = \wp(-z)$
4. $\wp'(z) = -\wp'(-z)$

$\wp(z)$ satisfies the Weierstrass differential equation

$$\wp'^2(z) = 4\wp^3(z) - g_2\wp(z) - g_3 \tag{9}$$

where

$$g_2 = 60 \sum_{\omega \in L'} \omega^{-4}, \quad g_3 = 140 \sum_{\omega \in L'} \omega^{-6} \tag{10}$$

What makes elliptic functions specially interesting for sound synthesis and analysis is the following theorem and its corollary:

Theorem. *Any even elliptic function is a rational function of $\wp(z)$. Any elliptic function $f(z)$ can be written uniquely in the form*

$$f(z) = g(\wp(z)) + \wp' h(\wp(z))$$

where g and h are rational functions.

Corollary. *There exists an algebraic relation between any two elliptic functions having the same periods. In particular, any elliptic function is connected with its derivative by an algebraic relation.*

4. CHOICE OF THE PARAMETERS AND COMPUTATIONS

In a sense, Weierstrass's $\wp(z)$ and its derivative $\wp'(z)$ are very much like two-variables analogues of cosinus and sinus. They are doubly periodic. Every elliptic functions can be expressed by means of them. The theorem recorded in the last section could be intrepreted as a sort of Fourier's theorem. Therefore the sound synthesis method that we propose can be interpreted as a nontrivial version of additive synthesis.

Orbits play a very important role in n-variable synthesis, their shape and geometrical characteristics affecting the final waveform. Close orbits will return periodic waveforms and open orbits (such as spirals) time-evolving sounds.

It might be worthwile to give some experimental corroboration to the observation that Jacobian functions are similar to trigonometric functions. Fixing a variable in an elliptic function, one obviously obtains a simply-periodic complex-valued one. Taking real or imaginary parts, or absolute values, gives a real periodic function of a real variable. It is enough to compare graphs of absolute values of Jacobi's functions restricted to one variable with those of trigonometric functions, to realize that $sn\ u$ and $cn\ u$ (called appropriately *sinus amplitudinis* and *cosinus amplitudinis*) are a sort of slightly distorted versions of sinus and cosinus. Distortion being more evident near the pole.

To put the above considerations in a geometric framework, one can view elliptic functions as functions on a two-dimensional torus. On the torus one can distinguish between two types of closed curves: curves which can continuously shrinked to a point, and curves which cannot. Restriction to simplest kind of curves of the second type, meridians and parallels, essentially gives classical periodic functions, while consideration of closed curves of the first type gives rise to completely new and unexplored phaenomena. One can also consider not-so-simple curves of the second type, namely curves winding around the toric surface *many times* (long period orbits). This gives rise to a concept of frequency, different from the one determined by the speed at which the curve is parametrized. This parameter can be represented geometrically as a slope on a period parallelogram. It might be interesting to define still another concept of *frequency*. Consider the function $\wp(kz), k \in \mathbb{N}$ (or any other elliptic function $f(kz)$). By the main theorem, one can write $\wp(kz)$ (or any other elliptic function $f(kz)$) as a rational function of $\wp$ and $\wp'$. By existing (complicated) multiplication formulas ([5]), one can explicitely write down these rational functions. One is naturally tempted to consider these expressions as elliptic analogues of Chebychev polynomials and try to understand the effect of such hierarchy of rational functions on the whole class of elliptic functions (elliptic nonlinear distortion). Investigations are in progress.

Experiments are carried on with Mathematica (both on Mac OS X and SGI-IRIX platforms). The procedure is:

(i) compute the graphs of a realization of an elliptic function (e.g. absolute value, real or imaginary part). Produce 3-dimensional plots and contour plots;

(ii) define an orbit on the torus;

(iii) compute the realization of the elliptic function on the orbit to obtain the waveform;

(iv) play the waveform;

(v) produce and store sound files.

The attitude is towards a systematic exploration of the sound quality of different elliptic functions, exploiting the different possible formulas. For example the "additive synthesis" formula offered by the main theorem: starting with $\wp(z)$ and $\wp'(z)$, to test rational functions of increasing degree:

$$\begin{array}{lcl} \wp(z) & + & \wp'(z)\wp(z) \\ \wp^2(z) & + & \wp'(z)\wp(z) \\ \wp^2(z) & + & \wp'(z)\wp^2(z) \\ \vdots & & \end{array}$$

5. REFERENCES

[1] Borgonovo A. and Haus G., Musical sound synthesis by means of two-variable functions: experimental criteria and results, *Computer Music Journal* 10(4) (1986) 57-71

[2] Roads C., *The Computer Music Tutorial*, MIT Press, 1996.

[3] Mitsuhashi Y., "Audio signal synthesis by functions of two variables", *Journal of the Audio Engineering Society* 30 (10)(1982) 701-706.

[4] Chandrasekharan K., *Elliptic Functions*, Springer-Verlag, 1985.

[5] Tannery J. et Molk J., *Eléménts de la Théorie des Fonctions Elliptiques*, Paris, 1893-1902.

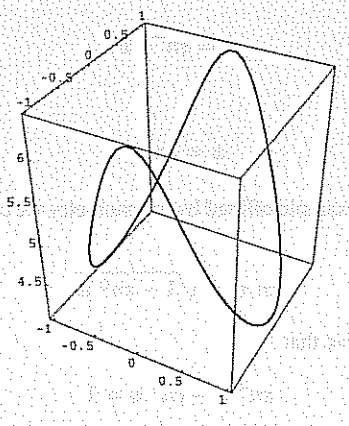


Figure 3: The Jacobi's $sn\ u$ function computed on the orbit: $x(t) = \cos(\pi t), y(t) = \sin(\pi t)$.

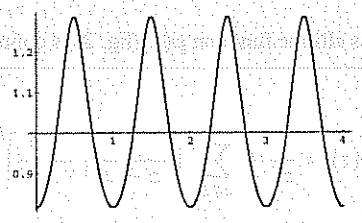


Figure 4: Waveform corresponding to the orbit of fig. 3.

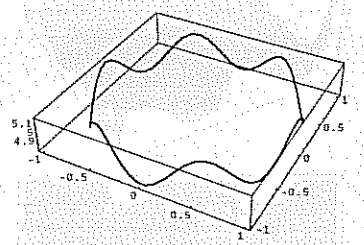


Figure 5: The Weierstrass's $\wp(z)$ function computed on the orbit: $x(t) = \cos(\pi t), y(t) = \sin(\pi t)$.

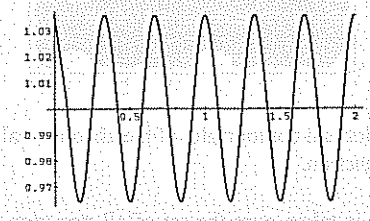


Figure 6: Waveform corresponding to the orbit of fig. 5.

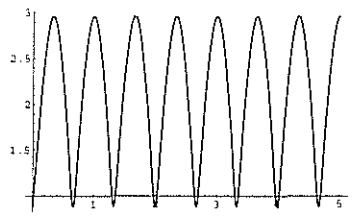


Figure 7: Waveform on the absolute value of the function $\wp(z) + \wp'(z)\wp(z)$ computed on the orbit: $x(t) = \cos(\pi t)$, $y(t) = \sin(\pi t)$

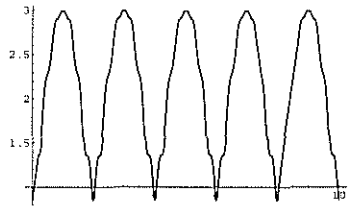


Figure 8: Waveform on the absolute value of the function $\wp^2(z) + \wp'(z)\wp(z)$ computed on the orbit: $x(t) = \cos(\pi t)$, $y(t) = \sin(\pi t)$

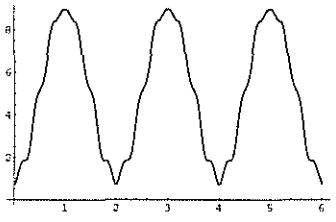


Figure 9: Waveform on the absolute value of the function $\wp^2(z) + \wp'(z)\wp^2(z)$ computed on the orbit: $x(t) = \cos(\pi t)$, $y(t) = \sin(\pi t)$

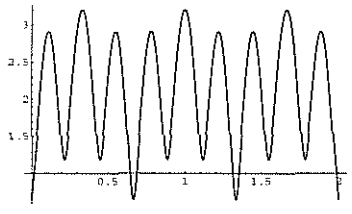


Figure 10: Waveform on the absolute value of the function $\wp^2(z) + \wp'(z)\wp(z)$ computed on the orbit: $x(t) = \cos(\pi t)$, $y(t) = \sin(\pi t)$

GenORCHESTRA: A MUSICAL BLIND WATCH MAKER

[1]Fabio De Felice, [1]Fabio Abbattista, [2]Francesco Scagliola

[1]VALIS Group - Dipartimento di Informatica - Università di Bari
febo74@libero.it, fabio@di.uniba.it

[2] Conservatorio di Musica "N. Piccinni" - Bari
fscagliola@libero.it

ABSTRACT

In this paper we present GenOrchestra, a project involving the Dipartimento di Informatica and the Conservatorio di Musica "N. Piccinni" in Bari. This project regards a Creative Evolutionary System, based on Evolutionary Computation (EC) techniques, applied to the field of western tonal music. The system is intented a useful tool for learning and widening the musical knowledge of a wide range of users, from novices to experts.

1. INTRODUCTION

GenOrchestra is a project aimed to the development of an e-learning web system applied to the field of western tonal music. In particular, it will provide the following functions: Automatic tunes composition, autonomous user's tunes evaluation, users evaluation of auto-composed or users let in tunes, suggestions upon variations and widening on users, and a *bibliographic area*.

The GenOrchestra (**Gen**etic **Orchestra**) sound engine is based on a modified version of the standard Genetic Algorithm (GA) [1]. GenOrchestra employes a novel way to evaluate the produced tunes, respect to the traditional ones adopted by other EC-based systems (GenJam [2], Conga [3], Vox Populi [4]): Indeed we adopt a hybrid solution composed of two kinds of fitness functions. The **technique fitness** evaluates the consonance degree between melodic, harmonic and rhythmic sections, moreover, it defines how well the rhythmic paths is organized into a coherent musical event. The second fitness function, called **human fitness**, determines how well the tunes are perceived from a human audience. The ultimate goal of this project, currently in progress, is the development of a human-like composer able to produce music in any musical genre, and to show a "personal style".

2. EVOLUTIONARY ALGORITHMS

Evolutionary Algorithms (EA) represent a family of problem solving techniques, inspired from natural evolution. In this section we will provide a brief description of Genetic Algorithms, as they represent the core of the GenOrchestra system.

The idea behind GA's is to exploit *Darwinian Evolution*, and transform it for application in mathematical optimization theory to find the global optimum in a defined *phase space*.

The three fundamental principles are: **Selection**, **Mating/Crossover**, and **Mutation**.

Selection means to extract a subset of genes from an existing population, according to any definition of *quality*. In fact, every gene must have a *meaning*, so one can derive any kind of a *quality measurement* from it, a "value". Following this quality "value" (*fitness*), Selection can be performed e.g. by Selection *proportional to fitness*:

1. Consider the population being rated that means: *each gene has a related fitness*. The higher the value of the fitness, the better.
2. The *mean-fitness* of the population will be calculated.
3. Every individual (=gene) will be copied as often to the new population, the better it fitness is, compared to the average fitness. All genes with fitness at the average and below will be removed.

The next steps in creating a new population are the **Mating** and **Crossover**: There exist also a lot of different types of Mating/Crossover. The most used is the single-point crossover (see Figure 1) in which, having chosen a random point in the individuals to be mated, the information after the crossover-point will be exchanged between the two individuals of each pair.

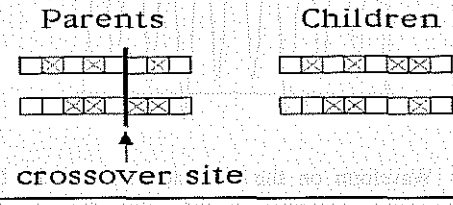


Figure 1: An example of the Crossover

The last step is the **Mutation**, with the sense of adding some effect of *exploration* of the phase-space to the algorithm. The implementation of Mutation is fairly trivial: *Each bit* in *every gene* has a defined Probability *P* to get inverted.

The main schema of a GA is the following:

g:=0 { generation counter }
Initialize population P(g)
Evaluate population P(g) {compute fitness values }

while not done do
 g:=g+1
 Select P(g) from P(g-1)
 Crossover P(g)
 Mutate P(g)
 Evaluate P(g)
end while

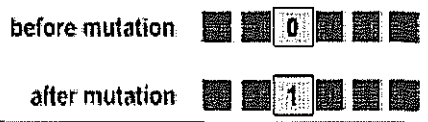


Figure 2: An example of the Mutation

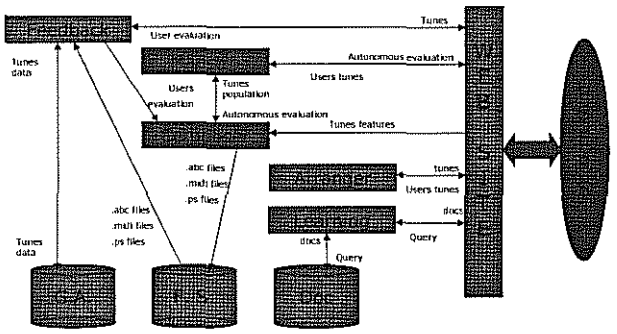


Figure 3: The general architecture of GenOrchestra

3. THE SYSTEM'S ARCHITECTURE

The GenOrchestra underlying evolutionary process has to be an Open-Ended one, i.e. a continued evolution to relative maximum points without nor temporal ending either evolution to a single absolute maximum point. The main feature of the system is its generality concerning the faceable musical genres and the capacities of self-judging the composed tunes by evaluating the melodic, harmonic and rhythmic qualities based on the ordering of consonance of musical interval. The aforesaid evaluations procedures are integrated with the web surfers and expert human composers evaluations, through the GenOrchestra site.
The architecture is composed of six main modules (Figure 3):

- **Composer**: Handles the system compositional process. This module inputs the composed tunes and receives their evaluations.
- **Maestro**: Embeds the overall consonance fitness evaluations.
- **Feedback**: Is responsible for the human evaluation of composer tunes.
- **Arranger**: Takes the user tunes, submitted via web, and applies musical transformations in order to arrange the musical materials.
- **Learning**: Handles the pure documentation side of the whole system.
- **Web Site:** Makes up the system Internet interface through which users can interact with the system.

4. THE CHROMOSOME FORMAT

GenOrchestra is based on a Steady-state GA with tournament selection and multi-cut points crossover. In a Steady-state GA, every new population presents an overlapping between the old and the new generations. In the tournament selection the population is grouped in several individual families, the best two individuals of each family mate with the crossover operator and the new solution substitutes those in the family with worse fitness. GenOrchestra applies a multi cut points operator. In our implementation, a cut point is chosen for every section in the tune structure and for every layer. The cut point has not been chosen at low level (such as the note or the chord), but at the measure level to avoid the production of new individuals with length measure that differs from the parents one.
In the genetic algorithms, the representation of a solution is usually called the chromosome. The goal of GenOrchestra is to continually evolve populations of tunes, so the chromosomes have to reflect the structure of a musical piece. We can simplify a piece of music as a significant set of sections, differing each other by the melodic theme and possible scale, time and beat variations. Every section can be repeated in the tune execution so if we have three sections, A, B and C, then the structure can be any disposition with repetition of these three sections. Furthermore a tune has some initial features such as: Scale, beat, note unit length (eight note, half note etc.) and playing tempo of the note unit length. These values can be different in a given section and from section to section. Every section is made up of a certain number of measures; a measure is a not unique set of notes where the overall length is equal to the beat value. What we hear in a piece of music is, usually, made up of three sonorous layers: a bass layer, a harmonic layer and a melodic layer. In the first layer we have the bass score, in the second layer we have the chords score and in the last one the tune theme or solo score. So the chromosome is defined as an array made up with so many components as the sections in the structure, each of these components points to a three-layered structure containing the aforesaid scores. The user can initialize the chromosome by setting the following tune characteristics:

- The tune structure, a string of sequential sections (ABABC, etc.).
- The number of measures per section.
- The initial beat.
- The initial scale.
- The tempo.

Moreover any of these features can be generated by the system automatically. The chromosome generation starts from the initial scale; on the ground of this scale the initial chord is built up with a random generated length that cannot exceed the length of the measure.

On this first chord a melody with the same length and bass score is generated, then a melodic note is randomly chosen for a new chord with the same initial scale. This process is repeated till the length of the measure is reached; then it is repeated for the number of measures of the current section and for any section of the tune (see fig. 4).

On the ground of the described chromosome, we decided to adopt the same approach as in [4] for the mutation phase that is musically meaningful mutations operators that work at measure level. These operators

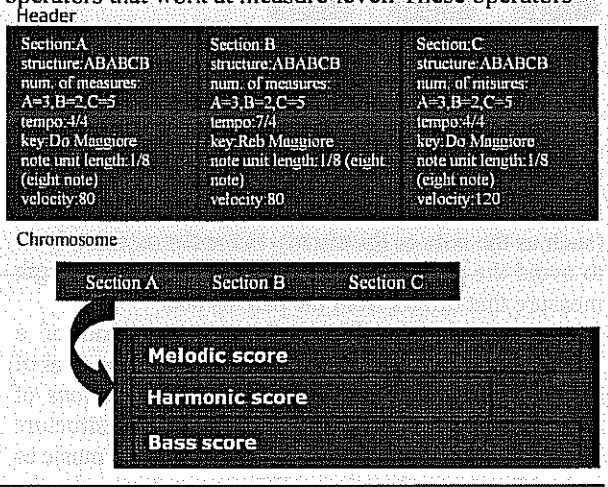


Figure 4: Chromosome structure for a three-section tune

implement classical compositions techniques. With a given mutation probability a chromosome is scanned and measures are chosen for the mutation in every layer. The mutation is randomly chosen among the following types:

- Transposition: Transposes notes and chords in a measure, by a random number of intervals in the given scale.
- Reverse: Reverses the events in a measure, rests included.
- Rotate–right: Rotates the events in a measure by a random number of positions to the right.
- Invert: Given an event in the measure, it evaluates the difference between the top position scale note (7) and the scale position of the current note.
- Sort up and sort down: sort the measure and preserve the rhythmic structure
- Invert-reverse: Given a measure, the invert and reverse operators are applied consecutively.

