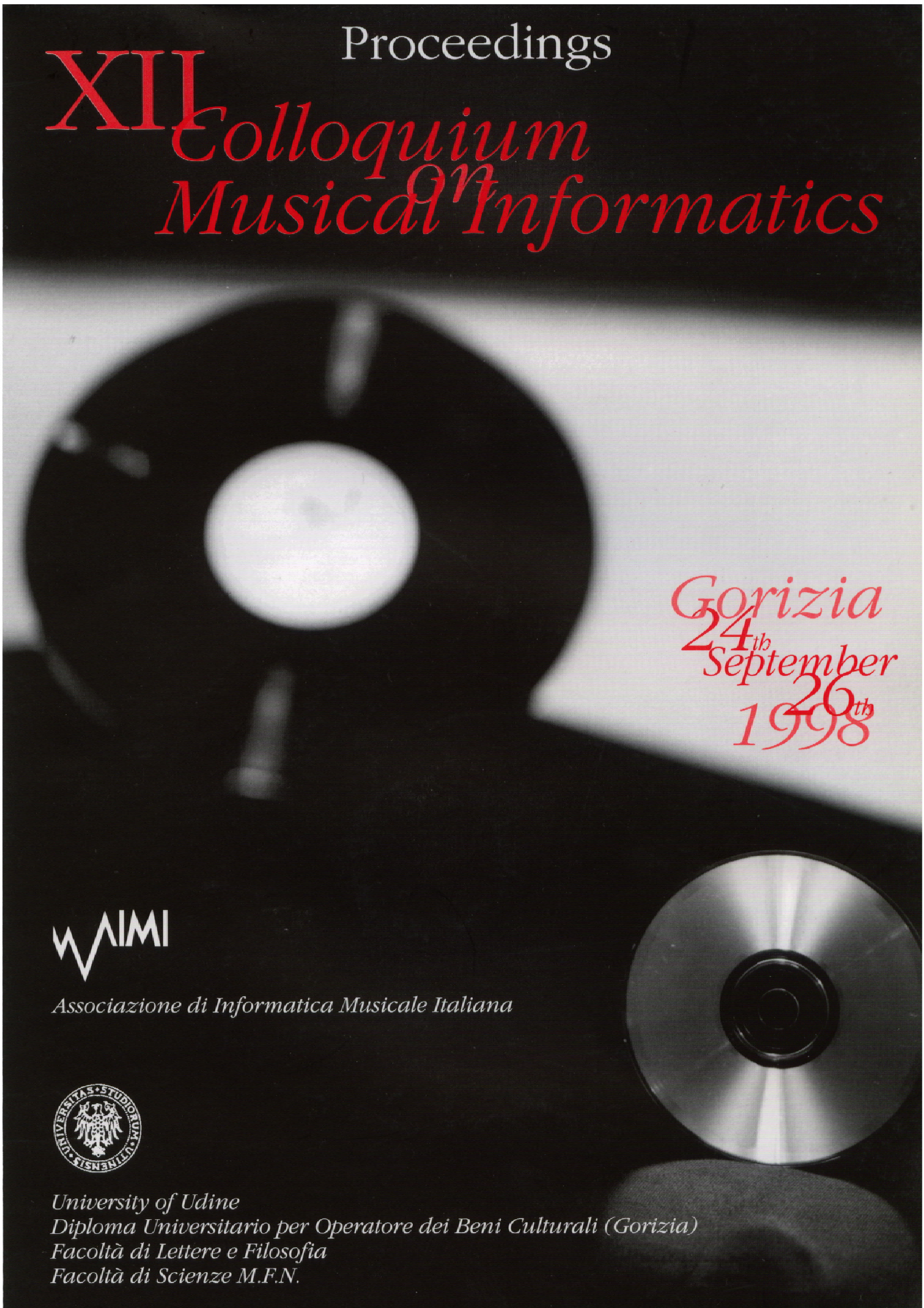
Proceedings

XII Colloquium on Musical Informatics



Gorizia
24th September
26th
1998

AIMI

Associazione di Informatica Musicale Italiana

University of Udine
Diploma Universitario per Operatore dei Beni Culturali (Gorizia)
Facoltà di Lettere e Filosofia
Facoltà di Scienze M.F.N.

Proceedings

XII

Colloquium Musical Informatics

Edited by:
Alessandro Argentini, Claudio Mirolo

Gorizia
24th September 26th 1998

CASTELLO
Sala del Conte e Carceri

AUDITORIUM DELLA CULTURA FRIULANA
Via Roma

AIMI *Associazione di Informatica Musicale Italiana*

University of Udine
Diploma Universitario per Operatore dei Beni Culturali (Gorizia)
Facoltà di Lettere e Filosofia
Facoltà di Scienze M.F.N.

Scientific Committee

Daniel Arfib	(LMA-CNRS - Marseille, France)
Denis Baggi	(IEEE CS TC on CGM)
Mario Baroni	(University of Bologna, Italy)
Antonio Camurri	(University of Genova, Italy)
Roger Dannenberg	(Carnegie Mellon Univ., USA)
Giovanni De Poli	(University of Padova, Italy)
Giuseppe Di Giugno	(Iris-Bontempi, Italy)
Gianpaolo Evangelista	(University of Napoli, Italy)
Shuji Hashimoto	(Waseda University, Japan)
Goffredo Haus	(University of Milano, Italy)
Furio Honsell	(University of Udine, Italy)
Marc Leman	(University of Gent, Belgium)
Alan Marsden	(Queens Univ. - Belfast, UK)
Livio C. Piccinini	(University of Udine, Italy)
Alberto Policriti	(University of Udine, Italy)
Stephen Pope	(Univ. of Calif. St. Barbara, USA)
Curtis Roads	(Univ. of Calif. St. Barbara, USA)
Robert Rowe	(New York University, USA)
Eleanor Selfridge-Field	(Stanford University, USA)
Xavier Serra	(Pompeu Fabra University, Spain)
Julius Smith	(Stanford University, USA)
Leonello Tarabella	(CNUCE-CNR, Pisa, Italy)
Ioannis Zannos	(Staat. Inst. für Musikforschung -Berlin, Germany)

Chairman of the Special Session on Restoration of Audio Documents:

Dietrich Schüller	(Akad. Wissensch.-Wien, Austria)

Musical Committee

Nicola Bernardini	(Conservatory of Padova, Italy)
Lelio Camilleri	(Conservatory of Bologna, Italy)
Agostino Di Scipio	(Conservatory of Bari, Italy)
Angelo Orcalli	(University of Udine, Italy)
Alvise Vidolin	(Conservatory of Venezia, Italy)

Organizing Committee

Fabio Alessi	(University of Udine)
Alessandro Argentini	(LIM - Gorizia, Univ. of Udine)
Giovanni De Mezzo	(LIM - Gorizia, Univ. of Udine)
Agostino Dovier	(University of Verona)
Claudio Mirolo	(University of Udine)
Angelo Orcalli	(University of Udine)
Luisa Zanoncelli	(University of Udine)
Paolo Zavagna	(LIM - Gorizia, Univ. of Udine)
Lucia Vinzi	(Secretariat)

XII Colloquium on Musical Informatics

Gorizia 24-26 September 1998

Promoted by

A.I.M.I. Associazione di Informatica Musicale Italiana

Organized by

Università di Udine
Diploma Universitario per Operatore dei Beni Culturali (Gorizia)
Facoltà di Lettere e Filosofia - Dipartimento di Scienze Storiche e Documentarie
Facoltà di Scienze M.F.N.

Co-sponsored by

I.E.E.E. CS Technical Committee on Computer Generated Music

Supported by

Comune di Gorizia
Assessorato alla Cultura

Provincia di Gorizia
Assessorato alla Cultura
Assessorato all'Istruzione e Polo Universitario

Regione Autonoma Friuli-Venezia Giulia

The publication of the proceedings contributed by

Consorzio Universitario del Friuli

INTRODUCTION

The town of Gorizia has the honour to host the 12th edition of the Colloquium on Musical Informatics, organized by the University of Udine – Diploma Universitario per Operatore dei Beni Culturali jointly with the Facoltà di Scienze M.F.N. As usual, this event has been promoted by the Italian Association on Musical Informatics, AIMI.

The contributions presented this year encompass not only the traditional fields of Digital Signal Processing and Music Programming Systems, but also some novel themes witnessing the growing interest in the informatic tools to analyze and support music languages, composition techniques, musical interpretation and performance. This trend leading to a more extensive use of computers to solve a wide range of problems was already highlighted by the previous edition of the Colloquium. This is a choice of a cultural nature for the "Diploma per Operatore dei Beni Culturali" (Archival Conservation), which recently instituted a Laboratory of Musical Informatics devoted to teaching contemporary music. Now the Laboratory is fully equipped for restoring audio documents, and then it is also trying to stimulate new synergies between researchers in music and computer science: this is indeed coherent with its main goals and has been fully supported since the beginning by the University of Udine.

Within this new context the link with Information Technology is not only a matter of choice, but rather stems from actual necessity. The musicology major of the Diploma Universitario per Operatore dei Beni Culturali deals with problems connected to the preservation and restoration of sound sources that are transferred onto digital memory, so that readability may be restored for documents that have been damaged or simply eroded by time. For instance, the Information Technology has developed very powerful tools to remove noise from the signals and for processing digital representations, but operating criteria and methodology need to be defined precisely.

It therefore becomes a question of restoring the information in a philologically correct manner, beginning with the reconstruction "in vitro" of the technical and operational features of the old equipment used to play outdated reproduction media, and arriving at the re-assembling of the information stored as bits in the computer into its original wholeness, with cognitive as well as informative integrity. This is an extremely complex task assigned to the joint-effort of computer scientists and professionals having specific historical-critical musical competence.

This is the main reason that spurred the University to suggest that a specific session of the CIM should be dedicated to the theme of preservation and restoration of sound sources. The response from both scholars and researchers has been impressive, as demonstrated by the high quality of contributions received in this area.

While promoting this Colloquium, the Organizing Committee hopes to give a further contribution to the methodological aspects of the conservation and restoration of sound documents. These are themes of strategic and cultural importance, given the growing number of electronically stored data, access to which is becoming more and more dependent on the technological evolution of information systems.

Gorizia, September 1998

Angelo Orcalli

PRESENTATION

Antonio Camurri

President
AIMI – Associazione di Informatica Musicale Italiana

The field of Musical Informatics changed significantly since the foundation of AIMI (Associazione di Informatica Musicale Italiana). It is therefore natural and necessary that AIMI adapts to the growth of the field: from the emergency of new research themes, to new interdisciplinarities, presence of the industry, etc. The main conference theme of this year Colloquium of Musical Informatics - Restoration of Audio Documents - bears witness of such deep transformations. "Restoration" is in fact a word which, until a few years ago, was referring to museums and to artworks of centuries ago.

Nowadays, there is the emergent need to restore also electronic and computer music works. A significant growth and transformation characterized the last few years of AIMI. For example, in 1997 AIMI offered its patrocine to the Intl. Workshop on Kansei - The Technology of Emotion, held in Genova in October, to the Workshop on sound signal processing, held in Florence in June, and to the Piero Grossi's 80 years event, in Florence in December. All these events received a very positive response from the public and confirmed the relevance of the Italian school within the international Musical Informatics community. AIMI has an internet home page (http://aimi.dist.unige.it), which has become a reference for AIMI members: it includes news on the activities of the Association and on the most relevant events in the field. The AIMI homepage has gradually enriched with new pages which bear witness of the vitality of the Italian community coordinated by AIMI: for example, the page on the sound synthesis based on physical models, which collects some national contributes to this important topic, the page on csound and related music software (ftp://musart.dist.unige.it/pub/CSOUND/ and http://aimi.dist.unige.it/AIMICSOUND/AIMICSOUND_home.html), and the page on the CEC project COST-G6 on Digital Audio Effects (http://aimi.dist.unige.it/AIMI_COST.html).

Furthermore, AIMI gives hopitality to the e-mail list of the DAT (Dipartimenti per l'Aggiornamento Tecnologico) of the national Music Conservatories, and actively participates to the activities of GATM (Gruppo di Analisi e Teoria Musicale), of which it is founding member.

The e-mail group of AIMI members is an active discussion forum and a service for a fast communication of national and international events. The links to the internet home pages of the single AIMI members allow them a better visibility and communication by means of the Association home page.

This year is yet more important for our association. The XII edition of the Colloquium on Musical Informatics will be held in Gorizia from 24 to 26 September. This conference has raised to an international relevance. It is characterized by a significant number of high-level scientific and musical contributes, as confirmed by this volume of proceedings. It is of particular significance that this year, the Colloquium will be for the first time in Gorizia, organized by the University of Udine, where an emergent research group on musical informatics is active. If from one hand the main theme of the Colloquium is this year "Restoration" , from the other hand it is important to notice that a relevant presence, in this edition of the Colloquium, of innovative contributes from other research themes: from digital sound signal processing to systems for the computer music, to emergent themes, e.g., analysis and synthesis of expressivity,

interactive systems and issues related to the integration of music language with visual, gesture, dance languages, and toward new generations of multimedia systems, including multimodal interaction and inhabited virtual environments.

In conclusion, I wish to thank the Organizing Committee, the Scientific and Music Committees, and the authors for their invaluable contribute to the success of the Colloquium.

Gorizia, September 1998

Thursday 24th **h. 9.20**

DIGITAL SIGNAL PROCESSING: SOUND ANALYSIS AND SYNTHESIS I

A System Based on Fourier Analysis / Synthesis for the Hybridisation of Sound Timbres

Raffaele de Tintis
LIM - Laboratorio di Informatica Musicale
Dipartimento di Scienze dell' Informazione
Università degli Studi di Milano
via Comelico 39, I-20135 Milano (Italia)
fax +39 2 5500637
e-mail: rdt@lalim.lim.dsi.unimi.it

1. Abstract

This paper presents audio hybridization using Morph, a software performing spectral modeling running on Windows95 platform. A description of the method and of the currently implemented features is given, followed by some guidelines for future development of the system. All of the algorithms are based on frequency domain processing in the Fourier space.
The hybridization algorithm is based on the deterministic plus stochastic decomposition proposed by Serra and Smith but Inverse Fourier Transform is used for re-synthesis.
Cross-synthesis was implemented adding the canonical method the possibility of changing, even drastically, the connections between pairs of spectral components.
A full description of the features actually performed by the system include: timbre hybridization, cross-synthesis, harmonic/inharmonic component separation, frequency-domain filtering, 3D visualization of spectra and sonogram.

2. Cross-Synthesis

Apart from special cases, cross-synthesis can not be considered valid for the production of perceptually good hybrids. For this reasons, some new features have been added to the basic method, which can be useful if the aim is that of producing totally new sonic matherials.
In addition to the classic features performed by cross-synthesis, the spectral connections between pairs of components of the two involved sounds, can be totally redefined. Neither component need necessarily be matched with that in the same frequency-bin in the other sound, but could also be connected with any other component.
Thus we can redefine, even drastically, the spectro-temporal characteristics of the two sounds involved in the processing. Perceptually, the presence of the two starting sounds may be lost and the sound generated may be completely new.
If, as a particular case, we don' t change any one of the basic spectral connections, the system performs cross-synthesis as described by the established rules.

3. Timbre Hybridisation

As a point of departure for the hybridisation method adopted, there is a technique of spectral manipulation (6) which, according to different variants, is nowadays used in a family of methods commonly known as spectral modeling. Among these, the techniques of hybridisation and morphing occupy a significant part.
The basis of the method is the recognition of the stable component of the spectrum, an approximation of the harmonic structure, and recognition of the unstable component which, from the perceptive point of view, constitutes the inharmonic part of the sound.
Seeking the largest peeks in the power spectrum and comparing them with the stable components in the phase spectrum, the stable component of the spectra is localised. In this way a segmentation of the starting spectra in different zones is obtained, each of which is centred around a peak. A rising number is assigned to every zone in such a way that, in phases of re-synthesis, the elaboration is obtained between pairs of zones with the same value.
This type of segmentation of the spectrum into different areas is a feature of the ideas known as "Auditory Image Formation" [1]. According to "Auditory Image Formation" we perceive groups of frequencies (formants,

energy patterns and so on), rather than individual components.
To improve elaboration possibilities, I have used 3 different coefficients of hybridisation which control moduli, phases and frequency-bins. In this way, it is possible to place the resonance areas characteristic of one sound at the positions taken by those ones of an other sound. The processing, consists in stretching or compressing of the distances between zones of the spectral envelope, independently for each zone. Thus, we can redraw the energy distribution of one sound, without modifying its moduli and phase values. The computation of new moduli and new frequencies, as of new frequency-bins, is done on a logarithmic scale.
The construction of the new spectrum is done by the simultaneous processing of pairs of spectral zones

Fig. 1

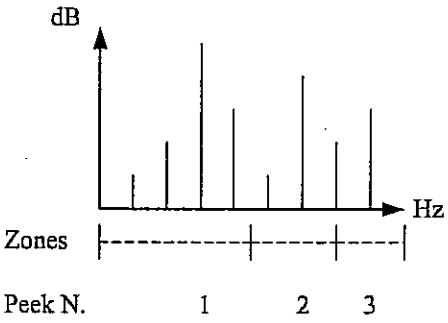


Fig. 2

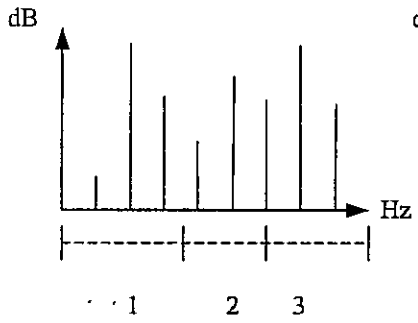
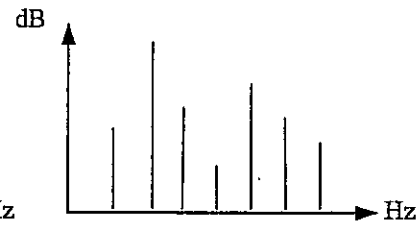


Fig. 3



belonging to the two initial sounds, which share the same number. One begins from peaks which distinguish each zone and moves on to the pairs of lateral components. Using the two spectrums in Fig. 1 and Fig. 2 we should match, on a logarithmic scale, components 3 and 2 respectively from the first and second spectrum. Then the second with the first, and so on.
The procedure can be schematized thus:

Mod[t] = exp(C1*log(Mod1[t])+(1-C1)*log(Mod2[t]))
Ph[t] = Ph[t-1]+exp(G*log((Ph1[t]-Ph1[t-1]))+
(1-G)*log((Ph2[t]-Ph2[t-1])))
FB[t] = exp(C3*log(FB1[t])+(1-C3)*log(FB2[t]))

Mod[t],Ph[t],FB[t] = Moduli, Phases and Frequency-Bins of the new Hybrid Component
Mod1[t], Ph1[t], FB1[t], Mod2[t], Ph2[t], FB2[t] = Moduli, Phases and Frequency-Bins of the Initial Sounds
C1, C2, C3 = Hybridization Coefficients of the Moduli, Phases and Frequency-Bins
t = Analysis Frame Index

Another algorithm was also implemented: spectral segmentation was applied only to magnitude spectrum. The phases in the frequency-bins sharing the same number were only interpolated.

Mod[t] = exp(C1*log(Mod1[t])+(1-C1)*log(Mod2[t]))
Ph[t] = G*Ph1[t]+(1-G)*Ph2[t]
FB[t] = C3*FB1[t]+(1-C3)*FB2[t]

G = C2*FC
FC = f(abs(Mod1[t]-Mod2[t]))

In both cases, problems occurred when two components with very different energies where matched. The amplification of low energy and instable components when matched with some high energy components produced noise resulting in inharmonic hybrids with most of their perceptive features destroied. In order to avoid this problem, the frequency content of the stable component was always given better result. For this purpose, a new multiplicative coefficient (FC, Frequency Corrector), function of the energy dissimilarity between the two components was introduced and used to correct C2.

4. Future directions in research into perceptive continuity

A work on hybridisation must confront the problem of the route of perceptively continuous transformation between any pairs of sounds, this being a fundamental quest. This involves seeing whether it is possible to identify an algorithm for the production of hybrids by locating the perceptive distance between any two points in a continuum of timbral space, or whether, a certain degree of quantization must be accepted.
In certain studies published up until about ten years ago, dubious arguments about the existence of a perceptive continuity can be found, while in the more recent literature a different position is assumed and the reasons for previous scepticism are explained.
Above all, the complexity of the principal mechanisms underlying perception and the difficulties of research into hybridisation methods research which do not propose restrictive hypotheses on the sound material to be elaborated, have been better explained and clarified.
Seeking to identify what the discriminating elements may be according to both the Ecological Psychology and the Information Processing approaches [1], we can see that the following, which induce the construction of a timbre space, play a fundamental role:

1) Property of resonance
2) Characteristics of the attack (duration and nature of the transient component)
3) Modulation laws in the sustain phase
4) Synchrony of fluctuations in frequency in each spectrum component.

For the purposes of recognition each of these elements can also, within limits, substitute for the others. The characteristics of attack (duration and attributes of the transient component), for example, are of great importance, often more so than the law of amplitude modulation. If the attack were cut, the latter attributes could play, for some kinds of instruments, a determining role for the purpose of perception and the recognition capability.
As a result of these observations we can see how movements within the timbre space producing a gradual attenuation of one of the above attributes do not lead to the occupation of points in the perceptive space in which our auditory system is totally unable to effect recognition.
The important thing is that the transformation executed on the material does not result in the complete destruction of the most important attributes of the sounds used. Some attributes which characterise one sound or the other should always be perceptible.

As well as this capacity of approximation, our perceptive system has another capacity which is well described in most psychoacustic essays. This makes a further contribution to the possibility of the existence of a perceptive continuum. It is concerned with the brain's ability to reconstruct visual and sound images.
Working on a sound image in such a way as to transform some of its perceptive attributes, the capacity for reconstruction and classification is not completely lost; rather, the whole image is recomposed, within limits, even when there is alteration or the elimination of some perceptive elements of secondary importance.
A correct method for executing the process of hybridisation can be considered that of using a classification of perceptive attributes, according to at least two different classes corresponding to the different levels of importance, and to process in synchrony and in inverse proportion, the attributes belonging to the same level in the two sounds. To take an example, let us suppose we want to hybridize a trombone with a voice.
The classification can be the following:

First level attributes:

Trombone:	Attack characteristics
	Laws of amplitude modulation
Voice:	Property of resonance
	Laws of amplitude modulation

Second level attributes:

Trombone:	Property of resonance
	Synchrony of frequency fluctuation
Voice:	Attack characteristics
	Synchrony of frequency fluctuation

In each point of the transformation it is necessary to take care not to destroy the first level characteristics of the two sounds. If, for example, the process of transformation begins with the trombone and ends with the voice, the attack characteristics of the trombone should not be completely destroyed before the resonant properties of the voice has clearly emerged.
The elaboration can be executed in two phases:

1) Classification of the attributes and weighting according to an estimation of the perceptive importance
2) Cross-processing between pairs of attributes of homogeneous importance

For the resolution of 1), one can consult tables such as those of Krimphoff [2].

The procedure, in the case of the hybridisation of a trombone with the human voice, can be summarised as follows:

First level weights:

Trombone: W11, Attack characteristics
W12, Laws of amplitude modulation
Voice: W13, Resonance property
W14, Laws of amplitude modulation

Second level weights:

Trombone: W21, Resonance property
W22, Synchrony of frequency fluctuations
Voice: W23, Attack characteristics
W24, Synchrony of frequency fluctuations

The transformation should correspond to a cross-processing between the pairs of attributes, while the perceptive presence of associated attributes must follow a proportional law as follows:

Wi1*HCoef+Pi2*(1-HCoef)
Wi3*HCoef+Pi4*(1-HCoef)

i = 1,2 Level of perceptive importance
HCoef = 0,1 Coefficient of Hybridisation

Final Notes:

The software for the hybridisation of the cited experiments is called Morph and was designed at LIM, Laboratorio di Informatica Musicale, Computer Science Department of the University of Milan, under the direction of Goffredo Haus and in collaboration with Civica Scuola di Musica of Milan. MORPH runs on Windows 95.

I would like to thank Goffredo Haus for the invaluable help and the support he gave me, Stephen McAdams for his suggestions and helpfulness, and Giovanni Cospito for his many insights.

References:

[1] "Thinking in Sound: the cognitive psychology of human audition", ed. S.McAdams and E.Bigand, Oxford University Press, 1993

[2] "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes", S.McAdams, S.Winsberg, S.Donnadieu, G.De Soete, J.Krimphoff, Psychol Res (1995) 58: 177-192, Springer-Verlag

[3] "Hearing musical streams ", S. McAdams, A. Bregman, CMJ, 3(4), 26-43

[4] "Timbre Morphing of Sounds with Unequal Numbers of Features", CERL Sound Group, University of Illinois, Urbana, USA, J. Audio Eng. Soc., Vol. 43, N. 9, 1995, September

[5] "Speech Analysis/Synthesis Based on a Sinusoidal Representation", R.J.McAulay, T.F.Quatieri, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.34, No.4, August, 1986

[6] "Extending the McAuley-Quatieri Analysis for Synthesis with a limited Number of Oscillators", K.Fitz, W.Walker, L.Haken, in Proc. 1992 Int. Computer Music Conf. (San Jose, CA), International Computer Music Association, San Francisco, CA, pp.381-382

[7] "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition", X. Serra, J. Smith III, Computer Music Journal, Vol. 14, No. 4, Winter 1990

[8] "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis", R.E.Crochiere, IEEE Trans. on Acoustics, Speech, and Signal Proc, VOL.ASSP-28,No.1, February, 1980

[9] "Analysis and Synthesis of Tones by Spectral Interpolation", M.H. Serra, D. Rubine, R. Dannenberg, J. Audio Eng. Soc., Vol. 38, N.3, 1990, March

[10] "Analysis and Synthesis of Musical Transitions Using the Discrete Short-Time Fourier Transform", J. Strawn, J. Audio Eng. Soc., Vol.35, N. 1/2, 1987, January/February

[11] "Analysis, Transformation, and Resynthesis of Musical Sounds with the Help of a Time-Frequency Representation", D.Arfib, in Representation of Musical Signals, MIT Press, 1991

[12] "Musical Transformation Using the Modification of Time-Frequency Images", D.Arfib, N.Delprat, CMJ, Vol.17, No.2, Summer, 1993

[13] "Rapid Measurement of Digital Instantaneous Frequency", IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP-23, N.2, April 1975

[14] "A Comparison of Computational Methods for Instantaneous Frequency and Group Delay of Discrete-Time Signals", J. Audio Eng. Soc., Vol.46, N.3, 1998 March

Automatic recognition of musical events and attributes in singing

Carlo Drioli
adrian@dei.unipd.it

Gianpaolo Borin
borin@dei.unipd.it

Centro di Sonologia Computazionale
Dipartimento di Elettronica e Informatica
Università degli Studi di Padova
Via S. Francesco 11
35121, Padova, Italy

Abstract

A framework for automatic analysis and segmentation of singing is presented. In order to provide a complete description of the signal for post-processing purposes, different levels of sound and performance analyses are integrated. A note-level description giving the position of notes within the signal, and a higher level musical attributes (like vibrato or glissando) description, are produced by means of a class of algorithms based on pitch analysis and score-performance matching. Examples of singer performance analysis is given where the note positions and durations, as well as higher level musical attributes, are recognized.

1 Introduction

Recently some powerful post-processing techniques have been presented which are well suited for voice and monophonic instruments analysis and resynthesis, and lead to attractive prospective in the field of expressive processing. Sinusoidal modeling of sound, for example, proved to be suitable for this kind of application ([1, 2, 3]). In fact, signal spectral modeling and basic modifications such as time stretching and pitch shifting can be effectively performed, while preserving the original quality of sound. These tools represent a good framework for expressive processing of digitally recorded performances. However, performing time stretching, pitch shifting or dynamic variations on musical notes or phrases is not a straightforward task. The related problems are mainly due to the continuos modulations in musical performance (*vibrato*, *portamento*, grace-notes and other high-level musical attributes), and to speech articulation (voiced/unvoiced nature of frames).

Some research on the modeling of continuos expression in performance can be found in [4]. Here, a formalism to represent continuos pitch, timing and dynamic modulations is presented, based on a decomposition of envelopes in terms of trigonometric functions or geometric figures.

The analysis system proposed here is intended to be integrated in an expressive processing framework ([5]), with the aim to produce a reliable musical description, thus permitting correct expressive manipulations and high quality re-synthesis of sound. The research focuses on singing voice, but monophonic and quasi-harmonic sounds such as wind instruments and solo string instruments have been considered as well.

The paper is organized as follows. First, a multi-level segmentation of the signal is proposed, which is a complete representation of musical events, from note level to high-level musical attributes like *staccato-legato* or *vibrato*. This allows the definition of a joined description of sound and performance. Next, the analysis step required for segmentation are illustrated, which involve the detection of performance higher level attributes, derived from basic parameters such as pitch, dynamic envelopes and timings. The higher level analysis is based on a class of algorithms for automatic recognition of musical events. Finally, an example of the analysis process applied to a pop-tune performed with different timing intentions is presented, in order to illustrate the procedure.

2 Sound analysis and multi-level description

The analysis performed is mainly based on pitch estimation of the signal. Thus, a robust analysis tool is required in order to achieve reliable results. Moreover, the results of our analysis are to be employed in a general manipulation framework, and share some of the analysis steps with the "analysis-resynthesis" subsystem. For these reasons it is a natural choice to rely on the analysis environment described in [3].

2.1 Multi-level description

The performance modification section (expressiveness model) requires a symbolic description of the signal. The extraction of this musical information can be considered at different levels. We will

refer here to a multi-layer structure composed by three layers, organized as follows:

- *Layer I*: Note-level segmentation. For this level a musical score-like representation of the signal is produced by means of a score matching procedure, which identifies the position, duration and pitch of the notes in the performance.

- *Layer II*: High-level musical attributes segmentation. As high level musical attributes, we consider pitch-related events (such as *vibrato* or *portamento*), timing-related events (such as *rallentando* or syncopation), and amplitude-related events (such as *tremolo* or accents). Timbre attributes should be considered in the second layer too, and will be investigated in future research.

- *Layer III*: Phonetic segmentation. The phonetic segmentation, which can be obtained with a speech recognition tool, or made by hand, will distinguish between consonants and vowels. Consonants will be further classified in pitched consonants (nasals and liquids), and non-pitched consonants (fricatives, plosives). This classification will clearly help the processing block to apply the correct processing algorithm for that frame. It shoulb be pointed out that comparing to the speech recognition phonetic segmentation, a fairly less amount of effort is required since we are interested in a gross classification of class of phonemes.

The importance of a high-level description of the performance has been recently pointed out in [2], where a detailed spectral description format is proposed for encoding all possible musical events. In order to automatically produce an high-level description analysis, part of the present work focuses on automatic recognition of events in *layer I* and *layer II*.

3 Automatic recognition of musical events

Since automatic multi-level musical description is our final goal, a class of procedures is proposed for detection of high-level musical attributes (*layer II*) and note onset and duration (*layer I*). The intention is to produce the *layer I* segmentation by means of a score-performance matching procedure, based on pitch information. This operation links note-events in a musical performance to the corresponding events in the score ([6]). We assume that the performer recorded his performance while listening to a musical accompaniment and by reading a score. With these constraints the score matching can be reasonably performed by cross-correlating the real pitch with the nominal pitch given in the score. After a first global alignment, the cross correlation window is progressively reduced (we can process musical phrases first, then semi-phrases, and so on), thus refining the local tempo alignment. The last step will refer to a single-note observation window, and at the end the local time alignment of note i is given by $\tau_i = \sum_{k=1}^{K} \tau_{i,k}$ ($\tau_{i,1}$ is the time index for the maximum of global cross-correlation, $\tau_{i,2}$ refers to the first phrase sub-division, and so on). When this step is completed, performed and nominal pitch are aligned and one can think of detecting the onset of the notes by comparison of the two. However, the pitch extracted by a singer performance is characterized by the presence of non-significant values corresponding to unvoiced frames, as well as by the presence of modulations corresponding to *vibrato*, glissando, or other musical attributes. Both aspects are responsible for a pitch curve that is quite dissimilar from the nominal pitch of the score. Thus, in our approach we first perform the detection of the high level pitch events for which an estimate of the nominal pitch can be easily obtained: vibrato, micropause (non-written rests), attack, etc. The same can be done for frames where pitch is not defined due to phonetic reasons. Once the pitch profile has been 'cleaned' from the influence of higher level attributes, the local analysis of onset and duration of notes will result sensibly improved (see figure 1).

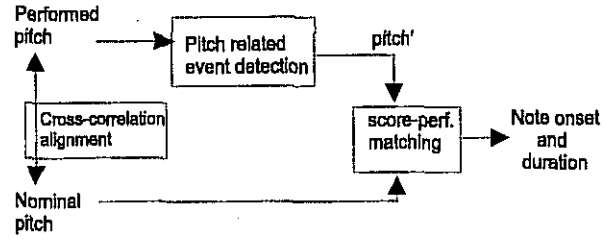


Figure 1: Schema of note onset and duration evaluation procedure

3.1 Basic high-level musical attributes detection

We will illustrate the high level musical attributes detection procedure, by considering an example of a singing phrase where a number of typical features are present. After the local tempo alignment, we may reliably estimate the desired pitch for the actual note under observation, and for the next note (under the assumption that there are no missing or added notes in the performance with respect to the score). A pitch tolerance region [1] is localized around each nominal value of the pitch, and it is now easy to evaluate the degree of fitting of the performed pitch whithin the two zones. This information will be used in order to delimitate the sustained part of the note, and to study the nature of the transition

[1] A ±80 cents band is adequate for most pop performances

from the actual note to the next note. The construction of the auxiliary pitch, *pitch'*, which is an estimate of the nominal score underlying the performed pitch, is performed by using the phonetic segmentation and by detecting some basic pitch related features. Figure 2 illustrates the construction of the auxiliary pitch from the performed pitch with respect to the class of considered features, which are summarized below.

- *Phonetics*: the main question with phonetic features is how to manage the duration of unvoiced frames when they separate two notes (as it is often the case). This is an important point in order to correctly interprete the timing of the sung performance. We can refer to some results obtained in the field of analysis and synthesis of singing ([7, 8]), which states that all consonants must be considered as part of the note of the preceding vowel. Thus, in the construction of the auxiliary pitch, the previous note's pitch will be extended to cover the duration of the consonant. For pitched consonants, the duration can be shared by the preceding and the actual note, depending on the nature of the pitch.

- *Pitch-related attributes*: we consider here a basic set of pitch-related attributes, namely *vibrato*, pitch attack, pitch decay, and *portamento*. In [5] a procedure has been proposed for vibrato detection, which relies on a STFT analysis of the pitch profile. This analysis can reliably highlight the portions where the pitch trajectory presents a sinusoidal behaviour, with a frequency lying in the vibrato range (i.e., 4.5 to 7 Hz). Attack, release and *glissando* can be detected by observing the regions where the pitch lies outside the tolerance zones. Decision on the nature of the attribute can be made by observing the derivative of the pitch. As for phonemes, the duration of attack and *glissando* is assigned to the preceding note (or pause).

Dynamic-related attributes and tempo-related attributes can be considered in a similar manner. A method based on perceptive temporal masking has been proposed in [5] for *staccato-legato* detection and a procedure similar to the one used for the vibrato detection can be used for *tremolo*. The dynamic-related attributes analysis can be done, as pitch-related attribute analysis, before the onset detection procedure. On the other hand, the recognition of tempo-related attributes, such as *accelerando*, *syncopation* and others, have to be performed necessarily after that the note onset and duration detection has been completed.

3.2 Score-performance matching

The decision on the i-th note onset time is taken by observing the derivative of the auxiliary performed pitch, *pitch'*, on the neighborhood of the expected onset time (which is $t_i = t_{nom,i} + \tau_i$, where $t_{nom,i}$ is the score onset time and τ_i is the time delay computed in the local tempo-alignment procedure). A gaussian-shaped observation window centered in t_i is used to weigth the derivative of *pitch'*, and a selection of the maximum weighted value with correct sign can be performed (the sign expected is given by the rising or descending pitch in the score).

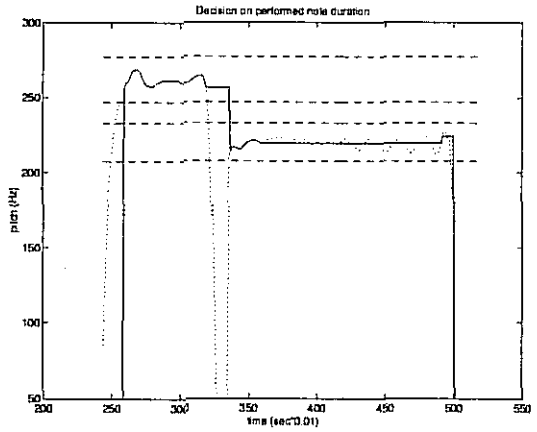


Figure 2: Performed pitch (dashed) of the sung phrase "Hey Jude", and auxiliary pitch (solid). The two 80 cent pitch tolerance zones are evidentiated (dashed horizontal lines).

Next, the score-performance procedure is applied to two different performances of the same musical phrase (figure 3): in the first case, the singer's intention was to respect the nominal onset and duration of each note, while in the second case a slightly less constrained performance was produced with respect to timing. By looking at the figure is clear that the second performance presents delayed onsets with respect to the score, as well as some note duration changes.

After the second layer segmentation procedure application, the auxiliary pitch signal is produced and a matching with the score pitch is performed in order to extract the onset and the duration of the notes. Figure 4 show the results obtained in the two cases exposed above.

In the upper figure, the attack delay of each note is reported. The solid line, referring to the "sharp" performance, exhibits little deviation from zero, while the dashed line, referred to the "loose" performance, clearly shows a higher amount of delay from the nominal score, in good agreement with figure 3 and with listener's perception. In the lower figure, the duration difference of each note is reported. The same convention is adopted for the "sharp" and the "loose" performance and the data can be once more compared to the performed pitch in figure 3. As a final note, we can be observe that the dashed curve of figure 4 (duration difference), can be used in

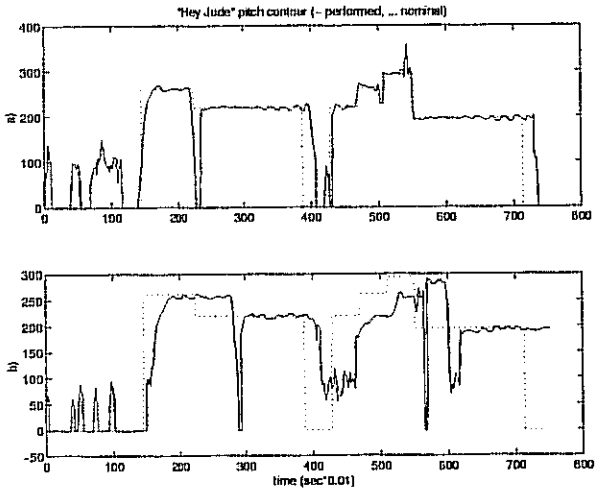


Figure 3: Two performances of the same sung phrase from the pop-tune *"Hey Jude"* (the pitch contour is shown).

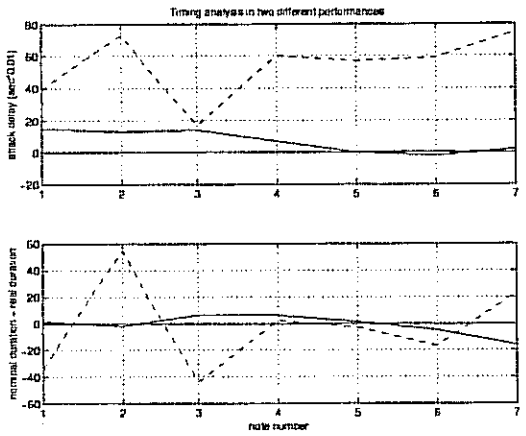


Figure 4: Onset of notes (upper figure) and note duration difference (lower figure) for the two performances (solid line: performance a); dashed line: performance b) of figure 3.)

order to study time-related musical attributes. The sinusoidal shape of this curve, for example, can be a clear indication of syncopation attributes.

4 Conclusions

The basic elements of a monophonic signal analysis framework are presented here. The system moves toward a completely automatized description of the signal, and a set of procedures for the recognition of the most common high level musical attributes is proposed. While the framework was primarly intended to be integrated in an expressiveness analysis and reynthesis system, it is designed to provide a joined sound and performance description of the signal. Thus, this system can be usefully employed as an aid in the field of musical style analysis.

Acknowledgments

This work has been supported by *Telecom Italia S.p.A.* under the resarch contract *"Cantieri Multimediali"*.

References

[1] J. L. Arcos et al., 1997. "SaxEx: a case-based reasoning system for generating expressive musical performances", in *Proc. ICMC97*, Thessaloniki, Greece, pp.329-336.

[2] X. Serra et al., 1997. "Integrating Complementary Spectral Models in the Design of a Musical Synthesizer", in *Proc. of ICMC97*, Thessaloniki, Greece, pp. 152-159.

[3] R. Di Federico and G. Borin, 1998. "An Improved Pitch Synchronous Sinusoidal Analysis-Synthesis Method for Voice and Quasi-Harmonic Sounds", in *Proc. of ICMC98*, Gorizia, Italy, to appear.

[4] P. Desain and H. Honing, 1995. "Towards Algoritmic Description of Continuous Modulations of Musical Parameters", in *Proc. of ICMC95*, San Francisco, pp. 393-395.

[5] R. Di Federico and C. Drioli, 1998. "An integrated System for Analysis-Modification-Resynthesis of Singing", in *Proc. IEEE SMC98*, San Diego, California, to appear.

[6] P. Desain et al., 1997. "Robust Score Performance Matching: Taking Advantage of Structural Information", in *Proc. ICMC97*, Thessaloniki, Greece, pp.337-340.

[7] G. Berndtsson, 1996. "The KTH Rule System for Singing Synthesis", in *Computer Music Journal*, 20:1, MIT, pp. 76-91.

[8] J. Sundberg, 1987. "The Science of the Singing Voice", Northern Illinois University Press, Dekalb, Illinois.

TIMBRE NUANCES OF THE ACOUSTIC GUITAR AND THEIR RELATION WITH THE PLUCKING TECHNIQUES

Nicola Orio
CSC-DEI Universitá di Padova
Via Gradenigo, 6/A, 35121, Padova, Italy
orio@dei.unipd.it

Abstract

A number of different sounds produced by an acoustic guitar, plucked with fingers, were digitally recorded. Samples have the same pitch and loudness, they differ only in the plucking technique. The timbre nuances obtained by the different techniques were analyzed using: time-frequency techniques, psychoacoustical analysis and Principal Component Parametrization of Mel Frequency Cepstral Coefficients. The aim was to investigate the timbre space of the instrument and to build a model for the automatic recognition of the plucking technique, to be used in a gesture interface. The main acoustic parameters were pointed out, in agreement with the theory on the vibrating string. Moreover it was shown that it is possible to give an estimate of the plucking technique from timbre analysis.

1 Introduction

Acoustic musical instruments are able to produce a number of distinct sounds that are characterized by small differences in timbre. Great part of musicians' training is spent in learning how to obtain different timbre nuances, or colors, from the instrument, and how these nuances are related to the playing technique. The control of sound, that is the control of its timbre, is the basis of any performance. It is well known that timbre is a characteristic of sound hard to study, because of its intrinsic complexity: timbre has a multidimensional feature, which varies in time. Even the common definition of timbre is somehow evasive: "timbre is the characteristic that permits to recognize differences in two sounds when pitch and loudness are equal".
Some studies were carried out on the timbre of different musical instruments. Gray [1] developed a study about the relation between the timbre characteristics of a set of sound samples and the subjective impressions of a group of listeners. Other works focused on the creation of a timbre space through sound parametrization and mapping techniques, like Neural Nets [2] and multivariate statistical analyses [3]. All these studies are about the timbre space of a group of different instruments, that is they are about the recognition of different instruments rather than the recognition of the nuances of a single instrument. In this paper it is presented a study on the space of these timbre nuances produced by an acoustic nylon string guitar, plucked by fingers with nails, like in the common technique taught in Music Conservatories. Two reasons suggested that the acoustic guitar could be a good testbed for this kind of analyses. From the one hand the guitar can produce a great variety of different timbre nuances, but at the same time it has a typical sound, always recognizable by listeners; hence timbre variations are inside some perceptual boundaries. From the other hand, guitar's timbre nuances mainly depend on the interaction between the string and the right hand's fingers, which is easily observable and measurable. Moreover the guitarist has a number of different plucking techniques that she can apply independently one from the other, as well as in different combinations.

2 Sound Material

Sounds were obtained by a classical spanish guitar, a First Class Ramirez of the 1988, with nylon strings. To avoid the problem of possible resonances, all the strings but the fifth (A at 110 Hz) were taken away. The pitch and loudness were the same for each sound: D at 150 Hz, played on the fifth fret, recorded at -10 dB level. Sounds were digitally recorded at a sampling rate of 48 kHz, with a DAT connected to a AKG cardiod microphone placed at a distance of 20 cm from the guitar hole.
The different timbre nuances were obtained by slowly changing the plucking technique. The variations of the plucking technique were made considering a *normal* position and inclination of the finger, seven centimeters away from its vibrating center and orthogonal to the guitar body. Samples were obtained by changing respectively:

- the finger's position along the string. Twentyeight samples were recorded, each one moving the finger of 1 cm from the 12th fret to the bridge; the finger had the *normal* inclination.
- the inclination between the finger and the string, moving the finger in a plane orthogonal to the guitar's body and parallel to the string.

Seven samples were recorded, each one changing the inclination of 15°, from −45° to +45°; the finger was in the *normal* position along the string.

- the inclination between the hand and the string, moving the finger in a plane orthogonal to the guitar body and to the string. Seven samples were recorded, each one changing the inclination of 30°, from −90° to +90°; the finger was in the *normal* position along the string.
- the degree of relaxation of the plucking finger. Seven samples were recorded, each one progressively decreasing the finger's tension, from *strappato* to *appoggiato*; the finger was in the *normal* position and inclination on the string.

A number of different sounds, obtained by mixing the different techniques, were added to these samples. Moreover they were sampled also seven more sounds, obtained by changing the rotation of the finger around its axis; since the differences among samples were not relevant, both perceptually and acoustically, they were neglected.

3 Analyses

All the sets of different samples were analysed, using time-frequency domain techniques, psychoacoustical parametrization, and Mel Frequency Cepstral Components extraction and mapping through Principal Components Algorithm.

3.1 Time-Frequency Analysis

Each guitar sound were analysed in ten different portions of the signal, spaced of 100 ms and beginning at the onset time, to test if there are relevant differences in timbre depending on the evolution of the signal. The Short Time Fourier Transform was used: the signal was windowed with a Hamming window 1024 points long.
The degree of inharmonicity of the sounds was the first measured parameter. As it is known a certain inharmonicity, with slightly sharper harmonics, is typical for the non-ideal string. The variations of this parameter were found significantly relevant only in relation to the finger position along the string.

Pos	2nd	9th	16th	23rd	30th
0	2.008	9.133	16.258	23.312	30.368
9	2.037	9.145	16.305	23.321	30.459
18	2.038	9.157	16.289	23.311	30.401
27	2.056	9.422	16.716	24.023	31.185

Table 1: Harmonic-to-fundamental ratio for some of the samples, obtained by moving the finger along the string; distance are measured in cm from the center of the string

In Table 1 the harmonic-to-fundamental frequency ratio is quote for some of the harmonics. It can be seen that this ratio increases when the finger plucks the string closer to the bridge. Other techniques were not found related to this parameter: the variations of inharmonicity are about 1/10 smaller and they probably are affected by noise, since a random trend was observed.
The measure of the harmonics amplitude shows that this parameter is related to the changes in all the different techniques. In particular moving the finger towards the bridge enhances the higher harmonics, coherently with the theory on plucked strings. Moreover the *normal* position has a high-pass spectrum if compared to the samples obtained by changing both finger and hand inclinations. Only the degree of finger's relaxation seems to have a slight correlation with the harmonics amplitude.
Attack time of the fundamental were found a good discriminant for the *normal* position in respect to hands movements. It has an attack time of 6.7 ms, while all the others are inside the range from 3.6 to 4.3 ms. This can be explained considering that an hand inclination greater that zero (in both direction) excites the resonant mode orthogonal to the guitar body, which transmits its energy to the bridge faster than the resonant mode parallel to the body (that is the only one directly excited when the string is plucked in the *normal* position). Changes in other techniques were not significantly related to the attack time, even if it was observed that samples obtained with finger at 4 cm, or less, to the bridge have a very fast attack time, from 2.4 to 2.7 ms.

3.2 Psychoacoustical Analysis

Another kind of analysis were carried out, for the extraction of the so-called psychoacoustical parameters [4]. Two parameters were calculated, the Center of Gravity of the Spectrum (CGS) and the Irregularity of the Spectrum (IRR), defined by the two formulas:

$$CGS = \frac{\sum_{k=1}^{N} k A_k}{\sum_{k=1}^{N} A_k}$$

$$IRR = \log \left| \sum_{k=2}^{N-1} 20 \log \frac{A_k}{\sqrt[3]{A_{k-1} A_k A_{k+1}}} \right|$$

where, in both equations, k is the number of the harmonic and A_k is the amplitude of the k-th harmonic. Results on CGS were coherent with the ones obtained from time-frequency analysis on harmonics amplitude. CGS, which varies from 1.7 to 6.2, has a linear trend when the finger is moved along the string; while, when the finger or hand inclinations are changed, it has a symmetric trend, almost parabolic with the maximum (CGS=4.7) in the *normal position and the minimums at the highest inclinations*. These two symmetric trends can be explained considering that, if the nail is symmetric, the timbre nuances depend only on the angle absolute value, not on its sign. CGS parameter is also

related to changes in finger degree of relaxation. Results are quoted in Figure 1. As it can be seen CGS decreases when the degree of relaxation increases. Changes in IRR were found significantly relevant

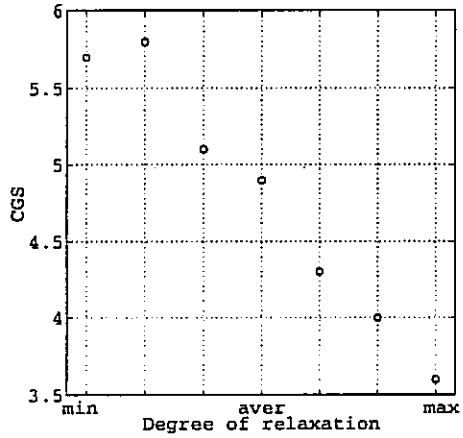


Figure 1: Values of CGS when the string is plucked with an increasing degree of relaxation of the finger

only when the finger is moved along the string. It has a maximum value (IRR=2.44) when the string is plucked in the middle, that is with maximum distance from the bridge; then it decreases to its global minimum (IRR=1.78) with the finger close to the bridge. IRR is almost randomly distributed in the interval from 2.03 to 2.21 when other plucking techniques are changed.

3.3 Principal Component Parametrization of Mel Frequency Cepstral Coefficients

The last used analysis technique was the extraction of the Mel Frequency Cepstral Components (MFCC), introduced in [5]. Its basic idea is to perform the cepstral analysis in *mel*, rather than in hertz, scale. Mel scale approximates how the frequency is perceived, hence it is linear with hertz below 1 kHz and then logarithmic. So the signal was filtered by a bank of triangular filters equally spaced in mel scale, obtaining a set of log-energies X_k, one for each filter. The set of log-energies is then transformed with the formula:

$$MFCC_i = \sum_{k=1}^{N} X_k \cos[i(k - \frac{1}{2})\frac{\pi}{N}]$$

where N is the number of triangular filters. In this analysis the filters were spaced by 150 mel, hence centered on the frequency of the harmonics under 1 kH; they were evaluated 30 coefficients.
It was chosen to reduce the redundancy of the information contained in these coefficients using a multivariate statistical technique [6] known as the Principal Component Algorithm (PCA). With this technique it is possible to map the space of MFCC parameters in a space with a smaller dimension and with uncorrelated axes. PCA performs a projection of the coefficients in this space giving the percentage of the global variance in the parameters explained by each new axis.
The samples, obtained by the plucking techniques, were separately mapped in four different spaces. Results were similar for all the techniques. In all the cases a single component was enough to explain more than 70% of the global variance; the components after the first were not statistically relevant. Results for each single technique are quoted in Table 2. The

Techniques	1st Comp.	2nd Comp.
Finger movement	72.8%	7.9%
Finger inclination	78.6%	13.3%
Hand inclination	87.4%	6.7%
Finger relaxation	77.3%	11.4%

Table 2: Percentage of global variance explained by the first and the second principal components, separately calculated on the different plucking techniques

fact that a single component is enough to explain most of the global variance, is coherent with the fact that, in each case, it was varied only one performing parameter. Hence the first principal component can be, probably, related to the effect in spectrum of the variations in the plucking techniques. To test this hypothesis, the MFCC can be projected, through PCA, in the first axis of the new space. In most of the cases it was found a good correlation among the positions of the samples on this axis and the variations in the plucking technique. In Figure 2 the dif-

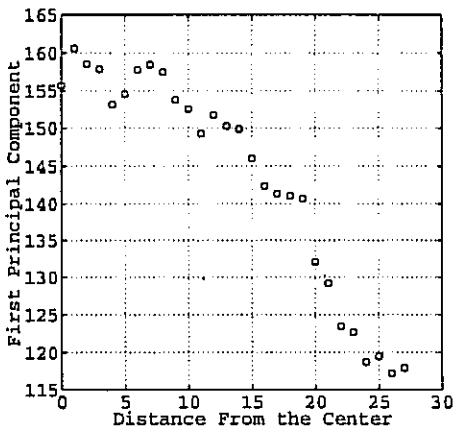


Figure 2: Projections along the First Principal Component of samples obtained moving the finger along the string

ferent samples obtained moving the finger along the string are plotted. How it can be seen the value of the first component decreases when the finger moves toward the bridge.
This correlation is maintained also when both finger and hand inclinations are varied. Results are respectively plotted in Figure 3 and in Figure 4. In both

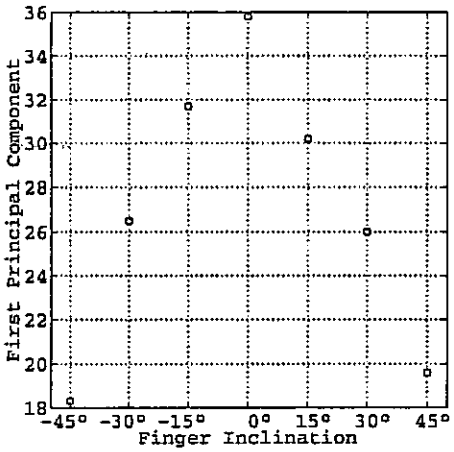


Figure 3: Projections along the First Principal Component of samples obtained changing the finger inclination

cases the first principal component has values symmetrical around the *normal* inclination. Also in this case this can be explained considering that the effect of these two techniques in timbre nuances depends only on the absolute value of the angle, not on its sign. The analysis developed on samples obtained by changing the degree of relaxation gave a trend similar to the one previously shown in Figure 1. Finally

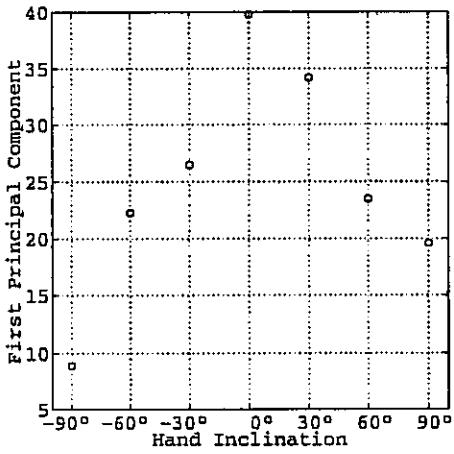


Figure 4: Projections along the First Principal Component of samples obtained changing the hand inclination

the whole analysis process, MFCC and PCA, was developed on the whole group of samples, including the ones obtained mixing different techniques. Unfortunately results were not as expected. Also in this case the weight of the first component is predominant, it explains 55.6% of the global variance while the second and the third components explain, respectively, 14.8% and 9.2%. It is interesting to note that each technique maintains its trend, if mapped on the first axis. Samples obtained with mixed techniques map coherently with the techniques used, even if they are more affected by changes on the finger position along the string. In order to separate the different timbre nuances, the samples were mapped on the plane of the first two principal components. Results were affected by the low variance explained by the second component: different techniques map close in the plane, even if it is possible to separate samples obtained by moving the finger along the string and samples obtained by varying the degree of relaxation of the finger. Changes on inclination, in both direction, are hard to discriminate.

4 Conclusions

A group of guitar sounds, played changing in four different ways the plucking techniques, were analysed using different techniques. The possible variations of guitar timbre were highlighted in relation to the different plucking techniques. From analyses it emerged that it is also possible to recognize the plucking technique from the sound analysis. The system needs some improvement when more plucking techniques are applied together. Nevertheless results show that it is possible to develop a recognizer of guitar sound nuances, which can be used as a new gesture interface for MIDI guitars.

References

[1] Grey, J.M., Moorer, J.A., "Perceptual Evaluations of Synthesized Musical Instruments Tones", *Journal of Acoustic Society of America*, Vol. 62(2), pp. 454-462, 1977.

[2] Cosi, P., De Poli, G., Prandoni, P., "Timbre Characterization with Mel-Cepstrum and Neural Nets", in *Proc. of International Computer Music Conference*, Aarhus, Denmark, pp. 42-45, 1994.

[3] Boatin, N., De Poli, G., Prandoni, P., "Timbre Characterization with Mel-Cepstrum: a Multivariate Analysis", in *Proc. XI Colloquium on Musical Informatics*, Bologna, Italy, pp. 145-148, 1995.

[4] Krumhansl, C.L., Why is musical timbre so hard to understand?, in *Structure and perception electroacoustic sound and music*, in S. Nielsen and O. Olsson (ed), Amsterdam, pp. 43-53, 1989.

[5] Davis, S.B., Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", *IEEE Trans. on Acoustic, Speech, and Signal Processing*, Vol. 28(4), pp. 357-366, 1980.

[6] Beyerbach, D., Nawab, H., "Principal Component Analysis of the Short Time Fourier Transform", in *Proc. of International Conference on Acoustic, Speech, and Signal Processing*, Vol. 3, pp. 1725-1728, 1991.

A DIGITAL DELAY LINE BASED ON FRACTIONAL ADDRESSING

Davide Rocchesso
Dipartimento Scientifico e Tecnologico, Università di Verona
Strada Le Grazie, Verona, Italy
rocchesso@sci.univr.it

Abstract

A novel realization of the digital delay line is proposed. It uses a memory buffer with a single fractional pointer rather than a reading and a writing pointer. The proposed method is about as efficient as the popular interpolated circular buffer, but it offers better performance in terms of frequency-dependent attenuation and response to delay-length modulations.

1 Prior Art

The classic implementation of the digital delay line uses a circular buffer, which is accessed by a writing pointer followed by a reading pointer [1]. When the delay length has to be made variable, the relative distance between the reading pointer and the writing pointer is varied sample by sample. In order to allow for fractional lengths and click-free length modulation, some form of interpolation has to be applied at the reading point [2, 3, 4]. The following properties should be ensured by the interpolation device:

1. flat frequency response
2. linear phase response
3. transient-free response to variations of the delay.

FIR filters, usually in the form of Lagrange interpolators [3], are widely used. Even though they can not satisfy property 1, they are certainly compliant to property 3 and can also satisfy property 2 to a great extent on a wide frequency range [3]. On the other hand, allpass filters, often designed to have a maximally-flat delay response at low frequencies, satisfy property 1 exactly but are quite nonlinear in their phase response [3]. Moreover, a rather complicated structure has to be devised in order to attain property 3 by means of allpass filters [5].

All of the previous realizations, as far as a fixed delay length is considered, are linear and time-invariant systems, thus being completely described by their frequency response. Vice versa, we are proposing a realization which is time-varying even in the case of constant delay.

2 A Fractionally-Addressed Delay Line

The key idea behind the proposed realization is that of using a single pointer for both the read and write accesses. If the delay line has fixed integer length B, it is possible to use a buffer exactly B-cells long and a single pointer whose entry is first read and then written. In the same buffer we can also implement any delay which is an integer fraction B/I just by incrementing the pointer at steps of I samples. We are going to show how this scheme can be generalized to non-integer fractions of the total buffer length. The resulting technique can be seen as an extension of the table-lookup oscillator [6, 7], such as every read is followed by one or more writes, in such a way that the waveform is continuously restored while it is being read.

Given a buffer size of B samples, and a sample rate F_s, a (fractional) increment of I samples gives a delay in seconds equal to

$$D = \frac{B}{I \cdot F_s} \tag{1}$$

Since this realization is related to waveform generation by fractional addressing [7], we call it the Fractionally-Addressed Delay (FAD) line.

2.1 Realization

As far as the value being read out of the delay line is concerned, the FAD line can behave similarly to the table-lookup oscillator, being possible to apply truncation, linear interpolation, or multirate interpolation techniques [1, 2]. More complicated is the injection of a new value, to be done right after the read, in such a way that no "holes" are left in the current pass through the buffer. For instance, for $I = 2$, two writes have to be performed for every read. A fractional increment would correspond to a variable number of writes at each step. Several interpolation techniques can also be applied at the write stage. Our quantitative analysis will be conducted on a realization where linear interpolation has been used in reading and quadratic interpolation in writing.

The pseudo-C code in general form looks as follows:

```
loop
      fph = floor(phase);
      output = interpolated_read(table[fph],
             table[fph+1], ...);
      ph = (phase_old + 1) % length_table;
      while (ph <= fph) {
        table[ph] = interpolated_write(
          ..., table[phase_old], input);
        ph = (ph + 1) % length_table;
      }
      phase_old = fph;
      phase = (phase + Increment);
      if (phase > length_table)
           phase = phase - length_table;
endloop
```

Notice that the `interpolated_read` uses samples following the (phase) pointer, while the `interpolated_write` uses samples preceding the pointer.

3 Input-Output Analysis

The FAD line is a time-varying system, and therefore it is difficult to characterize in terms of frequency response. However, it is worth noticing how it responds to a sinusoidal input. Fig. 1 shows the magnitude spectrum of a delayed $8 - kHz$ sine wave when linear interpolation is used in reading and quadratic interpolation in writing. It is clear that spurious components are added to the main spectral line. The magnitude of these components might be dependent on the frequency of the input sine wave and the initial (fractional) phase of the FAD-line pointer [1]. The signal-to-noise error ratio (SNR) as a function of these two parameters shows a very mild dependence on initial phase. Therefore, it makes sense to plot the average SNR as a function of the input frequency only (fig. 2). We can see that low frequencies are affected by a high SNR, thus indicating that the FAD line has an acceptable behavior for practical sounds. Fig. 2 also shows a comparison with the linearly-interpolated (FIR) delay line. The noise error has been computed as the sum of the squared differences between the input and output waveform samples [6] [2]. The noise error is larger in the FIR case even though that implementation has no spurious components in the output spectrum. This is due to the fact that the FIR attenuation is larger on average.

It should be noticed that the spurious components tend to cluster around the main peak, thus being likely to be below the masking threshold, whose slope is known to be less than 27dB/Bark [8].

[1] As an example of dependence on initial phase, consider the increment $I = 1$. If the initial phase is 0 the pointer always falls on samples. If the initial phase is 0.5 the pointer always falls between samples.

[2] A normalizing factor $\sqrt{2/N}$ has been applied, being N the number of samples per period.

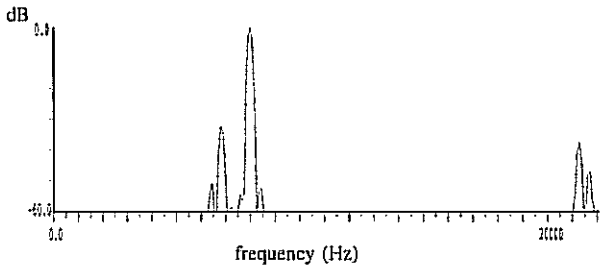


Figure 1: Magnitude spectrum of the output signal of a FAD line, where the input signal is a sine wave at 8000Hz, the delay is 0.74831s, the sampling rate is 44.1kHz and the buffer is $44100 -$ samples long.

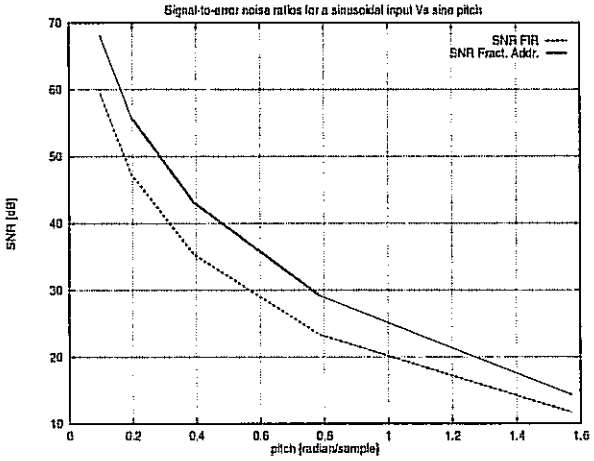


Figure 2: Signal-to-error noise ratio Vs. sine frequency for the FIR line and for the FAD line

Especially for applications such as waveguide modeling of musical instruments [9], it is important to consider the attenuation that different frequencies are subject to when fed into the delay line. The attenuation of the main peak of the output spectrum turns out to be highly dependent on the initial phase. Therefore, for the sake of comparison with the FIR line, we plot in fig. 3 the minimum, maximum, and mean attenuation as a function of the frequency of the input sine wave. Notice that at 2/3 of the Nyquist frequency the FIR line shows an attenuation of 1.2dB while the FAD line shows a mean attenuation of 0.5dB.

3.1 Behavior for time-varying delay

The FAD line shows an unconventional behavior when the delay length is dynamically varied. Suppose to vary the delay length as a linear function of time t, starting from the nominal delay τ_0 and decreasing it at the rate of k seconds per second:

$$D(t) = \tau_0 - kt \tag{2}$$

The FIR line responds with an instantaneous pitch shift in the output signal. In other words, we get a

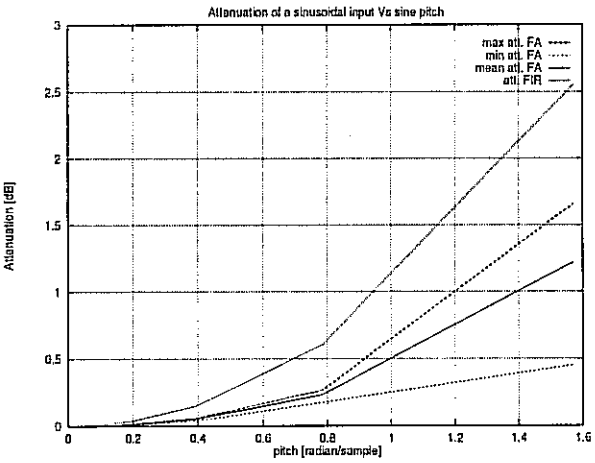


Figure 3: Attenuation of a sinusoidal input Vs. sine frequency for the FIR line and the FAD line

Doppler effect and the pitch shift is

$$\Delta f = 1 + k \tag{3}$$

On the other hand, the FAD line provides a steady pitch shift

$$\Delta f = e^k \tag{4}$$

after a transient time

$$\tau_i = \frac{\tau_0}{k}(1 - e^{-k}) \tag{5}$$

The transient time can be calculated by feeding the delay line with an impulse at time 0. It will come out of the line at the time instant τ_i such that

$$\int_0^{\tau_i} \frac{1}{D(t)} dt = 1 \tag{6}$$

The steady-state transposition can be calculated by observing that a second impulse entering the line at time T_i "sees" an instantaneous delay of $\tau_0 - k\tau_i$ seconds. It gets out of the line at time $\frac{\tau_0 - k\tau_i}{k}(1 - e^{-k}) + T_i$, exactly $T_i e^{-k}$ seconds after the impulse which entered at time 0.

If the delay ramp is applied for 1.11 seconds starting from an empty line, the response of the FAD line to a steady sinusoidal input is displayed as a sonogram in fig. 4. The transient is clearly visible in the output when the ramp is stopped. This sonogram should be compared to the one obtained with the FIR line (fig. 5).

A different behavior is also reported in response to sinusoidal modulations of the delay length. These modulations are essential for effects such as flanging or phasing. Modulations of the FAD and FIR lines are reported in fig. 6 and 7, respectively. The figures show that the FAD line is much less sensitive to artifacts clearly visible (and audible) as faint waves fig. 7, which are essentially due to amplitude modulation induced by frequency dependent attenuation.

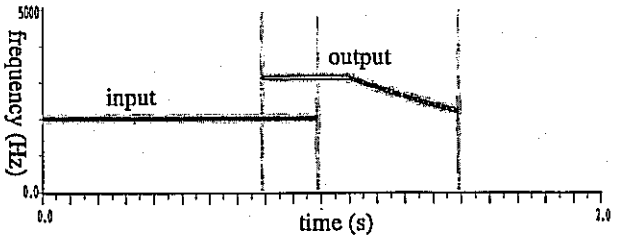


Figure 4: FAD line: delay ramp from 0.99 s to 0.5 s in 1.11 s; 1 s of sinusoidal input

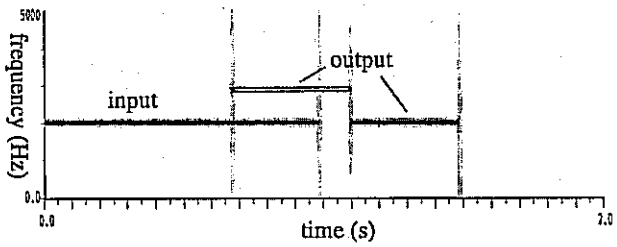


Figure 5: FIR line: delay ramp from 0.99 s to 0.5 s in 1.11 s; 1 s of sinusoidal input

3.2 Physical Interpretation

If the dynamic behavior of the delay lines is closely analyzed, we see that the FAD and the FID realizations actually simulate two different physical phenomena. In both cases, the lines can be thought of as a one-dimensional medium where waves propagate. However, when the delay length is dynamically reduced we have two physical analogies in the two cases. The shortening of the FIR line corresponds to the receiver getting closer to the transmitter, and therefore we have a tight simulation of the Doppler effect. On the other hand, the shortening of the FAD line corresponds to incresing the velocity of propagation in the medium while maintaining the same physical distance between the two ends.

4 Performance

The FAD line has only one pointer for accessing data in the buffer. It exhibits spatial locality because any short sequence of accesses spans over a small neighborhood of the pointed buffer cell. On the other hand, a FIR line has two pointers, thus exhibiting two distinct spatial localities. As a consequence, the FAD line makes better use of the cache in general purpose computer architectures. However, the FAD performs more writes than reads. In order to attain a 50% of delay variability, we have to accept up to two writes for every read. This overhead is compensated by the highest efficiency of write operations in modern architectures [10].

These two observations justify the fact that the FAD line, despite of its higher complexity, does not run much slower than the FIR line on a general pur-

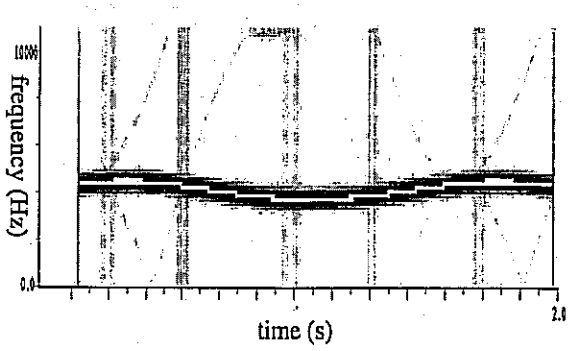


Figure 6: FAD line: delay-length vibrato

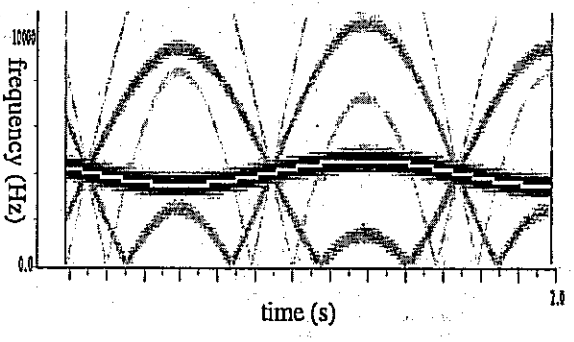


Figure 7: FIR line: delay-length vibrato

pose computer. A rough benchmark has been performed on an $AMD-K6$ architecture by repeatedly delaying a soundfile stored in an array. The first few repetitions were neglected because they seemingly involve instruction and data loading in the cache. An average of the following 10 repetitions gave us the results which are summarized in table 1. We see that the performance for the FAD line is even better than that of the FIR line if a simple interpolation scheme is used. This indicates that the proposed technique is particularly suited for modern general-purpose computers. On the other hand, the complicated control flow would make it difficult to implement the FAD line on a Digital Signal Processor.

Benchmark	
Delay Line	time (s)
FAD line, linear interpolation in write and read	1.00
FAD line, quadratic interp. in write and linear interp. in read	1.21
FIR line, linear interpolation	1.14

Table 1: Benchmark for different implementations of the delay line on a general-purpose computer

5 Conclusion

We have proposed a realization of the digital delay line which is based on an extension of the table-lookup oscillator. The proposed realization exploits the features of modern computer architectures and shows improved performance in terms of frequency-dependent attenuation and dynamic behavior. We expect this delay line will be considered as a building block for physically-based sound synthesis and for sound effects such as flangers and choruses.

References

[1] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice Hall, Englewood Cliffs, N.J., 1996.

[2] U. Zölzer, *Digital Audio Signal Processing*, John Wiley and Sons, Inc., Chichester, England, 1997.

[3] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay—Tools for Fractional Delay Filter Design", *IEEE Signal Processing Magazine*, vol. 13, no. 1, Jan 1996.

[4] S. Tassart and P. Depalle, "Analytical approximations of fractional delays: Lagrange interpolators and allpass filters", in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Munich*, Apr. 1997, pp. 455–458.

[5] V. Välimäki, *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*, PhD thesis, Elec. Eng. Dept., Acoustics Lab., Helsinki University of Technology, Dec. 1995, Otaniemi 1995 / Report 37.

[6] F. R. Moore, "Table lookup noise for sinusoidal digital oscillators", *Computer Music J.*, vol. 1, no. 1, pp. 26–29, 1977.

[7] W. M. Hartmann, "Digital waveform generation by fractional addressing", *J. Acoustical Soc. of America*, vol. 82, no. 6, pp. 1883–1891, 1987.

[8] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer Verlag, Berlin, Germany, 1990.

[9] J. O. Smith III, *Principles of Digital Waveguide Models of Musical Instruments*, vol. Applications of Digital Signal Processing to Audio and Acoustics, pp. 417–466, Kluwer Academic Publishers, 1998, M. Kahrs and K. Brandenburg, eds.

[10] V. J. Duvanenko, "Two writes make a read", *IEEE Computer*, vol. 31, no. 9, pp. 8, Sept. 1998.

Thursday 24th **h. 14.00**

DIGITAL SIGNAL PROCESSING: SOUND ANALYSIS AND SYNTHESIS II

OPTIMUM FREQUENCY WARPING OF PSEUDO-PERIODIC SIGNALS

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

Gianpaolo Evangelista

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

Abstract

Unitary transforms recently introduced by the authors, of which the Frequency Warped Wavelet Transform in its basic version and its Pitch Synchronous version are special cases, are explored as a new means for characterizing a large class of sounds. In these sounds the stiffness of the medium in which oscillations propagate results in a frequency dependent velocity and hence in a dispersive characteristic for the higher partials. . The frequency warping induced by a Laguerre Transformation modifies the structure of the partials, reverting the source sound pseudoharmonic features into a 'perfectly harmonic' sound, provided that a suitable choice of the controlling parameter is adopted. Besides the theoretical interest for the classification and definition of pseudo-periodic sounds, the adopted technique results in improvements in the characterization and separation of the periodic part from the aperiodic component of these sounds. This is performed in the domain of the frequency warped pitch synchronous wavelet transform, where the signal appears as a perfectly harmonic signal. In this paper we present a new method for matching the warping parameter to signals. This method is based on the adaptation of a normalized warped notch comb-filter choosing as a criterion that of minimizing the output energy of the filtered signal.

1 Introduction

A large class of sounds exhibits the feature, very relevant for sound quality and timbre identification, of having non uniformly spaced partials; in particular spacing increases with the partial order giving rise to a very specific pattern in the frequency domain. The observed spacing, as shown in the literature, is due to stiffnes in the medium where the oscillations propagate; stiffness results in frequency dependent propagation velocities and delays and therefore in a non-uniform spacing of the partials. Increased stiffness results in a very relevant inharmonicity of the partials, e.g., as found in low register piano tones or drums and percussive sounds. The observed feature has been thoroughly analyzed in different ways in [2,3,4].

Recently the authors introduced a novel class of unitary transformations which map a signal from the time domain to the frequency domain, to the time-frequency (wavelet) domain or to the pitch synchronous wavelet domain, altering in a controlled way the harmonic content of the signal and the tiling in the time-scale domain. Mainly the transformation, specified by a single parameter, produces a warping of the frequency axis, which may be used to compensate for the increased spacing of the partials in the signals under analysis. In what follows some strategies are developed, in order to adaptively revert the inharmonic signal into its periodic clone.

2 The Frequency Warped Wavelet Transforms

Starting point for this new class of transforms is the Laguerre Transform [1], a useful transformation which, due to the fact that the basis function are rational functions in the z-Transform domain, is suitable for an exact discrete implementation.

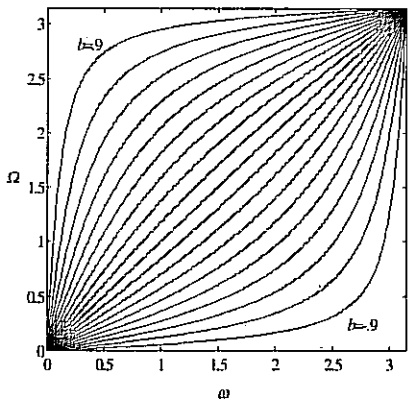


Fig. 1 The family of frequency warping laws

Moreover these basis functions are obtained by recursion of a single digital block, which makes the calculation more easily performed with simple difference equations. When cascaded with the usual Discrete Time Fourier Transform as well as with the Discrete Wavelet Transform, both ordinary and pitch synchronous [5], the overall transformation, inheriting all the positive features of the two cascaded unitary transform, exhibits orthogonality, completeness (perfect reconstruction property), power normalization in the various bands, in other words the unitarity property [6,7].

Both warping in the frequency domain (shown in Fig. 1) and arbitrarily shaped dyadic tiling in the time-scale domain (shown in Fig. 2 and Fig. 3) may be exploited in order to adapt the transformation to specific signals or class of signals (see also [8]).

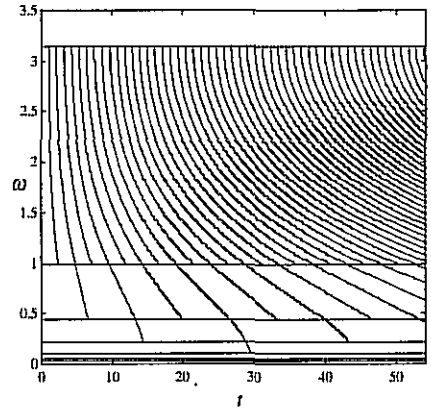


Fig. 2 *Tiling the time-frequency plane with frequency warped wavelets: Laguerre parameter b=.3*

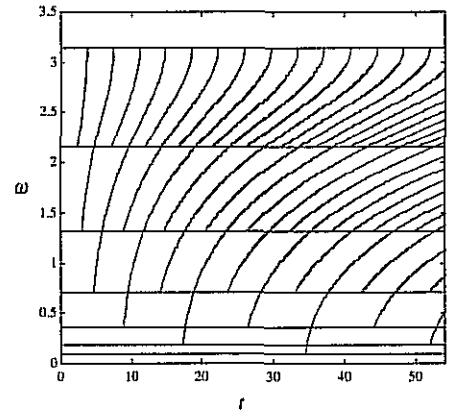


Fig. 3 *Tiling the time-frequency plane with frequency warped wavelets: Laguerre parameter b= -.3*

3 The peak picking algorithm

In the specific class of sounds that we are examining spacing among partials increases with partial order. From an experimental point of view the real curves can be very closely approximated by a curve of the family shown in Fig. 1. On the other hand it has been shown that the frequency dependent delay characteristic of the underlying physical medium due to the 4^{th} order term in the PDE describing the oscillating mean, may be well approximated by the phase response of a first-order all pass, or a chain of few different all pass filters [4]. Based on this fact, the choice of the suitable warping curve depends on a single parameter, the parameter of the Laguerre expansion. In order to achieve high quality in both the analysis and synthesis of sound one needs to carefully select this parameter, according to the dispersion curve of the partials. One method of selection is to perform a peak-picking algorithm in the frequency domain and choose the warping curve that best matches the given partial distribution.

Starting with a first choice of the fundamental pitch, at each step of the recursion, the next peak is selected as the abscissa of the maximum value in the frequency spectrum in a selected interval (a fraction of last pitch interval) where the next peak is expected. The size of this window in the frequency domain must be carefully adapted since a short window produces the artifact of an almost constant harmonic spacing, while a long window may preclude detection of the correct pitch sequence. One has also to check that the selected point is a local maximum and not simply an extremum of a strictly increasing or decreasing sequence; in this case the point is discarded from the final evaluation.

In Fig. 4 an ideal case is shown were the signal to be unwarped was produced by frequency warping a simple synthetic signal generated by the Karplus Strong algorithm. The number of partials which can be taken into account is very high for this ideal case and also the agreement with the curve is fairly good.

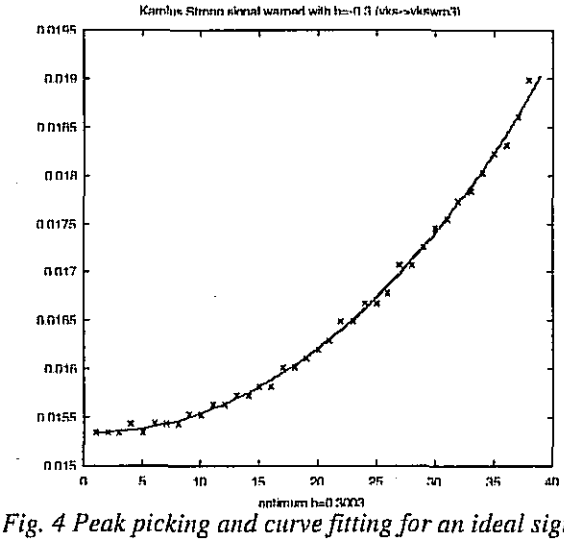


Fig. 4 Peak picking and curve fitting for an ideal signal

When real world signals are to be analyzed, the algorithm results are less accurate. For example, using a recorded signal from a piano, as shown in Fig. 5, the peak picking algorithm fails on higher order partials, whose energies are considerably lower. Furthermore, the dynamics includes starting transients due to initial percussion or pluck, rich in higher partials, beating due to strings coupling, and the well known feature consisting of a time varying spectrum rapidly evolving from a noisy excitation to a quasiperiodic steady state, with a practically resonant final decay. All these circumstances, definitely hide the underlying partial distribution law and prevent proper detection. On the other hand analysis of selected parts of the signal, far from the attack transient, does not provide better results. When the experimental points are properly detected, the Laguerre parameter may be obtained by iterating any simple gradient based optimization procedure in order to fit calculated points with the measurements using a mean square error criterion. The described procedure is shown in Fig.

6, where crosses are experimental data and the solid line represents the warping curve according to the detected value of the Laguerre parameter.

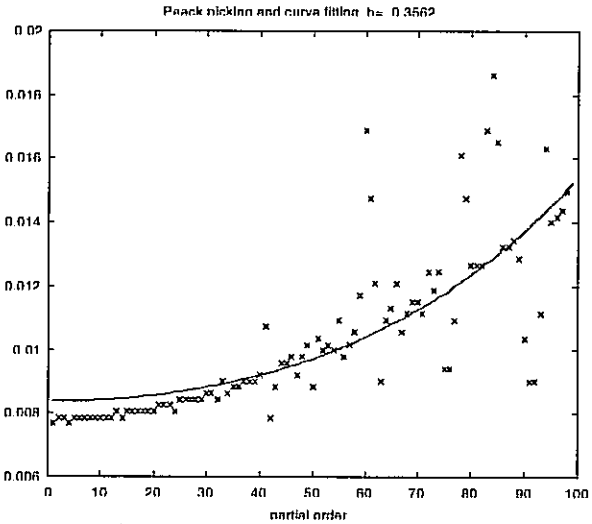


Fig. 5 Peak picking: higher partials are randomly detected. Subsequent fitting is failing and an incorrect value of b is selected.

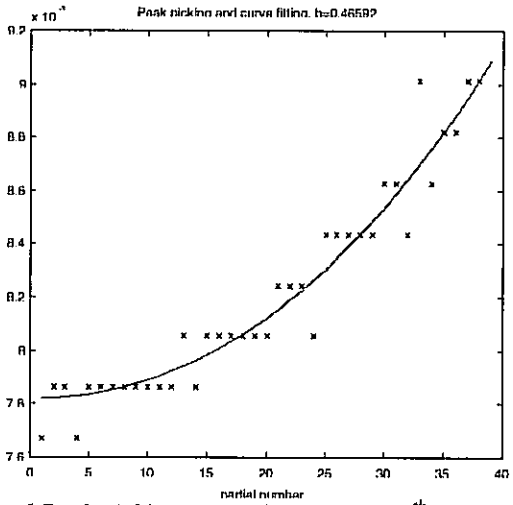


Fig. 6 Peak picking: partials up to the 40th are examined. Subsequent fitting reveals proper value of the Laguerre parameter.

4 A measure of the degree of disharmonicity of a pseudo-periodic sound.

The procedure described in the above is to be manually controlled, due to the difficulty met in the detection of higher harmonics. We face the absence of an objective and quantitative criterion.

Actually what is to be detected is the underlying comb structure of the sound, in its frequency domain representation. Proper warping is the one that lets our sound fit exactly with a comb structure with uniform spacing of the poles. The out of band energy actually constitutes a measure of the degree of disarmonicity of the signal, therefore it must be minimized. The value of the Laguerre parameter to be used in order to revert our signal to a 'perfectly harmonic' clone provides a parametric characterization of the disarmonicity of the signal, in the class of signals under analysis, for which the Laguerre transformation is appropriate.

5 A novel procedure for the adaptation of the Laguerre parameter to the signal

The above considerations suggest the idea of adaptively warping the signal by adjusting the Laguerre parameter using an optimization procedure, moving along the direction of the negative gradient of the out of band energy measure.

In turn, as it will be shown in the following, the $z^{-1}-b$ having unequal spacing of the poles, according to the inverse warping law having the opposite value of the parameter. In this case the inharmonic comb structure must be adapted to the source signal. This method is more convenient from a computational viewpoint. The out of band energy can be computed by warping a simple notch comb filter with pitch P and z-transform

$$H(z) = 1 - z^{-P}$$

by means of the transformation $z^{-1} \rightarrow A(z)$, where $A(z) = \dfrac{z^{-1}-b}{1-bz^{-1}}$ is the first order real all-pass filter. The filter is implemented by cascading P first order all pass sections, obtaining a dispersive delay line which simulates the line with frequency dependent parameters responsible for the peculiar spacing of the partials met in the class of signals under analysis [4,7]. The impulse and frequency responses of this filter are shown in Fig. 7, for a pitch of 40 samples.

By filtering the source signal with the given warped comb filter one obtains a signal whose energy will be minimum if the warping is canceling most of the energy of the partials of the source sound at the notch frequencies.

In order to properly set up the optimization procedure we numerically analyzed the error curve. It results that the out-of-band energy of the source signal filtered by the warped comb varies along a unimodal curve, at least in the region of interest, once the proper sign of the parameter is selected. A typical energy error curve is shown in Fig. 8, where we can see that for positive values of the parameter the curve has a single minimum. In the vicinity of the minimum we have an apparently erratic feature, with local extrema. This is due to the quantization of the pitch into integer values that in turn causes uncertainty also in the values of b. This simply imposes a limit to the precision of the Laguerre parameter b. We must point out that the optimization algorithm must be properly designed in order to overcome the above limitations. At each step of the recursion the warped pitch is recalculated by transforming the average pitch via the current warping map. In fact, after warping, the pitch frequency will move along the new warping map. Finally, in order to prevent overflow past the limit

values of the Laguerre parameter ±1, optimization is carried on the parameter $tan^{-1}b$, which limits the range of the unknown parameter.

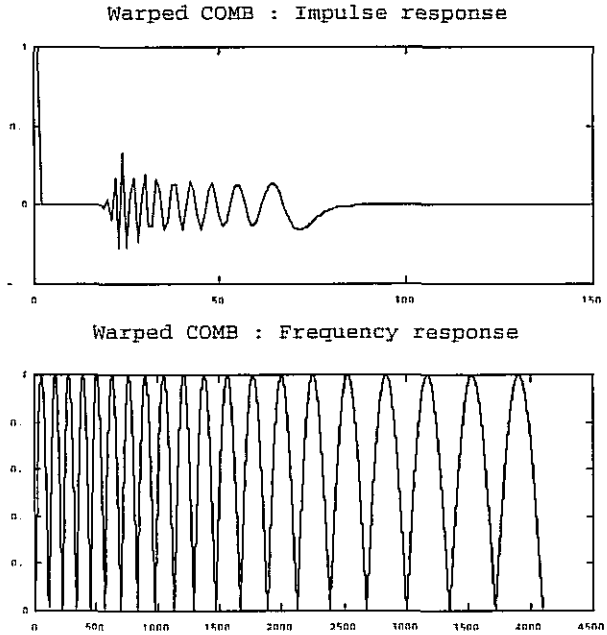


Fig. 7 Response of the dispersive delay line

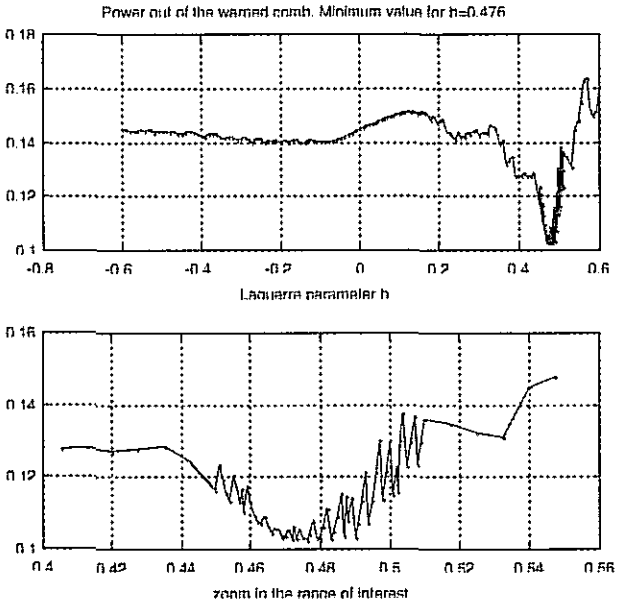


Fig. 8 Out-of-band energy of the warped comb vs. the Laguerre parameter b.

At each step the error norm is evaluated and the next trial parameter is selected by moving in the opposite direction of the error gradient.

With a similar procedure we obtained good results, as shown for example in Fig. 9 for the same piano signal used in the peak picking algorithm. Optimization is performed on the Laguerre parameter and assumes an initial estimate of the average pitch of the signal. Perfect adaptation is achieved when the position of the notches of the comb filter match the frequencies of the signal partials. The high-pass filter from which the notch comb filter is generated can be arbitrarily designed in order to improve performance, although the convergence of the method does not critically depends on this filter.

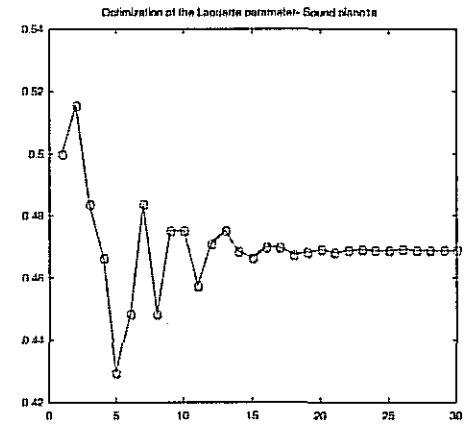


Fig. 9 *Optimization of the parameter b for a piano tone*

The main advantage of this method over the pick-peaking algorithm lies in the fact that the latter technique performs poorly with higher order partials, where noise energy is comparable to that of the partials, leading to unreliable estimates of their distribution.

References

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

[2] P. M. Morse and K.U. Ingard, "Theoretical Acoustics," Princeton University, 3rd ed., pp.187-191,1968.

[3] S. A. Van Duyne and J. O. Smith. "A Simplified Approach to Modeling Dispersion Caused by Stiffness in Strings and Plates", *ICMC Proceedings 1994*, pp. 407-410.

[4]I. Testa, G. Evangelista and S. Cavaliere, "A Physical Model of Stiff Strings," ISMA'97, Edinburgh, U.K., pp. 219-224.

[5] G.Evangelista, "Pitch Synchronous Wavelet Representations of Speech and Music Signals," IEEE Trans. on Signal Processing, special issue on Wavelets and Signal Processing, vol. 41, no.12, pp. 3313-3330, Dec. 1993.

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

[7] G. Evangelista and S. Cavaliere, "Frequency Warped Filter Banks and Wavelet Transforms: A Discrete-Time Approach Via Laguerre Expansion," to appear on IEEE Trans. on Signal Processing, Oct. 1998.

[8] R. G. Baraniuk and D. L. Jones, "Unitary Equivalence: A New Twist on Signal Processing," IEEE Trans. on Signal Processing, vol. 43, no. 10, pp. 2269 -2282, Oct. 1995.

ANALYSIS AND SYNTHESIS OF PSEUDO-PERIODIC 1/*F*-LIKE NOISE BY MEANS OF MULTIBAND WAVELETS

Gianpaolo Evangelista

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

Pietro Polotti

e-mail: polotti@na.infn.it

Abstract

The frequency spectrum of a large class of musical tones shows a 1/f behavior about each of the partials. The power spectrum of these processes can be modeled as a superposition of 1/f components centered on harmonic frequencies, i.e., as pseudo-periodic 1/f -like processes. The wavelet transform is a useful tool for the analysis and synthesis of the 1/f processes. The analysis coefficients of any 1/f process are approximately white noise processes at each scale level. Conversely, the output process obtained from white noise synthesis coefficients is approximately a 1/f process. We provide an efficient computational scheme for the analysis and synthesis of pseudo-periodic 1/f -like processes. Our analysis scheme is rooted on a P-band critically sampled filter bank whose channels are tuned to the sidebands of the harmonics. The dyadic wavelet transform is applied to the output of each channel of the filter bank. The overall structure is conveniently described in terms of projection onto a new set of orthogonal and complete multiband wavelets, i.e., the Harmonic Wavelet Transforms. The synthesis parameters may be conveniently extracted from the statistics of the Harmonic Wavelet analysis of the tone. The contribution of each of the parameters is highly intuitive and easily controlled interactively.

1 Introduction

Any real-life sound with a detectable pitch is not a periodic signal in a strict sense. This is not only due to the attack and decay transients. Even the steady part of any of these signals cannot be viewed as a deterministic process, rather as a stochastic one. This is mainly due to the complex of random elements present in the behavior of the exciters which produce voiced sounds in speech or music. The consequence is that such signals present stochastic micro-fluctuations compared to a pure periodic behavior. This is evident in the frequency-domain representation of these signals. In fact, their power spectra exhibit peaks centered on the harmonics but also relevant side-bands near these peaks. The shape of these side-bands have an approximate $1/f$ behavior. In the case of voiced sounds these chaotic but time-correlated micro-fluctuations are very relevant from the perceptual point of view. In fact, in sound synthesis, their simulation is indispensable to reproduce the dynamical evolution and therefore the naturalness of any sound with a detectable pitch.

In this paper we will introduce a new method for the analysis and synthesis of a class of signals, which we define as pseudo-periodic $1/f$-like processes and which is well suitable for the representation and reproduction of voiced sounds in speech and music. To this aim we will define and employ a special kind of wavelet packets, i.e. the Harmonic Wavelet Transforms (HWT). Our technique is based on the wavelet model for $1/f$ processes introduced by Wornell [2,3] and generalizes the pseudo-periodic $1/f$ analysis and synthesis scheme based on the multiplexed wavelet transform (MWT), previously presented by one of the authors [5,6]. In that scheme the exponential slope was the same for each partial. In the new model, each single side-band of the harmonics can be independently controlled in terms of peak value and exponential slope.

The claim of our method is that it is possible to model pseudo-periodic signals using simple white noise with proper scale-dependent variances as expansion coefficients on the HWT basis. Such variances are therefore the only parameters which one needs to specify in order to model the spectral behavior of the sidebands of a pseudo-periodic signal, i.e., of a highly complex stochastic process. The model is evolutionary in the sense that given a restricted set of parameters the signal waveshape is constantly updated, thus simulating the fluctuations present in natural sounds at each scale level. In this way one can control both the micro and macro structure of sound.

In section 2 we give a definition of pseudo-periodic $1/f$ -like processes. In section 3 we briefly introduce the Harmonic Wavelet Transforms. In section 4 we present the fundamental result of the synthesis of pseudo-periodic signals by means of HWT. Finally, we will discuss the applications results to several different kinds of instrumental sounds.

2 Pseudoperiodic 1/f-like noise

The frequency spectra of pseudoperiodic signals are characterized by harmonically spaced peaks at frequencies $\varpi_k = 2\pi k/T_p$, where T_p is the average period of the signal. In order to separate the contribution of each of the harmonic bands one can devise a set of ideal narrow-band

filters of bandwidth $\Delta\omega = \pi/T_p$ each fitting a single sideband of the harmonics. The magnitude frequency response of these filters for the right and left sideband respectively are given by

$$H_{k,R}(\omega) = \begin{cases} \chi_{\left[k\frac{2\pi}{T_p}, k\frac{2\pi}{T_p}+\frac{\pi}{T_p}\right[}(\omega) & k \geq 0 \\ \chi_{\left]k\frac{2\pi}{T_p}-\frac{\pi}{T_p}, k\frac{2\pi}{T_p}\right]}(\omega) & k < 0 \end{cases}$$

and

$$H_{k,L}(\omega) = \begin{cases} \chi_{\left[k\frac{2\pi}{T_p}-\frac{\pi}{T_p}, k\frac{2\pi}{T_p}\right[}(\omega) & k > 0 \\ \chi_{\left]k\frac{2\pi}{T_p}, k\frac{2\pi}{T_p}+\frac{\pi}{T_p}\right]}(\omega) & k \leq 0 \end{cases}, \quad (1)$$

where $k = 0, \pm1, \pm2, \ldots$, and $\chi_{[A,B[}(\omega) = \begin{cases} 1 & \text{if } A \leq \omega < B \\ 0 & \text{otherwise} \end{cases}$ is the characteristic function of the interval [A,B[and the indexes R and L correspond respectively to the right and left sidebands. The contribution of the each harmonic k is the sum of the outputs of the filters $H_{k,R}$ and $H_{k,L}$.

In order to model a pseudoperiodic signal with fundamental frequency $f_0 = 1/T_p$, we introduce and define the pseudoperiodic $1/f$-like noise. This model is essentially based on the superposition of band-shifted and bandlimited $1/f$ processes. The frequency shifts must be in an harmonic relation and the bandwidth BW of each process equals half the harmonic spacing, i.e., $BW = \pi/T_p$. Each process contributes to a single sideband of each of the harmonics and is characterized by two independent parameters: amplitudes $\sigma_{k,L}$ ($\sigma_{k,R}$ for the rigth sideband), and decays $\gamma_{k,L}$ ($\gamma_{k,R}$ for the rigth sideband). In such a way the average spectrum of our model for $\omega \geq 0$ becomes :

$$S(\omega) = \sum_{k=0}^{\infty} \frac{\sigma_{k,R}^2}{|\omega - k2\pi / T_p|^{\gamma_{k,R}}} \chi_{\left[k\frac{2\pi}{T_p}, k\frac{2\pi}{T_p}+\frac{\pi}{T_p}\right[}(\omega) + \frac{\sigma_{k,L}^2}{|\omega - k2\pi / T_p|^{\gamma_{k,L}}} \chi_{\left[k\frac{2\pi}{T_p}-\frac{\pi}{T_p}, k\frac{2\pi}{T_p}\right[}(\omega) \quad (2)$$

An example of pseudoperiodic spectrum is shown in Fig. 1.

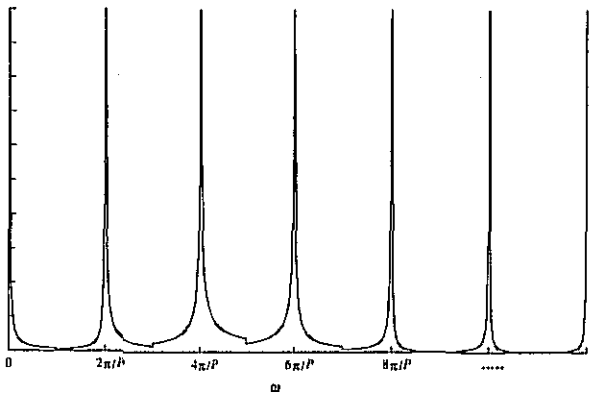


Fig. 1 : Pseudoperiodic power spectrum

We can provide an intuitive definition of pseudo-periodic $1/f$-like noise that can be viewed as an extension of the definition of the $1/f$ noise given in [3] (for a rigorous definition of pseudo-periodic $1/f$-like noise see [9]). We start by noticing that when any of the sideband of the harmonics is baseband shifted the resulting process is $1/f$, bandlimited to $[-\pi/T_p, \pi/T_p]$. By passing the shifted component processes through an ideal bandpass filter

$$H^{(\varepsilon)}(\omega) = \chi_{\left[-\frac{\pi}{T_p}, -\varepsilon\right]}(\omega) + \chi_{\left[\varepsilon, \frac{\pi}{T_p}\right]}(\omega), \quad (3)$$

where ε is arbitrarily small, the resulting process is finite-variance and wide-sense stationary. This is consistent with the $1/f$ noise definition given in [3]. Finally we can state the following

Definition 1 : A stochastic process $x(t)$ is said to be a $1/f$-like pseudo-periodic noise if it is possible to find a period $T_p > 0$ such that when $x(t)$ is filtered by means of the ideal passband filters (1) and conveniently baseband shifted, yields a collection of processes $x_{k,L}(t)$ and $x_{k,R}(t)$ that, if filtered through $H^{(\varepsilon)}(\omega)$, they become bandlimited and wide-sense stationary with power spectrum,

$$S_{x_{k,\bullet}}(\omega) = \begin{cases} \sigma_{k,\bullet}^2 / |\omega|^{\gamma_{k,\bullet}} & \text{if } \varepsilon < |\omega| < \pi/T_p \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

for some $\sigma_{k,\bullet}$ and $\gamma_{k,\bullet}$, where the symbol $\bullet$ indicates both the left and right sidebands.

3 Harmonic Wavelets

In order to introduce the harmonic wavelets we consider the following steps:

1) Since the resulting processes $x_{k,L}(t)$ and $x_{k,R}(t)$ in Definition 1 are bandlimited to $[-\pi/T_p, \pi/T_p]$, they can be sampled with sampling rate $1/T_p$.

2) It exists a set of functions defined as follows:

$$g_{q,r}(t) = g_{q,0}(t - rP), \quad (5)$$

with

$$g_{q,0}(t) = \frac{1}{\sqrt{T_P}} \cos\left(\frac{2q+1}{2T_P}\right) \operatorname{sinc}\left(\frac{t}{2T_P}\right),$$

such that, the samples $x_{k,L}(lT_p)$ and $x_{k,R}(lT_p)$, where $k = \left\lceil \frac{q}{2} \right\rceil$, are equal, up to a multiplicative constant $\sqrt{T_P}$, to the expansion coefficients of $x(t)$ over the set (5). This set is easily shown to be an orthogonal and complete set of functions [9].

3) In order to obtain an efficient scheme for the analysis and synthesis of discrete pseudoperiodic $1/f$-like noise we consider an approximation of the ideal filterbank with a perfect reconstruction structure [7]. Such approximation is provided by the class of Type IV cosine modulated bases, associated with an ideal P band filterbank:

$$h_{q,r}(l) = h_{q,0}(l - rP), \quad q = 0, \ldots, P-1; \quad r \in \mathbf{Z}$$

with

$$h_{q,0}(l) = h(l) \cos\left(\frac{2q+1}{4P}\left(l - \frac{M-1}{2}\right)\pi - (-1)^q \frac{\pi}{4}\right) \quad (6)$$

In (6) the lowpass impulse response $h(l)$ has length M and satisfies a number of technical conditions [8].

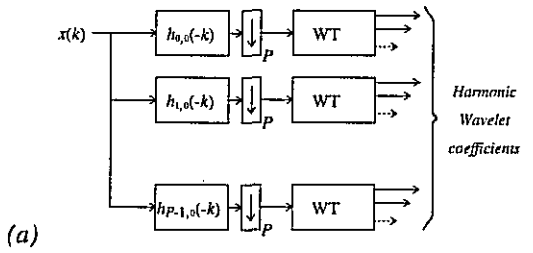
4) The samples of $1/f$ processes $x_{k,L}(l)$ and $x_{k,R}(l)$ can be synthesized adopting the method devised by Wornell [3], which consists of an ordinary discrete wavelet synthesis structure, with white noise as expansion coefficients. These four steps lead us to define the Harmonic Wavelets and the corresponding Harmonic Scale functions as:

$$\xi_{n,m,q}(l) = \sum_r \psi_{n,m}(r) h_{q,r}(l)$$
$$\vartheta_{n,m,q}(l) = \sum_r \phi_{n,m}(r) h_{q,r}(l), \qquad (7a)$$

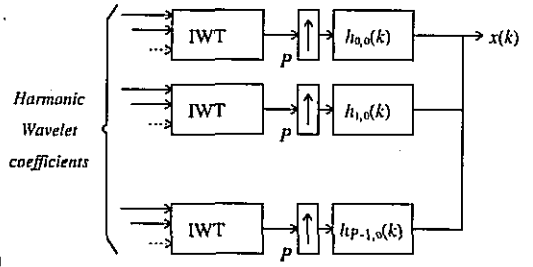
where $\psi_{n,m}(r)$ are discrete ordinary wavelets and $\phi_{n,m}(r)$ are the corresponding ordinary scale functions with Fourier transforms $\Psi_{n,m}(\omega)$ and $\Phi_{n,m}(\omega)$, respectively [4]. The Fourier transforms of the Harmonic Wavelets and of the Harmonic Scaling functions are comb versions of ordinary wavelets and scale functions, filtered by the filterbank frequency responses:

$$\Xi_{n,m,q}(\omega) = \Psi_{n,m}(P\omega) \mathrm{H}_{q,0}(\omega)$$
$$\Theta_{n,m,q}(\omega) = \Phi_{n,m}(P\omega) \mathrm{H}_{q,0}(\omega), \qquad (7b)$$

respectively. The action of filtering is essentially that of selecting a single sideband of the harmonics.



(a)



(b)

Fig. 2 : Harmonic Wavelet Transform: (a) analysis and (b) synthesis structures.

A structure for computing the Discrete Harmonic Wavelet Transform and its inverse is shown in Fig. 2. In the analysis structure, the signal is sent to a P channel filterbank and each output is Wavelet transformed (WT block). The number of channels is chosen according to the (integer) pitch of the signal. Signal reconstruction is achieved by separately inverse Wavelet transforming the Harmonic Wavelet coefficients and passing these sequences through the inverse P channel filterbank.

4 A synthesis scheme for 1/f-like noise

Our method is a sort of additive synthesis where one adds modulated $1/f$ signals instead of pure sinusoidal functions. The fundamental theoretical result supporting our synthesis scheme is based on the following result. Consider an arbitrary orthonormal Discrete Harmonic Wavelet set. Then the random-process

$$s_N(l) = \sum_{q=0}^{P-1} \sum_{n=1}^{N} \sum_{m=-\infty}^{\infty} \beta_q^{n/2} v_q^n(m) \xi_{n,m,q}(l) + \sum_{m=-\infty}^{\infty} \beta_q^{(N+1)/2} \mu_q^N(m) \vartheta_{N,m,q}(l) \qquad (8)$$

where the $v_q^n(m)$ and the $\mu_q^N(m)$ are zero-mean white-noise with normalized variance, the β_q are constants $\beta_q = 2^{-\gamma_q}$, $0 < \gamma_q < 2$, yealds an average power spectrum of the form:

$$\overline{S}_N(\omega) = \frac{1}{P} \sum_{q=0}^{P-1} \left(\sum_{n=1}^{N} 2^{n\gamma_q} \frac{|\Xi_{n,0,q}(\omega)|^2}{2^n} + 2^{(N+1)\gamma_q} \frac{|\Theta_{N,0,q}(\omega)|^2}{2^N} \right) \qquad (9)$$

which is approximately $1/f$ near each harmonic $k = \lceil q/2 \rceil$ with, $q=0,\ldots,P-1$.

$\xi_{n,m,q}(l)$ and $\vartheta_{N,m,q}(l)$ in (8) are the HWT as defined in (7a), while the $\Xi_{n,m,q}(\omega)$ and the $\Theta_{n,m,q}(\omega)$ in (9) are their Fourier Transforms as defined in (7b). It is possible to demonstrate this result in the continuos case with full generality [9].

5 Applications to speech and music synthesis

Our synthesis technique requires the estimation of three parameters per each harmonic partial of index k: the parameter $\sigma_k = \sigma_{k,L} = \sigma_{k,R}$, which controls the amplitude of the harmonics and the two parameters $\gamma_{k,L} = \gamma_q$ (where q is odd and equal to $2k$ -1) and $\gamma_{k,R} = \gamma_{q+1}$, which control the left and right side-band spectral slope respectively. The estimation method is based on the analysis of the signals one wants to reproduce. The parameters σ_k may be estimated from the frequency spectrum of the original signal by means of a peak-picking algorithm. For the estimation of the $\gamma_{k,L}$ and $\gamma_{k,R}$ we propose the following method rooted on the equality

$$\log_2\left(Var\left(x_{n,m,q}\right)\right) = \gamma_q \, n + \text{const}, \qquad (13)$$

where the $x_{n,m,q}$ are the coefficients of the Harmonic Wavelet analysis of the sample signal of the q^{th} channel.

