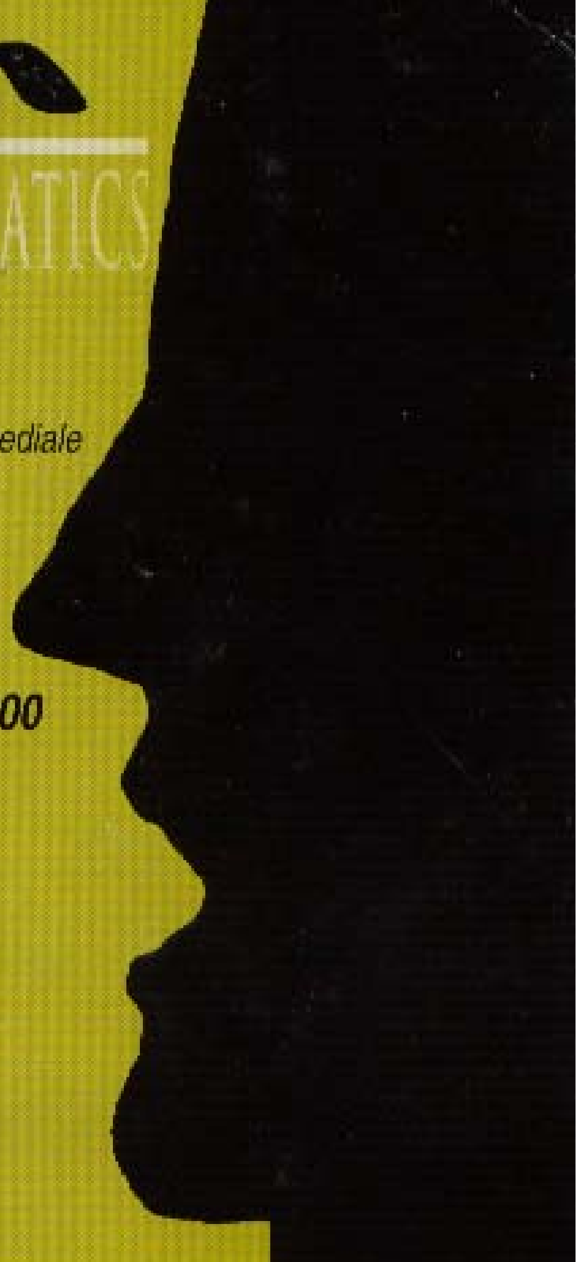
XIII CIM

COLLOQVIVM ON MUSICAL INFORMATICS

Spanish Castle *Forte Spagnolo*

Conference Hall *Sala Conferenze* Multimedia Hall *Sala Multimediale*



L'Aquila

2nd – 5th September 2000 ***2 – 5 Settembre 2000***

AIMI
Associazione di Informatica Musicale Italiana

Istituto GRAMMA

PROCEEDINGS

XIII CIM

COLLOQVIVM ON MUSICAL INFORMATICS

Spanish Castle *Forte Spagnolo*

Conference Hall *Sala Conferenze* Multimedia Hall *Sala Multimediale*

L'Aquila

2nd – 5th September 2000 *2 – 5 Settembre 2000*

Edited by *A cura di*

Maria Cristina De Amicis

AIMI Associazione di Informatica Musicale Italiana

Istituto GRAMMA

With the collaboration of *Con la collaborazione di*

Università dell'Aquila - Facoltà di Ingegneria
Dipartimento di Ingegneria Elettronica

Conservatorio di Musica "A. Casella" - L'Aquila
Parco Scientifico e Tecnologico d'Abruzzo
Federazione CEMAT - Roma
Motus - Francia
Soprintendenza B.A.A.A.S. - L'Aquila
te Musicale Società dei Concerti "B. Barattelli" - L'Aquila

CRM - Centro Ricerche Musicali - Roma
Computer Music Journal
A.P.T. Abruzzo Promozione Turismo

Sponsoring authorities *Con il contributo di*

Ministero per i Beni e le Attività Musicali
Dipartimento dello Spettacolo

Istituto Centrale per i Beni Librari,
gli Istituti Culturali e l'Editoria

Regione Abruzzo
Assessorato alla Promozione Culturale

Comune dell'Aquila
Assessorato alla Promozione Culturale

SCIENTIFIC COMMITTEE *COMITATO SCIENTIFICO*

Daniel Arfib	(LMA-CNRS - Marseille, France)
Chris Chafe	(CCRMA, Stanford, USA)
Roger Dannenberg	(Carnegie Mellon University, USA)
Giovanni De Poli	(University of Padova, Italy)
Giuseppe Di Giugno	(Federazione CEMAT, Italy)
Shuji Hashimoto	(Waseda University, Japan)
Douglas Keislar	(Computer Music Journal - MIT Press, Berkeley, USA)
Colby Leider	(Princeton University, Princeton, USA)
Marc Leman	(University of Gent, Belgium)
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)
Leonello Tarabella	(CNUCE-CNR, Pisa, Italy)

MUSICAL COMMITTEE *COMITATO MUSICALE*

Nicola Bernardini	(Conservatorio di Padova, Italy)
James Dashow	
Maria Cristina De Amicis	(Istituto GRAMMA, L'Aquila, Italy)
Agostino Di Scipio	(Conservatorio di Bari, Italy)
Francesco Giomi	(Conservatorio di Cagliari, Italy)
Michelangelo Lupone	(Conservatorio di L'Aquila – Istituto GRAMMA, Italy)

ORGANIZING COMMITTEE *COMITATO ORGANIZZATORE*

Maria Cristina De Amicis	(Istituto GRAMMA, L'Aquila, Italy)
Agostino Di Scipio	(Conservatorio di Bari, Italy)
Michelangelo Lupone	(Conservatorio di L'Aquila – Istituto GRAMMA, Italy)
Angela Panella	(Secretariat, Istituto GRAMMA, L'Aquila, Italy)
Laura Anselmi	(Secretariat, Istituto GRAMMA, L'Aquila, Italy)

XIII CIM

COLLOQVIVM ON MUSICAL INFORMATICS

Spanish Castle Forte Spagnolo

Conference Hall Sala Conferenze Multimedia Hall Sala Multimediale

L'Aquila

2nd – 5th *September* 2000 2 – 5 Settembre 2000

Promoted by Promosso da

AIMI - Associazione di Informatica Musicale Italiana

Organized by Organizzato da

Istituto GRAMMA - L'Aquila

Supported by Sostenuto da

Ministero per i Beni e le Attività Musicali
Dipartimento dello Spettacolo
Istituto Centrale per i Beni Librari, gli Istituti Culturali e l'Editoria

Regione Abruzzo
Assessorato alla Promozione Culturale

Comune dell'Aquila
Assessorato alla Promozione Culturale

Università dell'Aquila - Facoltà di Ingegneria
Dipartimento di Ingegneria Elettronica

Conservatorio di Musica "A. Casella" - L'Aquila
Parco Scientifico e Tecnologico d'Abruzzo
Federazione CEMAT - Roma
Motus - Francia
Soprintendenza B.A.A.A.S. - L'Aquila

Ente Musicale Società dei Concerti "B. Barattelli" - L'Aquila

CRM - Centro Ricerche Musicali - Roma
Computer Music Journal
A.P.T. Abruzzo Promozione Turismo
Compas - L'Aquila

This book of proceedings was made possible by kind cooperations of *Ha contribuito alla pubblicazione degli Atti*
Federazione CEMAT- Roma

Istituto GRAMMA • Via degli Scardassieri 14 67100 L'Aquila Tel. +39.0862.401209.414308 Fax +39.0862.412950
E-mail: info@istitutogramma.it Http://www.istitutogramma.it

INTRODUCTION

Maria Cristina De Amicis

Artistic Director
Istituto GRAMMA – L'Aquila

Three modes of cultural action

Popularisation

For the Italian community of researchers and musicians, the Colloquium on Musical Informatics (CIM) is a unique opportunity for mutual exchanges in the area of computer music technology. Over the years, the responsabilities over the organization have been growing; musical and scientific topics have been increasing in number and broadening in scope; the size of the audience attending concerts and lectures has become significantly larger. In my view, this means that the aims pursued are of significance to contemporary society, and that they establish a real interaction between science and art, at many levels. The many University and private centers that have so far hosted the Colloquium, have all contributed to enlarge the table of contents offered to attendees, but they also managed to let a composite and dynamical discipline such as musical informatics be better known by the people. Among the many elements that have contributed the most to the popularisation, I want to highlight a particularly strong integration between scientific sessions and concert performances. These two elements, together, proved able to reach a heterogeneous audience, and to give the non-experts a chance to get deeper in the technical details. Clearly, by providing a venue for the meeting of researchers and musicians, CIM also highlights common grounds and directions, as well as the overall social relevance that the discipline is able to gain.

Emergence

Since the beginning, the very mission pursued in organizing the CIM has been one of presenting and discussing advanced research works in the field. Both scientific developments and the opening of new fields of application, make this event an opportunity to check the current vitality of scientific and musical search. The themes peculiar to the colloquium, as well as the array of musical directions illustrated in the concert programmes, let the significance of present directions emerge in their most innovative theoretical and technological forms. This mission renews every two years, based on the careful screening of scientific and musical contributions, handled by the CIM international committees. However, it gets also developed by casting most of the contributions against a specific theme of actual interest to the international community.

Discussion

As its thematic focus, the 13rd CIM features a special session on the didactics of computer music, where paper presentations will give way to a discussion among all the attending delegates. This particular themes was proposed not by chance, as it will allow the CIM to be smoothly integrated with this year's La Terra Fertile ("The Fertile Land"), a conference held every two years in L'Aquila, providing Electronic Music students from all over Italy (and abroad) with an opportunity for illustrating their compositional and musicological work. Research developments should find their own terms in suitable educational programs. Timely reviewing of the goods and bads of educational programs is absolutely appropriate in this field, as it allows teachers and researchers to better keep in touch with what is relevant for educational purposes.

Personally, the organisation of CIM gives me one more opportunity to understand the vitality and the ideals that pervade the area of musical informatics. In preparing this event, I have already worked side by side, and incessantly, with Italian and foreigner researchers that share with Istituto GRAMMA the aim to offer a suitable venue for establishing closer connections between musical informatics and issues of relevance to contemporary society. I wish to sincerely thank all those friends, as I am confident that our work so far set forth the best premises for this meeting to become an opportunity of intellectual enrichment and insight.

L'Aquila, September 2000

INTRODUCTION

Maria Cristina De Amicis

Artistic Director

Istituto GRAMMA – L'Aquila

Three modes of cultural action

Popularisation

For the Italian community of researchers and musicians, the Colloquium on Musical Informatics (CIM) is a unique opportunity for mutual exchanges in the area of computer music technology. Over the years, the responsibilities over the organisation have been growing, musical and scientific topics have been increasing in number and broadening in scope; the size of the audience attending concerts and lectures has become significantly larger. In my view, this means that the aims pursued are of significance to contemporary society, and that they establish a real interaction between science and art, at many levels. The many University and private centers that have so far hosted the Colloquium, have all contributed to enlarge the table of contents offered to attendees, but they also managed to let a composite and dynamical discipline such as musical informatics be better known by the people. Among the many elements that have contributed the most to the popularisation, I want to highlight a particularly strong integration between scientific sessions and concert performances. These two elements, together, proved able to reach a heterogeneous audience, and to give the non-experts a chance to get deeper in the technical details. Clearly, by providing a venue for the meeting of researchers and musicians, CIM also highlights common grounds and directions, as well as the overall social relevance that the discipline is able to gain.

Emergence

Since the beginning, the very mission pursued in organizing the CIM has been one of presenting and discussing advanced research works in the field. Both scientific developments and the opening of new fields of application, make this event an opportunity to check the current vitality of scientific and musical search. The themes peculiar to the colloquium, as well as the array of musical directions illustrated in the concert programmes, let the significance of present directions emerge in their most innovative theoretical and technological forms. This mission renews every two years, based on the careful screening of scientific and musical contributions, handled by the CIM international committees. However, it gets also developed by casting most of the contributions against a specific theme of actual interest in the international community.

Discussion

As its thematic focus, the 13th CIM features a special session on the didactics of computer music, where paper presentations will give way to a discussion among all the attending delegates. This particular theme was proposed not by chance, as it will allow the CIM to be smoothly integrated with this year's La Terra Fertile ("The Fertile Land"), a conference held every two years in L'Aquila, providing Electronic Music students from all over Italy (and abroad) with an opportunity for illustrating their compositional and musicological work. Research developments should find their own terms in suitable educational programs. Timely reviewing of the goals and tasks of educational programs is absolutely appropriate in this field, as it allows teachers and researchers to better keep in touch with what is relevant for educational purposes.

Personally, the organisation of CIM gives me one more opportunity to understand the vitality and the ideals that pervade the area of musical informatics. In preparing this event, I have already worked side by side, and necessarily, with Italian and foreigner researchers that share with Istituto GRAMMA the aim to offer a suitable venue for establishing closer connections between musical informatics and issues of relevance to contemporary society. I wish to sincerely thank all these friends, as I am confident that our work so far set forth the best premises for this meeting to become an opportunity of intellectual enrichment and insight.

L'Aquila, September 2000

PRESENTATION

Nicola Bernardini

President
AIMI – Associazione di Informatica Musicale Italiana

Given its bi-annual cadence, the Colloquium of Musical Informatics has been a witness of changes in the field for over twenty years now – in such a time span, computers have evolved from a research instrument to an accounting tool and lately to a domestic commodity which manages most of our daily activites.

In the music field, the change in computer technologies has been all the more evident: from very dedicated programs operating in the contemporary music field to the pervasive use in all aspects of musical production. As a logical consequence, the interest in computer music has grown both in terms of people and of topics. Teaching and communicating electronic and computer music technologies and algorithms – the special topic of this colloquium – has become therefore a crucial element of growth and development in a very wide range of musical activities. The selection of this special topic has not been, however, a mere matter of choice: the Italian conservatories are currently undergoing a substantial change in status and legislation towards higher-level education; in this critical turning point computer technologies are bound to play a major role – and this Colloquium intends to be a major discussion forum towards a modern and better solution to the many problems that teaching computer music poses. Furthermore, the city of L'Aquila hosts another bi-annual symposium: that of conservatory teachers and students of electronic music (appropriatedly called 'La Terra Fertile', 'The Fertile Land') which, this year, directely follows the Colloquium: the choice of the topic makes explicit sense in this context.

However, besides the fulfillment of its institutional function (showcasing the latest advances in computer music research and production) this thirteenth Colloquium must tackle several other pending questions.

First and foremost (at least from my point of view), the changing role and functionality of our AIMI association. The pervasive spread of the commercial Internet and the use of computer networks in daily domestic activities has modified substantially the social composition of our associates and the way they use the association itself. The AIMI mailing list has grown in activity both in qualitative and quantitative terms, and AIMI functions more and better than ever as the coordinator of computer music activities in Italy. The status of AIMI as an association, however, dates back to the beginning of the eighties and does not suit any longer its current functionality: it is important to take the opportunity of this extended meeting to make the appropriate changes. Also, computers are functioning more than ever as significant vehicles between different forms of art (under the one-size-fits-all 'multimedia' term): it is important that AIMI opens up to these forms and these environments to gain a deeper insight in fields closely related to computer music.

Another important topic of discussion is the complicated relationship between research and production of computer music. Traditionally, research and production in Italy have followed closely interrelated paths, often leading to the very same people involved in both fields (and AIMI has played a major role in coordinating both activites). In the last few years however, these two activities had parted – for the most obvious reasons, of course: difference in time scheduling and finalities, different financial mechanisms, changing needs, etc. AIMI – created by researchers, scientists and musicians – has followed more naturally the research path, while several Italian production centers have teamed up in the CEMAT (Centri Musicali Attrezzati – technologically-enabled musical studios) association. This separation has indeed been important to gain better focus in the specific needs of each activity; now, however, it is important to find a way to cooperation to avoid missing the important foundations that both fields provide to each other. In this context, the contribution of CEMAT to the making of this Colloquium has been more than an (essential) financial help: it is a tangible sign of the interest in a renewed spirit of collaboration – and it will be important to use this occasion to tighten these bindings.

As it can be seen, many hopes are set on this Colloquium: I wish to thank the Organizing Committee, the Scientific and Music Committees, the authors and participants for their invaluable contribute to its success.

L'Aquila, September 2000

PRESENTATION

Nicola Bernardini

President
AIMI – Associazione di Informatica Musicale Italiana

Given its biannual cadence, the Colloquium of Musical Informatics has been a witness of changes in the field for over twenty years now – in such a time span, computers have evolved from a research instrument to an accounting tool and lately to a domestic commodity which manages most of our daily activities.

In the music field, the change in computer technologies has been all the more evident: from very dedicated programs operating in the contemporary music field to the pervasive use in all aspects of musical production. As a logical consequence, the interest in computer music has grown both in terms of people and of topics. Teaching and communicating electronic and computer music technologies and algorithms – the special topic of this colloquium – has become therefore a crucial element of growth and development in a very wide range of musical activities. The selection of this special topic has not been, however, a mere matter of choice: the Italian conservatories are currently undergoing a substantial change in status and legislation towards higher-level education; in this critical turning point computer technologies are bound to play a major role – and this Colloquium intends to be a major discussion forum towards a modern and better solution to the many problems that teaching computer music poses. Furthermore, the city of L'Aquila hosts another biennial symposium: that of conservatory teachers and students of electronic music (appropriately called 'La Terra Fertile', 'The Fertile Land') which, this year, directly follows the Colloquium: the choice of the town makes explicit sense in this context.

However, besides the fulfillment of its institutional function (showcasing the latest advances in computer music research and production) this thirteenth Colloquium must tackle several other pending questions.

First and foremost (at least from my point of view), the changing role and functionality of our AIMI association. The pervasive spread of the commercial Internet and the use of computer networks in daily domestic activities has modified substantially the social composition of our association and the very life of the association itself. The AIMI mailing list has grown in activity both in qualitative and quantitative terms, and AIMI functions more and better than ever as the coordinator of computer music activities in Italy. The status of AIMI as an association, however, dates back to the beginning of the eighties and does not suit any longer its current functionality: it is important to take the opportunity of this extended meeting to make the appropriate changes. Also, computers are functioning more than ever as significant vehicles between different forms of art (under the one-size-fits-all 'multimedia' term): it is important that AIMI opens up to these forms and these environments to gain a deeper insight in fields closely related to computer music.

Another important topic of discussion is the complicated relationship between research and production of computer music. Traditionally, research and production in Italy have followed closely interrelated paths, often leading to the very same people involved in both fields (and AIMI has played a major role in coordinating both activities). In the last few years however, these two activities had parted – for the most obvious reasons, of course: difference in time scheduling and finalities, different financial mechanisms, changing needs, etc. AIMI – created by researchers, scientists and musicians – has followed more naturally the research path, while several Italian production centers have teamed up in the CEMAT (Centri Musicali Attrezzati – technologically-enabled musical studios) association. This separation has indeed been important to gain better focus in the specific needs of each activity; now, however, it is important to find a way to cooperation to avoid missing the important foundations that both fields provide to each other. In this context, the contribution of CEMAT to the making of this Colloquium has been more than an (essential) financial help: it is a tangible sign of the interest in a renewed spirit of collaboration – and it will be important to use this occasion to tighten these bindings.

As it can be seen, many hopes are set on this Colloquium. I wish to thank the Organizing Committee, the Scientific and Music Committees, the authors and participants for their invaluable contribution to its success.

L'Aquila, September 2000

INVITED TALK

Sound Composition with Pulsars

Curtis Roads
Center for Research in Electronic Art Technology (CREATE)
Media Arts and Technology Program and Department of Music
University of California, Santa Barbara, California 93106 USA
clang@create.ucsb.edu, www.create.ucsb.edu

Pulsar synthesis (PS) is a method of electronic music synthesis based on the generation of trains of sonic particles. PS can produce either rhythms or tones as it criss-crosses perceptual time spans. The basic method generates sounds similar to vintage electronic music sonorities, with several important enhancements. The advanced method combines multiple pulsar trains and convolution with sampled aounds. Together with Alberto de Campo, the author has designed a program for pulsar synthesis called PulsarGenerator. Applications of pulsar synthesis in compositions by the author are noted.

0 Introduction

> *In July 1967 a young British astronomer detected in the sky by chance a radio signal in the form of a series of periodic impulses spaced every 1.33730113 seconds. The event was met immediately with incredulity. Deep in space, an object beat time with metronomic precision. The arrival time of the impulses was so regular that for a certain period it was believed that it was a message sent by an extraterrestrial civilization, destined for other beings in the universe.* – Luminet (1996)

All forms of music composition–from the freely improvised to the formally organized–are constrained by their sound materials. Thus the urge to expand the field of sound comes from a desire to enrich compositional possibilities. Much can be gained from the harvest of synthetic waveforms. Of special interest are those hybrids that crossbreed the richness of familiar sounds with unusual overtones.

Here we describe a powerful method of digital sound synthesis with links to past analog techniques. This is *pulsar synthesis* (PS), named after the spinning neutron stars that emit periodic signals in the range of 0.25 Hz to 642 Hz. By coincidence, this same range of frequencies–between rhythm and tone–is of central interest in pulsar synthesis.

PS melds established principles within a new paradigm. In its basic form, it generates electronic pulses and pitched tones similar to those produced by analog instruments such as the Ondioline (Jenny 1958; Fourier 1994) and the Hohner Elektronium (1950), which were designed around the principle of filtered pulse trains. Pioneering electronic music composers such as Karlheinz Stockhausen (1955, 1957, 1961, 1963) and Gottfried Michael Koenig (1957, 1959, 1962) used filtered impulse generation as a staple in their studio craft. Pulsar synthesis is a digital technique, however, and so it accrues the advantages of precise programmable control, waveform flexibility, graphical interface, and extensibility. In its advanced form, pulsar synthesis generates a world of rhythmically-structured crossbred sampled sounds.

PS belongs to a larger family of microsonic or particle synthesis techniques, one example of which is granular synthesis (Gabor 1946, 1947, 1952; Xenakis 1960; Roads 1978, 1991, 1996, forthcoming). These techniques stream or scatter acoustic particles in myriad patterns to produce time-varying sounds.

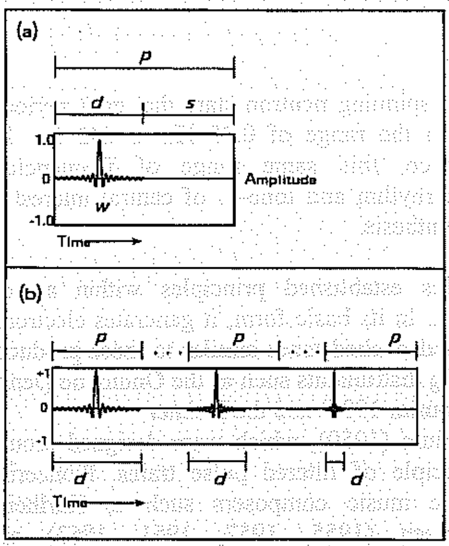
This article first presents the basic theory of pulsars and pulsar graphs. We then move on to the more advanced technique using pulsars to transform sampled sounds through cross-synthesis. We present musical applications of pulsar synthesis in compositions by the author. Near the end of the paper, we describe the features of a new interactive program called PulsarGenerator. The appendix presents a mathematical analysis of pulsar spectra.

1 Basic Pulsar Synthesis

Basic pulsar synthesis generates a family of classic electronic music timbres that are akin to those produced by an impulse generator connected to a bandpass filter. Unlike the classic technique, however, there is no filter in the basic PS circuit.

1.1 Anatomy of a pulsar

<Fig. 1; anatomy of a pulsar>



Roads, Fig. 1. Pulsar. (a) One pulsar (b) Pulsar train.

A single pulsar is a particle of sound. It consists of an arbitrary *pulsaret* waveform w with a period d followed by a silent time interval s (Fig. 1a). The total duration of a pulsar is $p = d + s$, where p is the *pulsar period*, d is the *duty cycle*, and s is silent. Repetitions of the pulsar signal form a *pulsar train*. Let us define the frequency corresponding to the repetition period as $f_p = 1/p$ and the frequency corresponding to the duty cycle as $f_d = 1/d$. Typical ranges of f_p are between 1 Hz and 5 kHz, and the typical range of f_d is from 80 Hz to 10 kHz.

In PS, both f_p and f_d are continuously variable quantities. They are controlled by separate envelope curves that span a train of pulsars. The train is the unit of musical organization on the time scale of notes and phrases. A pulsar train can last anywhere from a few hundred milliseconds to a minute or more.

Notice in Fig. 1b that the *duty ratio* or *d:s ratio* varies while p remains constant. In effect, one can simultaneously manipulate both fundamental frequency (the rate of pulsar emission) and what we could call a *formant frequency* (corresponding to the duty cycle), each according to separate envelopes. Lowering the fundamental means increasing s, and raising the fundamental means decreasing s.

<Fig. 2; Typical pulsaret waveforms;>

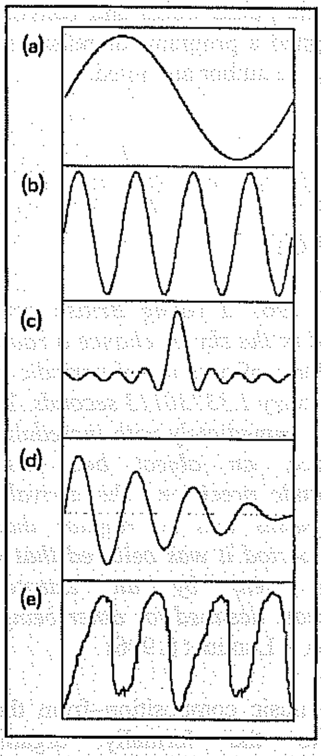


Fig. 2. Pulsaret waveforms

So far, the structure that we have described is similar to a standard impulse generator. Pulsar synthesis generalizes this configuration in several ways. First, it allows the pulsaret w to be any waveform. Fig. 2 shows some typical pulsaret waveforms, including those with multiple subperiods within their duty cycle (Fig. 2b and d).

<Fig. 3; Typical pulsaret envelopes>

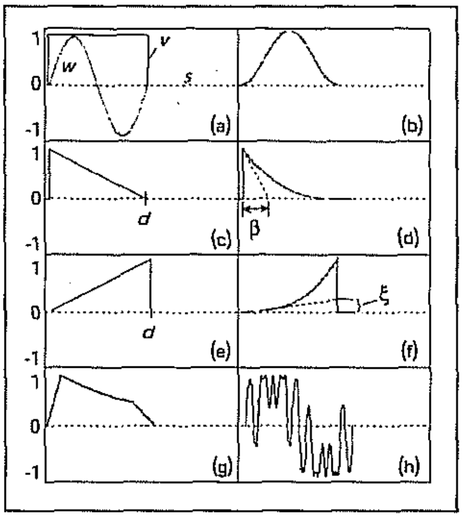


Fig. 3. Typical Pulsaret envelopes

Let us assume that w is a single cycle of a sine wave. From a signal processing point of view, this can be seen as a sine wave that has been limited in time by a rectangular function v, which we call the *pulsaret envelope.* An important generalization is that v can also be any shape. The envelope v has a strong effect on the spectrum of the pulsar train. Fig. 3 shows some typical pulsaret envelopes. A rectangular envelope (3a) produces a broad spectrum with strong peaks and nulls for any pulsaret.

Fig. 3g depicts a well known configuration for formant synthesis, an envelope with a sharp attack followed by an exponential decay (Kaegi and Tempelaars 1978; Rodet 1980). This configuration can be seen as a special case of pulsar synthesis. Later we look at the effects of the pulsaret envelopes on the spectrum of the pulsar train.

Keeping p and w constant and varying d on a continuous basis creates the effect of a resonant filter swept across a tone. There is, of course, no filter in this circuit. Rather, the frequency corresponding to the duty cycle d appears in the spectrum as a formant peak. By sweeping the frequency of this peak over time, we obtain the sonic equivalent of a time-varying bandpass filter applied to a basic impulse train.

1.2 Synthesis across time scales

PS operates within and between musical time scales. It generates a stream of microsonic particles at a variable rate, across the continuum spanning the infrasonic pulsations and the audio frequencies. When the distance between successive impulses is less than about one twentieth of a second, the human hearing mechanism causes them to fuse into a continuous tone. This is the *forward masking effect* (Buser and Imbert 1992). As Helmholtz (1885) observed, in the range between 20 and 35 Hz, it is difficult to distinguish the precise pitch of a sustained tone; reliable pitch perception takes hold at about 40 Hz, depending on the waveform. Thus for p between approximately 25 ms (corresponding to f_p = 40 Hz) and 200 μsec (corresponding to f_p = 5 kHz), listeners ascribe the characteristic of pitch to a periodic sustained tone.

As the rate of pulsar emission slows down and crosses through the threshold of the infrasonic frequencies (f_p < 20 Hz), the sensation of continuous tone evaporates, and we can perceive each pulsar separately. When the fundamental f_p falls between 62.5 ms (corresponding to the time span of a thirtysecond note at q = 60 MM) and 8 sec (corresponding to the time span of two tied whole notes at q = 60 MM), we hear rhythm.

1.3 Pulsaret-width modulation

<Fig. 4. PWM and PulWM waveforms>

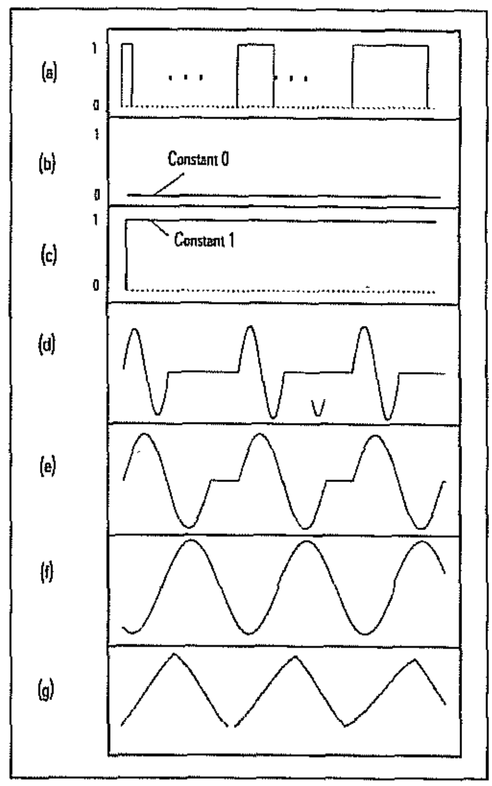


Fig. 4. Pulsaret-width modulation.

Pulse-width modulation (PWM) is a well-known analog synthesis effect that occurs when the duty cycle of a rectangular pulse varies while the fundamental frequency remains constant (Fig. 4a). This produces an edgy "sawing" quality as the upper odd harmonics increase and decrease over the course of the modulation. At the extremes of PWM, the signal is silent. For example, when $d = 0$, PWM results in a signal of zero amplitude (Fig. 4b). When $d = p$, PWM produces a signal of a constant amplitude of 1 (Fig. 4c).

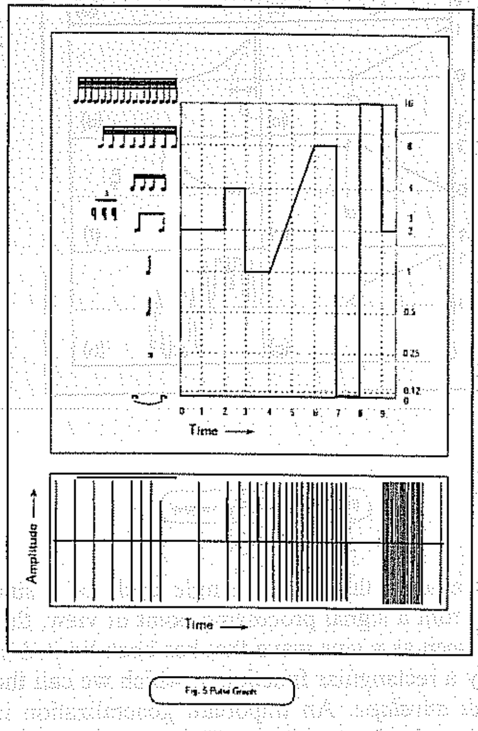
Pulsaret-width modulation (PulWM) extends and improves this model. First, the pulsaret waveform can be any arbitrary waveform. Second, it allows the duty cycle frequency to pass through and below the fundamental frequency. Here $f_d\ ^{2}\ f_p$. Notice in Fig. 4 how the duty cycle of the sinusoid increases from (d) to (e). In (f), $p = d$. Finally, in (g) $p < d$. That is, the duty cycle is longer than the fundamental period. Only the first quadrant of the sine wave repeats. The fundamental period cuts off the duty cycle of the pulsaret in mid-waveform. In our implementation, we apply a user-controlled crossfade time around this cutoff point, which we call the *edge* factor. When there is no crossfade, the edge factor is high.

We have also tested an alternative approach to pulsar-width modulation, which produces a different sound. In *overlapped pulsaret-width modulation* or OPulWM, the fundamental frequency is interpreted as the rate of pulsar emission, independent of the pulsaret duty cycle. That is, the duty cycle of an individual pulsar always completes, even when it crosses below the fundamental frequency. Whenever the fundamental period expires, our algorithm spawns a new pulsar. Thus when $d > p$, several pulsars overlap with others whose duty cycle has not yet completed. As d increases, the generator spawns more and more overlapping pulsars. For practical reasons, then, we stipulate an arbitrary overlap limit. In general, OPulWM results in a great deal of phase cancellation and thus tends to be a more subtle effect than regular PulWM.

1.4 Pulsar graphs

When the rate of pulsar emission crosses down into the infrasonic frequency threshold, we perceive each pulsar as a separate event in a rhythmic sequence. The fundamental frequency envelope becomes a graph of rhythm, as a function that is drawn onscreen (Fig. 5). Such a pulsar graph can serve as an alternative form of notation for one dimension of rhythmic structure, namely the onset time of events. The correspondence between the musical units of rhythmic structure (note values, tuplets, rests, etc.) can be made clear by plotting note values on the vertical or frequency scale. For example, assuming a tempo of 60 MM, a frequency of 5 Hz corresponds to a quintuplet figure. Note that the duration of the events is not represented by a two-dimensional pulsar graph, but could be represented by adding a third dimension to the plot.

<Fig.5 Pulsar graph >



In order to interpret the rhythm generated by a function inscribed on a pulse graph, one has to calculate the duration of the grain emission curve at a given fixed frequency rate. For example, a grain emission at 4 Hz that lasts for 0.75 seconds emits 3 grains. When grain emission switches from one value to the next, the pulsar corresponding to the new duration is immediately played, followed by a silence equal to the period of grain emission. Fig. 5 plots a rhythm that alternates between fixed-rate pulses, accelerandi, and silence.

2 Spectra Of Basic Pulsar Synthesis

The spectrum of the pulsar stream is the convolution product of w and v, biased in frequency by f_d and f_p. Since w and v can be arbitrary waveforms, and f_d and f_p can vary continuously, the range of spectra produced by PS is quite large.

When the formant frequency is set at a specific frequency, for example 1 kHz, this spreads energy in that region of the spectrum. Precisely how the energy is spread depends on w and v. The pulsaret waveform w can be considered a template of spectrum shape that repeats at the stipulated fundamental frequency f_p and is scaled in time by the duty cycle or formant frequency f_d. If, for example, the ratio of the amplitudes of the first five harmonics of w is 5:4:3:2:1, this ratio is preserved independently of p and d, when $f_p \geq f_d$.

<Figure 6; Effect of the pulsaret envelope on the spectrum>

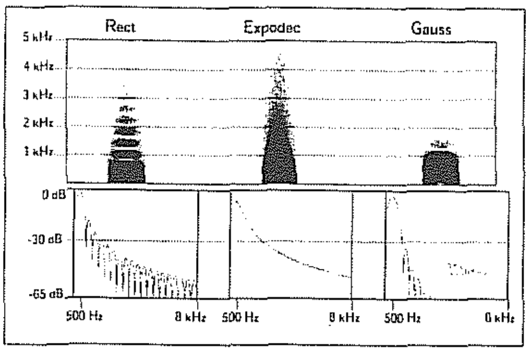


Fig. 6 Effect of pulsaret envelope on the spectrum

The pulsaret envelope's contribution to the spectrum is significant. Fig. 6 shows the spectra of individual pulsars where the waveform w is fixed as a sinusoid, and the pulsaret envelope v varies between three basic shapes. In the case of Fig. 6a, v is rectangular. Consequently, the formant spectrum takes the form of a broad sinc function in the frequency domain. The spectrum shows strong peaks at factors of $1.5f_d$, $2.5f_d$, etc., and nulls at harmonics of f_d. This is characteristic of the sinc function. An exponential decay or *expodec* envelope (such as in Fig. 3d) tends to smooth the peaks and valleys in the spectrum (Fig. 6b). The bell-shaped Gaussian envelope compresses the spectral energy, centering it around the formant frequency (Fig. 6c).

Thus by modifying the pulsaret envelope, one can alter the profile of the pulsar spectrum.

3 Advanced Pulsar Synthesis

The technique that we have presented thus far, basic pulsar synthesis, is the starting point for advanced pulsar synthesis. The advanced technique adds several features that take the method beyond the realm of vintage electronic sonorities. In particular, advanced pulsar synthesis is built on three principles.

1 Multiple pulsar generators sharing a common fundamental frequency but with individual formant and spatial trajectories
2 Pulse masking to shape the rhythm of the pulsar train
3 Convolution of pulsar trains with sampled sounds

<Fig. 7; pulsar synthesis schema>

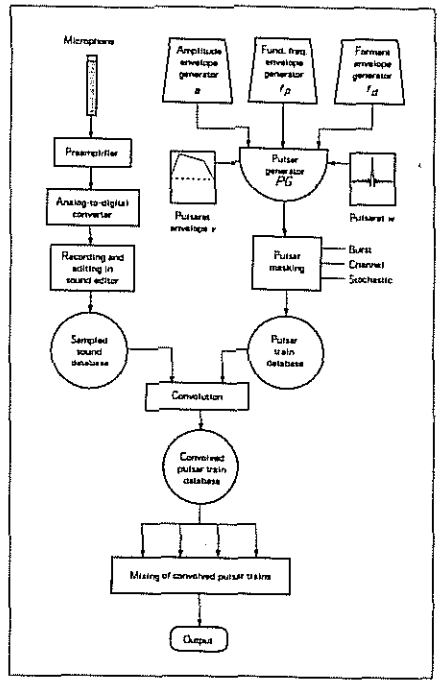


Figure 7. Schema for pulsar synthesis

Fig. 7 outlines the schema of advanced pulsar synthesis. The different parts of this schema are explained in the following sections.

3.1 Multiple pulsar generators

A pulsar generator has seven parameters:

1 Pulsar train duration
2 Pulsar train fundamental frequency envelope f_p
3 Pulsaret formant frequency envelope f_d
4 Pulsaret waveform w
5 Pulsaret envelope v
6 Pulsar train amplitude envelope a
7 Pulsar train spatial path s

The individual pulsar train is the simplest case. To synthesize a complex sound with several resonance peaks, we can add several pulsar trains with the same fundamental frequency but with different time-varying formant frequencies f_d. One envelope controls their common fundamental frequency, while two or more separate envelopes control their formant trajectories f_{d1}, f_{d2}, etc.

One of the unique features of pulsar synthesis is that each formant can follow its own spatial path. This leads to complex spatial interplay within a single tone or rhythmic phrase.

3.2 Pulsar masking

A pulsar generator emits a metronomic sequence of pulsars, where the rate of emission can vary over time according to the fundamental frequency envelope function f_p. *Pulsar masking* breaks up the stream by introducing intermittancies (regular or irregular) into the metronomic pulsar stream. It deletes individual pulsarets in a sequence, leaving an interval of silence in their place. This takes three forms: *burst, channel,* and *stochastic masking.*

<Fig. 8 a b c; pulsar masking 1>

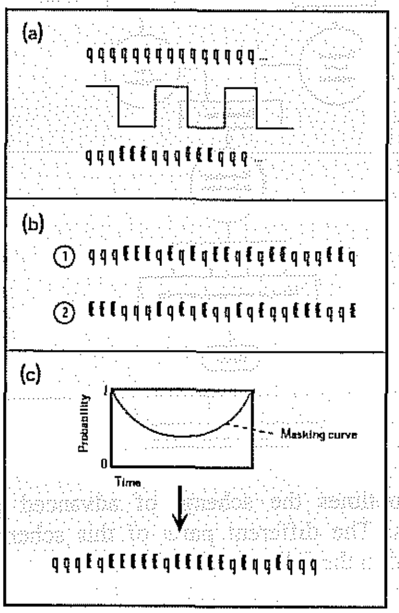
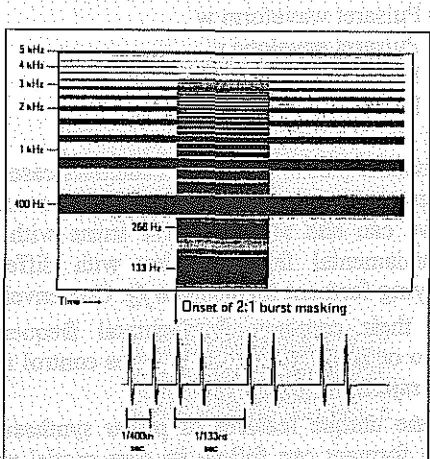


Figure 8. Pulsar masking.

<Fig. 9. Effect of burst masking in the audio frequencies>



Burst masking (Fig. 8a) models the burst generators of the classic electronic music studios. It produces a regular pattern of pulsarets that are interrupted at regular intervals. The on-off pattern can be stipulated as the *burst ratio b:r,* where *b* is the burst length in pulsaret periods and *r* is a rest length in pulsaret periods. For example, a *b:r* ratio of 4:2 produces an alternating sequence of four pulsarets and two silent periods: 111100111100111100111100, etc. If the fundamental frequency is infrasonic, the effect is rhythmic. When the fundamental is in the audio frequency range, burst masking imposes an amplitude modulation effect on the timbre (Fig. 9), dividing the fundamental frequency into a subharmonic frequency $b + r$.

Channel masking (Fig. 8b) deletes pulsars in alternate channels. By selectively masking pulsars in two channels 1 and 2, one creates a dialog within a phrase, articulating each channel in turn. Fig. 8b shows two channels only, but we can generalize this scheme to *N* channels.

Stochastic masking introduces random intermittancy into the regular stream of pulsars. We have implemented stochastic masking as a weighted probability that a pulsar will be emitted at a particular point in a pulsar train. The probability is expressed as an envelope over the duration of the pulsar train. When the value of the envelope is 1, a pulsar is emitted. If the value is less than 1, it has less possibility. A value of 0 results in no pulsar emissions. Values between 0.9 and 0.8 produces an *interesting analog-like intermittancy,* as if there were an erratic contact in the synthesis circuit (Fig. 8c).

3.3 Transformation of sampled sounds by convolution with pulsars

The technique of pulsar synthesis can be harnessed as a method

Convolution is fundamental to the physics of waves (Rabiner and Gold 1975). It "crosses" two signals, creating a new signal that combines the time structures and spectra of both inputs. Many transformations emerge from convolution, including exotic filters, spatialisers, models of excitation/resonance, and a gamut of temporal transformations (echoes, reverberation, attack smoothing, rhythm mapping). See Roads (1992, 1993b, 1997) for applications of convolution in musical sound transformation. Pure convolution has no control parameters. That is, the type of effect achieved depends entirely on the nature of the input signals.

The convolution of a pulsar train with a sampled sound causes each pulsar in the train to be replaced by a filtered copy of the sampled sound. In convolution, each pulsar represents the impulse response of a bandpass filter. Thus timbral variations can derive from two factors: (1) filtering effects imposed by the time-varying pulsar train, and (2) overlapping effects caused by convolution with pulsar trains whose fundamental period is shorter than the duration of the sampled sound.

<Fig. 10; filtering effects>

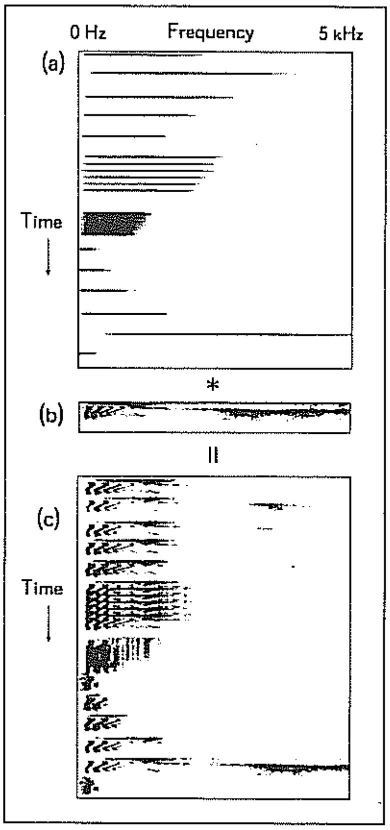


Fig. 10. Convolution w pulsar train.

Fig. 10 shows the temporal and filtering effects of convolution in the form of sonograms. The input signal (a) is the Italian word "qui" (pronounced "kwee"). It convolves with the pulsar train (b) with a variable infrasonic fundamental frequency and a variable audio formant frequency. The resulting convolution (c) combines the time structure and the spectra of the two signals.

A database of sampled sound objects can be stockpiled for crossing with trains selected from the pulsar database. If the goal of the synthesis is to retain the time structure of the pulsar train (e.g., to maintain a specific rhythm), the sampled sound objects should be of short duration (less than the fundamental period of the pulsar train) and have a sharp attack (a rise time less than 100 ms). These constraints minimize the time smearing effects of convolution (Roads 1992, 1993b, 1997). Thus a good starting point for a sound database is a collection of percussion samples. The constraints can be relaxed if one seeks a smoother and more continuous texture. Samples with long durations superimpose multiple copies of the sampled object, creating a rippling sound stream. Samples with slow attacks blur the onset of each sample copy, smearing the stream into a continuum. Thus by controlling the attack shape of the sample one has a handle on the sonic texture.

Sophisticated transformations involving rhythm and spatial mapping can be achieved through convolution. It is well-known that any series of impulses convolved with a brief sound maps that sound into the time pattern of the impulses. These impulses can be performed by a percussionist, or they can be emitted by a pulsar generator such as the one that we have implemented. If the pulsar train frequency is in the infrasonic range, then each pulsar is replaced by a copy of the sampled sound object, creating a rhythmic pattern. The convolution of a rhythmic pattern with a sound object causes each impulse to be replaced by a copy of the sound object. Each instance of the sampled object is projected in space according to the spatial location of a specific pulsar's position in space.

4 Composing With Pulsars

To interact with PulsarGenerator in real time is to experiment with sonic ideas. In the course of experimentation, a composer can save various settings and plan how these will be used within a composition. The PulsarGenerator program can also record the sounds produced in a real-time session. This session can be edited by the composer and possibly convolved or mixed with other material.

A final stage of pulsar composition is to merge multiple trains to form a composite texture. This is a question of montage, and is best handled by editing and mixing software that is designed for this purpose. Each layer of the texture may have its own rhythmic pattern, formant frequency envelope, choice of convolved objects, and spatial path. Working on a variety of time scales, a composer can apply signal processing transformations on individual pulsars, pulsar trains, and pulsar textures. These may include mixing with other sounds, filtering, modulations, reverberation, and so on.

5 Musical Applications Of Pulsar Synthesis

I developed pulsar synthesis in the course of realizing *Clang-tint* (Roads 1993a), an electronic music composition that was commissioned by the

Japanese Ministry of Culture (Bunka-cho) and the Kunitachi College of Music, Tokyo. The second movement of this work, entitled *Organic*, focuses on expressive phrasing. It combines bursts of insect, animal, and bird calls with electronic pulse-tones. The electronic sound palette is based on pulsar synthesis in multiple forms: pulsating blips, elongated formant tones, and clouds of asynchronous pulsars. For the latter, I first generated multiple infrasonic pulsar trains, each one beating at a different frequency in the range of 6 to 18 Hz. I then mixed these together to obtain the asynchronous pulsar cloud.

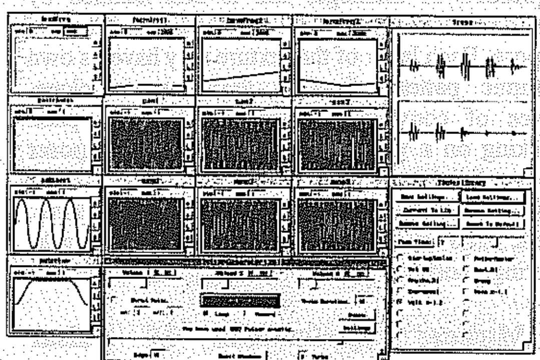
The raw material of my electronic music composition *Half-life*, composed in 1998 and 1999, is a one-minute pulsar train that varies wildly. Most sounds in the rest of the work were derived from this source. *Half-life* extends the pulsar material through processes of granulation, microfiltration, granular pitch-shifting, recirculating feedback echo, individual pulsar amplitude shaping, and selective reverberation.

We have begun to distribute the PulsarGenerator application to other musicians, so we expect that there will be more musical results in the near future.

6 Implementations Of Pulsar Synthesis

My original implementation of PS dates to 1991, using James McCartney's Synth-O-Matic, a programmable sound synthesis environment for Apple Macintosh computers (McCartney 1990, 1994). In 1996, Mr. McCartney replaced Synth-O-Matic with SuperCollider 1–an object-oriented programming language with a Power Macintosh runtime system (McCartney 1996). Using SuperCollider 1, Stephen T. Pope and I created a new implementation of basic PS in 1997.

<Fig. 11; PulsarGenerator screen>



Roads. Fig. 11. PulsarGenerator control panel.

Based on the improved SuperCollider 2 (McCartney 1998), Alberto de Campo and I developed a new realization of pulsar synthesis in 1999. We presented it that year in a summer course at the Center for New Music and Audio Technology, University of California, Berkeley. Further refinement of this prototype has led to the PulsarGenerator application, distributed by CREATE. Fig. 11 presents the graphical interface of PulsarGenerator, version 1. Notice the control envelopes for the synthesis variables. These envelopes can be designed in advance of synthesis, or manipulated in real time as the instrument plays. We have implemented a scheme for saving and loading these envelopes in groups called *settings*. The program lets one crossfade at a variable rate between multiple settings, which takes performance with PulsarGenerator to another level of synthesis complexity.

In our implementation, each pulsar is scheduled as a discrete event. The efficiency of synthesis is thus related to the rate of pulsar emission. A three-formant instrument running at an emission rate of 6000 pulsars per second (corresponding to the fundamental frequency of 2 kHz) consumes approximately 80% of the processor on an Apple G4 operating at a 500 mHz clock speed. It is a testimony to SuperCollider 2 that the entire implementation, including the graphical interface, required less than 1200 lines of code and comments. Our code builds the interface, defines the synthesis algorithm, schedules the pulsars, and handles file input and output. McCartney's SCPlay, an efficient real-time sound engine, calculates the samples.

7 Conclusions

Music transpires on multiple time scales, from high-level macrostructure down to a myriad of individual sound objects or notes. Below this level is another hierarchy of time scales. Here are the microsonic particles such as the classical rectangular impulses, grains, wavelets, and pulsars (Roads 1999). Impulse generation as an effective means of music synthesis was established decades ago in the analog electronic studio. By comparison, digital pulsar synthesis offers a flexible choice of waveforms and envelopes, increased precision, and graphical programmable control.

Unlike wave-oriented synthesis techniques, the notion of rhythm is built into techniques based on particles. Rhythm, pitch, and timbre are all interrelated but can be controlled separately. Pulsar synthesis offers a seamless link between the time scales of individual particle rhythms, periodic pitches, and the meso or phrase level of composition. Another novel feature of this technique is the generation of multiple independent formant trajectories, each of which follows its own spatial path.

As we have shown, the basic pulsar technique can be extended to create a broad family of musical structures: singular impulses, rhythmic sequences, continuous tones, time-varying phrases, and beating textures. The pulsar micro events can be deployed in rhythmic sequences or, when the density of events is sufficiently high, in sustained tones, thus allowing composition to pass directly from microstructure to mesostructure.

8 Acknowledgements

Pulsar synthesis was inspired by numerous conversations in Naples with my late friend Professor Aldo Piccialli and his colleagues in the Department of Physics at the Universit˜ di Napoli ÇFederico IIÈ (Cavaliere, Ortosecco, and Piccialli 1986; De Poli and Piccialli 1991; Cavaliere and Piccialli 1997). My deep thanks go to Alberto de Campo for his collaboration on the PulsarGenerator application, which he coded. I thank James McCartney for the excellent SuperCollider 2 software, which was the basis for our development of PulsarGenerator. I am grateful to Brigitte Robindor and Stephen T. Pope for their comments on an early draft of this manuscript. Dr. Luca Lucchese of the Department of Electrical and Computer Engineering at UCSB consulted on the analytic form of the pulsar spectrum equations. Throughout the period in which I originally developed this technique (1991-1995), I enjoyed the support of Dr. Gerard Pape of the Centre de Cration Musicale ÇIannis XenakisÈ in Paris, and Professor Horacio Vaggione of the Dpartement Musique at the Universit de Paris VIII. During the final preparation of this paper I enjoyed the generous support of Professor JoAnn Kuchera-Morin of CREATE at the University of California, Santa Barbara.

9 References

Bass, S., and T. Goeddel. 1981. "The efficient digital implementation of subtractive music synthesis." *IEEE Micro* 1(3): 24-37.

Buser, P., and M. Imbert. 1992. *Audition.* Cambridge, Massachusetts: MIT Press.

Cavaliere, S., I. Ortosecco, and A. Piccialli. 1986. "Synthesis by formants: a new approach." In *Proceedings of the 1986 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.* New York: IEEE Acoustics, Speech, and Signal Processing Society.

Cavaliere, S., and A. Piccialli. 1997. "Granular synthesis of musical signals." In C. Roads, S. Pope, A. Piccialli, and G. De Poli, eds. 1997. *Musical Signal Processing.* Amsterdam: Swets & Zeitlinger.

Cherry, C. 1949. *Pulses and Transients in Communication Circuits.* London: Chapman and Hall.

De Poli, G., and A. Piccialli. 1991. "Pitch-synchronous granular synthesis." In G. De Poli, A. Piccialli, and C. Roads, eds. 1991. *Representations of Musical Signals.* Cambridge, Massachusetts: MIT Press: 187-219.

Fourier, L. 1994. "Jean-Jacques Perrey and the Ondioline." *Computer Music Journal* 18(4): 18-25.

Gabor, D. 1946. "Theory of communication." *Journal of the Institute of Electrical Engineers* Part III, 93: 429-457.

Gabor, D. 1947. "Acoustical quanta and the theory of hearing." *Nature* 159 (4044): 591-594.

Gabor, D. 1952. "Lectures on communication theory." Technical Report 238, Research Laboratory of Electronics. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Helmholtz, H. 1885. *The Sensations of Tone.* Translated by A. Ellis. Reprinted 1954. New York: Dover.

Jenny, G. 1958. "L'Ondioline: conception et ralisation." Paris: *Toute la radio.*

Kaegi, W., and S. Tempelaars. 1978. "VOSIM--a new sound synthesis system." *Journal of the Audio Engineering Society* 26(6):418-426.

Koenig, G. M. 1957. *Essay.* Composition for electronic sounds. Score. Vienna: Universal Edition.

Koenig, G. M. 1959. "Studium im Studio." *die Reihe* 5.

Koenig, G. M. 1962. "Commentary." *die Reihe* 8.

Luminet, Jean-Pierre. 1996. "Musique avec pulsar oblig (È propos du *Noir de l'Etoile,* de Grard Grisey)." From the Internet.

McCartney, J. 1990. Synth-O-Matic version 0.06 software.

McCartney, J. 1994. Synth-O-Matic version 0.45 software.

McCartney, J. 1995. SuperCollider version 1 software.

McCartney, J. 1998. SuperCollider version 2 software.

Rabiner, L., and B. Gold. 1975. *Theory and Applications of Digital Signal Processing.* Englewood Cliffs: Prentice-Hall.

Roads, C. 1978. "Automated granular synthesis of sound." *Computer Music Journal* 2(2): 61-62. Revised and updated version printed as "Granular synthesis of sound" in C. Roads and J. Strawn, eds. 1985. *Foundations of Computer Music.* Cambridge, Massachusetts: MIT Press. pp. 145-159.

Roads, C. 1991. "Asynchronous granular synthesis." In G. De Poli, A. Piccialli, and C. Roads, eds. 1991. *Representations of Musical Signals.* Cambridge, Massachusetts: MIT Press. pp. 143-185.

Roads, C. 1992. "Musical applications of advanced signal transformations." In A. Piccialli, ed. 1992. *Proceedings of the Capri Workshop on Models and Representations of Musical Signals.* Naples: Dipartimento Scienze Fische, Universit˜ di Napoli ÇFederico II.È

Roads, C. 1993a. "Organization of *Clang-tint.*" In S. Ohteru and S. Hashimoto, eds. 1993. *Proceedings of the 1993 International Computer Music Conference.* San Francisco: International Computer Music Association.

Roads, C. 1993b. "Musical sound transformation by convolution." In S. Ohteru, ed. *Proceedings of the 1993 International Computer Music Conference.* San Francisco: International Computer Music Association. pp. 102-109.

Roads, C. 1997. "Sound transformation via convolution." In C. Roads, S. Pope, A. Piccialli, and G. De Poli, eds. 1997. *Musical Signal Processing.* Amsterdam: Swets & Zeitlinger.

Roads, C. 1999. "Time scales of musical structure." In F. Barri re and G. Bennett, eds. *Actes V. Acadmie Internationale de Musique ƒlectroacoustique.* Bourges: Editions Mnemosyne.

Roads, C. Forthcoming. *Microsound.* Cambridge, Massachusetts: MIT Press.

Rodet, X. 1980. "Time-domain formant-wave-function synthesis." In J. G. Simon, ed. 1980. *Spoken Language Generation and Understanding.* Dordrecht: D. Reidel. Reprinted in *Computer Music Journal* 8(3):9-14. 1984.

Stockhausen, K. 1955. "Actualia." *die Reihe* 1.

Stockhausen, K. 1957. "...how time passes..." *die Reihe* 3: 10-43. English edition 1959. Reprinted as "... wie die Zeit vergeht..." in K. Stockhausen. 1963. *Texte zur elektronischen und instrumentalen Musik.* Band 1. Cologne: DuMont Schauberg: pp. 99-139.

Stockhausen, K. 1961. "Two lectures." *die Reihe* 5. English edition. Bryn Mawr: Theodore Presser Company. pp. 59-82.

Stockhausen, K. 1963. "Die Einheit der musikalischen Zeit." In K. Stockhausen. 1963. *Texte zur elektronischen und instrumentalen Musik.* Band 1. Cologne: DuMont Schauberg. pp. 211-221. Reprinted as "The concept of unity in electronic music." E. Barkin, translator, in B. Boretz and E. Cone, eds. 1972. *Perspectives on Contemporary Music Theory.* New York: Norton. pp. 129-147.

Xenakis, I. 1960. "Elements of stochastic music." *Gravensaner BlŠtter* 18: 84-105.

Figure Captions

Fig. 1. Anatomy of a pulsar. (a) A pulsar consists of a brief burst of energy called a *pulsaret* w of a duration d followed by a silent interval s. The waveform of the pulsaret, here shown as a band-limited pulse, is arbitrary. It could also be a sine wave or a period of a sampled sound. The total duration $p = d + s$, where p is the fundamental period of the pulsar. (b) Evolution of a pulsar train, time-domain view. Over time, the pulsar period p remains constant while the pulsaret period d shrinks. The ellipses indicate a gradual transition period containing many pulsars between the three shown.

Fig. 2. Typical pulsaret waveforms. In practice, any waveform can be used. (a) Sine. (b) Multicycle sine. (c) Band-limited pulse. (d) Decaying multicycle sinusoid. (e) Cosmic pulsar waveform emitted by the neutron star Vela X-1.

Fig. 3. Typical pulsaret envelopes v. (a) Rectangular. (b) Gaussian. (c) Linear decay. (d) Exponential decay. (e) Linear attack. (f) Exponential attack. (g) FOF. (h) Bipolar modulator.

Fig. 4. PWM and PulWM. (a) Classical PWM with a rectangular pulse shape. The ellipses indicate a gradual transition between the pulses. (b) PWM when the duty cycle $d = 0$ results in a signal of zero amplitude. (c) PWM when the duty cycle $d = p$ (the fundamental period), the result is a signal with a

constant amplitude of 1. (d) Pulsar train with a sinusoidal pulsaret. (e) Same period as (d), but the duty cycle is increasing. (f) The duty cycle and the period are equal, resulting in a sinusoid. (g) The duty cycle is greater than the fundamental period, which cuts off the final part of the sine waveform.

Fig. 5. Pulsar rhythms. (top) Pulse graph of rhythm showing rate of pulsar emission (vertical scale) plotted against time (horizontal scale). The left-hand scale measures traditional note values, while the right-hand scale measures frequencies. (bottom) Time-domain image of generated pulsar train corresponding to the plot above.

Fig. 6. Effect of the pulsaret envelope on the spectrum. The top row presents frequency-versus-time sonograms of an individual pulsar with a sinusoidal pulsaret, a fundamental frequency of 12 Hz, and a formant frequency of 500 Hz. The sonograms are based on 1024-point fast Fourier transform plots using a Von Hann window, and are plotted on a linear frequency scale. From left to right, we see the sonogram produced by a rectangular envelope, an expodec envelope, and a Gaussian envelope. The lower row plots the spectra of these pulsars on a dB scale.

Figure 7. Schema of pulsar synthesis. A pulsar generator with separate envelope controls for fundamental frequency, formant frequency, amplitude, stochastic masking, and spatial position. In advanced pulsar synthesis, several generators may be linked with separate formant and spatial envelopes. A pulsar stream may be convolved with a sampled sound.

Fig. 8. Pulsar masking turns a regular train into an irregular train. Pulsars are illustrated as quarter notes, and masked pulsars are indicated as quarter rests. (a) Burst masking. The burst ratio here is 3:3. (b) Channel masking. (c) Stochastic masking according to a probability table. When the probability is 1, there is no masking. When the probability is 0, there are no pulsars. In the middle, the pulsar train is intermittant. Notice the thinning out of the texture as the probability curve dips in the center.

Fig. 9. Sonogram depicting the effect of burst masking in the audio frequency range. The pulsaret is one cycle of a sinusoid, and the pulsaret envelope is rectangular. The *b:r* ratio is 2:1. The fundamental frequency is 100 Hz and the formant frequency is 400 Hz. Notice the subharmonics at 133 Hz and 266 Hz caused by the extended periodicity of the pulse masking interval (400 Hz/3).

Fig. 10. Effect of convolution with pulsar train. (a) Sampled sound, the Italian word "qui" (pronounced "kwee"). (b) Infrasonic pulsar train with a variable fundamental and formant frequency. (c) Convolution of (a) and (b).

Fig. 11. Control panel of the PulsarGenerator application by Alberto de Campo and Curtis Roads. Copyright Regents of the University of California 2000.

SOUND ANALYSIS AND SYNTHESIS I

SOUND ANALYSIS AND SYNTHESIS I

REPRESENTATION AND MODIFICATION OF TIME-VARYING SOUND SIGNALS

Gianpaolo Evangelista

Audiovisual Communications Laboratory,
Swiss Federal Institute of Technology,
Lausanne, Switzerland.

e-mail: Gianpaolo.Evangelista@epfl.ch

Sergio Cavaliere

ACEL, Dipartimento di Scienze Fisiche,
Università "Federico II" di Napoli,
Complesso Universitario di M.S.Angelo,
Via Cinzia 80126 Napoli
e-mail: cavaliere@na.infn.it

Abstract

In recent papers the authors introduced in the range of acoustical signals the use of a powerful instrument for the analysis and modification of signals: the Laguerre Transform, mapping a signal space into another one whose frequency axis is warped in a controlled way. While altering the overall frequency content of specified signals is very useful in many applications, most real world signals show time-varying features, e.g., in both their amplitude and frequency content. A second step, therefore, has been that of extending the principles and the architecture of the Discrete Laguerre Transform to the time varying case so that the frequency content of a signal could be displaced over time to different values, realizing therefore a time varying warping transform. This transform has nice features and high regularity, it allows perfect reconstruction and it can effectively succeed in regularizing real world signals or in modifying in a controlled way some of their relevant parameters. These features well match those of ordinary sound signals whose frequency content is slowly varying with time, such as intonation in speech, glissando and vibrato in music. It also meets many processing needs for a wide range of sweep signals. In the paper, after briefly recalling the relevant features of the recently introduced transform, many examples are given, demonstrating the wide range of applications for which it seems to be well suited.

1 Introduction

Most real time signals depart significantly from the ideal model of a constant-pitch, oscillating signal. Due to many different physical reasons, the underlying oscillatory features are mostly time varying. Also, some relevant information is bound to these features, such as the perception of a pleasant timbre, the internal texture of a sound, interpretation cues and communicative or expressive contents. A transform able to compensate against these effects or, in turn, to add these features would be without doubt a strong tool for the analysis, the representation and the modification of sound signals.

2 Frequency Warping and the Discrete Laguerre Transform

Starting point of our work is the Discrete Laguerre Transform. It can be shown that by projecting a discrete time signal onto the orthonormal set of Laguerre sequences [1][2] we obtain a new signal whose spectrum $\hat{X}(e^{j\omega})$ is simply the frequency-warped version of the source spectrum $X(e^{j\omega})$:

$$X(e^{j\omega}) = \Lambda_0(e^{j\omega})\hat{X}(e^{j\vartheta(\omega)}).$$

Here $\Lambda_0(e^{j\omega})$ is a normalizing factor, which, due to the unitary property of the transform, preserves energy while passing from the source to the destination domain. The warping law $\vartheta(\omega)$ is a function of a single parameter, the real pole of a first order all-pass filter section. The transformation may be controlled by this parameter -- the Laguerre parameter – which allows a large degree of freedom in the choice of the warping characteristics. The authors used this transformation in conjunction with other transforms such as the Discrete Wavelet Transform and the Pitch Synchronous Wavelet Transform, introducing a new class of orthogonal transforms having the advantage of arbitrarily allocating the analysis bands [3][4][11]. They have also pointed out an entire new range of applications for the introduced transforms [8][9][12].
Finally, the Laguerre Transform can be computed by conventional DSP methods, by means of the all-pass cascade structure depicted in Fig. 1. The time-varying structure generalizes that of the constant case. In the constant case all the filters $A_i(z)$ are equal to a single all pass function whose phase $\vartheta(\omega)$ actually turns out to be the warping law described in the above.

2 The Generalized Discrete Time Laguerre Transform

The Laguerre transform may be extended to its time varying version. For this purpose, however, we must introduce a new parametric class of sequences, namely the Time-Varying Discrete Laguerre Sequences [10][11][12]. These sequences generalize the Laguerre sequences in that they allow for modification of the

frequency content of the signal by means of a time-varying frequency warping law.
Consider the sampled dispersive delay line shown in Fig. 1, consisting of a chain of real first-order all-pass filters

$$A_n(z) = \frac{z^{-1} - b_n}{1 - b_n z^{-1}} \quad \text{with } -1 < b_n < 1,$$

a sampling device closing at time k=0 and a shift-register loaded at k=0 with the outputs of the filters and outputting the sequence of samples $\hat{x}[k]$ at regular clock intervals. The dispersive line reverts to a linear delay line when all the parameters b_n are zero, and in the case of b_n constant from section to section, it reverts to the structure for computing the ordinary Laguerre Transform.

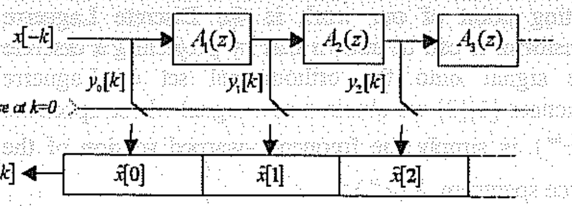


Figure 1. Sampled dispersive delay line.

Since the input sequence is time-reversed, the line implements the scalar product

$$\hat{x}[n] = \langle \varphi_n, x \rangle = \sum_k x[k]\, \varphi_n[k], \qquad (1)$$

where

$$\varphi_n[k] = a_1[k] * a_2[k] * \ldots * a_n[k].$$

Hence, the z-transform of the sequence $\varphi_n[k]$ is an order n all-pass filter with transfer function

$$\Phi_n(z) = \begin{cases} 1 & \text{if } n = 0 \\ \prod_{k=1}^{n} \dfrac{z^{-1} - b_k}{1 - b_k z^{-1}} & \text{if } n > 0 \end{cases} \qquad (2)$$

The output sequence $\hat{x}[k]$ may be interpreted as the set of coefficients of a suitable signal expansion. In fact, the set of sequences $\psi_n[k]$ whose z-transforms are

$$\Psi_n(z) = \begin{cases} \dfrac{1}{1 - b_1 z^{-1}} & \text{if } n = 0 \\ \dfrac{1 - b_n b_{n+1}}{(1 - b_n z)(1 - b_{n+1} z^{-1})} \Phi_n(z) & \text{if } n > 0 \end{cases}, \qquad (3)$$

can be shown ([11][12]) to be biorthogonal to the set $\varphi_n[k]$, i.e.,

$$\langle \psi_n, \varphi_m \rangle = \sum_{k=0}^{\infty} \psi_n[k] \varphi_m[k] = \delta_{n,m} u[n] \qquad (4)$$

and

$$\sum_{n=0}^{\infty} \psi_n[k] \varphi_n[m] = \delta_{k,m} u[k], \qquad (5)$$

where $u[k]$ is the unit step sequence. The set is complete over causal sequences. Property (5) requires some tecnical conditions on the asymptotic behavior of the parameters b_n. However, any finite selection of them within the specified range $-1 < b_n < 1$ leads to a set that can be embedded in a biorthogonal complete set; this is in fact the practical case where the signal to transform is causal and also has finite duration. Correspondingly, the signal $x[k]$ is expanded onto the set $\psi_n[k]$ as follows

$$x[k] = \sum_{n=0}^{\infty} \hat{x}[k] \psi_n[k], \qquad (6)$$

where the coefficients are given by (1).
There are several equivalent structures for implementing the inverse transform. The one shown in Fig. 2 is based on the following recurrence:

$$\Psi_n(z) = V_n(z) \Psi_{n-1}(z), \qquad n \geq 1$$

where

$$V_n(z) = \frac{1 - b_n b_{n+1}}{1 - b_{n-1} b_n} \frac{z^{-1} - b_{n-1}}{1 - b_{n+1} z^{-1}}$$

and we used the convention that $b_0 = 0$. The analysis coefficients $\hat{x}[k]$ are used as weights for the dispersive tapped delay line in a structure that generalizes Laguerre filters [1][2].
If the sequence of parameters $b_n = b$ is constant, the resulting transform, as expected, reverts to a biorthogonal variant of the Laguerre transform. The biorthogonal sequences $\psi_n[k]$ and $\varphi_n[k]$ may be used interchangeably for the analysis or for the synthesis, with obvious modifications of the structures.

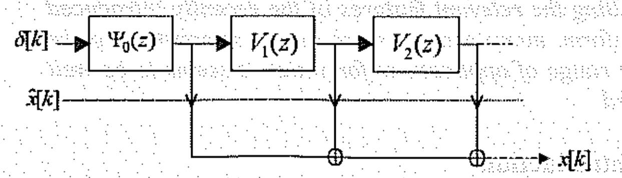


Figure 2. Structure implementing the inverse biorthogonal transform

4. Time-Varying Frequency Warping

Time-varying frequency warping is obtained by means of the analysis structure shown in Fig. 1, which, actually, may be seen as a special type of time-varying filter. Each of the y_n lines, n=0,1,..., carries a filtered version of the input signal $x(-k)$, the transfer function at index n being the cascade of the sections up to n. The signals $y_n(k)$ are therefore filtered versions of the input signal by the transfer functions

$$\Phi_n(z) = A_1(z) \cdot A_2(z) \cdots A_n(z).$$

For a real signal $x(k)$ we have in the frequency and time domains, respectively:

$$Y_n(\omega) = X^*(\omega)\cdot\Phi_n(\omega),$$

$$y_n(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X^*(\omega)\left(\prod_{m=1}^{n} A_m(\omega)\right)e^{jk\omega}d\omega$$

The *n-th* sample of the output sequence stored in the shift register is given by the sample of the $y_n(k)$ sequences, at time k=0:

$$\hat{x}[n] = y_n(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X^*(\omega)\left(\prod_{m=1}^{n} A_m(\omega)\right)d\omega = \langle x,\varphi_n\rangle \quad (7)$$

where,

$$\langle x,\varphi_n\rangle = \sum_k x(k)\,\varphi_n(k)$$

is the orthogonal projection coefficient of the signal over the analysis set.

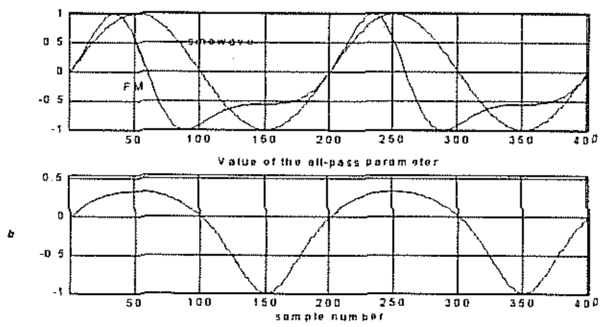


Fig. 3 Frequency modulation by means of warping

The $A_m(\omega)$ are all-pass functions characterized by a pure phase response: $A_m(\omega) = e^{-j\vartheta_m(\omega)}$, where for $|\omega| < \pi$

$$\vartheta_m(\omega) = \omega + 2\arctan\frac{b_r\sin\omega}{1-b_r\cos\omega} = 2\arctan\left(\frac{1+b_r}{1-b_r}\tan\frac{\omega}{2}\right). \quad (8)$$

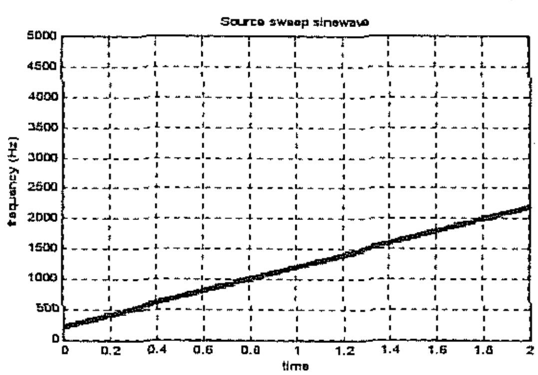


Fig. 4 A deeply swept sinusoid

At the n-th stage we obtain:

$$\Phi_n(\omega) = \prod_{r=1}^{n} e^{-j\vartheta_r(\omega)} = e^{-j\cdot\sum_{r=1}^{n}\theta_r(\omega)}.$$

For signals which are close to be periodic, as happens in most sound signals, the resulting frequency warping for each partial may be analyzed in the following way. For each partial we fall in the simple case of a single complex exponential tone

$$x(k) = e^{jk\omega_0},$$

with

$$X(\omega) = 2\pi\delta(\omega-\omega_0) \text{ for } |\omega| < \pi.$$

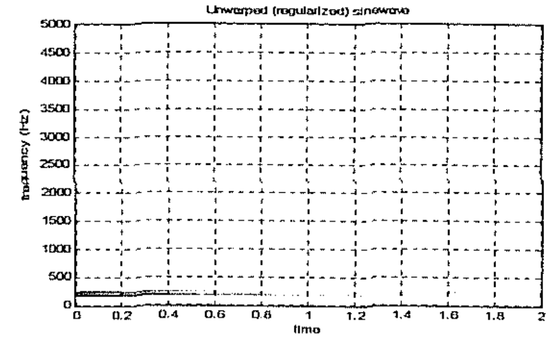


Fig. 5 The swept sinusoid is completely "regularized"

In this case, using the above inversion formula (7), we obtain for the output sample:

$$\hat{x}[n] = e^{j\cdot\sum_{r=1}^{n}\theta_r(\omega_0)} \quad (9)$$

This formula has a simple meaning when the all-pass sections are identical, with $\theta_r(\omega)=\theta(\omega)$, $r=1,2\ldots$. In this case we have: $\hat{x}[n] = e^{j\cdot n\cdot\theta(\omega_0)}$ and the output is a complex exponential whose frequency is the image of the source frequency ω_0 via the warping map $\theta(\omega)$, as expected in this case of ordinary Laguerre Transform.

If different warping laws are applied at each stage by using distinct values of the parameters b_r, then (9) corresponds to a phase-modulated signal, whose features and spectrum depend on the $\theta_r(\omega)$ laws. This signal corresponds to a frequency modulated sinusoid, depending on the choice of b_n, as shown in Fig. 3.

Frequency warping works properly even in case of drastic modification in the source frequency as shown for the linear frequency sweep in Fig. 4, which, after warping appears perfectly reduced to the single tone as in Fig. 5.

5. Acoustical applications of the transform

A first example concerns vibrato removal. Once we have identified the law of frequency variation over time in a sound, we may compute the sequence b_n of values of the Laguerre parameters by means of which the vibrato may be removed. Actually warping the sound by use of this sequence of parameters allows complete removal of vibrato, as shown in Fig. 6. In Fig. 7 we can see that the low frequency component (amplitude envelope) is practically unmodified under the warping transform. In fact the low frequency components, at the usual small values of the Laguerre parameters, move along an approximately 45-degree straigth line under the warping law.

In other applications we may want to increase the depth of vibrato in order to add some expression to our sound. The same warping law may be added, but now the Laguerre parameters will be chose to have opposite sign and will be scaled by a proper coefficient in order to impart any desired depth of vibrato: $-kb_n$. Finally the same depth may be easily modified over time, as in fact it happens in real signals. For example, the vibrato depth may be linked in an arbitrary way to the amplitude envelope of the signal.

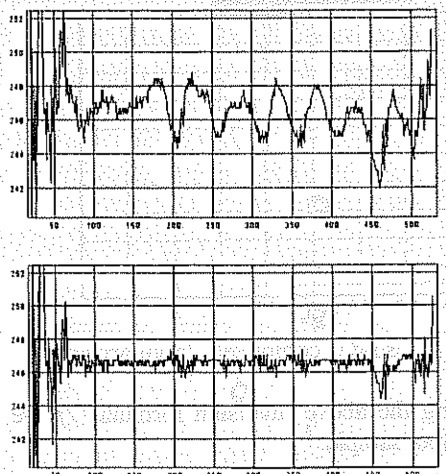


Fig. 6 The pitch of a flute sound showing vibrato and the same after removal by TVFW.

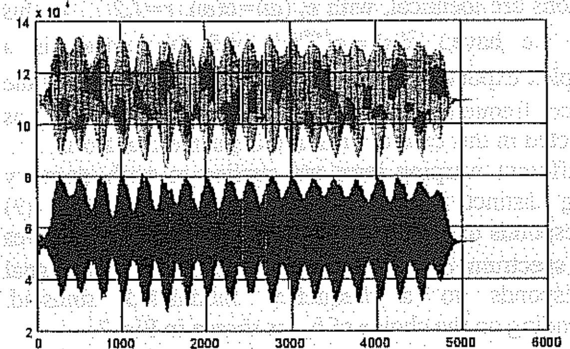


Fig. 7 Time domain flute sound before and after the removal of vibrato

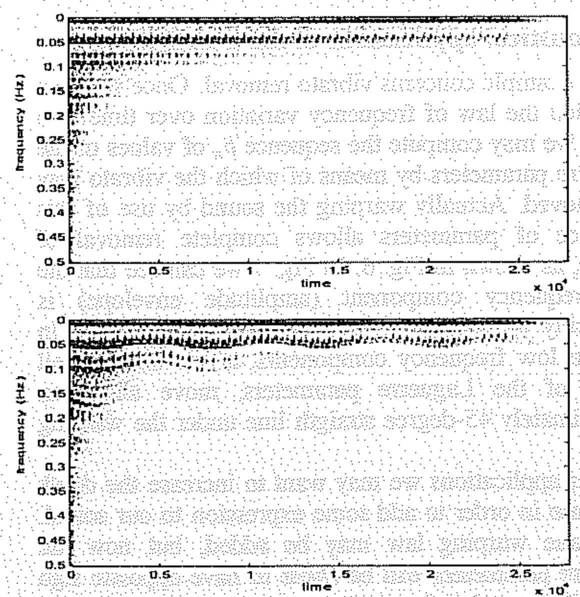


Fig. 8 Karplus-Strong plucked string original and with added vibrato.

A second example is provided, where we added to a simple sound produced by the Karplus-Strong algorithm a selected amount of vibrato (Fig. 8). In this case we may impart the desired amount of 'dispersion' to the harmonic structure (see [5]) and, in the mean time, also add vibrato, thus improving the naturalness of the sound. This is obtained by adding a non-zero mean value to the sequence of warping coefficients. In this way we obtain a mixture of effects typical of the constant and time-varying frequency warping transforms.
Other applications of time-varying frequency warping range from flanging-phasing to chorusing effects, when one mixes multiple time-varying warped versions of the same signal with the original signal, using suitable gains.

References

[1] P. W. Broome, "Discrete Orthonormal Sequences," J. Assoc. Comput. Machinery, vol. 12, no. 2, pp. 151-168, 1965.

[2] A. V. Oppenheim and D. H. Johnson, "Discrete Representation of Signals," Proc. IEEE, vol. 60, pp. 681-691, June 1972.

[3] G. Evangelista and S. Cavaliere, "Frequency Warped Filter Banks and Wavelet Transforms: A Discrete-Time Approach Via Laguerre Expansion," IEEE Trans. on Signal Processing, vol. 46, no. 10, pp. 2638-2650, Oct. 1998.

[4] G. Evangelista and S. Cavaliere, "Discrete Frequency Warped Wavelets: Theory and Applications," IEEE Trans. on Signal Processing, vol. 46, no. 4, pp.874-885, April 1998.

[5] I. Testa, G. Evangelista and S. Cavaliere, "A Physical Model of Stiff Strings," Proc. of the Internat. Symp. on Music and Acoustics, ISMA'97, pp. 219-224 Edinburgh, U.K., Aug. 1997 .

[6] G. Evangelista, S. Cavaliere, "Dispersive and Pitch-Synchronous Processing of Sounds," Proc. DAFX98, pp. 232-236, Nov. 1998.

[7] G. Evangelista, S.Cavaliere, "Arbitrary Bandwidth Wavelet Sets," Proc. ICASSP'98, vol. III, pp. 1801-1804, May 1998.

[8] S.Cavaliere, G. Evangelista, "Representation of Pseudoperiodic Signals by means of Pitch-Synchronous Frequency Warped Wavelet Transform," Proc. EUSIPCO 98, vol. II, pp. 625-628, Sept. 1998.

[9] G. Evangelista, S. Cavaliere, "Analysis and Regularization of Inharmonic Sounds via Pitch-Synchronous Frequency Warped Wavelets," Proc. ICMC '97, pp. 51-54, Sept. 1997.

[10] G. Evangelista, S. Cavaliere, "Time-Varying Frequency Warping: Results and Experiments," Proc. DAFX99 Nov. 1999.

[11] G. Evangelista, "Dyadic Warped Wavelets," invited paper to appear on Advances in Imaging and Electron Physics, Academic Press, Dec. 2000.

[12] G. Evangelista, S. Cavaliere, "Audio Effects Based on Biorthogonal Time-Varying Frequency Warping," to appear on Applied Signal Processing, Springer-Verlag, Dec. 2000.

Timbre Space and Technical Gesture: a study on the Traverse Flute sound.