5. THE TECHNICAL EVALUATION

The fitness function used in GenOrchestra is a hybrid solution formed by an autonomous evaluation, judging the consonance qualities among melodic, harmonic and bass layers and a human evaluation for the aesthetic qualities. Actual systems implements the fitness phase by two ways:

a) Delegating the individual evaluations to the human ears: this leads to great human-like musical production, but make up a heavy bottleneck for the system and a dull work for the human judge.
b) Adopting autonomous solutions like neural networks [5], implementing physiological aspects in listening music [4] and completely removing the fitness phase [6].

Both solutions do not fit the GenOrchestra general purposes but the former resolve the drawbacks of the latter. So, a hybrid solution seems to be a good alternative best reflecting real world situations.

Indeed, the GenOrchestra consonance fitness function is: *Consonance fitness = melodic-harmonic consonance + harmonic consonance + bass-melodic consonance + bass-harmonic consonance.*

To develop the fitness function we start from the fuzzy approach described in [4,7], where a consonance measure among notes in a chord was defined and we extend it to a consonance measure of notes over chords. By means of this approach we can represent a note as a compound tone consisting of its fundamental tone and upper harmonic series tones. It can be represented as a fuzzy set in which the membership degree of a given tone is proportional to its amplitude. Finally, a note is a fuzzy set made up of couples (x, y) in which x is a tone (also called partial), and y is the related weight in the note, corresponding to its amplitude. We can now define the consonance between two notes Sm and Sn as follows:

$$Co(S_m, S_n) = \sum_{(x,y)\in S_{m\cap n}} y \qquad (1)$$

The consonance measure between two notes is intended as the sum of the intersection of the partials weights, in the range [0,..,1]. Starting from this concept we have defined a set of evaluations to carry out the overall consonance of the tune. The consonance score functions are: note-chord consonance, chord-chord consonance, melodic-harmonic consonance; bass-harmonic consonance, melodic-bass consonance, harmonic consonance and total consonance.

Lets now describe the basic concepts and the resulting function for the rhythmic evaluations of the tunes composed. Listening to a piece of music we naturally organize the sound signals into meter groups. Furthermore we infer a regular pattern of strong and weak beats to which relate the actual musical sounds. GenOrchestra evaluates these patterns to judge how well the tune match the metrical structure defined by the starting meter input. It must be emphasized that beats do not have duration, and we can think about it as an idealization, utilized by the performer and inferred by the listener from the musical signal. To use a spatial analogy: beats correspond to geometric points rather than to the line drawn between them. But, of course, beats occur in time so an interval of time takes place between successive beats. For such intervals we use the term time-span. For the aforesaid analogy, we can represent beat by dots.

The two sequences differ in a crucial respect: the dots in the first sequence are equidistant but not those in the second. The meter function is to mark off, insofar as possible, into equal time-spans, this disqualifies the b)

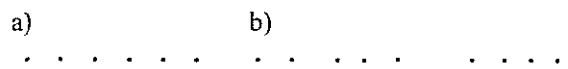


Figure 5: beat sequences example

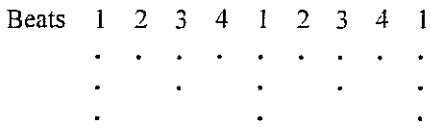


Figure 6: Metrical structure for a 4/4 meter

sequence from being called metrical. Another aspect of meter is the notion of periodic alternation of strong and weak beats, in a) sequence no such distinction exists. For beats to be strong or weak there must exist a metrical hierarchy. The relationship of strong beat and metrical level is simply that, if a beat is felt to be strong at a particular level, it is also a beat at the next larger level. This is shown in figure 6.

So given a note unit length and a starting meter we can built up the relative metrical structure as follows.

Define a first note unit length level formed by a number of beats calculated by the expression:

$$beats = round(\frac{num\_val * val\_mov}{unit}) \qquad (2)$$

Where *beats* is the number of beat per level, *num_val* is the number of movimento per meter, *val_mov* is the value of the meter, *unit* is the note unit length and finally the *round* function round up the ratio result to the next integer, if it is a float. While *beat* is greater or equal to 1 duplicate the *unit* and then calculate the (2) again. We refer to this structure as **perfect metrical structure.** Given this structure, the metrical patterns for every measure in a given tune defines how many pitches start time occur in a given time-span. We refer to this pattern as the **actual metrical structure**. Then the closer the **actual metrical structure** to the **perfect metrical structure** the better the evaluation.

6. THE HUMAN EVALUATION

So far we discussed about the automatic evaluation of the tune features. Nevertheless, a musical composition should be evaluated for its aesthetic qualities. GenOrchestra integrates the GA fitness function with the human evaluations of the produced tunes. The tunes produced by the GA are made available on the GenOrchestra Web site and, users accessing the site can listen them and provide their own scores. Users scores are averaged, for each of the evaluated tunes, and are summed up to the GA evaluations. The involvement of human users is an effective solution to the subjective evaluation of the tunes, but, on the other hand, it represents a bottleneck of the GenOrchestra system, due to the time consuming. To overcome this limitation, we assigned a fixed interval of time for the user evaluation of each generation of tunes. Indeed, not the whole population available on the Web will receive a user evaluation for the aforesaid drawbacks. This lead to a speciation of the initial population P after the evaluation phase: the population P_u made up with individuals evaluated autonomously and via web and the population P_t of autonomously evaluated tunes passed unseen on the web. As a consequence, each run corresponds to two separate evolutionary processes allowing the selection operator to work on individuals with comparable and homogeneous fitness. The separated populations merge into one after the mutation phase, ready for a new iteration of the GA.

7. CONCLUSIONS

In this paper we described a prototype of an evolutionary based system able to autonomously produce tunes presenting a good consonance degree, as confirmed by a human expert. Comparing GenOrchestra with other creative evolutionary systems we can conclude that it is a more complete system because of its goal (to generate a complete tune). However, a main weakness of the system is that the produced tunes do not yet correspond to a really human-like composition and a human composer is still needed to arrange the musical output into a finished thematic development. Further development will be:

- A user tunes consonance evaluation module, by which the system evaluates human composition with the aforesaid function, so an evolution to reach that value can be made.
- Formalization of a given musical genre in order to make the system able to compose in a given style, without any extensive knowledge.

8. REFERENCES

[1] J. Holland, "Adaptation in natural and artificial systems", Ann Arbor: The University of Michigan Press, 1975.

[2] J. Biles, "GenJam: A Genetic Algorithm for Generating Jazz Solos", Proc. of the International Computer Music Conference, 1994.

[3] N. Tokui, H. Iba, "Music Composition with Interactive Evolutionary Computation", Proc. of the Generative Arts Conference, 2000.

[4] A. Moroni, J. Manzoulli, F. Von Zuben, R. Gudwin, "Vox Populi: Evolutionary computation for music evolution", in P.J. Bentley, D. W. Corne, "Creative evolutionary Systems", Morgan Kaufmann, 2002, pp. 205-221.

[5] J. Biles, P. G. Anderson, L. W. Loggi, "Neural Network Fitness Functions for a Musical IGA", Proc. of the International ICSC Symposium on Intelligent Industrial Automation (ISA'96) and Soft Computing (SOCO'96), 1996, pp. B39--B44.

[6] J. Biles, "Autonomous GenJam: Eliminating the Fitness Bottleneck by Eliminating Fitness", Workshop on Non-routine Design with

Evolutionary Systems, Genetic and Evolutionary Computation Conference (GECCO01), 2001.

[7] G. Vidyamurthy, J. Chakrapani, "Cognition of tonal centers: A fuzzy approach", Computer Music J. Vol. 16, n° 2, 1992, pp. 45-50.

TOWARDS A SPECIFICATION OF MUSICAL INTERACTIVE PIECES

Myriam Desainte-Catherine *Nicolas Brousse*

SCRIME, LaBRI, université Bordeaux 1
351, cours de la Libération
33405, Talence Cedex, France

myriam@labri.u-bordeaux.fr brousse@labri.u-bordeaux.fr

ABSTRACT

We propose a formalism for specifying interactive musical pieces. In this paper, we focus on interpretation of musical pieces that are structured as temporal hierarchies.

1. INTRODUCTION

More and more musical creations involve interactivity during live performance. The conception of such musical pieces necessitates various levels of specification. Interactivity is often specified by the means of bindings between inputs coming from sensors, or gesture analysis, from one part, and musical parameters from the other part. Those musical parameters can be either parameters for sound synthesis, either attributes of low level objects, like sounds, or higher level objects like melodies, musical parts, and so on. Thus, composers use interactivity in various manners and provide very different musical pieces examples.

There is a need for a formalism to simplify the specification of interactive pieces. Such formalism should lead to an operational model providing:

- help the composer to specify bindings between input and output events, from sound parameters to musical ones.
- easy maintenance and diffusion, by adapting a creation to various types of sensors or gestures. It should be possible to substitute a gesture by another one in case of substitution of sensors, for example. Thus a musical piece should be able to evolve with technology.
- to adapt easily a piece to every musician, whatever his skill is. The musician may be a child [DCK02], a handicap person [HS98] or a maestro for a given sensor.

In order to reach all these objectives, a flexible and powerful model is needed. We propose in this paper abstract machines as formal basis for this model. This formalism comes from functional language interpretation (lisp language) and provides a way to specify formally the execution of a program depending on its instructions and its environment. We found out that a structuration of the musical material is necessary in order to provide a comfortable way to specify various levels of interactions (either sound or musical ones). Moreover, since those levels become explicit, they can be changed dynamically.

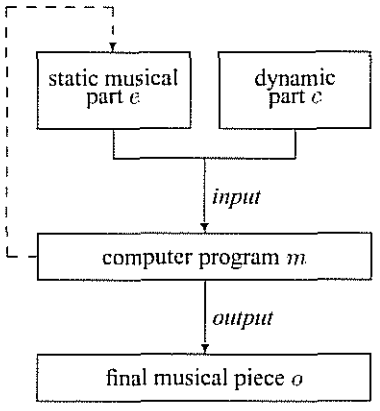
2. INTERPRETATION OF MUSICAL PIECES

Various levels of interactivity have to be taken into account. The lower level consists in just pushing a button in order to play a musical piece without any more interaction. The highest one consists in providing all data necessary for sound synthesis and musical structures. This level is the objective of many electroacoustic instruments like the *metainstrument* of Serge Delaubier or the *Esher system* [MWR98]. But they are generally more sound-oriented than music-oriented. Between these two extremities, we have the level of interpretation. Interpretation of musical pieces based on the operations of activating and releasing notes has been very well studied by Jean Haury [Hau]. In this case, the piece is entirely written, and the musician can activate musical events. He has the choice of the starting dates and the velocity of the events.

In this paper, we focus on interpretation of interactive musical pieces. That is, we consider that the piece is completely written and that the interaction is limited to the interpretation. This limitation permits to focus on the temporal aspects, putting aside (for now) sound synthesis.

3. DEFINITIONS

Interactivity process involves a static musical part e that is defined before performance, a computer program m which is intended to execute during the performance, a dynamic part c that correspond to the input of m during performance and a resulting piece o that is the piece which is finally heard.



For a given couple $< p, m >$, several resulting pieces can generally be obtained, depending on the input. Thus, we define in

this paper a musical interactive piece as an abstract machine called ECO which is defined in the following subsections.

3.1. ECO Machine

An ECO machine is an abstract machine such that:

- a state of the ECO machine is a 4-tuple (E, C, O, t) where:
  - E is an environment, which represents the static musical material;
  - C is a control string representing input time-stamped events;
  - O is the output string;
  - t is the time-stamp of the state.
- the operation of the machine is described in terms of state transitions that are synchronized on a clock. The first state is associated to the initial date 0. Let δt be the value of a cycle of the clock, transitions occur at that rate. Given the current state (E, C, O, t), the next state $(E', C', O', t+\delta t)$ is determined by the events of the current control string C whose time-stamp are greater than t and lower than $t + \delta t$. There are several cases to consider, depending on the type of the input event.

3.2. ECO States

Let us specify in this subsection the components of the ECO states.

Musical Environment E

We consider that the musical material is structured in a temporal hierarchy [BDC01]. The musical environment E can then be represented as a couple $< M, N >$ where M is the musical material and N a set of nodes of M. Leaves are sound material expressed in various models (temporal, spectral, etc.). Internal nodes admit a temporal operator of specific type: concatenation, superimposition, musical concatenation (see Balaban [Bal89, BN98]), and so on.

In this paper, we focus on the temporal organisation of the musical material. Since sound representation is another subject and it is not investigated here. Thus, let us consider that each node of the musical hierarchy is represented by a tuple:

$$n = < t, s, d, p >$$

where s is the starting date of the node n and d is the duration. Other elements of the tuple depend on whether the node n is a leaf or not:

- if n is a leaf, t indicates the model of the sound, and p is a list of parameters that are needed for sound synthesis.
- if n is an internal node, t indicates the temporal operator and p is the list of children of n.

The environment is closed. That is, every parameter is initialized to a default value, so that the musical material can be played as a complete piece without any interaction.

Since all durations of the environment M have a value, it is possible to associate to a date t a set $a(M, t)$ of active nodes of M. It is a tree of nodes starting from the root of the environment down to the active leaves at the date t.

Let t be the date, the set $a(M, t)$ is defined recursively from M in the following way:

- the root of M belongs to $a(M, t)$.
- if n belongs to $a(M, t)$, then let $\{n'_1, ..., n'_k\}$ be the children of n such that $s(n'_i) < t < s(n'_i) + d(n'_i)$ for $1 \leq i \leq k$, then n'_i belongs to $a(M, t)$.

Then, N is included in $a(M, t)$. The nodes of N are called control nodes. This set is initialized at the beginning of performance and is then modified dynamically by the means of commands from the set $\{UP, DOWN, LEFT, RIGHT, ...\}$. Those commands apply to the tree M and to a control node n of N. They move n to its parent, children, siblings according to the type of n and to the kind of command. Those commands provide a way to modify N during a transition, by, for example, replacing one note by its successor or predecessor in a melody, or by changing the level of interaction, from, for example, the level of a note to the level of a melody.

Control String C

Control string is a string of input time-stamped events which grows in an asynchronous way respecting to the abstract machine clock. Let S be a set of control symbols. If the symbol x belongs to S, then the symbol $\bar{x}$ belongs also to S. The projection of the control string on the set S must be a Dyck word. Then, we distinguish between two types of control symbols: continuous controls and discrete controls. The study of this paper is limited to discrete controls.

Let x be a symbol belonging to S, then, in the control string C, it is followed by a list of real parameters p coming from the input device and between two symbols x and $\bar{x}$, C admits any kind of control symbols.

A list of commands is associated to each symbol, acting on the set of control nodes N. Those commands are executed when the symbol is read from the control string C, and apply to the sets M and N of the current state. One command consists in building the string that has to be concatenated to the output string:

BUILD(symbol, input-parameters, N, t)

Other ones modify the set N. They are obtained by the function:

COMMANDS(symbol, N)

Output O

O is the empty string at the initial state. Then, each transition concatenates a string computed from E and C at the end of O. The control string is thus a mixture of strings that are dedicated to specific devices: sound devices, MIDI events or whatever.

3.3. ECO Transitions

Let us take for example a musical material that is a simple melody of three notes. Thus, M is reduced to one internal node m being a concatenation and having three children a_1, a_2, and a_3. Let us suppose

$$a_1 =< t_1, s_1, d_1, p_1 >$$
$$a_2 =< t_2, s_2, d_2, p_2 >$$
$$a_3 =< t_3, s_3, d_3, p_3 >$$

Then the following temporal relations holds :

$$s_1 + d_1 \leq s_2, s_2 + d_2 \leq s_3$$

Let us suppose that the type of the sounds is MIDI : $t_1 = t_2 = t_3 = MIDI$ and let us consider the channel C and the velocity V as the two parameters of the notes and that they have the same value for the three notes $p_1 = p_2 = p_3 = \{C, V\}$.

We also consider that the set of control nodes N is reduced to one node at one time.

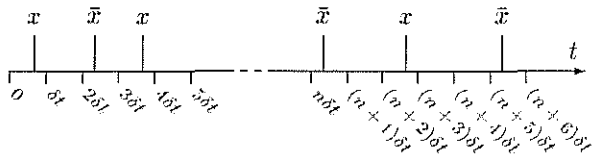
In this example, the physical access is reduced to one button that one can push and release. Thus only one discrete control is provided which is associated to the button : $S = \{x, \bar{x}\}$. The symbol x is sent when the button is pushed and the symbol $\bar{x}$ is sent when the button is released. In this example, we decide to ignore the parameters coming from the input device. The action of those symbols on the machine are defined as follows, where the symbol t denotes the date, and ON and OFF are the MIDI codes.

- BUILD(x, p, $\{n\}$, t) = $[ON, p(n), t]$, where $n = a_1, a_2$ or a_3, and p is the list of parameters of the node n;
- COMMANDS(x, $\{n\}$) = $\emptyset$, where $n = a_1, a_2$ or a_3;
- BUILD($\bar{x}$, p, $\{n\}$, t) = $[OFF, p(n), t]$, where $n = a_1, a_2$ or a_3;
- COMMANDS($\bar{x}$, $\{n\}$) = (RIGHT), where $n = a_1, a_2$;
- COMMANDS($\bar{x}$, $\{a_3\}$) = $\emptyset$.

In the case where the control node is m, the building of the output string is more complicated, because it is recursive: output strings of children have to be built by using the values of duration that are indicated in the nodes. This mechanism is not described formally here, but an example follows in this subsection.

Note-level example

Let us consider the following sequence of control symbols:



At the beginning, $t = 0$, O is empty and $N = \{a_1\}$. To simplify the presentation, we omit the list of real parameters coming from the input device with the symbol, because they are not taken into account during the transitions. Thus we obtain a sequence of transitions as follows:

$$(< M, \{a_1\} >, \emptyset, \emptyset, 0) \xrightarrow{\delta t} (< M, \{a_1\} >, x, \emptyset, \delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_1\} >, x, [ON, p_1, \delta t], 2\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_1\} >, x\bar{x}, < [ON, p_1, \delta t] >, 3\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_2\} >, x\bar{x}x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t] >, 4\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_2\} >, x\bar{x}x, < [ON, p_1, \delta t] >, [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t] >, 5\delta t)$$

$$\xrightarrow{\delta t} \ldots \xrightarrow{\delta t} (< M, \{a_2\} >, x\bar{x}x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t] >, n\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_2\} >, x\bar{x}x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t] >, (n+1)\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_3\} >, x\bar{x}x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t] >, (n+2)\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_3\} >, x\bar{x}x\bar{x}x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t] >, (n+3)\delta t)$$

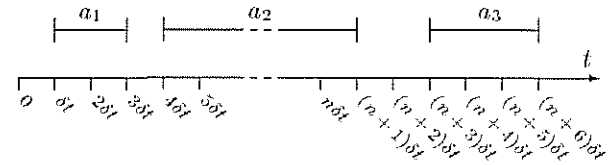
$$\xrightarrow{\delta t} (< M, \{a_3\} >, x\bar{x}x\bar{x}x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t] >, (n+4)\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_3\} >, x\bar{x}x\bar{x}x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t] >, (n+5)\delta t)$$

$$\xrightarrow{\delta t} (< M, \{a_3\} >, x\bar{x}x\bar{x}x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t], [OFF, p_3, (n+6)\delta t] >, (n+6)\delta t)$$

$$\xrightarrow{\delta t} (< M, \emptyset >, x\bar{x}x\bar{x}x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t], [OFF, p_3, (n+6)\delta t] >, (n+7)\delta t)$$

At last, the output sequence is the following.