The parameters γ_q can be thus evaluated by performing a linear regression based on (13). The experimental verification that voiced sounds have a pseudo-periodic 1/*f*-like noise behavior is given by the correlation coefficients of the linear regression. As an example we analyzed a sound of violin. We found that the principal harmonics, namely those with larger energy, are well approximated by a 1/*f*-like spectrum. Even more encouraging results were obtained with a sound of trumpet, as shown in Figs. 3 and 4.

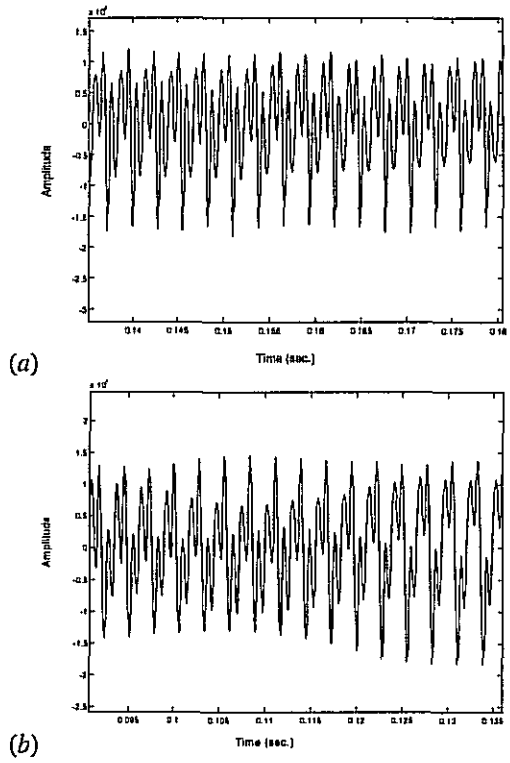


Fig. 3 (a) Real-life trumpet sound. (b) Synthesized trumpet sound.

Conclusions

In this paper we introduced a new method for sound synthesis that allows us to control and reproduce the micro fluctuations present in real life voiced sounds. This method is based on the experimental evidence that such fluctuations generate an 1/*f*-like behavior of the power spectrum near each harmonic of such sounds. We defined a new class of stochastic processes, i.e., the pseudo-periodic 1/*f*-like noise, adopting it as a mathematical model for voiced sounds. We introduced a powerful tool, i.e., the Harmonic Wavelet Transforms, as an extension of the Multiplexed Wavelet Transforms [5]. By means of the HWT we devised an efficient synthesis scheme able to generate pseudoperiodic 1/*f* -like noise. These processes can be controlled by means of few intuitive and completely independent parameters. The synthetic tones closely mimic the behavior of natural sounds. This method is also a powerful analytical tool, useful for separating for each harmonic the pure periodic behavior from transients and noise [5].

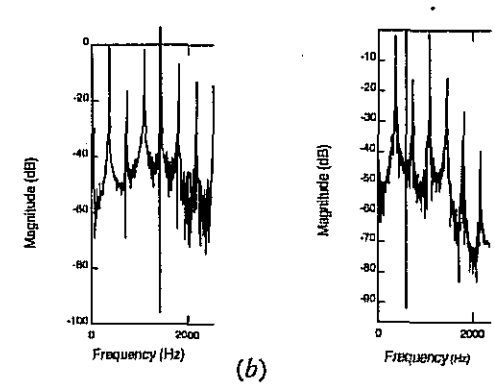


Fig. 4 : (a) Spectrum of the first harmonics of trumpet sound of figure 3a. (b) Spectrum of the first harmonics of the synthesized trumpet sound of figure 3b.

References

[1] M. S. Keshner, "1/f Noise," Proc. IEEE, Vol. 70, No 3, pp. 212-218, March 1982.

[2] G. W. Wornell and A. V. Oppenheim, "Wavelet-Based Representations for a Class of Self-Similar Signals with Applications to Fractal Modulation," IEEE Trans. Inform. Theory, Vol. 38, No. 2, pp. 785-800, March 1992.

[3] 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.

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

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

[6] G. Evangelista, "Pitch Synchronous Wavelet Representations of Speech and Music Signals," IEEE Trans. on Signal Processing, special issue on Waveletsand Signal Processing, vol. 41, no.12, pp. 3313-3330, Dec. 1993.

[7] P. P. Vaidyanathan, Multirate Systems and Filter Banks, Englewood.

[8] T. Q. Nguyen and R. D. Koilpillai, "The Theory and Design of Arbitrary-Length Cosine-Modulated Filter Banks and Wavelets, Satisfying Perfect Reconstruction", IEEE Trans. on Signal Processing, Vol. 44, No. 3, pp. 473-483, March 1996.

[9] P. Polotti, G. Evangelista, "Fractal Synthesis of Speech and Music by means of Harmonic Wavelets", manuscript in preparation, to be submitted to *IEEE Transactions on speech and audio processing.*

A Physically Based Model for Real-time Digital Synthesis of Analog-like Sounds

Michael Hamman
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
705 W. Nevada #4
Urbana, IL 61801
m-hamman@uiuc.edu

Abstract

This paper describes a physically-based model for synthesis of a large variety of analog-like sounds. The parameter space is relatively small and the algorithm allows for the creation and control of sounds with a surprisingly high degree of variety and interest. The model has been implemented in a real-time software music composition and synthesis system running on Pentium-based PCs under Windows 95 and Windows NT.

1. Background

Since the 1950s, composers have sought ways in which compositional principles could be employed in the construction of sound itself. This interest has carried over into computational synthesis of sound with the computer. By constructing algorithms for synthesizing sound, the composer constructs the very representations with respect to which designed acoustical patterns come to be imagined and articulated [1][2]. Such compositional motivation for the use of the computer redirects acoustical research from one which seeks ways to model already existing sounds and sound systems toward one which seeks to model possible *procedure* models of how compositional activity might itself be structured and constrained [3]. In this light, computational sound synthesis takes as its point of departure a desire for the possibility of modeling systems from which sound and musical pattern might be extrapolated [4].

Early efforts in non-standard synthesis (such as Brün's *SAWDUST* [5], Koenig's *SSP* [6], and Berg's *PILE* [7]) take this project to heart, as do more recent developments in non-standard synthesis [8][9], so-called "stochastic synthesis" [10], and synthesis based on iterative functions [11]. So-called physically-based models—which include Karplus-Strong models of the plucked string [12], modal synthesis[13], and chaotic systems[14]—provide a framework for constructing synthesis algorithms in that their behaviors may not be fully understood in advance. In this regard, as models of compositional process, they resemble non-linear systems. In this paper, I describe an approach to the development of a synthesis model, based on a simple physical model design, which can produce a large variety of analog-like acoustical behaviors.

2. Functional Overview of the System

The synthesis model breaks down into the two standard components: *excitor*, *resonator*.

The Excitor module generates a sequence of noise bursts. Both the duration of each noise burst and the rate of successive noise bursts are parameterised. Thus, one could construct excitors with but a single-sample burst to those that produce walls of noise to those producing various kinds of pulse-like behaviors. The output of the excitor is wrapped around by two additional components. The first is tunable comb filter while the second is hard clipping module.

The Resonator module contains two components, each being a tunable feedback delay module. Placing a first-order allpass filter at the end of the delay line effects the tuning. The allpass filter allows the resonator to produce frequency behaviors that are not only integral multiples of the sampling rate (as would be the case with only a delay line) [15].

Each feedback delay module is attached to oscillators that modulate the delay lengths and the feedback coefficient values. When turned on, these oscillators can be used to add a new dimension to the behavior of the resonator, thus generating new acoustical behaviors.

In the following sections, I discuss each module in more detail.

3. Excitor Module

Figure 1 depicts the Excitor module. The Excitor module contains three components. The *Initial Excitation* component generates the initial noise bursts. The *Filter* component filters the noise bursts using a comb-type filter. The *Clipper* component acts as a hard clipper to the final output of the Excitor.

The Initial Excitation component generates streams of noise bursts of a particular maximum amplitude. Each burst is set to a particular duration (parameter "D" in figure 1), while the rate of noise bursts is similarly parameterised (parameter "R" in figure 1).

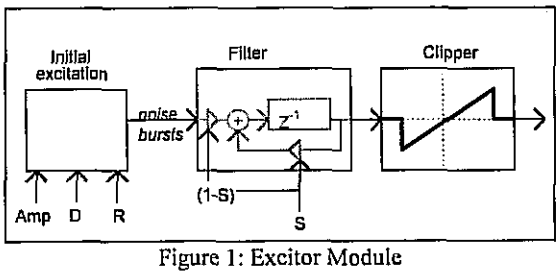


Figure 1: Excitor Module

One aspect of the behavior of the excitor has to do with the numeric relation between burst rate and the duration of individual noise bursts. When the difference between the duration of each noise burst and the burst rate falls within a certain range (in terms of numbers of samples), the excitor will exhibit oscillatory behavior within the audible frequency range. For instance, if the burst duration is 10 samples, and the burst rate is 210 samples then, the difference being 200 samples, the excitor will produce an output with a frequency of sampling rate / 200. With a sampling rate of 44.1KHz, the excitor output exhibits frequency behavior at approximately 220 Hz. Since burst duration and burst rate are independently modulated, this behavior can be exploited.

The output of the Initial Excitation component constitutes streams of noise bursts that are filtered through the Filter component. As is depicted in figure 1, the Filter component implements a comb filter. The particular comb filter implement used has the difference equation

$$y_n = (1 - S)x_n + Sy_{n-1}$$

where S is the only parameter to be set. S has the range {-1, 1}. When S is set equal to –1, the output of the filter easily becomes unstable with output samples exceeding unity. Otherwise the filter acts either as a highpass or lowpass, depending on the setting for S.

Any overflow from the Filter component is handled within the Clipping component. This component defines the function

$$f(x) = \begin{cases} 0; & x >= 1 \\ x; & -1 < x < 1 \\ 0; & x <= -1 \end{cases}$$

This function is a modified version of the standard hard clipper in order to avoid oscillation around the extremes should the comb filter start to oscillate.

The combination of the Filter and Clipper components produce two types of behavior, depending on the setting of S. When the value of S is in the range approximately {-.99, 0}, the filter acts as a high-pass filter. In the range {0, .99}, the filter acts as a low-pass.

However, within the range {-.99, -1.0}, a different kind of behavior emerges. Here, the filter will produce large numbers of overflows (outputs whose values are greater than unity gain). All overflows are trapped within the hard clipper, translating that overflow as a continuous DC (with an amplitude of 0). At the extreme (which S is equal to -1), very few non-zero samples ever leave the Excitor Module. However, as the value of S moves from -1 toward -.99, intermittent bursts are allowed through. As it moves closer to -99, the rate at which those intermittent bursts are allowed through increases.

By virtue of this dual behavior, S can act at once as an intermittence variable and a filtration variable.

4. Resonator Module

The Resonator module contains two resonators in series. Figure 2 depicts one such resonator. The grayed box delineates the allpass filter part of the resonator.

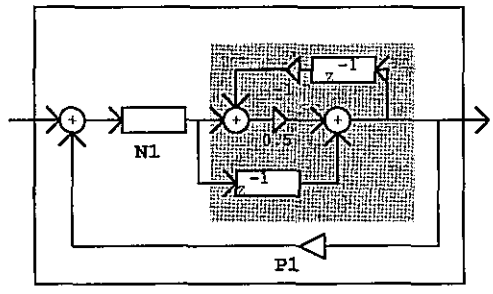


Figure 2: A single Resonator Module

Each resonator can be controlled through modulation of the delay length (depicted as "N1" in figure 1) and of the feedback coefficient (depicted as "P1" in figure 1). The synthesis model provides two means for their modulation: through direct manipulation (of N1 and P1) and through manipulation of oscillator whose outputs modulate N1 and P1.

Each resonator within the Resonator module has two oscillators attached to it. One oscillator modulates the delay length (including the allpass filter), while the other modulates the value of the feedback coefficient.

Figure 3 depicts this arrangement for a single resonator component. With the amplitude of an oscillator set to above zero, and frequency above zero, oscillation occurs around the current value for the parameter being modulated. For instance, if parameter "N1" is being modulated with the oscillator, and the current value for "N1" is 200 Hz, then oscillation would occur around 200 Hz. If the amplitude for the oscillator is .5 then the delay length would oscillate between 100 and 300 Hz. Oscillator frequency determines the rate at which such oscillation occurs. Complex timbres result when the frequency of oscillation occurs within the "audible" range.

5. Parameter Space

There are 17 parameters defined for the model. It is important to note however that, except in the most extraordinary circumstances, far fewer than all 17 would ever be manipulated simultaneously.

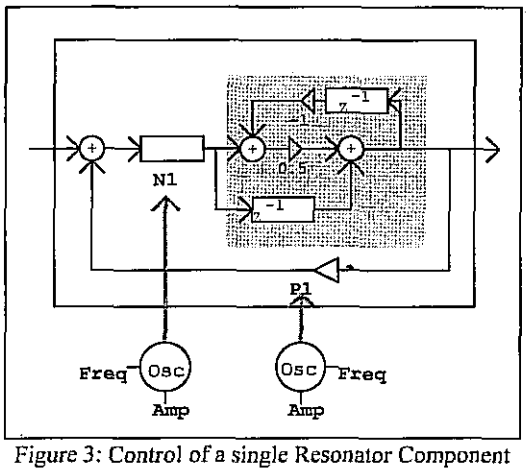


Figure 3: Control of a single Resonator Component

Two tools have been developed for the investigation of this model. The first tool is a standard *sliders interface*, one slider per parameter. With this tool, the composer can investigate the synthesis algorithm through independent control of various parameters.

The second tool is an *integrated control interface*. The integrated control interface allows manipulation of more than one parameter at once. It is predicated on the notion that when investigating the behavior of a system, one is often interested in the effect that various groupings of parameters will have on the behavior of a system.

Figure 4 depicts such an interface.

With this tool one clicks on a node and drags it around, observing acoustical feedback to one's movements. Movement of one node engenders movement of all nodes to which that node is connected (depicted by lines in the diagram). Each connection is defined by a "weight" according to which the movement of one node will effect the movement of its attached node. So, for instance, movement of the node labeled "N1" would cause movement of nodes "N2", "S", and "B" by weight factors of -2.0, .5, and 1.0 respectively.

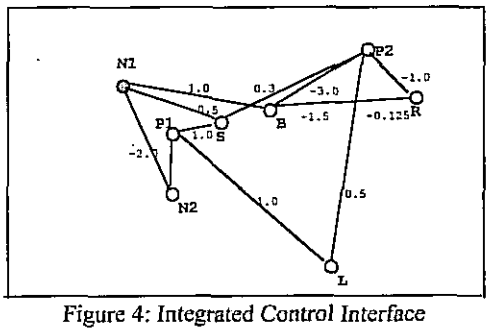


Figure 4: Integrated Control Interface

6. Implementation

The synthesis model is currently implemented within a real-time sound synthesis and composition system (described elsewhere in this Proceedings) and is built on top of the AREAL real-time synthesis library for Windows 95 and Windows NT [16]. Two to four simultaneous streams of sounds based on this algorithm can run comfortably on a Pentium-Pro based system running at 200 MHz.

7. Experiments

Several "classes" of acoustical behavior can be described. These range from tube-like behaviors to voice-like to intermittent.

Several examples are described and performed during conference presentation and can be obtained on-line.

9. Availability

Source code, executables, and examples are available on-line at:

`http://duracef.shout.net/~mhamman.`

10. References

[1] Hamman, M., "Interaction as Composition: Toward the *Paralogical* in Computer Music." *Sonus* 17(2), pp. 26-44. 1997.

[2] Hamman, M. "From Symbol to Semiotic: Computation as Interaction." *Proceedings of the 1998 ICMC*. San Francisco: Computer Music Association. 1998.

[3] Laske O. "Introduction to Cognitive Musicology." *Computer Music Journal* 12(1), pp. 43-57. 1988.

[4] Brün, H. "Infraudibles." in *Music by Computer*, ed. H. Von Foerster and J. Beauchamp. New York: John Wiley and Sons, pp. 117-121. 1969.

[5] Blum, T. "Herbert Brün: Project Sawdust." *Computer Music Journal* 3(1). 1979.

[6] Berg, P., Rowe, R., and Theriault, D. "SSP and Sound Description." *Computer Music Journal* 4(3), pp. 25-35. 1980.

[7] Berg, P. "PILE--A Language for Sound Synthesis." in *Foundations of Computer Music*, ed. C. Roads and J. Strawn. Cambridge, MA: The MIT Press. 1987.

[8] Chandra, A. "CounterWave: A Program for Controlling Degrees of Independence between Simultaneously Changing Waveforms." *Proceedings of the IAKTA/LIST International Workshop on Knowledge Technology in the Arts*, pp. 115-134. 1993.

[9] Corey, K. "My Algorithmic Muse." *Sonus* 17(2), pp. 81-93. 1997.

[10] Xenakis, I. *Formalized Music*. New York: Pendragon Press. 1991.

[11] Di Scipio, A. "Composition by exploration of nonlinear dynamic systems." *Proceedings of the 1990 ICMC.* San Francisco, CA: Computer Music Association. 1991.

[12] Karplus, K., and Strong, A. "Synthesis of Plucked-String and Drum Timbres." in *The Music Machine: Selected Readings from Computer Music Journal*, ed. C. Roads. Cambridge, MA: The MIT Press.

[13] Morrison, J. D., and Adrien, J.-M. "MOSAIC: A Framework for Modal Synthesis." *Computer Music Journal* 17(1), pp. 45-56. 1993.

[14] Choi, I. "A Chaotic Oscillator as a Musical Signal Generator in an Interactive Performance System." *Journal of New Music Research* 26, pp. 17-47. 1997.

[15] Jaffe, D., and Smith, J. O. "Extensions of the Karplus-Strong Plucked-String Algorithm." in *The Music Machine: Selected Readings from Computer Music Journal*, ed. C. Roads. Cambridge, MA: The MIT Press.

[16] Goudeseune, C., and Hamman, M. "A Real-Time Audio Scheduler for Pentium PCs." *Proceedings of the 1998 ICMC.* San Francisco: Computer Music Association. 1998.

Real-time Control of the Frequency-Domain with Desktop Computers

Cort Lippe & Zack Settel

University at Buffalo, Department of Music
Hiller Computer Music Studios
222 Baird Hall
Buffalo, NY, USA 14260
lippe@acsu.buffalo.edu

McGill University, Music Faculty
555 rue Sherbrooke Ouest
Montreal, Quebec H3A 1E3
CANADA
zack@music.mcgill.ca

Introduction

For some years, real-time general-purpose digital audio systems, based around specialized hardware, have been used by composers and researchers in the field of electronic music, and by professionals in various audio-related fields. During the past decade, these systems have gradually replaced many of the audio-processing devices used in amateur and professional configurations. Today, with the significant increase in computing power available on the desktop, the audio community is witnessing an important shift away from these systems, which required specialized hardware, towards general purpose desktop computing systems featuring high-level digital signal processing (DSP) software programming environments.

A new real-time DSP programming environment called Max Signal Processing (MSP) [1], was released this past year for the Apple Macintosh PowerPC platform. This software offers a suite of signal processing objects as an extension to the widely used Max software environment, and provides new opportunities for musicians and engineers wishing to explore professional-quality real-time DSP. Most important, MSP provides a number of frequency-domain processing primitives that allow for the development of sophisticated frequency-domain signal processing applications. Working in MSP, the authors have developed a library of frequency-domain DSP applications for cross-synthesis, analysis/resynthesis, denoising, pitch suppression, dynamics processing, advanced filtering, spectrum-based spatialization, and phase vocoding. Much of this library was originally developed by the authors on the IRCAM Signal Processing Workstation (ISPW) [2], and has been discussed in previous papers [3]. MSP is a direct descendant of ISPW Max [4], but provides greater portability and increased functionality. The authors have made improvements to the library, while developments in new directions have been made possible by features of MSP which ameliorate exploration in the frequency domain. Techniques and applications will be presented and discussed in terms of both general algorithm and MSP implementation, providing a concrete point of departure for further exploration using the MSP environment.

1. Fundamental operations

The standard operations which are used when processing audio signals in the frequency domain typically include: (1) windowing of the time-domain input signal, (2) transformation of the input signal into a frequency domain signal (spectrum) using the Fast Fourier Transform (FFT), (3) various frequency-domain operations such as complex multiplication for convolution, (4) transformation of the frequency-domain signals back into the time domain using the Inverse Fast Fourier Transform (IFFT), (5) and windowing of the time-domain output signal. This section of the paper will discuss some of the basic operations and techniques used in the various applications developed by the authors.

Implementation

The FFT object stores time-domain signals as buffers of samples upon which the FFT analysis is done. For the purpose of discussion, the examples given in this paper make use of buffers of 1024 samples. Unlike time-domain signals, a frequency-domain signal is represented by a succession of spectral "frames". Like frames in a movie, the frames of FFT data represent a "snapshot" of a brief segment of an audio signal. A frame consists of a certain number of equally spaced frequency bands called "bins". The number of bins is equal to the size of the FFT buffer, thus the frames of FFT data have 1024 bins. Each bin describes the energy in a specific part of the audio signal's frequency range.

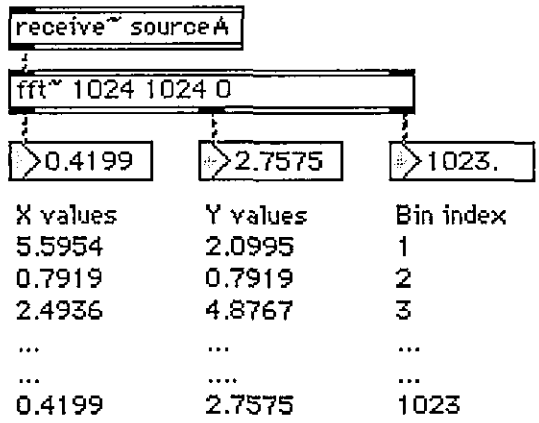


figure 1: sample-by-sample output of the FFT object

The FFT object in MSP, based on Miller Puckette's ISPW implementation, outputs each frame, bin-by-bin, using three sample streams running at the sampling rate. Thus, each bin is represented by three samples consisting of "real" and "imaginary" values, and the bin number (index). At any given instant, each of the FFT's three signal outlets, shown above, produce a sample describing the *nth* bin of the current FFT frame. The IFFT is the complement of the FFT and expects, as input, real and imaginary values in the same format as FFT output.

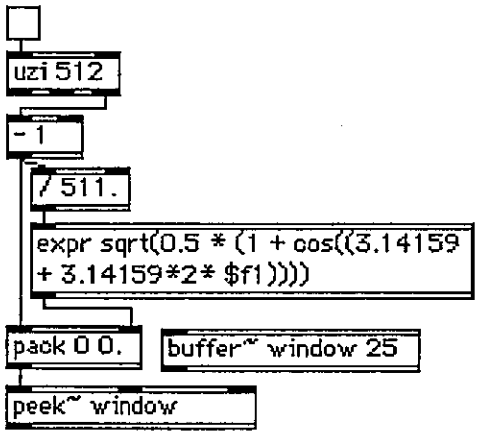


figure 2: windowing function generator

As seen in figure 1, the index values provide a synchronization signal, making it possible to identify bins within a frame, and recognize frame boundaries. The index values can be used to access bin-specific data for various operations, such as attenuation or spatialization, and to read lookup tables for windowing.

Windowing

It is necessary when modifying spectral data to apply an envelope (window) to the time-domain input/output of an FFT/IFFT pair, and to overlap multiple frames. For simplicity's sake, windowing operations are not shown here, but correspond to a two-overlap implementation (two overlapping FFT/IFFT pairs) which is easily expanded for use in a four or eight-overlap implementation. Because of the flexibility of Max and MSP, arbitrary windowing functions can be conveniently generated (see figure 2). The FFT frame index is used to read a lookup table-based windowing function in synchronization with the frame. The frame index is scaled between 0 and 1 in order to read a windowing function stored in an oscillator.

Bin-specific table lookup

Based on the implementation of MSP's FFT/IFFT objects, the authors have developed techniques for performing various operations on frequency-domain signals. The ability to access specific spectral components (bins) is central to performing frequency domain operations. This can be accomplished using lookup tables. In a simple application, a 1024-point FFT, using a lookup table, offers control of 512 equalization bands across the frequency spectrum.

2. Flavors of convolution

Multiplying one frequency-domain signal by another (convolution) involves the operation of complex multiplication. This operation is at the heart of the many frequency-domain processing techniques developed by the authors, and provides the basis for cross-synthesis, filtering, spatialization, phase vocoding, denoising and other applications. A technique often used with convolution is the reduction of a signal's spectrum to its amplitude and/or phase information. Four examples of convolution are shown below; each one corresponds to a particular type of signal processing application.

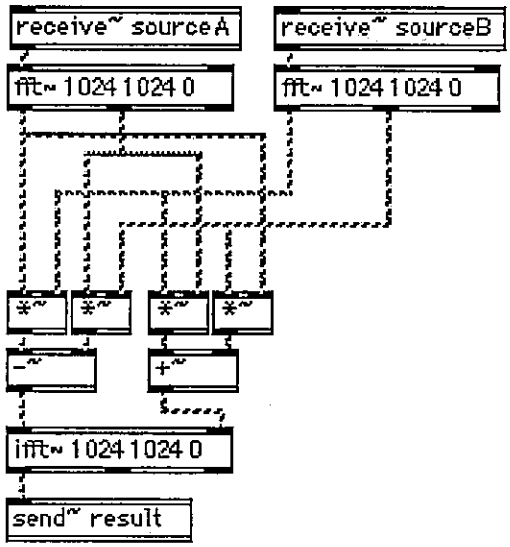


figure 3: simple convolution retaining the phases and amplitudes of each input source

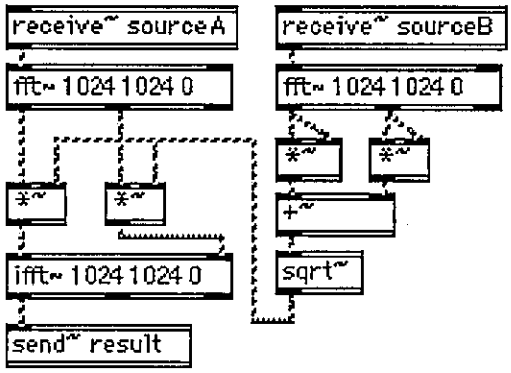


figure 4: phase/amplitude and amplitude-only spectra

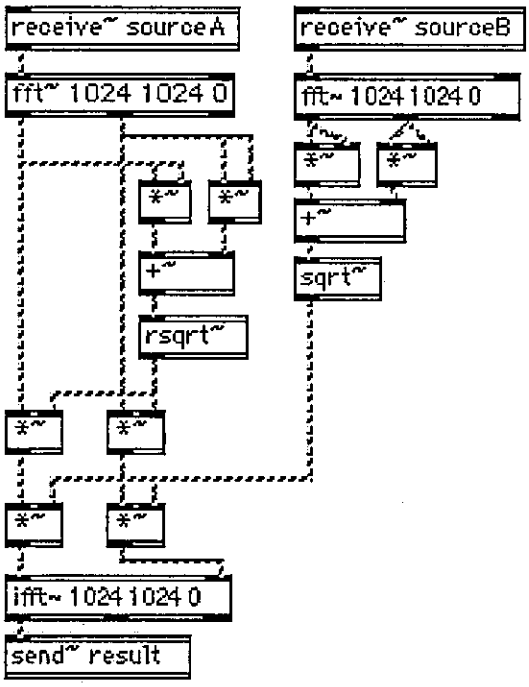


figure 5: phase-only and amplitude-only spectra

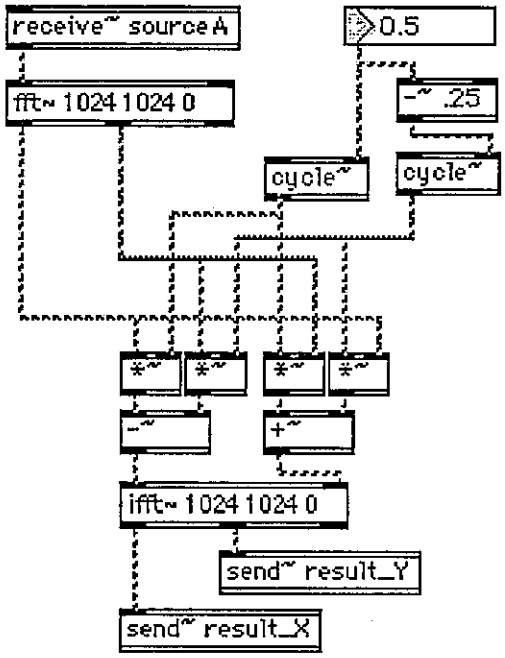


figure 6: phase rotation

3. Improvements to cross-synthesis: Increasing spectral intersection

Cross-synthesis is based on convolution (the multiplication of one spectrum by another). As anyone who has worked with cross-synthesis knows, the choice of the two input signals is critical to the outcome of the operation. Input signals with little common spectra (spectral intersection) will produce poor results. In the worst case, input signals with no common spectra will produce silence. By using filtering and dynamics processing (compressor/limiter) techniques to redistribute the energy in a given spectrum, it is possible to modify the degree of spectral intersection when convolving two frequency-domain signals with dissimilar spectra. Three approaches are presented below.

3.1 Amplitude spectrum smoothing

An amplitude spectrum with deep peaks and notches (the case for a pitched sound) makes a good candidate for spectral smoothing. With MSP's FFT implementation, it is quite easy to apply a second-order low-pass filter to an amplitude spectrum. The filter's cutoff frequency, and damping parameters control the degree of spectral smoothing, averaging the energy across empty bins, thereby reducing the sharp notches or peaks which are harmonically distributed across the spectrum of pitched signals. A smoothed spectrum will combine via multiplication much more effectively with the spectrum of another sound—particularly when the other sound also has pronounced peaks and notches distributed differently across its spectrum. For example, this technique can prove useful in crossing the amplitude spectrum of a singer with the spectrum of a pitched instrument, such as a saxophone.

3.2 Bin-independent dynamics processing

Boosting weaker components of an amplitude spectrum is another technique which increases the potential degree of spectral intersection in cross-synthesis. The authors have implemented a compressor/expander which operates independently on each frequency bin of a spectrum [5]. Each bin's amplitude is used as an index into a lookup table containing a compression/expansion function; the value read from the table is then used to alter (scale) the particular bin's amplitude. The degree of dynamics processing is controlled by biasing or scaling the lookup table index. Note that compression/expansion functions may also be used to attenuate stronger amplitude components. This can be an effective check against extreme amplitude levels which result when crossing two sounds with similar spectral energy distributions.

3.3 Constant-amplitude phase-only spectrum

Finally, it is possible to maximize a spectrum's potential for intersection with another spectrum by forcing the amplitude information in each bin of the spectrum to a constant value of one (unity). Phase information remains unmodified. The resulting constant-amplitude phase-only spectrum will combine with the spectrum of another sound wherever there is energy in the other sound's spectrum. Figure 5, shown earlier, illustrates a technique that maximizes the potential for spectral intersection: a phase-only constant-amplitude spectrum is crossed with an amplitude spectrum. In any spectrum, the phase of "empty" (near 0 amplitude) bins is undefined. Consider two spectra with dissimilar spectral energy distributions. When forcing the amplitude of one sound's empty bins to unity and performing cross-synthesis with the amplitude spectrum of another sound, the resulting spectrum will tend to contain components whose phase is undefined (random). Thus, the resulting sound will be stripped of any pitch or stable frequency information.

Controlling degrees of spectral intersection

By combining the dynamics processing technique with the constant-amplitude forcing techniques described above, the degree of amplitude forcing towards unity can be continuously controlled. Thus, it is possible to specify how much of a given spectrum's original amplitude information, if any, will be used in a cross-synthesis operation with another spectrum.

4. Audio-rate control of FFT processing

The Max/MSP environment has two run-time schedulers: the Max "control" scheduler, which is timed on the basis of milliseconds, and the MSP "signal" scheduler, which is timed at the audio sampling rate [6]. In FFT-based processing applications, where changes to the resulting spectrum are infrequent, MSP's control objects may be used to provide control parameters for the processing. This is both precise and economical. In the FFT-based equalization application mentioned in section 1, a lookup table is used to describe a spectral envelope that is updated at the control rate. However, updating lookup tables at the control rate has band-width limitations, since the rapidity with which a lookup table can be altered is limited, giving the filtering certain static characteristics. Using 512 sliders to control individual FFT bins, drawing a filter shape for a lookup table with the mouse, or changing the lookup table data algorithmically provides only limited time-varying control of

the filter shape. In addition, the amount of control data represented in a lookup table is large and cumbersome. Significant and continuous modification of a spectrum, as in the case of a sweeping band-pass filter, is not possible using MSP's control objects, since they cannot keep up with the task of providing 1024 parameter changes at the FFT frame rate of 43 times a second (at the audio sampling rate of 44,100 samples per second). Keeping in mind that the FFT data being filtered is signal data, a more dynamic approach to filtering is to update lookup tables at the signal rate (the audio sampling rate). Using simple oscillators for table lookup, well-known waveform generation and synthesis techniques can be used to provide dynamic control of filtering. FM, AM, waveshaping, phase modulation, pulse-width modulation, mixing, clipping, etc., all have the potential to provide complex, evolving waveforms which can be used as spectral envelopes to provide a high level of flexibility, plasticity, and detail to filtering. The use of operations such as stretching, shifting, duplication (wrapping), inversion, retrograde and nonlinear distortion (waveshaping) also provide comprehensive means for modifying these spectral envelopes (waveforms). Most important, the parameters of these waveform-based techniques are few, familiar and easy to control. Additionally, lookup tables can be read in non-linear fashion; control is not restricted to the linear frequency scale, thus making possible the implementation of a constant-Q bandpass filter.

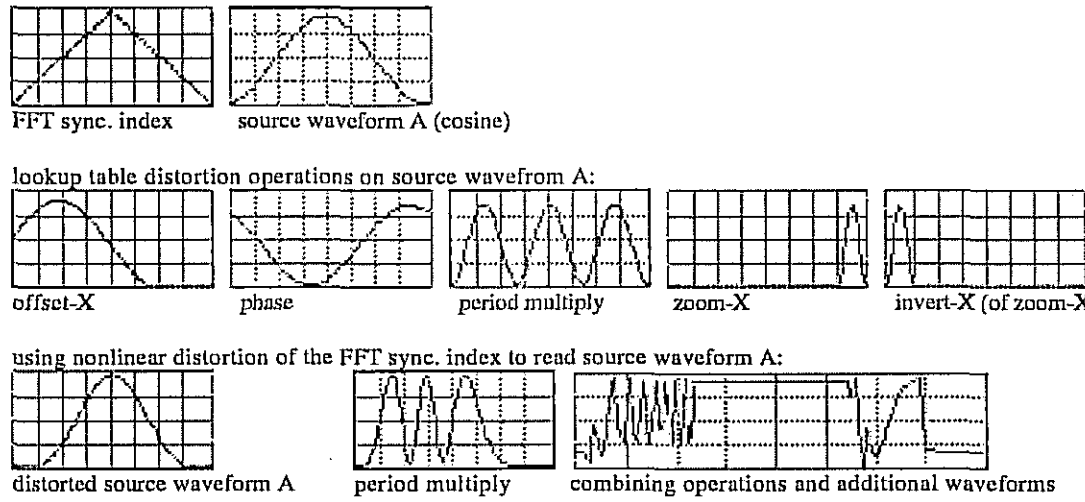


figure 7: generating spectral envelopes via simple operations on waveforms

While the discussion in this section has been limited to filtering applications, audio-rate control of FFT-based processing applies equally well to all FFT-based applications where a high degree of processing control is required at the frame rate. The authors have implemented the above mentioned techniques in the following applications: spatialization (bin-by-bin), denoising, dynamics processing, and cross-synthesis. We believe that these techniques hold great promise for control-intensive FFT-based applications.

Conclusion

The introduction of MSP offers new possibilities for musicians and researchers to develop real-timefrequency-domain digital signal processing applications on the desktop. In addition, the use of simple waveform-based techniques for low-dimensional audio-rate control of FFT-based processing has great potential: the parameters are few, familiar and easy to control, and direct mappings of real-time audio input from musicians to the control of FFT-based DSP is made possible.

Acknowledgments

The authors would like to thank Miller Puckette, David Zicarelli and the Convolution Brothers for making this work possible.

References

[1] Zicarelli D., 1997. "Cycling74 Website", www.cycling74.com.

[2] Lindemann, E., Starkier, M., and Dechelle, F. 1991. "The Architecture of the IRCAM Music Workstation." Computer Music Journal 15(3), pp. 41-49.

[3] Settel, Z., and Lippe, C., 1993. "FFT-based Resynthesis for Real-time Transformation of Timbre", X Colloquio di Informatica Musicale, Milano, Associazione di Informatica Musicale *Italiana, pp.* 214-219.

[4] Puckette, M., 1991. "Combining Event and Signal Processing in the Max Graphical Programming Environment." Computer Music Journal 15(3), pp. 68-77.

[5] Settel, Z., and Lippe, C., 1995. "Real-time Musical Applications using Frequency-domain Signal Processing" IEEE ASSP Workshop Proceedings, Mohonk New York.

[6] Puckette, M., 1991. "FTS: A Real-time Monitor for Multiprocessor Music Synthesis." Music *Conference. San Francisco: Computer Music* Association, pp. 420-429.

METAL STRING

PHYSICAL MODELLING OF BOWED STRINGS – A NEW MODEL AND ALGORITHM

Marco Palumbi

Centro Ricerche Musicali - Roma, Italy
Via Lamarmora, 18
00185 Roma, Italy
marco.palumbi@bigfoot.com

Lorenzo Seno

Centro Ricerche Musicali- Roma, Italy
Via Lamarmora, 18
00185 Roma, Italy
lorenzo.seno@bigfoot.com

Abstract

The paper introduces a new model of a bowed string, whose simulating algorithm is based upon a method quite different from the well know wave-guide approach – the currently used one in today musical research.
Our model implements both the viscous friction of the string with the air, and the internal friction, and a discontinuous bow, which includes also a parametric noise model (i.e., the noise is not simply added to the sound). The approach brings to an inherently time-variant system, so player can freely change any parameter without artifacts. The continuously controllable parameters are the tension/density ratio of the string (i.e- the pitch of the note), the two friction coefficients, the speed and the pressure of the bow. In today implementation, the bowing point (β) is variable but discrete.
Using a software based on our model, the Italian composer Michelangelo Lupone wrote the tape part of the string quartet "Corda di metallo" ("Metal string") (Kronos Quartett - Rome 1997).

1 INTRODUCTION

Musical Research in the field of musical instruments synthesis by Physical modeling is dominated by the tendency to use delay lines (or waveguides) as a solving algorithm [4][5][6][10]. This is mainly because of the low computational cost of delay lines, and may be because it is a form of historic tribute to the first method invented -- the Karplus-Strong algorithm [7][8]. Our approach is instead an evolution of the finite difference method, as an attempt to bypass its limitations. Finite difference method is equivalent to a physical model of a discrete mass-spring system, which inherently brings to partials with sub-harmonic ratios. Our string model is continuous, thus not suffering this inconvenient.
Non only the methods, by also the goals of our work are quite different from the prevailing one in this research field. Because of our personal tendency, and because of our bonds to contemporary composers, we were more interested in the search of new interesting sounds rather than in the imitation of true, traditionally played, instruments.
Our model (in the form of a computer software) was in effect used by the composer Michelangelo Lupone to get suggestions about new performing techniques and to compose the magnetic tape part of the "Metal string" quartet (from whom the title of this paper) for strings, tape and spatialiser (Kronos Quartett, Rome, 1997).

2 THE MODEL

2.1 The string

Without loss of generality, consider a string of unit length, whose free PDE is:

$$\frac{d^2}{dt^2} y(x,t) = \left(\frac{T}{\mu} \cdot \frac{d^2}{dx^2} y(x,t) - S \cdot \frac{d}{dt} y(x,t) \right) \ldots$$
$$+ - S_i \cdot \left(\frac{d^2}{dx^2} \frac{d}{dt} y(x,t) \right)$$

with boundary conditions:

$$y(0,t) = y(1,t) = 0$$

Where:

T	(Newton)	tension of the string.
μ	(Kg/m)	linear density of the string.
S	(sec^{-1})	coefficient due to the viscous friction with air.
S_i	(m^2/sec)	coefficient due to the viscous internal friction.

A few words about the presence (and the absence) of some terms. The classical dispersive term is absent:

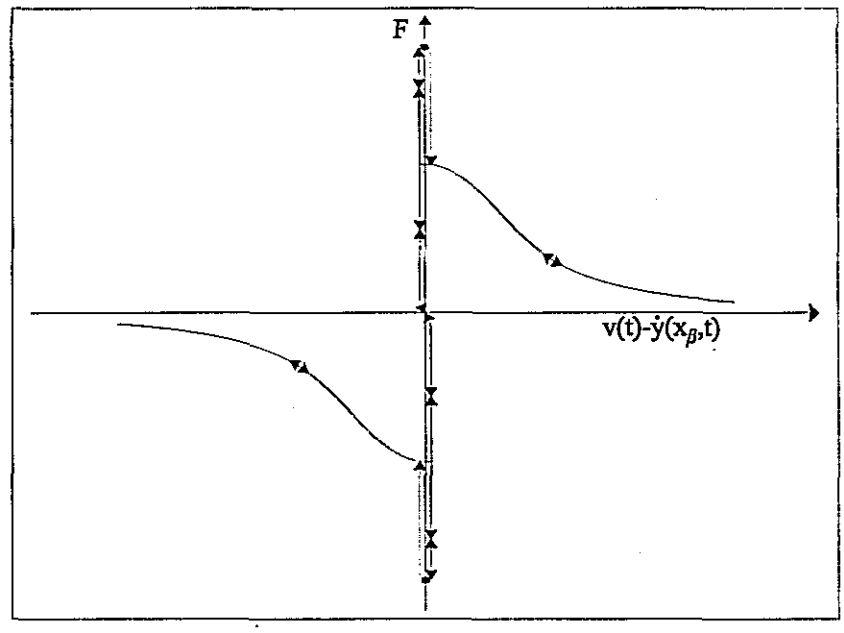
$$\frac{d^4}{dx^4} y(x,t)$$

This term is due to the stiffness of the string. It is responsible of the dependence of the propagation speed on the frequency. Because of these different speeds, partials are in super-harmonic frequency relationship.

This has important effects on the timbre of the sounds, especially with stiff strings –e.g. like in the low section of the piano.
Strings used in bowed instruments (as in the violin family) have low stiffness, but some researchers believe it is the reason of the so-called "rounding effect" [1][2].
Other researchers [3] suggested that this effect is due, for the most part, to the action of the bow, particularly to the hysteresis in the friction behavior of the rosin.
We skipped that term mainly because our integration method introduces at present some computational error whose effect is to slightly move the partial frequencies away from each other – the inclusion of this term in the computation being a straightforward task. On the other hand, strings used in the violin family have normally very low stiffness.
A few words now about the last, right-hand, mixed term generally neglected in the literature. It represents the effect of the energy losses due to internal, viscous, friction, which offer resistance to changes in the curvature. This phenomenon is responsible of a main behavior of actual, free running strings: the higher the frequency partial, the faster the dumping. For instance, if you pluck a string, during the transient you can hear many partials – a rich sound. Instead in the release part of the sound you can hear just the fundamental.
The pitch of the sound is due to the parameter T/μ. The way we integrate the evolution of the system – a finite difference method - does not make actually any assumption about the time-invariance of T/μ nor of any other parameter like internal and external dumping. You can thus modify in any desired way these parameters without any artefact.

2.2 The stimulus

There are several ways to excite a string: f.i. plucking, hitting or bowing.
At a first glance you can consider the excitation as a time-varying function applied to a specific point x_β of the string in the wise of a force, or a boundary condition for the speed or the displacement y.
A more realistic approach requires the consideration of the interactions between the excitation function and the string. This is the case of the exciting behavior of the bow. In our model we consider as excitation functions the pressure and the speed of the bow, and as interacting state variable the location of bowing point x_β, the speed and the acceleration of the bowing point.
The coupled systems is governed by $\sigma(t)$, a Boolean state variable describing the "stick" aut "slip" condition of the bowing point. The bowing point must be one of the sampling points of the string.
Our bow is in some way quite different from that usually implemented by means of continuous speed/force functions. During the stick state, the bowing point x_β is glued to the bow, and we force the speed of that point to be equal to the bow speed. During the slip state, the bowing force F is added to the forces balance in that point. The bowing force depends on the relative speed, in the way shown in the figure underneath.
As you can see, the "stick"→"slip" transition has a threshold that is twice the "slip"→"stick" transition (which happens for zero relative speed).



State diagram of the bowing point. Acceleration (force / linear density) versus point-bow relative speed.

The effect of the bowing pressure is to linearly scale the force (both threshold and continuous curve). One may explore other kinds of dependency.
We don't take explicitly into consideration the temperature of the point - a variable affecting the fluidity and thus the friction behaviour of the rosin (a natural polymer with a complex behaviour).
The shown threshold emulates however the cooling effect of the rosin during the stick time, during which the dissipation is zero. The noise due to the rubbing of the bow over the string is modelled by means of white noise added to the F curve, in such a way to preserve the curve as a "maximum value" of the random value. This is a true "noise model", and adds parametric noise to the model, which imparts a "chaotic" behaviour due to the random interaction between bow and string

3 THE SOUND

The sound is taken as the displacement of the spatial sampling point nearer to the bridge. One can think of this point as a "secant" approximation of the tangent to the string in the bridge position (i.e. the first spatial derivative in the origin), being the tangent the expression of the strain against the bridge.

4 THE METHOD OF CALCULUS

To obtain a continuous model of the shape of the string, we decompose it into a series of sinusoids of spatial frequency, which are multiples of that corresponding to twice the string length.

$$y(x)=\sum_{n=1}^{N} a_n \cdot \sin(n\cdot\pi\cdot x)$$

Thus we imagine the string as periodically and antisymmetrically infinitely expanded in the space. The series is truncated to the first N terms, which represents the maximum spatial frequency allowed to the string. For the free evolution, this limits to N also the partials of the sound. In order to derive the coefficients a_n of such series (i.e., to set-up a continuous model of the string shape) we must know the position of the string in N distinct points - the N spatial sampling points:
Given: $(x_1,y_1),(x_2,y_2) \ldots (x_N y_N)$
One can write:

$$y_i=\sum_{n=1}^{N} a_n \cdot \sin\left(n\cdot\pi\cdot x_i\right)$$

$$\vec{y}=\vec{a}\cdot M$$

Where M is a matrix whose elements are:

$$M_{i,n}=\sin\left(n\cdot\pi\cdot x_i\right)$$

So that:

$$\vec{a}=\vec{y}\cdot M^{-1}$$

Thanks to the knowledge of an analytic form of the solution, we can easily derive its differential properties in the unit interval; in particular we can formally get the curvatures (i.e. second spatial derivatives) in each of the sampling points.
Leaving out the boring algebra, the calculation of the curvatures can be performed through the product of a matrix (which is a function of the sampling points abscissas) by the positions vector y.

$$\vec{y}''=\vec{y}\cdot M'$$

Knowing the curvatures means breaking the equation of the string into N, second order, independent, motion equations of the sampled points. The left term (acceleration) results from the curvature, the velocities (which appear in the term of viscous friction with air) and from the time variation of the curvature itself (term of internal viscous friction). Our today software computes the motion of each point by means of a finite difference method, with an oversampling factor of 4, to reduce approximation errors.

5 RESULTS

The model has been widely tested and used with N=16 and a four times oversampling factor. In these conditions, it runs near to real time on a Pentium 266 system. Current researches are made on better methods of integration, in order to reduce the oversampling factor, reaching thus the real time on today's commercial systems. Research directions are also toward a global reduction of the complexity, to improve the number of harmonics generated, and to reach some degree of polyphony. The bowing point can be made continuous by a technique similar to polyphase filtering, thus requiring memory but no further computational complexity. In the same way can be also implemented supplementary bonds along the string, f.i. slight fingerings like those used to produce harmonics.
As to the excitations, we implemented simple models of pinch and percussion that brought very satisfactory results from the standpoint of imitation. As to the bowing, we made simulations with the specific goal to imitate sound emissions of actual instruments and they gave encouraging results in the case of little variations of the pitch, particularly as to the bow attack transients and the evolution of long-lasting sounds. Furthermore, the sounds generated with our model approximate very well, from an acoustic standpoint, the actual ones, especially in the variation of pressure, velocity and position also in non-standard conditions of execution.
The model has been used with "impossible articulations" (see [9]) not only in the sense that sounds with parame-

ter variations timings that can't be performed by humans have been produced, but also in the extended meaning that parameters, which are physically unreachable, like f.i. internal friction, have been varied during sound emission.

6 CONCLUSIONS

The evaluation of the various direction of research strongly depends on one's goals.
If we assume the imitation of actual instruments as a primary goal of the physical model simulation, we probably know how to interpret the results obtained. But if our goal is instead to provide a new synthesis instrument, both better controllable and more stimulating when compared to other methods, then the evaluation criteria become more complex and finally depend strongly on musical considerations.
In this latter case the imitation of actual instruments is relevant but only in an indirect way. It is just a guideline to verify the correctness of the hypothesis and the methods adopted, but is no longer, by its own, an interesting goal.
From the point of view of contemporary music, the modeling and implementation of features and behaving that are physically impractical may result more important and interesting even if unverifiable.
If we assume this point of view, the key of the evolution of our research lies in the hands of musicians and composers rather than in those of the researchers. Thus it's clear that such a research can be led but through a strong interaction between researchers and musicians.

ACKNOWLEDGEMENTS

Thanks to Michelangelo Lupone for his precious suggestions and for encouragement, and for composing and playing our instrument.
Thanks to Marco Giordano who reviewed the paper.

REFERENCES

[1] C.V. Raman, (S. Ramaseshan Editor), "Scientific Papers of C.V. Raman: Acoustics" 1989. *MIT Press*; ISBN: 0262031027

[2] L. Cremer, J.S. Allen. "The Physics of the Violin" 1985. *MIT Press;* ISBN: 0262031027

[3] Woodhouse, J. 1992. "Physical Modeling of Bowed Strings." *CMJ*, 16,4, pp. 43-56

[4] Smith, J.O. III 1996. "Discrete Time Modeling of Acoustic Systems with Applications to Sound Synthesis of Musical Instruments." *Proceedings of the Nordic Acoustical Meeting,* Helsinki, June 12-14

[5] Smith, J.O. III 1996. "Physical Modeling Synthesis Update." *CMJ*, 20,2, Summer 1996 pp.44-56

[6] Borin, G., De Poli, G, Sarti, A. 1992. "Algorithms and Structures for Synthesis Using Physical Models." *CMJ*, 16,4 Winter 1992

[7] Karplus, K., Strong, A. 1983. "Digital Synthesis of Plucked String and Drum Timbres." *CMJ* 7,2 pp.43-55

[8] Jaffe, D.A., Smith, J.O. III 1983. "Extension of the Karplus-Strong Plucked-String Algorithm." *CMJ* 7,2 Summer 1983

[9] Chafe, C. 1989 "Simulating Performance on a Bowed Instrument." *Current Directions in Computer Music Research* - Edited by Matew, M. and Pierce, J. MIT Press 1989

[10] Florensm J-L., Cadoz, C. 1991. "The Physical Model: Modeling and Simulating the Instrumental Universe" *Representation of Musical Signals* edited by De Poli, G., Piccialli, A., Roads, G. MIT Press ISBN 0-262-04113-8 (hc) pp.227-268.

Thursday 24th **h. 11.10**

MUSICAL INFORMATICS: EXPRESSION AND PERFORMANCE ANALYSIS I

The Other Way - A Change of Viewpoint in Artificial Emotions

Antonio Camurri, Pasqualino Ferrentino
music@dist.unige.it

Laboratorio di Informatica Musicale, DIST-Università di Genova (http://musart.dist.unige.it)

Abstract

We introduce a computational model of artificial emotions. The basic motivations and guidelines are discussed, then the paper focuses on two instances of our model we are working on: (i) interactive musical games for education and entertainment in science centers and museums, with particular focus to the Music Atelier at the Città dei Bambini (Porto Antico, Genova) we recently designed and developed; (ii) composition and performance in multimodal environments with emotional agents. We argue that models of artificial emotion and emotional agents can be conceptual tools to investigate new models and paradigms in music composition and performance.

1. Introduction

Roughly speaking, an *emotional agent* is a computer system, possibly including actuators (e.g., a robot), structured in a deliberative (or cognitive), reactive, and emotional modules, therefore able to interact with the external world, including humans and other agents. Agents can cooperate or compete to reach their goals. We are interested in agents able to process or simulate emotional behavior. From one hand, multimedia systems implemented as emotional agents can reach a more effective, natural, and stimulating interaction with humans [1,2], i.e., better user interfaces. From the other hand, agent architectures and models of artificial emotions might contribute to new paradigms in composition and performance. In music composition, our hypothesis - and a motivation to our work - is that models of artificial emotions might be both a conceptual and a realization platform to investigate new paradigms. For example, Grisey's music domain is that of "life". His view on sound as *être vivant* rather than *object* [6] seems close to our view on artificial emotions and emotional agents. In music performance, interactive art and museum applications, a typical domain scenario is the following. Imagine a sensorized space in which you can stand surrounded by music, and control the ebb and flow of that music by your own full-body movement (e.g. dance), without touching any controls. In addition, this space might react "emotionally", not only as a "virtual musical instrument", and might include "emotional" actuators (including physical actuators, e.g., on-wheel robots) behaving as actors or partners on stage [7], or as a mobile scenography. Further applications of this research include interactive museum exhibitions, science centers, theatre and music performance, psycho-motoric rehabilitation (e.g. of autism). Our model and related agent architecture [1,2] has been developed and applied in several multimedia and multimodal application environments [2], as well as in exploratory music composition and interactive performance environments.

2. The change of viewpoint

There are at least two different ways of approaching the creation of an emotional agent, corresponding to two different viewpoints to the problem. The first is to give it artificial emotions in order to communicate with other emotional agents (e.g., a simulated environment, an interactive game, in which there is one, or more, agents controlled by humans and others controlled by the computer). The second is to give it emotions *only* to communicate with humans.

In the former scenario the agent has to deal with two kinds of emotions: emotions from/to humans and from/to agents, and these two types of emotions have to be stored in a coherent way in the agent's mind.

In the latter scenario the only aim is to *convey* real emotions to humans by using *artificial* emotions. This is the relevant scenario for the music domain. The important aspects are the results of your experiment, rather than the "soundness" of the model, e.g., with respect to its analogy to the psychophysics of emotion in human. Such results are not the emotions that the agent has "felt" during the experiment, but the emotions the agent has conveyed to the audience.

In this paper we present a model useful to construct emotional agents using this latter approach. So we will not claim that this model is psychologically sound, since it has been conceived with the final goal in mind: e.g., a mobile robot that navigates on a stage or a museum, "talking" and playing with children in a science centre, showing itself and other interactive games in the exhibition.

The overall architecture of our emotional agent is depicted in figure 1 and described in detail in [1]. Let us here only discuss about the aspect of our model concerning the communication channel between the deliberative (or cognitive) and the emotional agent's components. In our model, the emotional state (later we will give a practical definition of that) can alter the way in which the agent "reasons", that is its deliberative part. This influence can occur in different ways: the deliberative part knows the agent's emotional state (in fact it has read access to the emotional part of the agent) and it has some rules to "reason" about it (e.g., if I am angry I will not do a certain thing). Alternatively, the deliberative part of the agent does not reason about its emotions. The emotional state acts as a perturbation in the inference engine (if I am angry probably I am blinded by anger

and do not reason at all). A more detailed discussion of the architecture shown in figure 1is available in [1].

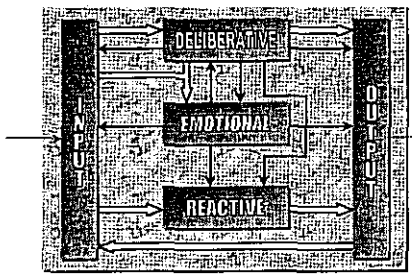


Figure 1: Our proposal of agent architecture embedding our model of artificial emotions (See [1]).

3. Artificial emotions in interaction agents

3.1 The model

We explain our model by means of an example in our museal project at *Città dei bambini* museum (see section 4). It consists of a group of agents - interactive music games. One of them is a "physical" agent (a Pioneer 1 robot from Stanford Research Institute enhanced with environmental sensors and other proprietary software and hardware: see [2] and our www site), acting as a guide and partner in interactive music games for children. It embeds the model of artificial emotions described in this paper.

Our model is inspired by the works of Frijda and Johnson-Laird [4] on human emotions. Frijda's approach says that emotions arise from an observed difference from the world we sense and the one we desire. Johnson-Laird's approach is practical: he propose five basic emotions and the others are only a combination. Our work lies something in the middle. We created thirteen different characters (regions) on a plane (see figure 2). A point in such space represents the *character* of the agent. The two axes represent the degree of affection of the agent towards itself and towards others, respectively. We call these two axes "Ego" and "Nos", from the Latin words "I" and "We". A point placed in the positive x (Ego) axis represents an agent whose character is in a good disposition towards itself. A point towards the left (negative) Ego would mean an agent fairly discouraged about itself.. Similarly for the vertical axis.

The emotion space is usually partitioned into regions: we divided the space into thirteen regions (figure 2), each labelled by the kind of character the agent simulates. A "believable" character does not change too much during time. A "happy" robot (that is, a robot whose character lies in the happiness region) can be angry for a short time (maybe because of an obstacle) but this does not alter its happiness (but if it continues to find obstacles it will eventually get angry). So, in our model we have two levels of rate of change of the character: the *character* that moves in the plane, and the *mood*, that is the "short term character" (very roughly, you can think of the character like of the *point of work* in a transistor and the *mood* like the *small signal perturbations* around that point of work).

Each region is defined by a maximum and minimum threshold for the quantities of the Ego and Nos character value. The minimum are called *Apathy-Ego* and *Apathy-Nos*, for the x and y axis respectively. The maximum are called *Maximum-Ego* and *Maximum-Nos*. The character of the agent is in this example a two-component vector named *Character-Ego* and *Character-Nos*. In other agents (for music applications) the character is a vector of N reals (where N is the number of regions defining the personality), each representing the degree of membership to the corresponding region. In this latter case, we can have a "fuzzy" definition of the character.

We give here a description for two of the thirteen regions available for the robotic agent.

- *Apathy:* (ABS(Character-Ego) < Apathy-Ego) AND (ABS (Character-Nos) < Apathy-Nos)) In this region the robot is apathetic, e.g., the robotic agent has a slow, lazy style of moving and its speech and audio output is fairly sad.
- *Happiness:* (Apathy-Ego < Character-Ego < Maximum-Ego) AND (Apathy-Nos < Character-Nos < Maximum-Nos)). This is the "normality" region, that is, the robot is extrovert and happy enough to meet other people. This region represent happiness *because* the degree of liking towards itself and towards the others is medium.

The other regions follow similar guidelines and are shown in figure 2. It is important to notice that the fact of being in a given region not only defines the character, but also influences the (short term) mood or disposition of the agent.

3.2 The dynamics

The character of an agent can change only by means of internal and external stimuli. This means that the character supports only four high-level messages: "Carrots" (or "candies", positive stimuli) and "Sticks" (or "pain", negative stimuli). Carrots and Sticks can come either from the agent itself or from the external world: so we have Carrots-Ego, Carrots-Nos, Sticks-Ego, Sticks-Nos.

The motivation for having only these two types of messages is rather simple: navigation in a two-dimensional space requires two non-aligned accelerations, and Ego and Nos are orthogonal. Possible extensions of the dimensionality of the emotion space, e.g., toward a 3D emotional space including, for example, a sort of "physical efficiency" axis, only would need a third independent stimulus, still consisting of the two ambivalent high-level inputs (carrots and sticks).

Our current model is not a case of reinforcement learning: the agent is not able to learn (a step in this direction is [7]). The four stimuli are used to explore the agent's emotion space. Our model tries to support also a realistic vision of the character in the sense that it may change abruptly (catastrophic) for a certain

stimulus received in a particular region in a certain state.

As an example, let us examine how the mood class behaves when the agent is in one region. A detailed description of the whole model for all the 13 regions is given in [2,3]:

- *"vanity" region*: here the robot's Ego component is small, while its Nos is over the maximum (see figure 2). This means that a vain agent will not take into consideration messages from himself, because its consideration is small (it is actually apathetic towards itself). Since it corresponds to a region already over the Maximum-Nos component, another Carrot from the outside would cause no effect: it is *insensitive* to others' carrots, it already knows that the world is considering it positively. But what about others' sticks? We are in a region where the agent is used to carrots, so a single stick would be catastrophic: it is *very sensitive* to sticks from the outside here.

The same catastrophic behavior occur in the other eight regions at the edge of the plane. For example, an agent in the Clown region is insensitive to Carrots-Nos and Sticks-Nos, but very sensitive to Sticks-Nos and Carrots-Ego.

4. The Emotional Agent at work

We experimented emotional agents in different real scenarios. The "Città dei Bambini" (Porto Antico, Genova) is a permanent museal exhibition for children, composed by modules corresponding to different visitors' age. We designed and developed the Music Atelier, an installation in the six-fourteen years-old module, about two hundred square meters wide. There is a number of agents-interactive games involving the exploration of sound and music by means of full-body movement. The *emotional robot* is a further game which tours around the space of other games and introduces them to the children, help them in understanding and playing, or simply plays with them. If they are playing good or not, if there is something that they have not found yet, etc. are information detected by the robot by means of environmental sensors and communication among games through a local network. These information are interpreted by the robot as external carrots or sticks. Examples of games-agents are "The real and virtual tambourines", "The virtual musical string", "The harmony of movement" (see our web page for more details). All the games involve movement in sensorized spaces to discover things and concepts related to music. The character of the robot therefore also depends on how children use other games. The emotional state of the robot is communicated to children by means of environmental coloured lights (yellow, red and blue: a happy robot will colour all the environment with full lights; angry / red; depressed/blue; etc.), style of navigation (tail-wagging, slow/fast, nervous sharp vs. smooth trajectories etc.), different speech sentences (describe the same thing with different inflections according to the current emotional state), music and sound.

5. Discussion

Our model is a starting point towards a practical synthesis of emotions. We must not forget that we are the target of the emotion synthesis. We did not create an emotional robot to make it communicate with other robots/agents, but with humans. So, the model must be evaluated for the *human* emotions that the *simulated* ones evoke. So, does our robot communicate emotions to the general audience? Our results from experimental work is encouraging. Our experience confirmed that children are a good test audience, because they surely adapt very quickly to new things, but they too are very good at discriminating the faults in your work and to say it to you more freely than adults. And what about the robot's reactions to children? The robot is *scared* by a big number of obstacles, so its performance gets low if there are more than three or four children around. The performance degenerates because it cannot reach the goal (reaching the next game and explaining it to children, or completing an evolution while navigating in the space). This is a negative stimulus that will eventually make it angry, and then depressed. We observed that small group of children (up to three) are prone to follow the robot and not to disturb it in its tasks. Small groups are often more impressed by the performance. Instead, if a full classroom enters our installation, the robot will not be understood, because it needs more attention. In our museal installation the robot is left alone with children, no intermediate person is active in the space. With a high number of children, the robot change its goal, and instead of accompanying to the other interactive games, it becomes a game partner, touring around with children until a more calm situation is reached (environmental sensors allow the robot to re-adjust its global position in the exhibition and to sense the behavior and activity of children).

Let us now address the question regarding the reusability of our model. If you want to use our model in another scenario, maybe not using a real robot but a software agent, the biggest problem (but also the biggest virtue, we believe) is that our model is rather abstract and high level, although practical. The Ego-Nos plane does only help in the first steps of the problem. The hard problem is how to make a, say, happy agent, i.e., how your agent will eventually show its happiness. You have to design thirteen different ways of behaving, one for each region, *then* our model introduces a complex dynamics between the different characters you have built. We think that the good impression our robot conveys is only in part due to the model that lies "under the hood". The hard part was to define how the robot had to change its behaviour. For example, when it is happy, it moves turning left and right quickly giving the impression of a dog that moves its tail. When it is in the apathy region all its movements are slow and the reactions are delayed (also

environmental coloured lights help visitors in understanding its emotional state). This is not included in the model but it is a main reason for the success or failure of an application using emotions.

5.1 Composition and performance

We experienced our model in music composition and performance [2,5,7]. A number of crucial issues arise. For example, in a live electronics, how many (and which) degrees of freedom can be left to the director of the performance on the agent's emotion dynamics? They can vary from a completely autonomous agent receiving carrots and sticks from the external world including the director, to a direct access of the director to the character point in the emotion space (figure 2), who therefore has a full control on emotions. In a performance for dancer and robot at the concert of the Kansei Workshop [5] (figure 3), a part of the choreography/score concerns the evolution of robot emotions and on stage navigation paths. The dancer and the robot are therefore interpreters of a dance script. In the case of the robot it is a perturbation of such script generated by its emotion dynamics. By means of their movement in the sensorized stage, both the robot and the dancer control - as in live electronics – also parameters of their music performance. Each of them has a different part to be played. The composition consists of a design of a multimodal "dialogue" between them. The composer can modify the robot character dynamics: e.g., a set of emotional attractors can be introduced to influence the dynamics in the emotion space.

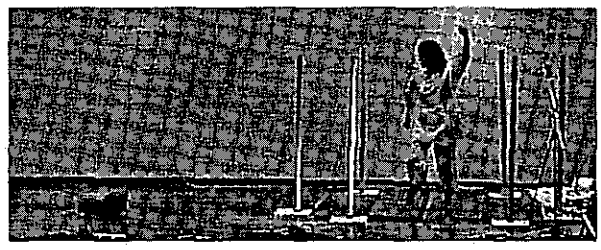


Figure 3. The performance for robot (left) and dancer at the AIMI Intl. Workshop on Kansei [5].

References

[1] Camurri, A., A.Coglio, P.Coletta, C.Massucco, An Architecture for Multimodal Environment Agents (1997) in *Proc. AIMI Intl. Workshop KANSEI: The Technology of Emotion*, 48–53, University of Genova.

[2] Camurri, A., P.Ferrentino (in press). Interactive Environments for Music and Multimedia. *ACM Multimedia.*

[3] Ferrentino, P. (1997) *La Macchina Museale Teatrale: modello e implementazione di un sistema ad agenti con stati emotivi.* Tesi di Laurea, DIST, Università di Genova.

[4] Frijda, N. (1986) *Emotions.* Il Mulino, Bologna.

[5] Camurri, A. (Ed.) *KANSEI: The Technology of Emotion - AIMI International Workshop, Proceedings*, DIST-University of Genova, Italy, October 3–4, 1997.

[6] Orcalli, A. (1993) *Fenomenologia della musica sperimentale.* Sonus Edizioni Musicali.

[7] Suzuki, K., Camurri, A., Ferrentino, P., Hashimoto, S. (1998) Intelligent Agent System for Human-Robot Interaction through Artificial Emotion. Proc. *IEEE SMC'98*, San Diego, CA, IEEE CS Press.

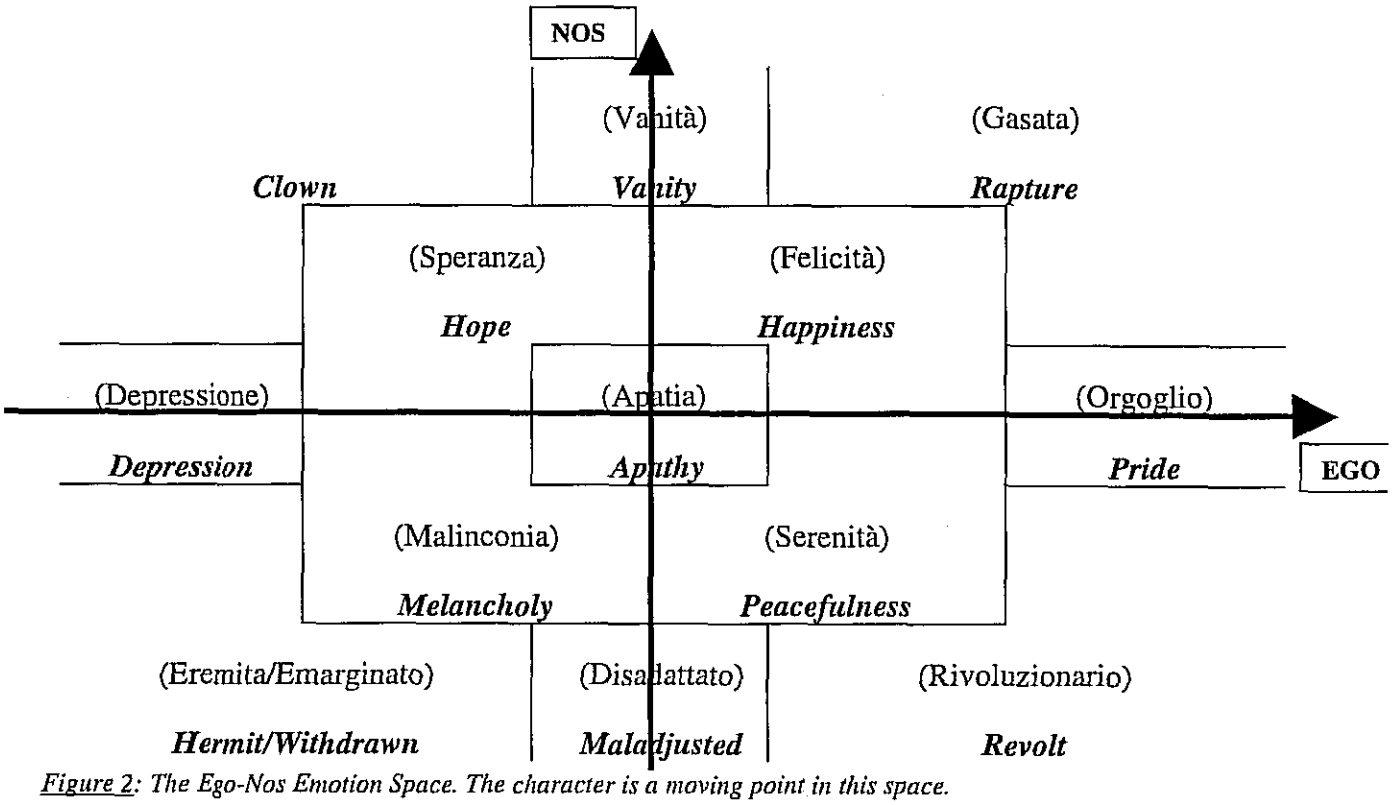


Figure 2: The Ego-Nos Emotion Space. The character is a moving point in this space.

EyesWeb – toward gesture and affect recognition in dance/music interactive systems

Antonio Camurri(*), Matteo Ricchetti(*)(**), Massimiliano Di Stefano(*), Alessandro Stroscio(*)

(*) Laboratorio DIST di Informatica Musicale, Università di Genova (http://musart.dist.unige.it)
(**) EidoMedia produzioni video, Genova (ricchetti@eidomedia.com)

Abstract

The *EyesWeb* project concerns the development of a system for real-time analysis of movement and gesture of one or more humans, with a particular focus to affect or emotion content. Such information is used to control and generate sound, music, visual media, and to control actuators (e.g., robots). The main goal is to explore extensions of music language toward gesture and visual languages. The paper describes the state of the art of the project, the design of the current EyesWeb system modules (hardware and software), their experimenting in public events, and dicusses ongoing developments.

1. Introduction

The general domain scenario include dancers on stage, performers in a live electronics environment, actors in a theatre performance, or the public in a museum or art installation. One of the main requirements is to go beyond the common interaction metaphor of "hyper-instrument" (e.g. [5]; see also the interesting web page of M.Wanderley on gesture research in music *http://www.ircam.fr/equipes/analyse-synthese/wanderle/Gestes/Externe/index.html*). We are interested in interaction metaphors like multimodal dialogue and social interaction including affect and emotional communication [1,2]. In [1,2] we also introduce our projects on *multimodal environments* and *interactive dance/music systems* as extensions of composition and performance environments, including inhabited virtual environments. The *EyesWeb* project is part of these projects. A general goal is the modeling of composition and performance environments in terms of communities of *agents*. In this context, an agent is a computer system, possibly equipped with a physical "body" (e.g., a robot), embedding deliberative, emotional and reactive capabilities, and able to interact with the external world, including other agents [2]. An agent integrates different components: input processing from the surrounding world (e.g., motion capture and sensing), cognitive (or deliberative) processing, emotional processing, output processing (generate sound, music, visual media, control actuators). The external world can be a stage, a museal exhibit – real, virtual, or a mix of both – populated by artificial as well as human agents able to interact, dialogue, cooperate, compete, etc. each contributing to the performance. This focus may imply new perspectives on performance environments and live electronics [9], as well as on composition models, as discussed in [6,1, 2] and in the next section.

The main goasl of the EyesWeb project is to develop a system able to analyse in real time movement and gesture of one or more humans, with a particular focus to affect or emotion communication. EyesWeb is based on multiple b&w (infrared) or color cameras, special electronics, and real-time multi-layered analysis software. We started experimenting with the interaction of gesture and music in 1985 with our system MANI (Music and Animation Interface) and camera-based sensors [3] (special purpose devices - Costel, MacReflex - originally designed for bioengineering). Several other interactive systems based on videocameras have been proposed: see for example Steim's BigEye, the Very Nervous System, and systems developed at Waseda University [7]. Our approach differs from most existing system: (i) we want to extract both two-dimensional and three-dimensional information; (ii) EyesWeb uses multiple video cameras, but is not based on the classical stereo vision approach. A typical on-stage configuration with two cameras consists of frontal and lateral views; (iii) our main focus is not only toward symbol recognition from a posture or gesture vocabulary, but to high-level parameters on expressive intentions in the performance. Existing systems are rather limited from this point of view. Ideally, we want a system able to distinguish the different expressive intentions from two performances of the same dance fragment. Our focus is on "Kansei" [8], i.e. on affect and emotion; (iv) the interaction metaphors, the requirements, and therefore the system architecture, are considerably different from existing systems.

2. Interaction metaphors

Let us consider the following example. An agent is able to extract from a human some gesture and movement features, thereby controlling the generation of sound and music. It is therefore capable of reconstructing "views", and interpret in some way movement and gesture. At the beginning, the agent is a "tabula rasa", nothing is evoked by movement; the system is observing the user. We can imagine that the agent is trying to identify features of the "style of movement" of the dancer. If he starts moving with nervous and rhythmic gestures in roughly fixed positions in the space, therefore evoking the gestures

of a percussionist, the agent, after a few seconds of observation, initiates a *continuous* transformation toward a sort of "dynamic hyper-instrument": a set of virtual drums located in points of the space where the dancer insists with his movement. "Continuous" means for example that neutral sounds begin to emerge and transform progressively into drums, e.g., gradually reducing the attack duration and moving from a default to a specific timbre. The number and the spatial position where the drums are located is decided by the movement of the dancer. Drums timbral and intensity features can be associated with the interpretation of dancer's movements. He is therefore now allowed to play the instrument he has built. Instruments not played for a certain period of time may begin to fade away and loose degrees of freedom of interpretation. Then, the dancer may change his "style" of movement, for example, by gradually reducing the speed of a harsh movement toward smoother gesture. The agent will adapt by continuously changing its behaviour toward a different *context* (again, *continuously* and in a time interval proportional to the amount of change of the dancer style of movement). Transformations can mean a continuous change both in the sensitivity, interpretation of gesture as well in the focus of attention, i.e. a change of the set of movements and gestures observed by the agents, as well as changes in the *causality* between movement and sound and music output.

In another example, the percussion agent emerges and learns behaviors and a character to live in that space (after a sufficient time interacting with the dancer): it really becomes a "clone" of the dancer, semi-autonomous, possibly including graphic animations (e.g., humanoid, see figure 3) besides sound. The dancer therefore can interact with such evoked clone as well as freely move in other places of the stage, and possibly create other clone-agents.

The agents evoked by the dancer should be able to generate and control a music output (and possibly graphic animations) coherently with the gestures and movements, the past state of the performance, and the music rules defined by the composer, including the degree of intervention left to the dancer. This does not mean necessarily strict causal relations as in hyper-instruments. In our example, a transformation might change the music output from the set of user-defined virtual drums into sound textures, where the movement, for example, controls the interpretation and the timbral contour of the textures. The designer of the performance introduces into the system the sound and music knowledge, the compositional goals, the aspects of integration between music and gesture (including a model of interpretation of gestures), and decides the amount of (possible) degrees of freedom left to the agent as concerns the generative and compositional choices. This may imply an extension of the music language toward action, gesture languages, and possibly visual languages. This example raises important issues about new perspectives on the integration of music and movement languages. For example, the director of the performance in a live electronics [9] now cooperates with dancers. Applications based on these ideas have been developed [1,2,6] and experimented with composers and choreographers in public events.

3. Overall System Architecture

EyesWeb consists of a special hardware electronics, a set of software modules including a software library for real-time motion capture and gesture analysis, running on Pentium workstations under the Windows NT operating system. The basic hardware configuration consists of two video cameras, a special proprietary electronics for the real-time capture of both camera signals and their sending to a single Matrox Meteor frame grabber board (model I or II). Figure 1 shows the overall hardware architecture.

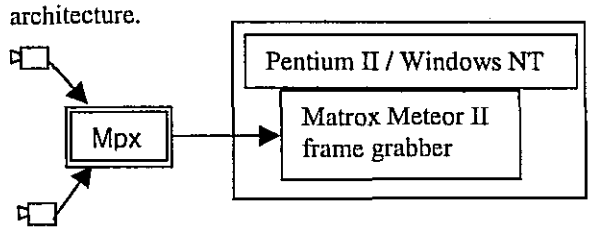


Figure 1. EyesWeb overall hardware architecture

The system supports both color and B&W (infrared) cameras. Interesting results have been obtained with infrared cameras and an infrared (therefore invisible) light system: we obtained significant stability and reliability also in cases of sudden changes in environmental lights during a performance, and even in darkness conditions (see figure 3).

Our proprietary special electronics board (*Mpx*, see figure 1) has been developed to capture concurrently the signal from two synchronized cameras. This board is based on the fact that we can multiplex 2 separate 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 and halfing the temporal resolution.

We use the signals from two camera to have two input views of the same scene: a typical use include a frontal and a lateral view of the stage, to extract

three-dimensional information. Our approach does not concern stereo vision techniques.
The EyesWeb layered software architecture is shown in figure 2. It consists of a *preprocessing module*, layered on *MIL-Lite* (Matrox Imaging Library). It is similar to a shared memory: it gives services to a set of hardware independent software modules for extracting high-level information about movement from the two camera signals. We call such modules *observer agents* (*OAs*). They can run in parallel and generate independently high-level outputs, each corresponding to a particular agent's point of view.

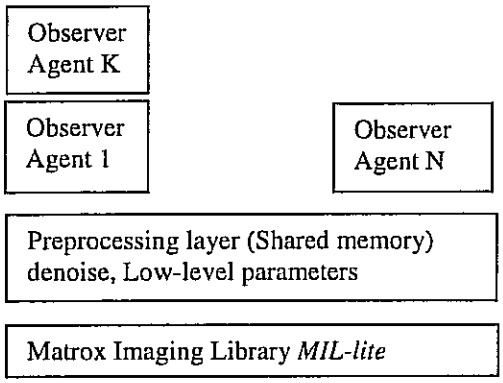


Figure 2. EyesWeb layered software architecture

OAs can read concurrently the data produced by the preprocessing module. OAs can communicate to cooperate or compete to analyse and understand high-level movement and gesture parameters. For example, a posture analysis OA can provide a gesture analysis OA with the recognized posture time marks on the input stream: the gesture OA can use them as candidate segmentation points to start the gesture recognition process.

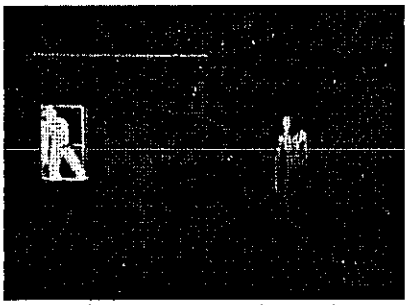


Figure 3. The dancer (right) and its "alter ego" (left), an animated figure projected on a large screen controlled in real time by EyesWeb (concert of the Laboratorio di Informatica Musicale, Teatro Carlo Felice, May 1998).

The low-level parameters and the processed image buffers are directly available to external OAs and to other modules, e.g. to control sound in a live electronics, to control animated human figures or other visual media for system testing or for multimedia performances. The picture in figure 3 shows a public performance with a dancer and its corresponding binary image, visualized on a big screen. The preprocessing module stores frame images as 320x200 buffers of color pixels for each camera. For each frame, it filters noise, extracts the occupation rectangles for each human (figure 4), extracts for each figure a set of low-level parameters (see text below), and converts such rectangles to an internal format. All these data are available to OAs.

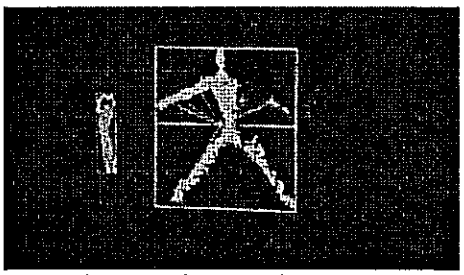


Figure 4. Tracking two dancers. The occupation rectangle is visible for both. Figure in the foreground: the absolute barycentre as the origin of vectors connecting to the relative barycentres of the peripheral parts of the body.

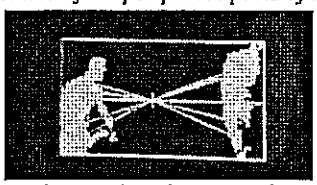


Figure 5. EyesWeb can also observe and analyse a group of dancers as a single entity.

The *posture analysis agent* is based on backpropagation neural networks (NNs). It simply extracts static information from each frame. Several instances of this agent can be active at the same time, each associated to a different camera and, for each frame, to each different human figure in a single frame extracted from the preprocessing module. The NN can be trained directly by users with sets of stereotypical postures. The observed human figure in the current frame is evaluated as belonging to one of learned stereotypical posture or clusters.

The task of the preprocessing module consists of the following steps (for each channel/camera):

0. Preliminary, off-line phase: background is acquired, filtered (denoise), and stored as a reference;
1. For each dancer, his minimum occupation rectangle is computed, available as a buffer to OAs;
2. Two-level quantization resulting from the difference between the current image and the background image (acquired in step 0);
3. Denoise of the resulting two-level image with a threshold based algorithm (using a mask);
4. Trace the sub-image(s) containing the whole human body figure(s) and computedimensions;

5. Compute parameters for each rectangle: percentage of black pixels, processing time, equilibrium, main and relative barycenter vectors, etc. (see text below)

Both color and digital image buffers for each dancer are available to OAs. The "equilibrium parameter" is defined in terms of the distance between the lowest extreme pixels in the rectangle defining the region occupied by the dancer. If he is standing, it corresponds to the last pixel – from left to right - of the right foot of the body and the first pixel of the left foot. The amount of pixels on the ground for each foot (or part of the body on the floor) and the amount of pixels (distance) among feet are also used to compute the equilibrium parameter. The heuristic algorithm takes into account different cases (e.g. crouched, lying down, etc.). Equilibrium is related (inversely) to the "tendency to immediate movement" parameter: roughly, the more distant are the feet and strongly placed on the floor, the more difficult it will be to have sudden and fast full-body movements in the immediate future. Other parameters in the preprocessing module include the minimum rectangle surrounding a human figure, a sort of 2D projection of Laban's kinesphere [8] for each camera view, the barycentre of the pixels occupied by the body in such rectangle, and relative peripheral barycenters (ideally, shoulder, elbow, hand, knee, ankle). We therefore have 10 vectors with origin in the main barycentre (see figure 4), whose module and phase are available to the OAs. Relative and absolute velocities and accelerations, kinetic energy of the movement (by means of heuristics approximation of the mass in terms of weighted pixels according to the body shape from the two camera images) are also available.

4. High-level observer agents

For each camera, more than 70 preprocessing parameters are currently available. Different subsets are useful to analyse different dance contexts. It depends on the type of movement, on which OAs are running, etc. The choice of the significant and reliable parameters in a context is a crucial issue, as well as the management of changes of contexts [2,6]. As for high-level analysis, we developed algorithms based on emotion models (see companion paper in these proceedings and [2]) and NNs (e.g., Kohonen's SOM). In an ongoing project we selected a number of dance fragments (10-15s each). For each fragment, a video with performances differing only in the expressive intentions has been recorded with multiple cameras. This video reference database is used as input for exploratory OAs to recognize expressivity. OAs differs on the type of NN, its input parameters, and its tuning. From a preliminary analysis, it seems that expression invariant and detector parameters emerge for the different performances of the same fragment. See our web site for more details. The approach is similar to the well known approach to perceptual analysis in psychology (e.g. music expression analysis [8]). In this exploratory work we are using the NeuralWorks software for a fast generation, training and tuning of NN-based OA prototypes.

5. Implementation and real time issues

EyesWeb can be integrated with other modules in our architecture [1,2] and with commercial products. It supports communication via MIDI, fast serial port, and a proprietary sockets software library: an agent community can be implemented as a network of computers. This is useful with computatonally intensive OAs, e.g., tasks concerning the control of real time sound synthesis and of graphics and animation. EyesWeb can be be integrated with our sensor-based modules (floor sensors, IR, ultrasound etc. See [1] and our web site) which are part of our architecture and are currently used in public events and museal installations. EyesWeb processing time varies from the number of active OAs and their computational complexity. A typical configuration with two cameras and a few (4-5) OAs takes about 50ms per frame on a Pentium II 333MHz / Windows NT, which corresponds to the processing of 2×20 fps.

Acknowledgements

We thank Riccardo Dapelo, Giovanni Di Cicco and his dance group Arbalete, Claudio Massucco, Giuliano Palmieri, and the students Luca Fraternale, Michela Rossi, and Riccardo Trocca.

References

[1] Camurri, A. (1995). Interactive Dance/Music Systems, Proc. *ICMC-95*, 245-252, Banff.

[2] Camurri, A., P.Ferrentino (in press). Interactive Environments for Music and Multimedia. *ACM MULTIMEDIA* special issue on *Audio and multimedia.*

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

[4] Maletic, V. (1987) *Body Space Expression.* Mouton de Gruyter.

[5] Machover, T., J.Chung, (1989) Hyperinstruments: Musically intelligent and interactive performance and creativity systems. *Proc. ICMC'89*, 186-190.

[6] Camurri, A., M.Leman (1997) Gestalt-Based Composition and Performance in Multimodal Environments. In Leman (Ed.) *Music, Gestalt. and Computing*, 495-508, Springer.

[7] Otheru, S., S.Hashimoto (1992) A new approach to music through vision. In *Understanding music with AI*, AAAI Press.

[8] Camurri, A. (Ed.) *KANSEI: The Technology of Emotion - AIMI International Workshop, Proceedings*, DIST-University of Genova, Italy, October 3–4, 1997.

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

Analysis of Affective Musical Expression With the *Conductor's Jacket*

Teresa Marrin and Rosalind Picard
MIT Media Lab
20 Ames Street, E15-491
Cambridge, MA 02139
marrin, picard@media.mit.edu

Abstract

The *Conductor's Jacket* is a wearable physiological monitoring system that has been built into the clothing of an orchestral conductor; it was designed to provide a testbed for the study of emotional expression as it relates to musical performance. We used the *Conductor's Jacket* to gather and analyze data from a professional conductor in Boston during rehearsals of Prokofiev's *Romeo and Juliet Suite No.2*. Our findings indicate that several forms of expressive communication can be measured and detected in physiological signals. These include the use of *handedness* to emphasize musical changes, the signaling of upcoming events with sudden changes in effort, the difference between information-bearing and non-information-bearing gestures, the indication of intensity and loudness with changes in muscular force, and the use of breathing to express phrasing in the music.

Introduction

Recent work in the domain of affect recognition and physiological monitoring has yielded important results on the nature and expression of human emotions[5]. For example, several early studies have pointed to the presence of a 'contagion effect' whereby emotions can be transmitted from one person to another[2]. The presence of this effect explains why stress can be communicated between people under various conditions; it has also been hypothesized that other states can be transmitted contagiously. The precise mechanism through which this transmission occurs remains unknown, although we suspect that gestures and body language play a big role.

One promising new direction for the study of contagious emotional expression is in the performing arts, particularly in music. Music has often been described as a direct conduit for the communication of emotion; it might be said to be an ideal carrier channel for the transmission of affective information. Correspondingly, musical performers might be said to modulate the structure of musical scores in order to convey affecting and dramatic performances. The act of performing for an audience often requires the performer to project amplified or enhanced emotional states, and to this end performers often train for years to be able to effectively and intentionally express these states. Several early and influential studies on emotional expression discuss this phenomenon with respect to performed classical music[1].

We chose to look at a very specialized form of musical performance, which is optimized for the transmission of emotional and dramatic expression: the role of the orchestral conductor. Conductors use a unique gestural language that combines both technical and affective information about a piece of music in real-time in order to aid those who are performing it. We hypothesize that conductors form expressive intentions for certain pieces that they then convey by means of gestures, and that the affective information is essentially encoded in the carrier signal of the beat-pattern. We hypothesized that the affective content of these signals might be decoded (as by musicians in an orchestra) by noting the difference between the conducted signals and the minimum amount of information that would have been required to execute an *unexpressive* (or minimally expressive) version of the same piece.

In order to test our hypotheses, we designed and built a system to robustly and unobtrusively sense expressive information from conductors under professional rehearsal conditions. This system had to be noiseless, light, not distracting or uncomfortable to wear for long periods of time, and able to withstand punishing conditions of extensive muscular activity, heat, and sweat. The resulting system, called the *Conductor's Jacket*, is a wearable network of physiological sensors that has been custom designed and embedded in clothing that is fitted to the wearer[4]. The jacket contains sensors for heart rate, respiration, skin conductance, temperature, and muscle tension. For muscle tension, we used four electromyogram (EMG) sensors, one on each bicep and tricep. These measure the small voltage created when the muscle generates force; the voltage is proportional to the instantaneous force output of the muscle. All of the sensors are held in place by elastic bands that have been sewn into the cloth of the jacket.

The data we collected supports three major features in the standard conducting technique: the use of the left hand to add emphasis and expressive information, the turning of pages so as to not attract attention or convey musical information, and the use of force in performing a beat gesture to indicate the volume and articulation with which that beat should be played. We also found some surprising results, including several instances where the muscles went limp right before a major event, which suggests that the sudden absence of information has been encoded to signal a 'heads-up' to the players in anticipation of an important future event.

Conductor Study

The first study using the *Conductor's Jacket* system took place during several weeks in February 1998, with a professional conductor during rehearsals of a youth orchestra in Boston. During the few minutes before each rehearsal, the subject fitted the jacket on himself, the sensors were adjusted, and the entire system was tested. Then for the duration of the three-hour rehearsal, the system was used to record numerous files of physiological data timed with the external video camera. Notes were taken during the data acquisition trials, which were used to correlate and analyze the data and video files afterwards.

Initial Data

Initial results indicate several promising features in the data, including clear separation between the expressive use of both hands, context-dependent variations in the respiration signal, and enticing indicators of emotional arousal in skin conductance. Out of more than twenty-two files that were recorded, four have been analyzed in detail and found to contain useful correlations between expressive parameters and the musical score.

In general, the quality of the signals was surprisingly very good. The four EMG signals demonstrated a particularly high signal-to-noise ratio; that is, if there was no observable motion, then the signal was generally almost completely flat. This signal clarity suggests that signal-processing algorithms could be developed to yield good results for the automatic recognition of the above features. Such work remains to be done; however, we present below some preliminary findings extracted from the data by inspection.

Among many observations of the data that have been documented, several features were found to be particularly noteworthy. The following section demonstrates these features with graphical data taken from several rehearsal segments; they have been analyzed for their correlations with known features of traditional conducting technique. These features indicate that our subject:

- used his left hand to provide supplementary information and expression
- suddenly withdrew gestural information when he intended to signal the onset of a major event
- showed fundamental differences in the way he made information-carrying gestures vs. non-information carrying gestures
- modulated the force output of his muscles when generating a beat gesture in order to indicate the overall loudness or intensity of the music at that beat
- modulated his respiration to express the phrasing in the music

In our first two examples, EMG signals from the right and left biceps demonstrate how the left hand was used to provide extra information to supplement the information given by the right hand. In the first example, our subject chose to modulate the meter from 4 to 2. At the moment just before he intended for the meter to change, he reached out his left hand (which was until that moment at his side) and reinforced the new meter with both hands. Figure 1, shown below, demonstrates how the previous faster meter (where only the right hand was used) transitioned to a slower meter as the left hand entered:

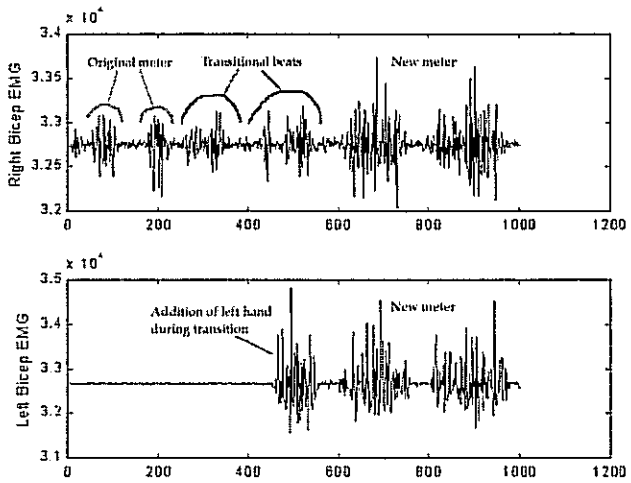


Figure 1. EMG signals from both biceps during a metrical shift.

The top graph shows the use of the right arm; in the first 200 samples of this segment, beats occur approximately every 100 samples. Then, during samples 220-600, the beats begin to transition to a new meter that is one-half as fast. These two beats are subdivided, as if to show both meters simultaneously. During the second of these beats, the left hand enters as if to emphasize the new tempo; this is shown in the bottom graph. Following

this transition, the slower meter comes into relief (beginning at sample 600), with the new beat pattern showing a clearly defined envelope again.

In another section, our subject used his left hand to indicate a drastic reduction in loudness at the very end of the movement. As shown in Figure 2, below, the right hand gave all the beats leading up to the ending, but at the last minute the left hand was used to indicate a quick volume change and a quiet ending:

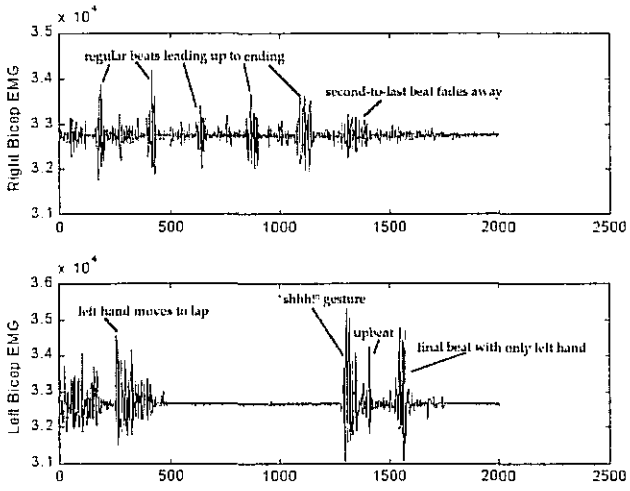


Figure 2. Use of the left hand to indicate drastic change in volume.

In this example, the right hand drops away at the very end and doesn't indicate the final beat. This drastic change in the use of the hands seems purposeful; the video shows that our subject looked directly at the wind section during this moment, as if he wanted to indicate a very different character for the final woodwind chords. As these first two examples have shown, the subject modified the *handedness* of his gestures in order to indicate something unusual.

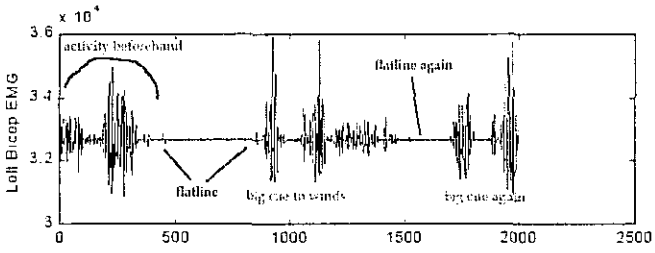


Figure 3. The characteristic "flatline" in the left bicep before a major event.

A second finding indicates that the subject would often withdraw gestural information suddenly before signaling the onset of a major event. That is, his EMG signals (particularly from the left bicep) often went to nearly zero before an important event or entrance. For example, it is very important for conductors to cue the woodwinds after they have waited silently for many measures; if the cue is not clear, they might not start playing in the right place. Such an event happens in bar number 32 of Prokofief's *Dance* movement; many of the winds need to play after many measures of silence. Leading up to this event, our subject used his left hand normally, and then, two measures before the wind entrance, stopped using it completely. Then, just in time for the cue, he gave a big pickup and downbeat with the left arm. In repetitions of the same passage, the same action is repeated. This is demonstrated in Figure 3.

A reasonable hypothesis for why this "flatline" occurs could be that the sudden lack of information is eye-catching for the musicians, and requires minimal effort from the conductor. The quick change between information-carrying and non-information-carrying states could be an efficient way of providing an extra cue ahead of time for the musicians.

A third feature we discovered in the EMG data is that the signals generated by the action of turning pages are inherently different in character from the signals generated by actions that are intended to convey musical information. That is, it seems as if page turns are done in such a way as to purposefully not attract attention or convey musical information. An example page turn is shown below in Figure 4:

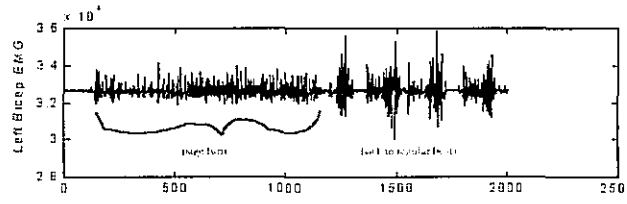


Figure 4. Difference between page turn gestures and information-carrying gestures.

We hope soon to be able to isolate aspects of these (and other non-useful) signals so as to teach a system how to distinguish an information-carrying from a non-information-carrying gesture.

A fourth feature found in the EMG signal is that the amplitude of a beat-generating spike seems to indicate intended sharpness of attack (or perhaps volume) of the notes to be played at that beat. This is compounded by what appears to be a kind of 'predictive' phenomenon, whereby the conductor indicates a very strong beat on the beat directly preceding the intended one. This is often discussed in the literature on conducting

technique, but has never been shown quantitatively to be true. Figure 5, below, shows a segment of Prokofiev's *Dance* movement score with the accents highlighted and aligned with the accents given by our subject:

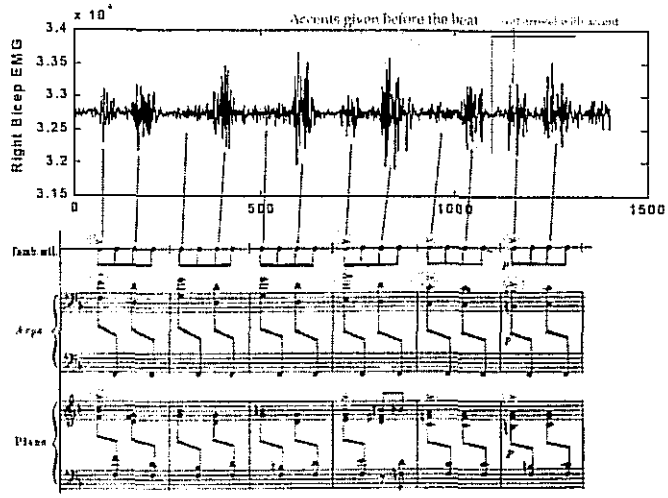


Figure 5. 'Predictive' accents and their relation to the score.

The dark, nearly-vertical lines show how the EMG signals line up with the beats in the score, while the lighter, slanted lines show the relationship between the conducted accent and the accent where it is to be played. The separation line around sample 1100 represents the barrier in-between the first, loud, section, and the second section, which is to be played quietly. The reduced amplitude of the EMG signal right before the separation line could indicate an anticipation of the new loudness level. Also, the existence of high-amplitude EMG signals on non-accented notes in this passage cannot be accounted for musically; perhaps they are due to conductor error. Alternately, they might be accounted for from the perspective of information theory, that once a pattern has been established it does not have to be indicated at every point.

Finally, we found correlations between the respiration signal and expressive aspects of the music. For example, in one musical section, our subject's respiration cycles matched the metrical cycles of the music; when the meter changed, so did his breathing patterns. Secondly, the amplitude of his respiration signal often increased in anticipation of a downbeat and sharply decreased right afterward. This might have been the result of the compression of the ribcage in the execution of the beat, but could also be an intentionally expressive phenomenon. For example, it is considered a standard practice among conductors to breathe in at upbeats and breathe out at downbeats, regulating their air flow relative to the speed and volume of the music.

Conclusions

The first results of the *Conductor's Jacket* project indicate several concrete findings. We are continuing to validate our preliminary results with additional data that has been taken from six other subjects who represent a range of abilities, techniques, and expressive styles.

In addition, a formal analysis of the data has been undertaken, which makes use of techniques from pattern recognition, signal processing (particularly short-time Fourier analysis), and semantic interpretation. Ultimately, we plan to synthesize models for musical performance and expression that incorporate affect and gesture, and that might be useful for the stage. Ideally, our results will be applicable both to professional conductors (enabling the composition of new works for conductors and orchestras where the conductor takes a more *instrumental* role) and to modern, technologically-augmented performers using physiological and gesture-capture systems.

Acknowledgements

We would like to thank the musicians who have participated in our studies, as well as Gregory Harman and Jennifer Healey for their technical assistance, and Professor Tod Machover for his advice. This work is supported in part by the "Things That Think" Consortium at the M.I.T. Media Lab.

References

1. Campbell, D. G. "Basal emotion patterns expressible in music." American Journal of Psychology 1942, 55, 1-17.
2. Hatfield E., J. Cacioppo, and R. Rapson. Emotional Contagion; Studies in Emotion and Social Interaction. Cambridge, UK: Cambridge University Press, 1994.
3. Healey, Jennifer and Rosalind Picard. "Digital Processing of Affective Signals." Appears in ICASSP '98, Seattle, WA.
4. Marrin, Teresa and Rosalind Picard. "The Conductor's Jacket: A Device For Recording Expressive Musical Gestures." Proceedings of the International Computer Music Conference, October 1998.
5. Picard, Rosalind. Affective Computing. Cambridge: M.I.T. Press, 1997.
6. Picard, Rosalind and J. Healey. "Affective Wearables." IEEE ISWC Proceedings vol. 1, no. 1, October 1997.

Friday 25th **h. 11.10**

MUSICAL INFORMATICS: EXPRESSION AND PERFORMANCE ANALYSIS II

HOW COMMUNICATE EXPRESSIVE INTENTIONS IN PIANO PERFORMANCE

Giovanni Umberto Battel
Conservatorio B. Marcello di Venezia
San Marco 2810 - 30124 Venezia
Tel. +39 (41) 5225604 - E-mail gubattel@adria.it

Riccardo Fimbianti
CSC-DEI University of Padova,
Via S. Francesco, 11 – 35131 Padova
phone: +39 -49-8273757 E-mail rf@csc.unipd.it

Abstract

To interpret, from the Latin *interpretâri*, literally means *to act as a mediator, to negotiate*. The performer is the *mediator*, the intermediary between the composition and the listener. Such *mediation* allows different degrees of freedom. Each musician *interprets* the musical sign translating it into sound, but thanks to the freedom this shift leaves him, he can, at the same time, communicate his own expressive intention to the listener. This research tries to establish if and in which measure a specific expressive intention is communicated and which degree of freedom the performer has at his disposal in this *mediation*.

With this purpose, we analysed several interpretations of the same excerpt, each one chraracterized by a different expressive intention to establish how the performer modified the performance parameters of his instrument. By sonological analysis, we found some parameters which permit to characterise different performances of the same piece.

1. Introduction

Music is defined as the art that expresses itself through sounds. Through sound, therefore, the musical work is conveyed to the listener. But the sound, the medium which allows the listener to enjoy the musical work, at least in classical music, is not directly generate by the composer. He uses a paper *algorithm*, the score, which the performer decodes through the instrument he plays to generate the sound and the musical work. In informatics, the *algorithm/performer* bond is deterministic, but in music it allows different degrees of freedom giving rise to diatribes about the definition of musical interpretation as a creative act *tout court*. It would not be possible to make an algorithm able to substitute the musician but, analysing various recordings, it would hypothetically be possible to understand the degrees of freedom that the score leaves to the artist. An understanding of such limit is important both for musicians and the scientific research in this field. The peculiarity of science is not to know perfect mechanisms allowing to control phenomena but to know and therefore control its own margin for error. The artist's creativity and the possibility to communicate his own expressive intention run along this margin. In this paper we took into consideration only some expressive directions, commonly adopted by musicians, which stand in this hypothetical creative space.

In order to study the sonorous result of each performance, we need to know on which parameters the performer intervenes. The performer often varies the interpretation of the same piece in relation to the more or less conscious definition of different expressive intentions. This research aims at the understanding of the modifiable executive parameters and the way it may be possible to intervene on them to make clear the different expressive characterisations to the listener.

Recent studies on the subject [5, 6] show how the performer, modifying the parameters available in accordance with the instrument, tends, consciously or not, to point out an understanding of the musical structure he plays. The communication of the composition structure does not exhaust and explain all interpretative choices.

Expressive intention analysis concentrates exactly on this aspect, evaluating on which parameters the performer intervenes to convey this *general musical intention* to the listener.

To this purpose, we chose eight adjectives typical of the sensorial and emotive sphere, already analysed in a previous significant study. This research completes and improves the previous one overcoming some limitations such as the study of a single performer and the analysis of a single melodic line. In this study we present the analysis of a complete performance (melody, accompaniment and pedals) of various pianists.

As in the previous research, the comparison is not between the mechanical performance and the player's one, but between a *natural* performance, produced by the artist following his own ability and experience and a performance characterised by a particular expressive intention.

1. Experiment

To acquire data we used a Yamaha Disklavier. This is a normal piano equipped with sensors able to gather three categories of movements in the mechanics, that allowed to obtain three kinds of parameters to be studied: 1) IOI (Internet Onset Interval) expressed in ms; 2) Dr (duration of the note key On/Off) expressed in ms; 3) Key velocity with a range 0-127. Furthermore, the disklavier allows to reproduce at the piano the synthesised versions made for musical checking. In this way, there is no information loss due to the use of sampled sound and during the performance the musician feels at ease, playing the instrument he normally uses.

Five students of Piano at the *Conservatorio Benedetto Marcello* in Venice took part in the experiment with the following procedure:

on January 10th 1997 they received a copy of the first sixteen measures of W. A. Mozart's 2° tempo Andante of the Sonata Facile K.V. 545 in C Major and were asked to prepare the performance for on January 22th 1997, correlating it to the following adjectives: **br**ight-clear, **d**ark-gloomy, **h**eavy-massive, **l**ight-gentle, **h**ard-strict, **s**oft-tender, **p**assionate and **f**lat. The students were informed that the recordings would be the object of a numerical analysis.

2. Sonological parameters analysis

Previous studies revealed how the expressive intentions are communicated to the listener [1,3,4]. On the other hand it exists some theoretical model of score analysis.
In the research here presented, we tried to correlate the sonological analysis with the models theorised in literature and to extend the results and the models to expressive intentions.
In the analysis of the sonata K545 the parameters studied were those available in the instrument and which the disklavier permits to analyse : IOI, Key velocity, Dro (Duration Offset) and pedal use. In the analysis of such parameters various levels have been differentiated: the overall level relating to the whole performance, the medium level relating to phrasing and the local or note level relating to the single deviations of each event.

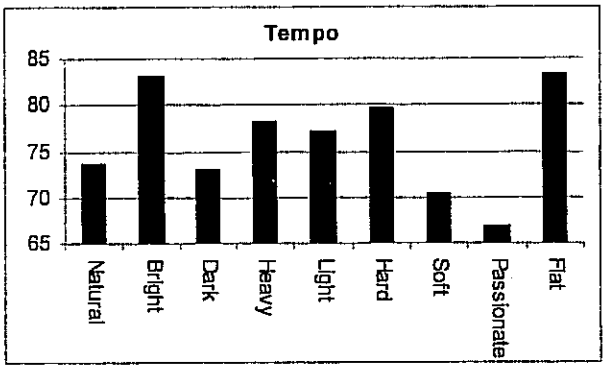


Fig.2.1: Average tempo in different expressive intentions.

With regard to the temporal parameters, the average metronome was studied for the overall level. To a statistics analysis this reveals to be a significant parameter ($p<0,0002$), that is to say, the tempo chosen by the pianist is directly connected to the expressive intention he wants to communicate.
Later, we studied the phrase temporal progress which is characterised by an accellerando/ritardando as already stated in literature. In the analysis of such parameter we referred to timing measured in quarter note. The sonata, in fact, has a ¾ tempo and we supposed the pianist to keep time in quarter note.

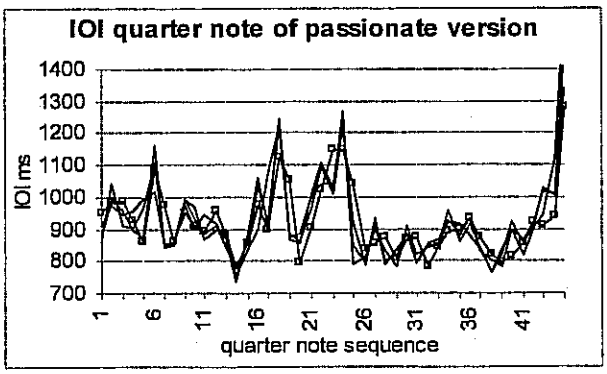


Fig. 2.2: Quarter note in passionate version. The three curves are relative to three subjects. The marked one is the real quarter note value of the pianist's passionate version.