Andrea Albanese, Nicola Orio

DEI - University of Padova; IRCAM - Centre Georges Pompidou

Via Gradenigo, 6/A, 35131 PADOVA; 1, pl. Igor Stravinsky, 75004 PARIS

andyv@dei.unipd.it; Nicola.Orio@ircam.fr

Abstract

This work investigates the possibility to characterise and identify through sound analysis, different tones produced by an acoustic music instrument when the player applies different technical gestures. We analysed flute recorded samples, where the different sound colours were obtained by consonant articulation in air-jet emission, legato, and noise coming from overblowing, plus a set o heterogeneous tones. We apply the concept of timbre space, owned by each acoustic instrument, and, in particular, we try to identify tones with different timbre due to different technical gesture. We also investigated if it is possible to summarise the acoustic perception of sounds with a limited number of parameters and characteristics

1 Introduction

The word *timbre* is used to identify the sound produced by a particular acoustic instrument. In the literature we can find works focused upon this interpretation, [1], [2], [3]. The concept of timbre sums up the whole set of psycho-acoustic experiences carried out by listeners, and for this reason it is hard to define parameters for measuring it. One direct consequence of this situation is the proliferation of descriptive and sectorial languages which aim to indicate timbre typologies. Often these languages are difficult to understand by non musicians. Moreover, there are projects about information retrieval of sound tones [4], which try to identify different tones, not only using descriptive labels, but by evaluation of parameters that represent the psycho-acoustic perception made by listeners.

We can extend the concept of timbre over the set of different sound that can be produced by the same instrument. Musicians use different techniques to obtain different tones that best represent their intentions and feelings. Like painters with colours and shades, musicians apply technical gesture in different way, which we indicate as "*technical nuances*".

Our work starts form the consideration that if we can measure the effect of each technical gesture in the sound of one instrument, we shall be able to define a timbre space whose axes are parameters of the technical gesture and in which each tone is identified by a set of co-ordinates, this approach is depicted in Figure 1.

Musicians select a particular set of nuances from the nuances space, and apply it to the acoustic instrument, which produce a particular sound that can be view like a point in the *timbre space*. This sound is perceived by listener and by the musicians himself who modifies the technical gesture according with the heard sound and his intentions. Evolution in sound productions due to expressivity can be represent in the Timbral Space as a trajectory.

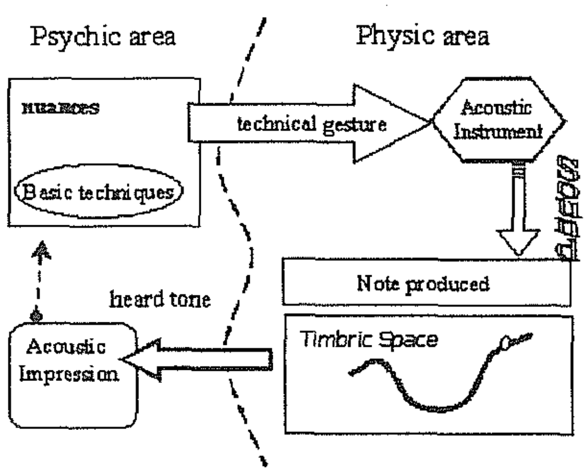


Figure 1 Model of interaction cycle between the musician and the instrument.

2 Timbre and Performing Techniques

The aims of this work are the followings: analysis of the physical properties of sound of the traverse flute; characterisation and identification of timbre modification produced by different technical nuances; highlighting of parameters that characterise different technical performing gestures. The flute, like other members of one family of wind instruments, is excited by air jet, flute players blow air inside a hole and hence the flow is free and hard to be measured it directly. For this reason, our work tries to estimate the excitation by analysis of only the acoustic sounds produced.

We focused on the flute for two main reasons: primary because it has a spectrum more simple than those of most of other acoustic instruments. Secondly there exist a relative steady vocabulary to indicate different techniques, which is universally accepted by all players.

Three recording session were performed at which two musicians (one female and one male) play three different instruments: MURAMATSU MFG-CO, MURAMATSU EX-2, and a YAMAHA studio flute. In each session we selected different microphones, and different analogue to digital chains. All the recorded material were digitised at a sample rate of 44.2 kHz with 16 bit linear coding.

The musicians were asked to play notes in each of the three octaves, using three different loudness: *piano*, *mezzo-forte* and *forte*, and applying different technical gestures. The first technique we wanted to study is the articulation of a consonant during the beginning of the air jet emission. Musicians were then asked to perform notes with a variable level of noise due to voluntary overblowing. Furthermore, musicians performed couple of notes with legato, the distance between the two notes was from a semitone to a fifth. Finally, musicians had the freedom to explore the timbres of their instrument.

All the analyses developed on the sampled sounds are based on the Short Time Fourier Transform [5], and the evaluation of psycho-acoustic parameters such as: Spectral Gravity Centre (SGC), Spectral Irregularity (IRR), [6] and Attack Time, plus En-Harmonic evaluation [7]. Secondly we evaluated Mel-Cepstrum coefficients (MFCC) and applied over them statistical techniques such as the one performed by the evaluation of Principal Components Analysis (PCA) [8]. All the analyses have been carried out using developed routines in Matlab environment aided by Signal Processing toolbox.

3 Consonant articulation in air-jet emission

The analysis of samples where the performers changed the air jet emission, represent a test for our routines and allowed us to gain knowledge about the flute sound physical properties. Musicians used 5 different articulations: d, t (dental consonants), g, k (velar), p (bilabial); together with the simple air emission. These technical gesture modify the excitation jet only in the first front waves and involve in the emission of air-flow the oral-cavity shape and mouth-tongue co-ordination.

The number of relevant harmonics is correlate with the note pitch: tones in the lower octave have up to 14 perceptually relevant partials, while in the second and third octave we identify no more than 10 or 9 partials. In the three cases the harmonic components end at about 10-11 kHz, over these frequencies the amplitude ratio between fundamental and the partials is 60 dB while the ratio with noise is 65 dB. In all the samples partials follow closely the fundamental behaviour and we did not find en-harmonic partials.

Psycho-acoustic parameters (SGC,IRR) and attack times evaluation seemed to us insufficient tools to identify different kind of attack. Hence we developed also other analyses on the harmonic component of the spectrum. We observed the presence of non-harmonic components (i.e. those that are not multiple of the fundamental). Often these new components do not present periodic behaviour, characteristic of harmonic series, however their behaviour is similar to that of non-harmonic partial at lower frequency and is quite different to the behaviour of the fundamental. For this reason we indicate them with the word "sub-partials". Ours non-harmonic partials seem to be localised near acoustic-impedance minimum, as we could see in [9], [10].

These sub-harmonics are affected by consonant articulation. In particular the maximum amplitude, amplitude behaviour in the attack interval, anticipation on the fundamental harmonic series are all characteristic modified by this technical gesture. Sub-partials seem to prevail in the attack and in the decay zone of each note while the ratio between them and the fundamental does not depend on loudness but is more a player characteristic. Moreover, we were able to identify an analogy between the signal attack envelope and the consonant articulation of the same typology like D, T or G and K, see Figure 2.

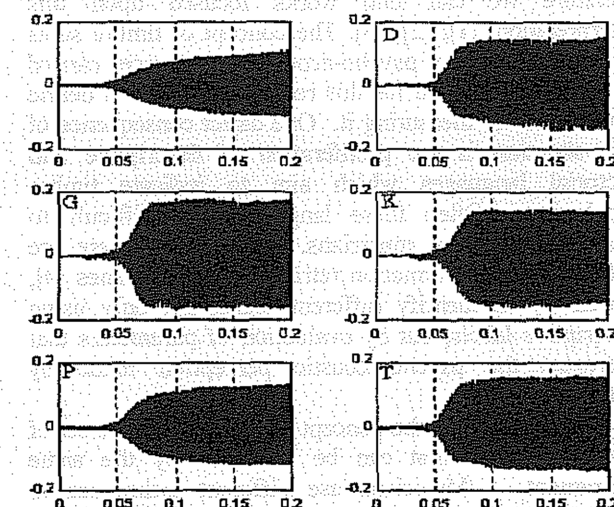


Figure 2: Amplitude envelopes of notes performed using the articulation of different consonant for a A5 forte

Referring to the P consonant, the only bilabial case, it can be seen that the envelope is more similar to the simple air emission than to all other cases, which corroborate ours previous interpretation.

In the tone's first instants listeners can hear an impulsive noise that we interpret as that produced by the first pressure wave fronts hitting the instrument's wall at the other side of the labium. Also if we were not able to identify relationship between the spectral characteristic of this noise, that we called SBF, from Italian word "sbuffo", and the articulation applied by the musician, there is a correlation with time intervals between SBF and attack envelope on-set, see Table 1.

The signal attack envelope, together with the presence of noise in the first milliseconds, could be *considered meaningful.* This was partially demonstrated by the reconstruction of these two properties upon steady sections of flute tones gives realistic synthetic samples.

	p	*mf*	*f*
no consonant	0.172	0.160	0.306
d	0.095	0.081	0.048
t	0.089	0.081	0.045
p	0.123	0.121	0.093
g	0.158**	0.110	0.073
k	0.096	0.781**	0.078

Table 1: Time intervals between SBF and note onset in seconds. ** samples with involuntary vibrato

4 Noise As a Performing Gesture

In this group of analyses, with the word noise, we mean the effect of overblowing made by musicians, and that can be modulated by him. In general players and listeners refers to this quality of sounds speaking of clearer or dirtier samples. Both consonant articulation and air overblowing modify the excitation jet that enter inside the flute's tube by the labium, but the second one modify the air flow during the whole performance of the tone, not only during the first wave's front moreover it can be changed with continuity up to the end of the note.

In this set of samples, the ratio between fundamental and partials and between fundamental and sub-harmonics decrease if noise level raise. We found a reduction of 2.5 dB with high presence of noise. The fundamental is the less amplified harmonic component while partials of higher order are the more intensified, this consideration is confirmed by the evaluation of spectral envelope calculated on the *deterministic* component derived from the originals by the software SMS-tool developed by Serra [11].

We calculated a modified version of SGC and IRR where amplitudes of all frequencies were involved, and the parallel between ours result and values in literature [4], give us a first evaluation about the ratio of noise inside the sample weather or not its amount changes within the sustain phase of the note. Original SGC and IRR on sinusoidal and original sample, as well as their modified version on deterministic sample give all similar values. We concluded that the noise due to overblowing does not modify the distribution of energy between harmonic components.

The presence of a richer spectrum give us the possibility to identify some regions where the mean energy is lower than in the neighbourhood at about 5 kHz, 10 kHz and 14 kHz. These region are present in all the samples but in this case they are more visible. Even if in these areas there are harmonic components, noise intensified less these partials than all the other.

We evaluated MFCC coming from a 29 filters bank and applied upon them the PCA. Projections on the axis realised with the only first principal component (that explain up to 90% of the variance) show a similar trend to the one perceived by listener when hearing samples of this typology of noise. Usually if noise level increases non-linearly inside the note, a saturation phenomenon happens: maximum noise position is heard before the real max level is reached, see Figure.3.

In spectral energy reconstruction, we highlighted the sensibility, both in the original and "residual" samples, of this first component to the minimum areas we referred in the previous paragraph. In order to focus PCA to slow varying components we used also less that the 29 MFCC, but these analysis did not bring any further information.

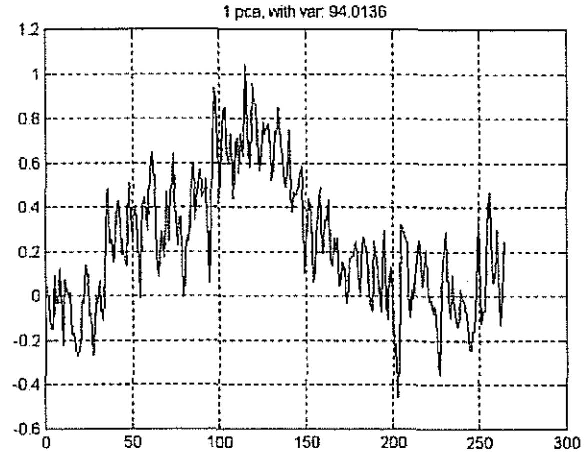


Figure 3: Projection on the first PCA axis, FFT frame number in abscissa

5 Performance with Legato

The *legato* technique is the more complex gesture we analysed, because it is made by sum of micro-gesture that modify continually the acoustic impedance and the excitation pattern. Samples were made by couples of notes, in the first two octaves, which tonal intervals go from 1 up to 6 semitones. Musicians apply legato to pass between notes softly, without stop in air emission and without silence insertion.

We saw that, during the passage from the first note to the second one, signal envelopes showed drop of energy, that became wider and deeper as the tonal distance grows, while the acoustic impressions remain quite similar to each other. In the CGS and IRR time evolution we identify a similar behaviour which evaluation can be used to estimate the tonal interval without refers directly to note's pitch.

In spectral domain, the legato technique brings some interesting characteristic: partials belonging to the first note of the couple tend to pursue in the second (we measured up to 80 milliseconds continuation). If the tonal intervals is less than 4 semitones, which means no more than 130 Hz gap between partials of same order *belonging to the couple, they resolve to their analogues,* but if the semitones became 4 and the couple is in the second octave, partials resolve to nearest harmonic component also if this one belong to different harmonic series: fundamental family partials to sub-harmonic components and vice versa. An example of legato is depicted in Figure 4.

In all cases, during the legato there are zones of high energy distribution from about 4kHz and 6kHz, in the first octave, and from 8 to 9 kHz in the second and third, that are not influenced by spectral configuration, tonal intervals or pitches. It seem to us that, in order to identify this technique, an useful tool would be the characterisation of partials and energy behaviours.

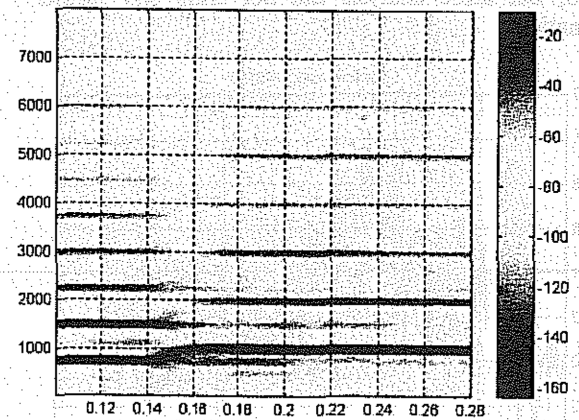


Figure 4 Spectrogram, in dB scale, of the performance of a legato between F5# and C5

6 Timbral and Dynamic Explorations

The dynamic influences both the signal amplitude (energy), and the spectral energy distribution together with the partials amplitude. On the other hand, musicians can change the timbre of the sound without modifying its amplitude. For this last set of samples, the evaluation of only one aspect is too limitative to resume the whole range of the acoustic perception. In order to obtain a single parameter that could resume this perception, we evaluated PCA and MFCC. These techniques were able to underline the noise level influence in timbre identification and characterisation. Projection of the first axes coming from PC analysis to 29 MFCC of dynamic exploration samples, give a good identification of the loudness and heaviness among samples with a similar noise level.

In the analysis of samples of timbre explorations only, we subdivided all the samples in two main sets, each one characterise by clearness. Projection on the first two axes which correspond to the first two PC show this subdivisions on a Cartesian plane.

As in previous case (overblowing , reconstruction of spectral energy distribution show that PC were sensible not only to harmonic components.

7 Conclusions

In the previous paragraphs we show several parameters or properties that can be used to identify the different technical gesture. In some cases we were able to identify not only the technical but also the nuances as in the case of consonant articulation, or in the case of noise coming due to jet-air overblowing. The analysis of audio signal via Time-Frequency techniques gives us sufficient information to identify each set of samples.

Many aspects remain out of our analyses as this is an initial work about timbre space and the primary objective was to investigate whether this kind of approach is useful and productive for research. We believe that our results are a step forward toward the comprehension of the links among the musicians gestures and the quality of the produced sound.

Moreover, results of this kind of work can be helpful in all the situation where different timbre must be identified or described between musicians or listeners and researchers; for instance we can refer at the simplicity of talk about a more T attack than G attack, unless talking of time attacks, envelope and noise. The former approach is surely more closer to the idea that musicians have of their instruments and their capabilities

8 References

[1] Wessel D.L. 1978 "Timbre Space as a musical control structure." *Rapport Ircam 12/78 1978* on IRCAM-CGP online server.

[2] Sandell G. J., Martens W. L. 1995 "Perceptual Evaluation of Principal Component Based Synthesis of Musical Timbres" *Journal of Audio Engineering Society 43(12) 1995*

[3] McAdams S., Cunibile J-c. 1992 "Perception of timbral analogies" *Philosophical Transactions of the Royal Society (vol 336) London Series B* on IRCAM-CGP online server

[4] Database SHARC 1999 available at the address: http://sparky.parmly.luc.edu/share, site written and maintained by Gregory J. Sandell.

[5] Kronland-Martinet R., Grossman A. 1988 "Application of Time Frequency and Time-Scale methods (Wavelet Trasform) to the Analysis, Synthesis and Transformation of Natural Sounds"

[6] Krumhansl, C. L. 1989 "Why is musical timbre so hard to understand?" in *Structure and perception electroacustic sound and music* (ed. E. Nielzen & O. Olsson Elsevier, Amsterdam, expcerpta medica 846), 43-53).

[7] Gray J. M. 1975 "*An Exploration of Musical Timbre*", report STAN-M-2, Standford University.

[8] Beyerbach D., Nawab H. 1991 "Princiapl Components Analysis of the Short Time Fourire Trasform" *IEEE ASSP-39(3) 1991*

[9] Wolfe J., Smith J., Tann, J. 1998 "Flute Acoustics: A Web Research" Musical Acoustic Group School of Physics The University of new South Wales. http://www.phys.unsw.edu.au/music/flute

[10] Wolfe J.,Smith J.R., Enrich N. 1997 "The Acoustic Impedance of the Boehm Flute: Standard and Some Non-Standard Fingerings" Proc. Inst. Acoustics, 19, 315-330

ECOLOGICAL MODELLING OF TEXTURAL SOUND EVENTS BY ITERATED NONLINEAR FUNCTIONS

Agostino Di Scipio

Scuola di Musica Elettronica, Conservatorio di Musica di Bari
labmuson@tin.it

Abstract

Ecological issues in computer-based sound modelling have so far been referred to the peculiar potential of granular synthesis. This paper outlines a different approach, using real-time *iterated nonlinear functions*. Based on the mathematical modelling of system dynamics, it illustrates for the generation of textural and environmental sounds hardly obtained with other methods. The question is raised as to how we establish a useful connection between the system dynamics and the perceptual attributes of sound textures.

A short overview is provided of synthesis by iterated nonlinear functions (Functional Iteration Synthesis). Examples are discussed of synthesized sound textures heard as different kinds of rain, thunderstorm, acoustic turbulences and other phenomena.

Overall, the approach outlined here may open to new experiments in electroacoustic music, and the creation of synthetic, but credible, auditory scenes in multimedia applications and virtual reality.

1. INTRODUCTION

The modelling of perceptual attributes of complex auditory images requires a number of *ecological* considerations. Here, I would like to discuss the computer generation of environmental sound textures, in terms of an ecological approach on sound modelling.

In the digital sound synthesis/processing community, ecological considerations have so far been mainly referred to sound design by granular synthesis (for example [1]). In the following I illustrate a different path, using iterated nonlinear functions as a sound synthesis engine. I will discuss examples including sonic textures reminiscent of rains, thunderstorms, and phenomena of acoustic turbulence.

The first part of the present paper, provides an overview of iterated nonlinear functions synthesis, and illustrates a particular model. The second part focuses on the relevance of nonlinear dynamics to an ecological approach on sound modelling, emphasizing the dependency of system components in terms of their *context* and overall emergent behavior.

While it builds on recent work in digital sound synthesis with iterated nonlinear functions [2][3][4][5][6], the present research is also inspired to work in *auditory scene analysis* [7]. However, it is more concerned with the holistic dynamics of sound synthesis algorithms than with an explanation of psychoacoustical details. The assumption is that one can study complex perceptual and cognitive phenomena by creating a perceptually viable simulation, and that for this to be possible nonlinear dynamics is especially valuable.

In a sense, this line of research takes on older issues in computer music research, such as pursued, a.o., by Jean-Claude Risset, i.e. the simulation of instrumental sounds based on psychoacoustical considerations. The present work enlarges that view to the synthesis of credible auditory scenes and their global emergent properties. This turns to be useful either in compositional work, or in sound design for multimedia and virtual reality.

2. SYNTHESIS BY *FUNCTIONAL ITERATION*

Consider these definitions: $A \subset \Re$ (a set of "init values" for some mapping process); $G \subset \Re^m$ (the parameters in the mapping), and $B \subset \Re$ (a set of samples in the generated digital signal). And consider the cartesian product $A \times G \subset \Re \times \Re^m$. Let our map, F, be defined as

$$\begin{array}{llll} F: & A \times G & \rightarrow & B \\ & (x, \{a_i\}) & \rightarrow & F(x; \{a_i\}) \end{array}$$

where ($\{a_i\} \equiv a_1, a_2, \ldots, a_m$). This represents a parameter-dependent function which maps from A to B, with a_i as a set of time-changing parameters. By fixing a set of m real parameters (a point in G) we have:

$$\begin{array}{llll} f: & A & \rightarrow & B \\ & x & \rightarrow & f(x) \end{array}$$

where $f(x) \equiv F(x; a_1 \ldots a_m)$. If $B \subset A$, then we can implement an iterated map process of by iteratively applying f to itself for n times. Finally, considering g_i a sequence of points in G, and provided that $x_{0,i} \in A$ and $g_i \in G$, we get a sequence of maps f_i. A discrete time series is generated where each sample is the n-th iterate of the same function using different parameters:

$$x_{n,i} = f^n{}_i(x_{0,i}) = f_i(f_i(...(f_i(x_{0,i}))...))$$

This is a general template for any method of *Functional Iteration Synthesis* (FIS). For interesting sounds, any well-chosen nonlinear f can be adopted, and smoothly changing control functions should be used to update the parameters value at each iterated process (= at each sample). Indeed, FIS is less a specific sound synthesis technique than a class of methods sharing the iteration of nonlinear maps as the basic operation. The crucial elementi here is more in the process of iteration than the function itself. [9]

3. ITERATING THE "SINE MAP"

Consider the following nonlinear map:

$$F : [-\pi/2, \pi/2] \times [0, 4] \rightarrow (-1, 1)$$
$$(x, r) \rightarrow \sin(rx)$$

The iterated discrete form is

$$x_{n,i} = \sin(r_i\, x_{n-1,i})$$

which is often found in the literature on "chaos theory" *[10][11][12]. W set* A = [$-\pi/2$, $\pi/2$] *because larger intervals* for the init values would result into trajectories also obtained starting from within [$-\pi/2$, $\pi/2$] (due to the periodicity of the sine function). The first iterate falls in the interval [-1,1], completely covered by $sin(rx)$ for x_0 within [$-\pi/2$, $\pi/2$] and $r \geq 1$. Also, G = [0, 4] because larger values for r would provide results similar to, if not identical with, those provided by $r < 4$.

In the dynamics of the iterated sine map, higher iterates usually yield richer (more active and dynamical) trajectories. These trajectories, often drawing oscillatory paths, are taken as output signals. The numerical relationship among these trajectories (and among their derivatives) seem not to have been addressed by any scientific study. We find ourselves in a territory where little help is provided by purely analytical means. Research here turns to more explorative, empirical methodologies. With higher iterates, the init value x_0 is soon *forgotten*: one cannot tell, by any analytical means, where in the numerical interval [$-\pi/2$, $\pi/2$] the iteration process was started. Transient fluctuations disappear and the overall system behavior could be rendered as a "bifurcation diagram" peculiar to chaotic systems.

4. EXPLORING THE PHASE SPACE

Considering the discrete sine map model, r determines the kind of waveform shape in the sound signal, ranging from very smooth curves (e.g. r = 2) to more intricate oscillations (e.g. r = 4). The latter signals always shows *continuing frequency-* or phase-modulations.

On the other hand, x_0 determines the actual shape in the output sampled waveform. Slight changes result in signals maybe overlapping at the outset but then gradually shifting apart from one another ("dependency on the initial conditions").

The number of iterations, n, mainly affects the output bandwidth. With larger values, the output signal goes through wider and wider oscillations, which might even have a fractal shape (signal contours are repeated at various time-scales). Eventually the spectrum gets denser and broad-band noise is obtained. In most cases, n = 9 already returns dense noise textures. The bandwidth also depends on the rate by which we are sampling the orbit in the phase space (i.e., on the actual "step" in walking across the system phase space).

At synthesis run time, we may (a) change r and keep x_0 *constant*; (b) change x_0 and keep r constant; and (c) change both r and x_0 (the iterate order, n, cannot change, as that would determine audible discontinuities). These changes create, respectively, (a) varying contours in the signal waveform (dynamical spectra); (b) different signal waveforms *of the same kind* (nearly-constant spectrum); and (c) a mixture often heard as articulated sound textures in a constant flux of change, including sudden "pauses" (sub-audio frequencies and DC offsets; these may require high-pass filtering to suppress).

Here, I cannot discuss the computer implementation of FIS. A straightforward C language implementation of the iterated sine map model is in [3] (audio examples are found at *http://www.swets.nl/jnmr/vol25_1.html#discipio25.1*). FIS Csound orchestras are discussed in [13]. A real-time *implementation* with the Kyma interactive workstation (™ SymbolicSound) is discussed in public documents available at *ftp://shout.net/pub/symsound/kymasnds/Agostino_DiScipio/*. In the latter case, the iterated sine map model is described as a kind of generalized waveshaping synthesis [14].

5. RELEVANCE OF SYSTEM NONLINEARITY TO ECOLOGICAL MODELLING

The time-dependent relationship between r_i and $x_{0,i}$ is crucial, and radically changes when different iterates (n) are considered. In a typical situation, one cannot modify any of the three values (r, x_0 and n) without causing, as a side-effect, a change in the way the others affect the overall result.

This phenomenon reflects reveals the non-integrability of chaotic systems to the ear. We are only left with the possibility of a *qualitative* characterization of the interdependency among parameter values. Clearly it is impossible to predict the precise output of any configuration of values. Put the other way round, it is impossible to fix the values to get a specific target sound, especially when the model is not forced to periodic behaviour. The exploration of the phase space, and the exploration of the parameter space as an ideal (but far from obvious) mapping of the perceptual space in the audible effects, should be left to interactive experiments (this is why a highly interactive real-time implementation is necessary in musical contexts [5][15]). Only *a posteriori* it becomes possible to fix some values and delimit phase space regions of special interest. That is the case with the sound examples illustrated in the following section.

What is significant in this approach, is that it provides the opportunity for a perceptual modelling of textural sound events based on a time-domain synthesis method. It is common understanding that sonic phenomena of textural nature – such as the sound(s) of the rain, cracking of rocks and icebanks, thunders, electrical intermittent noises, the sound(s) of the wind, various kinds of "sonorous powders", burning materials, rocky sea shores, certain kinds of insects, etc. – are best modelled by microstructural time-based design strategies, rather than spectral modelling strategies (see examples in [8][16], and musical considerations in [17][18], where, anyway, granular synthesis methods were employed).

It seems very appropriate that the effects generated with FIS require wholistic and largely indeterministic controls. The crucial element is that the approach provides cues as to take into consideration the dynamical relationship of as many elements contributing to such sonic events as possible. By establishing a structural link between those elements, it represents an approach of *ecological* validity. By considering the elements in their emergent sound properties, it provides a style of *perceptual modelling* (as distinct from, say, *physical modelling*).

6. SONIC TURBOLENCES AND THUNDERSTORMS

However "chaotic", the output of iterated nonlinear functions cannot be likened to fitlered white noise, however complex the filter system can be. The rate of change in the waveform of signals such as white- or $1/f^2$ noise is too fast for the ear and finally result into a sustained, non-articulated sound.

Sounds obtained with the iterated sine map model, instead, feature an internal articulation of their own, with random fluctuations which are anyway slower than white noise. Phase-modulations and amplitude curves are somehow "built in", they are innate to the system dynamics, and result into a micro-level activity which is useful in order to approximate environmental sound textures. In a real auditory scene, moving sources and reflecting surfaces cause all sorts of interferences among signals, which are for the ear as a peculiar element of realistic sound ambience. The rate of change in the amplitude of the generated signal usually ranges between shorter than 1/10th of a second to seconds. Which means, also, that applications can rely on the emergence of amplitude shapes "internal" to the sound (only some fade-in and fade-out is necessary to avoid discontinuities when the sound starts and ends).

The output signals obtained by visiting at random the phase space of the model, are heard as acoustical turbulences, very low-frequency (and even sub-audio frequency) rumbling sonorities. The actual frequency contents depend on the orbital velocity. Higher velocities result into larger bandwidths. Some effective but indeterministic control on the bandwidth is possible with carefully studied control functions or signals for r and x_0.

Taken *per se*, the low-frequency narrow-band noise of the iterated sine map model is a good starting point to model the perception of textures of, e.g., boiling water, sulphureous or volcanic areas, and the wind flapping against thin but large plastic or aluminum plates.

With higher-order iterates (e.g. $n > 8$), the signal waveform shows interesting correlations at multiple time-scales. Energy is scattered around in small, phase-modulated wavepackets of different lengths and amplitude (the sound is rather reminiscent of a large fire). With a high-order band-pass filter, it would be possible to isolate the energy at a particular time-scale. The effect would be that of a granular texture with a specific density of sonic microevents. However, more interesting is to submit the output of the iterated process to a simple 2[nd] order high-pass filter: by gradually lowering the cut-off frequency, sound droplets of larger and larger size are introduced, similar to the accumulation of raindrops when a rain shower is starting. What changes is the "sonic perspective", i.e. the auditory effect due to a change of the listener's position as relative to the physical phenomenon. There evidence emerges of an ecological view on sound modelling: what is modelled is the sonic effect of the mutual interplay of the physical components *in context*.

With rain-like sounds, it is important to introduce some perceptual cue as to the different surfaces impacted upon by the raindrops. This can be done by choosing the appropriate number of iterates, n. Upon listening, larger values of n result into more resonant materials, while smaller values are more reminiscent of "deaf", very compact surfaces.

Furthermore, by visiting the extreme regions in the system phase space ($r \leq 4$), we can control the perceived size of droplets, resulting either into a drizzling or a driving rain. As the droplet size gets larger, eventually a hailstorm is approached.

In this process, taking the cut-off frequency close to zero is like letting all energy at the different time-scales in the signal pass the filtering, resuming the characteristic turbulence of the "natural" state of the iterated sine map model. However, this time the sound texture is phase-modified by the high-pass, and might result into a surprisingly realistic thunderblot.

With just a few changes in the parameters value, that which so far was described as (various types of) rain or hail, becomes more similar to the sound of frying oil, or to boiling water burbles. In order to come closer to such sound effects, the underlying model may be re-configured either using an iterated function other than a sine, or adding some random numbers in the iterated sine map at each sample computation.

In musical contexts, I tend to use a large palette of finely-tuned parameter configurations in the synthesis, and to compose the emerging textures and gestures with the help of some higher-level iterated function system. That can be described an instance of algorithmic composition, but it uses chaotic textural materials instead of musical notes as the basic musical units. In a recent computer music composition of mine, *Natura allo specchio*, I created utterly synthetic sound environments that eventually are perceived as naturalistic, keeping the listener in an ambivalent situation where the sound is overtly artificial and still preserves something (a dynamical behaviour) that is realistically perceived as proper to natural sonic ambiances.

7. CONCLUSIONS

In this paper, I overviewed an approach on sound synthesis described as *Functional Iteration Synthesis*, and illustrated the particular case of the iterated sine map model. I showed that the nonlinear dynamics proper to the system dynamics can be an advantage when modelling sound percepts of textural nature. An ecological view is then required, taking into account the mutual dependency, *as heard by the human ear in context*, of several components playing a role in environmnetal sounds.

The subject discussed here represents a new area of concern in the field of digital sound synthesis, and is linked to an ecological approach in studies on auditory perception [7]. Due to the nonlinear dynamics of the synthesis models and the breath-taking complexity in the modelled phenomena, it is clear that more systematic research work is needed. Hopefully, the present paper delineated technological and scientifical aspects of relevance to the subject.

References

[1] Keller, D. and Truax, B. "Ecologically-based Granular Synthesis", *Proc. ICMC*, Ann Arbour Michigan, 1998.

[2] Di Scipio, A. "An Overview of Digital Sound Synthesis by Models of Nonlinear Dynamical Systems", *Sixth IMA Conference on Control: Modelling, Computation, Information*, Institute of Mathematics and its Applications, Sounthend-on-sea, 1992.

[3] Di Scipio, A. and Prignano, I. "Synthesis by Functional Iterations. A Revitalization of Nonstandard Synthesis", *Journal of New Music Research*, 25(1), 1996.

[4] Prignano, I. "Sintesi di eventi sonori complessi mediante iterazioni funzionali", Proc. *XI CIM*, AIMI, 1995.

[5] Di Scipio, A. "Composing with Iterated Nonlinear Functions in Interactive Environments", *Proc. XII CIM*, AIMI, 1998.

[6] Di Scipio, A. "Iterated NonLinear Functions as a Sound Generating Engine", *Proc. 1st Iteration Conference*, Melbourne, 1999. (Forthcoming extended version in *Leonardo Music Journal*).

[7] Bregman, A. *Auditory Scene Analysis*, MIT Press, 1990.

[8] Warren, R. and Verbrugge, R. "Auditory Perception of Breaking and Bouncing Events: a Case Study of Ecological Acoustics", *Journal of Experimental Psychology. Human Perception and Performance*, 10(5), 1984.

[9] Feigenbaum, M "Universal Behavior in Nonlinear Systems", *Los Alamos Science*, 1, 1980.

[10] Campbell, D. "Nonlinear Science. From Paradigms to Practicalities", *Los Alamos Science*, 15, 1987.

[11] May, R. "Simple Mathematical Models with Very Complicated Dynamics", *Nature*, 261, 1977.

[12] Collet, P. and Eckmann, J.P. *Iterated Maps on the Interval as Dynamical Systems*, Birkhauser, 1980.

[13] Di Scipio, A. "Sintesi del suono mediante iterazione di funzioni nonlineari", in (R.Bianchini and A.Cipriani, eds.) *Il Suono Virtuale*, Contempo, 1999.

[14] Arfib, D. "Digital Synthesis of Complex Spectra by means of Multiplication of Non-linear distorted sine waves", *Journal of the Audio Engineering Society*, 27(10), 1979.

[15] Choi, I. "A Chaotic Oscillator as a Musical Signal Generator in an Interactive Performance System", *Journal of New Music Research*, 26(1), 1997.

[16] Schumakler, M. and Gilden, D. "Auditory Perception of Fractal Contours", *Journal of Experimental Psychology*, 19(3), 1993.

[17] Di Scipio, A. "Micro-time Sonic Design and the Formation of Timbre", *Contemporary Music Review*, 10(2), 1994.

[18] MacKay, J. "On the Perception of Density and Stratification in Granular Sonic Textures", *Interface-Journal of New Music Research*, 13, 1984.

SOUND ANALYSIS AND SYNTHESIS II

SOUND ANALYSIS AND SYNTHESIS II

NUMERICAL STUDY OF THE ACOUSTIC PROPERTIES OF PIANO SOUNDBOARDS

Bernardo Bazzi (1) and Davide Rocchesso (2)

(1) Bernardo Bazzi, via Mascagni 28/1 - 41100 Modena - Italy - bernardo.bazzi@katamail.com
(2) Università di Verona - Dipartimento Scientifico e Tecnologico
Strada Le Grazie, 15 - 37134 Verona - Italy - rocchesso@sci.univr.it

ABSTRACT

Starting from measurements of the physical properties of piano soundboards, as they are found in the literature, we developed a finite-difference numerical model parameterized on actual boards. The underlying equation is the equation for bending waves in a thin (orthotropic) plate, where the displacement of the plate is normal to the propagation of the wave. The numerical stability has been assessed by means of the von Neumann analysis, which gives limiting values for the sampling frequencies as the physical parameters are varied.
We have focused our attention on mechanical impedance (magnitude and phase), in the frequency range 20-5000 Hz, and our numerical results match the experimental measurements quite well. For the purpose of modeling the piano sound, it is convenient to look at the conductance G, which is the real part of the mobility (reciprical of mechanical impedance). After some initial experimentation performed on a square table, the board has been shaped after a grand piano board (Steinway D-274), with the right orientation of the wood-grain for the case at hand. The model gives also the possibility to look at changes occurring if we modify some parameters such as thickness, density, damping factor (of the sounboard) number of ribs, and so on [1].

We also improved our model changing the boundary conditions from clamped ends to semi-hinged ends, to better simulate the real situation.

Then our model has been tested in this way: we set the physical parameters on standard values and with semi-hinged ends we found the first four modes of the soundboard of the Steinway D-274; they are very similar to experimental measures both quantitatively and qualitatively.

1. UNDERLYING EQUATION AND ALGORITHM

1.1. Underlying equation

The wood of a soundboard is strongly anisotropic: the Young's modulus along the direction of the grain is much larger than the cross-grain modulus. So we had to consider the equation for bending waves in an orthotropic plate, that is [2]:

$$\rho h\frac{\partial^2 z}{\partial t^2} = -D_x\frac{\partial^4 z}{\partial x^4} - (D_x\nu_y + D_y\nu_x + 4D_{xy})\frac{\partial^4 z}{\partial x^2\partial y^2} - D_y\frac{\partial^4 z}{\partial y^4} \tag{1}$$

where $z = z(x, y, t)$ is the displacement of the board in the direction normal to the plane xy, h is the thickness, ρ is the density and D_x, D_y, D_{xy} are the rigidity factors which are given by:

$$D_x = \frac{h^3 E_x}{12(1-\nu_x\nu_y)}, \quad D_y = \frac{h^3 E_y}{12(1-\nu_x\nu_y)}, \quad D_{xy} = \frac{h^3 G_{xy}}{12}$$

where E_x, E_y are Young's modulus, ν_x, ν_y are Poisson's ratio and G_{xy} is the shear modulus.

1.2. Algorithm

In order to treat equation (1) by means of the finite-difference method, we changed the continuous variables into discrete variables, so the displacement z changed in this manner:

$$z(x, y, t) \longrightarrow z(i\Delta x, j\Delta y, n\Delta t) \longrightarrow z(i, j, n) \tag{2}$$

where i, j and n are positive integers.

Moreover, to study the properties of the soundboard we added a damping force term having form $-R\Delta x\Delta y dz/dt$ (where R is a damping factor) and a transverse driving force $F = F(i, j, n)$ which simulates the transfer of energy from the string to the soundboard through the bridge. Then we obtained the following finite-difference equation:

$$\begin{aligned} z(i,j,n+1) &= a_1 z(i,j,n) + a_2 z(i,j,n-1) \\ &\quad + a_3 X + a_4 Y + a_5 H + a_6 F(i,j,n) \end{aligned} \tag{3}$$

where

$$\begin{aligned} X &= z(i+2,j,n) - 4z(i+1,j,n) + 6z(i,j,n) \\ &\quad -4z(i-1,j,n) + z(i-2,j,n)\,, \\ Y &= z(i,j+2,n) - 4z(i,j+1,n) + 6z(i,j,n) \\ &\quad -4z(i,j-1,n) + z(i,j-2,n)\,, \\ H &= z(i+1,j+1,n) - 2z(i,j+1,n) - 2z(i,j-1,n) \\ &\quad -2z(i+1,j,n) - 2z(i-1,j,n) + z(i+1,j-1,n) \\ &\quad +z(i-1,j+1,n) + z(i-1,j-1,n) + 4z(i,j,n)\,, \end{aligned}$$

$$\begin{aligned} a_1 &= \frac{2}{1+\beta}, \qquad a_2 = \frac{-1+\beta}{1+\beta}, \\ a_3 &= \frac{-D_x(\Delta t)^2}{(1+\beta)\rho h(\Delta x)^4}, \qquad a_4 = \frac{-D_y(\Delta t)^2}{(1+\beta)\rho h(\Delta y)^4}, \\ a_5 &= \frac{-(D_x\nu_y + D_y\nu_x + 4D_{xy})(\Delta t)^2}{(1+\beta)\rho h(\Delta x)^2(\Delta y)^2}, \\ a_6 &= \frac{(\Delta t)^2}{(1+\beta)\rho h\Delta x\Delta y} \end{aligned} \tag{4}$$

with $\beta = R\Delta t/2\rho h$.

Then we needed to add ribs to approach the real soundboard. The ribs are orthogonal to the grain of the board, they are tipically 2.5 cm large and up to 2.5 cm high; the distance between them is 10-15 cm. The cross-grain stiffness of a ribbed solid-wood soundboard comes predominatly from the ribs. To a first approximation, the ribs supply enough stiffness to make the cross-grain propagation characterisics of the soundboard similar to those in the direction of the grain. To include the presence of the ribs in the model, we had to assume that the rigidity factors in the location of the ribs are:

$$D_{x,rib} = \frac{h^3 E_x}{12(1-\nu_x\nu_y)}, \qquad D_{y,rib} = \frac{(h+h_{rib}^3)E_{rib}}{12(1-\nu_x\nu_y)},$$
$$D_{xy,rib} = \frac{(h+h_{rib})^3 G_{xy}}{12}, \quad (5)$$

where h_{rib} is the average thickness of the ribs and E_{rib} is the Young modulus in presence of the ribs which run parallel to y axis (the direction of the grain is along x axis).

1.3. Von Neumann stability analysis

The von Neumann stability analysis gave us a criterion for choosing the time and space discretization steps (2). The independent solutions of the difference equation (3) are all of the form [3, 4]:

$$z(i,j,n) = \xi^n e^{i(k_x i\Delta x + k_y j\Delta y)}, \quad (6)$$

where k_x and k_y are the real components of the bidimensional wave vector $\vec{k}$, and $\xi = \xi(\vec{k})$ is a complex number that depends on $\vec{k}$; i represents the imaginary unit. The von Neumann stability requests that [3]:

$$|\xi(\vec{k})| \leq 1 . \quad (7)$$

To find $\xi(\vec{k})$ in our case, we substituted (6) back into (3) and we obtained

$$\begin{aligned}\xi^2(\vec{k}) \; = \; & a_1\xi + a_2 + a_3\xi[2\cos(k_x 2\Delta x) - 8\cos(k_x\Delta x) + 6] \\ & + a_4\xi[2\cos(k_y 2\Delta y) - 8\cos(k_y\Delta y) + 6] \\ & + a_5\xi[4\cos(k_x\Delta x)\cos(k_y\Delta y) \\ & \quad - 4\cos(k_x\Delta x) - 4\cos(k_y\Delta y) + 4] . \quad (8)\end{aligned}$$

After lengthy calculations [1], we could derive the following formula for the right temporal sampling rate:

$$f_s = 0.32\frac{2\pi(h+h_{rib})}{(\Delta x)^2}\sqrt{\frac{E_{Max}}{3\rho(1-\nu^2)}}, \quad (9)$$

where $f_s = 1/\Delta t$ is the sampling rate and E_{Max} is the maximum value of Young's modulus (in our case E_{rib}). Such formula is very usefull because it gives the right sampling rate related to the physical parameters. In fig. 1, we report an example in which the (7) is not observed.

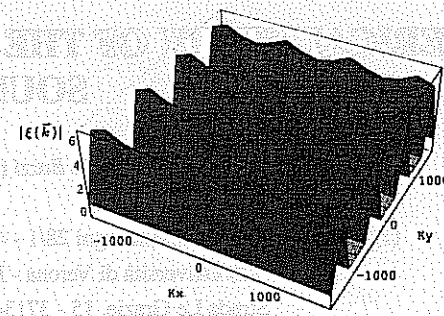


Figure 1: Stability factor magnitude for an orthotropic table with ribs for the case in which the sampling rate is half of the value requested by (9).

2. APPLICATIONS

2.1. Mechanical impedence

For semplicity we started considering square boards with clamped edges as boundary conditions. We obtained results for the mechanical impedence (magnitude and phase) similar to experimental measurements [5] and also to prior work similar to this [2]. In fig. 2 we report an example of such results [1]: the magnitude of mechanical impedence for a squared board with 12 ribs; the numerical results should be considered valid up to 5000 Hz [1].

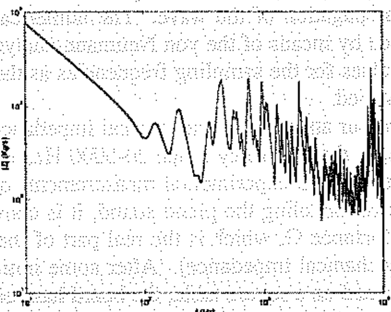


Figure 2: Magnitude of mechanical impedence for a square board 1×1 m^2 with 12 ribs large Δx=0.02 m and standard parameters [2]: h= 1 cm, ρ = 500Kg/m^3, $E_x = 1\times10^{10}$ N/m^2, $E_y = 4\times10^8$ N/m^2, ν_x = 0.4, ν_y = 0.01, $G_{xy} = 1.85\times10^8$ N/m^2, h_{rib} = 1 cm and $E_{rib} = 2\times10^{10}$ N/m^2; R = 1000 Kg/(m$^2\times$ s) (arbitrary). Driving force in $x = y = 0.3$ m (origin at one corner of the square).

2.2. Conductance

Mechanical impedence is frequently used in experimental measurements. However, from the viewpoint of modeling the piano sound, it is more convenient to look at the conductance G, which is the real part of the mobility (reciprocal of mechanical impedance). In fact if we consider the expression for the decay time of the vibration in the string we note the presence of G [6]:

$$\tau(f) = \frac{1}{(8\mu L f^2 G(f))}, \quad (10)$$

where f is the frequency, τ is the string vibration decay time, μ is the string mass linear density, L is the string length and G is the conductance at the bridge.

We want to underline that G is a function of frequency f, and this fact largely affects the sound generation process. In addition,

[1] Results that we obtained with an impulsive driving force and through the use of the Fourier transform

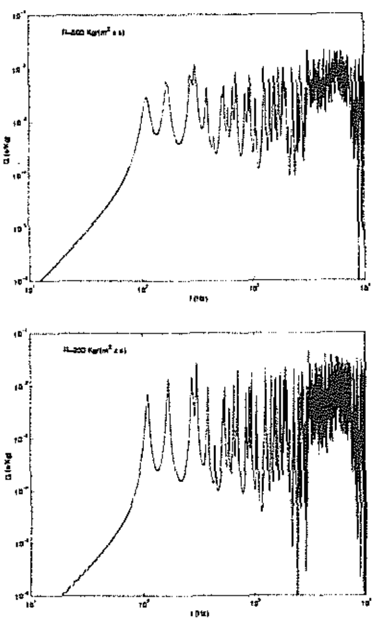


Figure 3: Conductance for a square board 1×1 m^2, with 12 ribs large Δx=0.02 m and same standard parameters of fig.2; measured in (0.3 m, 0.3 m) for two different values of the damping factor: R = 500 kg/(m$^2 \times$ s) and R = 200 kg/(m$^2 \times$ s).

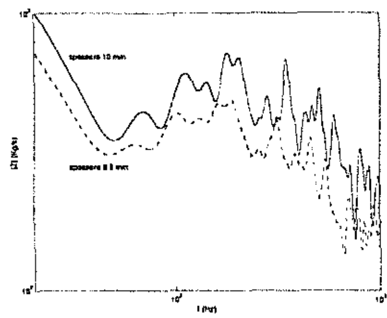


Figure 4: Magnitude of the mechanical impedence measured for two different values of thickness of a grand piano soundboard: h =10 mm and h=8.5 mm.

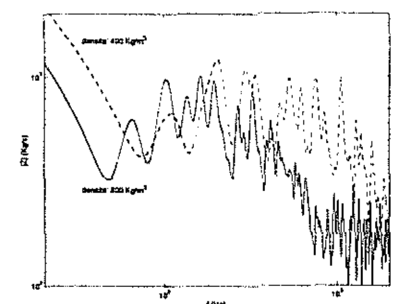


Figure 5: Magnitude of the mechanical impedence measured for two different values of density of a grand piano soundboard: ρ = 400 kg/m^3 and ρ = 800 kg/m^3. The different avarege behaviour in the range 500-1500 Hz is due to the different wavelength respect the distance between the ribs: if the wavelength is lower than such distance the stiffness effects of the ribs are less important.

G depends also on the position of the driving point: so it is important to know the behaviour of G in various points of the bridge, and our model gives the possibility for this kind of study.

In fig. 3 we report the results for the conductance for two values of the damping factor R. The damping factor is the most difficult quantity to estimate, but, as we can see, it does not change the average value of conductance and the positions of the resonances but only the quality factors of the resonances.

2.3. Variation of parameters

The model gives us the possibility to look at the changes produced by modifications of parameters such as thickness, density, number of ribs, and so on [1]. We report in fig. 4-5 two examples of this kind of study for a board shaped after a grand piano soundboard (Steinway D-274), with the right orientation of the wood-grain for the case at hand (around 45 degrees - see fig. 7 and remember that the grain is along x axis).

3. GRAND PIANO SOUNDBOARD MODES

To test the validity of our model we searched the first modes for the soundboard of the grand piano Steinway D-274. First of all we improved our model changing the boundary conditions from clamped ends to around "semi-hinged ends", to better simulate the real situation [7]. Then we fixed the physical parameters on standard values [8] and from the plot of the magnitude of the mobility function (fig. 6) we found the positions of the resonances. Afterwards, with a sinusoidal driving force at the frequencies of the resonances, we could find the displacement of our board in a single mode. In fig. 7-10 we can compare our results to experimental measurements (Chladni figures)[8]: they are very similar both quantitatively and qualitatively.

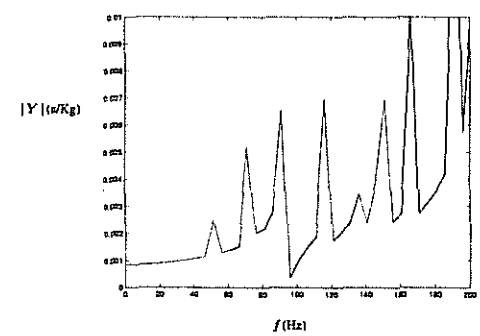


Figure 6: Mobility for the grand piano measured in (20;78), with semi-hinged ends and 17 ribs large Δx=2.45 cm; standard parameters [8]: ρ=418 Kg/m^3, h=8.5 mm, E_x = 1.2×10^{10}N/m^2, E_y = 2.7×10^{10}N/m^2, $G = 1.85 \times 10^8$N/m^2, $E_{cat} = 2 \times 10^{10}$N/m^2 and h_{cat}=10 mm; R=10 kg/(s$\times$ m^2) (arbitrary).

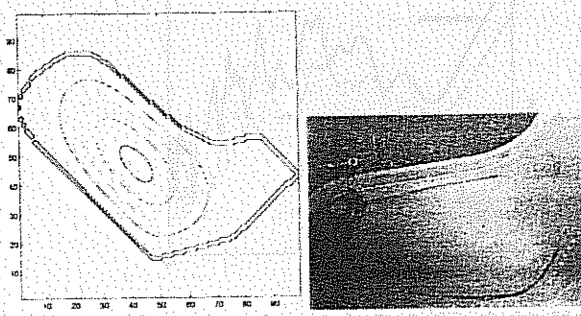


Figure 7: First mode at 49 Hz - 49 Hz exp. [8]

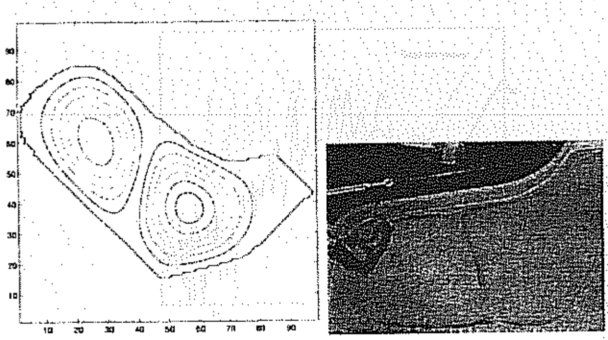


Figure 8: Second mode at 67 Hz - 66.7 Hz exp. [8]

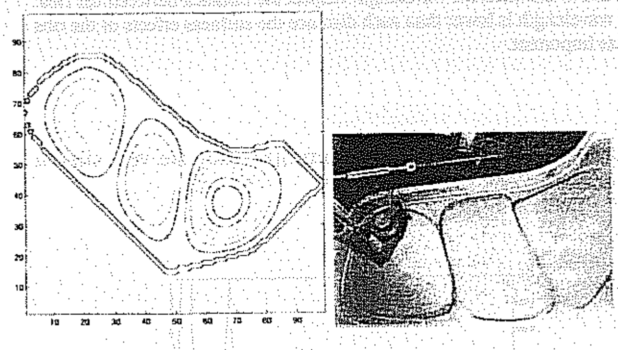


Figure 9: Third mode at 90 Hz - 89.4 Hz exp. [8]

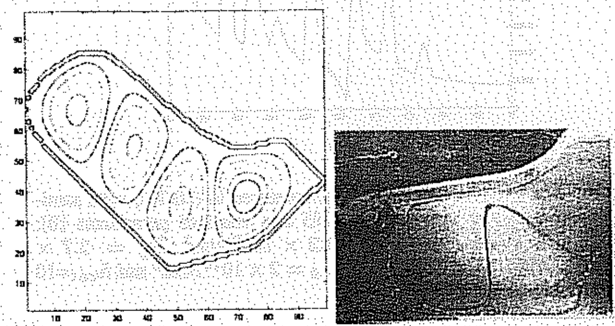


Figure 10: Fourth mode at 114 Hz - 112.8 Hz exp. [8]

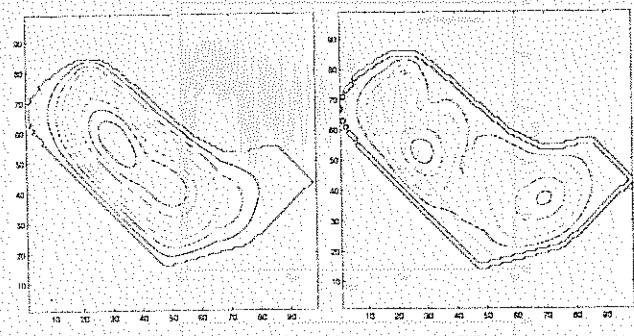


Figure 11: Sinusoidal driving force at frequencies: 60 Hz and 100 Hz, not corresponding to particular modes.

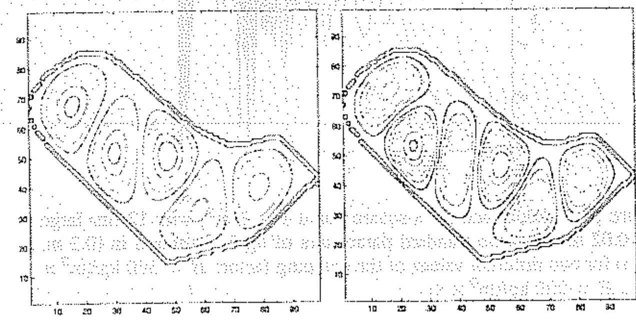


Figure 12: Modes at 150 Hz and 190 Hz

In fig.11 we report what happens if the frequency of the sinusoidal driving force is not corresponding to a particular mode. In fig.12 we report examples of higher modes.

4. REFERENCES

[1] B. Bazzi, "Studio numerico delle proprietà acustiche di una tavola armonica", *Tesi di Laurea*, Università di Modena e Reggio Emilia, 1999.

[2] N. Giordano, "Simple model of a piano soundboard", J. Acoust. Soc. Am. **102,** 1159-1168 (1997).

[3] W. H. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical recipes in C: the art of scientific computing*, Cambridge University Press, 1988.

[4] J. C. Strikwerda, *Finite-difference schemes and Partial Differential Equations*, Chapman & Hall, New York, 1989.

[5] K. Wogram, "The strings and the soundboard", *The acoustics of the piano* Royal Swedish Academy of Music Publication No.64, Stockholm, 1990, p. 83.

[6] T. D. Rossing, N. H. Fletcher, *Principles of Vibration and Sound* , Springer-Verlag, New York, 1995.

[7] I. Nakamura, "Vibrational and acoustic characteristics of soundboard (Acoustical research on the piano, Part 3)", J. Acoust. Soc. Jpn. **14,** 429-439 (1993).

[8] H. A. Conklin, Jr., "Design and tone in the mechanoacoustic piano. Part II. Piano structure", J. Acoust. Soc. Am. **100,** 695-708 (1996).

Sounds Modeling by means of Harmonic-Band Wavelets: New Results and Experiments

Pietro Polotti, Gianpaolo Evangelista

Laboratoire de Communications Audiovisuelles (LCAV)
École Polytechnique Fédérale de Lausanne, Switzerland
pietro.polotti@epfl.ch *gianpaolo.evangelista@epfl.ch*

Abstract

Musical signals contain both deterministic and stochastic components. The deterministic part provides the pitch and the global timbre of a sound; it is in a sense the fundamental structure of the sound. The stochastic part contains the "life of a sound", that is all the micro-fluctuations with respect to an electronic-like/non-evolving sound and noises due to the physical excitation system. These micro-fluctuations with respect to a pure harmonic behavior can be reconstructed from the power spectrum. A well-suited analysis and resynthesis tool of voiced sound spectra was introduced in [5] and [6], i.e., the Harmonic-Band Wavelet Transforms (HBWT).
The most attractive feature of the HBWT model is that resynthesis coefficients can be substituted in first approximation by white noise with proper scale-dependent energies [4]. At a more refined level one must take into account the little but non-zero correlation of the HBWT analysis coefficients of voiced sounds. This requires a pre-filtering of the resynthesis coefficients by means of AR filters. Furthermore an elementary waveform (wfs) model is employed for modeling the physical excitation system.
This method can be seen both as a musical tool for sound synthesis able to provide synthetic sounds with a natural timbre dynamic and as a compression technique.

1. Introduction

One of the most challenging aspects of sound analysis and representation is the definition of a good model for the noisy part of sounds. In other words we need a good representation of those components of sound whose spectra lie out of the frequency support of the partials.
In [4] we proposed a model for the particular case of voiced sounds, i.e., sounds with a harmonic spectrum, based on the Harmonic-Band Wavelet Transforms (HBWT). Thanks to the mathematical properties of the HBWT, the synthesis of signals with pseudo-periodic $1/f$ –like power spectra is straightforward and these spectra are very good approximations of those of real-life voiced sounds. In that model the only thing we needed was to control the energies of white noise coefficients, according to very few parameters derived from the analysis of real sounds.
In a more detailed perspective, the HWBT analysis reveals the existence of a little but not zero correlation between the coefficients. An AR analysis and resynthesis model, employing white noise as AR filters excitation and reproducing the above-mentioned loose correlation can substitute the trivial white noise coefficient model. As a further refinement of the technique we take into account scale-dependent time evolution of the resynthesis parameters. Also, our model cannot reproduce the excitation system of some instruments. In these cases we consider a second type of noise model, based on elementary waveforms (wfs) techniques [8].
The model does not fit the harmonic components. We preserve the restricted set of analysis wavelet coefficients corresponding to the narrow bands of the harmonics in order to have perfect reconstruction data for the deterministic part, i.e., the harmonics amplitude and their time envelopes. The attacks and decays are preserved as well.
The experimental results concerning all these refinements of the HBWT analysis and synthesis technique are the subject of this paper. In section 2 we make a short review of ordinary wavelets and Harmonic-Band Wavelets. In section 3 we review the pseudo-periodic $1/f$ model as developed in [4]. In section 4 and 5 we present the new developments of the synthesis method from the methodological and experimental points of view respectively. In section 6 we draw our conclusions.

2. Wavelets and Harmonic-Band Wavelets: a review

The wavelet transform provides a graded time-frequency representation of digital signals. In the particular case of audio signals, wavelets perform a time and frequency domain subdivision imitating the human perceptive system, i.e. a logarithmic tiling of the time-frequency plane (see Fig. 1-a). In other words we have a more detailed information in the low frequency area but a coarser sampling rate while a higher sampling rate but a coarser frequency resolution in the high frequency area.

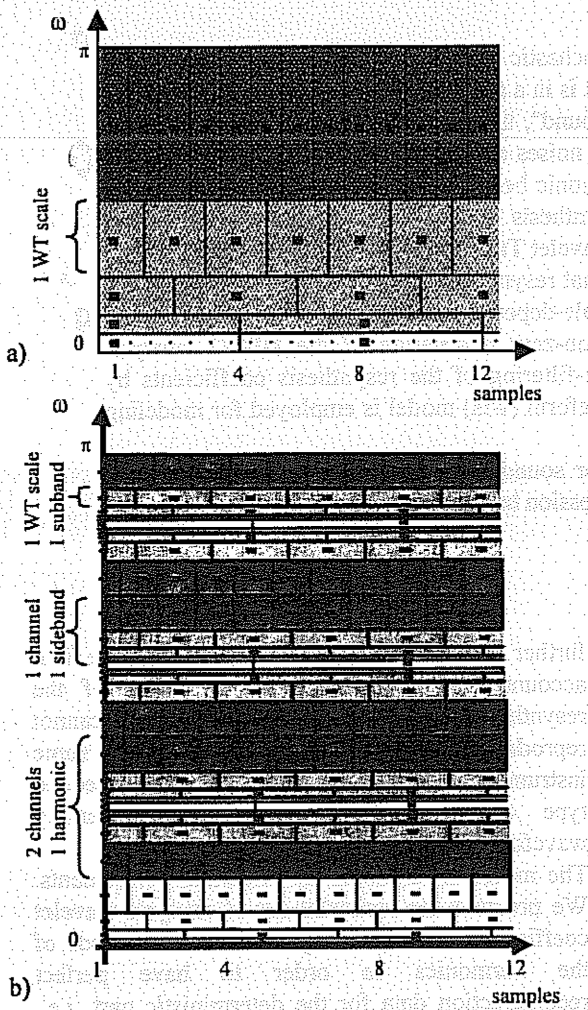


Fig. 1 Time-Fequency Plane Tesselation:
a) Ordinary Wavelets. Each dot corresponds to a wavelet coefficient. At each scale the sampling rate is divided by two.
b) HBWT. The tesselation is a frequency periodized version of a. The number of periods corresponds to the number of BP filters (that is, the number of channels) trapping each sideband of the harmonics.

The HBWT realizes a periodic version of the frequency domain subdivision of ordinary wavelets (see Fig. 1-b). This is obtained by means of the modulation and demodulation scheme described in [6]. With respect to ordinary wavelets, the HBWT provides a much more meaningful representation of voiced sounds. We can tune the frequency domain subdivision to the pitch of any given voiced sound, changing the number of channels of the HBWT filter bank (see [4]). In the ordinary wavelet representation the higher scales (corresponding to the low frequencies) represent the slow changes of a signal with respect to the "average" of the signal (0 in the case of an audio signal), i.e., with respect to a constant. On the other hand, lower scales (high frequencies) represent the changes with respect to the local mean at different rates. In the HBWT representation of voiced sounds the "local mean" is the average period, while the lower scales (the bands away from the harmonics) contain the information concerning the fluctuations with respect to the average period at different rates. In this way we are able to separate the harmonic part of voiced sounds from the different noisy components containing the "dynamics" of the sound.

3. The Pseudo-Periodic $1/f$–like Model

At a previous CIM meeting we presented a paper about the pseudo-periodic $1/f$-like noise analysis and synthesis [4]. The starting point was the experimental evidence, revealing an approximate pseudo-periodic $1/f$ behavior of the spectra of voiced sounds in music. The leading idea was to adapt the spectrum of a synthesized pseudo-periodic $1/f$-like signal to that of a real life sound (see Fig. 2 and Fig. 3). The synthesis process is controlled by means of a very restricted set of parameters, defining the $1/f$ shape of each harmonic of the synthetic spectrum. The analysis tool necessary for the extraction of the resynthesis parameters as well as the synthesis tool is provided by a HBWT filter bank and its inverse, respectively. This model has the advantage to be extremely concise. The lower limit of one parameter per channel of the analysis and synthesis scheme introduced in [4] is an extremely good result from the point of view of data compression. Actually some refinements are necessary in order to reach a good quality in sound reproduction at the cost of an increase of the number of parameters.

4. The new Analysis and Synthesis Method

The new method we are going to describe in detail is a refinement of the pseudo-periodic $1/f$-like spectral model [4]. The whole model is limited to the steady part of sound and is particularly well suited for long, sustained sounds. In our analysis and synthesis scheme the transients, that is the sound attack and decay, are just reconstructed from the complete set of HBWT coefficients resulting from their HBWT decomposition. A pitch detector based on the estimation of the autocorrelation is adopted in order to define the extension of the transient. Only the portion of sound where a steady pitch is detected is processed by means of the HBWT filter banks.

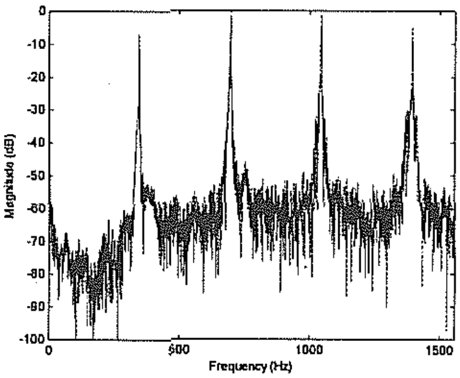
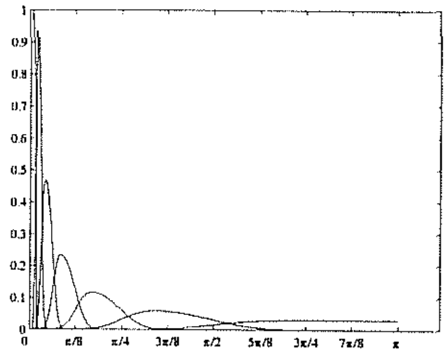


Fig. 2 Spectrum of a real-life trumpet

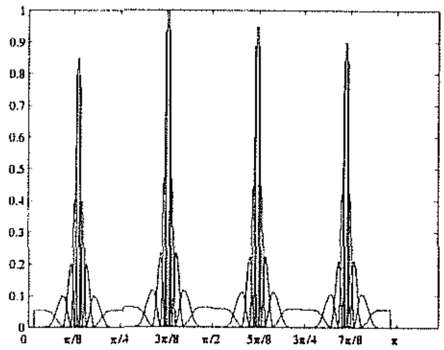
We subdivide the remaining portion of the time-frequency domain in different parts The first part corresponds to the deterministic components of the sound (see Fig. 1-b the white subbands). The analysis HBWT coefficients are preserved in order to obtain a perfect reconstruction of the harmonics and their time-envelopes. The spectrum portion close to the harmonics contains the micro-fluctuations with respect to pure periodicity (see Fig. 1-b the light gray subbands). Its behavior is approximately $1/f$. We estimate independent parameters controlling the energy of each HBWT subband separately. This allows us to find a better approximation of the spectrum. An even better spectral "design" can be obtained by means of a convenient and flexible analysis and resynthesis tool, recently introduced by one of the authors, i.e., the Arbitrary Bandwidth Wavelet Transforms [9]. These new wavelet transforms allow arbitrarily detailed, signal adaptable spectra subdivisions. This is necessary when the spectrum shape is not $1/f$–like. We employ a mean square error criterion in order to find the more suitable piecewise approximation of the spectrum. Furthermore, an LPC analysis is applied to the HBWT analysis coefficients. The AR filters so obtained are used to color the white noise used as input to the resynthesis filter bank, in order to reproduce the time-correlation in the subbands.

The third spectrum portion includes the first subbands of the HBWT decomposition (see Fig. 1-b the dark gray subbands). According to the analysis results is possible to see how these subbands, which lie far away from the harmonics, contain the most significant and non-masked information concerning the additional noise due to the excitation systems.

a)



b)



c)

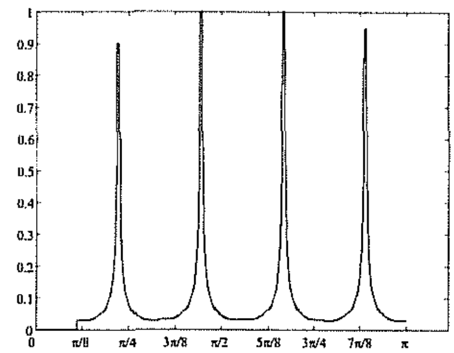


Fig. 3 1/f spectral model:

a) Ordinary Wavelets. The wavelet subbands with proper energies reproduce a $1/f$ spectral behavior.

b) HBWT. By means of a frequency periodized version of a) we are able to reproduce a pseudo-periodic $1/f$ -like spectral behavior. Each sideband has a $1/f$ behavior and is subdivided in wavelet subbands.

c) The resulting synthetic spectrum. This spectrum can be adapted to the spectrum of the sound we want to resynthesize. In this way both the harmonic components and the noisy components of sounds are reproduced.

In these frequency subdomains we look for the samples of elementary waveforms (wfs) which occur during the instrument continuos excitation. These noises include, for instance, saliva gurgling in wind instruments or bow noise in string instruments. These wfs are juxtaposed according to a statistical sampling and amplitude scaled according to the signal analysis itself. A more refined method should include the construction of a sort of codebook, i.e., a sample-case of elementary waveforms for each musical instrument.

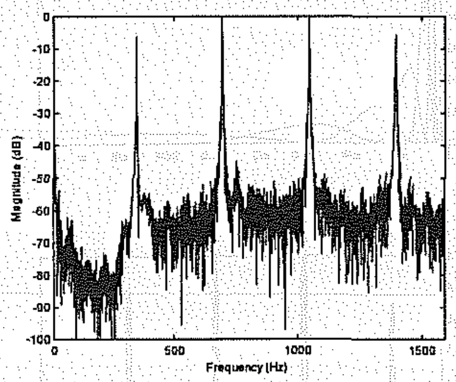


Fig. 4 Spectrum of a synthetic trumpet.

5. Experimental Results

The experimental results change significantly according to the instrument we analyze and resynthesize. The first step is the choice of the wavelet scale at which we stop the analysis. This defines which ratio of the spectrum we resynthesize by means of noisy coefficients and which ratio is perfectly reconstructed preserving the analysis coefficients. Normally 4 or 5 scales are the maximum values admitted in order to preserve the *main time characteristic* of the sound as the harmonics and their time envelope. The second step is to define the extension of the transients; the length of the attack and of the decay varies a lot according to the instrument, the pitch and the stabilization speed of the sound. Our algorithm gives good results for all the sounds we considered: a flute, an oboe, a clarinet, a bassoon, a trumpet, a french horn and a trombone. The AR filters employed are of the 10^{th} order for the second subband. The order diminishes with the order of the subbands. We performed a perfect reconstruction of each subband separately in order to compare them with the synthetic ones. The results from an acoustical point of view are very good.
A short-time version was also implemented, applying very short rectangular windows (20 coefficients) to the analysis coefficients. In this way we preserve the energy time envelopes of the subbands.
Finally we considered a dozen of elementary waveforms, extracted from the perfect reconstructed first subband, in order to synthesize the first subband. We mounted them in random order with amplitude following the short time energy analysis results.

Conclusions

We introduced an articulated method to deal with the different components of voiced sounds, focusing our attention on their noisy components, so important to maintain a real-life "color" in sounds. We think that our method provides a convincing noise model for voiced sounds in music, with a solid mathematical background [6] and very good experimental results.
The main improvement of the method, on which we *will work in the immediate future*, will be a pitch synchronous version, i.e., a time varying version of the filter banks. This would free the method from the limitations of a fixed number of channels, which restrict the set of sounds that we can analyze to those with a very well defined and stable pitch.

References

[1] G. W. Wornell, "Wavelet-Based *Representations for the 1/f Family of Fractal* Processes", Proc. IEEE, Vol. 81, No. 10, pp. 1428-1450, Oct. 1993.

[2] I. Daubechies, Ten Lectures on Wavelets, SIAM CBMS series, Apr. 1992.

[3] G. Evangelista, "Comb and Multiplexed Wavelet Transforms and Their Applications to Signal Processing", IEEE Trans. on Signal Proc., vol. 42, no. 2, pp. 292-303, Feb. 1994.

[4] G. Evangelista, P. Polotti, "Analysis and synthesis of pseudoperiodic 1/f-like noise by means of multiband wavelets", Proc. of XII CIM, pp.35-38, Sept. 1998.

[5] P. Polotti, G. Evangelista, "Dynamic Models of Pseudo-Periodicity", Proceedings of the 99 Digital Audio Effects Workshop, pp. 147-150, Trondheim, Dec. 1999.

[6] P. Polotti, G. Evangelista, "Analysis and Synthesis of Pseudo-Periodic 1/f-like Noise by means of Wavelets with Applications to Digital Audio", to appear in Applied Signal *Processing. The International Journal of* Analog and Digital Signal Proc., Springer-Verlag, Berlin, special issue, Dec. 2000.

[7] P. Polotti, G. Evangelista, "Time-Spectral Modeling of Sounds by Means of Harmonic-Band Wavelets", Proceedings of the ICMC 2000, Aug. 2000.

[8] C. d'Alessandro, G. Richard, "Random Wavelet Representation of Unvoiced Speech"

[9] G. Evangelista, S. Cavaliere, "Frequency-Warped filter Banks and Wavelet Transforms: A Discrete Time Approach via Laguerre Expansion", IEEE Transaction on Signal Processing, Vol. 46, No 10, Oct. 1998.

MUSICAL INFORMATICS: EXPRESSION AND PERFORMANCE ANALYSIS 1

MUSICAL INFORMATICS:
EXPRESSION AND PERFORMANCE
ANALYSIS I

Modelling Emotional Narrative Through System Dynamics: Applications To Computer Music.

Ian Whalley. musik@waikato.ac.nz
Music Department, The University of Waikato, Hamilton, New Zealand.

Abstract
Everyday narratives can be modelled and simulated using system dynamics tools, providing the basis of an emotional/ thematic approach to computer music composition. The same tools can then be used to model and simulate the interplay between musical thematic manipulation and structure. The generative compositional system is illustrated based on one narrative example.

Introduction
This paper stems from a concern about what appears to be a diminishing sense of music as a *dramatic art* in much current computer music. A grounding assumption is that the creation and control of dramatic tension through thematic juxtaposition is a worthwhile artistic goal in narrative art forms [1].

For the sake of example, computer music is narrowly defined here as being computer generated and intended for reproduction rather than performance. Characteristics include the abandonment of tonality and acoustic instruments as the basis for musical language. The emphasis is on timbral transformation and microtonal textures [2]. The compositional approach outlined in the paper could also be used to generate music based on other styles, or combinations of styles.

Recent work has dealt with applying system dynamics modelling to *structure* in computer music [3]. This work is summarised in the text when looking at the structural aspects of music. Milicevic [4] argues that the ideal reception of music is neither structural nor metaphorical, but lies in the dialectic between the two. The approach presented here extends my recent work by reflecting this view [5].

System dynamics
System dynamics thinking involves the application of non-linear models to simulate complex narratives [6]. The techniques have been popularised through commercial modelling packages such as *Vensim*, *Powersim*, and *Stella*. Simulation is based on influence diagrams, stocks, flows, and feedback loops. The application to music is limited to date [7].

System dynamics models rely on a three step process: drawing the structure of the narrative or situation being examined; making assumptions about the nature of the relationships between parts of the structure; and running the model in compressed time to illustrate the interaction of various parts. These tools can check assumptions about the situation being studied and run 'what if' scenarios to test different approaches to problems. As such, they provide a means to understand the dynamic experience of computer music.

In the structure/relationship model (figure 1) built using *Stella* software, the double lines with an arrow represent the direction of flows of information (like verbs). A box is an amount or stock of something (like a noun). The single arrow lines illustrate feed back or influence connections. The 'clouds' indicate zero activity. Relationships between parts are influenced by adding either graphs or formulas into the circles in the diagram to reflect assumptions made. This takes place 'behind the scenes' in the software by entering a lower level.

Once assumptions are added, the narrative can then be simulated in any timeframe. Output is graphed based on any part of the model selected. This process allows the user to trial several strategies by altering the driving formulas. Graphed outputs provide the results of the interplay between parts, or the behaviour of a single part.

In figure 1, the example narrative being modelled is a variation on a drifting goals archetype. This is encountered in everyday life where there is a need to move from a current situation toward a vision of a new situation. The contrast between the two states creates tension that can either force one to lower the vision, do something innovative to change the current situation, or fragment the tension through avoidance, willpower or devoting energy to other tasks (urgency).

The narrative here can have many outcomes. For example, in a positive outcome, the initial vision may be lowered slightly and there will be a period of turbulence due to tension fragmentation before the current situation rises to the level of the vision. Of course, the model can be tweaked any number of ways to replicate lesser outcomes or disasters.

System dynamics and music
Linking system dynamics modelling to music composition requires the modelling of the structural aspects of music (tension and relaxation) and the thematic/emotional drivers that provide the input to this structure. I will deal with the structural aspects first [8].

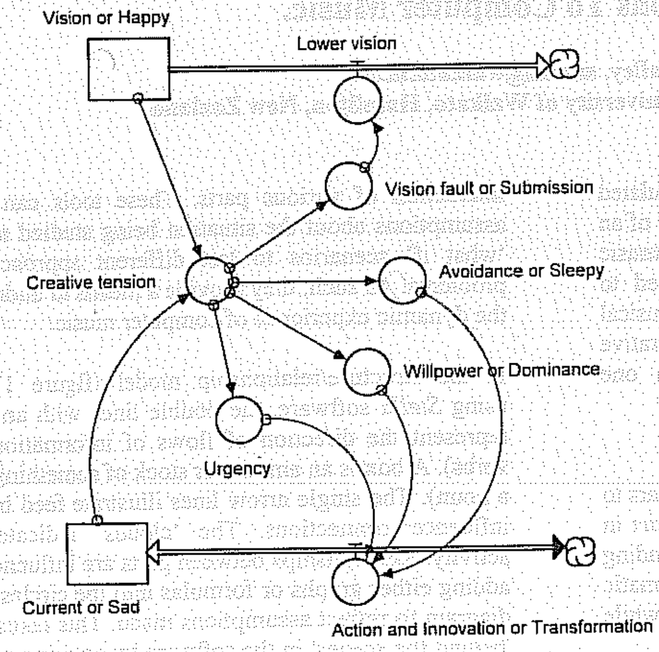


Figure I

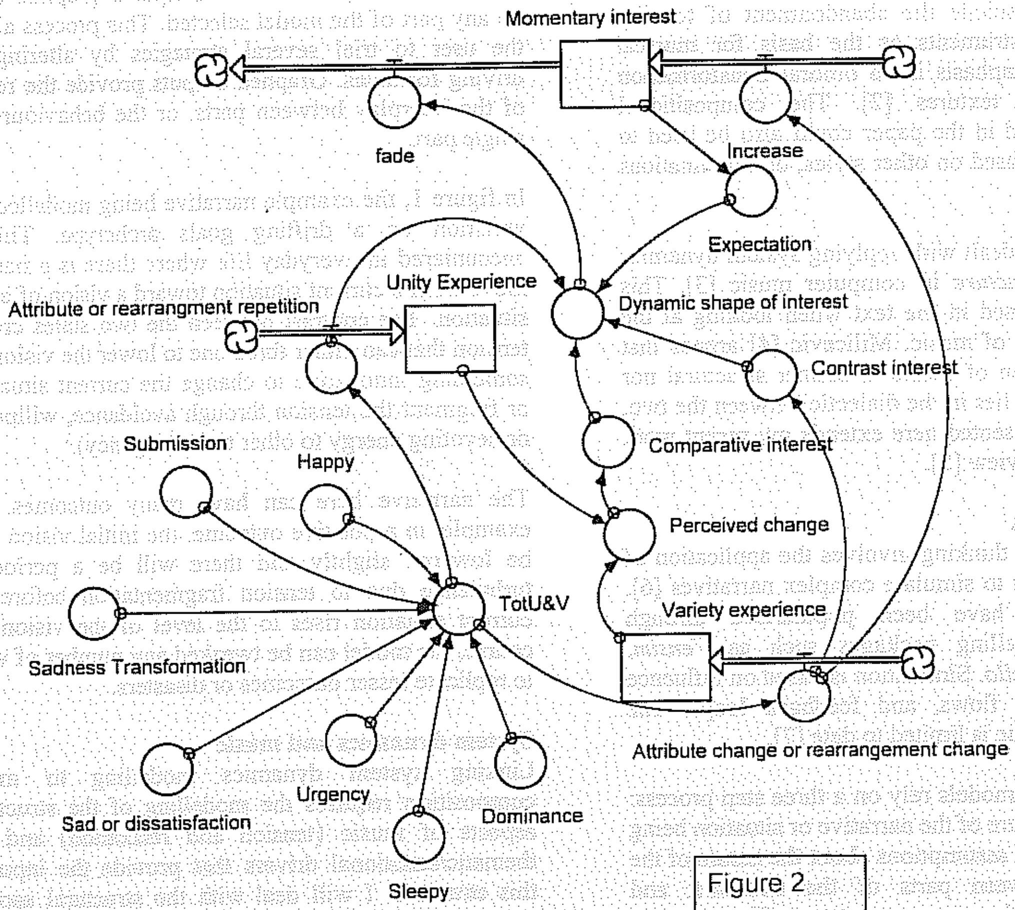


Figure 2

Music is a time based art form grounded in the experiential flow of events and relationships similar to other narrative art forms such as plays. Like other narrative art forms, it has a *dynamic shape of interest*. The *dynamic shape of interest* in most classical and popular drama, large-scale western art music, and popular song has a main climax about two-thirds to three-quarters the way through the work. This is played out within the framework of introduction, complication, climax, and resolution. Abstract music as narrative does not achieve dynamic interest through verbal/visual discourse, but through aural/physical discourse. Both approaches are kinetic but differ in that stories deal with the explicit and music is the province of the implicit.

The control of the structural dynamic of a musical work based on the manipulation of *unity and variety* of aural thematic material is a cornerstone of the development of conventional composition technique in music. Without a systematic and explicit way of illustrating this in real-time, acquiring this skill in computer music is usually by trial and error. This is not helped by the structural dynamic of computer music often being self-referencing rather than to an external referent such as tonality [9].

Since it is a *compositional* aid, the model (figure 2) is based on the *composer's* perception of how they would *hear* the work. It maps the *dynamic shape of interest* (right hand side) based on unity/variety manipulation (left hand side) as an abstract of the music.

The model is based on certain assumptions. First, that music as *dynamic structure* is a representation of a personal experience of time [10]; second, that this experience is only knowable through the *perception* of sound events [11]; and finally that the modelling of these events interactively in real-time will approximate the experience.

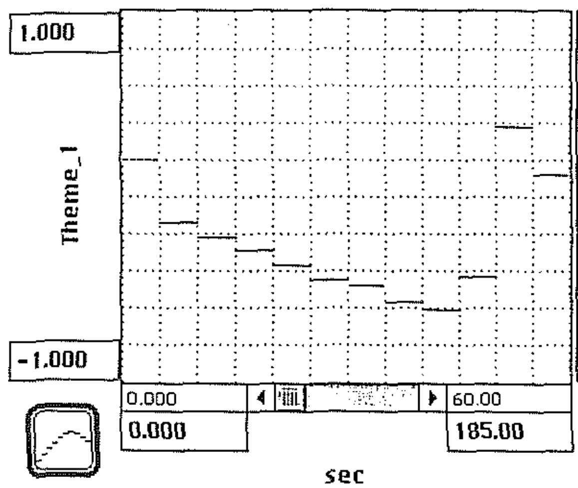
The structural outputs (from *TotU&V* or the total of unity and variety on figure 2) are based on Pressings notion [12] that there are two main types of change. '*Rearrangement* change occurs when time is expressed through motion or changes of position in unchanging sound objects. For example, through notes or recurring sound complexes. *Attribute* change occurs when time is expressed through alterations in the attributes of sound objects, such as parametric alterations in the qualities of a sound'. The 'momentary interest' stock on the diagram incorporates the novelty interest factor.

The *dynamic shape of interest* (figure 2) is the end for the structural aspects of the flow of events in the model, and the assumption is a perfect correlation between this and the contour of *tension and relaxation* in the piece. A limitation in the model is assuming that all listeners will be as fully cognisant and musically attentive and as informed as the composer; an assumption many composers have made historically.