Melody-level example

Let us consider the following sequence of control symbols: x between 0 and δt, and $\bar{x}$ between $(n + 5)\delta t$ and $(n + 6)\delta t$. At the beginning, $t = 0$, O is empty and $N = \{m\}$.

Let us suppose that the durations that are written are just the same than the ones that were played by hand in the preceding example. Then, the sequence that are input by the BUILD command on the node m will provoke exactly the same succession of transitions than we computed in the previous example. The transitions of the machine are the following ones (we have skip several transitions):

$$(< M, \{m\} >, \emptyset, \emptyset, 0) \xrightarrow{\delta t} (< M, \{m\} >, x, \emptyset, \delta t)$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x, [ON, p_1, \delta t], 2\delta t)$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x, < [ON, p_1, \delta t] >, 3\delta t)$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t] >, 4\delta t)$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x, < [ON, p_1, \delta t] >, [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t] >, 5\delta t)$$

$$\xrightarrow{\delta t} \ldots$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t], [OFF, p_3, (n+6)\delta t] >, (n+6)\delta t)$$

$$\xrightarrow{\delta t} (< M, \{m\} >, x\bar{x}, < [ON, p_1, \delta t], [OFF, p_1, 3\delta t], [ON, p_2, 4\delta t], [OFF, p_2, (n+1)\delta t], [ON, p_3, (n+3)\delta t], [OFF, p_3, (n+6)\delta t] >, (n+7)\delta t)$$

4. CONCLUSION

In this paper, we state that static musical material should be structured in order to clarify the correspondance between inputs and musical parameters. We have considered the temporal hierarchical organisation that permits to explicit the correspondance between static time and real time. Other kinds of hierarchical relations should be studied, involving other musical dimensions. For example, it would be interesting to provide the musician a way to dynamically move control from a sound object to its components or its parameters, or vice-versa.

A software based on the ideas that are presented in this paper is under development. Examples of the previous section are very simple. But several questions arise as we are going onto details. More complicated examples need several abstract machines that are interconnected in order to play automatically some parts of the musical piece (like in the second example) while events are still coming from input devices and modify the environment and the output string. For that purpose, it is necessary to structure several environments of execution. Thus, a more sophisticated model is needed in order to be able to specify all possible cases of interpretation of musical pieces. In such a model, continuous control should be taken into account. Next step consists in finding out suitable musical structuration for improvisation.

Acknowledgment

We would like to thank Jean Haury and György Kurtag for sharing with us their very precious experience with interpretation and improvisation.

This research was carried out in the context of the SCRIME[1] project which is funded by the DMDTS of the French Culture Ministry, the Aquitaine Regional Council, the General Council of the Gironde Department and IDDAC of the Gironde Department. SCRIME project is the result of a cooperation convention between the Conservatoire National de Région of Bordeaux, ENSEIRB (school of electronic and computer scientist engineers) and the University of Sciences of Bordeaux. It is composed of electroacoustic music composers and scientific researchers. It is managed by the LaBRI (laboratory of research in computer science of Bordeaux). Its main missions are research and creation, diffusion and pedagogy thus extending its influence.

5. REFERENCES

[Bal89] M. Balaban. Music structures: A temporal-hierarchical representation for music. *Musicometrika*, 2:1–50, 1989.

[BDC01] A. Beurivé and M. Desainte-Catherine. Representing musical hierarchies with constraints. In *Proceedings of CP'01, Musical Constraints Workshop*, 2001.

[BN98] M. Balaban and Murray N. Interleaving time and structure. *Computers and Artificial Intelligence*, 17(1):1–34, 1998.

[DCK02] M. Desainte-Catherine and G. Kurtag. La pédagogie de l'interactivité. In *conférence du 10e anniversaire de l'ESCOM, Liège, Belgique*, 2002.

[Hau] J. Haury. La grammaire de l'exécution musicale au clavier et le mouvement des touches. In *Manuscrit*.

[HS98] J. Haury and J. Schmutz. L'orchestre contre silence. In *Proc. of the JIM'98, Marseille, France*, pages D5–1, 1998.

[MWR98] N Schnell M. Wanderlay and J.B. Rovan. Escher - modeling and performing composed instruments in real-time. In *Proc. IEEE SMC'98*, pages 1080–1084, 1998.

[1] Studio de Création et de Recherche en Informatique et Musique électroacoustique, www.scrime.u-bordeaux.fr

BEAT AND RHYTHM TRACKING OF AUDIO MUSICAL SIGNAL FOR DANCE SYNCHRONIZATION OF A VIRTUAL PUPPET

Mario Malcangi, Alessandro Nivuori

LIM - Laboratorio di Informatica Musicale
DICO – Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano
malcangi@dico.unimi.it, nibou@tiscali.it

ABSTRACT

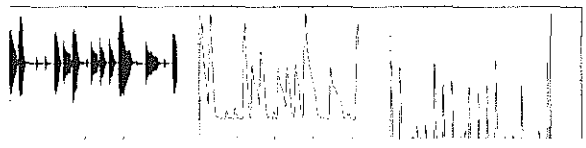
A system for extraction of metrical and rhythmic information from musical audio signal is described. The purpose is to drive a virtual 3-D dancer (puppet) directly from musical audio data.
Several approaches have been considered to synthesize a processing model robust enough to be extended to virtually any kind of music.
Scheires's and Sepännen's model has been chose as basic refers framework. An envelope extractor model has been developed to derive reduced data to be processed for onset detection.
A click-train signal is derived from a musical audio signal. This is the primary information to feed a comb filter bank resonator to detect the frequency of the beat.
A tatum-based search has been implemented to reduce comb filter bank computing complexity.
Fuzzy logic processing has been used to obtain a non linear model for accent detection.
A demonstration of the system is done using a virtual puppet droved to dance on baroque music.

1. INTRODUCTION

<<...Imagine that you are walking down a quiet city alley and enter a jazz club. The door opens and suddenly you hear the music. In just a second or two you have a strong impression of the "feel" of the music – in particular, the way it "swings" – its "beat" or "pulse"...>>

Richard Parncutt

Most people do not have any trouble to tap with their foot or clapping with their hand in time when hear a piece of music. It's an easy task, but not so easy to explain. This human activity is called **"foot tapping"** and the simulation of this activity with computer is called **"beat tracking"**.
Beat is the pulse of a musical piece, the same click that a musician follow with metronome when play a musical composition. When we talk about rhythm is the same as we talk about a pyramid and the beat is only the base of this pyramid.
Musical rhythm is a structure composed of accented beat and not accented beat, but we can consider more than one level of metrical structure.



A goal is to build a system that can understand more than one level of metrical structure from an acoustic signal and allow to a puppet or a virtual character to dance on it in time.

1.1. Previous Approaches

The extractions of metrical and rhythmic information from audio signal and/or symbolic representation of music have been a topic of active research in recent years [1]. Povel and Essen [2] proposed one of the first computational models of rhythm prediction.
They described an algorithm that could identify the clock, which a listener would associate with such a sequence of intervals, given a set of inter-onset intervals as input.
Parncutt [3] presents a model with a direct relationship between inter-onset intervals, durational accents, moderate tempo, and the perceived beat. In addition to the beat, Parncutt's model also estimates perceived meter, metrical accents, and expressive timing information.
Rosenthal [4] in his work called Machine Rhythm, formulated a symbolic meter analysis system for polyphonic music. The model compute an Inter-Onset Interval (IOI) histogram used for a first beat period prediction. The beat and its subdivisions and multiples are then used to construct the initial hypotheses for a multiple-agent algorithm. During the search of beat period, accentuation is attributed to onset events.
The Large–Kolen [5] model is one of the first models to use oscillator units for representing meter perception. When a non-linear oscillator unit is stimulated with a pulse train within its characteristic frequency range, it responds with synchronized pulsation. If we have more than one oscillator with different frequency, some of these synchronize with different metrical levels of a pulse input, while others fail to synchronize at all. The model is symbolic.
The algorithm proposed by Scheirer [6] is a signal-processing model of beat tracking from an acoustic

input signal. As the model of Large and Kolen, the beat tracking is carried out with an independent oscillator bank on each sub band is divided the audio signal, and the final beat tracking result is combined based on the energies of the sub band oscillators. Each sub band oscillator bank contains oscillators with identical characteristic frequencies, and the energies of identical oscillators are summed across bands. Scheirer introduced the idea of using comb filters as oscillator units, with the benefit that a comb filter oscillator will resonate at integral multiples of its characteristic frequency.

The model of Goto and Muraoka [7] process audio music signals by performing onset detection from the spectrogram of the audio signal. As Scheirer's model onset detection is performed independently on multiple frequency bands and the authors assign agents to operate strictly on the onsets coming from a specific frequency band. The agents compute an IOI histogram and determine the beat period based on it. In this model bass and snare drum are detected to distinguish strong and weak beat.

Dixon's model [8] is capable of processing acoustic musical signals in addition to symbolic data. The IOIs are clustered into a histogram-alike "class" representation. The beat period hypotheses are initialised as in Rosenthal's method. The beat positions are found by an iterative search through the onsets.

Temperley and Sleator published a hybrid harmony/meter recognition model [9]. Similarly to the heuristics of the other models, the rules specify e.g. that beats should be spaced regularly, beats should align with onsets, and strong beats should align with onsets of longer events. A score value is computed as a function of time, based on the fulfilment of the above rules, and the meter is recognized from the scores with the Viterbi algorithm. The Temperley–Sleator model is one of the models that produce a metrical grid with several levels. The model operates on symbols.

Sepännen [9] was the first to try to find more than one metrical level from an acoustic signal. His work is based on tatum [10] research, where tatum is the smallest rhythmic unit. The model comprises four stages: the detection of onsets of the signal, the tatum estimation, accents detection, and. the beat estimation. The system is fast enough to work in real-time.

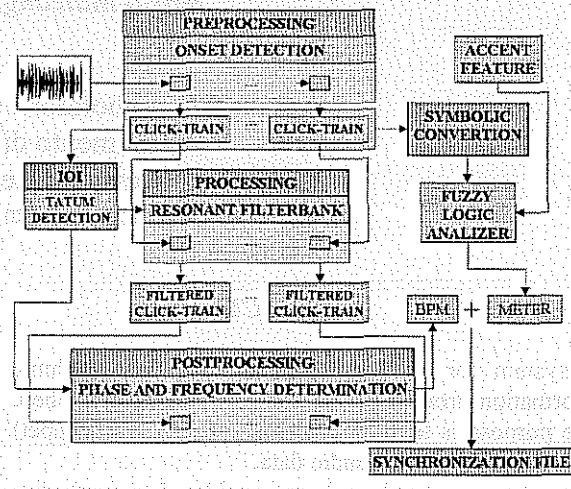
Finally we talk about Duden's algorithm [11] that use the Scheirer's model for real time beat-tracking.

2. PROPOSED MODEL

The model proposed is a mixture of precedent models. We have decided to use the Scheires's and Sepännen's model as basic refers work. The model comprises six main components: an onset detector, a tatum estimator, a resonator filter bank step, a phenomenal accent model, a Fuzzy Logic Analyser, and a measure estimator.

1.1 Sound Onset Detection

We use the term *onset detection* to refer to the detection of the beginnings of a note in an audio signal. This is the most important step for a beat tracking algorithm. If we can detect with high precision all the notes onsets in a musical performance we have the most important metrical information.



We know that this is not an easy task for polyphonic and multi tone audio music signal, but if we find a sufficient number of onset we are able to detect metrical features of the musical piece.

Scheirer was the first to propose a psychoacoustic solution to this problem. If we divide the signal into at least four frequency bands and the amplitude envelopes of the musical signal control the corresponding bands of a noise signal, the noise signal will have a rhythmic percept, which is significantly the same as that of the original signal.

The first step for onset detection is the frequency filter bank processing. The audio signal is divided into N frequency band and for each band s[n] we estimate the Envelope level using:

$$\mathrm{env}[n] = g[n] * abs(s[n]), \qquad (1)$$

where the function *abs* return the absolute value of the signal s[n] and g[n] is a low-pass filter. The low-pass filter reduces fast modulations and produces a signal that has the rough shape of the input signal. After this smoothing, the envelope can be decimated for furthers steps so that the signal env[n] under sampled. A low sample-rate allows fast processing, but it must be high enough to represent the rough shape of the envelope.

After the envelope, we calculate the first-order derivative. The derivative detects onsets of the signal. Sweeping over the sample buffer and subtracting the previous sample from the current sample easily implement it. So the derivative is actually approximated through the difference:

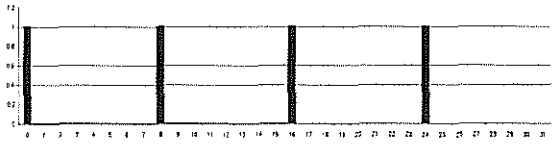
$$\mathrm{ClickTrain}(n) = \begin{cases} 0 & \text{if } Env(n+1)-Env(n)<k \\ Env(n+1) & \text{otherwise} \end{cases} \qquad (2)$$

where k is a constant value that define the level of threshold and **Env(0)=0**. This is the most important step of onset detection.

There are two important features that we must consider in onset detecting: exact time of note onset and the amplitude of the onset. To improve this step we use a process similar to Klaupuri's model [12] and for the thresholding problem we use a fuzzy logic computational model.
The output of this stage is a signal similar to a click-train. A click-train c with period $P(c) = \lambda$ or frequency $F(c) = 1/\lambda$, phase $\Phi(c) = \phi$ and peak C is a discrete signal with

$$c[x-\phi] \equiv \begin{cases} C \cdot u[x \bmod \lambda] & \forall x \geq 0 \\ 0 & otherwise \end{cases} . \quad (3)$$

We can see a representation of a click-train with $P(c)=8$ and $C=1$ in Figure 3.



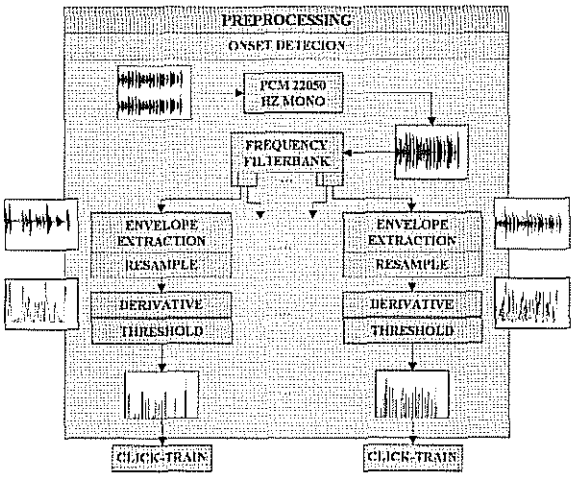
The final step of onset detection is to convert onset in symbol for accent recognition.

2.2 Resonator filter bank

We use a bank of resonator to detect the frequency of the beat. A comb filter has the feature to resonate with maximum peak only if the frequency of a click train filtered is that of the filter. If the frequency is not the same, the filter resonation is smaller or none. Another important feature of a comb filter is that it resonates with maximum peak at integral multiples of its characteristic frequency too.
Comb filters was introduced for beat detecting by Scheirer and optimised for real-time application by Duden. If we have a large number of comb filter we can detect a large number of frequency of a pulsation train. But a great number of filters mean more time for processing step. To improve the performance we search the tatum. Tatum is the smallest metric unit and was discussed by Sepannen in his thesis. The first step to find tatum is the IOI detection.

2.3 Inter Onset Intervals

To find the tatum unit we must consider the IOI. Given two onsets at times *t1* and *t2*, $t1 < t2$, the IOI between the onsets is defined as $o = t2 - t1$. The IOI's are not only computed between pairs of successive onsets; according to Sepannen the tatum is equal to the greatest common divisor approximation of the IOI's. We first define a IOIs histogram by which we can also retrieve information about the MRMU (most recurrent metric unit).
After the definition of this information we can set the resonant filter bank. Each filter delay will be a multiple of tatum or an integer around MRMU to decrease the filter's computing complexity.



2.4 Accent feature and fuzzy logic processing

In this stage we referred to more previous work: Rosenthal, Lee [13], Parncutt, Temperley-Sleator, Toivianen [15] and Sepannen. Task of this step is to select the feature that allows recognizing accent into the symbolic representation of onset.
According to Lerdahl and Jackendoff [14] there are some notational properties that constitute the phenomenal accent:

1. onsets of notes,
2. sforzandi (louder notes) and other local stresses,
3. long notes
4. sudden changes in dynamics or timbre
5. leaps to relatively high- or low-pitched notes, and
6. harmonic changes

For accent recognition we consider only durational accent. This means that we don't consider harmonic change, pitch accent and other accent that need a frequency analysis of the music signal.
After the selection of accent feature we use a fuzzy logic processor to detect accent from the symbolic representation of onset.
Fuzzy logic processing has been introduced at this step, as it is too complex to obtain a linear model of the accent detection. A fuzzy processor enables to transfer the expertise knowledge (musician) into its rules and membership functions.
A fuzzy logic processor is then trained to recognize accents in the audio musical sequence using as input harmonic changes, pitch accent and other extracted information.

2.4 Beat, accent and measure

The last stage has the task to use the information recognized to find the measure of the musical piece and to create a synchronization file. After the resonant filter bank step, the period of beat is evaluated selecting the frequency of the resonator that has the maximum peak. Other functions are applied to this step

to find the best BPM prediction. The second step has the task to try to find a high level of metrical structure. Accent recognition allows to determinate the meter in a musical piece. The combination of these two results allows evaluating the measure of a musical audio signal that for our virtual puppet is enough to dance on baroque music.