A test was made in order to prove this aspect. We asked some CSC researchers to listen to the K545 piano recordings and to beat time in a keyboard to obtain a sound feedback. They listened to their own beat (C4) and at the same time they listened to the sonata, synchronising themselves perceptively. The results proved the uniformity between the sonata tempo and the perceived one as shows Fig 2.2.
Phrase analysis, with reference to quarter note, proves that each phrase is characterised by an acc/rit. Besides, the analysis-by-synthesis method proved that the degree of acc/rit is significant to characterize expressive intentions. For instance, performing the passionate version of the Sonata, the musicians emphasized the acc/rit which underlines the phrasing.
From the results, it can be assumed that the performer varies the degree of acc/rit correlated to musical phrases to characterize the expressive intention. This supports the hypothesis that performer may use the musical structure to convey his own expressive intention.

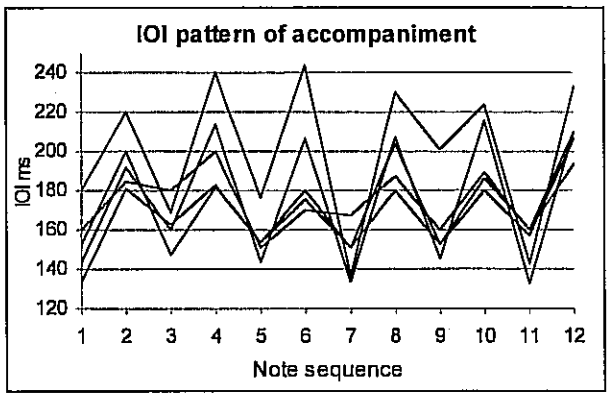


Fig. 2.3: Typical downbeat/upbeat pattern relative to the 5° bar of the sonata K545, bright version. Each curve refers to a different pianist.

In a subsequent analysis of the local temporal parameter we found that the pattern of accompaniment upbeat/downbeat is related to the different expressive intentions ($p<0.0001$).
The dynamics parameter was studied with the same method and the above mentioned levels used for the temporal parameter.

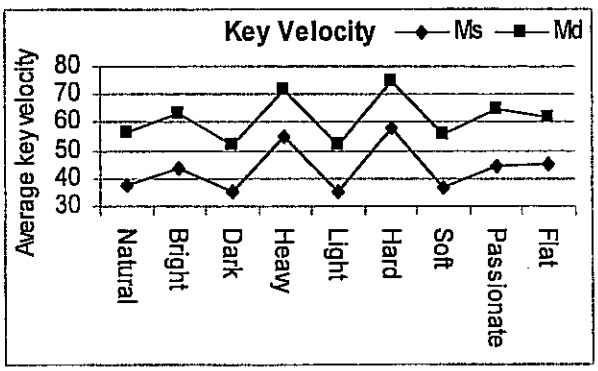


Fig. 2.4: Average key velocity in right (Md) and left (Ms) hand in expressive intentions analysed.

Average intensity, the crescendo/decrescendo degree and the downbeat/upbeat pattern ($p<0.0001$) allow to characterise different expressive intentions. In this context, average intensity was analyzed separately for melody ($p<0.0001$) and accompaniment ($p<0.0001$). Data proves the accompaniment average is always smaller than the melodic one, no matter which is the expressive intention. This is not a parameter the performer uses to communicate his expressive intention and it probably depends on the normal performance practice, at least in this context.
Studying articulation, we examined the Dro parameter, represented as the ratio between the note value (key on/off) and the IOI. The average Dro ratio, calculated on the whole performance, revealed to be a parameter characterizing the different expressive intentions ($p<0.005$). In this case no correlation was found between such parameter and the phrase conduction.

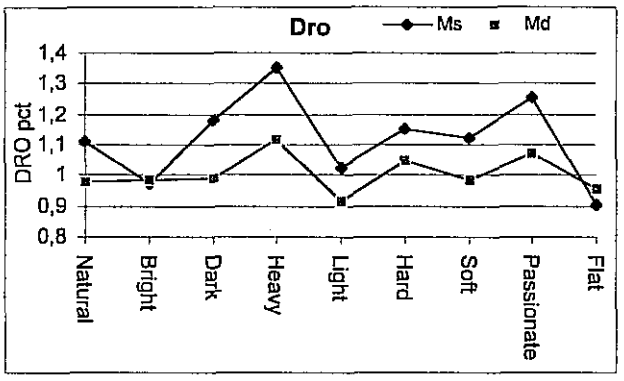


Fig.2.5: average DRO of right (Md) and left (Ms) hand.

The difficulty derives from the pause that is often found at the end of each phrases, and it is hard to discriminate from final retard or last note DRO. As in dynamics, we distinguished analysis between melody and accompaniment. In this context the ratio between melody DRO and accompaniment DRO seems to be significant ($p<0,025$) in the communication of the expressive intention as in Fig. 2.5. To emphasise his own expressive intention the performer not only varies the performance average DRO but also the ratio between the melody and the accompaniment.
Pedal analysis showed that they too are significant for expressive purposes. Putting in abscissa the onset and in ordinate pedal value (range 0-127) and integrating, the area becomes a significant parameter ($p < 0,0001$).
We also analysed the time synchronism between melody and accompaniment. It is here difficult to define a parameter valid for the whole performance. A first result shows that the global average (onset melody – onset accompaniment) is always negative, that is, the melody is normally in advance with respect to the accompaniment, but such average is not a significant one. On the contrary, the synchronism average related to each performer proved to be significant ($p<0,0001$). Eliminating the values that overcome standard deviation, the average of the remaining ones is statistically significant ($p<0,004$) and therefore correlated to the expressive intention conveyed. The values correlated to each pianist seems to be those overcoming the standard deviation. Moreover, these values presents a high degree of correlation ($r>0,5$, $p< 0.05$) within an expressive intention in different performers.
From these results we can infer that there are some parameters that can characterise the different expressive intentions and other that depend on musical structure, and don't change. The following table shows the results of characterising parameters:

	PA	BR	DA	SO	LI	HE	FL	HA
Tempo	↓	⇑	↓	↓	↑		⇑	↑
Acc/rit	⇑	↓	↑	⇑	↑		↓	↓
D/U beat	⇑			↑		↓	↓	
Int med		↑		⇓	↓	⇑	↓	↑
Cres/decr	⇑	↓			↑	↓		⇓
D/U int	⇑	↑			↑	↓	⇓	
Av. Dro	↑	⇓	⇑	↑	↓	↑	↑	↑
Dro m/a	↓	⇑	↓	↓		↓	↑	
Right pd	⇑	⇓	↑	⇑	↓	↑	↓	⇓
Left pd	↓	↓	⇑	↑	↑	↓	↓	⇓
Sincro		↓	↓			↑	↑	↑

Tab 1: the tab. shows the results of parameters relating to each expressive intention. The rows indicate the tendency of parameter to have high or low values. On the first column there are respectively the following parameters: tempo, degree of acc/rit, degree of downbeat/upbeat, average intensity, degree of crescendo/decrescendo, downbeat/upbeat intensity, average DRO ratio, use of right pedal, use of left pedal, synchronism.

Structural elements analysis

In the analysis of a musical performance sonological parameters, the starting point is given by the values of the notes on the score in order to compare them with the actual musician's performance ones. Above all, it is important to be able to split, at least ideally, the deviations due to the score from those depending on a different expressive intention. In this study the attention was driven towards those deviations depending on the different expressive intentions. To point out a possible correlation between the score and the different performances, we carried out an harmonic, rhythmic and melodic analysis. Lerdhal tension/relaxation method [6,7] was applied to the harmonic analysis. Each note is given a tension value and it is then possible to obtain the phrase segmentations to which correspond the accellerando/ritardando put in evidence by sonological analysis. Lerdhal analysis also showed a new correlation between harmonic tension value and local or note DRO. A close survey proves that the staccato/legato degree in each version depends on tension. Particularly, the DRO

average of all performance notes depends on the expressive intention while, for each note, local value depends directly on the harmonic structure. The performer increases the slur degree when the harmonic tension increases in each different intention, no matter the global average. Such characteristic is independent from the expressive intention and let us suppose that it is part of the musician's cultural background.
Other methods were applied to reveal the melodic and rhythmic progress of the score in order to find more significant correlations. The method which led to the best results is the Cambouropoulus' LBDM [2]. Applying Gestalt principles of similarity and proximity this method gives a weight to each note effecting a scansion of the score. Analysis proved a significant correlation between the value that the LBDM method attributes to the note and local DRO of each note.
In this context, too, the correlations were analyzed in order to understand if they could also depend on a different expressive intention, but, as in the previous case the result was negative. It seems that the performer increases the staccato value proportionally to the value attributed to the notes by LBDM, despite the expressive intentions and their legato value.

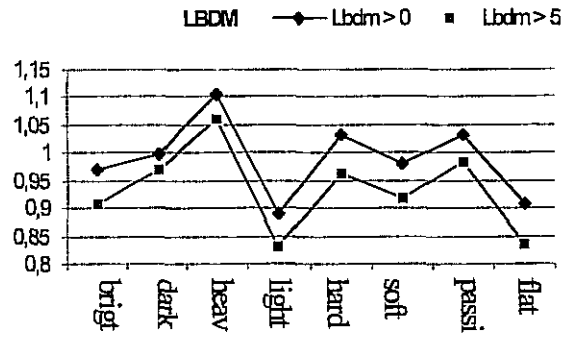


Fig. 2.5: different values of average DRO in different expressive intention, correlate with different value give by LBDM method.

Fig. 2.5 shows two series of average values. The Dro > 0 and Dro > 5 series refer to the Dro average of the notes presenting an LBDM value superior to zero and to five respectively, which shows how the curve translation value does not depend on the expressive intention.

3. Conclusion

Sonological analysis proved that some performance parameters are influenced by the different expressive intention and others are correlated to musical structure as accellerando/ritardando connected to phrasing and finally, others are independent from it.
The score gives various degrees of freedom to the performer but the musician cannot leave out of consideration the performance standards suggested by musical tradition and by his own culture. As previously hypothesized, the performer cannot leave out of consideration the musical structure, on the contrary he tends to underline it and communicate it to the listener in every expressive intention he wants to convey.
There are also parameters the performer can modify. The major result of this research is to have found that such parameters are modified in the same manner by different performers. This allowed us to reach the conclusion that there are parameters allowing to characterize the same excerpt with a different expressive intention. Moreover, the uniformity of the data obtained by the analysis seems to prove that there is an objective *sonorous climate* common to the performers: consciously or not, they operate on precise parameters to convey, for instance, brightness; the control of such parameters not always lies in what is psychologically regarded as consciousness. Analysis showed in which measure the different performance parameters are modify in order to convey a different expressive intention.
So, this research allowed an understanding of those objective performance parameters that make this characterization possible. The analysis of the results made by musicians arises a deeper understanding of the performance, and, at the same time, arises considerations that stray into perception and, at least for the time being, go beyond the limits of these researches.

References

[1] G. U. Battel, R. Fimbianti 1997. "Analysis of expressive intentions in pianistic performances" Proceeding of Kansei – The technology of emotion 1997, 128-134.

[2] Cambouropoulus E. (1996), "Musical rhythm: inferring accentuation and metrical structure from grouping structure". *Proceedings of JIC96-Brugge*, 15-22

[3] Canazza S., De Poli G., Rinaldin S., and Vidolin A. 1997. "Sonological analysis of clarinet expressivity," in M. Leman (ed) *Music, gestalt, and computing. Studies in cognitive and systematic musicology*, Berlin, Heidelberg: Springer-Verlag

[4] Canazza S., De Poli G., Roda' A., and Vidolin A. 1997. "Analysis and synthesis of expressive intentions in musical performance". Proc. ICMC '97, Thessaloniki

[5] Friberg A. 1991. "Generative rules for musical performance: a formal description of a rule system," Computer Music Journal, 15(2): 56-71.

[6] Krumhansl C. L. (1996), "A perceptual Analysis of Mozart piano sonata K.282 : segmentation, tension, and musical idea". *Music Perception*, Vol. 13, n°3, 319-363.

[7] Lerdahl F. 1996. "Calculating Tonal Tension", Music Perception, 13(3): 319-363

ADDING EXPRESSIVENESS TO AUTOMATIC MUSICAL PERFORMANCE

Canazza Sergio, De Poli Giovanni, Di Sanzo Gianni, Vidolin Alvise
Dipartimento di Elettronica e Informatica
Università di Padova - Via Gradenigo 6a – 35100 Padova – Italy
{canazza, depoli, vidolin}@dei.unipd.it.

Abstract

In this paper we will present an overview of a study about expressivity in music. Acoustic and perceptual analyses made on different performances, played it with different expressive intentions, suggested by a set of sensorial and affective adjectives, we will present. The acoustic analyses shown the acoustic parameters which separate the different performances of the same score. Perceptual analyses confirmed that listener's experience and performer's intention were basically agreed.

A model of expressiveness has been developed on the base of results of analyses. The model allows to obtain different performances, by modifying the acoustic parameters of a given neutral performance. The modification is performed by algorithms which use the hierarchical segmentation of the musical structure. Opportune envelope curves are applied, for every hierarchical level, to the principal acoustic parameters. Level's self-similarity is the main criteria for the envelope curves construction. The rendering steps can be realized both with synthesis and post-processing techniques.

1 Introduction

Different musicians, even when referring to the same score, can produce very different performances. The score carries information such as the rhythmical and melodic structure of a certain piece, but there is not yet a notation able to describe precisely the temporal and timbre characteristics of the sound. The conventional score is quite inadequate to describe the complexity of a musical performance so that a computer might be able to perform it. Whenever the information of a score (essentially note pitch and duration) is stored in a computer, the performance sounds mechanic and not very pleasant. The performer, in fact, introduces some micro-deviations in the timing of performance, in the dynamics, in the timbre, following a procedure that is correlated to his own experience and common in the instrumental practice. It is exactly for this great variety in the performance of a piece that it is difficult to determine a general system of rules for the execution. An important step in this direction was made by Sundberg and co-workers [1]. They determined a group of criteria which, once applied to the generic score, can bring to a *musically correct* performance. Further on, the performer operates on the microstructure of the musical piece not only to convey the structure of the text written by the composer, but also to communicate his own feeling or expressive intention. Quite a lot of studies have been carried on in order to understand how much the performer's intentions are perceived by the listener, that is to say how far they share a common code (see [2] for references). Gabrielsson & Juslin [2] in particular, studied the importance of emotions in the musical message. In this context, we tried to understand the way an expressive intention can be communicate to the listener and we realized a model able to explain how it can be possible to modify the performance of a musical piece in such a way that it may convey a certain expressive intention. A group of sensorial adjectives was chosen (hard, soft, light, heavy, bright, dark) which should inspire a certain expressive idea to a musician. A musician, inspired by appropriate adjectives, produces, systematically, different performances of the same piece. Perceptual analysis proved that the audience can indeed perceive the kind of intention he wanted to convey. Acoustic analysis confirmed that there are micro-deviations in the note parameters. We outlined models to connect such deviations with the intention wanted. Following the analysis-by-synthesis method, some musical synthesis were produced to verify and develop a model of musical expressiveness.

This paper, starting from the results of the acoustic and perceptual analysis, presents the design of a model able to add expressiveness to automatic musical performance. These studies on the model of musical performance are interesting not only from a scientific point of view, but also from an practical one, both in the field of automatic musical performance and in general in the multimedia systems.

This research was supported by Telecom Italia, under a research contract Cantieri Multimediali.

2 Perceptual Analysis

Perceptual analyses were carried out to observe the listeners' judgment categories and to verify if performers succeeded to convey to the listeners the expressive intentions.

Seven different interpretations (*light*, *heavy*, *soft*, *hard*, *bright*, *dark*, and *normal*) of a fragment of theme of Mozart's K622 Concert for Clarinet, were performed by a professional clarinet player. The recordings were carried out in three cycles, each cycle consisting of the seven different interpretations. The musician chose, then, the ones that, in his opinion, best corresponded to the proposed adjectives, trying to minimize the influence that the order of execution might have had on the performer. The recordings were carried out at the CSC of Padua University in the monophonic digital form at 16 bits and 44100 Hz.

We made an experiment [4] to determine the judgement categories used by subjects called in to listen

to the various interpretations of the same musical piece. The test was carried out on a group of 24 subjects, 12 musicians graduated at the Padua Conservatory, and 12 subjects without specific musical preparation.

The subjects were asked to describe the performances along 17 scales of evaluation adjectives of sensorial nature: **black** (nero), **oppressive** (greve), **serious** (grave), **dismal** (tetro), **massive** (massiccio), **rigid** (rigido), **mellow** (soffice), **tender** (tenero), **sweet** (dolce), **limpid** (limpido), **airy** (aereo), **gentle** (lieve), effervescent (spumeggiante), **vaporous** (vaporoso), **fresh** (fresco), **abrupt** (brusco), **sharp** (netto).

This list of adjectives did not contain those used in the performances and did not include their opposites.

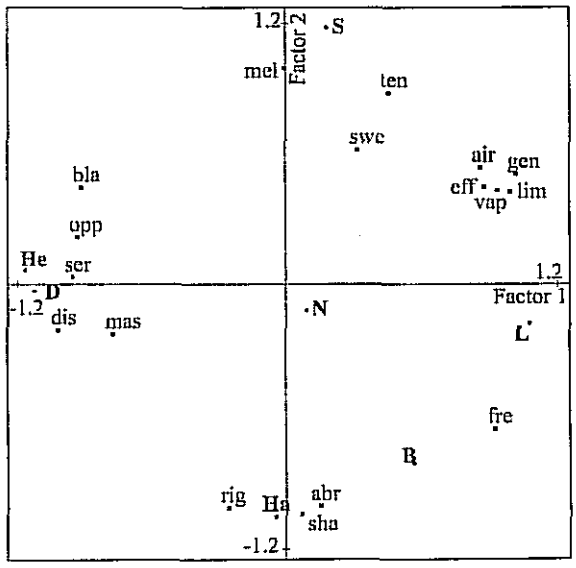


Fig. 1: factor analysis on the adjectives. The first factor explains 60% of the total variance, the second 27.2%.

They were chosen so as to offer the subjects a exhaustive sampling of a semantic space.

Factor analysis on adjectives allowed us to determine a semantic space defined by the adjectives proposed to the listeners. The performances could be placing placed on it, according to factor scores, in order to observe in which semantic sector they map. Two significant components (i.e. with eigenvalues greater than 1) accounted for 87.2% of the variance. Varimax rotation was used in order to simplify the factors' interpretation.

Fig. 1 shows the position of the evaluation adjectives in the resulting space. As can be seen, adjectives associated with the extremes of these factors were for factor 1 *dismal* vs. *limpid*, for factor 2, *mellow* vs. *rigid*. By means of factor scores, it was possible to insert the performances into this space. The comparison of the *positions of performances with* that of the evaluation adjectives demonstrated a good recognition of the performers intentions by the subjects. For instance soft (S) performance is placed near mellow, tender and sweet adjectives. Moreover it can be noticed that normal performance is placed near the center of the space, far from all the adjectives.

These results were similar both for musically trained and untrained subjects. However *Cluster analysis* of answers showed a behaviour distinction between the group of trained musicians which was highly cohesive and the group of non musicians which showed a greater variance in the judgements given.

3 Acoustic Analysis

The principle aim was to identify the relationship in the physical parameters modifications when the *expressive intention of the performer* was varied.

Every musical instrument has its own expressive resources (vibrato in strings, the tongue in wind instruments, etc.), which are used by the musician to communicate his expressive intention. It is inevitable, therefore, that the results of any sonological measure depend, not only on the score, but also on the characteristics of the instrument used and the choices effected by the musician. Consequently, it is necessary to compare the data relative to different scores, musicians and instruments, in order to identify the expressive rules that can be considered valid in a general way and which are specific cases.

Some acoustic analyses have been carried out on various musical pieces using different instruments and performers. Up to now, performances involving the clarinet [3], the violin [4] and the piano [5] have been analyzed.

Not all the results can be compared as some of the parameters measured were defined differently in the studies. Besides, quantitative comparisons are sometimes not very significant as the absolute values of parameters depend on the technical characteristics of the instrument used. Nevertheless, it is possible to make a qualitative comparison at least as far as the mean tempo (MM), legato (L), note attack time (DRA) and brightness (BR) of spectrum are concerned. Our main aim in this study is to determine, at least at a general level, those common strategies used by musicians to communicate their expressive intentions. Table 1 and 2 show the tendencies of parameters measured in the clarinet and violin performances and table 3 shows the qualitative results of parameters measured in piano performances. In the second case, it was only possible to compare MM and Legato parameters, considering the different technical characteristics of piano. It can be seen how, notwithstanding some differences, the pieces referring to the various expressive intentions have a similar behaviour in the different experiments. For instance, the adjective bright induced the musicians to perform their piece with a quicker metronome, a lesser legato, and a shorter attack time. In the piano, in fact, a high key velocity means a quicker attack. The main differences among the experiments have to do with a different choice in the expressive resources used, but not with a different use of these resources. In the soft version, for instance, the clarinet performer played with the values of the MM (low), DRA (high) and BR (low)

parameters significantly different from the other versions. The violinist played in the same way as far as the BR (low) parameter is concerned, but unlike the clarinettist, he modified the MD-MA (difference between the amplitude at the start of decay and end of attack) and UDR (ratio between upbeat and downbeat duration) parameters. It is worth noting that a high MD-MA value means an amplitude profile slowly raising, while a high legato value in piano, together with a low key velocity, leads to an equivalent qualitative result. The only conflicting result regards the heavy piece performed by the clarinet player with a different use of the parameter L. In this case, it seems that the clarinettist used a quick note attack time and a slow metronome, causing in the listeners a sense of heavy locomotion; but the violinist and the pianist tried to communicate a sense of effort in moving things.

	N	Ha	S	He	L	B	D
MM		high		low	high	high	
L				high	low	low	
DRA		low			high	low	low
BR		high	low	high	low		low
UDR			low		high		
MD-MA			high		low		high
VR						high	low

Tab. 1: Behavior of statistically significant parameters on varying expressive intentions in violin performances, Arcangelo Corelli's Violin Sonata in A Major, V Op. [4]

	N	Ha	S	He	L	B	D
MM			low	low	high	high	
L	high			low	low	low	
DRA		low	high	low		low	
BR		high	low	high	low		

Tab. 2: Behavior of parameters on varying expressive intentions in clarinet performances Mozart K622 [3]

	N	Ha	S	He	L	B	D
MM				low	high	high	low
L	low		high	high	low	low	high
KV		high	low	high	low	high	

Tab. 3: Behavior of parameters on varying expressive intentions in piano performances, Mozart k622 [5]

4 Architecture of the model

The researches we have been making [3] and [4] prove that the performance of a piece following a certain expressive intention can be described observing which variations take place with reference to a neutral and a nominal performance of the same piece. By *nominal performance* we mean the mechanic performance of the score where the metrical durations (the score) are accurately observed and by *neutral performance* we mean a literal human performance of the score without any expressive intention or stylistic choice.

The model developed [6] is able to obtain an expressive intention, transforming a neutral performance both with reference to the score and the acoustic signal itself. It must be underlined that our approach provides for the adoption of hierarchical structures similar to the spoken language ones (words, phrases), in the musical language. Once these structures are recognized, it is possible to modify the parameter of a group of notes (for example metronome or intensity) following a certain curve. Such curve describes the characteristic of the musical gesture on the group of notes. It is therefore convenient to describe (appropriately codified) the information about the neutral performance and the nominal performance (i.e. the score), the variations to be applied on the expressive traits of the single note (timing, intensity, timbre), the subdivisions of the piece into expressive units (words, semi-phrases, phrases) characterized by curves that modify one or more parameters of the notes that constitute them.

To this aim, we propose a new representation of the score, where the fundamental components and parameters of a musical piece are highlighted. Moreover it is provided with a number of controls on expressive parameters that allow the model to operate on the piece. Later we shall refer to this new score as metascore. The metascore is a file where the information about both a nominal performance and a neutral performance are codified. The parameters of the neutral performance are expressed as deviations from the nominal performance. The performances are read by a MIDI file and transcribed in the metascore. From the parameters of the

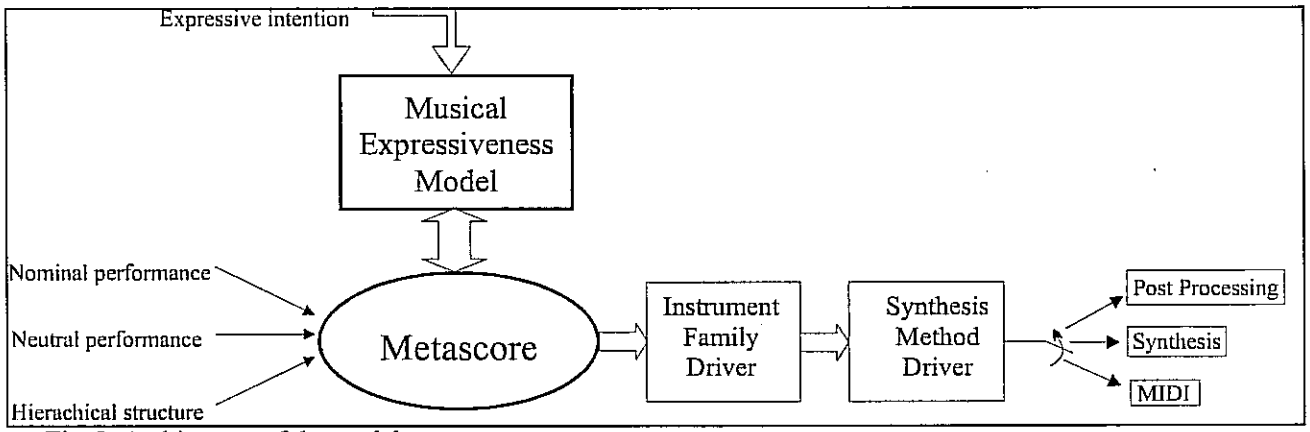


Fig. 2: Architecture of the model.

MIDI protocol, the model computes new parameters that describe the attack, the intensity, the spectral characteristics and other physical attributes of each note. Thus, the basic parameters of each note are immediately accessible. Each of these parameters is expressed in perceptual scale, in which the unit represents the difference between two perceptual levels (e.g. the difference between *f* and *ff* for loudness).

These parameters are independent from the particular musical instrument that will play the score. The importance of the notes in the piece can be specified by parameters such as the accent and the elasticity. This last parameter takes in account that expressive deviations are not made to the same extent in each note because the score sets technical and structural limits that prevent the musical phrase from being distorted [4]. The metascore also requires, as input, a description of the hierarchical structure of the piece (fig. 2). As already mentioned, we should not see the notes of a piece as units independent one from another: each piece has its own structure where the notes gathers to form the words or phrases of a musical discourse. For instance it is possible to set the metronome profile and the intensity on groups of notes following the division in words, semi-phrases and phrases. At the phrase level it is possible to specify an adjective inspiring the performance.

With reference to the adjectives and considering other factors such as elasticity, the model computes the deviations from neutral performance in order to render expressive synthesis. Output of this first computation is still independent from the instrument that will play the score. It is then necessary to adapt the modify metascore to a particular family of instruments and then to a specific synthesis method (fig. 2). The modular structure of the system defines an open architecture, where the rendering steps can be realized both with synthesis and post-processing techniques.

Different synthesis techniques, like FM or Wavetable, have been explored. Expressive synthesis of pieces belonging to different musical genres (European classical, European ethnic, Afro-American) verified the generalization of the rules used in the model.

As example of a piece codified and performed through the model we present A. Corelli's Violin Sonata in A Major, V Op (score in fig. 3). We shall show now the graphics of amplitude envelops of some clarinet performances obtained (through post-processing) thanks to the controls given by the model. In figure 3a the neutral performance is shown. In figure 3b, the hard performance and in figure 3c the soft performance obtained using time-frequency techniques in order to bring about the transformations calculated by the model.

Fig. 3 Violin original: **neutral** performance a). Violin obtained through post-processing, using time-frequency techniques, from the neutral performance: **hard** performance b); **soft** performance c).

5 Conclusions

In this paper has been presented an overview of a study about expressivity in music. Studies on musical expressiveness ([3], [4], [5], and [6]) made clear which are the choices made during performance in order to give a certain expressive intention. A new coding for the score (the metascore) suitable to the automatic performance of a musical piece was studied. The model was provided with a series of controls working on the single note. Besides, special attention was given to the importance of working on groups of notes, hierarchically ordered, and significant for the performance of the piece. The metascore thus obtained is not dependent on the instrument. The model processes this metascore in order to particularize it to a particular instrument family. Besides the fact that it was developed mainly for western classical music, the model showed a general validity in its architecture, even if it needs some tuning of the parameters.

References

[1] Friberg, A. "Generative Rules for music performance: a formal description of a rule system". *CMJ*, 15(2), pp. 49-55. 1991

[2] Gabrielsson, A., Juslin, P. "Emotional expression in music performance". *Psychol. of music*, 24. pp 68-91. 1996

[3] Canazza S., De Poli G., Rinaldin S., & Vidolin A. Sonological analysis of clarinet expressivity. In: M. Leman (Ed.) "Music, Gestalt, and Computing ". Berlin: Springer-Verlag. pp 431-440. 1997

[4] Canazza S., De Poli G., Roda' A., & Vidolin A. "Analysis and synthesis of expressive intentions in musical performance". In *Proc. of the ICMC 1997*. pp. 113-120. Tessaloniki: ICMA. 1997

[5] Battel G.U., Fimbianti R. "Analysis of expressive intentions in pianistic performances". In *Proc. of the Int. Kansei Workshop 1997*. Genova: Associazione di Informatica Musicale Italiana. pp. 128-133. 1997

[6] Canazza S., De Poli G., Di Sanzo G., Vidolin A. "A model to add expressiveness to automatic music performance". In *Proc. of the ICMC 1998*. Ann Arbouri: ICMA. 1998 [in press]

HOW ARE EXPRESSIVE DEVIATIONS RELATED TO MUSICAL INSTRUMENTS? ANALYSIS OF TENOR SAX AND PIANO PERFORMANCES OF "HOW HIGH THE MOON" THEME

Nicola Orio

CSC-DEI Università di Padova
Via Gradenigo 6A 35121, Italy
Email: orio@dei.unipd.it

Sergio Canazza

CSC-DEI Università di Padova
Via Gradenigo 6A 35121, Italy
Email: canazza@dei.unipd.it

Abstract

In order to understand the nature of expressive deviations introduced by the musicians, they were recorded two sets of seven different performances of the jazz standard "How High the Moon", respectively played on a tenor sax and on a piano. The musicians were asked to not introduce variations on the score like grace notes or modulations. Each performance was driven by a different sensorial adjective. Since expressivity has its sense only when there is a listener, some perceptual tests were made. Factor Analysis, Cluster Analysis and Multidimensional Scaling were carried out to evaluate the acoustic parameters that better characterize the different expressive performances. Subjects succeeded in recognizing the different expressive intentions; moreover acoustic analyses pointed out which are the parameters mostly relevant from the perceptual point of view.

1 Introduction

The process of composing, performing and listening to music can be viewed as a communication chain. The content and the form of the message, as well as the way the message is receipt, are strictly related to the musical culture shared by the elements of the communication chain. Every musical style refers to a common cultural heritage, which depends from different historical and geographical conditions. In Fig. 1 is summarized this process for classical western music repertoire.

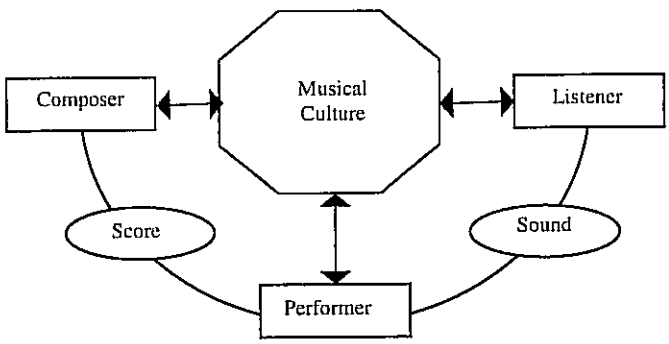


Fig. 1: Communication chain in classical western music

The sender is the composer, who translates her musical ideas into a score; the score is then interpreted by a performer, who makes a new translation of written score into musical performance. Finally the listener perceives and interprets the performance. In the scheme in Fig. 1 is also highlighted the role of the musical culture as a point of reference for all the elements of the communication chain. Not only the structure of the message, but also the structure of the communication chain itself varies with the repertoire. For instance in improvised music composition and performance are part of a single process. This is typical of jazz music: during the improvisation a jazz musician is continuously creating new musical phrases. Moreover, even during the exposition of the theme, the musician is free to modify completely the written score, adding notes or changing the rhythm structure, that is to compose a variation on the theme.

Even if the communication chain can vary and the message can have different forms, it is a well known fact that one of the main purposes of music is to communicate emotions. But this aspect is still not well studied from the scientific point of view.

To convey emotions, performers can introduce deviations from the indications written in the score in timing, in dynamics, in timbre, and in articulation; expressive deviations are, generally, different according to the musical genre, to the used instrument, and to the performer. From this the difficulty to create a system of rules for the automatic musical performance [1]. Kendall and Carterette [2], for instance, founded a considerable variety in different performances even on very short phrases, so that they deduced that measured data "failed to support something as strict and invariant as *the* musical grammar, performer grammar, or listeners grammar".

All of these works pointed out the need of studies about how a performer represents the music and how the related expressive intentions influence the performance. To analyze completely the communication chain it is also necessary to study the listener's experience, that is how performer's intentions are captured by the listener and if there is a common way to code musical emotions. Concerning this, the importance of emotional aspects in the musical context was deeply studied by Gabrielsson and Juslin [3].

This work presents a study on the communication between the performer and the listener in jazz music. Jazz repertoire is particularly suitable for this kind of study; in fact the performer has a lead role in the

communication chain, especially during the improvisation. Moreover jazz music is mainly based on the performance of *standards*: musical pieces, often composed for musicals in the first half of this century, that through the years loosed their own character to become a simple harmonic and melodic structural reference for the developing of improvisation. A jazz musician is free to make many modifications on the written score, to express her feelings better; hence the communication with the listener is mainly based on performer's expressive intentions rather than on the musical structure of the piece.

2 Experiment

With the aim of pointing out the expressive deviations introduced by a musician in the performance, two jazz players, the tenor saxophonist Maurizio Caldura and the pianist Marcello Tonolo, were asked to separately play seven different versions of the first eight bars of the jazz standard *How High the Moon* written by Hamilton and Lewis (plotted in Fig. 2).

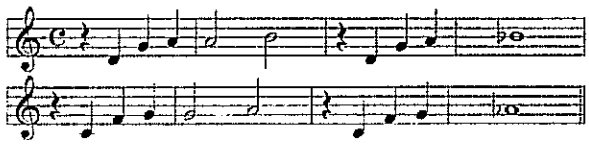


Fig. 2. How High the Moon (Hamilton-Lewis); first eight bars

The two musicians played in two different sessions: saxophone performances were digitally recorded with a sampling rate of 48 kHz, while piano performances were recorded in MIDI format. The choice of recording this two particular instruments, both typical of the jazz repertoire, allows to compare the expressive deviations performed on a monophonic and on a polyphonic instrument.

Some constraints were imposed to the performers. Musicians could change only the rhythm structure of the score, without introducing any variation on the written notes (like grace notes or temporary modulations); the pianist were free to choose a different accompaniment for each performance, in order to analyze the strategies offered by a polyphonic instrument.

Six of the performances were driven by a sensorial adjective. The proposed adjectives are: *bright, dark, hard, soft, heavy, light*. These adjectives were chosen to exhaustively sample a sensorial semantic space.

The seventh performance was driven by the term *normal*, meaning that the musicians had to play without a particular character and in complete concordance with the written score.

3 Methods

A group of 17 listeners was asked to score the performances. Perceptual tests were carried out on saxophone and on piano performances. Subjects who participated to the tests were 8 jazz musicians and 9 common listeners, that is without a particular knowledge of jazz repertoire.

First of all, the subjects listened to all the recordings of saxophone performances driven by sensorial adjectives in random order. After listening to all the stimuli, the subjects were asked to listen again, how many times they liked, to each performance and to give a description of their emotions using a group of 17 evaluation adjectives: practically they have to put a cross for each evaluation adjective on a graduated scale going from *nothing* to *extremely*. The complete procedure was repeated for piano performances.

The evaluation adjectives were chosen to exhaustively sample a sensorial semantic space. They were chosen among the synonyms of the driving adjectives. Evaluation adjectives are: *Airy, Abrupt, Sweet, Fresh, Serious, Oppressive, Gentle, Limpid, Massive, Black, Sharp, Rigid, Mellow, Effervescent, Tender, Dismal, Vaporous*.

Multivariate statistical analyses were performed on subjects' answers. First of all a Cluster Analysis were developed to homogeneously group the subjects. Hence a Factor Analysis was carried out on perceptual measurements of performances; the analysis allows to observe how the listeners disposed the musical stimuli in their mind, especially pointing out how many dimensions are used for the performances classification. In this way it is possible to see which are the judgment categories that are used by the listeners to differentiate the performances. Another analysis on data was made transposing the answers matrix and carrying out, again, a Factor Analysis. This to test how stimuli were mapped, with the factor scores, in the semantic space defined by the evaluation adjectives. Factor Analysis on evaluation adjectives is also a useful test on how they are able to correctly sample the semantic space. All these analyses were developed also in the groups of subjects that the Cluster Analysis pointed out as homogeneous.

Acoustic analyses were developed on the recorded performances to deduce the most relevant parameters that characterize the expressive intentions. Then the results obtained from perceptual tests were related with the measures obtained from the acoustic analysis, in order to observe which are the acoustic parameters that are more important from a perceptual point of view.

4 Perceptual analyses

Perceptual analyses were carried out to observe the listeners' judgment categories and to verify if performers succeeded to convey to the listeners the expressive intentions.

All the analyses were separately carried out on saxophone and piano performances. Cluster Analyses were developed to test the consistence of data obtained by perceptual measurements; a hierarchical method was adopted, precisely the centroid method, using the squared euclidean distance as a separation criterion. Both dendrograms, of saxophone and piano performances, showed the presence of a single cluster, in which are present almost all of the musicians plus, respectively, three and two untrained. This result is

analogous to what was highlighted in a previous work on expressivity in tonal western repertory [4]: musical praxis seems to give a common way to code expressivity. Musicians, which are used to express themselves with non verbal communication, give similar answers to non verbal stimuli. Further analyses were developed both on the whole subjects' answers and on cluster's ones. In this paper only results on the cluster's answers will be presented.

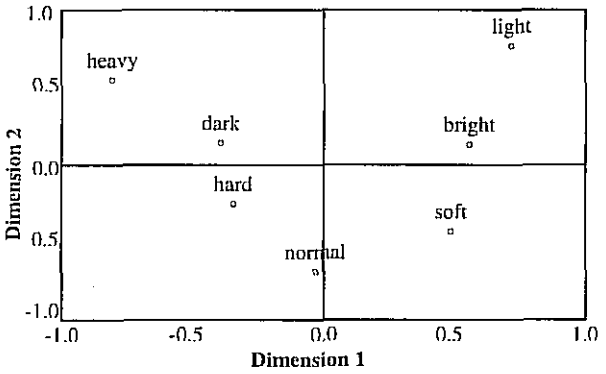


Fig. 3: Multidimensional Scaling on piano performances: there is a good separation among stimuli

A Factor Analysis was developed on subjects' answers, using the Principal Components Algorithm; the stimuli, that is the driven recordings, were

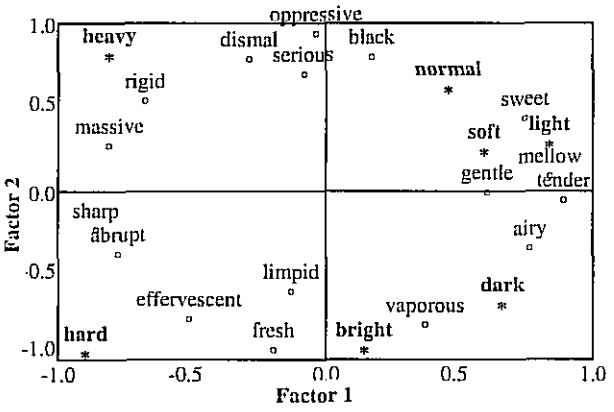


Fig. 4: Performances plotted, by factor scores, on semantic space derived from factor analysis on evaluation adjectives: performances map close to their synonyms.

considered the observable variables. In both analyses, saxophone and piano, emerged that two principal factors are enough to explain more than 80% of the global variance, meaning that listeners mentally organized the performances in a two-dimensional space. In saxophone analysis, the first axis is related to the separation between *heavy* and *bright,* and the second axis to *hard* versus *soft* and *light.* The *normal* performance is mapped close to this last two performances. In piano analysis there is not such a good correlation between axes and performances, which are mapped in two groups, *hard-heavy-dark* versus *soft-light-bright*, with the normal approximately in the middle.

Stimuli were plotted in a two-dimensional euclidean space through Multidimensional Scaling. In both cases there were a good separation among the performances, hence the musicians succeeded in giving a different interpretation to each performance and listeners understood the different expressive intentions. As an example, results on piano performances are showed in fig. 3.

Finally a Factor Analysis, using the Principal Component Algorithm, was developed considering evaluation adjectives as the observable variables. This analysis helps to define the semantic space to which the subjects were referring to; moreover it is possible to plot also the performances in this space, using factor score of the stimuli, in order to observe the way the subjects mapped the performances inside the semantic space. Both analyses point out that the evaluation adjectives are a good sampling of a semantic space of sensorial nature: they approximately map in a circle centered in the middle of the semantic space. There was a good recognition of the expressive intentions: stimuli are spaced and well distributed in the plane. Moreover the stimuli are usually mapped close to evaluation adjectives that are synonyms. Fig. 4 shows, for instance, the results for saxophone performances: **soft** is close to *mellow* and *sweet*; **bright** is between *limpid, fresh*, and *vaporous*; **heavy** has a big factor score both in the *black/oppressive* and in the *sharp/massive* factors.

5 Acoustic analyses

Acoustic analyses were developed on the recorded performances to evaluate the parameters that performers systematically changed to give a particular expressive intention. These results have to be compared with the ones obtained by perceptual analyses. In fact to understand what is important in the communication of expressivity, they have to be evaluated the parameters perceptually relevant to listeners, which is different than simply analyzing the acoustic signal [5], [6]. Hence, even if all the acoustic parameters were measured, here are presented only the ones which are more related to expressive intentions.

5.1 Sax analyses

The Tempo was the first measured parameter. The structure of the theme *How the High the Moon*, has the same rhythmic pattern repeated twice in the first eight bars; hence it was measured the average distance between the onsets of each couple of corresponding notes. This parameter is related to the opposition between *hard* and *soft,* moreover it maps the performances coherently with the second perceptual axis highlighted by Factor Analysis developed on the stimuli. In Tab. 1 are quoted the values in Beats per Minute (BPM).

The second acoustic parameter that was found related to perceptual axes is the Duration Attack (DRA). It was measured as the mean time, in ms, between the 10% to the 90% of the amplitude envelope of the notes;

in Tab. 1 are also quoted the values in ms. DRA is mostly related to the opposition between *heavy* and *light*. From the comparison with the results shown in Fig. 4, it emerges that DRA has a good correlation with the first factor. These results are in keeping with the ones obtained in tonal western music [7].

BPM		DRA (ms)	
Soft	129	Heavy	23.4
Normal	134	Hard	25
Heavy	150	Bright	26.3
Dark	155	Dark	33
Light	176	Normal	35.6
Bright	184	Soft	46.9
Hard	214	Light	94.3

Tab. 1: Beats per Minute and Duration Attack time of the performances

5.2 Piano analysis

Piano performances were recorded in MIDI format. Hence the acoustic analysis on data had to be developed differently, because the measurable parameters are only the KeyOn, KeyOff, and KeyVelocity values. Therefore it is not possible to analyze the DRA. Anyway it is well known that also the acoustic piano has a low amount of expressive parameters [5].

The most relevant parameter, from a perceptual point of view, is Tempo (see Tab. 1). It is related to the first axis obtained by MDS, as it is shown in Fig 3, where the opposition between *heavy* and *light* is pointed out.

Heavy	Section	Dark	Hard	Soft	Bright	Light
77	96	114	118	138	181	273

Tab. 2: Measure of Tempo in piano performances

Among the measured parameters, also articulation and loudness were found as perceptually relevant. The dispersion plot obtained by these two parameters is quoted in Fig. 5. From the figure it is clear that the pianist used both of them to differentiate the performances depending on the expressive intentions.

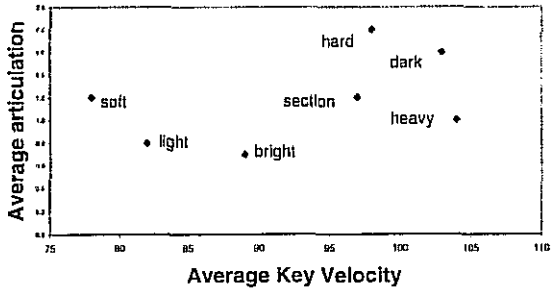


Fig. 5: Average key velocity vs. average articulation in piano performances

6 Conclusions

In this paper it has been presented an overview of a study about expressivity in jazz music. From the perceptual and acoustic analyses of a set of performances, driven by sensorial adjectives, it emerged that: a) musicians succeeded in differentiating the performances from a perceptual point of view (see, for instance, Fig. 3); b) the differences among performances were measurable, as acoustic analyses highlighted; c) it was possible to have a good sampling of a semantic space of sensorial nature, in which the listeners' judgment categories could be focused; d) the expressive intentions were recognized by listeners, in particular if they were musically trained: hence there is a common way to code expressivity in music, helped by musical praxis.

Acknowledgments

This work was supported by Telecom Italia, under the research contract *Cantieri Multimediali.*

References

[1] Friberg, A., Frydén, L., Bodin, L. G., & Sundberg, J. "Performance rules for computer controlled contemporary keyboard music". *Computer Music Journal*, 15(2), 49-55. 1991

[2] Kendall, R. A., & Carterette, E. C. "The communication of musical expression". *Music Perception*, 8, pp 129-164. 1990

[3] Gabrielsson, A., & Juslin, P. Emotional expression in music performance. *Psychology of music*, 24. pp 68-91. 1996

[4] Canazza, S., De Poli, G., & Vidolin, A. *Perceptual analysis of the musical expressive intentions in a clarinet performance*. In: M. Leman (Ed.) "Music, Gestalt, and Computing - Studies in Cognitive and Systematic Musicology". Berlin, Heidelberg: Springer-Verlag. pp 441-450. 1997

[5] Battel, G.U., & Fimbianti R. "Analysis of expressive intentions in pianistic performances". In *Proceedings of the Int. Kansei Workshop 1997.* Genova: Associazione di Informatica Musicale Italiana. pp. 128-133. 1997

[6] Repp, B. H. "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei". *Journal of Acoustic Society of America*, 92(11). pp 2546-2568. 1992

[7] Canazza, S., & Orio, N. "How are player ideas perceived by listeners: analysis of 'How High the Moon' theme". In *Proceedings of KANSEI - The Technology of Emotions*, AIMI International Workshop, Genova. pp 134-139. 1997

A MODEL OF DYNAMICS PROFILE VARIATION, DEPENDING ON EXPRESSIVE INTENTION, IN PIANO PERFORMANCE OF CLASSICAL MUSIC

De Poli Giovanni, Rodà Antonio, Vidolin Alvise
Dipartimento di Elettronica e Informatica
Università di Padova – Via Gradenigo 6a – 35100 Padova - Italy
depoli@dei.unipd.it; ar@csc1.unipd.it; vidolin@dei.unipd.it

Abstract

We asked five pianists to perform several different versions of the same score, inspired by a set of sensorial and affective adjectives. An analysis of the dynamics profiles shows that notable differences can be recognized between the different versions of the same score. In spite of that, same relation can be found between the dynamics profile of the different musical interpretations. This feature allowed us to formalize the relation between the profiles by means of a limited number of parameters. All the performances were mapped into a two-dimensional space, the dynamics parametric space. The results show that most of the performances, inspired by the same adjective, are grouped together in the same region. Each region of the space can be, therefore, associated with a specific adjective. The model was applied on various piano scores. The results show that the model has good generalization attribute and can properly render the dynamics characteristics of performances on varying of performer's expressive intentions.

1 Introduction

It is known that several performances of the same score often differ significantly, in particular when the musicians are instructed to play it with different expressive intentions [1]. In this context, expressive intention is taken to mean the inspiration given to musician through adjectives in order to obtain different expressive performances. According to the played instrument, the performer can use various musical means (timing, dynamics, amplitude envelopes, vibrato, tongue etc.) to express his/her interpretation of the score. This work deals with dynamics profiles, i.e. the values of note intensity during the performance. The aim of this paper is the discussion of the following questions: 1) how do dynamics profile change when a musician is asked to play drawing inspiration from a particular expressive intention? 2) is there any common performance strategy if different musicians are inspired by the same expressive intention?

2 Model

We asked five pianists (called pianist A, B, C, D, and E) to play the first 16 bars of the second movement of Mozart's piano sonata K545. The musicians performed several different versions of this score, inspired by a set of sensorial and affective adjectives: natural (na), bright (br), dark (da), hard (ha), soft (so), heavy (he), light (li), passionate (pa), and flat (fl). All the pianists played the Yamaha Disklavier and the performances were recorded in MIDI format.

Figure 1 shows key-velocity values measured in the nine performances of a single pianist. Each curve (called dynamics profile) represents the set of values measured in a single performance. In order to simplify the discussion, we reported only the pianist A's data, even if the following comments are true also for the other pianists. Dynamics profiles allow us to know the exact course in time of key-velocity. Due to the large amount and variability of data, however, they don't allow an easy comparison among the musicians' performance strategies. To this end, it is necessary to define a model that allows a parametric description of the different performances. By means of the model parameters, it will be possible to highlight and compare the main expressive characteristics of the performances.

The model is based on the observation that the score structure suggests suitable behaviors to the player. In order to emphasize some elements of the music structure (i.e. phrases, accents, etc.), the musician changes dynamics by means of expressive patterns as crescendo, decrescendo, sforzando etc.; otherwise the performance would not sound musical. Many works analyzed the relation or, more correctly, the possible relations between music structure and dynamics [2], [3], [4], [5]. The fact that there are many different interpretations of the same score [6], however, shows that musician keeps many freedom degrees beyond this relation.

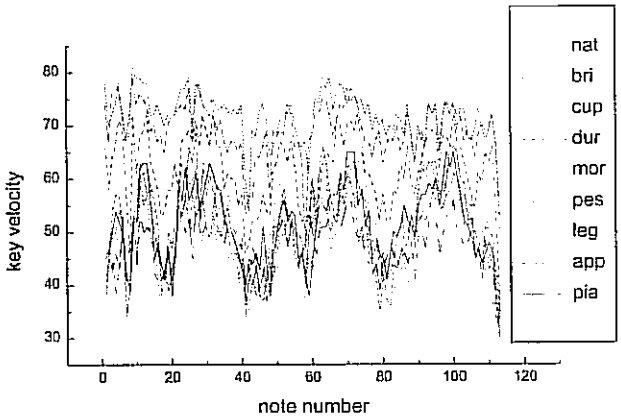


Fig 1: key-velocity in the nine performances of pianist A. The values were measured by means of Yamaha Disklavier.

	br	da	ha	so	he	li	pa	fl
na	0.55 \ 0.91	0.59 \ 0.87	0.45 \ 0.83	0.50 \ 0.82	0.28* \ 0.79	0.61 \ 0.87	0.58 \ 0.81	0.41 \ 0.84
br	-	0.35 \ 0.86	0.60 \ 0.85	0.42 \ 0.84	0.50 \ 0.79	0.51 \ 0.86	0.54 \ 0.82	0.45 \ 0.83
da	-	-	0.43 \ 0.82	0.52 \ 0.80	0.31 \ 0.77	0.54 \ 0.84	0.43 \ 0.80	0.51 \ 0.81
ha	-	-	-	0.38 \ 0.74	0.52 \ 0.86	0.42 \ 0.81	0.41 \ 0.79	0.45 \ 0.77
so	-	-	-	-	0.37 \ 0.62	0.57 \ 0.81	0.49 \ 0.75	0.42 \ 0.76
he	-	-	-	-	-	0.40 \ 0.76	0.27** \ 0.71	0.51 \ 0.74
li	-	-	-	-	-	-	0.61 \ 0.85	0.46 \ 0.84
pa	-	-	-	-	-	-	-	0.42 \ 0.69

Tab 1: minimum and maximum correlation coefficients calculated between the nine performances of each pianist ($p<0.001$ for all performances except * $p<0.003$ and ** $p<0.004$).

The hypothesis for the application of the model is that: when we ask to a musician to play in accordance with a particular expressive intention, he works on the available freedom degree, without destroying the relation between music structure and dynamics [7]. A proof of this hypothesis can be found in the dynamics profile of figure 1, where the structure of the score is the same for all the nine performances. If the relation between music structure and dynamics don't change, many common patterns could be observed among the profiles. To this end, the correlation coefficients between the different versions of each pianist were calculated.

Table 1 shows, for each pair of adjectives, the minimum and maximum correlation coefficients calculated for the five pianists. A significantly correlation can be noted between all the adjective ($p<0.004$). This result implies that all the profiles of each pianist have a similar shape, which we assume to be depending on music structure. Figure 2, in which dynamics profiles were normalized to zero mean and unitary variance, clarifies this observation. The relation between dynamics profiles and the main elements of music structure is particularly evident: for instance (see also figure 3) the musician emphasized with a *decrescendo* the end of the first *inciso* (bar 2), the first semi-phrase (bar 4), the first phrase (bar 8) and the period (bar 16).

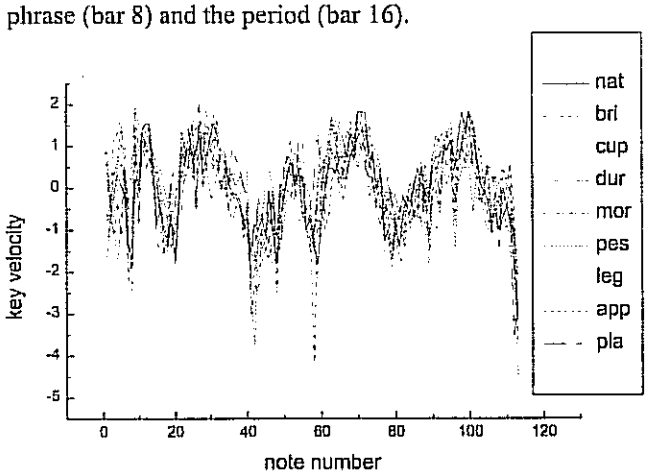


Fig 2: normalized dynamics profiles of pianist A.

The model exploits the idea that all the dynamics profiles can be obtained from an input profile (which agrees with the music structure) by means of some elementary transformations. We have now to define what is the input profile and which are the necessary transformations. We used, as input profile, the average of the dynamics profiles measured in the nine performances (figure 3). Since the mean calculation puts in evidence the performance common characteristics, which are supposed to depend by music structure, the average profile keeps intact the relation between structure and dynamics. Moreover, it represents the geometric center of gravity of the nine performances, which property we will discuss in the next paragraphs.

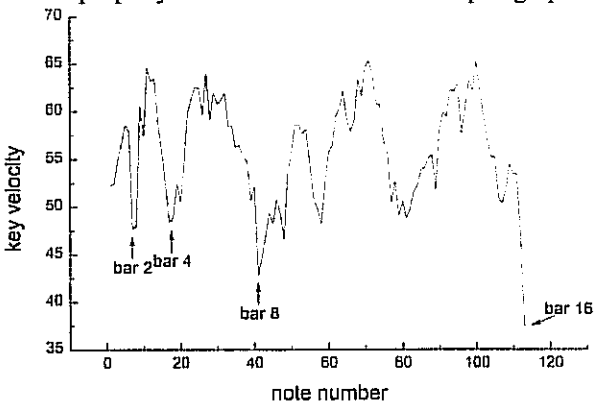


Fig 3: average profile of pianist A. It can be seen the relation between dynamics and the main elements of music structure

The transformations have to satisfy some conditions: 1) they have not to destroy the relation between structure and dynamics, 2) they have not to introduce too many parameters in order to not complicate the model unnecessarily. In order to represent the main characteristics of the performances, we used only two transformations: one shift and one expansion/compression of the values. The two above conditions are satisfied by a linear model, formally represented by the equation:

$$\tilde{y}_e(n) = k_e \cdot \bar{x} + m_e \cdot (x(n) - \bar{x}) \qquad \text{(Eq. 1)}$$

where $x(n)$ is the key-velocity of n-th note of the average profile, $\bar{x}$ is the mean of x, and are respectively the coefficients of shift and expansion/compression related to expressive intention e, $\tilde{y}_e(n)$ is the estimated key-velocity of the version related to expressive intention e. The parameters k_e and m_e, for each expressive intention, were estimated in order to

minimize the square error $\sum_n (y_e(n) - \tilde{y}(n))^2$, where $y_e(n)$ is the key-velocity of the n-th note, measured in the performance inspired by the expressive intention e.

3 Results

An average profile for each pianist was been calculated and the model parameters of his nine versions were estimated. Two values (m_e and k_e) are associated to each performance. So we can map the performances in a two-dimensional space, called Dynamics Parametric Space (DPS), which axes are defined by the two model parameters.

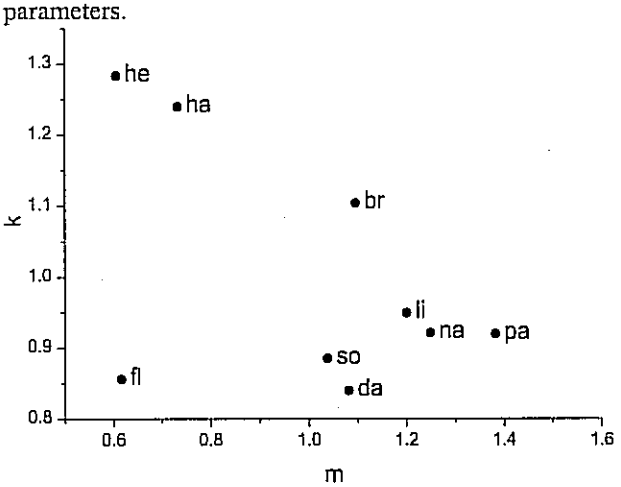


Fig 4: pianist A's performances, mapped in the Dynamics Parametric Space.

Figure 4 shows how the pianist A's versions are mapped in the DPS. By means of this space, we can easily obtain information about the musician's performance strategies: for instance the heavy version is characterized by higher key-velocity values (high k), in opposition to the dark version (low k); the passionate version is characterize by a large dynamics range (higher m), in opposite to the flat version (low m). So performances can be differentiated by means of two main characteristics: in this way, we answered to the first basic question.

Now we will discuss the second point, that is if there is any common strategy among the musicians. All the five pianists' performances were mapped in the DPS (figure 5).

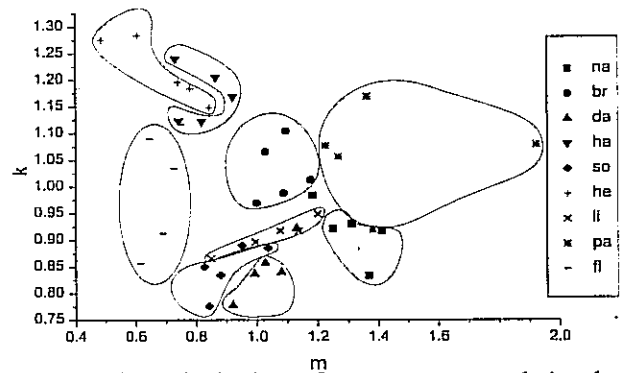


Fig 5: Five pianists' performances mapped in the Dynamics Parametric Space.

It can be seen that most performances, inspired by the same adjective, are grouped together in a region of the space. This fact signify that all the pianist have characterized, for instance, the soft version by means of a lower intensity (low k) and the bright version by means of intermediate values (k and m in the center of the DPS). These results suggest that there are some common strategies among the pianists, at least in relation to the proposed adjectives.

4 Discussion

In order to verify how the model works, the ratio between the variance accounted by the model and the total variance was calculated for each performance. Table 2 reports the mean, minimum, and maximum values, calculated among the performances of all the pianists. It can be seen that the mean variance accounted for by the model is about 67%, with a maximum value of above 90%. Only two performance have a variance below 50% (the hard and heavy version of pianist C).

	na	br	da	ha	so	he	li	pa	fl
Mean	78	70	67	61	65	53	73	73	60
Min	65	53	52	45	54	34	64	67	51
Max	90	91	86	83	78	73	86	79	77

Tab. 2: variance accounted by the model with two parameters (values are expressed as percentage of the total variance).

The values of table 2 are related to the two-parameters model (m and k). We developed a further analysis to test if both parameters are necessary. Table 3 reports the mean, minimum and maximum variance accounted for by a model, which have only the k parameter. It can be seen that, above all in the heavy and flat versions, the values are noticeably smaller. One performance has a negative value, which implies that in this case the model can't be apply. The second parameter allows a mean improvement of about 5%, with a maximum of 37%. This comments suggest that the two parameters are both necessary.

	na	br	da	ha	so	he	li	pa	fl
Mean	74	69	66	58	63	39	72	67	47
Min	60	53	52	40	52	-4	62	54	36
Max	88	90	85	82	78	64	85	75	62

Tab. 3: variance accounted for by the model with one parameter (values are expressed as perceptual of the total variance).

Another test of the model can be obtained by the production of computer performances, which key-velocity profiles are the estimated $\tilde{y}_e(n)$. So we can compare the original and computer generated performances and draw important observations about the model validity (analysis-by-synthesis method). The computer generated performances show that the model can well reproduce the global expressive characteristics

of the original performance. In particular the expressive intentions, which characterize the original performance, are clearly recognizable. Some local characteristics, however, are not very well reproduced by the model. This observation can be set in a hierarchical view of the musical discourse [8]: the followed approach can catch expressive characteristics as far as phrase level, but not lower. Model's goal is not a complete treatment of musical interpretation, but a study of the general performance strategies of musicians. The model, however, can be used as a good basis in order to study and apply other models, which can catch more local characteristics [9].

Now we will try to clarify the sense and the use of the DPS. Some outcome can arise by the definition of the input profile (we chose the average profile). By means of simple calculations, it can be showed that the average profile is the geometric center of gravity of the performances mapped in the DPS. The numeric values in the DPS, therefore, can not be considered in an absolute sense, but they are relative to their center of gravity, i.e. their reciprocal position. For instance, we can say that the mean key-velocity difference between the light versions ($k \cong 0.8$) and the heavy versions ($k \cong 1.2$) is about 40%.

It is interesting to find out if the DPS can be as well used in an inverse way. That is, we want to verify if the DPS can suggest how whatever input profile have to be changed in order to communicate a certain expressive intention (e.g. harder, softer, etc.). The verification was obtained by means of analysis-by-synthesis method, using both K545 Sonata and other piano scores. First, we need a human performance of the score, by which the input profile can be drawn. Then we chose a point of the space that correspond to a certain expressive intention and his coordinates (m and k) are used as parameters of the equation 1. We did it for all the adjectives and we obtained performances that reflect, in a relative sense, the chosen expressive intentions.

The DPS was obtained (see above) using a set of 45 performances, so represent a kind of sampling of the space. What do intermediate points of the space mean? We hypothesize that they can be used as an interpolation of the original samples: i.e. the points between heavy and light versions would have intermediate expressive characteristics. Analysis-by-synthesis method was applied choosing intermediate points of the space: the computer-generated performances have intermediate characteristics and show that all the points of DPS have an expressive meaning. These results imply that DPS can be used in order to render a kind of morphing between expressive characteristics. Generally, during the same performance, a trajectory that moves from a region to another one of the DPS can be drawn. The parameters, in that case, are functions of time and the performance will be characterized by changeable expressive features.

5 Conclusions

Starting from piano performance analysis, a linear model of dynamics variations depending on expressive intentions was developed. This model can be applied both to performance analysis and to the field of automatic performance. In particular, it is possible to draw trajectories in the DPS, which allow to control continuously the dynamics characteristics of the computer-generated performances. Analysis-by-synthesis approach showed that a linear model could properly render expressive characteristics and the defined parameters are suitable to describe different performances of the same score.

Acknowledgment

This research was supported by Telecom Italia, under a research contract "Cantieri Multimediali".

References

[1] Canazza, S., De Poli, G., Roda', A., & Vidolin, A. (1997). Analysis and synthesis of expressive intentions in musical performance. In *Proceedings of the International Computer Music Conference 1997* (pp. 113-120). Tessaloniki: International Computer Music Association.

[2] Palmer, C. (1996). Anatomy of a performance: sources of musical expression. *Music Perception*, 13(3), 433-453.

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

[4] Friberg, A. (1991). Generative Rules for music performance: a formal description of a rule system. *Computer Music Journal*, 15(2), 56-71.

[5] 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(2), 622-641.

[6] De Poli, G., Rodà, 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, Special Issue 1998.

[7] Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei. *Journal of Acoustic Society of America*, 92(5), 2546-2568.

[8] Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: The MIT Press.

[9] Windsor, W. L. & Clarke, E. F. (1997). Expressive timing and dynamics in real and artificial musical performances: using an algorithm as an analytical tool. *Music Perception*, 15(2), 127-152.

Friday 25th h. 14.00

MUSICAL SYSTEMS AND COMPUTER ASSISTED COMPOSITION

COMPOSING WITH ITERATED NONLINEAR FUNCTIONS IN INTERACTIVE ENVIRONMENTS

Agostino Di Scipio
LMS@aquila.infn.it

Survey of Functional Iteration Synthesis, FIS

The iterated mapping of nonlinear functions of a given interval can be utilized as a sound synthesis method particularly suitable in the digital synthesis of acoustic turbolences and other dynamical sound textures. In a previous paper this approach was defined *functional iteration synthesis* - FIS [ref.5]. FIS, however, is not one synthesis method, as is more a family of methods sharing the same basic mode of operation: the i-th sample in the digital sound signal is calculated as the n-th iterate of a given function. This can be expressed as follows:

$$x_{n,i} = f^n_i\,(x_{0,i}) = f_i\,(f_i(\ldots(f_i(x_{0,i})\ldots))$$

where i is the sample time index (integer), n is the number of iterations applied (integer) and f is a transfer function with its own set of m parameters.

When f is nonlinear the process can exhibit a complex dynamics, and the description of the system model thus implemented may profitably lean on the mathematics of chaos theory. The output values, $x_{n,i}$, is then a time series of peculiar behaviours, ranging from low-frequency turbulence to more regular patterns. However, I should stress that the relevant point here is more the *process* of functional iteration than the particular function: "*Yet, precisely because the same operation is reapplied* [...] *self-consistent patterns might emerge where the consistency is determined by the key notion of iteration and* not *by the particular function performing the iterates*" [ref.6].

I have focussed on a monoparametric map (m = 1) described as the cartesian product of two distinct spaces:

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

whose explicit iterated form would be

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

By adopting a sine function as f, and a scaling factor r, I call this the *sine map model* of FIS. The interval $[-\pi/2,\pi/2]$ accounts for the fact that, given the periodicity of the sine, any larger interval would only return time series already achievable within $[-\pi/2,\pi/2]$ (except for the time series of the 0-th iterates, $x_{0,i}$, as in fact the 1st iterate would fall in [-1,1], completely covered by $\sin(rx)$ for x_0 in $[-\pi/2,\pi/2]$ and $r \geq 1$). Moreover, the value range for r is [0,4] because any larger range would only provide results achievable with r within that interval.

For more insight the reader is kindly referred to other papers [ref.5, ref.8], which also include graphical examples and computer code. A simple Csound FIS instrument is in [ref.2]. The focus of the present paper is more on the musical application in interactive computer music works.

The parameter space of FIS

The output sounds of the FIS model in question, are dependent on the particular orbit in the phase space $[-\pi/2,\pi/2] \times [0,4]$, or one of its regions (see in fig.1 the phase space region [.2,.4] × [3, 3.4], where black = 1, white = -1). The orbit corresponds to a sound signal. In general, for any monoparametric map, the orbit is defined as the coupling of the two series, r_i and $x_{0,i}$ (the function parameter and the starting value for the iterated map). The resulting time series $x_{n,i}$ is dependent, in both its details and its overall evolution, from the particular orbit and its velocity across the phase space.

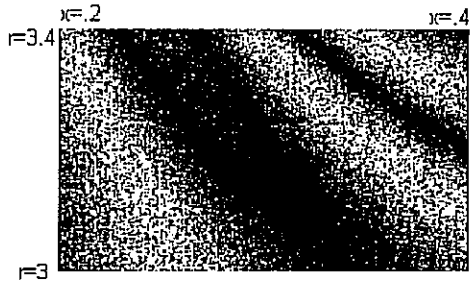


fig.1

If we sample an orbit in the phase space, we have, then, a 3-dimensional parameter space, with i (dicrete time), r and x_0 as the three coordinates.

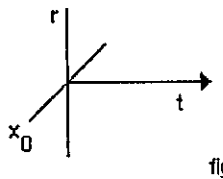


fig.2

Consider, then, that the number itself of iterations, n, heavily influences the result (see below). Hence, in effect we have a 4-dimensional space.

Broadly speaking, r determines the *kind* of behavior in the generated signal, ranging from very smooth curves (e.g. $r = 2$) to widely active and complex behaviors (e.g. $r = 4$), across many transitory phases (periodical cycles). On its part, x_0 determines the particular sample values. Slightly different values in x_0 would result in distinct signals which, after remaining identical, gradually shift apart one from the other (dependency on the initial condition, or "butterfly effect"). Finally, the number of iterations, n, also contribute to the spectrum width in the output sound, as with larger values the sound signal is more "active" and the spectrum gets richer in higher frequencies.

Once n is fixed, during the synthesis process we might want to (1) change r and keep x_0 constant; (2) change x_0 and keep r constant; and (3) change both r and x_0.

In the output sound, these different controls loosely correspond to (1) rapid changes of the signal waveform, i.e. highly dynamical spectra; (2) different signals sharing the very similar global properties, as captured by r (static spectra); and (3) a mixture of the former two, yielding articulated sound textures in a flux of continual change. If either r or x_0 are driven with a periodic driving signal (closed orbit in the phase space), the model is then forced to periodic patterns, and the output will sound patterned either in the sub-audio (rhythms) or audio frequency range (pitch).

The time-dependent relationship between r_i and $x_{0,i}$, and the way that relationship changes upon different values of n, are crucial as to compositionally interesting controls. One cannot modify one parameter without at the same time causing a change in the way the other parameteres affect the overall result. We can only *qualitatively* relate the parameters themselves with the perceptual properties in the synthesized sound. In a sense, this reflects the non-integrability of the underlying mathematical process and makes it perceivable to the ear. The rather uncommon sounds of the "raw" sine map model (i.e. before forcing it to periodic behavior) testify at this interesting situation.

With FIS I think a composer has to adopt an explorative, open attitude, and to learn about the potential musicality of the particular model in the course of empirical exploration of the parameter space. Therefore, s/he may turn to using this synthesis approach in interactive computer music environments, capable of real-time synthesis and allowing for empirical investigation of the audible effects emerging from the model.

Issues in the real-time implementation of FIS

Real-time implementation requires some key-conditions to be fulfilled. Both r and x_0 must be updated at sample rate, each sample being the n-th iterate of f applied to x_0. To do so, the calculation must include a loop of variable length, as the number of iteration, n, can be changed each time the synthesis process is called up (i.e. at event-rate). Hence, in principle the sample loop ends up being of variable duration, and must be forced to synchronize with all other operations going on in the computation.

Provided these conditions are fulfilled, one can implement the sine map model of FIS in a relatively straightforward way, just rewriting in DSP microcode the two nested loops required (the sample loop and its iteration sub-loop). I did this with the Kyma workstation, using the object icons (Sounds) in its graphical user interface (high-level implementation), or creating a short Motorola 56002 microcode to add to the microsound library (low-level implemenation).

However, a different approach exists, that I often use, and which is perhaps more general. This requires that we see FIS as a kind of generalized waveshaping synthesis, WS [ref.1]. In classical waveshaping a series of values (input signal) is mapped onto a given interval by a particular transfer (mapping) function, or waveshaper, the latter being usually composed of Chebichev polynomials. To iterate the operation, we then use the output as a new input value, i.e. as a new value to map rather than as the audio sample. This could be done again and again, for more iterations. Only the outcome of the last mapping will be taken as the audio sample. This iterated WS can soon become difficult to control, depending on the particular map, or waveshaper function.

It can be shown that a particular function exist that makes iterated WS become identical with the sine map model of FIS [ref.3, ref.5].

Indeed, classical WS could be seen as a special case (n = 1, f = some particular Chebichev polynomial summation) of the broader mathematical frame behind FIS (not the sine map model in particular). For me, the difference between the two approaches is a difference in conception. In principle, starting with WS one could achieve chaotic behaviors, while in actuality one is simply masking the periodicity inherent to the system. Starting with FIS one already is working in a condition of peculiar chaos (turbulence) and explores, and eventually finds, isles of order and periodicity, depending on how the parameter space is visited.

However, implementing FIS as iterated WS with *ad hoc* shaping functions is reasonably more handy and easier. In most of the compositional examples discussed below, I utilized this basic approach, thereby focussing on problems concerning compositional controls over the synthesis parameters.

FIS goes interactive. *SOUND & FURY* series

Iterated nonlinear functions have a relevant part in the author's compositional project *SOUND & FURY* (1995-1998). Shortly, this includes four different manifestations of the same low-level dynamics in different media: *SOUND & FURY (I)* and *(II)* are live computer music concert pieces, calling for one performer with MIDI faders. *SOUND & FURY (III)* is a kind of music-theatre (subtitle: "a theatre of noises, sounds, and voices"), based on fragments from Shakespeare's *The Tempest*, and featuring voices, percussions, interactive computer system and multiple slide projections. Finally *SOUND & FURY (IV)* is a permanent audio-video installation, with computer-controlled slide projection and interactive computer music.

All of these works involve (1) the sine map model of FIS (microlevel) and (2) more iterated functions as algorithms of automated musical structure generator (macrolevel). The latter calls for various instances of the former and passes to them the parameter values, according to the series of iteration of a nonlinear mapping, very similar to the microlevel (synthesis) process. Overall, this represents a kind of iterated function system distributed on two distinct hierarchical levels.

At every performance, the starting value, x_0, in the higher level iterations, i.e. in the automated music component, can be given a new value. That causes textural nuances and more evident global properties (time, density of events, etc.) to change at each different performance.

Significant differences exists among the various works as to two different layers of interactivity and their coupling: (1) man/machine, (2) machine/ambience.

Man/Machine interaction layer

The first interaction layer is captured in a straightforward "virtual control surface" (VCS) that I programmed for myself in Kyma. Each fader in this VCS is mapped onto a MIDI channel and made physically controllable with MIDI faders. Controllable parameters include the cycle time for x_0, which in the current implementation is the only cue to the orbital velocity in the phase space of the FIS. This eventually corresponds to pitch, provided velocity extends to audio range (just one possibility among others). In actuality, two faders are utilized to control the cycling of x_0, one for coarse tuning, the other for fine tuning. Cycle time ranges from several tens of seconds to 1 msec (1000 cps).

The sine map parameter r is not controlled interactively. The values in the interval [3,4] are provided at event time by the automated music component of the project, and are rescaled (enveloped) with simple envelope curves. The number of iterations n, too, is provided by the automated component at event time.

As the performer can hardly predict what value will follow for both r and n between overlapping sounds (that is: how active, broad-band or narrow-band, will be the sound), s/he has to continually adjust the x_0 cycle controls. Volume faders are also priveded in the VCS, so that s/he can compensate for different sound amplitudes as emerging from different values in r, n and x_0.

At this level, the performer acts more like a live "interpreter" of the chaotic, but structured flow of sonic information arising from the machine. That's a sort of "interactive interpretational design", reflecting a mode of algorithmic composition ("interpretational design" [ref.7]) where the composer has to make sense of the data output from his/her own computer program (Koenig's PR1 program is a classical example). However, in the present case interpretation is operated live, as from immediate perception, on the basis of parameter controls available in the VCS.

This picture appropriately reflects the performance situation in *SOUND & FURY (I)* and *(II)* (more should be added later for a complete picture). In *SOUND & FURY (III)*, however, man/machine interaction extends to the percussion part. In fact, in that piece the volume of the percussion sounds is employed to

modify, at each single moment in the performance, the cycling time values for x_0. This is done using simple amplitude followers. For some of the required percussions, the louder the sound and the faster the cycle (the higher the rate of timbre and pitch change in the synthesis). For other percussions, the inverse relation applies. Moreover, I made the amplitude followers to include a longer latency (or hysteresis) time. As a matter of fact, then, what is meant here with the percussion *volume* was in fact a more notion of *amount of energy per unity of time*: one short, loud stroke, followed by a silence of, say, 0.5", can have the same effect as a 0.5"-long soft roll.

Finally, the percussion volume was also utilized to change the timing of the automated musical structure generation: the louder the percussion sounds, the slower the scheduling of events from the automated composition component (controllable, dynamical scheduling is a relevant resource of the Kyma workstation). That makes sense in order to imbue into the musical evolution a kind of bias towards an average energy/time distribution, although it is precisely the deviations from that average that in the end articulate the musical flow in interesting ways.

Machine/ambience interaction layer

This second interaction layer plays a relatively important part in *SOUND & FURY (I)* and *(II)*, but it is central in the installation *SOUND & FURY (IV)*. The real-time generated FIS sounds being diffused in the performance place, or ambience, are captured by microphones and sent back to the computer, which utilizes their amplitude to modify the synthesis parameters and the timing of the algorithmic composition component (just as the percussion sounds in *SOUND & FURY (III)*). In short, both the sound fabricbeing generated and the hall's acoustic response to the former, become responsible for what is going to follow, in terms of timbre, pitch (cycling of x_0) and of density of events.

In such case we might see the musical work as an actual instance of *self-organizing system in permanent exchange with the environment* - a kind of *eco-system*. Changes in the ambience's response (caused, e.g., by visitors walking around the installation) induce adaptations in the behavior of the musical machine. Fig.3 shows an overall schema of interactions, illustrating the components of the *SOUND & FURY* project and the net of their interdependency (the role of the percussion part in *SOUND & FURY (III)* is reduced, for simplification, to that of the ambience in *SOUND & FURY (IV)*).

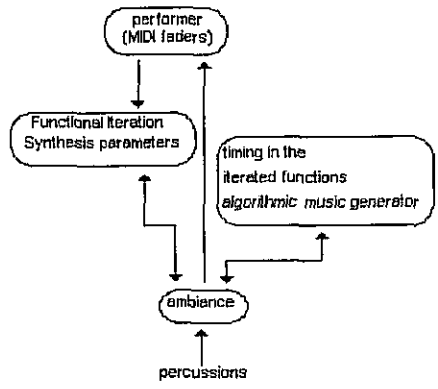


fig.3

Some conclusions

These observations seem to suggest that working with iterated nonlinear functions in interactive contexts leads to a sort of paradigm-shift from *interactive music composing* to *composing musical interactions* [ref.4]. This was at least my experience with the *SOUND & FURY* project, which extends my previous work with interactive methods of granular synthesis/processing (as in works like *Studio97-98*, in the CIM concerts, or *5 piccoli ritmi*). The musically fertile dynamics captured with interactive FIS illustrates a crucial but oblique relationship between the notion of *chaos* and that of *self-organization* as it emerges from simple, but iterated operations.

References

[1] D.Arfib "Digital synthesis of complex spectra by means of multiplication of non-linear distorted sine waves", *JAES*, 27(10), 1979

[2] R.Bianchini and A.Cipriani, *Il suono virtuale*, Contempo, Roma, 1998

[3] A. Di Scipio "Kyma Tips: Functional Iteration Synthesis", at site *http://www.SymbolicSound.com* (manuscript 1996).

[4] A.Di Scipio, "Interactive micro-time sonic design", *Journal of Electroacoustic Music*, 10, 1997

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

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

[7] O.Laske "Towards an Epistemology of Composition", *Interface-Journal of New Music Research*, 20(3-4), 1991

[8] I.Prignano "Sintesi di eventi sonori complessi mediante iterazioni funzionali", *Atti XI CIM*, AIMI/DAMS, 1995

Instrumented Footwear for Interactive Dance

Joseph Paradiso Eric Hu Kai-yuh Hsiao

MIT Media Laboratory
20 Ames St.
Cambridge, MA. 02139 USA
+1 617 253 8988
joep@media.mit.edu

ABSTRACT

We have instrumented a dance sneaker with an array of sensors that measure many parameters of foot, sole, and toe expression, continuously broadcasting them to a base-station and PC over a wireless link. This paper describes this system, reports its performance and outlines applications that we have developed for it in the field of interactive dance.

INTRODUCTION

Because of the comparatively high degree of manual dexterity in the general population, most human-computer and musical interfaces concentrate on precisely measuring gesture expressed by the hands and fingers, devoting little, if any, attention to the expressive capability of the feet. We have developed an interface that breaks this tradition by measuring many parameters that can be articulated at the foot of a trained dancer. Previous foot-sensing performance interfaces have generally been very simple, measuring only impacts at the heel and toe, usually with a piezoelectric pickup. Some of these were standalone shoes built for custom performances [1], while others, such as the Yamaha Miburi [2] are components of larger body-suit systems. Instrumented footwear has started to appear in virtual reality (VR) installations, for instance in the "Cyberboot" [3], an overshoe equipped with continuous pressure sensors to capture the dynamics of walking in CAVE (Cave Automatic Virtual Reality Environment) installations. Another foot interface for VR systems is the "Fantastic Phantom Slipper" [4], which uses an IR LED and PSD camera to track the position of feet atop a floor-mounted projection screen; here haptic feedback can also be generated by driving vibrators in the sole. Foot sensing has appeared in other application domains, such as in the research and development laboratories of major footwear manufacturers and biomechanical researchers, where detailed pressure profiles are obtained across the sole [5]. Although they collect fine-grained data, these interfaces are usually bulky, expensive and hardwired, inhibiting their application in areas such as dance. Some designers are beginning to deploy miniature inertial sensors in footwear for pedometry [6], producing jogging shoes that measure elapsed distance.

Our system, proposed initially in [7], was designed to go well beyond the simple tap detectors, pressure sensors, or pedometers sketched above and sense many of the different degrees of freedom that could be expressed at a dancer's foot. The system is entirely tetherless and self-contained; everything, including a small lithium battery that provides circa 3 hours of power, is mounted on the shoe, and data is offloaded from each foot over a wireless link.

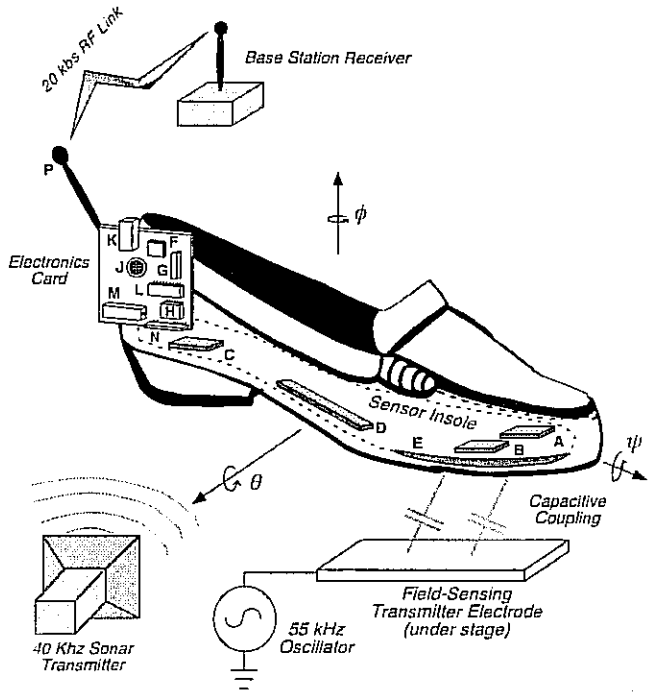


Figure 1: Diagram of the instrumented prototype shoe

THE PROTOTYPE SENSOR SUITE

Figure 1 shows the physical layout of our prototype system, including all sensors, while Figure 2 is an actual photograph of one of our early operational prototype shoes. We chose a Capezio "Dansneaker", which gave us sufficient room to mount our electronics without prohibitively impacting the dancer's movement. The only components inside the shoe are a set of tactile sensors and electrodes embedded in a standard "Dr. Scholl" foam insole sandwiched between the sole of the shoe and the sneaker's insole (the dancer is unable to feel any wires or objects through this). All other sensors and electronics are mounted on a small circuit card affixed to the outer side of the shoe, where it interferes minimally with the dancer's movements. In total, the prototype shoe measured 14 different parameters per foot, as described below.

The dynamic foot pressure is measured at two points under the toes (A,B) and one point at the heel (C) by piezoelectric film (PVDF) strips [8] mounted in the insole, as seen in Fig. 1. The bidirectional bend of the sole (Dansneakers are designed to allow large amounts of flex) is measured by a pair of back-to-back resistive bend sensors (D). When the shoe isn't quickly jerked, the two-axis tilt in pitch (θ) and roll (ψ) is measured by a 2-axis, 2G ADXL202 micromechanical accelerometer from Analog Devices (H). Large impacts and fast kicks

Figure 2: Photograph of the prototype sensor Dansneaker

are detected by a 3-axis, high-G ACH-04-08-05 piezoelectric accelerometer from AMP Sensors (F). In our prototype system, the twist angle (ϕ) was inferred when the foot is nearly level by an electromechanical compass made by the Dinsmore Co. (K), and the angular rate in ϕ determined by a small Murata Gyrostar vibrating-reed gyroscope (G). A 1-cm diameter, 40 kHz piezoceramic transducer from Polaroid (J) receives sonar pings from transmitters scattered about the stage, allowing the translational foot position to be determined by time-of-flight measurements. The bottom of the sensor insole is covered by an electric-field-sensing [9] receive electrode (E), which is tuned to detect low-level, *60 kHz radio signals transmit from conductive strips* placed beneath the stage. As the strength of this signal varies with the capacitive coupling into the shoe (hence distance from the transmitting electrode), it reflects the shoe's height atop the transmitting zones in the stage.

All signals are digitized by an onboard 16-MHz PIC16C711 microcomputer (L) that converts analog data into 8 bits and times the sonar and ADXL202 *signals. A zero-balanced data stream is generated* and broadcast to a base station up to 200 meters away by a small FM transmitter (N), a TXM series device from Abacom Technologies transmitting at 19.2 Kbits/sec. Our current embedded software updates all parameters at 25 Hz, which is conservative; when running at the maximum data throughput, we anticipate the performance system to be able to refresh all parameters at 50 Hz. Each shoe broadcasts continually at a different frequency (currently 418 and 433 MHz) through a stub antenna (P) and is powered by a small ($^1/_2$ AA) lithium 6-Volt camera battery (M). As the maximum current drain is 50 mA, these cells allow us to reach roughly 2-3 hours of useful battery life, sufficient for most performances (the shoes have an off switch to enable batteries to be conserved and send a binary battery status indicator across the RF uplink).

DANCE APPLICATIONS OF THE PROTOTYPE

In order to evaluate our prototype hardware, we have written a software application to map the data from a shoe onto a simple musical structure. This initial mapping was chosen to be extremely literal, enabling an improvisational dancer to quickly exploit such a novel interface. The music itself consisted of three voices: a drum voice, a bass voice, and a melody voice, articulated and controlled by a single shoe as outlined below.

Figure 3: Shoes in performance at MIT Wearables Conference

The drum voice ran steadily throughout the whole piece and gave a rough "techno" feel to the music. The volume of the bass drum and the bass voice were controlled by the θ tilt sensor, and the volume of the other drum instruments was controlled by the electric field sensor. The tempo was adjusted slightly by the bend sensor. The bass voice and melody voice were switched on and off in various combinations by the hi-G accelerometer. The bass voice itself produced a harmony effect, and the specific harmony was selected by rotating the shoe in ϕ. The bass voice was articulated by changing its octave based on input from the heel piezo sensor. The melody voice played harmonizing melody tones in upper registers; the range of the melody voice was controlled by the front piezos. Panning of both voices were controlled by the compass direction. Also attached to the hi-G accelerometer was

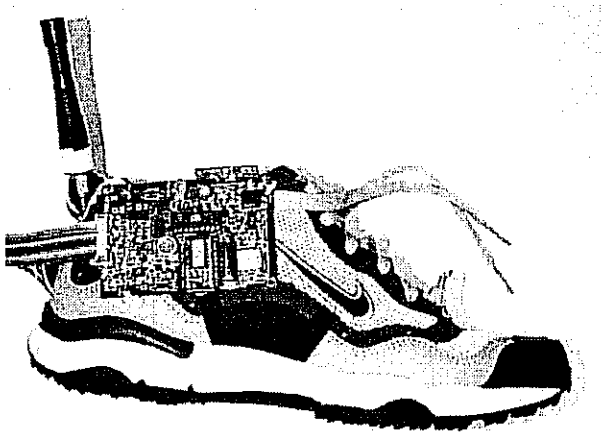


Figure 4: Photograph of the performance test design

an explosion sound, which triggered on heavy stomps and kicks. Finally, a panning wind sound was produced with quick ϕ rotations. A dancer practiced with these mappings, and performed [10] in the Wearables Fashion Show at the Media Lab in October 1997 (Fig. 3).

In another project [11], we have developed real-time classification algorithms that detect certain dance styles from the shoe data stream; e.g., discriminating between a waltz and a tango.

THE PERFORMANCE SHOE

After working with the prototype shoe for a few months, we modified and perfected our design to produce a shoe that will be robust enough for use in a professional dance performance. The sensor suite is identical, except that we have replaced the 2-axis electromechanical compass (which exhibited poor bandwidth and didn't hold up well enough to the shock and kinetics at the dancer's foot) with a solid-state, 3-axis magnetic field sensor from Honeywell (the HMC2003). Although these permalloy bridges can magnetize and drift over time, we have found them to be stable across several days after applying a single reset pulse to the device's common magnetizing strap, which we have made available at a connector on the edge of the card. We have also substituted force-sensitive resistors (FSR's), which provide continuous pressure readings, for the two PVDF strips at the toe, and added an additional FSR pressure sensor for measuring the pressure of the toes against the top of the shoe. We now measure an internal 3-Volt reference with the PIC A/D converter (which runs off the 5-Volt supply), giving a continuous indication of the battery state. These additions now bring the total number of transmitted analog channels to 17.

Fig. 4 shows the performance sensor card temporarily mounted onto a Nike "Air Terra Kimbia" sneaker, chosen by our collaborating choreographer for an upcoming performance. In our latest design, we have opted not to mount the battery on the circuit board, allowing greater application flexibility and freeing up layout space. We are now designing a robust, plastic, 3D-printed case to enclose the sensor card, electronics, and 9V Lithium battery, which should provide for over a day of continuous operation.

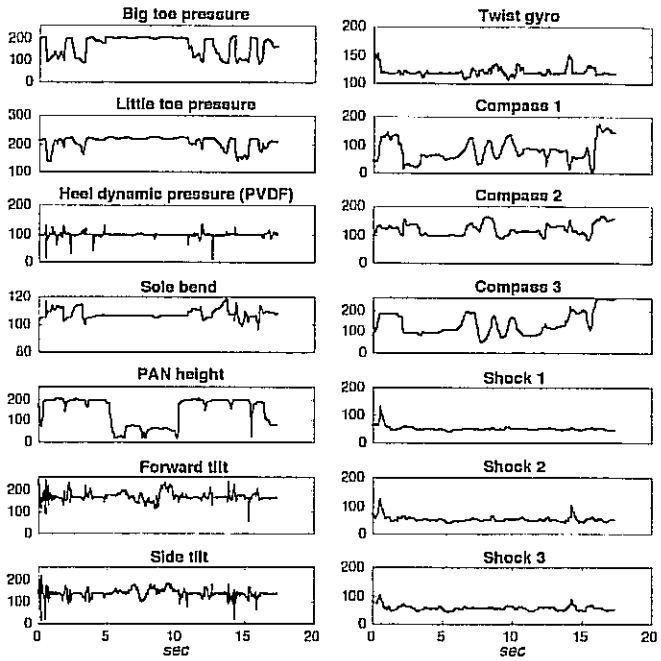


Figure 5: Actual data taken with the sensor shoe

Fig. 5 shows an 18-second snippet of data taken while the shoe of Fig. 4 was worn. The top-left pair of signals show the continuous pressure measured at the toes Here, one sees structure associated with footsteps at the beginning and end of the data; in the middle of the data stream the pressure data is much less active, as the shoe was elevated off the floor here and freely articulated. The dynamic heel pressure shows a similar profile (note that the PVDF strip measures pressure transients here, not continuous pressure), as does the sole bend (the Nike's sole is much stiffer than that of the Capezio used in the prototype). The footsteps and foot elevation are also seen very clearly in the electric field ("PAN" [12]) height data. The forward and side tilt accelerometer signals measure the pitch and roll of the shoe, as well as responding to transients from footsteps and kicks (giving the visible spikes). The twist gyro picks up fast yaw dynamics, and the magnetic sensor ("compass") signals respond well to attitude changes. The shock accelerometers are relatively quiet here, excepting at the beginning (when the user jumped) and at the end (another jump). As this run was taken in a small area, the sonar data is not shown. We have used the sonar successfully in larger volumes, however, measuring range out to 30 feet when using simple 40 kHz Murata PZT ultrasound drivers and attaining a resolution of circa 6 inches, limited mainly by the PIC's timing algorithm and the effects of background noise. We anticipate that this sonar will perform adequately for lower-bandwidth control (e.g., switching modes as the dancer moves into different regions of the stage).

FUTURE WORK

We are now concentrating on completing the performance shoe system, and are planning to finish a pair shortly with full base station support and faster sampling. Both shoes together will produce 32 parameters of useful gesture information, and the task of mapping these onto musically relevant and

choreographically interesting events is a challenging one, which we are now embarking upon in collaboration with colleagues who work in both dance and composition. The shoes produce a wealth of data on human gait, which likewise enable us to explore applications in other areas, such as sports and rehabilitative medicine.

The wireless sensor card that we have developed measures many general parameters, and is relevant to several applications beyond footwear. In this spirit, we have recently collaborated with our colleagues in the Synthetic Characters group at the MIT Media Lab to use our card in a multimodal interface recently shown at SIGGRAPH 98 [13]. Here, the card and an array of sensors were embedded in a stuffed toy "chicken"; by manipulating the toy, users were able to interact with similarly-appearing, semi-autonomous characters in an interactive animation.

ACKNOWLEDGMENTS

We are indebted to our MIT dancer, Yuying Chen, plus acknowledge Jack Memishian from Analog Devices and Kyung Park from AMP Sensors for providing technical support and donating the PVDF and accelerometers. We thank our many Media Lab colleagues who aided this project, especially Matthew Gray from the PIA group for the prototype software, Josh Strickon from Physics and Media for software and hardware support, and Andy Wilson and Zoe Teergarden from Synthetic Characters for their many helpful inputs. We likewise appreciate our developing artistic collaboration with David Borden and Byron Suber of Cornell University. We thank the Things That Think Consortium and our other sponsors at the MIT Media Laboratory for their support of this project.

REFERENCES

1. di Perna, A. Tapping into MIDI. *Keyboard Magazine*, July 1988, p. 27.
2. Paradiso, J. Electronic Music Interfaces. *IEEE Spectrum* 34(12), December 1997, pp. 18-30.
3. Choi, I., Ricci, C. Foot-mounted gesture detection and its application in virtual environments. *1997 IEEE International Conference on Systems, Man, and Cybernetics*. Computational Cybernetics and Simulation, Vol. 5, 12-15 Oct. 1997, pp. 4248-53.
4. Shirai, A., Sato, M., Kume, Y., and Kusahara, M. Foot Interface: Fantastic Phantom Slipper. Report ER14, Tokyo Institute of Polytechnics, 1998; see also *SIGGRAPH 98 Conference abstracts and applications*, ACM SIGGRAPH Press, p. 114..
5. Cavanagh, P. R., Hewitt, F. G. , Jr., Perry, J. E. In-shoe plantar pressure measurement: a review. *The Foot*. 2(4), 1992, pp. 185-194.
6. Hutchings, L.J., System and Method for Measuring Movement of Objects. US Patent No. 5724265, March 3, 1998.
7. Paradiso, J., Hu, E. Expressive Footwear for Computer-Augmented Dance Performance. *Proc. of the First International Symposium on Wearable Computers*, Cambridge, MA., IEEE Computer Society Press, Oct. 13-14, 1997, pp. 165-166.
8. Paradiso, J. The Interactive Balloon: Sensing, Actuation, and Behavior in a Common Object. *IBM Systems Journal*, Vol. 35, Nos. 3&4, 1996, pp. 473-487.
9. Paradiso, J., Gershenfeld, N. (1997). Musical Applications of Electric Field Sensing. *Computer Music Journal*, Vol. 21, No. 3, pp. 69-89.
10. *http://physics.www.media.mit.edu/danceshoe.html.*
11. *http://www.media.mit.edu/~mkgray/research/Fall97.html.*
12. Zimmerman, T.G. Personal Area Networks: Near-field Intrabody Communication. IBM Systems Journal, Vol. 35, Nos. 3&4, 1996, pp. 609-617.
13. Blumberg, B., *et. al.* Swamped! Using Plush Toys to Direct Autonomous Animated Characters. *SIGGRAPH 98 Conference Abstracts and Applications*, ACM SIGGRAPH Press, 1998, p. 109.

Motion Sensing and Realtime Sound Sampling Performance Systems and their Compositional Implications

Richard Povall
Director, Division of Contemporary Music
Oberlin Conservatory of Music (USA)
Visiting Researcher, Faculty of the Arts, University of Plymouth (UK)
richard.povall@oberlin.edu

Abstract:

This presentation profiles interactive compositions I have developed using two remarkable softwares developed at STEIM in Amsterdam. BigEye is a scriptable motion-sensing environment that essentially converts video input into MIDI output; LiSa is a realtime audio sampling system utilising the native audio processing capabilities of the Power Macintosh to capture, replay, and process multiple channels of digital audio in realtime.

Realtime software such as LiSa problematises numerous aesthetic and formal concerns that arise when working within realtime compositional environments. A notion of "performance composition" has been suggested, fundamentally different from traditional notions of composition. Relationships between composer and performer and audience are changed significantly when working with interactive systems. Is the notion of a through-composed interactive composition oxymoronic? How much control should a composer choose to relinquish? Can these interactive environments truly be described as compositions, or are they better described as instruments or even hyper-instruments? Is the composer merely an outdated modernist throwback, no longer able to function within new environments in which the performer is more empowered than ever before? Are the essential structural tools and methodologies also outdated? Is performance composition an entirely different process as well as a different product?

The software environments

1. BigEye

BigEye represents a significant step forward for those engaged in working with the moving body in performance. For some time, there have been a number of systems developed that attempt to capture the motion of the human body and use it to control musical or other environments. These have taken a variety of different approaches and utilise a variety of technologies. Systems that use hardware triggers — such as pressure pads, piezo sensors, and similar kinds of pressure-sensitive or light-sensitive devices — tend to suffer from too much specificity, and require the performer to be too precise. Systems that use infrared or ultrasonic beams to sense the moving body suffer from too much spurious information — infrared and ultrasonic transmissions tend to bounce around in space, and rarely give a truly accurate picture of movement. This may be fine for burglar alarms, but is not very useful when trying to sense accurately a moving body in space. Systems that use hardware triggers — typically stress or tension sensors — have become more successful since it had become possible to transmit the sensed data to the computer without wires, so that the performer is not encumbered by an umbilical cord to the computer. Composer/programmer Mark Coniglio of Troika Ranch in New York has developed MIDIDancer, arguably the most successful of these types of systems. The only remaining type of system — video-based systems — seem to show the most promise as performance systems because they leave the performer entirely unattached to the system, and do not require them to hit specific places in a specific way (as hardware sensors do) to be fully effective.

Many of the early video-sensing systems were based around the hardware sensor paradigm in that they required the performer to pass through a virtual trigger that existed in 3D space. Unfortunately, because of the nature of the two-dimensional video camera, the position of the trigger in three-dimensional space is highly skewed, and, if visible, would be seen as a cone shape. This is very difficult for a performer to grasp, particularly if there are a number of them in the space.

BigEye represents a shift away from this paradigm in that it is capable of viewing the entire virtual space of a 120 X 160 pixel video window. While it is still possible to draw conical virtual triggers within the BigEye window, it is also possible to use the entire space and attempt to interpret motion in a much more fluid way. Of course, it can be argued that each pixel will in fact have a conical shape as viewed by the camera, but this is less problematic when experienced on such a small scale and in such a tight consecutive matrix.

BigEye is capable of looking at motion in two distinct ways: colour and difference. The colour filter checks the incoming video stream against a colour look-up table that is defined by the user visually. It uses RGB colour space, and it is possible to add as many colours to the look-up table as are required. The difference filter compares every pixel to itself in the preceding frame, and is therefore capable of a quite accurate tracking of a moving object within the video space. The framerate of the system depends entirely on the hardware used. On a Macintosh 8500/180, it is possible to attain a framerate of between 15 and 20 frames per second. On a G3/300, full framerates of 30fps and faster are easily attainable. Actual framerate also depends on how much load is being placed on the system: how many objects/fields are defined within video space; how complex are the scripts within the file; how many objects are being tracked.

BigEye's scripting environment is powerful and relatively simple to use. The scripting enables the composer to interpret specific kinds of motion in specific ways, and to add, for example, speed filters that are so important and so necessary when working with choreographed motion. The output is MIDI, so it is easy to reach the outside world and control a wide variety of devices such as sound synthesis systems, lighting boards, laser disks, and digital video playback.

2. LiSa

LiSa is another groundbreaking program, also made at STEIM in the Netherlands. STEIM has been in existence for thirty years, and is best known for its development of extraordinary interactive technologies for performance such as the Spider, and the Data Glove. In recent years, STEIM's concentration has been not only physical instrument-making but also software instrument-making. Their stable of software products also includes Image/ine, a realtime video image processor with a mind-boggling array of MIDI control possibilities.

LiSa (Life Sampling) takes advantage of the native digital signal processing within the Power Macintosh (it will not run on 68000 systems, nor on non-PowerPC machines). It is capable of recording up to four voices, and playing back sixteen voices simultaneously. Once again, this is hardware dependent, but any of the faster machines can easily accommodate LiSa's maxima. I run LiSa on a 180MHz PowerBook 3400, and it works extremely well. The program also features a number of built-in DSP processes, and all of its parameters can be controlled via MIDI, once again an almost dizzying array of control possibilities.

A sample buffer holds the current collection of sampled sounds, and it is possible to read/write from the buffer at will, and under command of MIDI. The size of the buffer is limited only to physical RAM present in the host, so it is possible to run LiSa on a machine with 100MB of RAM or more and have an exceedingly powerful and large sample collection active at any one time. LiSa can grab pre-existing files into it's buffer, and can also record live into the buffer.

The system gives the composer/performer tremendous realtime power, and provides an extraordinarily broad palette. Difficult to tame, LiSa can be the performer's dream

- and his/her worst nightmare. There are numerous interactive programming environments in common usage, but LiSa's unique character comes from the fact that it is a sampling environment rather than a MIDI environment. It performs well for certain kinds of sound worlds and stylistic approaches, and poorly for others.

It is the flexibility and degree of external control that makes the STEIM environments so powerful as artistic tools. Because STEIM's history is centred on experimental music, and because they have traditionally made musical instruments, their focus in designing software tools is unique. There is little software available that has so much flexibility and that work so well as expressive tools for the composer and performer.

Problematising Interactivity, Performance, and Composition.

Realtime software such as LiSa problematises numerous aesthetic and formal concerns that arise when working within realtime compositional environments. A notion of "performance composition" has been suggested, fundamentally different from traditional notions of composition. Relationships between composer and performer and audience are changed significantly when working with interactive systems. Is the notion of a through-composed interactive composition oxymoronic? How much control should a composer choose to relinquish? Can these interactive environments truly be described as compositions, or are they better described as instruments or even hyper-instruments? Is the composer merely an outdated modernist throwback, no longer able to function within new environments in which the performer is more empowered than ever before? Are the essential structural tools and methodologies also outdated? Is performance composition an entirely different process as well as a different product?

These are huge, troubling questions, but they are the core questions, and no one who makes interactive environments — regardless of the software or hardware they use — can avoid asking them. Interactive composition must represent a fundamental paradigm shift because the composer is called upon to cede a large amount of control. In building an interactive performance environment, which I think of as my composition, I am really building an instrument upon which the performer(s) will play. Unlike most instrument builders, however, I must make the instrument with the performers in situ. Indeed, I really must work with the performers in order to build the instrument or have any concept of how it will sound. As I am primarily working with dance/theatre, I must then work with the performers, the choreographer, a dramaturg, and with all the other elements we are choosing to include in a given work before I can really build the environment. Particularly because I work collaboratively in developing content and material for inclusion in the environments, I am constantly faced with the questions I posited earlier.

In fact, many of these questions I consider answered years ago. I no longer expect to be able to control the "form" of a piece, nor to ever tightly control the final outcome of a piece. While I know how the piece will sound in general, I have little idea how it will sound specifically; while I know in a global sense what kind of shape the piece will have, I have relatively little idea what the innermost form will be, or what the specifics of the improvised form will be. These are parameters most composers hold dear, but in building a genuinely improvisational environment, it is essential to loose control over these most fundamental elements.

So what are my compositional decisions? Ultimately, all these discussions must come back to the work itself. Every technology is irrelevant if there is no discussion about the ways in which it can be used, and about the context in which it is used. I am drawn to using interactive technologies within performance because I want to root my sonic environments firmly in the body. If I were interested in purely musical environments, then a powerful controller would be all I need. I worked for some time with the Mattel

PowerGlove, but ultimately found its visual appearance too disturbing, and my lack of physicality too limiting. I have always been interested in dance, and in dance/theatre, so working with motion-sensing systems seemed to be an obvious choice. Here too, there were many problems with limitations of technology, but BigEye and LiSa have allowed me to interpret motion in sensitive and sensitised ways that go far beyond a literal interpretation of location in space. Using BigEye's scripting language, I am able to look deeply at the movement — how fast, how accelerated, how large, how many. These kinds of qualities are *intrinsic* to the movement, not external to it. I can really work with the language of the body, the language of movement and relate it to the language of the sonic or visual environment. The movement of the body *is* the music, or textual soundworld, it's not merely triggering specific samples or specific MIDI note sequences. Moreover, how I use this physical language is my composition. I craft with sound worlds created by gesture, and the larger shapes are determined in collaboration with the choreographer and the performers. I remember having a discussion with a composer who was just stepping into the world of motion-sensing all about a score-following object he was designing so that the dancers could "play" his score. This is compositional anathema to me. It makes little sense to ask a performer to work within an interactive environment if all s/he is doing is recreating a pre-existing score. Similarly, the kind of simple triggering of virtual triggers through motion is uninteresting to me compositionally.

So I repeat my questions. Is the notion of the composer some kind of modernist throwback when looked at within the context of interactive music, or interactive space? Should we be rethinking what we mean by music composition, and looking at entirely new paradigms of creation and formal thinking? Should we look again at what we mean by interactivity and what we mean by collaboration? Should we even challenge the fundamental notion of form as a musical language? As we enter into a world in which we interact increasingly via networks and computer systems, are we engendering a new notion of "composer".

Galileo, a Graphic ALgorithmic music language

Leonello Tarabella, Massimo Magrini

Computer Music Lab of CNUCE/C.N.R.
via S.Maria 36, 56126 Pisa - Italy
Tel. +39-50-593276 Fax +39-50-904052
email:L.Tarabella@cnuce.cnr.it
http://spcons.cnuce.cnr.it/music/cmd.html

Abstract

After the experience of the Real-Time Concurrent PascalMusic (RTCPM) language which gives the possibility of defining a composition as a Pascal program and of interacting with it at run-time, we recently designed a new language named *Galileo* which includes Graphical facilities and inherits from RTCPM the ALgorithmic approach to composition and to interactive performance. Depending on the number of active objects and on the length of code describing the functionality of each object, a composition can range from a pure algorithmic to a Max-like patch approach. It also gives the possibility to define Csound-like instruments to be executed in real time on the computer in use.

1 Introduction

In order to put at work the power of the algorithmic approach to composition and live performance, the RealTime Concurrent PascalMusic (RTCPM) [1][2], has been developed and used in the last five years; it consists of the standard Pascal language enriched with a library which implements the concurrence facility and deals with MIDI events.
It is based on the idea that a running program/composition issues MIDI messages which feed synthesizers, and that the global musical result depends on the predefined algorithms computation and on MIDI messages coming from the external. Concurrence is the vital mechanism that makes PascalMusic work.
A composition is organized as an usual Pascal program where the main program calls in sequence a number of *movements* (that is of procedures) and a movement is a collection of procedures concurrently activated, each describing a part and/or a subpart of the movement: notes, dynamics, effects, spatial position, etc.
Moving on, we recently developed a new language independent from whatsoever host compiler and from the machine architecture; this language, still providing the facilities and power of the algorithmic approach, greatly benefits of visual programming techniques. It allows to define programmable objects able to communicate to each other and provides tools for monitoring and setting parametric values and for grouping objects.
We called this language *Galileo* in Galileo Galilei's honor since he was born and lived in Pisa and the Domus Galilaeana [3] is located on the same side-walk 50m from the CNUCE Institute; incidentally, as an acronym, the first 3 letters of the name stands for Graphics & ALgorithmic. Once again, the textual part of the language is an *essential* Pascal-like language [4].

2 The graphic environment

As it happens in PascalMusic, in Galileo a composition consists of movements; however, in the new environment, a movement is defined inside a graphic window in terms of algorithmic elements connected via links for exchanging messages.