We now have to get information into the model through thematic/emotional drivers that provide the information to simulate the shape of interest (figure 2, inputs to *TotU&V*).

Entering and abstract of the music into the model relies on composers being able to auralise individual thematic elements of the composition in the first instance as they would do traditionally (or enter them in when a composition is completed). A parallel is an author being able to map a story imaginatively before setting words to paper. On the model (figure 2), this is done by entering information into the thematic inputs to *TotU&V*. Graphs are drawn for each theme so that each timeframe is decided based on what musical information proceeded it: whether information is the same or different to what has happened before, and to what degree (see example in figure 3). Inputs to *TotU&V* (happy etc.) then indicate a contribution to either unity or variety at any given interval through being either a negative or positive figure, with zero allowing for no contribution (no theme present).

Figure 3



To this point, the approach relies on composers taking a *thematic approach* to generating computer music, yet makes no assumptions about what themes may be made up of in terms of auditory material, or how they may relate to each other. Understanding this requires the linking of theme with emotion, and the original narrative (figure 1) to an auralisation of it. i.e. the structural and metaphorical aspect of music needs linking in the model.

As an implicit art form, music is primarily sensual rather than conceptual. To bridge this gap, conceptual notions must be translated into aural experience. This is made possible through the coding of emotion in the manipulation of musical thematic material.

The idea of composers' personalising general emotional/semiotic intention/response is well established in film music theory and practice [13], with innovative outcomes illustrated in the work of many experimental film music composers. The approach allows creativity within a broadly agreed framework of semiotic meaning with an *intended* audience [14][15] at a specific time and in a particular context.

Provided the music that has to be written is a 'stand alone' piece, a significant difference between writing for film and using system dynamics modelling as a means to generate scores, is that the composer selects the structure of the narrative and controls the dynamic shape. In this sense, unlike film music, one becomes author, director, and composer.

Linking figures 1 and 2 illustrates the process. *Figure 1* provides a narrative that can be manipulated by the composer to be a generator. In the diagram, parallel emotional states are attached to each part of the diagram (avoidance with sleepy, vision fault with submission etc). This is a routine part of film music compositional practice [16]. It allows a translation from the ideational to emotional/thematic process, and provides the thematic inputs to figure 2. The composer can then auralise how the thematic material may be used in musical terms.

The way that the themes respond over time is dependent on their behaviour in the simulation in figure 1. For example, it may be that in the narrative, avoidance (sleepy) is a strong initial response, which would result in this theme being introduced and manipulated early, and it then fading out.

A difference between figure 1 and 2 is that in figure 1 the happiness/sadness themes have to be first introduced as dynamic sensual ideas, since in figure 1 they are portrayed as static and conceptual. A parallel is in a novel when a 'current situation' and 'ideal situation' have to be related through linear dialogue or monologue rather than a picture.

A successful musical outcome based on this method relies on using emotional states that can be juxtaposed in some way during the work [17]: an essential element of most Western dramatic art forms, such as the protagonist and antagonist in a play. This requires contrasting two or more poles based on the primary emotions: aroused/interested, sleepy, sad, happy, fear/submission, and dominance/anger [18][19], although contrasts need not be extreme. The interplay of these emotional states in new art works then allows computer music to be reconnected to the dramatic narrative tradition and human emotional experiences from which it seems increasingly dislocated. The narrative here is one of many possibilities.

The approach here is a tool, not an end. It affords experimentation with different approaches as an aid to structural and semiotic expression once one has some musical/thematic ideas. In contrast to craft, a limitation of the approach presented here artistic. To use an analogy from literature, there are many literary critics with a good understanding of structure but who are unable to write an engaging novel.

References

[1] Whalley, I. (1999b) 'Thematic Juxtaposition and Resolution in Computer music: A System Dynamics Model'. *1st Iteration Conference in Generative Systems in the Electronic Arts*, pp. 158-174. Monash University, Australia.

[2] Smalley, D. (1997) 'Spectromorphology: Explaining Sound-shapes'. *Organised Sound 2(2):107-26.*

[3] Whalley, I. (1999a) "Beyond Pitch/Duration Scoring: Towards A System Dynamics Model of Electroacoustic Music." *17th International Conference of the System Dynamics Society*, pp 73-83. Wellington, New Zealand.

[4] Milicevic, M. (1998) 'Deconstructing Musical Structure'. *Organised Sound*, 3(1) 27-34.

[5] Whalley, I. (2000) ' Emotion, Theme and Structure: Enhancing Computer Music Through System Dynamics Modelling'. *ICMC Proceedings*, Berlin.

[6] Forrester, J. (1961) *Industrial Dynamics.* Productivity Press.

[7] Summary from Whalley (2000) Ibid.

[8] See Whalley (*Ibid.*) (1999a).

[9] Rudy, P. (1999) Music technology: two states. *Sound Ideas.* 2(3), 16-20.

[10] Imberty, M. (1993) 'The stylistic perception of a musical work: an experimental and anthropological approach'. *Contemporary Music Review.* 7: 33-48.

[11] Kramar, J. (1988). *The Time of Music.* New York: Schirmer.

[12] Pressing, J. (1993) 'Relations between musical and scientific properties of time'. *Contemporary Music Review. 7:105-122.*

[13] Gorbman, C. (1974) *Unheard melodies: Narrative Film Music.* New York: Garden City.

[14] Ibid 1974:71

[15] Frith, S. (1988) *Music For Pleasure: Essays in the Sociology of Pop.* Oxford: Polity Press p. 136

[16] Hagen, E. (1972) *Scoring for Films.* New York: Wehman.

[17] Whalley. Ibid. 1996b

[18] Descartes, R. (1649) 'Treatise on the Passions of the Soul' (in *The Philosophical Works of Descartes*: E. Halden and G. Ross (Trans) Cambridge University Press, 1967.

[19] Schubert, E. (1999) 'Measuring Emotion Continuously: Validity and Reliablily of the Two-dimensional Emotion Space' *Australian Journal of Psychology.* 51(3):154-165.

Analysis of the influence of expressive intention in piano performance of classical music

S. Canazza – G. D'Arduini – A. Rodà

Music performance is often viewed as part of a system of communication whose leading actors are the composer, the performer and the listener. Ambiguities in musical notation allow a player considerable freedom in deciding how to interpret the music content. Musicians modify lightly (introducing microvariations) the nominal value of the acoustic parameters to communicate moods and feeling. This work compares the values of acoustic parameters note by note in a series of piano performances characterized from different expressive intentions; moreover, the same type of analysis has been carried out relatively to the pedal's use

1 Introduction

Many investigations explore how musical structure influence music performance [1]. There are a few studies on how musician's expressive intentions are reflected in the performance ([2], [3], [4]). In this context, expressive intention is taken to mean how a musician's inspiration varies according to certain adjectives that have been given before each performance.
This area of investigation is stimulating ever greater interest not only from the scientific and cognitive point of view, but also from the applied one, both in terms of automatic music performance [5] and, more generally, in multimedia systems. Moreover, the study of musical performance and interpretation raises problems that go beyond the specific area of musicology, running into the wider field of non-verbal expressive communication involving various aspects of human perception.

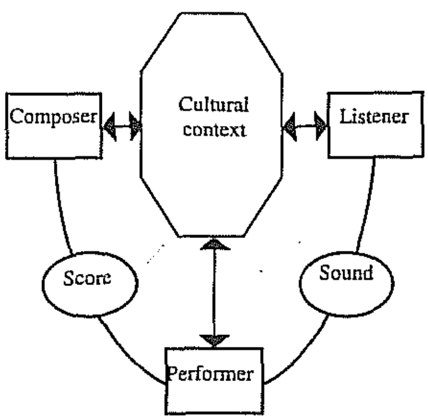


Fig. 1: Music as system of communication from composer to listener

Music performance is often viewed as part of a system of communication whose leading actors are the composer, the performer and the listener (fig. 1).
The performer uses the expressive means of his own musical instrument to communicate to the listener the contents of the music. Ambiguities in musical notation allow a player considerable freedom in deciding how to interpret the music content. Thus, the same musical score is performed differently by different performers, or the same performer may perform a piece differently in separate occasions.
Many studies about musical performance have demonstrated that musicians use small variations to communicate those aspects of the musical structure they wish the listener to perceive. These variations refer to deviations in timing, articulation, amplitude, *etc.* in relation to a literal interpretation of the score. Each expressive deviation introduced by the performer can be brought back to the expressive motivation that caused it. The first source of motivation deals with musical structural aspects such as phrasing, hierarchical structure of the phrase, harmonic structure and so on. The second involves those aspects that are indicated with the term expressive intention and which refer to the communication of moods and feeling.
Paying attention to the second source of expressive deviations, this work compares the values of acoustic parameters note by note in a series of piano performances; the aim of this work is to convalidate the preceding results of a strong correlation between the variations of these parameters and the musician's expressive intention [6]. Sound synthesis obtained by the model are used to validate the results of analysis following analysis by synthesis method.
Moreover, a further analysis concerning the use of the piano's pedals and its correlation with the player's expressive intention has been carried out.

2 Description of the experiment

Five professional pianists were asked to perform an excerpt (fig. 2) from Wolfang Amadeus Mozart's Sonata KV545, II movimento *Andante* [7].

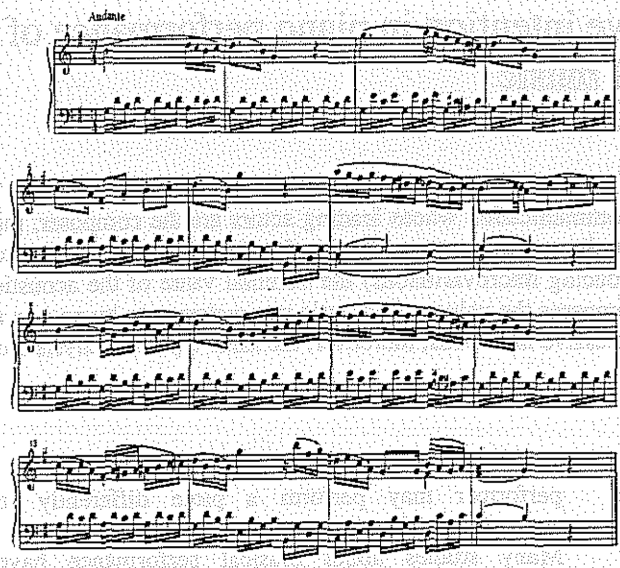


Fig. 2: excerpt from Wolfang Amadeus Mozart's Sonata KV545, II movimento Andante

The pianists have played the excerpt nine times, inspired by the following (Italian) adjectives: normal (*normale*), bright (*brillante*), dark (*cupo*), light (*leggero*), heavy (*pesante*), hard (*duro*), soft (*morbido*), passionate (*appassionato*), flat (*piatto*). The "normal" adjective has been used to mean a musically correct execution, but without any expressive intention; the "flat" adjective has been used to mean a performance as much as possible near to the score, even to the prejudice of musical correctness.
The pianists could study the excerpt for a week and could repeat every execution for three times; then they have chose, among these three interpretation, the one they judged more characterized.
The executions were played on a piano *disklavier* Yamaha. This piano has sensors which point out three categories of movement of mechanics of the piano: hammers, keys and pedals.
The data pointed out from the sensors were recorded and transformed, by means of an external operative unit, into digitals signals by MIDI standard. The data were obtained by means of the software *Adagio*, compatible with the MIDI protocol.
The Adagio files are text files; they then were converted into Excel format. A set of macro has been applied to these files in Excel format, to make easier the statistic analysis of the data so obtained
Therefore forty-five performances, nine for each pianist, have been recorded in musical format MIDI; this format allows to easily determine the acoustic parameters of each note, like onset, duration (DR), inter-onset interval (IOI), key velocity.
In the original file format MIDI were also recorded the pedal events. Each pedal's pressure has been measured according to a scale varying from 0 (no pressure) to 127 (total pressure).

3 Analysis of timing and intensity

A mean-value analysis has been carried out for the parameters tempo, key velocity and legato.

3.1 Tempo

The tempo has been measured by means of beat/minute.
Since the last measure of the studied excerpt is characterized from a marked *Rallentando*, the calculation of the Tempo for every execution was carried out without considering this measure. Moreover, the calculation was carried out considering the events relevant to the left hand, for, in the considered score, the latter gives the rhythmic reference.

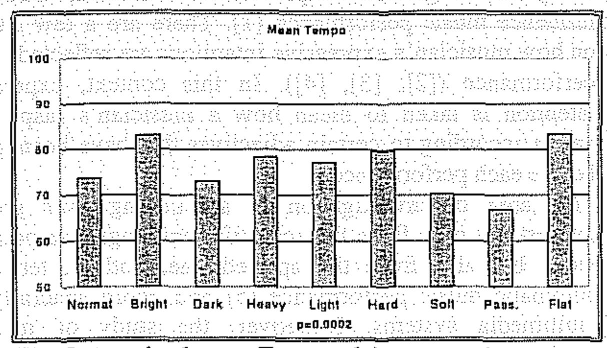


Fig. 3: graph of mean Tempo of the expressive intentions

The p parameter stands for the probability that the deviations of data are not correlated with expressive intention (or performer). The results of the statistic analyses have been considered reliable for $p < 0.05$.
The results of the analysis have demonstrated a correlation between the obtained values and the musician's expressive intention. The executions *flat* and *bright* have been shown the fastest tempo, the executions *passionate* and *soft* the slowest. No correlation with the executors has been carried out.

3.2 Key velocity

The key velocity has been measured according to a scale varying from 0 to 127. Separated analyses were carried out for the left and for the right hand.

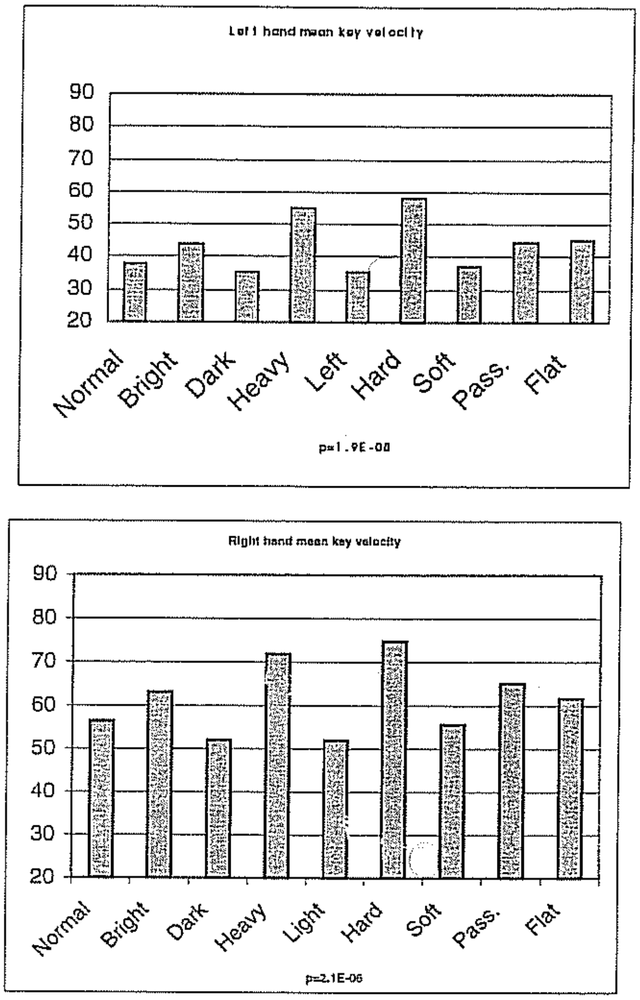


Fig. 4 – 5: graphs of mean key velocity of expressive intentions for the left and right hand

Once more the analysis has pointed out a correlation between the obtained values and the pianist's expressive intention. The performances *hard* and *heavy* have the highest values of key velocity, *light* and *dark* the lowest.
For the left and the right hand similar results have been founded. Moreover the key velocity of the right hand turned out to be lightly correlated with the excerpt's performer.

3.3 Legato

The parameter *Legato* has been defined as the ratio between the duration and the IOI (Inter Onset Interval) of the note. New separated analyses have been carried out for the left and for the right hand and again the analysis showed a correlation between the obtained values and the pianist's expressive intention.

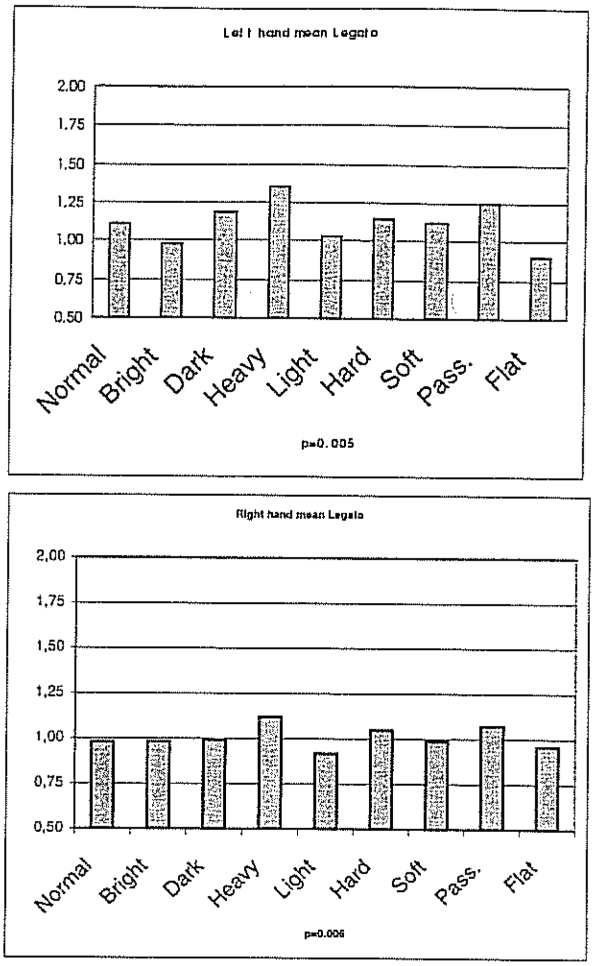


Fig. 6 – 7: graphs of mean Legato of expressive intentions for the left and right hand

The greatest value of legato has been measured in *heavy* and *passionate* performances, the smallest in *flat* and *light*. The legato values, both for the left and the right hand, turned out to be correlated also with the pianists.

4 Analysis of dumper

In the study of the pedal use (seldom considered in literature) the parameter "mean pedal" has been considered. Each variation of the pedal position is recorded with value in the scale 0 – 127; the parameter "mean pedal" stands for the mean value of pedal in the performance.
The Damper and the "Una Corda" have been separately analyzed. For both pedals the mean-value analysis has showed a correlation between their employ and the musician's expressive intention, and no correlation with the player.

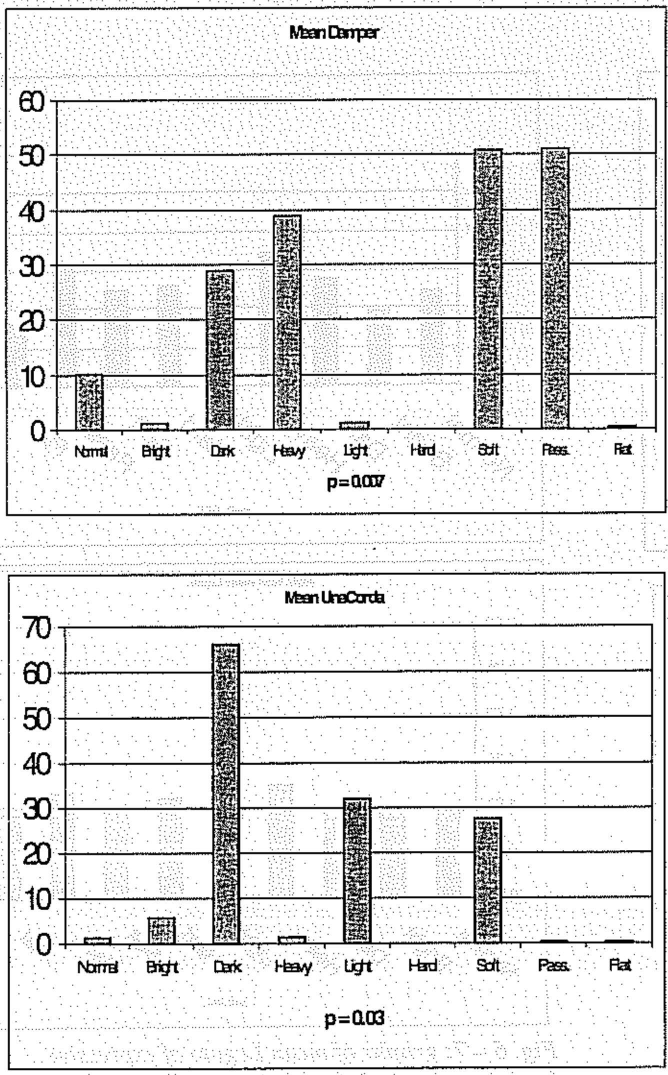


Fig. 8 – 9: graphs of mean pedal of expressive intentions for the Damper and UnaCorda pedals

The use of Damper shows that adjectives can be divided into two groups: *passionate*, *soft*, *heavy* and *dark*, characterized by a marked pedal employ, and *hard*, *flat*, *light* and *bright*, characterized by a very small pedal employ. The *normal* interpretation intermediate characteristics.
Also for the Una Corda pedal the expressive intentions have been divided into two groups: *dark*, *light* and *soft*, with a great use of this pedal, and the others, where this pedal is practically unused. No correlation between the use of pedal and performers has been carried out.

5 Conclusions

In this work we presented an analysis of the microdeviations (deivations from the nominal value of the acoustic parameters) introduced by the pianists in the musical interpretations.
The Tempo, the key velocity and the Legato are used by the musicians to communicate the wished expressive intention. For the key velocity and Legato, the analysis has been carried out separately for the left (bass) and the right (melody) hand.
Moreover, an analysis about the pedal's use has been carried out. Again, has been resulted a correlation between this parameter and the wished expressive intention.

References

[1] Repp, B. H. 1990. "Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists". *Journal of Acoustical Society of America*, 88, 622-641.

[2] Gabrielsson A. 1995. "Expressive intention and performance". R. Steinberg (Ed.), *Music and mind machine*. pp. 35-47. Heidelberg: Springer Verlag.

[3] Canazza S., Orio N. (1998). "How are expressive deviations related to musical instruments? Analysis of tenor sax and piano performances of 'How high the moon' theme". *Proc. of XII CIM*. Gorizia, 75-78.

[4] Canazza S., De Poli G., Rodà A., Vidolin A. 1997. "Analysis by synthesis of the expressive intentions in musical performance". *Proceedings of the ICMC 1997*, 113-120.

[5] Canazza S., De Poli G., Di Sanzo G., Vidolin A. (1998). "Adding expressiveness to automatic musical performance". *Proc. of XII CIM*. Gorizia, 71-74.

[6] De Poli G., Roda' A., Vidolin A., (1998). "Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance". *Journal of New Music Research*, 27, 3, pp. 293-321.

[7] Battel G.U., and Fimbianti R. 1997. "Analysis of expressive intentions in pianistic performances " *Proc. Int. Kansei Workshop*, Genova

MUSICAL INFORMATICS: EXPRESSION AND PERFORMANCE ANALYSIS II

Score Extraction from MIDI Files

Emilios Cambouropoulos

Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
emilios@ai.univie.ac.at

Abstract

In this paper a system that is designed to extract the musical score from a MIDI performance is described. The proposed system comprises of a number of modules that perform the following tasks: identification of elementary musical objects, calculation of accent (salience) of musical events, beat induction, beat tracking, onset quantisation, streaming, duration quantisation and pitch spelling. The system has been applied on 13 complete Mozart sonata performances giving very encouraging results.

1. Introduction

A system that attempts to extract the musical surface (i.e. a symbolic representation of notes in terms of quantised onsets and durations, and correctly spelled pitches) from a polyphonic MIDI performance is herein described. This system was developed as a means for obtaining the scores (in a symbolic machine-readable format) of a large number of performed piano works. In general, however, score-extraction techniques are indispensable for a plethora of applications that process performed MIDI input (e.g. music notation packages, interactive musical performance systems etc.).

The proposed system comprises of a number of modules that perform the following tasks: identification of elementary musical objects, calculation of accent (salience) of objects, beat induction, beat tracking, onset quantisation, streaming, duration quantisation and pitch spelling. The aim of this paper is to highlight a number of issues relating to score extraction and to give a quite broad understanding of problems relating to the development of and interaction among components that are necessary for score extraction tasks. Figure 1 gives an overall outline of the score extraction system.

The above system has been applied - to this date - to the midifiles of 13 complete sonatas by W.A.Mozart performed by a professional Viennese pianist. The system performs well in the above tasks (see, for instance, musical excerpt in figure 3b); mistakes can be corrected either interactively during the score-extraction process or manually after the symbolic representation has been generated.

2. Salience of Musical Objects

As a first task elementary musical 'objects' are identified in the MIDI file and corresponding accent values are calculated for each of them.

2.1 Identification of Elementary Musical Objects

Musical 'objects' such as chords, arpeggiated chords, trills, mordents etc. are initially identified in the raw midifile. It is hypothesised that such collections of notes are perceived as 'wholes' before being broken down to their constituent parts.

It is necessary, as a first step, to determine *when* a number of independent notes are close enough to be considered constituents of a chord. Empirical research has shown that two tones are heard as being synchronous if their onset differences are less than roughly 40 ms - for more than two tones this threshold is higher but usually not more than 70 ms [5]. We have used the threshold of 70 ms to convert the MIDI data into chords.

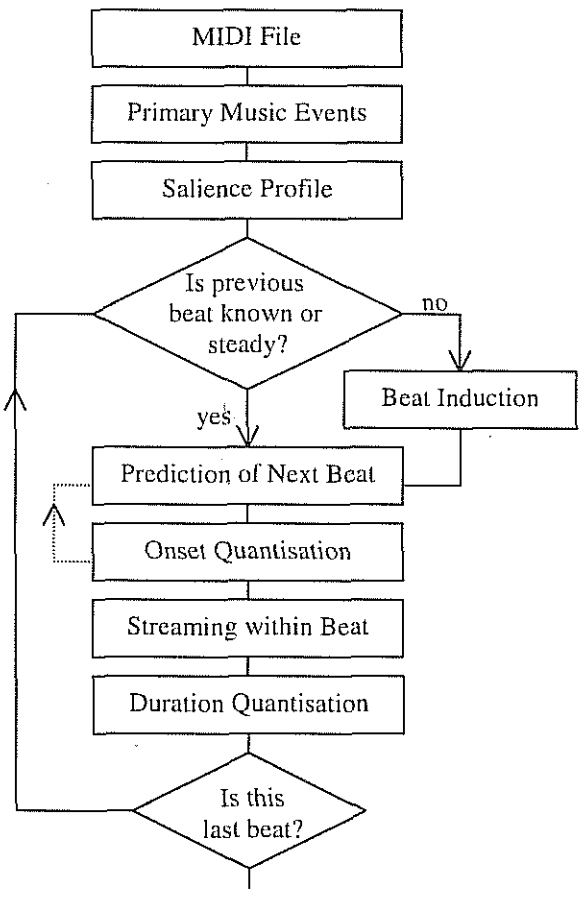


Figure 1 Overall score extraction system (pitch spelling is a separate module).

Musical knowledge is essential for identifying other kinds of objects such as trills, grace-notes, mordents, arpeggiated chords and so on (see, for instance, problems of transcribing arpeggiated chords in figure 3c). The exact definition of such objects and the development of routines that can compute them is not trivial. As this is a large topic in its own right it will not be discussed further in this paper.

2.2 Musical Accents

For each musical object an 'accent' strength is computed that indicates the relative salience of the object. Musical salience is determined by a number of factors such as note duration, dynamics, pitch, harmony, cadences and so on (for instance, longer notes or harmonically more stable notes are more salient). In the current version of this system the following three salience factors have been taken into account: a. duration in ms of notes in each event, b. dynamic value (MIDI velocity) and c. pitch value. Notes with longer durations, higher dynamic values or lower pitch are considered to be more salient. The rationale behind the last factor is that notes in lower voices (especially the bass line) tend to be less ornamented and more accurate in terms of keeping time whereas higher voices (e.g. melody) have more expressive flexibility and variation. The salience strength of events (e.g. chords) is the sum of the strengths of their constituent notes.

A possible salience strength function for each note is proposed:

$f(d,p,v) = (v/p) \cdot d$ where:

d is duration in ms (d = Offset - Onset),
v is dynamic value (MIDI-velocity), $30 < v < 90$
p is pitch (MIDI-pitch), $30 < p < 60$

The most significant factor into the above function is the duration of notes as this can take a wide range of values (from roughly 100 ms to a few seconds). The dynamic and pitch factors become more influential when notes have relatively similar durations.

Knowledge of such accent strengths can enhance beat-processing applications (see below).

3. Finding the Beat

Beat is taken to refer to a regular pulse perceived when listening to music - this is usually the temporal level at which listeners tap their feet or clap their hands. The action of finding an appropriate beat rate (tempo) will be referred to as *beat-induction* whereas the task of following the beat will be referred to as *beat-tracking*. Ordinary human listeners are competent in performing these tasks; equivalent performance on the computer has proved however remarkably hard to achieve.

Most beat-tracking and beat-induction systems rely solely on note onsets or inter-onset intervals [3,8,9]. It has often, however, been proposed in theoretical work that finding the beat (the lowest meaningful level of metrical structure) involves matching a regular grid to the *accentuation* structure of a music work [6,7]. It is herein hypothesised that beats tend to fall on more accented/salient events.

3.1 Beat Induction

Extracting an adequate beat is achieved by 'looking' into a pre-specified time window for the most common inter-onset interval. All the possible IOIs formed by all the onsets within the window (e.g. 3 seconds) are calculated. For each IOI, a salience value is attached that takes into account the accents of its two delimiting notes (in this implementation the IOI salience is simply the sum of the two accents). The IOIs are sorted by size and 'clustered' into small overlapping time bins (70ms). For each time bin the sum of all the IOI saliences is calculated. The cluster with the highest salience value determines the local tempo (i.e. inter-beat interval).

The IOI clustering algorithm is described in detail in [4]. The current version differs from that algorithm in that it takes into account the salience of each IOI.

The beat induction module is activated at the beginning of the piece and whenever beat variance of preceding beats exceeds a certain value (meaning that the beat-tracker is 'lost').

3.2 Beat Tracking

For a given moment in a piece a prediction is made for the next beat based on the previous few beats (e.g. weighted average of the last three beats). The next beat position is calculated using a Gaussian function that is centered at the predicted time point and that has an overall width equal to the beat value. The 'steepness' of the function depends on the observed variance of the previous beats. The accent strengths of musical objects within the prediction range are adjusted according to the function and the greatest value is selected. If there are no events in the predicted window, then a beat position is placed at the predicted point and beat-tracking proceeds from there on.

This method takes into account the relative salience of musical events. For instance, a weaker event that falls closer to the predicted beat position may be disregarded as the algorithm may select a stronger event that has a greater distance to it. This way, the beat tracker can track correctly sections with greater temporal deviations (more rubato). In a recent study [4] it is shown that the performance of a multi-agent beat tracking system improved from 75% to over 90% when musical event accents were taken into account (the system was applied to the same Mozart data).

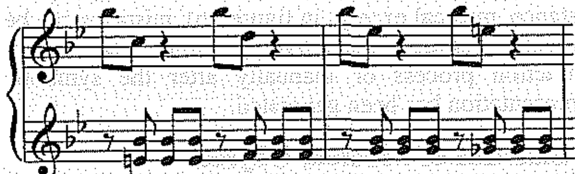


Figure 2 In this example, the beat may erroneously shift forward by one eighth note as the notes on every actual beat are less accented than their succeeding 3-note chords.

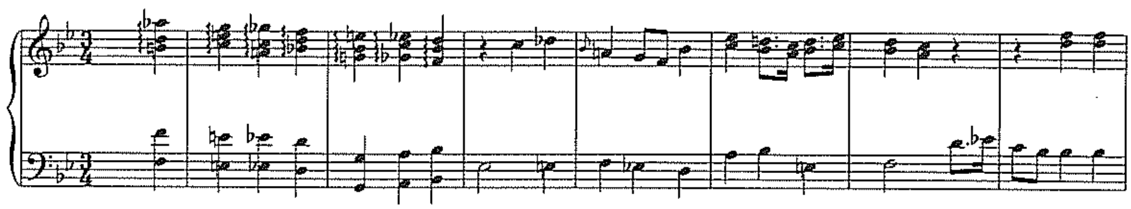
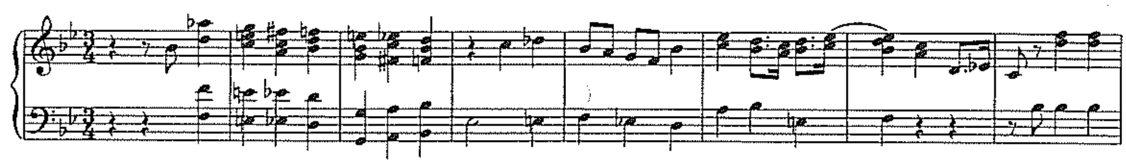
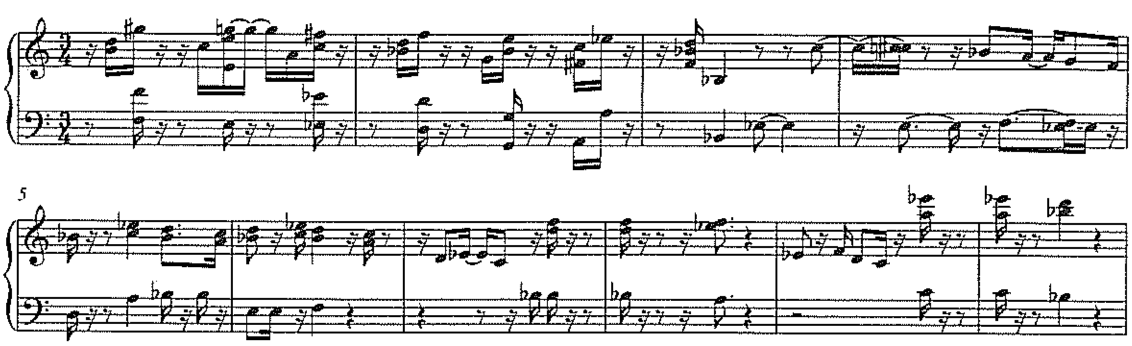
This beat-tracking algorithm can make mistakes especially in cases where weak events appear on beats and strong events appear very close to them. In figure 2 the beat-tracker may shift forward by one eighth note; if, however, the previous tempo has been very steady it is likely that this section will be tracked correctly as well as the prediction function will be very sharp. The beat-tracker usually recovers from mistakes as soon as strong musical events indicate the correct beat (there are cases however where the beat-tracker goes astray). Points where the beat tracker has gone wrong can be corrected interactively by the user; after a particular beat is corrected the system is allowed to track the beat again from that point onwards.

a. Original score

b. Score extracted from performed midifile by the system described in this paper

c. Performed midifile imported and notated in commercial notation software.

Figure 3 Beginning of second part of *Menuetto I, Sonata KV282* by W.A.Mozart: a) original score, b) extracted score by the current system, and c) the midifile imported and displayed by a music notation software package.

4. Time Quantisation

Once a beat is detected, an attempt is made to quantise all the onsets and durations of notes within the range of this beat and its preceding beat.

4.1 Onset Quantisation

Onsets within the range of two successive beats are quantised by selecting the time grid (multiples of 2 and 3) that fits best to the observed onset values. Rather than 'moving' the observed event onsets to the closest points of the quantisation grids and then selecting the grid for which the minimum deviation error is computed (as many commercial quantisers do), quantisation is applied to each inter-event interval. The last inter-event interval in the beat equals to 1 minus the sum of the previous quantised beat fractions (this way a round-off error is avoided). A goodness value that is inversely related to the deviation error is attached to each subdivision level. The quantisation subdivision level that gives the highest goodness value is selected (further discussion on the quantisation problem can be found in [4].

4.2 Duration Quantisation

The actual offsets of notes are of hardly any use in quantisation, especially for a percussive instrument such as the piano (even more so if the pedal is used). Performed durations of notes (at least for piano) are usually much shorter than the nominal score values; for instance, in figure 3c, quantised durations are much shorter than in figure 3a and there are many additional rests. Calculation of note durations has thus been based in the current system on an elementary 'streaming' algorithm whereby notes within a beat are split into independent streams (voice parts) and then durations are considered to be equivalent to the inter-onset intervals for the notes of each stream.

The streaming algorithm is based on the Gestalt principle of proximity and simply tries to find the shortest streams that connect all the onsets within a beat (figure 4). Crossing of streams is not allowed. The number of streams is always equal to the number of notes in the largest chord. The solution to this problem is not trivial and appropriate searching techniques are required for developing an efficient algorithm. The current elementary version of the algorithm makes mistakes but can be improved if other principles like 'goodness of continuation' are taken into account. Streaming is a large research topic in its own right [1] and the current algorithm is only a means for providing a crude streaming of notes so that durations of notes may be calculated as interonset intervals of each stream. Obviously this methodology avoids rests (rests, however, may appear at the beginning of beats if no note was held from a previous beat and if the beginning of the beat has fewer streams than its continuation).

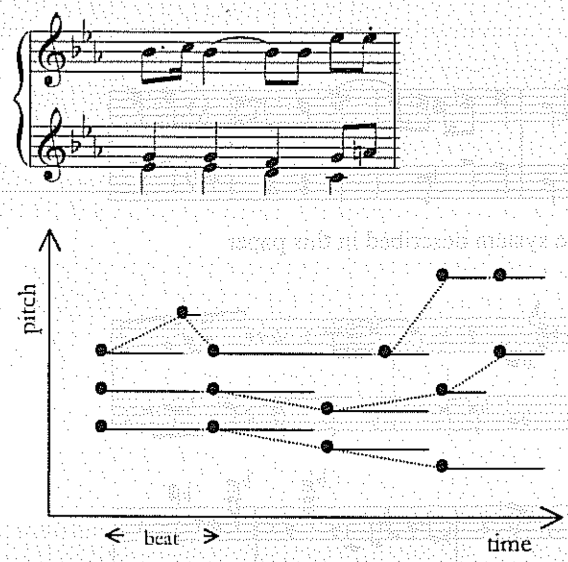


Figure 4 Duration quantisation relies on a streaming algorithm. Dots in the graph represent the onsets of the notes in the musical segment; dotted lines show the three streams detected by the algorithm; horizontal lines indicate the inter-onset intervals for each stream.

5. Pitch spelling

In [2] an algorithm is presented that finds correct diatonic spellings for the notes of a MIDI melody. This algorithm has been extended and can now be applied to polyphonic music. The algorithm is based on two principles:

a. notational parsimony (i.e. double-sharps and double flats are avoided) and

b. diatonic interval optimisation (i.e. common diatonic intervals are preferred whereas rare intervals such as augmented, diminished, chromatic intervals are avoided).

The above optimisation procedure is applied locally on a MIDI file using a shifting overlapping windowing technique. The spelling of the pitch classes within the window that gives an optimal solution for the above two rules is selected. This transcription algorithm is unique in that it requires no previous key- or tonality-finding mechanisms; on the contrary, it can be used as a precursor to key-finding algorithms. A different pitch spelling algorithm based on the closseness of pitches within the circle of fifths is presented in [10].

This algorithm has been tested on 3 complete Mozart piano sonatas (KV279, KV280, KV281) for which it gives 97% correct results (total number of notes is 13002; number of notes with accidentals is 3546; notes misspelled by algorithm is 111) - further extended tests are necessary on music from different tonal styles for assessing the full potential of the algorithm. Mistakes occur occasionally (see, for instance, bars 2 and 3 of figure 3b), mainly because of local conflicts introduced by voice-leading concerns.

The proposed pitch-spelling component, on the one hand, highlights the importance of the traditional diatonic pitch-interval system when dealing with tonal music and, on the other hand, provides a very simple and robust algorithm that can be easily incorporated in a number of music applications (e.g. music notation software, score-extraction systems, music analytic systems and so on).

6. Conclusions

In this paper a system that attempts to extract the musical surface from a performed MIDI file has been presented. A wide range of independent but interacting modules was described. The current score extraction system can be further improved by refining the existing modules and determining more accurately their interdependencies; additionally, new components may be introduced to cope with aspects of musical scores that depend on knowledge of musical structure.

Acknowledgements

This research is part of the project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science, and Culture in the form of a START Research Prize and support to the Austrian Research Institute for Artificial Intelligence.

References

[1] Bregman, A. S. 1990. *Auditory Scene Analysis*. Cambridge, Mass: The MIT Press.

[2] Cambouropoulos, E. 1996. A General Pitch Interval Representation: Theory and Applications. *Journal of New Music Research*, 25(3): 231-251.

[3] Desain, P. and Honing H. 1992. *Music, Mind and Machine*. Amsterdam: Thesis Publishers

[4] Dixon, S. and Cambouropoulos E. 2000. Beat Tracking with Musical Knowledge. *In Proceedings of ECAI 2000*, Berlin.

[5] Handel, S. 1989. *Listening. An Introduction to the Perception of Auditory Events*. Cambridge, Mass: The MIT Press.

[6] Lerdahl, F. and Jackendoff, R. 1983. *A generative Theory of Tonal Music*. Cambridge, Mass: The MIT Press.

[7] Povel, D. J. and Essens, P. 1985. Perception of Temporal Patterns. *Music Perception* 2:411-440.

[8] Rosenthal, D. 1992. Emulation of Human Rhythm Perception. *Computer Music Journal* 16(10):64-76.

[9] Rowe, R. 1993. *Interactive Music Systems (Machine Listening and Composing)*. Cambridge, Mass: The MIT Press.

[10] Temperley, D. 1997. An Algorithm for Harmonic Analysis. *Music Perception*, 15(1):31-68.

Analysis of Expressivity in Movement and Dance

Camurri Antonio, music@dist.unige.it
Trocca Riccardo, ghost@musart.dist.unige.it

Laboratorio di Informatica Musicale (http://musart.dist.unige.it)
DIST - University of Genova, Viale Causa 13
I-16145 Genova

Abstract
This paper shortly reviews some recent work from our research project concerning the development and experimenting of algorithms for expressive movement analysis. The method and computational models are inspired to several sources, including Rudolf Laban's Theory of Effort. The paper describes some of the ongoing experiments and results. The ultimate goal is to contribute toward a deeper understanding of the relations between gesture, music, and visual languages, and enable artists with conceptual as well as system tools for interactive performance.

1. Introduction

Our research aims at developing computational models and computer systems able to analyse human movement in the perspective of expressiveness, emotional content, KANSEI (Hashimoto 1997). In a broad perspective, research on "affective computing" (mainly in USA) and on "KANSEI Information Processing" (KIP) in Japan are consolidated research areas. In the scenario of our research project, we are aiming at a sort of third way, where research on expressivity and artificial emotions is reconsidered in the framework of European culture (Camurri 1999).

A gesture or a sequence of gestures can be performed in different ways and nuances, in order to convey different emotional and expressive content. The gesture "raising a hand" can be performed, for example, with different velocity and/or shape nuances of the trajectory.

In our research we first try to identify the crucial parameters that can be used to express emotional content. Then we study ways in which they can be used, and we select which of them can be studied reliably through sensors and computer systems. Our goal is to enhance man-machine communication by adding a new channel: expressivity. This level of communication may be useful for musicians, choreographers, actors to develop multimedia content for interactive performances with enhanced degrees of expressiveness. It is not only a gesture, recognised from a gesture vocabulary, to trigger a response in an interactive computer system, but also the way in which it is performed, its expressive content.

2. The EyesWeb System

In order to collect the movement data and to analyse them, we use our EyesWeb software system. It can use one or more videocameras, accelerometers, and other sensor systems. Sensor data are then processed according to user-defined patches, i.e., networks of modules. Results are available as MIDI or TCP/IP outputs. EyesWeb allows the user to access, through a visual environment, a wide library of software components (called blocks) and to connect them in a network performing various kind of operations (e.g., video signal processing, feature extraction, math and logic processing, etc.). Further, the real-time kernel supports different types of software modules (bricks in the visual language), including "active modules" and memory management (Camurri et al 2000). The system can be enhanced adding new blocks easily and it has been used also in several live performances.

3. Experimental Setup

In order to test the algorithms developed in the project, we created, with the help of professional choreographers, a library of dance fragments, called *microdances*. Each microdance is performed with different expressive content, for example "fluid" or "rigid" and also "neutral" (i.e., trying to avoid any expressivity). In this way we can test the differences among the same dance fragment performed with different expressive content and the similarities among different fragments performed with the same expressivity.

4. The Parameters

One of the most important steps of our research concerns the individuation of significant parameters of human movement that can be studied by a computer system. This has been accomplished trying to translate concepts from different movement disciplines into mathematical quantities. For example some of our experiments were inspired by the *Theory of Effort* (Laban and Lawrence 1947).

Usually, in the case of data collected through a videocamera, a measure is associated to each frame, e.g., limbs position or area occupied by the body. In order to study expressivity, however, it is necessary to study how the extracted measures vary in time, for example through derivative or FFT transform.

5. Experiment 1:Detecting Steps

One of the parameters we studied has been called "Stability". Stability is a number associated to a body posture that attempts to measure its stability (or instability). Roughly, it is computed by dividing the height of the body baricenter by the distance between the feet while they both are on the ground. Its value can be interpreted in several ways, for example if it is contained in a certain range it indicates a stable position, above or below that range it shows that the position may be unstable.

In this way it is possible to evaluate what kind of positions are used during a performance and how the performer skips through them. The most direct use, the simplest possible, is detecting peak values of the parameter, corresponding to a step being performed. Steps frequency is another parameter that can be studied. The picture shows the behaviour of the "stability" parameter during a dance fragment. Peak values correspond to feet crossing, making a step.

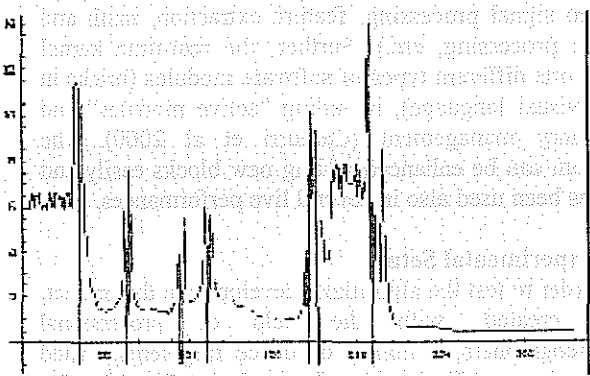


Figure 1 Plot of the "Stabilty" parameter. Peaks in the graph correspond to steps made by the dancer.

6. Experiment 2: Contraction and Expansion

The expansion or contraction of the space used by the dancer is an interesting parameter to study. According to Laban's *Theory of Effort* the human body is surrounded by a sphere, called *Kinesphere,* whose amplitude corresponds to the maximum extension of the limbs. During a sequence of movements the limbs can extend and touch the outer limit of the Kinesphere, or be kept close to the body. In several experiments we measured the space used by the body associating it with a rectangular area that covers the actual extension of the limbs. In figure 2 it is shown a sample of a motion capture session performed by EyesWeb with a rectangle surrounding the area occupied by the body.

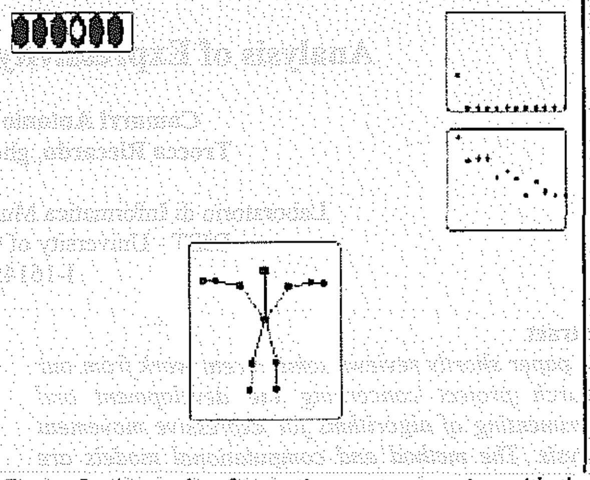


Figure 2: A sample of a motion capture session with the EyesWeb system. The plots on the right side are the FFT of the area of the rectangle surrounding the body, both in linear and logarithmic scale. On the left side there is a representation of a 6 units SOM with unit 4 activated.

Figures 3 and 4 show how this parameter varies during the same microdance performed with different expressive content. In order to detect the differences introduced in the various performances we used a Kohonen self organising map, applied to the FFT of the rectangle area. After training the net, we observed how the data relative to different performances were represented on it. We noticed that different kinds of movement (slow, fast, impulsive) activate different portions of the Kohonen map. For example, using a *simple monodimensional map* composed of 6 units, unit 1 is activated when there is no movement, while slow movements activates units 2 through 6 in sequence oscillating in a continuous way from one to another. Fast movements oscillate in a narrower region formed by unit 4 and 5, while impulsive movements are mapped by unit 2 and 3. An impulsive expansion followed by a slow contraction is mapped by units from 2 to 5, being a kind of superposition of the slow and impulsive movement behaviours.

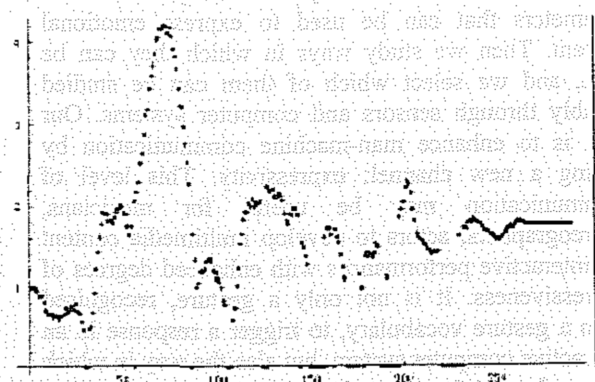


Figure 3 The variations of the area occupied by the body during a dance performed with "fluid" character.

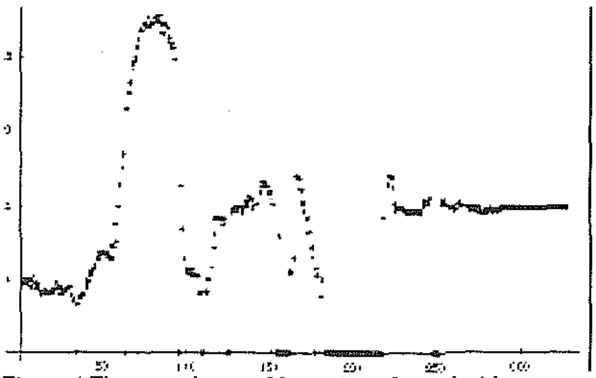


Figure 4 The same dance of figure 3 performed without any expressive intention("neutral")

7. Discussion

The paper raises issues and new perspectives on the integration of music and movement languages. For example, the director of the performance in a live electronics (Vidolin 1997) co-operates with dancers or music performers during the performance. Parameters based on qualitative analysis of movement such as those presented in the sample experiments in this paper, may be a step toward deeper interaction metaphors. We can conceive sophisticated examples of interaction in which the communication between a dancer or a music performer and the active space can evolve, basing for example on learning, imitation, dialogue: the interaction metaphor can therefore vary from the "musical instrument" to the "orchestra conduction", up to different kinds of dialogue and "social interaction". The role of the composer can become as "the moulder of a character/agent/active space", of "the perception system" of such a character/active space the dancers interact with, and can program its "learning capabilities". The composer can define in this way music objects as active entities, deciding which degrees of freedom he leaves to the performer, and for which kind of interaction and evolution of the music object. This does not imply necessarily degrees of randomness or stochastic processes in the interaction process.

Models of expressivity may contribute to new paradigms in composition and performance. Our hypothesis - and one of the motivations to our work - is that models of artificial emotions and expressivity might be both a conceptual and a realisation platform to investigate new paradigms. For example, theories of the composer Gerard Grisey define a music domain in term of living entities: his view on sound as *être vivant* rather than *object* seems conceptually close to our view on emotional agents as living beings during a performance.

Our choice of Laban's Effort Theory is another crucial point in our approach to face these issues.

8. Conclusion

Research on expressivity is in its infancy. Nevertheless, the preliminary results we obtained in the ongoing experiments outlined in this paper and in others are encouraging. For example, the simple algorithms outlined in this paper seem to provide reliable data on the amount of qualities of movement like "smooth/rigid", "slow/fast", "explosive". Further ongoing work includes extensions of our movement libraries, improvements and comparison of algorithms, further experiments and testing with dancers, actors, and musicians. A close work with psychologists is planned in order to compare the performance of our computational models with those of human observers of the movement.

References

Camurri A. P.Morasso, V.Tagliasco, R.Zaccaria. 1986. "Dance and Movement Notation." In Morasso & Tagliasco (Eds.), *Human Movement Understanding*, pp.85-124, Amsterdam: North Holland.

Camurri, A. 1995. "Interactive Dance/Music Systems." In Proceedings of the 1995 International Computer Music Conference. San Francisco: International Computer Music Association, pp.245-252.

Camurri, A. (Ed.) 1997. Proceedings of the International Workshop on KANSEI: The Technology of Emotion. Genova: AIMI (Italian Computer Music Association) and DIST-University of Genova, Italy.

Camurri, A. (1999) Music Content Processing and Multimedia: Case Studies and Emerging Applications of Intelligent Interactive Systems. *Journal of New Music Research*, 28(4):351-363, Swets&Zeitlinger.

Camurri, A., Hashimoto, S., Ricchetti, M., Suzuki, K., Trocca, R., Volpe, G. (2000) EyesWeb – Toward gesture and affect recognition in dance/music interactive systems. *Computer Music Journal*, 24(1), MIT Press.

Hashimoto, S. 1997. "KANSEI as the Third Target of Information Processing and Related Topics in Japan" in *Proc. of Intl. Workshop on Kansei - Technology of emotion*, pp.101-104.

Laban, R., F.C.Lawrence 1947. Effort, Macdonald & Evans Ltd, London.

Vidolin, A. 1997. Musical interpretation and signal processing. In C.Roads, S.T.Pope, A.Piccialli, G.De Poli (Eds.) *Musical Signal Processing*, Swets.

InterÉlastique

An interactive performance system for control of an audio-visual experience using novel stretchable sensors

Paul Nemirovsky, Dan Overholt

MIT Media Lab
Cambridge, U.S.A.

Abstract

Interactive installations today typically limit the interaction by allowing only one active user at a time, and are commonly based on traditional forms of user input (i.e., keyboards, trackballs, touchscreens, etc.). This paper describes InterÉlastique, a system which allows collaborative tangible interactions through a set of innovative stretchy sensors called eRopes[1].

The system represents an exploration of the relationship between music, poetry/text, visual elements, and physical space. The aim of the system is to allow users to learn about properties of music, and explore relationships between individual musical fragments and the overall composition – without having to learn a system of complex mappings. Users are presented with two projection screens that are "connected" with six eRopes. By pulling these eRopes, the users change the music as well as the relationship between the textual elements on each of the screens. Multiple users can collaborate, modifying poetry and music interdependently, based on the combination of the different eRope's signals.

The installation is composed of the eRopes and their related electronics, several MIDI synthesizers, two projectors, and a computer running the InterÉlastique application, which provides audio-visual feedback to the participants.

[1] eRopes (Patent Pending) are electronically active tensile sensors. We have developed this technology in cooperation with Saul Griffith, using a novel technique wherein conductive fibers are braided into the sheath of a bungee cord in order to determine the tension applied to it.

1 Introduction

Our goal is to create interactive art installations that are both educational and easily accessible to the general public and children. As such, we hope to persuade the user of the future to be much more proactive and creative while interacting with music, art, and technology. By using the eRopes as our interface, we can give the public a method of participating and learning through an innovative installation without feeling overwhelmed or intimidated.

Most installations today are based on trigger-mode sensors (such as buttons and other event-based inputs), whereas InterÉlastique utilizes the tension and elasticity of the eRopes to provide users with continuous control and feedback. The user input is mapped to the audiovisual components of the installation according to a number of predefined rules:

1. The aesthetic forms should change gradually and smoothly from one state to another.
2. The users' expressions should be directly mapped to an audio-visual effect to reflect their transformations.
3. Input from multiple users can be combined to produce predictable emotional/informative system responses.

2 Related Work

The last twenty years have been marked by an explosion of interactive installations, in both the musical and visual domains. Here are a few that inspired us: digital artist Chris Janney has created what he calls "an urban musical instrument" in a subway station in New York [1]. While waiting for a train, commuters could reach above their heads and break beams of light to trigger music. Passengers on each side of the tracks could not only play the 'instrument' among themselves, but also interact with people on the other side.

Composer Tod Machover created the Brain Opera [2], which explores a variety of multiuser-oriented interaction techniques. The Brain Opera connects a series of hyperinstruments designed for the general public with a performance and a series of real-time music activities on the Internet. Audiences explore the hands-on instruments as preparation for the performance, creating personal music that makes each performance unique.

George K. Shortess in his "Doorways of Meaning" (1996) explored the boundary between inner and outer experience by allowing viewers to generate voices using colored cord networks. Viewers moving in the room cast shadows on the photocells embedded in the installation and changed their resistances. When changes occurred, MIDI signals were sent to a variety of audio devices. The audio outputs were then mixed and played through the speakers in the room.

Finally, the Shape of Sound exhibition series [3] was aimed at creating dynamic environments with the audience defining its relationship to the art at any given moment. The artists involved came from both audio and visual backgrounds and collaborated to create multi-sensorial, immersive experiences which were often described as "ambient". The aesthetic emphasis of the work was "the fluid integration of wide-ranging elements, rather than juxtaposition and contrast".

3 Motivation

The InterÉlastique system is aimed at creating an environment of educational play. Visitors can use the eRopes as an intuitive interface to control the installation. In order to construct this environment, we began by defining our model of interaction. We chose to work in the domains of music, visuals, and poetry, and defined the system's responses for a variety of situations (e.g., idle state vs. a single eRope vs. several eRopes active at the same time).

In order to encourage users to explore the various aspects of the installation, we attempted to provide evocative content by designing an unusual combination of visuals and dynamic textual elements, as well as composing a set of emotive musical patterns. We then conducted user tests in order to refine the system's content. A brief overview of the Design and Testing process follows.

4 Interaction Design

The InterÉlastique system provides real-time response to the users, maintaining the perception of action-reaction for both participants and observers. The goal of InterÉlastique is not only to provide users with an innovative mode of control, but also to encourage collaborative learning through the installation. For example, a group of users pulling several eRopes simultaneously can achieve more complex musical and textual transformations than would be possible with just one person.

In an effort to develop a new way of interacting with musical and textual elements, we mapped the system's audiovisual output to the following emotions (determined by the response curves of the signals from the eRopes): Anger, Joy, Sorrow, and Surprise.

Anger is mapped to strong, irregular movements of the eRopes with large amplitude swings over time. Joy is recognized as a smooth, continuous movement of the eRopes with a rising speed of modulation. Sorrow is expressed by slow continuous changes in tension with an overall decline in modulation speed. Finally, surprise is mapped to the collaborative interaction such that, soon after one of the eRopes is triggered, several others follow in a similar manner.

All of these parameters affect the audiovisual elements presented to the visitors. These are: triggering of various musical and textual elements from pre-composed libraries, speed and direction of the text movement on the screens, color and intensity of the visuals, text sizes, controlling audio levels and filters, etc. The interactions are partially pre-scripted, but the interaction manner is left completely up to the users, dependent on the way they choose to control the eRopes.

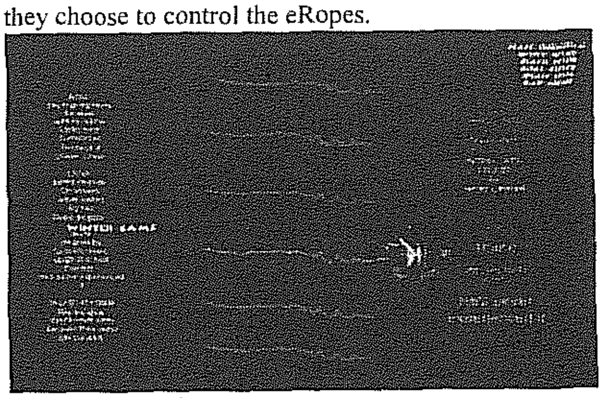


Figure 1: InterÉlastique software

The InterÉlastique software (fig. 1) has been designed using the Lingo programming language, with additional Visual C++ modules for serial port access, MIDI playback, and visual effects.

5 Structure of InterÉlastique

The InterÉlastique hardware (fig. 2) is based on the eRope technology that has been developing at the MIT Media Laboratory. The prototype eRope electronics system uses a custom-built circuit board that has a Microchip PIC microcontroller (with a built-in analog-to-digital converter), as well as analog signal conditioning components.

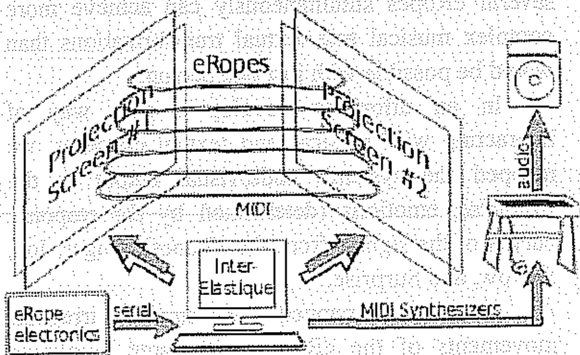


Figure 2: InterÉlastique hardware

The microcontroller runs a custom operating system, which does data acquisition for each of the eRopes, and provides serial RS-232 communication to the main computer. The tension in each of the eRopes is converted into an 8-bit value and sent to the main computer, which is running the InterÉlastique system. This computer controls several MIDI synthesizers which produce the music, as well as two video projectors which display the poetry and visuals.

6 Evaluation and Testing

The InterÉlastique system was installed in the central atrium of the MIT Media Laboratory, providing us with much informal feedback. In order to gather more formal data, a group of twenty-four users (four sessions each with six users) was asked to interact with the system, creating new collaborative musical / visual experiences. We asked the users whether the system was responding to their actions in a way they considered natural and engaging. Many of the users reported that the system led them to think about new and interesting ways to control music, verifying that the system has provided an engaging environment for learning.

We did not observe a significant difference in interaction between users that were given an explanation of the system before they tried it, and those that were asked to figure it out by themselves. The system facilitated interaction - users that were given an explanation first tended to be more proactive, asking others to join them in the creative process. Users have also reported that they enjoyed the process of learning as an integral part of their experience (rather than as an annoying first step). Additionally, the data seems to show that users tend to synchronize their interactions with the system after a short period of adjusting to each other.

7 Future Work

We would like to create a distributed model of the InterÉlastique system with multiple clients running simultaneously via a network. That will allow remote control of the installations, and will open the way to redefining collaborative learning and participation in an art installation setting.

We hope to expand the scope of our work by defining more generic modes of interaction based on non-discrete interfaces, with the ultimate goal of allowing each user to become an artist. We hope that this will serve a variety of artistic and educational purposes, and help blend the distinction between the passive audience model and the proactive model of artistic communication.

8 References

[1] Janney, C. - "REACH - New York", 1996.
[2] Machover, T. Hyperinstruments: A Progress Report. Cambridge: MIT - Media Lab, 1992.
[3] The Shape of Sound, www.lonestar.texas.net/~monorecs/bos/sos.html
[4] Roads, C. - "The Computer Music Tutorial", MIT Press, 1998.

MUSICAL INFORMATICS: EXPRESSION AND PERFORMANCE ANALYSIS III

MUSICAL INFORMATICS:
EXPRESSION AND PERFORMANCE
ANALYSIS III

Symbolic and audio processing to change the expressive intention of a recorded music performance

Sergio Canazza, Federico Cestonaro, Giovanni De Poli, Carlo Drioli, Antonio Rodà

University Of Padova, Center of Computational Sonology (CSC)
Dept. of Electronic and Informatics – Via Gradenigo 6/a – 35100 Padova, Italy
canazza@dei.unipd.it

Abstract

In this work we present a method for controlling the mutable musical expressiveness in an abstract way, not necessarily related to the acoustic parameters. Our approach integrates symbolic and audio processing in a real-time networked environment. We present a mapping strategy allowing to move in the control space and varying the expressiveness in a coherent way. All the variations computed by the model on the symbolic representation are successively managed by the signal processing sub-system which coherently transforms the sound by combination of different audio effects. The signal processing engine works in real time to render the desired expressive audio variations. The model is used interactively in a networked environment. The user controls the expressive character of the performance by moving within an appropriate control space.

1 Introduction

In multimedia products, textual information is enriched by means of graphical and audio objects. A correct combination of these elements is extremely effective for the communication between author and user. Usually, designer's attention is put on graphics rather than sound, which is merely used as a realistic complement to image, or as a musical comment to text and graphics. With increasing interaction in multimedia systems, while the visual part has evolved consequently, the paradigm of the use of audio has not changed adequately. In the usual interaction with audio media, transformations on audio objects are not allowed. A more intensive use of digital audio effects, will allow to interactively adapt sounds to different situations, leading to a deeper fruition of the multimedia product. It is advisable that the evolution of audio interaction leads to the involvement of expressive content. Such an interaction should allow a gradual transition (morphing) between different expressive intentions.

Recent researches have demonstrated that it is possible to communicate expressive content at an abstract level by changing the interpretation of a musical piece. In human musical performance, acoustical or perceptual changes in sound are organized in a complex way by the performer in order to communicate different emotions to the listener. The same piece of music can be performed trying to convey a specific interpretation of the score or the situation, by adding mutable expressive intentions. In a similar way we could be interested in having models and tools which allow to modify a performance by changing its expressive intention. In (Canazza et al. 1998, 1999) we presented models which are able to modify the expressive content of a performance in a gradual way.

We aim at using these models in applications of multimedia products; in particular, an extension of these models is presented, focused on auralization and web based applications.

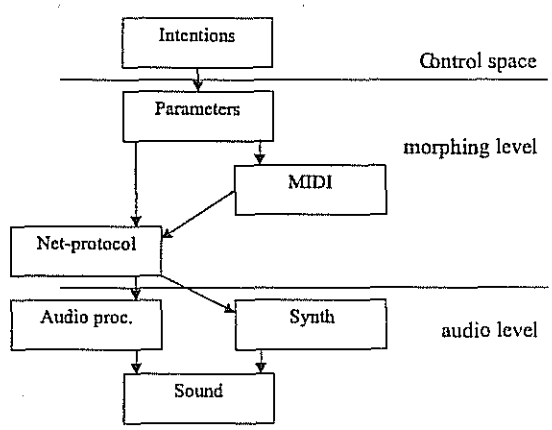


Fig. 1: Scheme of the system

2 Controlling expressiveness

For an effective control on expressiveness, three different levels of abstraction are needed.

The control space is the user interface, which controls, at an abstract level, the expressive content and the interaction between the user and the audio object of the multimedia product.

In order to realize morphing among different expressive intentions we developed two abstract control spaces: the first, perceptual parametric space (PPS), was derived by multidimensional analysis of various professionally performed pieces ranging from western classical to popular music. The second, synthetic expressive space, allows the artist to organize his own abstract space by defining expressive points and positioning them in the space. In this way, a certain musical expressive intention can be associated to the various multimedia objects. Therefore, audio is modified in its expressive intention, both when the user focuses on a particular multimedia object (by moving

with a suitable pointer) and when the object itself enters the scene. One can also take into account the possibility that the multimedia object tell the system information about its state, which can be exploited for expressive control on audio. For instance, in a virtual environment the avatar can tell the system its intentions which will be used to control audio expressiveness; one can thus gain a direct mapping between the intentions of the avatar and audio expressiveness, or a behavior chosen by the artist in the design step.
Using suitable mapping strategies allows the user to vary coherently and gradually the expressiveness (i.e. morphing among happy, solemn and dark), by moving inside the control space. Morphing can be realized with a wide range of graduality (from abrupt to very smooth), allowing to adapt the system to different situations. Analysis-by-synthesis method was applied to estimate which kind of morphing technique ensures the best perceptual result. The computer-generated performances showed appropriate expressive meaning in all the points of control space, computing intermediate points of the space using a quadratic interpolation.
It has to be noticed that expressive content of a performance is revealed on a time scale which is longer of that of a single event; therefore, in order to obtain a fruition which is coherent with the artist's intentions, the system allows to slow down the movements of the user, so to avoid unwanted "expressive discontinuities" in correspondence of abrupt movements. To this end, suitable smoothing strategies have been developed for movement data coming form the pointer.
The expressive parameter layer translates high-level information of the control space in order to modify the acoustical parameters used by the models for the expressive morphing. On the basis of performance analysis (Canazza et al. 1997) some of the parameters have turned out to be particularly important for the reproduction of expressive intentions, for instance Tempo, Legato, Intensity, Phrasing, etc. The models make use of these parameters in order to determine the deviation which have to be applied to the score for the reproduction of the desired expressive intention. The user can define an expressive intention by means of a suitable set of parameters, or can use a mapping Intentions-parameters derived by acoustical analysis.
The system can handle three different models for expressive morphing, which make use of different levels of information from the score. The first one (Canazza et al., 1998) has three inputs: the score, a description of its musical structure, and a "neutral" performance (i.e. human performance without any expressive intention). A second model (Canazza & Rodà, 1999) needs the score together with a neutral performance. Finally, a third model needs only a performance (nominal or neutral). The last two models work in real time, so in the next we will refer to them.
By means of the net-protocol, the model is used interactively in a networked environment. The user controls the expressive character of the performance by moving within the control space using suitable control devices.

3 Model

The expressiveness model acts on an abstract level (morphing level), computing the deviations of all musical parameters involved in the transformation.
For each expressive intention, the deviations of the acoustic parameters are computed using the following equation

$$\delta P(n)=\frac{P_{out}(n)}{P_{in}(n)}=k_P\cdot\frac{\overline{P_{in}}}{P_{in}(n)}+m_P\cdot\left(1-\frac{\overline{P_{in}}}{P_{in}(n)}\right)\quad(1)$$

where:

- n is the cardinal number of the n-th note of the score.
- P stands for the different parameters modified by the model. Up to now, the model can process the Inter Onset Interval between the current and the successive note (IOI), the duration of the current note (DR), the duration of the attack (DRA), the mean envelope energy (I), the time location of the energy envelope center of mass (EC).
- The subscript *in* indicates that the value of the parameter P is computed from the inputs. In fact, some parameters (e. g. IOI and L) were computed starting from both score and neutral performance.
- The subscript *out* indicates the value of the expressive performance, that is the output of the system.
- $\overline{P}$ stands for the mean of the values measured in the input performance (for the parameter P).
- k_P and m_P are coefficients that carry out two different transformations of the parameter P. The first one performs a translation and the second one performs a stretching of the values.

For each parameter P, the k and m coefficients are computed by means of a mapping strategy (Canazza et al., 1998, 1999) obtained by processing results from acoustical and perceptual analyses (Canazza et al. 1997). By means of these strategies, different points of the control space are associated to sonological parameters ("mid-level" acoustical parameters) such as intensity, tempo, articulation, phrasing, and so on. In Fig. 2 we show a mapping surface between points of a two-dimensional control space (PPS) and the parameter *Ktempo*, (a value greater than one stands for a *rallentando*, while a value lower than one stands for an *accelerando*). On the basis of movements of the pointer on the xy plane, the variations of the parameter *Ktempo* to be applied to the performance are thus computed.

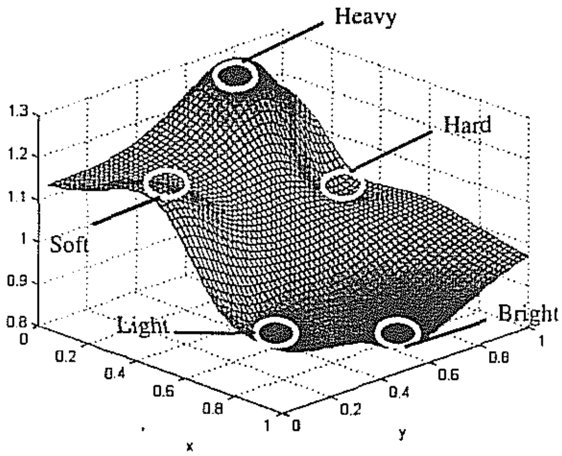


Fig. 2: Mapping between natural control space (see PPS in Canazza & Rodà, 1999) and *Ktempo* parameter.

4 Audio Processing

All the variations computed by the model on the symbolic representation are successively managed by the signal processing sub-system which coherently transforms the sound by combination of different audio effects.

The signal processing engine works in real time to render the desired expressive audio variations. It is based on a sinusoidal model analysis/resynthesis framework, a technique suitable for high quality sound transformation and higher level attribute recognition such as notes, grace notes and staccato-legato or vibrato. The input of the system consists of a digitally recorded performance with neutral expressive intention, and of a symbolic segmentation of the performance representing the musical-level attributes like note onsets, and offsets. This allows the definition of a joint description for sound and performance levels, and modifications in the symbolic level are reflected to the signal level through the audio processing techniques. To this purpose, a finite state automata (here called "Effect Manager"), responsible for the selection, combination and time scheduling of the basic audio effects, computes in real time low-level frame-rate control curves related to the desired expressive intentions. These time-varying curves are then used to simultaneously control the different audio effects during morphing.

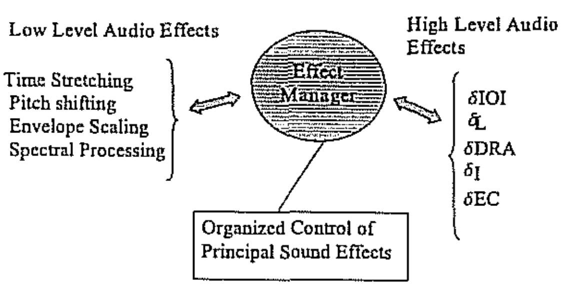


Fig. 3: Expressiveness-oriented high level audio effects (on the right) are realized by means of an organized control of basic audio effects (on the left).

The basic audio effects involved in the expressive processing are: *Time stretching, Pitch shifting, Envelope scaling,* and *Spectral processing*. The rendering of the deviations computed by the expressiveness model ("high-level" effects) may imply the use of just one of the basic sound effects, or the combination of two or more of these effects, as shown in Table 2 and as detailed in the following:

Local tempo processing (variation of *IOI* and *DRA*) is realized by computing different time-stretching factors for the attack, sustain, and release of notes.

Legato control relies on time stretching for changing the duration of the note, and on spectral interpolation between the release of the actual note and the attack of the next note, if overlapping occurs due to time stretching. More details on the management of local tempo and legato processing can be found in (Canazza et al. , 1999).

δIOI	Time Stretching
δL	Time Stretching & Spectral Processing
δDRA	Time Stretching
δI	Spectral Processing & Envelope Scaling
δEC	Envelope Scaling

Fig. 4: Basic audio effects involved in each high level audio effect.

The *musical accent* of the note, which is related to the amplitude envelope centroid, is usually located on the attack of notes for expressive intentions like *light* or *heavy*, or on the release for intentions like *soft* or *dark*. This parameter can be controlled by changing the position of the apex of a triangle fitted to the original energy envelope, and by scaling the energy of frames to fit the new target triangle.

Intensity control can be obtained by scaling of amplitudes of the harmonics. However, an equal scale factor for each harmonic would result in a distortion of the instrument characteristics. To preserve the natural relation between intensity and spectral envelope, a method based on accurate interpolation of spectra relatives to different intensities is used. From the analysis of performances with different expressive intentions, he mean spectral envelopes for maximum intensity I_{Max}, minimum intensity I_{min}, and intensity of the natural performance I_{nat}, are obtained for each note (Figure 5, upper frame, shows a single note). An interpolation schema is then designed in terms of a map $\mathcal{F}: \mathbb{R} \to \mathbb{R}^H$, where H is the number of harmonics, so that for a desired intensity change δI, $\mathcal{F}(\delta I)=[r_1 r_2 \cdots r_H]$ gives the magnitude deviations r_h, $h=1..H$, to be applied to each harmonic of the reference spectrum (i.e., the one relative to the intensity I_{nat}). For the purpose of having a parametric model to represent the behavior of the sound spectrum, Radial Basis Function Networks are used (Drioli, 1999). To resume the principal characteristics, RBFNs can learn from examples, have fast training procedures, and have good interpolation properties. The proposed interpolation schema is used to compute the spectral envelope related to a desired intensity in $[I_{min}, I_{Max}]$ (Figure 5, lower frame). The analysis step is due for each musical note in the score, in order to have different maps for notes with different pitch.

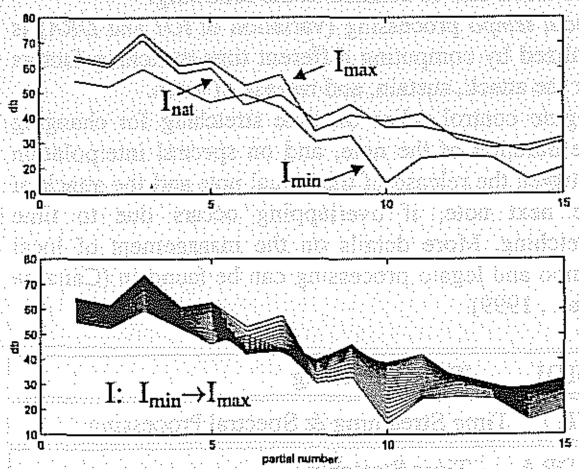


Fig. 5: Control of the intensity of sound, with spectral preserving interpolation. Upper figure: analysis step. Lower figure: synthesis of spectra relatives to a desired intensity I, with I in $[I_{min}, I_{Max}]$.

5 Once Upon a time

An example is given in this section, where an application for the fruition of fairy-tales in a remote multimedia environment is designed. This kind of applications allowed us to validate the system. Moreover, it allowed to select the sonological parameters that best agreed with the expressive morphing. In Table 1, the parameters used in this application are shown. With the selected parameters, an expressive character can be assigned to each individual in the tale and to the different multimedia objects of the virtual environement. In this way, the character of each individual of the tale become stronger due to the expressive content of audio.

	Snowwhite	Hunter	Witch	Sleepy	Grumpy
KTempo	0.85	1.03	1.30	1.24	1.04
KArticulation	0.57	0.96	0.84	1.42	0.66
KAttack	0.64	1.24	0.72	1.38	0.76
KIntensity	1.25	0.72	1.41	0.65	1.35
KBrightness	1.40	0.80	0.92	0.80	1.28

Table 1: The sonological parameters used to define the expressive characters in the tale called "Once upon a time..."

The application has been realized as an applet. The expressive content of audio is gradually modified with respect to the position and movements of the mouse pointer, using the mapping strategies described above.

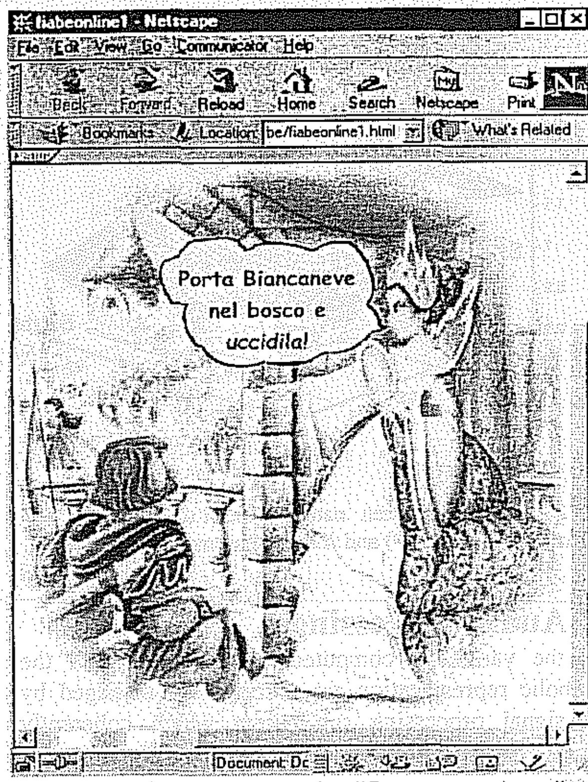


Fig. 6: screenshot from application "Once upon a time..."

6 Conclusions

In multimedia products, audio is becoming an active part of the non-verbal communication process. The proposed system gives to the user the possibility to build multimedia applications in which the expressiveness of audio is modified with respect to the user actions. The system is based on three levels and musical information is processed on an abstract level.

Acknowledgements

This work has been supported by Telecom Italia under reearch contract *Cantieri Multimediali*.

References

Canazza, S., De Poli, G., Roda', A., & Vidolin, A. 1997. "Analysis and synthesis of expressive intentions in musical performance." Proc. of the Int. Computer Music Conference 1997. pp. 113-120.

Canazza S., De Poli G., Di Sanzo G., Vidolin A. 1998. "A model to add expressiveness to automatic musical performance." Proc. of International Computer Music Conf. 1998. pp. 163-169.

Canazza S., Rodà A., Orio N. *1999*. "A parametric model of expressiveness in musical performance based on perceptual and acoustic analyses,". In Proc. of the Int. Computer Music Conference 1999, pp. 379-382

Canazza S. et al. 1999, "Expressive processing of audio and MIDI performances in real time." In Proc. of the Int. Computer Music Conf. 1999, pp. 425-428.

C. Drioli, 1999. "Radial Basisi Function Networks for Conversion of Sound Spectra", in Proc. of DAFx99 Workshop, pp. 9-12.

Synthesis of expressive movement in human-robot interaction

Antonio Camurri, Paolo Coletta, Matteo Ricchetti, Gualtiero Volpe
DIST, Laboratorio di Informatica Musicale - University of Genova (http://musart.dist.unige.it)
Viale Causa 13, 16145 Genova, Italy
music@dist.unige.it colettap@musart.dist.unige.it ricchetti@eidomedia.com volpe@musart.dist.unige.it

ABSTRACT

The paper focuses on experiments concerning the use of mobile robots in music theatre artistic productions and on their hardware and software implementation in concrete human-robot and human-virtual characters interaction.
Such experimental work aims at understanding how expressive content can be conveyed through the movement of robots and interacting with them (e.g. robots interacting with dancers, artists, musicians on a stage): a particular focus is on the expressiveness raising not only from the movement of a robot (i.e. its style of movement), but also from a global, multimodal interaction between a robot and an artist (e.g. a music performer or a dancer). From this point of view, the integration of music and movement is also a crucial issue presented in the paper.
As a concrete example of the work we carried out, we discuss the development of techniques for expressive movement synthesis starting from the lessons learned in experimenting the systems developed for a multimedia artistic production we partecipated with Virgilio Sieni Dance Company (L'Ala dei Sensi, Ferrara, November 1999).

1.INTRODUCTION

One of the main goals of our research is to explore paradigms of expressive interaction between humans and robots in the framework of multimodal environments in music, theatre, museal exhibitions, and art installations. In a previous work (Suzuki et al, 1998), we experimented a small mobile robot on wheels (the Pioneer 1 from Stanford Research Institute) as a semi-autonomous agent capable of communicating by means of several channels, including sound and music, visual media and its style of movement (smooth/nervous, tailwagging, fast/slow, etc.) with the visitors of a museum exhibit on contemporary art ("Arti Visive 2", Palazzo Ducale, Genova, October 1998).
In another experimental setup, we developed a "theatrical machine" for the performance of "Spiral", by K.Stockhausen, for trombone and radio output. The radio, audio amplifier and loudspeakers were installed on top of the robot navigating on the stage, thus creating effects of physical spatialization of sound due to the movement of the robot during the performance (trombone Michele Lo Muto, live electronics Giovanni Cospito, Civica Scuola di Musica, Sezione Musica Contemporanea, Milano, June 1996).

In this paper we investigate how the physical interaction of performers, dancers and robotic scenery can participate as a component of the music language. Interactive visual media and robotic scenery are therefore part of the compositional project of music theatre works, art installations, multimedia concerts. The paper explores these directions by presenting the performance environments developed for the multimedia performance "L'Ala dei Sensi" (literally, "The Wing of the Senses"), held in Ferrara (Italy) on November 1999.