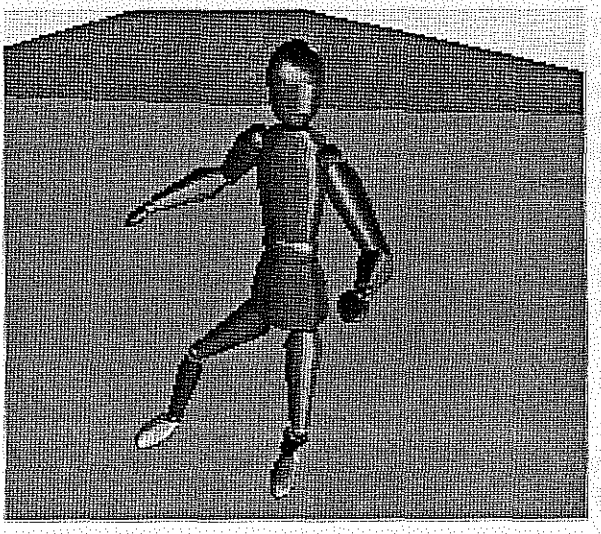
3. CONCLUSIONS

This work is a mixture of previous work for beat and meter recognition. The purpose is to drive the 3-D virtual character directly from an audio musical media. The output of this model is a synchronization file to allow a puppet to dance synchronised to a music execution.

To face this problem we have not considered pitch and chord recognition, but in future works we intend to improve the step of accent recognition using this feature too.

We intend to have extensive use of fuzzy logic as co processing in decisional process to extend this model to polyphonic and multi tone audio music performance.

A demonstration of the system is done using a virtual puppet droved to dance on baroque music (see fig. 5). The program reads an audio file, creates a synchronization file, and then the dancer executes the dance steps.



4. REFERENCES

[1] Camurri A., Morasso P., Tagliasco and V. Zaccaria R., Dance and Movement Notation, In P. Morasso and V. Tagliasco Editors, Human Movement Understanding, pages 85-124, North Holland, Amsterdam. 1986.

[2] Povel D. J. and Essens P., Perception of temporal patterns, Music Perception, 2(4):411–440, 1985.

[3] Parncutt R., A perceptual model of pulse salience and metrical accent in musical rhythms, Music Perception, 11(4):409–464, 1994.

[4] Rosenthal D. F., Machine Rhythm: Computer Emulation of Human Rhythm Perception, Ph.D. thesis, Massachusetts Institute of Tech., Agosto 1992.

[5] Large E. and J. Kolen , Resonance and the perception of musical meter, Connection Science, 6(2/3):177–208, 1994.

[6] Scheirer E., Tempo and Beat analysis of acoustic musical signals - J. Acoust. Soc. Am., 103(1): 588-601, January, 1998.

[7] Goto M. and Muraoka Y., Music understanding at the beat level:Real-time beat tracking for audio signals, School of Science and Engineering, Waseda University, Tokyo, JAPAN, 1995.

[8] Dixon, S., ''Automatic extraction of Tempo and Beat from Expressive Performance", J. New Music Research, 30(1), 2001.

[9] Sepännen, J., Computational models of musical meter recognition. Master's thesis, Tampere Univ. of Tech., Tampere, Finland, 2001.

[10] Bilmes A., Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm, Master of Science, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, September, 1993

[11] Temperley D. e Daniel Sleator - Modelling meter and harmony: A preference-rule approach - Comp. Music J., 23(1):10–27, 1999.

[12] Duden J. E., An Improved Approach to Real-Time *Beat-Induction from* Digital Audio Signal, Thesis, May 2002

[13] Klapuri. A., Sound onset detection by applying psychoacoustic knowledge, In Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP), volume 6, pages 3089–3092, 1999.

[14] Lee C. S., The perception of metrical structure: Experimental evidence and a model, In P. Howell, R. West, and I. Cross, editors, Representing Musical Structure, pages 59–127. Academic Press, London, United Kingdom, 1991.

[15] Lerdahl F. and Jackendoff R., *A Generative Theory of Tonal Music* Press, MIT Cambridge, MA, USA, 1983.

[16] Toiviainen P., Modelling the perception of metre with competing sub harmonic oscillators, In A. Gabrielsson, editor, Proc. 3rd Triennial ESCOM Conf., pages 511–516, Uppsala, Sweden, 1997.

Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003), Firenze, Italy, May 8-9-10, 2003

LILYPOND, A SYSTEM FOR AUTOMATED MUSIC ENGRAVING

Han-Wen Nienhuys, Jan Nieuwenhuizen

hanwen@cs.uu.nl, janneke@gnu.org

ABSTRACT

LilyPond is a modular, extensible and programmable compiler for producing high-quality music notation. In this article we briefly discuss the background of automated music printing, describe how our system works and show some examples of its capabilities.

1. INTRODUCTION

LilyPond was started by the authors as a personal project to investigate how music formatting can be automated. Over the years, the system has matured, and it is now capable of producing sheet music of respectable quality. LilyPond has not been designed with specific applications in mind, but has been used to print orchestral parts and scores, early music, as well as pop songs and piano works.

LilyPond is a modular, extensible and programmable compiler for producing high-quality music notation. The program produces a PostScript or PDF file by reading and processing a file containing a formal representation of the music to be printed. The output can be printed, or further post-processed, e.g., to produce images for web pages.

The system is partially implemented in the language Scheme [1] (a member of the LISP family of languages), and the program includes the GUILE Scheme interpreter [2], which allows users to override and extend the functionality of LilyPond. This ranges from adjusting simple layout decisions to implementing complete formatting subsystems.

LilyPond may be freely copied, used and modified under terms of the GNU General Public License, and can thus be described as "Open Source" software. In principle, users are not dependent on vendors to get bug-fixes, updates, and can download and use the program at no cost, virtually without any obligations.

LilyPond has an active community of users that offer support to newcomers, and a small band of developers that continue to improve the program on a voluntary basis. Documentation, downloads and typeset examples are available from the website, `http://www.lilypond.org`.

2. RELATED WORK

There are many music notation programs on the market, but most of these are proprietary systems, whose inner workings are kept secret. In the academic world, computerized music printing has received only little interest. The MusiCopy project [3], implemented and documented a system to typeset music notation. Unfortunately, the MusiCopy system is no longer available. TEX [4] is a programmable system for typesetting mathematics and text. It has become a basis for a number of macro packages to typeset music notation, of which MusiXTEX [5] is the most prominent. MusiXTEX puts formatting control almost completely in the user's hands, which makes it a powerful but hard to learn tool. Finally, we mention Common Music Notation (CMN) [6], a highly flexible notation system implemented in LISP. The input to CMN is also coded in LISP.

The input to LilyPond is a text file that encodes musical information. In other words, the format is a music representation that specifies music formally in terms of nested structures of pitches and durations. The format also allows for special instructions, which control layout of the printed output. In this sense, LilyPond resembles the GUIDO [7] and Haskore [8] format, which also contain primarily musical information in nested structures.

3. DESIGN AND IMPLEMENTATION

LilyPond is a batch program. When the program is invoked, it reads a file, which is then processed without any user interaction. Internally, the program executes the following steps.

1. The input is parsed and translated into a syntax tree.
2. Musical events are translated into of graphical objects; together they form the unformatted score. This step is called *interpreting*.
3. The unformatted score is formatted.
4. The formatted score is written to an output file.

Hence, LilyPond combines a music representation and a formatting engine. The conversion from music representation to graphical layout is done with a plug-in architecture. In the next subsections, we discuss these three concepts in more detail.

3.1. Input

The task of the program is to generate music notation with a computer given input in some format. Since the core message of a piece of music notation simply is the music itself, the best candidate for the source format is exactly that: the music itself.

Unfortunately, this observation raises a complex question: what really *is* music? Instead of pursuing this philosophical question, we have reversed the problem to yield a practical approach. We assume that a printed edition contains all musical information of a piece. Therefore, any representation that can be used to print a score contains the music itself. While developing the program, we continually adjust the format, removing as much non-musical information as possible, e.g., formatting instructions. At the same time the program is improved to fill in this information automatically. When the program is "finished" at some point, all irrelevant information will have been removed from the input. We are left with a format that contains exactly the musical information of a piece.

The input format was also shaped by practical concerns. LilyPond users have to key in the music by hand, so the input format

is the user-interface to the program. Therefore, the format has a friendly syntax. Producing music notation is a difficult problem, and difficult problems can only be solved if they are well-specified. Therefore we designed a format with a simple formal definition.

These ideas shaped our music representation. It is a compact format that can easily be typed by hand. It forms complex musical constructs from simple entities like notes and rests, in much the same way that complex formulas are built from simple elements such as numbers and mathematical operators. A simple example is given in the following fragment.

```
\notes { c'4 d'8 }
```

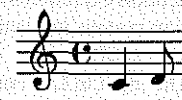
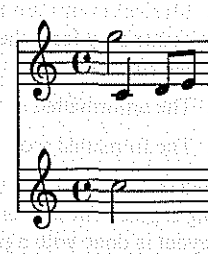


Figure 1: A simple LilyPond input fragment, with output.

The central concept of the input is formed by the *music expression*, a chunk of music with a specified duration. Notes (in this example `c'4` and `d'8`) form atomic music expressions. Simple music expressions can be combined to form more complex expressions, such as chords and voices. In this example, the braces combine both notes sequentially. The friendly syntax for notes is switched on with the `\notes` statement.

Similarly, music expressions can be combined parallel in time with the keyword `\simultaneous`. This construction is used both for parallel voices and for parallel staffs

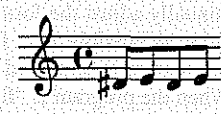
```
\notes \simultaneous {
 \context Staff = "1" {
  \simultaneous {
   { c'4 d'8 e'8 }
   { g''2 } } }
\context Staff = "2" {
  c''2 } }
```



In these examples, the keyword `\context` specifies how the following music expression should be interpreted.

With these basic constructors very complex music expressions can be formed. Large pieces need large music expressions. For example, a piano concerto can easily nest four levels deep (voice, staff, grand staff, score). Similarly, a 15 staff orchestral score will have a `\simultaneous` containing 15 sub-expressions. In practice, entering a such pieces in one large music expression is unwieldy. Therefore, the input format supports *identifiers*. Expressions can be entered separately and given names. A fragment can be entered as an identifier once, and used many times over. The following example uses an identifier (`seufzer`) to store two notes, and the fragment is repeated by using the identifier twice.

```
seufzer = \notes {
  dis'8 e'8 }
\score { \notes {
  \seufzer \seufzer
  } }
```



LilyPond has no concept of part-extraction, because there is no need for such a concept. Music fragments are assigned to identifiers. The music is then either combined into a full orchestral score, or it is used for creating the separate parts. Parts and scores

```
myMusic = \notes { c'4 d'4( e'4 f'4 }
\score { \notes {
    \myMusic
    \apply #reverse-music \myMusic
  }
}
```

Figure 2: Functions applied to music expressions. The first measure (named with identifier `myMusic`), is reversed by applying the `reverse-music` function, producing the second measure (The definition of `reverse-music` is omitted).

are derived from the same input, so changes in that input are always applied to both print-outs.

LilyPond includes a Scheme interpreter. It may be accessed from the input file by entering a Scheme expression preceded with a hash mark (#). For example, the following statement includes a Scheme expression (A list containing two symbols, `staff-bar` and `time-signature`).

```
\property Score.breakAlignOrder =
  #(list 'staff-bar 'time-signature)
```

When Scheme programming and music expressions are combined, they show the true power of the system. User-written functions can access and change all data in a music expression. This functionality can be used to analyze, change, generate and write musical data programmatically. A simple example is in Figure 2, where a piece of music is reversed by means of a user-defined function. A less frivolous example is in Figure 3. It shows the internal data representation of the example from Figure 1, dumped in XML syntax.

3.2. Interpreting

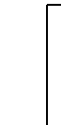
After parsing the input, musical contents are lined up and converted to graphical objects, resulting in an unformatted score. This step is called *interpreting* the input. The events are processed in the order that they would be performed. Events which would happen simultaneously are processed together, and end up at the same horizontal position. In this step, context sensitive information, such as key signature and measure subdivision, is computed and used to insert bar lines and print accidentals automatically.

Interpreting is implemented with a plugin architecture. These plugins are called *engravers*. Each each engraver performs one specific function in the conversion process. For example, there is a `Note_head_engraver`, that produces note-head objects for note events. Stems are created by the `Stem_engraver`. If the `Stem_engraver` notices a note head object at some point, it creates a stem object and connects both.

Engravers only have to perform one specific function. The interactions between the different plugins are handled by the architecture: it keeps track all events and graphical objects, and ensures that each engraver gets precisely the information it needs. This modular architecture makes maintaining and extending the program relatively easy.

```
<SequentialMusic>
  <EventChord>
    <NoteEvent>
      <duration log="2" dots="0"
                numer="1" denom="1">
      </duration>
      <pitch octave="0" notename="0"
                alteration="0">
      </pitch>
    </NoteEvent>
  </EventChord>
  <EventChord>
    <NoteEvent>
      <duration log="3" dots="0"
                numer="1" denom="1">
      </duration>
      <pitch octave="0" notename="1"
                alteration="0">
      </pitch>
    </NoteEvent>
  </EventChord>
</SequentialMusic>
```

Figure 3: The input format shown in an XML format. This output is generated directly from the parse tree of the example in Figure 1 using a short (100 line) Scheme function.

obj$_1$: `glyph-name="treble"`	obj$_2$:  `glyph-name="fourfour"`
obj$_3$: `position=-6` `stem=`→*obj*$_4$	obj$_4$: `line-thickness=0.12` `note-head=`→*obj*$_3$
obj$_5$: `position=-5` `stem=`→*obj*$_6$	obj$_6$: `line-thickness=0.12` `note-head=`→*obj*$_5$

obj$_7$:
`line-count=5`
`staff-space=1.0`
`line-thickness=0.1`
	obj$_8$: *container*
	`elements=`
	→*obj*$_1$, ..., →*obj*$_7$
	`height=`*group-height*

Figure 4: Graphical objects from the unformatted score for Figure 1. Each object stores style and layout settings in variables. These variables are generic, and can contain any type of object, including numbers, strings, lists, procedures and pointers to other objects. `obj8` is "abstract," i.e. it does not produce output.

3.3. Layout

The product of the interpretation step is a collection of graphical objects, the *unformatted score*. Each musical symbol in the score is represented by a graphical object. Relationships such as containment, alignment, or element spacing are also represented by *abstract* graphical objects. Figure 4 shows a simplified version[1] of the unformatted score for the example of Figure 1.

Objects contain variables that describe properties of the boject. These variables (*object properties*) are used in the formatting process in many ways.

- Global style settings are stored in properties. All objects share a global defaults, for properties. For example, the global default for beam objects has a property `thickness`, which is set to 0.48 staff space.
- Formatting adjustments are also stored in properties. A stem can be forced up by entering a simple command in the input file. This command adds `direction=UP` to the definition of a stem object.
- Properties containing subroutines define formatting procedures and other behavior of graphical objects. For example, in Figure 4, the height of the container object is given by a function *group-height*, stored in the property `height`. These functions may be replaced by user-written Scheme code.
- Objects can refer to each other. For example, the stem and note head objects have `note-head` and `stem` properties pointing to each other.

[1]The example in Figure 4 has been highly simplified. In LilyPond version 1.7.1, the file shown actually is translated into 33 different graphical objects. The line breaking process multiplies this to 59 objects, most of which are abstract.

Object properties are stored in Scheme data structures, and can be manipulated in user-written code.

In the formatting step, spacing and line breaks are determined, and layout details of objects are computed. For example, stem objects normally do not have a predefined length. During the formatting process, a length is computed and filled into a `length` property. The result of the formatting step is a finished score, which is written to disk. A helper program post-processes the output to add page breaks and titling, and produces a ready-to-view PostScript or PDF file. LilyPond by default outputs the notation in a TEX file, but other output formats are also available: there is experimental support for SVG and direct PostScript output.

4. EXAMPLES

Since none of the freely available fonts satisfied our quality demands, we have created a new musical font, called "Feta", based on printouts of fine hand-engraved music. A few notable aspects of Feta are shown in Figure 5. The half-notehead is not elliptic but slightly diamond shaped. The vertical stem of a flat symbol is slightly brushed: it becomes wider at the top. Fine endings, such as the bottom of the quarter rest, do not end in sharp points, but rather in rounded shapes. Taken together, the blackness of the font is carefully tuned together with the thickness of lines, beams and slurs to give a strong yet balanced overall impression.

The spacing of a piece of music should reflect the character of the music. A piece should not contain unnatural clusters of black nor big gaps with white space. The distances between notes should reflect the durations between notes, but adhering with mathematical precision to durations will lead to a poor result: the eye not only notices the distance between note heads, but also between

Figure 5: Three glyphs from the Feta font.

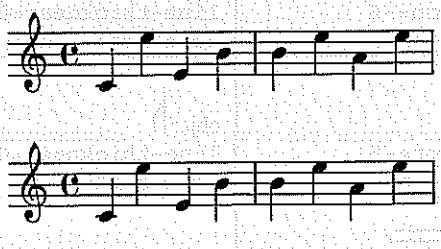


Figure 6: A fragment demonstrating spacing. The top fragment is printed with optical spacing. In the bottom fragment, all note heads are at equal horizontal distances. As a result, the down-stem/upstem note pairs form visual clumps.

consecutive stems. Therefore, the notes of a up-stem/down-stem combination should be put farther apart, and the notes of a down-up combination should be put closer together, all depending on the combined vertical positions of the notes [9]. Figure 6 demonstrates this optical spacing.

In engraved music, beams should cover staff lines as much as possible. This prevents small distracting wedges of white space, and uneven appearance of the beam thickness. In Finale, such beams are known as "Patterson beams" after the plug-in that offers this functionality. We call this *beam quantization*, as the vertical positions of the beam end-points are not continuously variable, but discrete. LilyPond also offers beam quantization. It uses a generic mechanism, where a large number of configurations for both beam endings are tested. For every configuration a penalty score is computed. For example, configurations that lead to very short stems incur a heavy penalty, and very long stems a small penalty. Similarly, a penalty is computed for the slope of a configuration, and for positions that lead to "forbidden" positions of secondary and tertiary beams. A weighted sum of the penalties measures the beauty of a configuration. After computing penalties for all configurations, the best scoring configuration is used as beam position. This approach is independent of the number of stems (Ross [10] lists many examples for beams with two stems, but gives no further rules), and adapts to different beam thicknesses. In addition, if more complex rules are needed, these can be integrated by adding more scoring functions to the code.

Some formatting procedures are based on other work. For example, Hegazy and Gourlay [11] describe a line breaking approach similar to TEX's line-breaking algorithm. This algorithm has been re-implemented for LilyPond. The spacing engine describes the desired spacing in terms of springs. When justifying a single line, force is needed to compress or stretch these springs. Very loosely and very tightly spaced lines require more force. A dynamic programming algorithm is used to find the configuration of line-breaks that keeps the total force as low as possible. This results in a set of line breaks that favors even and natural spacing across the entire piece.