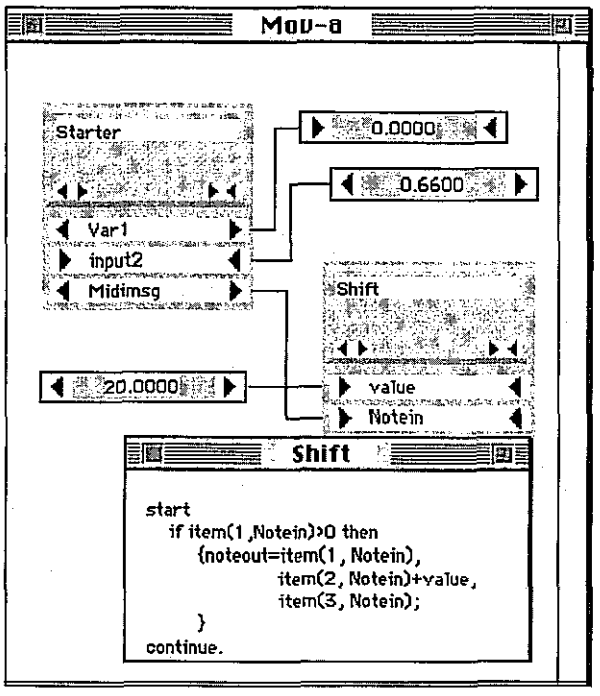
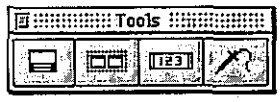


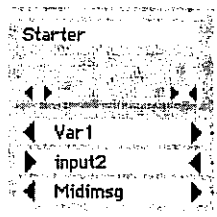
Figure 1: Example of a composition in GALileo

All the elements in a movement window are active simultaneously; elements may be connected to other elements with a link line meaning that a message can be transmitted between them: messages can be numeric

or Boolean values as well as strings of data. There exist three types of elements: process, constant and probe. What a process-element does, is defined by the user/composer by writing a C/Pascal-like procedure in the related text window which opens with a double click on the element. Text windows can be opened and closed individually or with a CLOSE ALL command.
A toolbar (mobile on the computer desktop) with four buttons allows to create elements to place in the movement window.



The first button (from left to right) creates a process element which can also have the meaning of Voice process that is a generator of note events such as Note-On messages in Midi or C-sound i-statements; generally speaking an element is a process concurrently active together with the others on the movement window.
A process element appears like in the following figure:



the upper part contains the name of the element given by the user; two buttons with inward and outward arrows allow to generate input and output variable outlets for communicating with other elements. Clicking on the buttons an input or output entry is physically generated in vertical sequence: the arrows appear on both sides for a better graphic appearance; in fact, for linking two variables of two elements it is sufficient to select the related arrows and a line is generated by an auto-routing algorithm which also takes into account the left or right arrow for the best path.
The second button collects all the selected elements of the movement in a single super-element.
The third button creates a constant-element to be connected with input variables of other process-elements. Data can be directly typed in the constant-element or, after double clicking on that element which pops up a dialog box, set by means of sliders ranging between definable limits with linear and/or logarithmic scales; the arrows of these elements are, obviously, outward.
The fourth button generates a probe-element useful for debugging the composition and at run-time. The arrows of this element are, obviously, inward.
The auto-routing rules seen for process-elements are also valid for constant and probe elements.

3 The language

As we said, the functionality of a process-elements is defined by writing a procedure in the related text window: a double click on that element opens a text window with the same name declared for the element and C/Pascal like code may be written in.
The language inherits from Pascal the control structures **for, if-then-else, while, repeat-until, case of,** and the assignment statement and from the C language freedom of variables definition and marking of blocks with { }.
There exist only two data types: numeric/real and booleans; it is also possible to define arrays and lists of both numbers and booleans.
Assigning a value to a variable defined as input-entry (as previously described) means to communicate that value to the linked element.
There exists a library of predefined mathematical functions (log, exp, sqrt, sin, cos, rnd, etc..); however the user/composer can define his/her own library in accordance to particular needs required by a specific composition.
Two very important functions are included in the standard library: MidiIn e MidiOut which make it possible to communicate with midi controllers and midi synthesizers. MidiIn gives a list containing the length of the message, the time-stamp and the midi message: Length=0 means no message in the midiIn queue.
Processes read messages from the midiIn queue without destroying them so that each process can read them independently from the others.
The MidiOut function enqueues a midi messages which is merged with those coming from other processes.
Processes are controlled by a scheduler which works following time-sharing techniques: midiOut messages, have a duration so that when a process issues a message that process is stopped and resumed after the proper time is over.
Even if there exists only one type of process, it is worthwhile to consider the process which issues Note-On messages as a Voice; other processes which control volumes, timbres, spatialization, etc. may be considered as auxiliary processes.

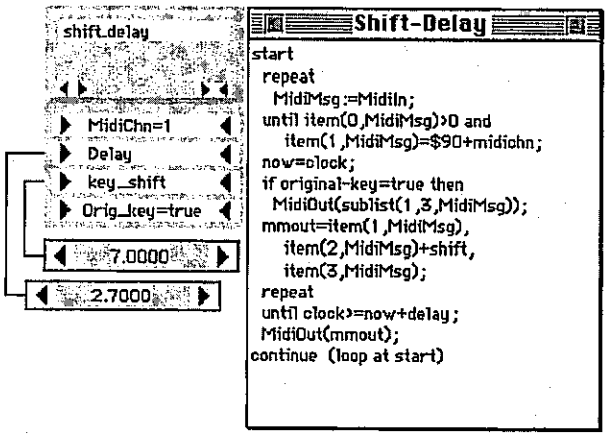
4 An example

Generally speaking the user can range between two limits consisting of, the first one in score or pure algorithmic composition (markov chains, fractals, chaotic functions...) and the second one in Max-like patch: in the first case a movement is characterized by few long-code process-elements and few links; in the second case a movement is characterized by many short-code process-elements and many links.
The following example shows a very familiar midi message processing: keynum-shift and delay of NoteOn messages. The problem is here tackled at two different

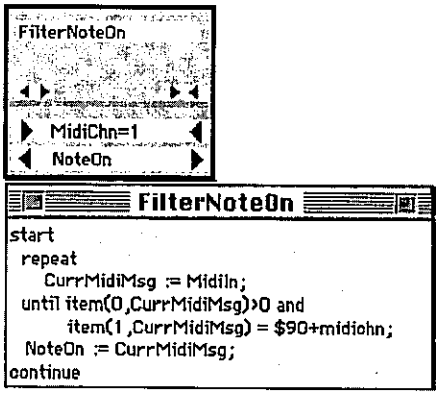
levels: single "long"-code process and multiple short-code processes.

First solution. The element has 4 input entries which specify: - the midi channel number of messages to be processed - the key-num shift value - the amount of time the message has to be delayed - whether or not to immediately play the original note.
While reading the code which implements the shift-delay functionality, take into account what follows:

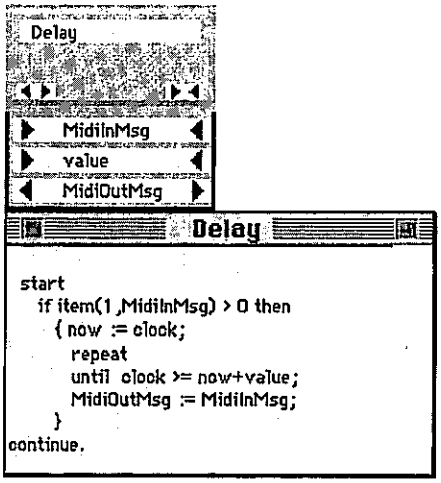
- only Note-On messages are considered;
- a MidiIn function generates a list-variable;
- single elements of a list are selected with the *item(i,namelist)* primitive function;
- the 0-item of a list specifies the length of the list
- *clock* is the global variable which counts time elapsed from the beginning;
- *MidiOut* function wants a list; a list is created by chaining numeric value with the , (comma);
- the *end* statement terminates the process; the *continue* statement sends the control to the start which causes a loop.



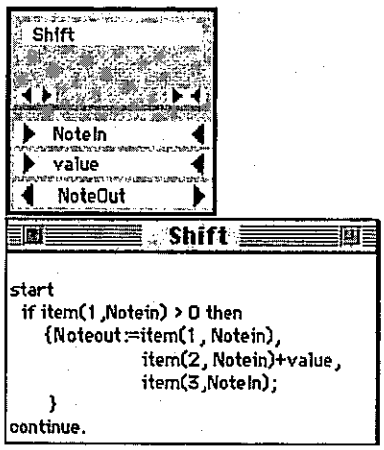
Second solution. Four short-code elements are created and the linked.



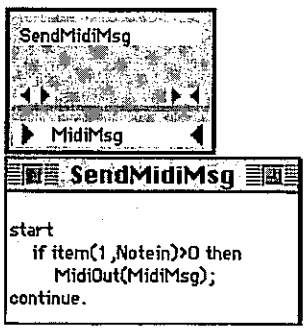
The first element filters only NoteOn midi messages with channel = 1;



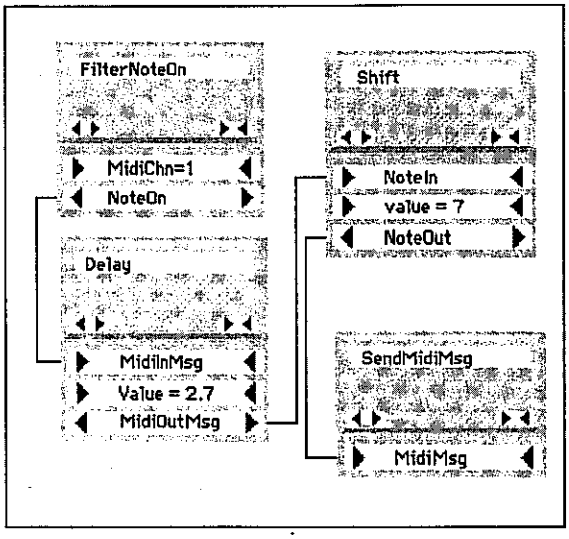
The second element receives a midi message and sends it out after a delay time found in accordance with the value found in the input entry.



This element receives a NoteOn message and sends it out with a key-num shifted by a value found in the input entry.



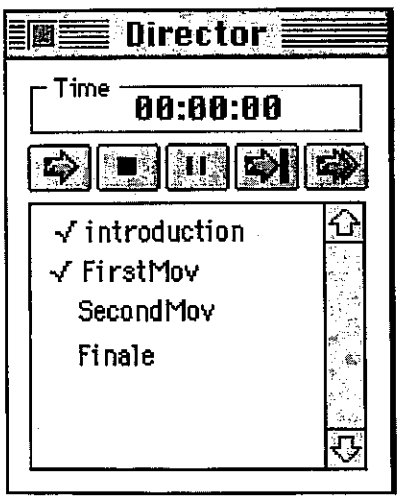
Finally, the fourth element transmits on the midi line the received message.
The following schema uses the four elements properly linked and without the related text windows.



We chose this example because it is simple and well known; besides, it gives the opportunity to introduce a very important and typical problem found in concurrent programming environment: the synchronization of processes. In Galileo this problem is solved by giving the user full control of communication between the active elements. The link lines are of two colors: red and black. Red lines represent a communication with acknowledgment; black lines represent a communication without acknowledgment. This means that an assignment to a red-linked output variable (which corresponds to send a message) stops the sending process until the receiver has read the value from the corresponding red-linked input variable; with black lines the involved element are completely independent and communication occurs asyncronously. It's up to the user to choose the type (color) of link depending on the specific situation.

5 Control of execution

The execution of the whole piece consisting of many movements is controlled using a tool named Director:



The Director allows to define the list of movements which are part of the composition and, for making it easier the debugging phase, to enable or disable them. In the upper part of the Director five buttons allow to: (from left to right) 1-start/restart the whole composition; 2-stop the execution; 3-pause; 4-step/debug; 5-fast_forward.

6 Improvements

The current version of Galileo is able to solve a large class of algorithmic and interactive compositions. At the moment many auxiliary graphic tools such as cursors and analogic displays for better entering and monitoring data, are under development.
As it is so far, Galileo is an event-processor language; however we planned to provide Galileo with facilities for audio signals processing too. On the tool bar a fifth button will enable the user to open a new kind of element able to accept C-Sound like code for describing instruments: in this case the input variables will correspond to the p-parameters of C-Sound to be directly linked to Process/Voice elements. On the basis of experiences [5] acquired in the previous years at the Lab in Pisa, a further button will give the user the possibility to define synthesis and sound processing schemes with graphic techniques.

7 Conclusions

The Galileo language has been developed for the Power Macintosh platform using the Metrowerks CodeWarrior Compiler. As a Midi event processor, Galileo can be used with all Midi controllers and expanders. As soon as the sound processor C-sound like facility will be ready, Galileo will be free-ware distributed for verifying the renspose of possible users. The very basic number of facilities provided by Galileo, will give the user the possibility to develop any kind of elements and compositions from simple to complex ones, edited within the same graphic and textual environment.

8 References

[1] Tarabella, L., 1992, "Real Time Concurrent Pascal Music", in Proceedings of the ICMC 92, San Jose.
[2] Tarabella, L., 1993, "Real Time Concurrent Pascal Music" in *INTERFACE*, vol.22 n.3 - Swets & Zeitlinger B.V..
[3] The Domus Galileana:http://galileo.difi.unipi.it/
[4] Tarabella, L., Sudio Report of the Computer Music Lab of CNUCE, in ICMC97, pp.86-88
[5] Tarabella L., Bertini G, A signal processing System and a Graphic editor for synthesis algorithms, in ICMC98 Procs, pp. 312-315

Intelligent Jazz Accompanist: A Real-Time System for Recognizing, Following, and Accompanying Musical Improvisations

Petri Toiviainen
Department of Musicology
University of Jyväskylä
P. O. Box 35
Jyväskylä, Finland
email: Petri.Toiviainen@jyu.fi
http://www.jyu.fi/~ptoiviai

Introduction

In jazz, blues, and rock, improvisations are based mainly on the harmonic structure of the tune. Listeners who are acquainted with this style can often identify the underlying harmonic structure and thus the tune even on the basis of an improvisation. This is also part of the basic skills of every improvising musician. The Intelligent Jazz Accompanist (IJA) is a working real-time system that models this skill. More specifically, it models (1) how we parse the rhythmic structure music and (2) how we recognize a piece of music in spite of the variation and improvisation present in the performance. Operating in a MIDI environment, the IJA listens to an improvisation played by a human performer and recognizes the tune on which the improvisation is played as well as the current location in the structure of the tune. Then it joins in with a synthesized rhythm section playing the accompaniment of the tune at the tempo of the performer.

The ability to infer beat and metre from music is one of the basic activities of musical cognition. It is a fast process: after having heard only a short fraction of music we are able to develop a sense of beat and metre and tap our foot along with it. Even if the music rhythmically complex, containing a range of different time values and probably syncopation, we are capable of inferring the different periodicities of it and synchronising to them. A rhythmical sequence usually evokes a number of different pulse sensations, each of which has a different perceptual salience. Furthermore, for a given piece of music, the most salient pulse sensation can vary between listeners. According to experimental literature [1], the range of most salient pulse sensations lies between 67-150 events per minute, corresponding to periods of 400-900 ms, the greatest salience being in the vicinity 600 ms.

In addition to the basic pulse level, we simultaneously perceive periodicities on higher hierarchical levels [2]; the longer periodicities are integral multiples of the periodicity of the basic pulse. This results to a percept of periodically alternating strong and weak beats, corresponding to the generally accepted definition of metre [3].

The theory of Lerdahl and Jackendoff [3] distinguishes between three kinds of accent contributing to the rhythmic organisation of music: phenomenal, structural, and metrical. They rely, respectively, on sensory, structural, and schematic sources of evidence [2]. Phenomenal and structural accents serve as perceptual input to metrical accents [3].

Attempts to model the perception of pulse and metre have relied on a diversity computational formalisms. These include, among others, rule-based systems [4], statistical approaches [2,5], optimisation approaches [1], and connectionist models [6]. A common feature of the above-mentioned models is that they deal with idealised rhythms, i.e., rhythms comprised of precise durations such as those found in a musical score. When attempting to parse the rhythmic structure of real musical performances, we are faced with additional challenges. These are caused by temporal deviations present in every musical performance. These can be either intentional, for instance, tempo changes intended to attract the attention of a listener, or unintentional, i.e., inaccuracies caused by the technical inabilities of the performer. Models that attempt to find rhythmic parsing directly from performances have been proposed in [7-12]. Large and Kolen [11] and McAuley [12] introduce nonlinear oscillators that entrain to an incoming signal by adjusting their phase and period.

Besides on the rhythmic level, the listeners also parse the structure of the music on a higher level. In an improvisation based on a given harmonic structure, the pitches of the tones carry information that can be used in both parsing the higher level structure and recognizing the tune. One reason for this is that for each type of chord there are tones which the improvisers tend to emphasize or use more often [13]. Furthermore, in different metrical positions the improvisers tend to emphasize different tones. For instance, on strong beats tones of the underlying chord are used more frequently than on weak beats. Consequently, the harmonic structure of an improvisation can be inferred using statistical learning and reasoning. This kind of reasoning has been used by Dannenberg and Mont-Reynaud [14] in a system for following improvisations. Their system listened to a 12-bar blues improvisation on a monophonic instrument, inferred the current location in the harmonic structure and joined in with the accompaniment. The reasoning scheme used by Dannenberg and Mont-Reynaud is somewhat similar to the one used in the IJA. The IJA, however, accepts polyphonic input and aims at identifying the tune among a number of possible alternatives.

Structure of the IJA

Figure 1 presents a general overview of the IJA. The input consists of an improvisation performed on a MIDI instrument. In the first stage, the basic rhythmic structure of the input is analysed. This includes inferring the basic pulse, or the beat, and tracking it. It is assumed that the note onset times are sufficient for finding the basic pulse, so pitch information is ignored in this stage. This stage is realized by means of competing adaptive oscillators and resonance dynamics. The output of the beat tracker at any time is the number of beats and fractions thereof elapsed since the beginning of the performance. In this stage the IJA thus maps absolute time, expressed in seconds, to relative time, expressed in numbers of beats. This step is justified, because there is evidence in the experimental literature that listeners represent rhythms with respect to an internal clock that is synchronized to the perceived rhythm rather than in terms of absolute intervals of time [15-16].

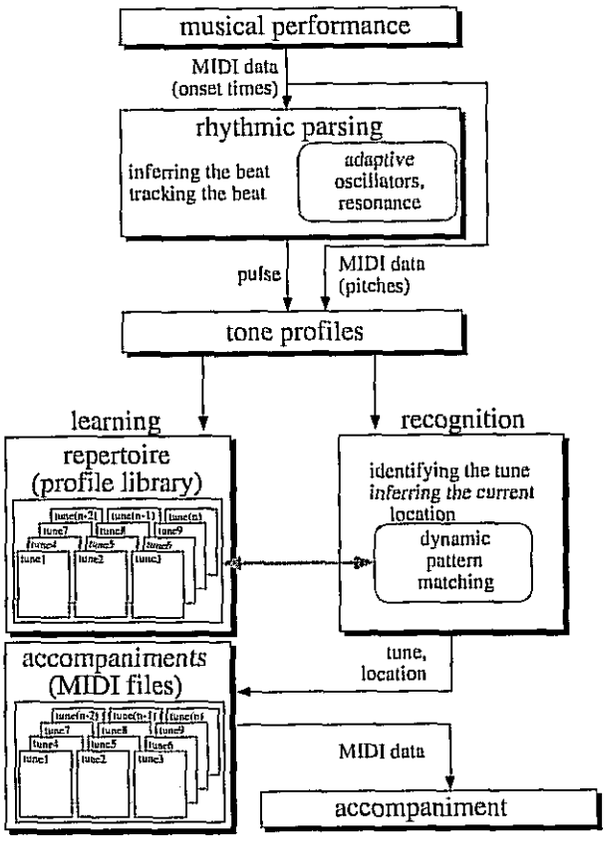


Fig. 1. A general overview of the Intelligent Jazz Accompanist.

In order to recognize music, the IJA has first to learn examples of it. For that, a repertory of tunes is played to it. The system uses the output of the rhythmic parsing unit and the pitch information of the MIDI data for constructing a set of tone profiles for each tune. Each profile is a vector representing the statistical distribution of pitches used at a given point of the structure. When attempting to recognize an improvisation, the system constructs tone profiles from the improvisation and compares them with the profiles of the library. The comparison is based on dynamic pattern matching. When the system finds it is at the top of the form of the tune it supposes the performer is playing, it starts playing an accompaniment of the respective tune.

Rhythmic parsing

The rhythmic parsing stage is based on adaptive oscillators similar to the one introduced by Large and Kolen [11]. The present oscillator, however, employs a more sophisticated adaptation mechanism than the original. It is based on the notion that onsets of perceptually salient notes, i.e., notes having a long duration, should give rise to stronger adaptation than those of perceptually less salient, or short, notes. As a consequence of the new adaptation scheme, the oscillator can follow fairly well even a rhythmically complex input, such as a melody with trills, grace notes, and syncopation. The adaptive oscillators adjust their phase velocities so as to become and remain phase- and frequency-locked to periodic components in the stream of note onsets. The main principles of adaptation are as follows: (1) the oscillator adapts to those note onsets only which occur within the peak of its output; (2) if it finds it is late, it speeds up, and vice versa; (3) there are two kinds of adaptation, namely, fast and slow adaptation (4) adaptation takes place gradually and a posteriori; as a consequence of this, short notes do not give rise to any significant adaptation. An example of adaptation to a stream of note onsets is presented in Figure 2.

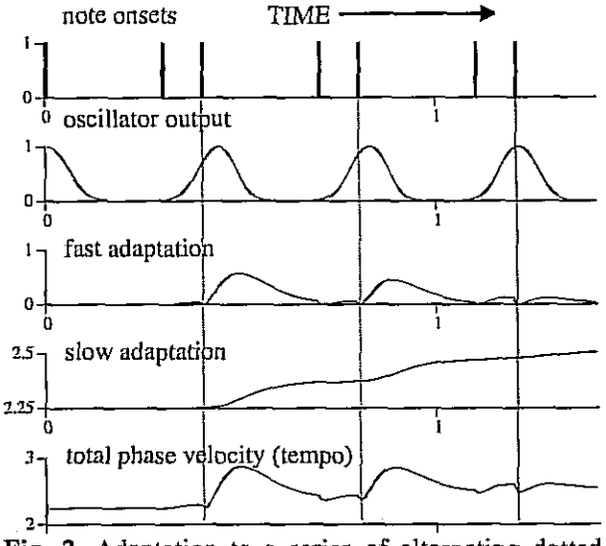


Fig. 2. Adaptation to a series of alternating dotted eighth notes and sixteenth notes. The total phase velocity is the sum of two parts, that is, slow and fast adaptation components.

From a psychological point of view, the adaptive oscillator models several aspects of musical rhythm processing. These include (1) perception, because the oscillator is capable of modifying its behaviour on the basis of some external input; (2) attention, because the oscillator responds to external input only when this input occurs during its output pulse; and (3) memory, because the oscillator retains its behavior even at the absence of any external input. A mathematical description of the adaptation mechanism is given in [17].

Before the IJA can track the beat, it has to infer where it is. This is not a straightforward task, because the performance can in principle start at any point within a beat and it can contain any time values whatsoever. The input sequence often contains a number of different periodicities. According to experimental literature [1,2], a pulse sensation has a high perceptual salience if (1) if the periodicity contains many note onsets; (2) the respective notes have a long duration; and (3) the tempo corresponding to the period is neither too slow nor too high.

The different pulse sensations evoked by the rhythmic sequence are modelled with a set of competing adaptive oscillators. The initial states of these oscillators are set so that each oscillator starts to oscillate at some note onset, and its initial period of oscillation is equal to the interval between that onset and some later onset. All possible combinations of starting points and periods are used. The perceptual salience of a pulse sensation is modelled by a resonance dynamics scheme. The main principle of this scheme is that, at each note onset, the oscillators that are at the peak of their output, start to increase their resonance up to the next note onset. The resonance values of each oscillator are weighted according to the phase velocity, or the speed of oscillation: oscillators having a phase velocity that correspond to the range of most salient pulse sensations receive the highest weighting. After a few note onsets, the oscillator with the highest resonance is sought. This oscillator represents the perceived pulse of the performance. For details of this resonance dynamics scheme, see [18].

Higher level parsing

Both the repertory of the IJA and the improvisation it listens to are represented as a series of profiles. These profiles are 12-dimensional vectors that represent the statistical distribution of pitch classes of the notes played at the respective temporal location. In this representation, octave equivalence is used, with the exception that low notes are weighted slightly more than high note. The reason for this is that low notes are supposed to be more important in deducing the underlying harmonic structure. Figure 3 provides an example of how the profiles drawn from a performance look.

When the system listens to an improvisation, the output of the rhythmic parsing unit tells the onset time of each note in units of quarter notes and fractions thereof. This information is used for deducing if a given note is played on a beat or on the eighth note between two beats. Notes which do not fall into either of these categories are ignored.

Learning. During the learning phase, the IJA constructs sets of tone profiles from performances of tunes it is presented with. These profiles are normalized so that for each tune the average length of profile vectors, defined as the sum of its components, is equal. The normalized profiles are stored in the profile library.

Recognition. When recognizing an improvisation, the IJA constructs tone profiles from the input that are similar to those in the profile library and uses them to test hypotheses about the tune being played and the current location in the harmonic structure. For instance, the first hypothesis is that the first tune of the library is being played starting at the first beat; the second hypothesis is that the first tune is being played starting at the second beat, and so on. Associated with each hypothesis is a likelihood estimate which is continuously updated. This is carried out by adding the dot product of the input profile and the current library profile as determined by the hypothesis to the likelihood estimate. At each instant, the hypothesis with the highest likelihood estimate in considered to be the correct one. When the system finds that, according to the correct hypothesis, the performer is at the top of the form, it starts playing the corresponding MIDI file at the tempo of the performer.

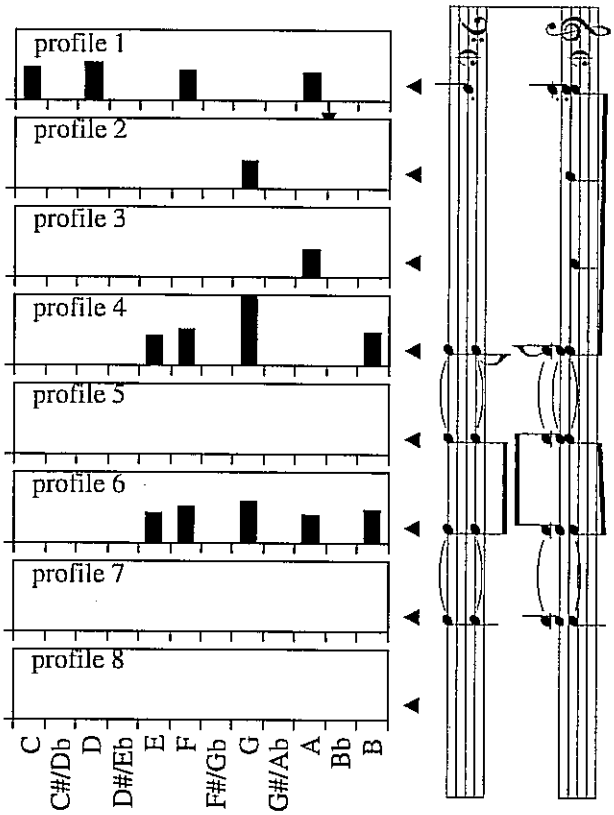


Fig. 3. Tone profiles constructed from an excerpt of a performance of *Satin Doll.*

Conclusion

Experiments carried out with the IJA show that it performs well, provided that the input it receives is not too complicated. If the performance begins with a rhythmically complex part, the inferred beat may not be the correct one. Further, if the improvisation contains plenty of syncopation, the beat-tracker can sometimes go astray and lose the beat totally. With monophonic input, the performer has to outline the harmonic structure carefully in order to be sure that the system makes the right guess. The pattern-matching part of the system seems to perform better with polyphonic input. It is so probably because polyphonic input contains more clues about the underlying harmonic structure.

The IJA has been tested with profile libraries consisting of not more than 10 tunes. It is clear that if

the number of tunes in the library is increased, the system is more prone to make mistakes. Besides, more computing time is needed for processing the incoming data, and this will worsen the temporal resolution of the system. One aim of future work is to explore how large a profile library can be used.

The recognition accuracy of the IJA depends crucially on the form of the tone profiles of the library, that is, how the tunes have been presented during the training phase. There are a number of choices for this. One can, for instance, play either monophonic or polyphonic improvisations on the harmonic structure of the tune, or the melody and the chords of the tune, or just the chords. A further possibility is to represent the tunes as artificially constructed psychoacoustics-based chord templates such as those presented by Parncutt [19]. Future work will focus on finding the optimal form of representing the tunes, that is, the form that would yield the best recognition accuracy over a set of improvisations played by different musicians.

Currently, the IJA employs an absolute representation of pitch. As a consequence of this, it recognizes a tune only if it is played in the right key. A goal of future work is to try to create some kind of a transposition invariant representation, which would make the system capable of recognizing a tune irrespective of which key it is played in.

The IJA does not take into account the hierarchical structure of meter. In other words, it ignores that we perceive metrical music as an alternation of strong and weak beats. Rather, it considers all beats as perceptually equally important. This fact has both technical and cognitive consequences. First, it leads to a computationally ineffective algorithm which limits its use with large profile libraries. Second, it leads to a model whose cognitive relevance leaves much to be desired. One aim of future work is to develop an algorithm which would infer the metric hierarchy of a musical performance. This would probably require that also second-order pitch information, that is, pitch transitions, were taken into account. If the metric hierarchy could be inferred, this would lead to a computationally less intensive algorithm, since the number of alternative hypotheses per tune would be reduced. Furthermore, the cognitive relevance of such a model would be better.

References

[1] Parncutt, R. 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11(4), 409-464.

[2] Palmer, C. & Krumhansl, C. 1990. Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.

[3] Lerdahl, F. & Jackendoff, R. 1983. *A generative theory of tonal music*. Cambridge, MA: MIT Press.

[4] Longuet-Higgins, H. C. & Lee, C. S. 1982. Perception of musical rhythms. *Perception*, 11, 115-128.

[5] Brown, J. C. 1993. Determination of meter of musical scores by autocorrelation. *J. Acoust. Soc. Am.*, 94(4), 1953-1957.

[6] Scarborough, D. L., Miller, B. O. & Jones, J. A. 1992. On the perception of meter. In M. Balaban, K. Ebcioglu & O. Laske (Eds.), *Understanding music with AI: perspectives in music cognition*. Cambridge, MA: MIT Press, 427-447.

[7] Chung, J. 1989. An agency for the perception of musical beats. M. S. thesis. EECS, MIT Media Laboratory.

[8] Desain, P. & Honing, H. 1989. The quantization of musical time: a connectionist approach. *Computer Music Journal*, 13(3), 56-66.

[9] Allen, P. & Dannenberg, R. 1990. Tracking musical beats in real time. *Proceedings of the 1990 ICMC*, 140-143.

[10] Rosenthal, D. 1992. Emulation of human rhythm perception. *Computer Music Journal*, 16(1), 64-76.

[11] Large, E. W. & Kolen, J. F. 1994. Resonance and the perception of musical meter. *Connection Science*, 6(2-3), 177-208.

[12] McAuley, J. D. 1994. Finding metrical structure in time. In M. C. Mozer, P. Smolensky, D. S. Touretzky, J. L. Elman & A. S. Weigend (Eds.), *Proceedings of the 1993 connectionist models summer school*. Hillsdale, NJ: Erlbaum Associates.

[13] Järvinen, T. 1995. Tonal hierarchies in jazz improvisation. *Music Perception*, 12(4), 415-438.

[14] Dannenberg, R. B. & Mont-Reynaud, B. 1987. Following a jazz improvisation in real time. In J. Beauchamp (Ed.), *Proceedings of the 1987 International Computer Music Conference*, pp. 241-248. San Francisco: International Computer Music Association.

[15] Essens, P. J. & Povel, D. 1985. Metrical and nonmetrical representation of temporal patterns. *Perception and Psychophysics*, 37, 1-7.

[16] Jones, M. R. 1987. Dynamic pattern structure in music: recent theory and research. *Perception and Psychophysics*, 41, 621-634.

[17] Toiviainen, P. (in press). An interactive MIDI accompanist. *Computer Music Journal*.

[18] Toiviainen. P. 1997. Modelling the perception of metre with competing subharmonic oscillators. In A. Gabrielsson (Ed.), *Proceedings of the Third Triennial ESCOM Conference*, pp. 511-516. Uppsala: Uppsala University.

[19] Parncutt, R. 1988. Revision of Terhardt's model of the root(s) of a musical chord. *Music Perception*, 6, 65-94.

Music Composition by means of Pattern Propagation

Kenneth B. M°Alpine
Dept. of Mathematics
University of Glasgow
G12 8QQ. Glasgow,
Scotland
km@maths.gla.ac.uk

Eduardo R. Miranda
Dept. of Music
University of Glasgow
G12 8QQ. Glasgow.
Scotland
eduardo@music.gla.ac.uk

Stuart G. Hoggar
Dept. of Mathematics
University of Glasgow
G12 8QQ. Glasgow,
Scotland
sgh@maths.gla.ac.uk

Abstract

Music can be thought of as a form of pattern propagation. In this paper we explore this notion, and show how it relates to the behaviour of a class of system known as cellular automata. We introduce the fundamentals of our work and present CAMUS 3D, an algorithmic composition system which uses cellular automata as the basis for its control system.

Key words: *composition, pattern propagation, cellular automata.*

1. Music as a form of pattern propagation

Music can be appreciated at many different levels. For some, it is sufficient that a composition possesses a pleasant melody that can be hummed along to. Others prefer to dig deeper and are concerned more with the temporal development of the piece – how the initial themes evolve from their exposition through to the work's conclusion. It is at this deeper level that the pattern-based viewpoint can be considered.

We may view each theme in a composition as a separate pattern. As the composition progresses, the patterns (themes) are subjected to certain transformations (such as straight repetition, transposition, inversion, augmentation and so on) according to the formal structure that the composer has chosen for the work. This structure can be rigidly adhered to or used as a general guiding principle, but so long as certain design constructs are in place to guide the temporal development of the composition, we can say that we have a system of pattern propagation according to some predetermined constraints.

Traditionally, composers have employed pattern propagation intuitively, but algorithmic composition techniques allow the pattern propagation to be formalised, albeit at a much higher level: Here, the composer does not, in general, apply specific transformations to a particular pattern. Instead, all of the musical patterns evolve according to the rules and constraints that have been specified at the design stage.

2. Cellular automata

A *cellular automaton* is a *dynamical system* (that is one that changes some feature as time progresses) over which space and time are discrete and all quantities take on discrete values. A cellular automaton is often viewed as an array of elements, referred to as *cells*, to which we apply some evolution rule which dictates how the automaton develops in time. When viewed as such, we see that cellular automata, like human composers, may be considered as a type of pattern propagator – the patterns here are the arrangements of the cells in the automaton, whose evolution is determined by the rules associated with the system. It is a logical step, therefore, to sonify the patterns that arise from the evolution of a cellular automaton, and compare the results with musical compositions that have arisen from other types of pattern propagator, including human composers.

The type and quality of the music which is produced by a cellular automata-driven composition system depends on a number of things, including the evolutionary rules employed, the initial configurations of the automata, and how the automata are mapped to musical output. We illustrate the use of cellular automata as music generators in the following section.

3. CAMUS 3D - a case study

CAMUS 3D (Cellular Automata MUSic in 3 Dimensions) is part of an ongoing research project at Glasgow University's Centre for Music Technology. The system is a development of an earlier two-dimensional system ([1], [2], [3], [4]), and uses three-dimensional extensions of the 'Game of Life' and 'Demon Cyclic Space' automata to generate compositions ([5]).

The Game of Life automaton consists of an array of ($l \times m \times n$) cells, which can exist in two states, alive (usually represented by **1**) or dead (represented by **0**). The rule which determines the development of the automaton is: *A cell will be alive at timestep* $t + 1$ *if and only if it has precisely* **3** *live neighbours at timestep* t. Figure 1 below shows 3 evolutionary steps of the 2-dimensional Game of Life.

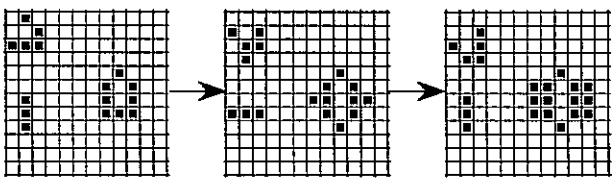


Figure 1 – Three successive timesteps of the 2-dimensional Game of Life.

The Demon Cyclic Space automaton is an array of ($l \times m \times n$) cells which can exist in k states. The evolution of the automaton is determined by: *A cell which is in state* j *at timestep* t *will dominate any neighbouring*

cells which are in state $j - 1$*, so that they increase their state to* j *at timestep* $t + 1$.

It is important to realise that the Demon Cyclic Space is *cyclic*, that is cells in state 0 dominate those in state $n - 1$, so that at any one time each cell is exerting an influence on every other cell.

When the initially randomised automaton is set in motion, the cells self-organise to give us a patchwork pattern like that of figure 2 below.

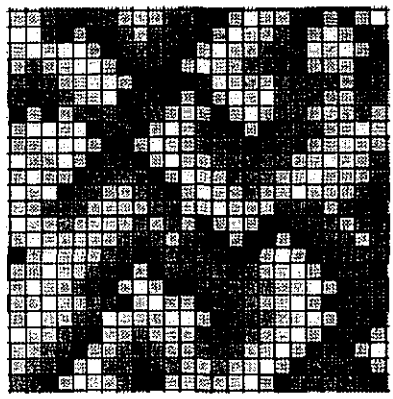


Figure 2 – The Demon Cyclic Space results in a self-organising system that produces patchwork patterns.

To begin the composition process, the Game of Life automaton is initialised with a starting cell configuration, the Demon Cyclic Space automaton is initialised with random states, and both are set to run.

At each time step, the co-ordinates of each live cell are analysed and used to determine a four-note chord[1] which will be played at the corresponding moment in the composition. The state of the corresponding cell of the Demon Cyclic Space automaton is used to determine the *orchestration*[2] of the piece.

This configuration is demonstrated in figure 3. Note that for the sake of clarity, the front two layers of the Demon Cyclic Space automaton have been omitted.

The cell at position (5, 5, 2) in the Game of Life is live and thus constitutes a musical event. The x cell-position (starting at 0 in the bottom left corner) defines a semitone interval from a fundamental pitch to the lower internal pitch of the chord. The y cell-position defines a semitone interval from the lower internal pitch to the upper internal pitch in the chord. The z cell-position defines a semitone interval from the upper internal pitch to the top note of the chord.

Thus, the cell (5, 5, 2) generates a chord in which the notes are 5, 10 and 12 semitones above the fundamental.

Note that if the cell position is 0 (corresponding to the first cell in each direction) the 'higher' pitch defined by the associated interval will be identical to the 'lower' pitch.

The corresponding cell in the Demon Cyclic Space is in state 4, and so the note data is sent to MIDI channel 4.

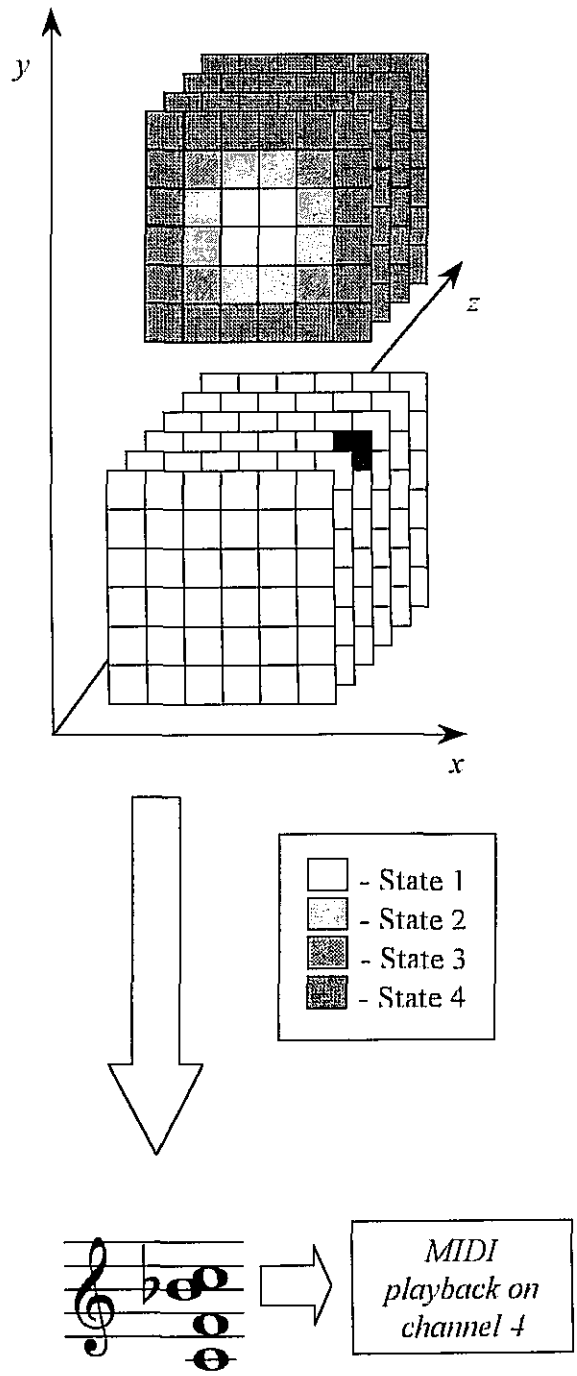


Figure 3 – Mapping the cells of the Game of Life and Demon Cyclic Space to music using the CAMUS 3D algorithm.

[1] In this paper, the term 'chord' is used to describe any set of two or more (not necessarily distinct) notes, which may or may not sound simultaneously.

[2] Similarly, the term 'orchestration' is taken to mean which instrument 'plays' the cells.

This configuration of the points in a discrete three-dimensional Euclidean space being used to represent musical intervals is an extension of the two dimensional *von Neumann Music Space* ([3], [4]).

Having established the intervallic content of the chord associated with a live cell, we must establish the fundamental note in order to specify fully each of the pitches in the chord. This can be done either manually, by reading note data sequentially from a user-specified list, or automatically, by using stochastic selection routines.

Now, in order to avoid a piece being composed entirely of block chords, we must implement a routine which staggers the starting (and possibly ending) times of each of the notes of the chord. It is a simple matter to calculate that there are 24 different ways of arranging the starting order of these 4 (non-simultaneous) notes. CAMUS 3D uses a stochastic selection routine which consults a user-specified table for the associated probabilities of each of the 24 possible starting arrangements.

When a starting arrangement has been chosen, CAMUS 3D calculates the precise note durations. This is achieved by means of a first order Markov chain, a discrete stochastic selection routine that retains a knowledge of the preceding event to aid in the calculation of the next ([6]).

A first order Markov chain may be specified by a state-transition matrix. This is a two-dimensional matrix whose rows and columns are indexed by the possible states. The probability of a transition from state i to state j is given by the entry in the ith row, jth column of the matrix.

We may also present the information contained within the state-transition matrix in the form of a directed graph whose vertices are the states of the Markov chain and whose edges are labelled with the transition probabilities. Figure 4 below shows the state-transition matrix of a Markov chain with three states, A, B and C, and the corresponding graphical representation.

	A	B	C
A	0.25	0.5	0.25
B	0	1	0
C	1	0	0

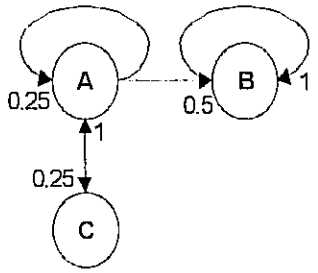


Figure 4 - State-transition matrix and corresponding directed graph for a Markov chain with three states.

A state, B, is said to be reachable from a state, A, if it is possible for the Markov chain to reach state B from state A after a finite number of steps. If state B is reachable from state A and state A is reachable from state B, then the two states are said to communicate. For example, in figure 4 above, states A and C communicate, since clearly C is reachable from A and A is reachable from C. However neither, A and B nor B and C communicate, since, although B is reachable from A (and thus C), states A and C are not reachable from B.

It can be shown fairly easily that the communication relation on a Markov chain is an equivalence, and thus, we can partition the states of the chain into equivalence classes of communicating states. One event may be reachable from an event of a different communication class, but the two cannot communicate.

We saw in figure 4 above that state B is reachable from state A but not vice-versa. Thus, once the Markov chain reaches state B, it can never return to state A. In a sense, the chain is trapped inside state B's communication class.

In general, we can further classify the states of the Markov chain in the following way. Those states which, given a sufficiently long timeframe are certain to occur again once they have been reached are called recurrent and the equivalence class to which they belong is known as that state's recurrent class. States which may never occur again (i.e. those that are not recurrent) are called transient, and belong to transient classes. It can be shown that every Markov chain

consists of at least one recurrent class and some number (possibly none) of transient classes.

Markov processes are exceptionally well suited for rhythm selection. Rhythmic lines often exhibit semi-cyclic behaviour in that short phrases often repeat exactly or in a slightly altered form as the line progresses – the human ear tends to like this sort of regularity. Similar behaviour can be engineered by careful manipulation of the recurrent and transient classes within the Markov chain.

The probabilities in the transition matrix of the Markov chain used by CAMUS 3D are, again, user-specified. Note lengths are quantised to semiquaver resolution, and simultaneous note events are catered for by allowing starting times of duration 0.

Finally, CAMUS 3D offers the user a palette of General MIDI sounds on which to perform the composition. Selection of the instruments is achieved by associating each possible state of the Demon Cyclic Space with the appropriate MIDI instrument. Patch changes are sent as MIDI patch numbers, so the files will work with non-GM instruments provided care is taken over the instrument setup.

When the composition process is started, the music is performed in real time, and can be saved as a type 0 standard MIDI file.

CAMUS 3D also allows the user to save the composition as a set of parameters which correspond to the states of the automata and the probability tables for the selection routines. Whilst this allows the composer to re-create the 'same' composition, the resulting music may sound quite different, since although the automata are wholly deterministic, and so produce identical chord sequences and instrumentations each time they run with identical initial configurations, the stochastic selection routines may lead to quite different fundamental pitches, note orderings and note durations.

4. Conclusion and further research

CAMUS 3D has proven itself to be a viable mechanism for driving musical composition, and has been successfully used to compose a number of works, including a prize-winning piece for chamber orchestra, *Entre l'Absurde et le Mystère*, by Eduardo Miranda. The music that is generated in general exhibits a specific formal style centred around a four-note phrase. For the composer who is prepared to put a little effort into the system, the results can be very pleasing, and often sound much more natural than compositions obtained using comparable algorithmic techniques.

The system is still under development, and we hope to address some of its limitations in the near future. For example, at present, the program does not have any knowledge of the idiosyncrasies of specific musical instruments, which may give rise to musical impossibilities, such as chords in a solo flute line.

Further developments for the system include implementing a number of different forms of pattern propagation, such as fractal zooms and other types of automaton, to drive the composition. More complex evolution rules and tools for manipulating the Game of Life cells along with a dynamics processor which will colour the music in a natural way are also planned.

A fully featured demonstration version of CAMUS 3D may be obtained by contacting the authors. An earlier version was published in issue 45 of the Mix, a British newsstand publication, and can also be obtained from the authors.

Acknowledgements

Many thanks to the Carnegie Trust for the Universities in Scotland, who generously provided the funding which has enabled this research. Thanks also to the University of Glasgow for providing both a research position and the facilities to undertake research in this field.

References

[1] K. B. McAlpine, S. G. Hoggar and E. R. Miranda, *Dynamical Systems and Applications to Music Composition: A Research Report.* Proceedings of Journees d'Informatique Musicale: 106 – 113, 1997.

[2] K. B. McAlpine, S. G. Hoggar and E. R. Miranda, *A Cellular Automata Based Music Algorithm: A Research Report.* Proceedings of IV Brazilian Symposium on Computer Music: 7 – 17, 1997.

[3] E. R. Miranda, *Cellular Automata Music: An Interdisciplinary Project.* Interface 22: 3 – 21, 1993.

[4] E. R. Miranda, *Music Composition Using Cellular Automata.* Languages of Design 2: 105 – 117, 1994.

[5] S. Wolfram, *Universality and Complexity in Cellular Automata.* Physics 10: 1 – 35, 1984.

[6] C. Roads, *The Computer Music Tutorial.* MIT Press, 1995.

Friday 25th **h. 9.00**

MUSIC ANALYSIS AND COGNITION

Musical Parallelism and Melodic Segmentation

Emilios Cambouropoulos

University of Edinburgh
Faculty of Music and Department of Artificial Intelligence
emilios@music.ed.ac.uk

Abstract

In this paper a formal model will be presented that attempts to segment a melodic surface based on both local discontinuities of the surface and higher-level melodic parallelism. Special emphasis will be given to the influence of musical parallelism on the segmentation of a surface; this is based on the assumption that the beginning and ending points of 'significant' repeating musical patterns influence the segmentation of a musical surface. This model has been implemented as a prototype computer system and has been applied successfully to melodies from diverse musical repertoires.

1 Introduction

Segmentation of a musical surface is a central part of musical analysis; an initial selected segmentation can seriously affect subsequent analysis as a great number of inter-segment musical structures are automatically excluded. Many systematic theories of music - such as pitch-class set theory [6], and paradigmatic analysis [11,12] - suffer on the issue of surface segmentation; they require some sort of pre-processing of the surface into segments which relies on explicit or implicit knowledge on the part of the human musician/analyst. Even the *GTTM* [8] that partially formalises the grouping rules at the local level avoids any systematic description of musical parallelism and its contribution towards the determination of the grouping structure of a musical surface. In the Implication-Realisation Model [9,10] the notion of 'closure' - on which grouping and transformation of notes to higher-levels are based - relies on the interaction of factors such as metre, harmony and musical similarity which are not fully described by the model.

Cambouropoulos [1,2] has proposed a formal model that attempts to define local boundaries in a given melodic surface. The *Local Boundary Detection Model (LBDM)* calculates boundary strength values for each interval of a melodic surface according to the strength of local discontinuities. The model suggests *all* the possible points for local grouping boundaries with various degrees of prominence attached to them (normalised from 0-100) rather than *some* prominent boundaries based on a restricted set of heuristic rules (see local maxima in Local Boundary strength profiles in figures 2, 3, & 4).

In this paper it is suggested that the segmentation of a musical surface is not only affected by local discontinuities but by higher-level processes as well. Perhaps the most important of these higher-level mechanisms is *musical parallelism*, i.e. similar musical patterns tend to be highlighted and perceived as units/wholes whose beginning and ending points influence the segmentation of a musical surface. For instance, a model for determining local boundaries would select the interval between the 3rd and 4th notes of *Frère Jacques* (see figure 1) as a local boundary (larger pitch interval in between smaller ones) whereas it is obvious that a boundary appears between the 4th and 5th notes because of melodic repetition.

2 The *SPIA* and *Selection Function*

The *String Pattern-Induction Algorithm (SPIA)* is a brute-force pattern-matching algorithm that can be applied to any sequence of entities - see brief description in [3] and a more extended formal description of an almost identical algorithm in [5]. The aim of the algorithm is pattern induction; more specifically *SPIA* discovers *all* the patterns that recur in a string of symbols.

The *String Pattern-Induction Algorithm (SPIA)* is employed in a bottom-up fashion, i.e. starting from the smallest patterns and extending them to maximum length (overlapping of patterns is allowed). For a given sequence of entities (e.g. a parametric profile of scale-step pitch intervals), the matching process starts with the smallest pattern length (2 elements) and ends when the largest pattern match is found. For a given pattern length, every possible pattern of the string (starting with the first) is matched against the remainder of the string by a shifting stepwise motion. The patterns for which at least one match is found are separated and labelled (melodic patterns may be matched in their original form or in their retrograde, inversion and retrograde inversion forms). Patterns for which no match is found are disregarded after the introduction of a *break* marker in their place. Pattern-matching cannot override such markers and the initial sequence is in essence fragmented into shorter sequences. As the matched patterns grow in size, the search space is usually reduced. When the last matching is found for the largest possible pattern, the matching process ends.

The *String Pattern-Induction Algorithm* is exhaustive, i.e. it discovers all possible matches, and although it is computationally expensive (polynomial time), it becomes more efficient through the reduction of the initial search space. (An efficient algorithm that computes all the repetitions in a given string is described in [4,7] - not as yet been implemented as part of the current prototype system. This algorithm takes $O(n \cdot \log n)$ time where n is the length of the string. It should also be noted that this algorithm does not match retrograde and inverted forms of patterns.)

The *SPIA* procedure can become significantly faster if *break* markers are inserted in the initial sequence for positions that are thought to be important boundaries in the sequence (e.g. for a melody, points suggested by the *LBDM* or positions marked in a score by breath marks, large rests, slurs, fermatas, and so on). It is also possible to pre-define a limited range of pattern lengths for which the *SPIA* will be employed. The *SPIA* is applied to as many parametric profiles as are considered necessary (e.g. pitch, duration, start-time, dynamic intervals and so on) for the melodic surface and/or reductions of it.

It is apparent that such a procedure for the discovery of parallel melodic segments will produce a very large number of possible patterns most of which would be considered counter-intuitive and non-pertinent by a human musician/analyst.

A complementary procedure has been devised whereby a prominence value is attached to each of the discovered patterns based on the following factors:

a) Prefer longer patterns
b) Prefer most frequently occurring patterns
c) Avoid overlapping.

A *Selection Function* that calculates a numerical strength value for a single pattern according to these principles may be:

$f(\mathrm{PL},\mathrm{F},\mathrm{DOL}) = \mathrm{F}^{a} \cdot \mathrm{PL}^{b} / 10^{c \cdot \mathrm{DOL}}$

where: PL: pattern length; F: frequency of occurrence for one pattern; DOL: degree of overlapping; a, b, c: constants that give different prominence to the above principles.

For every pattern discovered by *SPIA* a value is calculated by the *Selection Function*. The patterns that score the highest should be the most significant ones.

3 Segmentation based on parallelism

The computational model that consists of the *String Pattern-Induction Algorithm* and the *Selection Function* provides a means of discovering 'significant' melodic patterns. Figure 1 illustrates the most prominent pitch patterns for the song *Frère Jacques* highlighted by the *SPIA & Selection Function*. There is though a need for further processing that will lead to a 'good' description of the surface (in terms of exhaustiveness, economy, simplicity etc.). It is likely that some instances of the selected pitch patterns should be dropped out or that a combination of patterns that rate slightly lower than the top rating patterns may give a better description of the musical surface.

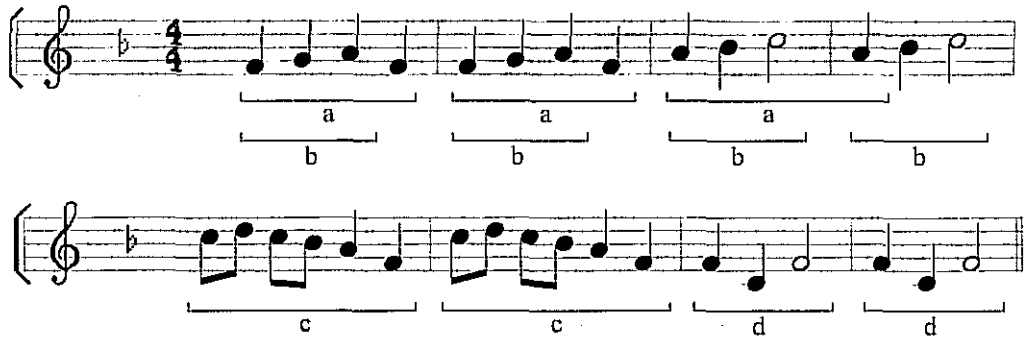


Figure 1 *Frère Jacques* - most prominent pitch-patterns highlighted by the *SPIA* and *Selection Function* (the *SPIA* is applied only on scale-step pitch profile, and the *Selection Function* constants are set to (a,b,c)=(3,3,4))

In order to overcome this problem a very simple but crude methodology has been devised. According to this, the *SPIA* algorithm is applied to as many parametric profiles of the melodic surface and reductions of it as required. No pattern is disregarded but each pattern contributes to each possible boundary of the melodic sequence by a value that is proportional to its *Selection Function* value. That is, for each point in the melodic surface all the patterns are found that have one of their edges falling on that point and all their *Selection Function* values are added together. This way a Pattern Boundary strength profile is created (normalised from 0-100). It is hypothesised that points in the surface in which local maxima appear are more likely to be perceived as boundaries because of musical parallelism. See, for instance, the local maxima in the Pattern Boundary strength profiles of figures 2, 3 & 4; note especially the points indicated by asterisks where boundaries are suggested because of musical parallelism rather than local discontinuities.

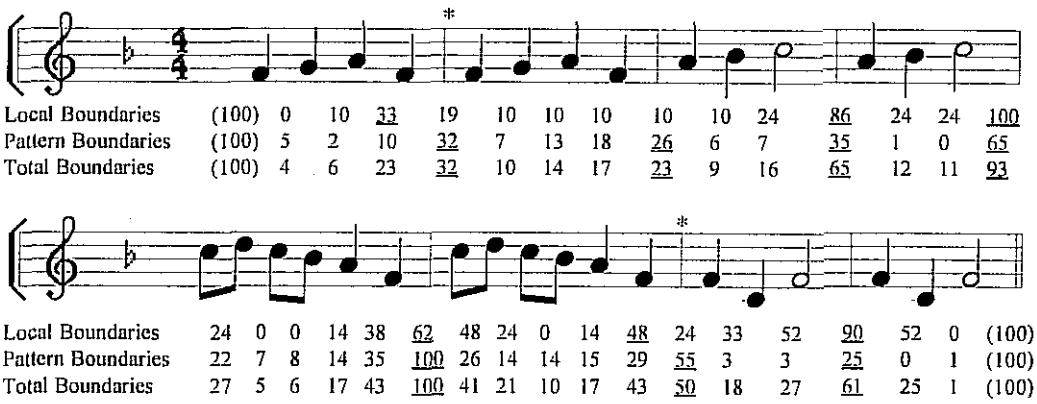


Figure 2 Local Boundaries strength profile, Pattern Boundary strength profile and the Total Boundary strength profile for the song *Frère Jacques.*

4 Interaction between *SPIA* and *LBDM*

Firstly, significant boundaries discovered by the *LBDM* can be used as a guide for inserting *break markers* in the musical surface (as suggested in section 2). The assumption underlying this procedure is that a listener may use strong local boundary cues as tentative points of segmentation which are unlikely to be overridden by a pattern.

Secondly, the boundaries discovered by the pattern-matching process may complement the local boundaries detected by the *LBDM* in defining the Total Boundary strength profile. In the melodic examples of figures 2, 3 & 4 the Pattern Boundary strength profile has been calculated by applying the *SPIA* to the scale-step, contour and duration profiles. The Total Boundary strength profile is calculated as a weighted average of the Local Boundary and Pattern Boundary strength profiles - in this implementation they contribute by 40% and 60% respectively. The local maxima in the Total Boundary strength profile can be taken as a guide for the segmentation of the musical surface (in figure 2 all the local maxima are underlined whereas in figures 3 & 4 only the local maxima that pass a certain threshold in terms of boundary strength and boundary sharpness are shown; further research is necessary to determine the exact way by which a final selection of boundaries should be made that may lead to one or more preferred segmentations).

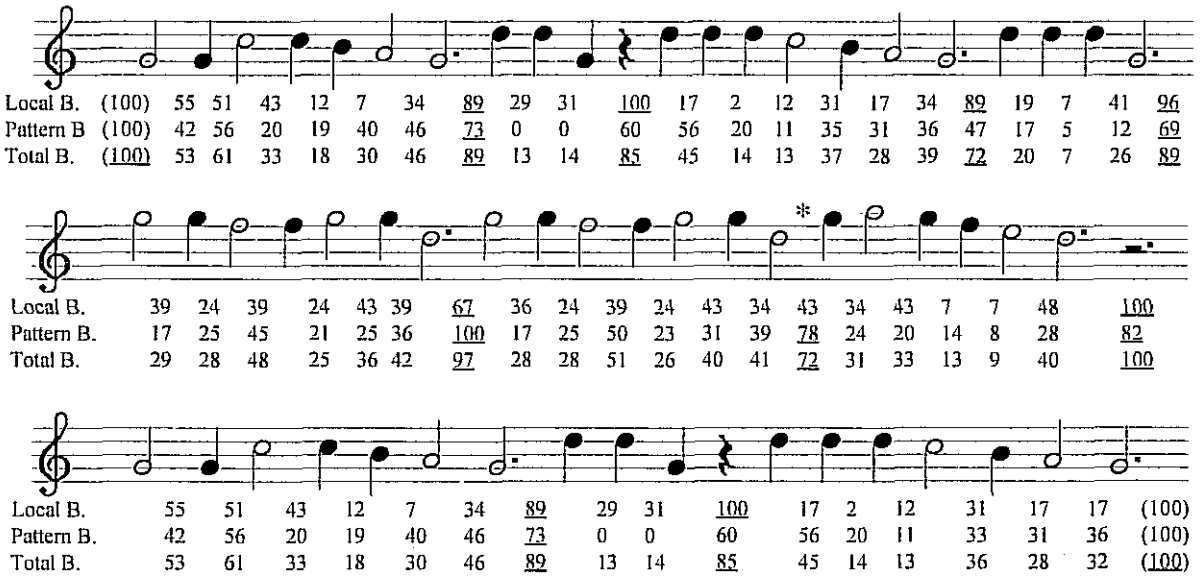


Figure 3 Local Boundaries strength profile, Pattern Boundary strength profile and a weighed Total Boundary strength profile for the melody *L'homme armé* (15th century melody as presented in the New Grove's Dictionary)

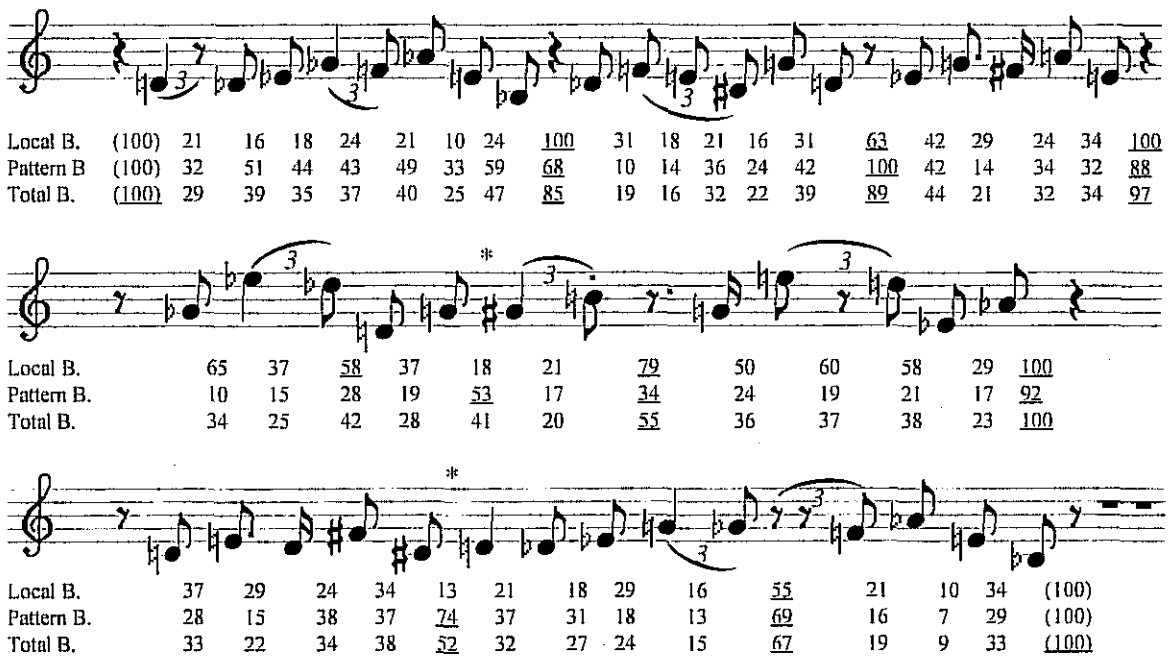


Figure 4 Local boundary strength profile, Pattern Boundary strength profile and the Total Boundary strength profile for the voice part of the first song from Webern's *Fünf Lieder Op.3*

Conclusion

A computational model has been introduced that discovers 'significant' patterns for a given parametric profile of a melody. This model can be applied to a number of parametric profiles of a melody and the results of each of these can be combined to produce a Pattern Boundary strength profile indicating the most prominent boundary positions due to musical parallelism. This, in conjunction with the local boundaries highlighted by the *Local Boundary Detection Model* leads to an integrated segmentation of a melodic surface.

References

[1] Cambouropoulos, E. (1997) Musical Rhythm: Inferring Accentuation and Metrical Structure from Grouping Structure. *Music, Gestalt and Computing - Studies in Systematic and Cognitive Musicology*, M. Leman (ed.), Springer-Verlag, Berlin.

[2] Cambouropoulos, E. (1996) A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface. *Proceedings of the III Journées d' Informatique Musicale*, Caen, France.

[3] Cambouropoulos, E. and Smaill, A. (1995) A Computational Model for the Discovery of Parallel Melodic Passages. *Proceedings of the XI Colloquio di Informatica Musicale*, Bologna, Italy.

[4] Crochemore, M. (1981) An Optimal Algorithm for Computing the Repetitions in a Word. *Information Processing Letters*, 12(5):244-250.

[5] Crow, D. and Smith, B. (1992) DB_Habits: Comparing Minimal Knowledge and Knowledge-Based Approaches to Pattern Recognition in the Domain of User-Computer Interactions. In *Neural Networks and Pattern Recognition in Human-Computer Interaction*, R. Beale et al. (eds), Ellis-Horwood, London.

[6] Forte, A. (1973) *The Structure of Atonal Music.* Yale University Press, New Haven.

[7] Iliopoulos, C.S., Moore, D.W.G. and Park, K. (1996) Covering a String. *Algorithmica*, 16:288-297.

[8] Lerdahl, F. and Jackendoff, R. (1983) *A generative Theory of Tonal Music*, The MIT Press, Cambridge (Ma).

[9] Narmour, E. (1992) *The Analysis and Cognition of Melodic Complexity.* The University of Chicago Press, Chicago.

[10] Narmour, E. (1990) *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realisation Model.* The University of Chicago Press, Chicago.

[11] Nattiez, J-J. (1990) Music and Discourse: Towards a Semiology of Music. Princeton University Press, Princeton.

[12] Nattiez, J-J. (1975) *Fondements d'une Sémiologie de la Musique.* Union Générale d'Editions, Paris.

Extraction of Music Harmonic Information Using Schema–Based Decomposition

Francesco Carreras
CNUCE/CNR
via Santa Maria 36
I-56126 Pisa, Italy
F.Carreras@cnuce.cnr.it

Marc Leman
Univ. of Ghent, IPEM
Blandijnberg 2
B-9000 Ghent, Belgium
Marc.Leman@rug.ac.be

Danilo Petrolino
CNUCE/CNR
via Santa Maria 36
I-56126 Pisa, Italy
marvid@tin.it

Abstract

This paper presents a computer model for describing the harmonic content of musical signals. The model consists of an auditory part by which musical signals are transformed into chord images, using virtual pitch recognition and onset detection. The chord images are then decomposed by mapping the chord images onto a schema of subchords from which the labels are inferred.

1 Introduction

In the last decades, much effort has been focused on attempts that aim at clarifying aspects of harmony and tonality from the point of view of perception and cognition. Experimental work in music psychology has thereby focused on aspects of consonance, virtual pitch perception, and context-based pitch perception. The results have revealed that dependency of pitch, harmony and tonality rely on both sensory and learned aspects. Effort has subsequently been undertaken to reframe the classical rules of harmony and tonality in agreement with the empirical findings (e.g. Parncutt, 1997; Balsach, 1997). Computational models have been built which aim to simulate the processes that underlay pitch and tonality perception. The signal–based models give an account of the complexities of context-based pitch perception from a physiological point of view, taking into account aspects of auditory information processing.

A study by Hanappe (1994) explored the possibilities of chord recognition using the schema architecture developed in Leman (1994). Camurri and Leman (1997) describe a system in which recognition of chords and tone centers are combined within a logical reasoning system in order to trace harmonic patterns such as cadences. The models assume that the recognition of general features of an object is easier than the recognition of the components that make up the object. Applied to music it means that the recognition of tonal centers is more straightforward than the recognition of chords types and the recognition of the individual pitches that make up the chords.

In this paper, extraction of harmonic information is based on chordal decomposition in combination with a schema–based approach. Being able to transform musical signals into a set of chord labels can be considered as a first step in a harmonic analysis task which, from that level on can proceed in terms of propositional reasoning.

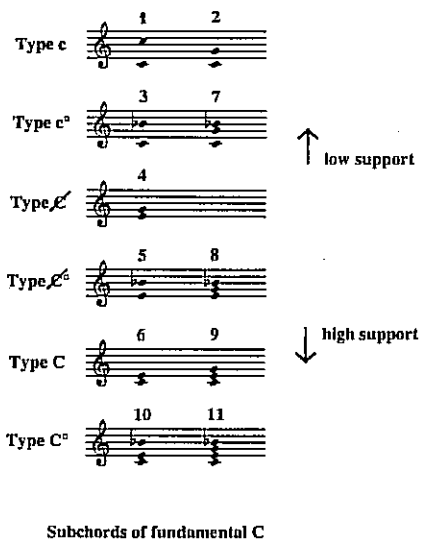


Figure 1: The schema-based decomposition model

2 The Modelling Framework

The framework for the present model is based of a theory of chordal decomposition and a theory of memory representation.

2.1 Chordal Decomposition

The technique of decomposition has been proposed by the Swiss mathematician and physicist L. Euler(1707-1783). His "gradus suavitatis" (or pleasantness of intervals) for the calculation of the degree of consonance is based on a decomposition of natural numbers into a product of powers of different primes. Recently, Tanguiane (1993) has proposed a decomposition method for chord recognition based on correlational analysis of sound spectra that make up chords. A drawback of both approaches is that they do not take into account the constraints of the auditory system nor the spectral complexity of real sounds.

The chordal decomposition model put forward by Balsach (1997) makes reference to the auditory phenomenon of virtual pitch, i.e. the fact that a collection of partials tends to be completed provided that the partials fit as harmonics of a low pitch (or fundamental). The low pitch does not necessarily have to occur as a partial of the collection. In the latter sense it is a real virtual pitch.

Since we have 11 subchords (Fig. 1) for each of the 12 pitches of the chromatic scale, there are 132 subchords onto which a given chord needs to be mapped in order to achieve the decomposition. They are labeled as $C4$(=Fundamental C, 4th chord as in Fig. 1), $C\sharp 6, D2, B6$, etc. In this set, some subchords may point to different fundamentals. For example, the chord containing the notes $E - B\flat$ has a support for both C and $F\sharp$. It is labeled $C6$ or $F\sharp 6$ (as $A\sharp - E$).

2.2 Schema Theory

Schema theory assumes that the memory for context-dependent pitch is somehow structured and organized. In the present model, we rely on the two-layered architecture developed in Leman (1995), nl. an auditory model connected to a two-dimensional topological memory structure which is called the schema. The latter is used here as a container of the 132 subchords and it is a model for the decomposition. That means that the decomposition of a given chord is achieved by a projection of the chord onto the map of subchords. The chord activates different regions on the schema from which we then infer the precise nature of the decomposition in terms of subchord labels. The latter provides the symbolic description of the musical signal.

The decomposition gives a description of the content of a piece in terms of harmonic content. An attractive feature of Balsach's proposal is that it provides a description for all possible combination of notes, unlike the classical description of chords which is restricted to a subclass of note combinations (mainly superpositions of major and minor thirds). The description of the harmonic content is therefore not restricted to the classical period but can in principle be applied to modern music (e.g. Bartòk, Strawinsky, ...) as well.

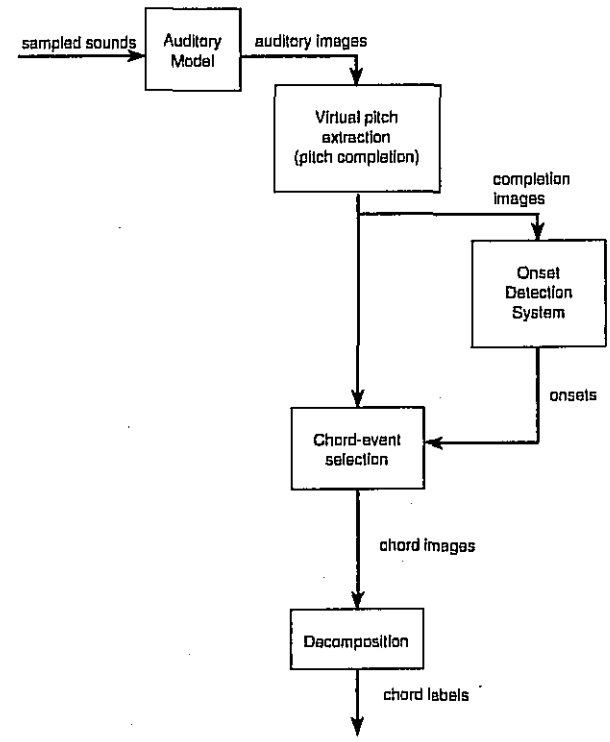


Figure 2: The schema-based decomposition model

3 System Architecture

Figure 2 gives an overview of the system architecture. A distinction is made between the auditory model, the virtual pitch extraction part, the onset detection part, a module for chord event selection, and a decomposition module.

- The auditory model takes as input the musical signal sampled at 20 kHz. The output are auditory images that represent the auditory nerve patterns of twenty auditory channels (Van Immerseel & Martens, 1992). The images are presented at intervals of 0.4 ms.

- The virtual pitch extraction part takes the auditory images as input and performs a periodicity analysis in terms of autocorrelation patterns along each of the twenty channels. Frames of 32 ms are analysed at intervals of 4 ms. The output images are called *completion images* (Leman,

1995).

- The onset detection system, based on (Moelants & Rampazzo, 1997), takes as input the completion images and produces as output a list of events containing (i) the time of the onset, (ii) the time index of the start of the stable regime of the chord, and (iii) the onset energy.
- The chord event selection module, takes as input two types of information nl., the completion images and the onset events and produces as output a chord image. Using the onset event list, a weighted mean image is calculated starting from the stable regime of a given chord to the onset of the next chord. This is done for every onset. The weight function is a decaying exponential. The chord images are obtained by normalizing the weighted mean images in an Euclidean way (so that the norm of each vector representing the chord image is 1).
- The decomposition module takes as input the chord images and maps these onto the schema. The output of this mapping is then processed into a list of chord labels which describe the harmonic content of the chord in terms of the subchords. Two aspects should be taken into account here: first the construction of the schema, and secondly, the mapping of the chord image onto the subchord container.

3.1 Construction of the Chord–Schema

The chord–schema is to be conceived of as a neural network based memory representation for 132 prototype subchords. The schema is constructed by means of a training procedure (Carreras & Leman, 1996). The data set for the training is constructed using 40 completion images for each subchord. The images are the outcome of processing the sounds of an electronic piano keyboard (Korg) with the auditory module. The training set thus comprised 5280 completion images (=11 subchords for each of the 12 fundamentals with 40 images each).

Two different schemata have been constructed and evaluated. They are called the "one–octave" and the "four–octave".

3.2 Decomposition as Mapping onto the Chord–Schema

Chord decomposition is here defined as the mapping of the chord image onto the chord–schema according to an algorithm which calculates the activation of each neuron (representing a prototype subchord) in the map according to its similarity with the chord image presented. The labeling for the decomposition is obtained by choosing for each of the 12 fundamentals the subchord (out of 11 subchords) with the highest activation. The highest scored subchord is given the value of 100 and the values of the other subchords are accordingly scaled.

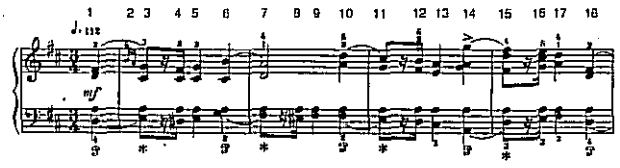


Figure 3: First measures of *Kuriose Geschichte* by Robert Schumann. The chord events are labeled from 1 to 18

4 Schumann's *Kuriose Geschichte*

The first five measures of Robert Schumann's (1810-1856) composition *Kuriose Geschichte* (events 1 up to 17) played by Martha Argerich on piano (Deutsche Grammophon 410 653-2) have been sampled and processed using the decomposition model. The onset model found the correct onsets except for the appogiatura where the notes are played too fast to be able to recognize them using the analysis window of 32 ms. The decomposition model provides labels as in Table 1.

5 Discussion

The output of the decomposition module selects for each fundamental only one subchord, i.e. the subchord with the highest activation. This generates some problems and the procedure may be improved in the future. In the first event, for example, f♯1 has the highest score followed by d6(34.1). However, other high scores of the subchords that point to the fundamental F♯ may be present. They are not selected. Out of the subchords that point to the fundamental D, the current procedure selects subchord d6 (see Fig. 1), yet the chord d9 may be quite close too (as it would cover the note A that is present in the score). In the current procedure, it is not selected either. Subchords pointing to other fundamentals have low scores and are not taken into account here. Recall, however, that d6 and d9 belong to the same chord type (see Fig. 1 transposed to the fundamental D) and that the support of a *Type D*–chord for the fundamental D is higher than the support of a *Type f♯*–chord for the fundamental F♯. A reasoning procedure will be developed in which these aspects are taken into account.

Table 1: Labeled events of Schumann's *Kuriose Geschichte*

event	onset time	energy	labels
1	0.156	146.049	f♯1 + d6(34.1)
2	0.520	16.155	g4 + e8(61.8)
3	0.612	39.440	a3 + c4(56.1)
4	0.960	100.370	f♯2 + b4(41.8)
5	1.076	29.351	a8 + c4(68.8)
6	1.516	43.107	e4 + g6(35.3)
7	1.940	64.803	d2 + a♯8(9.1)
8	2.296	22.489	f9 + a♯4(36.5)
9	2.384	23.259	d4 + f♯6(18.4)
10	2.816	62.043	d6 + f♯6(20.0)
11	3.216	62.292	a4 + f♯11(52.9)
12	3.560	59.236	b11 + e8(94.7)
13	3.692	13.814	a1 + f10(21.3)
14	4.128	76.760	a3 + d♯4(42.0)
15	4.536	68.535	d4 + b8(45.3)
16	4.864	91.454	a8 + c4(25.8)
17	4.992	34.229	d6 + a♯8(11.9)

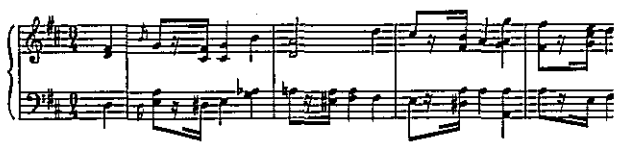


Figure 4: *Kuriose Geschichte* by Robert Schumann reconstructed from Table 1

The harmonic content extraction using the labels does not contain any information about the octave position of the notes. Nevertheless, it is possible, using the chord type, to reconstruct the score using the data contained in Table 1. The notes are taken from the information contained in the table, but the octave positions and doublings are inspired by the original score. The result is shown in Fig. 4.

Compared to the original score some notes are missing, and some notes are found which are not present in the score. In event 6, the model finds a G♯ and a G, whereas there is an A instead of a G♯. Also the C is missing. Event 12 is wrong in that the D♯ is heard instead of D. Subchord e8 gets a high score. This chord contains the notes G♯–B–D which shows that the model somehow hesitates between D–A (present in the score) and D♯–G♯ (not present).

Analysis of the results have revealed that the model could be improved at different levels. Future work will take these into consideration. Unlike the previous studies on tonal induction, no claims are made regarding the epistemological relevance of the present model.

References

Balsach, L. (1997). Application of virtual pitch theory in music analysis. *Journal of New Music Research, 26*(3), 244-265.

Carreras, F., & Leman, M. (1996). Distributed parallel architectures for the simulation of cognitive models in a realistic environment. In E. D'Hollander, G. Joubert, F. Peters, & D. Trystram (Eds.), *Parallel computing: State-of-the art perspective* (p. 585-588). Amsterdam: Elsevier. (ISBN 0 444 82490 1)

Hanappe, P. (1994). *Het herkennen van akkoorden in een muzikaal signaal.* Unpublished master's thesis. (Universiteit Gent, Vakgroep voor Elektronica en Informatiesystemen)

Leman, M. (1994). Schema-based tone center recognition of musical signals. *Journal of New Music Research, 23*(2), 169–204.

Leman, M. (1995). *Music and schema theory: Cognitive foundations of systematic musicology.* Berlin, Heidelberg: Springer-Verlag.

Moelants, D., & Rampazzo, C. (1997). A computer system for the automatic detection of perceptual onsets in a musical signal. In A. Camurri (Ed.), *KANSEI - The Technology of Emotion* (p. 140-146). U. Genova, Teatro C. Felice.

Parncutt, R. (1997). A model of the perceptual root(s) of a chord accounting for voicing and prevailing tonality. In M. Leman (Ed.), *Music, Gestalt, and computing: Studies in cognitive and systematic musicology.* Berlin, Heidelberg: Springer-Verlag.

Tanguiane, A. (1993). *Artificial perception and music recognition.* Berlin, Heidelberg: Springer-Verlag.

Van Immerseel, L., & Martens, J. (1992). Pitch and voiced/unvoiced determination with an auditory model. *The Journal of the Acoustical Society of America, 91*, 3511–3526.

Is there anisotropy in the acoustic representation of space?

Fabio Ferlazzo*, Clelia Rossi-Arnaud°^, Marta Olivetti Belardinelli°^

* Dip. di Psicologia, Università di Cagliari
° Dip. di Psicologia, Università di Roma 'La Sapienza'
^ ECONA, Interuniversity Centre for Research on
Cognitive Processing in Natural and Artificial System

Introduction

Speed of processing as well as attentional abilities are considered by information-processing theorists as important indicators on the mental manipulations of stimuli incoming from the external world, both in simple and in complex situations. However, the inference from these indicators is made troublesome by intriguing questions concerning the detection of isolated events in an external space. In perception, the segregation of events and objects (binding problem) is normally performed by means of different sensory information (sensory integration problem) which represents simultaneously the stimulus position at a given time in a surrounding space (temporal and spatial dislocation problems).

Moreover, these questions are strictly interconnected and for this reason psychologists formulated different types of explanations based on: a) different speed of sensory information incoming from different channels (starting with Exner [2]), b) fluctuation of attention (starting with Wundt [7]) or c) phenomenal relevance in perceived structure (starting with Benussi [1]). For a critical review, see Vicario [6].

Posner's recent studies connect directly the two hypotheses concerning speed of processing and attentional ability not only on the base of the distinction between two different attentional systems (the anterior essentially devoted to attending to word meaning and to selecting among alternative courses of action, and the posterior one principally activated in tasks involving visual spatial attention) but also as a consequence of findings that show that both systems may enhance neural activity in cortical areas involved in particular tasks (e.g. visual, auditory and motor ones), and that attentional activity occurs in three different forms (enhanced activation of attended items, inhibition of unattended items, or both) according to the nature of the task (Posner & Dehaene [4]).

The problem of attention activity in perception is connected also to the relationship between attention and consciousness by means of the mental representation derived from perception. From Egon Brunswick's and the New Look's first probabilistic conceptions of perception to Marcel's complete model [3], several psychologists referred to an unconscious matching process between sensory data and perceptual hypotheses that determines the final perceptual scene.

As regard the event scene, many experiments in cognitive sciences have tried to ascertain how the cognitive system creates a mental spatial representation and what is the relationship between this mental representation and the external space. One of the most intriguing aspects of the structure of the spatial representation regards its anysotropy. Spatial representation is not isomorphic to the spatial environment in that vertical and horizontal meridians of the visual field seem to be over-represented, since the time necessary for selective attention to move from one point of the visual field to another is greater if it has to cross the meridians (Rizzolatti et al. [5]). Most of todays literature deals with the visual spatial representation, largely neglecting the structure of the acoustic spatial representation and the relationship between the two. This problem is relevant, from both an applicative and a theoretical point of view, since it could shed some light on the problem of how a spatial representation is created from the environmental information and used in our everyday life.

Our aim is to investigate if the acoustic representation of space can be considered anysotropic in the same way as the visual one and, in this case, which relationships exist between the two.

Material and method

Subjects

Fifteen subjects aged 20 to 30 volunteered for the experiments. All subjects with normal hearing and normal or corrected to normal vision were naive with respect to the experimental procedures.

Stimuli

In both experiments the acoustic target stimulus was a 1000 Hz tone delivered through stereo-headphones. The amplitude of the right and left channels was varied in order to obtain four different subjective sources of the stimuli: tones were subjectively localized at the left ear (position A), at the right ear (position D), at the midpoint between the left ear and the vertex (position B), and at the midpoint between the right ear and the vertex (position C).

The visual cue was a black rectangle presented in one of four positions on the computer screen. Positions (A', B', C', D' from left to right) were horizontally arranged and evenly spaced. The combination of the four positions of the cue and the target gives a number of combinations: the positions of the cue and the target could be the same (e.g. A'A, «valid» trials) or different (e.g. A'B, «invalid» trials). In the «neutral» condition an uninformative central visual cue was presented to the subject.

Procedure

During the first experiment subject were required to maintain their gaze on a fixation point constantly displayed at the center of the screen. In this way the visual and somatic (head centered) vertical meridians were coincident. Each trial of the experiment started with the appearance of the visual cue at one of the four positions on the screen or at the center of the screen (neutral condition) for 1000 msec, then one of the four acoustic targets was delivered for 50 msec. Subjects were required to press a push button held in their right hand as soon as they detected the tone, independently from its position. They were also informed that the cue indicated the most probable position of the target so that the best strategy to achieve a fast reaction time was to allocate their attention to the position cued by the visual stimulus. The first experiment comprised two blocks of 80 trials each. The percentage of «valid» trials with respect to the «invalid» one was 75%. On invalid trials subjects had to shift their attention from the cued position to the position where the tone occurred. This shifting could include a crossing of the vertical meridian (for example when the cue was in position B' and target in position C) or not (for example when the cue was in position A' and the target was in position B). The inter stimulus interval was 2500 msec.

In the second experiment the stimuli and the procedure were identical to the first one, except for the position of the fixation point which was moved between positions A' and B' (in half of the trials) or between positions C' and D' (in the remaining trials) of the screen, while the head position was held fixed. In this way a dissociation between the visual and somatic meridians was achieved in that they were not coincident as in the first experiment. In this case on invalid trials subjects had to shift their attention from the cued position to the target position. In doing so, they could cross the visual vertical meridian (from A to B or viceversa) or the somatic vertical meridian (from B to C or viceversa) or none (from C to D or viceversa).

In both experiments mean reaction times to valid, neutral and invalid trials were analyzed by a one-way Anova design. Mean reaction times to invalid trials were further analyzed according to the crossing of the vertical meridian (first experiment) or to the crossing of visual and head centered vertical meridians (second experiment).

Results

Analyses performed on the data from the first experiment showed a significant effect ($F_{2,28}$=4.26, p=.02) of the condition (valid vs neutral vs invalid trials). Mean reaction times to valid trials were significantly faster than reaction times to invalid trials (178.71 msec vs 187.76 msec). This result was expected and confirms the validity of the experimental design. Furthermore, reaction times to invalid trials where subjects had to move their attention across the vertical meridian were slower then reaction times to invalid trials where subjects did not need to cross the vertical meridian ($F_{2,28}$=5.63, p=.008). The larger cost needed to move attention across the vertical meridian was small (about 12 msec) but significant. While the meridian effect in the visual space is well established in literature, these results confirm the existence of a meridian effect also in the acoustic space.

Analyses performed on the data from the second experiment showed again a significant effect ($F_{2,28}$=4.23, p=.02) of the condition (valid vs neutral vs invalid trials). Mean reaction times to valid trials were significantly faster than reaction times to invalid trials (105.79 vs 116.96). Analysis on reaction times to invalid trials showed a significant difference among uncrossed, visual crossed and somatic crossed conditions ($F_{2,28}$=3.44, p=.04). Reaction times to invalid trials where subjects had to move their attention across the somatic vertical meridian were significantly slower than reaction times in the other two conditions (122.01 vs 115.88 vs 109.23).

This result seems to suggest that in the acoustic space the anysotropy is determined by a natural boundary determined by the head position

References

□1□Benussi, V. (1907). Zur experimentelle Analyse der Zeitvergleichs. I: Zeitgroesse und Betonungsgestalt. *Archiv. Fur die gesamte Psychologie*. 9, 366-449. cit. in [5].

□2□Exner, S. (1875). Untersuchung ueber die einfachsten psychischen Prozesse. *Pfluger Archiv fur die gesamte Physiologie*.11. cit. in [5].

□3□Marcel, A.J. (1983). Conscious and unconscious perception: An approch to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*. 15(2), 238-300.

□4□Posner, M. & Dehaene, S. (1994). Attentional networks. *Trend in Neurosciences*. 17(2), 75-79.

□5□Rizzolatti, G., Riggio, L., Sheliga, B.M. (1994) Space and selective attention. In: C. Umilta' and M. Moscovitch (Eds.) *Attention and Performance XV*. Cambridge: MIT Press.

□6□Vicario, G (1988). *Tempo della fisica e tempo della psicologia*. in Vicario G. & Zambianchi E. (1998). La percezione degli eventi. Ricerche di psicologia sperimentale. Milano: Guerini Studio, 69-99.

□7□Wundt, W. (1983). *Grundzuege der physiologischen Psychologie*. Engelmann, Leipzig (IV ed.), cit. in [5].

A Learn-Based Environment for Melody Completion

Dominik Hörnel
dominik@ira.uka.de

Karin Höthker
hoethker@ira.uka.de

Institut für Logik, Komplexität und Deduktionssysteme
Universität Karlsruhe (TH)
Am Fasanengarten 5
D-76128 Karlsruhe, Germany

Abstract

We present an experimental environment for the investigation of melody space by learning and completion. Given any melodic fragment, e.g. the beginning, a melody evolves through application of musical knowledge learned by feedforward neural networks and nested Markov models from harmonic and motivic structure of the training examples. The system MELOGENET *illustrates our approach on folk melodies, and shows how it can be also used for melodic style recognition.*

1 Introduction

The investigation of melodic structure presents a challenging task. On the one hand, one can observe an astonishing consensus of humans assessing the quality of melodies in a certain style. On the other hand, there exist many schoolbooks about music harmonization and counterpoint, but few theories on how to write good melodies.

Existing theories fall into two categories: The top-down approach [2] assumes a hierarchical structure of the melody which is subsequently filled in with musical material. It can be realized using formal grammars [3]. The prediction approach [4] concentrates on musical expectancies for events following particular familiar sequences. Neural networks have proved useful to model this kind of structure ([6], [1]).

The main problem is to determine an appropriate assessment function for melodies taking into account the complex interaction of musical parameters like pitch, harmony and phrase structure. To automate the process, it is useful to learn this function from a given set of melodies that belong to a specific culture or style like traditional folk-songs, or from melodies invented by a specific composer. This is the idea followed in MELOGENET. The system analyzes a given set of training examples and retrieves musical information at various levels of abstraction. Then a set of neural networks is trained independently concentrating on specific aspects of melody. We introduce a nested Markov model to produce coherent global structure. The neural experts are put together to determine a fitness function which evaluates the quality of the melody by comparing it to the networks' predictions. In our system, the problem of finding good melodies is regarded as a fitness optimization problem and solved by an evolutionary algorithm.

2 Task Definition

To start with, we consider a subclass of rather simple melodies which we call *uniform* melodies. A melody is *uniform* if it can be divided "in a natural way" into segments of equal duration. Most folk-songs and classical themes fall into that class, whereas, for example, a Wagner melody does not. In our experiments, we consider diatonic melodies (no accidentals) in a two-four meter smoothed to eighth notes as smallest rhythmic units. The pitch range is from G below middle C to the C one octave higher, rests are ignored. All melodies are transposed to C major.

Given an arbitrary fragment, the task is to find a suitable completion in the style of the training examples. In our environment, one can also specify higher-level structure like harmonies or motif structure which may also conflict with each others.

3 Analysis and Learning of Melodic Structure

The analysis and selection of essential features for training the experts is the most important component of the system. We distinguish four types of melodic elements at various levels of abstraction: The motif component to capture typical melodic contours, the abstract motif sequence which determines the arrangement of those motifs, the harmony implied by the pitch sequence, and the pitch/interval level. These elements highly depend on each others. Therefore the task of the networks is to learn significant relationships from the training examples. Figure 1 displays the analytical and functional dependendies considered in our system. Boxes represent the musical entities at three different time scales, ellipses illustrate the relationships between them.

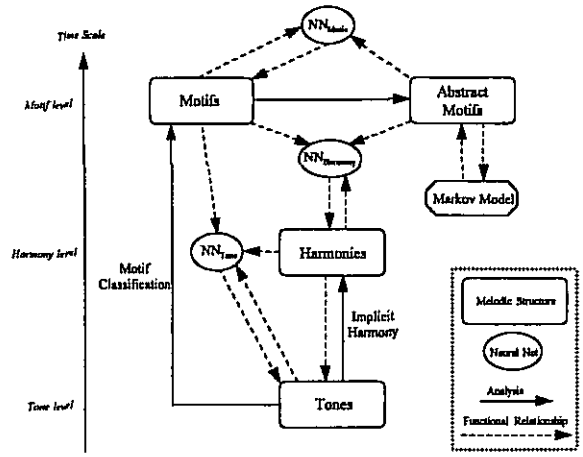


Figure 1: Analysis and Functional Relationships in MELOGENET

3.1 Motif Classification

The investigation of higher-level organized elements like motifs is necessary to capture essential melodic structure. The idea is to determine a segmentation of a piece of music and to classify those segments according to their similarity. In MELOGENET we use a hierarchical agglomerative clustering algorithm which was applied to the learning of baroque style melodic variations in MELONET [1]. The uniformity assumption allows us to consider motifs at a fixed time scale, in our case one measure per motif. As a result of motif clustering, we determine a sequence of motif classes for each melody. One neural network is then trained to learn the arrangement of these motif classes.

3.2 Abstract Motif Sequence

The abstract motif sequence represents the motivic structure of a melody abstracting from current motif classes. We propose a metric which measures common structure of two abstract motif sequences and allows to assess the quality of a sequence by computing the minimum distance between the sequence and a reference set.

Lemma: Given a finite set S,

$$d : \begin{cases} \wp(S) \times \wp(S) & \longrightarrow & [0,1) \\ (A,B) & \longmapsto & \frac{|A \triangle B|}{1+|A \cup B|} \end{cases} \qquad (1)$$

defines a metric on the power set $\wp(S)$.[1]

Proof: set calculus.

For an abstract motif sequence define an undirected graph by associating a vertex to each position and joining distinct vertices iff they represent equal motifs. Define the distance between abstract motif sequences M_1 and M_2 of equal length by $Distance_{AMS}(M_1, M_2) := d(E_1, E_2)$, where E_1 and E_2 denote the edge sets of the graphs associated to M_1 and M_2. Subsequences of abstract motif sequences may also be compared using the distance function, provided the sequences are restricted to the same motif positions.

In order to obtain suitable abstract motif sequences, we developed a mutation operator based on Markov processes. The idea is to delete abstract motifs from a sequence at random and determine a completion according to an appropriate training set of examples.

In folk-songs surprising events like new motifs alternate with familiar motivic elements in a balanced way. We assume that the appearance of new abstract motifs takes place on a flexible time scale. For example the analysis of uniform 49 children songs revealed that the occurrence of new abstract motifs could not be correlated with a periodic phrase structure.

This observation lead us to use nested linear first-order Markov processes. The model is a specialization of a Hidden Markov Model, where output symbols are produced by sub-processes rather than by constant probability distributions. A super-process operating on a flexible time-scale determines whether new motivic material is introduced, whereas sub-processes dispose current abstract motifs. Among all possible completions, a sequence is chosen which is generated by this model with maximum probability. The transition matrices are calculated using a reference subset of the training set which consists of sequences with minimum distance to the incomplete sequence. The distance is calculated on the subset of positions at which motifs are known.

We compared the Markov approach to other completion methods like local motif replacement based on the frequency of abstract motifs in the above reference subset and random completion with context. The hit rate for reconstructing abstract motif sequences was always best for the Markov model, independent of the degree of incompleteness.

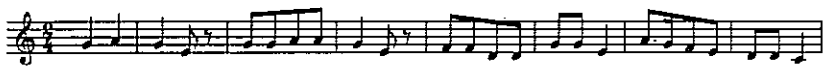
3.3 Implicit Harmony

The harmonic development of a melody is an important means to create a process of musical tension and relaxation. In the melodies considered here, we have a quarter note harmonic rhythm. Since we are considering diatonic melodies, each note can be harmonized by one or two of the basic harmonies *tonic*, *dominant* or *subdominant*. We use a simple heuristic to attribute a harmonic weight to each note. Summing up the harmonic weights, we obtain a winning harmony for each quarter. Another network is trained to predict the harmonic development within melodies.

3.4 Pitch Representation

Two networks are used for pitch prediction. For representing pitch, we use a special *complementary interval coding* developed in MELONET [1] which integrates musical knowledge about interval and har-

[1] $\triangle$ denotes the symmetrical difference of two sets.



Pitch Sequence	G	.	A	.	G	.	E	.	G	G	A	A	G	.	E	.	F	F	D	D
Harmony Sequence	T		S		T		T		T		S		T		T		D		D	
Motif Sequence	4				0				4				0				0			
Abstract Motif Sequence	a				b				a				b				b			

Pitch Sequence	G	G	E	.	A	.	F	E	D	D	C	.							
Harmony Sequence	T		T		S		D		T		T								
Motif Sequence	0				1				0										
Abstract Motif Sequence	b				c				b										

Figure 2: Analysis of German children song "Kreis Kreis Kessel"

monic relationships. The first tones of each motif, the so-called *reference tones*, are learned seperately to obtain coherence at a larger time scale. The remaining tones are predicted by another network according to current motif class, harmony, previous as well as next tone and the position within the motif. The functional dependencies to be learned by this tone network can be written as

$$Tone_t = f_{Tone}(Motif_{t/4}, Harmony_{t/2}, Tone_{t-1}, Tone_{t+1}, t \bmod 4) \quad (2)$$

Notice that we also base our decision on right context because this information is always available. In principle, arbitrary dependencies may be specified for the networks. This is not possible in context free grammars or sequential neural network composition.

In Figure 2 a children song and its analysis by the system are displayed.

4 Melody Completion with Evolutionary Algorithms

Genetic or evolutionary algorithms (EA) are widely used as an optimization technique in complex systems. They have also been applied to several musical tasks, e.g. for computer-assisted music composition.

```
pre_evolution();
repeat
  selection();
  mutation();
  optimization();
  evaluation();
  survival();
until stop_evolution();
```

Figure 3: Basic Evolutionary Algorithm

Figure 3 shows the basic structure of the EA. In the pre-evolution phase, the given melody fragment is copied to the initial population. The remaining gaps are initialized at random. In the evolutionary loop, a fixed number of offsprings is generated by first selecting their parents preferring the fitter ones, and then applying a mutation or crossover operator at various levels of abstraction. During local optimization, the melody notes of the offsprings are adapted to their context in a sequential manner. This is done in order to reduce the search space. Without local optimization, convergence is much slower and the results are weaker. The offsprings are evaluated according to the fitness function. Several parents might then be replaced by fitter offsprings. The loop is repeated until a user-defined stop criterion is fulfilled.

The most important component of an EA is the fitness function according to which the optimization is performed. In MELOGENET, the fitness is evaluated by the following formula:

$$Fitness = 1 - \sum_c \alpha_c Distance_c \quad (3)$$

where c indicates a melodic structure category like abstract motif, motif, harmony or pitch according to which the fitness is evaluated. The hyperparameters α_c assign an individual weight to each category and sum up to 1. For the categories motif, harmony and pitch the distance is defined as

$$Distance_c = \beta_c \sum_{t_c} \| C(t_c) - NN_c(t_c) \| \quad (4)$$

This term describes the euclidean distance between value C of category c and the prediction of the corresponding network NN_c summed up over time t_c. β_c is a normalization factor which guarantees that the distance lies between 0 and 1. The index c in t_c is used to denote the time scale at which c is defined (one measure per motif, one quarter note per harmony and one eighth note per pitch). Since the networks solve classification problems, after training their output should approximate the posterior probability of $C(t_c)$ given the input parameters $x(t_c)$:

$$NN_c(t_c) \approx P(C(t_c) \mid x(t_c)) \quad (5)$$

For the abstract motif level, $Distance_c$ is set to $Distance_{AMS}$ as defined in 3.2.

As an example consider the first harmony of a children song style melody in C major. The posterior probability $NN_{Harmony}(0)$ for *tonic* is 1.0, because each children song in our training set starts with C, E or G. Then the harmony distance $Distance_{Harmony}(0)$ reaches its minimum iff $Harmony(0) = NN_{Harmony}(0) = tonic$.

5 Experiments and Results

In our experimental environment, a set of melodies is analyzed as described in section 3, and

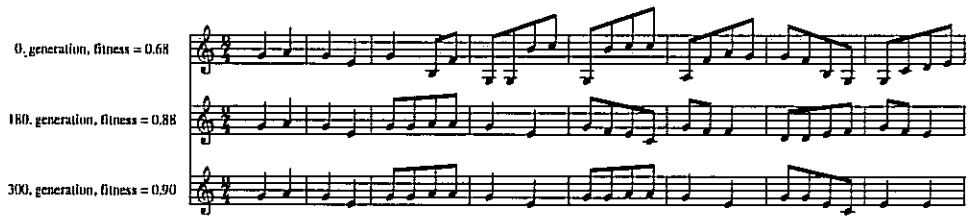


Figure 4: Evolved melody after 0, 180 and 300 generations

pattern files are generated for training, validation and testing of the corresponding networks. The networks are trained independently with the RPROP algorithm. Then a new melody fragment evolves until a given fitness value respectively a fixed number of generations is reached.

In our first experiment, the system was trained on 32 German children songs of the Essen folk-song data collection [5]. The melodies fulfill the uniformity condition and are eight to ten measures long. Some of them contain sixteenth notes and were therefore smoothed to eighth note resolution. Figure 4 shows the evolution of the melody depicted in Figure2 after 0, 180 and 300 generations, given the first two measures. The population size was 10 parents and 5 offsprings each generation using the Markov model for mutating the abstract motif sequence and random mutation for the remaining melodic structure categories. The weight parameters α_i were set as follows: 0.4 for tone and 0.2 for the other three categories. The musical impression resulting from several runs revealed that the melody tends to improve until a certain fitness threshold is reached. When going further, the evolutionary process converges towards smooth but less creative melodies.

Table 1: Recognition Rates for Children and Shanxi Song Style

	Recognized	Not Recognized
Children Song Test Set	91.3%	8.7%
Shanxi Song Test Set	87.2%	12.8%

Based on the homogeneous listening impression of folk-songs, we wanted to find out in the second experiment whether the fitness function (3) can be used to distinguish between melodies of different style. We trained the same networks on a set of 41 Chinese folksongs (Shanxi style) from the Essen database. Then we computed the children song fitness and the shanxi song fitness on independent test sets of the two styles (23 children and 39 shanxi songs). Each song was assigned to the style having the bigger fitness. Table 1 shows the recognition rates using this procedure.

6 Conclusions

We have dealt with relatively simple melodies until now and believe that a learn-based environment in combination with evolutionary algorithms opens a promising horizon towards the investigation of larger melody space. Future research will consider more complex melodies such as classical themes. We will also investigate how higher-level structure might be represented in melodies that do not satisfy the uniformity condition.

We have not yet developed a method for fitting the α_i parameters which weight the influence of the components on the melody fitness. One criterion for choosing them is the style recognition property examined in the second experiment. One could think of a meta-algorithm that determines the α_i in a way that maximizes the recognition rate.

The rhythm representation in MELOGENET seems to be inadequate for melodies of higher complexity. A hierarchical rhythm representation, like the rhythm grammar proposed in [3], will be more appropriate. Another interesting task is to examine the influence of the folk-song texts on melodic and rhythmic structure.

References

[1] D. Hörnel. "MELONET I: Neural Nets for Inventing Baroque-Style Chorale Variations", *Advances in Neural Information Processing 10 (NIPS 10)*, M.I. Jordan, M.J. Kearns, S.A. Solla (eds.), MIT Press, pp. 887-893, 1997.

[2] F. Lerdahl, R. Jackendoff. *A Generative Theory of Tonal Music*, Cambridge, Mass., MIT Press, 1983.

[3] D. Lidov, J. Gabura. "A melody writing algorithm using a formal language model", *Computer Studies in the Humanities 4(3-4)*, pp. 138-148, 1973.

[4] E. Narmour. *The Analysis and Cognition of Basic Melodic Structures*, University of Chicago Press, 1990.

[5] H. Schaffrath, B. Jesser. *Erfassungsregeln zur Melodiedatenbank*, Universität GH Essen, Handbuch, 1989.

[6] P. M. Todd. "A Connectionist Approach to Algorithmic Composition", *Music and Connectionism*, P. Todd and G. Loy (eds.), pp. 173-194, MIT Press, 1991.

FlExPat : A Novel Algorithm For Musical Pattern Discovery

Pierre-Yves ROLLAND
Laboratoire d'Informatique de Paris 6 (LIP6)
CNRS UMR 7606 Université Pierre et Marie Curie
4 place Jussieu, 75005 Paris, France
Pierre-Yves.Rolland@lip6.fr

Abstract: *Pattern discovery (or 'extraction') in sequences is a very general problem with diverse musical applications ranging from music analysis to music generating systems. In this paper we focus on automated discovery of patterns in corpuses of* melodic *sequences. A melodic pattern is defined by a set of either identical or 'equipollent' (i.e. significantly similar) sequence segments. In previous work and articles, we addressed such critical issues in musical pattern discovery as the* representation *of sequences and of their elements, and the definition of appropriate similarity metrics between (pairs of) sequence segments. Now we present a novel pattern extraction algorithm named FlExPat ('FlExible Extraction of Patterns'), which builds upon the concepts and techniques we previously introduced. FlExPat articulates in two phases,* factor matching *then* categorization, *with a theoretical worst-case complexity quadratic in the corpus' total sequence length. FlExPat has been implemented in our Imprology software system. Experimental results, a few of which are detailed here, clearly evidence FlExPat's qualities and performances.*

1. Introduction

Musical pattern discovery, and more specifically the *extraction* of melodic or harmonic patterns in given sets of composed or improvised works, is an important problem with many different applications ranging from musical analysis to music generating systems (see [10] for a review). For instance, in a voluminous study [5], musicologist T. Owens characterized the techniques and style of Charlie Parker through the nature and organization of recurrent melodic patterns ('formulae') in the jazz saxophone player's improvised playing. He extracted a lexicon of 193 such patterns from a corpus of about 250 Parker solos. We have used Owens' corpus and lexicon as validation material for our pattern extraction system.

2. Formalization

Melodic pattern extraction is a particular instance of *sequential data mining*. We represent melodies as sequences over generalized alphabets. We extended the notion of an alphabet, which is a finite sets of *symbols*, to that of a generalized alphabet — a finite set of entities of any nature, viz. structured representations of musical notes. Intuitively, a pattern is a set of things that display significant resemblance according to some similarity definition. Given a corpus, which is an ordered set of melodic sequences, we define a *melodic pattern* as a set, also called *block*, of passages — not necessarily phrases, in the closure-related sense. Each passage corresponds to a particular sequence *factor* (i.e. contiguous segment), and the 'significant resemblance' relation is dubbed *equipollence*. A pattern can be represented either (1) *intensionnaly*, through a *prototype* (melodic passage) or (2) *extensionnaly*, through its set of constitutive passages which are called *occurrences* of the pattern. Pattern discovery (or *extraction*) consists in finding all patterns in a given corpus, possibly imposing additional constraints such as a *quorum* threshold fixing the minimum number of different sequences in the corpus in which each pattern must appear (prototype or occurrences). So far, we have focused on monodic material, although a generalization to polyphonies or to harmonic (chord) sequences is entirely possible.

Equipollence is defined through a threshold on numerical similarity values. To test equipollence between two passages, they are *compared* using a numerical similarity assessment algorithm; if their similarity value is above the threshold, they are determined equipollent. Based on extensive experimentation on musical sequences [10], we got empirical confirmation that proper melodic passage comparison requires taking simultaneously into account multiple descriptions of melodies and of their individual notes. We have proposed, and implemented in our Imprology system (see below), the insertion of an automated representation enrichment — or change — phase at the begining of the pattern discovery process. Based on music perception and cognition work, user-choosable descriptions range from individual (pitch, duration, metrics, intervals...) to local/structural (characteristic contours, jazz-type chord shapes or harmonic cells, arpeggios...) to global (time signature, overall tonality...) [13].

3. FlExPat's Overall Features

FlExPat (Flexible extraction of patterns) is a general, combinatorial, algorithm[1] for extracting sequential patterns from sequences of data. FlExPat is structured in two main algorithmic phases (see Figure 1):

[1] To be totally accurate, we should speak of an algorithmic *family* as, for instance, extraction algorithms we designed take slightly different forms depending on whether the corpus is made of *one* or *several* sequences.

✧ The *factor matching* phase identifies in a computationally-economic fashion all equipollent factor couples. A graph, which we name *similarity graph*, is produced whose vertices each represents a distinct factor and edges each represents an equipollence relation between two factors. This graph is generally labeled with numbers, differentiating edges according to their respective similarity strength.

✧ The *categorization* phase extracts the actual patterns from the similarity graph. Among possible extraction paradigms, we implemented a method called "central star" which yields very good results despite a very satisfactory temporal complexity (see below).

Figure 1. FlExPat's overall algorithmic scheme

In terms of factor comparison model, FlExPat allows the use of our Multi-Description Valued Edit Model (MVEM). We already described the MVEM in detail elsewhere [10][13], so here we will just mention its key characteristics. The MEVM specializes the basic *edit model* of which the well-known *edit distance* is an instance. To compare two sequential entities, the optimal correspondence scheme between their respective elements (called *alignment*) is determined. Such an alignment is a series of *pairings*, each between individual [groups of] elements of each entity. This paradigm is particularly fit for comparing melodic sequences, as it neatly accounts for such important notions as ornamentation or variation. Any particular instance of the MVEM is characterized by the set of allowed pairing types (APTS), the standard set being {*Insertion, Deletion, Replacement*}. In a *valued* edit model, a *contribution function* is associated to each pairing type, so any pairing in an alignment gets a numerical evaluation reflecting its individual contribution to the overall similarity. The various melody/note descriptions are simultaneously taken into account in contribution functions using a weighted linear combination paradigm. The possibility, for the user, to dynamically choose the set of descriptions used and their respective weighs materializes the adoption of different *viewpoints* on melodic similarity and, hence, on the nature of extracted patterns. For instance, the user may at some point privilege temporal descriptions (durations, metrics, etc.) w.r.t. frequential descriptions (pitches, intervals, etc.) to achieve more rhythmically-oriented pattern discovery. We have also proposed and implemented methods for automatically optimizing description weighs [12].