Figure 1. The agent-robot in the art installation at the "Arti Visive 2" museal exhibit, Palazzo Ducale, Genova, Oct 1998.

2. THE MULTIMEDIA PERFORMANCE "L'ALA DEI SENSI"

We carried out experiments in the above described directions in the framework of the work "L'Ala dei Sensi" (Ferrara, Italy, November 1999). "L'Ala dei Sensi" is a multimedia performance consisting of a percourse on the theme of human perception from different perspectives. It aims at explaining scientific principles (on human perception mechanisms) by means of the language of art (dance, music, visual arts). Each episode focuses on a specific scientific issue which is explained and demonstrated in a performance.
The episodes involving interactive dance/music performances also made use of a small mobile on-wheel robotic platform Pioneer 2 from Stanford Research Institute. Our EyesWeb system for real-time gesture analysis and for the design of gesture-music interaction has been used (Camurri et al, 1999; 2000), together our wireless technology for movement sensors, visualisation

and live electronics hardware (see our web site http://musart.dist.unige.it).
These episodes consist of short dance performances in which the dancer himself is involved in a dialogue with the robot, with visual and musical clones on large videoprojection screens. In particular, three episodes are here described and then discussed.

2.1 Dancer - robot dialogue
This is the main episode, where we invested most of our efforts, experimenting and developments. The Pioneer 2 mobile robotic platform has been equipped with sensors, an on-board video projector and a videocamera. The robot was controlled by our proprietary supervision software application developed as an EyesWeb patch (Camurri et al, 1999; Camurri et al, 2000). The sensors allow the robot to avoid collisions with the scenery and the dancers. In the first part of this episode, the on-board video projector and the videocamera were directly controlled in real-time by the director Ezio Cuoghi (off-stage). He used in real time the images coming from the robot (robot's point of view) to assemble and mix overall images on a large screen managed by several computer controlled video projectors. He also planned and controlled in real-time the movements of the robot with the goal of causing proper movements of the images projected by the video projector on the robot. The movement of the robot is therefore amplified by its images projected on the large screen. That is, the robot, by its movement, causes the on-board video projector a dynamic video projection: by moving, the robot caused the migration, oscillation, etc. of fragments of the overall projected image from a part to the other in the overall large screen. In this first part of this episode, the robot is a sort of passive companion of the dancer: it follows the dancer and is controlled by the director, and captures in real-time video images of the dancer from an on-stage perspective. In this starting part of this episode, the movements of the robot are radio controlled from the director: the only degrees of freedom of the robot consisted of low-level reactive processes able to avoid collisions (in case of error from the human) and possible "interpretation" parameters (i.e. knobs) controlled by the director to add "nuances" to its movement (smoothness, directness, etc.). In this first part, the robot had the electric power cable attached, so it was not completely wireless. Its wireless part concerned sensors, audio, and video links. At a certain point, the dancer plugged-off the electric power cable of the robot, as a special important gesture. This automatically caused (i) the activation of an extra internal battery pack (to feed the on-board videoprojector) (ii) the link of the robot videocamera to its on-board video-projector (that is, the robot videocamera signal is not any more available to the director of the performance), (iii) the starting of a semi-autonomous behaviour of the robot, (iv) the activation of the wireless sensors on the body of the dancer. In this way, using the words of the director Ezio Cuoghi, "the robot comes to life". A deeper and more interesting dialogue therefore started between the dancer and the robot.

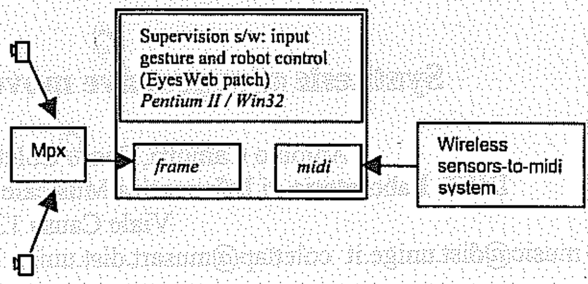


Figure 2. System architecture for movement analysis and for the "virtual mirrors" episodes, including our proprietary Mpx and Wireless-sensors-to-midi (see http://musart.dist.unige.it)

The dancer was equipped with two sensors (on the palms of the hands, whose signal is transmitted by a wireless radio link to the supervision software). By acting on the first sensor he was allowed to influence the robot toward one between two different styles of movement: a sort of "ordered" movement (aiming at a straight line constant speed movement) and a "disordered" movement. Through the second sensor the movement could be stopped and restarted. The dancer was also observed by a videocamera. Movements and gesture of the dancer then became further stimuli to the robot, which was able to react by changing (morphing) its style of moving. The language of the dancer to communicate with the robot was quite simple: intentional stimuli (start/stop the robot, positive/negative stimuli), and dancer's movement overall characteristics such as energy, equilibrium, and a number of time and space Laban's Effort-like parameters (Camurri, in press). Furthermore, the robot projected around in the environment (walls) the images coming from its on-board camera. In this way, the dancer was able to interact not only with the robot itself, but also with the projected images. Also in this case, even small movements of the robot were amplified by the changes in the visual feedback projected from the robot. The free space between the dancer and the robot is perceived as a sort of an invisible "elastic": in other words, the robot and the dancer became a sort of a whole entity.

2.2 The real and virtual dancer
Another episode concerned the real-time interaction between the dancer and its virtual clone visually perceivable on a large video screen. The dancer interacted with a computer generated silhouette changing over time its graphical appearance according to the style of movement of the dancer. In case of more dancers, an overall, whole clone corresponding to the whole group appeared.

2.3 The Virtual Mirror
This episode was based on our multiplexer hardware (see http://musart.dist.unige.it) developed for real-time multiple videocamera signals synchronised analysis. The effect was similar in some aspects to the video artists Vasulka's virtual mirrors performances. The dancer was

also doubled in real-time by this special system based on synchronised multiple videocameras.
Figure 2 shows the overall hardware architecture.

3. DISCUSSION

The main emerging issues from our experience in "L'Ala dei Sensi" are summarized in the following points.

3.1 Expressive movement or expressive interaction?

That is, in an artistic performance involving a robot moving on stage, can expressiveness arise from the movement of the robot itself or is it needed the interaction of the robot with one or more human interpreters? In fact, from the performance in "L'Ala dei Sensi" it resulted that a set of quite simple movements such as the "ordered" (near straight line constant speed movement) or the "disordered" movements were judged expressive and artistically interesting when performed by the robot interacting with the dancer and not by the robot alone. The dancer's ability to create an expressive performance conveyed expressiveness to robot movements. Therefore, it seems that the focus of future research should concern interaction and the mechanisms that make interaction expressive rather than on searching for expressive styles of movement. However, we have to remark again that the styles of movement utilized in these initial experiments were quite simple. An hypothesis is that further studies on expressiveness in human movement, taking into account the results obtained so far on the analysis side (Camurri, in press) together with the application to expressive movement synthesis of theories such as Laban's Theory of Effort, and with the utilization of more versatile robotic platforms, can lead to more effective and expressive styles of movement. This might allow a robot to move itself on a stage, alone or with other agents (robots, humans, virtual clones), conveying anyway in itself an expressive and artistically interesting content to the audience.

3.2 Synthesis of expressive movement in a multimodal perspective

The experiments carried out in "L'Ala dei Sensi" were mainly devoted to the synthesis of expressive movement and expressive interaction in a multimedia performance. The movement features analysed in real-time from the dancer were intentionally only a subset of those available from our system. As well, the synthesized movements of the virtual characters were very essential. This was a our choice due to (i) the need for a measurable experimental setup, and (ii) the resources available for the project. These aspects can be enhanced in future projects, with more complex interaction between real and virtual dancers. Further, the relationships between gestural and musical channels still need research efforts. Open problems concern for example extending the models for integrating expressive movement synthesis and synthesis of expressive content in other modalities like music and visual media. Furthermore, the mapping strategies of expressive multimodal input onto multimedia output, and in particular onto the synthesized movement, are another crucial research issues. A deeper discussion of these aspects can be found in (Camurri, in press).

3.3 Expressive autonomy

If a robotic agent is involved in an artistic performance can it make autonomous decisions? That is, has the robot to carefully follow the instructions given by the director, the choreographer, the composer, or is it allowed some degree of freedom in its movement?
This question can be extended to agents in general, including visual and music agents.
In the first section of the robot-dancer interaction of "L'Ala dei Sensi", the robot was an interpreter (that is, with some expressive content) of a predefined score of movements designed by the director. A number of rehearsals was devoted to obtain what the director wanted. Further, in most music and theatre performances, the performer's expressiveness is directed to transmit the expressive content that the director or the composer intended to communicate.
In general, an agent involved in a performance can have different degrees of *expressive autonomy*. We define as expressive autonomy the amount of degrees of freedom that a director, a choreographer, a composer (or in general the author of an application including expressive content communication), leaves to the agents in order to make decisions about the appropriate expressive content in a given moment and about the way to convey it. Expressive autonomy is therefore somewhat different with respect to *autonomy* as intended in Artificial Intelligence and Robotics: in fact, expressive autonomy doesn't concern the amount of built-in knowledge the agent contains nor its capabilities to make decisions on its own on the basis of the feedback coming from its physical sensors.
For example, if the agent has to strictly observe the directives given by a director, then it will result not autonomous from the point of view of expressiveness. That is, it is not allowed to make an autonomous decision about the expressive content to convey to its audience: such decision is made by the director. In the case of a robot, if during a performance the robotic agent is asked to perform in an effective and expressive way a sequence of movements (and possibly music and visual media) that the director predisposed and that has been repeatedly tuned and trained with the system during a number of rehearsals, then the robot is only minimally expressively autonomous or it is not expressively autonomous at all.
Of course, this is not always the case: if, as it happens for performers, some degrees of freedom are still present and the agent is flexible, versatile and rational enough as an interpreter, intervening when necessary to add nuances to its behaviour coherently with the performance, then the agent is expressively semiautonomous. That is, it plays the role the author or the director assigned to it, but it can still make decisions for example about the way of conveying the expressive content. For example, an expressively semiautonomous robot could choose what behaviour or, in particular, what style of movement it can show in order to

appear happy, in a part of a performance during which the director wants the robot to appear happy.
A high degree of expressive autonomy can be found for example in museum environments where a robot could be asked to play the role of guide for visitors or as a visitor itself (Camurri and Coglio, 1998). In such a case, although the robotic guide has to follow a narrative thread, however it can choose what expressive content to convey in order to increase the interest of its audience: the author of the application builds a narrative structure and process, and the agent is assigned the task to instantiate/interpret it in a suitable way given its current audience and context. The current degree of expressive autonomy, however, can depend on the structure and dynamics of the narration and can vary over the time during the visit.
Complete expressive autonomy implies that in a given moment the agent is completely free to choose the expressive content it wants to convey (i.e. the expressive content it judges to be more suitable given its current perceptions) as well as the way to convey it.
The requested degree of expressive autonomy is crucial when we deal with the implementation of the agent. In fact, a higher degree of expressive autonomy requires the agent to have more sophisticated capabilities in order to make its expressive choices. Thus, while the design and the implementation of an agent with limited degrees of expressive autonomy can result quite simple, a high expressively autonomous agent could need to be equipped with different kinds of components, such as components able to recognize the expressive content communicated by people interacting with it, components embedding artificial emotions and personality models, components providing rational capabilities able to make decision on the basis of the current goals of the agent.

3.4 Composition models: musical information organisms, agents, and physical interaction

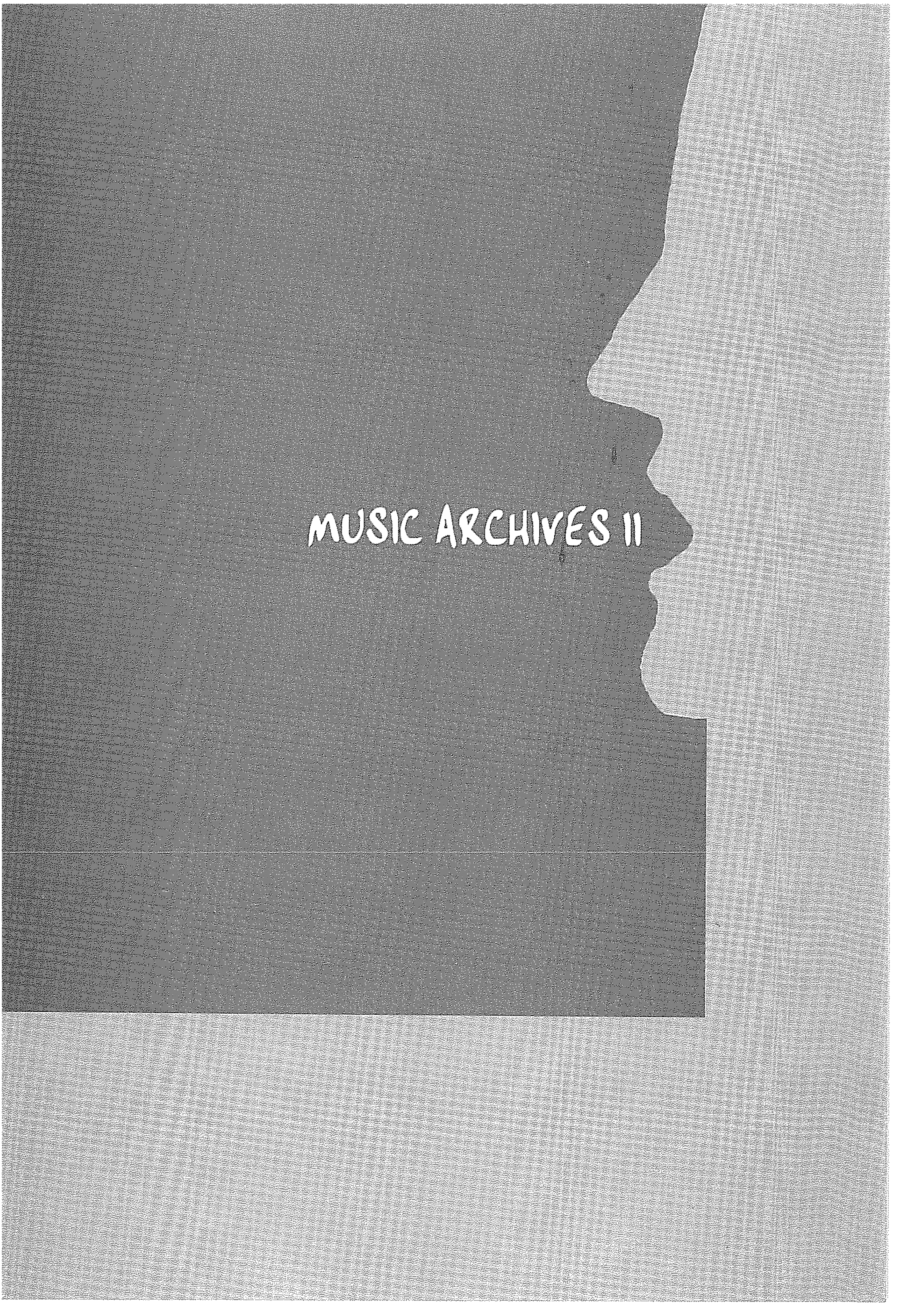
The previous considerations can be applied to the expressiveness of *music agents*. In (Camurri et al, 1994) we defined a cognitive model and an environment for composition and performance (HARP) based on agents and metaphors of energy fields. The use of metaphors as a "glue" between modalities, i.e. a means unifying the different languages involved (music, gesture, visual) was also one of the main motivation issues in HARP and now in EyesWeb.
Several composers are facing composition from similar perspectives, as discussed in (Camurri, in press). For example, Marco Stroppa's *Musical Information Organisms*, which can be represented at the microstructure level as *living being* and at the macrostructure level as *sources of energy*, and Gerard Grisey's *etre vivant* musical objects.
In our model of an interactive environment, such "organisms" (musical or, more in general, multimodal) should also be considered in an open-world perspective, i.e., to comprehend their real "sensors" and "actuators". That is, **the definition of a organism shoud be *situated* in the real-world (the real or virtual stage)**: an organism exists if it can act in the world it inhabits, and our proposal of model of organisms include real sensors and effectors. Organisms communicate either directly (e.g., Marco Stroppa's energy fields) or through the shared real-world, populated by performers, dancers, robots. Robots, environmental sensors and actuators become the sensors and the effectors of the perceptual and motor systems of such Musical Information Organisms.
This our last idea is one of the most intriguing issues on which our current research on interaction models is grounded.

ACKNOWLEDGEMENTS

We thank the Director Ezio Cuoghi, the Choreographer and Dancer Virgilio Sieni for the interesting discussions during the preparation of "L'Ala dei Sensi", and Andrea Ricci for his support on the EyesWeb software.

REFERENCES

Camurri, A., A.Coglio. 1998. "An Architecture for Emotional Agents." IEEE Multimedia, 5(4):24-33, Oct-Dec 1998, New York: IEEE Computer Society Press.
Camurri, A. In Press. "Music Content Processing and Multimedia: Case Studies and Emerging Applications of Intelligent Interactive Systems." Journal of New Music Research, 28(4), Swets&Zeitlinger.
Camurri, A., P.Ferrentino. 1999. "Interactive Environments for Music and Multimedia." Multimedia Systems Journal. Special issue on Audio and multimedia, 7:32-47, Association for Computing Machinery (ACM) and Springer.
Camurri, A., S.Hashimoto, K.Suzuki, R.Trocca. 1999. "KANSEI analysis of dance performance". In Proceedings of the 1999 IEEE Intl Conf on Systems Man and Cybernetics SMC'99, Tokyo, New York: IEEE Computer Society Press.
Camurri, A., R.Trocca, G.Volpe (1999). Full-body movement and music signals: an approach toward analysis and synthesis of expressive content. Proc. *Intl. Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop*. CEC-I3, Siena, October 1999.
Camurri, A., M.Frixione, C.Innocenti (1994). A Cognitive Model and a Knowledge Representation Architecture for Music and Multimedia. *Interface - Journal of New Music Research*, Vol.23, No.4, pp.317-347, Swetz & Zeitlinger, Lisse, The Netherlands.
Camurri, A., Hashimoto, S., Ricchetti, M., Suzuki, K., Trocca, R., Volpe, G. (2000) EyesWeb – Toward gesture and affect recognition in dance/music interactive systems. Computer Music Journal, 24(1), MIT Press.
Suzuki, K., Camurri, A., Ferrentino, P., Hashimoto, S. 1998. "Intelligent Agent System for Human-Robot Interaction through Artificial Emotion." Proc. IEEE Intl. Conf. On Systems Man and Cybernetics SMC'98, New York: IEEE CS Press.


MUSIC ARCHIVES II

A Multimedia System for Automatic Recognition, Indexing and Retrieval of Heterogeneous Music Documents

Alessandro Argentini
Lab. MIRAGE – Univ. di Udine, via A. Diaz, 5 34170 Gorizia

Carlo Combi, GianLuca Foresti, Claudio Mirolo, Angelo Montanari, Adriano Peron
Dip. Matematica e Informatica – Univ. di Udine, via delle Scienze, 206 33100 Udine

Abstract

This paper addresses a few problems related to transferring into the digital domain the information of multimedia documents. Indeed, it is desirable to realize archives easier to access and allowing a more widespread fruition of the contents. In this respect, the problem of automatizing as much as possible the transfer and retrieval operations is crucial. After a brief sketch of the work-in-progress activities being done in cooperation by Computer Scientists and Musicologists, some preliminary results will be presented in the fields of multimedia databases, optical music recognition and preservative transfer of sound documents.

1. Introduction

The transfer into the digital domain of the *information* of heterogeneous multimedia documents, such as texts, images and sounds, is intended to achieve two main goals. First of all, it is desirable to realize archives endowed with easier and faster access capabilities, as well as allowing a more widespread fruition of the information contents without risk of damage for the original, possibly fragile, sources. Second, it can support further information processing, such as that required in order to answer complex queries coming from domain-specific users and amateurs, or in order to automatically produce related documents (e.g., music transpositions, Braille scores, etc.).

In particular, to design a useful digital archive, it is necessary to address the following problems: automatic systems for placing material in the archives are needed, because the operation would otherwise be too long and costly to do manually; suitable representations have to be defined to capture and interrelate all the relevant items of musical, graphic and textual information; more efficacious ways have to be found to retrieve the desired material in answer to user requests.

This paper sketches a brief outline of the work-in-progress activities being done in cooperation by Computer Scientists and Musicologists of the University of Udine within an interdisciplinary project supported by the region Friuli Venezia Giulia. Furthermore, a few preliminary results in the fields of multimedia databases, optical music recognition and preservative transfer of sound documents.

The main purposes of the project can be summarized as follows:

- Preservative transfer of sound and paper documents onto digital support.
- Exploration of paradigms to design on-line music archives suitable for high-level inquiry.
- Exploration of new approaches to the optical recognition of music scores.
- More widespread dissemination and integration of musical knowledge and resources.
- Development of tools for useful multimedia consultation and interaction.
- Study of techniques to formalize and answer temporal queries in the field of music.

A reason for preserving digital copies of musical sources lies not only in the danger of losing the information conserved in traditional analog form, but also in the cultural and educational potential of extending the fruition of archives and libraries to a larger community of users. Unfortunately, the technology is not yet available for reliable systems of recognition, because of the high percentage of error in the reading and conversion of the musical information, which compels us to manually insert the data, with obvious expenditure of resources and energy.

2. State of the art

A critical problem that the sound archives face world wide is the disproportion between the amount of material to be properly documented and the available staff resources. Experience tells that the digital transfer and basic accessioning processes require at least the triple time of the duration of original sound documents. Only specialized institutions, therefore, can afford to aim at setting up structured data bases. Several institutions, therefore, would have a vital interest in all technologies leading to a more detailed segmentation of the archived material without increase of time for accessioning. Such technologies would, of course, also favorably assist in the retrieval of already archived material by researchers. Today the technology offers powerful devices for audio processing (24 bit A/D converters, 192 kHz samplers, 8,5 GB optical supports). Such systems, however, need correct protocols for a good exploitation, in order to avoid possible damages or, even worse, the loss of the original document, which is often available in a single copy of bad quality, [3, 13]. Unfortunately, the continuous technological evolution hinders the definition of standards for A/D conversion, preservation strategies, restoration, document manipulation.

The field of multimedia database systems is an emerging research area, [5, 8, 16, 6], whose topics of interest include: conceptual design, user interface, data models and logical design, query languages, retrieval and indexing techniques, data compression. An increasing interest in multimedia musical databases is witnessed by a number of recent contributions, e.g., [17, 10]. The conventional information-retrieval techniques are not sufficient in the case of multimedia information. In these cases, indeed, it is unreasonable to expect a sharp correspondence between the access-key and what is found. More sophisticated access techniques are therefore needed, which can somehow measure the degree of resemblance of non-textual information. Other problem are the size and complexity of non-textual document and the associations among heterogeneous information items (e.g., scores and recordings).

The digital processing of documents has until now been principally concerned with the recognition of textual characters (OCR), and only recently has the problem of the automatic recognition of music (OMR) been faced. In order to automatically recognize music scores, the main problem is the reliability of the systems, which may be improved both by developing suitable image-processing tools (segmentation, classification, etc.) and by exploiting *a priori* musical knowledge in a few refinement steps based on feed-back information. At present, automatic systems for the optical recognition of music scores are unsatisfactory because of the high percentage of errors and because of the lack of systematic information on their occurrences: also inconsistencies that are easy to detect are not reported by the commercial systems. Specific work in optical music recognition and related projects include, e.g., [1, 11, 12, 15]. Languages for music representation are reviewed and discussed in [14].

3. System Design

In this section we describe the design of the main components of the system.

3.1. Multimedia database

In the outlined framework, the main purpose of the design of a multimedia database is the integration and manipulation of textual data, sound documents and music scores (in symbolic form, whenever possible). In our view, in order to relate sounds, images, and textual information it is necessary to enrich the database with multimedia features (see Fig. 1). To address the issues concerning the representation and storage of multimedia information, we consider the following conceptual and logical data-modeling levels:

Segmentation and indexing techniques. Sound data coming from standard recordings (e.g., vinyl, tapes, CD) are often stored as an uninterpreted byte stream. A first operation is the segmentation of the raw data into basic items, which is done by identifying groups of basic items and by organizing them into meaningful subparts within the stream [8, 7]. Another fundamental issue which plays a major role in musical databases is the definition of non-standard indexing techniques aimed at retrieving the documents. The search can be based on keywords and/or more complex objects, obtained by standard (manual) classification, as well as by correspondence to given musical patterns, such as themes, rhythmic sequences, harmonic figures, etc., which exploit multi-dimensional indexes [4]. Such indexes are to be assigned to the documents by automatic or semi-automatic techniques. Sequences coming from different streams may also be arranged to produce new sound documents where suitable subparts are shared within the original streams.

Temporal dimension of sound data. The valid time of a fact is the time when the fact is true in the modeled realm [9]; the valid time of a sound data item is the time when the represented sound is to be listened. As for other kind of data, valid time of sound data could be given with different granularities (i.e., different time units) or with uncertainty [7]. Another temporal aspect that has to be considered is related to the *intrinsic* time of sound data: a CD track, for example, specifies a duration (a number of seconds) and it must be played according to the correct number of samples per second.

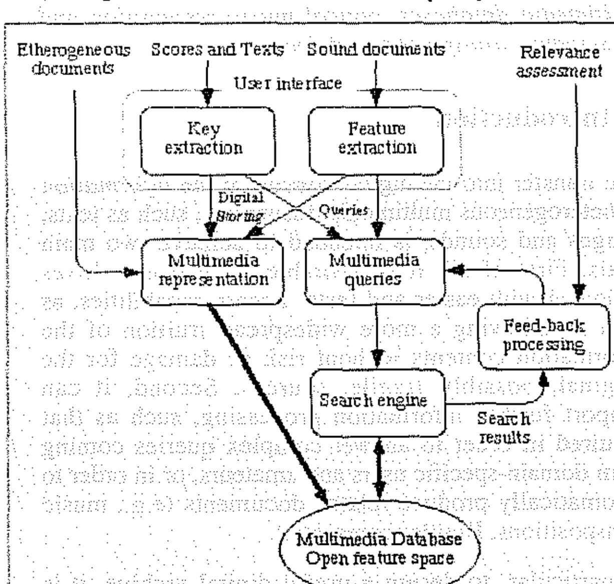


Fig. 1 – Main functional modules of the multimedia archive and their interrelationships.

Multimedia data integration. Sound documents are often associated to real world entities with some kind of (structured) textual description stored in the same database. It is important to allow the user to relate textual descriptions to (subparts of) sound data items, and to structure each textual description in a suitable way, according to its content [8]. A more difficult problem is related to the synchronization of scores and related sound pieces: in this case, the integration must be achieved by considering relationships between spatial locations in scores and the corresponding temporal position within the sound document.

Temporal relationships between text and sound. A further temporal aspect while modeling multimedia data is introduced by the fact that the textual information can be *temporally* related to the sound data in several different ways, according to the meaning of the textual data. For instance, suppose we are interested in a music sequence where a particular instrument, say a *cello*, plays a "*crescendo from pianissimo to fortissimo*", [8]. Then, we know that two properties hold for the whole sequence: (i) the cello plays and (ii) what is being played is a "*crescendo from pianissimo to fortissimo*". On the other hand, whereas property (i) also holds for any part of the sequence, it is no longer true that property (ii) holds for any piece of sequence too [7].

It is worth noting that some of the above issues, precisely those related to different temporal aspects in multimedia databases, have not yet been satisfactorily addressed in the literature on databases. Nevertheless, they can be addressed by defining a temporal multimedia object-oriented model, which can be implemented by means of object-oriented and object-relational database technologies, such as the Java language extended to manage persistent objects and the database management system DB2 by IBM.

3.2. Automatic music recognition

A consolidated approach to the recognition of a great variety of documents can be decomposed into three phases (see Fig. 2):

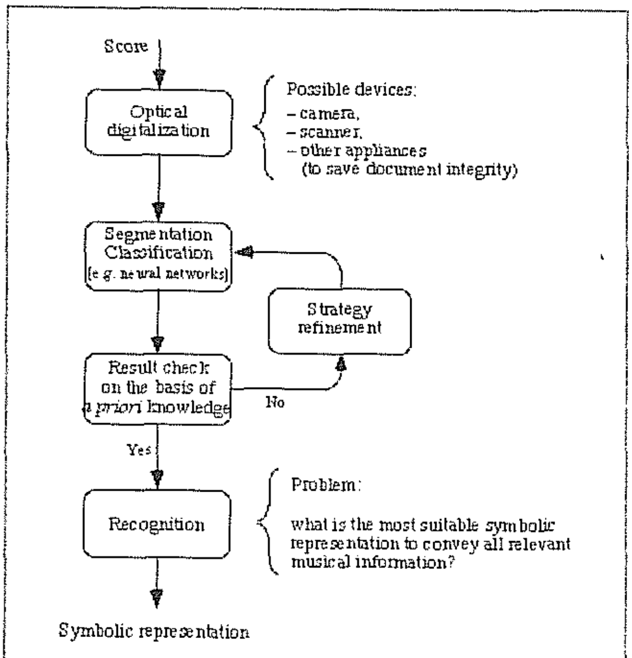


Fig. 2 – Main functional modules for the optical recognition of music scores.

1. *Image acquisition and preprocessing*: the source (paper) document is converted into a digital image by using either a scanner or a camera; then the image is processed by special purpose tools (pre-segmentation).

2. *Image segmentation*: the elementary parts that best represent the content of the document are recognized (musical notes, characters, other graphic symbols). This is an extremely critical phase in recognition.

3. *Recognition of the musical notation*: the elementary parts are processed and classified by a recognition module, e.g., based on pattern recognition or neural networks. The outcome of this processing phase can then be interpreted at a higher level and any inconsistencies w.r.t. the available *a priori knowledge* can be exploited in a feed-back loop in order to improve the results of the previous steps.

During our initial work, we have mainly been concerned with the design and development of segmentation techniques (stage 2) and with a preliminary recognition of a few symbols of the musical notation (stage 3). More specifically, here are the main steps on each music page:

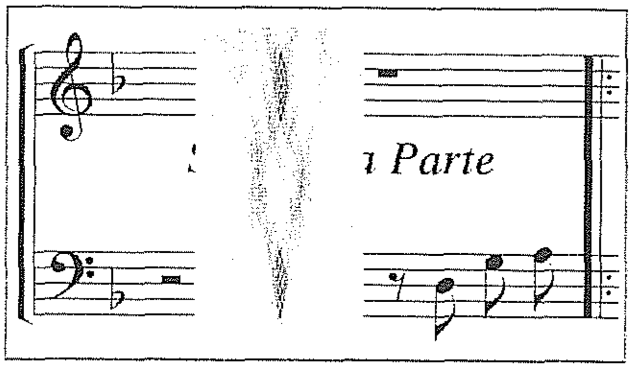


Fig. 3 – Stave recognition in the *Hough* space.

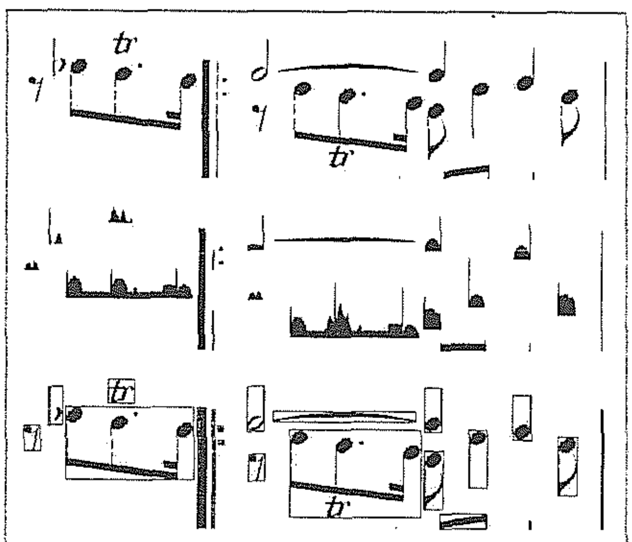


Fig. 4 – Horizontal projection and bounding boxes.

a) *Stave recognition* via Hough transformation (Fig. 3): an important step in order to get a reference frame.
b) *Stave removal*, to focus on the music symbols.
c) *Vertical segmentation* of the staves.
And then, after focusing on a bounding box within a single stave or bar unit (Fig. 4):
d) *Recognition of clefs*, by horizontal projection plus specific techniques.

e) *Recognition of bar lines*, by horizontal projection plus Hough transformation.
f) *Recognition of note stems*, by repeated horizontal and vertical projections plus Hough transformation.
g) *Recognition of note heads*, by repeated horizontal and vertical projections plus model matching.

To sum up, Table 1 summarizes the percentage of success obtained with the chosen segmentation techniques and provides a comparison with the results of a commercial product. Although this stage is still preliminary and only few components of the OMR system have been developed, the results encourage to study the considered tools in more depth.

	% Success			Commercial product		
	found	*missed*	*misread*	*found*	*missed*	*misread*
treble clefs	100%	—	—	100%	—	—
bass clefs	100%	—	—	100%	—	—
bar lines	92%	8%	6%	100%	—	—
note stems	98%	2%	2%	96%	4%	—
note heads	98%	2%	5%	98%	3%	—

Table 1 – Experimental results and comparisons.

3.3. Other aspects of musical interest

Besides the technical problems that have been outlined thus far, other activities are:

Preservative transfer of important sound documents. It is important to guarantee that the transfer process be carried out according to the methodologies adopted by the international archivist community.

Classification of the musical documents. Contributions of specific musicological interest are: the classification of the typographical kinds of musical notation; the formalization of the syntactic rules of musical writing; the classification of the different sources from the historical and the stylistic perspectives.

Philological analysis. A well organized digital archive would allow an easier comparison of a modern reissue with the original document, or a fast access to the related historical information (recognition of missing parts, cuts and interpolations). Moreover, different performances of the same work could be compared (length of internal sections, range and variation of dynamics, change of the perception of music in different historical periods).

Musicological analysis. The system should allow the user to define specific data links, depending on the focused contents. Although the musicological problems are rather difficult, a perspective work should also consider possible ways to give support to such activities as: search for identity or similarity, application of morphological models on 18th and 19th century music, semiological analysis, e.g., of the fragmentation of the musical *continuum* (including the contemporary non-tonal music), [2].

Conclusions

In this paper we have outlined the main topics and the work-in-progress within an interdisciplinary project. The results we have presented are only preliminary, but provide valuable hints for future developments. We are now studying in more depth the role of time granularity in musical databases and possible data fusion techniques to recognize more precisely the musical symbols.
Acknowledgments. We acknowledge the "Regione Friuli Venezia Giulia" for its financial support. We also want to thank Matteo Macoratti and Stefano Zuccolo for their contributions.

References

[1] Bainbridge, D., Carter, N., *Automatic Reading of Music Notation*, Handbook of Char. Recognit. and Doc. Image Analysis, Bunke & Wang Eds., World Scientific, 1997.
[2] Bent I., Drabkin W., *Analysis*, Macmillan Publishers, LTD, London, 1980
[3] Calas, M., F., Fontaine, J. M., *La conservation des documents sonores*, CNRS Editions, Paris,1996.
[4] Combi, Foresti, Franceschet, Montanari, *Indexing by Shape of Image Databases Based on Extended Grid Files*. IEEE Int. Conf. on Multimedia Compu. and Systems, 1999.
[5] Chu, Hsu, Cardenas, Taira: *Knowledge-Based Image Retrieval with Spatial and Temporal Constructs*. IEEE Trans. on Knowledge and Data Engineeering, 10(6), 1998.
[6] S. Christodoulakis, L. Koveos, *Multimedia Info. Systems: Issues and Approaches*, in W. Kim (ed): Modern Database Systems, ACM Press, 1995, pp. 318-337.
[7] C. Combi, *Modeling Temporal Aspects of Visual and Textual Objects in Multimedia Databases*, Goodwin, Trudel (eds.) Proc. 7th Int. workshop on temporal representation and reasoning, TIME 2000.
[8] J.D.N. Dionisio, A.F. Cardenas. *A Unified Data Model for Representing Multimedia, Timeline, and Simulation Data*. IEEE Trans. on Knowledge and Data Engin., 5, 1998.
[9] C.S. Jensen, C.E. Dyreson, et al., *The Consensus Glossary of Temporal Database Concepts*, in: Etzion, Jajodia, Sripada (eds.), Temporal Databases - Research and Practice, LNCS 1399, 1998, pp. 367--405.
[10] C. Liu, J. Hsu, A. Chen: *Efficient Theme and Non-Trivial Repeating Pattern Discovering in Music Databases*. Proc. of the 15th Int. Conf. on Data Engin., 1999.
[11] Ng, K.C. & Boyle, R.D., 1996, *Recognition and Reconstruction of Primitives in Music Scores*, Image and Vision Computing, 14, pp. 39-46.
[12] Reed, K., Parker, J., 1996, *Automatic Computer Recognition of Printed Music*, Proc. ICPR96.
[13] Schüller, D., *The Ethics of Preservation, Restoration, and Re-Issues of Historical Sound Recordings*. JAES, vol. 39, n.12, 1991, pp. 1014-1017.
[14] Selfridge-Field, E., Ed., *Beyond MIDI: the Handbook of Musical Codes*, MIT Press.
[15] Stueckelberg, M.V., Pellegrini, C., Hilario,M., 1997, *An Architecture for Musical Score Recognition UsingHigh-Level Domain Knowledge*, Proc. ICDAR97.
[16] Subrahmanian, V.S., 1998, *Principles of Multimedia Database Systems*, Morgan Kaufmann, San Francisco.
[17] C.L. Yip, B. Kao: *A Study of Musical Features for Melody Databases*. In: Bench-Capon, Soda, Min Tjoa (eds.) Proc. of the 10th Int. Conf. on Database and Expert Systems Applications, LNCS 1677, 1999, pp. 724-733.

A Novel Methodology for Music Information Retrieval

Massimo Melucci
DEI - University of Padova
Via Gradenigo, 6/A - 35131 PADOVA
melo@dei.unipd.it

Nicola Orio
IRCAM - Centre Georges-Pompidou
1, pl. Igor Stravinsky - 75004 PARIS
Nicola.Orio@ircam.fr

Abstract

This paper presents a novel methodology for content-based indexing and retrieval of music documents, and a prototype implementing the methodology. The main assumption of our work is that classical techniques for textual information retrieval can be applied also to the music domain, providing that they should be tailored to this specific medium [1], [2]. In particular, the problem of identifying the "lexical units" of the music language is addressed, together with the problem of dealing with variants in these lexical units. We proposed a methodology for fully automated processes, which provide: Segmentation of melodies in their lexical units that we called musical phrases, normalization of musical phrases into a common stem, indexing, retrieval, and construction of an hyper-media with links among music documents and phrases. We developed a prototype system based on our technology, which is available on the web. The prototype allows the user to query the system through performing a short melody excerpt and to browse the collection following links of "hyper-music".

Introduction

The growth of Web and the advent of search engines have been letting researchers in many disciplines to use and know information retrieval (IR) technologies since early nineties. However, research work in IR and operational IR system implementation dates back to early fifties when computers allowed to automatically process abstracts and bibliographic records. The main aim of a IR system remains the same – to retrieve in real-time the very small set of documents being relevant to user's queries from a very large collection of unstructured documents. Retrieval is content-based since the matching between queries and documents takes place on the basis of their own semantic content.

The reason why IR techniques can be used for music retrieval is the fact that IR has dealt with (multimedia) content-based description and searching since its early days. However, even if IR techniques may usefully be extended to different media, they should be tailored to any specific medium, and integrated with its specific processing techniques. The latter applies to music as well - hence, the combination with automatic music processing techniques is necessary. As a consequence, the potential for applying some of the standard principles of text information retrieval to music representations has been investigated, as proposed in [3]. We present a novel methodology for content-based indexing and retrieval of music documents, and a prototype implementing the methodology.

The Methodology

Our methodology is based on the assumption of a duality between the two domains of textual and music IR. In particular we suppose that, in both domains, it is possible to identify the lexical units of the respective languages that allows to describe the document content. While it is clear what is a lexical units in text, i.e. a *word*, and which are the separators between lexical units, e.g. *blanks*, *commas*, or *dots*, the same does not apply to music.

Music lacks of explicit separators. On the other hand, musicologists, like [4] and [5], proposed a number of theories on music structure, which are based on the segmentation of the continuous flow of events that constitutes the melody of a composition. Through the application of these models, it can be then possible identify the lexical units of music and introduce separators. Once lexical units of the melody, which we called *musical phrases*, are identified it is also possible to go further with the parallelism and explore the analogous of "stemming" in textual IR. Since musical phrases may appear with many variants, as in textual documents the same root may result in different words (i.e., from "muse" we have "music", "musician", "musicologists" and so on), we propose a method for conflating, through normalization, similar musical phrases into a common stem. After segmentation and normalization, the duality can be utilized by directly applying techniques developed for textual IR to music IR, providing that segmentation and normalization have to be applied both on music documents and on music queries. The process is depicted in Figure 1.

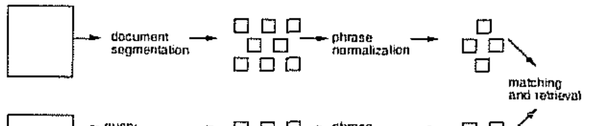


Figure 1: An illustration of the methodology

Our methodology consists of three main automatic methods – segmentation and normalization, indexing and retrieval, hyper-music construction. Automatic segmentation takes place on machine-readable notation-based scores which are transformed into content descriptive lexical units. Normalization conflates lexical units into a common stem. Segmentation and normalization take place on both whole music documents and final user's music queries. The role of musical lexical units in documents and queries is the one taken on by keywords of textual documents and queries. Indexing of documents creates music indexes being used at retrieval-time to match queries against the documents themselves. Hyper-music construction creates links among documents and lexical units to allow the final user browsing a hyper-media being made of music documents and queries. The *prototype* that implements the methodology was called SMILE, which stands for a *S*ystem for *M*usic *I*nformation retrieva*L* *E*nvironments.

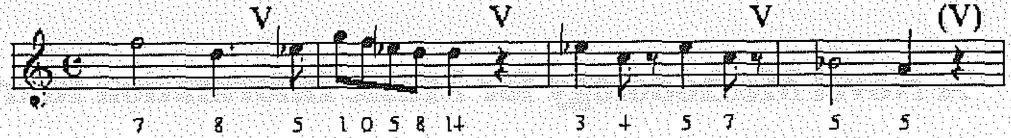



Figure 2: Mozart's Concert for Clarinet K622, for each note the respective weights are quoted

Melodic Segmentation for Automatic Music Indexing

Among all the dimensions that characterize music (melody, harmony, rhythm, structure, and so on), melody seems the most suitable for non-expert final users. In our methodology melody is used as the source of the content descriptors of a music document. But the melody, both in the notation and recording of a performance formats, lacks of explicit separators as text does; an example of a melody is reported in Figure 2. Anyway, there are a number of cues which musicians and listeners use to highlight the presence of the beginning (or the end) of a melodic surface. In our methodology the process of segmentation of melodies in their lexical units, which we called musical phrases, is performed by a totally automatic algorithm.

To perform the segmentation of melodies in lexical units, that is in musical phrases, it has been used the Local Boundaries Detection Model (LBDM) proposed by Cambouropoulos [6]. According to this model, listeners perceive the change from a musical phrases to another one, that is the presence of boundaries in a melodic surface, whenever there are some changes in the relationship among the notes. Hence a musical phrase can be bounded by a note with a duration, or which forms an interval, different from the surrounding notes. The LBDM detects the boundaries of a melodic surface by giving a weight to all the notes in a melody; the weight is related to the probability that a boundary may occur after that note.The weight values are evaluated depending on:

- the relationship among the *musical intervals*, which each note forms with the previous and the subsequent notes. If intervals are different – e.g.. a fifth versus a third – then a given value is added to the two weights – for example, 2 for the larger interval and 1 for the smaller one;
- the relationship among *note durations*. If durations are different a given value is added to both intervals – for example, 4 and 1 if the first note is longer than the second one, or 3 and 2 if the first note is shorter, respectively;
- the presence of *musical rests*. In this case a value of, for example, 4 is added to the weight.

Notes which have the higher weights, that is note which weight is a local maximum, are selected as boundaries in the melodic surface. Referring to the melody in Figure 2, for each note the respective weight is quoted and boundaries corresponding to local maxima are highlighted by a "V" sign. In this way it is possible to introduce separators of lexical units, like in textual documents. It is hence possible to apply methods developed for textual IR, considering the detected melodic surfaces as the analogous of terms in textual documents.

Normalization of Musical Phrases

One of the problems in the indexing of musical documents, is the presence of variants in musical phrases. Lexical units pointed out by the segmentation algorithm may not be equal but may be perceived similar by the user, that is they may carry the same content. In our methodology it is proposed to overcome partially this problem by introducing two kinds of normalizations in the representation of musical phrases. Normalization regards both note durations and note intervals. First of all it is performed a *Pitch Transposition*: all the notes are transposed, in order to have each musical phrase beginning with the same note. Then it can be performed a *Duration Normalization* – note durations are expressed in multiples of their Greatest Common Divisor, so that what becomes important is the relationship among durations. Since variants of musical phrases may regard small changes in the pitch contour, due for instance to modulations, also a *Pitch Normalization* can be performed – pitches are quantized in a number of different levels, related to the musical intervals (unison, from minor second to major third, from perfect fourth to major sixth, and so on). Finally it is possible to consider only the pitches, when *Duration Removal* is performed.

We propose to apply four different combinations of these normalizations. These combination correspond to an increasing level of generality of the musical phrases. The four combinations are: Pitch Transposition with durations unchanged, Pitch Transposition and Duration Normalization, Pitch Normalization and Duration Normalization, Pitch Normalization and Duration Removal.

Indexing and Retrieval

The method for music indexing takes place on the music documents and queries that have been transformed into text in a predefined format. The format allows for filtering different information – e.g., authorship, key and time signatures, and instrument. Hence, indexing can be applied to each of these information to produce different parallel indexes.

The IR model employed for constructing indexes and retrieving documents was the vector-space model (VSM). Accordingly to the VSM model, both documents and queries are represented as K-variate vectors of descriptor weights w_{ij}, provided K is the total number of unique descriptors. Then, document d_i is represented as $\vec{d_i} = (w_{i1}, ..., w_{iK})$, while query q is represented as $\vec{q} = (q_1, ..., q_K)$. For a document, a descriptor can be either an individual word, a sequence of words, or more complex word sequences, such as sentences. In our music domain, documents are notation-based music documents and descriptors are melodic phrases extracted via segmentation and normalization. The weight w_{ij} of descriptor j within document i can be expressed as $tf_{ij} \times idf_j$, where tf_{ij} is the frequency of descriptor j within document i, $idf_j = \log N/n_j$, N is the total number of documents, and n_j is the number of documents indexed by descriptor j. Query descriptor weights are usually binary values, then $q_i = 1$ if descriptor i occurs within query q, 0 otherwise.

The retrieval status value (RSV) is the usual cosine of the angle between the query vector and the document vector.

$$RSV(d,q) = \cos(\vec{d},\vec{q}) = \frac{\vec{d}\,\vec{q}}{\|\vec{d}\|\,\|\vec{q}\|}$$

where d, q represent the document and the query respectively, $\vec{d}, \vec{q}$ are the corresponding vectorial representations, and $||\vec{x}||$ is the norm of $\vec{x}$.

As the cosine function normalizes the RSV to the query and document lengths, the sizes of document and query have been controlled, and long documents have then the same chance of being retrieved than short documents. The $tf_{ij} \times idf_j$ weighting scheme gives higher RSVs to documents within which query descriptors occur with high intra-document frequency (tf_{ij}) and low intra-collection frequency (n_j). This means that the relevance of a document has been estimated as directly correlated to the number of times query descriptors occur within the document, and inversely correlated to the number of documents indexed by them.

As the cosine function normalizes the RSV to the query and document lengths, the sizes of document and query have been controlled, and long documents have then the same chance of being retrieved than short documents. The $tf_{ij} \times idf_j$ weighting scheme gives higher RSVs to documents within which query descriptors occur with high intra-document frequency (tf_{ij}) and low intra-collection frequency (n_j). This means that the relevance of a document has been estimated as directly correlated to the number of times query descriptors occur within the document, and inversely correlated to the number of documents indexed by them.

The RSV function can be applied also between pairs of documents since both queries and documents are represented as vectors. We can therefore compute the cosine between two document vectors to estimate the similarity between the document contents. Moreover, the RSV function can be applied also between pairs of descriptors since the latter can be represented as vectors of documents. Indeed, let us consider the $N \times K$ matrix consisting of document vectors as rows. The transposed $K \times N$ matrix consists of descriptor vectors $\vec{t_j} = (w_{j1}, ..., w_{jN})$, and therefore the cosine of the angle between descriptor vectors can be computed. Such a cosine estimates the degree to which two descriptors occur within the same sub-set of documents, and then the cosine can be used to compute a preliminary estimate of potential relationships between descriptors [7]. The latter way of computing similarities between music documents and descriptors will be the starting point for hyper-music construction as illustrated in the next section.

Hyper-music Construction

A retrieval system based allows for extracting from a collection a sub-set of documents judged as relevant for he final user information needs. However, it may be the case that the user finds very few relevant documents within the retrieved set. A method to partially overcome this problem is the construction of links between documents and phrases.

If links are available, the user can make up for the lack of relevant documents by following links from retrieved documents either to non-retrieved ones, or to phrases used by the system to index them. In the former case, the user may be able to retrieve additional, perhaps relevant documents. In the latter case, the user should be able to choose one of these phrases and then to retrieve new, perhaps relevant, documents being linked by them. Because the lack of relevant documents may occur again, the user should be able to carry navigation on through links among documents or phrases. Links have then to be constructed also between similar phrases, that is the ones used by the system to index the same documents, because the user should be able also to access to new phrases when he cannot find any adequate phrase describing his needs. The complete process encompassed by our methodology, which integrates querying with browsing, is depicted in Figure 3.

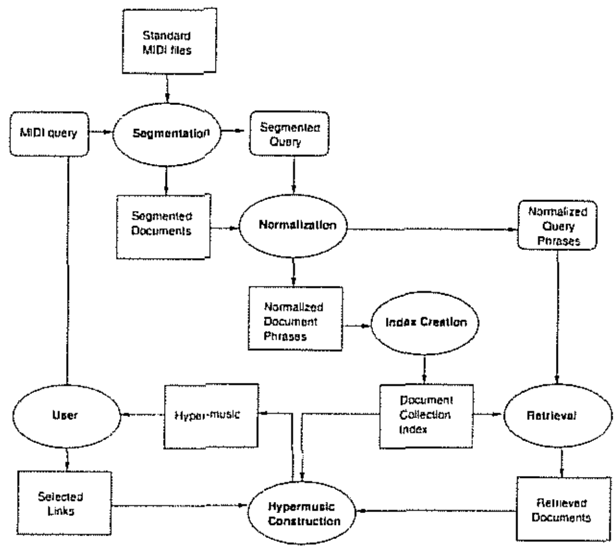


Figure 3: The process of hyper-music construction: documents are represented by "squares" and phrases are represented by "circles"

To make the navigation process possible, the construction of links among documents and phrases is required to make up the lack of relevant documents. Once these links are constructed, they can be used to set a network, called "hyper-music", of music documents and phrases, which the user can browse. The size and complexity of link construction make manual construction infeasible and then automatic methods are required. In this Section, we show how the methods for automatic melody segmentation and for automatic indexing can be combined together for automatically constructing hyper-music. The idea is based on four key considerations:

- the phrases being extracted by segmentation can be used as content descriptors like index terms do for textual documents;
- phrases can be used to build links between documents using co-occurrence information, i.e. the more two documents share the same phrases, the higher the chance of building a useful link between them;
- similarly, documents can be used to build links between index phrases using co-occurrence information;
- finally, bidirectional links can be build between documents and their most relevant phrases.

Automatic hyper-music construction hence allows for inserting four types of link – DD, DT, TD, or TT – depending on the type of linked nodes – D (document) or T (phrase). Each type of link corresponds to one possible way of expanding the initial user's music query. Links are weighted since phrases occur within documents with a $tf \times idf$-based weight, and documents (phrases) are linked to other documents (phrases) by means to the cosine-based RSV. Link weight provides with a threshold to decide whether a link should be inserted. Threshold can be either the weight itself, or the rank to which a document or a phrase is placed into a list ranked by weight. A DD (TT) link is inserted between documents (phrases) if the cosine-based measure of similarity, or the rank of document (phrase) is over a stated threshold. DT links are between a document and phrases, while TD links are between a phrase and documents. A DT (TD) link is inserted between documents and phrases if the $tf \times idf$-based weight, or the rank of the linked phrase (document) is over a stated threshold.

A Prototype for Music Information Retrieval

We developed a prototype system, based on the designed methodology for content-based music IR. We called it SMILE, which stands for "a System for Music Information retrievaL Environments". It implements all the features of the proposed methodology. SMILE runs on a client-server architecture, where the client side consists of a standard Java-enabled Web browser running a Java applet, while the server side consists of a music indexing and search engines, the segmenter, the normalizer, other that the usual HTTP server. The prototype is available at `http://livia.dei.unipd.it/smile/`, and it allows the user to retrieve music document within a collection of MIDI files. For a complete example of a search path, which includes searching and browsing hyper-music, the reader is suggested to refer to [8].

The **client side** is constituted by a *Java Applet* is the starting point for searching and browsing. The user can create interactively a music query through a virtual keyboard, which can be played with the mouse. Queries can be created both recording a short performance and inserting the notes one by one. Moreover the user can tune the retrieval by choosing the kind of normalization, corresponding to four different levels of exhaustivity-specificity. The applet then converts the music score to a text notation that is used to invoke the server through a CGI gateway script. A screen dump of a browser running the applet is depicted in Figure 4.

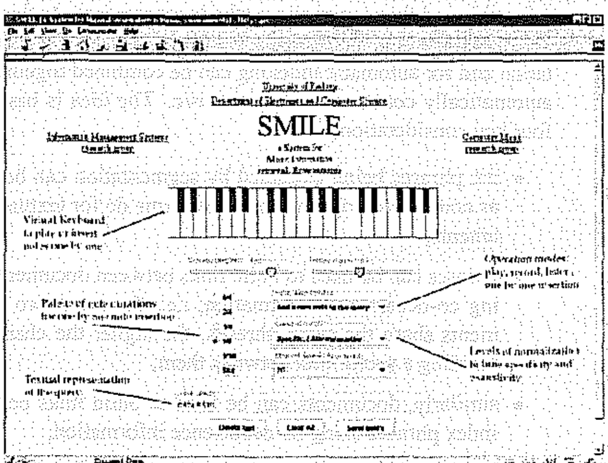


Figure 4: The music query panel of SMILE

The **Server Side** is constituted by a chain of specialized tools. The *Segmenter* takes an inner representation of a music piece and creates a list of its musical phrases. The same routine is used to segment both the documents and the query. The *Normalizer* creates four different representations, with an increasing level of generality, of the musical phrases produced by the segmenter. The phrases have different textual representations, which take into account the information of pitch and duration of the notes. The *Indexing Engine* takes the normalized textual music scores as input and produces the indexes being used at search time. Indexes were implemented as inverted files. The *Search Engine* takes the music query being played using the Java applet, runs the segmenter and the normalizer taking the query as input, and accesses the indexes to retrieve the matching music documents.

Conclusions

The proposed methodology is a step towards the design and implementation of a complete music IR system. The methodology is inherently modular because it is the result of the integration of two independent techniques – IR techniques and automatic music processing. Hence, it can be enhanced through classical techniques, such as data-based methods, to retrieve music information. The strength of this approach lies in the different retrieval capabilities of content-based techniques, and data-based techniques: While the former allow for partial match whenever the user is unable to use exact match, e.g. some musical phrases remembered by heart, the latter can provide the user with precise access means whenever he does exactly know what he is looking for, e.g. a Mozart's Sonata.

The methodology is also modular from the search point of view: The enhancement of querying and browsing can be implemented by adding new types of links from music documents or phrases to bibliographic fields, e.g. composer name or musical form. In such a way, hyper-music can be enriched through the automatic addition of new links and nodes. The new nodes can, for example, be those storing composer names or musical forms. For instance, links can be inserted between a phrase node of the T level and a node storing a composer name of a new level, by thus describing the relationship between a composer and his most characteristic musical phrases.

References

[1] Sparck Jones, K. and Willett, P. (1997). *Readings in Information Retrieval*. Morgan Kaufmann, San Francisco, CA.

[2] van Rijsbergen, C. (1979). *Information Retrieval*. Butterworths, London, second edition.

[3] McLane, A. (1996). Music as information. In Williams, M., editor, *Annual Review of Information Science and Technology (ARIST)*, volume 31, chapter 6, pages 225–262. American Society for Information Science.

[4] Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. University of Chicago Press, Chicago, MI.

[5] Lerdhal, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press, Cambridge, MA.

[6] Cambouropoulos, E. (1997). Musical rhythm: a formal model for determining local boundaries. In Leman, E., editor, *Music, Gestalt and Computing*, pages 277–293. Springer-Verlag, Berlin.

[7] Salton, G. and McGill, M. (1983). *Introduction to modern Information Retrieval*. McGraw-Hill, New York, NY.

[8] Melucci, M. and Orio, N. (2000). SMILE: a System for Content-based Musical Information Retrieval Environments. In *Proceedings of Sixth RIAO Conference on Content-Based Multimedia Information Access*, Paris.

TEACHING AND COMMUNICATING ELECTRONIC AND COMPUTER MUSIC

special session

TEACHING AND COMMUNICATING ELECTRONIC AND COMPUTER MUSIC

special session

Communicating Computer Music Memes

Dr. Mladen Milicevic, Author
Loyola Marymount University,
School of Film and Television, Recording Arts
7900 Loyola Bdoulevard, Los Angeles, CA 90045, USA
MMilicev@LMUmail.lmu.edu

What makes the primary difference between our species and all others is our reliance on cultural transmission of information, and hence on cultural evolution. Animals do interchange information but in a biological rather than a cultural context. Bird mating calls certainly fall under the category of sonicly transmitted information that is specific to a given species, but those species have limited intelligence and undoubtedly no bird culture. This does not mean that animals have no minds, it simply means that *by human standards* those "primitive" minds produce no cultural history.

Dawkins' meme has a peculiar but powerful role to play in our understanding of human culture. This is the way he defines it:

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperm or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via process that, in the broadest sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. [1]

The important rule for memes, as for genes, is that they must constantly replicate. This replication is a mindless process not necessarily for the good of anything; replicators that are good at replicating flourish—for whatever reason. Meme X spread among the people, because X is a good replicator. [2]

Let's take a moment and look at the case of one particular meme—the success of a four-note meme at the beginning of Beethoven's Fifth Symphony. Certainly, it has much less to do with the absolute value of its pitch-set "internal" design, i.e. the way a musical piece is compositionally structured, and much more to do with the design this meme presents to the world. What is important is its phenotype, the way it affects the minds and other memes in a particular socio-cultural environment.

It is very logical to assume that humans intelligently create musical pieces (potential memes) rather than producing them as random innovations. But, looking at the state of affairs in computer music, one sees more examples of innovations than creations. Let me clarify the distinction. The quality of computer music, and for that matter the quality of anything does not depend exclusively on its structural organization, but rather it is rooted in the transaction that occurs between the music and the audience. To evaluate music is to find the quality of transaction between the musical configuration and its cultural response. If that response is positive, meaning ecologically prudent for the given cultural environment, then it may be called creation. On the other hand, an innovation that is just randomly new and not holistically related to the environment, will certainly produce a negative cultural response.

In order to expand the audiences for computer music it is necessary to look at the minds of the people who constitute those audiences. Why in the minds? Because the minds are the habitats of the memes. Minds are in limited supply, and each mind has only a limited capacity for the support of memes; therefore, there is considerable "competition" among memes for entry into as many minds as possible. This competition is the major selective force in the infosphere, just as it is in the biosphere.

One way to look at music, is through the memetic competition among musical compositions for admission into the minds of the audiences. Randomly presenting new musical pieces (potential memes) to these minds can be compared to, playing by the rule of evolutionary biology known as *the survival of the fittest*. It is well known that this game is ruthless and will in its process discard the vast majority of its participants. Being the only intelligent species on this planet, it is certainly wise not to play by the rules of the very game that created us. If we continue to do so, we may easily become a casualty of the same mindless evolutionary process. For that reason, it would be preferable to turn this game around to our advantage, redefining it slightly, and call it *the survival of the wisest*. Since humans have the brain that is capable of intelligent thinking, it would be a tremendous waste not to use it in determining our own future, and the future of our music. How can this be achieved?

Wrestling with this problem may appear to be an insurmountable task, but the situation is not that hopeless albeit. For the sake of clarity, let us imagine continuation of our inquiry as an academic course in music composition. In this case, instead of teaching our students the elements of compositional structure such as counterpoint, harmony, orchestration, etc., we would teach them the basics of understanding the sonic entities towards which human minds show increased interest when listening. After the completion of such a course in music composition, our students would then be ready to learn the elements of musical structure that may represent the best tools in achieving our intended goal. In other words, students would use musical structure in order to compose pieces that will appear interesting to the musically inclined human minds of our audiences. As a consequence of this process, the successful musical pieces will have a much better chance of becoming the memes that reside inside the human minds. Once situated there, the memes will get a chance to replicate, which is their sole function, as it is the function of the genes.

It is dangerous to make generalizations, but I am going to offer the following anyway. Nobody cross-culturally educated would agree any more that music represents a universal language. However, research tangibly shows that there are some things that appear universally. In order to make sense from the vast sonic events that enter its auditory cortex, the brain had to become a master of simplification. This process is nothing like filtering unwanted information, because such a mechanism would be tremendously complicated and utterly inefficient. In actuality, the brain searches for familiar devices and patterns. [3] It latches on things that are in some respect already known, disregarding most of the unfamiliar information. The reason that human (but also animal) brain is doing this; lies in the fact that previously processed information can be very quickly reconstructed from the data stored in brain's long-term memory. Then, that reconstruction can be efficiently processed and compared with the similar incoming information, giving it the most pragmatic interpretation that fits the situation at hand.

To create a musical pattern in the human brain, it is necessary to have repetition of the sonic event in question, thus it can be remembered and used in the future. Composing a piece of music (a latent meme) which appeals to the minds of the audiences outside the narrow and highly specialized computer music niche requires the existence of some sort of a clearly recognizable musical reference. If there is a single negative point about the avant-garde approaches in music, then it is the lack of reference and use of discontinuity and disjunctness without any historical or compositional reflexivity. Human beings, in large part, will not find appealing anything which produces one unconnected innovation after another, never going back and reevaluating what has gone before in relation to what is going on now. Let me use an example to illustrate this point.

Digital sound sampling and computer technology of the nineties, readily allow computer music composers to manipulate and transform organic sounds through myriad methods and possibilities. This technological might inevitably renders countless numbers of musical compositions that deal with so-called sound exploration. Unfortunately, most of these pieces remain just that—sound explorations—and never present themselves as memes with faculty to catch on to the audiences. Why is this so? Most of these pieces suffer from a syndrome that may be called "no point of reference." What I mean is that if one is to make a computer music composition which deals, say, with sonic transformation of the sound of a baby crying, one may consider it interesting to do the following. First, it would be wise to ensure that the audience listening to such a piece can clearly discern where this sonic manipulation is coming from, and be able to hear and refer to from time to time to the original sound source. Second, throughout the piece the audience should be reasonably prepared for the direction these manipulations may take. The human brain perceives by anticipation. It formulates perceptual hypotheses and then confirms them. [3]

Thus, when the brain receives sequences of musical tones, it does what it does with other "new" information: it attempts to "interpret" it by using the "old" already processed and digested information stored in its long-term memory about previous, similar music experiences. This information may allow some aspects of a future musical signal to be anticipated—as it happens when we hear the first line of a familiar song. The ability for predicting incoming patterns of information, in our particular case musical information, on the basis of past experience is one form of what we call "intelligence"; it can dramatically enhance an organism's chances of survival. Knowing what is coming is always much more profitable than being caught by total surprise.

The way human brain handles this comparison-based process is grounded in the workings of neural mappings that correlate to a specific sonic event. For the sake of simplicity, let's say that there is a sound of an oboe playing A=440 Hz. Now, in most of the musicians' brains there is a neural correlate for A=440 Hz sound. These neural mappings are physical representation in the brain of a "formula" for the reconstruction of person's memory about an A=440 Hz. This "formula" will be put in use when the ear receives an external stimulus of an oboe playing and passes this stimulus to the auditory cortex. Then, in turn, the auditory cortex would use the "formula" to reconstruct the memory of the closest similar experience and its context stored in the long-term memory that matches the one of the external stimulus just being received. Through the process of semantically matching the new and the old mental image of an oboe playing A=440 Hz, the brain would assign the meaning to the event in question, based on the context in which the new information was received. If that context happens to be tuning an orchestra, rather than starting the first note of an A-minor scale, the brain will call up the semantic correlate for the given context and react appropriately.

Since we know that humans constantly judge by comparison, and our judgment of any item depends upon what we are comparing it to at that moment, let's be wise and use this knowledge in composing music. If there is a pattern that reflexively reoccurs throughout the musical composition it will create its memory in the brain and that will become a "point of reference" to which future transformations of the same pattern may be compared to. If humans are able compare, then in return, they will be also able to evaluate. If this evaluation process keeps going on, that probably means there is a growing interest in what is going on. This still does not mean by any means that a musical composition containing reoccurring patterns of some sort and their transformations will be a guaranteed recipe for the creation of a successful music. It simply means that unless there is a point of reference, which may be just about anything previously digested and recognizable to the ear-minds of the audience, there is a significantly much lesser chance for produced musical piece (a meme) to catch on and make people react with positive feelings. How successfully one may play with music patterns will, in the end, always remain a matter of human musical talent.

It very important to understand that it is no accident that the music memes which replicate tend to be good for humans—not for reasons of our *biological* fitness, but for whatever it is that we hold dear. This is an unsettling observation for a person, who believes in absolutes. However, the situation should not be viewed as totally desperate. Let me again put everything in computer music terms. It is amazingly fascinating to see what Super Collider, Csound, Cmix, Kyma, Sound Hack, Lemur, and countless TDM, VST and MAS plug-ins can do to sound samples, as well as what MAX can do to musical structure when applied to fractals, neural nets, fibonacci numbers, solar systems, palindromes, permutations, interpolations, pitch-sets, and population growth algorithms. It is crucial to make sure that those fascinations that we hold dear, do not remain the exclusive possessions of the composers who indulge themselves in playing with their technological marvels. It is extremely important that the audience also get to share some of our unique thrill. There must be some significant overlap between what the composer holds dear and what the audience holds dear.

This is a very demanding and difficult task to put into reality, but if it is not done soon, the computer music memes are not going to find their habitats and replicate themselves any further beyond the narrow, technically-oriented facilitators. Unless established compositional approaches are changed, the question will still remain: Do we compose only for our idiosyncratic selves or for the audiences of our culture?

On the other side of this argument Charles Rosen claims that survival of any music is independent of the audience's response.

> The music that survives is the music that musicians want to play. They perform it until it finds an audience. Sometimes it is only a small audience, as it is in the case so far for Arnold Schoenberg (and I am not sure if it will ever be a large one), but he will be performed as long as there are musicians who insist on playing him. [4]

As much as I personally disagree with Rosen's basic assumptions about survival of the music, I can pretend for a moment that Rosen is right having the following question rise: Who are the performers of the computer music? The answer: Computer music composers themselves. The only computer music composer that I came across who publicly performs other composer's computer music is Neil Rolnick. And here we arrive again at the inescapable conclusion that we are playing our music for ourselves.

Isn't it obvious that it is not feasible to continue clinging to the notion that music is exclusively a form of personal self-expression through which the composer produces a piece of music without regard to the response from its socio-cultural environment. Paradoxical notion of releasing musical memes in front of an audience and saying: "Take it or leave it, I did not write my music for you at all, but I want you to listen to it anyway" makes vary little sense. Evolution works on the same principle—the survival of the fittest—in which case 99.99% of the answers to the question raised above is certainly going to be "LEAVE IT!!!" On the other hand, if a composer is interested in a socio-cultural response to her/his music, it could make a wise decision and figure out what are the perceptual and cognitive mechanisms that make audiences to like something. Using the proposed approach may certainly prove to be more successful in sharing the computer music with people who don't know what word MIDI means.

REFERENCES:

- [1] Richard Dawkins. *The Selfish Gene.* Oxford University Press, Oxford, 1976.
- [2] Daniel C. Dennett. *Darwin's Dangerous Idea.* Simon & Schuster, New York, 1995.
- [3] Robert Jourdain. *Music, the Brain, and Extasy.* William Morrow and Company, Inc., New York, 1997.
- [4] Charles Rosen. *Classical Music in Twilight.* Harper's Magazine, Vol. 296, No. 1774, March 1998.

The Musical Fireflies - Learning About Mathematical Patterns in Music Through Expression and Play

Gili Weinberg
MIT Media Laboratory
20 Ames St., E15-491
Cambridge, MA 02139 USA
gili@media.mit.edu

Tamara M. Lackner
MIT Media Laboratory
20 Ames St., E15-020b
Cambridge, MA 02139 USA
tlackner@media.mit.edu

Jason Jay
Knowledge Learning Corp.
4340 Redwood Hwy., Bldg. B
San Rafael, CA 94903 USA
jason.jay@post.harvard.edu

Abstract

The Musical Fireflies are palm sized digital musical instruments that introduce mathematical concepts in music such as beat, rhythm and polyrythm without requiring users to have any prior knowledge of music theory or instruction. Through simple controllers, the Fireflies allow users to input rhythmic patterns, embellish them in real-time by adding rhythmic layers, synchronize patterns, and trade instrument sounds. Since interaction with other players increases the richness and complexity of the experience, the Musical Fireflies also motivate collaboration and social play.

1 Introduction

Traditional tools and methods for learning rhythm, as well as other musical concepts, usually separate the figural, intuitive experience from the formal, analytical internalization of the material [1]. When learning to play musical instruments in the formal mode, certain important musical aspects, which came naturally in the figural mode, may be temporarily hidden when students try to superimpose analytical knowledge upon felt intuitions. If this phenomenon is not acknowledged by a teacher or mentor and the gap between the different modes is not negotiated, it might ultimately lead students to cease altogether their participation in musical life [2].

This paper presents an interactive musical network of palm sized digital musical instrument called "Musical Fireflies," which are designed to help bridge the gap between the figural and formal modes. By employing digital interaction and wireless communication, the Fireflies provide players with expressive and fun rhythmical experiences that can be easily transformed into an analytical and formal exploration. Through simple controllers, the Fireflies allow players to input rhythmic patterns, embellish the patterns in real-time, synchronize patterns among different players, and trade instrument sounds. For a single player the instrument can provide figural as well as formal familiarization with musical concepts such as accents, beats, rhythmic patterns and timbre. During the multi-player interaction a wireless network is formed, which can provide novices, as well as professional musicians with an interactive group experience that leads to a deeper internalization of advanced musical concepts such as the correlation between monorhythmic and polyrhythmic structures.

1.1 Digital Manipulatives

The development of the Musical Fireflies is informed by the notion that interaction with digital physical objects, also known as digital manipulatives, can enhance learning [3, 4, 5]. The Musical Fireflies extend these studies to a musical realm by providing an expressive experience that can draw players into a meaningful musical exploration without requiring an exhaustive learning process, virtuosi performance skills, or an extensive knowledge of music theory [6, 7]. Access to and manipulation of Logo code for customizing the Fireflies also provides a basic and friendly introduction to Midi programming and electronic sound. Advanced players can therefore deepen their learning experience by reprogramming the Fireflies and adjusting their functionality to match personal musical interests and abilities.

1.2 Group Playing

The Fireflies are designed to provide simple and immediate musical interaction for single players at preliminary stages, which leads to richer, more complex musical experiences when multiple players, using multiple instruments, interact with each other. The instruments' wireless communication system allows players to explore new interactions by communicating and sharing their music with others. Through infrared communication, players can synchronize their instruments with other Fireflies, which are programmed in the same manner by other participants, and enhance their simple, monorhythmic patterns into a polyrhythmic experience. It is in these synchronized social interactions that the further mathematical aspects of the toy arise where individual users can obtain an understanding of their rhythmic patterns in relation to the group's composition [8]. Players can further explore their individual contribution to the group by trading their instrument sounds with their peers. This can be helpful for the perceptual separation of the timbre-oriented characteristic from the numerical aspects of the patterns.

2 Modes of Interaction

Interaction with the Musical Fireflies occurs in two distinct modes – the Single Player mode, where players convert numerical patterns into rhythmical structures, and the Multi Player mode, where collaboration with other players enhances the basic structures into polyrhythmic compositions.

2.1 Single Player Interaction

Each Musical Firefly is equipped with two default drum sounds that are operated by two buttons. When a Firefly is first turned on, it awaits the input of a rhythmic pattern from the buttons. The left button records an accented beat and the right button records a non-accented beat, using the same drum timbre. After two seconds of inactivity, the Firefly plays back the entered pattern in a loop, using a default tempo - ¼=80. This activity provides players with a tangible manner of entering and listening to the rhythmical output of any numerical pattern they envision, which leads towards an immediate conceptualization of the mathematical-rhythmical correlation. For example, the numerical pattern 4 3 5 2 2 would be entered and played back as follows:

● ○ ○ ○ ● ○ ○ ● ○ ○ ○ ○ ● ○ ● ○ (x loop)

(● = Accented note played by the left button
○ = non accented note played by the right button)

During playback players can input a second layer of accented and non-accented notes in real-time, using a different timbre. Each tap on a button plays a beat aloud and records its quantified position so that the beat becomes part of the rhythmic loop. Pressing both buttons simultaneously at any point stops the playback and allows the player to enter a different pattern.

2.2 Multiplayer Interaction

When two Fireflies that are playing different patterns using different timbres "see" each other (i.e. when their infrared signals are exchanged) they automatically synchronize their rhythmic patterns. (A similar interaction occurs when the firefly insects synchronize their light pulses to communicate in the dark). This activity provides participants with a richer, more complex rhythmical composition and allows for a fun and interactive introduction to polyrhythm. For example, if one Firefly plays a 7 beat pattern (● ○ ○ ○ ● ○ ○) and another plays a 4 beat pattern (● ○ ○ ○) then the players can hear the process of divergence and convergence as the patterns go in and out of phase every 28 beats, the smallest common denominator:

.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
7/4	●	○	○	○	●	○	○	●	○	○	○	●	○	○	●	○	○	○	●	○	○	●	○	○	○	●	○	○
4/4	●	○	○	○	●	○	○	○	●	○	○	○	●	○	○	○	●	○	○	○	●	○	○	○	●	○	○	○

While the two Fireflies are synchronized, players can also initiate a "Timbre Deal" in which instrument sounds are traded between the devices. Pressing either the left or right button trades both layers of the accented or non-accented timbre respectively. Each Firefly continues to plays its original pattern but with one button triggering the two new timbres that were received in the timbre deal. This provides players with a higher level of musical abstraction conceptualization since they now can separate the rhythmical pattern aspect of the beat from the specific timbre in which it is being played. Because the Fireflies network is richer after the interaction (i.e. each instrument now contains four different timbres,) the system can also encourage collaborative play where players are motivated by trading, collection and playing games by sending and receiving different timbres from different Fireflies.

3 Hardware and Software

The Firefly's casing is made of a 7.5''x5.5''x2.5'' 3-D printed fabrication, which is designed to be held with both hands while tapping the top-mounted buttons. The buttons are connected to two A/D converters on the embedded "Cricket" [9] - a tiny computer that is responsible for the musical algorithms.

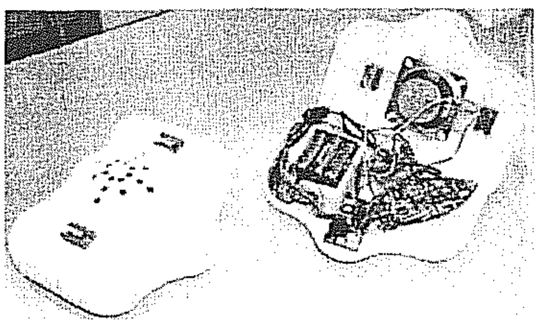


Figure 1. Two touch sensors on each Firefly are connected to a central microprocessor with infrared capabilities, which also serves as a driving force for the MIDI board, amplifier, and speaker. Each Firefly requires a power source of 6 AA batteries.

The Cricket, which is mounted at the front of the Firefly, is based on the Microchip PIC series of microprocessors. It can receive information from a variety of sensors and is equipped with an infrared system that allows for communication with other Crickets. The Cricket is programmed in a dialect of the Logo programming language. Application programs can be downloaded to the Cricket via its infrared communications port, allowing for players to easily rewrite and download applications and data to the Firefly. The entered rhythmic patterns are converted into musical messages using Cricket Logo general MIDI commands [10]. These are sent through the Cricket's serial bus port to the "MidiBoat" [11]

— a tiny General Midi circuit, which supports up to 16 polyphonic channels, 128 melodic timbres and 128 percussive timbres. In the current application, only the percussive timbres are being used. The audio from the MidiBoat is then sent to a top-mounted speaker.

4 Discussion

Several challenges were addressed in the process of designing a musical interaction that would focus on bridging the gap between the figural and the formal learning modes. One of the main challenges was to balance between the simplicity of operation and the depth of the musical interaction, between allowing for a fun, intuitive, and expressive activity as well as providing a meaningful educational experience. In order to address this challenge we tried to design a varied and rich infrastructure that will apply to a variety of players, who can be located in different places on the figural-formal axis. The Fireflies, therefore, were designed so that novices, with little formal education or experience would be able to experiment with stand-alone simple patterns using a single layer of rhythm. The instruments can also accommodate more advanced users, who can play with complex multi-layered interdependent patterns as well as reprogram their instruments using Logo. Our ultimate goal was to encourage players to advance from the simple basic interaction toward the rich, enhanced, and interdependent experience.

In order to address formal as well as figural musical aspects we had to come up with several compromises and trade-off modifications. For example, it was decided that the Fireflies would not capture the exact timing and rhythmic values of the entered taps. Rather, the algorithm merely records the sequence of accented and non-accented beats and plays them back in a default tempo. Although figural thinking would probably find exact rhythmic playback more intuitive and expressive, we reckoned that flattening the tempo would provide a better ground for comprehending the polyrhythmic collaborations, especially for children and novices. It was for this reason that we decided not to allow the input of rests. While it is clear that the addition of rests could have provided a richer more musical experience, experiments with a software-based version of the application showed that in the multiplayer mode players found it difficult and formally comprehend the polyrhythmic interaction. We faced a similar problem when deciding about the ideal value for the default tempo. For two-line rhythm patterns, a fast tempo sounded less mechanical and more fun to listen to than a slow tempo. However, when the Fireflies played their patterns too fast it became impossible to follow the divergence and convergence of different patterns in the multiplayer mode. We hope that the tempo chosen (¼ = 80) serves as a reasonable compromise between these two extremes.

As part of our efforts to balance between simplicity and depth of interaction we also had to decide about the number and complexity of controllers for players to interact with. We chose to implement only two discrete buttons and no continuous controllers in an effort to provide players with a simple, elegant and easy-to-learn interaction. This required us to impose a considerable amount of interaction onto these buttons while streamlining the software design. For example, instead of having a third mechanism to stop the rhythmic patterns we designed the Fireflies so that pressing the two existing buttons simultaneously would stop the music.

Another design challenge that we addressed was how to allow for an interesting and challenging multiplayer interaction that will stay coherent and will not confuse the participants. Several preliminary algorithms (such as trading the numerical patterns or mixing between patterns to create more complex ones) were ruled out since they failed to address this challenge by creating confusion and uncertainties among players. Finally we chose timbre as the parameter to be traded due to its "coloring" qualities, which do not complicate the already rich rhythmical texture, especially when more than two Fireflies are involved. The timbre parameter also seems to provide an educational value by helping to separate the instrument sound from the numerical patterns while maintaining the system's coherency.

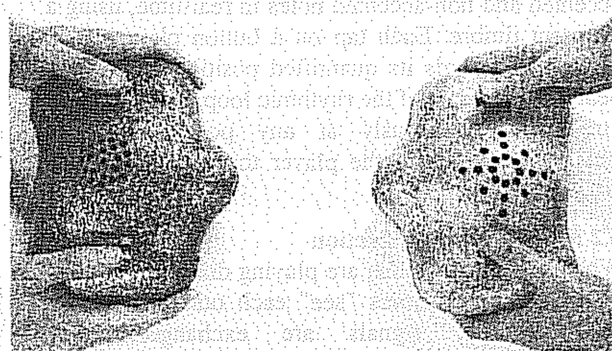


Fig 2: Two players interact with the finalized version of the Musical Fireflies.

5 Observation and Evaluation

Preliminary observations of group playing with the Musical Fireflies have been conducted, followed by discussions with the players. Participants were also asked to play with a Max-based [12] software version of the application and compare their experience with the tangible interaction that is provided by the physical Firefly objects. In general, players found the concrete aspects of playing with a physical object compelling in comparison to using a keyboard and a mouse. Subjects mentioned the unmediated connection that was formed with the instrument as contributing to the creation of personal involvement and relationship with the musical application. Tapping real buttons and listening to the music coming from distinct physical sources also helped players to comprehend and follow the trading interaction in a more coherent manner than listening to computer speakers, especially when more than two Fireflies were playing simultaneously.

The observations and discussions also led to the identification of several deficiencies in the interaction design. Although carefully considered, the timbre trading function turned out to create some non-coherent consequences especially for novices who found it confusing. These prevented the full internalization of the desired educational value. In addition, trading timbres, a discrete operation that does not provide long-lasting interaction, led players to lose interest in the interaction after a few taps. Both these deficiencies are addressed in the Future Work section.

Subjects also expressed their wish to interact in larger groups that are comprised of several simultaneous players. The current application allows synchronization among up to three players and timbre trading between only two players at a time. This limitation is imposed by the line-of-sight infrared technology. In order to provide a full multiplayer experience new technologies and applications should be developed.

6 Future Work

Several hardware and software improvements are currently being investigated for the Fireflies. In terms of hardware we are exploring the possibility of implementing radio frequency transceivers that will allow for a group of up to twenty players to interact simultaneously. This will obviously require the development of new applications that will take advantage of the enhanced multiplayer possibilities. Other hardware enhancements that are being considered are improving the speaker acoustic qualities by using different materials and enclosures as well as embedding a sample-based recording chip that will allow players to record and playback their own sounds. This will hopefully add a new dimension to the personalization of the instruments.

We are also looking to enhance the musical application in Single Player as well as Multi Player modes. The new interaction will include recording and playing back of pitch-based sounds in addition to the pitch-less percussive sounds that are currently being used. This will hopefully allow for a richer melodic and harmonic musical collaboration while enhancing the educational value of the interaction. Several software-based efforts in this direction are already under development.

Other enhancements under development are the implementation of continuous timbre control using an embedded digital signal processing engine, and the implementation of continuous rhythmic manipulation that would allow players to edit their entered patterns and explore new rhythmic variations. In the multiplayer interaction, we are currently experimenting with longer performance oriented collaborations like providing players with the possibility to record patterns using their collaborators' timbres and allowing participants to receive other players patterns, edit them and create their own variation before sending them back to the group.

7 Acknowledgments

We thank Fred Martin, Bakhtiar Mikhak, and Prof. Tod Machover for their contributions.

References

[1] Gardner, H. (1983) *Frames of Mind*. New York: Basic Books pp. 110-112.

[2] Bamberger, J. (1982) "Growing up Prodigies: The Mid-Life Crisis," *New Directions for Child Development* 17, pp. 61-78

[3] Papert, S. (1980). *Mindstorm*. New York: Basic Books.