We study engraved editions as a guide when implementing formatting algorithms. The most recent revisions of the the beaming and spacing code were guided by the Bärenreiter edition of the Cello Suites by J. S. Bach [12], in particular, measurements of the Sarabande of the second suite guided our current spacing algorithms. Figure 9 shows our rendering of this piece. LilyPond's default decisions for stemming, spacing and line breaking follow the printed edition, except in two places, where manual override of the layout was necessary. The layout quality of this piece is comparable with the original hand-engraved edition.

The best quality print-outs are attained for single staff, single voice music. Nevertheless, multiple staffs and polyphonic notation are also supported. Conflicts in notehead placement (*collisions*) between polyphonic voices are resolved automatically if possible. Figure 7 shows some some collisions in the context of piano music. LilyPond is not limited to classical music. There is also support for chord names, tablature, figured bass and medieval notation.

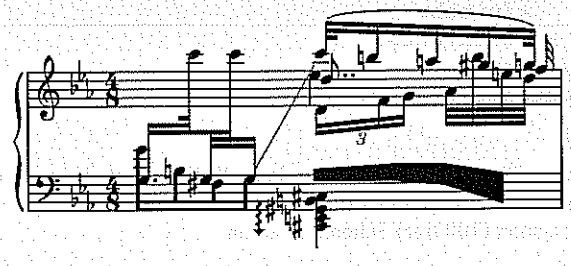
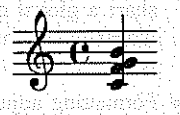


Figure 7: Random complex polyphonic notation. The lower left beam uses French beaming and different stem and beam thicknesses but its position is still quantized correctly.

The design of the program enforces a strict separation between content (music) and form (typography). A consequence is that the same piece music may be represented in different forms. Chords form simple example. In the following fragment, a chord is entered using the syntax `<< ...>>`. That same chord is then printed both in a staff and in textual form.

```
sus = \notes {
  <<c' f' g' b'>>4 }
\score { \simultaneous {
  \context ChordNames \sus
  \context Staff \sus
} }
```



Separation between form and content is also used in the support for transcribing mensural music. Mensural music uses different font shapes for notes, clefs and alterations. In addition, particular rhythmical patterns of notes are denoted by combining their note heads in special symbols called *mensural ligatures*. LilyPond does not add a separate music representation for this type of music. Instead, the music is entered as if it were modern notation, and ligatures are marked in the input. During print-out, a print style for mensural notation can be selected. Support for historic print styles is included, and can be used to check the transcription to modern notation. Figure 8 shows an example of this process. In effect, LilyPond transcribes from modern notation to mensural notation. As a consequence, there is a single input language representing both mensural ligatures and their transcriptions into modern notation. The separation between content and form is thereby maintained. Support for ligature notation is an experimental feature,

and current work focuses on implementing the variety of printing styles of Gregorian notation.

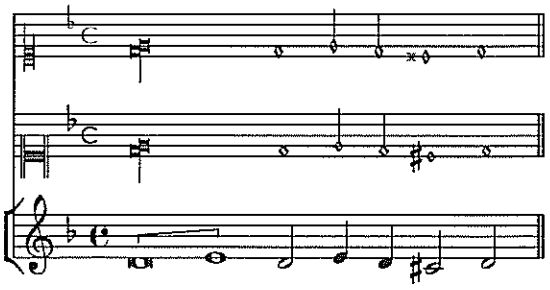


Figure 8: The same fragment of music in three ancient layout styles: historical print (top), contemporary mensural notation (middle), and modern notation with ligature brackets (bottom).

Finally, this paper itself shows an application of LilyPond: text and music can easily be mixed in the same document. The input format is ASCII based, so one can enter snippets of LilyPond input in other ASCII based document formats, such as LaTeX and HTML. With the aid of a small helper program, these fragments can be replaced in the output by the corresponding music notation, in the form of pictures (for HTML) or TeX (for LaTeX). For example, Figure 1 was created by entering the following in the LaTeX source file

```
\begin[verbatim]{lilypond}
  \notes { c'4 d'8 }
\end{lilypond}
```

5. DISCUSSION AND FUTURE WORK

We have presented our progress on LilyPond, a free music engraving system, which converts a music representation to high quality music typography. For some pieces, LilyPond output is comparable to hand-engraved music. The program is focused on producing high quality notation *automatically*. This makes it an excellent tool for users who are not notation experts.

LilyPond can run without requiring keyboard or mouse input. This makes it an excellent candidate for generating music notation on the fly, e.g., on webservers. The degree of automation also makes it a suitable candidate for transforming large bodies of music to print automatically: for example, LilyPond has been used to produce an automated rendering of a database of 3,500 folk songs stored in ABC [13]. This is helped by the fact that LilyPond includes (partial) convertors for a number of music formats, among others MusicXML [14], MIDI, Finale's ETF, and ABC.

Beaming, line breaking and spacing are the strong points of the formatting engine. In some areas the engine still falls short. For instance, placing of fingering indications, articulation and dynamic marks together is a complex problem. We plan to improve collision handling for these notation elements, so manual adjustments are no longer necessary in this case. Other plans for future work include improving formatting of slurs and adding page layout to the system.

The program has no graphical user interface, and always produces all pages of the final output. To see the result of a change, the program has to be rerun on the entire score. In effect, this transforms music editing into a debug-compile cycle, and fine-tuning layout details is a slow process. We plan to explore solutions that make manual adjustments with LilyPond a more interactive and efficient process.

6. ACKNOWLEDGEMENTS

Our sincere thanks go out to all our developers, bug-reporters and users; without them LilyPond would not have been possible. In particular, we would like to thank Jürgen Reuter for contributing ligature support, and providing the mensural notation example for this paper.

7. REFERENCES

[1] "Revised^5 report on the algorithmic language scheme," *Higher-Order and Symbolic Computation*, vol. 11, no. 1, September 1998.

[2] Free Software Foundation, "GUILE, GNU's Ubiquitous Intelligent Language for Extension," `http://www.gnu.org/software/guile/`, 2002.

[3] Allen Parish, Wael A. Hegazy, John S. Gourlay, Dean K. Roush, and F. Javier Sola, "MusiCopy: An automated music formatting system," Tech. Rep., Department of Computer and Information Science, The Ohio State University, 1987.

[4] Donald Knuth, *The TeXbook*, Addison-Wesley, 1987.

[5] Daniel Taupin, Ross Mitchell, and Andreas Egler, "Musixtex," 2002.

[6] Bill Schottstaedt, *Beyond MIDI. The handbook of musical codes*, chapter Common Music Notation, MIT Press, 1997.

[7] H. H. Hoos, K. A. Hamel, K. Renz, and J. Kilian, "The GUIDO music notation format—a novel approach for adequately representing score-level music," in *Proceedings of International Computer Music Conference*, 1998, pp. 451–454.

[8] Paul Hudak, Tom Makucevich, Syam Gadde, and Bo Whong, "Haskore music notation—an algebra of music," *Journal of Functional Programming*, 1996.

[9] Helene Wanske, *Musiknotation — Von der Syntax des Notenstichs zum EDV-gesteuerten Notensatz*, Schott-Verlag, Mainz, 1988.

[10] Ted Ross, *Teach yourself the art of music engraving and processing*, Hansen House, Miami, Florida, 1987.

[11] Wael A. Hegazy and John S. Gourlay, "Optimal line breaking in music," in *Proceedings of the International Conference on Electronic Publishing, Document Manipulation and Typography. Nice (France)*, J. C. van Vliet, Ed. April 1988, Cambridge University Press.

[12] Johann Sebastian Bach, *Sechs Suiten für Violoncello solo*, Number BA 320. Bärenreiter, 1950.

[13] Erich Rickheit, "Yet another digital tradition page," `http://sniff.numachi.com/~rickheit/dtrad/`.

[14] Guido Amoruso, "Xml2ly," `http://www.nongnu.org/xml2ly/`.

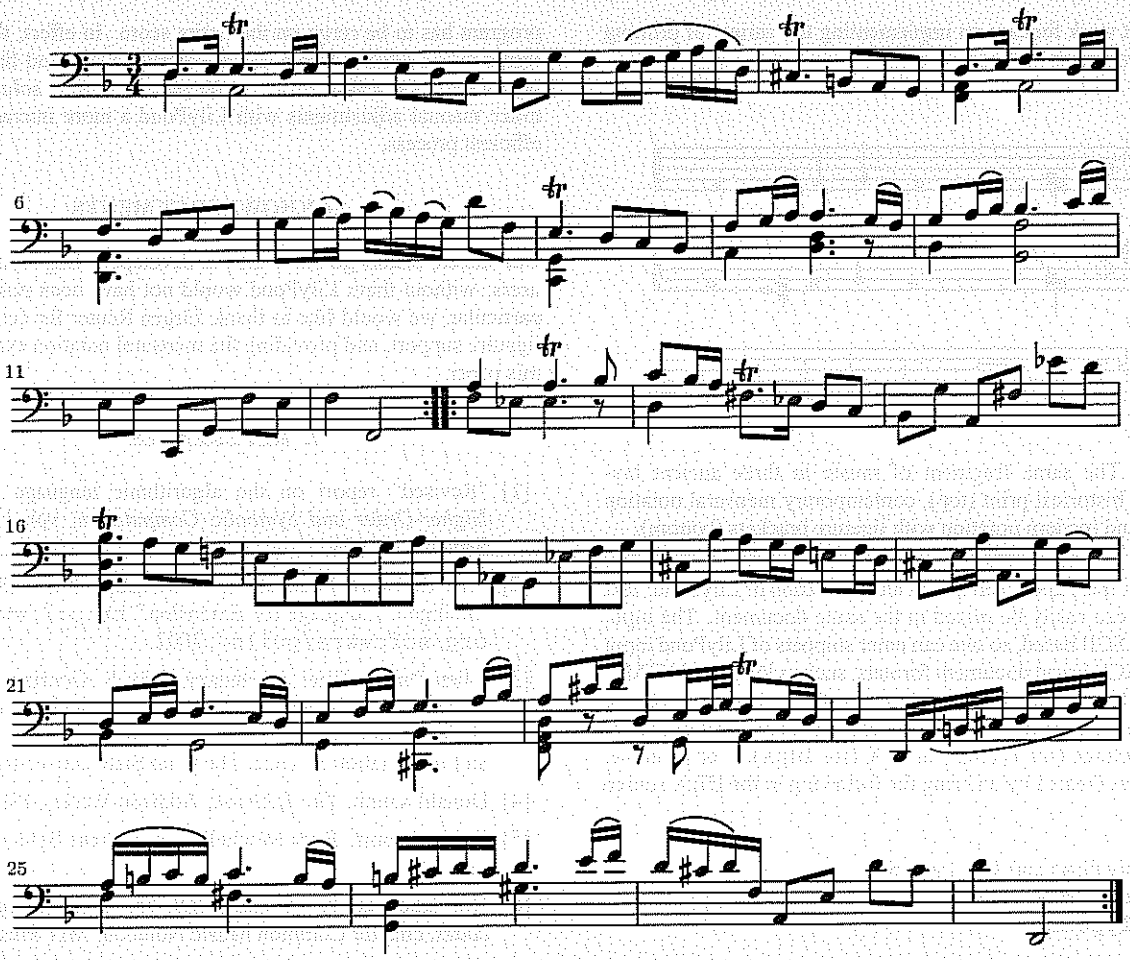


Figure 9: Sarabande of the second Cello Suite by J.S.Bach, after the Bärenreiter edition [12]. This example had manual adjustments in two places: the line break in the last line was forced, as were the stem directions on the last beat of measure 24. In addition, this example has been scaled by 80%, to fit in the format of this report.

A MATHEMATICAL MODEL FOR THE 'HOLOPHONE', A HIGH DIRECTIVITY ACOUSTICAL RADIATOR

Luca Mariorenzi, Lorenzo Seno

Centro Ricerche Musicali
CRM - Roma
info@crm-music.org
l.mariorenzi@inwind.it
lorenzo.seno@bigfoot.com

ABSTRACT

In this paper we will introduce a mathematical model of a highly directive auralization device, composed by a baffled driver whose acoustical field is reflected by a focusing paraboloid. Highly directive devices can substantially help in solving the puzzle of auralizing music and speech, both in outdoor and indoor performances, even in adverse conditions - as nowadays often happens in public shows.

The target of this work is to provide a first foundation for calculation and performance forecast methods, in order to make further improvements possible, together with a correct design, dimensioning, optimization and engineering of this device.

The model here introduced is based upon the Fresnel's approximation of the field integral formulation. The resulting "radiation diagrams" corroborate indeed the directivity gain obtained by means of the paraboloid, and qualitatively gives reason of the actual overall behavior of the device.

Further researches will make possibly use of a less restrictive method or of a BEM, avoiding the Fresnel's approximation restrictions (several wavelength of distance, namely frequencies not too low, and small axial angles). This way, it will be possible to cope with the analysis of the behavior of Holophone's clusters, for which an improvement in directivity is expected at the low end of acoustical spectrum.

1 INTRODUCTION

The composer Michelangelo Lupone, with the help of one of us (Lorenzo Seno), has recently developed prototypes of a focusing auralization device at the Centro Ricerche Musicali. The idea was to use a reflective paraboloid in front of a loudspeaker to improve directivity. These prototypes – called "Holophones" - are today currently used in several Italian and international indoor and outdoor music festivals. Perceptual results are generally considered as highly satisfactory.

These devices have been basically developed as an attempt to cope with the various problems encountered when auralizing music in ill environments and conditions, which are often typical of today's exploitation in public performances.

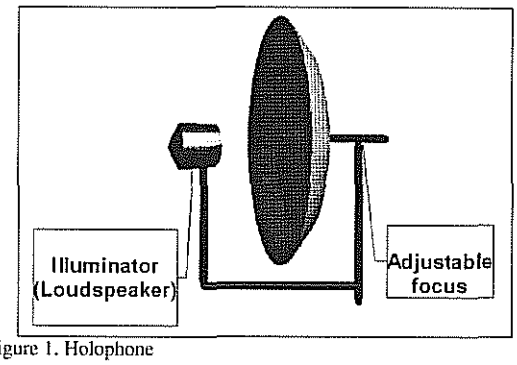


Figure 1. Holophone

A highly directive emission makes it easy the efficient transfer of large amounts of sound energy to the audience, saving in this way electrical power, and consequently weights, volumes, and costs. Directivity brings also, in indoor performances, to a more simple reverberation scheme – even in large premises - helping in this way an effective control of the perceived direction of sound and the perception of sources motion. The application of psychoacustical motion and spatialisation cues becomes also more effective. Briefly, such directivity makes it possible what the composer evocatively calls "wave-front sculpture" – considered as constitutive part of composition and performance. This is maybe the main attractive feature of this device from a musical standpoint.

Last but not least, with this device the finest sound details hold their intelligibility even at large distances and in reverberating premises. This suggests that interesting results can also be achieved in speech applications, such as Conference Halls, and in speech and music auralization in Churches.

2 DESCRIPTION OF THE METHOD

To cope with the large bandwidth of musical signals our approach is fully ondulatory, making use of the integral formulation of the linear field equations in terms of Green functions [1] [2] [3] [5] over the boundaries.

The prototype of the Holophone (Figure 1) is composed by an illuminator (Φ = 20 cm.), a metallic arm carrying the entire structure and the fiberglass reflector, which has the shape of a rotation paraboloid (Φ = 1.5 m.) with a focal length of about 83 cm. When appropriate, all the numerical calculations have been made taking these dimensions into account.

The source (illuminator) is made by a closed box loudspeaker of nose-cone shape, whose acoustical field is reflected and focused by the paraboloid.
The loudspeaker has been modeled as a plane circular piston over an infinite baffle (see Figure 2 for symbols and conventions).
The paraboloid surface can then be considered as a boundary surface (illuminated by the field radiated by the piston) over which we can integrate the appropriate Green functions. To ease the heavy computational load, we used the Fresnel approximation in the expression of the propagated field, taking advantage of the cylindrical symmetry of the problem to reduce the integration from 2D to 1D.
Fresnel's approximation keeps its validity only for a distance of more than some wavelength from the source (the paraboloid), and at small axial angles. This approximation is nevertheless good enough for our today purposes, which substantially consist of a first-time validation and estimation of the directivity gain around the axis.

3 THE SOURCE MODEL

The loudspeaker has been classically modeled as a baffled circular piston [1] [2] [5] , whose oscillation velocity is:

(1) $\vec{U}_0 = U_0 \cdot \hat{n}$

(see Fig. 2). The surface element dS in P' gives to the pressure field at the point P the contribute:

(2) $dp = \frac{j \cdot c \cdot \rho_0}{2\lambda} \cdot \frac{dQ(\vec{r}')}{|\vec{r} - \vec{r}'|} \cdot e^{j(\omega \cdot t - k \cdot |\vec{r} - \vec{r}'|)}$

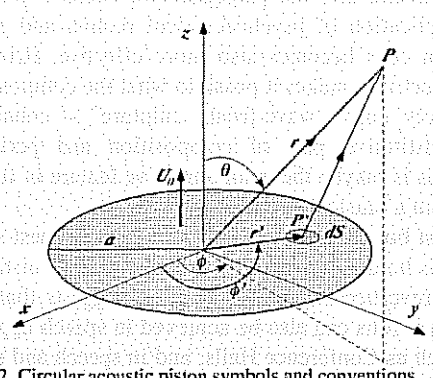


Figure 2. Circular acoustic piston symbols and conventions

Where:

(3) $dQ = \vec{U}_0 \cdot \hat{n} \cdot dS$

is the source strength of dS., $k = 2\pi \cdot \lambda^{-1}$ is the wave number, being λ the wavelength.
We are interested only in amplitudes, so that we can omit in (2) the temporal factor $e^{j \cdot \omega \cdot t}$.
The normalized acoustical pressure is thus:

(4) $p_p(\eta, \theta, \lambda) = p_{rif}(\lambda) \cdot p_n(\eta, \theta, \lambda)$

(5) where $p_{rif}(\lambda) = \frac{j \cdot \sigma \cdot 2\pi \cdot U \cdot c \cdot a}{\lambda}$

(6) $p_n(\eta, \theta, \lambda) = \int_0^1 \int_0^{2\pi} \frac{\eta_1 \cdot e^{-j \cdot k(\lambda) \cdot a \cdot \Delta\eta}}{\Delta\eta} d\phi' \, d\eta_1$

(7) $\eta_1 = \frac{r'}{a}, \quad \eta = \frac{r}{a}$

(8) $\Delta\eta = \sqrt{\eta_1^2 + 2\eta \cdot \sin(\theta) \cdot \eta_1 \cdot \sin(\phi') + \eta^2}$

being a = piston radius, σ = air density, c = sound propagation velocity. In these formulae we made use of the adimensional normalized distance η - bringing this to results not dependent on the piston dimensions.
The integral in (6) can be numerically calculated using usual methods such as those contained in numerical software packages.
As expected, the trend of the acoustic pressure versus the distance shows two regions (Figure 3). In the near-field region, in which vanishing waves prevail, we have the classical oscillatory trend.
At distances of some wavelengths the trend becomes asymptotic, tending to the r^{-1} law.
The reflector lies in the near-field region of the field emitted by the source at any acoustic frequency. For this reason a fully ondulatory approach is a must when estimating the illuminating field, and the usual far-field approximations cannot be used.