After this overview of FlExPat, we will now present its key characteristics and properties whose detailed descriptions or proofs are given in [10].

4. FlExPat In More Detail

4.1. Input

FlExPat's main inputs are (1) a sequence corpus; (2) integers m_{min} and m_{max} controlling the minimum (resp maximum) length of pattern factors; (3) an integer function Δm controlling the maximum possible length difference between compared passages; (4) a fully specified instance of the MVEM: description set and weights, set of allowed pairing types APTS along with associated contribution functions; (5) a similarity threshold defining equipollence between passages.

4.2. Factor Matching Phase

First, the corpus' sequences are concatenated altogether in order, yielding a *global sequence* **S** of length **L**. After an initialization phase including the creation of an empty similarity graph, the space of all acceptable factor pairs is explored in a specific order, viz. increasing factor positions combined with increasing factor lengths. For each considered pair of passages, their similarity value is computed, and if it is above the threshold an edge is created between the corresponding two vertices, denoting equipollence. The core property in FlExPat's factor matching algorithm is the recurrence relation shown in Equation 1 (for clarity, we assume that the set of allowed pairing types is the standard set). It expresses the similarity $Sim_{i'\ m'}^{i\ m}$ between factors F and F' — of respective positions i and i' in the corpus (i.e. in global sequence S) and respective lengths m and m' — as a function of the similarity values between three other factor pairs. Each factor in those pairs is either F or F', or *prefixes* of F or F' of length m-1 (resp. m'-1), i.e. the results of the removal of the last element out of F or F'. Because of the specific exploration order of the factor pair space, it is guaranteed that all three similarity values have already been computed (and stored). Thus computing $Sim_{i'\ m'}^{i\ m}$ takes constant time (which is at most proportional to #APTS), as opposed to the normal O(m.m'.#APTS) time required for computing $Sim_{i'\ m'}^{i\ m}$ using efficient sequence comparison algorithms (e.g. [4]). This recurrence property allows the use of specific Dynamic Programming techniques, and a dramatic enhancement in algorithmic efficiency is achieved (see subsection 4.4).

$$Sim_{i'\ m'}^{i\ m} = \max\begin{cases} contrib\begin{pmatrix} S[i+m-1] \\ S[i'+m'-1] \end{pmatrix} + Sim_{i'\ m'-1}^{i\ m-1} \\ contrib\begin{pmatrix} S[i+m-1] \\ - \end{pmatrix} + Sim_{i'\ m'}^{i\ m-1} \\ contrib\begin{pmatrix} - \\ S[i'+m'-1] \end{pmatrix} + Sim_{i'\ m'-1}^{i\ m} \end{cases}$$

Equation 1

4.3. Categorization Phase

Among possible approaches for extracting patterns from the similarity graph, we have proposed the Star Center algorithm, which is influenced by an approach proposed in molecular biology [3]. Figure 2 shows an example of such a star, with numbers (edge values) each denoting a prototype-occurrence similarity value. The extracted pattern's prototype "is" the star center while its occurrences are the star's branches' ends. The corresponding algorithm can be sketched as follows:

1. For each vertex v in the similarity graph, compute:

$$totalValuation(v) = \sum_{v' \in adj(v)} value(v, v')$$

2. Sort the set of stars by decreasing totalValuation.

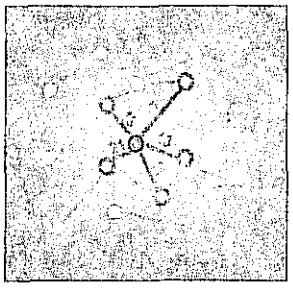


Figure 2. Example « Star » in the Similarity Graph

This yields a list of patterns (stars) ordered in decreasing prominence, the prominence of a pattern being naturally defined as its corresponding totalValuation. We have proposed various online/offline filtering approaches for fitting the above-mentioned possible additional pattern constraints. Specific advantages of the Star Center approach are the following: (1) it yields *de facto* a prototype for every pattern (the factor corresponding to the Star's central vertex); (2) a quantitative prominence evaluation of every pattern is obtained; and (3) its algorithmic complexity is satisfactory. In particular, its time complexity is proportional to the number of edges in the similarity graph, which is (loosely) bound by $L^2 \times (m_{max} - m_{min}+1)^2$. This is to be compared to other possible approaches for extracting pattern from the graph; for instance, finding [maximal] cliques is known to be an NP-complete problem [1].

4.4. Algorithmic Complexity

For FlExPat's two phases' cumulated complexity, the following (loose) upper bounds can be given. **Time complexity**: $O(L^2 \times mmax^2 \times ||APTS||)$, where ||APTS|| denotes the *effective size* of the allowed pairing type set — for instance, this parameter is 3 for the standard pairing set. **Space complexity**: $O(L^2 \times mmax^2)$, as the dominant term is the memory required for storing the similarity graph.

In practice, m_{min} and m_{max} are always set to small, constant values compared to L, so both complexity bounds merely rewrite to $\mathbf{O(L^2)}$.

5. Implementation and experimental results

Our Improlology system [12][10] is implemented in the Smalltalk-80 (ParcPlace VisualWorks) object-oriented language. It reuses, adapts and extends the MusES collection of Smalltalk *classes* and *methods* [6] for representing and manipulating basic tonal music concepts. A few example results will now be shown. The considered corpus, referred to as C1, is made of 10 Parker solos in the 'C major - Blues' category of Owens' corpus: 3 takes on "Cool Blues", 3 takes on "Relaxin' at Camarillo" and 4 takes on "Perhaps". C1's cumulative length is 2000. Among FlExPat's parameter values are the following: m_{min}=3, m_{max}=27, and the APTS is the standard set. Figure 3 shows the prototype of an example pattern extracted by Improlology/FlExPat, which we will refer to as Pat1; the next three figures show its 3 occurrences throughout corpus C1, ordered by decreasing computed similarity with the prototype. Respective similarity values are 213, 171.5 and 171. Each passage is delimited visually by the dark bar on top of each staff and with the usual <bar No.>:<beat No.> notation in captions.

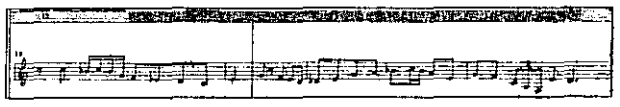


Figure 3. Prototype of pattern Pat1: 19:3-23:1 passage of *Relaxin' at Camarillo Take 4* solo

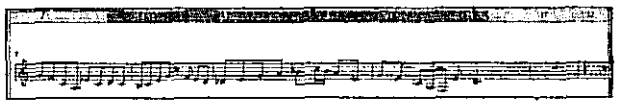


Figure 4. First occurrence of Pat1: 7:4-11:1 passage of *Relaxin' at Camarillo Take 4* solo

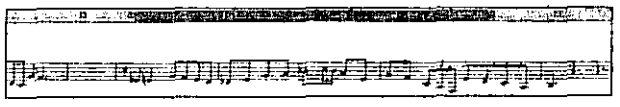


Figure 5. Second occurrence of Pat1: 32:1.75-35:1 passage of *Perhaps Take3* solo

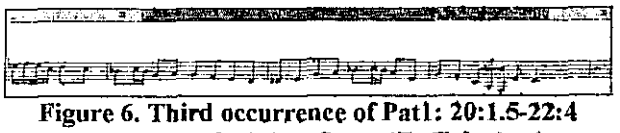


Figure 6. Third occurrence of Pat1: 20:1.5-22:4 passage of *Relaxin' at Camarillo Take 3* solo

Of course, this is no more than one particular example of pattern, clearly making musical sense, among the dozens extracted by Improlology/FlExPat from such a corpus. However, it does illustrate several key points which will be discussed in section 6. Particularly, Pat1 was not signaled by Owens, thus being effectively 'discovered' by the system.

Among extracted patterns are also a number of patterns that are part of Owens' lexicon. Figure 7 shows three such patterns of various lengths and nature, for instance the second one is a mere 8th-note chromatic descent.

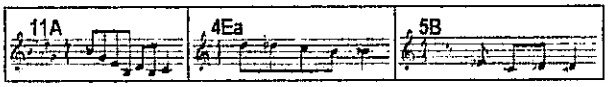


Figure 7. Examples of Owens' lexicon patterns which were extracted by Improlology/FIExPat from C1

6. Discussion and related work

Overall, based on the whole set of results we obtained with different [sub-]corpora and different description and description weighting sets, we have made the following key observations :

- A large number of Owens' patterns are satisfactorily extracted by the system, which conforms to our validation objective. Even more interestingly, a number of new and musically meaningful patterns, such as Pat1 above, are also 'discovered' by the system.
- Significantly different sets of melody/note descriptions or description weights — i.e. different viewpoints w.r.t. musical similarity — lead to the extraction of significantly different patterns.
- Extracted patterns are very often specific to particular melodies (sequences) of the corpus, or to related groups of melodies. For instance, the best occurrence of Pat1 above is *internal* (same tune and take as the prototype), and 3 out of the 4 passages belong to the same tune (*Relaxin' at Camarillo*). This opens a clear perspective in automated pattern-based identification of style or author.

The systems whose pattern extraction paradigm are most similar to ours are Cope's EMI [2] and Pennycook et al.'s system [7]. Due to its real-time functioning capability, the latter skirts around combinatorial aspects inherent to sequential pattern extraction, using an online phrase segmentation algorithm on which it strongly relies. EMI's pattern extractor's main limitations lie in its major combinatorial and algorithmic redundancy and in its imposed equipollence criteria (Hamming distance). To our knowledge no other general pattern discovery algorithm than FIExPat allows the use of an edit model with non-constant contribution functions and user- or system-ajustable descriptions set and weigths.

7. Future work

As musical pattern discovery is a very general problem, a number of perspectives to this work can be drawn, some of which were touched upon in this paper. A major one is the merger of Improlology and the *ImPact* jazz bass accompaniment system [8], a work we already-started [9]. That fusion allows that both the initial constitution and the incremental enrichment of ImPact's Musical Memory — the system's main source of generative material — now be automatically undertaken by Improlology, as opposed to manually (user).

References

[1] Aho, A.V., Hopcroft, J.E., Ullman, J.D. The Design and Analysis of Computer Algorithms. Addison-Wesley, Reading, MA. 1974.

[2] Cope, D. 1991. Computers and Musical Style. Oxford: Oxford University Press.

[3] Gusfield, D. 1993. Efficient Methods for Multiple Sequence Alignment with Guaranteed Error Bounds. ***Bull. Math. Biol.***, 55:141-154.

[4] Needleman, S., Wunsch, C.D. 1970. A general Method Applicable to the Search for Similarities in the Amino-Acid Sequence of Two Proteins. J. Mol. Bio. 48:443-453

[5] Owens, T. 1974. Charlie Parker: Techniques of Improvisation. Ph.D. Thesis, Dept. of Music, University of California at Los Angeles (UCLA).

[6] Pachet, F. 1994. The MusES system: an environment for experimenting with knowledge representation techniques in tonal harmony. In First Brazilian Symposium on Computer Music - SBC&M '94, pp. 195-201.

[7] Pennycook, B., D.R. Stammen, & D. Reynolds 1993 Toward a computer model of a jazz improviser. In *Proceedings of the 1993 International Computer Music Conference*, pp. 228-231.

[8] Ramalho, G., & Ganascia, J.-G. 1994. Simulating Creativity in Jazz Performance. In Twelfth National Conference on Artificial Intelligence, pp. 108-13. AAAI Press.

[9] Ramalho, G., Rolland, P.Y., and Ganascia, J.G., 1998. An Artificially Intelligent Jazz Performer. ***Journal of New Music Research*** (to appear).

[10] Rolland, P.Y. 1998. Découverte Automatique de Régularités dans les Séquences et Application à l'Analyse Musicale. Thèse de Doctorat en Informatique de l'Université Paris VI. July 1998.

[11] Rolland, P.Y., Ganascia, J.G. 1998. Musical Pattern Extraction and Similarity Assessment. ***Contemporary Music Review***. (to appear).

[12] Rolland, P.Y., Ganascia, J.G. 1996. Automated Motive Oriented Analysis of Musical Corpuses: a Jazz Case Study. In Proceedings of the 20th International Computer Music Conference, (ICMC'96, Hong Kong) pp. 240-243.

[13] Rolland, P.Y., Ganascia, J.G. 1996. Automated Identification of Prominent Motives in Jazz Solo Corpuses. In *Proceedings of the 4th International Conference on Music Perception and Cognition* (ICMPC'96, Montreal) pp. 491-495.

[14] Rowe, R. 1993. Interactive Music Systems. MIT Press.

Saturday 26th **h. 9.00**

MUSIC ARCHIVES

An evaluation about relations between musical, technical and perceptive environments in AFS project*

A.Borgonovo, A. Paccagnini, D. Rossi, D. Tanzi
E.mail: tanzi @dsi.unimi.it
Laboratorio di Informatica Musicale - Dipartimento di Scienze dell' Informazione
Università degli Studi di Milano - Via Comelico, 39
Milano (Italy)

Abstract

This contribution intends to point out a few discussion items appeared in the course of methodological set up in AFS (Archivio Fonico Scala) project, with regard to tape degrading determination and conformity of criteria carried in rescuing operations, besides some consideration in perspectives of restoring tapes and musicological media spaces.

Introduction

In the sphere of enhancement of cultural layers associated with La Scala Board activity, AFS project's goal is the rescuing of the phonic archive of La Scala theatre, through re-organization, digitization and preservation of the theater's musical works dating back 1951. Project has been divided on different moments, wich include different layers of evaluation, in particular regarding: a) musical wealth preservation state; b) rescuing methods; c) proceedings and intermediate phases; d) operative conditions; e) conformity criteria selection; f) results evaluation; g) data coherence planning and data storage. In AFS project, happening relactions between kinds of technical, perceptive and fruitional activities are constantly subjected to checks and settlements, in function of materials distinctive features and diversity of events whenever considered.

Methodological set up

The basic stages involved in AFS project can be represented as an analytical data collection, where the position of each of four subset area (columns) identify a kind of ordered succession, starting from key structural elements gathering the Repertory grids, and followed by Competence areas, evaluation areas, activity areas. Top-down direction indicate a kind of increasing complexity, in order to feed-back controls point of view.

Repertory grids	Competence areas	Evaluation areas	Activity areas
tapes inspection	audio, acoustic, electronic	defects correction	processing protocols
recording classification	elettroacoustic musical	signal location and de-masking	mapping and data extraction
verification transferring	elettroacustic, data processing	perceptive variances	correction, standardization
saving and masterization	audio, data processing	congruence editing	classification, hearing test
overall evaluation	musicological, elettroacoustic	conformity difference	hearing protocols
database project	linguistic data management	consistency accessibility	data extraction quering
aestehetic evaluation	musical, musicological	event destination artistic fruition	reinstatement and/or restoration
scientific and general direction	methodological set up	strategy, resources, skills, research	validation and goals planning

Figure 1: Basic stages in AFS project

* *This project has been partially supported by the Italian National Research Council in the frame of the "Metodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneus magnetic media" research, finalized project "Cultural Heritage" (subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1).*

This approach might be integrated with complementary perspectives, appaeared from surveys where skilled people needed to be faced with real situations (or situations became meaningful for them), often solving problems at first caused by imprecision of natural language or sensory devices, often when it was not clear how some particular rules might be understood and expressed. In a general kind of construction of problem-solving actions, where different roles and types of knowledge are involved [1], the AFS staff had to recognize how various requirements for knowledge modelling environments and heuristic classification was at the same time concerned. As in dynamic scheme below [2], varied kinds of experience play a different role in improving the performance of humans problem solving:

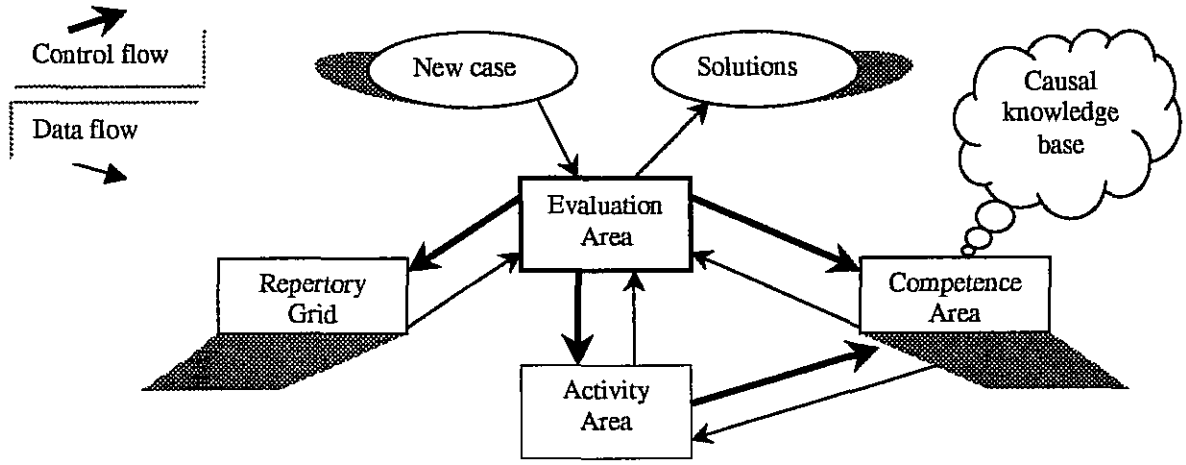


Figure 2: A problem solving scheme

Tape degrading determination

The analog tapes preservation state advised staff to drawing up a board of steadiness conditions. After an inquiry into main alterations and defects, a wide range of preservaton processing have been settled up to minimize every kind of information loss before digitalization phase. In fact, non-homogeneity in preservation state regarding many lots of tapes have binded the team to define, besides the distinctive features, further differential factors [3]. Following a strategy aimed to harmonize and document all the safeguard actions, the AFS staff deemed to had to isolate, for each problematic case, a core solving connected with the information about intervention proceedings qualified to state of every single tape.

Among the large case study regarding magnetic tape defects and damages [4,5], AFS staff often runned into the relative humidity storage conditions or tapes bad quality, wich caused a treacherous condition in binder integrity: as a consequence of mechanical sliding, tapes physical condition might be altered to the point that progressive deposit of magnetic particle on record heads may cause a progressive sound darkening with essential data losing:

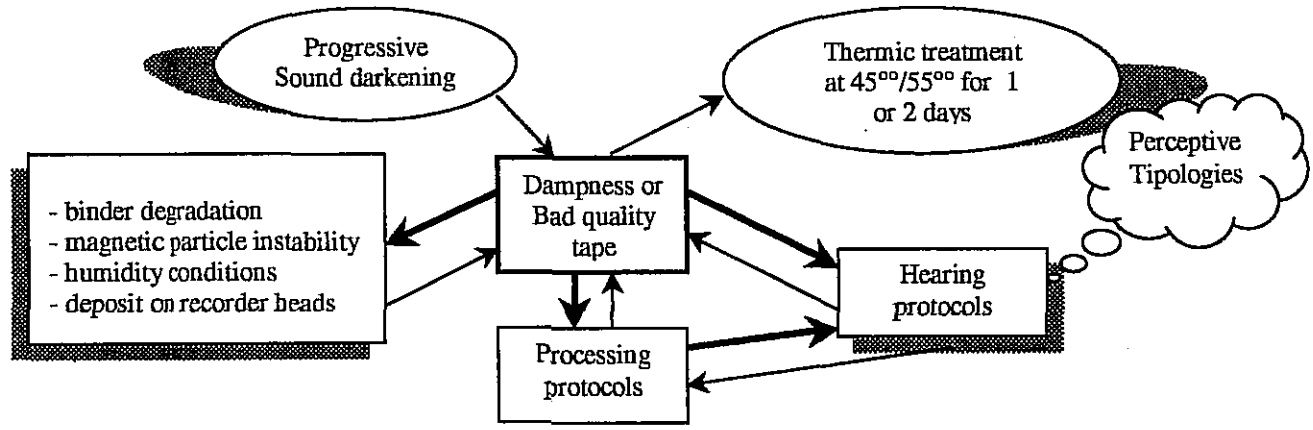


Figure 3: scheme in sound darkening correction

Conformity in digitization processing

To deal deep the topics in order to degrading notion, the competences in position to stimate as well storic/aestethic value events have been urged to express theirselves. Spreading of these competences towards whole staff allowed a careful specification of hearing protocols, carried during digitization, masterization and control phases. In particular, the staff had to examine performances during the analog sounds digitization processes, since it has been noticed that some of the outcomes might be caused by subjective bound variables during trasferring phase: indeed all the people involved in trasferring processes might give rise to some shifting from the standards. All the same, awareness of meaningful differences in personal fruition offered further specifications in order to digitization processing and hearing protocols set up, to the point that the AFS staff has become able to prefigure, recognize and audit incidental changes during passage from the old to new one support in the same event fruibility. Likewise, all rescuing project phases have been faced with awareness of data and procedural phormulas and exchange, wich allowed quality standards and technological comparisons method definition for various hardware testing. Besides, comparisons effected in order to recording techniques had been able to allow the necessary deepenings for valuation scale about recording quality, even if the partial incompleteness of technical data coming from the archive recording cards had to make news settlements necessary about important elements of signal features. Finally, next to musical wealth value verification and next to tape degrading and damage suffering determination all conditions [6,7] have been foreseen (and cautions prepared) to make AFS staff to be able to express documented suggestions for adequate restoring processes according to stated models of fruition.

Data source	Effects	First actions	Rescuing auditing	Restoring suggestion
Tapes defects	Digitization trouble	Pre-restoring tape	Tape quality scaling	Itemized treatments
Signals defects	Unbalanced signal	Equalizing signal	Spectra evaluation	Spectral editing
Sounds pick up	Spatial hearing disturb	Mapping aural source	Contextual analysis	Hearing optimization
Background noises	Masking signals	Mapping noises band	Noise power analysis	Denoising
Impulsive noises	Missing signals	Mapping clicks	Revealing tasks	Declicking

Figure 4: main restoring ideas

Conformity in audio data planning

Through all operating phases documentation and phonic auditory-perceptive suitable consideration, individual and collective team's capabilities growt, step after step, in direction of a strategy wich hold in due consideration the musical stereo-listening conditioning and recording tipologies, especially for music wrote for execution and hearing in accordance with non-amplified music common customs [8]. Dedication in describing relations between perceptive tipologies and musical destinations, represented a basic step from define the main fruition areas according with query strategies and online database project planning, wich potentially may enable extensive use and distribution of the recordings via Internet browser by means of a multiplatform and user-friendliness system that is needed for the network [9]. But AFS staff was already setting out for enrich contextual leading techniques in order to improve, in query mechanisms, the "perception-planning-carrying out" cycle [10], because the junction between artificial and natural perceptive elements play a critical role whenever users chose to perform pattern searches on the database even by own singing espression into a microphone linked to the computer, with the objective of bringing up all the similar pieces of music that are stored within the system. For that, next music information science, LIM's interests are turning in order to the artificial sensority, (even if held with natural sensority studies in music perception), according to determine an operative behaviour between present world and past musical treasures, also transferring experience acquired in musical analysis tools design, with *Intelligent Music Workstation* (IMW), a CNR PF 12 Project [11].

Musicological media spaces

Features of media spaces have significant implications for mediated perception and interaction: therefore, AFS designers having to define a set of patterns able to influence the system performance, because different responses of different users (musicians, researchers, teachers, students) are conditioned by unlike experience, learning histories or expectancies. AFS staff attempt to extend musicological media spaces to integrate richer forms of perception and interaction, in view of facilitate musical users transition into a new medium by supporting conventions already familiar to them, but also creating templates for new potential conventions that take advantage of the new medium's capabilities. In both cases, however, AFS designers even though knowing that musical conventions are existing by virtue of their enactment by a community of users, and knowing that innovations in design have to be aware of maningful changes that can happen in cognitive skills till modify motor-sensory reactivity [12]. Since musicological

practices are constituted of a complicated and mutating set, the musicological media space in design is also seen as mediating the relations between musical heritage and professional users, as being engaged in comparing esthetic trends. Since mediation role of artifact (sistem in design) could be criticized on grounds that mediators may have too partial knowledge to represent users and developers adequately to one another, every step of implementing tools had been toke by focusing task in musical, perceptive and technical relations so that no semplification may diminish the role of the retrieval work in rules, vocabulary, history and actuality in musicological media spaces. Thus, musical users may develop a representation of relational data not only supports inference mechanisms, but also supports acquisition of new knowledge [13] by being activated whenever corpus of musical heritage are reviewed and remapped in a complex information-processing domains by integration of low and high level data: from records quality to subsequent score revisions; from signal reconstruction tasks to conductors interpretations; from agogical analysis to picking up techniques consideration, as far as extracting MIDI file from a particular recorded piece.

References

[1] Haus, G. 1998. *Rescuing La Scala's Music Archives*, Computer, Volume 31, Number 3. IEEE Computer Society.

[2] Borgonovo, A. 1998. *Digitizing and Preservation of the Audio archive of Teatro Alla Scala di Milano*, CD-R Forum, Utrecht, Netherlands, CD-R Application session.

[3] Torasso, P., Portinale, L. 1997. *Combining Experiential Knowledge and Model-Based Reasoning for Diagnostic Problem Solving*, in: *Knowledge based Systems - Advanced Concepts, Techniques and Applications*, editor by S. G. Tzafestas, World Scientific Publishing, Singapore.

[4] Van Bogart, J.W.C. 1995. *Recovering of Damaged Magnetic Tape and Optical Disc Media*, National Media Laboratory, St. Paul, Mn.

[5] Van Bogart, J.W.C. 1995. *Media stability Studies Final Report*, National Media Laboratory, St. Paul, Mn.

[6] Niedzwiecki, M., Cisowski, K.1996. *Adaptive scheme for elimination of Broadband noise and impulsive disturbances from AR and ARMA signals*, IEEE Trans. On Automatic control, vol. AC – 24, n. 1.

[7] Rossi, D. 1995. *Attenuazione numerica dei disturbi nei segnali audio*, Atti dell XI Colloquio di Informatica Musicale, Bologna, AIMI - Conservatorio di Musica G. B. Martini - DAMS Università di Bologna, a cura di L. Finarelli e F. Regazzi.

[8] Paccagnini, A. 1992. *Tecnologie musicali elettroniche e nuovi modi di ascolto*, Scienza & Tecnica – Estratti Annuario della EST.

[9] Record, S. 1998. *Brava, La Scala!*, Oracle Magazine, May/June, http://www.oramag.com/

[10] Mangili, F., Musso, G. 1992. *La sensorialità delle macchine*, McGraw-Hill Libri Italia srl, Milano.

[11] AA.VV., 1994. *IMW CD: Intelligent Music Workstation (IMW)*, mixed mode CD-ROM (Macintosh HFS + CD-DA), IEEE Computer Society Press, Washington.

[12] AA.VV., 1994. *Commentary on Borderline Issues*, Human-Computer Interaction, Volume 9, pp. 37-135, Lawrence Erlbaum Associates, Inc.

[13] Harvey, L., Anderson, J.1996. *Transfer of declarative Knowledge in Complex Information-Processing Domains*, Human-Computer Interaction, Volume 11, pp. 69-96, Lawrence Erlbaum Associates, Inc.

Designing Music Objects for a Multimedia Database

Elena Ferrari Goffredo Haus
LIM - Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico 39/41, 20135 Milano (Italy)
email: {ferrarie,haus}@dsi.unimi.it

Abstract

In this paper we describe a large application project whose goal is the development of a multimedia database for the La Scala Theater. La Scala is one of the major theaters worldwide and has a huge amount of information on operas and musical performances. Most of those information are either on paper or stored on a variety of media, such as disks, tapes, or CD-ROMs. The goal of the project is to build a multimedia database, using the object-relational DBMS Oracle 8, making available all this information in digital form.

1 Introduction

The La Scala Theater is probably one of the best-known musical temples in the world. Built in 1776, La Scala is particularly famous for opera performances in that it has staged the openings of some of the most famous operas, including Bellini's Norma, Verdi's Otello, and Puccini's Turandot. Destroyed during World War II, the La Scala Theater was reopened in 1946 due to the effort of the great conductor Arturo Toscanini.

During La Scala life a huge amount of information on operas and musical performances have been collected. Most of those information are either on paper or stored on a variety of media, such as disks, tapes, or CD-ROMs. The availability of this information in digital form would be very useful for both external users and people working at La Scala Theater. For instance, when a new performance must be prepared, the musicians can easily access all the materials (such as CD-ROMs, video, photos, and scores) of previous editions of the same performance. On the other hand, external users can query part of the information stored in the database (for instance on the web), thus acquiring information on the La Scala activities.

This project has been partially supported by the Italian National Research Council in the frame of the "Methodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneous magnetic media" research, Finalized Project "Cultural Heritage" (Subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1).

In this paper we describe a project, currently under development at the Laboratorio di Informatica Musicale (LIM) of the University of Milano, having the ultimate goal of recording and organizing all the La Scala information into a multimedia database. This task is currently carried out as an incremental process involving several steps: the first step was the cleaning and digitizing of all the La Scala's tapes dating back to 1951 and the storing of such tapes on CD-ROMs [2]. This activity is necessary since most of the La Scala original tapes are rapidly deteriorating. The goal of the digitizing process is therefore to preserve the integrity of the music heritage of the La Scala Theatre.

The second task, which began in 1997 and is currently under development, is the recording of audio and music score materials into a database, known as the *Phonic Archive*. The Phonic Archive is based on the Oracle 8 object-relational technology [4] and it will make available nearly five decades of La Scala history to both internal and external users. Developing such a database poses several interesting challenges [1, 3]. First setting up the database schema is a complex activity in that, unlike conventional database environments, our environment is characterized by an extraordinarily large amount of music and multimedia objects stored in a variety of formats and characterized by tightly relationships.

Multimedia data are inherently different from conventional ones. The main difference is that information about the content of multimedia data is usually not encoded into attributes provided by the data schema (like traditional *structured data*). Rather, text, image, video, and audio data are typically *unstructured*. Therefore, specific methods to identify and represent content features and semantic structures of multimedia data are needed. Another distinguishing feature of multimedia data is their large storage requirements. One single image usually requires several Kbyte of storage, whereas a single second of video can require several Mbytes of storage. Moreover, the content of a multimedia data is difficult to analyze and compare, in order to be actively used during query

Figure 1: *Phonic archive's main window*

processing.

The peculiar characteristics of multimedia data have required the design of both ad-hoc data structures to efficiently store objects, such as audio or music scores' objects, and keep track of their relationships, and the development of data-entry tools and associated methods for populating these data structures. In the remainder of the paper we give a brief description of the most innovative features of our project.

2 The Phonic Archive

The Phonic Archive is a multiplatform (Unix, Windows NT and Macintosh) and distributed (10 workstations) database developed using the Oracle 8 object-relational technology. Our platform is also complemented with a CD-ROM juke-box able to contain up to 224 CD-ROMs which enable their fast online retrieval.

At the current stage, our prototype system, whose main window is illustrated in Figure 1, consists of two main applications. The first, is a data entry application by which the user can enter the information on the La Scala performances and on the associated audio materials, such as tapes or CD-ROMs. The second application is a visual query tool by which users can submit their queries to the underlying database and look at the query results. This tool provides an integrated support for both conventional alphanumeric queries and content based queries on the objects stored in the database. In the following, we describe both those applications.

2.1 Data entry

Our system is centered around la Scala nights. For each night we record more than 60 attributes. Such attributes are divided into the following five groups: *night attributes*, which store information on the La Scala nights, such as the date, the place where they were performed, or the type of the performance (e.g., opera, concert, recital and so on); *performance attributes*, keeping track of the cast of the performance, such as the singers of the performance, the conductor, or the players; *CD attributes*, storing information on the CDs of a given night, such as the number of tracks of the CD, the CD duration, and the CD content; *tape attributes*, storing information on the tapes corresponding to a given night and on their quality. Quality information are very important since they allow the retrieval of all the tapes and corresponding CDs with a particular quality level (for instance,this could be useful for merchandising purposes). Tape's quality is measured according to three different parameters (see Figure 2): *state of preservation*, that measures how the source tapes are preserved from the phisical viewpoint; *microphonic take*, that measures how many and where microphones were placed to pick up sound from the stage; *quality of recording*, that measures how source audio signals were mixed and processed during recording of the original tape.

the final group of attributes are the *digitalization process attributes*, storing information on the digitiz-

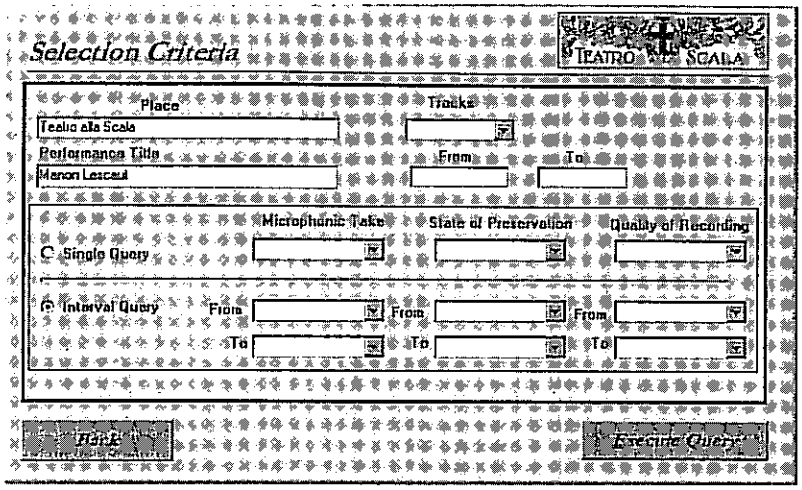


Figure 2: *Tape's window*

ing of La Scala tapes.

Besides traditional alphanumeric data, used to characterized La Scala performances and their associated audio materials, the Phonic Archive makes available up to 5 terabytes of audio and graphical data which are used to store both the audio materials of the La Scala nights and the associated music scores. At the current stage, the Phonic Archive contains the La Scala audio materials from 1951 up to 1978.

2.2 Queries

Our query tool is divided into three parts (see Figure 1). By pressing the *CD* button on the main window you can query the CDs stored into the Phonic Archive. Selection criteria for CDs include several parameters, such as the title of the CD, the name of the conductor, the names of the singers and of the players, or the title or author of a particular piece of music in the CD. Examples of queries you can submit are: "*Find all the CDs of the Nozze di Figaro sung by Placido Domingo and conducted by Riccardo Muti*", or "*Find all the CDs in which Maurizio Pollini plays*".

Another important feature of our query tool is that it allows the retrieval of the encore contained in the CDs. Such information is very important from a musicological point of view and it was missing in the original tapes of La Scala Theater. Once the CDs have been retrieved according to the submitted criteria, the user can browse the CD content or he/she can play it by issuing a command to the CD-ROM juke-box.

By pressing the *Tape* button from the Phonic Archive main window, the user activates the window shown in Figure 2 by which he/she can query the tapes recorded into the database. Queries can be both on the tape content (such as "*Find all the tapes recording the Manon Lescaut representations of 1978*) and on the tape quality.

Finally, the most innovative feature of our query tool can be activated by pressing the *Score* button from the main window. Our tool for querying music scores provides an integrated support for both standard and content-based queries. By means of this tool, a user can formulate both traditional queries such as: "*Retrieve all the scores written by Mozart*", or content based queries such as: "*Retrieve all the scores written by Mozart and containing a particular sequence of notes*". This latter functionality is particularly useful because often you remember the sound of a particular aria but not the title and/or the author. At the current stage, the query by content functionality is activated by submitting to the system an audio file containing the sample sequence of notes. However, we plan to extend our system to make the users able to directly sing a few bars into a microphone connected to the system and find all the similar pieces of music stores into the database.

Music scores returned by the query are then presented to the user along which their similarity degree with the input audio file. Then, the user can select a music score, views its graphical representation, or excerpts from it (see Figure 3), and simultaneously plays the music.

The matching between the input audio file and the scores stored into the database is based on a Pitch-Tracking method. We refer the interested reader to [5] for details on such method.

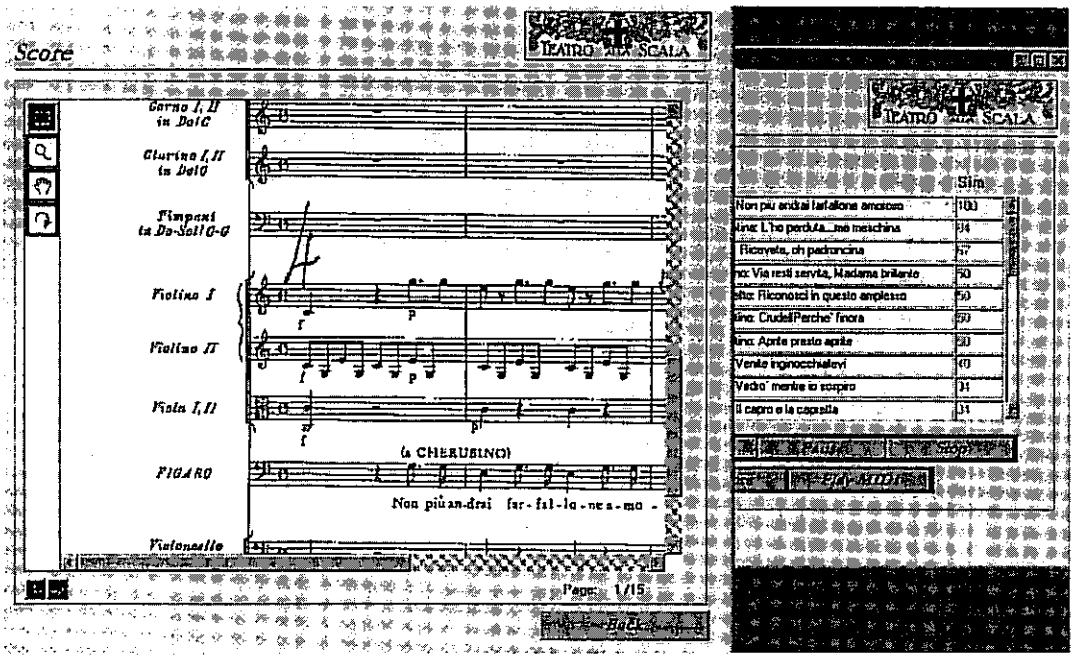


Figure 3: *The results of an audio query on music scores*

3 Conclusions and future work

In this paper we have presented a project, currently under development at the Laboratorio di Informatica Musicale of the University of Milano whose ultimate goal is the development of the multimedia database of the La Scala Theater. At the current stage our prototype system supports a number of innovative features including: digital-audio recording of operas and excerpts; graphical representations of music scores; representation of the scores in coded sound files; content-based queries on music scores.

The work presented in this paper is part of a larger on going project. Future work include the integration of additional musical media and information in the database, such as videos and photos of the performances, the multimedia representation of historical scenography and costumes, the development of more advanced tools for querying both audio and music score objects and the integration of such tools with the Oracle multimedia data cartridges.

Acknowledgments

The authors wishes to thank Luisella Rotolo and Laura Vaiani for their great contribution to the development of the Phonic Archive.

References

[1] E. Bertino, B. Catania, E. Ferrari. Multimedia IR: Models and Languages. In *Modern Information Retrieval*, Addison-Wesley, to appear.

[2] G. Haus. Rescuing La Scala's Music Archives. *IEEE Computer*, vol.31 n.3, pages 88–89, March 1998.

[3] S. Khoshafian and A.B. Baker. *Multimedia and Imaging Databases.* Morgan Kaufmann Publisher, Inc., San Francisco, CA, 1996.

[4] Oracle Corporation. *Oracle8 Server Concepts*, June 1997.

[5] E. Pollastri. Memory-Retrieval based on Pitch-Tracking and String-Matching Methods. In *Proc. of the 12th Colloquium on Musical Informatics*, Gorizia, Italy, September 1998.

Automatical acquisition of orchestral scores: the "Nozze di Figaro" experience*

Giuseppe Frazzini, Goffredo Haus
LIM, Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
via Comelico, 39
I-20135 Milano (Italy)
Fax +39 2 55006373
e-mail:giusfra@globalnet.it, haus@dsi.unimi.it

Abstract

From the cooperation between the LIM and the "Teatro alla Scala" in Milan, aiming at the computerization of the musical archive of the theatre with the project of a multimedia database for the organization of the whole archive material, it's emerged the question of the automatical and efficient acquisition of a huge number of orchestral scores.

In the period from February to June 1997 an important experimentation has been made through the acquisition of the whole orchestral score concerning "Le Nozze di Figaro" of Wolfgang Amadeus Mozart. On this occasion it was used the 1973 Barenreiter Kassel edition [1] in two different versions: the original score without changements, and the score modified with some annotations by the M° Riccardo Muti, on the occasion of the performance of "Le Nozze di Figaro" in May 1997, of the Teatro alla Scala.

This experimentation has been executed in close cooperation with the theatre archive musicians, so to have the opportunity to test and elaborate a way of working satisfying their needs.

1. The problem definition

The whole opera pratically consists of 600 B4 sized pages, is divided into 4 acts and a Symphony, or more precisely, into a Symphony, 10 Aria, 6 Duettini, 3 Cavatine, 2 Terzetti, 3 Cori, 1 Arietta, 3 Finali (of large dimensions), 1 Sestetto and 34 Recitativi.

In Figure 1 we can see a typical fragment of an orchestral score (incipit of the Aria 24 "L'ho perduta, me meschina").

Currently the staff of the musical archive effects every work of musical publishing trade using the commercial application Finale of the Coda Software [11] in its latest version (Finale 97) and giving the input of the notes through a MIDI keyboard. The addition of

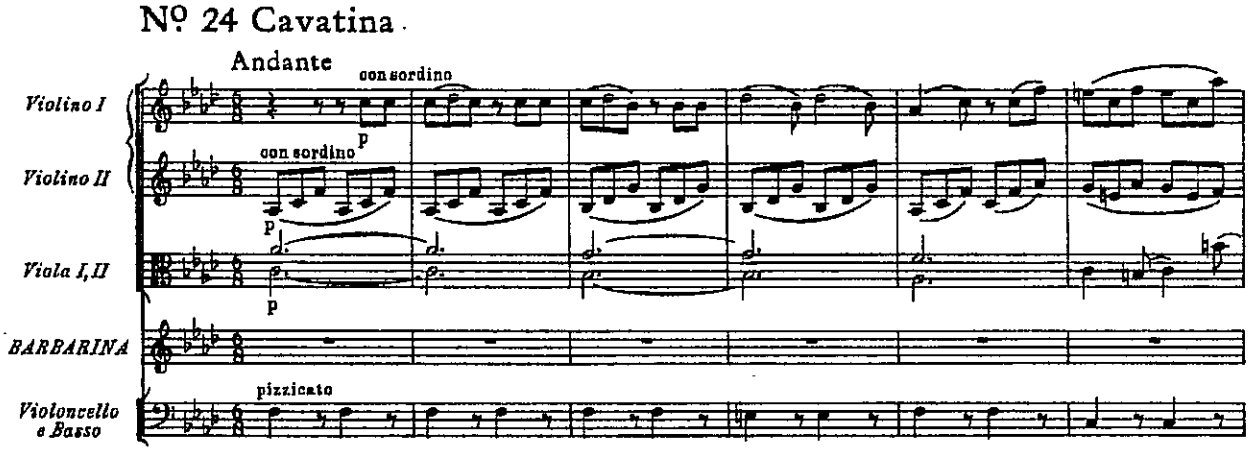


Fig. 1: *Example of orchestral score, the Cavatina 24 incipit*

* This project has been partially supported by the Italian National Research Council in the frame of the "Methodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneous magnetic media" research, Finalized Project "Cultural Heritage" (Subproject 3, Topic 3.2.2, Target 3.2.1).

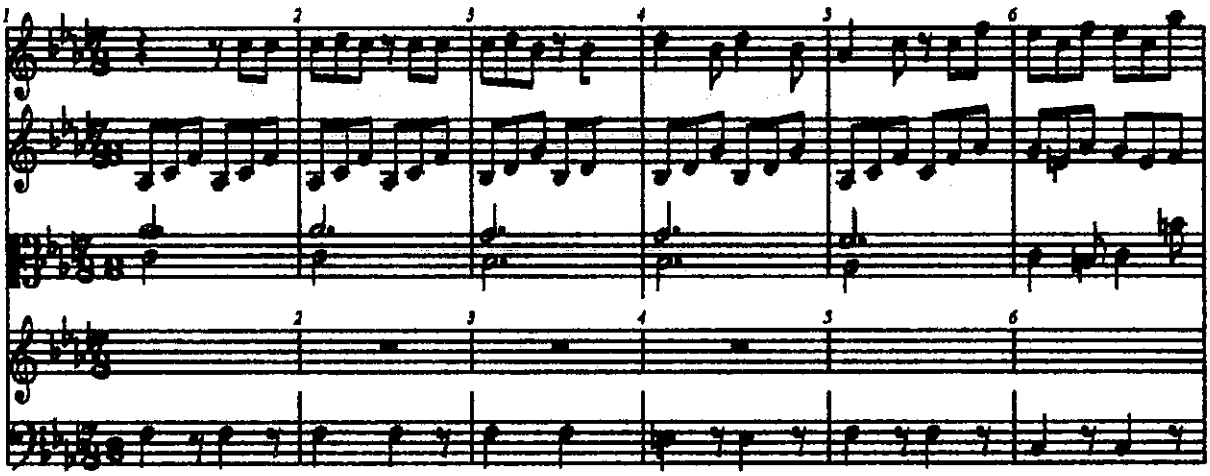


Fig. 2: *the score of Fig. 1 as recognized by the OMR application*

the conductors' annotations occurs by hand (sometimes it's a very heavy work).

So, to maintain a certain continuity in the way of working has been necessary to produce the following material for every score (in Finale Enigma format):

- a complete orchestral score;
- the parts for every instrument;
- the song-piano part (used during the rehearsals by the singers, the producer, the choreographer, and so on).

In regard to the building and the filling of a database of scores it has been necessary to face also the problem about the automatical extraction from the symbolic codification of the main themes of a whole excerpt, themes on which to do a comparison during audio or symbolic query.

2. Starting approach

The starting approach to this question (then in conflict with the current restrictions of the software) provided for an automatic loading of the orchestral scores through the scanning and the following conversion. Particularly the beginning work could be divided into 8 passages:

1- division of the score into some basic units (aria, duets, terzetto, ... , 20 pages long on average);
2- scanning of the basic unit to reproduce a copy in graphic format (TIFF);
3- conversion of the basic unit from graphic format into symbolic format (MIDI) through the application of the commercial OMR (Optical Music Recognition) application Midiscan 2.5.1 [10];
4- correction, within the same Midiscan session, of possible mistakes of symbolic recognition;
5- inlet in Finale 97 of the obtained symbolic score with conversion into the proprietary Enigma format;
6- editing under Finale (with manual adding of directorial annotations, but not in a high number in the case of this opera);
7- automatical extraction followed by a make-up of the single instrument parts;
8- correction of some possible mistakes made during the previous passages.

It has also been studied, using the automatical procedures given by the notation program, a method about the extraction of song-piano part from the orchestral score. This part of the project is still in course of study because of the high quantity of needed intelligence. Of course it cannot use rough approaches of stave implosion reproducing too scanty or unplayable scores.

In this phase it has been very important the aid of the M° Angelo Paccagnini, in pointing out the importance of some particulars compared to some others and in the correction of mistakes escaped even to the score listening.

3. The limitations of the starting approach

In the Figure 2 we can see the outcome of OMR applied to the system of Figure 1. Even if the excerpt doesn't present in this case particular difficulties, emerge some tipical mistakes such as the wrong time signature (7/8 instead of 6/8), lacking notes (particularly in the first and the second stave), additional notes (at the beginning of every stave), lacking augmentation dots and ties (in the third stave).

The deep analysis of every passage, made in the same time of its execution, has laid to the definition of the limitations of this method and to the following development of some appropriate solutions to overcome them.

The two main reasons of problems are:

- limitations of carefulness in the procedure of OMR [2] [3] [4], (especially in the case of complex scores);
- limitations of inadequacy of the adopted format of symbolic codification [5] (the MIDI turns out to be unsuitable to the representation of scores and the following conversion in Enigma introduces a further important loss of information).

In regard to OMR some "critical cases" have been found, i.e. the configurations of scores such that cause mistakes then hard to correct or irretrievable mistakes. These "critical cases" can be so summarized:

- slight inclination (also changeble along the width of the page) of the scores;
- very high or very low notes as to the staves;
- changeble distance between the scores;
- systems with changeble dimensions.

The latter case in particular forces to a further division of the basic units as far as to use the same system as the smallest unit (a following work of rebuilding is the result).

In regard with the codification format we have the complete inadeguacy of the MIDI to the reproduction of scores, because information essential for print and for reading is not included in the standard (among other things not arisen for this purpose).

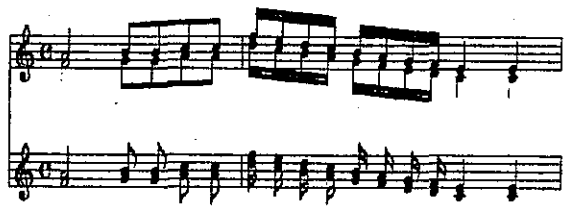


Fig. 3: *Differences beetween MIDI and NIFF coding*

For example in the case of two voices written simultaneously on a single score, the direction of the stems plays an important part for the correct assignment of every note to the right voice. Figure 3 shows the difference between a correct visualization of a similar case (see the first score) and what results from the conversion into Standard MIDI File (see the second score): the notes belonging to two different voices are grouped in a chord and the information concerning the beaming is lost.

4. The modified approach

The problems learnt and described in the preceding paragraph have forced to the definition of an alternative method of acquisition:

1. division of the score into smaller basic units (system);
2. conversion of the basic unit from the graphic format into the symbolic format;
3. building in Finale of a template valid for the whole excerpt;
4. editing under Finale;
5. automatical extraction and page make-up of the instrument parts;
6. correction of some possible mistakes made during the previous passages.

A careful valuation of the acquisition times with this modified approach has laid sometimes to prefer the direct inclusion in Finale through a MIDI keyboard.

5. Development of new solutions

The first decision has been that of getting over the MIDI codification and taking up to a codification studied expressly for a graphic return of the score. We have so decided to use the new "potential standard" of notation provided by the NIFF [6] (Notation Interchange File Format).

In Figure 4 we can see the same fragment of Figure 1 visualized by importing the corresponding NIFF file through the application of notation called Lime 5.03 [12].

This decision and the almost complete current lack of applications supporting this format has as a result the development of conversion applications: in particular it has been developed in Tonino Mendicino's graduation thesis [8] a MIDI-NIFF translator capable of locating in the MIDI files hidden musical structures, that can be represented in NIFF in an appropriate way, and in Paolo Mandatelli's thesis [7] a NIFF-Enigma translator that produces an Enigma file starting from the NIFF file given by the OMR program, and simulates the actions of a Finale operator (the Enigma format is not yet published and Coda has no intention to adopt the NIFF).

A deep knowledge of this format give us now the possibility to use it as a base for the development of further applications in the general project of the archive, such as the automatical segmentation of scores (necessary in the database to do the search for excerpts using short audio or score fragments), and the automatical adding of the conductors' annotations.

The track undertaken in the same time and now developing in Andrea Bandera's thesis consists in originating, after identifying the main problems of OMR systems, an application of pretreatment of the figure able to correct in the starting graphic files every imperfection causing serious problems. It must have the ability to:

- individuate the position of the staves so as to order them correctly, adding empty staves at the right places, when the system becomes composed of a smaller number of staves;
- correct inclinations (also changeble) along the width of the page (typical for photocopied pages);
- equidistance, reduce or widen the staves in order to solve the problem of notes very far from the staves.

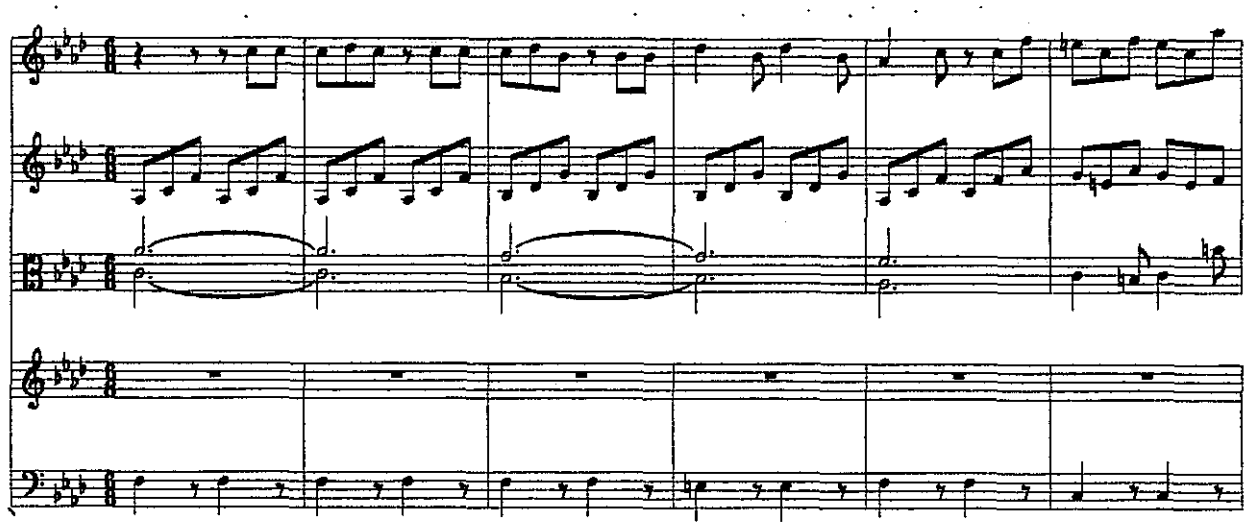


Fig. 4: *NIFF representation of the figure 1 fragment*

The combination of these solutions allows us to adopt the following acquisition procedure:

1. division of the score into basic units (aria, duets, terzetto,... 20 pages long on an average);
2. scanning of the basic unit to reproduce a copy in graphic format (TIFF);
3. graphic preprocessing of the basic unit;
4. conversion of the basic unit from graphic format into symbolic format (NIFF) through the application of commercial OMR (Optical Music Recognition) Midiscan 2.5.1;
5. automatical inlet in Finale 97 of the obtained *symbolic score*;
6. editing under Finale;
7. automatical extraction and following page make-up of the scores.

Acknowledgements

Authors are pleased to thank both researchers and graduate students working at LIM in the frame of this project. Thanks also to the team working at the Scala Music Archive for their precious help in defining the requirements of the project: Carlo Tabarelli, Cesare Freddi, Laura Serra.

A special thank is due to Angelo Paccagnini for his fundamental role in formalizing and designing all the automatic procedures for OMR and music formatting.

References

[1] Wolfgang Amadeus Mozart : "Le Nozze di Figaro", Barenreiter Kassel, 1973

[2] David Baindbridge, Nicholas P. Carter : "Automatic Recognition of Music Notation", in Handbook of Optical Character Recognition and Document Image Analysis, H. Bunke e P. Wang (editors), World Scientific, pp 557-603

[3] Ichiro Fujinaga : "Adaptive Optical Music Recognition", Faculty of Music, McGill University, Montreal, Canada1

[4] Eleanor Selfridge-Field: "Optical Recognition of Musical Notation: A Survey of Current Work", Computing in Musicology 9 (1993-94), pp. 109-145

[5] Eleanor Selfridge-Field (editor) : "Beyond MIDI: the Handbook of Musical Codes", MIT Press

[6] Cindy Grande: "NIFF 6a.1, Notation Interchange File Format"

[7] Paolo Mandatelli: "Un sistema per la trascrizione automatica di partiture musicali da formato NIFF a formato Enigma", LIM, Università degli Studi, Milano, 1998

[8] Tonino Mendicino: "Codifica dell'informazione musicale: progettazione e sviluppo di un ambiente prototipale integrato di codici MIDI e NIFF", LIM, Università degli Studi, Milano, 1998

[9] Maurizio Longari: "Codifica integrata di informazione audio digitale, MIDI e NIFF basata su SMDL/HyTime", LIM, Università degli Studi, Milano, 1998

[10] Musitek: "Using Midiscan 2.5 for Windows", http://www.musitek.com, 1993

[11] Coda Music Technology: "Finale 97 User Manual", http://www.codamusic.com, 1997

[12] Lippold Haken, Dorothea Blostein, Paul S. Christensen: "Lime User's Manual", http://datura.cerl.uiuc.edu, 1998

Coding Music Information within a Multimedia Database by an integrated description environment*

Goffredo Haus **Maurizio Longari**
LIM Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
via Comelico, 39 I-20135 Milano (Italia)
fax +39 2 55006373 e-mail: maurizio@lim.dsi.unimi.it

Abstract

The symbolic representation of musical information involves several aspects of computer science. A piece of music can be represented in several ways such as: digital signals, sets of time-dependent events that a sequencer (or other devices) can translate into an audio performance (e.g. MIDI), sets of graphic images that represent the score (e.g. TIFF, GIF, JPEG), a notational representation of the score that respects its hierarchical structure (e.g. NIFF, DARMS) or a representation of the logical information (for logical information we mean the information common to all the above music aspects) that the composer intends to put in the piece (SMDL). The fact that SMDL is an HyTime application, makes it a powerful environment in which all the other formats can be integrated and organized by means of the hypermedia ability. Since SMDL's cantus domain can represent the logical information of a musical piece, we can extract from files coded in the above mentioned formats the logical information, translate them into SMDL, and then use the SMDL representation within a database.

The representation of the musical information of a whole piece of music within the database is quite expensive with respect to memory occupation and computing time. Therefore it is necessary to devise methods for reducing the cost. A possible solution to this problem, we are currently working on, is to automatically extract from the SMDL files the significant parts of a piece of music, and then code these parts into the database instead of SMDL file. By this approach, most of the queries on pieces of music can be solved directly into the database without requiring to access the files.

Therefore in this environment SMDL plays two roles: the role of the common descriptor for the other file formats with the ability to organize them as hypermedia objects, and the role of the front-end to the database system.

SMDL's origins

The standard music description language (SMDL [1]) was born in 1984 based on an idea of Charls Goldfarb for standardizing the computer representation of music in an SGML format. The idea was to create an environment in which the several aspects of music representation can be considered and used.
SMDL is an instance of HyTime (Hypermedia Time/Based Description Language [2][3]) and so it inherits the functionality of this standard [6][7]. It divides the aspects of music representation in four domains:

- Logical (cantus)
- Visual
- Gestual
- Analytical

The Logical domain can represent the intent of the author when it thinks the musical piece. The Visual domain incorporates the graphic information and score notations. The Gestual domain incorporates the performance and executional information. The Analytical domain incorporates all of the music background information.
In the Logical domain you can write in an SGML format the music information while the other domains use the hyperlink capabilities of HyTime to point to and integrate the other formats that can represent the particular information [5].
The Logical domain is called *cantus* and the SGML syntax of SMDL is defined in the standard draft ISO 10743 [4]. The structure of information is developed over a system of time dependent axis called FCS (Finite Coordinate Space an architectural form of HyTime) which is composed by *cantus events* (notes, rests, symbols). The other domains can refer to this representation to link the several piece of music information.

SMDL vs musical standard formats

As s prototypal study of relations between SMDL and other formats we consider the well known formats:

* *The project has been partially supported by the Italian National Research Council in the frame of the "Methodologies, techniques, and computer tools for preservation, the structural organisation, the intelligent query of musical audio archives stored on heterogeneous magnetic media" research, Finalised Project" Cultural Heritage" (Subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1).*

- MIDI: as an instance of Gestual domain;
- AIFF: as another instance of Gestual domain;
- NIFF: as an instance of Visual domain.

The essential information that must be coded in SMDL is the representation in time of the notes, so during the analysis of the formats we try to individuate if we can extract this information from the files coded in each of the above formats.

In the MIDI format notes are coded in various tracks and channels by means of MIDI messages NOTE ON and NOTE OFF so it's easy to extract notes, duration and parts from the MIDI files.

The AIFF format is an example of audio file format. All this formats store, in different manner, the sample of audio signals. The difficult to manage this kind of information isn't the way in which they are stored (*e.g.*: by means of chunks) but to extract the musical information. In fact there are a lot of study about audio segmentation. So, for the moment, it's only possible to include this type of information as a compact object that represents a particular instance of the musical piece. The AIFF format is interesting because it is a standard format designed by means of chunks (the IFF format) and so it allow the standard to be very handy, the new version permits also to include several type of compression algorithms: a very useful facilities for files of those dimension.

The NIFF format is a new standard designed to code the notational information of music. It is the result of an effort of the major music software-houses to allow an easier interchange between the several different program commercially available. Its structure is made by chunks in which are coded the objects that compose a score. The score is divided in pages, systems and staffs in which are contained the musical symbols. The way in which they are coded allows an easy extraction of the information about notes, duration and parts so it's possible to integrate NIFF format in an SMDL environment.

Translation and integration in SMDL

Based on the above discussion we have developed a software that extracts the information of notes, duration and part from the MIDI and NIFF files, the AIFF format is not currently considered.

We did start from a NIFF to an SMDL translator made by Steve Mounce of Bradford University [11]. This software is composed by two stages that execute the steps of translation. The first stage extracts from a NIFF file, by parsing of file, the essential information for the translation and writes them in another file in an intermediate SGML format using the tools provided by the NIFF SDK toolkit. The second stage takes the intermediate file and translate it in SMDL syntax by means of the HyMinder System (the only software-tool available for can managing the HyTime Standard [10]).

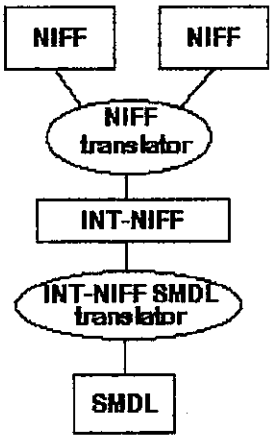


Fig.1: NIFF to SMDL translation

Starting from this two stages we have developed a second branch of translation that takes MIDI files (in both format0 and format1), translates it in another intermediate SGML format (INT-MIDI) and by means of HyMinder system translates it in SMDL syntax.

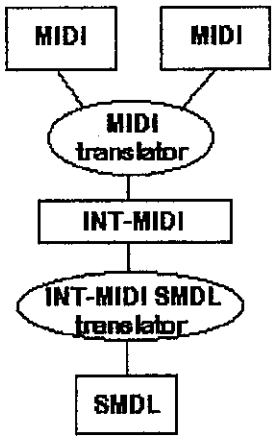


Fig.2: MIDI to SMDL translation

The two intermediate formats are both written in SGML syntax because the HyMinder system can read them, translates them in its internal class-representation and so writes them as SMDL files. But their structure is different because each of those reflects the structure of the respective source file: that is INT-NIFF file reflects the structure of NIFF file and INT-MIDI file reflects the structure of MIDI files.

At this point, if we have the version of the same piece of music in NIFF and in MIDI format, we can translate them in two, quite different, SMDL files. A third stage makes a merge of this two files comparing the information of notes present in each file and writes an SMDL file, the result is an instance of cantus and the hyperlinks to the source file from which we have extracted information are the instances of Visual and Gestual domains.

Since the software developed is a prototypal it works, for the moment, only with pre-prepared file. It can be

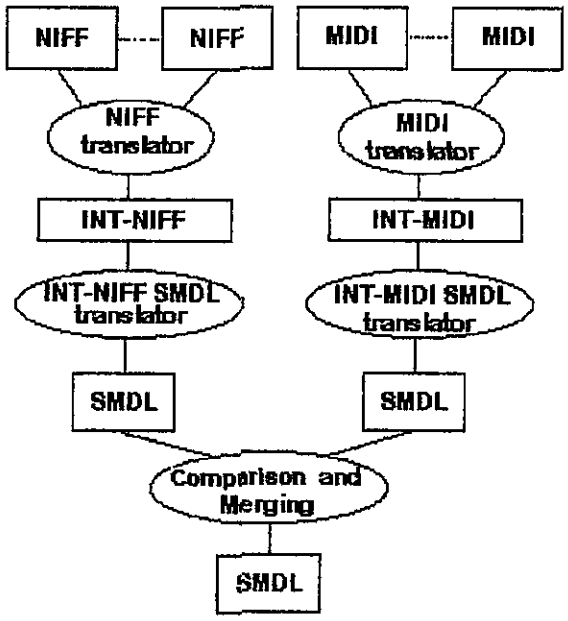


Fig.3: Translation and Merging

extended to work with all kind of file by means of some changes.

Integration with multimedia database

The problem that arises in the treatment of music files in a database environment is that they aren't directly representable in database tabular forms. This implies the conversion and the extraction of some type of data in order to allow the system to do the queries in an acceptable time [14][15].

The conversion of music files in SMDL permits to accumulate all the information in one file upon which we can extract the desired information only one time for each piece of music.

The kind of information that we extract from SMDL files is the parts, by means of which we recognize the piece. This is obtained by the implementation of score segmentation algorithms.

Now we'll take a look to a brief description score segmentation algorithms and the structure of an example database.

Score segmentation algorithms

The segmentation algorithms are based on the observation that most composition (quite all) have some source parts (*i.e.* brief sequence of notes) to which a great part of the piece can be brought back by means of several *music functions* [17][18][19][20]. This function can be a transpose, a repetition, a specular inversion or other transformations. The source parts, in most cases, are the melodies of the piece so they allow distinguishing the piece from other pieces also with respect to the human perception [13].

From this view point we can use the segmentation algorithms to extract the source parts from SMDL files and use them within a tabular form as indexes of pieces for intuitive queries in the database. As intuitive queries we mean the queries where the user can input the information (melody of the piece) playing a keyboard or whistling in a microphone [23].

An example of database

The example we have sketched is based on an object relational database management system (Oracle8) and in particular upon some of its tools as "Nested Tables" and large data type. Nested tables are particular objects that allow the treatment of some tables as row of other tables [16].

During the study of musical information in relation with database management we have noticed that there can be different level of information at which the user could access. For example a query by whistling must access a different kind of information (*i.e.* more detailed) from a catalogue query or from a musicological query (*e.g.* search how many times this sequence of notes is repeated in the piece).

So, we have divided the structure of the database into three levels:

- Database Level: information in tabular form.
- Logical Level: information in SMDL form.
- Code Level: information in source forms (NIFF, MIDI, AIFF, etc..).

In the Database Level the catalogue information is contained in tabular form and the source parts of pieces, extracted by means of segmentation algorithms, are also contained in tabular form and stored in nested tables. In the tables below we show an example of the score's database: the audio database need a separate (but joined) treatment because often the several pieces refer to a "show event".

In Logical Level are contained the SMDL files extracted from source files. As explained before there can be more source files to which a single SMDL file refer.

In the Code Level are contained all source file from which are extracted the SMDL files. They are the actual materials of the multimedia database.

SCORE Table

Field	Description
Id	Score index
Nome Autore	
Cognome Autore	
Data Pubblicazione	Day , Month, Year
Luogo Pubblicazione	
Edizione	Editor Surname, Date (Day, Month, Year)
Tipo	Opera, Concert, ...
Strumenti	Instrument's list as set of character
Nomi file	Nested Table referring to code files (FILERIF)
Elementi Significativi	Nested Table referring to the tables containing source parts (ELEMENTI_SIGNIFICATIVI).

FILERIF Nested Table

Field	Description
Strumento	Score's Instrument name.
File NIFF	Pointer to NIFF file.
File SMDL	Pointer to SMDL file.

ELEMENTI_SIGNIFICATIVI Nested Table

Field	Description
Strumento	Scores's Instrument name.
Tabella Note	Pointer to NOTE table.

Source part Nested Table.

Nested Table NOTE

Field	Description
Accento	1 character.
Altezza	Number from 0 to 127; 0 is NULL value.
Durata	Integer; fraction.
Dinamica	String of character (max. 4).
Tempo inizio	Integer defining start time (in virtual units).
Tempo fine	Integer defining end time (in virtual units).

Conclusions

The discussion above is only a prototypal draft, many things can be developed and implemented:

- in 1997 a new release of HyMinder have been approved by ISO, it's necessary to incorporate this new features [8][9];
- SMDL is still in draft status and since it will change because of the changes in HyTime it's necessary to implement them;
- more complicate musical structures;
- to allow the translator to recognize the more kind of files as possible (a generic file MIDI or NIFF);
- to implement the translation and integration of audio files;

Many other things can be developed and implemented in order to make the system more powerful and flexible. This article is intended as a starting study of an application of SMDL as an actual tool to manage the musical information in a database environment.

Acknowledgements

We would thank Steve R. Mounce of Bradford University for him help and for the start up information that he gave us.

References

[1] D. Sloan, "Aspect of Music Representation in HyTime/SMDL" Computer Music Journal, Cambrige, MIT Press 17:4, pp 51-59,Winter 1993

[2] S.R. Newcomb, N. A. Kipp, V. T. Newcomb "The HyTime. Hypermedia/Time-based Document Structuring Language". Communication of the ACM 34:11, November 1991.

[3] C.F. Goldfarb, S.R. Newcomb, W.E. Kimber, P.J. Newcomb "Information Processing – Hypermedia/Time-based Structuring Language (HyTime) - 2nd edition", ISO/IEC JTC 1/SC 18 WG8 N1920rev, 9 Maggio 1997.

[4] S.R. Newcomb, Standard Music Description Language ISO/IEC DIS 10743 ftp.techno.com/pub/SMDL/10743.ps, June 1995.

[5] S.R. Newcomb, "Standard Music Description Language complies with hypermedia standard" *COMPUTER, vol.X N.Y. 24:7, pp. 76-79, July 1991*

[6] "A Brief History of the development of SMDL and HyTime" http://www.techno.com/history.html

[7] S.J. Derose, D.G. Durand, "Making hypermedia work : a user's guide to HyTime" Boston, Kluwer Academic, 1994

[8] E. Kimber, "An Excerpt from Practical Hypermedia: An Introduction to HyTime Property Sets and Groves", http://www.hight.com/IHC96/ek8.htm, 1996.

[9] "The GroveMinder System – Technical IHC 1997 Presentation" http://www.techno.com//gminder3.htm

[10] "HyMinder User Guide 0.8.4", June 1996, Techno Teacher Inc., http://www.techno.com

[11] S.R. Mounce "NIFF to SMDL translator Documentation", Software Documentation for WP3, Project Cantate

[12] Project CANTATE "Computer access to notation and text in music libraries" Project LIB-CANTATE-4-2016, Commission of the European Communities. Work Packages: WP1, WP2, WP3, WP5, Final Report.

[13] Ghias, J. Logan "Query by humming: musical information retrieval in an audio database" Proceedings *[of the] ACM Multimedia 95 : San Francisco*, California, November 5-9, 1995

[14] B.M. Eaglestone, "Composition tools integration with a music database system" Database and expert systems applcations : 6th International Conference ; proceedings/dexa 1995

[15] B.M. Eaglestone, "Extending the relational database model for computer music research" Computer representations and models in music, 1992

[16] M. Marchesi, "Le basi di dati orientate agli oggetti", Informatica oggi & Unix, May 1992.

[17] F. Lonati, "Metodi, algoritmi e loro implementazione per la segmentazione automatica di partiture musicali" Master Thesis in Computer Science University of Milan 1990

[18] D. Wilson, "The Role of Patterning in Music" Leonardo, Journal of the International Society for thr Arts, Science and Technology, vol. 22, Pergamon Press, Oxford, 1989

[19] G. Haus, "Elementi di informatica musicale" Jackson, Milano, 1984.

[20] F. Lerdhal - R. Jackendoff, "A Generative Theory of Tonal Music" MIT Press, Cambridge, USA 1993.

[21] E. Pollastri, G. Haus, "Metodi e prototipi software per la classificazione e il reperimento di brani audio in archivi musicali basati su tecniche di elaborazione numerica dei segnali", Master Thesis in Electronic and Computer Engineering, Politecnico di Milano, 1998.

Characterization of Music Archives' Contents. A Case Study: the Archive at Teatro alla Scala.°

Goffredo Haus, Angelo Paccagnini, Maria Luisa Pelegrin Pajuelo
LIM Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
via Comelico, 39
I-20135 Milano (Italy)
fax +39 2 55006373
e-mail: haus@dsi.unimi.it

Abstract

In this paper we try to characterize special features of music archives. To do this we consider a representative case study: the Music Archive at Teatro alla Scala. We summarize the music contents ot the archive considering all the kind of materials involved - i.e. audio tapes, videos, etc. - with particular emphasis on audio media and scores.
Both qualitative and quantitative aspects of these materials are considered in order to allows us to define a basic schema for designing a computerized processing of information within a music database framework. So, for each kind of material we consider the state of conservation, the historical, musicological, and economic-publishing value parameters, how it is involved in the usual activity for the realization of new productions of the theatre.
From this analysis we get mainly two kind of information: about the contents and their artistic meaning, and the relations of these contents with the activities of the theatre.

Music Archive contents

Although Teatro alla Scala is often associated with operas, concerts, recitals, it is probably not known that it possesses a historical music archive of great significance. It includes three kinds of music media:
- audio recordings;
- video recordings;
- music scores.

Audio recording of live performances has kept on with no interruptions since 1951. This activity has produced, in close to five decades, over 5,000 media: analog open reels from 1951 to 1990, DATs from 1991 to 1996, CD-R from 1997, all of which contain the work of the most famous musicians of this century (singers, conductors, performers, etc.).

Video recordings cover approximately the same period, while they were always taken by a single fixed telecamera. So, they are not considered in the current project.

Music scores include:
- handwritten original scores;
- commercially available published scores;
- elaborations of previously published scores; for examples, personalizations of orchestral scores by conductors (they generally have special directives for dynamics and expression).

Both the first and the third have great historical value. Thousands of these documents are available and need to be preserved, and organized for efficient enjoyment. Furthermore, computer methods allows the Archive to enhance the quality of its products, and reduce the costs of many activities. For example, the printing of transposed parts.

Rescuing the Phonic Archive

Unfortunately, most of the audio tapes have not been well preserved and are actually deteriorating at a quick pace. For this reason, a panel of international sponsors has funded in 1997 a project for the preservation of this precious music heritage. The scientific direction and the execution of the project has been entrusted to the LIM - Laboratorio di Informatica Musicale of the Computer Science Department at the University of Milan [1].

The projects consists of many tasks and activities, quickly summarized as follows:
- cleaning of analog tapes;
- heat treatment for those tapes that have softened due to age;
- digitization of analog tapes to preserve and manipulate their contents;
- mastering of more than one CD-R for each original recording;

° This project has been partially supported by the Italian National Research Council in the frame of the "Methodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneous magnetic media" research, Finalized Project "Cultural Heritage" (Subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1).

- classification of the original recordings and of the new media thanks to about 60 attributes which are entered in a distributed and multiplatform database under Oracle8.

Here follows a short explanation of the techniques mentioned above. While many tapes need to be cleaned by hand, turn by turn, with a special liquid, those which have suffered from softening of the oxide coating and have absorbed moisture during their long term storage require a special heat treatment if they are to be restored to playable condition - without squeaking, nor sticking to the guides and heads of the recorder - and transferred to digital form. This treatment consists in putting the tapes into a heated oven or incubator at a temperature of 45°C - 55°C for about three days. Once the tapes reach a satisfactory quality, their contents is digitized at the standard sampling rate of 44.1 kHz and with a 20-bit, and stored on computer hard disks. The resulting digital audio files are edited and structured as digital tracks that correspond to single musical pieces. Meaningless heads and tails are removed. The main goal of this step is to keep the original information exactly "as is", postponing choices for a possible restoration - hence, extra noises are also carefully preserved. The digital tracks are copied on CD-R's, for each of which there are two initial main copies: one for the Music Archive of the Theatre and a second one as a backup copy stored in the vault of the Italian Commercial Bank.

While these activities take place, information is collected about the contents of the tapes, both from the artistic and from the technical point of view. In addition, all information acquired during cleaning and heating, digitalization and mastering, contributes to the definition of roughly 60 attributes about the original recordings and the new media, entered in a database built with Oracle 8 and running on a multiplatform (UNIX, NT, Windows95, Mac OS) and distributed (about 10 workstations) system [2] [3]. At the same time, the research team at LIM is designing and developing special software modules for Oracle, to process musical and multimedia objects: digital sound, scores and performances, photos, videos, etc [4]. In this way, musicians who are currently preparing a performance which has been previously given at the Scala can easily access all CD-R's of previous performances, under Oracle8, located in a CD-ROM juke box (224 CD online, 4 CD mounted).

Relevant features of the Archive audio media

Original recordings concerns the period 1951-1996 with a very different density with respect to periods 1951-1973 and 1974-1995. Only few tapes of the former and quite the complete covering of the theatre nights of the latter. In Table A you can see how original tapes are distributed with respect to their dating. Open reel 1/4" mono/stereo tapes were used since 1951 till 1990, while DAT stereo tapes were used since 1991 till 1996. Starting with the first night of december 1996 the theatre are directly recorded onto CD-R.

All the main genres of theatrical and music events are represented within the Music Archive. You can see in Table B the numbr of tapes available for each of the main genre represented within the archive. A number of different rehearsal sessions were recorded so that a historical and musicological heritage is present together with the performances recordings. Particularly interesting the dress rehearsals and the singers' training sessions. Most of the famous conductors, singers, performers, orchestras, dancers, etc. are well represented and, even if recorddings are not of the top technology class, all the audio media of the Scala' archive are of interest for both cultural and publishing purposes.

1951	1
1952	0
1953	0
1954	2
1955	3
1956	1
1957	0
1958	12
1959	23
1960	33
1961	31
1962	48
1963	25
1964	36
1965	6
1966	6
1967	0
1968	7
1969	16
1970	4
1971	26
1972	12
1973	18
1974	185
1975	403
1976	603
1977	409
1978	529
1979	336
1980	276
1981	309
1982	280
1983	249
1984	227
1985	273
1986	400
1987	341
1988	416
1989	372
1990	416
1991	560
1992	754
1993	736
1994	813
1995	454

Table A

Concerts	5770
Operas	2527
Balletts	970
Recitals	571
Lectures and meetings	186
Various events	12
Melodramas	2

Table B

For example, there are 263 different conductors within the recordings. Most of them have conducted only few times while the most "present" are shown in Table C. The right column gives the number of tapes in which conductors occur.

Muti Riccardo	844
Abbado Claudio	353
Gavazzeni Gianandrea	197
Maazel Lorin	178
Pretre Georges	158
Sawallisch Wolfgang	151
Giulini Carlo Maria	123
Chailly Riccardo	108
Sasson Michel	100
Mehta Zubin	89
Gatto Armando	76
Pesko Zoltan	74
Chung Myung-Whun	69
Kleiber Carlos	64
Patane' Giuseppe	62
De Mori Enrico	60
Ranzani Stefano	60
Urbini Pierluigi	59
Ozawa Seiji	58
Sanzogno Nino	58
Boulez Pierre	56
Gatti Daniele	56
Scimone Claudio	55
Solti Georg	55
Orizio Agostino	53
Gabbiani.Roberto	51
Ferro Gabriele	45
Mueller Edoardo	41
Sinopoli Giuseppe	39
Benini Maurizio	38
Renzetti Donato	37
Rozhdestvenskij Ghennadi	37
Schippers Thomas	37
Florio Ermanno	35
Letonja Marko	34
Berio	32
Campanella Bruno	31
Gandolfi Romano	31
Weller Walter	30
Fedoseyev Vladimir	29
Thielemann Christian	29
Inoue Michiyoshi	24
Von Karajan Herbert	24
Theuring Guenter	23
Bernstein Leonard	22
Penderecki Krzysztof	22
Votto Antonino	20

Table C

The theatre is well known mainly for the performances of operas. There are 318 different operas within the archive. Table D shows the operas which were more frequently performed. The right columns gives the number of tapes concerning that particular opera.

Similar tables can be shown concerning singers, performers, dancers, etc.

Giacomo Puccini	La Boheme	81
Giuseppe Verdi	La Traviata	67
Giacomo Puccini	Madama Butterfly	59
Giuseppe Verdi	Don Carlos	51
W. A. Mozart	Don Giovanni	49
Francesco Cilea	Adriana Lecouvreur	41
Gaetano Donizetti	Lucia di Lammermoor	41
Jules Massenet	Manon	40
Gioacchino Rossini	Guglielmo Tell	36
Carl Maria Von Weber	Oberon	36
Richard Wagner	Parsifal	36
Giuseppe Verdi	Rigoletto	35
Ferruccio Busoni	Turandot	33
Gioacchino Rossini	La Cenerentola	32
Giuseppe Verdi	Aida	31
Vincenzo Bellini	Beatrice di Tenda	31
Gaetano Donizetti	Don Pasquale	31
W. A. Mozart	Le Nozze di Figaro	31
Giuseppe Verdi	Simon Boccanegra	31
Gaspare Spontini	La Vestale	30
Giuseppe Verdi	Nabucodonosor	30
Gioacchino Rossini	Il Barbiere di Siviglia	28
Gioacchino Rossini	Otello	27
Giuseppe Verdi	I Vespri Siciliani	26
Gioacchino Rossini	Maometto II	25
Giacomo Puccini	Tosca	25
Giuseppe Verdi	Falstaff	24
Giacomo Puccini	La Fanciulla del West	24
Umberto Giordano	Fedora	23
Ernest Bloch	Macbeth	23
Ruggero Leoncavallo	Pagliacci	23
Giuseppe Verdi	La Forza del Destino	22
Gioacchino Rossini	Tancredi	22
Giuseppe Verdi	Un Ballo in Maschera	22
W. A. Mozart	Cosi' Fan Tutte	21
W. A. Mozart	Die Zauberfloete	21
W. A. Mozart	Idomeneo, Re di Creta	21
Sergei Prokofiev	L'angelo di Fuoco	20
G. B. Pergolesi	Lo Frate 'Nnamorato	20

Table D

The general characteristic of the Scala's audio media is the heterogeneity with respect to:
- the quality of recordings,
- the state of conservation of magnetic media,
- the state of the copyright topic.

While transferring from analog to digital, we have found and catalogued sonic materials not previously known; particularly, a great number of encores from concerts, recitals, and even operas have been identified.

The description of the performances consists of 60 attributes which belong to the following major classes:
- Audio Archive Identifiers,
- Event Descriptors,
- Opera/Piece Descriptors,
- Technical Features of the Original Recording,
- Technical Features of the Transferring Process,
- Technical Notes,
- Digital Transfer Identifiers.

Perspectives

The main goal of our project is the design and development of a unique database environment for the whole multimedia database of Teatro alla Scala i.e. for simplifying communication, organization and management for many internal tasks.

From a quantitative viewpoint, the audio component of the project will be approx 5 terabytes; the other components (scores, photos, videos, etc.) are under estimation. Anyway we can consider an approximation for the complete DB around some 20 terabytes. Audio and most probably also other multimedia data will be taken out of the DB, but obviously linked and well-known within the DB. Some of them - the ones concerning events under production - will be kept online, the others are stored offline and indexed online.

Specifically, we are researching, designing, and developing formal and applicative tools to integrate symbolic and subsymbolic music data together with traditional alphanumeric data; even methods associated to those data have to be integrated; specific methods for browsing will be developed for PCM, MIDI, NIFF and SMDL components. Recently - 16 june 1998, in the frame of a press conference held at Teatro alla Scala - we have shown the first realization of music queries to an Oracle8 database i.e. asking by sound patterns to the database to achieve both audio files and digital scores.

Acknowledgements

Authors are pleased to thanks both researchers and graduate students working at LIM in the frame of this project. Thanks also to the team working at the Scala Music Archive for their precious help in defining the requirements of the project: Carlo Tabarelli, Cesare Freddi, Laura Serra.

A special thank is due to Fiorenzo Galli, Secretary General of Foundation Milano per La Scala, for his fundamental role in designing and supporting the whole project.

Then, many thanks are due to the sponsors of the project: AEM, Andersen Consulting, Banca Commerciale Italiana, Hewlett-Packard, Oracle, TDK, Teatro alla Scala Foundation USA.

References

[1] Goffredo Haus: "Rescuing La Scala's Music Archives", IEEE Computer, 31(3), 88-89, 1998

[2] Sara Record: "Brava La Scala", Oracle Magazine, May/June, 70-74, 1998

[3] G. Haus: "The Database Environment of the Scala Project", Multimedia Today, Milano, november 1997

[4] Goffredo Haus, editor: Technical Report Series in the frame of the finalized project "Cultural Heritage" of the Italian National Research Council, see site http://lim.dsi.unimi.it/PFBC_Musica/, 1997-1998

Melody-Retrieval based on Pitch-Tracking and String-Matching Methods*

Emanuele Pollastri
LIM Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
via Comelico, 39
I-20135 Milano (Italy)
Fax +39 2 55006373
e-mail: lele@lalim.lim.dsi.unimi.it

Abstract

The emergence of audio data types in databases and the rapid increase in speed and capacity of computers require new information retrieval methods dedicated to audio files. Our goal is to let users search musical pieces by melodic audio-content.

We developed a pitch-tracking system based on RMS-power segmentation and harmonic gathering; this software can measure pitches from monophonic sources in the range of 50 Hz to 20 KHz, with over 90% of successful estimates, and can set the basis for a potential polyphonic pitch recognition.

Resulting pitches derived from two different audio files can be compared by means of another software tool which is able to measure their degree of melodic similarity. For this purpose, three approximate string matching algorithms and a string distance calculation algorithm were implemented; all these algorithms are based on dynamic programming. Several different representations of melodic lines are available: absolute notes, intervals, relative intervals to a given note, exact and relative rhythm. The comparison can be carried out using only notes, only rhythm or both notes and rhythm. This system returns a ranked list of matching points and their degree of similarity.

Preliminary results show that it's actually possible to achieve a concise distance-measure between melodies acquired from audio files and to realise systems for melody-matching or melody-retrieval.

1. Introduction

Conventional Information Retrieval Systems are mainly based on (computer-readable) text. Tipically, text documents can be found by locating query keywords within them. Unfortunately, for audio and especially for music, this approach is useless due to the simple lack of identifiable entities comparable with words. In addition, the increase in dimension of digital musical libraries requires powerful algorithms to retrieve music information (audio files and/or scores).

In this paper, the difficult problem associated with learning new features for melody recognition in complex audio sources (eg. orchestral music) will not be considered (for a recent overview on this subject see [1]). Instead, we will be focused on analysing, representing and comparing monophonic sources. As an extension, we will describe a system designed to retrieve musical pieces stored in a database in the form of music notation by means of similarity to an acoustic input. In particular, a complete melody retrieval prototype has been recently implemented in Milan's La Scala Theater Digital Archives [2].

The remainder of this paper is organized as follows. Section 2 describes previous works on melody retrieval. In section 3 and 4, we present respectively audio analysis stage and comparison stage. Section 5 is dedicated to results and conclusion.

2. Background: Melody-Retrieval

By 'Melody Retrieval' we mean looking for items that contain a given theme or a sequence of notes in music databases. To our knowledge, only two integrated systems for melody retrieval has been attempted. The first was developed by Ghias et al. at Cornell University [3]; here, a melodic query based on few hummed notes is quantized to three levels, depending on whether each note was higher, lower or similar pitch as the previous one. Then, a MIDI-files database can be searched for similar melodic contour using an approximate string matching algorithm. The other application has been developed by the University of Waikato in New Zealand [4,5]. This system transcribes an acoustic input from the user into common music notation; then, it searches a database of folk tunes for those containing the sung pattern. Different search criteria are implemented:

* This project has been partially supported by the Italian National Research Council in the frame of the "Methodologies, techniques, and computer tools for the preservation, the structural organization, and the intelligent query of musical audio archives stored on heterogeneous magnetic media" research, Finalized Project "Cultural Heritage" (Subproject 3, Topic 3.2, Subtopic 3.2.2, Target 3.2.1).

melodic contour, musical intervals and rhythm. Tests were carried out using both approximate and exact string matching algorithms.

3. Pitch Tracking and Note Segmentation

Pitch-Tracking algorithms convert an acoustic input into note-like attributes. Unfortunately, this is possible for anything but simple music pieces. Monophonic transcription is well understood and can be obtained with algorithms usually classified by whether they work in time-domain, frequency-domain or hybrid-domain [6]. Polyphonic transcription is much less usual, even if some systems capable of transcribing music with more than two voices have been recently attempted (see Keith Martin [7] and Kashino et al. [8]); due to the great emphasis set on perceptual organization of sound, we will not consider these new methods. Instead, our aim is to provide a software enviroment in which users can develop a robust pitch-tracking algorithm based on conventional frequency domain analysis [9,10]. Figure 1 shows a diagram of the signal analysis path.

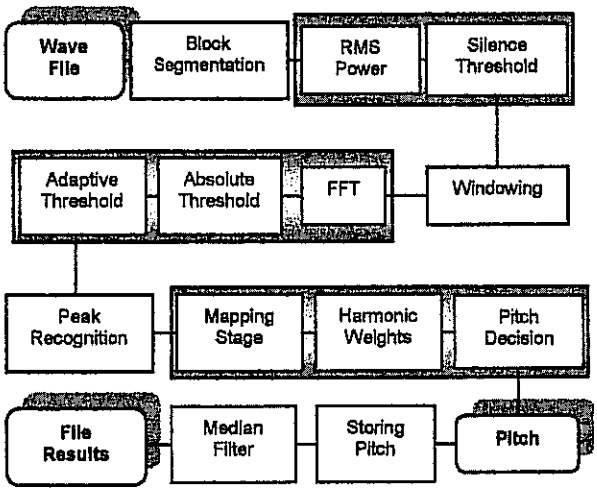


Fig.1: *Pitch-Tracking: diagram of the signal analysis path.*

A wav or raw file is taken in input and it is divided into overlapping frames; both frame and shift dimension are user-defined (in number of samples). For each frame, RMS-power is compared with a threshold to decide whether the frame is silence or not. Next step involves windowing; five different window types are implemented (Rectangular, Hanning, Hamming, Blackman, Bartlett). A Fast Fourier Transform of the windowed frame is performed. Resulting FFT-partials are compared with the lowest between an absolute and an adaptive threshold; partials below this threshold are set to zero. The adaptive threshold is calculated as in the equation below, where TRatio is a user defined value (good values are proven to be 3÷7) and MaxValMean is the average of the ten greatest partials.

$$Adaptive\_Threshold = \frac{MaxValMean}{TRatio}$$

Among all partials we are interested only on prominent ones; we call these partials Peaks. Peak-Recognition algorithm works on a list of schemas, each one representing a possible configuration of meaningful values. For example, a peak is the partial between a monotonically increasing and a monotonically decreasing sequence of partials.

Every peak is mapped first on a frequency in Hertz and then to the corresponding nearest MIDI-note number. This set of MIDI-note numbers represents the harmonic content of the current frame and we call it Detected_Harmonics. Identifying the fundamental frequency (or fundamental frequencies, in polyphonic case) is a matter of gathering all the Detected_Harmonics. This operation is achieved by calculating for each harmonic how many overtones we can find in Detected_Harmonics; moreover, the presence of an overtone is log-weighted depending on its amplitude. The harmonic that produces the greatest harmonic gathering weight is the pitch (of that frame). After pitches are determined, we pass them through a median filter of user-defined dimension.

This system provides graphical interface and can represent all computational steps. Resulting pitches and the most important information from the source file (sampling frequency, frame and shift dimension) are stored into an ascii file.

4. Melody-Matching

The pitch-tracking system translates an input audio file into a sequence of note-numbers, each of these representing the absolute pitch of a frame. Comparing these sequences is essentially a matter of string representation [11] and string matching.

Several different representations are avaible for both rhythm and melody information:

- absolute notes
- absolute rhythm
- interval (*note - previous_note*)
- relative rhythm (*duration / previous_duration*)
- relative interval to a given note
- relative rhythm to a given duration

To avoid symbol fragmentation, each relative representation is quantize to a set of suitable values. For example, we consider only intervals included in -15÷15 half-tones; intervals out of this range are set to 'Undefined', which is a valid symbol in our alphabet.

For performing the string matching we need an efficient approximate pattern matching algorithm [12]. By "approximate" we mean that the algorithm should be able to take into account various forms of errors. In fact, music is usually performed with a great variability, due to player's errors (memory deficiences or performing errors) or to musical variants.

In general, our problem is mainly based on measuring the distance between two given strings. We will use the Levenshtein Distance which is the minimum number of

insertions, deletions and substitutions needed to transform one string into another; this problem is addressed as Approximate String Matching with K-differences [13]. Sellers demonstrated that Dynamic Programming methods for exact string matching can be adapted to approximate string matching problems with a time which is proportional to the product of the lengths of the strings compared [14]. A great deal of work has been done to speed this operation; from the computational point of view, complexity in the worst case is always quadratic, but in case of random strings it is always less. In particular, we will use Ukkonnen's Cut-Off algorithm [15] which is always O(m) in space and O(kn) in time for random strings (m,n string dimensions; k is a distance threshold) and Chang-Lampe's algorithm [16] which is the fastest dynamic-programming-based technique at present. Finally, Wagner-Fisher's algorithm will be adopted as a standard string distance calculation method.

We developed a software tool in which all the above representations and algorithms are implemented. Users can choose a suitable representation, define a distance threshold and start a match between two files derived from the pitch-tracking system. The comparison can be carried out using only notes, only rhythm or both notes and rhythm. This system returns a ranked list of matching points and their degree of similarity. Scores are calculated by normalizing dynamic programming distance to pattern string dimension; 100% score indicates maximum similarity, ie. the entire pattern matched with text string from a given point.

5. Results and Conclusion

First, we tested the pitch-tracking system under several different conditions:

- synthesized sound / monophonic source.
- acoustic instruments / monophonic source.
- acoustic instruments / real performance.
- singing voice.
- polyphonic source.

We chose a collection of timbres from a sound-expander on the market and from McGill University Sound Library. Results showed more than 98% of successeful estimates for both synthesized and acoustic sources in monophonic case. Then, a real performance of solo flute caracterized by long time reverberation was analysed; we registered about 10% of unsuccesseful estimates, above all determined by room-acoustics and by noise from instrument's keys. The typical features of human voice caused a great increase in error rate (about 25% of unsuccesseful estimates); this is mainly due to the absolute musical scale applied in the mapping stage and to the unvoiced frame that are analysed as voiced frame. In case of polyphonic sources, present system is unable to execute a reliable transcription; many improvements could be made for adjusting peak-recognition especially to avoid meaningless peaks.

Then, experiments were carried out to test melody-matching and melody-retrieval. The right hand part of the third tempo of Beethoven's Sonata Op.13 was played with a sampled-sound of piano in the following variants:

- time-quantized performance without trills and grace notes.
- non time-quantize performance without trills and grace notes.
- real performance with expression.
- real performance, slow time and with some wrong and misplaced notes.

After the pitch-tracking stage, we matched intervals and relative rhythm extracted from all these examples with each others. Resulting comparisons returned a high degree of similarity; this confirms system capability in recognizing the same piece of music under different performance conditions. Then, we played only four bars from the same piece by Beethoven with a flute sound and we matched it with previous perfomances; the system returned the maximum degree of similarity connected with the right bar, ie. the system can localize the beginning of a pattern everywhere in a piece. To test melody-retrieval, about thirty pieces of music in different genres (soundtrack themes, one voice from Bach's Invenzioni, right hand part of Beethoven's Sonata Op.13) were processed by the pitch-tracking system. We simulated an audio query playing "by ear" few bars from one of the previous pieces.

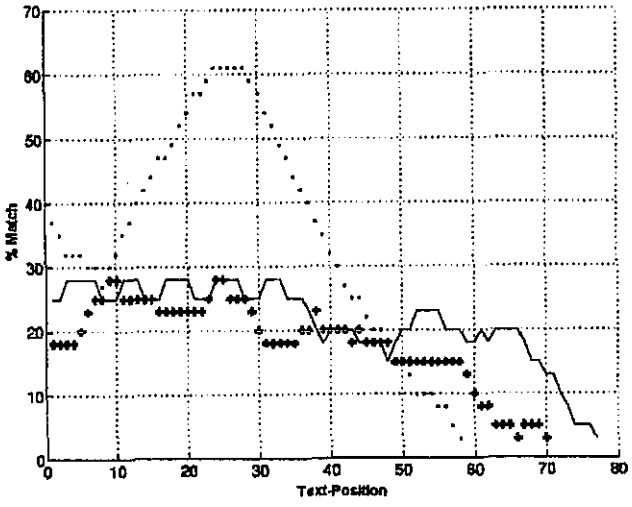


Fig.2: *Graphical representation of resulting matching scores in a melody-retrieval test; beginning from position 24, a piece of music (dotted-line) is very similar to the query.*

In fig.2 it is displayed the degree of similarity between the audio query and the three most similar pieces from the test-database; at position 24 of the dotted-line piece we have the right maximum similarity.

At present, the most slowly link in our system is pitch tracking, but its performance could be easely improved by means of better FFT algorithms. As expected, Chang-Lampe's algorithm was proven to be

the fastest; it took only few milliseconds (1÷10 msec on a Pentium 200) for short string-comparisons; we measured the worst performance when whole Beethoven's piece in the slow version was matched with the expressive version (262 msec). We plan to test this algorithm on a large database of scores based on Milan's La Scala Digital Archives which is currently under construction; an audio-query module has been implemented to extend the information retrieval feature of the database. How much our system is able to successefully discriminate music on the base of few notes is a matter still opened. At this time, we can point out that string-matching algorithms do not discriminate the various forms of errors, but consider them collectively. Though, preliminary results show that it's actually possible to achieve a concise distance-measure between melodies acquired from audio files and to realize systems for melody-matching or melody-retrieval.

7. References

[1] Martin, K. 1998 "Toward automatic sound source recognition: identifying musical instruments", to be presented at the NATO Computational Hearing Advanced Study Institute, Il Ciocco, Italy, July 1-12, 1998 (ftp://sound.media.mit.edu/pub/Papers/kdm-comhear98.pdf).

[2] Haus, G. 1998 "Rescuing La Scala's music archives", *IEEE Computer, Vol.31, N.3, March '98.*

[3] Ghias, A., Logan, J., Chamberlin, D., Smith, B.C. 1995 "Query by humming-Musical information retrieval in an audio database", in Proc. ACM Multimedia, San Francisco 1995 (http://www.cs.*cornell.edu/ Info/People/ghias/publications/query*-by-humming.html).

[4] McNab, R. 1996 "Interactive applications of music transcription", Master of Science in Computer Science Thesis, University of Waikato-New Zeland, 1996 (http://www.cs.waikato.ac.nz/~rjmcnab /publications.html).

[5] McNab, R., Smith,L.A., Witten, I.A. 1996 "Towards the digital music library: tune retrieval from acoustic input", in Proc. ACM Digital Libraries Conf.'96, Bethesda, Maryland, 11-18.

[6] Rabiner, L.R., Cheng, M.J., Rosenberg, A.E., McGonegal, C.A. 1976 "A comparative performance study of several pitch detection algorithms", IEEE Trans. on Acoustics, Speech and Signal Processing, vol ASSP-24, no.5, Oct'76, 399-418.

[7] Martin, K.D. 1996 "Automatic transcription of simple polyphonic music: robust front end processing", M.I.T. Media Lab Perceptual Computing Section, Technical report No. 399.

[8] Kashino, K., Nakadai, K., Kinoshita, T., Tanaka, H. Japan "Organization of Hierarchical Perceptual Sounds", Proceedings of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI-95), Vol.1, pp.158-164 Aug. 1995 (http://www.mtl.t.u-tokyo.ac.jp/~optima/Paper/ijcai95p.kashino.ps.gz).

[9] Schroeder, M.R. 1968 "Period histogram and product spectrum: new methods for fundamental-frequency measurement", J. of Acoustical Society of America, 43(4), 5pp, '68.

[10] Seneff, S. 1978 "Real-time harmonic pitch detector", IEEE Trans. on Acoustics, Speech and Signal Processing, vol ASSP-26, no.4, 358-365, Aug.'78.

[11] Lindsay, A. 1997 "Using contour as a mid-level representation of melody", Master of Science in Media Arts and Sciences Thesis; M.I.T. Media Lab (http://sound.media.mit.edu).

[12] *Sankoff, D. e Kruskal, J.B. 1983 "Time-warps,* string edits and macromolecules: the theory and practice of sequence comparison", Addison-Wesley'83.

[13] Stephen, G.A. 1994 "String searching algorithms", *World Scientific Publishing'94.*

[14] Sellers, P.H. 1980 "The theory and computation of evolutionary distance: pattern recognition", J.of Algorithms, 1, 359-373, '80.

[15] Ukkonnen, E. 1985 "Finding approximate patterns in strings", J.of Algorithms, 6, 132-137, '85.

[16] Chang, W. e Lampe, J. 1992 "Theoretical and empirical comparisons of approximate string matching algorithms", in Proc. Combinatorial Pattern Matching 1992, 1-14.

Saturday 26th **h. 11.00**

Invited Talk by Dietrich Schüller

Saturday 26th **h. 11.40**

RESTORATION OF AUDIO DOCUMENTS I (SPECIAL SESSION)

Can You Retrieve the Original Studio Acoustics In Pre-1925 Recordings?

George Brock-Nannestad

The Initiative for AV Preservation, Denmark
Resedavej 40
DK-2820 Gentofte, Denmark

Abstract

The paper opens a discussion of the retrievability of the actual sounds present in early recording studios, meaning not only the performance as such, but also the acoustic environment represented by intensity/distance relationships and reverberation. Heavy use is made of demonstrations of early recordings.

1 Introduction

Pre-1925 recordings (acoustic recordings) all made use of a diaphragm in a soundbox coupled by means of tubing to one or several conical horns directed towards the source of sound. This means that the receiver of sound was resonant and that any studio acoustics had the complex transfer function of the recording system superimposed on it. It is possible to perceive the original studio acoustics hidden behind the recording equipment, provided the linear influence of the equipment is eliminated. This is possible in many ways, although the masking influence of filtered background noises competes for the attention of the listener. Possibly this is the reason why there are so many opinions of taste in this field.

2 Approaches to the problem

The earliest investigations into the complex interaction between a sound source in a room, a horn terminated by a diaphragm, and a mechanism for registering the movement of the diaphragm were made by D.C. Miller [1]. Using essentially a 1/12 octave approach he determined that it was possible to subtract the inverse of the transfer function of the horn-diaphragm combination from the spectra he measured from a number of sounds in order to obtain correct spectra. The problem with undocumented horn/diaphragm combinations was dormant until the mid-1930s when both RCA-Victor and the Gramophone Co. constructed filters to enhance the voice of Caruso on his acoustic recordings and to suppress the orchestral accompaniment in order that it could be replaced with a modern accompaniment recorded electrically.

Thomas Stockham [2] discussed the problem in 1971 and introduced comparison with an all-electric recording of the same selection performed by a "similar" voice type. The results were not convincing in all respects. The present author has approached the problem from various angles since 1982, including re-construction of early recording situations [3].

The earliest existing live operatic recordings which were made in the far field of the recording horn and which have an immense documentary value are the cylinder recordings of Lionel Mapleson 1900-03. Listening to reissues from The Rodgers and Hammerstein Archives of New York Public Library does provide acoustic environment even through a high noise level.

3 Listening to early recordings

The language is not yet very good at describing the complex listening experiences obtained from early recordings. The trade of High Fidelity has refined its special language which enables discussions relating to the use of valve amplifiers versus transistor amplifiers (irrespective of linearity or bandwidth!), and obviously musical criticism also has its vocabulary. Hence, for progress to be made in the present field, a number of listening experiences, using primary source material, must be shared. During the presentation a number of the following recordings will have been presented and subjected to controlled manipulation.

1) Melba Distance Test (Matrix No. 4195f, recorded 11 May 1910) provides an unprecedented input in that Nellie Melba steps back between repeats of the same brief selection. The piano is stationary.

2) Adelina Patti: "La Calasera" recorded 1906

3) Paderewski: "Minuet G-major" recorded 1911

4) Barrère Little Symphony: "La Féria" (La Reja) recorded ca. 1915

5) Gas Shell Bombardment recorded 1918

6) Hennebains (flute): Allegretto (B. Godard) recorded ca. 1905

Experimental evidence demonstrates that it is possible to obtain a sense of depth in the reproduction of these and many other acoustic recordings, meaning that it is not just the perceived volume of e.g. a soloist performing directly into a horn in contrast with the perceived lesser volume of an accompanying instrument, but also that it appears that the group delay through the system may be corrected. This means that it is possible to perceive a spaciousness through or behind the noises which normally accompany an acoustic recording.

4 Requirements for listening

Many of the present noise-reducing procedures proposed and discussed [4] appear to remove the perception of spaciousness in early recordings, and hence it is imperative, either to use high quality originals or to use procedures which truly only attack the noise signals [5]. The gramophone (or phonograph in the case of cylinder recordings) must be of a high quality, as must be the pickup and its arm, and the signal must be realised to be a MONO signal which should preferably be sent from one loudspeaker only (listening experience has shown that transmitting the same mono signal from several loudspeakers adds a layer of confusion).

5 Conclusions

The paper does not wish to impose a conclusion on the audience or readership - it is a question of agreeing on perceived phenomena. It is, however, hoped that the importance of the *sound* as a document will be realised, even when discussing a more complete experience than mere performance.

References

1 Miller, D.C., "The Science of Musical Sounds", The Macmillan Co., New York 1916, 1922.

[2] Stockham, Jr., T.G.: "Restoration of old acoustic recordings by means of digital signal processing", AES Preprint No. 831, Audio Engineering Society 41st Convention 1971 October 5-8.

[3] Brock-Nannestad, G.: "The Objective Basis for the Production of High Quality Transfers from Pre-1925 Sound Recordings", AES Preprint No. 4610, Audio Engineering Society 103rd Convention 1997 September 26-29, New York.

[4] Picaut, J., Valière, J.C., Simon, L.: "Subjective Comparisons of Various Noise Reduction Techniques", Proc. 2nd Int. Conf. on Acoustics and Musical Research CIARM 95, Ferrara 1995, pp. 389-394. The conclusion as to the spaciousness is the present author's, however.

[5] *vide* the paper Brock-Nannestad, G.: "'The Requestor Decides' - the Fundamental Ethical Issues When Dealing With Sound Recordings", this volume.

"The Requestor Decides" - the Fundamental Ethical Issues When Dealing With Sound Recordings

George Brock-Nannestad

The Initiative for AV Preservation, Denmark
Resedavej 40
DK-2820 Gentofte, Denmark

Abstract

Sound recordings as artifacts contain information of two kinds: intended content and ancillary information. Any transfer which is not a digital cloning causes both kinds of information to be modified in the process. In many cases the ancillary information is of a nature that may reveal the modifications, and attempts to compensate in order to re-create the ancillary information may have a beneficial influence on the reproduction of the intended content. There is a wide range of compensatory options available to the researcher or publisher of recorded material from the last 120 years, and they must be chosen wisely.

1 Introduction

Audiovisual media are unique in the storeable arts in that they represent functions of time and in that they are totally dependent on machines for their reproduction. Many of the considerations in the present brief text will apply equally to moving images; however sound recordings are the subject here.

All restorative and preservative activities regarding sound recordings are concerned with "making available", either in the historical present or in some future. Replay of an original carrier can only occur by means of suitable equipment, preferably of new design and materials, however very firmly guided by knowledge about the original recording conditions. Such knowledge-heavy situations occur in the *primary transfer* and in the *utilisation*. Any intervening digital copying or cloning for preservation requires no knowledge about matters other than digital technology. The *utilisation* is a very rich field for experimentation without any risk of losing historic material.

Ethics enter much more into the considerations of actions concerning the media than other storeable art, because the use of machines makes the manipulations much less transparent, even to the professional. As usual, *ethics* and *truth* are related. The risk lies in the presentation of the result as a general *truth*. Here the concept of *source criticism* is an essential tool.

2 Components related to sound recordings

The proper stimulus for the ear is sound [1], but the storage means is the medium, and sound is only created when the medium is presented to the apparatus or machine. This machine converts the signals of the medium to a sound. It should be remembered that Thomas A. Edison (and Charles Cros) did not invent sound recording in 1877 but the *reproducible sound*. Recording of an arbitrary sound was invented 20 years earlier by Léon Scott de Martinville. At any time, the recording machine, manufacturing of a record, and its replay on contemporary machines constituted *one* system for the recording and reproduction of sound.

When we find recordings today it is obvious that we cannot change anything at the *recording* stage or at the *manufacturing* stage for the recording, only at the reproduction stage. We may thereby create any number of other systems which may provide sound that is adapted to a particular purpose. We may choose to use old equipment for the reproduction, in order to obtain an impression of the sounds available to our predecessors, however thereby endangering the medium. We may choose to learn as much as possible about the recording and manufacturing processes of early years in order to counteract various distortions that influenced what the medium was made to hold [2]. Modern signal processing holds great potential for manipulating early sound recordings. However, any manipulation should be properly documented to enable proper later interpretation of the value of this manipulation.

One means of documentation is to keep track of what happens to the ancillary information in the various processes. The ancillary information is of a type which is not necessarily detected by the replay equipment when it is used for its intended purpose: to reproduce the intended content. The information may be certain low frequency vibration patterns (mechanical recordings), mains hum, or the bias frequency present on analogue magnetic recordings.

3 A digression on preservation of sound

Long term preservation of sounds is a difficult subject, because the medium is entirely dependent on a machine for its reproduction. Hence it would seem

more relevant to speak about long term preservation of *a sound system*. In many, many cases it is necessary to perform a transfer from one medium to another in order to preserve. One school of thought prefers a very robust medium adapted for the simplest of reproduction machines - machines that any culture may manufacture in a simple way if required to start from scratch [3]. Another school of thought prefers a digital medium with its inherent capability for cloning - however requiring human generation after human generation to provide high technology cloning facilities *ad infinitum*. The short-term advantages of using the latter approach are immense. A third school of thought would even accept physiological data reduction, thereby storing not a linear signal, but one which only presents sufficient stimulus to ears to simulate a high quality sound. The digital storage capacity will thereby far exceed the capability for digesting all the information by future generations, and indexing in order to increase retrievability becomes a major issue. Storage of digital sound data is no different from storing any other digital information; however, the time reference aspect has to be taken into account. The digital information must be presented at the correct cadence [4].