[4] Resnick, M., Martin, F., Sargent, R., & Silvernan, B. (1996). Programmable Bricks: Toys to Think With. *IBM Systems Journal*, 35 (3 & 4).

[5] Turkle, S. & Papert S. (1992) "Epistemological Pluralism and the Revaluation of the Concrete," *Journal of Mathematical Behavior* vol. 11, pp. 3-33.

[6] Weinberg, G., Orth M., and Russo P. (2000) "The Embroidered Musical Ball: A Squeezable Instrument for Expressive Performance," *Proceedings of CHI 2000*. The Hague: ACM Press.

[7] Weinberg, G. (1999) "The Musical Playpen: An Immersive Digital Musical Instrument," *Personal Technologies*. London: Springer-Verlag Publication, 3:132-136.

[8] Handel, S., (1984) Using Polyrhythms to Study Rhythm. *Music Perception*. Summer 1984, Vol. 11, No 4, 465-484.

[9] Martin, F. (1998) The Cricket web site. http://fredm.www.media.mit.edu/people/fredm/projects/cricket.

[10] Martin, F., Mikhak, B, and Silverman, B. (2000) Metacrickets – A designer's kit for making computational devices. To be published in *IBM System Journal* 2000.

[11] Smith, J. (1998) The MiniMidi web site. http://www.media.mit.edu/~jrs/minimidi.

[12] Winkler T. (1998) *Composing Interactive Music – Techniques and Ideas Using Max*. Cambridge Massachusetts: The MIT Press.

EDUCATING A NEW GENERATION OF WOMEN: WOW'EM

http://music.dartmouth.edu/~wowem

Kristine H. Burns
Florida International University
School of Music, University Park Campus
Miami, FL 33199 USA
burnsk@fiu.edu

ABSTRACT

Computer-Assisted Instruction (CAI) developed from a need to engage students in a one-on-one interactive teaching paradigm. While materials may be presented from a multi-user classroom or distance-learning situation, the most important concept is that students learn at their own pace.

The Internet has provided not only sources of reference and research, but also means of personal support for many women. Because the Internet offers an environment in which gender is hidden, women are capable of thriving both economically and socially as well as men. Traits such as age, sex, race, and gender may remain anonymous when using the Internet, so discussion may focus on content, ideas, and meaning [1]. As a group, women comprise some 25–30% of those using commercial online services among users in the United States. More web-based resources for electro-acoustic composition and multimedia technology are being developed in order to meet the growing needs of female artists. It is with this in mind that WOW'EM (Women On the Web—ElectroMedia) [2] was developed.

History

Early CAI programs such as Ars Nova Software's Practica Musica and MacGAMUT Music Software International's MacGamut focused on rote drills for college-level ear training. More recent tools have concentrated on developmental learning among younger students in musicianship, composition, and even electro-acoustic music. Morton Subotnick's *Making Music* [3] and *Making More Music* [4] are CD-ROMs designed to teach very young students fundamentals music notation and basic ear training skills. NOTAM's *DSP for Children* [5] guides students through basic principles of sound synthesis and electro-acoustic composition. Courseware such as Jeffrey Stolet's *Electronic Music Primer* [6] provides college students with an interactive on-line environment for electro-acoustic composition that includes text and audio examples.

Inspired by this new online educational philosophy and tools such as these, the WOW'EM World Wide Web site was created to fill a void in the education and support of young women with interests in digital media arts, audio, and video. WOW'EM has a twofold purpose: (1) to creatively engage students in learning about electro-acoustic music, digital video, and digital imaging; and (2) to foster a sense of community and support among young women with interests in these interrelated fields.

OVERVIEW

Released in December 1996, WOW'EM (Women On the Web—ElectronMedia) provides information for students with interests in music and visual art, including primary and secondary school subjects such as orchestra or chorus, painting or pottery, and math, science, or computers [7].

The site demonstrates that the audio and visual arts may indeed be integrated into a hybrid artform. With the arrival of inexpensive computer hardware and software, multimedia has garnered much attention in recent years. Students today are the first generation to grow up with media such as music videos and computer games. While students are familiar with intermedia arts, what they are often lacking is a sense of artistic guidance and education. Second, WOW'EM provides information of special interest to young women, thereby creating a kind of "cybercommunity." The site includes discussions on numerous topics in technology, interviews with women and men working in technology, as well as young women's Internet sites. Now that so many families own computers, especially in the United States, an increased number of children know much more about the Internet than do their parents. Distributing the information is not the problem; the dilemma is showing these young women that they, too, may have every ability that young men do in arts and technology.

WOW'EM is divided into nine sections: *Interviews and Viewpoints*, *Schools*, *Homework*, *Hardware and Software*, *ElectronMedia*, *TeleTalk*, *In-Forms*, *No Boys Allowed!*, and *Techno Teachers*.

Interviews/Viewpoints features prominent women and men in the intermedia arts. These pages include conversations with professionals from the fields of recording engineering, electro-acoustic composition, graphic design, and video art. Experts tell their stories: what do they do for a living, how they got started, and their educational background.

The *Schools* section allows users to explore topics such as choosing a high school or college, summer workshops, and international research centers. Departmental contact information is provided whenever possible.

Homework provides information on applying for internships and writing a résumé. Short articles appear in the "What Do I Do" section. Professionals of all ages—from recent college graduates to widely respected international figures—have described their jobs and backgrounds.

Software/Hardware begins with an extensive list of major audio and video hardware companies and includes contact information and product listings. This section also lists software and applications in the sections called "Digital Audio," "MIDI," "Software Sound Synthesis," "Digital Video and Imaging," and "Authoring Tools." These sections also contain articles explaining basic concepts of these software categories.

ElectronMedia provides information from many areas including books, recordings, museums, scores, and educational and artistic videos and DVDs. The section "Electronic Music" covers the history of electro-acoustic music, significant books (including history and aesthetics, composition, MIDI, sound synthesis, and various composers), recordings, and scores. The "Visual Media" pages focus on fundamentals such as multimedia history, along with performance spaces, festivals, and museums (both online and traditional). "Related Areas" include subsections on "Dance," "CD-ROM," and "Film."

TeleTalk teaches users how to use the Internet. Beginning with a brief overview of the Internet, the pages also define basic principles of electronic mail and file transfer protocol (FTP). This section also discusses the design of personal web pages, plug-ins, and other current developments.

In-Forms addresses five main areas: professional organizations, magazines, distributors, festivals, and web sites.

No Boys Allowed! is a section devoted exclusively to women's issues. Articles by and about women working in music and visual art, as well as women's studies and gender areas, appear in this section. This section also includes biographies of women composers and visual artists, competitions, study grants, and other material on women's issues. Additionally, a reprint of Pauline Oliveros' article "And Don't Call Them 'Lady' Composers" appears in this section.

Techno Teachers is the newest addition to the site. Primary and secondary educators are requested to post innovative projects and curriculum suggestions. This section is the only section devoted to teachers rather than to students.

Table 1 shows a selection of the contributed articles available in WOW'EM.

Section	**Contributor**
Hardware and Software	
Audio Plug-Ins	Ed Gray
An Overview of Algorithmic Composition	Kristine H. Burns
Sequencing	Onche Ugbabe
Music Notation	Jody Nagel
Digital Video	Peter Swendsen
3D Imaging Programs	Jeff Seitz
Homework	
How To Release Your Music	Kym Serrano
Electronic Composition	Judith Shatin

Multimedia	Eric Somers
Graphic Design	Eric Somers
The Life of an Intern	Stephanie Baker

In-Forms

Professional Organizations	
Web Sites	
Distributors	

Magazines and Journals	
Festivals and Concerts	

ElectronMedia

Music	
Visual Art	
Mish-Mash	
Science Stuff	
Related Areas	

Schools

Out of the frying pan...	James Phelps

TeleTalk

Color on the Web	Peter Swendsen
Web Copyright Issues	Peter Swendsen

Viewpoints and Interviews

Interview: Beverly Grigsby	
Viewpoint	Cort Lippe

No Boys Allowed!

And Don't Call Them 'Lady' Composers	Pauline Oliveros
Women's Centers	Giavanna Munafo
Women's Studies Programs	Giavanna Munafo
Discographies of Women's Music	Elizabeth Hinkle-Turner
International Alliance for Women in Music	Deon Nielsen Price
Sophia Corri Dussek (1775-1847)	Suzanne Moulton Gertig

Techno Teachers

A Sound Education	Jeff Peller

Table 1. *Selected WOW'EM sections and articles*

CONCLUSION

The majority of information contained in WOW'EM may benefit anyone and is not necessarily gender-specific. While the site as a whole is intended for young women, it is not meant to segregate. Both boys and girls, women and men, will find many similar areas of interest. The philosophy of the site is not to separate, but to integrate.

REFERENCES

1. Sinclair, C. 1996. *Netchick: A Smart-Girl Guide to the Wired World*. New York: Henry Holt and Company, Inc.
2. Burns, K. H. *WOW'EM (Women On the Web—ElectronMedia).* Available online at http://eamusic.dartmouth.edu/~wowem.
3. Subnotick, M. 1995. *Making Music.* CD-ROM. Voyager Co.
4. Subotnick, M. 1998. *Making More Music.* CD-ROM. Voyager Co.
5. Rudi, J. 1997. "DSP for Children." In *Proceedings of the 1997 Computer Music Conference.* San Francisco: International Computer Association, 268–270.
6. Stolet, J., et al. *Electronic Music Primer Main Module.* Available online at http://nmc.uoregon.edu/emi/emp_mac/main.html
7. Burns, K. H. 1999. "WOW'EM: Women On the Web—ElectronMedia." *Leonardo Journal* 32(1): 61–62.

The Electronic and Computer Music Course in the Italian Conservatories. History, Status, Perspectives

Alessandro Mastropietro
Via Colle Pretara, 51/B, 67100 L'Aquila (Italy) – e-mail: ale_mastropietro@hotmail.com

Doctorate in History and Analysis of Musical Cultures, Università degli Studi "La Sapienza" – Roma, Italy
Istituto Gramma – L'Aquila, Italy

Abstract

This paper aims at giving an historical and reasoned survey on the activities and the outline of the electronic and computer music course in the Italian Conservatories, now denominated "Extraordinary Permanent School in Electronic Music". It represents the main educational structure in Italy for this specific field, even if not the only one. The paper will try to give an overlook: 1) on the history of the Conservatories' courses, particularly on their very beginning (between the late 60^{th} and the early 70^{th}); on the context of their birth, marked by the Utopian idea of a new and renovated musical and compositional world, a well-spread and deeply-rooted idea at that historical moment, in which electronic music had a fundamental role; on the figures of their early teachers, mostly quite important pioneers in the Electronic and Digital Music; 2) on the actual status of the courses, and (if possible) on their capability and specific attitude in encouraging and developing non strictly-compositional activities. This is, perhaps, the main expansion front of electronic music teaching in Italian Conservatories, above all after their recent school reformation towards a University-status: the perspectives of this develop, which has come out yet in some past didactic conferences, will be discussed, considering the present shape of the course internal program (as established by the state recognition) and its previous history.

1. History

1.1 The experimental beginnings

In lacking of a specific university-degree in music education (the Accademia di S. Cecilia in Rom excepted, an institution for perfecting in music), in Italy the Electronic Music state teaching is practised in the Conservatories of Music (Conservatori di Musica). Such structures attends to the musical teaching for the very beginner as well as for the composer, who has engaging himself to reach a personal language by following the tracks of the most important and advanced compositional experiences in this century. As a part of this last field, Electronic Music was introduced in such a didactic system since the second half of 60^{th}, when the Conservatories were less then 20, rather then the present 70 ones. This penetration rose in spite of – or better, thanks to – an outdated (1933) general study-prospectus (which is still in force); it was initially accomplished according to the opened formula of free experimental courses (whose teachers were often already in office with different teaching-rolls), and took place aside others didactic experimentation in composition, whose need was felt at that time so urgent to jump over the ministerial study-prospectus. In a so utopian and strongly experimental compositional environment, the electronic music was then widely considered the fulcrum of the engine that was pushing ahead the musical thought, and so a pivotal component in building and ripening the compositional figures of the future. Some Conservatory headmasters, in spite of their solid but traditional formation, realized this crux, and their broad-minded consciousness about this was a central factor in the starting of electronic music courses.

The very first course was activated by the "L. Cherubini" Conservatory in Firenze (scholastic term 1965-66) thanks to the availability of the headmaster Antonio Veretti (who let Grossi draw and print a graphically-beguiling call-poster) as well as to the outstanding dynamic and experimental attitude of Pietro Grossi, who was its first teacher and became available to transfer his own studio (*Studio di Fonologia Musicale di Fierenze - S 2F M*, founded in 1961 and well-equipped within 1963) inside the Conservatory (on this singular figure see: F. Giomi, M. Ligabue, *L'istante zero. Conversazioni e riflessioni con Pietro Grossi*, Firenze, SISMEL-Edizioni del Galluzzo, 1999). Pietro Grossi (Firenze, 1914), cellist and composer, is a pioneer of electronic and, above all, digital music in Italy, inaugurating at the end of 60^{th} early researches for controllability and producibility of sound events by digital computer of great factories and university departments. In such a field, he promoted the foundation (1969) of the Musical Computing Department of CNUCE-C.N.R. in Pisa, still existing; moreover, he attended to the organization of the first International Meeting for Experimental Studios in Electronic Music (*XXXI Maggio Musicale Fiorentino*, 1968). The Firenze Conservatory was able to profit by digital computing machines since 70^{th} (in 1973 the teaching-roll was meanwhile assumed by Albert Mayr, an important figure for experimentation in electronic improvisation, who tried to change the course label from "Musical Phonology" into "Experimental music") thanks to a co-operation with CNUCE-C.N.R., thus permitting to its students to acquiring knowledge and experiences that were at that time hardly workable. The attention of this Conservatory in musical applications of computing had another proof in 1984 with the activation of the sole teaching of Musical Computing in Italy, once again promoted by Grossi.

In the planned and carried out projects by the Florentine studio, an evident "work-in-progress" character comes to light, and allow to use a former project and its results as a material for a following project; the collective to-develop-availability of this projects and their frequent automation gave rise to a lacking of individual intellectual property of the musical works, a characteristic in common with the coeval studios in Torino and Padova (see below).

The next activation of an electronic music course took place just in Torino ("G. Verdi" Conservatory) in 1968; its first teacher was Enore Maria Zaffiri (Torino, 1928), previously teacher of compositional subjects in the same institute since 1953. He's another pioneer of Italian electronic music, to which he devoted himself since 1965 by his *Studio di Musica Elettronica di Torino* (SMET); documentation of his own and his apprentices' projects has been published in the 1968 volume (E. Zaffiri, *Due scuole di musica elettronica in Italia*, Torino, 1968, Silva editore) together with a writing by Grossi on the Florentine course and documentation on it. So the birth of Torino course too seems to be a natural continuation, inside the Conservatory, of a previous external studio activity, in which didactics had already a certain significance. In this studio mostly geometrically-planned projects were worked out (f. e. the EL/25 project – 1967, built on a sinusoidal-glissandi net, that grows from the focuses of an ellipsis) according to Zaffiri's interest in geometric shapes; but also researches on sound poetry and ambient music were there developed.

The year after (1969-70), the electronic music course in the "G. Verdi" Conservatory in Milano began his activities, thanks to the spirit of enterprise of his first teacher, Angelo Paccagnini: engaged composer, he was at that time director of the famous *Centro di Fonologia della RAI di Milano*, founded in 1955 by Maderna, Berio and others, but then already outdated in its instrumentation, notwithstanding the considerable compositional outcome Luigi Nono was achieving there. Documentation of this birth was published in the 1969 Conservatory year-book and illustrated as a prominent and up-to-dating event: the whole class was able to work inside the *Centro di Fonologia* thanks to a special convention. In ten years of teaching, Paccagnini (unfortunately dead in June 1999) developed projects in co-operation with external university- and cultural partners, mostly in the field of psycho-acoustics and psychology of music, and in that of primary-educational applications of electronic music, working even inside prisons. So the course structured itself as a multidisciplinary department, in which also external researcher (such as University teachers) were active.

In the same year a course was open by the "Accademia di S. Cecilia" in Roma (which had established it in 1968) and granted to the teaching of Franco Evangelisti (Roma, 1924-1980). He was the first and sole teacher in this course, closed in 1972; however, he taught electronic music also in the "A. Casella" Conservatory in L'Aquila (from 1970-71 to 1975-76, then closed and re-activated in 1980-81 with Michelangelo Lupone as its teacher) and then in the "S. Cecilia" Conservatory in Roma (from 1974-75 until his death). Evangelisti was one of the most important New-Music figures since 50[th], beside composer as Stockhausen, Boulez, Nono, Maderna, Berio, Ligeti and so on; probably the most outstanding international figure among Italian electronic music teachers, he composed a very important work (from both musical and historical points of view) such as *Incontri di fasce sonore* (1956-57, carried out in the WDR Studios, Köln); then he made use of a magnetic tape in the musical theatre piece *Die Schachtel* (1962-63), and of electronics in *Spazio a 5* (1959-61), as well as within his activity in the *Gruppo d'Improvvisazione di Nuova Consonanza*, which he founded in 1964.

In the same 1970-71 the electronic music course in the "G. Rossini" Conservatory in Pesaro started a seminar activity acted by Domenico Guaccero, Mario Bertoncini and then Walter Branchi, the sole teacher in the course from 1971-72 until 1979-80 and the first author of a specific volume in Italian mother-speech on electronic music (*Tecnologia della musica elettronica*, Roma, 1975). This course, activated during the direction of Marcello Abbado, was - during 70[th] - one of the most productive and updated in Italy, so editing an own journal named *Tecnomusica*. One of the promoters of this course was a pivotal figure for the didactics of electronic music such as Guaccero (Palo del Colle, 1927-Roma, 1984): although he formally never thought electronic music, he continually stimulated his students to deepen the study of electronic music and its coeval and future potential; this intense energies- and curiosity-producing around electronic music, made the birth of the electronic music courses in Pesaro and, later, in Frosinone (1974-75, Giorgio Nottoli as teacher, still in roll) possible; a brochure from the *Centro di Musica Sperimentale* in Roma (founded by Guaccero in 1972) informs us about some seminars on electronic music he gave at the "F. Morlacchi" Conservatory in Perugia, where a real course started only in 1979: a VCS3 synthesizer (used by Guaccero in his electronic improvisation groups), bought by the institute for the seminars, is still existing in that Conservatory.

The following new courses will be in the order: "G.B. Martini" Conservatory in Bologna from 1971-72 (the teacher was almost unceasingly till 1991-92 Gianfelice Fugazza, cellist and expert in electronic musical instruments), whose course was initially named "Musical instruments" (an evidence of the open didactic experimentation of this early period, not closed in a compositional profile). Then, the year after (1972-73), "B. Marcello" Conservatory in Venezia and "C. Pollini" Conservatory in Padova; in the former, the teacher was for three years the famous orchestral-conductor Giuseppe Sinopoli, whose lessons were necessarily theoretic because of the lack of any suitable instrumentation. On the contrary, the Padova course was able to benefit of a co-operation with the local University Computing Centre, which the *Centro di Sonologia Computazionale* was officially in 1979 born of. Protagonist of this agreement and teacher in the course till 1979-80 was Teresa Rampazzi (Vicenza, 1979), pianist, attendant at the Darmstadt *Ferienkurse*, promoter of C.S.C. and, above all, another Italian pioneer in electronic music within the *Nuove Proposte Sonore* group, active from 1965 to 1972 with an own studio attended by some future researcher in C.S.C. (De Poli, Tisato, Vidolin...). Later on, also the Venezia course was able to join this agreement under the teaching of Alvise Vidolin (Venezia, 1949, teacher from 1975-76 till today, electronic performer of great experience and importance especially for the later electronic works of Luigi Nono), so starting its early real productions.

1.2 Growing of the courses. The law context within the Conservatories

These early years went by without a precise law context and a unifying study-prospectus, until a law ordinance about experimentation in Conservatory of Music was issued in 1974: it ratified a three-years extraordinary course (in accordance with 1969 agreement upon the programme between the Milano Conservatory – Paccagnini and the headmaster Jacopo Napoli – and the central Ministry), a final after-examination certificate and the possibility of entries for external-to-Conservatories students. But this regularization didn't cause a wide flourishing of new courses: besides the above-mentioned in Frosinone, a new course started in 1974 in the Conservatory "L. D'Annunzio" in Pescara (Riccardo Bianchini as its teacher until 1979-80, then Michelangelo Lupone from 1980) under Firmino Sifonia's headmastership; unfortunately, this course was closed in 1981 because of the burglary of the most important equipment of the studio (two modular Moogs and a professional tape recorder); afterwards, the course in Perugia Conservatory was activated in 1979-80 (Luigi Ceccarelli as teacher, today still in roll). This can be evidence that the experimental pushes in the previous years were authentically independent of any bureaucratic recognition.

The law context got worse between 1982 and '83, when two Ministerial ordinances degraded the course to an annual auxiliary one, so preventing it from granting a final certificate and from accepting external students. Notwithstanding this, a new course (in the "P. da Palestrina" Conservatory in Cagliari) was born in 1984. It's important to lay stress on the untiring (didactic as well as productive) exchanges between electronic music teachers, who recognized soon themselves as a unitary group. All this led to realize *Musica ex Machina*, the first Meeting of Italian Electronic Music Courses, which took place in Firenze from 22 to 26 April 1980 in accordance with a programme (co-ordinated by the Florentine course of Albert Mayr) of reports and listening sessions.

A new and more solid recognition as "Extraordinary Permanent School in Electronic Music" (1992, with 1994 improvements in examination programme, which has been due to a collective teachers' initiative from the first meeting La terra fertile) has made possible new and evident outcome for the courses: six new professorships (Latina, 1992; Catania, Bari, 1993; Parma, Sassari, Trieste, 1999); a wide up-to-dating of the equipment in every course, due to the great progress (also in terms of costs) of digital technology; an incredible flourishing of public events (concerts, congresses), carried out by Conservatories themselves or in co-operation with important organizations (the biennial International Meeting "La Terra Fertile", created in 1994 by Conservatory of L'Aquila and Istituto Gramma as a reference point for Italian electronic music courses, must be especially mentioned); a large number of various professional figures (sometimes international-prises winners), active in different musical fields such as composition, direction and assistance in the most important European Research Centres, musicology, performance and so on.

2. Present Status. Perspectives

At the present time, 18 courses (with an aggregate of about 200 students) are activated in the Italian Conservatories. Their list (with respective teachers) follows: Firenze (1965-73, Grossi; 73-91, Mayr; 91-92, Lelio Camilleri; 92-97, Alfonso Belfiore; 98-2000, Paolo Zavagna). Torino (1968-82, Zaffiri; 82-86, temporary posts; 86-, Ruggiero Tajè). Milano (1969-80, Paccagnini; 80-85, Bianchini, previously teacher in Pescara from 1974; 85-, Riccardo Sinigaglia). L'Aquila (1970-75, Evangelisti; 80-, Lupone, in 1980 and '81 also teacher in Pescara). Pesaro (1970, seminars; 71-80, Branchi; 80-, Giordani). Bologna (1971-73 and 74-92, Fugazza; 92-, Camilleri). Padova (1972-80, Rampazzi; 80-92, Belfiore; 92-, Nicola Bernardini). Venezia (1972-75, Sinopoli; 75-, Vidolin). Roma (1974-80, Evangelisti; 80-87, Branchi; 87-, Bianchini). Frosinone (1974-, Nottoli). Perugia (1979-, Luigi Ceccarelli). Cagliari (1984-89, Bernardini; 89-92, Roberto Doati; 92-95, Serena Tamburini; 95-, Francesco Giomi). Latina (1992-99, Doati; 99-, Francesco Galante). Catania (1993, Agostino Di Scipio; 94-, Alessandro Cipriani). Bari (1993-, Di Scipio). 1999: Parma (Giovanni Cospito), Sassari (Tamburini), Trieste (Doati). Many teachers have been student in previously some courses: Mayr was student under Grossi's teaching; Doati, Camilleri, Giomi and Ligabue (present Musical Computing teacher in Firenze) under Mayr's; Nottoli, Ceccarelli and Giordani under Branchi's; Lupone,

Galante and Tamburini under Nottoli's; Di Scipio under Lupone's; Vidolin under Rampazzi's; Doati, Cospito and Zavagna under Vidolin's; Bianchini, Sinigaglia and Tajè under Paccagnini's; Cipriani under Bianchini's. Further important students to be named are Ivan Fedele, Gilberto Bosco, Francesco Pennisi, Lorenzo Ferrero, Sandro Gorli, Mario Baroni, Nicola Sani, Enrico Cocco, Franco Sbacco, Guido Baggiani, Giovanni Piazza, Tonino Battista Marco Di Bari, Claudio Ambrosini, Marco Stroppa, Wolfgang Motz and many others that can't be mentioned only because of space lack.
After its 1992 recognition, the course is 4-years long with the annual programmes formulated every year by the teacher. However, a common pattern is based on the gradual knowledge of the history of electronic music, on physic-, psycho- and electro-acoustical cognitions, on the step-by-step sound-synthesis, digital signal theory and algorithmic programming learning. As the final examination proofs are in analysis and in composition (a first session of 8 hours to produce e project or a score on the basis of starting points and materials furnished by the board of examiners, then 30 days to realize it) and the entry qualifications prescribe – but not exclusively – compositional degrees, the course has presently a prevalent compositional profile: the principal aim of electronic music teachers is that of diversifying the study-programmes contours, to make possible to prepare performers, programmers and other professional figures without delegating it to other private (mostly public-financed) associations which have developed such special courses. In order to achieve this, the favourite perspective is that of a like-university transformation of the course structure: such an aim goes towards the whole debated status-changing of Conservatories from a school- to an university-status, which electronic music has an actual inclination because of its multidisciplinary foundation for. The Vidolin's proposal of a many-years common term, followed by some different few-years specializations (with a single teacher for each individual subjects) can be a good solution to give chance to better supervise the professional growth of the students, but also to come back (in a deeply different outline) towards the experimental roots of pioneers (such as Paccagnini and Guaccero) and towards the open and dynamic educational structure they had prefigured.

3. Bibliographical references

[AA. VV.], *Di Franco Evangelisti e di alcuni nodi storici del tempo*, Roma, 1980, Nuova Consonanza ed..

[AA. VV.], *Il Conservatorio di Musica "B. Marcello" di Venezia. 1876-1976. Centenario della fondazione*, catalogo della mostra, Palazzo Pisani – Venezia, 1977.

Conservatorio di Musica "G.B. Martini", Bologna, *Profili di un decennio, 1964-74*, Bologna, 1975.

Conservatorio di Musica "G. Verdi", Milano, *Annuario dell'a. a. 1969-70*, Milano, 1970.

Conservatorio di Musica "L. Cherubini", Firenze, *Annuario relativo agli anni 1983-84 e 1984-85*, Firenze, 1985.

R. Aru, S. L. Manca, *La musica elettronica in Sardegna: storia ed attualità, in La terra fertile. Incontro Nazionale di Musica Elettronica – II edizione. Atti del Convegno 1996*, edited by M. C. De Amicis e Ignazio Prignano, L'Aquila, 1996, Conservatorio di Musica "A. Casella" – Istituto Gramma, pp. 142-147.

A. Basso, *Il Conservatorio di Musica "G. Verdi" di Torino. Storia e documenti dalle origini al 1970*, Torino, 1971, UTET.

R. Doati, *Esaltata l'arte, affermatasi la fantasia, quale educazione?*, in *Il complesso di Elettra*, Roma 1995, CIDIM – CEMAT, pp. 53-56.

F. Evangelisti, *Verso una composizione elettronica, estratto dall'Annuario Accademia Nazionale di S. Cecilia, 1969*, 18 pp..

F. Giomi, *Scuole storiche italiane di musica elettronica*, report (20-11-1999) at the congress *Musica e Tecnologia, domani*, Milano, 20-21 novembre 1999, Foyer del Teatro alla Scala, Musica/Realtà, in collaboration with Teatro alla Scala, issue forthcoming in the proceedings.

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

A. Mastropietro, F. Sebastiani, *Didattica: fondamenti e prospettive*, in *Atti de "La Terra Fertile '98", Incontro Biennale di Musica Elettronica – III Edizione*, edited by M. C. De Amicis, pp. 69-72.

G. Nottoli, *Musica elettronica in Conservatorio*, in the concerts-booklet of *Suoni d'argento. 25 Anni di Musica Elettronica*, Conservatorio di Musica "L. Refice", Frosinone, 5-8 june 2000.

N. Sani, *[Musica elettronica] espulsa dal Conservatorio*, in *Paese Sera*, 15-11-1983, p. 7.

S. Sapyr, A. Vidolin, *Importanza della didattica per l'evoluzione dell'informatica musicale*, in *Atti del VII Colloquio di Informatica Musicale*, Roma, 1988, Associazione Musica Verticale, pp. 113-118.

A. Vidolin, *La tradizione e il rinnovamento*, report (7-6-1996) at the Congress "Il Conservatorio e le nuove Scuole", Ferrara.

E. Zaffiri, *Due scuole di musica elettronica in Italia (Con una nota di Pietro Grossi)*, Torino, 1968, Silva editore.

L. Zattra, *Da Teresa Rampazzi al Centro Di Sonologia Computazionale (C.S.C.): la stagione della Musica Elettronica a Padova*, Degree Dissertation in Letters, Università di Padova, Facolta' di Lettere e Filosofia, Dipartimento di Storia delle arti visive e della musica, A. A. 98/99.

POSTER SESSION 1

Quad DSP based board for real time sound synthesis and processing

Giovanni Costantini[1], Giorgio Nottoli[2], Mario Salerno[1], Patrick Serra[1]

[1]Department of Electronic Engineering, University of Rome "Tor Vergata"
via di Tor Vergata 110, 00133 Roma, Italy, e-mail: giovanni.costantini@uniroma2.it
[2]Conservatorio di Musica "L. Refice"
via Roma 25, 03100 Frosinone, Italy, e-mail: consmusfr@tin.it

Abstract

In this paper a new system for real time sound synthesis and processing is proposed. The system is based on a standard PCI full length card that can be used as a standard add–on board in a PC or as a standalone unit.
The board consists of four ADSP-21060 by Analog Devices clustered in a multiprocessor architecture sharing up to 512 Mb of DRAM, 4 Mbytes of Flash memory, two high speed serial interfaces to provide real-time control and an expansion module site to allow incorporating digital audio interfaces.
The system can be programmed using C standard or assembly languages or by the Csound language downloading the operating kernel in the memory of the processors.

1 Introduction

It's a matter of fact that computer music evolution is possible as a result of the maturity of certain technical components like Digital Signal Processors (DSPs).
The low processing power available few years ago, was the first limitation to real time execution due to an intrinsic delay between the setting of the composition parameters and the presentation of the result.
Nowadays the underlying technologies have improved. DSPs processing power exceeds mainframe computers of the recent past and new filtering algorithms allow to manage a wide variety of music synthesis applications.
The composer can rely on systems for building and for executing music while using them with traditional instruments.
Who needs to develop a system for real-time processing must takes in account the following requirements:

1- a sufficient level of interaction between the system and the composer
2- a total absence of delays during the execution.

While the first requirement needs the development of dedicated user interfaces the second one can be satisfied by a high processing power.
The latter can be achieved combining several DSPs into a single system. At the same time the algorithm will be broken into sequential steps, each processor performing one of the steps in an "assembly line" and interacting with each other through a single shared global memory, accessed over a parallel bus.
The proposed system goal is to offer the support for the most common methods of sound synthesis [1,2], like the additive synthesis, the subtractive synthesis, the PCM synthesis, the frequency modulation synthesis and the synthesis by physical models. A calculation subsystem composed by four processors connected in a multiprocessor architecture allows to obtain 1 Gflop of computing power.
System inline programming will be performed through the Csound language who became a standard "de facto" in the computer music environment.
It allows to implement all of the synthesis methods known today, but it allows adding new software modules for new processing algorithms too, thanks to its modular structure.

2 SHARC DSPs

One of the biggest bottlenecks in executing DSP algorithms is transferring information to and from memory. This includes data, such as program samples from the input signal and the filter coefficients, as well as program instructions. Super Harvard Architecture (SHARC) by Analog Devices [3] is an improvement of the Harvard Architecture based on separated memories and buses for data and program instructions.
Two areas of improvements are important: the instruction cache and the I/O controller.
The instruction cache is a small memory that contains about 32 of the most recent program instructions. In typical DSP algorithms, like FFT or filtering, most of processor execution time is spent doing loops. If we previously relocate part of the data in the program memory, after the first transfer necessary to load instructions in the cache, additional loops program instructions are fetched

from the cache while data are fetched from data and program memory at the same time.

The I/O controller, is a dedicated logic unit that manages data transfers to and from the CPU. Serial communications are managed by two synchronous serial ports that operate at 40 Mbits/second with a 40 Mhz clock speed, while six parallel ports (link ports) provide a 240 Mbytes/second data transfer. Dedicated hardware allows these data stream to be transferred directly into memory via the DMA.

The main buses (program memory bus and data memory bus) are also accessible from outside the chip via the external port, providing an additional interface to off-chip memory and peripherals. This allows the SHARC DSPs to use four Gigaword (16 Gbyte) of memory, accessible at 160 Mbyte/second on a 48 bit wide bus. The ADSP-21060 contains 4 Mbit of on-chip SRAM, organized in two blocks of 2 Mbit each one, which can be configured for different combinations of code and data storage.

Each memory block is dual-ported for single-cycle, independent access by the core processor and I/O or DMA controller on separate on-chip buses. This allows two data transfers at the same time.

Distributed bus arbitration logic is included on-chip for simple connection of systems containing up to six SHARCs. The unified address space allows direct interprocessor accesses of each ADSP-21060's internal memory.

Also mapped in the unified address space is the host interface that allows easy connection to standard microprocessor buses, both 16-bit and 32-bit, with little additional hardware required. Four channels of DMA are available for this interface; code and data transfers are accomplished with low software overhead. Data and program transfers are controlled by two units called DAGs (Data Address Generators), one for each memory, specifying where the information is to be read from or written to.

The data register section of the CPU is used in the same way as in traditional microprocessors. In the ADSP-21060 SHARC DSPs, there are 16 general purpose registers of 40 bits each.

The math processor works on those registers and it is broken in three sections, a multiplier, an arithmetic logic unit and a barrel shifter.

Another interesting feature is the use of shadowing for all the CPU key registers. These are duplicated registers that can be switched with their counterparts in a single clock cycle. They are used for fast context switching, the ability to handle interrupts quickly.

3 Hardware architecture

To describe the system architecture we refer to Fig.1. On a block based description, all the subsystems can be viewed as stand-alone modules each one dedicated to a specific use. We'll start analysing the PCI interface.

Like previously stated, the system can be used as stand-alone unit or as an expansion board into a PCI based personal computer.

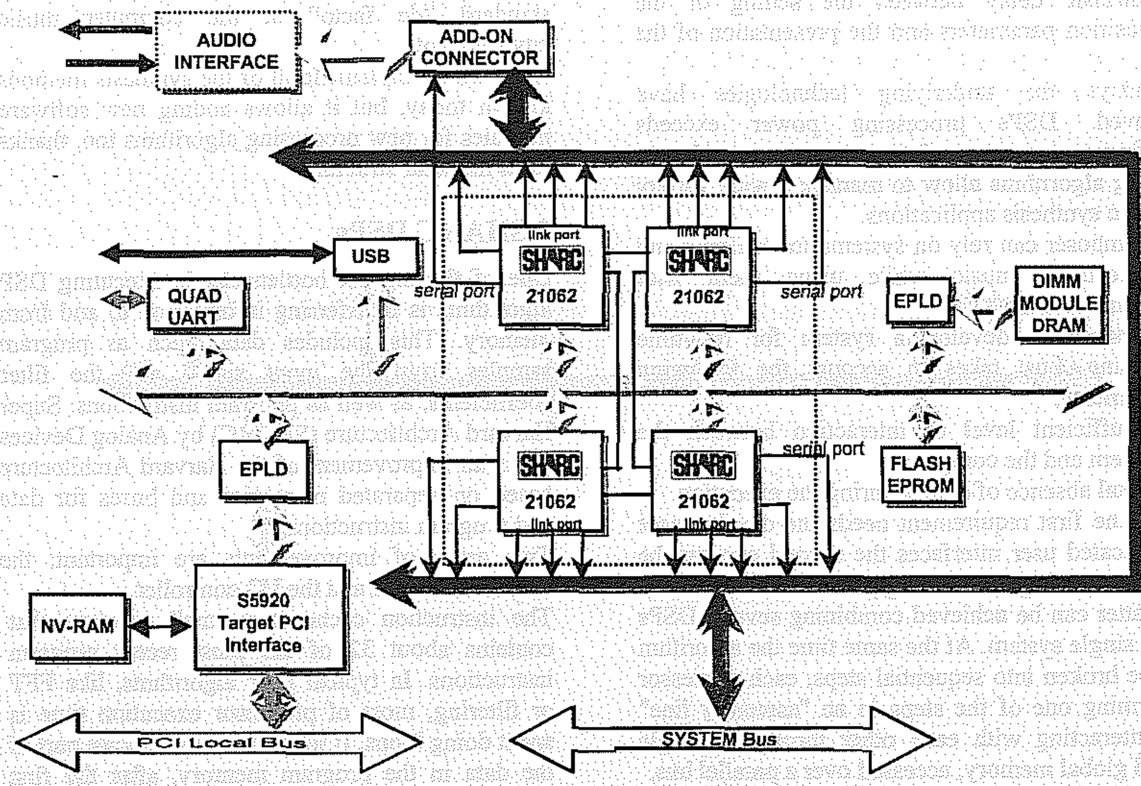


Fig.2 - Hardware architecture

3.1 The PCI interface

To support the required transfer rates to and from the multiprocessor subsystem the hardware uses the S5920 high performance PCI slave controller from AMCC [4], with data streams up to 132 Mb/s . The complex PCI bus signals are converted through the bridge S5920 into an easy-to-use 32 bit bus (the Add-on bus) which operates asyncronously with SHARCs bus. This allows the PCI based host computer to access all shared system resources like DRAM and FLASH memory, peripherals and internal DSP configuration registers.
Since the bridge is designed to interface to several different types of microprocessors, some glue logic has been added using a Programmable Logic Device (PLD) between SHARC processors and the bridge chip.

3.2 Multiprocessing

As we mentioned previously, the SHARC's on-chip multiprocessing arbitration logic provides interfacing to a variety of microprocessors, so no other hardware was needed to control bus accesses and timings.
To maximize system performances, all inter-processor communications (IPC) take place via the ADSP-21060's four 40 Mb/s link ports, removing a IPC bottleneck created by using non deterministic shared busses or the PCI bus. The on-board link ports are connected in a clustered point-to-point configuration.
A total of twelve SHARC links are taken off board, providing 480 Mb/s maximum interconnectivity to adjacent DSP hardware. Additional boards can be cascaded by interconnecting link communication ports (System Bus) available at the expansion connector located at the top of the board via a flat ribbon cable to create multiple board processing arrays.

3.3 The memory

The memory subsystem uses a shared memory architecture with a standard DIMM module and a 4 Mbyte FLASH PROM to keep the code necessary to boot the whole board via the root processor, when the board is used as stand-alone unit. All those resources are mapped into SHARCs addressing space and into the PCI memory space, so they are accessible by the PCI based host computer.
To provide control and timings for the DIMM module, a second PLD was used.

3.4 Serial interfaces

Also the two asyncronous serial communication controllers available for remote interfacing are memory mapped in SHARC's addressing space.
By a high performance UART controller, four MIDI protocol compliant units can be easily connected to the system to control SHARC behavior while real-time control is requested.
RS-232 protocol for remote console control can also be programmed.
A second 12 Mb/s USB controller make the system compatible with the last generation Pentium based boards serial standard. JTAG connectors are available for code development, system debugging, and program loading.

3.5 Audio interface

The system provides an expansion module site that allows it to incorporate digital audio interfaces (SP-DIF or ADAT protocols) or I/O converters like audio codecs. The six synchronous serial ports available on board connectors can drive up to 32 digital audio channels, to allow sound sources localization and moving through 32 loudspeakers settled on an half dome around the listener.

4 Software architecture

Two programming modes are possible, each one working at a different level of interaction with the user.
The low level interaction allows the use of Analog Devices assembly language or C language to implement algorithms and compile the object code [ELF executable] to be loaded into the SHARC DSP. Code compilation and loading can be executed by tools supplied by Analog Devices as well as by other tools we purposely developed; it is also possible to develop custom interface tools for PC, using suitable libraries, available both for Windows 9x and Linux.
The high level interaction allows to directly use CSOUND programming language. A software package for the system has been developed taking as a model the CSOUND language developed at MIT by Barry Vercoe, and available on most operating systems. The system is then capable of reading CSOUND-written files, achieving great flexibility, expandability (related to the possibility of implementing new modules) and, furthermore, compatibility with the large amount of available material; we must also consider that easy-to-use features and popularity made CSOUND become almost a standard among musicians.
CSOUND uses two types of files: the "orchestra" file, and the "score" file. The orchestra file contains the definitions of all instruments involved; an instrument is an algorithm for sound synthesis or sound processing, described using predefined opcodes like oscillators, filters, etc. appropriately connected together. Every opcode, when executed, calls a processing module. It is possible to easily add user-designed modules to the existent modules.

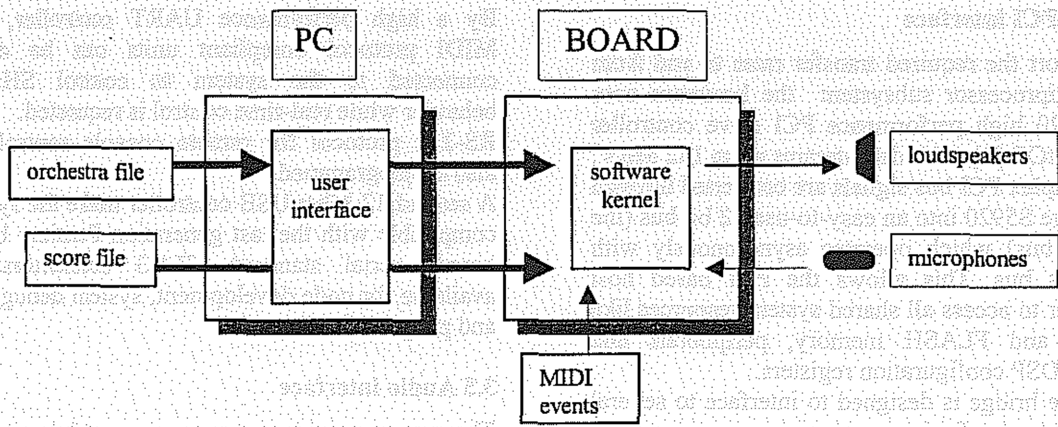


Fig.2 - Software architecture

The score file contains the data for runtime control of instruments, such as attack time of sound events, the duration, amplitude, frequency and other parameters necessary to the definition of a particular sound event.
A program running on PC reads the orchestra file, compiles it showing any error occurrence, and generates a data structure that can be efficiently stored in the DSP external memory to be used for sound generation. When executing a piece, the system can use both score file and MIDI commands sent to the system in real-time by external controls. The sound output uses the system audio channels, so the user can manage up to 32 output channels, useful for sound spatialization, e.g. placing 32 loudspeakers in a dome around the listener.

5 Acknowledgments

The authors wish to thank Daniele Casali for his valuable collaboration in the software development.

6 References

[1] *Computer Music synthesis, composition and performance*, Charles Dodge, Thomas A. Jerse, Schimmer Books, 1985
[2] *The computer music tutorial*, Curtis Roads, MIT Press.
[3] *ADSP-2106x SHARC DSP Microcomputer Family Manual*, Analog Devices Inc.
[4] *S5920 PCI Target/Slave Interface Databook*, AMCC, Applied Micro Devices, Inc., 1997

Music via Motion

Kia Ng

Interdisciplinary Centre for Scientific Research in Music (ICSRiM)
Department of Music School of Computer Studies
University of Leeds
Leeds, UK

mvm@compmus.com www.leeds.ac.uk/icsrim/mvm

Abstract

This paper presents an ongoing research prototype – Music via Motion (MvM), which aims to create an interactive audio-visual augmented environment to provide the users real-time control of musical events using their physical movements.

1 Introduction

There are many associations between music and movement, especially in dance and film music, where synchronisation is also an important issue [8]. In many cases, such as in a ballet performance, there is much interaction between the conductor (and hence the orchestra and the music) and the dancer (the movement) who may lead or follow each other to produce a coherent performance.

MvM is designed to be a dynamic real-time performance tool and is a motion sensitive system intended to track meaningful activities in the visual domain and *translate/map* them onto musical events.

There has been considerable research of sensor-based gestural control for interactive performance, for example the AtoMIC sensor/MIDI-interface from IRCAM [1] and the DIEM Digital Dance system [12, 13, 14]. In this project, we attempt to minimise the constraints on movement introduced by body-mounted sensors by using small and non-intrusive devices and concentrate on applying techniques from computer vision [2, 6, 7, 15]. This enables freedom of movement for the participants, whether dancers, actors, musicians or the audience itself.

There has been an increasing growth of interest in this research area. [4] presents a comprehensive background survey of related projects, including STEIM's BigEye [3] and Rokeby's Very Nervous System [17, 18].

This paper outlines the MvM prototype and discusses interesting applications of the system, including a performance/installation design with dance, costume design and music. Future directions, including a multiple camera set-up, are briefly discussed.

2 Music via Motion (MvM)

MvM makes use of input from a video camera, and processes video frames acquired in real-time. The software detects and tracks visual changes of the scene under inspection, and makes use of the recognised gestures to generate interesting and '*relevant*' musical events using an extensible set of predefined mapping sub-modules. The prototype is portable, can be set-up easily in a public environment and is designed to be intuitive and user-friendly to minimise the time needed for familiarisation.

MvM uses a differencing tracker to detect motion. The tracker is sensitive under a range of lighting conditions and it is convenient to use since the user does not need to wear any sensors or markers.

2.1 Modules

MvM consists of five main modules:

- A data acquisition module, which is responsible for communication with the imaging hardware.
- A motion-tracking module, which detects visual changes. Currently MvM uses a differencing tracker, involving the subtraction of the current frame from the previous frame to detect changes between contiguous frames.

- A music-mapping module, which consists of an extensible set of mapping sub-modules, for translating detected visual changes onto musical events.

- A graphical user-interface module, which enables online configuration and control of the musical mapping sub-modules, and provides overall control of the scale type (tonality), note filters, and pitch and volume ranges.

- An output module, which is responsible for the audio and graphical output.

The main window of the system offers a graphical user-interface for the configurable parameters, the choice of the tracking algorithms, and other options. There is also a live video window, displaying the camera view, and a motion tracker window, highlighting the areas with detected movements.

The system is intended to be lightweight, portable and efficient. It is implemented in C++ with Microsoft Video for Windows (VFW) and it has been successfully tested with various commercially available VFW compatible frame-grabbers, including web cameras with parallel and USB interfaces.

2.2 *Default Musical Mapping*

With this system, the user can be both the audience and the performer, controlling the events in visual and musical domains. Currently, MvM has been equipped with several mapping functions, including a simple distance-to-MIDI-events mapping with many configurable parameters, such as scale type, pitch range and others. Parameters of motion such as proximity, trajectory, velocity and direction can also be tracked and mapped onto musical parameters such as pitch, velocity, timbre and duration. By default, the mapping module translates horizontal movement onto pitch. Imagine a virtual keyboard in front of a user: by waving his/her hand from left to right, the user plays a series of notes from a lower pitch to a higher pitch. The vertical axis is used to control volume – the height at which the activities are detected is mapped onto loudness. Motion at a higher position is translated to a louder sound and motion at a lower position is mapped onto a softer sound.

MvM also offers user configurable '*active regions*' where detected visual activities in certain areas can be mapped onto different MIDI channels. By default, the system divides the scene under inspection into a number of equal size regions, and translates any detected visual changes in each region onto a user-definable MIDI channel. Figure 1 illustrates a user controlling different MIDI channels (with different sounds) in four active regions (left and right hands in different regions).

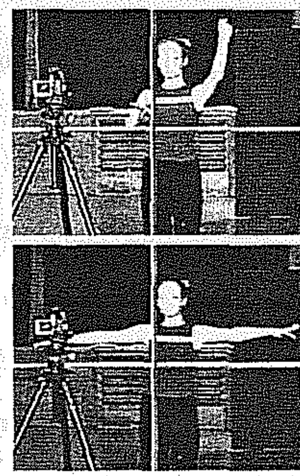


Figure 1: Active regions

Several visual feedback sub-systems of MvM are currently under development, to provide users with a graphical representation of what the system sees and detects, so that they can make any necessary adjustment when controlling and interacting with MvM. Future work includes background music generation using video-data from surveillance cameras, and virtual instrument design and interfacing inside an augmented 3D virtual environment [5, 7, 9, 10, 11].

3 Applications

There has been much interest in MvM as a tool to explore new directions, from a variety of disciplines. This include:

- Choreographers and dancers who are interested in exploring new choreographic possibilities inspired by MvM technology, and real-time control of the sound using their physical movement. MvM also enhances the dancer's awareness of movement and space, not only the movement of her/him-self, but also the movement of other performers (who may be obstructed in view), by listening to the sounds. Figure 2 shows some snapshots of dancers using the MvM system.

- Designers, employing interactivity to enhance design with added dimensions. In a later section, we briefly present an ongoing collaborative project called CoIN (Coat of Invisible Notes) which uses colour and motion tracking of specially designed costumes to trigger sound and musical phrases.

- Composers can explore new compositional frameworks, offering real-time control, with a collection of pre-composed short musical segments.

- There may also be applications for music therapists, to encourage movement, using this motion-sensitive system to provide interactivity and creative feedback.

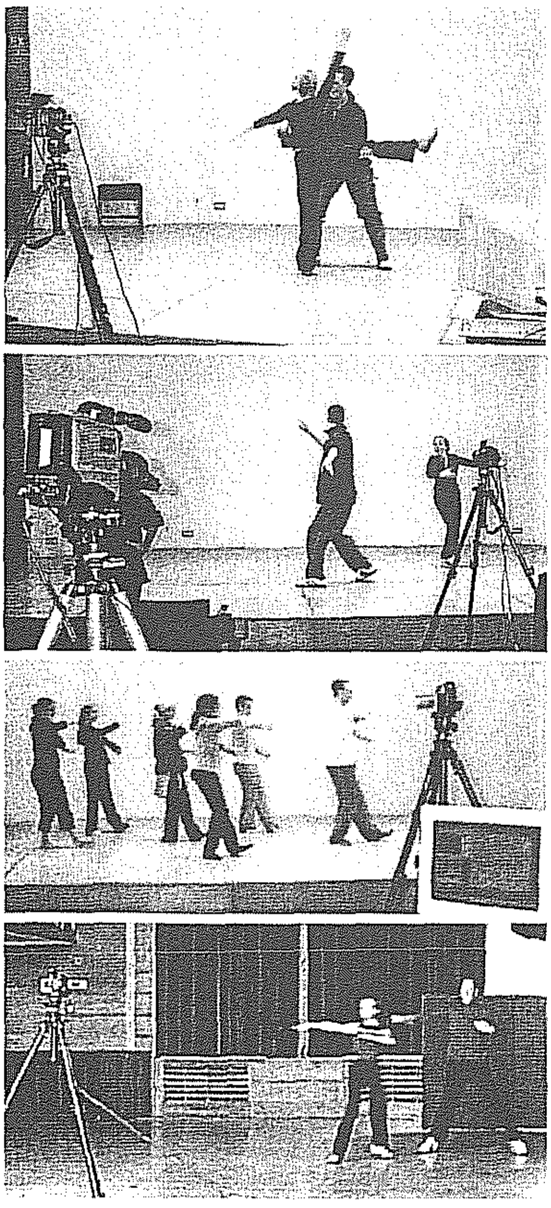


Figure 2: Dance with MvM

3.1 CoIN (Coat of Invisible Notes)

This is an ongoing project, exploring a creative application of the MvM technology with costume designs, dance and music. A particular feature of the costumes is that they are reversible and can be split apart into sections allowing the users to *re-assemble* and *re-configure* them so as to design their own image and to achieve different visual effects. These various changes in turn will be detected by MvM and will be used to alter the character of the musical responses.

In tune with the costume design, which will make use of everyday objects, the composition of the music will feature sound derived from these and other similar sources. The intention of the music is to bring familiar sounds into the performance so as to encourage the audience to perceive them differently in this artistic context. The relationship between music and sound will be explored, with the aim of expanding the audience's conception of music.

Musical phrases will be composed for use with the MvM software so that certain forms of physical movement will result in distinct musical responses. The phrases will be designed so as to be completely *re-configurable* (as with the costume) so that the performers/audience may re-arrange coherent musical structures from the musical materials that have been prepared. Phrases will contain melodic and rhythmic elements, sampled sounds, and electronically mutated versions of everyday sounds, which will allow for humorous or ironic juxtapositions.

4 Future Development

The MvM system is currently being extended to track visual activities in more than one view, with multiple cameras. A distributed MvM system with a music server is currently under development, which will offer additional control and features to the interactive environment.

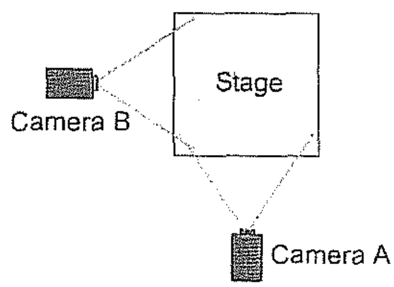


Figure 3: MvM set-up with two cameras

Figure 3 illustrates a two cameras set-up. Camera A could control the pitch and dynamic (as the default basic setting, discussed earlier) and camera B could control the MIDI output channel (selection of sound). The music server collects the resultant streams from all the motion trackers and performs the musical mapping functions. With this setting, a user/performer could *play* a violin when s/he is near to the main camera, but play a set of timpani when s/he is located further away from the camera. With multiple cameras, other enhancements may include stereo vision and 3D tracking.

5 Conclusion

MvM brings together multiple creative domains to create an interactive and augmented environment, providing the users with real-time control of musical sound by their physical movement. In front of the camera, the users seem to be able to *swim with the wave of musical sound* and pick invisible musical notes from the air. With the advancement in science and technology, it is hope that systems like the MvM will integrate art and science to offer artistic and creative sensory experience.

Acknowledgement

I would like to thank all the dancers and Sita Popat (Choreographer) from the Bretton Hall College, Eddie Copp (Head of Dance at Airedale Community Campus), Claire Nicholson (Freelance Dancer/Choreographer), Christine Hughes (Costume designer), David Cooper and Ewan Stefani (Composers, University of Leeds) for their time, interests and collaborations in the project. This project is receiving sponsorship from the Yorkshire Arts.

References

[1] AtoMIC: www.ircam.fr/produits/techno/real/Atomic-e.html, Ircam.

[2] A. Baumberg and D. C. Hogg: "An efficient mehod for tracking using active shape models." *In Proceedings of the IEEE Workshop on Motion of Non-rigid Objects*, 1994 pp. 194–199.

[3] BigEye: www.steim.nl/bigeye.html

[4] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe: "EyeWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems." *Computer Music Journal*, MIT Press, Vol. 24 N. 1, 2000 pp. 57–69.

[5] O. Faugeras: Three-Dimensional Computer Vision: A Geometric Viewpoint, MIT Press, 1993.

[6] L. Harris and M. Jenkin (eds): Spatial Vision in Human and Robots, The proceedings of the 1991 York Conference on Spatial Vision in Human and Robots, Cambridge University Press, 1993.

[7] N. Johnson, A. Galata and D. Hogg: "The Acquisition and Use of Interaction Behaviour Models." *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1988 pp. 866–871.

[8] F. Karlin and R. Wright: *On the Track*, Schirmer Books, 1990.

[9] K. C. Ng, V. Sequeira, E. Bovisio, N. Johnson, D. Cooper, J.G.M. Gonçalves and D. C. Hogg: "Playing on a Holo-stage: Towards the Interaction between Real and Virtual Performers." *Digital Creativity*, Swets & Zeitlinger Publishers, Vol. 11 N. 2, 2000 pp. 109–117.

[10] K. C. Ng, V. Sequeira, S. Butterfield, D. C. Hogg and J.G.M. Gonçalves: "An Integrated Multi-Sensory System for Photo-Realistic 3D Scene Reconstruction." *Proceedings of ISPRS International Symposium on Real-Time Imaging and Dynamic Analysis*, Hakodate, Japan, 1998 pp. 356–363.

[11] V. Sequeira, K. C. Ng, E. Wolfart, J.G.M. Gonçalves and D. C. Hogg: "Automated Reconstruction of 3D Models from Real Environment." *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, Vol. 54 N. 1, 1999 pp. 1–22.

[12] W. Siegel: "Two Compositions for Interactive Dance." *Proceeding of the International Computer Music Conference*, 1999 pp. 56–59.

[13] W. Siegel, W and J. Jacobsen: "The Challenges of Interactive Dance, An Overview and Case Study." *Computer Music Journal*, Vol. 22 N. 4, 1998 pp. 29–43.

[14] W. Siegel and J. Jacobsen: "Composing for the Digital Dance Interface." *Proceeding of the International Computer Music Conference*, 1999 pp. 276–277.

[15] D. Terzopoulos and R. Szeliski: "Tracking with Kalman snakes." *Active Vision*, 1992 pp. 3–20.

[16] M. Wanderley and M. Battier, eds.: "Trends in Gestural Control of Music." Ircam – Centre Pompidou, 2000.

[17] T. Winkler: "Creating Interactive Dance with the Very Nervous System." *Proceedings of the 1997 Connecticut College Symposium on Art and Technology*, New London, Connecticut: Connecticut College, 1997.

[18] T. Winkler: "Composing Interactive Music: Techniques and Ideas Using Max." Cambridge, Massachusetts, MIT Press, 1998.

Scorebot: Toward an Automatic System for Film Music Composition

Steve Pierce, David Cooper, Kia Ng

Interdisciplinary Centre for Scientific Research in Music (ICSRiM)
Department of Music, University of Leeds
Leeds, UK

scorebot@compmus.com www.leeds.ac.uk/icsrim

Abstract

This paper describes a system which manipulates music to be used in a film score. Given a musical theme and information which describes a scene (emotional content and timing and importance of events), the software can create a variation on the theme which conforms to the scene.

1 Introduction

Scorebot is a tool for a composer to use to aid in the realization of a film score. It is a musical assistant whose primary function is to compose variations on themes which fit specific scenes. Its decisions are by no means final. It is up to the end user to decide whether certain variations will actually be used, or even to do further work on the variations to make them conform to their own vision of the score more accurately.

The ultimate goal of this project is to develop a software program that automatically composes a film score based on input about timing of visual events and subjective mood changes. Such a program would be useful to producers and directors from any musical background when a quick and easy solution is required. This application can take input from a composer, but it would also work effectively without explicit musical input.

Scorebot, at its core, is an original database application which stores and allows manipulation of data on two levels: the 'higher' functions which manipulate or generate the musical elements that will evoke certain moods, and the 'lower' functions which control data related to specific projects (the timing, mood, and thematic information that forms the basis for the actual score).

The higher functions will be carried out by a potentially endless number of external programs, or 'agents' which may be running on the same or a different machine. Each agent will be responsible for a single task, such as: shortening or lengthening a musical phrase to fit a scene; or, altering a musical phrase to evoke a different mood. These external programs will be required to conform to a protocol for communicating with the main *Scorebot* engine.

The lower functions will form the basis of the main application and will be responsible for storing the information and managing the communication with the external programs. An Object Oriented (OO) approach is preferred because it encapsulates both the data structure and the functions which operate on that data.

Together these elements comprise a compositional tool which can compose variations on themes which are designed to be dropped into a soundtrack and fit perfectly.

It is also understood that film is not the only medium for which this application may useful. Other visual media, or even audio media (any situation where music is required to accompany some kind of non-musical presentation), can benefit. Perhaps the most important of these moving forward is interactive media where music is required to respond to unpredictable input.

2 Theory

'Many prominent composers increase the quantity of their musical output by employing musical assistants.' [9] *Scorebot* is just such a musical assistant, employed to increase a composer's output by helping the composer to develop themes.

Film music is particularly well-suited to a computer application because of the need to fit music more or less precisely to the action on the screen. The accuracy required is often down to the frame ($1/24^{th}$ of a second in film).

The composer likely to be best served by *Scorebot* is probably a computer-literate, traditional composer. This is firstly because the application is obviously computer-based and will store musical information in digital form, and, secondly, the types of functions that will be applied to the manipulation of musical data are likely to be based on traditional rules of melody, harmony, and rhythm which are more easily applied to traditional music. Functions which perform other, more avant-garde types of manipulation will certainly be possible, however.

Scorebot's functions can be broken down into two main categories: timing functions which manipulate of the timing of musical events in relation to the screen events and 'mood adjustment' functions which alter the quality of the music to evoke a different mood or 'feeling.' The timing functions will make changes such as repetition of sections, slight tempo adjustments, lengthening or shortening of notes, etc., but will attempt to do so without altering the character of the piece. Mood adjustment functions, on the other hand, may change the key or mode of a piece, make tempo adjustments, lengthen or shorten notes, but will do so to change the character of the piece and not attempt to alter the overall length. This is because a piece which is being changed for emotional character may have already been modified to fit a scene, and overall length changes will not be desirable in this circumstance.

Under normal circumstances, a film is 'spotted' (viewed) in its 'final cut' stage (the final edit) by the composer, music editor, producer and director. 'The discussion which takes place at this session ... aside from the actual composing of the score, is probably the most critical aspect in the process of providing music for motion pictures.' [8]. The result is a cue sheet (what Prendergast calls 'timing breakdown notes', 'timing notes', or 'breakdowns') which outlines the points in the film at which there should be music and details the timing of the cues required. *Scorebot* will map musical information to this cue sheet.

The mood adjustment functions, on the other hand, deal with the emotional content of musical pieces, which is highly subjective. Meyer describes emotion as the frustration of expectations. [6] These expectations are likely to be similar in people who share musical experience, as the composer's initial inspiration is surely based on his or her own experience. [2] *Scorebot*'s extensible design will allow for as many different functions as there are subjective opinions about the emotional content of a piece.

Scorebot functions will have to apply their own rules as to the assignment of meaning in musical pieces. Musical analysis will be brought into play in these circumstances. This will affect the design of both timing and mood adjustment functions. Each individual function will have to decide for itself which aspects of meaning are important in order to understand how to process the piece correctly.

Nicholas Cook describes music analysis as 'the practical process of examining pieces of music in order to discover, or decide, how they work.' [1] Because *Scorebot* is, in essence, a 're-composition' tool, analysis will play an important role. A piece will have to be analysed for its fundamental qualities in order for it to be re-presented in another context while maintaining those qualities. This might include, for example, a desire to maintain a composer's style in the variation which is created.

Music must not only be analysed for some notion as to its meaning or emotional significance, but it must also be 'recomposed' or varied to fit the scene. This means that algorithmic composition techniques must be applied. Notes must be chosen which are appropriate under the circumstances. Broadly, this is done through either deterministic or stochastic principles. [9] Deterministic procedures generate notes using fixed functions which take 'seed' data as input. Stochastic functions on the other hand choose notes based on random functions, but then notes are kept or discarded depending on limiting factors which are defined by the composer. Both deterministic and stochastic procedures are likely to find a place in *Scorebot*'s function repertoire. In fact, many other types of computer-based composition aids exist as well, from the simple style templates of accompaniment programs to the use of fractals and neural networks [9] and any number or combination of techniques for recomposing a musical piece will be possible.

3 Design

Scorebot is a framework. In the broadest sense, it is a mechanism that can store, manipulate, and output musical information in relation to another set of information which describes events that co-exist with or are being supported by the music. These other events are usually visual events in the context of a film, but can be other sound effects or dialogue, or any other medium for which music can act as accompaniment.

The framework must allow for the following with regard to both the musical and non-musical event information:

- Import/input
- Store
- Display
- Edit
- Export/play/print

Additionally, the framework must provide an Application Programmer's Interface (API) for functions, to be defined by the user community, which will manipulate the musical information. This API must pass to the function non-musical information (a scene) for reference and musical information (a theme) for manipulation. It must also accept as return data the modified musical information and store this in its database as a variation on the theme.

A modular approach is preferred. On the larger scale, this will keep the main, essential functioning of *Scorebot* (storage, input and output of information) separate from the individual processes which manipulate musical information. By doing so, management of the *Scorebot* application will be facilitated and more efficient extension of the application will also be possible.

On a smaller scale, keeping the individual musical manipulation functions separate from each other will allow composers and developers to focus on discreet elements of musical manipulation and mix and match functions for a virtually infinite variety of manipulation. 'Composition problems are notoriously difficult to define precisely and completely, so satisfying one composer's needs may not lead to a universal solution. Sometimes it is better to provide a flexible toolkit...' [9]

There are a number of ways in which the API can be realized, but, in the interest of simplicity and flexibility, it will be important to limit the discussion to those techniques which are easiest to implement, both in terms of the development of the *Scorebot* application itself and the extension of the application by composers and developers.

In terms of flexibility, it is desirable that external functions can be written in any language. Each one should be a separate application which is executed by the main *Scorebot* application. Therefore the *Scorebot* application will need to be written in a language which can execute and receive data from external programs.

It would also be possible to have applications run as daemons (a 'background process that performs a system-related task' [7]) which would mean that execution would be much faster because 'start-up costs' would be eliminated. This not ideal, however because it would unnecessarily complicate the development of *Scorebot* functions.

In the interest of simplicity, the passing of data between *Scorebot* and its functions will be done locally, initially through the file system. This should increase efficiency of *Scorebot* development because developers will not have to be versed in communications protocols or the intricacies daemon programming. If running functions on external machines later became desirable, a future incarnation of *Scorebot* could use an existing protocol, such as HTTP, to exchange information. HTTP would have the added advantage of allowing the external programs to be run behind a web server, in which case the question of whether to run it as a daemon or an application (which needs to be executed each time it is run) could be a decision entirely left up to the function programmer. HTTP also provides a number of mechanisms through which to communicate with external applications including the Common Gateway Interface (CGI), Java Servlets, and many more. [4]

Scorebot will write a file to the file system, then execute the program chosen by the user. This program will then rewrite the file and return an exit status to *Scorebot*, indicating its success or failure. Common programming style would dictate that this be a '0' on success, and another number, corresponding to an error code, on failure.

The file will list the musical information defining the theme to be worked on and the visual information defining the scene to which it is being applied in some standard format which can easily be read by the function. The function can simply append its new musical information to the file, which would then be re-read by *Scorebot* after it has determined that the function executed successfully. This new musical information, produced by the function, can be stored as a variation on the original theme, which can then be previewed by the user.

Other methods of function invocation can be added at a future date.

Because the application will make use of both musical and visual (or non-musical) information, it will need a language (both internal and external) for describing each.

Internally, the application will use an OO approach to store data. In the OO model data and methods will be codified into classes.

A scene will be defined by the class Scene, which has as one of its attributes an array of objects called ScreenEvents. A ScreenEvent will be defined by its start and stop time, level of importance and emotional classification.

The musical theme will be defined by a class called MusicalPiece. A MusicalPiece can be a theme, a variation on a theme (as created by one or several functions) or a 'cue' which is a musical piece that has been designated for use as final accompaniment to a particular scene. Theme, Variation and Cue will, therefore, be subclasses of MusicalPiece.

A MusicalPiece will have as one of its attributes an array of objects called MusicEvents. A MusicEvent is, in essence, a musical note. A MusicEvent, therefore, will be defined by its start and stop time, pitch, amplitude and articulation. Other attributes of a MusicalPiece object will include base tempo, key, etc.

The musical language of choice for purposes of communicating with external programs (input and output) is MIDI. MIDI is ubiquitous and many tools exist for manipulating MIDI information in sundry ways. [10]

A human-readable language for describing musical information would also be useful for import and export from the main *Scorebot* application. ABC will be considered appropriate as it is easily read and defines aspects of music which correlate with the attributes of *Scorebot*'s MusicalPiece and MusicEvent classes. [11] ABC to MIDI converters also already exists, which make these languages compatible with each other.

Input and output of visual, or non-musical, information will be done directly through the *Scorebot* interface. A CueSheet object will contain an array of Scenes and will provide methods for displaying and printing this information. The cue sheet concept was used by Lucasfilm in the development of the SoundDroid application in the early 1980s [9] and is designed to be analogous to the type of notes from which film composers traditionally work [5, 8].

At the highest level will be the Project class, whose attributes will include a CueSheet object as well as arrays of Themes, Variations, and Cues (MusicalPieces). Other Project attributes will include project name, file name and user name. Object serialization, such as that provided by the Java Programming Language, will be defined as a method of the Project object. This will enable the entire project to be saved to disk or sent over the network to another *Scorebot* user, as the Project object will contain all the data relevant to the project in the form of the CueSheet and MusicalPiece objects.

4 Summary

Computer manipulation of music will become increasingly important as the quantity of material which requires accompaniment increases. This will be particularly true in interactive applications where the unpredictable factor of real-time user input will create a virtually infinite number of possibilities.

Scorebot provides an extensible framework for manipulating musical themes to fit specific scenes in a film or other medium. It aims to provide an interface which can be easily understood by composers, and is extensible in any programming language.

References

[1] N. Cook: *A Guide to Musical Analysis*, Oxford University Press, 1987.

[2] D. Cooke: *The Language of Music*, Oxford University Press, 1959.

[3] D. Flanagan: *Java in a Nutshell, 2nd Edition*, O'Reilly & Associates, Inc. 1997.

[4] J. Hunter and W. Crawford: *Java Servlet Programming*, O'Reilly & Associates, Inc., 1998.

[5] F. Karlin and R. Wright: *On the Track*, Schirmer Books, 1990.

[6] L. B. Meyer: *Emotion and Meaning in Music*, The University of Chicago Press, 1956.

[7] E. Nemeth, G. Snyder, S. Seebass, T. R. Hein: *Unix System Administration Handbook*, Prentice-Hall, Inc., 1995.

[8] R. M. Prendergast: *Film Music: A Neglected Art: A Critical Study of Music in Films*, W. W. Norton & Company, Inc., 1977.

[9] C. Roads: *The Computer Music Tutorial*, MIT Press, 1998.

[10] E. Selfridge-Field: *Beyond MIDI: The Handbook of Musical Codes*, MIT Press, 1977.

[11] C. Walshaw: *The ABC Musical Notation Language:* *www.gre.ac.uk/~c.walshaw/abc/* University of Greenwich website.

POSTER SESSION II

Decorrelation as a By-Product of Granular Synthesis

Chris Rolfe, Damián Keller

Third Monk Software / CCRMA
Vancouver, Canada / Stanford, CA

Abstract

Several researchers (Kendall, 1995; Truax, 1992) have noted that decorrelation occurs as a by-product of granular synthesis (GS). Decorrelation between grain streams is responsible, for instance, for the unique stereo and panning effects that have been achieved under GS transformation. The correlation measure itself, however, is not generally explicit or variable within existing synthesis models. The following paper describes a systematic approach to granular decorrelation, relating individual parameters to their effect upon grain-to-grain, cross-channel (stream) and instance (event) signal correlation.

1 Introduction and Definitions

The cross-correlation measure of two signals is a significant predictor of many perceptual phenomena including spatial imagery [3], constructive and destructive interference [1], echo suppression [4] and externalization under headphone listening [2]. In multi-channel loudspeaker reproduction, inter-channel cross-correlation (ICCC) predicts the likelihood that a given signal will, all else being equal, suppress an earlier reflection arriving from a different direction (the precedence effect) [2].

The cross-correlation measure of two signals y1 and y2 is:

$$F(t) = \lim 1/2\ T \quad y1(t) * y2(t+t)dt$$

To obtain a useful, single measure of cross-correlation, we take the peak value of the cross-correlation function normalized to a range -1.0 to 1.0.

Expressed statistically, the cross-correlation function is the covariance between two signals divided by the product of their standard deviations. In audio signal processing, it may be more familiar to describe cross-correlation as passing an input signal y1 through a weighted, moving-average filter, y2. The value, k, is adjusted for DC offset, RMS and group delay, and is presumed to provide a convenient measure of similarity between two sounds.

Decorrelation is defined here as *any* technique that reduces the absolute value of the cross-correlation measure between two signals y1 and y2 (not only allpass filters designed explicitly for that purpose).

For practical purposes, given an arbitrary function F(x), and excluding the trivial cases of identity F(x) -> y=x, delay F(x) -> y[t]=x[t-z] and phase inversion F(x) -> y=-x, the cross-correlation(x,y), is always $-1.0 > k < 1.0$, so we further restrict our definition of decorrelation to include only those techniques that permit some means of influencing the degree of correlation across the entire range for k, -1.0 to 1.0.

2 Importance of Correlation in Granular Synthesis

Granular synthesis (GS) of sampled sound, or, as it is increasingly known, time-scrambling granulation, is essentially a statistically-controlled mixing scheme that recombines thousands of short snippets (grains) into multiple channels of output. In such mixing systems, constructive and destructive interference is of paramount importance not only in determining subjective outcomes, but also in controlling overall output signal levels.

Furthermore, because grain rates usually lie on the boundary between audio and event rates, GS output may be best described as depending upon emergent perceptual properties, wherein correlation measures have been shown to be useful predictors. Correlation in GS successfully predicts chorusing and echo, volume and diffuseness, and the robustness of spatial effects.

Most published granular synthesis (GS) models also allow for stochastic variation of control parameters, usually as a parameter range specification, as does our model. One reason for introducing randomness is to distribute or 'smear' artifacts caused by windowing and granulation both in the time-domain (pulsing, or beating) and the frequency-domain (amplitude modulation and comb-filtering).

For example, an amplitude modulation of a complex source signal by a 40 millisecond triangular window at a scanning ratio of 1:1

distributes the input signal's energy equally between the original spectrum and generated AM sidebands. The sound is subjectively mechanical because the artifacts are bound, perceptually, in the frequency domain. The AM sidebands are phase-correlated with the input signal, and thus perceived in toto as a spectral modification to the original input.

By focusing on those parameters to the GS model that help to decorrelate the signal, we can contrive strategies to avoid or minimize such artifacts, or, indeed, tailor the artifacts to our musical purposes.

Finally, although it is correct in one sense to describe decorrelation as a by-product of granulation, it is one of the most characteristic outcomes of GS processing, and thus deserving of closer examination.

3 (De-)correlation at Various Levels

In granulating sampled sound for various applications, it is useful to consider three related, but distinct, cross-correlation measures (described below). In characterizing each level, we will make reference to two phases of granulation, *analysis*, during which the choice of position within the input stream(s) is made, and *synthesis*, during which the source is manipulated and recombined into multiple grain streams. For reasons of computational efficiency, analysis in real-time GS systems tends to be limited to advancing a pointer according to a set of parameter controls. Particularly in real-time systems such as ours, we speak of decorrelation as a by-product of the synthesis stage, without implying that an actual correlation measure is to be calculated during run-time.

The first level, grain-to-grain correlation, is defined as the cross-correlation measure between successive grains within a single stream. Given a sine tone input to the process, for example, grain-to-grain correlation varies as a function of source synchronization, that is, measures how closely the phase of the grain envelope function aligns with the sine tone's frequency. Variation in correlation in this simple case is slight, affected mainly by boundary and base-frequency considerations.

In more complex examples, however, or in the case where input has been analyzed to form a pre-selected pool of grains rather than a continuously sampled sound event, or finally, when the grain duration is smaller than the audio rate boundary (< 20 msec), then grain-to-grain correlation becomes more important. In this sense, grain-to-grain correlation predominates during the analysis phase, and is significant primarily in matters of synchronization.

The second level, stream-to-stream correlation, is defined here as the cross-correlation between grain voices, or, *streams*, each stream representing an independently controllable channel to be mixed or routed to a given output(s).

Stream-to-stream correlation is important during the synthesis phase and has the most predictive value in terms of how multiple streams will be perceived after they are combined. It is here that we find the characteristic GS effects such as increased volume and diffuseness.

And thirdly, output instance correlation is stream-to-stream correlation extended to include non-contemporaneous GS output streams, either at the event level, or generated on different occasions.

It is important to mention decorrelated instances in order to emphasize the profound ecological difference between identical and similar events, and the general desirability of decorrelation in the electroacoustic domain. A wholly deterministic synthetic process, for example, produces identical output from a given parameter set; serendipity and variation is removed, and the resulting output is exactly repeatable. In contrast, most GS systems rely upon stochastic variation of parameters resulting in similar, but nearly always unique outputs.

Not all parameters to a GS model affect output instance correlation equally, however: amplitude may be varied stochastically at a grain rate, for instance, but does not decorrelate one instance from another. The effect of individual parameters and their role in decorrelation is considered below.

4 Synthesis Parameters

Our granular synthesis model presents the user with the following parameters, each controllable independently per stream:

advance_rate:	controls time-expansion or -compression;
delay:	initial sample delay;
delay_range:	varies sample delay;
mod_phase:	initial phase of amplitude modulator;
envelope:	amplitude modulator function
duration:	grain duration (AM frequency);
duration_range:	varies grain duration;
amplitude:	grain amplitude.

Grain streams can be synchronous or quasi-synchronous according to the setting of delay_range parameter.

As granulation begins, each stream's read pointer is set to its initial delay +- delay_range. If no time-expansion or –compression is chosen, that is, if the advance rate is equal to the original sampling rate, then the process resembles a multi-tap delay line. Each delay tap, however, is also amplitude modulated by the grain envelope function, which in our case is usually a triangle or variable trapezoidal window with a peak equal to grain amplitude, and edges set to zero.

Each delay tap is updated at the end of each grain cycle according to a random value whose range is controlled by the delay_range parameter.

At the end of each cycle, the modulation function is zero, thus allowing us to move to a new value without encountering discontinuities.

The sum of the grain streams is described by the instantaneous impulse response shown below:

Careful selection of initial delays and phases allows for the creation of many traditional effects, such as comb-filtering and reverberation, although usually with the addition of a "beating" or amplitude modulation caused by the grain enveloping. This common side-effect of granulation, however, can be avoided simply by pairing grain streams such that, for each pair, the modulation functions combine to a constant amplitude [5]. Under this pairing scheme, the total number of taps available is equal to the number of grain streams divided by 2, since it requires two output streams to, in effect, cross-fade between successive delay values.

If all delay tap values are constant, then the process is time-invariant and determinate. Grain-to-grain correlation will depend upon the grain durations chosen and the nature of the input signal. Cross-correlations between paired grain streams will equal 1.0, and the cross-correlation between output instances will be also be constant at 1.0. The subjective impression when using short delays is that all streams tend to fuse into a single percept and location, and repetitions (instances) of the process result in exact duplicates.

When a stream's grain duration is allowed to wander by even a few samples, however, the cancellation of amplitude modulation within stream pairs quickly breaks down as the modulation phases within pairs are randomized. Desynchronizing the modulation phases, however, only decorrelates the output streams very slightly: amplitude modulation by any unipolar function, such as a triangle window, combines the input signal with generated sidebands and thus results in a highly correlated, if non-linearly distorted, output. Another way to look at it: as the number of streams approaches infinity, the combination of the modulation functions approaches a constant. For practical purposes, then, 24-64 streams is sufficient to distribute AM effects, although we note that amplitude variation does not, per se, decorrelate a signal (although in practice, quantization error can introduce some calculable, if not desirable, decorrelation).

The chief means of controlling decorrelation in the described GS implementation is to adjust the delay_range parameter in conjunction with the stream amplitudes. Varying a given stream delay by a random amount introduces phase-shifting causing the value k to vary dynamically from $-1.0\ldots1.0$. The precise amount of decorrelation depends upon the relation between grain duration and source content, but can, with practice, be tuned by ear to the desired result.

Perceptually, introducing randomness into the delay taps creates a chorusing effect between streams. By controlling the distribution, amplitude and range of these delays, we exercise reasonably effective control over both the stream-to-stream and output instance correlations. Generally, the rule of thumb is that greater delay variation in more prominent (louder, earlier) streams increases output instance correlation.

The final parameter considered within our model is the advance rate. Advance rate controls the average grain hop as the input buffer is scanned. As Jones and Park [5] note, transparent (high-quality) time-compression and –expansion require that grain-to-grain correlation be maximized to maintain source phase synchronicity and thus to minimize boundary and overlap artifacts.

Because we selectively introduce random variation into our stream delays, however, we are in a sense forgoing the goal of transparency in our GS model in favour of a thickening or chorusing, and thus decorrelating effect. Transparent time-expansion requires maximizing correlation at all levels, best suited to one or two determinate streams, while we are interested equally in useful applications of decorrelation and multiple streams.

Generally, we avoid undue artifacts by selecting grain durations and envelopes by ear, rather than by calculation. As noted earlier, the grain-to-grain correlation is more important during the analysis than the synthesis phase of GS, which in a real-time system where efficiency is a concern.

Another parameter commonly found in GS applications but not considered as part of the above implementation, is pitch-shift:

Pitch-shifting is best counted as a form of stream pre-processing, for the practical reason that the amount of decorrelation is strongly dependent upon the choice of algorithm: a simple, non-interpolating drop-sample technique, for example, introduces highly uncorrelated error into the signal, whereas more sophisticated algorithms may produce correspondingly less error. As with other uncorrelated errors, such as those caused by amplitude quantization or jitter, the error may *not* be in fact desirable, and is, in any case, not integral to granular synthesis.

5 Applications

5.1 Late-field Reverberation

Keeping in mind our impulse response above, we can model reverberation by decreasing the amplitude but increasing the delay range of later taps. As a result, streams with shorter term delays will be highly correlated with the original input signal, while those with longer delays will tend to be decorrelated. The result is an intuitive model of late-field reverberation, in which later signals are more diffuse. Two useful additions to this reverberation model include the introduction of a feedback mechanism to add IIR resonance, and varying the choice of modulation function to add control over spectral content.

5.2 Multi-channel Spatialization

Each decorrelated GS stream can also be routed to independent outputs, creating a unique type of spatialization. Merely routing copies of an input signal to multiple speakers would, owing to the precedence effect, create a phantom image at best, or collapse onto the speaker nearest the listener. Decorrelated grain streams, on the other hand, will tend to create a more complex spatial field. In conjunction with the stream-to-stream decorrelation described above for simulating late-field reverberation, rich spatial reverberators can be easily constructed from the relatively simply GS model described.

6 Conclusions

Cross-correlation measures have enormous predictive value when dealing with complex mixing operations, and in particular, when working at near-audio or audio rates. Our consideration of correlation measures above demonstrates that merely stochastically varying a parameter to a GS process may increase distortion or other effects but does not necessarily affect decorrelation. This has suggested several strategies for minimizing distortions and artifacts as well as articulated a basis for several techniques unique to GS. Additionally, because decorrelation in GS is achieved primarily through manipulation of time-delays, other parameters to the process can be fairly freely selected according to the purpose at hand.

7 Footnotes

[1] Augspurger, G., et al. 1989. Use of Stereo Synthesis to Reduce Subjective/Objective Interference Effects: The Perception of Comb Filtering, Part II. Preprint 2862, the 87th Convention of the Audio Engineering Society, October 1989, New York.

[2] Begault, Durand. 1994. 3D Sound for Virtual Reality and Multimedia. Boston: Academic Press.

[3] Kendall, Gary S. 1995. The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. Computer Music Journal, 19(4):71-87.

[4] Clifton, R.K. 1987. Breakdown of Echo Suppression in the Precedence Effect. Journal of the Acoustical Society of America 82:1834-1835.

[5] Jones, Douglas L. and Thomas W. Parks. 1988. Generation and Combination of Grains for Music Synthesis. Computer Music Journal 12(2):27-34.

[6] Truax, Barry. 1992. Discovering inner complexity: time-shifting and transposition with a real-time granulation technique. Computer Music Journal 18(2):38-48.