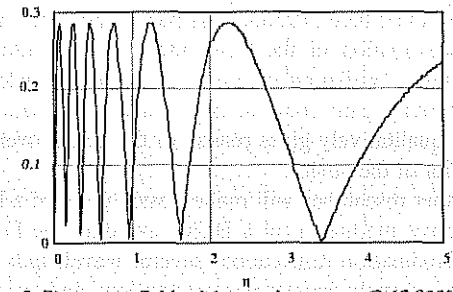


Figure 3. Piston near field axial acoustic pressure @ 12.000Hz.

4 THE HOLOPHONE MODEL

The field reflected from the paraboloid has been computed using the first Rayleigh-Sommerfeld form, considering the surface as illuminated by the piston field. The second form is null because we have considered the surface as a perfect reflector, so that the particle velocity is null on its boundary.
As already said, we used also Fresnel's approximation (which is the usual approach in Optics [3]).
Putting the origin in the paraboloid vertex, the generatrix parabola has the equation:

(9) $\rho = \alpha \cdot \sqrt{\xi}$

the focal point is thus at $z_{focus} = 0.25 \cdot \alpha^2$

The Rayleigh-Sommerfeld expression for the Holophone's field, taking into account the paraboloid shape and the field of the illuminating piston (see eq. (4)), is as follows:

$$(10)\quad p_h(z,\psi,\lambda) = jk \cdot \int_0^{\frac{k\cdot R}{2}} \frac{p_p(\varsigma,\pi-\arcsin(\frac{\rho_n}{k(\lambda)\cdot\varsigma}),\lambda)}{z_n - z_n'} \cdot e^{-j\cdot\gamma} \cdot J_0(\rho_n \cdot \tan(\psi)) \cdot \rho_n \cdot d\rho_n$$

Where $k = k(\lambda)$ and :

$$(11)\quad \varsigma(\rho_n, z_p, \alpha, \lambda) = \sqrt{z_p^2 + \frac{\rho_n^2}{k^2} + \frac{\rho_n^4}{k^4\alpha^4} - 2z_p\frac{\rho_n^2}{\alpha^2 k^2}}$$

$$(12)\quad \gamma(z_n, z_n', \rho_n, \psi) = z_n - z_n' + \frac{(z_n \cdot \tan(\psi))^2 + {\rho_n}^2}{2\cdot(z_n - z_n')}$$

$$(13)\quad z_n'(\rho_n, \lambda) = \lambda \cdot \frac{\rho_n}{2\pi\cdot\alpha^2}\ ,\ \rho_n = k\cdot\rho \text{ and } z_n = k\cdot z$$

Here z is the distance of the observing point from the paraboloid vertex, z_p is the axial position of the piston, ψ is the angle from axis of the observing point. J_0 is the zeroth-order Bessel function. The integration variable ρ is the parabola abscissa (see eq. (9)).

5 SIMULATIONS

Figure 4 shows the typical swings of near-field vanishing waves. The near field region ends at about the last peak. @ 5000 Hz this means 7 m. of distance from the Holophone. In this Figure we made use of the normalized distance $z_n = z \cdot k$. Amplitude in arbitrary dB units.

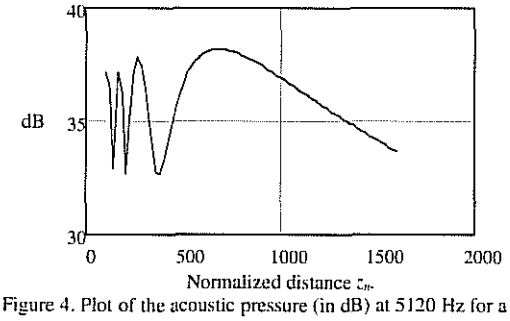


Figure 4. Plot of the acoustic pressure (in dB) at 5120 Hz for a Holophone with an 150 cm exit aperture (Φ), a Φ 20 cm driver and focus @ 83 cm

Figure 5 and Figure 7 allow a comparison between the Holophone and its piston. Figure 9 shows this comparison in the form of a ratio. Figure 6 shows the behavior of a piston of the same aperture as the Holophone, outputting the same power at the top frequency as the 20 cm. Piston. Figure 10 shows the ratio between the Holophone and this piston. The two gain diagrams show quite clearly that the effect of the parabolic reflector is not limited to the directivity improvement with respect to a smaller source. It shows actually an improvement also in comparison with a plane aperture of the same size, excited with the same power. This means that the use of a parabolic reflector is not merely equivalent to an increase in size of the aperture. Of course, this gain decreases with frequencies, becoming more and more negligible when approaching the low frequency limit (see Figure 10). This is an expected feature, because the acoustic mirror curvature becomes more and more negligible in comparison to the wavelength, when frequency decreases. All the plots involving the Holophone has been limited to ±10°, as a consequence of Fresnel's approximation.
In all these polar diagrams, frequency grows as the axial (peak) values do. Frequencies are 320, 640, 1280, 2560, 5120, 10240 Hz. Reference pressure is arbitrary, but it is the same for all plots, allowing direct comparisons.

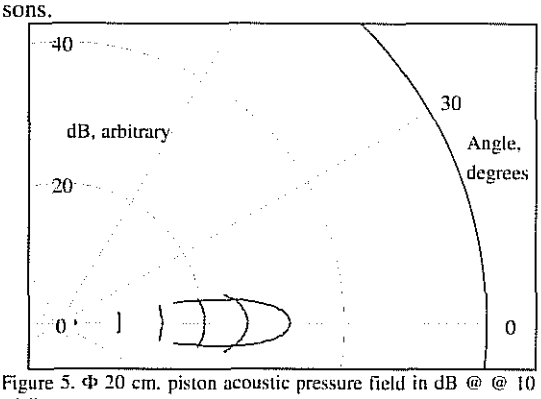


Figure 5. Φ 20 cm. piston acoustic pressure field in dB @ @ 10 m. of distance.

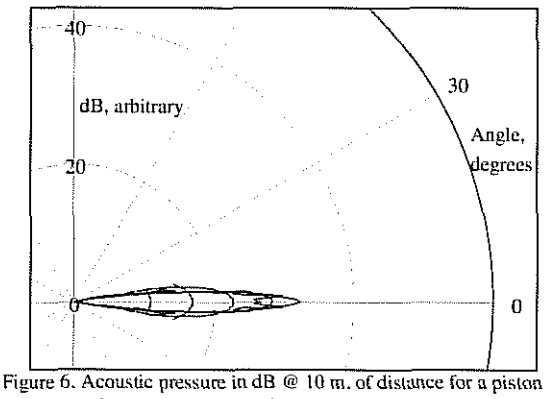


Figure 6. Acoustic pressure in dB @ 10 m. of distance for a piston of Φ 150 cm, the same aperture as the Holophone.

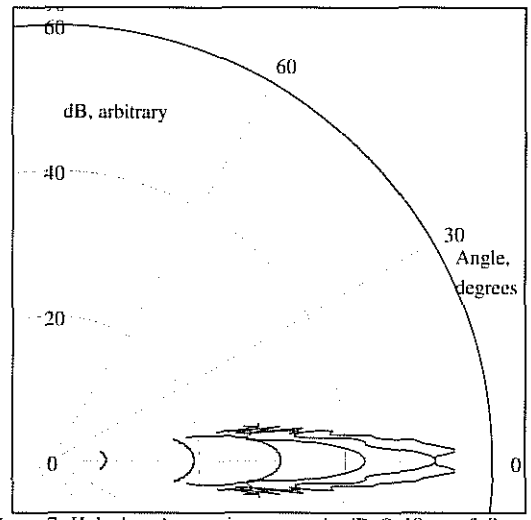


Figure 7. Holophone' acoustic pressure in dB @ 10 m. of distance from the reflector, for the same set of frequencies. The piston is at focal point.

Figure 8 shows how defocusing can smooth the frequency characteristics of the Holophone, widening the useful angle. They are actually used close to these conditions

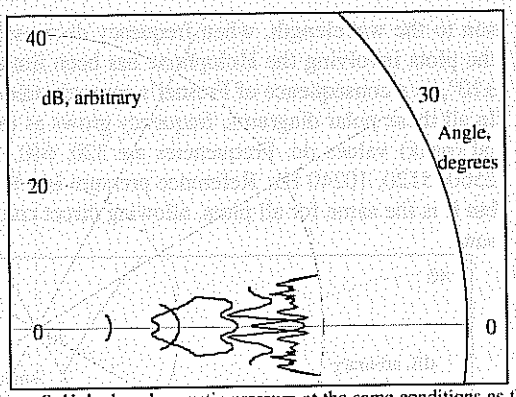


Figure 8. Holophone' acoustic pressure at the same conditions as the previous figure, but with the piston in intrafocal position (40 cm.)

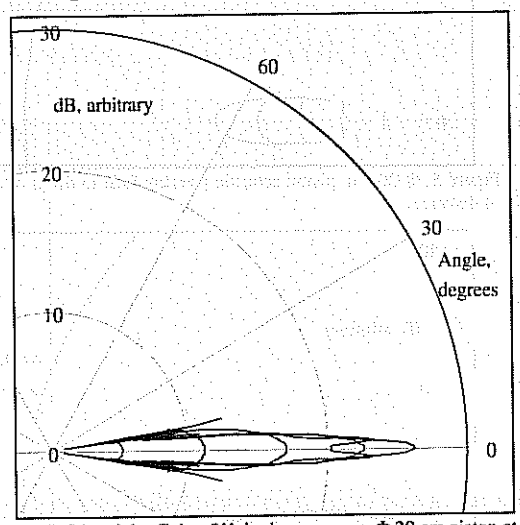


Figure 9. Directivity Gain of Holophone versus Φ 20 cm piston at a distance of 10 m. from both.

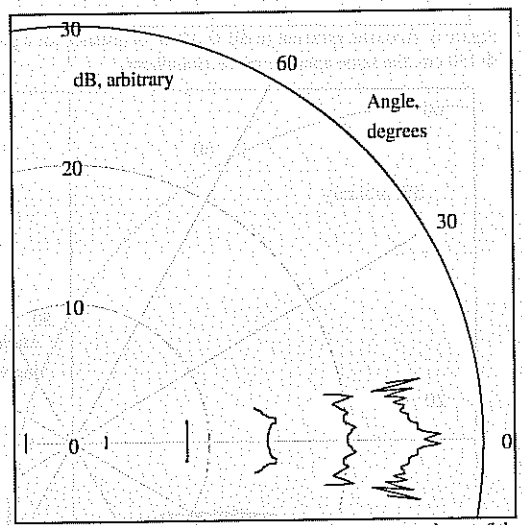


Figure 10. Directivity Gain of the Holophone versus a piston of the same aperture (Φ 150 cm.) at 10 m. of distance from both.

6 CONCLUSIONS AND FUTURE DEVELOPMENTS

The high directivity pattern, especially at high frequencies, can explain the reason why the Holophone sound is so detailed and crisp even at large distances. In the mid to high frequency range the quasi-planar radiation pattern gives also reason of the large amount of energy transmitted to the audience, and of the high intelligibility degree even in large reverberating premises. This device can substantially improve the efficiency of the auralization, considered as the ratio of the acoustic energy transmitted to the audience versus the electrical power spent.

From the energy standpoint, a reduction of the aperture ratio of the reflector can further improve the power transmitted back to the audience. This can be accomplished increasing the sagitta in order to intercept a wider solid angle from the loudspeaker, shortening the focal length.

A second problem is how directivity improvements can be obtained even at lower frequencies. A viable way to cope with this goal is the use of Holophone's arrays. This technique is substantially borrowed from the radio-astronomy field, in which this method is used to achieve synthetic apertures using several distant radio-telescopes. The problem is formally the same, with the exception that radio-astronomy signals are in a very narrow band, while acoustic signal have a very large spectrum. For this reason, the computation of the field radiated by such arrays requires the drop-out of Fresnel's approximation, because it looses its validity as far as the low frequency end of the acoustic spectrum and wide axial angles are approached.

7 ACKNOWLEDGMENTS

Thanks to Marco Giordano for review, criticism and discussion.

REFERENCES

[1] Philip McCord, Morse, K. Uno Ingard (Contributor) – "Theoretical Acoustics" - Princeton University Press – 1987 – Ch. 7, Ch. 10.

[2] Bruneau, Michel – "Manuel d'acoustique fondamentale" – Hermes – 1998 – Ch. 6

[3] Gori Franco – "Elementi di ottica" – Accademica – 1995 – Ch. 1, Ch. 2.(2.6)

[4] S. Solimeno, B. Crosignani, P. Di Porto – "Guiding, Diffraction and Confinement of Optical Radiation" – Academic Press – 1986 – Ch. 1, Ch, Ch.3, Cap.4.

[5] Lawrence E. Kinsler & others – "Foundamentals of Acoustics" – John Wiley & Sons – 1982 – Ch. 8.

[6] C.R.M. dati tecnici sugli Olofoni di M. Lupone – dispense – 1999-2000.

[7] Malcolm J. Crocker – "Encyclopedia of Acoustics" – John Wiley & Sons – 1997 – Vol. 1.

[8] L. Marston – "Quantitative ray methods for scattering" in "Encyclopedia of Acoustics" – John Wiley & Sons – 1997 – Vol. 2.

[9] I. Malecki – "Physical foundations of technical acoustics" - "Reflection of acoustic waves"; "Single reflection of acoustic waves" - Pergamon Press – 1969.

VM-Zone: a tool for interactive didactical experiences in Music

S. Cavaliere - G. Sica

Università di Napoli Federico II - Dipartimento di Scienze Fisiche
Gruppo ACEL
giancarlo.sica@na.infn.it

ABSTRACT

The present paper highlights the evolutions concerning the VM-Zone (Virtual Music zone) project, whose purpose is to make available a set of hardware/software tools to explore both acoustical and musical worlds by means of sensors and virtual instruments developed via Max/MSP software environment and I-Cube sensors to MIDI interface. These tools may constitute a valid didactical approach in order to investigate about acoustical phenomena as well as musical micro and macro-structures. Moreover, they offer the possibility of exploring the more or less subtle, intertwined relations existing between objective (physical) and subjective (perceptual) world by means of sensors; the purpose is to make the users free to experiment with various kind of physical properties like pressure, proximity, speed and so on. Here we will present some practical experiments concerning this project, developed in a real didactical environment. The target were pupils from an elementary and medium school.

1. Introduction

One of the common limitations in the production and execution of computer music is the lack of naturalness and full interactivity with timbres and compositional structures [12].
Students in the classroom can interact with the system described below in a way that is determined by the virtual interconnections between sensors and mechanisms of sound production. These mechanisms may concern the microstructure of sounds, thus allowing modifications of their timbral features, their weight and their time evolution. The overall sound-scape is accordingly modified.
On the other hand the control may be exercised and directed to compositional rules, altering in real time a predetermined musical structure or creating an entirely new one.
Based on didactical experiences devised and developed during the past fifteen years, this new project is bound to the following main educational goals:
1) acquiring a basic knowledge of sound and music parameters by means of a "physical" interaction with sound
2) creating awareness of the link between gesture and music and sound production in a 3D space
3) increasing the expressive power of both individuals and coordinated groups.

1. The VM-zone project

The VM-Zone (Virtual Music Zone) project consists in the configuration and realization of a workstation for the production of sound and music for didactical application: the real time control of sound and music synthesis is accomplished by means of sensors [4,5].
The underlying key idea is that of augmenting the input channels for sound and music control and also to give full cooperative control of the sound production mechanism to students in a classroom.
The system is built around a Personal Computer equipped with a sound synthesis card, which may be upgraded to a more powerful sound system such as a midi polyphonic module.
The system is connected to a set of sensors and actuators by means of a MIDI controller (in the actual realization an I-CUBE sensor to MIDI interface by Infusion Systems).
All the sensors are intended for remote motion or presence detection. They include ultrasonic and infrared detectors as well as pressure detectors. System modularity allows the use of any kind of sensor up to the provided number of input channels.
The task of sound synthesis is accomplished in the PC under Max/MSP environment, a powerful object oriented software, both for real time synthesis, MIDI control and even music composition; intensive use is made for the purpose of the *timeline* object [9,10,11].
The sensors used in the present project were developed in our laboratory for inexpensive applications in order to allow its use in any school environment.
In the present case we use inexpensive LDR sensors for both triggering events and controlling sound parameters. Since the sensor may detect, at least in a robust way, only a binary information (on/off) special patches were developed in order to obtain an analogue performance, by means of integration. Therefore you may increase or decrease a chosen parameter just by triggering continuously an accumulator.
Main use of the system was in the range of basic musical education, at an elementary level.

2. The actual implementation: sensors for the VM-zone

The inexpensive LDR sensors are shown in figure 1. They are simply used in a resistive partition net. The

output signal is sent to the analog input of the I-Cube interface, which in turn provides, under program control, proper values for chosen parameters of the virtual instruments.

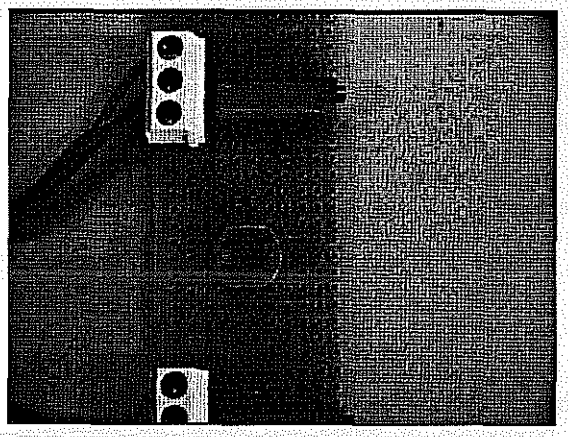


Figure 1: An LDR sensor

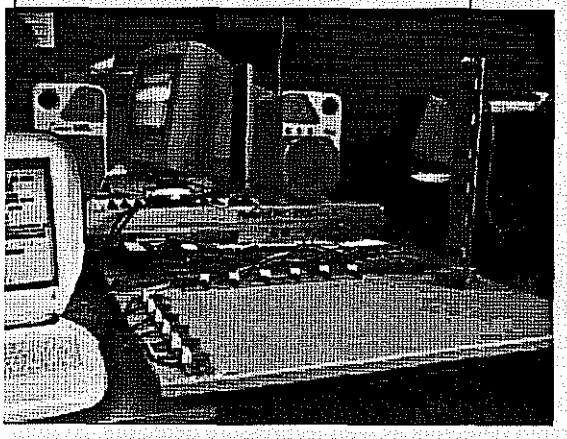


Figure 2: The array of sensors as a frame of reference

The array of sensors, arranged as in figure 2 along a three axis system, allows detection of the position of hands in the chosen space; it thus allows control of parameters by movements in the selected space. Also, by use of a derivative patch, it allows the control by means of speed and acceleration . As a reference for this kind of interactions and the possibility that it may provide for sound control and real time composition we may refer to the pioneeristic work by Leonello Tarabella [6,7,8].