4 Stages in a chain of transfers

We only need to be concerned with the concept of *primary transfer* in connection with analogue recordings, because cloning will replace it when the source material is digital (there may be sampling rate conversions, but these are usually well-behaved). Primary transfer occurs when the original medium is subjected to a suitable machine in order to create either another time-function for a different medium or for creating a sound (or a signal adapted for analysis by machine). The fundamental requirements of the medium have to be observed, obviously (e.g. for a magnetic tape a suitable replay head in accordance with the track lay-out adjusted to correct azimuth (i.e. the azimuth of the signal of the tape, not absolute) and the correct speed of the tape). However, the equalization provided in the replay amplifier may influence the high frequency content or timbre of the signal considerably. The trained ear would be able to discern this where natural sounds were originally recorded, but in the case of electronic music other approaches must be taken [5]. In the absence of any comment from the composer the studio technician at the original composition/performance was interviewed, and the intended differences between e.g. filtered noise and the sound or ring modulators were brought out in the primary transfer.

For mechanical media problems occur which are different but perhaps even more difficult. The signal is presented as an undulating track created in the surface of the medium, and the geometric problems in correct tracing of these undulations are formidable. To this are added the elastic and plastic properties of the medium which causes the track to deform during mechanical replay. It is a tribute to the care with which our predecessors worked that the results of proper replay may recreate such life-like sounds.

Let us imagine an analogue medium with a signal in the form of continuous variations of some variable (vertical distance from the surface, Δr (a change of instantaneous radius of a disc), magnetic remanence of a tape, opacity of a film). This variation over a distance may be converted into a variation over time by translation or rotation of the medium or the replay point. The signal is thus converted into another form, typically electrical (whereas the earliest used the power of the reproducing turntable to generate vibrations directly which became sound when transmitted via air). Once the signal is in this form, it may be subjected to a wide variety of modifications before being fed to a power amplifier driving a loudspeaker or headphone. One type of modification is linear, consisting of changes in the frequency characteristics of the signal, preferably in accordance with a prescribed inversion of a change introduced at the time of recording onto the medium [6]. Other types of modification are non-linear, consisting in an expansion or compression of the signal in accordance with a defined standard [7], and further modifications will try to separate desired signals from undesired noise contributions. This is frequently followed by an artificial change of the reverberation of the signal, and hence these modifications merit a separate discussion (section 5).

Some processes ideally really only influence the ancillary signals and may be used to reduce the noise content without influencing the intended content. Such processes make use of the fact that at any one instant the intended content in one groove flank of a lateral *mechanical* (mono) recording is identical (or 180° to it, dependent on the phase convention) to the signal in the other groove flank, whereas the noise contributerd by minute grit particles is different. When a stereo pick-up is used, signals from both groove flanks are made available, and is possible to use selection criteria to give access only to that flank which does not (according to said criteria) contain a contribution of grit noise. For the time being, such processes are only available as analogue equipment [8], but the present author would invite cooperation to obtain digital versions which would be much more versatile.

5 "Embellishments"

This term is used to describe a phenomenon which is known from the commercial re-issue business. Many originals are only available in a form in which the ancillary information *click*, *crackle*, and *hiss* are

present. Although the trained human ear is very efficient at concentrating on the intended content, noise reduction is standard in the re-issue business. Dependent on the sophistication and setting of the equipment, a lifeless and dull sound may be the result. If, for instance a singing voice is in effect low-pass filtered at ca. 2 kHz, then the "singer's formant" (ca. 2.5-3.3 kHz) indicating mastery and brilliance in the voice will have been removed. The "remedy" in many cases has been to increase the amplification in the octave 800-1600 Hz and the addition of artificial reverberation, providing a powerful sound which is saleable to the general public. The provision of such transfers is not *per se* un-ethical, but presenting it as the "true sound" or "original recording" is definitely wrong. A scientific user has a chance of discarding such transfers as sources of lesser value [9], but the general consumer is not openly presented with this choice.

6 What is the responsibility involved?

In general, it must be stated that the more the "manipulator" has performed a digestion of the intended content, the greater the responsibility that the result is correct according to the stated criteria.

In the times when the technology was primitive, it was only possible to reduce background noise by low-pass filtering, and this is the way that many early recordings were re-issued until ca. 1971 [10]. It is possible to reverse this "manipulation" by simple inverse filtering, and hence the responsibility is less than in the case of more recent but irreversible processes. Hence we have introduced the concept of *reversibility*. This is term which is much more useful in the context of signal processing than in the world of physical restoration, because true reversibility is obtainable under certain provisions.

The reversibility in practice comes back to the S/N ratio of the medium used for holding the manipulated version. If the manipulation has reduced certain signal components by more than 60dB it is not realistic to expect that they may be reconstructed faithfully (in view of the ca. 90 dB dynamic range available from present-day 16 bit systems). This is true, even if the equipment does not contribute any self-noise. If the attack transients have been removed from piano or harpsichord renderings, there is no way that they may be reconstructed, except in the case where only time delay or phase manipulations have occured. The energy will not have been removed, only the timing of its components.

Even in the case of totally irresponsible dealings with our recorded heritage we must remember that even a faint trace of a sound document may be better than no document at all!

7 Conclusions

The provocative title identifying "the Requestor" attempts to distinguish the end-user (the proper requestor) from the "manipulator", i.e. the person controlling the modification of the signal retrieved from a medium. However, in practice, outside of a purely scientific environment, the end-user will have to accept what is presented to him, and only after a growing suspicion followed by a thorough study will he be able to protest and request less manipulated transfers via the normal commercial channels. One may hope that future interactive distribution formats will provide such less manipulated transfers *and* a choice of suggested manipulations that the user may select in the *utilisation* situation.

References

[1] However, MD, DCC (format which is now defunct) and DAB all make use of physiological properties of ears in order to provide not sound, but *simulated sound* at a much reduced information bandwidth (albeit with full analogue bandwidth and dynamic range!) which is sufficient for all direct listening purposes.

[2] It has been previously pointed out that the process of generation of a *commercial* recording is controlled by feedback, whereas the generation of a *scientific* recording is a documented, possibly calibrated process: Brock-Nannestad, G.: "Applying the Concept of Operational Conservation Theory To Problems of Audio Restoration and Archiving Practice", AES Preprint No. 4612, Audio Engineering Society 103rd Convention 1997 September 26-29, New York.

[3] This also covers the situation where it has been established that the *original* medium is very robust (shellac pressings, good vinyl pressings, metal precursors to commercial pressings, certain magnetic tapes, certain films) so that *no primary transfer* takes place for preservation purposes. The problem is then one of replay of the *original* medium.

[4] In simple soundcards this is obtained by changing the sample rate during replay. In the imagined case of the signal electronics of a CD-player being re-constructed as a breadboard, using only low-scale integrated circuits, the physical distances would be such that the transmission time inside the replay circuits would not permit the high data rates required for reproduction of the original. Hence a subsequent step of data rate change would be required on the digital signals obtained.

[5] Scaldaferri. N. "La conservazione della musica per nastro magnetico: un problema tecnico e musicologico", Proc. 2nd Int. Conf. on Acoustics and Musical Research CIARM 95. Ferrara 1995. pp. 395-400.

[6] Also known as pre-emphasis and de-emphasis. These transfer functions are simple first-order functions. The much more severe second-order functions contributed by the acoustical recording process are properly dealt with by means of signal processing.

[7] Such systems were previously known as companders. and may be exemplified by Dolby-B (which however operates on a split frequency spectrum) or Dolby SR which divides the spectrum into a large number of channels which are individually treated.

[8] The earliest equipment of this type which is still in production is the Packburn Audio Noise Suppressor. by Packburn Electronics Inc.. Dewitt. New York. U.S.A.

[9] The concept of source criticism for sound recordings was introduced by the present author in Brock-Nannestad. G: "Zur Entwicklung einer Quellenkritik bei Schallplattenaufnahmen", 'MUSICA', Vol. 35, No. 1. pp. 76-81 (January. 1981).

[10] Stockham. Jr., T.G.: "Restoration of old acoustic recordings by means of digital signal processing", AES Preprint No. 831. Audio Engineering Society 41st Convention 1971 October 5-8. This paper deals predominantly with the spectral imbalance, however it marks the entry of the digital computer into this field.

WAVELET BASED DECLACKER OF MUSICAL RECORDINGS

Alvaro Tuzman, Sergio Chialanza and Eduardo Pena.

Instituto de Ingeniería Eléctrica, Universidad de la República,
Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay.
tuzman@iie.edu.uy

Abstract

We present a wavelet-based algorithm for restoration of musical signal recordings. While this algorithm is suitable for most types of impulsive noise degradation, the present work studies its behavior when optimized as a declacker. We design the detector based in the local properties of the music signal and of the noise. The detection stage of our algorithm consists in the decomposition of the signal via a wavelet transform followed by a decision stage. Both the analyzing basis and the thresholding operation are locally designed. A cost function is devised to obtain the best projection basis. This cost function involves a nonlinear constraint minimization problem. This task is translated into another nonlinear minimization problem of much lower computationally complexity. We compare the performance of the new detector with one that uses a generic wavelet (D_4) to build the projection basis, showing the superior performance of the local approach.

1. Introduction

In the past few years we have witnessed an increased interest in the restoration of musical recordings. There are two main reasons for this new situation: on one hand the new production and marketing strategies of the record labels, and on the other hand the new processing systems, which provides better and less expensive restoration techniques for widespread use. We can verify the above not only by looking at the academic literature, but mostly in the large quantity of commercially available restored recordings ([1], [2]).

As it is well known, there are several types of distortions that may take place in a musical signal. Among all possible signal problems, clacks present one of the most challenging ones. Broadly speaking, we can define clacks as impulsive noise of very short amplitude (compare to the musical signal) and very short temporal support. Impulsive noise appears in many other disciplines, and several methods have been proposed through the year to deal with it. While some of these methods are effective when applied to musical signals, there are some others that are not. The former includes methods based in different flavors of linear predictors and derivative operators. The latter include methods developed for statistical outliers, among others ([3], [4]).

Since a good characterization of impulsive noise in musical signals is not available, most methods relay on general considerations of the signal and/or the noise. As usual, and in the best case, general methods trade robustness for performance. For the problem at hand, researchers have proposed methods that are very effective with impulsive noise of large amplitude, but whose performance drops considerably when the amplitude is very small compared to the musical signal. In other words, when applied to musical signals, these methods are a good solution to clicks or scratches, but their performance shows a substantial decrease when applied as a declacker (see e.g. the work of Vaseghi and Frayling-Cork [5]).

In order to achieve a better performance, we believe that it is very important to exploit the local characteristics of the signals. Based in this idea is that we decided to base our algorithm in a wavelet transform (WT) decomposition of the signal. This transform is introduced in section 2 of the paper. As it is well known, the WT is especially well suited to deal with transients and non-stationary signals. In previous work, other researchers used wavelets as detectors [6]. Valière *et al.* have used the WT to detect Clicks and scratches in musical recordings [7]. Kronland-Martinet *et al.* among others, used the WT to analyze sound patterns [8]. In most cases however, the choice of the particular basis used in the decomposition had no relationship to the type of signal under analysis. Arguing that the analysis stage of the system should be local, our algorithm designs a new basis for different segments of the musical signal. The objective is to find in each case a basis such that the projection of the signal and the projection of the noise are best differentiated. In order to choose the best basis, the strategy followed was to design a cost function that reflects the performance of the detector. The problem becomes that of finding the basis that minimizes the cost. This is described in section 3. The use of dynamic thresholds is another important aspect of the algorithm. Its inclusion yielded a big improvement in performance. In section 4 we developed a closed form equation to determine the thresholds as a function of the local characteristics of the signal. Once the detection is performed, the support of the clack is estimated for each occurrence (section 5). Linear interpolators are then used to replace the missing samples. The length and direction of the interpolation are adjusted in order to use the maximum amount of information available. When it is possible, forward and backward interpolation is used.

The length of both predictors is adjusted independently (section 6). Finally, a short discussion of the results is presented.

2. The Wavelet Transform (WT)

The Wavelet Transform decomposes the signal into different scales. The construction of the basis is built by adding dilations and translations of a certain mother function. The mother function is defined by a two-scale difference equation. Regarding decomposition, it was shown that any square integrable signal $f(t)$ can be represented in terms of translates and dilates of a single wavelet $W(t)$ as:

$$f(t) = \sum_{j,k} 2^{j/2} f_{j,k} W(2^j t-k),$$

where $f_{j,k}$ is the inner product of $f(t)$ with the (j,k) basis function defined by $2^{j/2}W(2^j t-k)$. The sum $\sum_{j<J,k} 2^{j/2} f_{j,k} W(2^j t-k)$ gives an approximation of $f(t)$ up to scale 2^J, and it can be shown that equates to a lowpass filtered version of $f(t)$ (general texts on multiresolution and wavelet transforms include [9] and [10]).

There are many mother functions that fulfill the necessary conditions that make the above expansion span L_2, and in general, a particular choice of wavelet is used. It has been shown, however, that when the number of coefficients to be used in the decomposition is fixed, a proper choice of the basis can minimize the error in the span. In other words, the approximation properties of the decomposition depend on the analyzing basis. This is quite reasonable, since the spaces spanned by the different basis are different. After the wavelet conditions are fulfilled there are a number of degrees of freedom left that can be used to design a specific wavelet for each application. There are several design strategies, which range from the numerical minimization of the set of equations to the introduction of new types of basis. For example, Tewfik et al. [11] presented a discussion of this subject in the case of a given deterministic signal. Other researchers have addressed this issue from a statistical point of view. Coifman et al. have also addressed the problem in a different context [12]. Recently, a number of papers appeared with applications that exploit this local character [10]. These techniques generalize the wavelet transform and suggest the use of more general bases. While the use of more powerful representations is very appealing from a theoretical point of view, it usually implies a very large computational cost. At any rate, we should keep in mind that the main reasons to use the wavelet transform in the application at hand are its local character in time and scale, the possibility to build filters with good regularity properties (which lead to bases with good approximation characteristics), and its reasonably low computing cost (depending on the actual strategy chosen).

3. Wavelet Design

As mentioned, there are several approaches available for wavelet design. The first one is based in optimality with respect to a certain cost function, while the second approach chooses between all wavelets with maximum number of vanishing moments. As we shall see, the last approach is equivalent to choosing the phase of the filters in the filter bank.

An orthogonal wavelet basis of is defined imposing certain conditions on a set of K coefficients. The usual conditions imposed are: the dilation equation, normalization, orthogonality, and regularity. If we look for wavelets whose coefficients minimize certain cost function, the problem becomes one of nonlinear constraint optimization.

As specified by Tewfik et al. [11], this search over the constrained sequences $\{c_k\}$, k=0,...,K-1, is equivalent to an unconstrained search over $\{\psi_k\}$, $k=0,...,(K/2)-1$, with $0 \le \psi_k \le 2\Pi$. Let $\Psi = [\psi_1, \psi_2, ..., \psi_{K/2-1}]$ be the vector of trigonometric coefficients that determine the wavelet and Ψ_i the value of Ψ for the ith heartbeat. The wavelet design is performed based in the above parameterization. By design we mean, of course, to locally fix the trigonometric parameters $\{\psi_k\}$. We define a cost as a function of the wavelet coefficients, the actual local data, and certain error criteria. This function defines a surface $S_i = f(\Psi_i, x_i)$, where x_i is the data that defines the *ith* heartbeat. Our objective for each heartbeat, is to find Ψ_i, that yields a minimum of S_i. In other words, we want to solve

$$\Psi_i = min_{\Psi} COST(x_i, f(\Psi_i, x_i)),$$

for certain cost function *COST*.

In order to be able to perform this task we still have to define the error criteria. It was found that a good choice of such a criteria is to choose the wavelet that minimizes the energy of the transform, subject to a constraint in the number of wavelet coefficient to be kept.

Then, the problem at hand becomes:

$$\Psi_i = min_{\Psi} \{ x_i - WT_{\Psi}^{-1}(cod[WT_{\Psi}(x_i)]) \}^2,$$

subject to $N = \Sigma number\ of\ projection\ coefficients\ kept$

where WT is the wavelet transform using the wavelet Ψ_i, , x_i is the WT of a neibourhood of the clack, and N is the given number of wavelet projection coefficients to be kept. The numerical minimization is done by exhaustive search followed by a gradient type algorithm.

It is well known that the usual conditions that lead to orthonormality are not enough to construct wavelets that perform well for different tasks. An important condition usually included in the design in useful wavelets is that

the wavelet coefficients decay to zero as fast as possible, i.e., to have a high degree of regularity. General regularity characterization includes estimation of Sobolev (spectral approach) or Hölder (time domain limit) indexes and are based on properties of the scaling function $\phi(t)$. Given the relationship between $\psi(t)$ and $\phi(t)$, it can be shown that both functions share the same regularity properties. Based on intuition we could argue that it is convenient to have well-behaved filters in the filter bank, if we want to obtain well-behaved outputs, and as it turns out, this is actually the case.

Figure 1 below shows an example of the influence of the wavelet on the detection performance. In this caption, two clacks are detected by one wavelet (middle caption) while no clacks are detected by the second wavelet (lower caption). In this example, the second wavelet was chosen as D_4.

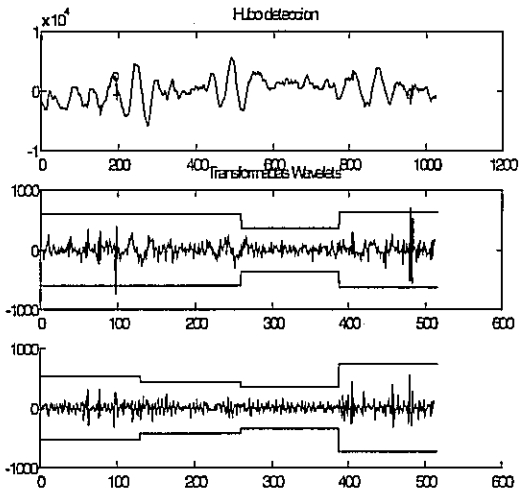


Fig. 1. Comparison of wavelet performance. The top figure is the original signal, with clacks. The middle figure show the threshold detector with the locally designed wavelet, while the lower figure show the detector with the D_4 wavelet.

4. Dynamic Threshold

An important issue in the design of the clack detector is that of the detector thresholds. We will argue here that the provision for dynamic thresholds, locally designed, is of utmost importance for reasonable performance of the system. As it was mentioned before, one of the main reasons of the difficulty involved in clack detection is the small amplitude of the distortion, and the variability of this amplitude. Another reason is the dynamic nature of the music signal, where the usual assumptions of stationary cannot be made.

Figure 2 below shows an example of the operation of the dynamic threshold. It is clear that a fixed threshold would yield a much lower performance.

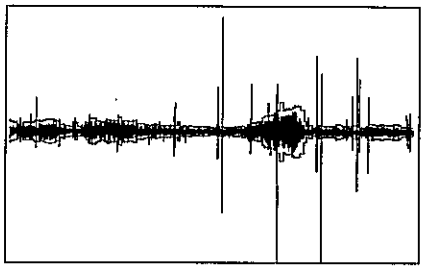


Fig. 2. This caption shows the output of the detector with the dynamic threshold superimposed.

It should be clear from the above discussion that the thresholds on which the detector operates should be dynamic, and local to the musical signal. We compute then the threshold value for small windows around each point in the signal. The value for the musical segment $\underline{x}_i$ is computed by the following expression:

$$TH(\underline{x}_i,b) = \alpha . (\| (WT(\underline{x}_i),b) \|^2 + \varepsilon)^{\beta},$$

where: b is an exclusion band used to compute the energy of the transform, α is an amplitude weighting factor, ε is an offset set for the energy, and β is a compressor factor.

5. Clack Detection and Support Estimation

After the above discussion, we are finally in a situation to perform the clack detection operation. The musical signal is divided in segments of equal length $\underline{x}_i$. For each such segment, the best wavelet is designed by minimizing the function COST described earlier. Then the signal is decomposed in the wavelet designed. At the same time, the dynamic thresholds are calculated. The thresholds are applied in the transformed domain and the detection is performed. Note that in a working system, the final selection of the threshold parameters should be of easy access to the technician, as their values represent a compromise between the probability of error and the probability of missing an event.

An important step in the algorithm is that of estimating the temporal support of the clack. This is an important step, which involves different possibilities, such as multiple clacks. Special care should be taken in order to avoid classifying good samples as part of the clack, since valuable information could be thrown away. We implemented an ad-hoc solution that decides between possible scenarios.

6. Signal Reconstruction

In our system, the signal is reconstructed using linear interpolators in the time domain (as opposed to the transform domain). This step is also local to the signal in order to maximize the information available and hence, the performance. The length of the interpolator is

parameterized and the parameters are designed for each clack occurrence.

7. Results and Discussion

The algorithm presented wqas tested against our own implementations of linear filtering based methods and showed much improved performance. The nonlinear character (from the bases design point of view) of our technique gives potential advantage that should be the subject of further study.

A most important aspect of the algorithm is the low computational complexity. The bases design as well as the detection and the interpolation are very inexpensive to implement. This is particularly true when compared with techniques such as best basis or matching pursuit.

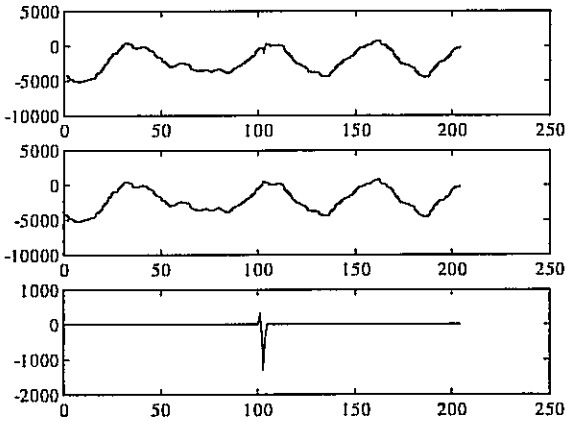


Fig. 3. Top caption: original signal with clack. Middle caption: restored signal. Bottom caption: difference between the original and the restored signal.

The algorithms described above were implemented in MATLAB on a Windows based system. The algorithms proved to work robustly and were applied to several short pieces of musical signals. The main drawback of our implementation is that it is not possible at the time to adapt the parameters in real time, as it should be done in a real application.

Ackowledgements

We thank Ing. Rafael Abal, from Sondor S.A., for supplying us with musical recordings.

References

[1] G. Wiener, "On the Acetate Groove", MIX Magazine, December 1995.

[2] Gramophone Magazine, Historical Recordings Section, General Gramophone Publications, London.

[3] S. Ofanidis, "Optimum Signal Processing", 2nd. Edition, MacMillan, New York, 1988.

[4] V. Poor, "An Introduction to Signal Detection and Estimation", Springer-Verlag, New York, 1988.

[5] S.V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings", J. of the Audio Engineering Society, Vol.40, No.10, Oct. 1992.

[6] F. Tuteur, "Wavelet Transformations in Signal Detection", Proc. Int. Conf. ASSP, pp. 1435-1948, Apr. 1988.

[7] J.C. Valière, "La restauation des enregistrements anciens par traitement numérique", Ph.D. Dissertation, Université du Maine, 1991.

[8] R. Kronland-Martinet, J. Morlet and A. Grossmann, "Analysis of Sound Patterns through Wavelet Transforms", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 1, No. 2, 1987.

[9] G. Kaiser, "A Friendly Guide to Wavelets", Birkhäuser, Boston-Basel-Berlin, 1994.

[10] M. Vetterli and J. Kovacevic, "Wavelets and Subband Coding", Prentice-Hall, New Jersey, 1995.

[11] A. Tewfik, D. Sinha, and P. Jorgensen, "On the Optimal Choice of a Wavelet for Signal Representation", IEEE Trans. on Information Theory, Vol. 38, pp. 747-765, March 1992.

[12] R.R. Coifman and M.V. Wickerhauser, "Entropy Based Algorithms for Best Basis Selection", IEEE Trans. on Information Theory, Vol. 38, March 1992.

Substitution-Oriented Digital Audio Document Restoration and Editing

yeeOn lo and dan hitt

e-mail: acoustic@netcom.com

Abstract. We present a software-based method to perform a certain class of digital audio document restoration and editing (DAREs). The method is local and substitution-oriented (SO). It has a rather complex analysis component and a relatively simple synthesis component. The analysis attempts quantitative source separation with the intervention of a musically trained human operator and with the assumption that waveforms for all the interacting components are available. The requirement of human-machine interactivity, *i.e.*, as opposed to automation, is pragmatic but has strong implications for the software design. We offer methods for dealing with the quantitative aspects of the analysis and a software architecture to accomplish the complex sequence of operations. Our experience allows us to conclude that a) the methods of quantitation are perceptually adequate for high quality music editing and b) the performance enhancement using the software architecture introduced in this paper is significant where fault instances are numerous, human monitoring is crucial, and cost minimization is desired.

1 Motivation. By an audio document we refer to a file of audio data in some known representation. For instance, we might have a discrete waveform of 16-bit samples derived from the original of a vinyl record of Gieseking playing Debussy's *Hommage à Rameau.*

Suppose that some imperfections render the recording unsuitable for further release. What we have is a problem of audio document restoration (ADR).

The authors have had some peripheral but nevertheless relevant experience in dealing with this problem in conjunction with making compositional changes to an existing performance of piano music on recording as well as removing flaws in computer-generated music, for a CD production. We have designed and written efficient software which solved these problems successfully. It is fairly obvious that the same method, although not necessarily the same software, can be used as a basis for the restoration of the Gieseking recordings.

We believe that the ADR methods and software described in this paper are useful to recording engineers as well as modern composers who need better control over the music they create.

2 Model. For simplicity, we refer to an audio document as data.

Conceptually ADR amounts to altering data, subject to a constraint imposed by a model or reference which is not available in the same form as the data[1].

By assuming that a transformation from the reference form to the form of the data exists and calling, for clarity, the image of this transformation an ideal or ID, we can then think of ADR as a process of altering the data with the aim of bringing the difference between the ID and the data itself to within some threshold of tolerance. One might view ADR as *altering data* according to a *reference.*

In reality, either because we have insufficient knowledge about the transformation, i.e., its details, or because the transformation is too complex to implement or the task too daunting to carry out, the ID is not synthesized or computed.

What is computed in ADR is an approximation or PROXI, starting with the data and modifying it using substitution or some other means, with the reference guiding it.

To judge the result of this computation, we need some way to measure success. Normally for audio problems, we expect to use our ears to determine the distance between the synthesized and the target, or ideal. This distance must fall within some threshold of tolerance; therefore typically we need to iterate the computation.

Since in reality the ID does not exist, we seek alternative criteria. The criteria we found manifest themselves as a series of continuity tests with respect to an array of perceptual dimensions within the altered data itself[2]. Bringing to within the threshold the distance between the ID and the PROXI becomes, ideally, a logical 'and'-ing of bringing to within threshold distances on perceptual dimensions projected from the two vectors.

For example, in the afore-mentioned Gieseking recording of Debussy's *Hommage à Rameau*, a target of correction might be a phrase beginning at location x. Assuming we have the piano waveforms, either from synthesis, recording, or extraction, then continuity in tempo, tunings, and tone quality, etc., across the boundaries of substitution

[1] If it were, then we would have no need for ADR since we could use the model directly.

[2] With SO-DARE, waveform continuity is a constraint automated into the software.

will ensure that the synthesized replacement is close to the ID in the neighborhood surrounding location x.

Sophistication might actually call for *predicting* the articulation of the phrase, along the dimensions of note-dynamics, -duration, and -onset, as well as the manner of attack, sustain, release, etc.[3], on the basis of how the music is being articulated before and/or possibly after the location of substitution, much like video is synthesized[4].

3 Methods. We first describe the approach, then outline the methods.

3.1 Substitution - a local approach. A simple way to remove faults in data is via some form of global filtering. For instance, low-pass filtering or moving average applied to a waveform can "wash out" spots and speckles, or pops and clicks. Yet it also degrades the signal. For instance, the brilliance of a tone might be "washed out" along with the pops and clicks. This is the case because algorithms of this type is essentially not target-specific, i.e., they are ignorant of what is signal and what is not: they provide a global treatment not sensitive to different kinds of signals.

As a result, the treated data typically sounds dull, or lifeless. An alternative to this approach is to make the treatment target-specific. In other words, this approach analyzes the data into signal and errors and directs treatments only to errors.

Substitution-oriented techniques generally provide higher-quality restoration/editing. This is so because they are target-specific and therefore the alteration is local. Local alteration means that it can avoid compromises with competing minimization demands from other parts of the data, especially those which are actually not intended targets in the first place, as in the case of a global solution.

On the other hands, substitution-oriented techniques require more assumptions and resources than a global solution. For example, we need to have all the materials available for any replacement involved in the substitution. We also need to know where each component begins and ends as well as its strength. In other words, we need some sort of analysis on the data to identify all the parameters necessary for the substitution.

3.2 Outline. The method at the core level is actually rudimentary. Conceptually and for a first explanation, at each location with an error between the data and the model, we *take out* the undesirable element $\mathcal{U}$ and replace it with something else, say, $\mathcal{V}$. $\mathcal{U}$ and $\mathcal{V}$ may have different lengths, but they must both satisfy the same smoothness condition at the boundaries[5].

In the simplest form, *taking out* actually involves applying an amplitude envelope to the location. Substitution involves superposing the substituents onto the location, using an inverse amplitude envelope in a cross-fading fashion. The sequence of operations is performed as steps of a waveform mixing operation.

Space constraints motivate the desire to apply the mix operations locally. Consequently, the overall SO-DARE operation involves

1. cutting out a block containing the fault;
2. *taking out* the fault $\mathcal{U}$ (from the block);
3. cross-fading the desirable $\mathcal{V}$ onto the remainder (of the block), maintaining invariance on the boundaries and making sure the the boundaries are in the interior of the block;
4. pasting the block (which now consists of the superposition of the remainder of the block and the desired substitution) back to the original document.

3.3 Details. It would be instructive to describe the method by first considering the case of a single voice: $\{A, B, C\}$, i.e., event A is followed by B, which is in turn followed by C.

We assume, without loss of generality, that A, B, and C are single notes although they could just as well be chords, i.e., chord A is followed by chord B, which is in turn followed by chord C.

Typically, the waveforms that represent the sound events A, B, and C overlap one another.

Now suppose that B is the event that is our editing target, i.e., we desire to replace B by something-else, as indicated in our model or reference. Since the waveform of B overlaps with its neighbors, the editing process involves dealing with the waveforms of A and C as well.

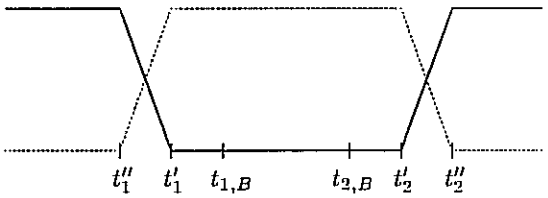
For clarity and emphasis, we say that $\{A, B, C\}$ is to be replaced by $\{A, B', C\}$ where B' can differ from B in pitch and duration.

[3] Some of these parameters are actually complex dimensions which require further decomposition for meaningful comparison.

[4] In fact, there are indications that acoustical instrument tones, including some speech sounds, can also be synthesized this way with significant data reduction [4], [5], [6].

[5] In polyphony, one must make sure that corresponding durational changes in concurrent voices are acceptable and specify how they should be changed beforehand; otherwise heuristics will be used to adjust the lengths and onsets of the notes or audio events.

Let's say B begins at time $t_{1,B}$ and finishes at $t_{2,B}$. We first construct an envelope in the shape of an inverted, or "notched" trapezoid, and apply it to the waveform which contains $\{A, B, C\}$ (in order to remove the waveform of B)[6]. Next we use waveform mixing to string together $\{A, B', C\}$, keeping in mind that the elements in the sequence generally overlap either as a consequence of the composition's intent or as a consequence of the acoustics. Then, we construct a "peg"-shaped envelope which is the exact complement of the "notched" trapezoid window, and apply it to $\{A, B', C\}$. Finally, we superpose the two waveforms together at the place where the notched and peg envelopes exactly match up:



The breakpoints of the envelopes t_1'', t_1', t_2', t_2'', satisfy the conditions:

$$t_1'' < t_1' < t_{1,B} < t_{2,B} < t_2' < t_2''.$$

The case of multiple voice is manifested by the model configuration of $\{A, B, C\}$ superposed on D where D can be another voice or a superposition of two or more voices that are concurrent with B, and may possibly extended beyond A and C. For the purpose of substitution, we need to have that portion of D that interacts with B and its surroundings. Consequently, we may not need to know when D begins or ends if it extends beyond the 'notched' window of substitution. We just need to know its dynamic levels within the window, the determination of which is discussed in the next section under analysis and identification of substitution parameters.

3.4 Analysis and synthesis. Unlike any global method which gives each unit of data the same treatment, there is an analysis and a synthesis aspect to a local method such as our substitution-oriented approach. Analysis is necessary because we want to find out how the current unit of data is different from the ID so as to try to match the best correction possible to the data. Synthesis is necessary because we must supply the "match" between the data and the ID, or the best "match" for the difference between them. Naturally, we assume that the discrepancy or difference between the ID and the data is not uniform. And that is why we need to do unit-by-unit or window-by-window analysis and synthesis.

For synthesis, the tool involves waveform mixing and time-varying amplitude scaling to ensure all transitions in a polyphony are smooth and all components mix at the correct level at every perceptible moment.

Furthermore, because of the interactive nature of analysis using the human ear, the same synthesis tool necessary for generating the replacement for substitution is also needed for analysis. And it is used even more extensively there than for replacement synthesis because we often need to try out a succession of waveforms in order to ascertain the best estimate of the parameters. Therefore, a capable, well-designed waveform mixing tool is essential to the success of large-scale SO-DARE operations[3].

For analysis, we need to identify the components as well as ascertain the parameters which define these components. It is a much more complicated task than synthesis because any DARE involving polyphony requires at least an implicit notion of source separation which the computer cannot do very well at the moment. As a result, a key component for SO-DARE is human-machine interaction.

Analysis, in so far as DARE is concerned, has two fundamental dimensions: timing and (scalar) amplitude. In other words, we need to know, in a piece of polyphony, where a sound event begins and ends. We also need to know the strengths of the interacting sound events: not just the event to be substituted in, but also its neighbors and those which are present concurrently.

For timing, we have two strategies: 1) cross-correlate the waveform under repair (WUR) with the waveform that coressponds to the note in question and look for a sharply defined peak. 2) If such an indicator is not readily available, then we can resort to human intervention: We listen for a double attack as we "slide" the waveform of the component note across the WUR; when the double attack vanishes, we have gotten the sliding note aligned with its image in the WUR, to within our perceptual limit of resolution. In our example involving $\{A, B, C\}$, we need to ascertain the onset timings for A and C, as well as B.

These methods do not provide great precision for end time estimation. Fortunately, the ear is also less sensitive to the exact termination under polyphonic conditions. In certain situations, such as one involving a sequence of notes played on the same instrument, the attack of the next note often determines the end of the one under consideration provided we allow for a fixed amount of overlap, due to acoustical conditions.

The other dimension pertains to amplitudes. Here we propose a method involving listening for unnatural dynamic behavior in a superposition given by the form

$$f(t) = \lambda(1 - h(t))Y(t) + h(t)X(t)$$

where Y represents a WUR, X, a waveform for the com-

[6]The problem of removing B when there are one or more voices superposed on it is discussed under the case of multiple voices below.

ponent in Y whose relative strength in Y we want to discover, and $h(t)$, any non-constant time function extended over the course of X. For example, assuming a domain of the closed unit interval $[0, 1]$, $h(t)$ could be the (continuous) piecewise linear function which starts with a value of 1 for $t = 0$, declines to 0 at $t = 0.3$, and remains at 0 for the rest of the interval:

$$h(t) = \begin{cases} 1 - \frac{t}{0.3} & \text{if } 0 \leq t \leq 0.3 \\ 0 & \text{if } 0.3 < t \leq 1 \end{cases}$$

The λ for which the superposed waveform $f(t)$ is not abnormal in the attack (or otherwise) is the correct scalar to use for synthesis of the replacement.

4 Implementation Issues. For the type of DARE where fault instances are numerous, where human monitoring is crucial, and where costs minimization is desired, it is important to design the software in such a way as to streamline the process. This can be accomplished via a collection of interacting GUI-based object-oriented computer programs including a sound file playback program with precise timing information ready for display and modification; a waveform mixing program; a graphic function editing program; and an efficient and cautious block substitution program, all of which are programmed to "know" each other and each other's data structures. These can be arranged in a *polyhedral* architecture [8][7][2].

The applications perform the requisite tasks of fault identification, instance location and other parametric quantitation, substituent preparation, interpolation, substitution, and audio-visual-tactile monitoring and control for fault identification, parametric quantitation, and verification.

For variable-length substitution, the audio document could be optimally represented internally as a linked list of waveform segments together with onset and length information. On the other hand, for search purposes, an array data structure is best when the user can supply precise sample indices to the computer program as a result of an interactive monitoring effort.

A structure which has reasonable performance for both search and insertion is an AVL-tree[1][9]. More exactly, both search and insertion are $\log(n)$ operations in time (so they are not quite as good as constant time operations, but very nearly that good because of the very slow growth of $\log(n)$). In fact, AVL-trees are how some modern operating systems solve the problem of providing processes with the illusion of a contiguous address space even though the physical memory a process has access to is actually composed of fragments spread across the RAM. The operations typically involved in memory management in a modern OS and the space-time issues arising from them are analogous to those that face SO-DAREs.

5 Conclusion. From actual experience, we can conclude that the methods proposed for deducing the timing and amplitude information are perceptually accurate enough for high quality musical editing. The software implemented according to the architecture described above performs adequately and eliminates unnecessary wait-arounds.

References

[1] Adelson-Velskii, G. M. and Landis, E. M., *Doklady Akademia Nauk SSSR*, **146**, (1962), 263-66.

[2] Hitt, D. & Lo, Y., "An Alternative Digital Environment for Music Synthesis", *Proceedings of the Delphi Computer Music Conference/Festival*, 1992.

[3] Hitt, D. & Lo, Y., "New Facilities for Digital Audio Mixing", to appear in *Proceedings of 1998 AES Convention in San Francisco*, 1998.

[4] Lo, Y., "A Technique for Timbre Interpolation", *Proceedings of the ICMC*, 1986.

[5] Lo, Y., *Toward a Theory of Timbre*, Stanford U. Technical Report STAN-M-42, 1988.

[6] Lo, Y. & Hitt, D., "Uniform Treatment of Sounds and their Syntheses on Digital Computers", *International Workshop on Models and Representations of Musical Signals (Capri)*, 1992.

[7] Lo, Y. & Hitt, D., "Modern Synthesis and Sampled Sound", *Proceedings of Synaesthetica '94*, 1994.

[8] Pinson, Lewis J., and Wiener, Richard S., *Objective-C: Object-Oriented Programming Techniques*, Addison-Wesley, 1991.

[9] Wirth, Niklaus, *Algorithms + Data Structures = Programs*, Prentice-Hall, 1976

Saturday 26th **h. 14.00**

RESTORATION OF AUDIO DOCUMENTS II (SPECIAL SESSION)

"Audiorestauro": a Digital System for Audio Signal Restoration

L. Bazzanella,
Centro di Sonologia Computazionale,
Universitá di Padova
049/8276981 - lb@csc1.unipd.it

G.B. Debiasi
Dipartimento di Elettronica ed Informatica
Universitá di Padova
049/8277675 - debiasi@ibm.net

Abstract

The matter concerning the restoration of audio recordings can be divided in three stages for practical purposes: removing bursting noises (so called "clicks"), background noise reduction , removing the amount of distortion caused by the recording system. There are many software applications which can solve the above mentioned problems.

Available solutions try to reconcile contrasting necessities and they often bring one to accept compromises with the aim of providing the best global solutions; moreover, one should also keep in mind that procedures should be strongly automatized.

"Audiorestauro" is meant to solve the above mentioned problems separately, by adopting interactive data processing methods which allow the user to optimize the choice of operations to carry out according to different necessities.

Introduction

Audio signal recordings on cylinders or gramophone recordings, optical or magnetic film soundtracks on tape recordings or similar recording means turn out to be affected, more or less heavily, by distortion and noises which deteriorate the "quality" of the recorded signals compared to the original sources. This variety of situations and the drawbacks one has to deal with require us to give a more detailed definition of the word "restoration" usually used in these situations. It means every intervention aiming at "improving the quality" of the processed audio signals from the an auditive point of view. This implies that it is necessary to increase the signal-noise ratio, paying attention not to introduce further distortion. In filtering side effects are fairly common; this implies timbre modification and transient extension. In addition to this even when this phenomena are of small importance, the processing method is likely to modify the spectral peculiarities of residual noise; this outcome can be even worse for a hearer. On the contrary, even though a certain amount of noise still remains, this is not likely to be a problem if the noise cannot be heard, thanks to the masking effect by the sound components.

1. Overview of restoration methods

The matter concerning the restoration of audio recordings can be divided in three stages for practical purposes: removing bursting noises (so called "clicks"), background noise reduction , removing the amount of distortion caused by the recording system. In order to deal effectively with bursting noises time-domain processing is used; it consists in automatical detection of the noise and subsequent removal by replacing lost samples. Detection techniques make use of linear models [1] [2], spectral characteristics [3] and neural nets [4]. Interpolation of damaged data can be implemented in various ways: with linear prediction [1] [2], wavelet decomposition [3] or neural nets [5]. The problem concerning the removal of background noise has been studied in depth as far as voice signals are concerned and then it has successfully been adopted in the field of music. A general overview of several classic filtering methods can be found in [6]; other more elaborated methods eliminate the noise from the short-time spectrum of the signal which has to been cleaned [7], or they can make use of the probability function that at a given frequency there is only noise: this probability is estimated by means of an instantaneous maximum likelihood calculation [8] or by observing a series of subsequent short-time intervals [9] . Another technique which implements signal recovery by mimizing the mean-square error [10] turned out to be particularly useful to avoid "musical noise" [11]. As for the problem of eliminating the distortion caused by the recording system, it is mathematically solved by deconvolution of the desired signal from the impulsive response of the distorting system. When the impulsive response is known, the solution is not difficult. In real situations, nevertheless, this rarely happens. A method was however proposed [12] to carry out this operation, which makes use of a particular transformation which calculates spectral complex logarithms [13].

2. Removal of bursting noises in "Audiorestauro"

Bursting noises are commonly known as "clicks" (caused for example by a speck of dust on the surface of a LP) or "scratches" (they last longer, they are caused by scratches). They are caused by the damaging of the recording means; the acoustic effect is short and very relevant perceptively; hence the definition of bursting noise. The general procedure to repair this damage can be divided in two stages: the first is detection, which consists in recognizing the impulse over the desired signal; the second is removing the clicks which have been found. In this way, the original sound is restored in the most faithful way.

2.1 Detection of bursting noises

As we have already pointed out, the first step is detecting the instants which feature bursting noises. Since old recordings have plenty of them, this operation has to be carried out automatically, insofar as manual detection by the restorer would be too long and toilsome. In "Audiorestauro" click detection is carried out on the bases of the following principle. An acoustic signal, which does not contain noisy characteristics, can be adequately modeled by means of a linear

prediction system; However, this model is not suited for impulsive signals.
Let be:

$$y_n = x_n + d_n$$

the value of a sample n of the sound disturbed by a click and digitally acquired; let be x_n the corresponding value of a noise free sample and let be d_n the noise contribution. Let us consider as for x_n modeling with an order p linear predictor:

$$xn = \sum_{k=1}^{p} a_k x_{n-k} + e_n$$

where the a_k are the predictor's coefficients and e_n the non-disturbed excitation sequence. If we transform y_n into its corresponding disturbed excitation signal with reverse filtering :

$$vn = yn - \sum_{k=1}^{p} \hat{a}_k y_{n-k} =$$

$$= x_n + d_n - \sum_{k=1}^{p} \left(a_k - \tilde{a}_k\right)\left(x_{n-k} + d_{n-k}\right)$$

with $\hat{a}_k$ estimation of coefficients and $\tilde{a}_k$ deviations from real values, we find a signal which is made up mainly of the noise sequence d_n. Since sounds can be considered stationary only for short time intervals, it is clear that the filter shall be adaptive, i.e. it will need periodical recalculation. In "Audiorestauro" the autoregressive model is recalculated at regular intervals by hypothesizing stationarity over intervals of ten ms or so. Windows of analysis partially overlap; the purpose is to avoid clicks which are somewhat in between two windows not to be correctly detected and subsequently removed. The autocorrelation function is calculated once for each window; thanks to this the linear system is calculated with the proper recursive relation. The same parameter evaluation algorithm allows one to evaluate also the expected variance of the excitation signal. Click isolation is then effected by means of an ordinary threshold detector, calculated with reference to the estimated variance for non-disturbed signals. The proceeding described above works well, but it is not always effective: in those cases in which the window of analysis in use contains an uncorrupted fragment of a sound, the expected value of the variance generally results smaller; as a consequence, the detecting threshold can become so low so as to generate a certain amount of false alarms. This is to be avoided, because a false alarm causes a "good" fragment to be replaced with an artificial one, obtained by interpolation. The device which "Audiorestauro" uses to solve this problem is that of inserting, before calculating autocorrelation, a fake click in a known position; its amplitude is half the maximum amplitude of the samples in the same window. Obviously, after calculating the filter this device is removed, so as to enable data processing. In this way the detector is made not to be too sensitive; at the same time the quality of the linear model is kept on a high standard. Since one has to deal with a considerable variety of sounds, including the noisy ones of film soundtracks, it is important for the user to interact with the system during the stage of automatic detection. Some waves which are detected as clicks might not be caused by corruption; on the contrary they may be one of the constituents of the original signals (special effects, percussive sounds, noise of falling rain, etc.), so their removal or modification is to be avoided. In "Audiorestauro" it is possible to modify the saved list of detected impulses. The detected clicks are scanned, they are displayed on the screen and it is possible to listen to them; the user can then decide whether they are to be removed or not.

2.2 Removal of bursting noises

Only after removing background noises, according to the method which will be explained further on, is it possible to remove bursting noises. At this stage the final list of detected impulses is analyzed, so that they are removed one by one. Since one click lasts more than the delta function , it is necessary to replace not only the data corresponding to the detected position, but also a certain amount of neighboring samples in order to have it completely removed. An interval having a click in its middle is taken into consideration. Through interpolation this bulk of data is replaced; it is therefore important that it accurately corresponds to the actual size of the damaged portion of the signal. In this way one does not have to process portions of signals which do not need it. Here, too, a linear model for the sample is used, but this time it has a higher degree of complexity, so that a more accurate interpolation can be effected. While scrolling the list this model is updated at regular intervals, so that it can fit the signal variations. The interpolation method adopted is based on the hypothesis that the sequence to be calculated may be considered realization of a autoregressive model. In general this hypothesis is correctly verified as far as the production of voice or musical signals is concerned. The algorithm therefore applies to the above mentioned instances. The samples to be processed consist in a sequence x_n with n=0,1,...,L-1; the lost ones because of bursting noises fall in correspondence with the known instants t(1), t(2),..., t(m) considered in increasing order. The problem is that of evaluating the unknown samples $x_{t(1)}$, $x_{t(2)}$,..., $x_{t(m)}$ and the parameters of the linear model from the acquired data, so that the interpolated fragment minimizes the mean-square error. As for the order p of the model, there are proper methods to evaluate it, but it is also possible to achieve good results by choosing it on the basis of the length of the fragments which are to be interpolated. The above mentioned minimization is a very complex problem, because it involves fourth degree terms. A suboptimal solution can be achieved if one takes into consideration that the number m of the unknown samples is usually smaller in comparison with the length L of the fragment. It is therefore possible to choose an iterative method: firstly an arbitrary evaluation of the fragment

to interpolate is chosen (for instance, each sample corresponds to zero); then, through minimization, the coefficient of the autoregressive model are obtained. Now minimization can be repeated, and a more plausible evaluation of the fragment taken into consideration can be obtained. Generally, iteration converges very quickly and it is stopped as soon as the solution achieves the desired stability within the boundaries of a fixed tolerance. It should be pointed out that the aim is not only interpolating samples, but also evaluating the linear system which allows to do this. As for the algorithm employed, the model is obtained through autocorrelation applied to noisy data and subsequently the Levinson-Durbin recursion. Since click density in the windows of analysis is usually fairly low, it is normal to expect that ergodic autocorrelation is not very sensitive to them. Experimental results have verified this hypothesis and they have produced high precision evaluations, together with the advantage given by the low level of computational complexity of the Levinson-Durbin.

3. Removal of background noise

Background noise removal is likely to generate a disturbing side effect, because it is never possible to know exactly the spectral power of the noise; one can only get to know a time averaged evaluation. The actual evolution of this variable is characterized by random fluctuation around this average value; a certain amount of noise then is not removed by the restoration proceeding: it consists of the peaks of the spectral density of the noise which randomly exceed the threshold of the time averaged evaluation. In time domain this means that sinusoid-like components appear and they are then perceived as pure tones (this phenomena is also known as "musical noise"); they have random frequencies and short duration. To a listener, this kind of corruption in the signal can be even more disturbing than the original noise. The method implemented in "Audiorestauro", which is able to minimize the appearance of this phenomena, falls within the so called STFA (Short Time Spectral Attenuation) i.e. the group of algorithms based on short-time spectrum calculation of the signal which is to be restored and on the subsequent attenuation of those components which are more seriously corrupted by noise. Practical realization of background noise reduction is not very difficult but for two aspects: system gain calculation and the calculation of the average spectral power of the disturbing noise. As for G calculation there are algorithms which give excellent practical results, even though they are not very good from a mathematical point of view. As for the evaluation of the average spectral power of the disturbing noise, it is worth noticing that the effectiveness of any filtering algorithm depends on the possibility to evaluate the disturbing signal correctly. It is often possible to know the disturbing signal by making use of the muted zones in the signal which is to be restored. In "Audiorestauro" it is the restorer's task to specify interactively which are the samples to be taken into consideration to evaluate the noise. If there are no muted zones available, an evaluation system has been achieved; it is based on the assumption that the desired signal should have slowly varying spectral characteristics, whereas noisy signal should not behave like this. Another characteristics is the possibility the user has to manually adjust the estimated spectral power envelope.

4. Experimental results

So as to experiment the restoration of a soundtrack we used the audio track of a documentary by Resnais and Hessens entitled "Guernica"; the track includes music and voices both of man and women; therefore it offers a fairly wide range of sounds having different characteristics. We only removed background noise; the few clicks in recording were not relevant, so that filtering was enough to improve the quality. Fig. 1 and fig. 2 show the spectral analysis of a short extract from the signal before and after being processed.

Fig. 3 and fig. 4 show the temporal evolutions concerning the restoration proceeding of a 78 rpm LP having a scratch.

5. Conclusions

As we have already explained, "Audiorestauro" allows the user to have both a high degree of automation in the bursting noise detection proceeding and the possibility to keep everything under control: the user can decide which bursting noises are to be removed and the background noise can be adjusted to fit the required needs. Experimental tests proved that the method employed is effective and we can state that "Audiorestauro" provides the user with an efficient tool for audio signal restoration.

References

[1] S.V. Vaseghi, P.J.W. Rayner "Detection and suppression of impulsive noise in speech communication systems" Proc. IEEE, vol. 137, pt. 1, n.1, pp. 38-46, Feb 1990.

[2] S.V. Vaseghi, R. Frayling-Cork "Restoration of old gramophone recordings" J.A.E.S., vol. 40, n.10, pp. 791-801, Oct. 1992.

[3] S. Montresor, J.C. Valiere, J.F. Allard, M. Baudry "The restoration of old recordings by means of digital techniques" A.E.S. Preprint 2915 (G4), Montreux 1990.

[4] A. Czyzewski, C. Supron "Learning algorithms for the cancellation of old recordings noise" A.E.S. Preprint 3847 (P11.7), Amsterdam 1994.

[5] A. Czyzewski "Artificial intelligence-based processing of old audio recordings" A.E.S. Preprint 3885 (F6), San Francisco 1994.

[6] J.S. Lim, A.V. Oppenheim "Enhancement and bandwidth compression of noisy speech" Proc. IEEE, vol. 67, n.12, pp. 1586-1604, Dec. 1979.

[7] S.F. Boll "Suppression of acoustic noise in speech using spectral subtraction" IEEE Trans. A.S.S.P., vol. 27, n.2, pp. 113-120, Apr. 1979.

[8] R.J. McAulay, M.L. Malpass "Speech enhancement using a soft-decision noise suppression filter" IEEE Trans. A.S.S.P., vol. 28, n.2, pp. 137-145, Apr. 1980.
[9] O. Cappé "Enhancement of musical signals degraded by background noise, using long-term behavior of the short-term spectral components" Proc. IEEE I.C.A.S.S.P., pp. I.217-I.220, 1993.
[10] Y. Ephraim, D. Malah "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator" IEEE Trans. A.S.S.P., vol. 32, n.6, pp. 1109-1121, Dec. 1984.
[11] O. Cappé "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor" IEEE Trans. A.S.S.P., vol. 2, n.2, pp. 345-349, Apr. 1994.

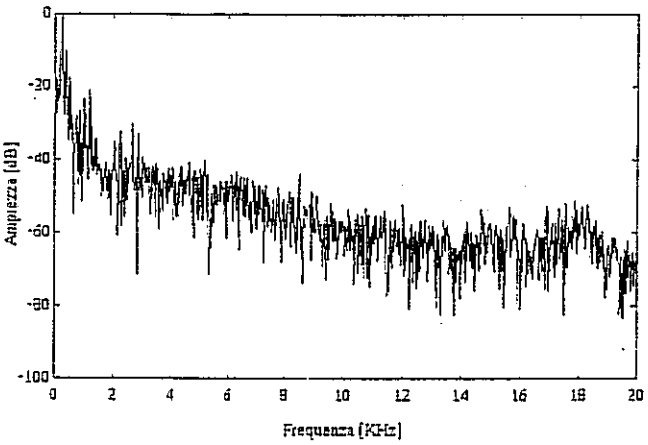


Fig. 1

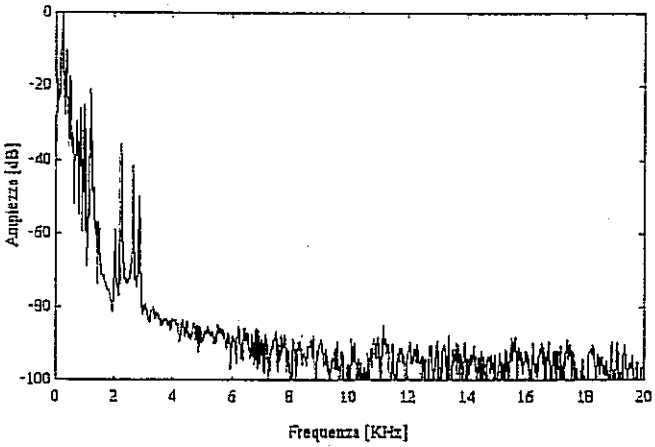


Fig. 2

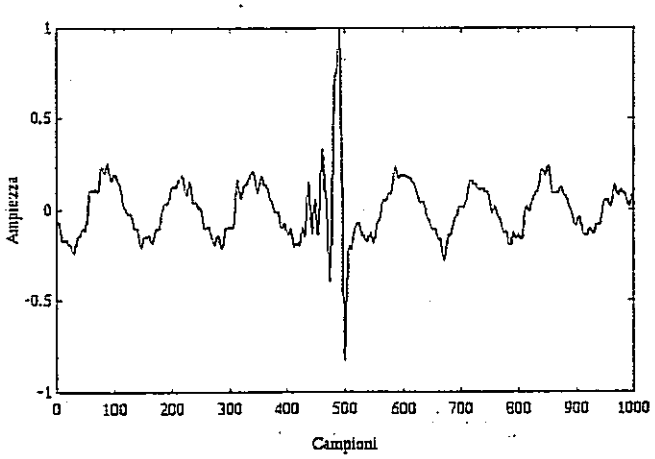


Fig. 3

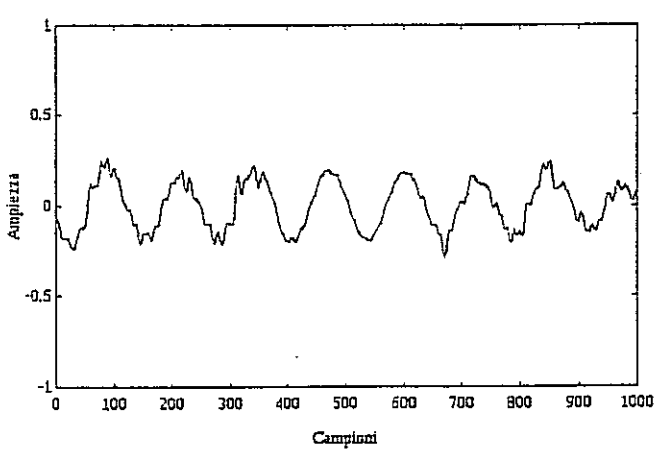


Fig. 4

Issues on training of operators in the field of restoration of audio documents

Canazza Sergio, De Poli Giovanni, Mian Gian Antonio, Vidolin Alvise
Dipartimento di Elettronica e Informatica
Università di Padova - Via Gradenigo 6a – 35100 Padova – Italy
{canazza, depoli, mian, vidolin}@dei.unipd.it.

Abstract

In order to achieve a good knowledge in the field of audio restoration, it is necessary to have both musical and scientific skills. In fact, it will be pointed out that it is not sufficient to give a simple skill of use commercial software for denoise and declick; this has to be combined with the knowledge of all the aspects related to the handling of audio recording media. For this reason an expert in audio restoration is required to have autonomy in the use, adaptation, and development of audio restoring tools.

1 Introduzione

The cultural heritage of musical recordings is slowly fading away since the physical media, where the audio signal information is stored, are exposed to deterioration. In this scenario, it's getting more and more important to devise some techniques aimed at digital audio restoration. The evolution of audio digital storage technology, leading to the development of the Compact Disc and the Digital Audio Tape, yielded high quality recordings. But this potential quality can not be reached through a simple sampling of audio data stored in old media like wax cylinders, vinyl records (78rpm, LP, 45rpm, EP, and so on) [1] or magnetics tapes. In fact all this media are exposed to physical degradation thus compromising the quality of the final product. Causes of degrade can have local influence (as scratches and dust) or global influence (as white noise, wows and flutters, and distortions). Restoration is required in any case in which the noise due to deterioration compromises the artistic value of the recording. The ideal goal of the restoration should be the *perfect* reconstruction of the original audio signal, as it was received by the transducer (i.e. a microphone) during the original recording.

Digital signal processing theory developed very sophisticated techniques for noise reduction. But, as in the case of visual arts restoration, the knowledge of practical techniques have to be combined with a great artistic sensibility: an expert audio restorer should have an interdisciplinary culture, both scientific and humanistic, in order to use the more suitable techniques and to judge the results. The audio restoration field lacks of a specific apprenticeship which could prepare a skilled expert able to operate with precision and independence. In fact restoration projects are usually developed in humanistic departments, with a consequent inadequate knowledge about the facilities offered by the new technologies. In this presentation the attention will be focused on the competences required an operator in this field, with particular attention to the scientific expertise. The schedule of an hypothetical course in audio restoration will be presented in detail. Particular attention will be paid to subjects that are essential for the comprehension of the digital techniques for audio restoration [2], [3], [4], [5], and [6]. It will be pointed out that it is not sufficient to give a simple skill to use commercial software, this has to be combined with the knowledge of all the aspects related to the handling of audio recording media. A suitable background on chemical composition of the media will be required, in order to reduce as possible further deterioration. Considering that the actual trend is to store audio information in big digital archives, in which audio information is catalogued, an expert in this field should also have the competencies to manage a multimedia archive. These competencies should include the storage and retrieval of the information, with particular attention to archives of audio documents.

A balanced scientific training is required to allow the student to face the audio restoration in all aspects. For this reason we supply a list of the necessary class programs with suitable basic theory.

A preliminary hypothesis about the didactic organization of a course on audio restoration technologies takes into account a subdivision in two levels of formation: post-laurea and post-diploma. Depending on the level, the course will respectively create expert researchers and technicians, skilled on the safeguard and the valorization of the acoustic and musical heritage and with a particular knowledge in the field of restoration and audio archives maintenance.

This specialistic formation implies the institution of classes on the following subjects, to which practical laboratory experiences are added:

- Elements of calculus
- Computer science laboratory
- Musical acoustics and psychoacoustics
- Electroacoustics and sound signal manipulation
- Methodologies and techniques of storage
- Sound restoration techniques
- Architectonic acoustics measurement techniques
- Acoustic virtual reality techniques
- Cataloguing, archiving, and fruition techniques for multimedia material

2 Class programs

ELEMENTS OF CALCULUS

Scope:

Audio restoration is strictly related to digital signal processing. This requires a knowledge on basics mathematics and physics, usually taught in scientific university faculties. This class will introduce the students to these concepts and techniques, in order to give the basic knowledge to understand the concepts of sound processing techniques. Starting from sets theory concepts, the definition of function will be given, together with a number of polynomial algebra concepts. Students will deal also with concepts on algebra of matrices and trigonometry. Some fundamental instruments will be explained, like the logarithmic and the exponential functions, and the polynomial series, together with the derivative and the integration operands for continuos functions. The concepts will be exemplified using a number of short Matlab® scripts.

Program:

- Classes of numbers
- Variables and functions
- Polynomials
- Vectors and matrices
- Exponential e logarithmic functions
- Trigonometric functions
- Polynomial series
- Derivative and integration operands

COMPUTER SCIENCE I

Scope:

Audio restoration, in its practical applications, is achieved through the use of computer programs. Since each degradation implies different restoration problems, sometimes it can be asked to the restorer to develop new digital processing algorithms. This asks for an adequate knowledge of computer science techniques. To this end the class will be teach, at first, the computer architecture and its hardware and software components. Concepts on operating systems will be explained; in particular these concepts will be exemplified with the use of Ms-Dos, Windows95, Mac-OS operating systems. The computer arithmetic will be introduced. Starting from the definition of algorithm, a number of linguistic concepts for the creation of algorithms will be explained. In particular conditional and iterative constructs will be highlighted. The definition of multidimensional array will be given in parallel with the definitions of vector and matrix taught in the Elements of Calculus class. A number of algorithms will be developed in C language.

Program:

- Computer architecture
- Operating system: files, directories, and disks
- Binary system. Algebra of Boole.
- Algorithms
- Data types
- Expressions and operands
- Control structures: conditional and iterative constructs
- Data structures
- Functions
- Word processing – Compiler - Linker
- Standard libraries in C
- Examples

COMPUTER SCIENCE II

Scope:

The working of computer system for audio will be explained. Si ripercorrerà idealmente l'evoluzione dei linguaggi di programmazione che ha portato all'approccio *object-oriented.* The concepts of class, encapsulation, inheritance, polymorphism, and overriding will be illustrated. Implementation in C++ language of algorithms illustrated during the *electroacoustics and sound signal manipulation* course will be made.

The methods of information retrieval and their presentation by hypertext will be studied.

Program:

- Digital sound storage system
- Computer systems for audio
- Class
- Encapsulation
- Inheritance
- Polymorphism
- Design of a system for digital audio-signal processing
- The hypertext
- The WEB
- The HTML language
- Information retrieval on the net

MUSICAL ACOUSTICS AND PSYCHOACOUSTICS I

Scope:

In order to correctly judge the various possibilities offered by the different restoration techniques, it is fundamental to have a skilled knowledge about the physics of sound production and about the human perception of sound and music. Basic concepts of acoustics will be explained to the students, after a brief introduction on kinematics and dynamics laws regarding the movement of a point in the space. The fundamental notions of psychoacoustics will be explained in detail.

Program:

- Kinematics of the point
- Relative motion
- Dynamics of the point
- Friction

- Impulse, work and energy
- Undulatory motion: elastic waves
- Reflection and refraction
- Sound wave characteristics
- Sound wave speed
- Sound attributes
- Vibration in acoustic tubes

MUSICAL ACOUSTICS AND PSYCHOACOUSTICS II

Program:
- Acoustic waves properties.
- Auditory system: primary sensations
- Sound pitch: first order effects; critical band
- Second order effects. Auditory system structure
- Sounds intensity; isophonic curves. Masking. Intensity perception theories.
- Sounds timbre. Spatial effects.

METHODOLOGIES AND TECHNIQUES OF STORAGE

Scope:
The musical cultural heritage is physically vanishing, for instance because of the deterioration of the media containing audio information, but it runs also the risk to become unavailable, because of the fast evolution on storage technologies that make obsolete the old supports and hard to retrieve the analogic and digital information. To overcome these problems the subjects related to physical media preservation will be explained.

Program:
- Audio e video information: from physical media preservation to information retrieval
- Electronic memory and a new definition of the preservation
- History of audio storage of this century
- Technical problems of existing archives

ELECTROACOUSTICS AND SOUND SIGNAL MANIPULATION

Scope:
Some concept of electromagnetic interaction will be presented. The use of microphones and analogic and digital tape recorders will be explained, in order to give to the students an adequate skill in the use of recording instruments. The concept of signal will be introduced, as well as the basic tools of signal manipulation like convolution and Fourier transform. Since the audio signal will be digitally processed, the concepts of sampling and interpolation will be explained in detail. All these concepts will be exemplified through the use of Matlab® scripts.

Program:
- Electrical interactions
- Circuits
- Magnetic interaction
- Undulatory motion: electromagnatic waves
- Systems for sound recording and sound reproducing
- New technologies for sound recording and transmission
- Signal
- Convolution
- Fourier transform
- Analysis in the time and frequency domains
- Signal transformations
- Filters
- Sampling theorem
- Signal reconstruction errors: signal to noise ratio
- Short Time Fourier Transform
-

AUDIO RESTORATION TECHNICS I

Scope:
To produce the necessary capabilities to complete audio restoration works. To overcome these problems the different audio deterioration will be explained. The best methods to removal audio noise will be illustrated (see [6] for references).

Program:
- The audio restoration problem.
- Local and global noise: impulsive noise, broadband noise, non linear distortions
- Audio signal modeling.
- Description of the best methods of declicking and denoising

AUDIO RESTORATION TECHNICS II

Scope:
The best professional software about audio restoration will be presented. In order to achieve a good knowledge in the field of audio restoration, noise reduction algorithms in Matlab® language will be explained and implemented. The audio restoration technologies will be applied in audio data stored in old media like wax cylinders, vinyl records (78rpm, LP, 45rpm, EP, and so on) or magnetics tapes [1], in order to study the particular characteristics.
The problematic of different audio signal types like speech, elettro-acoustics music, western classical music, and afro-american music, will be investigated.

Program:
- Professional commercial software
- Design of a software to audio restoration
- Examples of audio restoration of different media types
- Examples of audio restoration of different audio signal types (speech, music)

CATALOGUING, ARCHIVING, AND FRUITION TECHNIQUES FOR MULTIMEDIA MATERIAL

Scope:

To achieve the main techniques to administrate multimedia databases. To develop instruments in order to employ the databases from remote.

Program:

- History of audio media
- Computer system for cataloguing and archiving
- Criteria for a correct media preservation
- Commercial enterprise linked to databases spread
- Relational Databases
- Administration of multimedia databases
- Multimedia Databases on the WEB

3 Classes organization

For instance, a two years course is proposed, subdivided in semesters lasting 12 weeks. A final exam will test the students' knowledge acquired during the course. The exam will consist in a practical work, which is supposed to be developed in the school laboratory in the last 20 hours. Class organizations is quoted in the table 1.

Year	Semester	Class	Hours	
			Lecture	Lab
I year	I sem.	Elements of calculus	6	2
		Computer science I	1	3
		Musical acoustics and psychoacoustics I	2	2
	II sem.	Methodologies and techniques of storage	5	1
		Electroacoustics and sound signal manipulation	6	4
II year	I sem.	Musical acoustics and psychoacoustics II	2	2
		Computer science II	1	3
		Sound restoration techniques I	2	6
	II sem.	Cataloguing, archiviation, and fruition techniques for multimedia material	2	6
		Sound restoration techniques II	2	6

Tab 1: Example of class organizations.

Conclusions

In this paper has been pointed out the problem of training of operators in the field of restoration of audio documents. We have been focused on specialized training. However, the proposed class program is *open*, that is it allow also to achieve a good knowledge in the field of multimedia databases administrating and man-machine interfaces based on non-verbal communication.

Acknowledgments

This work was supported by *Consiglio Nazionale delle Ricerche* (CNR), under the project *Progetto finalizzato Beni Culturali*

References

[1] Godsill S., Rayner P., Cappé O. "Digital audio restoration". In *Applications of digital signal processing to audio and acoustics*. Kahrs – Karlheinz Brandeburg (ed.). Kluwer Academic Publishers. 1998

[2] Godsill S. J. "The restoration of degraded audio signals". *Ph.D. thesis* Cambridge University Eng. Department Cambridge, England, 1993.

[3] Schueller D. "The ethics of preservation, restoration and re—issues of historical sound". *J. Audio Eng. Soc.* 39(12). pp. 1014-1016. Dec 1991.

[4] Veldhuis R. "*Restoration of lost sample in digital signals*". Prentice-Hall, 1990

[5] Vaseghi S. V., Frayling—Cork R. "Restoration of old gramophone recordings" *J.Audio Eng. Soc.* 40(10). pp. 791-801. Oct. 1992.

[6] Cappé O. "Noise reduction techniques for the restoration of musical recordings", Ph.D. Thesis, Ecole Nationale Supérieure des Télécomunications, Paris, 1993.

PERFORMANCE OF THE EXTENDED KALMAN FILTER FOR RESTORATION OF AUDIO DOCUMENTS

G. De Poli, G.A. Mian, G. Re
Dipartimento di Elettronica e Informatica
Via Gradenigo 6/a, 35100 Padova (IT)

Abstract

The problem of removing impulsive and background (white or coloured) noise from audio recordings is considered. The algorithm used simultaneously solves the problems of wideband noise filtering, signal parameter tracking and impulsive noise elimination by using the Extended Kalman Filter theory (EKF), as proposed by M. Niedzwiecki and K. Cisowski [5, 6]. Results, obtained with the proposed method for significant cases, are presented. Moreover, features and performance of the method are compared with other existing techniques.

1 Introduction

The introduction of high quality digital media, combined with an increasing awareness of the historical importance of "audio heritage", has led to a growing requirement for the preservation and restoration of old recordings [7]. In this work we present some results on the restoration of magnetic tapes and vinyl records, carried out within the project " Beni Culturali " of the CNR [1], whose aim is the preservation and fruition of all Italian cultural assets.

The types of degradation common in audio sources can be broadly classified into localized and global degradations [3]. The former are finite duration defects which occour at random in the waveform and include clicks, scratches, clipping, ... (in the sequel they will be simply referenced to as "clicks"). The latter affect all the audio recording and include background noise (perceived as "hiss"), wow, flutter and some types of linear and nonlinear distortion.

In this context, we consider the problem of the reduction of impulsive and background noise from audio signals. This task is usually carried out using different methods for detection/restoration of impulsive noise and for broadband noise reduction [3, 8].

In this work we employ an algorithm whose objective is to simultaneously solve the problems of filtering/parameter tracking/elimination of the outliers ("clicks") by using the Extended Kalman Filter theory (EKF), as proposed by M. Niedzwiecki and K. Cisowski [4, 5, 6]. In particular the algorithm in [6] can be interpreted as the nonlinear combination of two Kalman filters: the first is used to follow the slow variations of the signal time–varying AR model parameters, while the second takes part in the reduction of background and impulsive noise.

2 Problem statement

Let the audio signal $s(t)$, $t = 1, 2, \cdots$, be modelled by a p order *time varying* autoregressive (AR) model

$$s(t+1) = \sum_{i=1}^{p} a_i(t)s(t-i+1) + e(t) \qquad (1)$$

driven by the gaussian zero–mean white noise sequence $e(t)$ with variance σ_e^2.
The time evolution of the time varying coefficients $a_i(t)$ is modelled by the random walk model

$$a_i(t+1) = a_i(t) + w_i(t) \qquad (2)$$

with $w_i(t)$ zero–mean gaussian white processes of variance σ_w^2 mutually uncorrelated, i.e., $E[w_i(t)w_j(t)] = 0$ for $i \neq j$, and independent of $e(t)$. Moreover, let us assume that the original signal $s(t)$ is corrupted by a mixture of a broadband noise $z(t)$ and impulsive noise $v(t)$ (independent of $e(t)$ and $w_i(t)$), so that the available signal $y(t)$ can be written as

$$y(t) = s(t) + z(t) + v(t). \qquad (3)$$

The noise $z(t)$ is assumed gaussian zero–mean white noise (see later for relaxing this hypothesis) of variance σ_z^2, while $v(t)$ is assumed gaussian zero-mean noise with $\sigma_v^2(t) = \infty$, if a click is present, or $\sigma_v^2(t) = 0$, otherwise. As a consequence, if a click is revealed at time t, the corresponding sample $y(t)$ must be discarded since it not bears information on $s(t)$ and $s(t)$ must be recovered from $\{\cdots, y(t-1), y(t+1), \cdots\}$.

In [5] it is shown that under the hypothesis made, the problem of recovering the signal $s(t)$ based on the noisy measurements $\mathbf{Y}(t) = \{y(t), y(t-1), \cdots, y(1)\}$ can be optimally handled by the extended Kalman filter (EKF). To this purpose it is convenient to represent signal $s(t)$ in eqn. (1) in the non–minimal state space form

$$\mathbf{s}_q(t+1) = \mathbf{A}_q[\mathbf{a}_p(t)]\mathbf{s}_q(t) + \mathbf{b}_q e(t) \qquad (4)$$

where $\mathbf{s}_q(t) = [s(t), ..., s(t-p), \cdots, s(t-q+1)]^T$, $q \geq p$, is the signal vector, $\mathbf{a}_p^T(t) = [a_1(t), \cdots, a_p(t)]^T$

is the vector of the AR model coefficients, $\mathbf{b}_q^T = [1, 0, \cdots, 0]$, and $\mathbf{A}_q(t)$ is the companion matrix associated with the extended parameter vector $\mathbf{a}_q^T(t) = [\mathbf{a}_p^T(t), \mathbf{0}_{q-p}^T]$. The provision of a nonminimal state-space description: $q > p$ will allow one for two-sided reconstruction of up to $q - p$ samples corrupted by impulsive noise.

Notice that to remove noise an accurate signal model is needed and to obtain a reliable signal model the signal should be noiseless. The problems of filtering and parameter tracking are strictly tied and are to be jointly solved. The solution to their combined treatment is obtained by combining the unknown AR model coefficients and the signal vector in a $p+q$ "state vector" $\mathbf{x}^T(t) = [\mathbf{s}_q^T(t), \mathbf{a}_p^T(t)]$ and by rewriting (1–3) as

$$\begin{cases} \mathbf{x}(t+1) = f[\mathbf{x}(t)] + \mathbf{u}(t) \\ y(t) = \mathbf{c}^T\mathbf{x}(t) + \zeta(t) \end{cases} \tag{5}$$

where

$$f[\mathbf{x}(t)] = \begin{bmatrix} \mathbf{A}_q(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \cdot \mathbf{x}(t), \quad \mathbf{u}(t) = \begin{bmatrix} \mathbf{b}_q e(t) \\ \mathbf{w}(t) \end{bmatrix}$$

with $\mathbf{w}^T(t) = [w_1(t), \cdots, w_p(t)]$ and

$$\zeta(t) = z(t) + v(t), \quad \mathbf{c}^T = [\mathbf{b}_q^T, 0^T] = [1, 0, \cdots, 0].$$

The problem of estimating the model parameters $\mathbf{a}_p(t)$ and the noisefree signal $\mathbf{s}(t)$ is reduced to a nonlinear filtering problem in the state space. A (suboptimal) solution to the problem can be based on the theory of extended Kalman filter (EKF)[5, 6] and is obtained linearizing (5).

Let us denote with $\hat{\mathbf{x}}(t|t)$ the estimate of the state at time t from the measurements $y(\tau) : \tau \leq t$ and with $\hat{\mathbf{x}}(t|t-1)$ the state prediction at time t from the measurements $y(\tau) : \tau \leq t-1$. Let $\mathbf{F}(t)$ denote the state transition matrix of the linearized system

$$\mathbf{F}(t) = \frac{\partial f[\mathbf{x}]}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\hat{\mathbf{x}}(t|t)} = \begin{bmatrix} \mathbf{A}_q(t|t) & \hat{\mathbf{s}}_p^T(t|t) \\ \mathbf{0}_{\mathbf{p}\times\mathbf{q}} & \mathbf{0}_{\mathbf{q}-\mathbf{1}\times\mathbf{p}} \\ & \mathbf{I}_p \end{bmatrix} \tag{6}$$

where $\hat{\mathbf{x}}^T(t|t) = [\hat{\mathbf{s}}_q^T(t|t), \hat{\mathbf{a}}_p^T(t|t)]$ is the filtered state trajectory given by the EKF algorithm, $\mathbf{A}_q(t|t) = \mathbf{A}_q[\hat{\mathbf{a}}_p(t|t)]$ and $\hat{\mathbf{s}}_p(t|t)$ is the vector made up with the first p components of $\hat{\mathbf{s}}_q(t|t)$. Moreover, let:

$$\Omega = cov[\mathbf{u}(t)]/\sigma_e^2 = \begin{bmatrix} \mathbf{b}_q\mathbf{b}_q^T & 0 \\ 0 & \xi\mathbf{I}_p \end{bmatrix} \tag{7}$$

with $\xi = \sigma_w^2/\sigma_e^2$.

The EKF equations become for the **prediction** step:

$$\begin{cases} \hat{\mathbf{x}}(t|t-1) & = f[\hat{\mathbf{x}}(t-1|t-1)] \\ \Sigma(t|t-1) & = \mathbf{F}(t-1)\Sigma(t-1|t-1)\mathbf{F}^T(t-1) + \Omega \end{cases} \tag{8}$$

and for the **update** step:

$$\begin{cases} \hat{\mathbf{x}}(t|t) & = \hat{\mathbf{x}}(t|t-1) + L(t)\varepsilon(t) \\ \Sigma(t|t) & = (I_{p+q} - L(t)\mathbf{c}^T)\Sigma(t|t-1) \end{cases} \tag{9}$$

where $\Sigma(t|t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))^T]$ is the state estimation error covariance and $\Sigma(t|t-1) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))^T]$ is the state prediction error covariance. Moreover in (9)

$$\varepsilon(t) = y(t) - \mathbf{c}^T\hat{\mathbf{x}}(t|t-1) = y(t) - \hat{s}(t|t-1)$$

is the *prediction error* (Kalman filter innovation) and $L(t)$ is the Kalman gain, whose value depends from the click indicator function $\hat{d}(t)$:

$$L(t) = \begin{cases} \frac{\Sigma(t|t-1)\mathbf{c}}{\mathbf{c}^T\Sigma(t|t-1)\mathbf{c}+k(t)} & if \quad \hat{d}(t) = 0 \\ 0 & if \quad \hat{d}(t) = 1 \end{cases}$$

and $k(t) = \sigma_z^2/\sigma_e^2(t)$.

The EKF can be started with the values

$$\hat{x}(0|0) = 0, \quad \Sigma(0|0) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \delta I_p \end{bmatrix}$$

with δ a large positive constant (~ 100) to account that nothing is known in advance about $\mathbf{a}(0)$.

The corresponding algorithm has a complexity $O((p+q)^2)$. In [6] a reduced complexity split EKF algorithm is presented and it was used in the actual experiments referred to in the next section.

In addition, it can be noticed that it is not difficult to drop the hypothesis of a white noise $z(t)$. In case of coloured noise $z(t)$ it suffices to model it as an AR process and to increase the state dimension accordingly [2]. Such a provision was found quite effective for the noise reduction of some old vinyl records.

2.1 Click detection

The detection of clicks is based on the value assumed at each t by the prediction error

$$\hat{d}(t) = \begin{cases} 0 & if \quad |\varepsilon(t)| \leq m\hat{\sigma}_\varepsilon(t) \\ 1 & if \quad |\varepsilon(t)| > m\hat{\sigma}_\varepsilon(t) \end{cases} \tag{10}$$

In (10)

$$\hat{\sigma}_\varepsilon^2(t) = \eta(t)\hat{\sigma}_e^2(t) \quad \text{with} \quad \eta(t) = \mathbf{c}^T\Sigma(t|t-1)\mathbf{c} + k(t)$$

is the estimated innovation variance, m is the parameter determining the threshold for impulsive noise detection (in practice $m = 3 \div 5$) and $\hat{\sigma}_e^2(t)$ is the *local* estimate of the model input noise $e(t)$ variance

$$\hat{\sigma}_e^2(t) = \begin{cases} \lambda\hat{\sigma}_e^2(t-1) + (1-\lambda)\frac{\varepsilon^2(t)}{\eta(t)} & if \quad \hat{d}(t) = 0 \\ \hat{\sigma}_e^2(t-1) & if \quad \hat{d}(t) = 1 \end{cases} \tag{11}$$

In (11) $0 < \lambda < 1$ determines the adaptation speed. In actual experimentation we used $\lambda = 0.98$, except in the case of signals with fast dynamics as in the guitar case, where a smaller value (0.7) was used.

Fig. 1 shows a segment taken from an old 78 rpm gramophone disc and the corresponding innovation ($p = 12$), which takes on greater values in corrispondence with signal discontinuities.

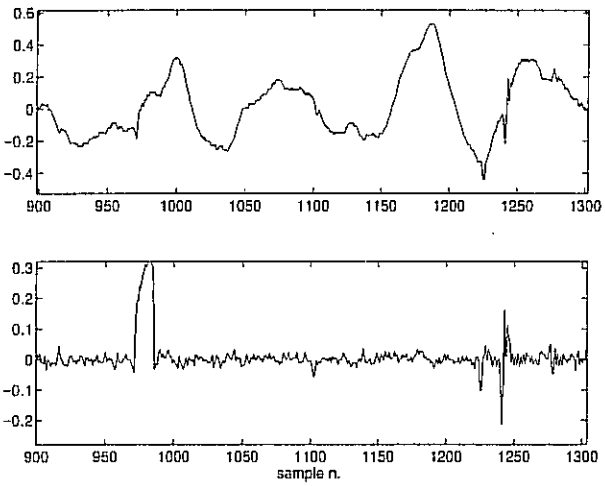


Figure 1: Click detection: noisy signal (top) and corresponding innovation (bottom).

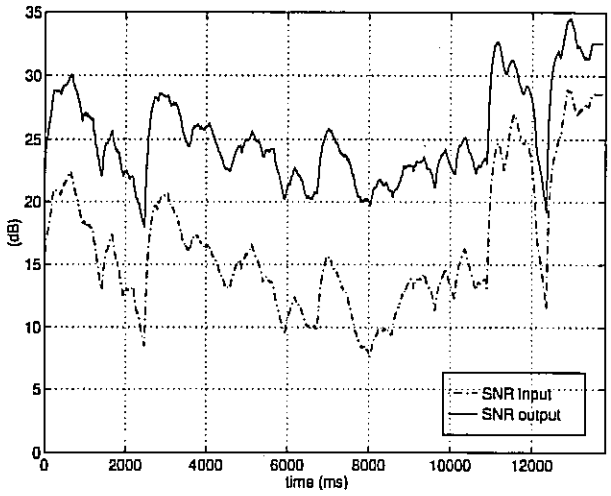


Figure 2: Example of segmental SNR improvement

2.2 Smoothing and reconstruction

It can be noticed that $\hat{\mathbf{s}}(t|t) = [\hat{s}(t|t), \cdots, \hat{s}(t - q + 1|t)]^T$ represents the optimal (mean square) smoothed estimate of $s(t), \cdots, s(t-q+1)$ given $\mathbf{Y}(t)$, i.e., all the measurements available up to time t. To make full use of the available information, it is convenient to use, at time t, $\hat{s}(t-q+1|t)$ as an estimate of $s(t - q + 1)$, i.e., it is convenient to introduce a delay of q samples.

As a result, for signal smoothing it is enough to use $q = p$. In presence of clicks it can be shown that, for a p order AR process, a block consisting of at least p "good" future successive samples is needed for good reconstruction [4]: for a group of n successive samples corrupted by a click, a value $q \geq p + n$ is required.

This consideration can be exploited to derive a variable order EKF [6], which usually uses $q = p$ and, in presence of clicks, increases q until the filter innovation corresponding to the "corrected " signal becomes "white" noise or q does not reach a predetermined threshold. Thus the length of the replaced signal is incremented until this condition is true. During the interpolation step the order of the filter is temporarily increased, in order to allow for a better estimation, and both past and "future" measurements are employed, so as to carry out a " forward - backward " interpolation. Such a provision offers a significant computational reduction over the use of a fixed large q value.

3 Experimental results

To evaluate the performance of the EKF algorithm it is necessary to identify noise on the input recording. As a preliminary step, we used computer generated noise (white/coloured and impulsive) and added it to some test CD quality recordings, supposed to be "noise free". This helped us to gain some insight into the method and confidence on the choice of parameters ξ, $k(0)$, m and p, which determine the ultimate performance of the algorithm.

The parameter $\xi = \sigma_w^2/\sigma_e^2$ (see (2) and (7)) should be chosen in accordance with the degree of nonstationarity of the signal at hand. We found a constant value $\xi \simeq 10^{-4}$ adequate in most examples, the most noticeable exception beeing an old Segovia excerpt. The fast guitar attacks required a greater value $\xi \simeq 10^{-2}$. In the future it is planned to use a time–varying value $\xi(t)$ for ξ.

The parameter $k(0)$ allows one to obtain an initial estimate of $\sigma_e^2(0) = \sigma_v^2/k(0)$ and to start the recursive estimation of $\hat{\sigma}_e^2(t)$ via (11) (σ_v^2 can be measured during silences). Its value was found not critical and in most cases we used $k(0) \simeq 2$.

As for parameter m, a small m value, say $2 \div 3$, allows one to detect small clicks but introduces many false alarms. This gives rise to the substitution of many samples that would be better dealt with by the EKF smoother. As a rule of thumb, we found preferable to use a high m value (i.e., $m = 4 \div 5$) and, in any case, to iterate the declicking process starting from, e.g., $m = 5$ and forcing a high $k(t)$ value during the first iteration/s to reduce smoothing effects accumulation.

To evaluate the white noise reduction performance, controlled amounts of "white" noise were added to "clean" recordings. The SNR_o of the output signal produced by the smoothing algorithm was measured and related to the SNR_i input signal. It was found that equation

$$SNR_o \simeq 12 + 0.8\, SNR_i \quad (dB)$$

well represents the measured values for $0 \leq SNR_i \leq 40$ dB and $p \geq 10$, i.e., the algorithm provides an average SNR improvement of about 10 dB.

Fig. 2 reports the segmental SNR_i and SNR_o *vs* time for a 20 dB overall SNR_i (the segmental SNRs were computed every 10 ms on a 20 ms window).

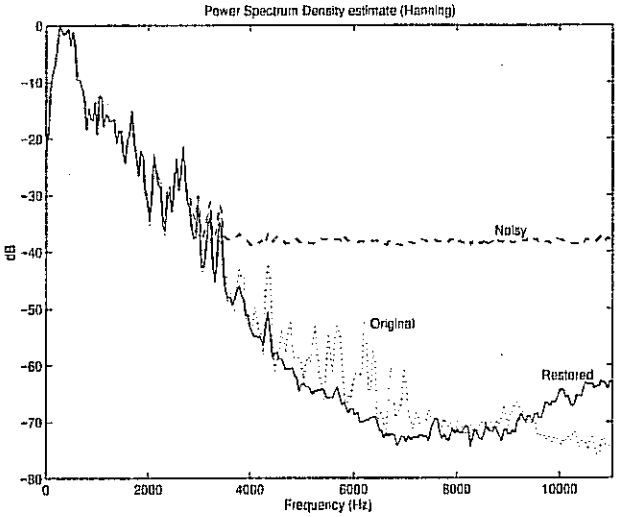


Figure 3: Power spectrum density estimate.

From the figure it is apparent that the SNR gain is approximately uniform (actually it is greater in lower SNR_i regions).
Fig. 3 gives (limitedly to $0 \div 11$ kHz) the (Welch) power spectrum estimate of a 5 s long Schubert piano piece taken from a CD, the one corresponding to its noisy version ($SNR_i = 20$ dB) and that corresponding to the restored version. From the figure it can be appreciated that the restored version spectrum strictly follows that of the original up to about 3 kHz, i.e., up to frequencies at which the white noise power density equals the one of the clean recording. In addition, differently from what would be obtained by a simple low-pass filter (with cutoff at 3 kHz) or by spectral subtraction, beyond 3 kHz the restored version "follows" the original spectrum. This property results to be perceptively important and appreciated by experienced listeners.

In order to evaluate the role of predictor order p, the algorithm was tested on artificially degraded recordings ($SNR_i = 20$ dB) with p varying between 2 and 30. The general conclusion was that the SNR gain ($SNR_o - SNR_i$) quickly increases up to $p = 8 \div 10$ and then it remains approximately constant. (The value $p = 12$ was used in all the prsented figures).

Finally, Fig. 4 shows a segment taken from a noisy piano recording (kindly supplied by S. Godsill) and its restored version. The proposed method seems to have dealth properly with clicks and wideband noise.

Acknowledgments

This work has been carried out with the finantial support of the C.N.R.-"Progetto Finalizzato Beni Culturali".

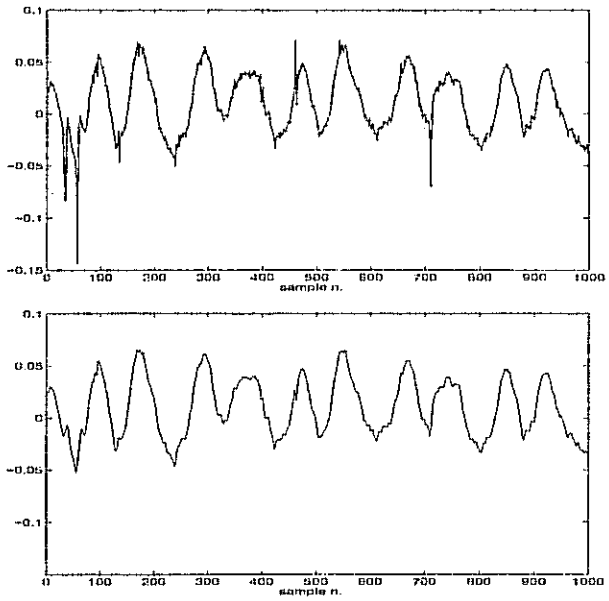


Figure 4: Segment of noisy audio (top) and its restored version (bottom).

References

[1] G. Adamo, G.B. Debiasi, G. De Poli, P. Gina, G.A. Mian, M.C. Sotgiu, A. Vidolin, "*Problemi di conservazione e restauro di archivi sonori*", in Atti Convegno A.I.A., Perugia, 1997, pp. 106-113.

[2] B.D.O.Anderson, J.B.Moore, "*Optimal filtering*", Prentice-Hall, 1979.

[3] S. Godsill, P. Rayner, O. Cappé, "*Digital audio restoration*", in M. Kahrs and K. Brandenburg (Eds), "Applications of digital signal processing to audio and acoustics", Kluwer, 1998.

[4] M. Niedźwiecki, "*Steady-state and parameter tracking proprieties of self-tuning minimum variance regulators*", Automatica, pp. 597-602, 1989.

[5] M. Niedźwiecki and K. Cisowski, "*Adaptive scheme for elimination od broadband noise and impulsive distubances from AR and ARMA signals*", IEEE Trans. Signal Processing, vol. 44, March 1996.

[6] M. Niedźwiecki, "*Identification of time-varying processes in the presence of measurement noise and outliers*", Proc. 11th IFAC Symp. System Identification, 1765–1768, Tokyo, 1997.

[7] D. Schueller, "The ethics of preservation, restoration and re-issues of historical sound", J. Audio Eng. Soc., 39, 12, pp. 1014-1016, Dec 1991.

[8] R.Veldhuis, "*Restoration of lost samples in digital signals*", Prentice-Hall, 1990

SOUND RECOVERY OF COMPUTER MUSIC WORKS PRODUCED WITH LOW SAMPLING RATES: THE CASE OF *Traiettoria*[1]

Marco Stroppa
Composer
11, Rue des Rosiers - F-93220 Gagny
Tel. ++33 1 43086064, Fax: 43091373
Email: marco@ircam.fr

Alvise Vidolin
Centro di Sonologia Computazionale
Università di Padova
Via San Francesco, 11 - I- 35121 Padova
Tel. ++39 49 8273757 - Fax: ++39 49 659975
Email: vidolin@dei.unipd.it

Abstract

This text will delve into the main problems we had to face while recovering the electronics of Traiettoria *by Marco Stroppa, a piece for piano and computer-generated sounds produced at the Centro di Sonologia Computazionale of the University of Padua between 1982 and 1984. This work, synthesized at a sampling rate of 15 kHz, has therefore an audio range limited to 7.5 kHz. In addition, since in 1988 it was analogically recorded onto a DAT tape, the new digital master contained also some defects due to the conversion. We will describe the various phases of the process of audio recovery and will focus particularly on the algorithm devised to restore brightness to the original signal by generating synthetic events that are coherent with it in the frequency range superior to 7.5 kHz. This algorithm is based on two steps: the analysis of the input signal and the synthesis of the high-frequency region according to parameters deduced from the analytical data. The newly generated region is then added to the original signal.*

1 Introduction

It might seem inappropriate to apply the issue of audio recovery to musical works as recent as those using computer-generated sounds [1], but many pieces produced in the 70's and early 80's occasionally had to pay for the limitations of the technology at that time. Quite frequently, in fact, they were synthesized at sampling rates much inferior to 44.1 kHz, which has then become the musical standard since the introduction of compact discs. In such cases, the electronic sounds have a limited frequency range and must be recovered by giving them more brightness.

As far as the support chosen for the master recording is concerned, some works were digitally recorded and therefore should not contain more defects, let alone the ones due to the low sampling rate; others, however, were recorded onto an analog tape or had to be converted from a digital to an analog format one or more times before being digitally recorded. In both cases their audio recovery is mandatory because of analog tape hiss and other troubles due to a mediocre digital-to-analog and analog-to-digital conversion.

The tape of the piece discussed here contained both defects and is therefore emblematic of the issue of sound recovery of many computer-music pieces composed at that time. *Traiettoria* is a 45-minutes work for piano and electronic sounds generated by computer composed by Marco Stroppa. It consists of three movements (*Traiettoria...deviata*, *Dialoghi*, and *Contrasti*) and was produced at the Centro di Sonologia Computazionale (CSC) between 1982 and 1984 [2, 3]. The performance of *Traiettoria* requires the diffusion of the electronic sounds recorded on a compact disc via a multi-channel amplification system, made of several speakers placed on the stage and around the audience and of one speaker located underneath the piano and facing upwards towards the sound board.

In order to maximize musical interaction and sound quality, the electronic sounds were recorded at a high dynamic level even in the soft sections of the piece. It is the interpreter in charge of the sound diffusion who has the task of finding the right levels in tune with the pianist and of shaping the sounds in space as a function of the acoustics of the hall and of the characteristics of the sound diffusion system.

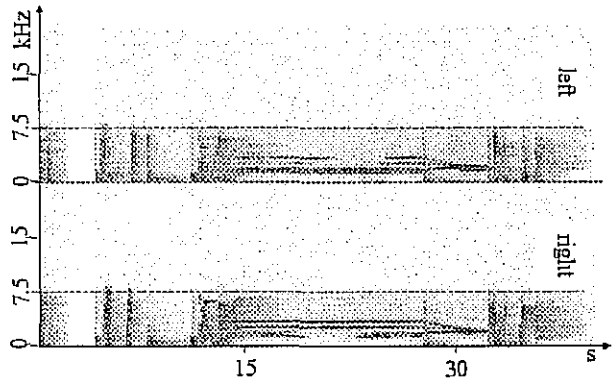


Fig. 1. Beginning of *Contrasti*

Although *Traiettoria* is a fairly recent piece, the electronic sounds require being recovered for the following reasons. First the system for the digital-to-

[1] This work has been carried out with the financial support of the CNR - Progetto finalizzato Beni Culturali

analog conversion used at the CSC in the early '80's allowed a maximum sampling rate of 15 kHz, thus limiting the audio range to 7.5 kHz (fig. 1).

Second, when in 1988 the composer decided to make a digital master of the piece the computers of the CSC did not have yet a digital output. It was therefore necessary to record it on a DAT tape analogically. As a consequence, the recovery of the electronic sounds requires two sorts of operations: the reduction of the noise due to the double conversion and the generation of the missing frequency region starting at 7.5 kHz.

This procedure is greatly different from the one employed for the recovery of analog works that have such defects as impulsive noises (clicks due to scratches of vinyl records or magnetization problems of tapes), hiss and audio distortions [4]. In our case, the background noise was limited to imperfections of the D/A and A/D conversion, which have appeared only since the adoption of digital technology. The problem of generating the missing high-frequency region is also new and is a direct consequence of digital sound (low sampling rate). However, it could also be applied to analog materials recorded with a limited spectral range.

2 Reduction of the defects of the conversion

The system used at the CSC in the '80s to transfer the electronic sounds of *Traiettoria* on a DAT cassette was not a commercial product, but a 16-bit prototype built with discrete elements by the Department of Electronic Engineering of the University of Padua [5]. The DAT was one of the first semi-professional portable machines. The transfer thus introduced some noise, albeit quite moderate, which was not present in the original sounds. The noise was more perceptible in low sounds with a limited spectrum, but was not very serious. The software plug-in DINR from Digidesign [6] (version 1.01), running under Sound Designer II (version 2.6) on a Power Macintosh computer could efficiently reduce it using the Broadband Noise Reduction (BNR) module. BNR uses a proprietary technique (called Dynamic Audio Signal Modeling) to analyze a segment of noise in the audio file and to build an internal dynamic spectral model of what the noise and the desired audio "sound" like. It then attempts to "pull apart" the two models, separating the noise from the desired audio. Fig. 2 shows the spectrum of the conversion noise of *Traiettoria*. The bump in the low register corresponds to the pitch of a low B flat. Since the noise was produced by the converters, it was only present with some sound and not during silent sections. It was therefore impossible to isolate a segment of noise alone in any sound file. The line connected by squares is the actual spectral model: it was obtained by creating a model of an area containing only noise (above 500 Hz approximately) and expanding it to the lower region of the spectrum.

To obtain a sufficiently clean signal without provoking spectral artifacts a total amount of noise reduction of 10.17 dB was considered satisfactory (fig. 2).

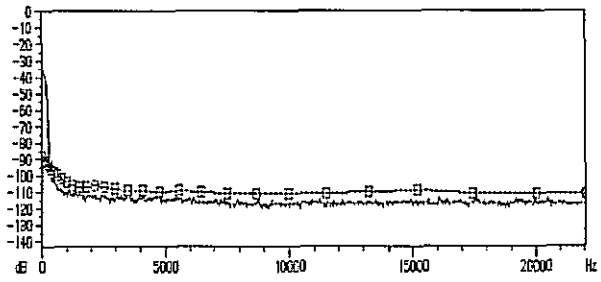


Fig. 2 Spectral Graph Display of the conversion noise of *Traiettoria*.

3 Generation of the missing high-frequency signal

Traditionally, the audio recovery of old analog materials requires the reduction of various noises masking some areas of the sound spectrum. Here, however, the process is totally different: it is not a question of making "audible" previously masked spectral areas, but of generating a high-frequency signal that is missing in the original, although it has to be coherent with it.

In order to find a model of high-frequency signals to be used as a reference, we analyzed the spectral energy above 7.5 kHz of some instrumental sounds: piano, string quartet, large orchestra and percussion ensemble. After listening to these filtered sounds, we observed how varied and differentiated both their energy and their timbral characteristics were: piano sounds do not contain a lot of energy in this region, while the string quartet is much richer and sensitive to changes in frequency, especially to glissandi. The orchestra shows a homogeneous distribution of energy over the whole area. The percussion ensemble, finally, is the richest and the most diversified, probably because several instruments generate a spectrum that is mainly situated in this area, while other instruments also cover it in many different ways.

From the standpoint of timbral quality, we identified two main morphologies: the high-frequency range consists either of clusters of sinusoidal sounds that are sometimes slightly modulated by an amplitude jitter (random variations), or of narrow bands of white noise.

An aural test led us to choose the first morphology for *Traiettoria*, since it seemed more coherent with the type of synthetic sounds used in the piece from both a perceptual and a theoretical perspective (all the sounds were generated using additive synthesis or frequency modulation).

The next problem was then how to generate these clusters of sinusoids. We started from the

psychoacoustical consideration that only notes with a frequency below 4-5 kHz are musically meaningful (the last note that an orchestra can play is about 4.1 kHz). We have thus chosen to analyze the energy of the spectral region between 4 and 7.5 kHz (middle-high range) and to project it in the region between 7.5 and 22.05 kHz (very-high range, taking the standard sampling rate of 44.1 kHz as a reference). We divided the two spectral regions into their critical bands [7] and controlled the energy of the very-high one with data coming from the analysis of the middle-high region.

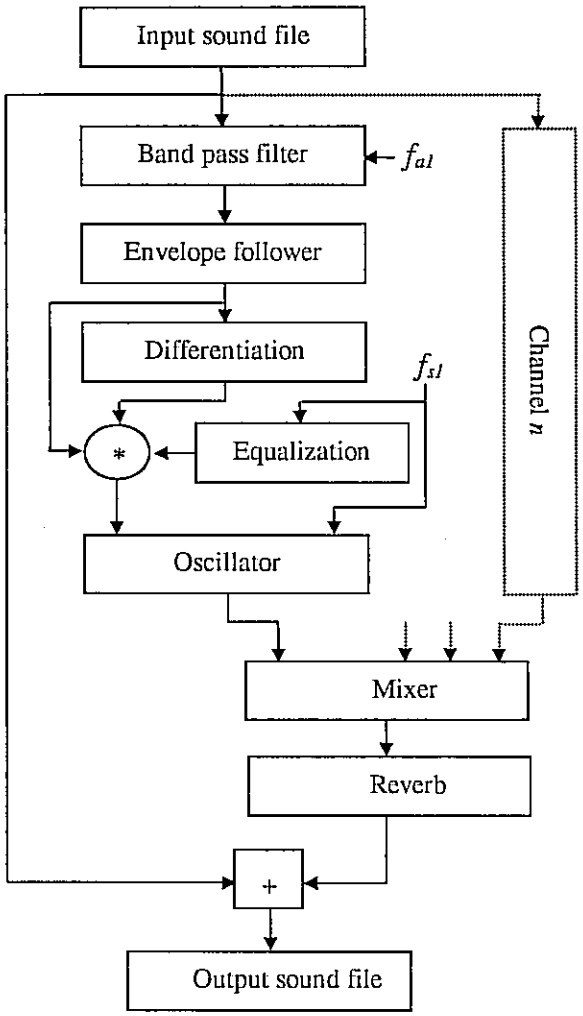


Fig. 3: Flowchart of one channel of the analysis-synthesis algorithm

For each band, we extracted the amplitude envelope and its derivative. The latter proved to be particularly useful to detect fast transients, which, as is well known, contain a lot of energy in the high-frequency range. During the experimental phase of our work, in fact, we observed that synthesizing the high-frequency range with data coming just from the amplitude envelope wasn't very successful. Too much energy was present and this was inconsistent with both our analytical data and with our perceptual simulations. However, when multiplying the amplitude envelope with its derivative we obtained optimal results.

Conversely, using the derivative alone yielded less balanced temporal contours, since they were dependent only from the rate of change of the amplitude and did not take into account its absolute value. This led to an excessive strengthening of the high-frequency range for rapid transients with a soft dynamic level.

Our analytical data also showed that it was essential to slightly vary the frequency of the sinusoidal clusters. A random variation within the critical band proved to be enough. The change of frequency was performed when the amplitude was zero. In this way, potentially harmful and perceptible discontinuities were eliminated. Finally, a reverb of 2.5 sec was applied exclusively to the newly generated high-frequency signal in order to give it a more homogeneous timbre. Both these additions were judged adequate and sufficient when tested aurally. Fig. 3 shows the flowchart of one channel of the analysis-synthesis algorithm used to generate the high-frequency signal.

The algorithm was very effective in recovering the missing high-frequency region of *Traiettoria*. It performed well with all the sound morphologies of the piece, such as percussive sounds, granular glissandi, continuous layers, and the like. Fig. 4 illustrates the sonogram of the beginning of *Contrasti* with the newly generated high-frequency region.

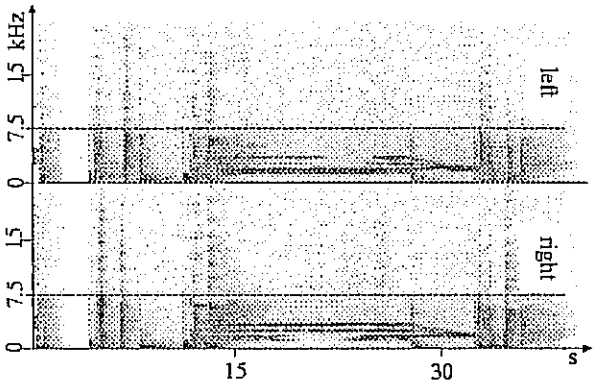


Fig. 4: Beginning of *Contrasti* with the new high-frequency region

However, every process of audio recovery has a component of craftsmanship that cannot be totally automated, especially when the sound files are as long and varied as in *Traiettoria*. The electronic sounds last more than half-an-hour and are divided into 8 sections whose duration is between 3' and 7'. In order to obtain a more musical result, we finely tuned the control

parameters of the algorithm for each section. Sometimes, the high-frequency region of one section was generated more than once with slightly dissimilar control parameters and recorded onto separated sound files. We then superposed all the available sound files to the original signal, but we always kept them in the different channels of a multi-tracks direct-to-disk system (ProTools from Digidesign, version 4.1). In this way, the intensity and the choice of one high-frequency file or another was made heuristically while listening in order to obtain the most convincing musical solution.

4 Conclusions

Although the algorithm described above was conceived for audio recovery, it might also be used as a compositional tool. When generating electronic sounds, the musically pertinent frequencies, that is the frequencies to which a given compositional thought can be applied and be perceived (provided that the composer has an approach based on musical pitches), are limited to the range of the orchestral instruments. Psychoacoustic research shows that the frequencies above this range are not perceived as pitches [7]. As a consequence, applying the same compositional process to so high frequencies often create problems of energetic balance (the highs tend to be too loud and to last too long), because what is good for musical pitches might not work as well for high-frequency spectra. On the other hand, restricting the maximum frequency of a composition to the beginning of the high-frequency region will probably produce an overall sonic result lacking brightness. The algorithm described above might therefore be useful to improve the acoustical quality of such works. For example, the compositional process might determine the frequencies of the synthetic sounds up to a region of, say, 6÷8 kHz and then recur to this algorithm to enrich the spectrum according to psychoacoustical and not only compositional rules.

We also imagine that the algorithm used in *Traiettoria* might also work in other contexts as well, or could be used to recover historical recordings in which the frequency range is reduced for various reasons: defects during the recording phase or deterioration of the medium used as a support of the magnetizing particles (as in old records, tapes or cassettes).

It is obvious that our cognitive system is much more exacting when dealing with acoustical instruments, since it already possesses a model of how they behave and sound like in the real world. The re-creation of the missing regions will presumably require that we also take into account phase information during the analysis step, that we extend our algorithm so as to include these data and that we find a way to project them onto the missing areas.

References

[1] Vidolin Alvise. "Conservazione e restauro dei beni musicali elettronici." *Le fonti musicali in Italia - Studi e Ricerche*, CIDIM, 6, pp. 151-168, 1992.

[2] Stroppa Marco. *Traiettoria.* Ed. BMG-Ricordi. Milano 1983.

[3] *Traiettoria*, program notes of the CD released by Wergo, WER 2030-2. Pierre-Laurent Aimard, piano, Marco Stroppa, sound projection.

[4] Shüller Dietrich. "Informazioni audio e video. Dalla preservazione nei supporti fisici alla preservazione delle informazioni." In T. Gregory and M. Morelli eds *L'eclisse delle memorie.* Laterza, Roma-Bari 1994.

[5] Rubbazzer Maurizio. "Convertitori D/A a 16 bit per audio professionale." *Atti del V CIM.* Ancona 1983.

[6] DINR, Digidesign Intelligent Noise Reduction. User's Guide, 1992.

[7] Moore, Brian C. J. *An Introduction to the Psychology of Hearing.* Academic Press, London 1982.

The preservation and restoration of audio documents: two practical examples

Paolo Zavagna - University of Udine - LIM Gorizia
E-mail: paolo.zavagna@ud.nettuno.it

Abstract

This paper will discuss problems concerning the preservation and restoration of audio documents on magnetic tape, using two practical examples wich cover a wide range of cases.

The first example is the electronic part of a piece of music, Musica su due dimensioni *by Bruno Maderna, 1958; the second is an old recording (1948) of an ethnic piece. The steps listed below will be presented, explained and discussed in the two audio documents:*

1) identification of the sources;

2) definition of the criteria to adopt for the source choice in the preservation backup;

3) choice, calibration and settings of the playback equipment;

4) preservation backup;

5) compositional process reconstruction and original identification of procedures, techniques and apparatuses;

6) identification of restoration problems and definition of priorities;

7) restoration; choice and production of the 'definitive' support medium on which the 'finished' product will be backed-up;

8) filing.

Each case contains some unusual details which give different degrees of importance to the various aspects listed above.

1. The electronic part of *Musica su due dimensioni* (1958) by Bruno Maderna[1]

Musica su due dimensioni (1958) by Bruno Maderna is a piece of music for flute and magnetic tape; it was one of the first "mixed" compositions including live instruments and a pre-recorded electroacoustic part; it was composed at the RAI Studio di Fonologia at Milan, with the collaboration of Marino Zuccheri as technician. The tape lasts 11'23".

1.1. Identification of the sources and material for the preservation backup

The pieces of electroacoustic music from the '40's and '60's are the result of a series of complex mixing procedures involving various sound material and consequently it is not always easy to identify the different sources which may have contributed to the composition of the original piece [12, 13]. Bibliographic research starting from [1] and a search in the files where the Maderna works are kept – at Basilea (Paul Sacher's Foundation), Bologna (Fondo Maderna), Milan (RAI Studio di Fonologia and the Suvini Zerboni publishing house), Florence (Tempo Reale) – has lead to the identification of the tape (two tracks, stereo, 1/4 inch, 38 cm/s) preserved in the RAI Studio di Fonologia Archive at Milan.

1.2. Setting up and calibrating the equipment for the preservation backup

This stage – probably the most delicate from the point of view of the documentation [3, 10] – can vary depending on the capabilities offered by the technology available at the time of the execution of the work. As there is no precise standard to refer to, a possible approach involves ensuring the best longevity, resistance and preservation of the information and readability by choosing one or more formats [11, 4]. In our case, a 16-bit DAT with sampling frequency at 48 kHz has been adopted as the standard[2]. We chose a modern playback equipment wich allowed the tape tension to be adjusted in order not to damage the carrier. The calibration was carried out with the tape test more close to the time of the carrier.