The effect of inharmonicity on the perceived quality of piano tones

Bruno L. Giordano (1), Davide Rocchesso (2)

(1) Department of General Psychology, University of Padova, Italy (email: brungio@tin.it)
(2) Department of Science and Technology, University of Verona, Italy (email: rocchesso@sci.univr.it)

Abstract

The relationship between spectral inharmonicity and the perceived timbre naturalness was investigated. Quantitative methods are used to establish a measure of this relationship for different fundamental frequencies. Qualitative datas are collected to study in depth the perceptual dimensions used by the subjects to judge the perceived naturalness in piano tones.

1 Introduction

Inharmonicity of piano tones is due to the stiffness of the strings, which causes the frequency of vibration modes to be higher than the frequency of perfectly harmonic partials. The simulation of the dispersion of partials by means of physical modelling could require the use of high order filters, so that computational resources are excessive for real-time synthesis. As a consequence, measurements of the minimum amount of inharmonicity required in order to preserve a piano timbre judged as natural or non synthetic are needed. Back in time Fletcher [1] pointed out how inharmonicity in piano tones is essential in order to preserve naturalness and timbre warmth. Rocchesso and Scalcon [2] found recently that the amount of inharmonicity required to preserve a natural piano timbre decreases with increasing fundamental frequencies. The general purpose of our investigation is to reply the findings of the cited studies enlarging the qualitative observations about the perceptual dimensions used by subjects to discriminate between natural and non natural piano timbres. This will follow the acquisition of quantitative data by means of adequate psychophysical techniques. In analysing collected data we will then verify the weight of the spectral centroid in determining subjects responses. A general finding in timbre perception studies is in fact that the spectral centroid, a measure of the displacement of the high energy spectral components positively correlated with perceived brightness, explains most of the variance of subjects responses in timbre discrimination tasks [3, 4, 5, 6, 7]. As this has been found using different sets of stimuli, our purpose is to verify the weight of this physical/psychological dimension in discriminating between the timbre of synthesised piano tones.

2 Synthesis

The synthesis of the stimuli has been driven by the sinusoidal representation of the signal extracted by analysing the notes C1, C2, C3, C4 and C5, with a fundamental frequency respectively of 32, 65, 130, 261 and 523 Hertz, played with a Schulze-Pollmann piano. For the analysis we used the program SMS based on the Deterministic plus Stochastic Model of Serra [8]. All stimuli were synthesised keeping the formant structure, the spectral bandwidth and the micro-variations in amplitude and frequency of the original reference piano tones. Inharmonicity has been decreased lowering the average frequency of all the partials above a variable cut-off frequency. The difference between adjacent partials above the cut-off frequency have been kept constant and equal to the difference between the last two non manipulated partials. As independent variable we used the percentage of lowered partials out of the number of partials found in analysis; we will refer to this index as PH or percentage of harmonised partials. We preferred to use a relative variable, as the PH, in order to be able to compare the results obtained from stimuli with different fundamental frequencies and with a different number of partials. We used PH's of 0, 25, 50, 75 and 100%. So a stimulus with a 0% PH is the original reference tone, while in a stimulus with a 100% PH the partials form an harmonic series, being their average frequencies integer multiples of the fundamental frequency.

The elimination of the inharmonicity causes the perceived pitch to be lower than the one of the original tone. The deviation in the perceived pitch, consistent only for low fundamental frequencies and for high PHs, isn't however greather than 50 cents under the best conditions. For this reason we didn't equalise the pitch of the stimuli synthesised from the same original piano tone, as in previous works it was found that little pitch differences do not affect the timbral relationships among the stimuli (i. e. their positions inside the computed *timbral space*) [9]. Overall duration of all the stimuli has been equalised to 2 seconds applying a linear 200 ms. fadeout.

3 Experimental Procedure

To collect data we used the method of paired comparisons plus an additional procedure similar to those used in classical psychophysics to measure absolute thresholds. With the method *of paired* comparisons, we can obtain an order and a distance relationship for all the stimuli with the same fundamental frequency *with respect to* the psychological continuum of the perceived timbre naturalness. With the second procedure we estimate an *absolute threshold for* timbre naturalness in terms of PH, the threshold itself defined as the PH below which the stimuli are judged as having a *natural piano* timbre at least 50% of the times. In the first case subjects have to say which of the two coupled stimuli have the more natural timbre, while with the second procedure subjects judge one stimulus at a time saying if the listened tone have a natural piano timbre or not. Coupled stimuli were separated by 1000 msec. After the two procedures had been applied subjects were interviewed about the criterions followed in executing the preceding tasks. We were particularly interested in the perceptual/phenomenical nature of the main dimensions of timbre used by subjects in order to discriminate between natural and non natural tones. For all the tests we used the same 13 subjects, all Piano graduates of Padua Conservatory. The overall duration of the experiment was almost of 30 minutes.

4 Results and discussion

Apart from the particular fundamental frequency, the first evident effect of increasing the PH is a progressive fall of the related scalar values, which stands for a decrease of the perceived timbre naturalness (Figure 1).

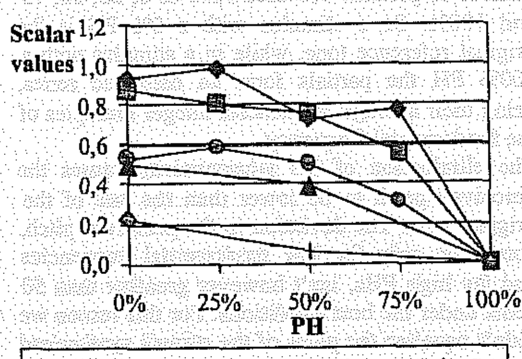


Figure 1: Scalar values for all the tested stimuli.

All the correlations between the scalar values and the PH, calculated for the different subsets of stimuli, synthesised manipulating the same original tone, are significant (Table 1).

Note	Correlation (r)	p-value
C1	*-0.890*	*0.043*
C2	*-0.892*	*0.082*
C3	*-0.884*	*0.046*
C4	*-0.930*	*0.022*
C5	*-0.892*	*0.042*

Table 1: Correlation between the scalar values and the PH for the different subsets.

Another information can be derived from the range of variation of the scalar values associated to all the stimuli synthesised by manipulating the same original piano tone. Defined the range as the difference between the highest and the lowest of the considered scalar values, the lower the range the harder is to discriminate between the stimuli in respect to the considered psychological attribute. The tendency in our data is that as the fundamental frequency increases the range of the scalar values decreases. The measured correlation between these two variables is −0.855 (p=0.002). So we can conclude that for higher fundamental frequencies the lowering of the spectral inharmonicity *has little* or no effect on timbre naturalness and on timbre itself.

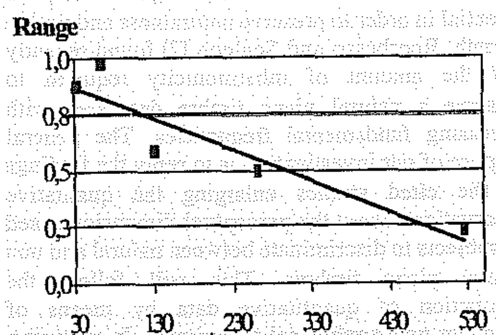


Figure 2: Range of variation of the scalar values versus fundamental frequency. A tendency line is displayed.

The decreasing in timbre naturalness with increasing PH *is consistent with* Rocchesso and Scalcon data [2]. The decreasing of the range of the scalar values with increasing fundamental frequencies agrees with Fletcher's observation that the effect of inharmonicity in determining the liveliness of the perceived piano timbre is important *only in the lowest* three octaves [1]. The results concerning the determination of an absolute threshold indicate in general that for higher fundamental frequencies we can eliminate an increasing amount of inharmonicity without any appreciable effect on the timbre naturalness. The general tendency found in data is in fact a relation of direct proportionality between the threshold in PH and the fundamental frequency. Nevertheless all the subjects judged as having a particularly unnnatural piano timbre all the stimuli synthesised from the notes C3 and C4 (medium-high range). For this reason thresholds obtained for C3 and C4 are lower than zero, which is a nonsense.

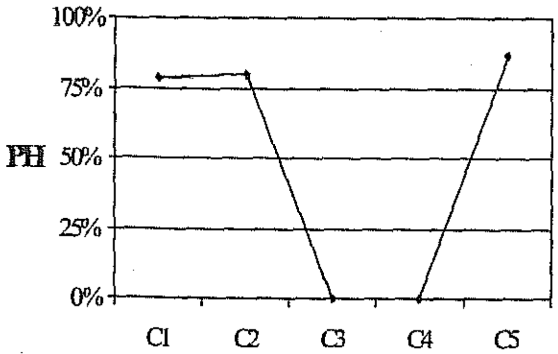


Figure 3: Absolute threshold for the perceived timbral naturalness in PH for the different notes.

This measurement artifact reflects a peculiarity of the brand Schulze-Pollmann used for the recordings of the original piano tones: some of the subjects recognised the brand used saying that it's timbre is tipically similar to the one of the clavichord in the middle-high range. This uniqueness of the piano used is an obstacle in the measurement of the relationship between the amount of spectral inharmonicity and the perceived timbre naturalness: for this reason further measurements with a different piano model are required.

The correlations between the average values of the spectral centroid of all the stimuli and the relative scalar values, measured for all the subsets, are not significant. This indicates that subjects in judging the perceived timbre naturalness used different perceptual dimensions than the one of the spectral brightness.

Note	Correlation (r)	p-value
C1	-0.806	0.099
C2	-0.714	0.176
C3	-0.857	0.063
C4	-0.682	0.522
C5	0.356	0.768

Table 2: Correlation between the scalar values and the mean spectral centroid

This conclusion is consistent with the analysis of the interviews conducted after the experimental phase. The two most important indexes used in discriminating between natural and non natural piano timbre have in fact been the rapidity of the attack (medium-high notes were judged as having a plucked-like timbre, similar to that of the clavichord) and the amount of time-fluctuations of the signal (subjects, especially for PHs of 100% reported a strong presence of, as they defined them, "beats"). Even if difference in perceived brightness among the stimuli were clear and present it is evident that subjects disregarded them as not being fundamental in discriminating between the timbre of the listened stimuli. These latter findings lead us to think, as Bregman himself did [10], that the indexes used by subjects in timbre discrimination or timbre scaling tasks are heavily dependent on the particular set of stimuli used.

5 References

[1] H. Fletcher, E. Blackham, R. Stratton: "Quality of piano tones", *J. Acoust. Soc. Am.*, 36, 749-761, 1962.

[2] G. Borin, D. Rocchesso, F. Scalcon, "Subjective evaluation of the inharmonicity of synthetic piano tones", 1998.

[3] J. M. Grey, "An exploration of musical timbre", Ph. D. dissertation, Stanford University, Stanford, CA, 1975.

[4] J. M. Grey, "Multidimensional scaling of musical timbres", *J. Acoust. Soc. Am.*, 61,1270-1277, 1977.

[5] R. Plomp, "Timbre as a multidimensional attribute of complex tones", in *Frequency Analisys and Periodicity Detection in Hearing*. R. Plomp e G. F. Smoorenburg editors. A. W. Sijthoff, Leiden, 1970.

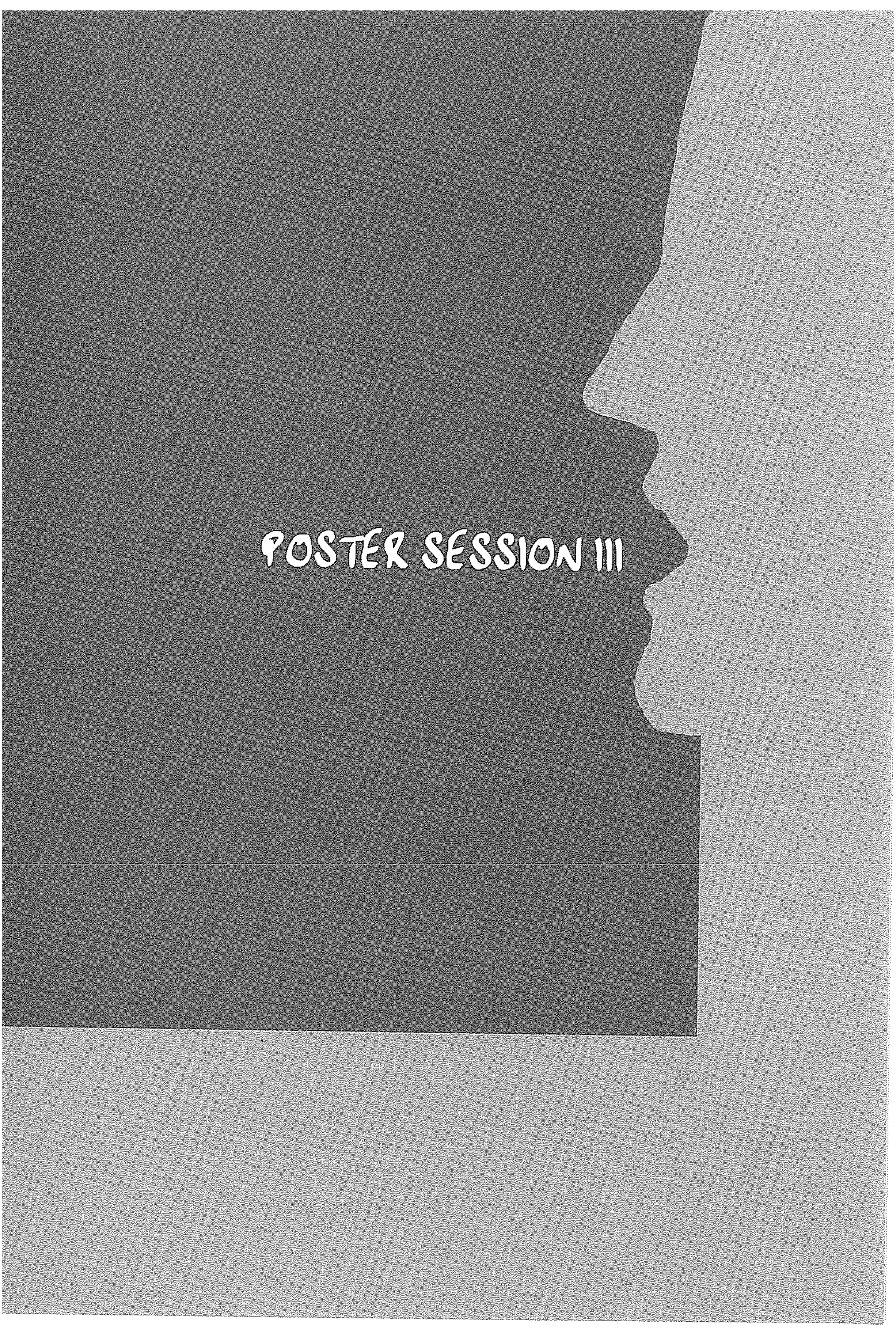
[6] R. Plomp, *Aspects of tone sensation*, Academic Press, London, 1976.

[7] G. Von Bismarck; "Timbre of steady sounds: A factorial investigation of its verbal attributes", *Acoustica*, 30, 146-159, 1974.

[8] X. Serra, "Musical Sound Modeling with Sinusoids plus Noise", in *Musical Signal Processing*. C. Roads, S. Pope, A. Piccialli, G. De Poli, Ed. Swets & Zeitlinger, 1997.

[9] J. M. Hajda, R. A. Kendall, E. C. Carterette, M. L. Harshberger, "Methodological issues in timbre research", in *Perception and cognition of music*, I. Deliege, J. Sloboda Editors, Psychology Press/Erlbaum (Uk), pp. 253-306, 1997.

[10] A. S. Bregman, *Auditory Scene Analysis*. MIT Press, Cambridge, Massachusetts, p.238, 1994.


POSTER SESSION III

POSTER SESSION III

THE MAKING OF *PASSIONE SECONDO MATTEO* BY ADRIANO GUARNIERI: AN OUTLINE OF SYMPHONIC LIVE-ELECTRONICS

Nicola Bernardini
Conservatorio "C.Pollini"
35100 Padova, Italy
nicb@axnet.it

Alvise Vidolin
Conservatorio "B.Marcello"
Venezia, Italy
Centro di Sonologia Computazionale
Padova, Italy
vidolin@dei.unipd.it

Abstract

This presentation describes the many steps involved in the production of a complex orchestra, soli, choir and live-electronics composition (Passione secondo Matteo by Adriano Guarnieri) which calls in its score for layered sound spatialization and motion, parallel digital processing of multiple sources, design, construction and sound reinforcement of special acoustic instruments, etc.

In this case, the term 'symphonic' (from συν—φονοσ, 'sound together', 'play together') is applied to live-electronics indicating a closely-knit network of interactions in a highly heterogeneous hardware/software environment.

1 Introduction

After the experiences of many significant works in the last four decades, live-electronics techniques in musical performance can be considered to have reached some degree of maturity. Several papers have been written on the subject and may be referred to for general techniques and implementations (cf. for example[2]).

However, the dramatic size increase of current compositions in terms of live-electronics requirements, together with the rapidly changing world of concert hall environments call for new solutions towards the interaction of several performing environments.

2 *PASSIONE SECONDO MATTEO*

Passione secondo Matteo is indeed one such piece: commissioned by the Teatro alla Scala of Milan and composed by Adriano Guarnieri in 1999, it is a 40 minutes long work for soprano, countertenor, 2 soli (C, alto and bass flute, contrabass flute), choir, medium-sized orchestra, percussion, amplified iron cables and live-electronics.

3 Concert premises

The commission was assigned to Guarnieri as the central work of the Easter concert of the Jubilee year 2000, and was expressely conceived for the Basilica of San Marco in Milan. The Basilica of San Marco is a large rectangular church which measures approximately 28 meters wide and 70 meters long. The size and importance of the concert premises along with its particular acoustic properties have implied a number of additional difficulties, such as:

- the placement of the acoustic instruments
- the number of diffusion sources
- the remote positioning of the live-electronics equipment, which in turn implies
- a reduced-size remote control system from the center of the hall

Here's a plan of the hall that shows the displacement of instruments, speakers and live-electronic control:

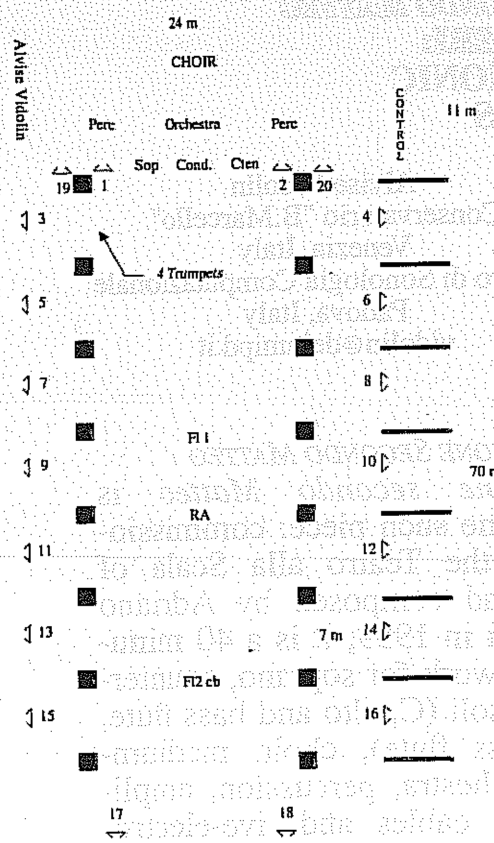


Figure 1. *Passione secondo Matteo*: premises plan

4 Live-electronics setup

In particular, the live-electronics of *Passione secondo Matteo* calls for:

- general 'transparent' sound reinforcement
- active sound reinforcement
- simultaneous sound motion in space of multiple sources
- simultaneous processing of multiple sources:
  - — the trumpet section
  - — the orchestral flutes
  - — the string section
  - — the bass drum
  - — the iron cables
- long feedback delay loops
- reverberation

As a general outline, the processing used in *Passione secondo Matteo* involved:

- one static sound reinforcement processor (*BSS Omnidrive 2x8 delay processor*)
- four different sound motion processing systems
  - a. permanent motion
  - b. automated motion
  - c. dynamic (amplitude-following) motion
  - d. hand-followed motion

  these were implemented with two simultaneous spatialization systems controlled by different versions of the same control software (*spAAce*, cf.[1])
- five different typologies of live sound processing:
  - a. derivative pitch shifting on orchestra flutes (code-named *flexatone* processing)
  - b. ring/FM-modulation combination controlled by the amplitude envelope on bass drums (code-named *tam-tam* processing)
  - c. critical-bandwith multiple harmonizing on orchestra trumpets and on iron cables (code-named *metallization* processing)
  - d. envelope remodeling on high strings pizzicati (code-named *rock-rain* processing)
  - e. infinite loop delay extensions on instrumental sounds and choir

  these were implemented with a mixture of processing systems (essentially, a *MAX/MSP* G3 workstation and a *Kyma Capybara 320 8x8* processor)

4.1 Transparent sound reinforcement

The capability of amplifying acoustical signals in a hidden way (that is without any noticeable effect) is called *transparent* sound reinforcement. Generally, this process is noticeable only by immediate comparison between the non-amplified signal and the amplified one.

It is achieved by progressively delaying the signal on speakers that lie closer to the listeners, so that the amplified signal arrives to the ears of the listeners *after* the (much softer) direct one. In this way, the *precedence* effect (also called *Haas* effect, cf.[3]) will produce the illusion that the amplified signal coincides in its position with the original source: the acoustical signal will sound louder without a noticeable contribution of the speakers.

4.2 Spatialization

The score of *Passione secondo Matteo* calls for a variety of different spatialization techniques according to the function ascribed to the movement of a given source. What follows is a schematic layout of the spatialization path of some instrument/voices:

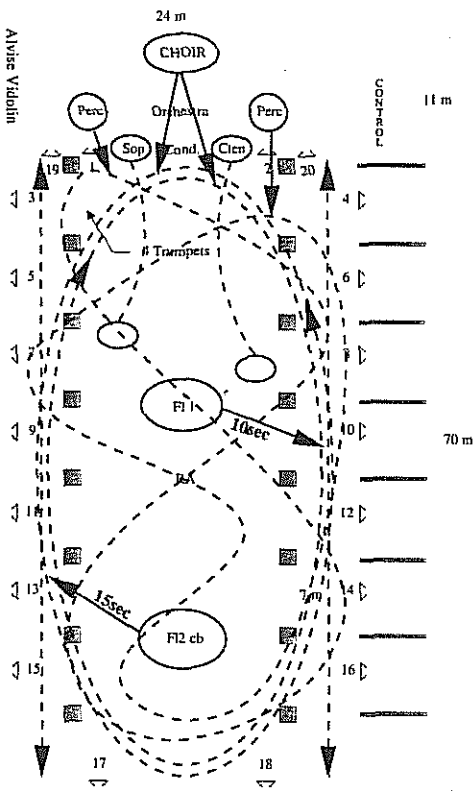


Figure 2. Spatializiation path

We have:

- the soprano and countertenor which step down from the stage and move into the public; the reinforced sound is supposed to move with them, so their position is hand-followed by a live-electronics performer by means of two x-y cursors
- the two solo flutes in the hall whose sound moves up and down the hall (longitudinally) with a fixed periodic rate
- the metallic percussions whose sound move randomly in the hall with different transition timings
- in the two choral passages of *Passione secondo Matteo*, the sound of the two choirs is sent around in the hall with slow circular movements
- the pizzicato strings sound is moved randomly in the hall with swift transition times
- the trumpet section sound and the iron cables sound get moved longitudinally in the hall calculating their position upon their instantaneous amplitude

(the pizzicato strings and the trumpet/iron cable spatialization paths are missing from the above picture for the sake of clarity).

Since most of the techniques used here for sound spatialization have been described elsewhere (cf.[4]) we will describe here only the last technique (the dynamic sound spatialization based on the amplitude of the trumpet section and of the iron cables).

Here's a schematic description of its functionality:

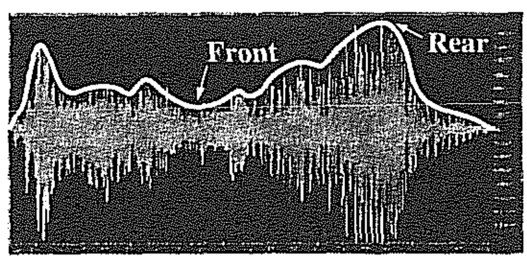


Figure 3. Dynamic sound movement

The amplitude of the trumpet section and of the iron cables get extracted and routed — properly filtered — to a multi-speaker front-rear controller.

4.3 Live sound processing

Here's a short description of some of the real-time processing of acoustical sounds.

The amplitude envelope of the orchestral flutes is detected at two different rates, a fast one and a slower one, and the following algorithm is applied:

$E_{pitch} = (A_{fast} - A_{slow)} \times K$

where E_{pitch} is varying pitch shifting factor controlled by the difference value given by the two envelope detectors. The resulting effect is that of a sort of *flexatone* instrument.

The amplitude envelope of the bass drum is picked up and properly gated to drive a combined FM/ring modulator. The frequency modulator drives a ring-modulated carrier; the index of the frequency modulator is controlled by the amplitude envelope of the bass drum. The effect is that of a *tam-tam* like sound which gets summed to the bass drum one.

Multiple harmonizers at microtonal intervals are used on trumpet sections and iron cables to multiply the dissonances of such instruments. This results in a strong metallic ringing of these instrumental sounds.

Envelope extraction is used also with the string section in a specific pizzicato passage: the onset time is detected and a very short clicking envelope is created from it — the resulting sounds (combined with their movement) give the impression of a 'rain of rocks' of some sort.

5 Conclusion

As this paper tried to describe (in a very synthetic form), the score of *Passione secondo Matteo* calls for a number of interacting live-electronics processes.

Consequently, the live-electronics environment required is extended and complex. Rather than single all-encompassing applications and machines, all processes are distributed onto several smaller applications and devices which 'play together' in a symphonic way like acoustical instruments do. This way, environment debugging and accidental malfunctioning are limited to single, small processes and are easily solved (even during rehearsals and concerts).

6 Aknowledgements

Such a complex system requires the hard work and collaboration of several people which we would like to thank here: first of all, Adriano Guarnieri without whose music none of all this would exist, then everybody at BH Audio Srl. (Dino and Massimo Carli, Enrico Dall'Oca) which have built the extraordinary reinforcement systems which we use, and last but not least Giovanni De Poli, who allowed us total freedom during experimentation and preparation at the Centro di Sonologia Computazionale of the Padova University.

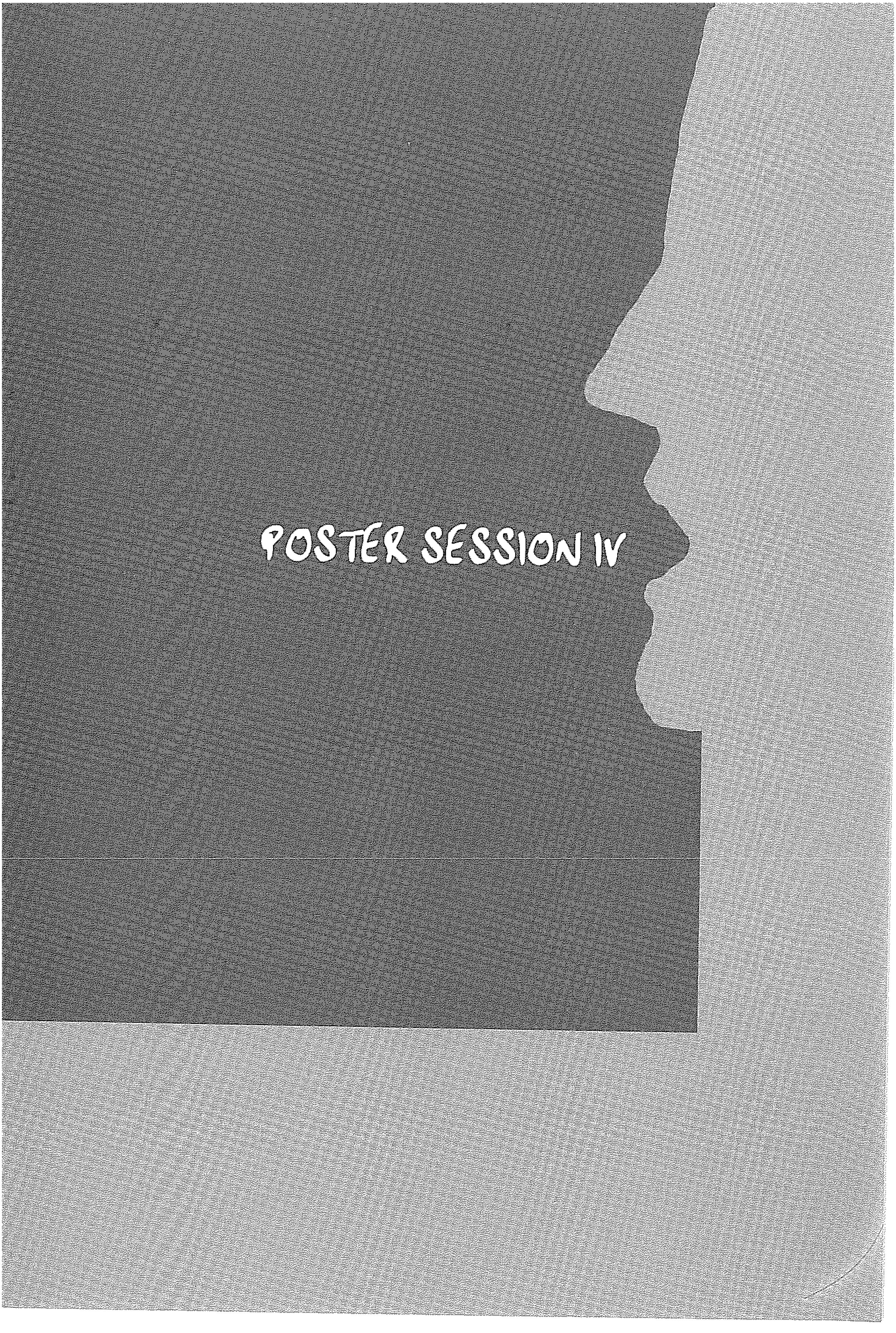
References

[1.] Belladonna, Andrea and Vidolin, Alvise, "Applicazione MAX per la simulazione di sorgenti sonore in movimento con dispositivi musicali a basso costo" in *Atti del X Colloquio di Informatica Musicale,* 351-358, Milano (1993).

[2.] Bernardini, Nicola, "Il Live Electronics" in *La Nuova Atlantide - Il continente della Musica Elettronica,* Vidolin, Alvise and Doati, Roberto eds., LIMB-Biennale di Venezia - Edizioni ERI. (Novembre 1986).

[3.] Haas, H., "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache," *Acustica,* **I,** 49-58 (1951).

[4.] Rocchesso, Davide and Vidolin, Alvise, "Sintesi del movimento e dello spazio nella musica elettroacustica" in *La Terra Fertile. Atti del Convegno 1996,* De Amicis, Maria Cristina and Prignano, Ignazio eds., 27-31, L'Aquila (1996).


POSTER SESSION IV

Enhancement of Optical Music Recognition using Metric Analysis

Kia Ng [1,2], David Cooper [1]
Interdisciplinary Centre for Scientific Research in Music (ICSRiM)
[1] Department of Music [2] School of Computer Studies
University of Leeds, Leeds LS2 9JT, UK
omr@compmus.com www.leeds.ac.uk/icsrim/omr

Abstract

The computer has become an increasingly important device in music. It can not only generate sound (audio synthesis) but is also able to perform time consuming and repetitive tasks, such as transposition and part extraction, with speed and accuracy. However, a score must be represented in a machine-readable format before any operation can be carried out. Current input methods, such as using an electronic keyboard, are laborious and require human intervention. Optical Music Recognition (OMR) provides an efficient and automatic method to transform paper-based music scores into a machine representation. In this paper, we outline the low-level pre-processing techniques and discuss the high-level musical rules, focusing on metric analysis, which may be applied to enhance recognition.

1 Introduction

The potential benefits of an Optical Music Recognition system were recognised over thirty years ago [1]. A robust OMR system can provide a convenient and time-saving input method to transform paper-based music scores into a machine-readable format [10] for widely available music software, in the same way as Optical Character Recognition (OCR) is useful for text processing applications [2,3,4,5,8,11]. As with other forms of optical document analysis, such as OCR, imperfections introduced during the printing and digitising process that are normally tolerable to the human eye can often complicated the recognition process. Musical symbols are highly interconnected: they may connect horizontally (for example in beams), vertically (for example in chords) or be overlaid (for example, in slurs cutting through stems or bar lines). Furthermore, when symbols are grouped (beamed), they may vary in shape and size, for example, consider the shape of isolated semi-quavers and the many possible appearances of four-semiquaver groups.

2 Low-level Approach

We propose a phased approach which divides and conquers complex notation by sub-segmenting the symbols into primitives before recognition and applying syntactical rules during reconstruction. After low-level processing and classification, we attempt to detect global information such as time and key signatures, and use this higher-level information to reconfirm earlier results. The prototype takes a digitised music-score grey image (300 d.p.i. with 256 grey) as input. An iterative thresholding method [6] is used to obtain a threshold value, and the image is binarised. Using the binary (black and white) image, the skew can be automatically detected by reference to the music-typographical feature of the roughly parallel stave lines, and the image deskewed by rotation. Stave lines form a grid system for musical symbols, most of which are related to the geometry of the staves: for example, the height of a note head must approximate the distance between two stave lines plus the thickness of the two stave lines. The sum of average distance between two stave lines and the average stave-line thickness form the fundamental unit used by the classification process [7].

3 Music Primitives

After segmentation using an image-labelling method, a rule-based classifier recognises primitive musical symbols that are isolated (e.g. dots and bar lines), and symbols that are normally located at certain positions with respect to their staff (e.g. clefs and rests). To deal with the problem of the inter-connected and overlapping features of musical symbols, we propose a sub-segmentation method to decompose them into lower-level graphical primitives, such as dots, note heads, vertical-bars and others, before recognition [8]. Figure 1 illustrates the process of sub-segmentation, where one beamed two-quaver group, one beamed four-semiquaver group, three chords and two overlaid curves, are sub-segmented into fourteen noteheads, nine stems, three beams, two curves, and isolated symbols (a crotchet rest, two bar lines and a sharp) which remain unbroken since they were classified during the initial pass. A basic graphically-reconstructed result is also illustrated. Each musical feature is passed to an iterative loop of sub-segmentation and recognition which repeats until classification, or termination by one of a set of other criteria. Details on the sub-segmentation processes can be found in [7,8].

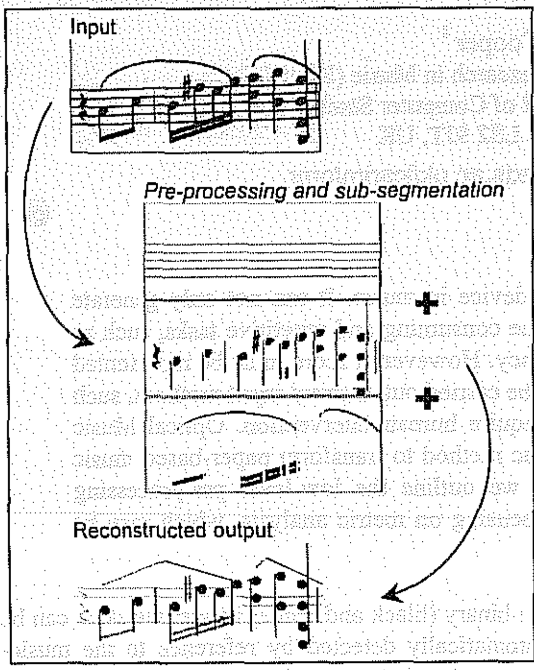


Figure 1: Pre-processing, sub-segmentation and reconstruction.

After primitives are recognised they are reconstructed using basic musical syntax. At this stage, extensive heuristic and musical rules may be applied to reconfirm the recognition. After reconstruction, we attempt to detect global information such as the key and time signatures and use them to provide evidence in the detection and correction of possible mis-recognition. It is clearly important that the OMR system should accurately detect the time signature of a piece or section of a piece of music in order to produce an accurate representation of the musical score. The intermediate results may be ambiguous or incomplete because of defects or artefacts in the source image – input images may often be extracts from a score, and not contain any or all such symbols. An approach to key-signature detection using note distribution can be found in [9]. In this paper we concentrate on time-signature detection and the use of rhythmical grouping to enhance the results.

4 Metre and Rhythm Clarification

Many possible time signatures can be unambiguously resolved using a simple count of the linearly disposed note values in a single musical strand lying between pairs of barlines (for domain-level reasons, the first and last bars of a piece are ignored). The value of a quaver is selected as the most appropriate unit for an initial estimate because 8 is usually the lowest value of the denominator in a time signature though 16 is sometimes found. The methods discussed below are as applicable to semiquaver beats as quaver ones. If the count produces a fractional number, the lower unit should be used for the initial estimate.

Quaver-count	Possible time-signatures	Comments
2	2/8, 1/4	1/4 unlikely
3	3/8	
4	2/4, 4/8, 1/2	1/2 unlikely
5	5/8	
6	3/4, 6/8	Both equally likely
7	7/8	
8	4/4, 2/2, 8/8	8/8 unlikely. 2/2 prioritises minims
9	9/8	
10	5/4, 10/8	10/8 unlikely
11	11/8	
12	12/8, 3/2, 6/4	All equally likely
13	13/8	
14	7/4	
15	15/8	
16	2/1, 4/2, 8/4	2/1 and 8/4 very unlikely

Table 1: Possible time signatures for 2 – 16 quaver bars (with no unresolved tuplets).

Two of the quaver-count totals produce problems of resolution – 6 and 12. In the case of a 6-quaver count, one simple (3/4) and one compound (6/8) reading is possible, and beaming information is required to identify the correct metre. A consistent feature of much music in 6/8 is the presence of groups which have the duration of a dotted crotchet starting on the first or fourth quaver of the bar. Grouping in 3/4 tends to favour 2-quaver groups starting on the first, third or fifth quaver of the bar, although streams of six quavers in a bar are often beamed together. Usually the ratio of the count of 3-quaver groups to the count of 2-quaver groups will give a strong indication of the metre, experimental data (taken randomly from a number of music scores) suggesting that a high ratio (> 2) implies compound metre and a low ratio (<0.5) implies simple metre. Unfortunately, this tendency is not strong enough to form a single unambiguous rule, and where the ratio lies between 0.5 to 2 further tests must be performed to resolve the ambiguity.

As was indicated above, compound-duple times divide the bar into two 'beats' such that beamed groups rarely cross the white space which separates them. Thus the boundary between the third and fourth quavers of a bar forms an important area for investigation. A count is made of all beamed groups which traverse the third and fourth (C_C) quaver boundary, and of beamed groups which connect the second and third quavers (C_m) and the fourth and fifth quavers (C_n). The ratio of $C_c + 1 : C_m + C_n + 1$ is computed (1 is included to avoid division by 0). A high ratio (> 2) indicates a simple time, a low ratio (< 0.5) indicates a compound time. In case of no clear ratio (0.5 to 2), further clues may be obtained using the notational conventions discussed below. If a time signature cannot be detected, a highest mean count of semiquavers per bar is taken as the most likely bar duration and used to make necessary correction only in the event of unclassified note- and rest-type; sub-bar level corrections are not possible.

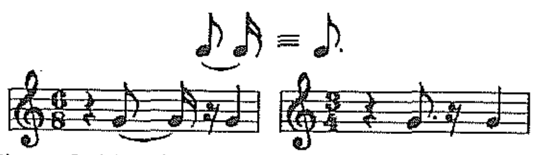
A semiquaver tied to a quaver has exactly the same duration as a dotted quaver, and the notation differentiates the metric organisation of simple and compound time such that white space is preserved between beats in a compound time (see Figure 2). A final test based upon this observation looks for individual notes that straddle the white-space border between the third and fourth quavers. Untied notes of duration longer than a quaver beginning at the third quaver are not often found in sections in compound time. Their presence can be regarded as a strong but not invariable indicator of a simple time signature (3/4).


Figure 2: Notational conventions for the depiction of a dotted quaver in compound and simple times.

Similar criteria to those discussed above apply when dealing 12-quaver counts. An initial test for 4-quaver groups starting on the fifth quaver is performed – if this grouping structure is found, the metre is almost certainly 3/2. To distinguish 6/4 and 12/8, the bars are subdivided into two equal halves and the tests described above to demarcate 3/4 and 6/8 applied. The correct metre will be:

$$\left(\frac{2 * \text{numerator of resolved meter}}{\text{denominator of resolved meter}}\right)$$

For example, if the resolved numerator is 6 and the resolved denominator is 8, the time signature is 12/8. The ratio thresholds chosen work by overall domination of beam and note type. Typically the algorithm needs at least half a page (A4) of music to work. It is not sensitive to possible errors introduced by early classification processes.

5 Problems and Ambiguities

Although the rules outlined above can accurately resolve most time signatures, there are cases where a composer has deliberately introduced ambiguities. The second movement of Ravel's String Quartet is prefaced with the time signature 6/8 (3/4), and throughout the movement the composer superimposes the characteristics of the two metres (see Figure 3). The second movement of Beethoven's Piano Sonata Op. 10 No. 2 has several passages in which the left hand plays in a clear 3/4 rhythm while the right hand plays in what could be interpreted as a 6/8 metre (see Figure 4). Composers sometimes rely on conventions such as the omission of the triplet sign; this can lead to confusion, and crotchets may be mistakenly converted to dotted crotchets. Figure 5 demonstrates an example from Beethoven's Piano Sonata Op. 7 (second movement) which may be interpreted as 9/8 by the algorithm.

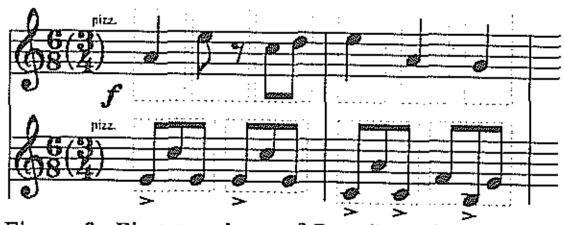



Figure 3: First two bars of Ravel's String Quartet, second movement, with double time signatures.

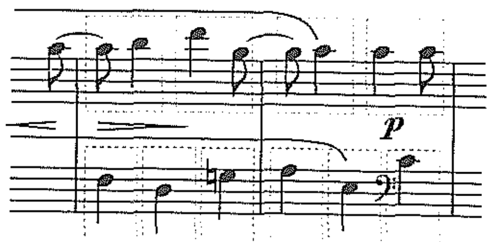


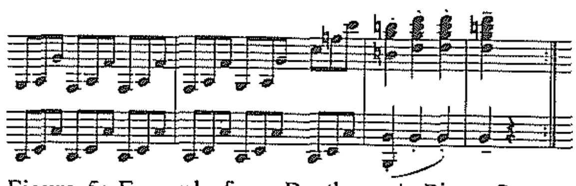
Figure 4: Apparently conflicting 3/4 and 6/8 meters in the second movement of Beethoven's Piano Sonata Op. 10 No. 2.


Figure 5: Example from Beethoven's Piano Sonata Op. 7 (2nd movement) with the omission of triplet signs.

6 Resolution of Ambiguously-notated groupings

Where the triplet sign is omitted, grouping may be similar to that found in compound times: for instance, two groups of triplet quavers notated in 2/4 will produce a pattern identical to that found in 6/8. See Table 2 for a comprehensive list of ambiguous groupings. Resolution of such incompletely notated groups may be clarified by:

(a) The presence of two-quaver or 'whole-beat' groupings in a number of non-contiguous bars - this will tend to imply a simple-time reading;

(b) The presence of what appear to be additive metres, particularly if their sums are odd (e.g. 3+2 quaver groups or 2+3+2 quaver groups, where the sums are 5 and 7 respectively) - such irregular groupings will also tend to imply a simple time signature particularly if they are not found consistently throughout a section of a piece.

Additive metres will tend to have other contextual musical clues, for example in the case of a piece of music in 5/8, other instrumental voices may have features such as regular crotchet-dotted crotchet rhythms which articulate the 2+3 division of the bar.

Quaver-count	Possible time-signatures
5	5/8, 2/4
6	6/8, 2/4
7	3/4
8	3/4
9	9/8, 3/4, 4/4
10	4/4
11	4/4
12	4/4

Table 2: Possible time signatures for 5 – 12 quaver bars with unmarked triplets.

Tuplets of higher value than 3, for example quintuplets, will generally be beamed together as groups where this makes typographic and musical sense (e.g. five quaver quintuplets beamed as a minim-value group). In some cases a triplet may be beamed with a note equal to the combined triplet value (e.g. a quaver followed by triplet semiquavers) to produce a grouping typographically similar to that of a quintuplet. Such groupings will usually be clarified as triplet-based by subsequent groupings of notes in which six notes are beamed together.

7 Conclusion

The results of this process provide important evidence for the internal organisation of notes and rests within bars of the music being parsed. It is possible to detect errors in terms of missing or mis-classified symbol types and intelligently replace them. With a known time signature, the total duration of a bar is established, and any discrepancy between the estimated duration and the expected duration indicates missing or misread events. The conventions of grouping as discussed earlier in the detection of the time signature allow us to work at the sub-bar level and provide evidence of missing or misread duration information. A limitation of the current implementation of the algorithm is that the system assumes that all bars of the segment under investigation have the same time signature.

Work in hand includes a scale-up of the system to deal with handwritten manuscripts [11] in which ambiguities and uncertainties are even more likely to occur than in printed scores. Any confirmation or enhancement from the above-mentioned algorithm will contribute to better recognition performance.

We believe that domain-knowledge enhancement is essential for complex document analysis and recognition. Other possible areas of development include melodic, harmonic and stylistic analysis to improve recognition results further.

References

[1] D. H. Pruslin: *Automated Recognition of Sheet Music*, Doctor of Science dissertation, MIT, 1966.

[2] D. Blostein, H. S. Baird: "A critical survey of music image analysis", *Structured Document Image Analysis*, eds. H. S. Baird, H. Bunke and K. Yamamoto, Springer-Verlag, 1992 pp. 405 – 434.

[3] E. Selfridge-Field: "Optical recognition of musical notation: A survey of current work", *Computing in Musicology*, Vol. 9, Center for Computer Assisted Research in the Humanities (CCARH), Stanford, Menlo Park, CA, USA, 1994, pp. 109 – 145.

[4] A. T. Clarke, B. M. Brown, and M. P. Thorne: "Inexpensive optical character recognition of music notation: A new alternative for publishers", *Proceedings of the Computers in Music Research Conference*, ed. A. Marsden, University of Lancaster, 1988 pp. 84 – 87.

[5] A. Ruttenberg: *Optical reading of typeset music*, Master thesis, MIT, 1991.

[6] N. B. Venkateswarlu and R. D. Boyle: "New segmentation techniques for document image analysis", *Image and Vision Computing*, Vol 13 No 7, 1996 pp. 573 – 583.

[7] K. C. Ng and R. D. Boyle: "Segmentation of music primitives", *Proceedings of the British Machine Vision Conference*, Leeds, UK., Springer-Verlag London Limited, 1992.

[8] K. C. Ng and R. D. Boyle: "Reconstruction of music scores from primitives subsegmentation", *Image and Vision Computing*, 1996.

[9] K. C. Ng, R. D. Boyle and D. Cooper: "Automatic Detection of Tonality using Note Distribution", *Journal of New Music Research*, Swets & Zeitlinger Publishers, 1996.

[10] E. Selfridge-Field: *Beyond MIDI: Handbook for musical code*, MIT Press, 1997.

[11] K. C. Ng, D. Cooper, E. Stefani, R. D. Boyle, N. Bailey: "Embracing the Composer: Optical Recognition of Hand-written Manuscripts", in *Proceedings of the International Computer Music Conference (ICMC'99) - Embracing Mankind*, Tsinghua University, Beijing, China, 1999 pp. 500 – 503.

A method for an objective comparison of rule systems for expressive rendering in musical performance

Patrick Zanon, Sergio Canazza, Antonio Rodà
Dept. of Electronics and Computer Science (DEI) – Centre of Computational Sonology (CSC)
University of Padova – Via Gradenigo 6/a – 35100 PD - Italy
email: patrick@dei.unipd.it, canazza@csc.unipd.it, ar@csc.unipd.it

Abstract

The use of rule systems for studying expressive deviations in musical performances is complicated by the absence of an objective method for their tuning to emulate a particular human performance. In this work will be presented a solution of the problem that also allows a comparison between the synthesis produced by different rule systems. To achieve best fit the theory of Hilbert space is used, by representing a performance or a rule as a vector in a "performance space", in which distances are defined according to the perceptive characteristics of the human ear and the fitting is obtained with an orthogonal projection. The results confirm this methodology, and give a numerical idea of how near the selected rule system can approach to a human performance.

1 Introduction

The studies of analysis of performances made in the last few decades, often use the word "expressiveness" to indicate, and somehow also to justify, the systematic presence of infractions to the prescription of the musical notation: these infractions would in fact be justified by "expressive reasons". However behind the term "expressiveness", lies a great part of a world which is still to explore; therefore it happens that a single word must summarise in a generic way the causes that go from the physiological necessities of the movements to the emotions of the performer, leading them all to the contrast with the notation, which for reasons of graphical economy, neglects them [1]. The analysis of the infractions has led to the formulation of some models that describe their structures, with the aim to be able to automatically synthesise what the interpreter makes unconsciously on the score.

Several models of expressive deviations in musical performance has been developed (e.g. [2], [3], [4], [5], [6]) with different characteristics and various degrees of flexibility, but all of them have the purpose of covering a range of "expressive" variations as wide as possible. However, if on one hand of the scale the elasticity allows a more and more kaleidoscopic and human synthesis, on the other the complexity of the models costs in terms of proliferation of parameters that make it difficult to use the models themselves. In other words it's very hard to tune the system in order to emulate a given human performance.

2 Methodology

In this work will be presented a method for evaluate the parameters of a system rule in an optimal sense, beginning from an interpretation of a given score by a professional pianist. Of course the meaning of "optimal evaluation" lies in the perceptive characteristics of the human ear.

The methodology can be applied to every rule systems which introduce variations in an additive way, but in this study, the model of the KTH [2] has been selected for its modularity and the existence of a good software (Director Musices) that allows to easily generate a different synthetic performance for each rule.

As indicated in Fig. 1, rule systems start from the score , called *nominal performance* (i.e. a literal or mechanic interpretation of the pentagram), and add some variations producing *synthetic performances* according to the specified parameters. In the first step a *unitary set of parameters* is provided to generate different synthetic performances in each of that only one rule with a unitary coefficient is applied per time. This whole of synthetic performances is called set of *rule performances.*

A program called "Interpretazioni"[1] has been developed to compare the rule performances with the *sample performance*, which is provided by a professional pianist, and generates the *tuned parameters*, used to spawn the final synthetic

[1] This software has been written in "Visual Basic for Application" provided with Microsoft Excel and can be obtained via email contacting the authors.

performance that would be the best approximation of the human one.
In conclusion, it is possible to relate the final synthetic and the sample interpretation to find differences that may suggest new rules or corrections to the model.

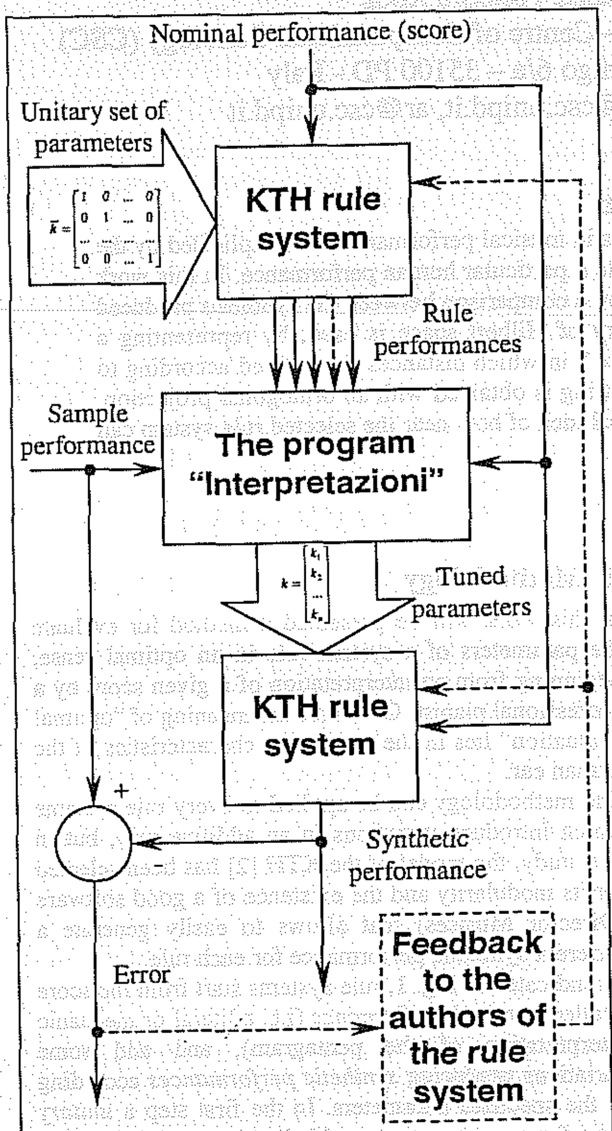


Fig. 1: methodology for parameter evaluation to fit a sample performance using the KTH rule system.

3 The performance space

The problem has been resolved at first by observing that the considered models start from the nominal performance introducing, in an additive way, duration and volume variations on some notes according to particular musical principles, each modulated by a characteristic multiplicative coefficient. It is just the set of these coefficients that will be the object of the research. The synthetic interpretation that will turn out will have to be as similar as possible to the sample performance. This suggests the way to go: by representing the performances with suitable vectors and formalising their concept of distance with a particular shape of the Euclidean norm, it will be possible to access to the results of the theory of Hilbert space, and in particular to the theorem of the projection, that is the best approximation in the mean square sense.
It is important to underline that each musical principle (or rule) is represented by a suitable performance; each of them is obtained from the nominal one to which the variations are added that are related to the rule modulated by a unitary coefficient.

Every performance of n notes corresponds to a vector in a ($3n$-1)-dimensional vector space **W**, in which the elements are:

- n sound intensities (Sound Level SL);
- n durations (Dr);
- n-1 intervals between notes (Inter On set Interval IOI).

W will be called "performance space". This vector space as it is, does not evidence the infractions inserted by the performer or by the particular principle of the expressiveness model; therefore a translation is operated by moving the origin in the point characterised by the vector of the nominal performance. So in the following paragraphs "performances" and "rules" in **W'** are to be intended as the difference between the vector that represents them in **W** and the vector of the *nominal performance*.

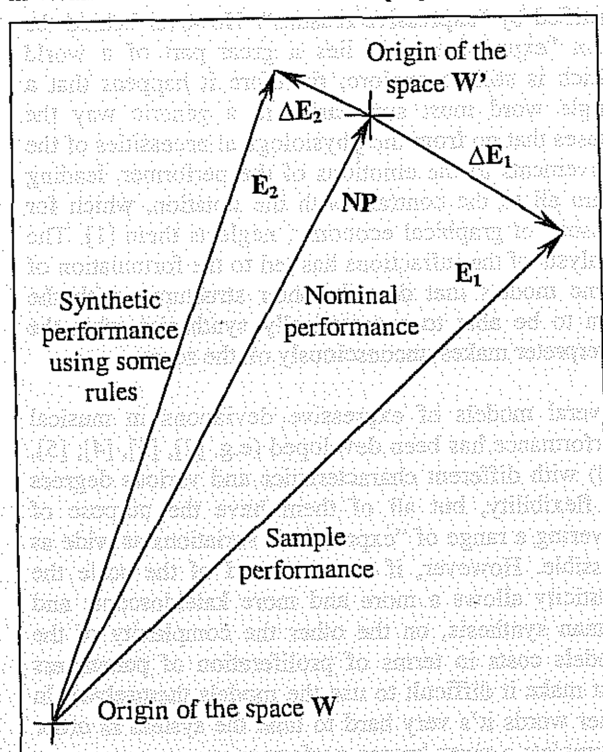


Fig. 2: the translation of the origin in the performance space **W**, operated to underline the infractions inserted by the pianist or by rules.

The Euclidean norm must take in account characteristics of the human ear, so several psychoacoustic principles were used. The main one is that variations of volume must be expressed in dB and variations of duration in percentage. At this point a criterion about how considering volume variations and time variations together must be selected. There are many possibilities and, after some attempts, it has been preferred to weight this different kind of variations so as to balance the effects of the just noticeable variations for human ear (we used 0.5 dB on average for sound levels and 5% approximately for the durations):

$$|\Delta \mathbf{E}|^2_{ear} = \alpha_{SL} \cdot \sum_{i=1}^{n} \Delta SL^2_{dBi} + \alpha_{Dr} \cdot \sum_{i=1}^{n} \Delta Dr^2_{\%i} + \\ + \alpha_{IOI} \cdot \sum_{i=1}^{n-1} \Delta IOI^2_{\%i}$$

where α_{SL}, α_{Dr} and α_{IOI} are the weights.

By providing the vector space **W'** with the inner product that induces the above-mentioned norm, a pre-hilbertian space is obtained, in which the terminology and the result of linear algebra can be used:

$$\langle \Delta \mathbf{E}_1, \Delta \mathbf{E}_2 \rangle_{ear} = \alpha_{SL} \cdot \sum_{i=1}^{n} \Delta SL_{1dBi} \cdot \Delta SL_{2dBi} + \\ + \alpha_{Dr} \cdot \sum_{i=1}^{n} \Delta Dr_{1\%i} \cdot \Delta Dr_{2\%i} + \\ + \alpha_{IOI} \cdot \sum_{i=1}^{n-1} \Delta IOI_{1\%i} \cdot \Delta IOI_{2\%i}$$

Under these conditions the rules constitute in **W'** a base (in general not orthonormal) of a vector subspace **S'** that represents the totality of the possible variations that the model in analysis can bring to the nominal performance.

In order to find the performance, produced by this model, as similar as possible to the sample performance, the theorem of the best approximation is used: this means that the solution is simply the orthogonal projection of the sample vector onto **S'**. The multiplicative coefficients (or the tuned parameters, represented by the vector k, see below) are the components of **ΔHP** (see Fig. 3) according to the base of the rules (**ΔR1** and **ΔR2**), and can be easily obtained with a rotation with dilatations and shears from the components of the orthonormal base (**B1** and **B2**), which is the result of the procedure of Gram-Schmidt.

More precisely, let be j the vector of the components of **ΔHP** according to the orthonormal base **B**={**B1**...**BN**}, and let A the upper triangular matrix, whose columns are the vectors a_i, representing the components of **Ri**, then

$$k = A^{-1} \cdot j$$

where k is the vector of the components of **ΔHP** according to the base of the rules **R**={**ΔR1**...**ΔRN**}, that is to say the tuned parameters.

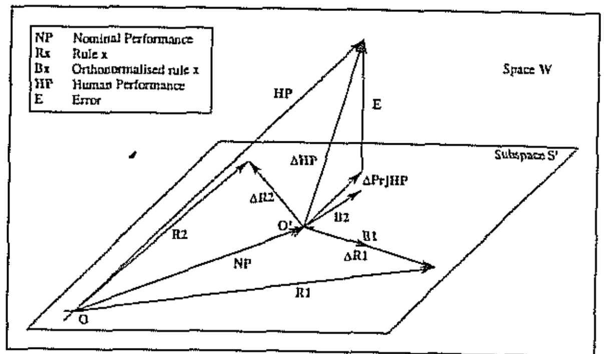


Fig. 3: interpretation in the performance space, with the best approximation in norm, that is an orthogonal projection.

The introduction of the inner product to evaluate the parameters, allows also some geometric considerations on the performances in **W'**, that is the possibility of measuring the angle between them (e.g. to verify if two different performances of the same score are more or less close in the performance space).

4 Results

After the tuning of the rule system, it is possible to evaluate how near the model can approach to sample performance. The use of a Hilbert space allows this, simply measuring the "size" of the vector **E**, representing the error in the best approximation in norm. One can also objectively compare different rule systems, relating the norm of each vector E after every projection.

The method has been tested with different human interpretations of the "Sonatina in sol maggiore" of Ludwig Van Beethoven.

The first group of performances are 5 neutral ones, in wich the pianist was asked to play without any expressive intention and interpreting the score in an as similar as possible way. As illustrated in **Fig. 4** the parameters are rather stable among the 5 performances and this means that the performances are very close in the performance space, how we should have known.

The second group of performances have been recorded asking the pianist to play with 6 different expressive intentions, inspired by six affective adjectives. As reported in Fig. 5, in this case some parameters vary considerable among the performances, suggesting that the rule system and the estimated parameters are suitable to represent different the expressive characteristics in musical performances inspired by different expressive intentions. For example, the estimated k values show that the rule named *Melodic charge* can differentiate the performances inspired by the adjectives *happy* (the highest value) and *sad* (the lowest value), and the *Punctuation* rule put in evidence the differences between the *fear* and *no-expression* performances.

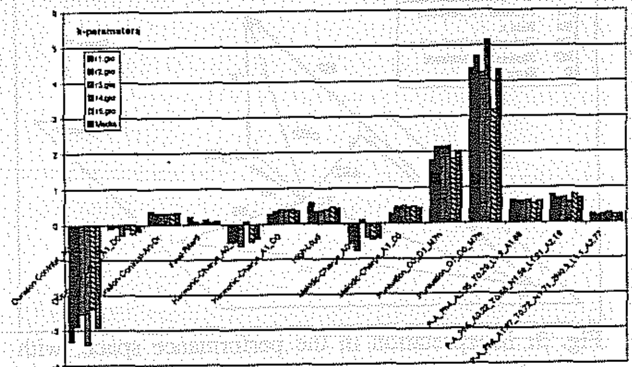


Fig. 4: the set of the tuned parameters calculated to fit 5 neutral interpretations of the "Sonatina in sol maggiore" of Ludwig Van Beethoven; the values are quite stable.

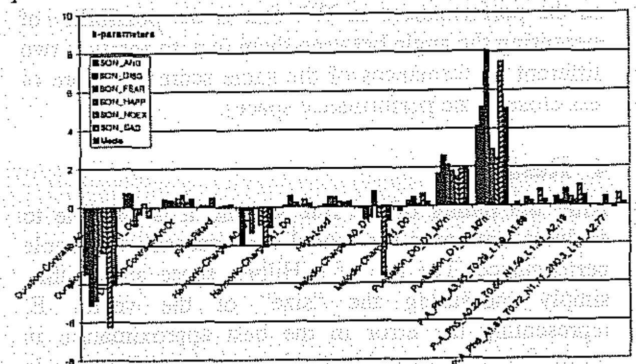


Fig. 5: the set of the tuned parameters evaluated to fit 6 performances of the "Sonatina in sol maggiore" of Ludwig Van Beethoven with different expressive intentions; the value are related to the different expressive intentions.

Fig. 6 shows the profiles of the Sound Levels of a human neutral performance and a performance generated by the rule system with the parameters automatically estimated by the software *Interpretazioni*.

5 Conclusions

A methodology that allows the tuning of rule systems for studying expressiveness in musical performance was presented. The use of rule systems, in fact, allows to model a lot of the expressive deviations introduced by the musician. The estimate of the parameters, however, can be an hard job, because of the high number of parameters to be considered.
The results obtained with several piano performance showed that, defining a performance space with an appropriate Euclidean norm, the theory of Hilbert space can be used to obtain a quite good estimate of the parameters.
The method was applied to the rule system developed by the KTH of Stockholm, but it can be used with every model which introduce expressive deviations in an additive way, and can be a good means to develop and improve rules for musical expressiveness.

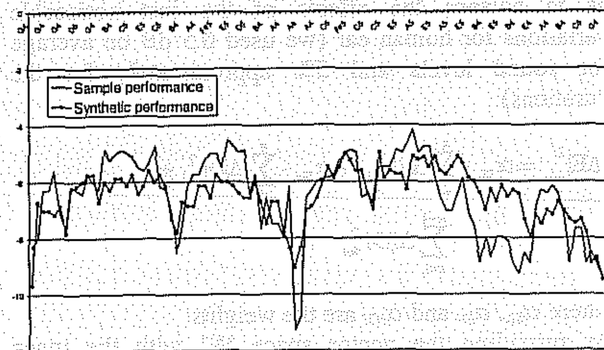


Fig. 6: sample performance vs. synthetic performance; profiles of the Sound Level in a neutral performance of the "Sonatina in sol maggiore" of Ludwig Van Beethoven.

References

[1] Baroni M.: "L'esecuzione musicale e la psicologia delle emozioni", *Bollettino di analisi e teoria musicale*, GATM, Vol. IV, N.1, June 1999, pp. 9-37.

[2] Friberg, A.. "Generative Rules for music performance: a formal description of a rule system". *Computer Music Journal*, 15(2), 1991, pp. 56-71.

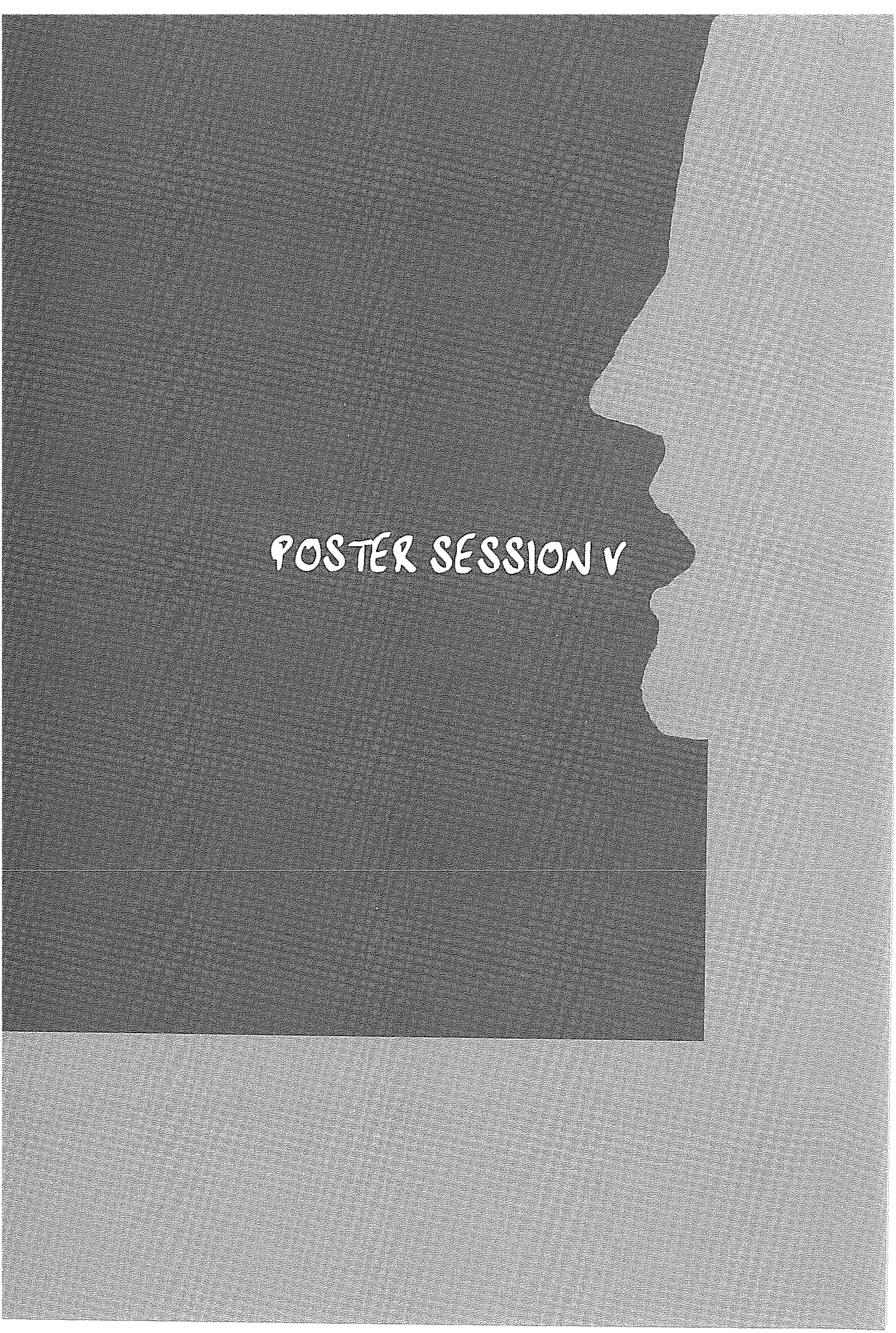
[3] Bresin, R.. "Artificial Neural Networks Based Models for Automatic Performance of Musical Scores". *Journal of New Music Research*, Vol. 27, N. 3, 1998 pp. 239-269.

[4] 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, N. 3, 1998, pp. 293-321.

[5] Clynes, M.. "Microstructural musical linguistics: composers' pulses are liked most by the best musicians". *Cognition*, 55,1995, pp. 269-310.

[6] Todd, N. P. McA.. "The kinematics of musical expression". *Journal of Acoustical Society of America*, 97(3), 1995, pp. 1940-1949.

POSTER SESSION V



POSTER SESSION V

TEACHING MUSIC AND ACOUSTICS : INTERACTION BY USE OF SENSORS IN A VIRTUAL MUSIC ZONE

Sergio Cavaliere[0] Gianpaolo Evangelista[e] Giancarlo Sica[*]

[0] ACEL, Dipartimento di Scienze Fisiche
Università "Federico II" di Napoli
Complesso Universitario di M.S.Angelo,
Via Cinzia 80126 Napoli
e-mail: cavaliere@na.infn.it
[e] Audiovisual Communications Laboratory,
Swiss Federal Institute of Technology,
Lausanne, Switzerland.
e-mail: Gianpaolo.Evangelista@epfl.ch
[*] Responsabile Musicale ACEL (Dip. di Scienze Fisiche, Napoli)
e-mail: sica@na.infn.it

Abstract

In the following we present work in progress consisting in the configuration of a workstation for the production of sound and music for educational purposes: the VM-Zone (Virtual Music Zone).
The aim of the project is to make easier and more intuitive the approach to some important aspects of sound and music production and control in an educational project.
*The goal may be achieved by setting up a proper space where students may interact in a physical way with sound and music. Sensors and various kinds of controllers allow them full control of the sound/music production by computer, with the aim of gaining for them deep knowledge of the meaning of relevant parameters of sound and music. The realized system the **VM-zone** is intended to be used in public schools at both elementary and intermediate levels*

1 Introduction

As a matter of fact one of the well known limitations in the production and execution of music by computer is the lack of naturalness and full interactivity with timbres and compositional structures, as well as the complete lack of gestural performance. This is true even if experiences in this direction opens up completely new and unexpected horizons[2][3].
This is especially true in the educational case, where pupils may greatly benefit from a 'physical' interaction with sound and music. In fact postural and gestural aspects are significant in any musical performance, both in the production and control of sound.
Another problem experienced while teaching sound and music is the complex manuality and skills required to the student by an acoustical instrument. This may create a deep barrier against largely experiencing music, sounds and noises, as required in order to promote interest in music and sounds.
We may refer, for these aspects,to the instrumentarium by Carl Orff, which in fact is aimed at filling this gap. In our case a virtual space as an integration to real instruments may also help in this same direction, just easing the approach to music and sounds, as a first step towards full disclosure of the creative and amusing world of music performance.
Finally we have to consider that movement as well as singing and making music all of them are connected and integrating activities which deliver an high degree of creativity. The possibility to create sound-scapes, music, rhythm by means just of presence in a space and movements in it, probably coordinated in a group activity, may definitely render amusing the experience of sound and music at school.

2. The VM-zone

The system we have set up, called VM-zone Virtual Music Zone, is built around a personal computer properly equipped, with enough speed for real-time synthesis and enough disk space. The computer is equipped with different sound synthesis boards. We have used for sound production a powerful synthesis board, the PULSAR I by Creamware, with 4 DSP Shark by AMD. The system is equipped with a special midi Interface the I-Cube by Infusion Systems which will be

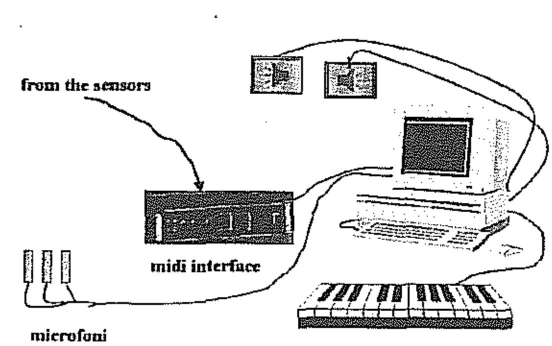


Fig. 1 The VMzone: the computer and peripherals

described later. This interface is in charge of collecting information from the sensors and sending it to the computer by means of a standard MIDI interface.

2. The I-Cube Midi Interface

The I-Cube interface collects data from a number of channels up to 32 (8 of which double as Digital Outputs). It digitizes the analog inputs with 12 or 7 bits (n. of bits is selectable), with a sampling rate which depends on the number of inputs and on the required precision, and ranges from a minimum of 50Hz to a maximum of 225 Hz, enough to detect slow changes in the ambient, while avoiding possible overload of the MIDI interface. Each sensor input may be arbitrarily mapped to either Control Change (7 bit resolution), Pitch Bend (12 bit resolution, but only available for 16 sensors - each one on a separate MIDI channel) or Note On messages (1 bit resolution, with programmable note value and velocity). The I-Cube may be operated in a stand alone or in a computer controlled mode. It in fact is programmed by the personal, using the connection in Fig. 2; it saves the programmed configuration after power down; therefore after being configured it may be used in a stand alone mode.

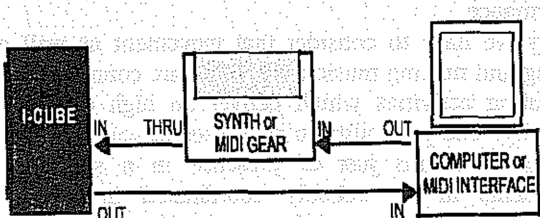


Fig.2 Interfacing the I-cube

The I-Cube also provides sensor data processing: inversion (increasing sensor values result in decreasing MIDI data values), thresholds (the sensor values within

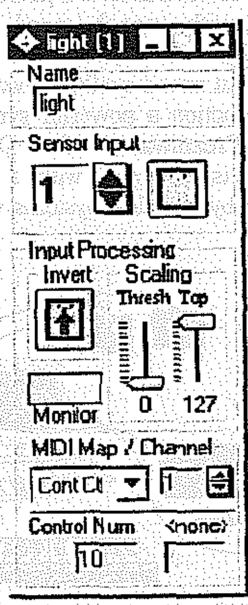


Fig. 3 Programming the single sensor

the specified range are offset and scaled to the MIDI data output range), zooming (set the sensitivity of an input to a specific voltage range), as shown in fig. 3. These programmability reveals invaluable in order to use different kind of sensors.

In addition the I-Cube can output voltages (to control actuators), a feature which will be used in the future.

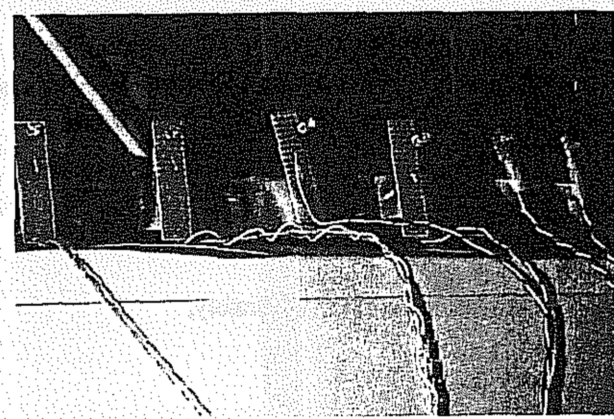


Fig 4 Some light sensors

4. The sensors

Connected to the I-Cube Midi interface a number of sensors can be used in order to detect movements and presence in the ambient under control; they may include ultrasound devices, infrared detectors photo-transistors and others. An arrangement for tests showing the light sensors is shown in fig. 4.

A console of potentiometers, detectors based on touch sensitive resistors and a MIDI keyboard are also connected to the MIDI interface of the computer by means of a MIDI-merge. These devices allow for local control by a 'director' which may also interact with the ongoing experiment.

5. The physical space

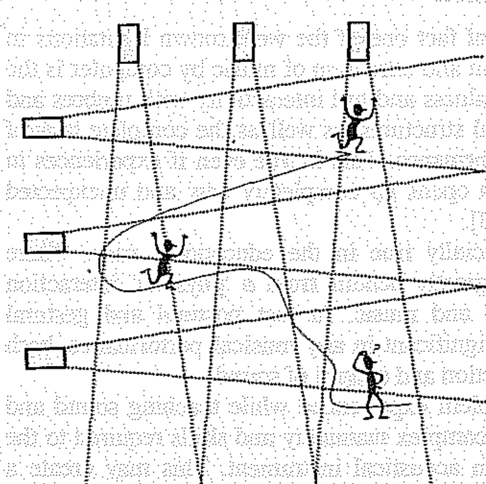


Fig.5 The physical space

The sensors are allocated in a space of approximately 20 square meters, with the topology in fig. 5, forming a matrix, in order to control the whole space.
Any other topology may be as well adopted, also in consideration of the use of different sensors.

5. The director program

What is relevant in the overall organization is the connections between data collected from the classroom/ambient/stage regarding what happens in the controlled space and the synthesized sound/music. The connection is readily made by use of the continuous controllers provided by the MIDI protocol or also NOTE ON/NOTE OFF in case of on/off triggers from the stage. The instrument built in this way embodying the virtual interconnection between peoples in the controlled space and music describes the sound/music experiment.
In the case of use of Direct Csound [1][7],the new MIDI controller module, **MIDIC7**, reveals quite efficient for the purpose.

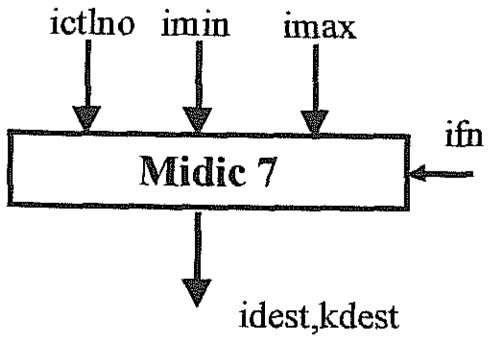


Fig. 6 The **midic7** module by Maldonado

This opcode in fact makes it possible to assign data coming from the sensors (properly mapped to instruments by means of the I-CUBE software) to various Direct Csound modules. This allows driving in real time the different functions of the module. For example we may control amplitude and frequency of a virtual oscillator, such **oscili**, using a couple of **midic7** modules, connected to different sensors. For example a sensor may control the amplitude while another controls the center frequency.
Syntax for the instruction is:
idest **midic7** ictlno, imin, imax [, ifn]
kdest **midic7** ictlno, kmin, kmax [, ifn]
where:
ictlno = continuous controller number (1-127)
imin = minimum output number (real number defined by the user)
imax = maximum output value (real number)
ifn (opzionale) – table to be read when indexing is requested. Values in the table must be normalized: output data are scaled to fit the range **imin, imax**.
In the **midic7** opcode minimum and maximum values may be updated at K-speed.

Following we give an example of the use of sensors to

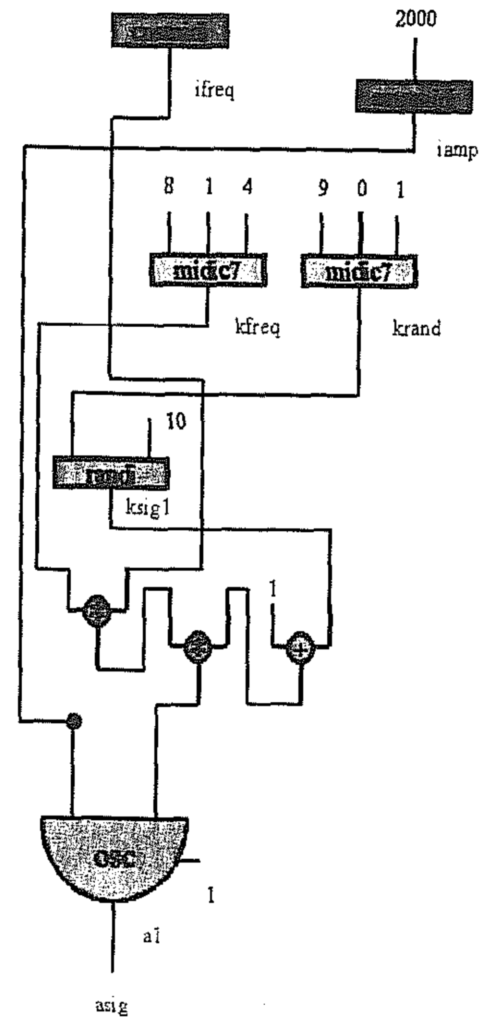


Fig 7 A virtual Csound instrument using the **midic7** opcode of Direct Csound.

control via the I-CUBE interface a simple virtual Csound instrument, realized in Direct Csound (see fig.7)

```
sr=44100
kr = 441
ksmps = 100
nchnls = 2
instr 1
ifreq cpsmidi
iamp  ampmidi      2000
kfreq midic7       8,1,4
krand midic7       9,0,1
ksig1 randi krand,10
a1    oscili       iamp,ifreq * kfreq *
(1 + ksig1),1
asig = a1
out   asig
```

```
out    asig
endin
```

In the above instrument two sensors (mapped to controllers 8 e 9), control the amount of randomness added to pitch frequency (**cpsmidi**) and the amplitude of the random generator module ***randi*** (the modulation depth).

5. Some experiments

Some experiments have already been carried on while others may be foreseen to prove useful.

- Create and modify a complex timbre by means of additive synthesis; in this case optoelectronic sensors control the activation of a variable number of oscillators, the degree of detuning of the partials against the multiples of the fundamental frequency, amplitude of each partial, overall sound volume and above all the fundamental frequency.
- Altering sounds by means of subtractive synthesis, by means of the control of pass-band filters. Both center frequency and bandwidth are controlled by means of the sensors.
- Realize the fundamental acoustical experiments: based on the physical mechanism of sound production and propagation such as beats, doppler effect, echo, reverberation;
- Realize psicoacoustics and perceptual experiments, such as critical bands, range of audible frequencies, acoustical illusions, sound spazialization.

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 [5][6].
- Control may be also exercised on the process of executing a score, by means of interactive modification of execution parameters [4].

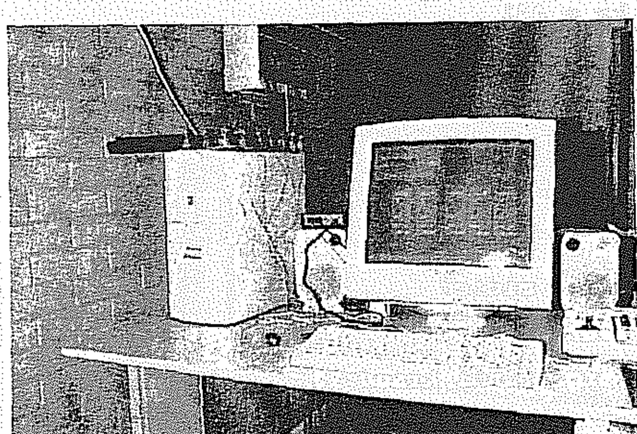


Fig. 8 The computer in the laboratory setting with some sensors and the I-cube interface

Conclusions

Goal of the didactical project are:

1) acquire a basic knowledge of the parameters of sound by means of a "physical" interaction with sound
2) acquiring consciousness of the interaction between gesture, music and sound in the 3D space.
3) augmenting expressive power both for each pupil and also for groups, in a coordinated way.

Some experiments have already carried on, but more must be experienced in practical situations in schools.

The project was funded by Regione Campania (the regional Government Institution) and is intended to be used in public schools at both elementary and intermediate levels. The project is being developed in the context of a larger project aimed at setting up inter-school Laboratories for Musical Education.

Finally the same system is also suited for installations and interactive performances.