Figure 3: The I-CUBE interface

3. The didactical experiences

In what follows we describe some experiments that we found very significant for our research, mostly for the *interaction with the pupils and the resulting sounds.*
We firstly developed a Max/MSP patch, shown in fig. 4, to simulate bell-like sounds, by means of an array of simple FM instruments. Sounds were to be activated by a pupil just moving his hands over LDR sensors, in order to trigger predefined events, made of single sounds.
Using this patch, the *Imaginary bells*, the pupil was able to build more or less complex sequences, making more dense or sparse chains of events.
The richness of the sounds was able to induce deep interest both in the pupil and in the whole class, due to the physical interaction with both the sensors and the synthesized sound.

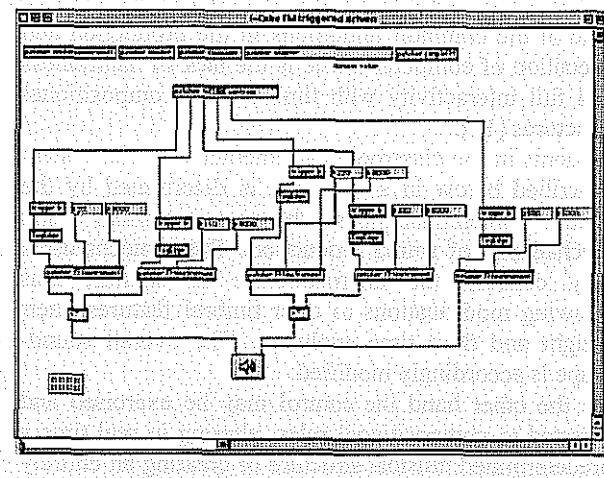


Figure 4: Imaginary bells

In a second example, *playing with beating waves*, we developed a patch based on beating sinusoids, controlled in real time by means of the same set of sensors, used, in this case, not only as a means to trigger events but also in order to simulate controlling potentiometers to modify the pitch of each oscillator.

The pupils could control the frequency of each of the oscillator pair, both incrementing or decrementing it in time. This patch was very efficient in order to develop finer pitch evaluation, and a deep understanding of the underlying acoustical phenomenon.

The third experiment used a patch for synthesizing a cluster of 128 sinusoidal oscillators. The pupils were given control of the fundamental frequency of the cluster, of its bandwidth and also updating rate.
The result was the creation of complex and rich sound textures, whose complexity and evolution changed over time. Also in this case the complexity and the novelty of the resulting unusual sound, far from the daily listening , was a rich experience for all of them.

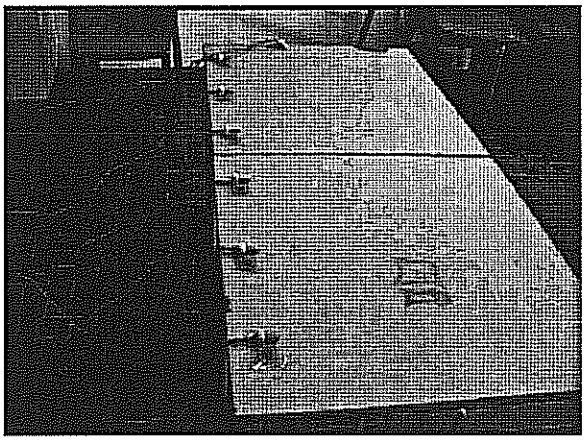


Figure 5: The sensors on the school desk

4. Further developments

As a development we are building a matrix of photodiodes, in order to provide bi-dimensional detection of controlling movements. The resulting image will be acquired by a patch in order to be translated into control parameters for the virtual instruments.
Future application will practice also real time control of sounds and even of composition, in accord with presence and movement of actors on the stage. We may refer for this kind of use to the work on a theatrical automaton [1].
These sensors are displaced in a room of approximately 10 square meters, in order to detect the movements and the shadows of the people in the room.

Finally, with the same purpose we are going to use a digital video camera, in order to obtain more complex control on the events, including presence, absence of motion, speed and others. The connection to Max/MSP environment will allow easily developing proper patches for this purpose.
As far as regards mode of use of the system as a development a second range of controls may regard high level synthesis processes:

- algorithmic composition: information from the sensors are used to modify the compositional rules of an undergoing real-time composition[2], [13][14].
- Control may be also exercised on the process of executing a score, by means of interactive modification of execution parameters [3].

5. REFERENCES

[1] S. Cavaliere, L. Papadia and P. Parascandolo, "From Computer Music to the Theatre: the realization of a theatrical automaton," Computer Music Journal, Cambridge, MA: MIT Press, vol 6, no.4, Winter 1982, pp.22-35.

[2] Rowe,R. 1993. Interactive Music Systems. Cambridge,Mass:MIT Press

[3] G. Sica, Notations and interfaces for musical signal processing, in Musical Signal Processing, Piccialli, De Poli, Roads and Pope Edts., Swets and Zeitlinger, Amsterdam, 1997.

[4] Cavaliere S., G.Evangelista, G.Sica. "Teaching Music and Acoustics : interaction by use of sensors in a Virtual Music Zone" Proceedings of the XIII CIM 2000 L'Aquila, settembre 2000.

[5] S.Cavaliere, G.Evangelista, G.Sica." Didattica della musica: un sistema basato su controlli gestuali," Proc. of the XXVIII Convegno Nazionale dell'AIA (Associazione Italiana di Acustica) , Trani, Italy, June 2000.

[6] Tarabella L., guest editor, "Special issue on Man-Machine Interaction in live performance" - INTERFACE, vol.22 n.3 - Swets & Zeitlinger B.V., 1993.

[7] Tarabella L., Bertini G., Sabbatini T., "The Twin Towers: a remote sensing device for controlling live-interactive computer music", Proc' of Workshop on Mecatronical Computer Systemes for Perception and Action ", 1997, Pisa

[8] http://www.cnuce.pi.cnr.it/tarabella/cART.html

[9] http://www.cycling74.com/

[10] Puckette, M. 1989. "Max reference manual for version 2.0." IRCAM internal documentation, 31 pp.

[11] Puckette, M. 1991. "Combining Event and Signal Processing in the MAX Graphical Programming Environment." Computer Music Journal 15(3): 68-77.

[12] J. Paradiso, Electronic Music Interfaces: New Ways to Play. IEEE Spectrum Magazine, Vol. 34, No. 12, pp. 18-30 (Dec., 1997).

[13] G. Sica, Reti neurali e loro applicazioni musicali. Konsequenz ,Liguori Editore, Napoli, 1999.

[14] G.Sica, "Linguaggi e sistemi per la Computer Music: reti neurali semplificate" Atti del workshop "Sistemi Intelligenti per l'Arte e l'intrattenimento" Napoli, Aprile 1999 su Notizie AIIA, settembre 1999.

XIV CIM - CONCERTS

PROGRAM NOTES

Jacopo Baboni-Schilingi
Le Château **(2003)**
for voice, women's choir and live electronics

Original text by Yannick Liron
Commissioned by Centro Tempo Reale

Le Château originated from a commission assigned by Luciano Berio to Baboni-Schilingi in 2000. It is a work created in collaboration between the poet and writer Yannick Liron and the Milanese composer. *Le Château* is one "scene" of a larger project for voices, instruments and electronics on which Baboni Schilingi and Yannick Liron have been working for some time. In the scene presented at Florence the main character (played by Nicholas Isherwood - solo bass vocalist) recalls a castle (from which derives the title *Le Château*) which he saw years ago. He remembers a lake beside the castle, the noise of the water, as well as the park and the windows opening onto the garden. This gentle recollection is mingled with other memories of an impossible love (represented by women's voices) for a woman he saw years ago and has never found again. As regards technical aspects, the composition employs the technique of Composition through Interactive Models - for the part dedicated to the voices - which provides a control of morphological type over the composition material. For the electronics instead Baboni Schilingi has utilized a system of granular synthesis through which the voices are processed in such a manner as to give the impression that the electronics anticipates the voices and not - as is frequently the case - the contrary.

Elisabeth Bossero
Dino

Composed in 2000, *Dino* was technically re-elaborated in 2002. The starting sound for the entire composition is merely a bandoneón chord taken from a composition by Dino Saluzzi called *Memorias.*
The formal project of the piece is based on a slow "recomposition" of disintegrated elements, far removed in timbre from the source, which break out into the evocation of a tango.

Matthew Burtner
S-Morphe-S
for a singing bowl soprano saxophone

S-Morphe-S explores the coupling of a disembodied soprano saxophone with the virtual body of a singing bowl. The saxophone signal is used as an impulse to the physically modeled bowl by Stefania Serafin. The result is a hybrid instrument with the articulatory characteristics of a soprano saxophone but the body of a singing bowl.
The saxophone uses varied articulations such as key clicks, breath, trills and sustained tones. The shape and material properties of the bowl are varied in real time creating a continuously metamorphosizing body.
In Greek Morphe means form and in Greek mythology Morpheus was the god of sleep, of disembodied forms. The english word commonly used for a transformation between two objects is morph, a shortening of metamorphosis, derived from the Greek. The title of this piece is meant to evoke all of these meanings - dreamed images, transformative bodies, and disembodied forms.

Nicola Buso
(...) wo sterbend die Sonne (...)
for flute, violin and live electronics

"[...] wo sterbend die Sonne [...]"
(G. Trakl, Die Schwermut)

"[...] Now hast thou but one bare houre to live,
And then thou must be damn'd perpetually.
Stand still you ever moving Spheares of heaven,
That time may cease, and midnight never come. [...]"
(Ch. Marlowe, Faust, vv. 1927-1930)

"[...] Long stay'd he so;
At last, a little shaking of mine arm,
And thrice his head thus waving up and down,
He raised a sigh so piteous and profound
That it did seem to shatter all his bulk
And end his being; that done, he lets me go,
And with his head over his shoulder turn'd
He seem'd to find his way without his eyes;
For out o' doors he went without their help,
And to the last bended their light on me. [...]"
(W. Shakespeare, Hamlet, Act II, Scene I, Ophelia)

"[...] Or thou might'st better listen to the wind,
Whose language is to thee a barren noise,
Though it blows legend-laden through the trees. [...]"
(J. Keats, The fall of Hyperion, Canto II, vv. 4-6)

"[...] Als der Abend des ersten Kampftages gekommen war, ergab es sich, daß an mehreren Stellen von Paris unabhängig von einander und gleichzeitig nach den Turmuhren geschossen wurde. [...]"
(W. Beniamin, Über den Begriff der Geschichte, XV)

John Cage
Aria
for voice

The score consists of 20 pages of music, each page being sufficient for 30 seconds in performance. Pages may be performed over longer or shorter time-ranges to create a program of a determined time-length. The text employs vowels and consonants and words from Armenian, Russian, Italian, French and English. The notation consists basically of wavy lines in different colors and 16 black squares denoting "non musical" vocal noises. The colors denote different singing styles, to be determined by the singer. Cage used *Fontana Mix* as a composing means to create this Aria.

Fabio Cifariello Ciardi
Prologo

Prologo is the first of several short sonic landscapes. Listeners are invited to the experience of "passing through". Possible spaces, invisible places? Is the narration of concrete memories or the surrealistic concealment among a "forest of symbols"? "... the casual meeting on a operating table of a sewing-machine and an humbrella"?

Riccardo Dapelo
Two studies on digital synthesis of images and sound
(Installation version)

For some time now I have sensed the need to explore the interaction between image (generally abstract) and sound. Naturally it is not a case of adding sound to pre-existing images or the reverse. My interest is the exploration of the generative moment; that is, an environment-system in which these two worlds come into existence and interact simultaneously at the level both of conception and of perception. A hazard may well be hidden in this intention: in our social system of communication image has predominance (if not absolute power) over sound. My intention is not to align myself with this tendency but rather to refine the perceptive characteristics of a system of communication through the resonances of visual and aural stimuli, the possibility of transferring techniques of sound synthesis/transformation to the image and vice-versa. It could be said, from the point of view of the musician (sound designer, sound sculptor), that the aim (as far as listening is concerned) is to insert the parameter image in the global composition. The problem immediately becomes complicated, since the image (either dynamic or static) is already in its turn "composition". However, it is a problem (in the case of image in movement) of "composition" as function of time; consequently it can be treated with compositional parameters that are generally applied to sound composition: dynamics, tension, texture, accumulation, dispersion, density. These two short studies are the first stages of this exploration.

Studio I - Variations of complex numbers over a green space
This study explores n-dimensional spaces, generated by means of complex numbers over a continuously changing green background. The application of complex numbers states the overall dynamic shape of the images, as well as shadowing and lighting position

Studio II - Dancing line patterns on the borderline between darkness and light
The second studio is based on dancing line patterns, trying to create anthropomorphic (or simply expressive) movements with a simple element (a line). This study is a preliminary exploration for a system in which the line movements will be controlled in real time by a dancer.

Installation version
In this version the two studies are interpolated, mixed and super imposed each others with particular compositing effects and different sound generation processes, with the aim to explore every nuance of the composing materials (the relationship in the image/sound synthesis).

Agostino Di Scipio
Ecosistemico udibile n.1 (Impulse Response Study)
for live electronics

A little "audible ecosystem", made of sound powders, residual sonic events, abrasive textures of variable density, precisely arranged over several time scales. A "sonorous niche" is created during the performance, emerging from the interaction between a computer-controlled DSP unit and the room or hall hosting the performance and the audience. All exchanges between DSP unit and the external ambience takes place *in sound*, as it is mediated by microphones placed around the hall. "Ecosystem": a composed network of objects and functions considered in their symbiotic/adaptive relationship to the surrounding ambience. *Ecosistemico udibile n.1* is the implementation of a real-time process which makes something to the space and, at the same time, is subjet to what the ambience makes to it. Not a naturalistic image, not a virtual space, but the unfolding of sound in close contact to the material and historical place, the unfolding of the structural coupling of a system to its ambience (the two, together, make up the specific "ecosystem"). Almost a microsonic "chronicle" of the real space, which is actually "listened to" as the dynamical shapes of sound, as timbre - *here and now*, space is not the object of representation, not something out there that we presume to be able to describe, to represent, to name. Space is experienced in what it makes to the piece, and the latter is only experienced in what it makes to the host ambience.
At the core of the performance there is a feedback process at sub-audio range, mainly concerning the real-time analysis and generation of control signals. The "sensitivity" of the DPS unit to the room resonances changes in time, and self-regulates as it constantly adapts itself to the different emergent properties of the overall sound fabric. A good performance of this work is one where the interconnection of the system components - room resonances, the microphone mediation, analysis and

feature-extraction processes, and the audio signal processing - develops across a large variety of system states. That is heard as micro-time variation (timbre variation, density in textural constructs, etc.)
The raw sonic material - the impulse construction used to solicit and excite the room resonances - was synthesized with Curtis Roads' *PulsarGenerator* program, at CCMIX (Centre de Creation Musicale Iannis Xenakis), Paris, April 2002. The implementation of the real-time process (all signal processing, including the signal analysis and feature-extraction methods, and the real-time synthesis of control signals) was programmed by the composer with the Kyma5.2 orkstation, Summer 2002. First performance took place in Stoke-on-trent, October 2002.

Roberto Doati
Il domestico di Edgar
A ruled improvisation for alto saxophone and tape (*Octandre* ad libitum)

If we could hear all the sounds of the world,
we'll immediately go crazy.
Charlie Parker

In 1995 I was requested from Claudio Ambrosini to write a work with instrument and electronics to be performed by his Ex Novo Ensemble. It had to be part of a collection of works commissioned to different composers with peculiar indications: they could have been arrangements from pop or jazz standards, with or without improvisation. I decided to work with a well-known italian jazz saxophone player -Pietro Tonolo- to a piece where improvisation and electronics were closely connected each other, but within the classical XX century musical language. The concert with the first performances was fixed in the following year at Sale Apollinee, in the Gran Teatro La Fenice in Venezia. When the Teatro burned in January 1996 I was at the beginning of my work, and the bitterness and despondency that took me because of the loss of such a cultural and affective heritage were so strong that several times during the following years I tried to conclude it. In 2002 I can consider it finished but not complete, just as a "work in progress" could be, exactly as the theater rebuilding works are: until today they are not completed.
It is well known that in the last years of his life, Charlie Parker was more and more interested in classical XX century music. Once he called Edgar Varèse, asking him to have composition lessons. His wish was so strong that he volunteered to be Varèse waiter in case the money he offered had been considered not enough from the french composer. Finally Varèse accepted, but starting after his imminent trip to Europe. When Varèse came back, Parker was dead.
In my piece I try to bring this never happened meeting The electronic part is based on Varèse *Octandre*, both from the sonological and formal point of view and on this tape -or the recording or real perfomance of *Octandre*- the saxophonist has to play following a score containing improvisation outlines often recalling be-bop style. Before the performance the solo saxophone is recorded, and some fragments are computer transformed and added to the electronic part for a next performance. So when a new saxophone player will perform the piece, he/she will improvise also on a previous improvisation. This rebuilding and "ruins" overlaying process will proceed while there will be a new saxophone player wishing to perform the piece.

Franco Evangelisti
***Incontri di fasce sonore* (1957)**

This is perhaps the most lively composition of that early Cologne period. The "encounters" of "sound-bundles" unveil themselves as dramatic confrontations, intensified by biting chains of impulses. *Incontri* is an outstanding example of a dialectic attitude towards serialism. Formal rigour (serial determinism) and freedom of decision penetrate each other in such a way that Evangelisti's composition rises above the static "pointillistic" level of early serial music.