Finally it was possible to operate safely on the digital copy by transferring the data from one support medium to another without substantial deterioration. The piece was recorded on hard disk and, after converting the sampling frequency to the CD-A standard (16 bit, 44.1 kHz), a copy on optical support was made from the hard disk. In this way, it was possible to avoid, as far as possible, problems due to obsolete technology (particularly outdated digital methods) and 'standards' [11]. As another backup copy in a different format and on a different support [4] was available, all the desired operations of editing and restoration could be made, after normalising the amplitude of the sound file.

1.3. Preliminary listening test, reconstruction of the compositional process, techniques and equipment used by the composer and the technicians involved

A first listening test revealed no serious deterioration problems, such as the magnetic printing, the presence of physically damaged parts, impulsive

[1] I thank the editing house Suvini Zerboni, which has the copyrights, and the RAI of Milan, particularly the archive of the Studio di Fonologia Musicale, where the original tape is filed and where the DAT copy was made, with the collaboration of Maddalena Novati, Giovanni Belletti, Massimo Bozzoni and Fabio Ferrarini.

[2] We can already talk about a 24 bit resolution with a sampling frequency of 96 kHz.

noises. The worst problem was represented by the broad band noise, which was very high, discontinuous and considerably different between the two channels – in some points the broad band noise on the left channel exceeded that of the right channel by as much as 7dB (fig.1).

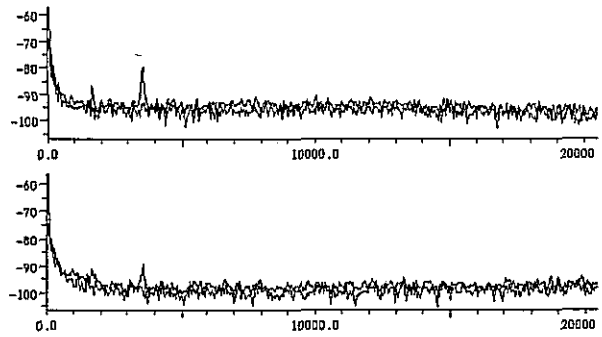


Fig.1: broad band noise at the very beginning of *Musica su due dimensioni*, left channel (above), right channel (below).

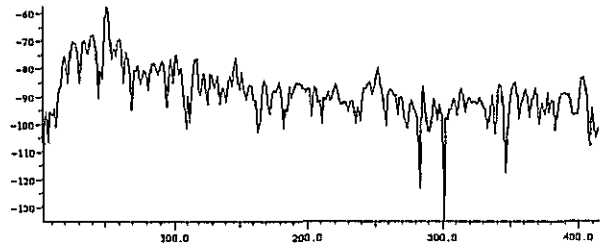


Fig 2: AC 'hum' in *Musica su due dimensioni*, left channel.

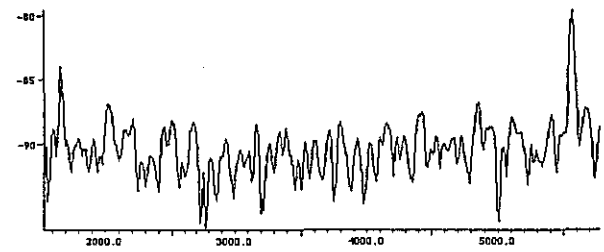


Fig. 3: Narrow band foreign frequencies (1656 and 5562 Hz) detected in a noise sample at the end of *Musica su due dimensioni*, left channel.

A preliminary analysis aimed at identifying the sound materials and the technology [8, 9] used by the composer, demonstrated the massive but not exclusive presence of flute sounds which were either natural or manipulated. The manipulations included the editing, mixing, transposition with and without speed variation, ring modulation (characterised by a strong presence of noise), and reverberation in an echo chamber. In the central section of the music piece there are sounds with non-periodic characteristics, both percussive and sustained.

During the tape cutting, the usual procedure to make the pauses involved using a blank tape [13], basically free from noise. However, in some cases, to obviate the difference in sonority between the pure silence and the sound recorded on the tape (with the broad band noise essentially, but not exclusively, due to over-recordings), pieces of tape were used.

1.4. Identification of the 'introduced alterations' and restoration processes

The following restorable 'defects' were found[3]:

1) rare impulsive noises with small amplitude

2) 'hum' at 50 Hz with some harmonics in evidence

3) constant frequency with a very narrow band around 1656 Hz (figs. 1 and 3);

4) high frequency in a very narrow band passing from approx. 3560 Hz (fig.1) to approx. 5562 Hz (fig.3);

5) noise in the pauses;

6) broad band noise of many different types.

The few impulsive noises detected by ear were eliminated manually.

The two fixed frequencies of 50 Hz (with some harmonics: 100, 150, 200, 250, 400) and 1656 Hz were eliminated with a very narrow band notch filter operating at 7 Hz (for AC hum and the most evident harmonics) and 10 Hz respectively. The *glissando* was eliminated with a dynamic notch filter (3560 to 5562 Hz in 11'23") with variable pass band 12 to 25 Hz.

The frequent and fairly long pauses present in the tape were replaced with silence. These choices were madeconsidering that the piece of the music contained also a live instrument, which often covered the 'silences' of the tape.

The frequent pauses allowed several noise samples to be taken for analysis from various points of both channels. It was possible to note immediately that each single fragment had different noise characteristics to those present in the music parts, whose handling, over-recordings and cutting had introduced characteristic noise bands. This entailed the music piece fragmentation in several sections and the evaluation of the noise in each single section for each single channel.

From the musicological analysis point of view, it is interesting to point out that the 'colour' of the background noise corresponded to different types of electroacoustic processing and different sections of the piece of music.

It was decided to reduce the elimination of the broad band noise as much as possible, by not using it in the parts in which the music masked it completely or contained it as its integral part.

2. Bosa resuscitada

The music piece under consideration is particularly interesting from the ethnomusicological point of view. It was taken from a recording on acetate at 78 rpm by the RAI of Cagliari (Sardinia) in 1948. The copy on

[3] The procedures and the individuation of the values of the parameters adopted in this work for the two pieces of music are similar in many digital systems of editing, analysis and sound restoration present in the market.

which the restoration was carried out is quite recent. It is a two-track magnetic tape[4], mono, 1/4 inch, recorded at 19 cm/s. The backup copy was made in the same way as the first example work, calibrating the playback equipment with the test tape actually used at RAI. The duration of the piece of music is 1'22".

2.1. Identification of the 'introduced alterations' and restoration operations

The material contains almost all the typical defects of recordings on 78 rpm records:

1) line frequency ('hum');

2) isolated impulsive noises of various duration, amplitude and spectral content ('click');

3) superposed impulsive noises ('crackle');

4) broad band noise ('hiss');

5) incorrect equalisation;

6) limited band.

With an highpass filter centred on 60 Hz, all the disturbances in the low frequencies were eliminated without causing any change to the musical material, whose lower basic frequency was around 107 Hz. With a series of notch centred at frequencies of 100, 200 and 400 Hz, the line frequency with the more substantial harmonics was eliminated.

The localisation and removal of the 348 clicks was carried out at different times. As the equalisation of the tape copy limited the energy at high frequencies, it was difficult to find all the clicks contained in the material to restore. Where the automatic algorithms were incapable of signalling all the impulsive imperfections, it was necessary to operate partly by ear and partly filtering the material during analysis, to find the majority of the imperfections introduced, depending on their energy contents at high or low frequencies. This was the longest phase, due to the heterogeneity of the impulsive noises (figs. 4-6).

Two points of the tape were chosen to evaluate the broad band noise: before the beginning of the piece and a brief central pause. The result is that the broad band noise, as for the entire piece, is particularly coloured (fig.7).

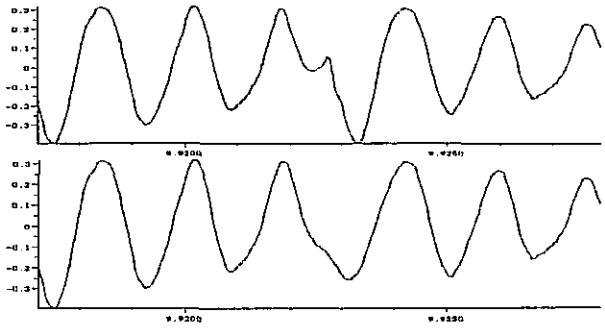


Fig. 4: click of approx. 0.7 ms before (below) and after (above) the interpolation

[4] The tape was supplied by Professor Pietro Sassu, who edited the CD *Bosa and the Planargia*, NOTA 2.52 (1998), in which the restored version of the piece of music was published.

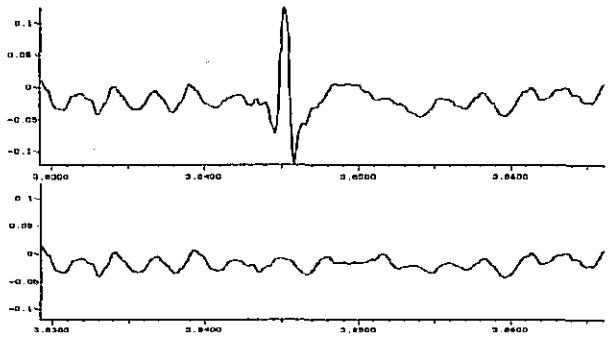


Fig. 5: click of approx. 3.5 ms before (above) and after (below) the interpolation

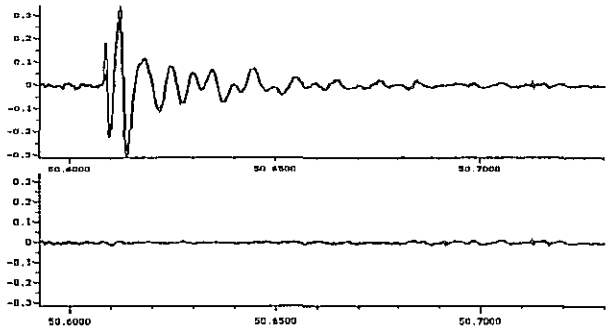


Fig. 6: typical click followed by 'ringing' before (above) and after (below) the interpolation

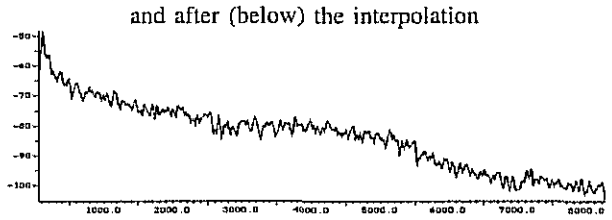


Fig. 7: coloured broad band noise of the *Bosa resuscitada*.

We believe – and the colouring of the noise confirms this – that the equalisation of the tape copy was made following the RIAA standard (turnover at 500 Hz, 10 kHz rolloff = -13.7 dB) emphasising the frequencies below 1000 Hz (0 dB as reference) and attenuating those above.

However, before proceeding with a re-equalisation, it was decided to remove the broad band noise, as the latter certainly did not pertain to the original sound material. This choice, however questionable, was taken considering the poor information at our disposal on the equalisation curves of the recordings on disc (record) before 1956, the year in which the RIAA standard was adopted [6].

To attenuate the broad band noise, the same technique applied for the previous example was used, with different problems for the colouring of the noise and the possibility that it might mask the sound content. Therefore we operated on few sections, diversifying them only where we found particular problems.

Finally, assuming a turnover of 450 Hz and a 10 kHz rolloff equal to 0 dB, the zone below 450 Hz was attenuated and the zone above 1000 Hz was

overemphasised to re-equalise the music piece correcting the RIAA curve introduced [6].

Considering that the spectral range of the recording did not exceed 5 kHz in the point of maximum content at high frequency (fig.8), we can suppose the reconstruction of the spectral range lost (see Stroppa and Vidolin in this volume) – which was not carried out – also comparing the material with more recent similar recordings therefore having more information on the spectral range.

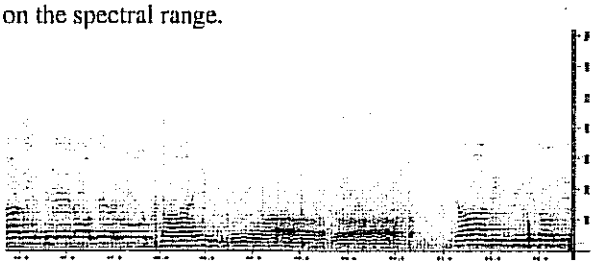


Fig. 8: sonogram of a portion of the *Bosa Resuscitada* which shows the spectral energy content reduced below 5 kHz.

3. 'Finished' restored product

Two different types of backup of the restored material was carried out:

1) audio data, as files in raw format (16 bit, 48 kHz);

2) sound in CD-A format preceded, on the same track, with various test signals. Both the copies were then backed-up on optical support.

4. Filing

The two works were part of detailed documentation concerning the restoration and filing (see table below) for the preservation backup.

AUTHOR	BRUNO MADERNA
TITLE	MUSICA SU DUE DIMENSIONI (TAPE)
DATE	1958
FORMAT AND SPEED	TAPE, 1/4 INCH, 2 TRACKS, STEREO, 15 IPS
MARK	BASF LR 56 P
EQUALIZATION	UNKNOWN
RECORDING EQUIPMENT	AMPEX [?]
PLAYBACK EQUIPMENT	STUDER A 812
PLACE	MILAN, RAI STUDIO DI FONOLOGIA, E3
NOTES	1) TECHNICIAN: Marino Zuccheri 2) RAI TAPE TEST - 1982 TO CALIBRATE THE PLAYBACK EQUIPMNENT

Carrier

DATE	1998
FORMAT AND SPEED	DAT, SR 48 KHZ
MARK	BASF DAT MASTER
EQUALIZATION	SEE THE DOCUMENTION ENCLOSED WITH THE TAPE TEST
RECORDING EQUIPMENT	SONY PCM 2800
PLACE	MILAN, RAI
NOTES	TECHNICIANS: Massimo Bozzoni, Giovanni Belletti, Maddalena Novati, Fabio Ferrarini.

Preservation copy

4. Conclusions

The rapid degradation of magnetic support [2] calls for the need of preservation copies of the sound documents contained in several archives. However, the manual operations necessary to restore even a few minutes of corrupted audio are still too numerous to allow the restoration of large quantities of sound material. The expectations on the restoration of a tape of electronic music and a 78 rpm record indicates the difficulties arising from automatic procedures, especially for the broad band noise [5].

To restore a sound document – with all the limitations that this operation involves – there must be a reason justifying the work, as a new version, mass diffusion or the use of it as an example for scientific use.

References

[1] M. Baroni and R. Dalmonte (eds.). *Bruno Maderna documenti*. Suvini Zerboni, Milano, 1985.

[2] J.W.C. Van Bogart. *Magnetic Tape Storage and Handling Guide*. Commission on Preserv. and Access and NML, Washington, 1995. (http://www.nml.org)

[3] G. Boston (ed.). *Guide to the Basic Technical Equipment Required by Audio, Film and Television Archives*. UNESCO, Paris, 1991.

[4] M.F. Calas and J.M. Fontaine. *La Conservation des documents sonores*. CNRS Editions, Paris, 1996.

[5] O. Cappé. *Techniques de réduction de bruit pour la restauration d'enregistrements musicaux*. PhD thesis, ENST, Paris, 1993.

[6] Gary A. Galo. Disc Recording Equalization Demystified. *ARSC Journal*, 27, 2, pp. 188-211, 1996.

[7] S.J. Godsill, P.J. W. Rayner, O. Cappé. Digital Audio Restoration. In M. Kahrs and K. Brandenburg (eds.). *Applications of digital signal processing to audio and acoustics*. Kluwer, 1998.

[8] A. Lietti. Gli impianti tecnici dello Studio di Fonologia Musicale di Radio Milano. *Elettronica*, III, pp. 116-122, 1956.

[9] P. Santi. La nascita dello "Studio di Fonologia Musicale" di Milano. *Musica/Realtà*, XIV, pp.167-188, 1984.

[10] D. Schueller. The Ethics of Preservation, Restoration, and Re-Issues of Historical Sound Recordings. *JAES*, 12, pp. 1014-1016, 1991.

[11] D. Schueller. Informazione audio e video. Dalla preservazione dei supporti fisici alla preservazione delle informazioni. In T. Gregory and M. Morelli (eds.). *L'eclisse delle memorie*. Laterza, Bari, pp. 21-32, 1994.

[12] A. Vidolin. Conservazione e restauro dei beni musicali elettronici. In Le fonti musicali in Italia - Studi e ricerche. CIDIM, 6, pp.151-168, Roma, 1992.

[13] P. Zavagna. Thema (Omaggio a Joyce) di Luciano Berio: un'analisi. I quaderni della Civica Scuola di Musica, special issue dedicated to Bruno Maderna, 21-22, pp. 58-64, Milano, dicembre 1992.

Thursday 24th **h. 12.10**

POSTER SESSION I

Measuring and Analyses Carried out on Some Historical Pipe Organs in Rome

L. Bazzanella
Centro di Sonologia Computazionale
Universitá di Padova
049/8276981 - lb@csc1.unipd.it

G.B. Debiasi
Dipartimento di Elettronica ed Informatica
Universitá di Padova
049/8277675 - debiasi@ibm.net

Abstract

Whenever the restoration of a historical pipe organ is executed the best thing to do should be to gather an amount of records as wide as possible including any information concerning its acoustic features before restoration work, so that it is possible to verify the quality and regularity of the results achieved.

Nevertheless, if the pipe organ to restore is so damaged that it does not allowed any acoustic measuring, it is then necessary to gather information about instruments of the same age and having the same building features. This is what we did together with I.C.R.-Istituto Centrale per il Restauro (National Institute for Restoration) in Rome.

Introduction

We sum up here the outcome of a series of measuring and analyses which have been carried out by working on the sound of some mechanical transmission pipe organs in Rome. These pipe organs were chosen upon suggestions of the I.C.R. in Rome; the purpose was that of acquiring audio documentation concerning historical instruments which provides us with useful descriptions for restoring similar instruments. The instruments taken into consideration are listed below.

1) Positivo su base 4' - National Museum of Musical Instruments - inv. 2771: date and author unknown; 4 stops; most pipes were recent.

2) Positivo ad ala - National Museum of Musical Instruments - inv. 900: roman school (perhaps manufactured in the circle of the Testa family) - end XVII - beginning XVIII cent.; 7 stops; hand-bellow only peg windchest.

3) Positivo Neapolitan school - National Museum of Musical Instruments - inv. 895: author unknown; first half XVIII cent.; 8 stops.

4) Organo Testa-Alari - Chiesa S. Giovanni dei Fiorentini: 1673-1680; 15 stops; restored e partially rebuilt by Bartolomeo Formentelli in 1994.

During the recordings two condenser microphones with directional characteristics AKG C414 B ULS were employed in cardioid configuration, with a proper microphone preamplifier (Symetrix). The microphones were placed at the same level as the front organ pipe mouths; they were 20 cm distant from each other (this is approximately the average distance which separates human ears). The output signal from the preamplifier was recorded by a DAT unit (Denon Dtr2000).

2. Analyses

The study carried out on the characteristics of the sound analyzed concerned both transient and steady conditions.

Transient conditions

The study on the transient conditions consisted in analyzing the amplitude envelope of the first six harmonics with different "touches"; these analyses were executed with a software application based on heterodyne filter. Moreover, we inquired the possibility for "touch" to exert an influence on the transient harmonicity. For this purpose we obtained the dispersion curves concerning the transient which show the deviation of the actual frequencies of the harmonics with regard to their theoretical value. There are many written contributions which in most cases confirm this peculiarity in presence of mechanical transmission, especially with direct mechanical transmission [1][2][3]. In this instance it can happen, however, that if the block chain linking the keys to the valves is fairly long (large pipe organs), the possibility one has to find at the valves the features of the excitation given to the key is quite weak. Because of their slimness the metal bars are put under a torsional stress and they store elastic energy until the resistance opposed by the spring and the compressed air is balanced: at this point the energy is released causing the valve to open suddenly. The result is that the transient is insensitive to "touch". The same happens when squares and compensators are part of the mechanism, because of the noteworthy static attrition and the inevitable clearances.

Both after studying the writings dealing with this problem and previous tests we can argue that very often touch effect on pipe organ sounds mainly causes different amplitude envelopes for one or more harmonics of the sound taken into consideration.

2.2 Steady-state conditions. Pseudoformants

As for steady-state conditions, we obtained the spectrum of the sounds; a problem came up, i.e. verifying if in particular circumstances these sounds presented a pseudoformantic structure. Formant studies were introduced to explain some acoustic phenomena such as that concerning vowel sounds; for each vowel it is possible to draw a so called formant curve linking all the harmonics peaks. With vowels the formant is constant even though the frequency at which the vowel is pronounced varies. In general instruments having a resonator present frequency envelopes which show the presence of fixed formants, insofar as the resonator filters the signal produced by the oscillating string or by whatever produces the vibration originating sound. Pipe organs, on the contrary, do not have any element which can be compared to a resonator, since each note is produced by one single pipe; this means that spectral envelope of pipe organ sounds should not have a fixed

formant. Sometimes, however, the organ builder builds the pipes of some stops, especially those of the *Ripieno*, so as to make the sound particularly pleasing to the ear. In these situations the timbre keeps specific features even though the frequency of the note varies. Sometimes it is possible to recognize within the spectra of these sounds the presence of one or more formants, called pseudoformants.

3. Analysis results

An analysis of the results gathered allows us to make some observations about the way in which the tested instruments behave. They are presented below, with reference to the figures corresponding to each instance.

Fig. 1.a-b National Museum of Musical Instruments - Positivo di accompagnamento, Principale 4', D2, fast and slow "touches". With regard to the steady-state condition, with the fast "touch" we noticed a stressed overshoot as for the 2nd and 4th harmonics. Even though the latter has lower amplitude than the 2nd, it is however relevant: from a psychoacoustic point of view ear judgment is strongly influenced by the gradient and the effect of time priority of each harmonic.

Fig. 2.a-b National Museum of Musical Instruments - Positivo di accompagnamento, Principale 4', C2, fast and slow "touches". With regard to D2 (see fig. 1.a-b) we can notice a lower enhancement of the 2nd harmonic and a stressed overshoot of the 5th. This confirms that during the transient harmonic behavior in the same stop depends also on the tuning given to each pipe.

Fig. 3.a-b S. Giovanni dei Fiorentini - Positivo Napoletano, XIX cent., Principale 8', C3, fast and slow "touches". Dramatic enhancement of the 2nd harmonic with the fast "touch" preceded by two peaks (4th and 6th harmonics). Because of the reasons explained above, these last two peaks (especially the 6th harmonic) play an important role in the sensorial evaluation of the timbre composition of the transient.

4. Conclusions

The tests carried out allowed us to gather an impressive amount of data concerning the acoustic features of several pipe organs and to examine them closely.

The results we obtained were used during restoration work on the Altemps pipe organ by Filippo Testa (1701) in S. Maria in Trastevere, supervised by the Istituto Centrale per il Restauro.

References

[1] N.H. Fletcher "Transients in the speech of Organ Flue Pipes - A theoretical study" ACUSTICA, vol. 34, pp. 224-233, 1976.

[2] T.L. Finch, A.W. Nolle "Pressure wave reflections in an organ note channel" J.A.S.A., vol. LXXIX-5, pp. 1584-1591, 1986.

[3] L. Bazzanella, G.B. Debiasi "Analysis of touch effect on the transient of pipe organs with mechanical transmission" Proc. ICMC, pp. 485-487, Aarhus 1994.

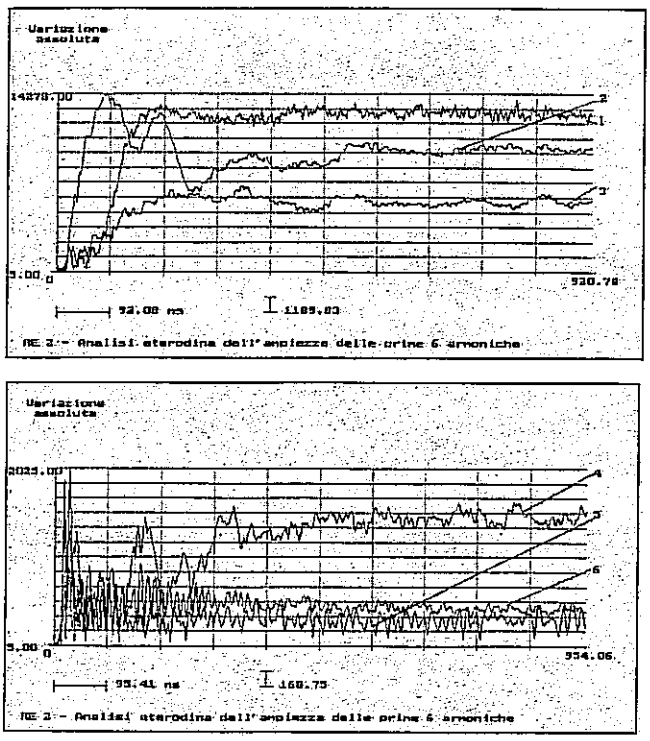


Fig. 1.a

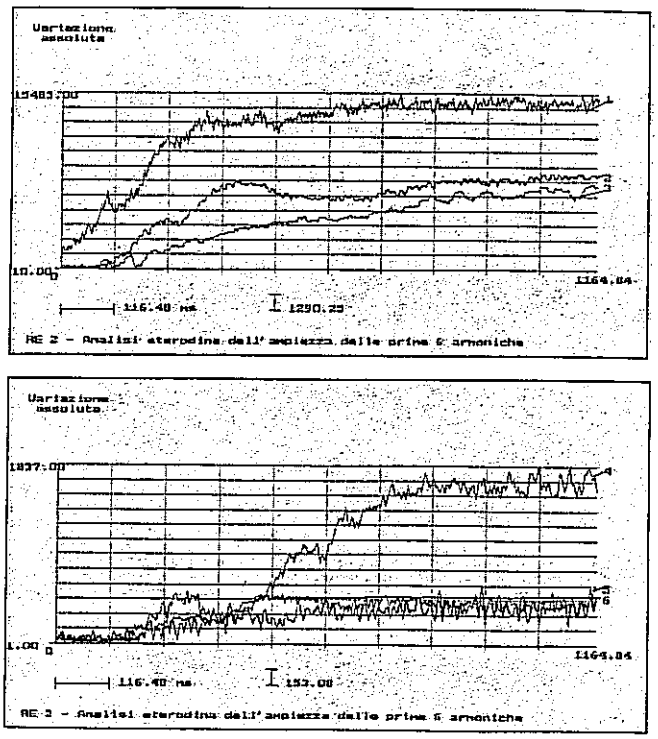
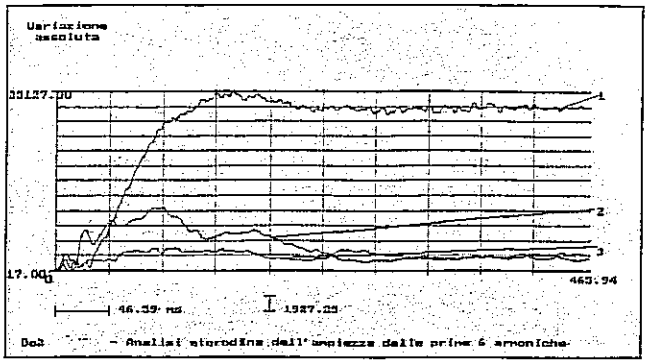


Fig. 1.b



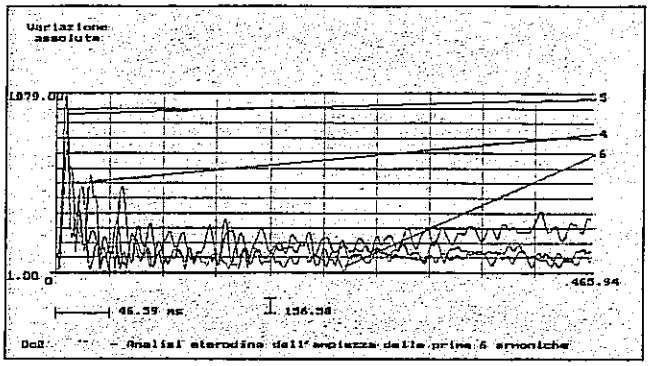
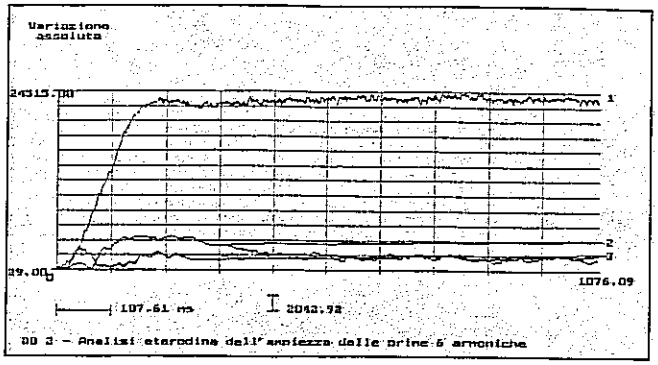


Fig. 2.a



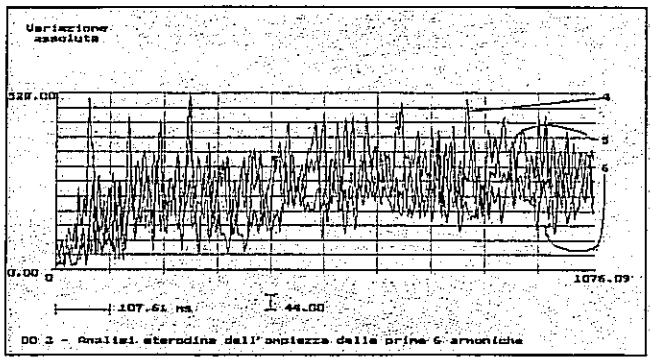
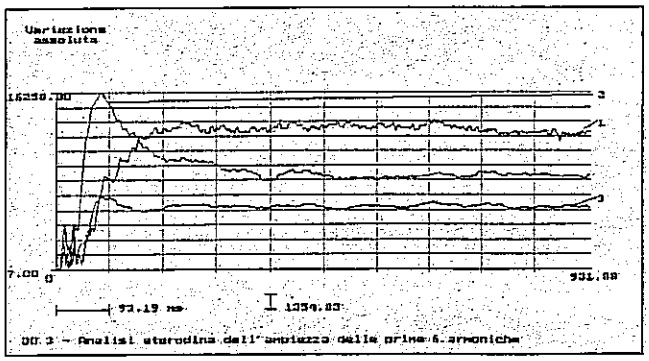
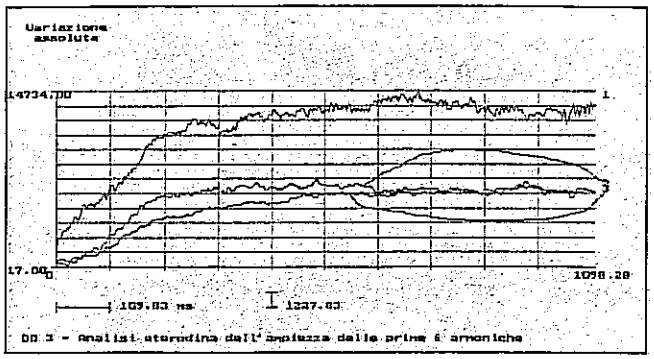


Fig.2.b





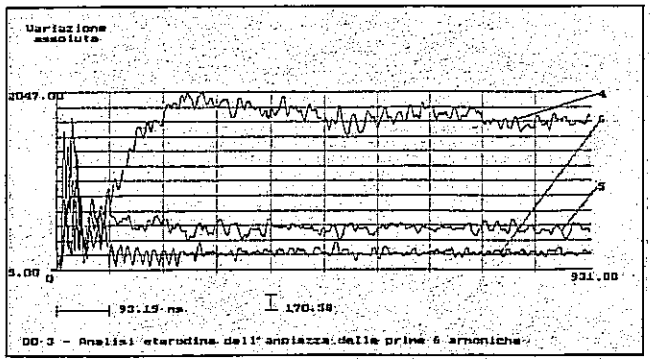


Fig. 3.a

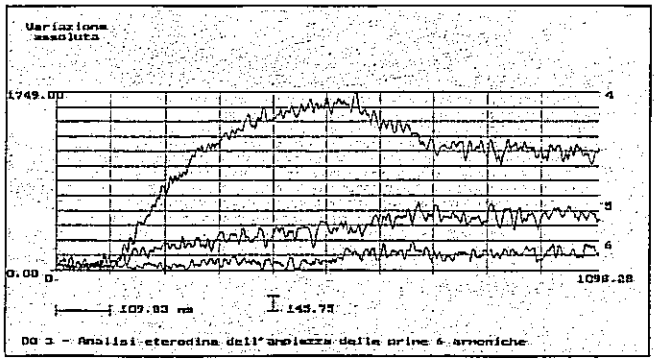


Fig. 3.b

Artificial Life, Embodiment and Computer Music

Jon Bedworth
Centre for Advanced Inquiry in the Interactive Arts
University of Wales, College Newport
j.m.bedworth@newport.ac.uk

Abstract

In relation to musical activity, the computer can be seen to impact in three inter-related ways, these being: augmenting human creativity; understanding compositional processes; and creating artificial musicians. If a computer is to be creative it has to be able to not merely generate but also be able evaluate its productions. Composing with timbre highlights the importance of the phenomenal experience of sound in such an evaluation and ethnomusicological evidence points to alternative embodied and decentralized music making. The implications of these are discussed in relation to aspects of Artificial Life and 'embodied cognition', particularly in regard to the pursuit of artificial musicianship.

Introduction

Computers have opened up new sound worlds, and ways of working, both in composition and music understanding. We are able to model aspects of reality, and then conduct experiments that would not otherwise be possible. Thus the computer alters our perception of, and potential for action in, the world. As technology, the computer is both a result of social concerns and values, and a causal agent in our perceptions. We see and manipulate the world through technology[1]. A computer music system embodies an abstraction of a particular musical approach [2] and yet also transforms 'what can be considered music', the computer acting as "...a dynamic device that transforms the imagination" resulting in a 'circular, cybernetic relationship' [3]. Given that musical developments "...have paralleled important changes in social, cultural and intellectual life" [4], this paper considers some perspectives within the fields of Artificial Life (A-Life) and 'embodied cognition', which have musical implications. Such models and metaphors inspired by biology "... will have their greatest influence when they spread outside of the scientific community and into the general culture." [5]

For the purposes of discussion here, it seems important to first delineate three inter-related aspects of how the computer could be considered to be transforming musical practice. These are: (a) augmenting human creativity in assisting the exploration of new musical approaches; (b) understanding compositional processes and (c); creating 'artificial musicianship', involving emulating, equalling or even replacing human musical creativity.

(a.) *Augmenting human creative activity.* The computer may be perceived as an instrument [6], medium or tool, in relation to which it is important to consider the mapping of human input (gestural, haptic and so forth) to musical output [7]. Also it may act as a suggester of ideas and device for 'organising complexity'[8] within certain compositional approaches.[9]

(b.) *Understanding compositional processes.* Computer music pioneers Lejaren Hiller and Leonard Isaacson " ... desired to simulate the composing process itself with computers, rather than use computers as an aid to composition."[10] Subsequently computers have been applied to 'cognitive musicology'[11], which "..derives from reflections upon the structure of real-time processes made possible by, and monitored by, computers." Here "thinking is situated in some social and technological context ... A form of action, rather than a disembodied reflection."[12]

(c.) *Emulating, equalling or replacing human creativity.* One motivation is to give computers 'greater musicianship', "... making the program more aware of musical implications in the performances of its partners, and ... better able to perform itself more musically."[13] Here many problems are still not clearly articulated at the conceptual level, let alone the level of implementation.[14] Opinions about the computer's potential in this regard range from those who see such attempts to mimic human musicians as stage performers as 'misplaced'[15] to implications that computers could make music 'understood only by other computers'[16] The latter echoes beliefs that computers could become sentient beings.[17]

In attempting to automate music making, it is useful to be aware of the social conditions that lead to such a desire. Music's social technology can pave the way for music's machine technology, where "...machine technology involves the processing of tools and machines, and social technology involves the skills and means of organizing people to get work done."[18] The 19th century sees a growing desire for automating music, and the "...abstraction of the performer from the performance takes place when a social technology allowing such abstraction prepares the way for machine technology.... Current aesthetic arguments which view technology in utopian terms reflect values by which man has abstracted from his productions"[19]

Musical Systems and Processes

In considering the 'abstraction of the performer from the performance', it is useful to discuss the relationship of musical systems to musical process and function. What is the relative importance given to music as the construction of objects and products on the one hand, and the type and quality of human engagement with the musical process, on the other? The rise of the classical composer 'genius'[20] and the development of more formalised music notation furthered the sense of an independently existing 'eternal' art object, that can be abstracted, recorded and analyzed. Computers, like previous technologies such as the phonograph[21], have further helped to freeze music in time.

Abstracting away musical forms from their origins in specific musical activities makes detached analysis of music's structure possible, aiding our understanding of musical forms. However, traditional musicology's primary focus has been on such musical texts to the virtual exclusion of performance[22]. Western tonal music, codified into written notation, can give the

impression that we are dealing with a closed, entirely encodable system of finite harmonic relationships, transformable by clearly definable rules. Trevor Wishart believes the 'fundamental thesis' of the analytic system of notation is:
"...that music is ultimately reducible to a small, finite number of elementary constituents with a finite number of 'parameters', out of which all sounds possibly required in musical praxis can be notated by the combination of these constituents."[23]
It can appear as if the music itself resides in the notation.

Such a view finds resonance with traditional Artificial Intelligence's view of cognition (excluding connectionist approaches) as the representation of explicit symbols and intelligent behaviour as the manipulation of those symbols; the so-called Physical Symbol System Hypothesis[24]. Symbols could be represented in the computer as data structures and intelligent behaviour by algorithms, which act on the data structures. Crucially, it assumes that such intelligence resides wholly within the mind, and therefore could equally reside in the computer. However, certain musical approaches question such a view. Here I refer to timbre composition and evidence from ethnomusicology.

Timbre Composition

Composing with timbre[25] has consequences for the nature of the distinction between the sound materials (data structures) and the rules for sound manipulation (the algorithm); i.e. between musical design and materials, between 'composing-the-sound' and 'composing-with-sounds'.[26] This challenges the distinction between computer as instrument (able to encode, process and transform sound) and as having the "... potential for designing the compositional process".[27] For example, Di Scipio and Prignano's Functional Iteration Synthesis "...provides an *indeterministic* model of sonic material: the composer must learn his/her strategy by interacting with a source of structured information, at the level of the microstructure of music - within and through the sound."[28] Such an approach implies, along with the application of intelligence, a response to the phenomenal, experienced aspects of sound.

That computer software can produce surprising results leads to speculation that the computer itself could be creative in this situation. However, it is important to distinguish the difference between the generation and evaluation of creative ideas. It is not enough to produce something novel and interesting. Matthew Elton says:
"If we are aiming at for artificial creativity, then our main worry concerns not what the machine can produce, but the way in which it evaluates its productions"[29]
In evaluating its productions, is the computer capable of experiencing sound in the same phenomenological way as humans?

Will A-Life produce artificial entities capable of not just generating, but evaluating such musical productions? Christopher Langton, one of Alife's originators, has defined Artificial Life as:
"... a field of study devoted to understanding life by attempting to abstract the fundamental dynamical principles underlying biological phenomena, and recreating these dynamics in other physical media - such as computers - making them accessible to new kinds of experimental manipulation and testing."[30]

And a variety of techniques and technologies used within A-Life research have been used as generators of musical ideas and compositions, arguably opening music up to 'new kinds of experimental manipulation and testing' using computers. These include fractal mathematics, cellular automata, genetic algorithms and dynamical systems, which seemingly provide fertile ground for musical explorations. In generating musical form, A-Life systems, in contrast to connectionist approaches, could attempt to 'develop their own principles of organization' and '... model multi-level order'[31]

Is such order emergent or is it imposed? Margaret Boden sees that "[t]he central concept of A-Life ... is self-organization [which] ... involves the emergence (and maintenance) of order, or complexity, out of an origin that is ordered to a lesser degree ... This development is 'spontaneous', or 'autonomous', following from the intrinsic character of the system itself (often, in interaction with the environment) instead of being imposed on the system by some external designer. In that sense, A-Life is opposed to classical AI, in which programmers impose order on general-purpose machines"[32]

Equipped with more realistic biological models, 'recreating these dynamics in other physical media' in the form of computer software, it would appear that we may be a step closer to an 'artificial musician' able to generate and evaluate its own productions. However, what is the status of such biologically inspired models? As Stevan Harnad argues that "Computational modeling ... can conquer the formal principles of life, perhaps predict and explain it completely, but it can no more be alive than a virtual forest fire can be hot. In itself, a computational model is just an ungrounded symbol system; no matter how closely it matches the properties of what is being modeled, it matches them only formally, with the mediation of an interpretation."[33]
Similarly, an 'epistemological danger' exists in "... the belief that a high-quality simulation can become a realization – that we can perfect our computer simulations of life to the point that they come alive."[34]
The same could be said for the modelling of emotions and creativity. What would make the computer feel like it wanted to participate in the musical situation, to *be* musically aware rather than just be a model of musical awareness?

Ethnomusicology

What type of musical situation do I have in mind? Ethnomusicology reveals other models of musical participation, examples being the Venda[35] and Kaluli[36]. Understanding such musical practice may in turn have implications for our own, at least in reminding us what we may have lost. Such studies point to decentralized, distributed musical activity. Michael Chanan writes that "in surviving oral cultures .....[t]here are no composers in such societies set apart from other musicians in a separate caste, and music is far from an exclusive activity of specialised performers. .... tribal community encourages and sustains a degree of musical ability in

virtually all its members through the widespread use of informal music. Moreover music enters into the widest range of activities."[37]
In such societies "The logics of a society's relationships are not mediated through the being of any one individual ... The musical mediation of the relations between the individual and society and these instances emphasise the importance of individual autonomy and creativity to the social process."[38]

And such participants have a knowledge of what has been termed by Roland Barthes as *musica practica*, who described it as "...a muscular music in which the part taken by the sense of hearing is only one of ratification, as though the body were hearing."[39] For Chanan "... *musica practica* is nothing but the form that musical knowledge takes directly from musical practice. Theoretically filtered or not, fundamentally it has no need of theory or even notation. It is the musical equivalent of the way the baby learns to talk."[40] If music is viewed as physically situated, collective human activity, is it then just a case of encoding the particular music system's observed rules into a machine? Where does the musical knowledge reside?

Distributed, Embodied Cognition

An interesting perspective on this question emerges when we reconsider cognition as embodied and decentralized; when we view cognition as distributed between brain, body, and world.[41] Within such a viewpoint perspective structure does not lie in any one place, and is not entirely represented in any one thing (or mind), but emerges through complex interactions. Furthermore, in contrast to Cartesian mind-body distinctions which gave rise to an idea where "...the mind controls the body as a captain pilots a ship"[42] the phenomenologist Merleau-Ponty considered that it "...is not brains that think, it is bodies. The brain is one part of a larger system, the nervous system and ultimately the entire body."[43]

Such a view is finding technical application in the design of robots. The emerging field of Embodied Artificial Intelligence highlights the necessity "... to study intelligence as a bodily phenomenon",[44] in constructing autonomous robots. Furthermore, such robots "...react directly (bottom-up) to environmental cues rather than (top-down) to internal world-models or representations."[45] Here cognition "...is no longer viewed separately from its bodily substrate ... cognition is also the study of bodily action and perception in the system's environment and cannot be viewed separately from either of the three (body, action, environment)."[46]

In seeing cognition as embodied, we can consider artistic activity less as the conscious, top-down, application of symbolic rules, and more as embodied, situated human experience, form arising out of physical interaction with the environment. Analysis should consider how structure emerges through process, increased understandings of which will impact on our relationships to computer technology, and its interface.[47]

Decentralized Models

How might we consider the use of A-Life in relation to embodied musical experience? Otto Laske believes that in contrast to the narrow view of technology "...as tools for producing artifacts", it is "..ultimately a tool for self-knowledge deriving from forms of situated, technologically embedded cognition."[48] In his article *Learning about Life* Mitchel Resnick talks about how the "...growing interest in Artificial Life is part of a broader intellectual movement toward decentralized models and metaphors." He suggests that the methodology of Artificial Life can help people move away from a centralised mindset and "...develop intuitions about decentralized phenomena"[49] and he suggests providing "...opportunities to create, experiment, and play with decentralized systems."

Conclusion

The arguments above point to new techniques to augment musical creativity, and new considerations in attempting to understand the compositional process. However, attempts at developing computer based autonomous 'musicians' may be inappropriate given such considerations, reflecting both a questionable interpretation of the computer's capabilities, and its relationship to social action. In relation to the various aspects of musical process, it becomes important to question the appropriateness and feasibility of computer programming; to make judgements regarding those aspects that can be abstracted such that the computer performs an integrative, rather than distancing, role in the creative process. Most importantly, in defining our creative relation to the computer we should not limit ourselves "... to the categories and procedures represented in the computer, without realising what has been lost."[50] It is perhaps more useful to see the computer more as augmentation to compositional and improvisatory activity, "...as *habitats* rather than mere *tools*",[51] and as a medium for communication, than as a potentially autonomous 'musician', attempts at which may serve to further abstract ourselves from embodied social functions of music.

References

[1] For example, D. Ihde, *Technics and Praxis*, Reidel, 1979 and D. Rothenberg, *Hand's End: Technology and the Limits of Nature*, University of California Press, 1994.
[2] A. Gerzso, "Paradigms and Computer Music", *Leonardo Music Journal*, Vol. 2(1), pp73-79, 1992.
[3] D. Rothenberg, "Sudden Music: Improvising Across the Electronic Abyss", Vol. 13(2), pp23-46, 1996. p28
[4] J. Shepherd, "Music as Cultural Text", in J. Paynter, T. Howell, R. Orton, P. Seymour (eds.) *Companion To Contemporary Musical Thought, Vol. 1*, Routledge, 1992. p128
[5] M. Resnick, "Learning About Life", in C. G. Langton, (ed.), *Artificial Life: An Overview*, MIT Press, 1995, p230).
[6] J. Pressing, "Cybernetic Issues in Interactive Performance Systems", *Computer Music Journal*, Vol. 14(1), pp12-25, 1990.
[7] S. Emmerson, "'Live' versus 'Real-Time'", *Contemporary Music Review*, Vol. 10(2), pp95-101, 1994.
[8] B. Truax, "Computer Music Language Design and Composing Process", in S. Emmerson (ed.), *The*

Language of Electroacoustic Music, MacMillan Press, 1986, p157
[9] D. G. Loy, "Composing with Computers – a Survey of Some Compositional Formalisms and Music Programming Languages", in M. Mathews & J. R. Pierce (eds.), *Current Directions in Computer Music Research*, MIT Press, 1989, p308. See also Iannis Xenakis, *Formalized Music: Thought and Mathematics in Composition*, Indiana University Press, 1971.
[10] *ibid*, Loy [9], p309. Also Lerajan Hiller & Leonard Isaacson, *Experimental Music*, McGraw-Hill, 1959.
[11] O. Laske, "A Search for a Theory of Musicality", *Languages of Design*, Vol. 1, pp209-228, 1993.
[12] O. Laske, "Knowledge Technology and the Arts: A Personal View", *Computers and Mathematics with Applications*, Vol. 32(1), pp85-88, 1996, p85.
[13] R. Rowe, "Incrementally Improving Interactive Music Systems", *Contemporary Music Review*, Vol. 13(2), pp47-62, 1996, p47.
[14] *ibid*, Rowe [13]
[15] *ibid*, Emmerson [7], p100
[16] S. R. Holtzman, *Digital Mantras: The Languages of Abstract and Virtual Worlds*, MIT Press, 1994.
[17] H. Moravec, *Mind Children*, Harvard University Press, 1988.
[18] J. Frederickson, "Technology and Music Performance in the Age of Mechanical Reproduction", *International Review of the Aesthetics and Sociology of Music*, Vol. 20(2), pp193-220, 1989, p 194. This distinction derives from R. Peterson, *The Industrial Order and Social Policy*, Prentice Hall, 1973.
[19] *ibid*, Frederickson [18], p215.
[20] E. T. Harris, "Handel's Ghost: The Composer's Posthumous Reputation in the Eighteenth Century", in J. Paynter, T. Howell, R. Orton and P. Seymour (eds.), *Companion To Contemporary Musical Thought, Vol. 1.* Routledge, 1992.
[21] M. Chanan, *Repeated Takes: A Short History of Recording and its Effects on Music*, Verso, 1995.
[22] For example, S. G. Cusick, "Feminist Theory, Music Theory and the Mind/Body Problem", *Perspectives of New Music*, Vol. 32(1), pp8-27, 1994.
[23] T. Wishart, *On Sonic Art*, Imagineering Press, 1985.
[24] For example, A. Newell, "Physical Symbol Systems", *Cognitive Science*, Vol. 4, pp135-83, 1980.
[25] For example, Wishart [23] and A. Di Scipio, "Micro-Time Sonic Design and Timbre Formation, *Contemporary Music Review*, Vol. 10(2), pp135-148, 1994.
[26] A. Di Scipio, "Inseparable Models of Materials and of Musical Design in Electroacoustic and Computer Music", *Journal of New Music Research*, Vol. 24(1), 34-50, 1995.
[27] *ibid*, Truax [8], p156. See also B. Truax, *Acoustic Communication*, Ablex Publishing, 1984.
[28] A. Di Scipio & I. Prignano, "Synthesis by Functional Iterations: A Revitalization of NonStandard Synthesis", *Journal of New Music Research*, Vol. 25, pp31-46, 1996, p42.
[29] M. Elton, "Towards Artificial Creativity", *Languages of Design*, Vol. 2, pp207-222, 1994, p217.
[30] C. G. Langton, "Preface", in C. G. Langton, C. Taylor, J. Farmer, J. Rasmussen, S. (eds.) , *Artificial Life II*, Addison-Wesley. 1992, p xiv.
[31] M. A. Boden, "Introduction", in M. A. Boden (ed.), *The Philosophy of Artificial Life*, 1996], Oxford University Press, 1996, p4.
[32] *ibid*, Boden [31], p3
[33] S. Harnad, "Levels of Functional Equivalence in Reverse Bioengineering", in C. G. Langton, *Artificial Life: An Overview*, MIT Press, 1995, p293.
[34] H. H. Pattee, "Simulations, Realizations, Theories of Life", in M. A. Boden (ed.) *The Philosophy of Artificial Life*, Oxford University Press, 1996, p384.
[35] J. Blacking, *How Musical is Man?*, University of Washington Press, 1973.
[36] S. Feld, "Aesthetics as Iconicity of Style, or 'life-up-over-sounding: getting into the Kaluli groove", *Yearbook for Traditional Music*, Vol. 20, pp74-113.
[37] M. Chanan, *Musica Practica: The Social Practice of Western Music From Gregorian Chant to Postmodernism*, Verso, 1994, p24.
[38] *ibid*, Shepherd [4]
[39] cited in Chanan [37], pp27-28.
[40] *ibid*, Chanan [37], p28
[41] See, for example, F. Varela, E. Thompson, E. Rosch, *The Embodied Mind*, MIT Press, 1991. Also A. Clark, *Being There: Putting Brain, Body and World Together Again*, MIT Press, 1997.
[42] L. A. Loren & E. Dietrich, "Merleau-Ponty, Embodied Cognition, and the Problem of Intentionality", *Cybernetics and Systems*, Vol. 28(5), pp345-358, 1997.
[43] *ibid*, Loren & Dietrich [42]
[44] E. Prem, "Introduction to the Special Issues/Epistemological Aspects of Embodied Artificial Intelligence", *Cybernetcs and* Systems, pp v- xi, 1997.
[45] *ibid*, Boden [31], p4.
[46] *ibid*, Prem [44], p vii.
[47] I. Hybs, "Beyond the Interface: A Phenomenological View of Computer Systems Design", *Leonardo*, Vol. 29(3), pp215-223, 1996.
[48] *Ibid*, Laske [12], p86.
[49] *ibid*, Resnick [5], p231.
[50] J. Lanier, "Agents of Alienation", *Journal of Consciousness Studies*, Vol. 2(1), 76-81, 1995.
[51] *Ibid*, Laske [12], p85.

Three levels of education in Electroacoustic Music: The Virtual Sound Project

Riccardo Bianchini* - Alessandro Cipriani**

*Conservatorio di Musica "S.Cecilia", Rome, Italy
**Istituto Musicale "V.Bellini", Catania, Italy
EDISON STUDIO - Rome

*via Ternana, 108 - 02034 MONTOPOLI S. (RI) ITALY
**via Voghera, 7 - 00182 ROMA ITALY
rb@fabaris.it a.cipriani@agora.stm.it
*http://www.geocities.com/Heartland/Acres/4768 **http://www.axnet.it/edison

ABSTRACT

This is a presentation of the first published Computer Music Education project in Italy, that includes a series of printed textbooks, interactive courses via the Internet and a series of CD-ROMs. The aim is to create a deeper and more wide-spread knowledge in the fields of electroacoustic composition, digital sound synthesis and processing.

1. INTRODUCTION

The Virtual Sound Project project consists of three main areas: a series of printed textbooks in Italian (English and possibly Spanish translations are planned), a set of courses via the Internet (complementary to the printed textbooks) and a series of CD-ROMs. The authors teach Electroacoustic Music in the Conservatories of Rome and Catania, and have for some time felt the need for such a project. The project is aiming to create a bridge from the very beginning to an advanced level for non-English speakers in Italy in order to provide them with a level of knowledge that will let them feel comfortable with electroacoustic music fundamentals and more.

2.1 *IL SUONO VIRTUALE*

The first result of this project is the book *Il Suono Virtuale*, the first textbook on sound synthesis and sound processing in italian (with a CD-ROM enclosed), published by ConTempo in May 1998 and available through its website (www.axnet.it/contempo).

The special feature of this textbook is that it includes *theory* and *practice* together, by means of lessons and examples with Csound. For each kind of signal processing and synthesis technique there are discussions about theory and history followed by immediate practice with Csound exercises. The textbook is, therefore, not only a tutorial on synthesis, but also the most extensive tutorial on Csound in any language.

In almost every chapter the student finds a theoretical part (covering the basics of a particular synthesis/processing method), a practical part (with Csound orchestra and score examples and exercises), and an "in depth" part, including more advanced uses and ideas concerning each method, not essential for the comprehension of the chapter, but useful in helping the student to broaden his/her knowledge.

Il suono virtuale includes an introductory chapter on the fundamentals of Csound (*Csound: how it works*); additive, subtractive, amplitude modulation, frequency modulation, waveshaping, FOF, granular and physical modelling synthesis; the use of sampled sounds, analysis/resynthesis (by means of *hetro/adsyn*, *pvanal/pvoc* and *lpanal/lpc*), delay, echo, reverberation, chorus, flanger, phaser, convolution; the basics of flow-charting; the fundamentals of digital audio; the use of MIDI files; the real-time use of Csound; Csound from the viewpoint of a programming language, etc.

The first Appendix is dedicated to WCShell (a Windows® NT/9x front-end for Csound written by Riccardo Bianchini); the second Appendix concisely covers the mathematical and trigonometrical fundamentals required for a "well behaved" use of Csound (or any other synthesis program).

The book concludes with five advanced *readings* about: real time granular synthesis for Csound (by E.Giordani); sound synthesis by means of non-linear functions iteration (by A.Di Scipio); Max and Csound (by M.Giri); generation and modification of Csound scores by means of general purpose programming languages (by R.Bianchini); Csound and Linux (by N.Bernardini). James Dashow kindly wrote the introductory preface.

Il Suono Virtuale provides also a listing of the main Csound Web sites.

A double format (Windows/PowerMac) CD-ROM includes all orchestras, scores, soundfiles, analysis files used throughout the book. Most importantly, the CD-ROM contains Csound for Windows and PowerMac, WCShell for Windows, and the HTML version of Web sites file.

The first five paragraphs of Chapter 1 can be downloaded from Contempo web site.

Since there was no other tutorial in Italian about synthesis, and most of the material for learning synthesis techniques is in English, italian students had to rely primarily on lessons in class if there was an opportunity to study electroacoustic music in school. The lack of a published project for electroacoustic music education in

the publishing field left most of italian computer music enthusiasts in the hands of the MIDI keyboards market, which hardly promotes a systematic education. It was our notion that a deeper and more wide-spread knowledge about computer music fundamentals and a direct practice with a software like Csound that helps to secure the appropriate skills can make a difference.
The difference between our approach and the commercial approach lies primarily in promoting awareness of each single act involved in the compositional process. The commercial approach merely creates user-friendly black boxes as far as technical issues and the creation of each single sound is concerned (*push this button and you'll obtain that; don't ask why, just learn all the buttons! And the presto-patches!*).
We believed that some action had to be taken in order to promote such awareness, and *Il Suono Virtuale* is our first step. We chose to present the material in as clear and simple a style as possible, to start from the very beginning and always to provide the English for the most basic terms: this should also ease the reader in approaching the literature in English.

2.2 *WCSHELL* AND INTEGRATION WITH THE TEXTBOOK

As the power of personal computers increases, Csound becomes faster and faster. Two of the main problems with Csound, however, are the non-graphical approach (inherited from Unix text-only user interface) and the text writing of the score on a *note by note* basis.
The use of *WCShell* for Windows NT/9x helps to create a different approach to Csound. This front-end for Csound includes the possibility of graphically creating functions, to convert a Csound score into a spreadsheet and generally work in a faster way, or to write a score by drawing lines in a field that represents durations, frequency values etc. In the spreadsheet software called *Scorex*, for example, the notes and the functions are put in the rows and parameters fields in the columns. Therefore larger groups of notes can be created at once filling spaces with particular values, adding, multiplying, applying functions to groups of cells values, using linear or exponential interpolations between initial and final values, converting amplitudes and frequency values from/to any format etc. Via *WCShell* it is also possible to edit the analysis file produced by *hetro*: for example you can graphically re-design the frequency or the amplitude values of all the harmonics. It is also possible to save the flags configurations and eliminate a significant amount of other time-consuming activities.
R.Bianchini is now working on *CSGraph* (see Fig.2), a new Windows software by which the user can generate Csound orchestras by using a graphic, object-oriented interface. These are just some examples of how this set of programs can help the student to have a more intuitive approach to synthesis and thereby help him/her overcome some of the difficulties of Csound's steep initial learning curve. With the new implementation of real-time Csound by Gabriel Maldonado (this uses DirectX, a Windows method to drastically reduce latency time in sound input/output), the interaction with real-time MIDI control becomes very appealing, and we believe that the combination of a textbook like *Il Suono Virtuale,* plus the use of a front-end for Csound like WCShell combined with the use of satisfactory real-time control can provide an environment for education in which the interaction between theory and practice achieves a rare degree of integration.

2.3 *VIRTUAL SOUND*

The book *Il Suono Virtuale* is now being translated into English, since the interest for this kind of approach goes well beyond that of an exlusively italian audience.

2.4 *CINEMA PER L'ORECCHIO*

The first step of the educational project will continue with a series of printed textbooks called *Cinema per l'Orecchio (Ear Cinema)* (series editor Alessandro Cipriani, publisher Contempo). As well as the above mentioned textbook, this series on electroacoustic music education and production, multimedia and intermedial art includes:
a) A textbook on acoustics and psychoacoustics
b) A textbook for ear-training, practice of analysis of the "*texte sonore*", study of the interrelationships between new sounds for acoustic instruments and electroacoustic sounds seen from the spectromorphological and compositional point of view
c) A textbook on studio techniques, multimedia practice, intermedial art
d) A textbook on the history and analysis of electroacoustic music

3. THE SECOND STEP: COURSES VIA THE INTERNET

Before approaching an interactive expansion and re-issuing of the first project on CD-ROM, we are experimenting with students' interaction via the Internet. We are working with the Edison Studio in Rome which will host the project in its web site.
This step, which we plan to start in November 1998 and which should continue for two months, will consist of the following:

3.1 A series of on-line lessons, which will take as a basis the first few chapters of the book *Il suono virtuale*, and will exploit the same arguments, treated in deeper detail, and with a much larger quantity of examples (Csound orchestras and scores).

3.2 At the end of each lesson the student will practice with incomplete orchestras and scores , orchestras and scores with syntax errors, orchestras and scores without syntax errors, but containing subtle bugs. Furthermore the student will find a series of tests. He/she will send by E-Mail his/her

orchestra and score corrections, and his/her test answers, and will be graded accordingly.

3.3 The student can ask a maximum of 20 questions, which will be put on line on a special page of FAQ (Frequently Asked Questions).

3.4 At the end of the course, the student, if he/she has a sufficient cumulative grade, will be rewarded with a certificate allowing him/her to start the next course.

4. THE THIRD STEP: INTERACTIVITY AND CD-ROM

This is obviously the most problematic of the steps, expecially regarding its funding. We are only just beginning to work on this idea, and its realization depends on various factors, including the availability of a greater number of experts.

In our project the CD-ROM reissue of *Cinema per l'Orecchio* should be interactive, multimedial and multilingual.

Interactive: the CD-ROM should exploit the maximum of interactivity, with real-time generation of sounds by means of Csound, and interactive correction of tests, so that the student can use a trial-and-error approach. An on-line method of updating should allow the user to enrich his/her repertoire of orchestras, scores and sounds.

Multimedial: the CD-ROM should contain animations, sounds and text.

Multilingual: the CD-ROM should be in Italian, English and other languages such as Spanish.

SUMMARY

The Virtual Sound Project has just been started by the authors, Contempo and Edison Studio. The first book in italian is already available, the interactive course will start in November 1998, the English version of *Il Suono Virtuale* should be available for the ICMC '99. We hope to be able to publish other two textbooks within year 2000.

A REAL TIME ALGORITHM FOR STEREOPHONIC LOCALIZATION OF MOVING SOUND SOURCES

Riccardo Dapelo dapelo@belva.infomus.dist.unige.it
Simone Macelloni smmac@tin.it

Laboratory of Musical Informatics – http://musart.dist.unige.it
D.I.S.T. Università degli Studi di Genova, Viale Causa 13 16145 Genova, Italy

Abstract

This paper presents a real time stereophonic version of the model proposed by Moore (1983). The listener is supposed to stay in the center of a room (the inner room) defined in meters, with a loudspeaker in each front corner. The loudspeakers seems like "holes" in the walls that allows to listen a moving sound source in an illusory acoustic space (the outer room). For each sound source is computed a series of direct and reflected paths, simulating the early reflections of the source against walls. This realization is slightly different from Moore' model because of the addition of a filters bank for enhancing binaural perception of side or rear movements of the sound source. The algorithm (first developed in Csound language) requires the previous definition (in meters) of the inner and outer rooms and allows the dynamic control of the position and movement of a sound source. Actually has been ported on the Iris-Mars workstation for real-time development in a stereophonic version using ARES (Audio Resource Editing System). The algorithm, described before, is subdivided into nine modules that cooperate in an orchestra file fulfilling the spatial processing of sound. Two modules determine the source position, communicate it by an internal bus to the other algorithms and filter the input signal if the source position is behind the inner room. Four algorithms compute the distances covered by the sound signal with its reflections over the outer room four walls, determine the delays and apply them to the sound source. Another one processes the calculation of the delays of the direct sound's path to the holes of the inner room (the loudspeaker positions) and delays the input sound by an interpolated delay. At last two modules sum up direct and delayed signals giving separated outs to the right and left loudspeakers.

Introduction

The aim of the present work is to develop a real time system for stereophonic localization (spatialization) of moving sound sources. The algorithm is intended for live performance and obviously requires portability and speed of calculation time. The chosen model is useful for this purpose because can simulate the illusion of sound sources in motion with few calculations. The computing of the early reflections in a room with a simple geometric shape, in this case a square, gives sufficient perceptual informations and allows the illusion of a moving sound source.

The Model

As well known, the Moore model (Moore 1983) is a general model of spatialization, based on the computing of the early reflections of sound. It uses the metaphor of the two rooms (see Fig.1) and computes the path of a sound with a network of tapped delay lines. For each sound source are calculated a direct path and four reflected paths over the outer room walls to each loudspeaker, according to Moore formula: $N_{path} = N_{vect} N_{spkr} (1+N_{wall})$, where N_{pat} is the total number of computed paths, N_{vect} is the number of vectors (the sound sources), N_{spk} is the number of loudspeakers and N_{wall} is the number of reflecting surfaces (the walls). The length of each path (in meters) is obtained by means of the "image method" (see Fig.2) and determines the delay and the gain of the reflections. In this implementation are used only two loudspeakers placed in the left and right front corners of the inner room and so obtaining, for a single sound source, a total number of 10 paths (2 direct and 8 reflected).

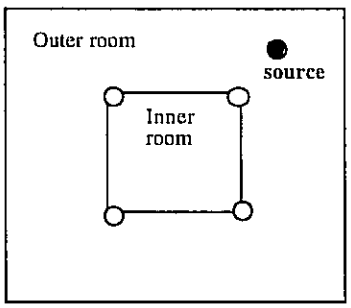


Fig.1
The two rooms model. The white circles represent the "holes" in the wall of the inner room (the loudspeakers), that transmit the sound, direct and reflected, over the wall of the outer room, to the listener.

Features

In this project some different features, tested in different configurations and reasonably approximated by practical checks, have been added to the original model (see Fig.3):

- The choice to think the walls of the inner room not absorptive, allowing the sound paths passing trough the walls in every position of the source. This choice was made during several tests about circular motion, in which

was found a consistent loss of intensity when the source moves around the x-axis (behind the rigth or left walls)

- Two dynamic filtering zones for enhancing side (left and right) movements of the sound source. Experimental tests demonstrate that such filtering seems to compensate the perspective distortions due to the relation between the loudspeakers position and the listener, specially for the critical case of a direct path perpendicular to the ear of the listener (around the x axis). These zones are obtained with two band-pass filter, centered at 3 kHz with a bandwidth of about 2 kHz, which are dynamically linked to the position of the source (see Fig.3). This filtering is reserved to the direct paths.
- A rear-filtering zone, simulating the perception of a sound source reaching the outer ear of the listener from behind. These are obtained with a high-pass filter dynamically linked to the source. The filter range start from 150hz (maximum filtering) to 20000 Hz (no filtering) according to the source position. This filtering is applied both to the direct paths and to the reflected paths on the wall 3.

The sum of the direct and reflected paths (filtered or not) is simply routed to a global reverberator, according to Chowning formula for the intensity of reverberated signal: *1/sqrt (dist)*, where *dist* is the distance (in meters) of the sound source. The reverberated signal is finally mixed with the dry signal, with variable percentage and then sent to the output.

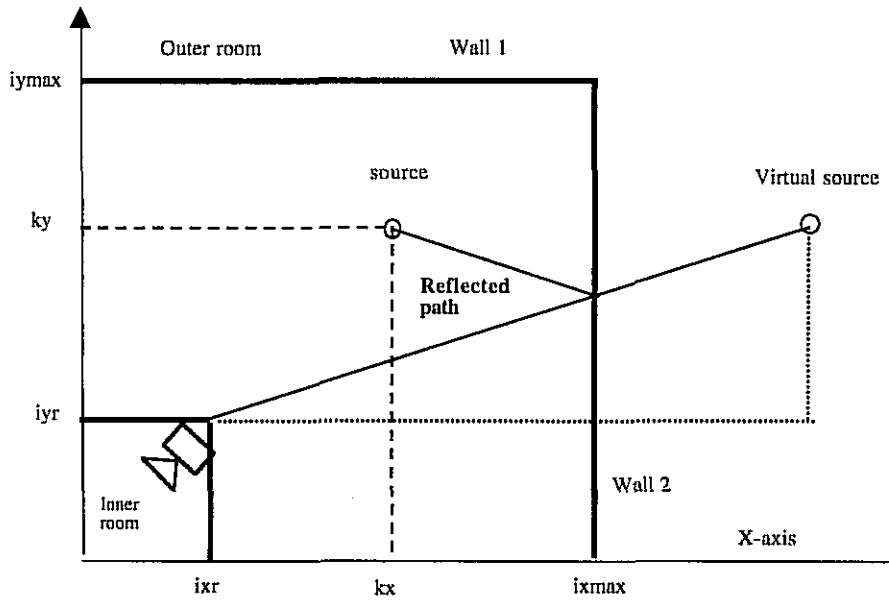


Fig.2
Computing of a reflected path by means of image method. The virtual source is considered symmetrically placed in relation to the reflecting surface.

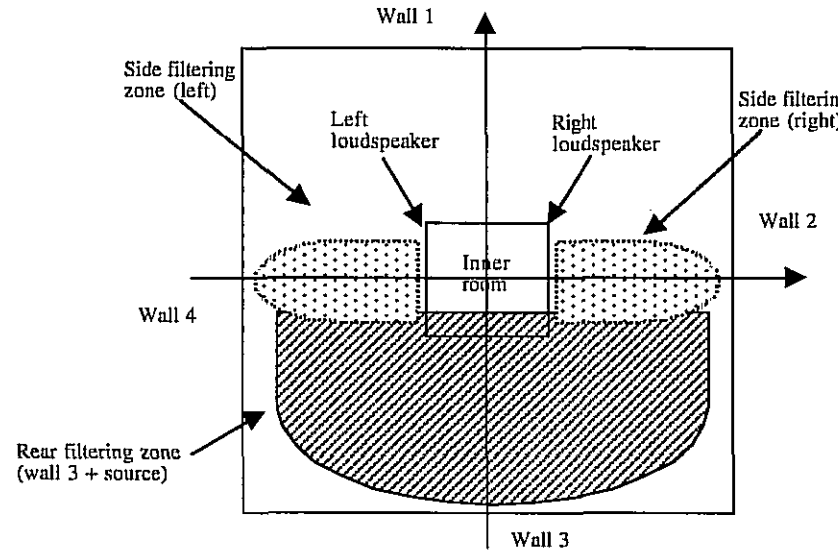


Fig.3

Reference model of the algorithm.

The porting on the Mars workstation

Since this algorithm is intended for live performance, it has been subdivided into two parts:

- an application running on a Win95 pc, called *SpaceSound*, that allows to control the sound source position;
- an *ARES* (*Audio Resource Editing System*) orchestra file, running on *MARS Workstation*, that executes the spatial processing of the sound source.

SpaceSound is an application, fulfilled in C++ programming language, that supplies an easy-to-use visual interface to set the dimensions of the inner and the outer room and to control the position of the sound source, moving it across the room using the mouse (see Fig.4).

To reduce the computational load of *MARS Workstation*, improving the efficiency of the real time processing, *SpaceSound* carries out the calculation of the distances between the sound source and the loudspeakers positions (left and right), along the direct and the reflected paths.

The communication of these distances, the dimensions of the inner and the outer rooms and the Y-coordinate of the source position (used to determine the dynamic filtering zone) to *MARS Workstation*, is executed using a *Midi* connection.

The *orchestra file*, carried out using ARES tools, is constituted by a set of algorithms linked in the orchestra environment. As you can see in Fig. 5, the distances computed with SpaceSound, are the input of five modules :

- Th first one receives the distances of the direct paths and calculates the time used to run along the paths simply multiplying the distances by *1/velocity of sound*; then it delays the input sound source for the time calculated using the *interpolated delay* module that ARES supplies. If the sound source is in the left or right *side filtering zone* (see Fig. 3) the signal is filtered accordingly to the model described before. Moreover this module dynamically filters the input sound source if the source position is in the *rear filtering zone* as you can see in Fig. 3.
- The wall's reflection modules of wall 1, 2, 4 delay the input signal.
- The module "*Third Wall Reflection*", delays the input sound, and low pass filters it accordingly to the model described before.
- Each module operate a gain adjustement according to the formula $1/dist$, where *dist* is the distance (in meters) of the computed path

All the output signals are added up on two different busses (left and right) and sent to two ARES reverberation unit (slightly different).

Future development

Other developments, such as the control interface software and hardware for real-time performance (with joystick and/or sensors) and the porting on the MAX-MSP environment of this algorithm are actually in progress.

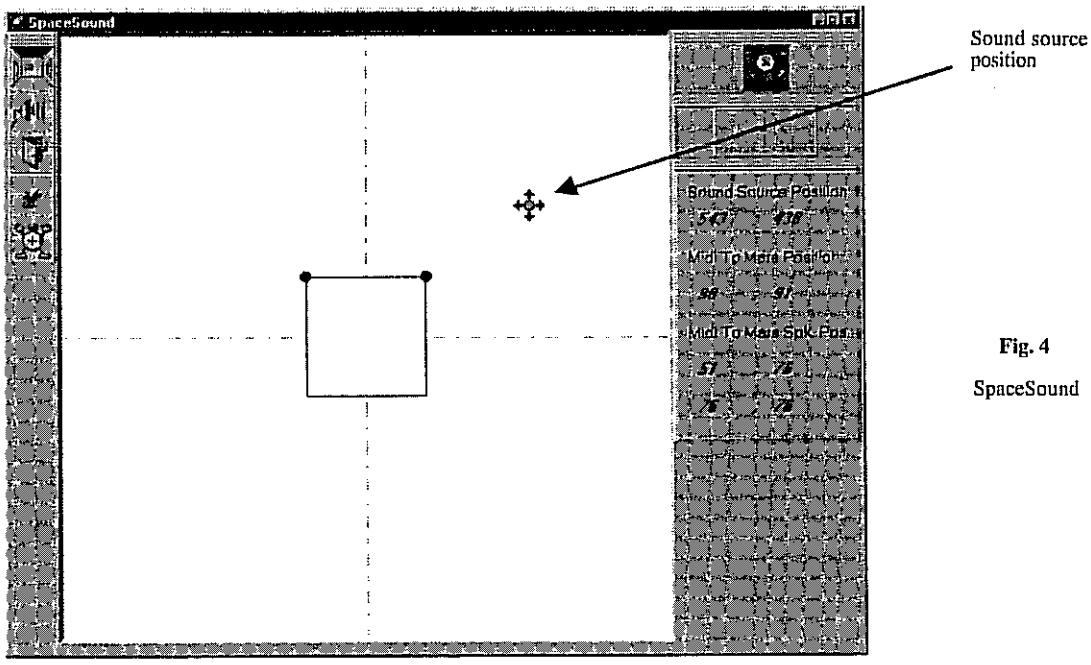


Fig. 4

SpaceSound

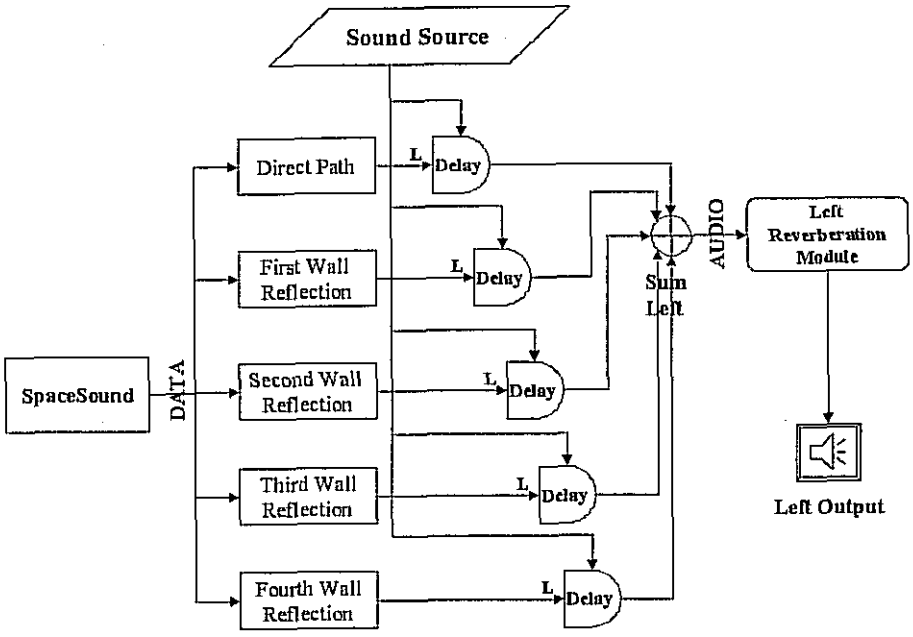


Fig. 5

Structure of the algorithm for the computation of the left audio output signal (the same is used for right)

References

Chowning J., *The simulation of moving sound sources,*
Journal of the Audio Engineering Society, vol. 19, n.1, 1971

[Moore 1983]
Moore F.R., *A General Model for Spatial processing of sounds,*
Computer Music Journal, vol. 7, n.3, Fall 1983, MIT Press

Mozzoni L., *Realizzazione di un sistema di spazializzazione del suono in tempo reale,*
Tesi di laurea, Relatore De Poli G., Correlatore Rocchesso D.,
DEI-Università di Padova, PADOVA 1995

Roads C., *the computer music tutorial,*
Cambridge, 1996, MIT Press.

Rocchesso D., *The Ball within the Box: A Sound-Processing Metaphor,*
Computer Music Journal, vol. 19, n.4, Winter 1995

AA.VV. *Musical Audio research Station – User's guide,*
IRIS S.r.l., Edition March 1994, version1.1

Gramma: the new music in an old architecture

Maria Cristina De Amicis, Mauro Cardi
Istituto GRAMMA
Via degli Scardassieri 14 67100 – I – L'Aquila
e.mail: mc.deamicis@usa.net; mcardi@iol.it; gramma.it@usa.net

Abstract

The composers working at GRAMMA Institute have conceived and realized some events where they tried to experiment with new forms of performances in connection with a dual research, concerning listening and viewing.
The presence of a not traditional instrumentation, somehow connected to the *imaginary* and *fantastic*, has characterized the Institute's activity, from the beginning orientated to the realization of a project untitled CORPI DEL SUONO. This project, based on a festival and an exhibition of very special instruments, proposed the most complex aspects in the scientific and musical researches in a spectacular way, in order to simplify the understanding of them and, at the same time, verifying the result in the audience about an aesthetic intervention.
In the following article we want to illustrate the different fields in the research and production by means the GRAMMA Institute carries on its own activity.

1. Substantial relation: music and technology

GRAMMA Institute was established in 1989 with the restoration of the baroque church St. Caterina d'Alessandria, located in down-town in L'Aquila, in order to promote contemporary music, making use of the most advanced technologies.

The project, ambitious but stimulating, was to carry out CORPI DEL SUONO based on the contemporary presence of both a festival and an exhibition of not traditional instrumentation (Fig.1). This experience has shown the most difficult aspects of the scientific-musical research in a spectacular way that has made easier its understanding and checked its expertise of aesthetical participation.

In every edition of CORPI DEL SUONO we have shown a study of the fundamental work themes of GRAMMA Institute: the use of advanced technologies both for composition and concert performance and the interpretation of contemporary music. Both the themes are strongly connected and many contemporary works require a qualified approach, both from the organitation and from the audience, so that each part of the artistic message finds coherent communicative mediation.
Just from the point of view of the organitation, the first problem we must have faced and settled, was to adapt the acustic-scenic features of the place where the event takes place to the needs of the musical works in program. We make use of sophisticated systems of hearing, based on many loud-speakers, of the placement of interpreters in the best points of the theatrical space and of the building of variegated, visual environments; this tecnique has permicted us to support image with sound through coherent criteria according to the expressive will of the musical work.

The versatile space of St. Caterina d'Alessandria has proved itself as precious for this kind of events.

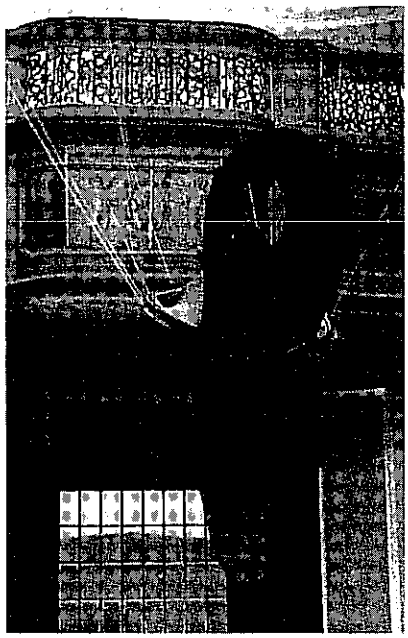


Fig. 1 ST. CATERINA D'ALESSANDRIA CHURCH
Arpa Eolica - Corpi del Suono 1990

The whole work by GRAMMA Institute is strongly linked with the idea of experiment, one of the essential starting-points also for musical creation with the

involvment of traditional interpreters, who have become, thanks to the technological development, the guided authors of processings in real time (Fig.2).

We have lived at GRAMMA Institute what, for some time, was discussed in the most important meetings on the evolution of Electroacustic Music: the birth of new professionals.

Two fundamental and connected phases are included in the musical production: both the creative and the executive phase. For the composer nowadays, this means that the creative phase is constantly accompanied by an experienced person able to add the musical knowledges to the scientific ones. The need to train these new professionals has lead us to the carrying out of an event capable to check the state of didactics in the school Institutes dealing with Electroacustic Music.

Fig. 2 COURTS DRAGONETTI PALACE
Corpi del Suono 1997

With the biennal event LA TERRA FERTILE, we have planned and carried out a Symposium devoted both to the training of young musicians and to the professional specialization in informatics applied to music: composition, interpretation and performance, sound effects [1].

2. Performance

In the different editions of the event CORPI DEL SUONO, a more developing trend of experimenting the new spectacular forms of the event-concert has grown up contemporary to a program, so to say, more regular, based on the proposal of mainly electroacustic, historical or contemporary works. In these events GRAMMA Institute defines its artistic activity through the whole human e technical resorts. The main part of these events are idealized and produced by GRAMMA Institute (Fig. 3).

They based themselves on the use of new technologies in support of the spectacular performance that the composers and the other artists each time require. The research today, more than in the past, is a basic step for the activity of the production-centres in the field of Electroacustic Music; it is more and more aimed at the production of works and the carrying out of projects taking on account, since their idealization, the fundamental questions concerning musical events in the round. We begin from the thoughts on the audience who one is addressed to, the matters of hearing psicology, the problems of the new music perception, till the considerations on the acustic and the architectural space where music is purposed.

The scientific research, lead in the Electroacustic field, then, is necessarily linked with the expressive research of the artists, the first, on the contrary, finds incentives, tests and legitimations in the second one. In our centre, like in many others, composers, interpreters, visual artists on one hand, scientists on the other hand, work in close contact. We are sure of the validity of this way and the whole activity of GRAMMA Institute is addressed to this aim and to the consequent idea of an open-centre, living and enriching of several contributes.

In order to extend the reading levels and deep the hearing of the purposed works in the various editions of CORPI DEL SUONO, we have developed a theme for each concert and studied an environment in unusual places, cutting the space in several parts, through panels, lights and projections. The fascinating but hard hypothesis is to carry out concerts where the audience would have the possibility to move without loosing the information coming from the performance, on the contrary, amplifying the perceptive experience as regards sounds and performed music.

We have carried out the concert distribuitable, that is to say, thought to be listened continually in all places guesting the concert.

The audience could follow a sort of free route, where several equipped stations with texts and projections, have permicted to follow the development of the concert.

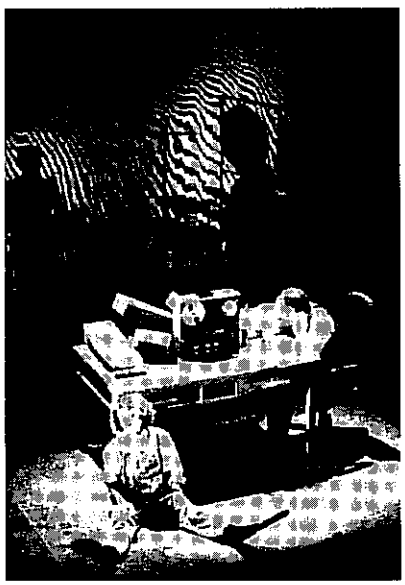


Fig. 3 MUSICAL THEATRE
Il Mondo di Sopra, di Sotto, di Dentro - Corpi del Suono 1997

Our inventions are addressed to the multiple sonority of the environments and to their acustic specialization. In this way, we have exalted the specific resonance of places so that the musical work, spread contemporary in the whole environment, could be appreciated from the audience according to different hearing formalities (Fig. 4). The reflections coming from each room carachterized by architectural geometries of stairs, ceilings, stone and marble walls, have been handled with many loud-speaker systems, of different size and power, placed so as to permict that the audience could appreciate the variations of the same work in different acustic spaces.

Fig. 4 METAMORFOSI DEI CORPI
Performances produced by GRAMMA Institute in 1998

3. Music applied research

The high professional level audio technologies of the service Agorà, put at our disposal, have made possible the experimentation, for years, on sophisticated techniques of recording, spatializzation and audio installations.

The team of our researchers is now able to supply the musical productions with both technological and theoretical relevant resources giving the musician a way for new and deeper electroacustical experiences.

The our group have gained a considerable algorithmically experience on real time digital signal processing with use of the sophisticated informatical systems available at the institute. these system are designated to process the musical signal at any level and are completely programmable. This great flexibility allow the composer to redirect the algorithmically research to his own needs.

The acquisition of the application software developed at the Ircam permits a more flexible access to midi instrumentation.

Every collaborator of GRAMMA Institute is able to develop programs oriented to each individual composer, because his familiarity with many informatical languages and software systems.

4. Communication

The constant test and the exchange of ideas inside the work-group of GRAMMA Institute has permicted to face another aspect of music spreading (particularly contemporary music) not connected to the hall where the concert takes place.

If someone is interested in hearing the purposed works in the different performances of CORPI DEL SUONO, could connect with an Internet site and listen in real time the concert or part of it. The idea is to purpose, through the hearing in net, a virtual image of what is happening where the concert take place. As regards the technical point of view it was simple: it was sufficient to be connected by a modem and use the real Audio software. More complex was the work that GRAMMA Institute has produced in order to permict the transmitted sound not to loose its original brightness.
We have carried out an algorytmh by the System Fly30 of CRM, capable to codify the signal without losses.
The selected tecnology for the implementationof the initiative is based on the encoders and on Progressive Network Real Audio and real Video server.

[1] M.C. De Amicis *"La Terra Fertile: didactis and innovation"*, in the Proceedings XII CIM, 1998

References

M. C. De Amicis, M. Lupone, "*Note di presentazione*", Corpi del Suono, Ed. Gramma, L'Aquila 1989
M. C. De Amicis, "*Corpi del Suono*" in Regione Abruzzo, L'Aquila 1990
M. C. De Amicis, "*Note di presentazione*", in Corpi del Suono, Ed. Semar, L'Aquila 1990
M. C. De Amicis, "*La Terra Fertile: una necessità musicale*" in Suono Sud n.22 1994
M. Cardi, "*Il grado xerox della cultura. La musica contemporanea nell'epoca (pubblicitaria) della fotocopia*" in Suono Sud n.22 1994
M.C. De Amicis, "*Suoni in un piatto di rame. L'Aquila: una città laboratorio*" in Suono Sud n.24, 1995
M. Cardi, L. Ceccarelli, "*Live electronics*" in Il Complesso di Elettra, Ed. Cidim-Cemat 1995
M. Cardi, M. C. De Amicis, M. Lupone, "*Note di Presentazione*", in Corpi del Suono, Ed. Gramma, L'Aquila 1995
M. C. De Amicis "*Note di Presentazione*", in Corpi del Suono, Ed. Gramma, L'Aquila 1996
M. Cardi, M. C. De Amici, I. Prignano, "*Gramma: uno spazio per la musica contemporanea*", in the Proceeding La Terra Fertile, Ed. Gramma, L'Aquila 1996
M. C. De Amicis "*Note di Presentazione*", in Corpi del Suono, Ed. Gramma, L'Aquila 1997

La Terra Fertile: Didactics and Innovation

Maria Cristina De Amicis
Istituto GRAMMA
Via degli Scardassieri 14 – 67100 L'Aquila
e-mail: mc.deamicis@usa.net; gramma.it@usa.net
http://www.webaq.it/gramma

Abstract

The narrow correlation among the scientific and technological research sets the Electronic Music in a circle where different disciplines collaborate. The didactic aspect of this peculiar discipline needs to make people aware of how much and which professional levels it must invest. That's why we need a consistent professional formation able to satisfy the demands of the production of the research and of the musical interpretation. LA TERRA FERTILE is useful for the discussion of the programs and the results achieved from different Italian Conservatories, pointing up a renewal of the programs that through the computer, physical and musical studies distinguish three moments: the composition moment, concerning the creation of sounds and structures, the interpretative moment, concerning the control of the sound given out by electronic or electroacustic sources, the historian moment concerning the process of analysis and resynthesis. The conference based on the presentation and the debate around the didactic works gave evidence to the strong coincidence of the intents and the methodology among the different schools. This means that the cultural patrimony that the Electronic music has produced until today, could find a coherent institutional affirmation in Italy and spread, through the school in a more capillary way, the idiom and the meanings of the scientific-musical research that characterizes this discipline.

Didactics and Innovation

Six years have elapsed since when, during an examination session of Electroacustic Music, the heterogenity of the compositive, analytic, informatic and technological subjects, stimulated the commisssion to purpose a meeting among all the Electroacustic Music schools of the Italian Conservatoires in order to check the state of this particular discipline.
The basic aspect of the proposal concerned didactics, with the aim of producing what occurred in the compositive and technological circle, that is an exchange, a comparison and a choose of trend lines.

LA TERRA FERTILE, that is the title of the meeting, would have been directed on methods and didactic contents, supported by a thick thread of scientific papers, the presentations of the Italian Research Music Centres and the most recent musical technologies.

The importance of creating an involvement of young forces and the opportunity of comparison and elaboration of the didactic, artistic and research experiences, drove us to realize the idea of this meeting; so we began to manage the work, conscious of the precious support we would have received by the teachers of the Electroacustic Music Courses coming from the Italian Conservatoires.
In taking contact with the Conservatoires, the Universities and the Research Centres we took on account the thick thread of both personal and institutional relationships that scientists and musicians had already woven among them.

The first edition, in 1994 (Fig.1), counted 180 participants including audience and reporters coming from each part of Italy, to witness the successful initiative, mainly for the natural connection established among the different geographic and institutional realities. Fourteen Conservatoires and eighteen Research Centres and Musical Production presented their activities inside five sessions:

- Analysis and Didactics
- Composition
- Scientifics
- Electroacustic works listening
- Presentations and Demonstrations

The papers produced at the Symposium attracted attention on the different approaches of the Schooles, but also a natural convergence on the organizing and content themes.

The session of Analysis and Didactics is concentrated particularly on the analytic formalities of historical and rilevant technical-scientific passages; the compositive session alternated the papers describing musical projects to those ones of aesthetical and poetic genre; the scientific session focused the synthesis sound signal aspects and the hardware and software systems devoted to it, even if balanced the number of the theoretical and applicative papers. Many papers of the Centres treated deeply the historiography of the first Italian experiences on Electroacustic and

Computer Muisc. The presentation of the production activities of the Centres put conveniently on evidence their different technological and realizing statement, soliciting consequently an important thought on the limitations that the heterogenity gives both the composer and the researcher.

The most important issues of this first Symposium are emphasizes as follows:

→ the contact among the students of the different schooles has favoured the cultural exchange and some common projects;

→ the presentation of young musicians to the community and the spreading of their work;

→ the contact between the training reality (Conservatoires) and the Research and Production Centres has allowed to several students to establish a work relation.

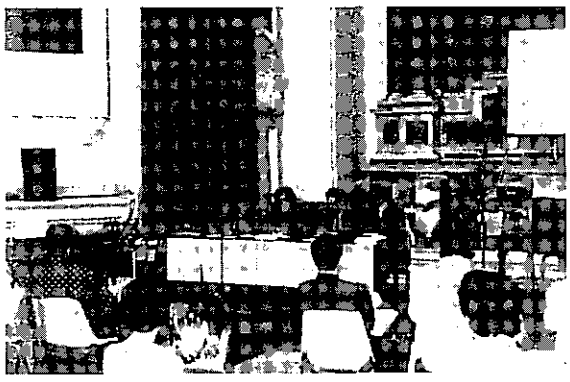


Fig. 1 LA TERRA FERTILE '94
from right G. Di Giugno, M. Lupone, S. Prodigo, G. Verna, D. D'Alfonso

At distance of two years, in 1996 the second edition of LA TERRA FERTILE, focused its attention on the pedagogic criteria and on the most advanced technologies of research, particularly related to the systems of sound spatializer.

The whole event was planned relating the demonstrative and workshop sessions to the Symposium ones.

The different hardware and software systems produced at audience's disposal (*Mars*, *Smart*-IRIS, *Fly*30-CRM, *Kyma*-SYMBOLIC SOUND, *Spatializer*-GRISBY MUSIC) gave a great contribution as regards the experimentation that students, teachers and researchers carried out on the algorithms of synthesis and sound processing.

A substantial part of the Symposium explored the complex aspects concerning sound spatializer through the theoretical interventions and work applied analysis. This prevalent theme, was integrated through high profile didactic papers produced by the students attending the Electroacustic Music schools; they are directed to a vast range of subjects so that we grouped the Symposium works in the following sessions:

- Analysis and Didactics
- Scientifics
- Listening
- Scientific-music centres Activities
- Informatic assistment to the music composition
- Historiography, Writing and Documentation
- Integrated Systems for Composition and Music Performance
- Technologies for sound spatializer
- Theory and praxis on sound spatializer

The variety of arguments treated during the second Symposium put in evidence the interactivity among the Electroacustic Music School and many aspects of music, scientific and technological fields.
The natural evolution of musical language converges with the development of theories in the scientific circle and this aspect joins many experiences increases the formative value of this school. If this represents a recognizable pedagogic principle, more substantial was the meeting in that occasion of the whole Italian teachers (Fig.2). They discussed and choose a common strategy of interventation for the study programme defined by *Education Ministry* and for a more homogeneous structure of diploma examinations.

Among the aspects related to the studio programme, the need to give space and content also to new specialization forms, parallel with the compositive one. New professionals have been emphasized, such as "sound production interpreter, informatic-music assistant", assuming a preminent position in the work contest of Electroacustic Music.

Fig. 2 LA TERRA FERTILE '96
Electroacustic Music teachers meeting
(*from left* L Camilleri, R. Bianchini, E. Giordani, A. Vidolin, A. Di Scipio, R. Doati, M. Lupone, A. Cipriani, G. Nottoli, N. Bernardini, L Ceccarelli)

The different pedagogic competence that each specialization requests, is supported by choosing didactic programs suitable to the scientifc, informatic progress and to the musical language development.
Taking on account this prospective, the teachers of the Electroacustic Music Courses by choosing the study

programs, have regarded all the aspects giving this discipline a formative expertise in the compositive, executive and technical field. The evidence of the pedagogic task in the direction of multidisciplines is given by the theoretic and practic contribution offered by the students of Electroacustic Music Schools to the works of the event LA TERRA FERTILE.

The didactic works and the scientific contribution produced, emphasize as the way of Electroacustic Music is strictly connected with the following definition:

→ new expressive means
→ new aesthetical meanings
→ more complex linguistic articulations

Just these elements create the training of new professionals, suitable both to the interpretative needs and to the coherent mediation between compositive thought and electroacustic or informatic realization.

The Electroacustic Music is going on its own evolution and occupies innumerable fields of arts and science; sometimes it's the cheapest means for the musical production addressed to the artistic areas such as: cinema, theatre and dance.

The operators in the didactic field of the electroacustic music are requested of a necessary consciousness of the numerable professional levels that can be occupied. This demands pedagogic chooses capable to satisfy the production, research and musical interpretation needs.

The reflections made till now have been deducted by the LA TERRA FERTILE '94-'96 works.

The event, created with the aim of promoting the largest spreading of informations in the didactic field, represents nowadays the unique example in that area and it's an incentive and a work tool for everyone who is involved in the musical and scientific research directed to the professional training.

The meeting among all the Italian Schooles, through this biennal Symposium, has focused the strong coincidence of common aims and methodologies.
This means that the cultural heritage that the Electroacustic music has produced till now, can find in Italy a coherent institutional assertment and spread through school institution in details, the language and the meanings of the scientific-musical research characterizing this discipline.
An essential contribution to the diffusion is offered by the performance of musical works.

LA TERRA FERTILE dedicates the largest space to the performance of the works produced by the students attending the Italian Conservatoires.
In order to increase the spectacular aspect and put in evidence the recent productions of young composers, all the concerts have been inserted inside an event that annually presents in L'Aquila contemporary music concerts, performances and multimedia installations.

This has permicted to offer to a large audience the multiplicity of contemporary languages, suond ideations, instrumental tecniques, the innovative technoligies used by students.

The third edition of LA TERRA FERTILE takes place in 1998 in the impressive scenary of the Spanish Castle in L'Aquila and in the church of St. Caterina d'Alessandria.

The constant presence of all the Italian Conservatoires is amplified this year for the participation of some of the most important European realities involved in the didactic field and in the electroacustic music research.

GRAMMA Institute, in order to offer to the musicians taking part to the event, a valid aid for the performance of their works, has projected in the center guesting the Concerts, two specific sound equipments.

The first, based on a wave guide system , makes easier the concentration of energies permicting to realize localizations and evident movements of sound signal; the second one, based on special small diffusers, presents a reflecting screen capable to adapt the sound signal to the acoustic features of the hall.
The variety of scientific and musical contributions, present in this third edition too, has permicted the distribution of the works in large pertinent areas.

The subjects treated have been inserted in five sessions:

- Tecnique and Expression
- Training and Specialization
- Research and Application
- Project and Activity
- Idea, Structure, Means

Conclusions

LA TERRA FERTILE, established in L'Aquila on the work of Electroacustic Music group class, has found since the beginning the collaboration of the whole more representative forces; this expresses the collective need to spread and go deep in the disciplines converging in this field; our aim, so as our hope, are addressed to the achievment of these goals and to the possibility of creating a permanent event, corresponding ever more to the expectations of both the musicians and the researchers.

References

M.C. De Amicis "*La Terra Fertile: una necessità musicale*" in Suono Sud, n.22 1994
M.C. De Amicis "Suoni in un piatto di rame. *L'Aquila: una città laboratorio*" in Suono Sud 1995
A.A.V.V. Proceeding La Terra Fertile, 1996
R. Doati, "Esaltata l'arte, affermandosi la fantasia, quale educazione?" in Il Complesso di Elettra, 1995
M. Lupone "Civiltà del suono. Le questioni del cambiamento" in Il Complesso di Elettra, 1995

AN IMPROVED PITCH SYNCHRONOUS SINUSOIDAL ANALYSIS-SYNTHESIS METHOD FOR VOICE AND QUASI HARMONIC SOUNDS

Riccardo Di Federico
difede@dei.unipd.it

Gianpaolo Borin
borin@dei.unipd.it

Centro di Sonologia Computazionale
Dipartimento di Elettronica e Informatica
Università degli Studi di Padova
via S. Francesco 11
35121 Padova - Italy

Abstract

We present an improved pitch based sinusoidal analysis technique. Traditional sinusoidal analysis involves three basic stages: Short Time Fourier Transform (STFT), peak picking and peak tracking (or continuation). By limiting our target to quasi harmonic signals we developed a new pitch based sinusoidal analysis method which skips the partial tracking algorithm: a partial is assumed *to be the spectral maximum nearest to the position predicted by the pitch (partial order times the pitch). Unlike previous methods the pitch is calculated before any decision on partials and therefore the problem of its estimation has been carefully considered. The pitch detection method that we propose works as follows: after a peak detection on the short time spectrum of the signal is performed, a weighted measure of the mean frequency distance between peaks is taken. The fundamental frequency is then chosen to fit this measurement. The proposed algorithm is able to provide a pitch estimate even when the fundamental is missing and it is very robust to octave errors.*

1 Introduction

The two main issues related to audio transformations are the analysis-synthesis method and the modification algorithms. In the last decade time-frequency techniques such as phase vocoder [5] have gained popularity over time domain systems due to their flexibility in modeling and transforming musical signals. Among time-frequency techniques sinusoidal modeling [1] [2] [3] [4] allows the widest range of high quality transformations on musically relevant features as well as a very powerful framework for signal analysis. In the classical sinusoidal analysis, evaluation of the parameters of the signal involves three main steps: a STFT, a peak picking algorithm and a partial identification process which recognizes the significant spectral peaks and tracks them over time. The last step is definitely the most critical since it must decide whether a peak is just noise or it belongs to a partial track. Many approaches are possible for this problem, from a heuristic set of rules as in [1] [2], in which partials can be born or die depending on their amplitude and frequency deviation, to statistical methods as in [6], where Hidden Markov Models are employed. These methods are intended for a very general class of sounds, including polyphonic and inharmonic signals. Here we propose to skip the tracking step whenever the sound is quasi harmonic and monophonic, such as singing and many solo instruments. The basic idea is that if we can reliably decide on the pitch we can also predict the (theoretical) position of the harmonics. Based on this estimate, all the harmonics up to a previously decided order are detected, whether they are present or not, that is whatever their amplitude is. This assumption, that can appear expensive in terms of memory resources and weak when estimating low level noisy partials, has some advantages. First, the used approach dramatically reduces the computational burden of the analysis. Second, when modifying the sound, for example by pitch shifting, low level partials can be brought to high energy spectral regions; if they were disacarded by the analysis, the modified sound would be incomplete. Third, whenever the sound is pitched, its full bandwidth is represented, thus avoiding the lowpass artifact, found for example in SMS [2], due to non well developed partials near the sound attack.

The remaining part of the article is organized as follows: section 2 briefly introduces the classical sinusoidal model, section 3 describes the proposed pitch detection algorithm and section 5 shows some results of the analysis process.

2 The sinusoidal model

In the classical sinusoidal model [1] [2], the sound is approximated by a sum of N sinusoids, also called deterministic part, which retains most of the energy of the signal:

$$s(t) = \sum_{i=1}^{N} A_i(t) cos[\theta_i(t)] + r(t) \qquad (1)$$

The error term r(t) is the so called *residual*, which, as long as N is high enough, represents the noisy part of the sound. Each partial is characterized by time varying amplitude and phase. The instantaneous phase is taken to be the integral of the instantaneous radian frequency of the partial:

$$\theta_i(t) = \int_0^t \omega_i(\tau)d\tau + \theta_0 \qquad (2)$$

Aim of the analysis process is to evaluate A_i, ω_i, θ_i at predefined time intervals, or *frames*, that must be short enough to allow an accurate signal reconstruction by interpolation of these sampled parameters. As seen before, in the classical sinusoidal model the evaluation of the parameters is carried out on the base of a STFT followed by a peak picking algorithm. The choice of the STFT settings (window type and length, hop size, etc.) is decided on the base of the signal characteristics (variability, pitch) and the desired accuracy (compromise between spectral resolution and the bandwidth of the main lobe and side lobes). Here we assume a Blackman - Harris 74dB window, whose length is made pitch synchronous, set to three times the average pitch period over the last few frames.

3 Pitch detection algorithm

Since our partial search is heavily based on the pitch, the algorithm for pitch extraction must be as robust as possible. Many strategies have been proposed for extracting the pitch from a set of given spectral maxima [8] [7]. In this paper we will introduce a method which demonstrated to be very robust to common errors such as pitch doubling and halving, and which is able to return a value for the (perceived) pitch even though the fundamental is not present in the spectrum, thus allowing a correct partial positioning almost in every spectrum configuration. The basic idea is to define a weighted mean frequency distance between adjacent peaks. This distance is intended as a first estimate of the pitch, which will be refined by searching, if present, the maximum located in the neighbors of the pitch estimate. The main steps, given the FFT of the current frame, are as follows:

I Detection of the spectral maxima between 50 Hz and 6000 Hz. In this range, for normal pitched sounds at least five or six partials should be present, which is adequate for the algorithm.

II Detection of trivial cases. These are:

1. No maxima are found. An error (no valid pitch) is returned.
2. One or no minima are found (but exactly one maximum is present). The pitch is set to the frequency position of the maximum.

If the algorithm reaches this point, at least one maximum and two minima are present.

III Each peak is then characterized by its dB amplitude M_i, frequency F_i and mean dB difference between its amplitude and preceding and following minima, $pv_i = \frac{2M_i - m_{i-1} - m_i}{2}$. A high value for this parameter (here refferred to as *peak to valley* ratio) means that the peak is well separated and therefore clearly distinguishable from the others.

IV The absolute dB maximum M_M and its frequency F_M are stored for reference (see below).

V Elimination of the spurious and noise corrupted peaks. A peak is declared 'good' if it accomplish the following conditions on its relative amplitude and *peak to valley* ratio:

- $M_i > M_M - \alpha$
- $pv_i > \beta pv_{M_M}$

We found that adequate values for α and β are 40dB and 0.1, respectively. The set of chosen peaks will be indicated with $\tilde{M}_i$

VI If after the previous selection process just one peak survived (necessarily the absolute maximum), it is declared as the fundamental and the algorithm returns. If more than one peak, say $\tilde{N}_i$, remained, the algorithm attempts to estimate the mean frequency difference between adjacent maxima. This is achieved by the construction of an histogram $H(\Delta\tilde{F})$ of the differences, each weighted by the dB product of the two related peaks:

$$H(\Delta\tilde{F}_i) \leftarrow H(\Delta\tilde{F}_i) + \tilde{M}_i\tilde{M}_{i-1} \quad i = 1...\tilde{N}_i$$
$$H(\tilde{F}_0) \leftarrow H(\tilde{F}_0) + \tilde{M}_0^2$$

where

$$\Delta\tilde{F}_i \doteq \tilde{F}_i - \tilde{F}_{i-1} \quad i = 1...\tilde{N}_i$$

There are a few cases in which a difference is not included in the histogram:

- The distance between the maxima is greater than the frequency of the first maximum:
$$\Delta\tilde{F}_i > \tilde{F}_i$$
- The difference is outside the possible range for the pitch $[p_{min}\ p_{max}]$:
$$\Delta\tilde{F}_i > p_{min} \quad \Delta\tilde{F}_i < p_{max}$$

This histogram is then averaged so as to cluster near high level areas. Finally, the maximum is searched and its position assumed as a first esimate of the pitch (see figure 2).

VII The estimate is refined by searching the spectrum for the closest maximum; if no maxima are found in a narrow range around the estimate, $\pm$ 1 bin, the estimate itself is returned as the pitch. Otherwise, the estimate is refined by calculating the position of the selected maximum, using a quadratic interpolation over the three highest bins.

VIII An optional final step is the voiced / unvoiced / silence (tonal / non tonal / silence) decision, which marks the frame as pitched or non pitched. The decision is taken on the base of the frame energy, zero crossing and high frequency energy to low frequency energy ratio.

IX A non linear smoothing algorithm is applied in order to correct possible gross errors. This is obtained by putting a threshold on the maximum pitch variation and by correcting only those pitch samples which are preceded *and* followed by exceedingly high or low values.

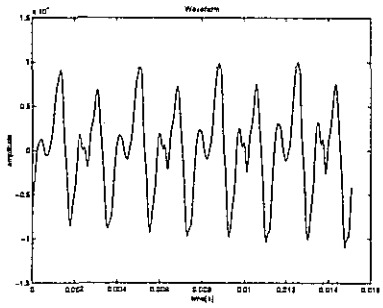


Figure 1: Example of signal waveform (singing)

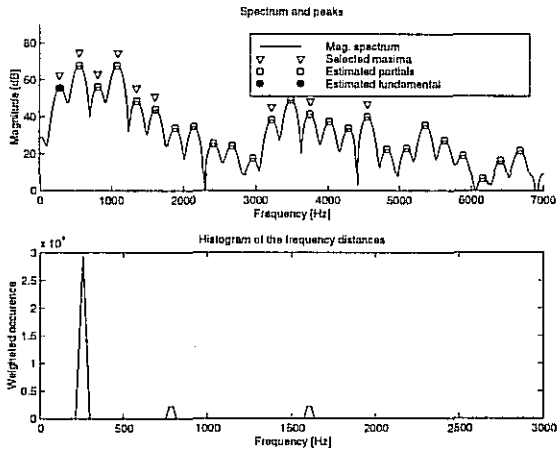


Figure 2: Spectrum with detected partials, pitch, and selection of strong peaks for the algorithm (upper part). Weighted histogram of the frequency peak distance

4 Detection of the remaining partials

Once the pitch has been computed, partials are searched in spectral region around the associated harmonic. If we denote the pitch with p, the i^{th} partial is searched in the frequency range $[ip - p/2, ip + p/2]$. If no maxima are found in the interval, the partial amplitude is set to zero, its frequency to ip and its phase to a random number in the interval $[-\pi, +\pi]$. When one or more maxima are found the tallest is chosen and the final values for amplitude and frequency are calculated by a quadratic interpolation, while phase is computed as the weighted average of the most prominent bins around the peak.

4.1 Synthesized signal and residual related issues

Here we want to point out some issues related to the resynthesis process and the treatment of the residual as a completion of the system. When synthesizing the signal from its sampled parameters, an interpolation over the duration of the frame is needed. Various schemes have been proposed, from the classical cubic phase and linear amplitude partial interpolation [1], to a smoother quadratic polinomial phase interpolation [3]. These schemes allow a good approximation of the waveform between the frames so that it is possible to subtract the synthetic waveform form the original sound, thus obtaining the residual. However, if we are not interested in

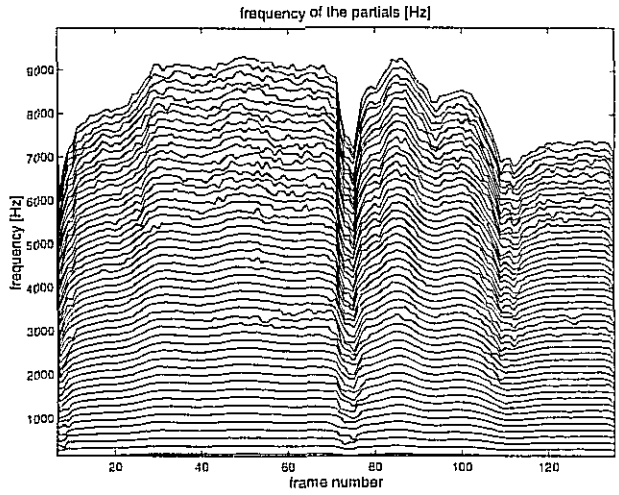


Figure 3: Example of tracked partials (singing). The roughness in the upper part is due to the background noise that becomes comparable with the signal energy

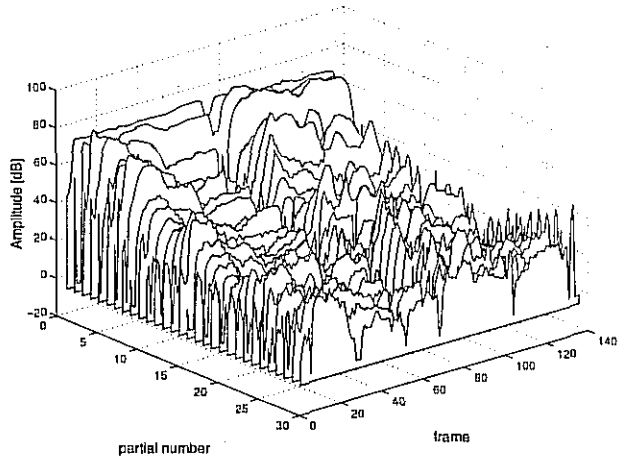


Figure 4: Amplitude evolution of the partials

very accurate waveform reconstruction it is possible to use much faster synthesis algorithms, as proposed in [9], based on the inverse FFT. A problem which is closely related to synthesis is that of residual modeling. The question arises because in many applications, for instance sound transformation, residual cannot be neglected and must be treated separately from the sinusoidal part. Since the residual is mostly noise, a widely adopted approach is the one proposed by Serra [2], by which the residual is represented by filtered white noise. Unfortunately, this residual model is unsuitable for sharp attacks, when the reconstructed noisy part is not coherent with the deterministic sound. Another approach is to use a high number of sinusoids, closely spaced in frequency, so that the noise will be included in the deterministic part [1]. Other models have been proposed but, at present, none seems to be fully satisfactory in terms of flexibility and/or sound quality.