References

[1] Richard Boulanger, editor. *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing,and Programming* . MIT Press, 2000

[2] J. Paradiso, *Electronic Music Interfaces: New Ways to Play*. IEEE Spectrum Magazine, Vol. 34, No. 12, pp. 18-30 (Dec., 1997).

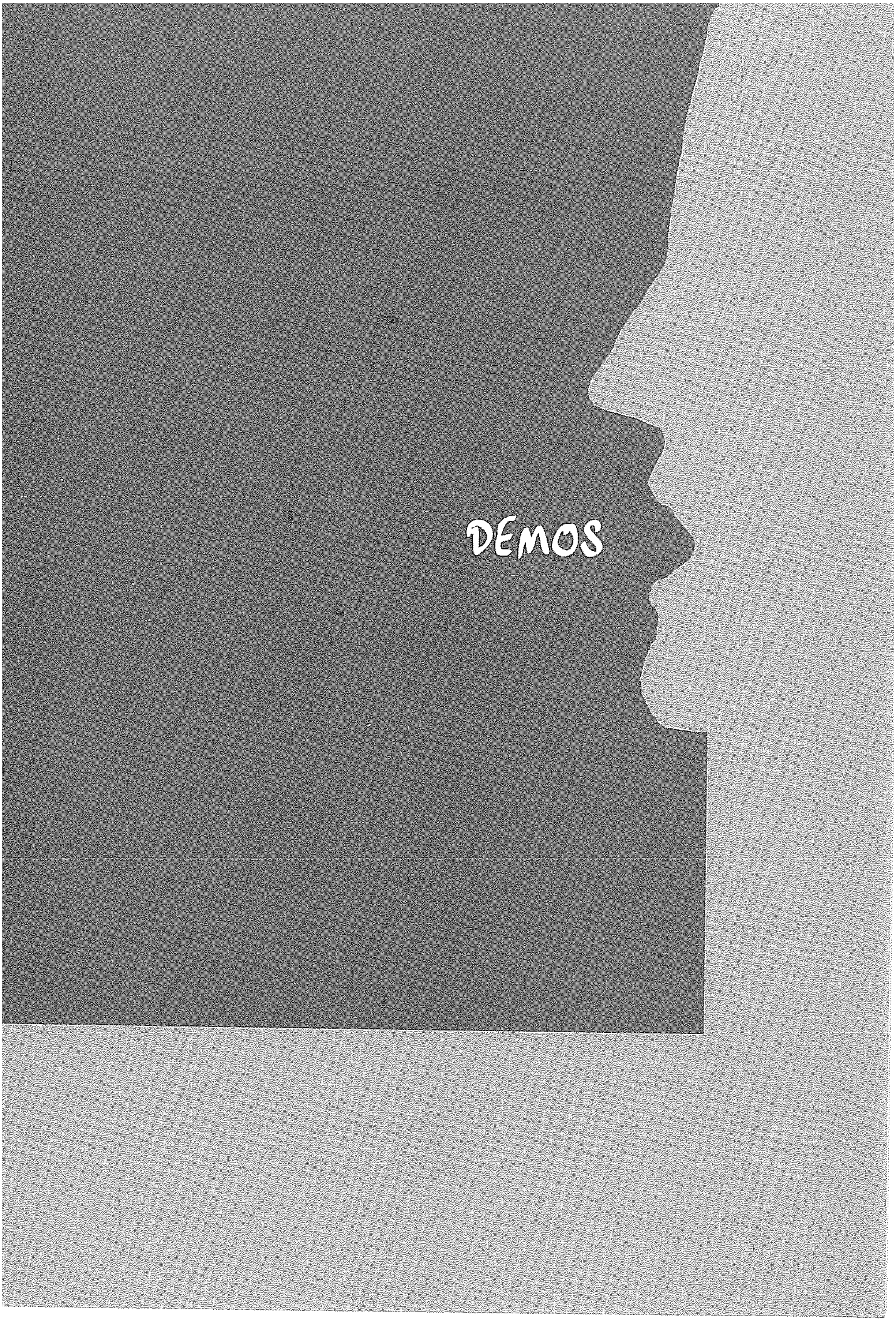
[3] S. Cavaliere, L. Papadia and P. Parascandolo, *From Computer Music to the Theatre: the realization of a theatrical automaton.* Computer Music Journal, Cambridge, MA: MIT Press, vol 6, no.4, Winter 1982, pp.22-35

[4] 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.

[5] G. Sica, *Reti neurali e loro applicazioni musicali.* Konsequenz ,Liguori Editore, Napoli, 1999.

[6] 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.

[7] Gabriel Maldonado: *Midi designs for real time performance* in Richard Boulanger, editor. *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing,and Programming* . MIT Press, 2000


DEMOS

A real-time platform for interactive dance and music systems

A.Camurri, P.Coletta, M.Peri, M.Ricchetti, A.Ricci, R.Trocca, G.Volpe
Laboratorio di Informatica Musicale (http://musart.dist.unige.it)
DIST - University of Genova, Viale Causa 13
I-16145 Genova

Abstract
EyesWeb is a research project aiming at providing models and tools for integrating movement, music and visual languages in a multimodal perspective. The project includes a hardware and software platform to support the user (i) in the development and experimenting of computational models of expressive content communication and of gesture mapping strategies, and (ii) in fast development and experiment cycles of interactive performance setups. An intuitive visual programming language allows to map – at different levels - gestural components with integrated music, visual, and mobile scenery. System original features, such as its real-time patch scheduler supporting *active modules* and *user-defined datatypes*, are shortly described in the paper. The EyesWeb platform consists of a serie of integrated hardware and software modules which can be easily interconnected and extended by users.

1. Introduction

Several systems and applications are currently used for interactive performance, e.g., the several applications using Max and Supercollider, or a few special-purpose systems based on videocameras mapping low-level image data to MIDI. Several conceptual basic issues remain currently open. Typical state of the art applications concern quite simple cause-effect mapping of (low-level) movement into MIDI parameters, in many cases characterised by weak control strategies, no memory, poor interaction metaphors, and low-level interaction design. The problem of gesture mapping as well as the individuation of effective gestural integration in a multimodal perspective is an important goal of current research.

In (Camurri 1995; Camurri and Leman, 1997) we proposed new interaction metaphors as an attempt to go beyond the "musical instrument" metaphor: from orchestra direction, to metaphors like artificial potential fields and "maps" (Camurri et al 1994), to dialogue based on agent models embedding models of comunication of artificial emotions (Camurri and Coglio 1998; Camurri and Ferrentino 1999) and KANSEI (Camurri, Hashimoto et al. 1998, 2000).

In this paper we propose a flexible and powerful platform to support research in these directions, from both an artistic and scientific perspective.

The motivation for designing an original platform derives from the requirements we collected from several artistic projects involving real-time interaction we participated in the last years, and from previous research projects (Camurri et al 1986) including HARP (Camurri et al 1994, 1995). For example, as for real-time movement analysis, a main focus is on the extraction of parameters on **expressive content** in the performance (Camurri and Trocca, 2000). Existing systems are rather limited from this point of view. Ideally, we want a system able to distinguish the different expressive content from two performances of the same dance fragment or the same expressive content from two different dance fragments. The system is designed to support this and other research issues.

2. The EyesWeb software

The EyesWeb platform consists of a number of integrated hardware and software modules which can be easily interconnected and extended. The EyesWeb software consists of a development environments and a set of libraries of reusable software components which can be assembled by the user in a visual language to build patches as in common computer music languages inspired to analog synthesizers. A patch can be used as a module in a higher-level patch. The software runs on Win32 and is based on the Microsoft COM standard. An example of a simple EyesWeb patch is shown in fig. 1. The patch shows a typical application on movement analysis based on videocameras. Let us shortly examine the patch modules of fig. 1 (from left to right). A frame grabber module sends its output to a *splitter* module: it takes the input signal and separate the odd and even lines from each frame (this because this patch uses the EyesWeb **Mpx** hardware to use two cameras in a single video input – see below). The output from Mpx goes to two identical sub-patches: in figure 1 only one is shown. Then a simple mechanism based on background subtraction is used, followed by a "binarizer". The output of the two-level "silhouette" is sent to feature extraction modules trying to extract barycenters. In the upper part of Fig.1 user interface widgets available also at run time are shown. Fig. 2 shows the dialog window for the barycenters module. In fig. 3, the outputs from two test modules show the results of the real-time analysis from this patch. MIDI output modules end the chain of this patch. Besides the "silhouette" model, further graphical representations of movement parameters have been developed (3d models and in general visual metaphors).

EyesWeb is a multi-thread application based on the Microsoft COM standard. At run-time, an original real-time patch scheduler supports **active modules** and **user-defined datatypes** for link types of data streams among modules. A patch is automatically splitted by the scheduler into several threads according to the topology and the presence of active modules. Active modules are software modules used in patches characterised by a special behavior: they have an internal dynamics, i.e., they receive inputs as any other kind of modules but their outputs are asynchronous with respect to their inputs. For example, an "emotional resonator" able to react to the perceived expressive content of a dance performance, embedding an internal dynamics, may have a

delay in activating its outputs due to its actual internal state, memory of past events. This is one of the memory mechanisms explicitly supported by the system to implement interaction metaphors beyond the "musical instrument".
New datatypes used to communicate data (e.g. on expressive content) among modules in a patch can be defined by the user. Data links come from two different root types: signals and controls. Signal type taxonomies can be defined by the user.
Multiple versions of modules (versioning mechanism) is also supported by the system, e.g., allowing the use in patches of different versions of the same datatype or module. The compatibility with future versions of the systems, in order to preserve the existing work (i.e. patches) in the future is also supported.
An open library of basic module libraries include input and movement capture, signal conditioning, filters, active modules, observers, and output modules. Movement capture and input modules have been developed for different sensor systems: both environmental sensors (e.g. video cameras) and the wireless on-body sensor technology we developed (see below in the paper). Low-level filters (e.g. preprocessing, signal conditioning, etc.) as well as medium-level filters (e.g. the module to extract the barycentre coordinates of a human figure; the module for evaluating equilibrium) are available.
Observers can be high-level filters, active modules, patches able to extract high-level information, typically concerning expressive content. Output modules include MIDI, TCP/IP, DMX outputs, including the communication of expressive content to external applications (e.g. on Virtual Environments inhabited by avatars and clones/characters).
A current research project concerns the development and experimenting of EyesWeb software modules for movement analysis inspired to a computational model of Laban's Theory of Effort (Camurri et al 1999; 2000). To this aim, we are developing analysis modules of medium level features. For example, figure 4 shows the output of a module fo analysis of *instability* of a movement (in the figure: a walking): peaks in the graphic correspond to steps (a foot is raised from the floor).
End users can directly assemble modules from the available libraries to build patches implementing interactive performance setups.

3.1 The EyesWeb Wizard

Modules are standard COM modules. Users can build new EyesWeb modules as standard COM modules, and use them in patches. In order to hide to the user the complexity of COM programming, we developed the **EyesWeb Wizard**. The user can develop automously (i.e., possibly independently from EyesWeb) the algorithms and the basic software skeletons of their own modules. Then, the Wizard supports the user in the process of transforming his algorithms in integrated EyesWeb modules.

4. EyesWeb Hardware Modules

Here we list a few main hardware systems we developed during the last years, that we use in EyesWeb-based interactive performances and experiments.
Wireless Sensor-to-MIDI is a wireless, small, battery-equipped, easily wearable system designed to capture signals in real-time from on-body sensors. Signals are sent by this system to a remote receiver by means of a wireless radio link. A proprietary redundant data transmission protocol makes the system wireless communication robust and error-free. The receiver convert data into MIDI signals. System latency is less than 5ms.
The system consists of two hardware unit boxes:
- a small data acquisition, conversion, and wireless transmission unit (to be worn on-body),
- a receiver and converter to MIDI unit.

Video Multiplexer (***Mpx***) for connecting and sync two videocameras to the same framegrabber is our proprietary special electronics developed to capture the signal from two synchronized cameras. This board is based on the fact that we can multiplex two separate input video signals in only one, by switchings between the two video signals at the field rate (50 Hz). In this way we obtain a new interlaced signal in which odds and even fields contain the two different signals. We can then acquire the signal using an ordinary full frame single channel acquisition board. At this point, we have in the frame memory buffer the two original signals, just missing half vertical resolution but maintaining the same temporal resolution.
Hardware for human-robot communication: We developed since 1991 a number of setups for museums, music theatre, and art installations. We recently extended the robots Pioneer 1 and Pioneer 2 from SRI with audiovisual interfaces, sensors, interactive audio I/O, for real-time interaction with users (dancers, public, actors, music performers). In "L'Ala dei Sensi" multimedia event (see our web site) we equipped a Pioneer 2 with a video camera, a video projector, a microphone to interact with a dancer (Virgilio Sieni). The dancer wears on-body sensors (Accelerometers and FSRs connected to our previously described *wireless-sensors-to-midi box*) to send "stimuli" to the robot.
A MIDI-controlled audio matrix 8x8 channels and a ***Long distance MIDI signals Tx/Rx*** (which uses the standard audio/cannon cables commonly available to send audio signal on stage to send MIDI signals, at long distances) are other examples of special hardware developed for interactive setups.
Previous systems developed for HARP (Camurri et al 1994) a few years ago, can be used in EyesWeb. For example, ***DanceWeb*** is a low-cost sensor system based on ultrasound (US) technology. Up to 64 ultrasound sensors and up to 32 digital I/O can be connected and multiplexed in groups (e.g. to avoid interferences between US sensors). The system provides fast serial and MIDI outputs. The system consists of an external programmable (via MIDI) rack unit with a microcontroller and electronics for sensor data acquisition and signal conditioning.

5. Conclusion

The EyesWeb platform has been experimented and used in music theatre projects, e.g., Luciano Berio Opera "Cronaca del Luogo", which opened the 1999 Salzburg Festival, as well as in other multimedia events and museum exhibits, and in science centers interactive games-experiments. It is currently experimented in a high-school for music education. Recent minor but useful improvements concern the extension of the system to support not only Matrox Meteor (I and II) frame grabbers but any Video For Windows compatible video input board. Further, the system does not use any more the Matrox MIL libraries, so no expensive software license is needed to use the software. This because we are planning to start distributing the system to a selected group of users.

References

Camurri A. P.Morasso, V.Tagliasco, R.Zaccaria. 1986. "Dance and Movement Notation." In Morasso & Tagliasco (Eds.), *Human Movement Understanding*, pp.85-124, Amsterdam: North Holland.

Camurri, A. 1995. "Interactive Dance/Music Systems." In Proceedings of the 1995 International Computer Music Conference. San Francisco: International Computer Music Association, pp.245-252.

Camurri, A. (Ed.) 1997. Proceedings of the International Workshop on KANSEI: The Technology of Emotion. Genova: AIMI (Italian Computer Music Association) and DIST-University of Genova, Italy.

Camurri, A., M.Leman. 1997. "Gestalt-Based Composition and Performance in Multimodal Environments." In Leman (Ed.) *Music, Gestalt. and Computing*, 495-508, Springer.

Camurri, A., A.Coglio. 1998. "An Architecture for Emotional Agents." IEEE Multimedia, 5(4):24-33, Oct-Dec 1998, New York: IEEE Computer Society Press.

Camurri, A., P.Ferrentino. 1999. "Interactive Environments for Music and Multimedia." Multimedia Systems Journal. Special issue on Audio and multimedia, 7:32-47, ACM and Springer.

Camurri, A., S.Hashimoto, K.Suzuki, R.Trocca. 1999. "KANSEI analysis of dance performance". In Proceedings of the 1999 IEEE Intl Conf on Systems Man and Cybernetics SMC'99, Tokyo, New York: IEEE Computer Society Press.

Camurri, A., Hashimoto, S., Ricchetti, M., Suzuki, K., Trocca, R., Volpe, G. (2000) "EyesWeb – Toward gesture and affect recognition in dance/music interactive systems." Computer Music Journal, 24(1), MIT Press.

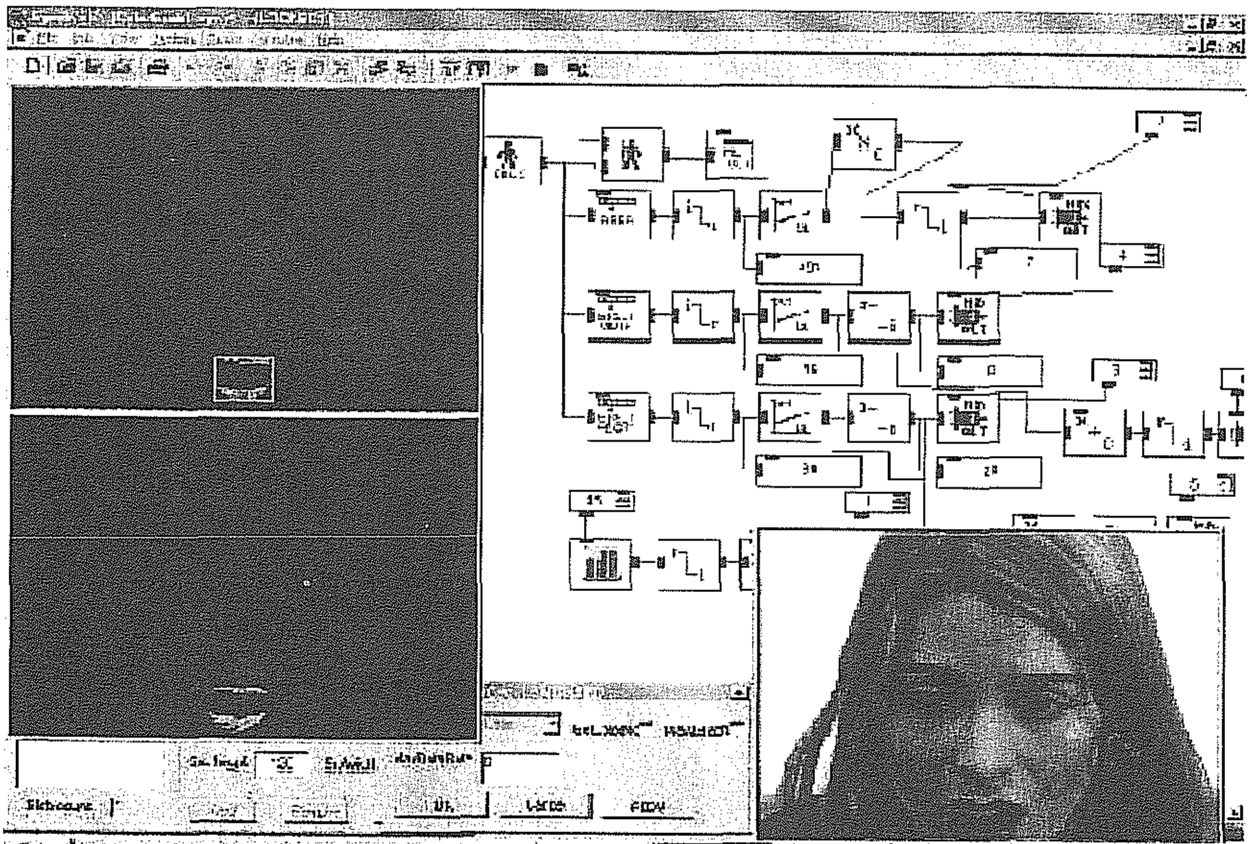


Figure 1: EyesWeb at work. The figure shows an example of a patch at run-time implementing a simple movement analysis of the lips of an actress, based on a videocamera. The patch has been developed for early experiments on a piece by the composer Roberto Doati during the Workshop on Music and Multimedia (18-22 June 2000) that we organised in our Laboratory at the Opera House Teatro Carlo Felice, Genova. The two black windows on the left show the output from the lip detection. The right window shows the original image. The patch generate MIDI signals related to lip movement, which control in real-time the voice of the actress Francesca Faiella.

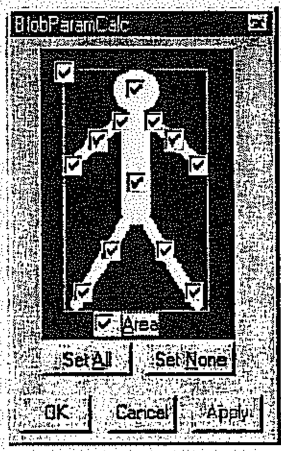


Figure 2: the module for extracting baricentres: the window for the (off-line) setting of parameters.

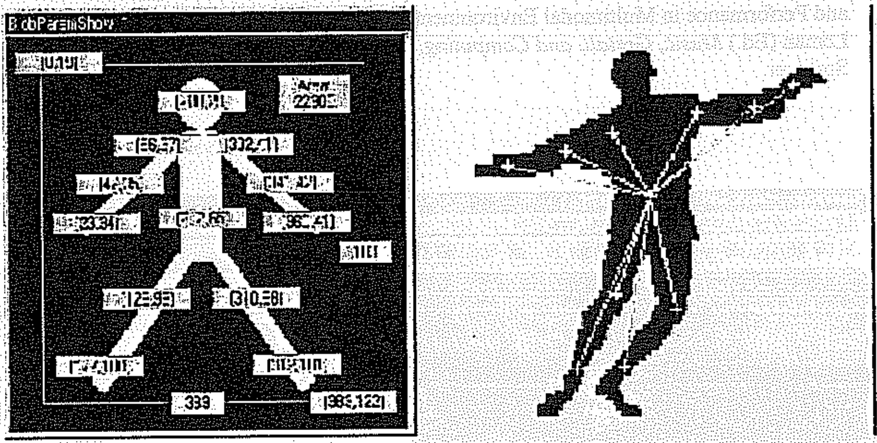
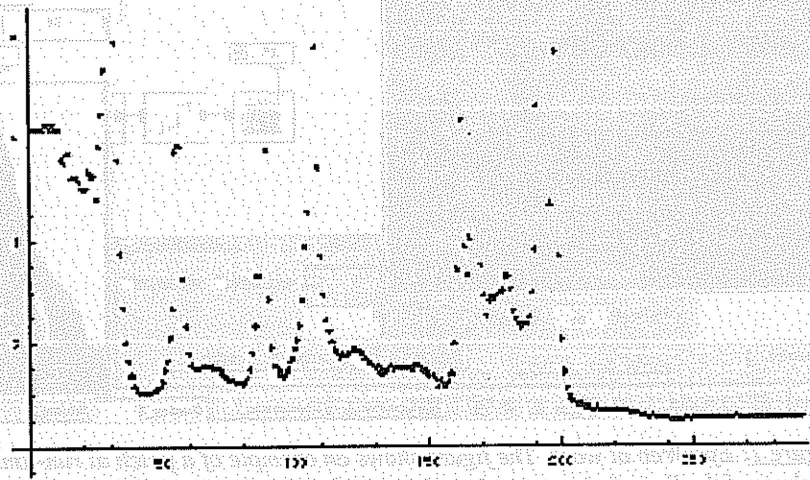


Figure 3: the module for extracting baricentres at run-time. Numbers are coordinates in pixels. The picture on the right is the output of a visualisation module.

Figure 4



'SDP - SONOROUS DRAWING PLANE' and Laboratory 'THE SONIC DRAW'
Timbric Exploration in Real Time through Expressive Gesture Tracks

Silvia Lanzalone, Fausto Sebastiani, Sandra Fortuna

CRM – Centro Ricerche Musicali
via Lamarmora 18 – 00185 Roma – Italy
phone +39-064464161, fax +39-064467911; e-mail: crm.it@crm-music.org

Abstract

The software 'SDP - Sonorous Drawing Plane' (S. Lanzalone), is developed in order to manage live electronics, allowing the real-time performance by computer in a more versatile way, that is to say using the computer as an 'instrument'. SDP reads and converts computer mouse data as the operator creates lines corresponding to performance gestures thus creating both visible and audible output. This software allows a single gesture to control more than one parameter thus creating complex changes in the audio program output.
The article will deal with a description of 'The Sonic Draw', a didactic laboratory for computer music that S. Lanzalone, F. Sebastiani, S. Fortuna, have experienced using SDP.

1 Introduction

The relationship between man and machine, seen from the point of view of live electronics, can be seen as a search for freer, more versatile ways in which to use a computer as an actual 'instrument' in the classic sense of the word. In this relationship the computer would use the operator's actions (or gestures) to create sounds in real time. The concept of 'instrumental performance' includes the concept of 'gesture', a concept that electronic music, be it analog or digital, has ignored for some time. It is now possible to re-introduce the concept of gesture through the use of real-time electronics. In order to extemporaneously explore through gesture one needs an effective, efficient environment that allows control of as many independent sound parameters as possible while presenting the user with an uncomplicated interface. There is also need to allow for the interaction of gestures (giving the operator feedback generated by the actual acoustic events) allowing there to be alteration and interaction at even the most detailed levels.

2 'SDP - SONOROUS DRAWING PLANE': a program for real-time performance by computer

The software 'SDP - Sonorous Drawing Plane' (S. Lanzalone), is developed in order to better manage live electronics, referring to the area of philosophy of 'complex thinking'. It allows the creation of computer music in real-time thus enabling the computer operator/performer to react to and influence the work as it is unfolding. This allows interpretation to become an integral part of the electronic composition.
The program, written in Visual Basic, interfaces with the Fly30 (CRM)[1] and Mars (IRIS)[2] systems by calling up external procedures contained in DDL (Dynamic Link Library), files created by CRM and IRIS. These files extend the functionality of the system allowing, for example, the assignment of data to a certain algorithm. This assignment of data is controlled by mouse as it operates within the white background that delimits the user interface. The mouse movements (or gestures) are seen by the system as excursions along the x (horizontal) and y (vertical) axes. The coordinates of every point correspond to different values of the sound parameters: they can be assigned to x and/or to y axe and then vary coherently with the variation of the mouse position. As it moves, the mouse creates a graphic representation of it's path giving the user visual feedback. Before passing this data to the algorithms, the software translates the single x/y coordinates into numerical values that the chosen algorithms can understand. It interacts with the parameters according to three conditions pre-selected by the user:

- the axis associated with each parameter;
- relationship between the length of the axis and the variation supplied by the data itself within a predefined minimum and maximum range;

[1] The Fly30 is a system operating in real time, designed and realized at CRM of Rome. It is based on the floating point processor TMS320c30 by Texas Instruments.
[2] The Musical Audio Research Station (Mars) is a system designed and constructed at IRIS of Frosinone. Sound processing is realized with the processor X20, designed by Giuseppe Di Giugno.

- the mathematical function associated with each parameter. If, for example, an exponential function is used, a linear movement of the mouse will create a straight line that corresponds to a data series that varies according to an exponential sequence.

This last condition is extremely powerful allowing the independent control of more than 100 parameters while creating a framework that allows interaction between the different parameters. Assigning the axis, functions and limits to each parameter (decided before the performance) is called 'configuration', a file of data that can be called up at the keyboard by using the fkeys. Calling up a new configuration does not interrupt any processes already underway. This means that the operator can switch configurations at any time without causing glitches in the audio. The speed with which the operator moves the mouse effects the processing speed of the system adding another dimension to the sound quality. This is all done without any appreciable time delay. When the mouse stops moving, the values of the parameters remain constant. The construction of an SDP configuration depends on the type of gesture desired. Some configurations can be relatively static in which large gestures create small variations in the sound. In contrast, one can also program a configuration in which a small mouse movement corresponds to large variations in the sound parameters. This last condition (most configurations use values that fall between the possible minimum and maximum) makes it very difficult to define the gesture accurately.

The form that the score must take in order to accurately describe the execution in a musical context becomes extremely complex and is necessary to create a new kind of musical score for the symbolic representation of gesture. Unlike the performance gestures available when using a keyboard, an audio mixer or a pedal – which have no real spatial characteristics and which vary only in terms of speed and quantity (amplitude) of movement – the characteristics of a mouse gesture are much more complex. They include:

- the spatial coordinates of the gesture
- the direction of the gesture
- the form of the gesture
- the length of the gesture
- the speed of the gesture
- the succession of all gestures

Not all of these variables are easily defined in a way that can be understood by the operator: the resulting representation on the computer monitor can show the form, length and starting and ending points but cannot describe direction, velocity and events that precede or follow the gesture. The notation used for scores of pieces that use SDP include:

- graphic representation of the actual SDP screen showing only the necessary information: length, x/y coordinates and type of gesture;
- indications of speed, direction and sequence of gestures distilled to a simple series of icons that the operator can easily read.

This software environment has been used by some composers underlining different ways in which the program can be useful, from music-dance interaction to other problems that deal with the man-machine interface[3]. The 'expressive gesture track' have evolved through several compositions, each one of which explores different aspects of gesture[4].

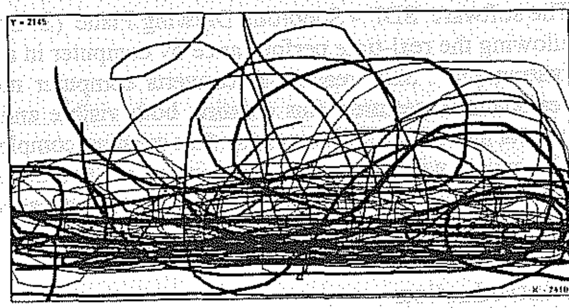


Figure 1. 'SDP - Sonorous Drawing Plane': a screen shot after a performance.

3 Laboratory "THE SONIC DRAW": musical computer games for children

The relationship between movement, computer mouse drawing and sound can be startling, so much so that is suggests other uses for SDP, including games for children in which they use the computer to create electronic compositions. At the Laboratory for Computer Music " The Sonic Draw", done in collaboration with musician Fausto Sebastiani and pedagogist Sandra Fortuna, in Rome last June in the Festival "La Festa della Musica" at the "Il Punto di Svolta" art gallery, SDP was used in an experiment with children between 5 and 15 years of age, with appreciable results.

Even smaller children, who were not yet familiar with computers, had no difficulty in

[3] See the article Lanzalone S., "The Manipulation of 'Body Sound', Interaction between Music And Dance. Analysis of 'Contropasso' by Michelangelo Lupone.", in proceedings *XIII Colloquium on Musical Informatics*, L'Aquila 2000, to have an example of music-dance interaction and an example of notation used for scores.

[4] Silvia Lanzalone, 'Rhysmos', sounds in movement for real - time systems (1998); Michelangelo Lupone, 'Contropasso' for real - time computer and three dancers (1998 - 1999); Silvia Lanzalone, 'Tracciati' for percussion, planephones and live electronics (1998); Fausto Sebastiani, 'Melodie II' for tape and real - time computer (1998); Silvia Lanzalone, 'Dis-trazioni', interaction for two real - time computers (1999).

learning to use the mouse. In addition, some program modifications had been made so as to make the use of SDP easier and more pleasant for the children, for instance:
- use of the mouse does not require pressing the buttons but only shifting the mouse on the flat surface of the table;
- the drawing reproducing the mouse movement can be realized with lines of various sizes and various colours.
The three principal objectives of the laboratory were:
1. to discover the possible relationship between gesture, graphics and timbric variation;
2. to emphasize and make clear the synaesthesic perceptions linked to the perception of timbre so as to define this parameter through attributes connected with other sensorial fields, which are not only auditory (visual, thermic, pressure, spatial, tactile, etc.)[5];
3. to identify and experiment *codes* on the basis of which the sound event is connected to the sequence of images and to their relative significance.

The children were encouraged to create a composition by making an abstract work of art. The goal was to expose the children to the language of contemporary music by exploring the relationship between gesture, design and sound. As one uses the system, this relationship is easily understandable, often intuitive. The setting up of the system and creating the different configurations is not so intuitive. For this reason the children were presented with a choice of pre-programmed sounds[6]. They could then use the mouse to change selected sound parameters. The children were asked to create a composition based on a drawing or the 'sonic background' for a cartoon[7]. The games were called as Sonic Draw and Sonic Cartoon.
For the *Sonic Draw* the children were instructed in three stages, the first two of which were aimed at helping the children to achieve an intuitive understanding of the way the system acted on the sounds and the effect that different gestures would have on these sounds. In the third phase the child was left free to experiment on his own.
1. The operator invites the children to make graphic designs on the screen while noting how the sound changes in relation to these designs. The next step was to take the children through each sound parameter one by one enabling them to understand in more detail the effects their designs had on the sound. In this way the children discover which types of gestures they must create in order to achieve the results they want. Some gestures change pitch level while others effect intensity, density and timbre.
2. Once they have understood the effect of various gestures, the children are invited to 'explore' the video screen and try out different graphic and musical gestures.
3. In this final phase the children were asked to create a composition using the techniques they had learned. In this way they created compositions that had both a visual and musical dimension.

Cartoon Sound was influenced by the audio tracks heard everyday on cartoon programs. Once the children had learned the possibilities available to him he was able to create a sonic background to cartoons that were played for him on the computer monitor. Given the wide range of ages of the children, a wide range of cartoons and sounds were provided. Each child chose three sounds which he used in the creation of his musical accompaniment guided by the emotional reaction to the images to which he was exposed. Given the wide range of ages of the children, a wide range of cartoons and sounds were provided. For children from 5 to 8 years old, we suggest two short stories of six pictures each, extracted from Donald Duck's comics. Each child can choose one of the two following sequences: a swarm of moths in feasting the city of Paperopoli and Paperinik chasing a gangster during the night. The child can choose, among the sounds previously 'exsplored', the most adequate one to add a sound track; moreover he can try to create the right sound-track for each picture elaborating at the moment the chosen sound so as to follow the development of the story. The pictures arranged for children from 9 to 15 years old have been extracted from comics of 'Corto Maltese'. They offer, for contents and graphic, a more complex elaboration on the soundtrack: the main subject of the first sequence is a man followed by his own shadow, the second is a whirling fight between two figures. Also in this case, children can choose among the sounds previously known and can elaborate a sound sequence to associate with the chosen pictures; obviously the sound elaboration will be prevalently leaded by the simbolical and emotional suggestions that the child receives looking over the pictures. In this second game timbric exploration was enriched by musical sessions based on the use of the voice and by the children creating sounds using didactic instruments and everyday objects, as well as by narrative interludes on the part of an operator. In this way it was possible to realize a dramatization of the cartoon foreseeing more areas of intervention: narration, sonorisation

[5] A timbre can be qualified as hot, cold, rough, soft, dark, light, luminous, opaque, dry, etc. The explanation of these synaesthetic perceptions constitute a premise for subsequent correlation between the sensorial sensations produced by the sound and those evoked by the images.
[6] Sounds configurations were excerpts from the composition "Contropasso" by Michelangelo Lupone.
[7] Compositions were based on improvisation techniques.

with vocal and instrumental interventions, sonorisation with electronic sounds processed in real time. Vocal and instrumental sound exploration was not limited to the production of descriptive or onomatopoeic sounds, but was designed in such a way as to underline from the expressive, allusive, symbolic point of view the psychological influences and suggestions provoked by the images, in the same way as with timbric exploration on the computer.

The object of this computer music laboratory was to show that children could consciously make a symbolic sound event through the manipulation of the timbre, bypassing the formal study of conventional music theory. Now that is possible to process sound in real - time with the use of a computer, we now have at our disposal a tool that will enable us to use the language of sound in new contexts. A child uses speech and pictures before learning the organized and hierarchic rules of language and art. In a similar fashion we retain that he might use a not verbal language consisting of sounds by creating spontaneous correlations between images, sounds, meanings. The SDP approach is intuitive and, above all, allows timbric manipulation in a simple way. This is a powerful tool since timbre is one of the most important sound parameters contemporary music deal with. By making these manipulations easy to achieve - by linking them to gestures and divorcing them from traditional music study - it allows children to approach music from an improvisational point of view by leaving them free to explore the language as well as the substance of music, an important stimulus for his subsequent research development.

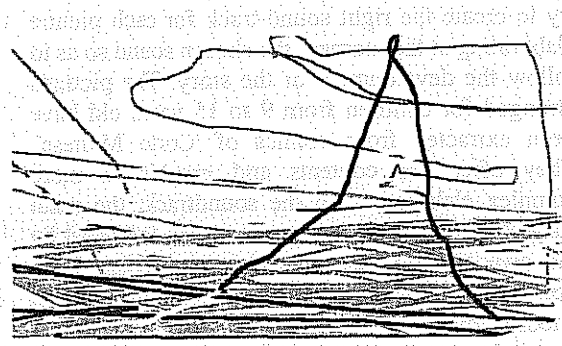


Figure 2. The game 'Sonic Draw' with SDP: drawing done by Simone (aged 8).

4 Improvements

The current version of SDP is able to solve a great many performance problems relative to the management of sound processing in real time; it is also an excellent compositional and didactic instrument. Looking back on the experiments that have been done, the advantages of the program and its limitations are easily understood, and this is an important factor when approaching a project in which one desires to enhance the possibilities of his 'instrument'. The traditional concept of 'musical instrument' implies the concept of 'technical limitations' that make its use limited to a specific range of actions. At the moment some program modifications are in course of development, including the transformation of the performance environment from bidimensional space to tridimensional space and the intensification of the possibility of controlling the algorithms during performance. The interface will be modified so that other hardware input devices (such as graphic tablets, joysticks, keyboards, analog digital converters, pedals and sensors of various kinds) can be used in the real-time manipulation of sound. One will be able to create gestures that will go far beyond the possibility of the simple computer mouse.

References

[1] AA.VV. *Nuova Atlantide. Il continente della musica elettronica 1990 – 1986*, editor Doati R., Vidolin A., La Biennale di Venezia, Settore Musica, Venezia 1996.

[2] Armani F., Bizzarri L., Favreau E., Paladin A., "Mars – DSP environment and applications", in proceedings ICMC, San Jose, 1992.

[3] Asta V., Chaveau A., Di Giugno G., Kott, J., *The real time digital suynthesis sistem 4X*, automation and instrumentation, n° 2 February 1980.

[4] Cavaliere S., Di Giugno G., Guarino E., "Mars: X20 architecture and SM1000 sound generation board description", in proceedings ICMC 1992.

[5] Delalande F, *Le condotte musicali*, ed. CLUEB, Bologna 1993.

[6] Del Duca L. M., *Musica digitale*, Franco Muzzio Editore, Padova, 1987.

[7] De Vitis A., Pellecchia A., "Fly30: un sistema programmabile per l'elaborazione numerica dei segnali musicali in tempo reale", (Fly30: a programmable system for sound processing in real time), in *proceedings XX Convegno Nazionale "20 anni di acustica in Italia*, AIA, Roma 1992.

[8] Paynter J., *Sound & Structure*, Cambridge University Press, 1992 (trad. it. *Suono e Struttura*, EDT, Torino 1996)

[9] Staccioli G., "Rumori grafici e graffiti sonori", in *Atti del Convegno dell'Associazione Corale Goriziana*, 1991.

[10] Tafuri J., Spaccazzocchi M., Stefani G., *L'educazione musicale nella scuola di base*, ed. La Scuola, Brescia 1979

A Didactic Tool for Composition in Continuous Space

Colby N. Leider[1] and Kristine H. Burns[2]

[1] *Princeton University*
Department of Music
Princeton NJ 08544 USA
http://music.princeton.edu/~colby
colby@music.princeton.edu

[2] *Florida International University*
School of Music
Miami FL 33199 USA
http://www.fiu.edu/~burnsk
burnsk@fiu.edu

Abstract

We describe a new compositional tool for drawing control curves and function tables for use in software synthesis languages and compositional environments. The program, curvePainter, allows users to create and graphically modify curves using a variety of techniques. Furthermore, curvePainter can generate curves from spectral features of sound files for use with other programs. This paper first presents historical motivation for the program and then describes its features. Finally, brief examples of didactic uses of curvePainter for sound processing and stochastic composition are presented.

Background

With the advent of software synthesis languages came the notion of *control functions* or *control curves*—a paradigm that enabled composing within a continuum. Iannis Xenakis' seminal book *Formalized Music* (1971) [1] and other writings (e.g., [2]), the UPIC system (1977), the work of Julio Estrada (e.g., [3]), and the *gen* functions of Music-N-style languages all illustrate applications of composing with control curves in continuous spaces. Recent programs like Cecelia [4], CurveControl [5], and StochGran [6] supplement the notion of graphical generation of control functions with robust interfaces to a target software synthesis environment.

Motivation and Overview

curvePainter is a new graphical music composition tool that generates control functions and score files for use with software synthesis languages and composition programs. The motivating force behind curvePainter was to create such a program that is not bound to any particular language, method, or platform. Additionally, curvePainter is well suited to introduce students to basic concepts of software synthesis languages and computer-mediated composition.

The program does not produce sound itself; it creates ASCII files of numbers for use in environments like SuperCollider, Max/MSP, Cmix, OpenMusic, and Csound. It allows the user to create and view up to twenty-four independent vectors of floating-point numbers—"curves"—using a variety of techniques. Additionally, curves may be extracted from various characteristics of sound files, such as the spectral centroid, pitch, amplitude envelope, and spectral peaks. Graphical interfaces to the legacy *gen* functions are also provided.

Many processing functions are supplied for shaping individual curves, including filters, a compander, and a phase shifter. Furthermore, individual curves may be used to warp the shape of other curves. For example, one curve may "morph" into another to create a third curve, or several curves may be averaged over time to create another curve.

Because the program makes no assumptions about the target environment, files generated with curvePainter can be used to control any parameter in any kind of synthesis model or to control compositional parameters in a score. Curves might be used to control the time-varying amplitudes of harmonics in an additive synthesis instrument; the density, duration, and frequency of grains in a granular synthesis instrument; the trajectories of point sources in a spatialization instrument; or the breath pressure, embouchure, pitch, and vibrato of a flute physical model. The curves may also be used to create score files.

Figure 1 illustrates the Curve Palette. The program draws each of up to 24 curves in separate colors.

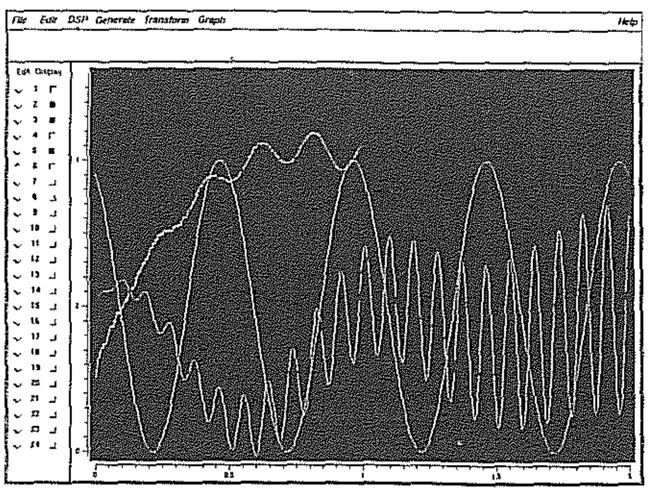


Figure 1. *The Curve Palette.*

THE GENERATE MENU

Users may generate curves with the basic tools provided under the *Generate* menu. The primary curve generation techniques currently available are *Gaussian*, *Impulse*, *Ramp*, *Random*, *Sinc*, *Sinusoid*, *Square*, *Triangle*, *Average*, *Arbitrary Function*, and *FM*. When the user selects one of these from the Generate menu, a window of options is immediately presented. Most of the functions simply require data such as frequency, amplitude, phase, etc., to generate a new curve, as required by that particular function. The Average method, however, is unique in that other curves must be currently active. It simply creates a curve whose values represent the arithmetic mean of the curves selected to average. Also, users may enter a mathematical formula in terms of x and y (for the horizontal and vertical dimensions of the graph, respectively) to generate a curve by selecting "Arbitrary Function."

The curves generated using these techniques are by nature often quite simple. They serve as source material to create more complex curves via warping and modification in the *Transform* menu.

THE TRANSFORM MENU

Once basic curves have been created, they may be modified using various techniques. The Transform menu, shown in Figure 2, allows users to operate on previously drawn curves. Operations such as "add clipboard" and "morph to clipboard" use the current curve in the clipboard's memory to transform the currently selected curve.

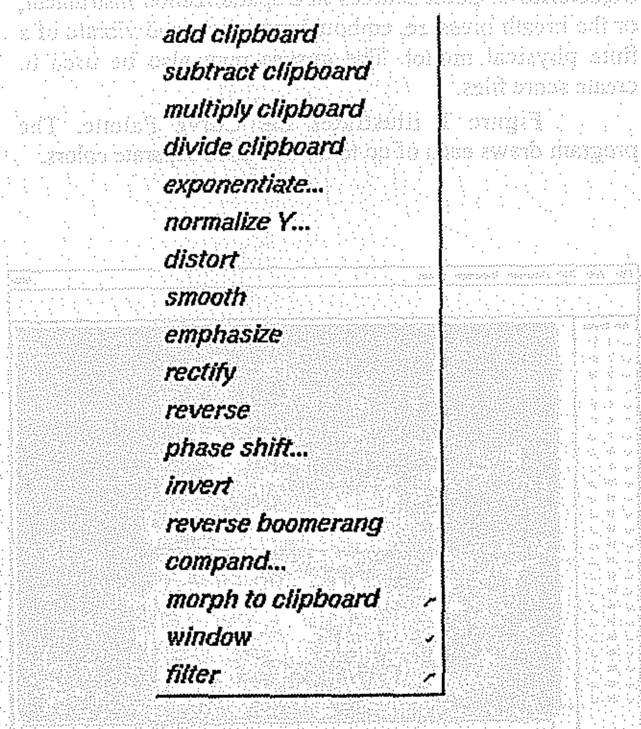


Figure 2. *The Transform menu.*

THE DSP MENU

curvePainter can extract curves from spectral features of sound files, such as the spectral centroid, pitch, amplitude envelope, and spectral peaks. When a user opens a sound file, a spectrogram immediately appears, along with a time-domain representation. The user then may select an operation to perform from the *DSP* menu that will extract a particular feature as a curve. Once the spectral feature has been turned into a curve, it may be modified and used like any other curve.

THE GRAPH MENU

The *Graph* menu features controls for the physical appearance of the curves. The user may also toggle a grid overlay and choose between logarithmic and linear axes.

THE FILE MENU

The *File* menu allows users to save sets of curves for future modification in curvePainter. The *Export* function presents the user with various options for creating ASCII text files of floating-point numbers derived from the curves. Controls over the layout of the text file and the sampling frequency of the curves are also available.

A COMPOSITIONAL EXAMPLE: SOUND PROCESSING

Beginning students of electroacoustic music sometimes find it difficult to integrate unified local and global features of a composition. One way to present this concept is to use a single compositional feature—such as the desired global amplitude envelope or frequency range—to inform local choices of timbral modification. For example, a simple one-to-one compositional mapping could be designed around a particular dynamic curve (e.g., loud-soft-loud, represented by a triangle window function). That particular window shape can easily be mapped to the envelope of individual sound events. Conversely, the dynamic shape of a particular sound object might suggest certain choices for the dynamic shape for the entire piece.

Both of these examples are easy to realize in curvePainter. Here we will examine the second one, that of mapping the dynamic curve of a sound object to a global dynamic parameter, more closely. We will assume the student has created a score file to order the sound objects in a meaningful way.

First, we load the sound into curvePainter. Then, we select *Amplitude Envelope* from the DSP menu. A curve is then drawn which represents the amplitude. Now that we have a curve of the amplitude envelope, we can export it as a file of floating-point numbers into a software sound synthesis or composition program. Once the curve is then loaded into the target environment, it can be used to effect any desired parameter—in this case, the global amplitude shape of the piece.

ANOTHER EXAMPLE: STOCHASTIC COMPOSITION

curvePainter is particularly applicable to stochastic composition. For example, if a student composer is writing for string quartet, he or she may identify several parameters to change throughout the piece: density of events, pitch dispersion, tessitura, dynamics for each instrument, and overall acoustical consonance derived from some given measure, for instance.

Then student simply creates the desired curves for these parameters within curvePainter. Once in curvePainter, the curves can even interact: the tessitura curve might slowly change into the shape of the pitch dispersion curve, and vice-versa. The student can then export the curves to a file for processing in another language or notation program.

STATUS AND FUTURE WORK

curvePainter was developed in Tcl/Tk and C and uses Kåre Sjölander's Snack sound extension [7] and the Tcl/Tk BLT extension [8]. It runs under Linux and can also run under Irix and Windows 98/NT. A version for MacOS X is planned. Future work might also include a World Wide Web-based version.

A freehand drawing tool should be added to allow users to intuitively create arbitrarily shaped curves. Currently, users are bound to the built-in curve generation and transformation techniques, although it is straightforward to write plug-ins in TCL or C that provide new functionality.

A final area of improvement relates to the program's ability to calculate statistics of curves. Currently, curvePainter provides a "Show Statistics" feature, available under the Edit menu, that calculates the average deviation, skew, kurtosis, maximum value, minimum value, mean, median, standard deviation, and variance of the selected curve. The window is shown in Figure 3.

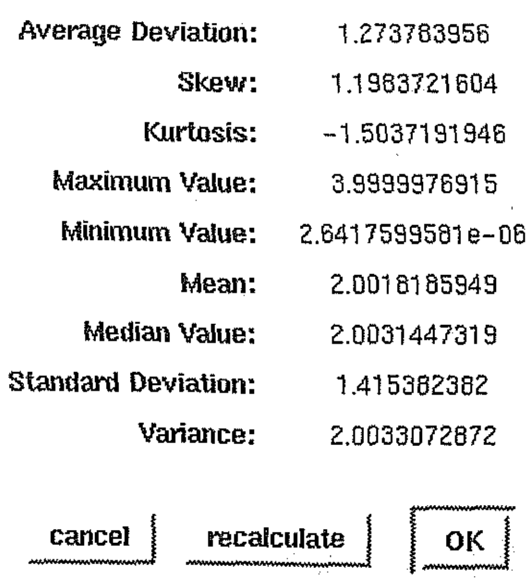


Figure 3: *The Statistics window.*

A helpful feature of the program would be the ability to warp curves so that they exhibit specified statistical values. For example, the user might want to increase the standard deviation of a curve's values over time. Although there is of course no single method of uniquely changing a curve's statistics, selected options might be computed and displayed.

WORKS CITED

[1] Xenakis, I. 1971. *Formalized Music.* Bloomington, IN: Indiana University Press.

[2] Xenakis, I. 1986. "Arhitektura, informatika, originalnost / L'architecture, l'informatique, l'originalite." In *Skladateljske sinteze osamdesetih.* Zagreb: Muzicki Biennale; Muzicki Informativni Centar KDZ, 85–90, 201–06.

[3] Estrada, J. 1994. "Le continuum en musique: Structure et ouvertures en composition, ses derivations esthetiques." In *Asthetik und Komposition: Zur Aktualitat der Darmstadter Ferienkursarbeit.* Mainz, Germany: Schott, 50–65.

[4] Piche, J., and A. Burton. 1998. "Cecilia: A Production Interface to Csound." *Computer Music Journal* 22(2):52–55.

[5] Bencina, R. 1995. *Curve Control.* Available online at http://www.audiomulch.com/~rossb/.

[6] Helmuth, M. 1998. "StochGran on Silicon Graphics IRIX." In *Proceedings of the 1998 International Computer Music Conference.* San Francisco: International Computer Music Association, 467–470.

[7] Sjölander, K. N.d. "The Snack Home Page." http://www.speech.kth.se/snack/.

[8] —. BLT home page. Available online at http://www.tcltk.com/blt/index.html.

Lip*Sync*™ 1.0:
A System for Real-Time Virtual Characters Lip-Synching and Facial Modeling

Mario Malcangi, Raffaele de Tintis

DSPengineering
Milan, Italy
malcangi@dspeng.com
rdt@dspeng.com
www.dspeng.com

Abstract

In the present paper we present Lip*Sync*™, a specific application of real-time phonemic recognition dedicated to virtual characters lip-synching.
This system introduces some innovative elements in important application areas such as virtual sets, characters animation and cartoons industry.
We will hereby present it and discuss some of the implemented design choices. Moreover, we will make a comparison with other systems which are based on different technologies and which are presently available on the market.
Lip*Sync*™ 1.0 runs on Windows 95/98/2000™ and Windows NT™ platforms and can produce lip-synching data for Softimage™, Maya™, 3D Studio MAX™ and, in general, for every system with *MIDI functionalities.*

1. General description

The system analyses speech signals, then uses a phonemic detector based on a real-time neural network to identify the speaker's lips position. At each analysis frame, a coding of the identified phoneme is transmitted and it can be used by 2D and 3D modeling systems – such as Softimage™, Maya™, 3D Studio MAX ™ and so on – to control lips characteristics and, partly, virtual characters facial modeling. Fig.1 describes the Lip*Sync*™ recognition/coding framework. Once the gender of the speaker is specified, this system is able to carry out the coding on an unlimited vocabulary and this is completely independent on the language of the speaker (speaker independent, vocabulary illimited coding).

Presently, Lip*Sync*™ can be connected to other external devices through MIDI interface. In this way, the user is not forced to choose a specific modeling environment and, moreover, lip-synching data can be easily transmitted among different systems and platforms. It is also possible to exploit all the visual and editing features of any sequencer, for lip-synching data processing. What is more, it is possible to add audio and video effects which can be triggered by the recognized events and consequently be in synch with the vocal flow. The use of the latter might be especially useful when it comes to music applications where it can be integrated with environments such as MAX/MSP. These extensions are likely to be developed in the near future.

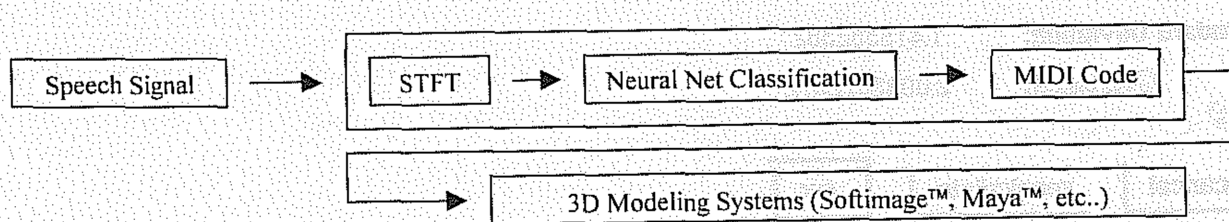


Fig. 1: The Lip*Sync*™ 1.0 Recognition Framework

2. Recognized events

Several models for speech animation have been proposed so far and each one has different complexity levels. One of the first Madsen's studies in 1969 showed that a number of parameters smaller than 10 should be considered as the starting point of every serious speech animation project [1]:

- Recognition of the vowels *a*, *e*, *i* which are pronounced with open lips
- Recognition of the consonants *p*, *b*, *m* which are pronounced with closed lips
- Recognition of *u*, *o*, *w* which are pronounced with round-shape mouth
- Recognition of *f*, *v* in which the lower lip touches the frontal teeth
- Possibility of interpolation between these positions so that any abrupt changes in movements should be avoided.

This very simple model is sufficient for cartoons production. Yet, some other more sophisticated models were proposed by Di Paola, Pearce, Cohen and many others.

The events recognized by Lip*Sync*™ are the following:

1. Volume
2. No-speech frames
3. Vowels
4. Consonants

Volume is continuously tracked and can be used to control mouth opening strength and some facial movements.
By executing simple experiments using loud intensity voice, you can observe that:

- Facial muscles become stretched proportionally to speech amplitude
- Eyebrows tend to frown
- Forehead tends to wrinkle
- Nostrils tend to extend

You can use speech amplitude data to produce facial movements embedding this control in your model.
It should be also noted that a correct use of energy, continuously produces modulated lips shapes, resulting in more natural mouth movements. In fact, modulation always produces different instances of the same lips position so the mouth shape corresponding to different occurrences of the same phoneme will always be different, according with what happens in real-life mouth movements.

Vowels are individually recognized: a, e, i, o, u

Consonants are grouped in 5 classes corresponding to similar lips positions:
[(l, n, r), (b, m, p), (v, f), (d, t), (c, s, z, g)].

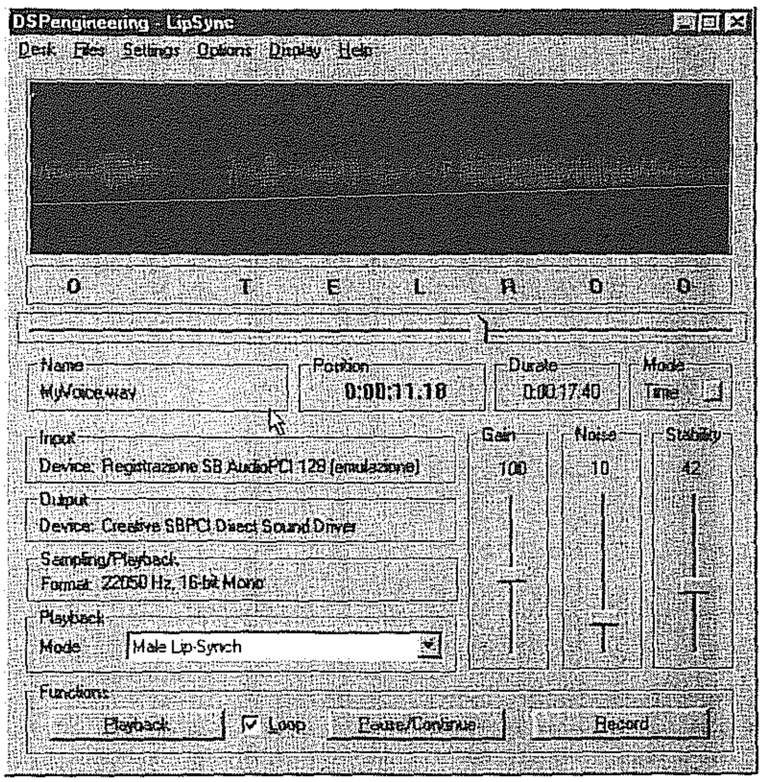


Fig. 2: Lip*Sync*™ 1.0 Main Panel

3. Transmission specifications

Coding results are transmitted on MIDI transmission channel 1, using the following 12 continuous controllers :

- Silence, transmitted on control 1 (Modulation Wheel)
- Vowel A, transmitted on control 2 (Breath Controller)
- Vowel E, transmitted on control 3 (Undefined)
- Vowel I, transmitted on control 4 (Foot Controller)
- Vowel O, transmitted on control 5 (Portamento Time)
- Vowel U, transmitted on control 6 (Data Entry MSB)
- Class of consonants L, N, R, transmitted on control 7 (Main Volume)
- Class of consonants B, M, P, transmitted on control 8 (Balance)
- Class of consonants V, F transmitted on control 9 (Undefined)
- Class of consonants D, T transmitted on control 10 (Pan)
- Class of consonants C, S, Z, G transmitted on control 11 (Expression Controller)
- Energy, transmitted on control 30 (Undefined)

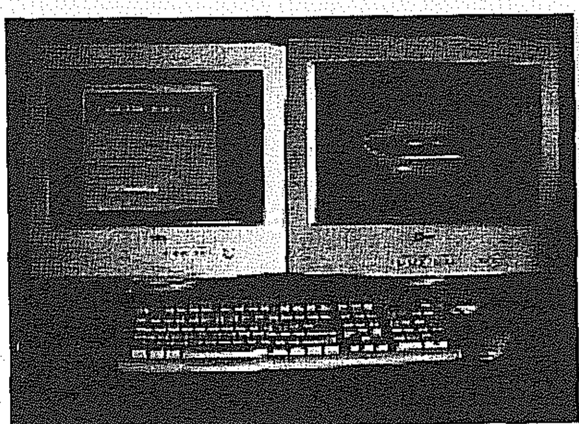


Fig.3: Lip*Sync*™ controlling a Softimage™ 3D model

4. Speech recognition/synchronization for virtual characters lip-synching

The phonetic recognition problem is hereby presented in an innovative way which is quite different from the traditional approach of speech-recognition in telephony applications, or in speech-to-text coding.
The first major difference is that it not necessary to have a semantic comprehension of words. As the coding is based on phonetic units, it is no longer necessary to wait for a complete words recognition before transmitting the results. Moreover, such a phonetic recognition also permits to have a total language-independent system so that distinct training sets during the neural net learning phases are unnecessary.

The key to have good quality synchronization is to respect all the movements of the mouth and the articulation limitations which characterize the natural mouth shape during continuous speech phases.
Some phonemes have unique spectral characteristics but their differences are mainly due to a different position of the tongue which does not really affect the mouth modeling [1].
A good recognition of the different consonants classes (plosives, fricatives, nasals, semivowels etc.) [8], is essential to produce a convincing synchronization.
B, m, and p are very important consonants because, from a visual point of view, they can be considered the most important accents of speech. They are basic elements [1], to ensure a good quality synchronization process.
One of the very first problems we met during the initial phases of the system development was due to the fact that we needed to soften all the transitions in lips movements.
As codings are transmitted at each frame, there was actually no way to make the system wait for the end of the word before starting the recognition of it, and this could have created errors and random coding transmissions in case of very similar phonemes.
This generally occurs when the analysis window switches from a particular phoneme to an other one, during fast speech transitions. Because of blurred spectral informations, the differences between similar vowels and consonants tend to disappear and this affects the entire recognition process.
To avoid any abrupt lips movements, we decided then to continuously adjust the interpolation degree among mouth positions, proportionally to the average coding uncertainty level.
By increasing the interpolation degree between mouth positions during fast speech transitions, where coding errors are more frequent, it was possible to have a good quality lips synchronization. Infact, during fast speech phases, mouth physical limitations tend to minimize the differences in lips movements and the higher interpolation rate is not really evident.

5. A non-invasive fast system use

The entire creation of the Lip*Sync*™ project, was made possible thanks to technologies designed to be both non-invasive and easy to use.
The speaker does not have to wear any sort of body devices or instruments which allow all the movements to be completely unlimited or uninfluenced by the system.

Contrarily to what normally happens with devices based on the use of different technologies, either magnetic or laser sensors or based on optic approaches, all initial operations of calibration and markers positioning on the face, are unnecessary.
Furthermore, Lip*Sync*™ is based on technologies which permit a speaker-independent approach so that also the phases of initial neural net training on the speaker's voice, which are sometimes present, can be eliminated.
Thanks to all these characteristics it is possible to reduce the time of preparation and use of the system.

6. Future directions

Currently the costs for virtual characters animation are so high that only a few industries, working with the animation pictures industry or the advertisement sector, can afford them.
These high costs are also due to the lack of technologies that could make facial animation systems automatic while some remarkable progress has been made for body movements motion-capture. To fill this gap it is necessary to reach a good feature-extraction level for continuous speech flows so that it is possible to control not only lips movements but also facial modeling.
The recognition of prosodic characteristics, for instance, or the identification of question and positive sentences as well as of the speaker's feelings (anger, peace of mind and so on) could make the whole facial modeling process automatic and it could drastically reduce production costs.

Trademarks

Windows 95/98/2000™, Windows NT™ are registered trademarks of Microsoft Corp.
SOFTIMAGE™ is a registered trademark of Softimage Inc.
Maya™ is a registered trademark of Silicon Graphics Ltd.
Lip*Sync*™ is a registered trademark of DSP*engineering*
All other products are trademarks or registered trademarks of their respective owners.

References

[1] F.I.Parke, K.Waters, *Computer Facial Animation*, A K Peters, Ltd., 1996.
[2] M.Cohen and D.Massaro. "Synthesis of visible speech" in *Behavioral Research Methods and Instrumentation*, 22(2):260-263, 1990.
[3] M.Cohen and D.Massaro. "Modeling coarticulation in synthetic visual speech" in N.M.Thalman and D.Thalman editors, *Models and Techniques in Computer Animation*, pp.141-155. Springer-Verlag, Tokyo, 1993.
[4] A.T.Dittman. "The body movement-speech rhythm relationship as a cue to speech encoding". In *Nonverbal Communications*, Oxford University Press, Oxford, 1974
[5] S.Duncan. "On the structure of Speaker-Auditor interaction during speaking turns" in *Language in society*, 3:161-180,1974
[6] M.Cohen and D.Massaro. "Development and experimentation with synthetic visual speech", *Behavioral Research Methods, Instrumentation and Computers*, 26:260-265, 1994.
[7] W.S.Condon and W.D.Osgton."Speech and body motion synchrony of the speaker-hearer" in *The perception of Language*. Horton and Jenkins, 1971.
[8] J.A.Markowitz. *Using Speech Recognition*, Prentice Hall, 1996.

Hardware Real-Time Sound Spatialization with Csound

Gabriel Maldonado
g.maldonado@agora.stm.it - http://web.tiscalinet.it/G-Maldonado
Rome - Italy

Abstract

This demo concerns some new features of ***DirectCsound*** related to sound spatialization (DirectCsound is a low-latency real-time-oriented version of Csound). At present time several hardware-accelerated audio cards are available in the market which support Microsoft's ***DirectSound3D API*** in hardware. Some of these cards have quadraphonic output, allowing four-channel surround audio. Even if DirectSound3D API was created primarily for 3D games, its implementation in new generation of audio cards allows to generate high-quality sound indeed. So new possibilities are opened in musical field too, especially in situations where real-time is required. New opcodes of DirectCsound can take advantage of hardware-acceleration of such cards, allowing real-time surround capabilities; they support almost all ***DirectSound3D*** APIs, as well as Creative's ***EAX 2.0***, which provide additional environmental effects such as high quality reverberation, reflections, sound obstruction/occlusion by intervening objects etc. Musicians can use the surround opcodes of DirectCsound to create separate sound sources that can realistically move around in the 3D aural space, by controlling them in real-time.

1 Introduction

3-D audio generally needs a huge amount of calculation, so, even using latest-generation CPUs, realtime spatialization would leave only a small amount of processing percent to the computer CPU to be used for other purposes. For this reason, several manufacturers have proposed some hardware-accelerated models of audio cards. Since audio quality of such DSP processing is very good (all stages are processed with 32-bit precision), it is possible to use the provided hardware features for computer music applications, even if they were mainly designed to obtain realistic 3D video-games audio effects.
I added to DirectCsound an interface to API for direct access to hardware-acceleration provided by Microsoft® and Creative®, making them useable by means of a special set of *opcodes*.

2 Spatialization APIs

Latest versions of DirectCsound provide a set of opcodes that support two spatialization-related APIs:

1. DirectSound3D
2. EAX (Environmental Audio Extensions)

Microsoft's ***DirectSound3D API*** mainly deals with the following factors (related to listener perception of sound sources displaced in determined positions of three-dimensional space):

- *Volume.* The farther an object is from the listener, the quieter it sounds. This phenomenon also called *rolloff.*
- *Arrival offset.* A sound emitted by a source to the listener's right will arrive at the listener's right ear slightly before it arrives at the left ear. (The duration of this offset is in the order of milliseconds.)
- *Muffling.* The orientation of listener's ears ensures that sounds coming from behind the listener are slightly muffled compared with sounds coming from in front of the listener. In addition, if a sound is coming from the listener's right, the sounds reaching the left ear will be filtered by the mass of the listener's head.

Creative's ***Environmental Audio Extension API (EAX)*** mainly deals with the following perceptual factors:

- the apparent *size* of the room surrounding the listener;
- *duration and tonal color* of the reverberation's decay;
- the *amount and delay* of reflections and reverberation;
- natural *statistical behavior* of reverberation in enclosed spaces;
- a reverberation *rolloff factor* that allows to customize or modify the statistical reverberation model;
- an adjustable *air absorption* model that filters out high frequencies with increasing distance (to simulate moisture in the air such as high humidity or fog);
- *occlusion* (the phenomenon that occurs when a wall that separates two environments comes between source and listener);

- *obstruction* (the phenomenon that occurs when source and listener are in the same room, but there is an object placed directly between them).

3 The DirectSound3D model

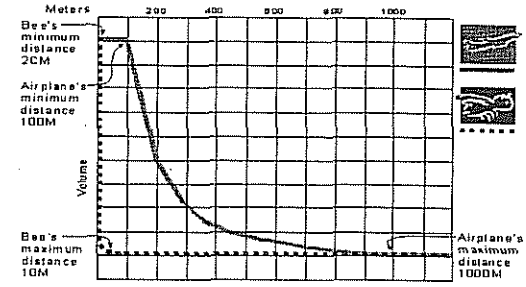
DirectSound3D API introduces two classes of aural objects: the ***Listener*** and the ***Sound Sources***. Even if many sound sources are possible in the same sonic field, only one listener can be possible at a time.
The ***Listener***, as well as every ***Sound Source***, has a ***position*** i.e. their location in three-dimensional space.
Listener experiences an identical sonic effect when an object moves in a 90-degree arc around them or if they move their heads 90 degrees relative to the object. However, when several sound sources are present, it is much simpler to change the position or orientation of the listener than to change to positions of every other object in a scene. For this reason listener has a ***position*** and an ***orientation***.
A *Sound Source* is *punctiform*, and normally *omni-directional*; the farther the listener is from the sound, in any direction, the quieter the sound. However there are many kind of sources which don't send the same level of sound to all directions equally (for example speakers). In this case we need a sound source with a position and an orientation (named *sound cone*).

3.1 Sound Cones

In DirectSound3D, sound cones include an ***inside cone*** and an ***outside cone***. Within the inside cone, the volume is at the maximum level for that sound source. For example, by setting the outside volume to -10000,the sound source will be inaudible outside the outside cone. Between the outside and inside cones, the volume changes gradually from one level to the other.
Every sound source is a sound cone, but normally these sound cones behave like *omni-directional* sound sources. For example, the default value for the volume outside the sound cone is zero; unless user changes this value, the volume will be the same inside and outside the cone, and sound will not have any apparent orientation. Additionally, you could make the sound-cone angles as wide as you want, effectively making the sound cone a sphere.

3.2 Minimum and Maximum Distances

As a listener approaches a sound source, the sound gets louder. Past a certain point, however, it isn't reasonable for the volume to continue to increase; either the maximum (zero) has been reached, or the nature of the sound source imposes a logical limit. This is the minimum distance for the sound source. Similarly, the maximum distance for a sound source is the distance beyond which the sound does not get any quieter. The minimum distance is especially useful when an application must compensate for the difference in absolute volume levels of different sounds. For example, though a jet engine is much louder than a bee, for practical reasons these sounds must be recorded at similar absolute volumes. An application might use a minimum distance of 100 meters for the jet engine and 2 centimeters for the bee. With these settings, the jet engine would be at half volume when the listener was 200 meters away, but the bee would be at half volume when the listener was 4 centimeters away. This concept is shown in the following picture:



The default values for the 3D sound effects mimic the natural world. User could choose to change these values, however, to make the effects more dramatic.
DirectSound3D API uses ***meters*** as the default unit of distance measurements but this unit can be changed by the user.

4 The EAX model

EAX implements a set of three groups of features:

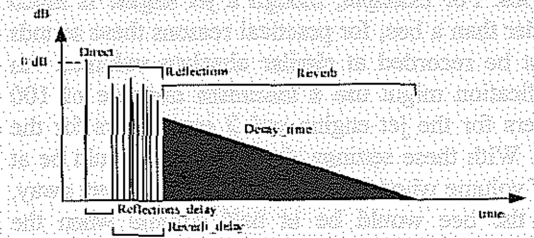
1. Reverberation control
2. Occlusion and Obstruction
3. Source directivity

4.1 Environmental Reverberation Control

EAX offers a complete set of ***reverberation parameters*** that allow the user to control the intensity and delay of reflections, to continuously adjust the environment size, and to set the direct path and reverberation filters for each sound source. These direct controls allow to tweak environmental reverberation to get precisely the aural surroundings.
The reverberation response is defined by the following parameters:

- the energy in each of the three sections *Direct*, *Reflections*, and *Reverb* at low frequencies;
- the *Reflections Delay* and the *Reverb Delay*;
- the "*Direct filter*", a low-pass filter that affects the Direct component by reducing its energy at high frequencies;
- the "*Room filter*", a low-pass filter that affects the *Reflections* and *Reverb* components identically by reducing their energy at high frequencies;

- the *Decay Time* at low and high frequencies;
- the *Diffusion* and the *Size* of the room;



EAX reverberation response model

To help to provide a better perception of listener-to-sound-source distance, EAX provides several different modes of automatic ***distance-effects*** management. They manage attenuation and filtering effects that change according to the distance between the listener and a sound source. EAX provides two methods for automatically attenuating the reflected sound (Reflections and Reverb) according to source-listener distance:

- The first (and default) method simulates the natural *rolloff* of reverberation vs. distance in typical rooms. This method is based on a physically proven statistical reverberation model that allows predicting the intensity and tonal characteristics of reflections and reverberation in natural environments according to source-listener distance, room size, reverberation decay time and source directivity.
- The second method for controlling the rolloff of reflected sound vs. distance uses a parameter called ***Room Rolloff Factor*** that acts like DirectSound3D's ***Rolloff Factor*** but affects the reflected sound component instead of the direct-path sound component. The default value of Room Rolloff Factor is 0.0, which implies that the reflected intensity does not vary with distance, while the default value for the (Direct) Rolloff Factor is 1.0. Setting Room Rolloff Factor to 0.5 instead would imply that the reflected sound is also attenuated with increasing distance, although not as fast as the direct path.

EAX also provides an additional parameter to control the *attenuation at high frequencies* caused by the propagation medium. This parameter is called ***Air Absorption***. It is possible, for example, increase it in order to simulate higher humidity in the air.

4.2 Occlusion and Obstruction Effects

Occlusion occurs when a wall that separates two environments comes between source and listener. There's no open-air sound path for sound to go from source to listener, so the sound source is completely muffled because it's transmitted through the wall. EAX occlusion properties provide parameters that adjust wall transmission characteristics to simulate different wall materials and thickness. You can, for example, use occlusion properties to make a voice or noise sound very realistically as if it were coming from behind a door or from outside the listener's house.

Obstruction occurs when source and listener are in the same room but there's an object directly between them. There's no direct sound path from source to listener, but the reverberation comes to the listener unaffected. The result is altered direct-path sound with unaltered reverberation. EAX obstruction properties can simulate sound diffraction around the obstacle or sound transmission through the obstacle, which provide rich aural cues about the nature of the obstruction. You can, for example, use obstruction properties to make a voice sound as if it were coming through a thin curtain or from behind a large pillar.

4.3 Source Directivity

DirectSound3D API already offers control of source directivity, where a sound source is stronger in one direction than in all others—like a trumpet, for example, that's strongest in the direction of the bell and attenuated to the sides and behind the bell. DirectSound3D sets the directivity across all frequencies equally, which usually not the case in the real world.

EAX makes source directivity sound much more natural by allowing the user to make sound sources more directive at high frequencies than at low frequencies. This makes the sound of directional sources such as a voice, a loudspeaker, or a horn sound much more realistic as the listener moves around the source, or as the source rotates or flies by the listener.

EAX's processing model for each sound source comprises an attenuation and a low-pass filter that are applied independently to the direct path and the reflected sound. All the sound-source properties in EAX have the effect of adjusting these attenuation and filter parameters relative to the Environment settings.

The important thing is that all parameters can be adjusted independently for each sound source.

The EAX 2.0 sound-source property set defines a variation relative to the baseline setting described by the listener properties. This variation results from several combined effects:

- Distance-dependent effects (statistical reverberation model, Room_Rolloff factor, and Air Absorption);
- Frequency-dependent source directivity;
- Occlusion and obstruction effects;
- Direct and Room attenuations and filters;

5 Csound opcodes

DirectSound3D and EAX APIs have been embedded in DirectCsound, in the form of opcodes. A list of opocde names follows:

- ***Initialization and outputs:***
  Init3dAudio,
  InitEAX,
  Out3d
- ***DirectSound3D Listener properties:***
  DsListenerPosition,
  DsListenerOrientation,
  DsListenerRolloffFactor,
  DsListenerDistanceFactor,
  DsListenerSetAll
- ***DirectSound3D Sound Source properties:***
  DsMode,
  DsPosition,
  DsMinDistance,
  DsMaxDistance,
  DsConeAngles,
  DsConeOrientation,
  DsConeOutsideVolume,
  DsSetAll
- ***EAX 2.0 Listener properties:***
  EaxListenerEnvironment,
  EaxListenerEnvSize,
  EaxListenerEnvDiffusion,
  EaxListenerRoom,
  EaxListenerRoomHF,
  EaxListenerDecayTime,
  EaxListenerDecayTimeHfRatio,
  EaxListenerReflections,
  EaxListenerReflectionsDelay,
  EaxListenerReverb,
  EaxListenerReverbDelay,
  EaxListenerRoomRolloff,
  EaxListenerAirAbsorption,
  EaxListenerAll
- ***EAX 2.0 Sound Source properties:***
  EaxSourceDirect,
  EaxSourceDirectHF,
  EaxSourceRoom,
  EaxSourceRoomHF, EaxSourceObstruction,
  EaxSourceObstructionRatio,
  EaxSourceOcclusion,
  EaxSourceOcclusionRatio,
  EaxSourceOcclusionRoomRatio,
  EaxSourceRoomRolloff,
  EaxSourceAirAbsorption,
  EaxSourceOutsideVolumeHF,
  EaxSourceFlags,
  EaxsourceAll

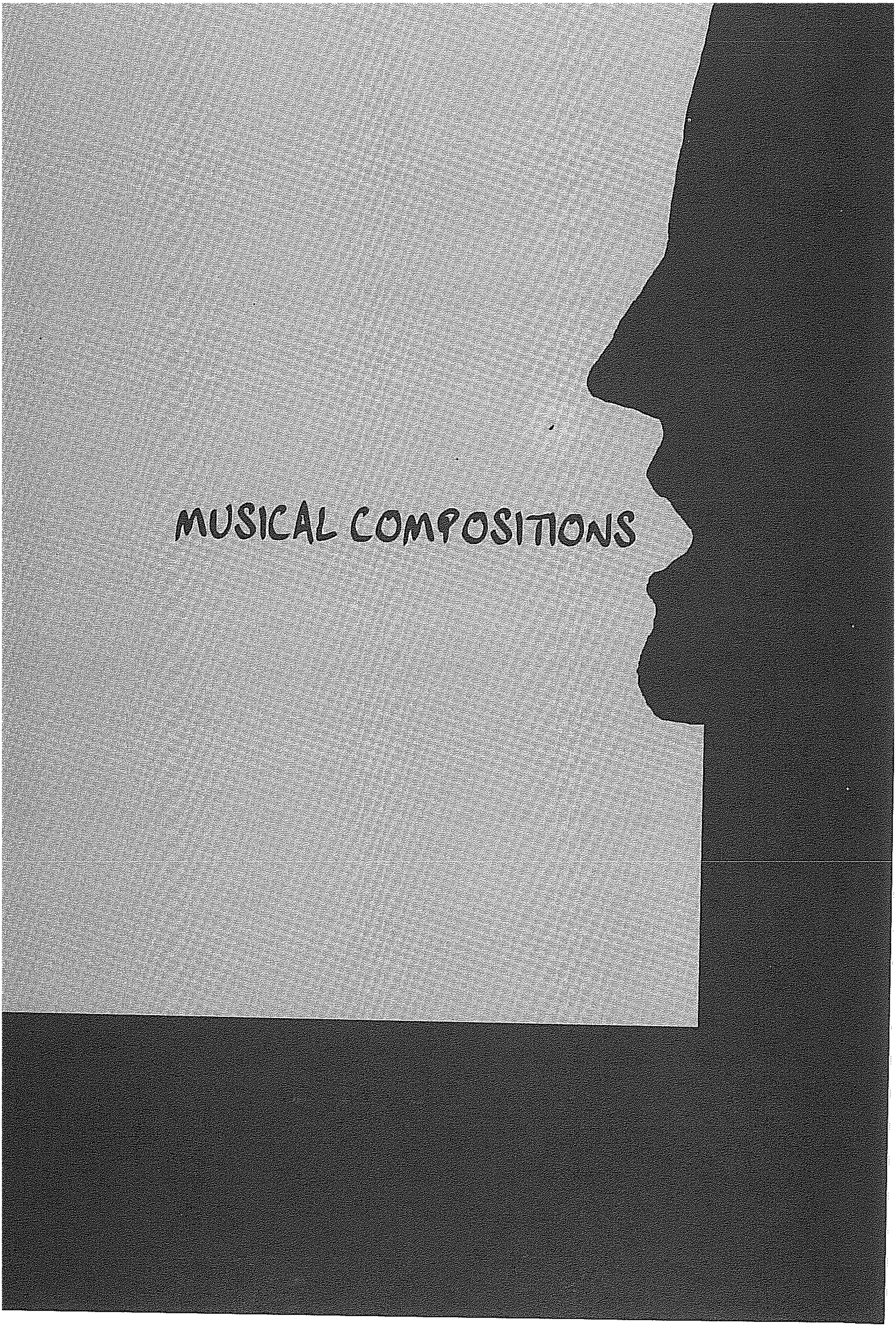
For more information about these Csound opcodes see the manual.

6 Conclusion

Though these opcodes are far from supporting a complete implementation of 3D sound spatialization, they allow to the user a convincing realtime performance and, when used with some specially designed cards (for example Sound-Blaster Live!), audio quality is very good. The main advantage is that using these opcodes only subtracts a minimal amount of CPU time to the other realtime synthesis processes, because almost all the calculation is done by card hardware. Unfortunately, at present time DirectCsound3D and EAX APIs are only available under Windows-based platforms. However other similar APIs will be available soon even under Linux (Open AL). Adapting DirectCsound 3D opcodes to the new APIs will probably be a trivial task.

At last, remember that DirectCsound is an Open Source project, so it is freely downloadable at the following URL:

http://web.tiscalinet.it/G-Maldonado



MUSICAL COMPOSITIONS

ACOUSMATIC MUSIC

Mathew Adkins
Mapping (1997 rev. 1999)

Mapping is a piece that depicts the slow evolution of a landscape: it leads the listener through unknown territory, arriving at landing points from time to time by means of transformed sounds from the real world. The work unfolds structurally as material emerges and is incorporated into the musical fabric. Some of the material is slowly transformed throughout the work, while other parts remain unchanged, serving as points of reference. More generally the work deals with the notion of 'beginning' and the perception of objects over time. The work was awarded 1st Prize at the EAR 95 Competition (Hungary) and in its revised version was a Laureate at the 5th Prix International Noroît.

Mapping è un pezzo che descrive la lenta evoluzione di un paesaggio. Esso guida l'ascoltatore attraverso un territorio sconosciuto, arrivando di tanto in tanto a punti di approdo per mezzo di suoni trasformati che provengono dal mondo reale.Il lavoro si sviluppa strutturalmente come del materiale che emerge ed è incorporato nella struttura musicale. Parte del materiale viene lentamente trasformato attraverso il lavoro, mentre altre parti rimangono invariate, servendo come punti di riferimento. Più generalmente il lavoro ha a che fare con la nozione di "inizio" e la percezione degli oggetti nel tempo. Il lavoro ha meritato il 1° premio al EAR 95 Competition (Ungheria) e la sua versione rivisitata è stata premiata al 5° Premio Internazionale di Noroit.

Michele Brugnaro
Migrazion (1998-1999)

The basic materials have been mostly drawn from a series of recordings of vocal and percussive sounds. The manifold transformation to which these materials have been subjected manifest themselves, in the last analysis, in the action of some compositional strategies, that may be singled out, in symbolic terms, as belonging to a conceptual sphere, eventually linked to the phenomenon of 'migratory streams'. The very concept of 'migration' must be understood here as an evocation of an existential path, which requires firstly an *eradication* – i.e. the loss of an initial identity, a traumatic event which usually occurs when one is confronted with the inevitability of a *change*, of an internal process of transformations directed toward the other-from self – and, secondly, a gradual *adjustment* to the new and unfamiliar situation. This adaptation may exhibit some features denoting irreversibility, or may oscillate, more or less cyclically, between refusal and acceptance of any change. These different approaches have many related consequences, relevant specifically for the organization of one's own horizon of thought. Under such conditions, this horizon is obliged, so to speak, to recontextualize itself and at the same time to extend its 'boundary limits', often experiencing a development following different (and sometimes completely unexpected) directions.
In such a way it becomes possible to see in an appropriate perspective the complex set of mutations of the basic materials, according to a formal logic founded on two principles: i) from one side, the progressive surfacing of unexpressed *affinities* between groups of sound events which usually are associated with well distinct and differentiated timbral-sonorial categories: starting from the oppositional pair voice-percussion – whose opposition is more seeming than substantial - , other infinite ramifications are easily obtained, which constitute the working background by means of which the first opposition may be intensified, specified and, at the same time, dissolved. These further conceptual dyads may either describe some relationships, internal to the single categories – e.g. voice (vowel/consonant, sung/spoken intonation, solo/choral effects); percussion (metal/wood/membrane, presence/absence of resonance, tuned/untuned instruments, individual/composite events) – or may conversely establish some intersections or connections between the two regions, in order to, for example, give a 'vocal' allure to percussive instruments, or extrapolate the 'percussive' articulations enclosed in a vocal gesture.
ii) This first group of transformations – comparable with the phase of 'eradication' and consequently becoming a potential source of concealed (afterwards, more and more explicit) tensions – has been brought to extremes by employing specific techniques of electroacoustical treatment (which assume, in particular circumstances, a prevailing role), namely convolution and granulation. This process of elaboration may lead to the creation of a variegated universe, crowded with 'chimerical' sonic images, which ultimately reveal, through their ephemeral appearance in the temporal scenario, not only all the perceptual and structural ambiguities which derive from their living presence in a sort of 'no man's land', but also show, in some cases, a kind of osmotic equilibrium, interpretable as a utopian realization of that intimate fusion toward which a truly *integrated* 'society' is directed.