Francesco Galante
Retroscena - memoria di una voce

The theater of *Phonè*, therefore as it was meant from the famous italian author and actor Carmelo Bene, is the pretest in order to realize a possible example of "acousmatic theater" by means of electroacoustic music. Some audio samples of his voice are used and they come from the entitled Carmelo Bene's opera *Maiakovskij* (1980). The sound voice is granulated and various noise-organized morphologies were produced in that way. Then I digitally explorated the granulated audio files as one sculptor impresses space and energy forms to the matter. In addition the sampled voice appears in the piece as a "sound photogram", a sequence of photograms able to model a reinvented dramaturgical line within an episode of electronic music. I believe there are many analogies between the acousmatic thinking in electronic music and the 70's advantgarde theater by C. Bene. He has explored a sort of "theater of the listening" by using of the sound of word-text, and his main objective was to cancel the visual performance and the scene.
On this hypothesis I have realized this homage to the memory of Carmelo Bene. By an electronic thinking I tryed to expand his idea of theater. The form of the piece is like an "ostinato", but the sound is broken by actor's voice or by silence, and the dramaturgia it is developed by the use of the both original and transformed voices, and it is transformed until the pure decomposition. It becomes a sound texture while the noiseorganized structures developes themselves in a progressive sequence of high density energy microforms in similar way to one molecular reaction.
Granular synthesis, amplitude granulation, filtering, pitch altering, reverberation are used in the piece and my work has been one exploration, perhaps, in the backstage of C. Bene search. Carmelo Bene died in March 2002.

Thomas Gerwin
Feuer-Werk (Fire Work)

The sound material of this concert piece originates almost solely from fire. Every structural elementsof the composition are inspired from a pure fire sound, which can heard in its simplicity and with all its inharent variaty in the first third of the piece. The longer I listened inside this sound, the more complex and interesting sound conglomerations, hythmic patterns and even wide reaching, long sound movements and mutations appeared in my âinner ear'. This inner point of view(listening) was even made deeper then by filtering the material in different
manners. To open this hermetic perspective, I introduced other sounds, which relate to heat and fire later, such as lighting a match, floating lava or sounds from an earthquake. These sounds are also structured by following the found forms in the formerly mentioned âpure form of fire'. This way the whole piece is the discovering and modelling of the inner musical circumstances and possibilities of
a simple fire sound - from its smallest sound particles up to the overall form.
I am aware of the fact, that with this piece I only started a longer process of the investigation of âfire as musical phenomenon'.
Production: inter art project - studio for media art Berlin

Pietro Grossi
Polifonia mix
Tempo Reale version

During the Seventies the pioneer of Italian computer music Pietro Grossi disegned a series of automated sonic processes that leaded to the creation of single compositions (as Polifonia, Monodia, Unending music) or series of different pieces of the same logic (Sound Life, Unicum, Mixed Unicum). During the recent inauguration of the Renzo Piano's Auditorium in Rome (april 2002), Tempo Reale staff included Grossi's music in the big sound installation realized in the foyer where three of the above-mentioned pieces were mixed and moved in space; this musical structure was given the name of *Polifonia mix*.

Pietro Grossi
***Collage* (1965)**

This piece is particulary representative of the period of the Sixties when Grossi founded the S 2F M electronic music studio in Florence. Several "pieces of tape" (excerpts of electroacoustic compositions received from all over the world) were used to create a musical collage that is possible to perform in different version and durations, as many of later Grossi's pieces.

Panayiotis Kokoras
Response

The composition "Response" for tape was composed during the summer 2001 at both studio "Métamorphose d' Orphée" M&R, Ohain/ Brussels and Postgraduate Electroacoustic Studio at York University, UK. It is the second piece of a project in process called "Grand Piano Trilogy". The main characteristic of this trilogy is that the source of the sounds/ samples comes from piano only. I decided to have a unified sound source in order to go deeper in the sound and the structure of it, to find out its own gravities and tensions to explore the phenomenology of the sound/ timbre itself.
A great variety, of artificial and natural responses triggered out by energetic impulses and resonators, characterize the piece. The response vibration may be a simple harmonic motion based on a minor second or some more complex action by distorted, inharmonic textures. The response's impulse may be as short and simple as a click of a spire along the string, a cluster by a modified hammer inside piano, a damped or pizzicato note. In that case the sound material manipulated in the time domain via convolution, granular processes, time- stretching, etc. Moreover, the response may be an elaborate resonant structure itself. The energy is applied as a repeated stream of pushes functioning as sound generator. The sources used are circular sweeping and accelerated strumming sounds inside the piano with different material like glass, plastic. The processing techniques applied are in frequency domain FFT-based cross synthesis and analysis/ resynthesis, as well as more standard signal processing such as harmonizing, frequency shifting, phasing, specialization etc.
The form structure of the piece is an ongoing development and transformation of the initial idea. Sections with pitch implementation followed by inharmonic one and so on. Most of the samples are recorded in small fragments; thus, it is necessary to work very close to sound, in a kind of micro montage-mixage, to develop unique timbral interactions between them.

Beatrice Lasio
...di vetro

I like to play with marbles.
I find it irresistible to watch these shiny multicoloured glass marbles that spin and knock against each other...besides, the artificial light gives a magical effect to their movements. I also find it irresistible to have some of these smooth and cold marbles in my hand and barely move and squeeze them to cause a slight sound as I bring them close to my ears.
And then?
I can capture the sound of these marbles with a microphone and transform it in various ways and even combine these transformations.
This game is like magic. It astonishes me.
All sorts of objects appear. It seems as though Silence produces them.
Nothing to see, smell, taste or touch. But I feel consistencies, thicknesses, colours, movements, lights, transparencies, distances...
The objects correlate with each other. What are the criteria?
How can I make use of identities, resemblances or contrasts?
I aim at a coherence. What do these sounds suggest? I try with varied repetitions of some element throughout the construction.
The sounds I have obtained have similar and common features: they are iterative and irregularly crumbled, and occasionally they fluctuate slightly in pitch; besides certain of their modulations evoke some prosodic trait of the declamation. I try to emphasize these aspects and impose a principal rule: the melodic fluctuation in the range of a second.
The acrobatics of flocks of birds in a piece of the city sky at dusk come to my mind. I believe this image suggests something to me.
To start with, I'd like a dense and rich flow of somehow chaotic contrasts, which gradually tends to thin out. And I'd like

this movement to be disguised by is contrary, just like a figure and a background that switch roles inconspicuously. Then I'd like surprise to bring a whole new different context: a vast space with a persistent iteration where the variations are slow and predictable. Therefore, I would still make use of surprise to re-evoke the beginning, with vague hints, and eventually generating expectation. Of what? Of a juxtaposition of identity and contrast in which order is paced, in a wonderful chaos, with a subtle touch of childish ingenuity.
I listen and listen again with headphones. I never have the same impressions and sometimes these are contrasting...
What a curious game!
I try with loudspeakers. I have a pair. Let's see what happens...other sounds. Sometimes entirely unknown objects appear but I refuse acceptance of all of them. I try with a different pair of loudspeakers and, then again, I surprise myself approvingly and disapprovingly.
I'd like to try with many other different loudspeakers, experiment with a large number of them, in different positions with a different setting, so as to organize the sounds.
I would also like to try with strange sculptures of any possible material or form, which diffuse and project sounds in their own particular nature.

Bruno Maderna
Musica su due dimensioni
for flute and tape
Musica su due dimensioni for the first time draws live music near music modified by electroacoustic effects. A synthesis of both the existing possibilities, that I call "dimensions", seems to me especially fruitful, due to the fact that the performer - meeting the sonorous materials fixed on tape, created by the composer or controlled by himself - reaches a deeper contact with the author (in fact he not only reads the score, but at the same time hears what the composer wants). On the other hand the author must to fulfil the same synthesis in himself if he wants to create such a complex form, in which the immediate interpretation and what he fixed meet theirselves. (B. Maderna).

Matteo Malavasi
Sordo grido nel ricordo
For hotebois, saxophone and tape

As breath turns to a breeze, and the breeze to nature, the ear plunges into the soul of the woods, into the heart of the storm, into the echo of something - or someone - wafting through the music. Then, suddenly, it is silence again. And so the short season ends, as lightly as it started.
(Riccardo Giandrini)

Marco Marinoni
In the blinded room

In the blinded room is part of a triptych of works for tape alone, which have the peculiarity of assuming as starting material for sound processing the recording of some poems read by the author. In the case of *In the blinded room* the poem is *Soledad* by Rober Hayden (the other two works, *And sound alone* and *The dying of the light* involve materials from, respectively, *The idea of order at Key West* by Wallace Stevens and *Do not go gently into that good night* by Dylan Thomas).
The morphologic characteristics determining the structural direction of these works are:

1. (micro-formal - gestural level) ascending negation of source bonding through non-linear (adirectional) processes of
    a. fragmentation (*confraction*),
    b. stratification (*diffraction*),
    c. diffusion (modulation of concentration gradient);
2. (macro-formal - textural level) monodirectionality and non reversibility of directional motions;
3. (perceptive - esthesic level) integration of vertical dimension and horizontal dimension (*gesture framing*)

    In particular, two are the ideas at the basis of *In the blinded room*:
    a. defocusing: fragmentation of time axis and superimposing of more time axes;

b. interaction between circular structure (cyclic iteration of objects exposed to micro-variations) and linear structure (process' mono-directionality)

SOLEDAD by Robert Hayden
Naked, he lies in the blinded room
chainsmoking, cradled by drugs, by jazz
as never by any lover's cradling flesh.
Miles Davis coolly blows for him:
O pena negra, sensual Flamenco blues;
the red clay foxfire voice of Lady Day
(lady of the pure black magnolias)
sobsings her sorrow and loss and fare you well,
dryweeps the pain his treacherous jailers
have released him from for a while.
His fears and his unfinished self
await him down in the anywhere streets.
He hides on the dark side of the moon,
takes refuge in a stained-glass cell,
flies to a clockless country of crystal.
Only the ghost of Lady Day knows where
he is. Only the music. And he swings
oh swings: beyond complete immortal now.

Pino Monopoli
BrACe

BrACe (Brano con Automa CEllulare) is the result of my personal experience of interaction and integration between a rigorous construction of musical form and an empirical attention to the acoustic properties of the sonic material. The compositional process of the piece is in fact founded on a continuous exchange, at several levels, between algorithmic procedures and the hearing of the sonic material. The algorithm used is based on cellular automata, and it holds a central role in the composition. First of all, it is a process for generating musical forms: the automaton made the initial sounds evolve in time, generating musical structures. At a higher level of formal organization, moreover, the structures have been connected with one to another according their own sonic properties, 'disregarding', in one sense, the formal aspect below: hence, in this meaning, the algorithm is also a process for generating sounds. The first step in the realization of the piece has been the design and the implementation of eight digital synthetic instruments, using the Csound language. All instruments have a sonic common characteristic: the 'noisy sound', for some instruments intended as 'white noise', filtered by some types of filters; for other instruments intended as 'glitch', as 'digital noise' generated by the discontinuous points in non-sinusoidal sound waves. The instruments are ruled by Csound scores, with an undergoing cellular automata logic, generated by a computer program purposely designed and implemented by me. Several tasks have been assigned to the cellular automaton, making different types of associations between it and some characteristics of the eight digital instruments: instruments activation and inactivation, stereo positioning of the sounds, selection of amplitude and frequency gestures, initial and final values for these gestures. First of all, a certain number of audio fragments have been generated; then, I start creating the piece by selecting, assembling, superimposing and placing side by side them, according to sonic relations suggested by hearing, in a certain conformity with differentiation and integration criteria. Furthermore, new form-oriented fragments have been generated, where the growing piece requested it. This is the bivalent role of the automaton: on the one hand as a tool for generating fragments which intrinsically suggested several formal relations; on the other as a tool for generating fragments which can be oriented by specific formal needs. So, the automaton has been the presupposition for the formal structuration of the piece, even if in a merely aural and non-algorithmic sense, at this level. In the end, the generated audio materials revealed a certain affinity with the sonic material of *Concret PH*, piece for solo tape realized by Iannis Xenakis in 1958. Different sound generation methods, different formal construction, but very clear perceptual similarities. This historical electronic piece have been explicitly quoted in my work: some fragments have been selected and assembled in an evident perceptible manner, as a reverent tribute to Xenakis. Which said: "someone thinks that my music originates entirely by calculi, but before declaring an opinion it should be necessary to listen to it. In no music, in fact, the calculus can exercise a total control."

NPS (Alfonsi/Chiggio/Marega/Rampazzi)
***Ricerca 4* (1965)**

The *Ricerche 1, 2, 3, 4* are realized with close groups of frequencies that proceed maintaining the constant duration of 10 seconds each, departing gradually from the middle zone of the sound spectrum and diverging so far as to reach the two audible extremes in the highest and lowest zones. The groups proceed without pause. *Ricerca 4* has been completely reverberated and the dynamics results naturally from the density of the groups.

Teresa Rampazzi
***Fluxus* (1979)**

"We step into the same river
but not into the same waters.
Impetuous currents
from opposite directions
have struck us
but have not swept us away.
Tamed, the great
wild river
has become a form." (Heraclitus)

Fluxus was born of the ambitious desire to search for and find a form that would be more suited to the means we possess today for making music, to the technological level and world view that have together called for and led to the creation of these means.
The computer suggests forms that can repeat no past and that reflect our philosophy of Heraclitean stamp, where nothing can ever turn backward.
The acoustic signals thus flow one from another and one after the other in a ceaseless process of evolution.

Giuseppe Rapisarda
Almaquae

Imagine you hear a drop...
...a single drop of water could seem unimportant because it is like only one star in the sky...
...but it's a door that allows you to go inside the soul of water: a virgin world that keeps many undreamablesecrets.
A drop of water sometimes lights the most intense feelings and raises life from desert.

Jean-Claude Risset
Saxatile
for soprano saxophone and tape

Saxatile is dedicated to Iannis Xenakis on the occasion of his seventieth birthday. The tape was realized at the Ateliers UPIC in 1992. On the UPIC synthesizer, the frequency curves of the music can be specified by drawing. The tape includes some graphic allusions to *Metastasis*.
The title means "living among the rocks": it evokes the relations between the saxophone and the tape as encounters between the biological and the mineral realm. Initially the tape sounds revolve around a given pitch, then they glide and drift, and finally they are dispersed as grains. Despite this diversity of morphology, they come from the same realm, like strata, rocks and sand. Meanwhile the saxophone lines keep a quasi-biological suppleness.
The author thanks Daniel Kientzy, Gerard Pape, Didier Rocton, Brigitte Robindoré, Marie-Hélène Serra for their help on UPIC, and Solenn.
The tape for *Saxatile* was entirely realized from drawings realized on the tablet of the UPIC graphic computer music

system, built and programmed according to the specifications of Iannis Xenakis. UPIC makes it possible to specify waveforms, amplitude and frequency envelopes by drawing them. One can also use as waveform sine waves or portions of a recorded wave. I used the latter possibility - from saxophone tones recorded by Daniel Kientzy. Mainly, I specified curves for the evolutions of frequency - initially horizontal straight lines with constant pitch; later, drifting pitches along straight line glissandi as well as other curves; toward the end, patterns similar to the spectral analyses of chords which I used in computer pieces such as *Little Boy* or *Mutations*.
Saxatile performed by Daniel Kientzy, appears on the CD in Homage to Xenakis: Xenakis/UPIC/Continuum, C.D. Mode, New York, distribution Abeille.

SMET (fondatore Enore Zaffiri)
***EL/25* (1966/67)**

SMET: Project EL/25 (1966-67)

The piece known as "Project EL/25" was created in the experimental course in Electronic Music conducted at the Studio di Musica Elettronica of Turin founded by Enore Zaffiri, during the academic year 1966-67. To the young students approaching electronic music for the first time, a linguistic-formal research methodology was proposed. The basic premise was that of carrying out an investigation based on extremely simple material, reduced in the course of all of the experiments conducted in the first year of study, to a single element provided by electronic means: the sine wave, processed as a sonic beam, as pulse, as glissando sound, along with the parameters of intensity and time. The methodology applied consisted of formulating a basic structure of geometric nature, from which to obtain all of the relationships necessary to carry out the research. The "project" is an organization of sounds in which certain sounds relate to one another according to determined laws; laws which must be considered on the same level as arbitrary postulates.
Project EL/25 is a study of 25 glissando sinusoidal sounds, organized by the figure of the ellipse. The sound event consists mainly of two contemporaneous glissandi for opposite movements which start from two focuses of the figure and, in completing their trajectory, encounter established points, from each of departs a glissando traveling in the direction opposite that of the main one. (A detailed description of the Project is found in the publication "Duescuole di Musica Elettronica in Italia", published by Silva, Genoa 1968.)

Martin Supper
Fragment
Speaker: Hanns Zischler

Durations, dynamics, spatialization, and degree of transformation of the voice were all determined by structural elements of several texts chosen from the works of Roland Barthes und Michel Serres.

"Die rauschende Masse einer unbekannten Sprache bildet eine delikate Abschirmung; sie hüllt den Fremden [...] in eine Haut von Tönen, die alle Entfremdung der Muttersprache vor seinen Ohren haltmachen läßt..." (Roland Barthes) "Die Dinge sprechen in Zahlen. [...] So war denn die Musik die erste strenge Physik, de erste von der Physis beherrschte Linguistik. Wen kann es da wundern, daß die Grundlagenphysik es uns ermöglicht, eine andere Musik, oder sagen wir besser, das Rauschen der Dinge selbst zu schaffen oder zu verstehen." (Michel Serres).

"The murmuring mass of an unknown language constitutes a delicious protection, envelops the foreigner ... in an auditory film which halts at his ears all the alienations of the mother tongue." (Roland Barthes). "Things speak in numbers... and so music was the first strict physics, the first naturally ruled linguistic. Who can be surprised that the fundamentals of physics made it possible for us to create a new music? Or perhaps we should say that it allowed us to create the very murmurings of things themselves, or to understand them." (Michel Serres).

Stefano Trevisi
Swallow
for female voice and electronics

Swallow is an open project about phonetic sounds: at the moment it consists of two independent sections, the first one which is rather gestural, and the second one which is characterized by a gradual cyclical movement. Both sections are a study on how sound materials can be perceived by overlapping them, basing on different timestretching processes, to create different levels of energy from constructive organic processes (when different layers have the same phase) and destructive ones (coinciding with opposite phase).
The research is based on phonetic gesture which results from spoken materials, excluding components of singing. It has been developed from a poetical text, Masticazione, by Claudia Castellucci, which is characterized by a pregnant phonetic rhythm, although it is not an onomatopoeic text, and by a strong gesture. The poem contains a detailed description of the chewing and the swallowing of a fig, and it has been completely crumbled and then re-coagulated basing on phonetic movements generated by the performer and on the violent gesture which bursts out from the text.
Tape materials have been recorded at Tempo Reale (Florence); the tape has been composed atthe Electronic Music Lab of the Conservatory of Music "A.Boito" (Parma); the voice has been recorded and mixed at the GrocLab (Barcelona).

Regione Toscana

Comune di Firenze

RAI - Radiotelevisione italiana

Ministero per i Beni e le Attività Culturali

AIMI - Associazione Italiana di Informatica Musicale

TEMPO REALE