I materiali fondamentali sono stati ricavati da una serie di registrazioni di suoni vocali e di percussioni. Le molteplice trasformazioni a cui questi materiali sono stati soggetti si manifestano, in ultima analisi, nell'azione di alcune strategie compositive, che possono essere distinte, in termini simbolici, come appartenenti a una sfera concettuale, eventualmente collegata al fenomeno di 'flussi migratori'. Il concetto esatto di 'migrazione' deve essere compreso qui come un'evocazione di un percorso esistenziale, che richiede principalmente uno sradicamento – come la perdita di una iniziale identità, un evento traumatico che accade normalmente quando si è confrontati con l'inevitabilità di un cambiamento, di un processo interno di trasformazione diretto verso l'altro sé – e, secondariamente, un graduale adattamento alla nuova ed insolita situazione. Questo adattamento può mostrare alcune caratteristiche che denotano irreversibilità, o possono oscillare, più o meno ciclicamente, tra rifiuto ed accettazione di qualunque cambiamento. Queste differenti approcci hanno molte conseguenze correlate, attinenti specificatamente all'organizzazione dei propri orizzonti di pensiero. In queste condizioni, l'orizzonte è obbligato, per così dire, a ricontestualizzare se stesso e allo stesso tempo ad estendere i suoi 'limiti di confine', spesso subendo uno sviluppo che segue diverse (e spesso completamente inaspettate) direzioni. In tal modo diventa possibile vedere in una prospettiva appropriata il complesso gruppo di mutazioni dei materiali fondamentali, in base a una logica esplicita fondata su due principi: i) da un lato, la progressiva emersione di affinità inespresse tra gruppi di eventi sonori che normalmente sono associati con categorie timbiche-sonore ben chiare e differenziate: partendo dalla coppia opposta voce/percussioni – la cui opposizione è più apparente che sostanziale – , si ottengono facilmente altre infinite ramificazioni. Queste coppie possono anche rappresentare alcune relazioni, interne alle singole categorie – ad es. voce (vocale/consonante, cantato/parlato, assolo/effetti corali); percussioni (metallo/legno/membrane, presenza/assenza di risonanza, strumenti accordati/non accordati, eventi individuali/compositivi) – o possono, al contrario, determinare alcune intersezioni o connessioni tra i due campi, in modo da dare, ad esempio, un andamento 'vocale' agli strumenti di percussione, o estrapolare le articolazioni di 'percussione' incluse in una gestura vocale.
ii) Questo primo gruppo di trasformazioni – paragonabile alla fase di 'sradicamento' e di conseguenza diventando una potenziale fonte di tensioni celate (dopo, sempre più esplicite) – è stato portato agli estremi dall'impiego di specifiche tecniche di trattamento elettroacustico (che assumono, in circostanze particolari, un ruolo dominante), cioè avvolgimento e granulazione. Questo processo di elaborazione può condurre alla creazione di un universo variegato, popolato da 'chimeriche' immagini soniche, che in definitiva rivelano, attraverso la loro effimera apparenza nello scenario temporale, non solo tutte le ambiguità percettive e strutturali che derivano dalla loro viva presenza in una specie di 'terra di nessuno', ma mostrano anche, in alcuni casi, un tipo di equilibrio osmotico, interpretabile come realizzazione utopistica di quella fusione intima verso cui una società veramente integrata è diretta.

Chen Chin-Chin
Points of No return (1997)

Points of No Return shifts between two different environments or landscapes. To achieve this, sounds were divided into two categories according to their nature and timbre; but as the piece goes on, some sounds from one environment also occur in the other one. *Points of No Return* employs music concrete techniques and digital editing and processing. Only at a very large stage is electronically generated sound incorporated to color some dramatic moments. *Points of No Return* is divided into 5 sections, alternating between two different landscapes.

Points of No Return si sposta tra due diversi ambienti o paesaggi. Per compiere ciò, i suoni sono divisi in due categorie secondo la loro natura e timbro; ma come il pezzo avanza, alcuni suoni da un ambiente si spostano nell'altro. Points of No Return impiega tecniche concrete di musica, elaborazione e trasformazione digitali. Solo in una fase più ampia viene generato elettronicamente un suono che colora alcuni momenti drammatici. Points of No Return è diviso in 5 sezioni, che si alternano in due diversi paesaggi.

Fabio Cifariello Ciardi
Due piccoli paesaggi (1999)

Two short sonic landscapes invite listeners 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 an operating table of a sewing-machine and an umbrella"?

Due piccoli paesaggi sonici invitano all'esperienza dell'"attraversare".

Spazi possibili? Luoghi invisibili? La narrazione di memorie concrete o l'occultamento surreale in una "foresta di simboli"?.
"....l'incontro casuale su un tavolo anatomico, di una macchina da cucire e di un ombrello"? Il lavoro è stato commissionato dal CRM per il Festival Musica e Scienza 1999.

Riccardo Dapelo
I suoni che distilli (1996)

The work is derived from the poetry of Eugenio Montale "Resta lontano...". The text of the poetry was initially read (only few significant words were sung) by the soprano Daniela Aimale. Like the title "I suoni che distilli" (the sounds that you distill) explains, the attempt of this composition is to extract the essence, the pure sound of poetic language. This was made with different techniques of processing the human voice (spectral and dynamic granulations, vocoder). Together with this processes mimetic and synthetic sounds variously interpolated with voice grains (some original csound instruments for morphing and granular synthesis – see papers enclosed) have been used. To obtain a sensation of movement between foreground/background the piece is developed in different levels (each with its own dynamic grade of reverberation and/or volume) laid one upon the other. The piece is dedicated to Alvise Vidolin.
The piece should be played over at least 4 loudspeakers with spatialization (usually midi-controlled by MAX patches).

Il lavoro è tratto da una poesia di Eugenio Montale "Resta Lontano...". Il testo di questa poesia è stato inizialmente letto (solo poche parole significative venivano cantate) dal soprano Daniela Aimale. Come spiega il titolo "I suoni che distilli", il tentativo di questa composizione è di estrarre l'essenza, il suono puro del linguaggio poetico. Ciò è stato fatto con diverse tecniche di elaborazione della voce umana (granulazioni spettrali e dinamiche, vocoder). Insieme con questo processo sono stati usati suoni mimetici e sintetici diversamente interpolati con granelli di voce (alcuni strumenti originali Csound per la sintesi granulare). Per ottenere una sensazione di movimento tra primo piano e sfondo il pezzo è sviluppato in vari livelli (ognuno con il suo proprio grado di risonanza e/o volume) posti uno dopo l'altro. Il pezzo è dedicato ad Alvise Vidolin.
Il pezzo dovrebbe essere eseguito con almeno 4 altoparlanti con spazializzazione.

Thomas DeLio
Decker (1998)

P. Inman has been described as "a high modernist's high modernist". He is the author of some of the most beautiful, stimulating poetry written over the past twenty years. In his work language is decontextualized to a point where each word and sound becomes concrete and palpable, each line becomes a force reaching out to embrace a multiplicity of meanings, and language itself becomes fluid and unfixed. The present composition "Decker" is based on a text by P. Inman of the same time. It is based on the poet's own reading of his text. I am indebted to P. Inman for his permission to use his poem for the sake of my musical musings and hope that, if nothing else, this sonic rendition will introduce his marvelous poem to a larger audience.
It seems clear that the page has superseded the line as the most important structural unit in the design of much recent poetry. This is certainly true of Inman's work (see *ocher, "waver", "annette", "smaller",* and *ply).* As Inman himself has stated: "The general organizational push to my poetry has become page-specific. I tend to write in pages...". In "Decker" he seems to play with the very notion of what constitutes a page in a written text. The poem is constructed from two types of pages; yet, it also consists of fifteen "virtual" pages – fifteen sections which, within the text itself, are labeled "pages". These fifteen virtual pages tend to make the reader quite conscious of the presence and function of the seven physical pages, creating a counterpoint of different page types, each type imposing a different order on the text. In my setting of "Decker" I basically add a third type of page which I superimpose upon the original two. The musical composition is in seven sections (or "sonic" pages) reflecting the original seven physical pages of the text. Each of these sonic pages consists of transformations of the one or more virtual pages while two of them (the third and the seventh) also contain some very clear, unaltered and identifiable statements of several physical pages.
Certainly this emphasis on the page is only one aspect of Inman' s remarkably rich poetry. It seems to me that in this poetry we become aware of language in two different ways. At times words and phrases seem to move from opaque to transparent – to be caught in the process of taking on referential associations to objects and ideas ("speak in from black knock" or "man immense off clough") and thus caught on the verge of losing their alternate identity as concrete sounds and black lines on a pieces of white paper. At other times Inman's words and phrases seem to move in the opposite direction; they seem to become so opaque that they lose any such associations ("meig crag"). At these times his words seem to achieve the state of pure sound; a bird call, a drum roll, the buzz of a lawnmower (music). In my musical setting I try to heighten the

tension between these opposite states of opaqueness and transparency. This work (a song?) starts with the text and moves in two directions. At times, the music surrounds the sound of text with other, non-vocal sounds: fitting the sounds of the words into the larger world of sound (section four). In these cases the text usually remains clear and apprehensible: its transformations effected more by changes in context (both sonically and structurally) than changes in the sounds of the words. At other times, however, the words themselves are broken up, stretched, and dissolved electronically to such an extent that their purely sonic attributes are enhanced, while their function as elements of language is lost. Words become pure sounds, unrecognizable as elements of language, and the text – both its words and structure – is lost.

P. Inman è l'autore di alcune delle più belle e stimolanti poesie scritte negli ultimi vent'anni. Nei suoi lavori il linguaggio è decontestualizzato ad un punto in cui ogni parola e suono diventa concreto e palpabile, ogni riga diventa una forza che si estende per abbracciare una molteplicità di significati, ed il linguaggio stesso diventa fluido e stabile. La composizione "Decker" è basata su un testo di P. Inman che ha lo stesso titolo. È basata sulla lettura di questo testo da parte del poeta.
È evidente che la pagina ha soppiantato la riga come l'unità strutturale più importante in molta della poesia recente. Ciò è vero nell'operato di Inman. In Decker egli sembra giocare con la nozione di ciò che costituisce una pagina in un testo scritto. L'opera è costruita per due tipi di pagina: pubblicata, stampata su sette fogli di carta – le sue sette pagine "fisiche"; ancora, consiste in quindici pagine "virtuali" – quindici sezioni che, nel testo stesso, sono qualificate "pagine". Queste quindici pagine virtuali tendono a rendere il lettore completamente cosciente della presenza e della funzione delle sette pagine fisiche, creando un contrappunto di vari tipi di pagine, ognuno di essi imponendo un ordine diverso nel testo. Nella messa in opera di Decker da parte di DeLio, questi ha aggiunto un terzo tipo di pagina che ha sovrapposto alle altre due originali. La composizione musicale è in sette sezioni (o pagine "soniche") che riflettono le sette originali del testo. Ognuna di queste pagine soniche consiste in trasformazioni di una o più pagine virtuali mentre due di esse (la terza e la settima) contengono anche alcune chiare, inalterate e identificabili affermazioni delle varie pagine fisiche.
Sicuramente questa enfasi sulla pagina è solo un aspetto della ricca poesia di Inman. Sembra che in questa poesia si diventa consapevoli del linguaggio in due modi diversi. A volte parole e frasi sembrano muoversi dall'opaco al trasparente – per essere catturate nel processo di assunzione di associazioni referenziali per oggetti e idee, e ciò porta al limite della perdita della loro identità alternata come suoni concreti e linee nere su un pezzo di carta bianca. Altre volte le parole e le frasi di Inman sembrano muoversi nella direzione opposta; sembrano diventare così opache da perdere qualunque delle associazioni. In questi casi le sue parole sembrano raggiungere lo stato di suono puro; un richiamo di uccello, il rullio dei tamburi, il ronzio di una falciatrice. DeLio, nella sua messa in opera, ha cercato di intensificare la tensione tra questi stati opposti di opacità e trasparenza. Questo lavoro (una canzone?) inizia con il testo e si muove in due direzioni. A volte, la musica circonda i suoni del testo con altri suoni non-vocalici: adattando i suoni delle parole ad un più ampio mondo di suoni (sezione quattro). In questi casi il testo di solito rimane chiaro e comprensibile. Le sue trasformazioni sono state determinate più dai cambiamenti nel contesto (sia sonico che strutturale) che dai cambiamenti nel suono delle parole.
Altre volte, tuttavia, le parole stesse vengono distrutte, distese e dissolte elettronicamente verso un'estensione in cui i loro attributi puramente sonici vengono accresciuti, mentre la loro funzione come elementi di linguaggio viene persa. Le parole diventano suoni puri, irriconoscibili come elementi di linguaggio, ed il testo – sia le parole che la struttura – è perso.

Francis Dhomont
En Cuerdas (1998)

To Arturo Parra
Commissioned by the Colombian guitarist Arturo Parra, "En Cuerdas" is the acousmatic version – it does not contain the instrumental part that is played life – of a work for guitar and tape co-written by A. Parra and myself. Completely independent from the version for guitar and tape, this piece is to be projected on an orchestra of loudspeakers and is an autonomous work. its sonic environment however remains that of strings that are plucked, rubbed and struck, made virtual and transformed by computer processes and expanded by the use of "electroacoustic writing". Their recognisable shape is combined with sound objects from a variety of sources that may, in some cases, be closely related in timber, pitch, etc., in order to create sonic ambiguities.
"En Cuerdas" is not inspired by an external idea but instead tries to bring together heterogeneous sound objects. During the first part, most of the sound material of the entire work are introduced. These are morphological variations that have been realised through the use of a variety of processing "instruments" (software) found in the real-time digital system SYTER at the INA-GRM, in Paris (France).

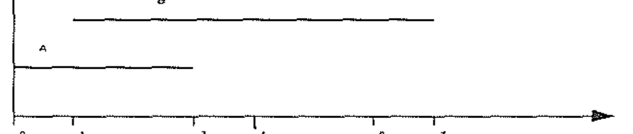
Within this formal context and the preliminary choices made with this software, I have, as always, left myself some room for improvised "jeux-sequences"; I'm referring to my relationship with the machines that I try to treat as musical instruments or resonating bodies that the hand brings to life; playing with parameters in real-time, with the mouse, fader, using gestures instead of programmed sequences, using my hands to control dynamics, etc. Nevertheless, these moments of pure intuition do not carry the same irreparable risks that are found in real improvisations, for the choice of which material to keep will follow the performance, and only those "magical" moments will remain. Half-way between chance and will, this is an attempt to reconcile Pascal's "spirit of fineness and spirit of geometry".

Ad Arturo Parra
Commissionato dal chitarrista Colombiano Arturo Parra, En Cuerdas è la versione acusmatica – non contiene la parte strumentale che è suonata dal vivo – di un pezzo per chitarra e nastro scritto in collaborazione da A. Parra e F. Dhomont. Completamente indipendente dalla versione per chitarra e nastro, questo pezzo viene indirizzato ad un'orchestra di altoparlanti ed è un lavoro autonomo. Tuttavia il suo ambiente sonico rimane quello di corde che vengono pizzicate, stropicciate e colpite, rese virtuali e trasformate dai processi del computer e dilatate dall'uso di "scrittura elettroacustica". La loro forma riconoscibile è combinata con il suono di oggetti che possono, in alcuni casi, essere collegati nel timbro, tono, ecc., in modo da creare ambiguità sonore.
En Cuerdas non è ispirato da un'idea esterna ma prova invece a unire suoni eterogenei di oggetti. Durante la prima parte, viene presentata la maggior parte dell'intero materiale sonoro. Queste sono variazioni morfologiche che sono state realizzate attraverso l'uso di una varietà di "strumenti" di elaborazione (software) trovati nel sistema digitale di tempo-reale SYTER all'INA-GRM, a Parigi, Francia.
In questo contesto formale e nelle scelte preliminari fatte con il software ho lasciato alcuni spazi per "sequenze-di gioco" improvvisate; ciò si riferisce alla relazione di Dhomont con le macchine che egli prova a trattare come strumenti musicali o corpi risonanti che le mani portano in vita; suonando con parametri in tempo-reale, con il mouse, usando i gesti invece di sequenze programmate, usando le mani per controllare le dinamiche ecc.
Tuttavia, questi momenti di pura intuizione non comportano gli stessi irreparabili rischi che si trovano nella improvvisazione reale, e solo quei "magici" momenti rimarranno. A metà via tra il caso e la volontà, questo è un tentativo di riconciliare lo "spirito di raffinatezza e lo spirito di geometria" di Pascal.

Roberto Doati
Forma di nebbia (1995)

the past must be Invented
the future Must be
revIsed
doing boTh
mAkes
whaT
the present Is
discOvery
Never stops
John Cage

The strong and geometric form used comes from the chosen poem metre here represented:



The above structure is presented 15 times - one for each strophe. Each time superposing to the proceeding one which is time stretched. So the first presentation starts at the beginning and lasts until the end, the second one starts at 70" and lasts to the end too, and so on every 70".
I covered this form with mist to break the poem tale. The purpose is to experience difficulty in understanding the form built up by 15 different ancient vocal styles. One style for each strophe, from the Jewish Salmody to Andrea and Giovanni Gabrieli. Most of the models used come from the Western sacred vocality, above all

from the Venetian one. Not by chance, because the Biennale di Venezia music festival - that commissioned the work - occurred within the Basilica di San Marco Dedication Nineth Century Celebration and was titled "L'ora di là dal tempo. Moments of spirituality in contemporary music".
There are nevertheless other vocal or instrumental allusions whose choice is due sometimes to the meaning and character of the verse, sometimes to recall friends. So the Gregorian Chant is represented by the *sequenza* "Victimae Paschali Laudes" because the fact taled by von Droste-Hülshoff happens during Eastern night, and the *ballata* by Ciconia - born in Liège in 1340 and dead in Padova, 1411 - is used to allude to my friendship with Marianne Pousseur.
The computer is used basically to multiply Marianne voice. Ideally to increase it as many times as the woman in the poem crosses her own image, to achieve an unpleasant state for the singer to hear her own voice unceasingly rising in hugely magnified details.
Sounds and noises ideally coming from the periods evoked by the different styles (harp, fiddle, lute, organ, strings, brass, market noises, ...) trickle into the mist of this vocal polyphony.
This electroacoustic version is preferable performed through the 8 tracks tape. It allows a quite large range of interpretation as concerns the mixing of its complex compound of elements.
Forma di nebbia is the third part of the cycle for female voice and electronics "L'olio con cui si condiscono le parole", a chamber theatre musical work commissioned by La Biennale di Venezia 1995, never produced for the scene.

La forma è quella geometrica e forte, derivata dal metro della poesia utilizzata.
Questa struttura viene proposta 15 volte (una per ogni strofa), ogni volta sovrapponendosi alla precedente dilatata nel tempo. Così la prima struttura parte all'inizio del pezzo e dura fino alla fine, la seconda comincia dopo 70" e dura anch'essa fino alla fine, e così via ogni 70".
La nebbia è quella che provoca interruzioni nello svolgersi della narrazione poetica. Impedisce di riconoscere chiaramente la forma cui danno vita i 15 stili vocali (uno per ogni strofa, in ordine cronologico, dalla salmodia ebraica ai Gabrieli) usati per la composizione.
La maggior parte dei modelli storici utilizzati riguarda la vocalità sacra occidentale, con numerosi riferimenti a quella veneziana. Non a caso, poiché il festival della Biennale di Venezia che ha commissionato il lavoro si svolgeva all'interno del Nono Centenario della Dedicazione della Basilica di San Marco e portava il titolo "L'ora al di là del tempo. Momenti di spiritualità nella musica contemporanea". Ma vi sono altre allusioni a musiche, vocali e non, la cui scelta è stata determinata talvolta dal carattere e del significato della poesia (ad esempio, il canto Gregoriano è rappresentato dalla sequenza "Victimae Paschali Laudes" perché la vicenda, narrata da von Droste-Hülshoff, si svolge durante la notte di Pasqua), talvolta da richiami a persone amiche (ad esempio, l'uso di una ballata di Ciconia, autore del '300 nato a Liegi e morto a Padova, allude al suo legame con Marianne Pousseur).
La funzione svolta dal computer è fondamentalmente quella di moltiplicare la voce di Marianne Pousseur idealmente per quante volte la protagonista della poesia incrocia la propria immagine, fino a rendere quasi un incubo per la cantante, il continuo emergere della propria voce con particolari enormemente amplificati.
Tra le maglie di questa polifonia vocale si insinua il trattamento di suoni legati ai diversi modelli storici: arpa, viola, percussioni, liuto, organo, archi, ottoni, rumore e voci di mercato, artigiani...

Denis Dufour
Terra Incognita (1998)

Danced in four stages (taking up again the four stages of our development: discovery, exploration, conquest and servitude) *Terra incognita* unfolds (treading carefully, little certain of the adventure, alternately rushing into rapids, canyons and dead-end streets and sometimes leading up a pleasant valley). As though coming from an orbiting library, some majestic echos come unstrung "on your lips..." (1) and its quintessential coughing fit. Running through the uninhabitable spaces of our cities and minds, we follow the tracks of a primitive man, projected in the future, like all of us. Only instinct guides us here, provided that we have not shut it away in our linguistic compartment of our brain. In short, the gesture speaks.
As "it is, we believe, unique that a radical revision of generally accepted ideas has ever been embarked upon with so light a heart", one need, as monsters know, all the resources of ingenuousness, all the energy of innocence in order to get moving again in these worrying and seductive worlds. Once more taking up Descartes' asceticism, as Schaeffer suggested in the preface of his *Traité*: "to strip oneself of all received opinions and beliefs and to begin again with completely new foundations".
1) "sur tes lèvres...", one of Schaeffer's fragments of "found" sentences, used in one of his earliest compositions of *concrete* music *Etude pathétique* (*Etudes de bruit*, 1948).Tom Aconito (1998)

"Contingent and fundamental" – it's all there and there lies the difficulty: in this double nature of acousmatic music, which is at once angelic and terrestrial. "Our kingdom is not of this world" musicians were in the habit of thinking, until our and Schaeffer's era. Schaeffer quotes this remark from Hoffmann at the beginning of his

Solfège des Objets Sonores – which inaugurates an era where music, on the contrary, takes up the option of including all sorts of things... Schaeffer is the first to take offence at this saucepan that he has hung up at the tail of the music... He attempts to make it disappear by substituting – in the discourse – a descriptive abstract with worldly references; and by recommending the avoidance of sounds which – in order to make music – he calls "anecdotic"... Waste of time and effort, or almost. His neat *Etude aux Casseroles* – which happens to be quite sublime – has opened Pandora's box: what will appeal is just to make *concrete* music to the ultimate. A paradoxal music which vaporizes – undoubtedly by the mystery of the sense of hearing, at once disembodied and carnal – the most real world.
Jean-Christophe Thomas (1999)

Divisa in quattro fasi (considerando le quattro fasi del nostro sviluppo: scoperta, esplorazione, conquista e servitù) Terra Incognita si sviluppa (trattando attentamente piccole avventure, precipitando alternativamente nelle rapide, nei canyon e in strade senza fine che talvolta conducono a piacevoli valli). Come se provenissero da una biblioteca in orbita, alcune eco maestose arrivano senza corda "sulle tue labbra" e il loro quintessenziale attacco di tosse. Attraversando gli spazi non abitabili delle nostre città e delle nostre menti, seguiamo le tracce di un uomo primitivo, proiettato nel futuro, come tutti noi. Solo l'istinto ci guida qui, a condizione che non lo segreghiamo nel compartimento linguistico del nostro cervello. In breve, parlano i gesti. Poiché "crediamo sia straordinario che, una radicale revisione delle idee che sono generalmente accettate, non sia mai stata intrapresa in modo così spensierato", si ha necessità, come i mostri sanno, di tutte le risorse dell'ingegnosità , di tutta l'energia dell'innocenza in modo da ridare movimento a questi tormentati e seducenti mondi. Riprendendo l'ascetismo di Descartes, come Schaeffer suggeriva nella prefazione al suo Traité, "per spogliarsi di tutte le opinioni e credenze ricevute e per iniziare da capo con basi completamente nuove".
1) "sur tes lèvres...", uno dei frammenti delle frasi "forgiate" da Schaeffer, utilizzate in una delle sue prime composizioni di musica concreta Étude pathétique (Études de bruit, 1948).
Tom Aconito (1998)

"Contingente e fondamentale" – è tutto qui e qui sta la difficoltà: nella doppia natura della musica acusmatica, che è contemporaneamente angelica e terrestre. " Il nostro regno non è di questo mondo", così i musicisti erano abituati a pensare, fino all'era nostra e di Schaeffer. Schaeffer cita questa osservazione da Hoffmann all'inizio del suo Solfège des Objets Sonores – che inaugura un'era in cui la musica, al contrario, si prende la libertà d'includere ogni tipo di cosa...Schaeffer è il primo ad offendersi per questo calderone che "ha attaccato alla coda della musica"...Cerca di farlo sparire sostituendo - nel discorso – un riassunto descrittivo con riferimenti temporali; e consigliando lo sfuggire dei suoni che – per fare musica - chiama "aneddotico"...spreco di tempo e di forze, o quasi. Il suo accurato "Étude aux Casseroles" – che si dà il caso sia quasi sublime – ha aperto il vaso di Pandora: creare musica concreta all'estremo. Una musica paradossale che fa evaporare - indubbiamente tramite il mistero del senso dell'udito, disincarnato e carnale allo stesso tempo – il mondo più reale.
Jean- Christophe Thomas

Michael Edward Edgerton
Wassermann (1999)

The piece is an attempt to present a warmth of spirit within an explorative expression.
Technically, the piece asked of its environment to do things it was not designed to do, such that I attempted to manipulate source material in such a way as to often disguise its original property if it would lend to the overreaching formal and expressive properties imagined.
Emotionally, the piece was to pay homage to the gifts of the featured voice. The homage was designed as a vehicle in which her instruments would prompt technology into dramatic, crucial and vivid responses. In this way the work suggests certain relationships between our increasingly controlling "information age" and the citizens who willingly are being blindly led into a corner of decreased control and freedom.

Il pezzo costituisce un tentativo di presentare un ardore di spirito in un'espressione esplorativa.
Tecnicamente, il pezzo chiede al suo ambiente di fare cose per cui non è designato, tanto che io ho tentato di manipolare del materiale originale in modo tale da camuffare spesso le sue proprietà se si fosse adeguato alle proprietà formali ed espressive immaginate.
Emozionalmente, il pezzo doveva rendere omaggio alle doti della voce narrante. L'omaggio era progettato come un veicolo in cui i suoi strumenti avrebbero indotto la tecnologia a risposte drammatiche, cruciali e vivide. In tal modo il lavoro suggerisce alcune relazioni tra la nostra crescente "età dell'informazione" che si controlla e i cittadini che spontaneamente sono condotti ciecamente ad un angolo di libertà e controllo ridotti.

Frank Ekeberg
Intra (1999)

The title suggests something from within, introspection. *Intra* deals with psychological solitude and the relationships between external and internal worlds. The inspiration for this work came as a result of my interest in psychopathology, particularly in conditions which manifest themselves in an inability to differentiate between self and other. I use space and spatial cues as central tools for portraying such a state. Environments of internal and external nature are set up and appear both separately and juxtaposed throughout the work. Transitions between different virtual spaces are guided by the development of the sound material, and take place sometimes gradually, other times abruptly. As with the state of the mind, one cannot always anticipate what happens next.
Intra was composed during the summer and autumn of 1999 in the composer's private studio and in the studios of the Center for Electroacoustic Music Studies at City University in London.

Il titolo suggerisce qualcosa che viene dall'interno, introspezione. Intra ha a che fare con la solitudine psicologica e le relazioni tra il mondo esterno e quello interno. L'ispirazione per questo lavoro scaturisce dall'interesse dell'autore per la psicopatologia, in particolare in condizioni che si manifestano nell' incapacità di differenziare se stessi dagli altri. Uso spazio e suggerimenti spaziali come strumenti principali per ritrarre tale stato. Le condizioni di natura interna ed esterna appaiono sia separatamente che giustapposte nel lavoro. I mutamenti tra differenti spazi virtuali sono guidati dallo sviluppo del materiale sonoro, e si verificano in modo graduale, altre volte improvvisamente. Come per lo stato della mente, non si può sempre anticipare ciò che accadrà in seguito.
Intra fu composta tra l'estate e l'autunno del 1999 nello studio privato del compositore e negli studi del Centre for Electroacoustic Music Studies nella City University di Londra.

Beatriz Ferreira
Rio de los pàjaros azules (1999)

This piece is the third movement of a triptych called "Rios del Sueño" (the dream's rivers). This composition is a pale and vague transposition of signs and tracks from a very intense musical sensation, imperfectly seen, one night long ago, in a tropical latino-american dream.

Questo pezzo rappresenta il terzo movimento di un trittico intitolato "I fiumi del sogno". Questa composizione è una pallida e vaga trasposizione di segni e tracce che provengono da una sensazione musicale molto intensa, sentita in modo imperfetto, una notte di molto tempo fa, in un sogno tropicale latino-americano.

Lawrence Fritts
Doctrine of Chances (1999)

Doctrine of Chances, the title of which is taken from an obscure 18th century mathematical treatise on probability, re-examines the relation between statistical distribution and form in music in the latter part of the twentieth century. The so-called chance composers of the fifties and sixties, represented by Cage, Wolff, Brown, and others, used probabilistic procedures to create sound worlds that were free from what they would regard as doctrinaire approaches to formal organization, as represented by Babbitt and the east coast academic composers. From our vantage point at the end of the century, these differences between these two approaches appear to be not so great as once imagined. Serialists, it has often been argued, sometimes too casually applied higher mathematical processes without a clear understanding of their structural implications in order to create expressive patterns of richness and complexity. At the same time, chance composers came to incorporate increasingly elaborate sets of rules and conditions governing how chance operations would be employed in their music. Both approaches to creating pattern and complexity are integrated in *Doctrine of Chances*, which combines the structural clarity of digitally-generated sound with the richness of digitally-processed musical instruments.

Technical realization
Two types of sounds are used in this piece: synthesized and processed. The synthesized sounds were algorithmically generated by a Kyma system and C-Sound on SGI and Machintosh computers. The processed sounds originated as recordings I had made of a dozen wind and brass instruments in the University of Iowa Anechoic Chamber. These sounds were mixed, cross-faded, convolved, cross-synthesized, and morphed using a Kima system, C-Sound, and other Machintosh and SGI applications. Certain features of these sounds, such as transients and harmonic spectra, were extracted, processed, then

placed into a stereo mix with other components of the sounds. The final work was edited and mixed in Pro Tools.

Doctrine of Chances, il cui titolo è preso da un vago trattato matematico sulle probabilità del 18° secolo, riesamina la relazione tra la distribuzione statistica e la forma nella musica nell'ultima parte del 20° secolo. I cosiddetti compositori delle probabilità degli anni cinquanta e sessanta, rappresentati da Cage, Wolff, Brown ed altri, utilizzarono delle procedure probabilistiche per creare mondi del suono che erano liberi da ciò che essi consideravano approcci dottrinari all'organizzazione formale, così come rappresentati da Babbitt e dai compositori accademici della costa est. Alla fine del secolo, le differenze tra questi due approcci apparvero non essere così grandi come si immaginava. I compositori di musica dodecafonica, è stato spesso sostenuto, a volte troppo casualmente applicavano processi matematici più alti senza una chiara comprensione delle loro implicazioni strutturali in modo da creare modelli espressivi di ricchezza e complessità. Allo stesso tempo, i compositori delle probabilità incorporavano in modo crescente un gruppo di regole e condizioni che riguardavano il modo in cui le operazioni del caso sarebbero state impiegate nella loro musica. Entrambi gli approcci per creare modelli e complessità sono integrati in Doctrine of Chances, che combina la chiarezza del suono digitale con la ricchezza degli strumenti musicali elaborati dal digitale.

Realizzazione tecnica
Due tipi di suono vengono utilizzati in questo pezzo: sintetizzati ed elaborati. I suoni sintetizzati sono stati generati in modo algoritmico da un sistema Kyma e computer C-Sound su SGI e Machintosh. I suoni elaborati sono stati mixati, avvolti e sintetizzati usando un sistema Kyma, C-Sound, e altre applicazioni di Machintosh e SGI. Alcuni di questi suoni sono stati estratti, elaborati, e posti in uno stereo con altri componenti del suono. Il lavoro finale è stato edito e mixato in Pro Tools.

Diego Garro
Six dreaming jewels (1999)

Bright sounds, crisp resonance, sharp sonorities. Meta-materials as honed as diamonds; glowing like crystals. The *Dreaming Jewels*, in Theodore Sturgeon's novel, have the magic power of synthesizing and duplicating living beings as by-products of their mysterious vagaries and nightmares. They are material. They are alive. They feel pain and go their own inscrutable way, leaving behind completed men and animals, along with unfinished jobs, maimed cats, blind mutants, freaks that society will always despise.
I decided to compose six sketches of tape music while reading this fairytale. I was struck by the sort of 'mysticism of things' which does not need a supernatural creator to transcend reality and explore what lies beyond it. Each object can be a riddle. Every instant can carry an enigma. Each being is a puzzle, for itself and for its fellow beings. And every sound is a conundrum when we sculpt it with our own hands...
Each of these abstracts miniatures can be considered as self-contained electroacoustic chiseling. They can be played separately or in a different order, even though they work together as part of a whole, punctuated by cross-references, glimpses of call and echoes of response.

Technical Notes:
As in most pieces of mine, the materials were sculpted using different techniques: computer-synthesis, FM timbre-design, processed voice and 'close-microphoned' events taking place in the recording room.
A rig comprising MIDI-sequencing software running on a PC driving a set of MIDI-synthesizers and samplers was used to create some motives, subsequently processed and re-arranged.
The sound processing of most materials was carried using the 'Composition-Development-Performance' (CDP), a command-line bundle of music-software tools developed as an international cooperative effort by a group of sound artists based in the UK. Further timbre design, along with the final montage, were realized using a ProTools® system running on a Machintosh platform.

Suoni brillanti, vivace risonanza, sonorità acute. Semi-materiali levigati come diamanti, splendenti come cristalli. I Gioielli dei sogni, nel romanzo di Theodore Sturgeon, hanno il potere magico di sintetizzare e duplicare gli esseri viventi come se fossero prodotti dalle loro misteriose stravaganze e incubi. Sono materiali. Sono vivi. Sentono il dolore e vanno per la loro imperscrutabile via, lasciando dietro uomini e animali, procedendo con lavori incompiuti, gatti mutilati, mutanti ciechi, bizzarrie che la società disprezzerà sempre. Ho deciso di comporre sei pezzi di musica per nastro mentre leggeva questo racconto. Sono stato colpito dal tipo di 'misticismo delle cose' che non hanno bisogno di un creatore soprannaturale per trascendere la realtà ed esplorare ciò che c'è dietro. Ogni oggetto può essere un mistero. Ogni istante può essere un enigma. Ogni essere è un enigma, per se stesso e per i suoi compagni. Ed ogni suono è un enigma quando lo scolpiamo con le nostre mani...

Ognuno di queste miniature astratte può essere considerata come cesellatura elettroacustica autonoma. Esse possono essere suonate separatamente o in ordine diverso, anche se operano insieme come parti di un insieme, punteggiate da riferimenti incrociati, tracce di chiamata ed echi di risposta.

Note tecniche
Come in molti pezzi dell'autore, i materiali sono stati scolpiti utilizzando varie tecniche: sintesi col computer, progettazione del timbro FM, voce elaborata ed avvenimenti a 'microfono chiuso' che avvengono nella stanza di registrazione.
Un'attrezzatura comprendente il software MIDI-ordinamento di sequenza che guida un gruppo di sintetizzatori MIDI e campionatori è stata usata per creare alcuni motivi, in seguito elaborati e riarrangiati.

Paul Goodman
Mirror Images (1999)

The first step is always to create or find a basic set of sound material. From 1981 to 1989 I made use of a sound synthesis program called MIDIM/VOSIM developed by Mr. Werner Kaegi at the Institute of Sonology in Utrecht, the Netherlands. In these years a very strict aesthetic was adhered to, namely, that all sound material had to originate by means of sound synthesis, that is built solely within the boundaries of the MIDIM/VOSIM system. A basic set of material was operated upon to create families of related sounds and then structured. Since 1989 it is, for various reasons, no longer possible to work with MIDIM/VOSIM therefore I have opted for an eclectic approach. Now sound material comes from anywhere, regardless of its origins as long as it is expressive and its quality is acceptable. Sound Synthesis is still used but not to the exclusion of other means of sound production or reproduction. Whereas previously one worked with synthetic sounds which had to be made to sound non-artificial it is now necessary when working with a 'concrete' sound to make it sound not too mundane. The challenge was formerly to bring sounds into the world while at the present it is to take them out of the world so to speak. From the beginning the term 'associative' was applied to describe the music. The sounds were conceived of as 'magnetic fields' (in the sense in which André Breton and Philippe Soupault used it) which held aural and visual associations. The sound fields overlap each other and therefore by operating upon the one we could lead it in the direction of other sound fields creating associative complexes. For example if I take the sound of a cat as my basic material and apply a series of operations to it I not only transform the sound itself but the associations (aural and visual) called up in the listener. The sounds also serve as image-generators. The tape entitled 'Mirror Images' was created for an open project at the Bourges Festival 1999 and was given its premiere there.

Il primo passo è sempre quello di creare o trovare un gruppo fondamentale di materiale sonoro. Dal 1981 al 1989 l'autore ha utilizzato un programma di sintesi del suono chiamato MIDIM/VOSIM sviluppato da Mr. Werner Kaegi all'Institute of Sonology ad Utrecht, Paesi Bassi. In questi anni un'esatta estetica consisteva nel sostenere tutto quel materiale sonoro che si doveva originare dai mezzi di sintesi sonora, che viene creata esclusivamente nei limiti del sistema MIDIM/VOSIM. Un gruppo fondamentale di materiale ha influito a creare famiglie di suoni collegati e strutturati. Dal 1989, per vari motivi, non è più possibile lavorare con il MIDIM/VOSIM e per questo ho optato per un approccio elettrico. Ora il materiale sonoro proviene da qualunque parte, senza cura della sua origine per tutto il tempo che è espressiva e la sua qualità tecnica accettabile. La Sintesi del Suono è ancora utilizzata ma non fino all'esclusione di altri mezzi di produzione o riproduzione del suono. Laddove in precedenza si lavorava con suoni sintetizzati che dovevano essere suonati in modo non-artificiale, è ora necessario, mentre si lavora con un suono 'concreto', farlo suonare in modo non troppo comune. La sfida un tempo era di far entrare il suono nel mondo mentre oggi è di cacciarlo dal mondo per farlo parlare. Dall'inizio il termine 'associativo' era usato per descrivere la musica. I suoni erano concepiti come 'campi magnetici' (nel senso in cui André Breton e Philippe Soupault lo usavano) che contengono rapporti auricolari e visivi. I campi del suono, sovrapposti l'uno all'altro, influiscono su di uno che noi possiamo guidare nella direzione degli altri campi del suono creando complessi associativi. Ad esempio, se si prende il verso di un gatto come materiale fondamentale e si applica ad esso una serie di operazioni non solo si trasforma il suono stesso, ma i rapporti (auricolare e visivo) richiamano l'ascoltatore. I suoni servono anche come generatori di immagine. Il nastro intitolato Mirror Images è stato creato per un progetto del Bourges Festival 1999 e lì fu eseguito per la prima volta.

Michael Hamman
Discourse/Current (2000)

This piece concerns a framework for making time through the deployment off minimally articulated acoustical events. These events are structured around four elemental materials: spoken voice, sine tone, broadband

noise, and silence. Three short excerpts are taken from a recorded performance by poet Chris Mann of his poems *in representing.* These excerpts are processed through filtration, compression, and very slight stretching, so that only very brief fragments – in some cases, little more than *erasures* – are left.
These fragments are interleaved with other material elements according to matrix operations through which minimal musical materials are shaped. The combination of elemental materials is executed through simple logical operations: logical-and (selection by like elements only); logical-or (overlapping of elements); and logical-xor (cancellation of one element by another).
The resulting musical fragments are presented over the course of a series of ten segments. Each segment in the series presents a barely structured unfolding; each is announced by the sounding of a single DTMF signal. At a higher temporal level, there are four longer partitions. Each such partition projects a slightly directed movement from broadband to narrowband sounds.
As is the case on various structural levels within the piece, most sounds are characteristically flat in structure: flat envelopes, extremely low amplitude (some sounds bordering on the barely audible), sudden onsets and cut-offs, spectrally simple sounds with narrow bandwidth, and long silences. The intent is for an overall musical coherence that articulates simple surfaces rather than fully developed morphologies.

Questo pezzo riguarda una composizione di marcatura del tempo attraverso lo spiegamento di eventi acustici minimamente articolati. Tali eventi sono strutturati intorno a quattro elementi materiali: voce parlata, tono sinusoidale, rumore a banda larga, e silenzio. Tre brevi estratti sono presi da un'esecuzione registrata del poeta Chris Mann della sua composizione In representing. Questi estratti sono elaborati attraverso filtrazione, compressione, e stiramento molto leggero, cosicché solo dei brevi frammenti – in alcuni casi, poco più che cancellature – rimangono.
Questi frammenti sono interposti con altri elementi materiali secondo operazioni matrici attraverso le quali i minimi materiali musicali vengono adattati. La combinazione di materiali fondamentali viene eseguita attraverso semplici operazioni logiche: logica-e (selezione elementi solo); logica-o (sovrapposizione di elementi); e logica-xor (cancellazione di un elemento da parte di un altro).
I frammenti musicali risultanti sono presentati attraverso il corso di una serie di dieci segmenti. Ogni segmento nella serie presenta uno spiegamento appena strutturato; ognuno viene annunciato dal suono di un singolo segnale DTMF. Ad un livello temporale più alto, ci sono quattro partizioni più lunghe. Ognuna di queste partizioni progetta un movimento lievemente diretto dai suoni della banda larga a quelli della banda stretta. Come nel caso dei vari livelli strutturali nel pezzo, la maggior parte dei suoni sono caratteristicamente piatti nella struttura: involucri piatti, estensione estremamente bassa (alcuni suoni che rasentano a malapena l'udibile), immediati attacchi ed esclusioni, semplici suoni con strette larghezze di banda, e lunghi silenzi. L'intento è per una totale coerenza musicale che articola semplici superfici piuttosto che morfologie pienamente sviluppate.

Hideko Kawamoto
Night Ascends from the Ear like a Butterfly (1999)

Night ascends from the Ear like a Butterfly, composed in 1999 and dedicated to my grandmother, Tami, was inspired from Haruo Shibuya's poem, *Coliseum in the Desert*. The words Shibuya uses in this poem such as 'night', 'a time of music', 'rain', 'black fountain', 'piano string', 'useless choir' and 'butterfly' gave me compositional ideas. These images were developed in my imagination separately from Shibuya's poem, and they were transformed into music. To me it is very interesting that once one finishes a piece, it leaves the creator, and it grows inside somebody on its own, maybe not as same as the creator's mind. The piece has its own life. I hope my piece has left me...
These are two types of visual images; visual images do not possess the sounds and others do possess the sounds. For instance, 'night', 'butterfly' and 'a time of music' belong to the former type, and 'rain', 'black fountain', 'useless choir' and 'piano string' belong to the latter type. To realize the sound of which the visual image does not possess was challenging during the compositional process. The intention of the butterfly sound is to depict the surrealistic vision of a butterfly flying away from the ear. To me the sound had to be shimmering and transparent. To create the butterfly sound, a tremolo passage from Maurice Ravel's piano piece *Noctuelles* (Night Moths) from *Miroirs* was sampled and processed in the computer using variations techniques including filtering, reververation and pitch shift. On the contrary to the pitched tremolo sound, I also used the sound of small pieces of aluminum foil shaking up and down in a metallic bowl, which is non-pitched, to create the surrealistic vision of butterfly staying at one place, not flying, but moving its wings delicately as it breathes.

Night ascends from the Ear like a Butterfly è stata composta nel 1999 ed è stata ispirata dal componimento di Haruo Shibuya Coliseum in the Desert. Le parole che usa Shibuya, come 'notte', 'un tempo di musica' 'pioggia', 'fontana nera, 'corda del piano', 'coro inutile' e 'farfalla' hanno dato all'autrice le idee per la composizione. Ci sono due tipi di immagini visive: quelle che hanno il suono e quelle che non lo hanno. Ad *esempio, 'notte', 'farfalla', e 'tempo di musica' appartengono al secondo tipo, e 'pioggia', 'fontana nera', 'coro inutile' e 'corde del piano' appartengono al primo gruppo. Realizzare il suono di quelle immagini che non lo hanno è stato suggerito durante il processo di composizione. Lo scopo del suono della farfalla è di rappresentare la visione surrealistica di una farfalla che vola via dall'orecchio. Il suono deve essere brillante e trasparente. Per creare il suono della farfalla, un tremulo passaggio dal pezzo per piano di Maurice Ravel, Noctuelles (Farfalla notturna) da Miroirs, è stato campionato ed elaborato al computer utilizzando varie tecniche come filtrazione e mutamento di tono. Al contrario, al suono tremolo, ho utilizzato anche il suono di piccoli pezzi di fogli di alluminio agitati dentro un contenitore di metallo, per creare la visione surrealistica di una farfalla che sta in un posto, non vola, ma muove le sue ali delicatamente come se respirasse.*

Elsa Justel
Pieza en forma de tè (1999)

The technical resources used in the composition intend to observe, by means of spectral analysis, the characteristics of natural sounds in order to resynthesize them and to exploit their better qualities.
On the other hand, the title of the piece refers to its form. In other words, the form of the tea...but that is a long story...

Le tecniche utilizzate nella composizione intendono osservare, per mezzo di un'analisi spettrale, le caratteristiche di suoni naturali in modo da sintetizzarli e sfruttare le loro qualità migliori.
Dall'altro lato, il titolo del pezzo si riferisce alla loro forma, cioè la forma del tè...ma questa è una lunga storia....

Lillios Elainie
Arturo (1998)

Arturo was written based on an interview with a tarot card reader living in Denton, Texas. Arturo requests visitors to "...please let me answer the questions before you ask them", and claims that most times, he does. After year of palmistry and tarot readings, Arturo has interacted with many people and has learned any life lessons. This piece reflects some of his views on life and the casting of cards to reveal future possibilities.
Arturo was composed in Electroacoustic Studios at The University of Birmingham, England. Recordings of Arturo were engineered by Michael Thompson at the CEMI (Center for Experimental Music and Intermedia) studios at The University of North Texas, Denton, Texas, in August 1997.

Technical Description:
Arturo is a composition for tape alone, which combines acousmatic principles of gesture and development with documentary elements of text and interview. Material for the piece was derived from conversations with Arturo, a tarot card reader, along with samples of the environment in which he reads palms and tarot cards. My goal was to present a documentary-style view of Arturo and his philosophies along with an electroacoustic environment, using his stories and observations as catalysts for acousmatic gestures. Thus the text and voice present information, yet assume the additional role of serving as abstract gestural elements and developmental material.

Arturo è stato scritto sulla base di un'intervista con un lettore di tarocchi che vive a Denton, in Texas. Arturo richiede ai visitatori "........per favore lasciatemi rispondere alle domande prima che io ve le faccia". Dopo anni di lettura di mani e di tarocchi, Arturo ha interagito con molte persone ed ha appreso molte lezioni di vita. Questo pezzo riflette alcune delle sue vedute sulla vita e sulla disposizione delle carte per rivelare possibilità future.
Arturo è stato composto negli Electroacoustic Studios alla University of Birmingham, Inghilterra. Le registrazioni di Arturo sono state preparate da Michael Thompson agli studi CEMI (Center for Experimental Music and Intermedia) alla University of North Texas, Denton, Texas, nell'agosto 1997.

Descrizione tecnica:
Arturo è una composizione per solo nastro, che combina i principi acusmatici di espressioni gestuali e di sviluppo con elementi documentari di testo ed intervista. Il materiale per questo pezzo deriva dalle conversazioni con Arturo, un lettore di tarocchi, insieme con esemplari dell'ambiente in cui egli legge le mani e i tarocchi. Lo scopo dell'autrice era di presentare una veduta documentaria di Arturo e le sue filosofie insieme in un ambiente elettroacustico, utilizzando le sue storie ed osservazioni come catalizzatori per gesture acusmatiche. Così il testo e la voce presentano informazioni, assumendo perfino il ruolo supplementare di servire come elementi astratti gestuali e di sviluppo materiale.

Mitani Norikazu
Piece for Mukkuli (2000)

"Mukkuli" is one of the Japanese traditional instruments. I recorded one sampling sound from mukkuli. And I try to rebuild sampling sound by some digital signal processing language.
I take some processing parameter from database of tidle. I choice some interesting movement of tidle from over 13,000 record by perl language.

"Mukkuli" è uno strumento tradizionale giapponese. L'autore ha registrato una campionatura del suono proveniente dal mukkuli, ed ha provato a ricostruirla con alcuni segnali digitali di elaborazione del linguaggio. L'autore ha preso alcuni parametri di elaborazione da un database di tidle. Ha scelto alcuni movimenti interessanti del tidle da oltre 13.000 registrati dal linguaggio perl.

Francesco Scagliola
Clause mates. El mar del tempo (2000)

The project "Clause Mates. El Mar del Tiempo" is contemporary to another I'm thinking of doing for a long time: writing a Requiem. A Requiem can't be written on commission, neither one's nor someone else's commission, because writing it has nothing to do with musical composition; you write a Requiem because you miss someone or something. I think everybody has his own Requiem to tell. "Clause Mates" isn't the piece I would have written, but it's born in that context.

Il progetto "Clause Mates. El Mar del Tiempo" è contemporaneo ad un altro sul quale intendo lavorare: scrivere un requiem. Un requiem non può essere scritto su commissione, poiché lo scrivere non ha niente a che fare con la composizione musicale; si può scrivere un requiem perché si perde qualcuno o qualcosa. Penso che ognuno ha un suo proprio requiem da dire. "Clause Mates" non è il pezzo che Scagliola avrebbe voluto scrivere, ma è nato in questo contesto.

Michael Thompson
Miniatures (2000)

Miniatures is a piece for solo stereo tape (CD). The piece is comprised of smaller pieces intended for performance as a complete work. *Miniatures* was composed in the composers home studio, the University of Birmingham in Birmingham England and the University of North Texas Center for Experimental Music and Intermedia in Denton, Texas.

Technical details:
Miniatures can be performed on a basic stereo sound system and it can also be performed over a sound diffusion system with 8 or more speakers. The only technical needs are a sound system and a CD player.

Miniatures è un pezzo per nastro stereo solo (CD). Esso è costituito da pezzi più piccoli da eseguire come un lavoro completo. Miniatures è stato composto negli home studio della University of Birmingham a Birmingham, Inghilterra, e nel Center for Experimental Music and Intermedia della University of North Texas a Denton, Texas.

Dettagli tecnici:
Miniatures può essere eseguito su un sistema stereo e può essere eseguito con un sistema di diffusione del suono con 8 o più altoparlanti. Le uniche esigenze tecniche sono un sistema del suono e un lettore CD.

MUSIC SCORED FOR INSTRUMENTALISTS WITH TAPE/LIVE ELECTRONICS

Laura Bianchini
Fiaba (1997)
Sect. 1 - Ein Echtes Maerchen
SEC. 3 - Die Stimme
for percussion, tape and live electronics

FIABA for percussions, tape and live electronics is a cycle composed by four short compositions. Each of them devoted to a different family of percussions (metal, wood, drums, hand drums). The last one (FIABA 5) includes also a voice. It is possible to perform one or combinations of two, three compositions as well as the whole cycle.
One may tell a story or represent it; it is possible to do it using words or actions or a combination of words, actions, sounds. The sound mixture in the work FIABA (Fable) started out with the assumption to trace a path which allows the listener to freely define an imaginary world of characters, actions, places. As in the telling-stories where the scenes and the actions are consciously deformed, so that the musical path of FIABA, constructed with a combination of compositional timbrical materials are transformed to become "grotesque". The sound mixture in this work derives from the transformation of instrumental sounds and from the dialectics of recorded and direct sounds. There is no underlying text; the narrative trail is constructed entirely by the listener. The sound materials evolve accordingly to a metamorphosis which, in some cases, makes it impossible to recognize their original characteristics. The sounds are moved through the acoustic space defined by the loudspeakers generating a perception of movement or localization. The piece was produced using techniques for sound synthesis and processing in real time. At this occasion the two compositions Ein Echtes Maerchen, Die Stimme will be presented.

FIABA è un lavoro per percussioni, nastro magnetico e live electronics in quattro sezioni, ognuna dedicata ad una famiglia di strumenti a percussione; l'ultima (Fiaba 5) prevede anche una voce (mezzosoprano). Il lavoro può essere eseguito integralmente oppure in singole sezioni.
Una fiaba può essere narrata, rappresentata; la si può narrare con parole o con gesti o con una strumentazione di parole, gesti, suoni.
Come nella narrazione fantastica, in cui le immagini vengono deformate e gli accadimenti ingigantiti da descrizioni grottesche e caricaturali, così l'intreccio sonoro di FIABA è sottoposto a processi di trasformazione dinamica tali da rendere irriconoscibili, indefiniti e ambigui gli elementi caratteristici dei suoni originali.
Ogni episodio è caratterizzato dall'uso di una diversa famiglia di percussioni (metalli, legni, pelli, pelli naturali); gli strumenti e il nastro magnetico rappresentano due sorgenti sonore di cui una è estensione o deformazione dell'altra. Alla fusione timbrica di eventi, si alterna una riconoscibilità di questi, resa possibile grazie anche all'uso di figurazioni ritmiche che ne evidenziano le modalità di attacco, di scansione temporale e alla caratterizzazione dello spazio acustico.
Il lavoro è stato prodotto e realizzato al Centro Ricerche Musicali - CRM di Roma; per la generazione ed elaborazione dei suoni è stato utilizzato il sistema digitale in tempo reale Fly30, progettato presso il Centro.
E' dedicato a Michael Marschall von Bieberstein.

Agostino Di Scipio
Natura allo specchio (1998-2000)
2 or more percussionists, 8-track digital tape, and interactive processining computer with 8-channel output

Natura allo Specchio is based on two short fragments from Shakespeare's *The Tempest* and was initially composed as the introductory section to the composer's interactive music theatre work SOUND & FURY.
Through the Shakespeare fragments, this work speaks of noise and solitude. Noise, however, is considered here as the opposite to the negative attitude it is usually attached: it means the richness and complexity of life, it means the overwhelming intricacy of human perception before the understanding of its own environment, its own ambience. Solitude, on the other hand, means not only separation and loneliness, but also self-appropriation, self-observation and understanding. An epistemological metaphor. These two conceptual elements are rendered in sounds, not descriptively, not as narration. Quite the opposite. Each sound, each sonic event in this music, is a unique phenomenon emerging from a hidden, chaotic, pre-linguistic universe. Timbre is musical form. Musical gestures and the rhythms and pace of the piece emerge from an underworld which constitutes an entirely synthetic, but dynamically rich and veritable, soundscape.
The 8-track digital tape was generated with Functional Iteration Synthesis, used here to create an environment or ambience, which – though utterly synthetic in nature – is heard as animated by small agencies causing all sorts of acoustic turbolences, intermittence noises, etc.. Two voices are also featured in

the tape (thanks are due to Simon Emmerson and Peppe Servillo for lending voice, speaking the Shakespeare fragments.
The number of percussionists being involved in the performance is not fixed. Two should be the minimum. Required instruments include: woodblocks, tin drums and rototoms, all played with jazz-brushes. There is one only musical part, and all of the percussionists play from that, each with his/her own timing and interpretative variations. This arrangements results into a kind of micro-time polyphony, letting the details of macroscopical gestures emerge from the performers' interaction more than the notation itself. Also, the percussionists play only for 2 minutes, right in the middle of the piece. The percussion sound is processed in real time with granular processing algorithms and diffused over 8 loudspeakers. The processing parameters are computed in real time as a function of the percussion sounds, as well of the sounds travelling across the concert hall. Therefore, the particular sonic details in the processed sounds bear traces of the material and historical enironment where performance takes place.
Texts

[English speaker, loud:]
hang you,
hang you whoreson,
insolent noise-maker...

[English speaker, whispering:]
there they hoist us
to cry to th' sea that roar'd to us...to sigh
to th' winds, whose pity, sighing back again,
did us but loving wrong

[Italian speaker, whispering:]
e qui ci lasciamo, a gridare
al mare che ci ruggiva contro,
al sospirar dei venti, la cui pietà, ricambiando i sospiri,
ci faceva soffrire per troppo amore

Natura allo specchio è basata su due frammenti tratti da "La Tempesta" di Shakespeare e fu inizialmente composta come introduzione all'opera dell'autore stesso Sound & Fury, un'opera musicale teatrale interattiva.
Attraverso i frammenti di Shakespeare, quest'opera parla di rumore e solitudine. Il rumore, tuttavia, è qui considerato come l'opposto dell'attitudine negativa che normalmente gli viene attribuita: significa la ricchezza e complessità della vita, la complessità opprimente della percezione umana prima della conoscenza del proprio ambiente, della propria atmosfera. La solitudine, dall'altro lato, significa non solo separazione e solitudine, ma anche auto-appropriazione, auto-osservazione e discernimento. Una metafora epistemologica. Questi due elementi concettuali vengono resi nel suono, non in modo descrittivo, non come narrazione. Del tutto opposti. Ogni suono, ogni evento sonico in questa musica, è un unico fenomeno che emerge da un universo nascosto, caotico, pre-linguistico. Il timbro è forma musicale. I gesti musicali, i ritmi e l'andatura del pezzo emergono da un inferno che costituisce una scappatoia del tutto artificiale, ricco e veritiero in modo dinamico.
Il nastro digitale è stato generato con Sintesi Funzionali di Iterazione, usati qui per creare un ambiente o atmosfera che – benché totalmente sintetica in natura – è sentita come animata da piccole forze che provocano ogni sorta di turbolenze acustiche, rumori intermittenti, ecc. Anche due voci compaiono nel nastro (ringraziamenti sono dovuti a Simon Emmerson e Peppe Servilio per aver prestato le loro voci), cantando i frammenti di Shakespeare.
Il numero di percussionisti coinvolti nell'esecuzione non è fissato. Dovrebbero essere minimo due. Gli strumenti richiesti includono: blocchetti di legno, bidoni di latta, tutti suonati con spazzole da jazz. C'è solo una parte musicale, e tutti i percussionisti suonano da questa, ognuno con le proprie sincronizzazioni e variazioni interpretative. Questo arrangiamento risulta in una specie di polifonia micro-tempo, lasciando che i dettagli di gesti macroscopici emergano dall'interazione dell'esecutore più che dalla notazione stessa. Inoltre, i percussionisti suonano solo per due minuti, giusto a metà del pezzo. Il suono delle percussioni è elaborato in tempo reale con algoritmi di elaborazione granulari e diffusi da 8 altoparlanti. I parametri di elaborazione sono computerizzati in tempo reale come funzione del suono di percussioni, così come i suoni viaggiano attraverso la sala del concerto. Pertanto, i particolari dettagli sonici nei suoni elaborati portano traccia dell'ambiente materiale e storico dove ha luogo l'esecuzione.

Silvia Lanzalone
Tracciati (1998)
for percussion, planephones and live electronics

The work was inspired by the book "*Le città invisibili*" (*The invisible cities*) by Italo Calvino: the percussionist and the computer operator are explorers in an imaginary city as they follow paths created through use of different timbres according to criteria which oppose, exchange, assimilate and fuse as the piece progresses. The percussion is divided into three set-ups spread throughout the hall. These set-ups are based on the construction materials used in the instruments, wood, metal, membranes. The *Planephones* corresponding to each set-up are constructed of similar materials. These Planephones, designed by Michelangelo Lupone at CRM-Roma, are panels made using different construction materials that are set in vibration by the use of transducers attached to them. The wood, plastic and metal from which they are constructed transform the sound and give it different sound qualities coherently to the materials used. The panels, divided in three groups (metal, plastic, leather) are set on the stage besides the three stations of the percussion instruments. The percussionist is free to move between the different positions insuring a different outcome at each performance. The score consists of 9 non numbered pages, each page corresponding to different materials and therefore different sonorities. The performers can decide to perform any of three to nine pages in any order creating performances of different length and complexity.
The percussion sounds are altered by diffusion through the Planephones and by the use of the Mars system. The Mars is controlled via SDP (S. Lanzalone) and contributes another sound element to the composition. This material remains in evidence when the percussionist changes stations. Formal flexibility corresponds to the interpretative freedom given to the performers. They interact freely, creating a dialog that is further amplified as the percussionist carries on his interaction with the live electronics.

Tracciati è un lavoro ispirato al libro "Le città invisibili" di Italo Calvino. Il percussionista e l'esecutore al live electronics sono protagonisti di un processo di esplorazione all'interno di città immaginarie, attraverso percorsi sonori che mettono in relazione mondi timbrici diversi, secondo criteri di opposizione, scambio, corrispondenza, fusione, realizzando un progressivo divenire. L'organico delle percussioni è suddiviso in tre sezioni che sono disposte in tre punti differenti della sala e raggruppate secondo i materiali di cui sono composti gli strumenti – legno, metallo, membrana -, in corrispondenza con i materiali dei planofoni. I planofoni, pannelli vibranti progettati da Michelangelo Lupone e realizzati presso il CRM di Roma, trasformano il segnale sonoro loro inviato coerentemente alle caratteristiche vibratorie del materiale di cui sono composti. Il legno, il metallo e la plastica dei pannelli impongono le loro formanti timbriche al materiale sonoro che le mette in vibrazione; lo scambio dei materiali consente di ottenere molteplici mescolanze: il legno-metallo, il metallo-plastica, la plastica-legno, e così via. Il percussionista raggiunge ed esplora le diverse postazioni strumentali e la possibilità di variare l'itinerario dei tracciati, in modo differente per ciascuna esecuzione, consente di realizzare un percorso formale di volta in volta nuovo. La partitura è composta di 9 pagine non numerate. Ciascuna pagina ha un titolo che fa esplicito riferimento ai materiali utilizzati e di conseguenza alle sonorità ricercate. Gli esecutori possono scegliere di eseguire in concerto da 3 a 9 pagine, disponendole nell'ordine che ritengono opportuno. I criteri che devono seguire per operare la selezione devono tenere conto, oltre che della durata complessiva, soprattutto di un tipo di percorso formale che metta in un particolare relazione le pagine scelte (es.: opposizione, evoluzione, ecc.). Il suono delle percussioni viene trasformato secondo due livelli di elaborazione, il primo effettuato direttamente dai planofoni, il secondo effettuato con ulteriore contributo del sistema Mars. L'elaborazione del suono, realizzata con algoritmi di granulazione, si pone come un secondo piano sonoro che si fonde con i suoni strumentali, oppure si pone in evidenza durante gli spostamenti del percussionista. Alla flessibilità delle scelte formali corrisponde una flessibilità delle scelte interpretative, per cui l'esecuzione dei percorsi e di ciascuna pagina prevede la possibilità di un margine di dilatazione o di contrazione temporale che è perfettamente assecondato dall'interprete al live electronics, con cui il percussionista instaura un rapporto dialogico.

Seungyon-Seny Lee
Chuk-Won I (1999)
For percussion, tape and live elctronics

This piece is based on "Samul nori" which is a traditional form of Korean percussion music. Samuel means "four things" in English and nori means "performing". The ensemble's members consist of two skins and two metals.
The instruments symbolise earth (skins) and the heavens (metal). The instruments are identified with a constantly changing natural world. The metal instruments represent (1) Spring/lightening/thunder and (2) Summer/wind. The skin instruments represent (1) Autumn/rain and (2) Winter/clouds. It is said that if people

play on these four instruments together, the resulting vibrations will harmonize earth and heaven into one universe.
Sounds for this piece originate from recordings of skin and metal instruments used in the performance of Samul nori. Sounds for the 1st movement records of stones, metal instruments, ceramic bowels, and insect, etc.. The recording sounds are used of sound editing program such as Sound Editor and Granular Synthesis. The piece will from the first part of a four-movement composition entitled Chuk-won which roughly translates as "invocation".

Questo pezzo è basato sul "Samul nori", che è tradizionale della musica a percussione coreana. 'Samuel' significa "quattro cose" e 'nori' significa " rappresentazione". L'insieme degli elementi consiste in due pelli e due metalli.
Gli strumenti simboleggiano cuore (pelli) e cieli (metallo). Essi sono identificati con un mondo naturale in continuo cambiamento. Gli strumenti in metallo rappresentano (1) Primavera/illuminazione/tuono e (2) Estate/vento. Gli strumenti in pelle rappresentano (1) Autunno/pioggia e (2) Inverno/nuvole. Si dice che se si suonano questi quattro strumenti insieme, le vibrazioni risultanti armonizzeranno terra e cielo in un unico universo. I suoni per questo pezzo traggono origine da registrazioni di strumenti di pelle e di metallo. I suoni per il primo movimento registrano sassi, strumenti di metallo, ciotole di ceramica, insetti ecc.. I suoni registrati vengono usati in programmi di elaborazione del suono come Direttore del Suono e Sintesi Granulare.
Questo pezzo verrà fuori dalla prima parte di una composizione in quattro-movimenti intitolata Chuk-won, tradotto approssimativamente come "invocazione".

Zlatko Tanodi
Air (1997)
For bass clarinet and tape

My inspiration was the polysemy of the word air (air, atmosphere, area, aria, song with accompaniment. The motto of the whole work is a paraphrase: "air you were, air you are and into you shall return" treating sound as a special physical condition of the air needed for acoustic perception, as well as for starting of the resonance of the bass clarinet's body. While a naked instrument looks for its own expression, offering building material heartlessly exploited by sampling machines, a cause-effect of man and machine appears: it is not possible to distinguish who is who. The acoustic instrument would like to sound like an electronic one, and electronic instruments more faithfully imitate acoustic ones. However, at the end, all these are just vibrations in the air, all is but air.
Other association lines are also possible (e.g. building, history, money ways, etc.).
Composition has four movements: RespIRAtion, PrepARAtIon, Aria and Gone with the Wind. The electronic part recorded on the DAT tape (or CD) is made by electronic transformations of the bass-clarinet sound (except well-known rhythmic quotation). I tried to make virtual interaction with soloist: listener is not sure if soloist goes after tape or, on the contrary, tape goes after soloist. Consequently, listener is not quite sure about the origin of the sound: is it recorded or is it alive? For that reason it is necessary to have quality amplification (stereo). In the case of live performance, a microphone for soloist is needed only in big concert hall. Once started, tape is running till the end of piece. The relation of loudness between performer and tape is defined by the first sound on the tape: it must sound equally as a soloist.

L'ispirazione è stata la polisemia della parola aria (aria, atmosfera, area, aria, canzone con accompagnamento). Il motto del mondo intero è una parafrasi: "aria eri, aria sei e nell'aria tornerai", trattando il suono come una speciale condizione fisica dell'aria necessaria per la percezione acustica, così come per l'inizio della risonanza del corpo del clarinetto basso. Mentre uno strumento spoglio cerca la sua propria espressione, offrendo materiale per la creazione insensibilmente sfruttato da campionature di macchine, appare una causa-effetto di uomo e macchina: non è possibile distinguere chi è chi.
Lo strumento acustico suonerebbe come uno strumento elettronico, e gli strumenti elettronici più fedelmente imitano quelli acustici. Tuttavia, alla fine, tutti questi sono vibrazioni nell'aria, tutto tranne aria.
Sono possibili altre linee di associazioni (ad es. costruzioni, storia, soldi, ecc.)
La composizione ha quattro movimenti: RespIRAzione, PrepARAzIone e Via col Vento. La parte elettronica registrata su nastro DAT (o CD) è prodotta da trasformazioni elettroniche del suono del clarinetto basso. Ho cercato di produrre un'interazione virtuale con il solista: l'ascoltatore non è sicuro se il solista segua il nastro o, al contrario, se il nastro segua il solista. Di conseguenza, l'ascoltatore non è abbastanza sicuro circa l'origine del suono: è registrato o è dal vivo? Per questo motivo è necessario avere un'amplificazione di qualità (stereo). In caso di esecuzione dal vivo, solo nelle grandi sale da concerto è richiesto un microfono per il solista. Una volta partito, il nastro scorre fino alla fine del pezzo. La relazione di sonorità tra l'esecutore e il nastro è determinata dal primo suono sul nastro: deve sembrare come un solista.

Riccardo Vaglini
September – Exercises (1992)
For 7 sax and tape

September-exercises is a ritual exploration of the realms of contradiction, of multiplicity and inconsistency. A small number of definite instrumental and electronic signs, both symbolic and conceptual is almost scenically employed, which become not only an exterior conflict but particularly one within an own context. In the relationship between the soloist and *the other* as a projection, the ritual to summon the *September* and the opposite prayer or exorcism blend into each other and with the hues of desire and death.

September-exercises è un'esplorazione rituale dei regni della contraddizione, della molteplicità e dell'inconsistenza. Un piccolo numero di definiti segni strumentali ed elettronici, sia simbolici che concettuali, è quasi scenicamente impiegato, e diventa non solo un conflitto esteriore ma in particolare nel proprio contesto. Nella relazione tra il solista e l'altro come una proiezione, il rituale per convocare Settembre e la preghiera opposta o l'esorcismo si mescolano l'uno nell'altro e con tinte di desiderio e morte.

MUSIC FROM PERFORMER

Antongirolami Gianpaolo, saxsophones
Michael Edwards
Flung me, foot trod

Gianpaolo Antongirolami has received a first-class degree at the Conservatory of Fermo under the direction of Federico Mondelci, later specializing with Jean-Marie Londeix. Immediately launched his brilliant musical style in various concerts all around Italy and abroad with various groups and has also made several recordings.
Being particularly interested in contemporary music and experimentation, he has played many world premieres such as "Voliera" of Sylvano Bussotti. His fascination for new technologies has led him to combine contemporary music and electronics applied to electroacoustic music. In 1999 he received a diploma in Electronic Music at the Conservatory of Pesaro under the direction of Eugenio Giordani with whom in 1998 he created a new saxophone real-time digital system version of Karlheinz Stockhausen's "Solo". Parallely, he has been working as a professor since 1987, running specialization courses and seminars in electronic music and saxophone. He has co-operated for many years with the flautist Roberto Fabbriciani, teaching 20th-century chamber music in his courses in Italy and Greek. *He has been also invited to teach in workshop in the Musikhochshule in Freiburg im B. where he will return the next year.*

Gianpaolo Antongirolami ha studiato sassofono presso il Conservatorio "G. Rossini" di Fermo con Enzo Veddovi e Federico Mondelci: sotto la guida di quest'ultimo si è diplomato nel 1987 con il massimo dei voti; ha seguito corsi di perfezionamento con Jean-Marie Londeix. Ha intrapreso una brillante attività concertistica con varie formazioni in tutta Italia e all'estero, suonando anche da solista con l'Orchestra "Ukraina" di Kiev; ha effettuato varie registrazioni radiofoniche e discografiche. Particolarmente interessato alla musica contemporanea e alla sperimentazione, ha eseguito numerose prime esecuzioni, come quella della composizione "Voliera" di Sylvano Bussotti – per flauto sax e piano – realizzata nell'ambito del X World Saxophone Congress. Nel suo repertorio figurano, oltre ad opere a lui dedicate, le composizioni dei più importanti autori del nostro secolo: Berio, Bennett, Bryars, Cage, Carter, Donatoni, Gentilucci, Grisey, Kolb, Manzoni, Nyman, Nono, Pousseur, Sciarrino, Stockhausen, Torke, Xenakis. Rilevante è inoltre il suo impegno nel campo della musica elettronica. Nel 1999 ha conseguito il diploma presso il Conservatorio di Pesaro sotto la guida di Eugenio Giordani. Ha seguito corsi perfezionamento con Di Scipio, Lupone, Risset. Ha tenuto seminari presso i conservatori di Bari, Pesaro, Foggia, e un corso di formazione professionale sull'audio digitale; affianca all'attività di ricerca quella di interprete. Nel 1997 è stato selezionato, tra settecento varie proposte, ad esibirsi nell'ambito di ISEA-97, a Chicago. Nel 1998 ha realizzato insieme ad Eugenio Giordani, nell'ambito delle attività del Laboratorio Elettronico per la Musica Sperimentale del Conservatorio di Pesaro, una nuova versione per sassofono contralto e sistema digitale di elaborazione in tempo reale (Kyma/Capybara) di "Solo" di Karlheinz Stockhausen; questa versione è stata presentata nel corso del 1998, in occasione del 70° compleanno di Stockhausen, durante l'incontro "La Terra Fertile" (L'Aquila), nell'ambito del convegno internazionale "Il sassofono, strumento inevitabile per il 2000" svoltosi al Conservatorio di Milano, e nell'ambito del "Progetto TecnoArte 2000" presso il Conservatorio di Pesaro. Sempre nel 1998 ha partecipato al XII Cim (Colloquio di Informatica Musicale), tenutosi a Gorizia.
Suona in prevalenza da solista e collabora con varie formazioni cameristiche. Nel 1993 è stato invitato a tenere una workshop presso la prestigiosa Musikhoschule di Freiburg dove ha anche eseguito un apprezzato concerto.
Parallelamente all'attività concertistica svolge dal 1987 quella didattica; ha collaborato per vari anni ai corsi tenuti dal flautista Roberto Fabbriciani in Italia e in Grecia insegnando musica da camera del '900; tiene corsi di perfezionamento per vari enti quali l'ASI. È titolare della cattedra di sassofono al Conservatorio di Foggia.

Michael Edwards
Flung me, foot trod (1994)
For sax alto and tape

Flung me, foot trod takes its title from the Gerard Manley Hopkins sonnet, Carrion Comfort. This is urgent, violent, exciting poetry, but it was not until I read some of Hopkins' own notes to the verse that I felt particularly drawn to pilfering a title from him. He writes of one word, "rude," that must be enunciated with force, "in an uncouth, violent, barbarous manner." This, if anything, summarizes the articulation necessary to interpret my piece. In preparing the tape, I sampled selected portions of the solo part. In particular I concentrated on some of the more unorthodox sounds an alto saxophone can make, key clicks, breath noise, growling etc. For demonstrating these sounds, I am very grateful to Gary Scavone who gave freely of his time and tolerated my often outlandish requests. Indeed, the whole piece is aimed at utilizing his slick virtuosity. Armed with these samples, it was my intention to create sounds that go far beyond the timbral qualities of the saxophone. Although the tape sometimes presents recognizable saxophone sounds, on the whole it is in its own sonic realm, marrying itself with the solo part only in its presentation of similar material types (driving rhythms, scurrying textures etc.). It was not my intention to create the effect of an "orchestra of saxophones," or to have the saxophone play against itself on tape. On the contrary, flung me, foot trod takes its precedent more from the solo concerto, pitting two unequal forces against each other, their only common ground being material and, hence, structure. On the more technical side, the samples were processed using Bill Schottstaedt's "Common Lisp Music", the note lists were created with Heinrich Taube's "Common Music", and the mixing was accomplished with Paul Lansky's "Real Time Mixer" application-all on the NeXT computer.

Flung me, foot trod prende il nome da un sonetto di Gerad Manley Hopkins, 'Carrion Comfort'. Si tratta di una poesia pressante, violenta ed eccitante, ma non me ne sono reso conto finché non ho letto alcune note dello stesso Hopkins riguardanti il verso che mi sembrava più adatto per fornire il titolo alla composizione. Egli scrive che la parola "rude" deve essere pronunciata con forza in "an uncouth, violent, barbarous manner" (in modo rozzo, violento e barbaro). Questa frase riassume in maniera esemplare lo spirito necessario per interpretare il brano per sassofono e nastro. Nella preparazione del materiale elettronico ho utilizzato parti del solo sassofono da lui selezionate. In particolare mi sono concentrato su alcuni dei suoni meno convenzionali del sax alto: colpi di chiave, soffi intonati, growling; tutto il pezzo mira a sfruttare un virtuosismo pirotecnico del sassofono. Grazie a questo materiale campionato, ho creato suoni lontani dalle qualità timbriche convenzionali del sassofono. Tuttavia la parte del nastro presenta qualche volta suoni chiaramente riconoscibili, propri del suo regno sonoro, che si sposano con la parte a solo. Non è stata mia intenzione creare un effetto tipo 'orchestra di sassofoni' o do avere il sassofonista che suona con una sua copia impressa nel nastro. Per la preziosa collaborazione, sono particolarmente grato al sassofonista Gary Scavone.
Il pezzo è stato creato presso il laboratorio del CCRMA della Stanford University in California. La prima esecuzione è stata realizzata a Stanford dal sassofonista Gary Scavone.

Susanna Borsch, recorder
Mike Vaughan
In memoriam...(layer 1)

Susanna Borsch was born in Hamburg in 1974 and in 1983, began recorder lessons with Silker Kühner. She followed master-classes with the leading recorder players of Europe, including Kees Boeke, the members of the ALSQ, Marion Verbruggen, Han Tol, Peter Holtslag, Gerd Lünenbürger a.o. after many years of taking part with great success in Germany's "Jugend musiziert" – competition, she founded in 1993 the recorder trio "Les Doux Siffleurs" which in 1995 won top prizes at both The German Music Competition in Bonn and the International recorder symposium in Calw. Three years later, she won third prize in the solo competition at the same event.
In addition, she was a founding member of the recorder quartet "Malle Symen" and in 1998, took a major part both as soloist and ensemble member in the 2nd International Recorder week' at "de IJsbreker", the internationally renowned modern music venue in Amsterdam.
In the summer of 1994, she began studying at the Amsterdam Sweelinck Conservatorium with Walter van Hauwe, finishing her teaching diploma in 1998. The same year, saw her begin post graduate studies, specializing in the combination of the recorder and live electronics. This project will cumulate in the year 2000, with her final solo examination, featuring new works written especially for her and this combination.
Her teaching experience includes giving master-classes in many European countries.
In 1999, she became a member of "The Interval Chamber", a larger ensemble of diverse musicians, specializing in micro-tonal music. She contributed to their recent production of "Wölfli" a multi-media opera written by Rafael Reina. Also in 1999, she deputized at very short notice, in the celebrated "Amsterdam Loeki Stardust Quartet".

Susanna Borsch è nata ad Amburgo nel 1974 e nel 1983 ha cominciato a seguire lezioni di flauto dolce con Silker Kühner. Ha seguito master con i principali musicisti di flauto dolce d'Europa, incluso Kees Boeke, i membri dell' ALSQ, Marion Verbruggen, Han Tol, Peter Holtslag, Gerd Lünenbürger. Dopo aver preso parte per molti anni con grande successo alla competizione "Jugend Musiziert" in Germania, nel 1993 ha fondato il trio "Les Doux Siffleurs" che nel 1995 ha vinto il primo premio sia a "The German Music Competition" a Bonn che all' "International Recorder Symposium" a Calw. Tre anni dopo, ha vinto il terzo premio nella gara per solisti nella stessa manifestazione. In più, è stata membro fondatore del quartetto di flauto dolce "Malle Symen" e nel 1998, ha preso parte sia come solista che come membro di un gruppo alla II International

Recorder Week' a "de Ijsbreker", ad Amsterdam, il convegno di musica moderna conosciuto a livello internazionale.
Nell'estate del 1994 ha iniziato gli studi all'Amsterdam Sweelinck Conservatorium con Walter van Hauwe, completando il suo diploma d'insegnante nel 1998. Nello stesso anno, ha iniziato gli studi post-diploma, specializzandosi nella combinazione di flauto dolce e live electronics. Questo progetto vedrà nel 2000 il suo esame finale da solista, rappresentando nuovi lavori scritti per lei e per questa combinazione.
Nel 1999 è diventata membro di 'The Interval Chamber', un vasto gruppo di differenti musicisti, specializzato in musica micro-tonale. Ha contribuito alla loro recente produzione di "Wölfli" , un'opera multimediale scritta da Rafael Reina.

Mike Vaughan
In memoriam...(layer 1) (1999)
(1994)
For recorder and tape

In Memoriam is a working title for a cycle of pieces which are all based on an analysis and projection of fragments of improvised material performed by the jazz artist Eric Dolphy on the Blue Note recording 'Out to Lunch'. The original idea was developed over a period of time, following a request for an ensemble piece by Roger Dean and austraLysis based around a nucleus of bass clarinet, percussion and bass.
The projected cycle now includes instrumental solos for bass, bass clarinet and saxophone, two ensemble works (one including triggered soundfiles and live processing), a piece for tenor recorder and live electronics (completed in 1999) and two pieces for solo instruments and tape – bass clarinet (completed in March 2000) and bass – (next to be done). Multiple readings of the same material are a feature of the composition process; all works share a common pool of material, which is recontextualised in each piece. The two completed works, for tenor recorder and live electronics and bass clarinet and tape, have the most material in common.
The first member of this set for tenor recorder, (currently layer 1), was written for Susanna Borsch in Autumn 1999 and was requested as part of an on-going project to expand the repertoire for this instrument in relation to live electronics. Although much of the live processing is relatively uncomplicated in itself, the rapid transitions from one state to the next (hopefully) creates a means whereby differences in the microstructures within the original material source can be explored more fully.

"In Memoriam" è il titolo di un ciclo di pezzi tutti basati sull'analisi e la proiezione di frammenti di materiale improvvisato, eseguito dall'artista jazz Eric Dolphy nella registrazione Blue Note "Out to Lunch". L'idea originale era di sviluppare un periodo di tempo, seguendo una richiesta di un pezzo per gruppo di Roger Dean, basato su un nucleo di clarinetto basso, percussioni e basso. Il ciclo comprende assolo strumentale per basso, clarinetto basso e sassofono, due lavori per gruppo, un pezzo per tenore e live electronics (finito nel 1999) e due pezzi per strumento solo e nastro – clarinetto basso e basso. Molteplici letture dello stesso materiale sono caratteristiche del processo di composizione; tutti i lavori condividono un insieme comune di materiali, che viene ricontestualizzato in ogni pezzo. I due lavori completi, per tenore e live electronics e clarinetto basso e nastro, hanno la maggior parte del materiale in comune.
La prima parte di questa raccolta per tenore è stata scritta per Susanna Borsch nell'autunno 1999 ed era stata richiesta come parte di un progetto per espandere il repertorio per questo strumento in relazione al live electronics.

INDEX • INDICE

Contents Contenuti

Invited talk

Sound Analysis and Synthesis I

Sound Analysis and Synthesis II

Musical Informatics: Expression and performance analysis I

Musical Informatics: Expression and performance analysis II

Musical Informatics: Expression and performance analysis III

Musical Systems

Music Analysis and Cognition I

Music Analysis and Cognition II

Music Archives I

Music scored for instrumentalists with tape/live electronics

Music from performers