5 Results

The proposed algorithm was tested on a variety of harmonic and slightly inharmonic sounds, such as

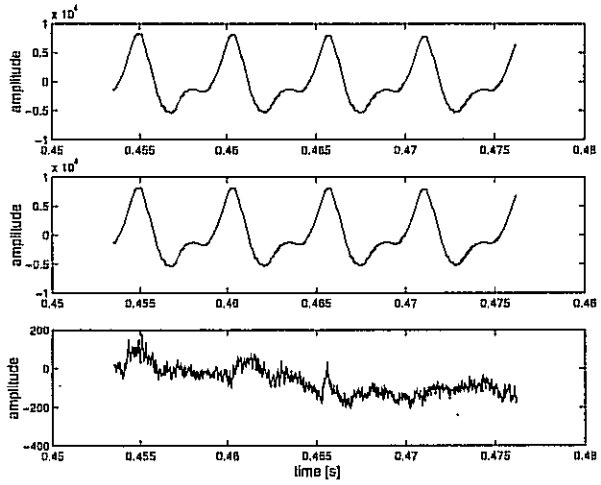


Figure 5: Residual extraction. Original sound (top), sinusoidal part (middle), residual (bottom).

singing, violin and piano, looking for pitch errors, quality of the reconstruction, robustness to modifications such as time stretching and pitch shifting. As an example, figures 3 and 4 show the partial tracking and amplitudes for the sung italian word 'quando' (excerpt from the song 'Il cielo in una stanza' performed by a professional singer). The analysis was carried out with 50 partials and a 100Hz frame rate. The deterministic part of the signal is completely represented even in the neighborhood of the attack because all partials exist whenever the pitch exists. We could say that the analysis is *instantaneous*; no past estimates or default values for the pitch are needed. Moreover, since no continuation (dependence on past frames) algorithm is needed, the system is free from pitch error propagation.

In figure 5 an example of sinusoidal + residual separation is showed. The residual is obtained by subtracting the sinusoidal part from the original sound.

It should be noted that the proposed procedure ensures that particular cases such as a pure sinusoid or a sound without fundamental are treated correctly. The use of a measure of the mean frequency distance between spectral peaks makes the presence of the fundamental unnecessary, while in the case of a sinusoid (one single prominent maximum plus some noise) the histogram is dominated by the related peak, making the decision process straightforward.

6 Conclusions

This work introduced an improved algorithm for partial tracking of quasi harmonic sounds using sinusoidal models. The basic idea is to skip the traditional continuation step and replace it by a pitch estimation followed by a partial search. Partials are expected to be found close the harmonic series generated by the current pitch. Since the whole analysis was based on the pitch, an effective pitch detecting algorithm has been designed. The main features of the pitch detector are the robustness to octave errors, the memoryless estimate (no past values or default values are requested) and the possibility of detecting the pitch even for sound with very low, or completely absent, fundamental [1]. Since all partials are supposed to exist whenever the sound is pitched, partial tracks start immediately after the attack, thus producing a full representation of the sound. This avoids the characteristic delays found in the classical continuation process, in which a track is grown from silence only when it becomes sufficiently steady.

We remark that the analysis procedure presented here does not include any model for the residual. Future research will focus on a high quality parametric model for the residual, capable of preserving, even after sound modification, the waveform coherence with the deterministic part.

Acknowledgements

This work has been supported by Telecom Italia S.p.A under the research contract "Cantieri Multimediali".

References

[1] R. J. McAulay, T. F. Quatieri "Speech Analysis / Synthesis based on a sinusoidal representation" *IEEE Trans. ASSP* vol. 34 No. 4 August pp.744-754 1986.

[2] X. Serra "Musical Sound Modeling with sinusoid plus noise" in *Musical Signal Processing* Ed. by C. Roads S. T. Pope A. Piccialli and G. De Poli Swets and Zeitlinger Publ. pp. 91-122 1997.

[3] Y. Ding, X. Qian "Processing of Musical Tones Using a Combined Quadratic Polynomial-Phase Sinusoid and Residual (QUASAR) Signal Model" *J. Audio Eng. Soc.* vol. 45 No. 7/8 July/August pp.571-584 1997.

[4] E. B. George, M. J. T. Smith "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model" *IEEE Trans. ASSP* vol. 5 No. 4 September pp.389-406 1997.

[5] M. Dolson "The Phase Vocoder: A Tutorial" *Computer Music Journal* vol. 10 No. 4 Winter pp.14-27 1986.

[6] Ph. Depalle, G. Garcia, X. Rodet, "Analysis of Sound for Additive Synthesis: Tracking of Partials Using Hidden Markov Models" *proceedings* of the ICMC, pp.949-97, 1993.

[7] B. Doval, X. Rodet, "Fundamental Frequency Estimation Using a New Harmonic Matching Method" *proceedings* of the ICMC, pp.555-558, 1991.

[8] R. C. Maher, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure", J. Acoust. Soc. Am., vol 95(4), April, pp.2254-2263, 1994

[9] M. Goodwin, X. Rodet, "Efficient Fourier Synthesis of Nonstationary Sinusoids" *proceedings* of the ICMC, 1994.

[1] In some particular cases such as almost perfectly odd sounds, the pitch detector fails. For this purpose some extra heuristics is under developement in order to deal with such situations.

Thursday 24th **h. 16.00**

POSTER SESSION II

Tuning to the rhythm through the circle map

Rosalia Di Matteo*, Brunello Tirozzi^°, Manuela Imperiali^, Marta Olivetti Belardinelli*°

* Department of Psychology, Via dei Marsi 78, 00185 Rome
^ Department of Physics, Piazzale Aldo Moro 7, 00185 Rome
° ECONA, Interuniversity Center for Research
on Cognitive Processing in Natural and Artificial Systems

Introduction

The emergence of a spontaneous cognitive rhythm during the performance of two different and competing tasks (motor tracking and backwards counting) was firstly observed by Valentini several years ago [16],[17]. Following the same line of research Fraisse [5] also suggested the existence of a specific organisation of the cognitive activity, essentially based on a rhythmic structure. In the following years several works performed by Olivetti Belardinelli [8],[12],[13],[14] showed a link between spontaneous rhythmic organisation and cognitive strategies adopted by subjects in solving simple non rhythmic problems. On the basis of the experimental findings by Jones & Boltz [6] it is possible to regard cognitive processes, such as perception, attention, and memory, as rhythmically organised processes. From this point of view the cognitive functioning can be interpreted as the *attunement* of the cognitive system to sequential events occurring in the environment. In a more general systemic framework Olivetti Belardinelli [9],[10],[11] proposed to consider the human cognitive system as characterised by one or more spontaneous rhythms. By consequences when the cognitive system is engaged in processing rhythmic impulses coming from the environment, it has to synchronise its internal rhythms to the external ones.

Entrainment models

A rhythmic structure is a configuration that emerges in time, as its single units are not simultaneously present in the perceptive field; it is perceived as a repeating alternation of strong and weak beats originated by changes either in the temporisation or in the qualitative features (loudness, tonality, timbre, etc.) of the sequence, or both. For this reason the attempt to model rhythmic processing results a very complex task.

Among the different models used to explain how a cognitive system is able to process rhythmic events, entrainment models look well suitable to account for this behaviour.

Entrainment dynamics is the process by means of which two oscillatory events can achieve synchronisation. Synchronisation between two coupled oscillators occurs if one of them (*driver oscillator*) is able to affect the other (*driven oscillator*) by altering either its phase (*phase-tracking*) or its period (*frequency-tracking*) or both, in such a way that they come regularly into phase (*phase-locking*).

Among the mathematical theories capable to describe entrainment processes we choose to reproduce and verify [2],[3] a model proposed by Large and Kolen [7]. In this model the basic oscillatory unit was able to synchronise its continuous output to an incoming periodic and discrete signal slightly adjusting its phase and period at each step. This behaviour was obtained by minimising the error function and by using the circle map theory in order to define the system parameters. The model gives a reasonable account for rhythmic processing as the internal driven oscillation is able to detect (perception), selectively respond to (attention) and retain (memory) the periodic components of the driver signal. Nevertheless it reveals some ambiguities in linking entrainment dynamics with the circle map theory [4].

The circle map theory

The circle map theory is a mathematical construction devised in order to describe in a discrete fashion the synchronisation process between two coupled oscillators. We analysed in more detail the circle map theory and developed a model explicitly based on the circle map dynamics (exactly the *sine circle map*).

Reiterating a sine circle map we calculate the phase of the driven oscillator at strobed intervals t of the driver oscillator:

$$\theta_{t+1} = \theta_t + \Omega + g(\theta_t) \qquad (1)$$

In this way the phase of the driven oscillator at any time depends on the old phase (θ_t), on the uncoupled ratio relating the two periods ($\Omega = q/p$) and on a non-linear function ($g(\theta_t) = K \cdot sin 2\pi\theta_t$) representing the dynamics of coupling (q represents the period of the driver oscillator, whereas p indicates the periods of the driven one).

According to this theory the dynamics of two coupled oscillator can be summarised by a rotation number $W(K,\Omega)$ that identifies some stable mode-locking states, acting as attractors or resonances. When W is an irrational number the dynamics of coupling remains quasi-periodic, whereas when W is a rational number the dynamics tends to be asymptotically periodic, i.e. the driven oscillator tends to modify its period in such a way as to complete q cycles every time the driver one has completed p cycles.

$$W(K,\Omega)= \lim_{t\to\infty}\left[\frac{\theta_t-\theta_0}{t}\right] \quad (2)$$

The rotation number W depends on K, a parameter expressing the extent of the mode-locking states. It depends also on Ω, a parameter relating the periods of two oscillators (*q/p*) and expressing the starting degree of tuning between the driver oscillator and the driven one.

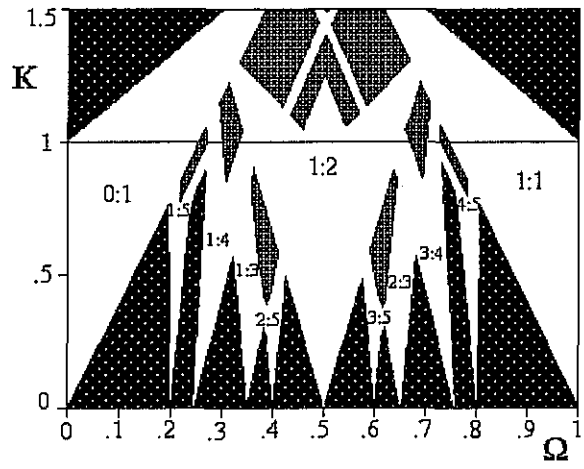


Fig. 1 Regime diagram showing Arnold's tongues for different ratios Q/P.

An elegant account of the sine circle map has been given by Arnold [1]. According to this author there are some areas A_i in the (K, Ω) plane such that if (K, Ω) ∈ A_i then the ratio *q/p* converge to a well defined ratio Q/P (where Q and P are small integers). The areas A_i are called *Arnold tongues* from their characteristic shape. Reiterating the coupling dynamics, we can observe that synchronisation occurs every time, given an appropriate selection of K, the parameter Ω falls in a stable mode-locking state of W = Q/P [15].

An attempt to modelling

This observation allows us to build a network of 15 oscillators based explicitly on the circle map dynamics such that at least one of the driven oscillators of the network achieve asymptotically a phase-locking with the driver one.

In order to obtain this result we had to find explicitly the interval of existence and stability of Ω = *q/p* such that it converges to a rational W = Q/P, by reiterating the map. Let us call f_Ω the sine circle map:

$$\theta_{t+1}=\theta_t+\Omega+K\sin 2\pi\,\theta_t \quad (3)$$

The existence and stability for the phase of the driven oscillator characterised by a rational winding number W is given by the two following conditions:

$$f_\Omega^{Q}\theta_1=\theta_1+p \quad (4)$$

and

$$\frac{df_\Omega^{Q}}{d\theta_1}=\theta_1+p \quad (5)$$

where $f_\Omega^{Q}\theta_1$ means Q iterations of the f_Ω.

The values of Ω which satisfy these two conditions can be found using the Newton method. By means of this numerical method we calculated the boundaries x_i of the intervals of stability for Ω given a fixed value to K.

We realised the net of 15 oscillator in such a way that their periods can range from 10 to 596 cs., on the whole. As we were interested to achieve a mode-locking 1/1 we chose K=1 because corresponds to the largest width for this ratio.

We can see in figure 1 for the 1/1 ratio we get *phase-locking* if $x_1 < \Omega < 1$ (with x_1= 0.7614). This means that, by reiterating the circle maps for each one of the *n = 15* oscillators, a driver signal with period *q* entrains just the driven oscillator having a starting period that satisfy the following condition:

$$p_n < q < \frac{p_n}{(x_1)^2} \quad (6)$$

This simple construction has been realised because we use the powerful sine circle map theory with all its implications. In the more detailed and complicated model of Large and Kolen the same approach would be more difficult.

Conclusion

Our model represents a first attempt to formalise entrainment dynamics through the circle map theories. Although requiring a more extensive verification, the model reproduces the entrainment dynamics avoiding some mathematical ambiguities contained in other similar models. On the other hand the simplification we adopted does not imply a sort of computational reductionism. Indeed the model we proposed manifests a behaviour characterised by a selective response to the temporal features of the external signal.

Moreover it is coherent with a systemic view because the intrinsic periodic activity spontaneously adapts itself to the external stimuli. This process represent a kind of *attunement* characterising the interaction between the living organisms and the environment and constitutes therefore a solid ground for comprehend the adaptive functioning of the human cognitive system. In this view our model represents a reasonable account for cognitive *attunement* to sequential events and processing of rhythmic structures.

References

[1] Arnold, V.I. (1965). *American Mathem. Society Trans.*, Ser. 2, 46, 216.

[2] Di Matteo, R., Olivetti Belardinelli, M., & Tirozzi, B. (1997). Strutture temporali e meccanismi di trascinamento. *Scientific Contributions to General Psychology*, (in press).

[3] Di Matteo, R., Rossi-Arnaud, C. & Tirozzi, B. (1997). Rhythm processing and entrainment processes. Proceedings of the *3rd Triennial ESCOM Conference*, Uppsala (Sweden), 7-12 June, 1997.

[4] Di Matteo, R., Tirozzi, B., Imperiali, M. & Olivetti Belardinelli, M. (1998). Modelling entrainment of rhythm processes through circle map theory. Proceedings of the *International Conference on Contribution of Cognition to Modelling*. Lyon (France), 6-8 July, 1998

[5] Fraisse, P. (1978). Time and rhythm perception. In E.C. Carterette e M.P. Friedman (eds), *Handbook of Perception*. Academic Press, New York, 8, 203-254.

[6] Jones, M.R., & Boltz, M. (1989). Dynamic attending and response to time. *Psychological Review*, 96, 459-491.

[7] Large, E.W., & Kolen, J.F. (1994). Resonance and the perception of musical meter. *Connection Science*, 6, 177-208.

[8] Olivetti Belardinelli, M. (1979). L'influenza degli stati interni nell'apprendimento di ritmi visivi. *Scientific Contributions to General Psychology*, 6, 7-43.

[9] Olivetti Belardinelli, M. (1986). *La costruzione della realtà come problema psicologico.* Boringhieri, Torino.

[10] Olivetti Belardinelli, M. (1993). Research on the factors determining preference for human or computerised rhythmic performance. *Scientific Contributions to General Psychology*, 10 n.s, 169-192.

[11] Olivetti Belardinelli, M. (1996). Preference for human or computerised rhythmic performance: Research on determining factors and attempt on modelling. *9th ESCOP Conference*, Würzburg, 4-8 September.

[12] Olivetti Belardinelli, M., & Besi, M. (1993). Spontaneous rhythms and cognitive strategies in problem solving situation. *Scientific Contributions to General Psychology*, 10 n.s, 43-58.

[13] Olivetti Belardinelli, M., Del Miglio, C., Reali, G. & Paluzzi, S. (1993). L'influenza del ritmo spontaneo nella elaborazione modulare della sagoma del corpo umano. *Scientific Contributions to General Psychology*, 10 n.s., 31-42.

[14] Olivetti Belardinelli, M. & Pessa, E. (1981). Simulazione dell'elaborazione di inputs ritmici con successiva deformazione dello stimolo. *Scientific Contributions to General Psychology*, 8, 51-81.

[15] Treffner, P.J., & Turvey, M.T. (1993). Resonance Constraint on Rhythmic Movement. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1221-1237.

[16] Valentini, E. (1951). La funzione del ritmo nell'esecuzione di due compiti differenti ed antagonisti. *Atti del IX Convegno degli Psicologi italiani.* Rome, November.

[17] Valentini, E. (1956). Modalità d'interferenza nell'esecuzione contemporanea di due compiti diversi e antagonisti. Contributo dell'Istituto di Psicologia dell'Università di Roma, XII.

REPORT OF THE *COMDASUAR*: A SIGNIFICANT AND UNKNOWN CHILEAN CONTRIBUTION IN THE HISTORY OF COMPUTER MUSIC

Martín Alejandro Fumarola
LEIM, National University of Córdoba
Estafeta 56, RA-5001 Córdoba
ARGENTINA
maralefo@hotmail.com

ABSTRACT

This paper describes the Intel 8080 microprocessor-based musical instrument called ***COMDASUAR*** (it stands for ***COmputador Musical Digital Analogico Asuar***), which was devised and developed in 1978 by the Chilean engineer and composer *José Vicente Asuar*, one of the pioneers of electroacoustic music in Latinamerica. Both the hardware and software (with 26 subprograms) components are further explained pointing out the computer-related compositional possibilities. This paper embosses *COMDASUAR*'s uniqueness and pioneer role in the world history of computer music since it does not register any equivalent at the end of the seventies and serves for the revalorization of Latinamerica in the world context of electroacoustic and computer music.

I. INTRODUCTION

Chilean engineer and composer José Vicente Asuar is, along with Juan Amenábar, one of the pioneers of electroacoustic and computer music practise in Chile and in the whole Latinamerica region [1]. In 1958 he founded the *"Estudio de Música Electronica"*, the cronologically second one in Chile and in Latinamerica. His electroacoustic piece *"Variaciones Espectrales"* is one of the pioneer works in Latinamerica together with Juan Amenábar's *"Los Peces"*.

José Asuar was extremely aware of the computer music experiences that had been carried out in the USA, Canada and Europe from 1955 onwards, as he summarizes in his Paper *"Un sistema para hacer música con un Microcomputador"* [3]. A sharp knowledge of the early and posterior achievements of Lejaren Hiller, Leonard Isaacson, Iannis Xenakis, and Max Mathews, among others, is undoubtedly present in the way he entered upon the *COMDASUAR*, and not only it; his background in the musical applications of computing science and programming is clearly noticeable in several of his previous pieces (both instrumental and electroacoustic), and in his articles published in the *"Revista Musical Chilena"*. His historical awareness ,is untypical in Latinamerica since most practitioners of electroacoustic music in Latinamerica in the time tended to work isolated with respect to their colleagues in North America and Europe. If we are going to apply the distinction between 'electroacoustioc music' and 'computer music', it is very wise to state that whereas Juan Amenábar is the pioneer of electroacoustic music in Latinamerica, José Asuar is the pioneer of computer music in that subcontinent.

His formation as engineer helped him out, evidently. The *COMDASUAR* puts into evidence, among other things, Asuar's deep understanding of the electronic side of computers' architecture. Besides, the differenciation between supercomputing and microcomputers is mentioned in his articles, something that no other composer did in Latinamerica at that time. On the other hand, his musical training was very profondous, making easy for him to develop a very particular compositional thought, in which his influence from the serialism as well as from the stochastic approach are distinctive qualities. The *COMDASUAR* is the first computer music instrument in Latinamerica, bringing computing sciences applications and music together.

II. PREVIOUS REALIZATIONS

Before to the design and implementation of the *COMDASUAR*, José Asuar had several important accomplishements in electroacoustic and computer music, including meaningful instrumental pieces devised by computer music methodologies. Above all, the most important is *"Formas I"* (1970) produced in collaboration with the *"Grupo de Investigaciones en Tecnología del Sonido"* [2]. That work makes application of probabilistic processes to composition, algorithms based on serialism but with probability, "directed probability", histograms, sequences, and it is programmed in FORTRAN IV. Two vinil disks with Asuar's pieces had been released before to the birth of the COMDASUAR: *"El Computador Virtuoso"*, and *"Música Electrónica de José Vicente Asuar"*.

III. GENERAL DESCRIPTION OF THE COMDASUAR

The COMDASUAR is able to perform any musical score in an authomatic manner (i.e., with no human intervention), having a poliphonic capacity of up to 6 voices. It is completely tuned and sincronized, with the feasibility to choose the timbre for each voice. Its most outstanding compositional features include, for example: the possibility to develop heuristic programs,

and to propose musical ideas based on probability and musical gambles. The *COMDASUAR* was the first musical instrument based on a microcomputer developed in Latinamerica. Some of its general characteristics are the following:

-unexpensive: the total cost of its components was around U$ 1,000 in 1978.
-broad range of use: from a home use replacing a piano until a concert performance including its use as a tool for a composer.
-sounds produced in real time. It is possible to modify the musical results while they are listened to. Its frequency range is equivalent to the audible range (8 octaves).
-poliphony of up to 6 voices
-standard QWERTY keyboard for data entry as well as a TV-like monitor
-sounds obtained from the microcomputer are square waves, which are passed on to analogical units which transform them in the resulting timbres. It gives a balanced and equalized mixing result of the 6 voices that must be able for performance or recording.
-it can perform any musical score, offering heuristic possibilities.

IV. EXPLANATION OF THE SOFTWARE

it is written in machine language and ocupies 5 Kbytes of memory. It has 26 subprograms, each one named with the letters of the alphabet from A to Z. Those subprograms are divided into:
A1 COMMANDS FOR THE COMPUTER: 1)displays in the screen the content of the memory, 2)erases the screen (2 pages), 3)stores information in memory, 4)changes data in memory, 5)moves memory allocation, and 6)executes programs (2 subprograms).
A2 Operations with musical data: 1)introduces data (3 subprograms), 2)changes data, 3)removes data, 4)interpolates data, 5)moves pitches, and 6)moves durations.
A3 Heuristic: 1)Canon, 2) Retrogradation, 3)transmutes tones, 4)transmutes durations, 5)probability, and 6)inserts groups of durations.
A4 Convertion to sound: 1)with sincronism (2 subprograms), and 2)without sincronism.
A5 Control of peripherals: 1)it records cassettes, and 2)plays cassettes.
All those programs of A are recorded in the EPROM so it is possible to resort them at any moment. In addition, other programs were developed, specially heuristic ones, which are stored in cassette and can be utilized from a certain location of RAM memory. In those cases, it is necessary to trespass the desired program from the cassette to the foreseen memory location. In fact, this allows to extend the memory capacity to an undefinied quantity of programs.

B MUSICAL CODES, each one expressed by its pitch and duration:
B1. Pitch, expressed in 3 ways:
Octaves: a number from 0 to 7, which implies 8 octaves
Grade in the scale: with letters according to the standards of American format: A, B, D, E, F, and G. R for the silence
Cromatism: the standard, S, W, and Q
Quarter tones alterations: U = ascending quarter tone, T = three ascending quarter tones, V = descending quarter tone
R = three descending quarter tones
B2 DURATION. It is expressed according to the terminology in Spanish: R = *redonda*, B = *blanca*, C = *corchea*, N = *negra*, S = *semicorchea*, F = *fusa*, M = *semifusa*, L = *lunga*, P = *punto* (it multiplies the previous value by 1.5). Any duration value can be obtained by the addition of the aforementioned values. Besides, the following ciphers are defined: 0 = normal, 3 = triplet, and so on for the irregular values.
B3 FINAL DENOMINATION. For the reading on the screen, the computer numbers automatically each tone. As soon as it is finished to write the pitch or duration of a tone, the composer presses the space bar, what is represented in the screen with the symbol /. Example of a score written for the *COMDASUAR*:

0001 5BW/ 0NF/ 0004 6E/ C/
0002 3BQ/ 3C/ 0005 2DU/ 7B/
0003 R/ 0S/

B4 REDUNDANCY. In order to simplify and speed up data introduction, the *COMDASUAR* takes advantage of the enormous redundancy of musical data. In this sense, only it is indicated to the computer those elements that change from tone to tone. Any element of notation like an octave, grade, cromatism, duration that keeps constant, it is not necessary to mention it, the computer assigns the value it had in the previous tone. Example:

0001 3C/ C/ 0004 / / 0007 E/ N/
0002 / / 0005 G/ /
0003 E/ / 0006 / /

The *COMDASUAR* also takes advantage of redundancy by defining different modes of introducing musical data:
Mode 0 (J0): it is the usual mode, as seen in the previous examples. For each tone, its pitch and duration are indicated.
Mode 1 (J1): Constant duration. At the beginning, the value of the common duration to a tones series is indicated. After that, only the pitch of each tone is indicated:

J1 3S 0003 E 0006 4A 0009 D
0001 3C 0004 F 0007 B 0010 E/ 0B/
0002 D 0005 G 0008 C J0

Mode 2 (J2): Constant pitch. At the beginning, the value of the common pitch to a rhytmical sequence is indicated. After that, only the duration of each tone is indicated:

J2 4C 0004 NS 0008 J0

0001 CP	0005 S	0009 0C	0012 3G/ 0B/
0002	0006 3	0010 5/	
0003 S	0007	0011 7/	

Mode 4 (J4): Reiteration. This mode is utilized when a tones succession is repeated at least once. First, it is indicated the number of times it repeats, and then, the tones succession, which is limited by another mode indicator. This mode is specially useful for representing trills, tremoli, etc.

Mode 5 (J5): Repeats a sequence. This mode inserts a sequence that has already been written before. There is no limitation in the extension of the sequence to be repeated. Both the first and the last tone of the sequence to repeat are indicated:

0001 3C/ N/	0004 4C/	J0	0010 4C/ R/
J1 C/	0005 B/	0008 C/ B/	
0002 E/	0006 A/	0009 3G/ /	
0003 G/	0007 B/	J5 1/ 7/	

B5 TEXTURE. Asuar defines the texture of a sound to "its variation like a glissando or a vibrato".

Glissando. It is expressed like Mode 3 (J3) and defined indicating the pitches of the beginning and of the end of the glissando and its duration. The glissando is obtained as a fast succession of tones whose difference of frequency is very little and follows the direction of the glissando. The computer calculates the number of tones it must output and the frequency of each one, based on a fixed speed for the succession of tones that is the 16^{th} part of a semifuse. By employing this procedure it is possible to obtain any design of frequency variation:

Vibrato. It is calculated by points of a sinusoide whose axis is the central tone and its amplitude the 16th part of a tone. The frequence of repetition depends on the position of the Tempo regulator. For a Tempo N = 60, it is of 8 cicles per second. The computer calculates 32 points of this sinusoide according to a table. Due to the speed those sounds are issued, it happens something similar to the glissandi and a continuous vibrato is heard. The beginning of a vibrato is indicated as Mode 6 (J6) and its end as Mode 7 (J7). Between those 2 indications, any quantity of tones can be placed:

C HEURISTICAL PROGRAMS. For José Asuar, heuristical programs are those ones in which the computer has a creative intervention (it issues a score in the memory) or acts as a performer (it manipulates a score stored in memory). The following programs are stored in the EPROM of the *COMDASUAR*: C1. CANON; C2. RETROGRADATION. It is necessary to indicate the initial and the ending tones of the section that is retrograded.

C3. TONES TRANSMUTATION. It can be run with at least two voices in memory. It allows to exchange the pitches of the tones corresponding to the 2 voices while the durations remain unchanged. It is mandatory to indicate from which tone in each voice the transmutation is realized.

C4. DURATIONS TRANSMUTATION. Similar to the previous one. It exchanges durations while keeping the pitches. José Asuar states that he developed these 2 programs influenced by his former Professor Boris Blacher. At this point, we have a wide spectrum of possibilities for composing music.

C5. PROBABILITY. For this program it is necessary to define and quantify the musical elements that will be affected by probability. The *COMDASUAR* allows the simultaneous and independent probabilistic organization of registers, pitches, durations, harmony, and sound texture (vibrato and glissando).

C6 INSERTION OF GROUP OF DURATIONS. Normally, the tones lists sprouted by probabilistic programs are monotonous and unarticulated. With this program is possible to interpolate pauses and to make possible that tones groups of the probabilistic series have the same duration. The quantity of tones, its location and the constant duration are also calculated by probabilities.

C7 ANOTHER HEURISTICAL PROGRAMS. José Asuar developed several additional heuristical programs for meeting the requirements of specific musical fragments. One of the best examples is the piece *"Así habló el Computador"*, appearing in the LP of the same name, which was composed based on a numerical series and in various arithmetical gambles that determine the rhythm of some of its voices. In other compositions, José Asuar utilized aleatoric combination of groups of durations with groups of pitches. Of special interest are the weird mixtures, such as, of modal series with rhythmical series belonging to the serialism as well as using the melodies of the Argentinian vidalita, as in pieces of the LP *"Así habló el Computador"*. Several procedures of the serialism were utilized.

V. HARDWARE

D TONES OBTAINMENT All tones are attained from an only quartz oscillator. This oscillator resounds at the frequency of 2,048 kilocicles and the different tones are obtained by divisions or subarmonics of this generating frequency. Let's take the example of an oscillator that resounds at 28,160 cicles per second: it is outside of the audible spectra but if it is divided by 64 (the subarmonic of order 64), we obtain a pitch of 440 c.p.s. (A). Between 64 and 128 we have 64 possibilities of division or subarmonics, and if we choose all those that are nearer of the values of the tempered scale, we can obtain a musical scale of aproximate tuning. Therefore, the worse tuning, which corresponds to the high pitches, has a definition in the subarmonic 70, which is equivalent to a quarter tone, aproximately. The lower the pitch the better the tuning.

Tones are get by a system of division of frequencies supplied by the INTEL Timer 8253. This Timer is capable of obtaining simultaneously 3 different subarmonics from the same generating frequency. With that purpose, the microprocessor delivers the dividing cipher, with a definition of 16 bits, to each one of the 3 dividers of the

Timer, attaining the 3 tones as subarmonics of the same oscillator, having stable intervalic relations. In the *COMDASUAR*, it is possible to work either with a quartz oscillator that delivers a fixed frequency of 2,048 kilocicles or with an oscillator of variable frequency, which can be used manually or with voltage control. Therefore, there are two possibilities for obtaining the tones: by fixed tuning and by variable tuning.

E. RHYTHM OBTAINMENT. For producing rhythm, the *COMDASUAR* makes use of a quality of the microprocessors: the possibility to be activated by interruptions of the external media. The inner logics of the COMDASUAR has a program for interruptions waiting. When an interruption comes, the microprocessor goes to execute the main program, which consists in decrementing the counters in charge of counting the duration of each tone. As soon as the counters are decremented, the microprocessor goes again to the program of interruptions waiting in order to repeat the process. The *COMDASUAR* employs a multivibrator of variable frequency for generating the interruption pulses which allows to manually obtain accelerandi and retardandi. The maximal spped is around the 19 or 20 tones per second and per voice.

F. ANALOGIC EQUIPMENT For each voice, one filter with voltage control, one envelope generator, and one amplifier with voltage control are built. The control voltages for the filters are obtained from six digital-analogic converters connected to 7 bits of each port of two paralel interfaces (Programmable Peripheral Interface, INTEL 8255). The 8th. bit is used as trigger for launching the envelope generator. Three of the voices have at their disposal a wave form generator, which consists in a demultiplexor that divides the square wave in a wave of 8 segments. The original frequency is divided in eight parts as well, i.e., the pitch of the tone decreases in 3 octaves. By passing this stepped wave through the filter, it is possible to attain different spectrums, in which it is possible to gamble with the relative magnitudes of the first 8 armonics. With this procedure it is possible to achieve a very convincing synthesis of many well-known timbres, such as of some acoustic instruments.

The *COMDASUAR* also includes additional analogic equipment for the generation of effects: a white noise generator, a rose noise generator, two ring modulators used for getting inarmonic espectra like bell sounds, two tremolo generators, two generators of functions oscillating at low frequency with the goal of generating sinusoidal, triangle, and square voltages, which control filters and amplifiers for attaining different effects (vibrato, tremolo, etc.). A dephaser for accomplishing spatial-like effects as well as complementary units like inverters, multipliers, mixers, reverberators, etc. are also available in the *COMDASUAR*.

G. SCORES STORAGE. The RAM memory where musical data are stored has a size of 2 Kbytes. Because of the simplicity of many codes and to the information reduction by redundancy, this memory size allows to store up to 2,000 tones in the 6 voices. This information is physically placed within a cassette. The transference of music data from the cassette to the memory of the *COMDASUAR* is very fast, it takes only a few seconds.

H. SYNCHRONISM. As it was stated before, the poliphonic capacity of the COMDASUAR is up to 6 voices. For producing performances of more than 6 voices, it allows to syncronize them by resorting multi-track equipment. In the case of a 4 track recorder, one track is appointed to record a pulse for originating the interruptions. In the remaining 3 tracks, up to 18 voices (up to 6 in each individual track) can be recorded with total sincronism. For the recording of the pieces featured in the LP *"Así habló el Computador"*, a 2 tracks REVOX A77 tape recorder was utilized. The sincronism in the 2 tracks was obtained by recording a reference tone at the very beginning of track holding the first recording. After the issuance of that tone, the COMDASUAR counts certain amount of time and starts to perform the score. For syncronizing the second recording, it is necessary to play the track with the reference tone, and after this last one, the computer begins to play the new score.

CONCLUSION

The *COMDASUAR* has been useful not only for high-end computer-based composition but also for pedagogical and teaching purposes. Many of its features were an advance of computer music developments that came years later. José Asuar then suggested several new additions and improvements that made the COMDASUAR an authentical tool for composing and real-time performing music, for instance, the inclussion of piano-like keyboards, firstly one monophonic, then two monophonic, and finally one polyphonic. Sensors and optoelectronic devices have also been tried and this could have the most important potential field for its development. The COMDASUAR is a pride for Latinamerica and should be included in all the manuals and handbook listing and describing pioneer computer music instruments. The uniqueness of the COMDASUAR as a sequencer, a score editor, an algorithmic composition tool, and a sound synthesis program puts it in the same seat of honor as the key computer music developments of the USA and Europe.

REFERENCES

[1] Martin Fumarola. *"Interview with Juan Amenábar - More than 40 years of electroacoustic music in Latinamerica"*, to be published in Computer Music Journal 22 (3).

[2] José Vicente Asuar. *"Música con Computadores: ¿Como hacerlo?"*, *Revista Musical Chilena* N° 118, Pages 36-66, April-June 1972

[3] José Vicente Asuar. *"Un sistema para hacer música con un Microcomputador"*, Revista Musical Chilena N° 151, pages 5-28. July September 1980.

MISTUNED SCALES

Massimo Grassi

Department of General Psychology
via Venezia, 8
35131, Padova, Italy

Abstract

Twenty-six subjects were asked to judge the presence of mistunings during the execution of eleven major diatonic scales. The stimuli set was composed by the tempered scale, five scales obtained by compressing the equal tempered intervals, five scales obtained by stretching the dimension of the same intervals. Subject's task was to stop the scale execution as soon as he/she perceive the scale out of tune. The duration of this execution was measured. A 2 (kind of mistuning) ×5 (greatness of mistuning) ANOVA showed that: (a) the subjects successfully discriminated the stimuli set: the greater the scale distortion, the smaller the time of the stimulus execution; (b) for each degree of mistuning, the duration of execution of compressed scales was shorter than the duration of execution of stretched scales. The second result concerns the note on which the subjects stopped the scales. The most chosen notes to stop the ascending scales were the fifth (sol) and the third (mi). Contrary, the listeners choose different notes to stop the descending scales: the descending third (la) and the fourth (sol). This study confirms the good tolerance to the stretched mistuning predicted by the Terhardt (1974) model on the pitch perception. The second result shows new evidences for the template theory purposed by Shepard and Jordan (1984). In the current research the tonal schema is different for the two direction of execution.

1 Introduction

Traditionally the study on the perception of fine pitch differences has focused on the discrimination of chords [1], [2], [3] or intervals [4], [5], [6] in categorical perception task. Differently, other investigators have used more complex stimuli as scales [7], [8] or different harmonic or melodic situations [9], [10]. Those researches are concentrated on the effects of the instrument played on the conception of accurate tuning [7], [8], the best tuning for the execution [9], the effects of musical acculturation on the perception of mistuning musical patterns [10].

The purpose of the current research is to study the perception of mistunings during the execution of the major diatonic scales. Twenty-six subjects were asked to stop the scale's execution when they perceived the scale out of tune. The mistuning scale set was composed by ten scales: five scales equally compressed and five scale equally stretched. The correct scale was the major diatonic tempered. This study is focused on two problems: (a) the possibility that there are differences on the perception between log compressed and log stretched melodic intervals (b) which set of notes will be preferentially used to stop the scale.

a) With regard to the perception of log stretched and log compressed intervals, Burns and Ward [11] think that the general finding "evident in the data of all experiment [is the] tendency for the observer to hear narrow intervals (intervals less than a fourth) as perceptually wider (i.e. a compression of the scale relative to equal temperament) and to hear wide intervals (greater than a fourth) as perceptually narrower (i.e., a stretch of the scale)". Contrary Terhardt and Zick [9] suggest that "the subjective optimum [tuning] depends on the structure of the musical sound: if the constituent (...) tones of a musical sound cannot interact strongly in the auditory system because they are played in sequence or because they are widely apart in frequency, stretched intonation is desirable; if the musical sound's nature is such that some interaction of spectral components may give rise to beats and roughness, normal intonation will be optimum; in musical chords of high spectral complexity even contracted intonation may be suitable". Those results are predicted by the Terhardt model on the pitch perception [12].

b) The structural representation of the musical pitch, set forth by Shepard [13] by developing the ideas of Drobish [14] and Révész [15], provides some hypotheses to predict the set of notes used by the listeners to stop the scale. Shepard and Jordan [16] suggest that the "musical tones, though physically variable along a continuum of frequency, tend to be interpreted categorically as the discrete notes (...) of an internalised musical scale. We suggest that the internal schema may act as a template that, when brought into register with a tonal input maps the unequally spaced physical tones into the discrete step of the schema, with a resulting unique conferral of tonal stabilities on the tones". In the graphical description of the template, Shepard and Jordan represent the schema by attaching to each position a bar whose length is proportional to the importance and the stability of that tone ([16], fig. 1). My hypothesis is that the more stable tones inside the major diatonic scale (the fifth, dominant, and the third, mediant, according to Shepard and Jordan [16]) are more significant to conduct the subject at the mistuning detection. In fact, if those tones are the most

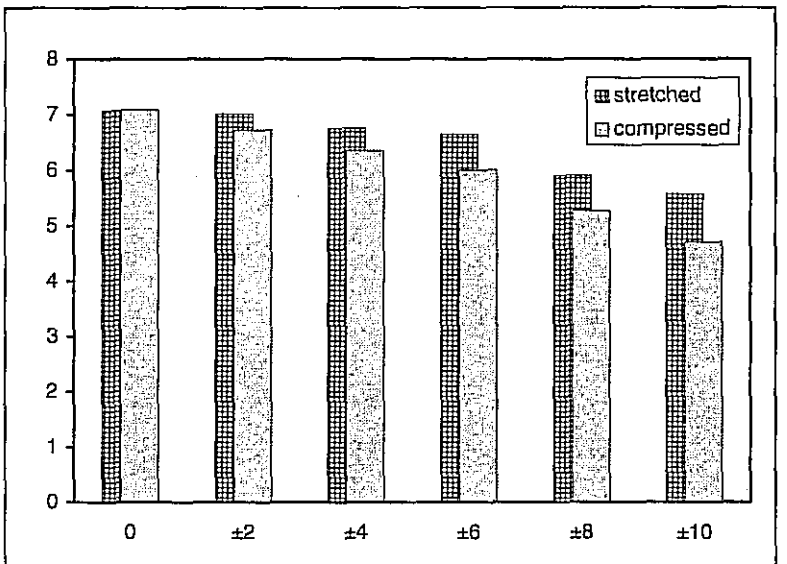


Fig. 1: duration of the execution (in seconds) by function the kind of mistuning and the amount of mistuning (in cent).

stable, they probably become the most noticeable in case that they are compressed or stretched. By using the method of my research it is possible to corroborate the results obtained by the probe-tone method [17], [18], [19], [20], [21], [22] and by different methods [23], [16]. Furthermore, it is possible to analyse in a finer way the hierarchic structure inside of the major diatonic scale.

2 Method

2.1 Subjects

Twenty-six undergraduate students participated to the experiment with basic musical education. None of the subjects has reported hearing loss or others difficulties with their hearing.

2.2 Stimuli

The stimuli were a set of eleven scales: the correct one was the diatonic tempered major; five stretched scales (+2, +4, +6, +8, +10 cent for tempered semitone); and five compressed scales (-2, -4, -6, -8, -10 cent for tempered semitone). The stimuli were presented either in ascending or in descending execution. The pitch value of each tones was obtained using the following equation:

$$f_a(i) = f_0 R_a^{i/12}$$

where f_0 is the starting tone frequency of the scale, $f_a(i)$ is the frequency of the stretched-compressed tone placed i semitones away from f_0 (where i = 1, 2, 3,... for ascending scales and i = -1, -2, -3,... for descending scales) and $R_a = 2^{12 \times n/1200}$ (where n = 100, the equal tempered semitone; 101, 103, 105, 107, 109, the set of stretched semitones and 99, 97, 95, 93, 91 the set of the compressed semitones) is the ratio frequencies of a stretched-compressed octave. The pitch of the first note was: do4 (261.6 Hz) for the ascending scales, do5 (523.2 Hz) for the descending scales. The single tone duration was 1 second. The complete scale duration was 8 seconds.

2.3 Apparatus

The timbre of the scale was a sine wave generated by a Sound Blaster 64 AWE Gold. The signal output, analogic and monophonic, was amplified by a McIntosh amplifier and two JBL speakers. The listener was placed 1.5 meters in front of the speakers. The stimuli sequence was controlled by the subjects with a computer keyboard.

2.4 Procedure

Each listener heard five scales, randomly chosen from the ascending stimuli set. These scales were used to familiarise the subject with the ascending stimuli and to adapt the scales volume according his preferences. Then the listener heard the random test sequence of fifty five ascending scales. After a pause the listener heard five descending scales then the random test sequence of fifty five descending scales. In every test sequences each different scale was repeated four times. The presentation order of the ascending and descending sequences was balanced within subjects. The complete session duration was approximately thirty minutes. Before the experiment the subject was informed that the stimuli started always at the same pitch frequency (do4 261.6 Hz or do5 523.2 Hz) and he must to respond just after the first note. The task was to stop the scale's execution as soon as the listener perceived the scale out of tune. The subject interrupt the scale by pressing a keyboard key. The key pressure immediately stopped the scale execution. The duration of the execution was measured.

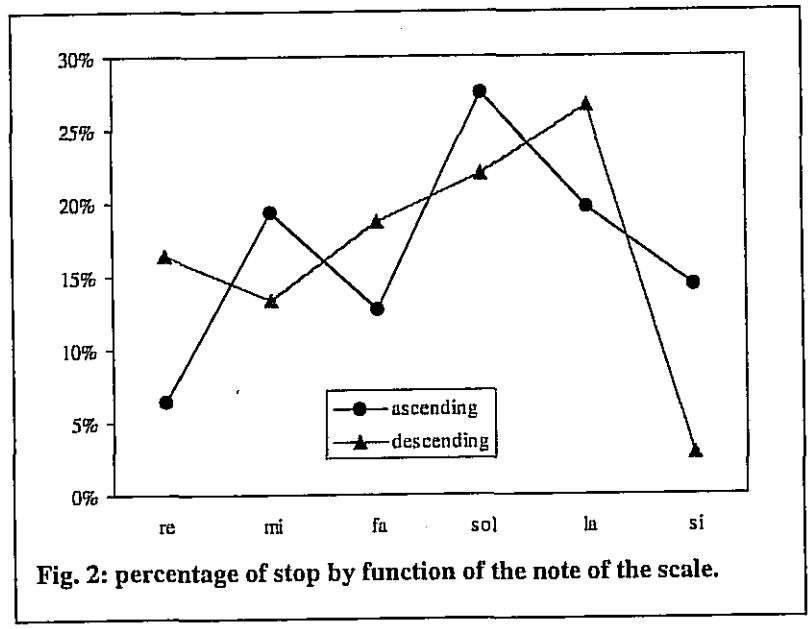


Fig. 2: percentage of stop by function of the note of the scale.

3 Results

a) A 2 (kinds of mistuning) ×6 (greatness of the mistuning) ANOVA was conducted on the duration of stimuli execution. The factor greatness of the alteration demonstrates that the subjects successfully discriminated the stimuli set: the greater the scale distortion the smaller the time of the stimulus execution: $F(5, 25)=48.871$, $p<.0001$. The listeners perceived in different times the two kind of mistunings: $F(1, 25)= 11{,}393$, $p<.0024$. The interaction between the factors, greatness of the alteration and kind of alteration, shows that, for each dimension of distortion, the compressed mistuning is always perceived before the stretched mistuning: $F(5, 125)=4.185$, $p<.0015$ (see figure 1).

b) When the subject stopped the scale the responses fell especially in the second half of the note in hearing. In fact the subject spends a little time to detect the mistuning and to respond: first half stops 460 (27%), second half stop = 1183 (73%) (I have used the answers that fell in the second half of the note to make the second graph; see figure 2).

4 Conclusion

a) This research confirms the prediction of the Terhardt model [12] on pitch perception. The compression and the stretch of the musical intervals are perceived in different ways. In the melodic execution the stretched mistuning is tolerated better than the compressed one. Differently, in complex harmonic situation the presence of the beats and roughness induces the listeners to prefer the compressed intonation or the equal intonation [9], [1], [2].

Usually the musicians think that the stretch of the intervals gives more brilliance to the execution. This preference is well know also by many listeners. Than, if the musicians habitually stretch the dimensions of the played intervals it is possible that the listeners, the subjects of this experiment, will be induced to perceive this kind of intonation as more correct. Furthermore, the stretch of the intervals is normally used in the piano [24] and harp intonation. So, the light compression of just intonation can be considered as "too flat" by the expert musicians [8] for the melodic execution. The present result can be used for the design of the musical instruments. In the electronic keyboard flexible intonation can be realised. By "control circuits", the intonation can be adapted to what actually is played on the keyboard at each moment, i.e., to the complexity of the musical pattern: melodic, harmonic and so on.

b) The listeners preferentially use a little set of note to stop the scales. In the ascending execution the subjects use the third (mi) and the fifth (sol). The evidence of the mistunings is maximum in the fifth. During the experiment some subject has reported that the most part of the scales were out of tune on the same note (the fifth). The subjects use different notes to stop the descending execution: the descending third (la) and fourth (sol). The evidence of the mistuning is maximum in the descending third, than in the descending fourth. The great importance of the third (mi) to stop the scale in ascending execution is minimum in the descending execution. This result is different from that predicted by using the Shepard and Jordan template. In the schema of those authors each degree of the scale have the same importance either in ascending or in descending execution. As the ascending and the descending scale have two different successions of whole tones and semitones, we could think that the direction of execution influences the mental schema of the major diatonic scale.

Acknowledgement

I would like to thank the professor Giovanni B. Vicario for his helpful suggestions.

References

[1] Biasutti, M., Vicario, G. B. (1993). Prestazioni di soggetti competenti e non competenti in un compito di valutazione di accordi musicali. Giornale italiano di Psicologia, 20 (3), 453-473.

[2] Biasutti, M., Grassi, M., Vicario, G. B. (1997) in preparation.

[3] Zatorre, R. J., Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. Perception & Psychophysics, 26, 384-395.

[4] Burns, E. M., Ward, W. D. (1978). Categorical perception - phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. Journal of Acoustical Society of America, 63 (2), 456-468.

[5] Siegel, J. A., Siegel, W. (1977). Categorical perception of tonal intervals: musicians can't tell sharp from flat. Perception & Psychophysics, 21 (5), 399-407.

[6] Siegel, J. A., Siegel, W. (1977). Absolute identification of notes and intervals by musicians. Perception & Psychophysics, 21 (2), 143-152.

[7] Loosen, F. (1994). Tuning of diatonic scale by violinist, pianist, and non musicians. Perception & Psychophysics, 56 (2), 221-226.

[8] Loosen, F. (1995). The effect of musical experience on the conception of accurate tuning. Music Perception, 12 (3), 291-306.

[9] Terhardt, E., Zick, M. (1975). Evaluation of the tempered tone scale in normal, stretched, and contracted intonation. Acustica, 32, 268-264.

[10] Lynch, M. P., Eilers, R. E., Oller K. D., Urbano, R. C., Wilson, P. (1991). Influences of acculturation and sophistication on perception of musical interval patterns. Journal of Experimental Psychology: Human Perception and Performance, 17 (4), 967-975.

[11] Burns, E. M., Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), The psychology of music (pp. 241-269). New York: Academic Press.

[12] Terhardt, E. (1974). Pitch, consonance, and harmony. Journal of the Acoustical Society of America, 55, 1061-1069.

[13] Shepard, R. N. (1982). Structural representation of musical pitch. In D. Deutsch (Ed.), The psychology of music (pp. 343-390). New York: Academic Press.

[14] Drobish, M. W. (1855) Uber musicalische tonbestimmung und temperatur. Abhandl Math. Phys. Kl. Konigl. Sachs, Ges. Wiss. 4, 1-120.

[15] Révész, G. (1954). Introduction to the psychology of music. Norman, Oklahoma: University Oklahoma Press.

[16] Shepard, R. N., Jordan D. S. (1984). Auditory illusions demonstrating that tones are assimilated to an internalised musical scale. Science, 226, 1333-1334.

[17] Krumhansl, C. L., Shepard, R. N. (1979). Quantification of the hierarchy of tonal function within a diatonic context. Journal of Experimental Psychology: Human Perception and Performance, 5, 579-594.

[18] Krumhansl, C. L. (1985). Perceiving tonal structure in music. American Scientist, 73, 371-378.

[19] Krumhansl, C., L. Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation musical keys. Psychological Review, 89, 334-368.

[20] Kessler, E. J., Hansen, C., Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the West. Music Perception, 2, 131-165.

[21] Jordan, D. S. (1987). Influence of the diatonic tonal hierarchy microtonal intervals. Perception & Psychophysics, 41 (6), 482-488.

[22] Jordan, D. S., Shepard, R. N. (1987). Tonal schemas: Evidence obtained by probing distorted musical scales. Perception & Psychophysics, 41 (6), 489-504.

[23] Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. Cognitive Psychology, 11, 346-374.

[24] Martin, D. W., Ward W. D. (1961). Subjective evaluation of musical scale temperament in pianos. Journal of the Acoustical Society of America. 33, 582-585.

Design and Implementation of the New GENDYN Program

Peter Hoffmann, Skalitzer Str. 96, D-10997 Berlin, phoffman@inf.fu-berlin.de

Abstract: The New GENDYN Program (="GENération DYNamique") is a new implementation of Iannis Xenakis' composition algorithm Dynamic Stochastic Synthesis in a distributed real-time environment. The New GENDYN program is backward compatible to Xenakis' original concept of a self-contained sound producing automaton creating music "out of nothing". However, the step towards real-time processing extends the Stochastic Synthesis into the domain of "interactive composition", turning the GENDYN into a stochastic composition instrument interactively controlled in real-time by a musician/composer. The consequences of this evolution will be discussed, along with a presentation of the design and the prototype implementation of the software.

1. Introduction

This paper discusses the design of a new implementation of Dynamic Stochastic Synthesis [10] in the context of software engineering paradigms such as real-time sound processing and distributed object systems on the one hand, and artistic paradigms such as "interactive composing" [2] on the other. The Xenakis algorithm is an extreme example of what can be called rigorous algorithmic composition, in the sense that a whole musical artwork, in its final acoustic shape, is generated "direct to disk" by a self-contained computer program. This radical concept is linked to a specific aesthetical approach to computer composition which, for want of a better term, is generally referred to as "non-standard synthesis" [6]. In this approach, the composer not only creates the computational model of the compositional macrostructures of the music but also invents his/her own model for the creation of sound, without taking recourse to any existing acoustical or physical model.

In Xenakis' algorithm, all sound and structure is created by exploiting probability fluctuations when generating random numbers with a set of different distribution characteristics. These fluctuations are accumulated in time and thus form the random walks of a waveform's breakpoint "coordinates" (amplitude values and sample time points) as well as a stochastic "patchwork" structure of durations that create a multiple "counterpoint" of sounds in time. All sound starts with perfect silence, i.e. a "degenerated" wave form polygone with all its breakpoints set to zero, and gradually "inflates" to a constantly changing jagged waveform as the fluctuations displace the waveforms's breakpoints towards positive and negative amplitude values and different spacings in time.

Xenakis realized his implementation of Dynamic Stochastic Synthesis at the Center for Mathematics and Automated Music (CEMAMu), Paris, with the assistance of Marie-Hélène Serra [7]. In 1991, one of the versions of this program, written in BASIC, calculated the composition "GENDY3", a piece of about 20 minutes, in two program runs (one for each channel) lasting each for about 2 days. With the new implementation, the same music can in principle[1] be generated in real time, making it possible, for example, to compute it "on the fly" during a concert instead of playing a tape.

2. Interactive Composing

But there are more ramifications to the fact that the computation of GENDYN sound has become "faster than sound". The pure computation speed radically changes the way of composing as it becomes possible to change the synthesis parameter settings while listening to the sound output.

The Dynamic Stochastic Synthesis in itself is powerful enough to generate complex sound with several temporal layers of sonic evolution (microscopic and macroscopic structures. The idea of a computer algorithm so efficient that it generates sonic data of a complexity comparable to that of music has indeed been formulated by Xenakis himself.

"The question that arises [...] is to know which mathematical construction to specify to the computer so that what is heard will be as interesting as possible - new and original. Without dwelling too long on this, I will cite an interesting example belonging to a case I was able to discover some time ago by using the logistic probability distribution. For certain values of its parameters α and β and its elastic barriers, this distribution goes through a sort of stochastic resonance, through a most interesting statistical stability within the sound produced. *In fact it is not a sound that is produced, but a whole music in macroscopic form.*" [9]

In Xenakis' original program, the set of synthesis parameters can be thought of as being "hardcoded" into the program since they are defined before program execution and not altered until the entire music is calculated. This classical notion of "automated music" is of a very different quality than the notion of

[1] On a Pentium system, 5 tracks in parallel, on a Pentium II system, 10 tracks and more. GENDY3 is 16-track.

interactive composing where the composer enters into a feedback dialogue with the computer, altering the boundary conditions of algorithmic calculation while the algorithm is in action.

As a matter of fact, Xenakis himself wanted the synthesis parameters to change during sound generation, and in 1994, he devised a version of GENDYN where each parameter is "modulated" by either a deterministic or a stochastic function, resulting in a piece called "S709". However, the change of the parameters in S709 is a programmed change, so it is now the set of parameters driving the modulation functions which has to be considered "hardcoded". Moreover, without the real-time control of the human ear it was almost impossible to foretell the impact of those changes on the stochastic generation of sound, and S709 must be regarded as suboptimal in this respect.

The notion of algorithmic composition, as it has been realized by Xenakis in his GENDYN project, is intimately linked with the pure mathematical notion of computation, often presented in the operational model of the "Turing machine", a computational formalism that conceives of an idealized, abstract programmable machine to carry out programmed mechanical action. Xenakis applied the strength of computation to the creation (in a mechanical sense) of the huge number of sound samples defining the aural shape (and as an emergent by-product also the underlying "deep structures") of a musical composition [4]. In the interactive approach, however, the notion of the Turing machine computation must be replaced by the model of an interactive Turing machine that reacts to asynchronous events [8]. It can be thought of as a Turing machine with ever-changing input tape or as a robot moving through ever-changing terrain.

With the means at his disposal, Xenakis was not able to transform his GENDYN sound-producing automaton into such a sound producing robot navigating through an ever-changing terrain set up and changed in real time by the composer. With the Dynamic Stochastic Synthesis now working in real time on a standard computer, the interactive model of composition, where the human and the machine interact in real-time and together form a more complex unity, has become possible. The composer, in reaction to what he/she perceives, pushes the system through a trajectory of states that represents his/her individual way of "playing" the composition algorithm, creating interesting or even surprising results.

3. Reactive Systems

With the advent of powerful and inexpensive computing machinery, the computation paradigm has changed from the classical notion of batch processing to reactive systems. Monolithic systems are split into slim components that are easier to develop, test, evolve and maintain. Smart systems such as "intelligent" tools are designed to interact with the user in a constructive way. If such a system is controlled by another such system, or a human, an interaction loop is established which extends the system beyond its own inherent computational power by harnessing external input [8]. Consequently, a composer who enters a feedback loop with a system of which he has actively participated in developing establishes a dialogue with his complex self. The master-slave relationship in the classical use of the computer is replaced by a more cooperative approach of interaction, challenging human intelligence by contributing genuine computational elements of unpredictability and surprise. It is different from the romantic will to power where the computer is merely used to maximize productive efficiency in the spirit of industrial automation. The new, interactive approach requires that the composer be willing to explore and conquer new sound worlds, instead of asking for the comfort of having preconceived compositional thinking faithfully executed by a machine, a dream doomed to fail due to fundamental differences between the algorithmic nature of machine action on the one hand and human creativity on the other [5].

Composer and engineer Angelo Bello has recently transformed Xenakis' UPIC system into a "composition instrument" by setting up a complex FM "algorithm" with multiple feedback loops that build a complex nested system out of UPIC's 64 hardware oscillators sending the UPIC onto the road to chaotic oscillation[1]. It is striking to compare some of the sonorities of his music with GENDYN sound. In fact, the classical pseudo-random number generators that drive the random walks in GENDYN do nothing else than folding and stretching numbers within a modulo interval in order to create a chaotic number sequence. Indeed, the GENDYN algorithm can be viewed as a combination of stochastic frequency modulation (where the waveforms are not sinusoid, but complex, and the modulator not a periodic, but a stochastic signal) and stochastic amplitude modulation. Depending on the parameter settings, the speed and the impact of the stochastic modulators reaches from a minimum (fixed rigid "tones") to perfect Brownian noise.

4. Distributed Objects

Interactive soft- and hardware systems tend to be no longer developed as closed systems with a rigidly defined overall functionality but composed of small autonomous entities flexibly cooperating for the fulfillment of a superordinated task. To achieve the overall function, software components are plugged together in a specific way. Such systems are easily reconfigured, scaled, extented, and adapted to evolving needs. Components are self-contained, functional entities with a well-defined behavior and interface.

Typically, a component hides implementation details to the outer world, but not the parameters of its functioning: it can be fully automated and monitored through the control of remote components. Since components are autonomous computing entities, they are even more reusable than software objects. Not only are

they reusable by the developing programmer but also by the user himself. The user (=artist) therefore gains considerable independence and emancipation from the system designer, much in the sense of the participatory design paradigm.

If components implement local-remote transparency and multi-threadedness, computation, control and monitoring can be flexibly assigned to different hardware platforms and/or input/output devices, without changing the code. If portable code is used (e.g. Java), components can be made to run on different hardware architectures, and be controlled by various means, e.g. through a Web browser. The access from the outside ("online-studio"), then, comes as a natural consequence of the transparent distribution in the component approach.

5. Design

The component approach is used to seperate the interface and the synthesis engine of the New GENDYN program into two processes (tasks), each implemented in a different language: C++ for synthesis, for sake of efficiency, and a RAD language for the interface (currently Visual Basic). The communication between the components is done by remote object creation and invocation, instead of low-level communication like streams or sockets. The ORB used is currently Microsoft's OLE, but the design can easily be adapted to a CORBA architecture.

The new GENDYN program implements the Xenakis algorithm as a hierarchy of small objects, each fulfilling a specific task, and cooperating for the production of sound. Enumerated from low to high level, objects implement distributions, elastic mirrors, random walks, polygones, sounds, tracks, sequences, the piece and a global control of I/O and playback. The objects encapsulate the generation routines together with the variables and the parameters of the synthesis[2]. All objects are controlled in real time by the graphical interface which contains the control counterparts of the synthesis objects. At program startup, each control object is linked to its remote partner in the synthesis engine. The engine runs as a backgound process (currently implementing the Windows MFC "idle loop"), and it computes the sound as a sucession of sample chunks. All computation is done on a sample-by-sample basis; sound is only buffered for output to the DAC device. In order to speed up computation, the C++ inline declaration feature is used for the routines of the synthesis calling tree (up to 7 levels deep into the object's hierarchy).

During synthesis, the distribution functions are displayed as graphs, and the random walks as billiard balls moving in 2D space. Moreover, either the waveform or a pitch curve (the evolution of the wavelength as a representation of the sound's fundamental) can be plotted. All plotting is done by pulling the current states of the synthesis variables in a timer loop (i.e. there is no callback from the synthesis engine).

The synthesis object hierarchy can either be remotely created by the interface's control objects or by the engine itself parsing a data file. Whereas the stubs of the creation/invocation are integrated into the Basic language (and implemented in the Visual Basic Object Adapter) the C++ skeletons have to be explicitely coded by the server application. (Fortunately, the MFC framework does most of the dirty job.) What may be interesting from the point of view of current ORB implementations is the fact that in the GENDYN server, the skeletons inherit from the implementation classes, a design that allows almost 100% transparency[3]. (However, for the server to expose its synthesis objects to remote control, object creation has to be implemented in virtual methods that can be overridden by the skeletons.) Thus the synthesis engine can also be used stand-alone in "command line" mode without changing a single program line, because the skeleton classes are only externally linked to the server.

6. Software Engineering Issues

The difference in a nutshell between the original GENDYN program and modern programming techniques can be seen in the fact that the new GENDYN program comprises less than half the size of source code of the old program (7000 vs. 15000 lines), although a graphical interface has been added. At the same time, its binary file is fifteen times as large (about 1 MB, without counting the dynamically linked system libraries). That means that there is a lot more support by the programming environments today used to implement additional functionality (like e.g. remote calls).

The load of interactive graphic processing may be assigned to a different computer in a (possibly heterogeneous) network or even across the Internet, if standard protocols (like CORBA) are supported. Moreover, the distributed design allows for an independent development of the "purely algorithmic" and the interactive part of the program. For example, it will be easy to reintroduce the idea of time-variant parameters that Xenakis tried in the further development of his program. It could be done in the interface part of the new GENDYN program, without interfering with the sound computation algorithm.

[2] There are actually 2 layers of objects: one encapsulates the synthesis state variables, the other the synthesis parameters, in order to keep them logically seperate.

[3] This is different from e.g. Visigenic's CORBA implementation, where the skeletons either are base classes of the implementation classes (default) or delegate the incoming calls to the implementation objects ("tie-mechanism").

7. Future Work

It would be interesting to explore the effect of stochastic transformation to existing sound sampled into the program. The Dynamic Stochastic Synthesis would thus becomes a novel sound *transformation* tool. Depending on the boundary conditions imposed onto the Dynamic Stochastic Synthesis process by the user, transformation could vary from slight stochastic enrichment of the sound spectrum to complete independent complex sonic evolution. The degree of faithfulness of the transformation would also depend on the rate and the frequency (regular or irregular) with which the original sound is sampled. In one extreme, the stochastic signal could be defined to faithfully follow in the (harmonic and pitch) tracks of the original sound; in the other extreme, it would just be triggered by the input sound and take on its own route.

Controlled by a human during live performance or in the process of composing, the gap between instrumental and computer sound could be bridged by adopting a radically new approach, where neither the computer is made to "imitate" instrumental sound, nor instrumental sound used to "embellish" or "humanize" genuine computer sound. Mutual penetration of human and computational rendering of sound would be based on equal grounds, with the computer taking benefit of the specific complexity of human creative action on the one hand and the human taking benefit of the specific complexity of computational algorithmic action on the other.

8. References

[1] Angelo Bello, "An Application of Interactive Computation and the Concrete Situated Approach to Real-Time Composition and Performance", International Computer Music Conference, Ann Arbor, MI, 01.-06.10.98, forthcoming.

[2] Joel Chadabe, "Interactive Composing: An Overview", *Computer Music Journal* vol. 7 (1983), pp. 22-27

[3] Peter Hoffmann, Implementing the Dynamic Stochastic Synthesis, *Troisièmes journées d'informatique musicale JIM 96* (=Les cahiers du GREYC , 4, 1996), p. 341-347.

[4] Peter Hoffmann, "Evaluating the Dynamic Stochastic Synthesis", *Journées d'informatique musicale JIM 98* (=Publications du LMA no. 148, mai 1998), CNRS Marseille, pp. F4-1-F4-7.

[5] Hoffmann, Peter, "Music Out of Nothing? The Dynamic Stochastic Synthesis: A Rigorous Approach to Algorithmic Composition by Iannis Xenakis", Ph. D. Dissertation, Technische Universität Berlin, forthcoming.

[6] Holtzman, Steven R., "An Automated Digital Sound Synthesis Instrument", *Computer Music Journal* vol. 3, no. 2 (1979), pp. 53-61.

[7] Marie-Hélène Serra, "Stochastic Composition and Stochastic Timbre: Gendy3 by Iannis Xenakis", *Perspectives of New Music 31* (1993), pp. 236-257.

[8] Peter Wegner: "Why Interaction is More Powerful than Algorithms", *Communications of the ACM*, vol. 40, no. 5 (May 1997), pp. 80-91.

[9] Iannis Xenakis, "Music Composition Treks", in: Curtis Roads (ed.): *Composers and the Computer*, William Kaufmann. Los Altos, CA, 1985, pp. 172-191 (quotation on p. 180, italics by me).

[10] Iannis Xenakis, *Formalized Music*, Pendragon Press, Stuyvesant, NY, 1992.

The Corner Effect

Damián Keller , Chris Rolfe
damian_keller@sfu.ca, rolfe@sfu.ca
http://www.sfu.ca/~dkeller

We discuss some theoretical and practical aspects of real-time granulation of sampled sounds, such as windowing, grain overlap, synchronicity, and control through high-level events. Our analysis of the trapezoidal window has shown that it approximates the response of a Gaussian window, with the addition of comb-shaped spectral effects. Zeros are proportional to the position of the 'corners' of the window. Therefore, we call the artifacts as the 'corner effect.'

Keywords: real-time granular synthesis, windowing, ecological models.

The introduction

This paper discusses some of the processes involved in granular synthesis (GS), in an effort to identify relevant variables in granular temporal and spectral transformations. Windowing, AM effects, grain overlap and their interaction produce complex time-varying spectral profiles. We address these issues in relation to the implementation of MacPod [11], a real-time GS system for the Macintosh PowerPC which is based on Truax's (1988) POD system. Furthermore, we discuss some new concepts and techniques relevant to the development of ecologically-based sound resynthesis, namely, the use of local or global parameters to define granular events, the control of phase-synchronicity among streams, and the simplification of windowing by using pre-stored grains [5].

The implementation of a real-time granular synthesis (GS) system on a personal computer presents two basic challenges: (1) an efficient use of computational resources to generate high grain densities, (2) a simple and intuitive organization of synthesis parameters to facilitate real time control. The first issue is directly related to the synthesis engine of the system, that is, how the source sounds are windowed and mixed. The second issue corresponds to the control level of the system, which concerns how the synthesis parameters are generated and how the user-performer-composer interacts with them.

The complex

The interaction among processes in asynchronous GS generates rich sound results with fairly little source material. We have identified four causes for the increased complexity of granulated sound. (1) By applying an envelope, or window, we produce a signal equivalent to the convolution of the impulse response of the window and the sampled sound. In other words, the window applied on the original signal causes a resonant main lobe and several spectral sides lobes which smear the original spectrum. (2) At subaudio grain rates, amplitude modulation adds upper and lower components to the granulated sound. The spectral modifications are proportional to the spectral content of the signal and the grain rate applied. (3) The overlap among grains in different voices produces time-varying cancellation and reinforcement which also modify the spectrum of the original signal. (4) When time-stretching is applied to a single sound file, time-delayed copies of the granulated signal are overlaid. This process produces temporal and spectral effects that depend on the stretch-ratio being used.

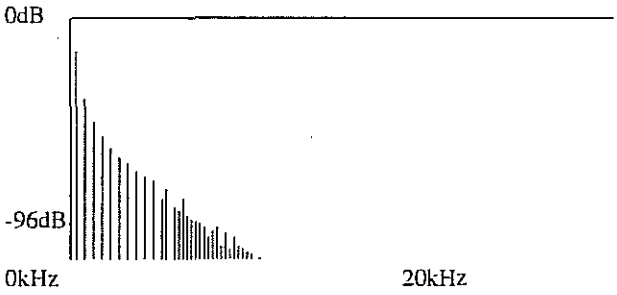
The window

Windowing effects in audio signal processing are generally well understood. Window functions used in spectral analysis, such as von Hann and Hamming, minimize unwanted artifacts but increase the computation time [7, p. 149]. To reduce computational cost, the earliest real-time granular synthesis systems [12] used simple trapezoidal windows to good aural effect.

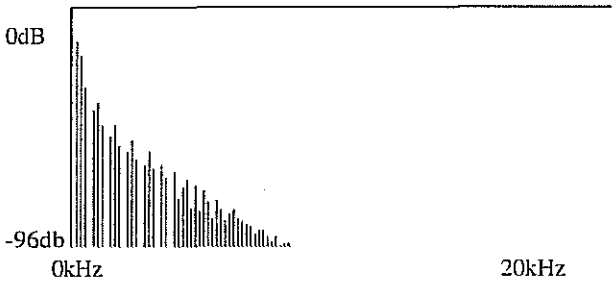
We have focused our research on the effects of the lowly trapezoidal grain window, within the context of asynchronous granular synthesis. The trapezoidal window, in fact, resembles the popular Gaussian window, with the addition of ripples that produce an effect aurally similar to comb-filtering. Zeros are proportional to the position of the 'corners' of the window and hence, we call the artifacts as the 'corner effect.' (Fig. Spectrum of a trapezoidal window).

While undesirable for most signal processing applications, this filtering effect is unobtrusive in GS. As we will discuss in 'The Overlap' section, aurally similar modifications of the signal are inherent in the GS technique due to the delay between overlapping grains. Therefore, we can confidently state that complex windowing is unwarranted for granular synthesis at medium-to-high grain densities. We invite the reader to compare the spectral effect of a triangular window with a

trapezoidal window using identical synthesis settings. Both spectrograms show very similar results, with a slight 'smearing' of the spectrum when the trapezoidal window is used.



Spectrum of a triangular window.



Spectrum of a trapezoidal window.

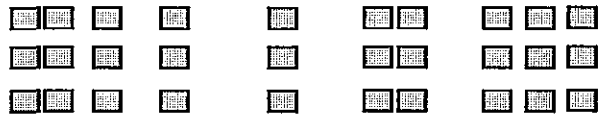
Our particular focus is the application of GS to modeling environmental sounds. High grain densities (approaching one thousand grains per second) are needed to model complex, time-varying sound events. The chief objection to densities of this magnitude in real-time systems is the inefficiency of windowing and mixing the grains [2]. Using the trapezoidal function, however, we achieve real-time GS with the required density on a standard Macintosh PowerPC.

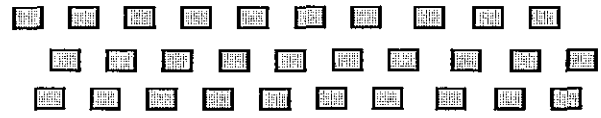
The overlap

The grain overlap is defined as the time interval during which two or more grains are sounding simultaneously [4]. An average grain overlap can be estimated by the difference between the average grain rate and the grain duration. If grain duration is longer than the grain rate, overlap occurs. Thus, there are three possible configurations: (1) negative overlap, there is a delay between the end of a grain and the onset of the following grain, (2) no overlap, a grain starts when the previous ends; (3) positive overlap, before a grain ends the next one starts. In batch implementations, there can be as many overlapping grains as memory and patience allow. On the other hand, real-time constraints place a limit on the number of simultaneously sounding grains. MacPod can achieve up to 20 simultaneous grain streams, with a minimum grain rate of one millisecond.

Although some GS systems combine several grain streams into a single voice [1], it is conceptually clearer to conceive each voice as a separate stream. Thus, overlap can be controlled from a unique parameter which stands for the coincidence [3], or phase-synchronicity, among grain onsets in all active voices. Following the central limit theorem [8, p. 174), it is reasonable to state that if each grain stream is defined as an independent random process, the overlap distribution will eventually approach a Gaussian probability distribution.

Careful control of phase-synchronicity among grain onsets in different streams produces transformations in the temporal and spectral profile of the granulated sound. With very fast grain rates - under 5 ms. - using pitched sample material, we obtain formants akin to those produced by FOF synthesis. A small delay between grain onsets adds volume (as defined in [13]) to the original signal, producing an effect akin to early reflections in a reverberant space. Of course, we must keep in mind that all these processes are independent from the asynchronous grain rate established for each stream.


Phase-synchronous streams.


Phase-asynchronous streams.

Within the context of ecologically-oriented resynthesis, phase-synchronicity is especially meaningful in the simulation of attacks. In stricking a solid object, most resonant frequencies will be excited in the first fifty milliseconds or less. Contrastingly, if the excitation is produced by several small objects, each impact will excite different frequencies at various time delays causing a granular sound texture. This type of sound can be heard when walking on glass pieces or on snow.

The stream

A grain stream generator produces a series of grains with a given frequency, amplitude and duration. These parameters can vary in time. The concept of grain generator implies that only a single grain can be

produced at a time. Thus, when more than one *simultaneous grain is desired* (to produce overlaps) several grain generators have to be used. This introduces the need to define the phase relationship between the grain streams. The phase-asynchronous implementation, as found in asynchronous GS, produces streams which are completely independent. If the time among the grains in different streams is to be controlled, a phase-synchronous approach is necessary. As we stated before, *in this case the grain onsets can be synchronized across* streams or a short delay may be used. Therefore, there are three possible configurations: (1) a single stream generator, (2) multiple phase-asynchronous stream generators, and (3) multiple phase-synchronous stream generators.

The waveform

GS techniques have used different types of source material: (1) sine waves, in FOF synthesis [10]; (2) FIR filters derived by spectral analysis, in pitch-synchronous granular synthesis; and (3) sampled sounds, in asynchronous GS [12], FOG, and pulsar synthesis [9]. Ecologically-based resynthesis adds the option of using pre-stored sample grains [6].

More especifically, in ecologically-based GS we create a grain pool before the synthesis stage, instead of retrieving arbitrary segments of the sound file. The samples keep the spectro-temporal characteristics of the short original sounds, avoiding the 'blurring' effect that occurs in asynchronous GS [9]. These samples are placed on a time-frequency grid according to meso-level time patterns which are, in turn, designed to match the temporal characteristics of naturally occurring sounds, e.g., bounce [6]. Given that this approach simplifies the windowing process, it may provide a good alternative to existing real-time methods.

The pointer

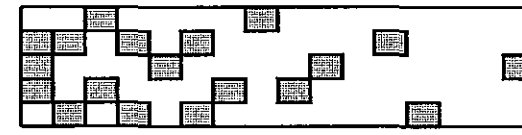
GS systems access the sound database contents in four different ways to: (1) incremental, the file is read from beginning to end; (2) loop, the file is read repeatedly from beginning to end; (3) cycle, the file is read repeatedly from beginning to end and backwards; and (4) random, the file is read at random locations.

The current implementation of MacPod, following the POD model, uses a single pointer to source material. Interestingly, the effect of the overlapping grains can be simply explained as a comb-filter delay. If one assumes a fixed grain envelope, an asynchronous grain six milliseconds later than the original is simply a six-millisecond delay mixed in with the original signal. By keeping the resolution at a sample level, we are able to explore a variety of spectral transformations - at subaudio rates - and reverb-like effects at slower rates.

The event

A logical implication of the ecological approach to sound resynthesis is to establish the sound event [6] as a high-order unit of sound generation. Resynthesis parameters are thus directly linked to a finite time length. Rate of change is scaled according to the length of this event. Instead of fine-tuning unrelated parameters *(such as amplitude or frequency of a given grain stream)*, transformations of a sound event are carried out along correlated variables within ecologically valid time ranges.

We point out two possible strategies: (1) High-level events are defined by global settings. These settings define ranges of possible values for the local parameters. (2) Local parameters determine the overall behavior of the high-level event. For example, the density of an event can be defined by two global parameters: duration and quantity of grains. If grains with fixed duration are evenly scattered along a predefined time span, we get an invariant average density. But let's say that we want to have a dense distribution that changes linearly to a sparse one:



Tme-varying grain distribution.

If synthesis parameters vary independently, we will spend several trials until we find the right amount of grains and the right rate of change in distribution. On the other hand, by using grain overlap as the only control variable and letting the quantity of grains and the overall duration change accordingly, we will be dealing directly with the relevant perceptual parameters. In this example, the only high-level variable that needs to be defined is the rate of change in grain overlap.

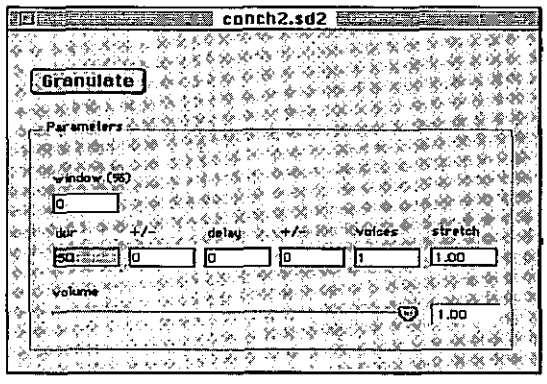
The conclusion

We have investigated several issues involved in the implementation of a real-time granular synthesis application. The focus of our work has been the efficient use of computational resources, and a simplified method for synthesis parameter control.

Our results point to two effective approaches to windowing: (1) the use of a trapezoidal function, as suggested by Truax (1988), (2) the use of a grain sample pool, as implemented in ecological sound resynthesis. By applying a trapezoidal window, we obtain aurally *effective results with a drastic reduction of* computational time. This type of window produces a

spectral profile which depends on the placement of the 'corners' of the trapezoid. Thus, what has been regarded as an unwanted artifact by DSP theory, becomes a useful parameter for sound synthesis.

Our current efforts are concentrated on bringing the ecological perspective to the real-time realm. By using events instead of low-level control parameters, we pave the way to a more intuitive interface between user input and sound output. At the other end, the independence in grain-rate control and the resolution of grain overlap at a sample level permit not only to work on the temporal characteristics of the sound, but also to shape its spectral profile.



MacPod: real-time granular synthesis for the Macintosh PowerPC.

The References

[1] Behles, G., Starke, S., & Röbel, A. (1998). Quasi-synchronous and pitch-synchronous granular sound processing with Stampede II. *Computer Music Journal*, 22(2), 44-51.

[2] Cook, P.R. (1997). Physically informed sonic modeling (PhISM): synthesis of percussive sounds. *Computer Music Journal*, 21(3), 38-49.

[3] Dziech, A. (1993). *Random Pulse Streams and their Applications*. Warszawa: Elsevier.

[4] Jones, D.L., & Parks, T.W. (1988). Generation and combination of grains for music synthesis. *Computer Music Journal*, 12(2), 27-33.

[5] Keller, D. (1998). ". . . soretes de punta." *Compact disc Harangue II*. Burnaby, BC: Earsay. (http://earsay.com)

[6] Keller, D., & Truax, B. (1998). Ecologically-based granular synthesis, *Proceedings of the International Computer Music Conference*. Ann Arbor, MI: University of Michigan.

[7] Lynn, P.A., & Fuerst, W. (1998). *Introductory Digital Signal Processing with Computer Applications*. Chichester: John Wiley.

[8] Mix, D.F. (1995). *Random Signal Processing*. Englewood Cliffs: Prentice Hall.

[9] Roads, C. (1997). Sound transformation by convolution, *Musical Signal Processing*, C. Roads, S.T. Pope, A. Piccialli, & G. De Poli (Eds.). Lisse: Swets & Zeitlinger, 411-438.

[10] Rodet, X. (1984). Time-domain formant wave-function synthesis. *Computer Music Journal*, 8(3), 9-14.

[11] Rolfe, C. (1998). *MacPod*. Real-time asynchronous granular synthesis software for the Macintosh PowerPC. Vancouver, BC: Third Monk Inc.

[12] Truax, B. (1988). Real-time granular synthesis with a digital signal processor. *Computer Music Journal*, 12(2), 14-26.

[13] Truax, B. (1992). Electroacoustic music and soundscape: the inner and outer world, *Companion to Contemporary Musical Thought*, Vol. 1, J. Paynter, T. Howell, R. Orton, & P. Seymour (Eds.). London: Routledge, 374-398.

FEEDFORWARD NEURAL NETWORKS FOR PIANO MUSIC TRANSCRIPTION

Matija Marolt

Faculty of Computer and Information Science
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia

Abstract

The paper presents our first experiences with the use of feedforward neural networks for transcription of polyphonic piano music. Recently, some attempts of building systems that would successfully transcribe polyphonic music with more than two voices ([4], [8]) have been made. However, the systems presented were built using conventional techniques, which require no learning, but manual tuning of system's parameters. Artificial neural networks have been used for speech recognition and other pattern recognition tasks for a long time. They are especially suitable for such tasks, because of their ability to learn from examples, generalise and robustness to noise. We present preliminary results obtained by using feedforward neural networks for piano music transcription, where by transcription we mean recognising the note and time when the note occurred (length and dynamics are not taken into consideration). The paper presents the system and results in more detail and gives some ideas for further work.

1 Introduction

1.1 What is transcription?

Music transcription could be defined as an act of listening to a piece of music and writing down music notation for the piece. For each note its starting time, duration and loudness (dynamics) need to be determined.

Music transcription is a difficult cognitive task. It can be (to some extent) easily performed by trained humans, but it is a very difficult problem for current computer systems to solve. We could in a way compare it to speech recognition, where we convert an audio signal to phonemes, syllables and finally words; with music transcription we convert an audio signal into notes, their starting times, duration and loudness. Speech recognition has recently been quite successfully solved (at least for single words) and we see more and more applications coming to the market. For polyphonic music transcription this is not yet the case.

First attempts of transcribing polyphonic music have been made by Moorer [6]. His system was limited to two voices of different timbres and frequency ranges and had limits on allowable intervals. Later other systems have been developed. [4] for example, uses a blackboard system for piano music transcription (see also [4] or [8] for more references).

Until now we have still not encountered any system that would employ machine learning algorithms (such as neural networks) for transcription. Since these algorithms are successfully used in other pattern recognition tasks, it is our main motivation to study the usability of neural networks for transcription.

1.2 Artificial neural networks

Artificial neural networks are a class of machine learning algorithms, inspired by biological models of neurons and their interconnections. Artificial neural networks are of course very simplified models, which account for only the most basic neural processes, but as it turns out, they can provide good solutions for many practical problems, such as classification (classifying inputs into one or many classes), noise reduction (recognition of patterns corrupted by noise), prediction,...

There is a very large variety of neural network models in existence. A typical structure of a neural network, although not the only one (there are also many other models in use), is as follows. A network can have one or more (usually many) inputs, which are commonly called input neurons (although they perform no processing of input data). A network also has one or more output neurons, which provide results to the world. Between the input and output neurons is a so-called "black box", containing additional layers of neurons and their interconnections. The exact network model determines the nature of this black box. Some models have only feedforward connections, some can have feedback (recurrent) connections, some models connect the input neurons directly to the output neurons,...

When we define the basic structure of a neural network, we usually train it in order to build certain "knowledge" into it. There are three ways of training a neural network. The most common is supervised learning, where we collect many input samples to serve as exemplars. These samples constitute the training set, which completely specifies all inputs, as well as outputs for the network. We present the inputs and outputs to the network and update its connections in order to reduce a measure of the error in the network's results.

The second training method is unsupervised learning. Here, we also collect some sample inputs, but we do not provide the network with outputs for those samples. The network itself tries to find some features

in the training set and group input samples into classes it finds distinct.

There is a third training method called reinforcement learning. This method is a hybrid between supervised and unsupervised learning. It is unsupervised in the sense that the exact outputs of the network are not specified. At the same time, it is supervised in that when the network responds to a sample in the training set, it is told whether its response was good or bad.

2 Transcription with artificial neural networks

This paper presents our first attempt of using artificial neural networks for music transcription. To simplify the domain, we limited ourselves to piano music, because piano notes have discrete pitches, which do not modulate. Our the goal is to correctly determine the starting times and notes of a polyphonic piano performance.

The basic structure of our current system is shown in Figure 1.

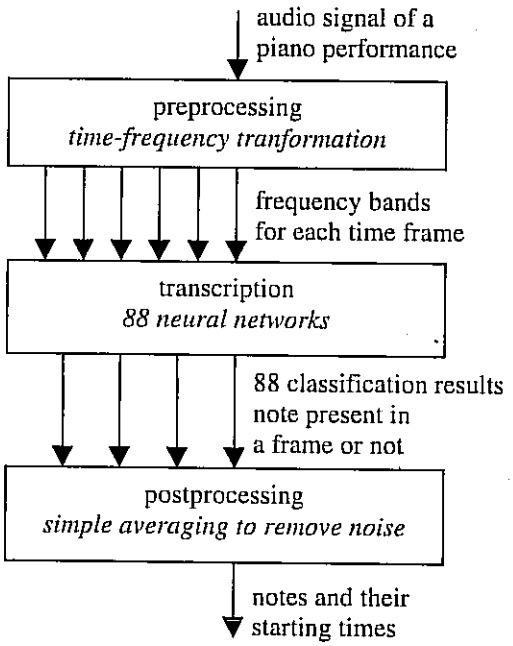


Figure 1: structure of the system

The system has three main parts:

- preprocessing stage takes the digital audio signal (time-amplitude) of a piano performance and performs a time-frequency transformation to obtain the time-frequency-magnitude spectrum.
- 88 neural networks (one for each note) process the output frames of the preprocessing stage and produce outputs showing whether a network considers a note to be present in the time frame or not.
- postprocessing stage takes the outputs of neural networks and performs simple time averaging to produce the final output.

2.1 Preprocessing – time-frequency transformation

To transform the input audio signal into time-frequency space, we chose a simple correlation-based transformation. Results of the transformation are 290 logarithmically spaced frequency bands, the highest band located at 5000 Hz. Time resolution of the bands ranges from 45 ms at lower frequencies to 11 ms at higher frequencies, Hamming window is used for windowing. Individual bands overlap quite a lot and there is a lot of redundancy in the transformation, but this is not a problem, since the type of neural networks we used for transcription can handle correlated and redundant inputs very well.

2.2 Transcription

To perform the transcription, we used a class of neural networks called multilayer feedforward networks. Such networks consist of a set of neurons arranged into two or more layers - they may also have one or more "hidden" layers between the input and output layer. Feedforward means, that all information flows in only one direction, from input to the output neurons (there are no feedback connections). In our case we used one hidden layer where each neuron in the hidden or output layers was connected to all of the neurons in the previous layer (layers are fully connected).

Each neuron calculates a function of all of its inputs to produce its output. Let's assume that a neuron has n inputs, labelled from 0 through n-1. A fictional input called bias is also present and is always equal to one. Each neuron is then characterised by $n+1$ weights (n inputs and bias) and an activation function f. It computes its output as:

$$out = f(\sum_{0}^{n-1} x_i w_i + w_n)$$

For the activation function f, we chose the standard logistic activation function, which is very commonly used.

In the learning process, the network's weights are updated, to reduce the mean square error of the network's outputs. There are many training algorithms for multilayer feedforward networks; we used the scaled conjugate gradient method, which proved to converge effectively and produced good results.

Each network has been trained to recognise a single piano note, so we had to train 88 networks altogether. To train the networks, a large training set of approximately 15000 chords was constructed. The chords were constructed from individual piano notes, obtained from several commercially available piano synthesiser patches and sampling piano CD-ROMs. For each note, a set of 160 chords with that note were constructed (50 with 2 notes, 40 with 3, 30 with 4, 20 with 5 and 20 with 6 notes) by randomly choosing accompanying chord notes. In the final training set, each note was present in approximately 600 chords.

For each chord, a time frame of 45 ms after the attack portion of the chord was extracted and transformed with the transformation described in section 2.1. These transformed parts were then scaled to a lesser magnitude and used for supervised training of the networks. Each network has been trained to classify whether its particular note is in the input chord or not (its input is a member or is not a member of a class). Therefore, each network has been trained to an output of 0.9 if its note was present in the chord or 0.1 if not. We used SNNS (Stuttgart Neural Network Simulator) v4.1 for training the networks – see [10].

The trained networks were then used for transcription by presenting them with consecutive time frames from the frequency transformed piano performances as their input. The training results on chords and transcription tasks can be found in section 3.

2.3 Postprocessing

As mentioned before, a simple postprocessing algorithm was used for processing the outputs of all 88 networks. The outputs are numbers from 0 to 1, indicating how each network classified its inputs. Postprocessing currently consists of a very simple time averaging of outputs to prevent a single high neural network activation to cause the system to mark a note as present. Several high neural network activations are needed for a note to be declared present in the input audio signal.

3 Results

3.1 Chord recognition results

As mentioned before, we have trained our networks on a set of approximately 15000 chords, obtained by mixing single notes of 10 different piano patches (synthesizer and sampled pianos). Each network has been trained to recognize whether a particular note is present in a chord or not. To determine, which neural network architecture gives good results, we tested several network architectures with different number of input and hidden neurons.

- Networks with 290 and networks with 69 neurons in the input layer were tested. Networks with 290 input neurons take the entire spectrum of the time-frequency transform described in section 2.1 as their input. The input of networks with 69 input neurons is not the entire spectrum, but only: frequencies around the note which the network is trained to recognize, frequencies 1 octave up and down (2. harmonic), 19 semitones up and down (3. harmonic) and 2 octaves up and down of the note (4. harmonic).
- Networks with different number of neurons in the hidden layer were tested. We trained networks with 2, 5, 10, 18, 30, 50 and 88 neurons in a single hidden layer. This was done to establish the approximate optimal number of neurons in the hidden layer, regarding the speed of training and accuracy of test results.

All of the networks have been tested on an independent training set, which consisted of approximately 4000 chords obtained by mixing single notes of 16 different piano patches (six new were added in regard to the training set). All of the chords in the test set were different than the ones in the training set.

Average classification accuracy of all 88 networks (for all notes) is 99%. The accuracy ranges from 97% for the lower octaves to over 99.5% for higher octaves. This was expected due to better frequency resolution at higher frequencies.

Networks with 69 input neurons performed slightly better than networks with 290 input neurons, although the difference is very small.(usually less than 0.5%).

The optimal number of hidden neurons that gives best results is 18. This number is sufficient also for lower octaves, where less hidden neurons (10) produced worse results, while for higher octaves 10 neurons would also be sufficient.

We also looked at the types of errors made by our networks. Most of the errors made (over 70%) are "missing note" errors, where the note was present in the chord, but was not recognized by the network. This was expected, since the training set for a note includes more chords in which the note is not present, than chords in which it is. "Octave errors" where a note an octave below (or above) causes a misclassification represent around 20% of all errors made, while "halftone errors" represent around 15% of all errors. Other errors occurred due to other more or less obvious reasons, some of them also because of the "holes" in the training set – randomly generated chords may insufficiently represent some features necessary to correctly recognize notes in some chords.

3.2 Transcription results

After training, we tested the trained networks in the context of the system described in section 2. We used several MIDI files of solo piano pieces and rendered them with different sampled piano patches. The pieces ranged from very simple Bach's Two-part Inventions, samples from the Well Tempered Clavier to more complex excerpts from Tchaikovsky's Nutcracker Suite.

piece	right	late	missed	false
2 Pt. Invent.	50.00	42.59	7.41	7.76
3 Pt. Sinfon.	51.09	47.45	1.46	5.33
Engl. Suite	62.03	28.88	9.09	8.69
Overture	50.46	14.68	34.86	13.95

Table 1: Transcription results

Transcription results for four pieces can be seen in table one. The second column (*right*) represents the percentage of correctly transcribed notes. The third column (*late*) represents the percentage of notes, which were correct, but their starting times were not estimated correctly (they were placed some time frames later as the real note started). The fourth column (*missed*)

represents notes, which were not found by the algorithm, while the fifth column (*false*) represents notes, that were found by the algorithm, but did not exist in the original score.

Results are given for four pieces; J.S. Bach's Two-part Invention No. 1 (polyphony mostly 2), Bach's Three-part Sinfonia No. 1 (two or three voices polyphony), Bach's English Suite No. 1 (polyphony 4) and an excerpt of Tchaikovsky's Nutcracker Suite Miniature Overture (polyphony over 4).

As can be seen, results get progressively worse as the polyphony increases. There is also a very large percentage of notes, for which the starting times were transcribed later than they actually occurred. We believe that these errors occur due to the fact, that the networks were trained only on steady portions of chords in the training set. The attack portion was left out, so the networks react unpredictably until the sound of each note settles down to its steady portion.

4 Conclusion and future work

In the paper we presented our first experiences with the use of multilayer feedforward neural networks for transcription of piano music. Results obtained so far are not brilliant, but there is a lot of space for improvements.

Within the current system, several things could be improved. The training set could be better constructed to give a more balanced ratio of positive and negative cases (chords with and without a given note). Also, randomly constructed chords might not be the best solution. A chord constructing procedure taking into account the way the piano is played (we only have two hands) and the fact that extreme notes (bass or treble) are not so commonly played as notes in the middle part of the keyboard might result in better trained networks with better performance.

Other time-frequency transformations should be tested for the preprocessing part of our system. Musical wavelets [7] or correlogram techniques [1] are good possible alternatives.

Time plays an essential part in music. Each note is starts on a certain moment and has certain duration. One of the main problems of our model is that time only plays a role in the postprocessing stage of our system. Neural networks perform transcription on single time frames and do not "know" what will happen next or at least what has happened before. This is also one of the main problem of the currently used neural network model. Multilayer feedforward networks do not have mechanisms for dealing with data evolving through time. Of course, we could always put more than one time frame in the input layer (have more input layer neurons for more time frames), but this would not be the best solution. Other neural network architectures that do have some notion of time should be tested. Architectures used in speech recognition systems for phoneme recognition could be taken as examples. Two such good alternatives are time delay neural networks [9] or partially recurrent neural networks [2].

In the current system, networks are trained on a different domain (chords) than the one they are intended to be used on (transcription). This should be changed, so that the networks would be trained on piano performances instead of chords.

Finally, the postprocessing stage of our system should be changed. The current version is too simple to be able to filter out noisy outputs and network errors well. A blackboard system (see [4], [3]), integrating results directly from the preprocessing stage and from neural network outputs could be useful.

References

[1] Ellis, D.P.W. *Prediction-driven computational auditory scene analysis*. Ph.D. thesis, Department of Electrical Engineering & Computer Science, M.I.T., 1996.

[2] Kirschning, I., Tomabechi, H. "Phoneme Recognition Using a Time-Sliced Recurrent Recognizer," *Proceedings of the 1994 IEEE International Conference on Neural Networks*, USA, pp.4437-4441, 1994.

[3] Klassner, F. *Data Reprocessing in Signal Understanding Systems*. PhD. Thesis, University of Massachusetts Amherst, 1996.

[4] Martin, K.D. A Blackboard System for Automatic Transcription of Simple Polyphonic Music. MIT Media Laboratory Perceptual Computing Section Technical Report No. 385, 1996.

[5] Masters, T. *Practical Neural Networks Recipes in C++*. Academic Press, San Diego, CA, USA, 1993.

[6] Moorer, J.A. *On the segmentation and analysis of continuous musical sound by digital computer*. PhD thesis, Department of Music, Stanford University, CA, 1975.

[7] Newland, D.E. "Harmonic and musical wavelets," *Proc. R. Soc. Lond. A.*, vol. 444, pp. 605-620, 1994.

[8] Nunn, D., Purvis, A., Manning, P. "Source Separation and Transcription of Polyphonic Music," http://capella.dur.ac.uk/doug/icnmr.html, 1997.

[9] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. and Lang, K. "Phoneme Recognition Using Time-Delay Neural Networks", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 3, March 1989.

[10] Zell, A., et al. *Stuttgart Neural Network Simulator User Manual.* http://www.informatik.uni-stuttgart.de/ipvr/bv/projekte/snns/snns.html. 1995.

PHYSICAL MODEL AND INSTRUMENTAL SOUND. AN ANALYSIS OF LUPONE'S "CORDA DI METALLO"

ALESSANDRO MASTROPIETRO

Via Colle Pretara, 51/B, 67100 L'Aquila, Italy
Tel. 0039-862-317167 -
Fax. 0039-862-412950
ale_mastropietro@hotmail.com

Abstract

The paper will focuses on *Corda di metallo* (1997), a work for string quartet, tape and live electronics by the Italian composer Michelangelo Lupone and written for Kronos Quartet. This work employs a new synthesis algorithm for the synthetic sounds of the tape and a spazialitation system with a relevant task in the compositional design. Indeed, the tape sound are synthesized by a new model of physical simulation of the string; this new model renews the previous ones, by inserting new control parameters (like the internal deadening due to the attrition of the string with the colophonia on the bow) and introducing a refined control of the ways for string excitation. The non-linear complexity of this model is reflected in Lupone's score by the richness and multiplicity of the instrumental techniques experienced, so designing sound textures with a chaotic articulation. Such a musical thought is strengthened by the formal path of the work and the employment of spazialization: the formal path is described by the gravitation of the compositional effort around definite linguistic attractors (timbral masses, chords, thematic and rhythmic cells); for each one of these compositional nuclei, the composer designs specific spatial behaviours of sound, that is static, in approaching or removing movement along precise trajectories, in chaotic motion, in which case the spatial parameter becomes the central one.

1. The Work. Sound-Synthesis of magnetic tape and instrumental Sound

Corda di metallo for string quartet, magnetic tape and live electronics, represents in the catalogue of Michelangelo Lupone (Solopaca, BN, Italy, 1953) a crossroads of various experiences and research streams. Besides his favourite aesthetic and compositional themes, it's possible to meet in this 1997-composed work (performed, in his wordl premiere, by Kronos Quartet in Roma) his individual researches on the experimentation of new techniques on bow instruments, the sound spazialization by avanced live-electronics tools and - a relatively new element in his musical path - the employment of a synthesis algorithm according to a physical model.

The physical model is here a string of a string-instrument excited by the rubbing of a bow. The synthesis algorithm for the tape sounds has been realized by Marco Palumbi and Lorenzo Seno by CRM (Centro Ricerche Musicali), and provides for the parametrical control for tension/density ratio of the string, the attrition coefficients, the bow pression and velocity, the point of bow application on the string and finally the lenght of the string [1]. A cross-reference to the relation of the algorithm designer, inside this congress, is recommended for the algorithmic and computational specifics. A total novelty of this algorithm in comparison with the Karplus-Strong one - an historical landmark for the string-instruments simulation - is certainly remarkable. The last one, with his articulation in two segments with defined functions (respectively excitation and resonance of the instrument), favoured especially the plucked string, but it is substantially far from the structure of a physical model, and so from his conceptual end computational complexity too.

The *Corda di metallo* algorithm (so called in homage to the work which has been projected for) introduced this complexity by considering first some new factors connected to the friction of the bow on the string, and this factors generates complex and chaotic sound manners. These are besides in conceptual syntony with Lupone sonical and compositional thought, and they are also well joined with his individual experimentation on bow instruments, aimed at generating sound textures marked by a chaotic complexity which transform the traditional sound of the instruments until their unrecognizability. The project of sound spazialization integrates with

coherence the aesthetic-compositive scene of this work: spazialitation (both for quartet- and tape-sound) is here intended to move sound along approaching or removing trajectories or chaotic motion.

Going back to the tape, it consists of five sequences introduced in exact points of the score, with the task of extending the parallel sound-research on the four bow-instruments and precising as well as enlarge infromative degree and richness of the complex texture determined by combining acoustic and synthetic sounds. Lupone realizes infact the sounds synthesis by assigning to control parameters some around- or out-of-bounds values according to the normal physical reality of bow instruments. The instrumental experimentation too is carried to the extremes of organological powers; besides the use of a widest range of harmonics, timbral research is deepened by a continous and detailed moving of the bow friction-point on the string according positions and modalities of bow conduction (over the fingerboard, circular motion of the bow, highest pression and lowest velocity of the bow) not limited to the traditional ones (fingerboard, normal, bridge; gettato, martellato, tremolo, col legno...). In the particular case of harmonics with the bow over the fingerboard, an interference between the harmonic node and the bow node takes place, that causes a sound with a non-determinated timbral grain and pitch-reference, similar to a grotesque distorsion of human voice yet employed by Lupone in the radiodrama *In un grattacielo*.

Another interesting experimental field on the instruments, important for the sonic character of the tape, is that one on glissandi and micro-fluctuation of pitch (that is micro-tonal vibrato); such a field, familiar to the last 50 years music (Xenakis, Scelsi...), is functional, in Lupone, to embody that fine clinamen which causes chaotic structures of a non-deterministic, but form-owner kind (f. e. strange attractors). Glissandi and micro-fluctuation are obtained by Lupone not only with traditional manners, but pushing with the left hand the active string towards the lower, so transforming his length and tension. These researches, taken ahead by Lupone since his previous works for violin and electronics (*Ciclo Astrale parte II* - 1986, excerpts from *In un grattacielo*), reflect themselves in the contents of magnetic tape; the first four sequences especially, give rise to complex textures, centred on some reference pitch and dense of pitch- and timbre-fluctuation; nevertheless timbre remains adherent to a harsh and rough sound-grain, related to the "metallic" character of simulated string, chosen for the contemporarity of its sound-material (metallic steel) in comparison of the traditional catgut [2]. The acoustic sound of live string-quartet integrates itself in these textures both by quartettistic writing, and by that live-electronics, whose task I will discusse later. Pitch-reference, adopted also for synthesis of digital mixtures, has been chosen by Lupone according some intervallar-accordal model, distributed on the peculiar octaves for string instruments and selected on their results in musical variety and distintivity dued to perceptive tolerance.

The last sequence shows a more various behaviour in timbre and frequencies. Particularly, timbral palette becomes more articulated, and enriches itself of a kind of sound not far from a quasi-isochronous plucking. The reason of this variety, opposing previous timbral unitarity, rests on expliciting the formal direction of the work, that exalts in his last section the informative degree of timbral component, stopping at the same time the evolution of the others.

2. Compositional behaviour. Role of sound-space and live-electronics

The compositional path of *Corda di metallo* (but extensible to the whole Lupone's opus) is regulated, also in sound-form and structure of the work, by the gravitation of compositional effort around some linguistic attractors [fig. 1] that act the strange attractors of the Theory of Chaos. Such a gravitation in fact doesn't involve a pre-determination of the length and order of this path, but it's able to characterize the subsequent parts of the piece and organize his diacronic structure, as well as the presence of a strange attractor doesn't preordain the motion of a particle, but gives form to its possible paths. Every attractor owns a nucleus that associates it to a compositive dimension, and around every nucleus a confluence-field visualized by overlapping areas. Timbral masses, chords, thematic and rhythmic cells are the four primary musical categories, from which is possible to cross through such a compositional space. They takes place around the areas according to their conceptual adhesion to one of the four dimension at the center of confluence-field (reflexive, punctual, logical or chaotic). At the times of greatest closeness of the compositional path to the centers of corresponding areas, they humble themselves in the informative worth, so letting another category emerge.

The characteristics of mobility and internal complexity of sound, and this organizing the diacronic-formal path through a non-predeterminated gravitation, let a musical poetic of Chaos come out in Lupone's opus. This poetic is both ground and effect in the use of the new electronic and digital technologies in sound syntesis and control, of which Lupone is recognized as a profound expert and innovator. Nuclear and organic complexity of sound, diacronic complexity of formal path and development, have been enriching, since a decade in Lupone's opus, by complexity in treating and governing space as a linguistically meaningful component. Space represents for Lupone not only a new frontier where

new technologies - particularly the live-ones - are improved. It is infact a sound parameter that opposites itself on the whole to all the others, in his operating in a qualitative - and not quantitative - dominion, that is in a sincronic - and not diacronic - dimension too. On the contrary, the other temporal parameters refer to diacronic (and measurable) one, representable with a linear - or better poli-linear - thought, whereas space is capable to represent sound as a hierarchic, associative, therefore qualitative, and to do this representation in a synchronic paradigm, before it disposes itself in the diacronic order of sound-events. Space so assumes an informative and expressive high-value in Lupone opus, and his task in *Corda di metallo* will furnish us an evidence of this. The four general modality in spazializing sound (localized, approaching from stage towards public, removing from public to stage, in chaotic motion above the public) correspond with the four compositional dimension (reflexive, punctual, logical and chaotic) and emphasize their formal centrality.

Spazialization is controlled, in *Corda di metallo*, through Smart system, realized by IRIS, and Kronos program, realized by Carlo Galletti and Felice Cerone (spazialization algorhythm by Marco Giordano) for CRM. Both system and program control an eight-channel sound-equipment, corresponding to the four loudspeakers placed [see fig. 2] near the stage, to two loudspeakers in a mean position and to the remaining two behind the public. These sound-movements are obtained by a live sound-processing suggested and consacrated by psycho- and physico-acoustic researches on spatial hearing: ratio between direct (non-reverberated) and reverberated signal and spectral density (for distance simulation), interaural differences obtained by phase- and intensity-modulation (for lateral localizing), spectral structure (for median-vertical localizing) [3], time-differences in emission and attack-transient (preceding-effect), energies distribution and phase-differential between different sources (motion).

Fig. 3 shows a score-page of control program for spazialization, corresponding to a situation of chaotic motion of sound. Also space, nay overall space, contributes in precising perception of chaotic sound-structures, an imaginific expression of author's individual poetic. In those points where the chaotic dimension becomes the central one, the space rises to principal parameter both in constructive and perceptive side.

Live-spazialization is the showier chapter in the live sound-processing in *Corda di metallo*. Live electronics has been regarding by Lupone, since the starting point of his career, as a necessary and technically achievable of his own technological and compositional efforts. The presence of a live-executor, with the rising to this rank of the sound-technician himself, has in fact completely transformed the status of the electro-acoustic opus. Besides, confering a warmer and more spectacular to the sound-event, the interpretative gesture has given to electronic music some qualities, essential also for composer's role, that is: 1) depth and prismatic-kind of gesture, 2) high feed-back capacity, 3) synchronization of musical thought and technique, 4) expressive coherence of gesture.

Besides spazialization, sound-processing consists, in *Corda di metallo*, in a reverberation and - seemingly less significative, but really important - in a equalization of acoustic sound of the four instruments taken ?recorded? by a micro at bridge (while micros at floor are used for a widespread diffusion of sound) too. Such an equalization, emphasizing unusually high formants for a bow-instrument - its harmonic body exhalts infact low-medium formants - exhalts on the contrary the noise-component typical of the bow-friction on the string, very perceptible in the high frequencies of harmonics, so confering to the instruments a rough and metallic sound-grain, in coherence with that one of the tape.

3. The last section

Rather then following the formal path of the work since the starting phase (in which, after the briefest exposition of the viola, the transformation and development of compositive nuclei begins around the pitch reference of D, a common sound of "a vuoto" strings), I'll describe the final section, previously mentioned as relevant for rising of timbral category end consequent exhalting of "reflexive" dimension.

The entrance of this long section is represented by the sequence of suspensions at b. 156-165, during which the harmonic contents remain almost blocked on a notes-aggregate (F-G-C-D-A-B) that reflects frequential content of two pitch-classes ("d" and "e") employed for the tape sound-synthesis. In spite of the harmonic stasis and temporal suspension, timbre (and intensity in a range pp/mf) of each instrument is varied through a continous shifting of bow position (bribge-normal-fingerboard, with intermediate positions). The collapse of this section happens during b. 163-5, when cello begins to rotate circularly and with increasing velocity the bow in the string, so creating a chaotic timbral vortex, while the other instruments shatter the pitches fixity by mean of a gradually wider vibrato (from a quarter- to an half-tone). A directionned glissando leads the instruments to the new pitch-reference, bichord A/G; the area of supremacy of this bichord is very large (b. 166-184), even if internally speckled by a reticle of many punctual events, referring both to rhythmic and/or intervallar cells, and to sharp dyinamic (sforzati) or timbral prominences in the texture. These contributes generally to point out significative changes in timbral behaviour of the instruments (individually or in the

whole), so causing a "swapp" with timbral dimension, still leader in articulating the texture. The clearest example of such a writing is situated at b. 179-180, when aleatory sequence of sforzati on the last quarter of b.179 inverts the direction of boow shifting, for all instruments, from bridge>fingerb. to fingerb..>bridge. Also b. 173-177 can be considered as a timbral of the bichord A/G: the instruments the "a vuoto" string on the A/G position, yet obtaining a different bichord of harmonics (D/G three octaves higher).

Vertical pitch-reference is going to change, until the end of the work, other three times: D/Eb (b. 185-190), B/G/A (b. 190-195), G/F (b. 195 to the end). At b. 184 - when the reference is still A/G - another sequence of sforzati preannounces the entry of magnetic tape, which extend the timbral palette of string-instruments through a timbral and frequential behaviour more dynamic and readable (as made of much more distinguished and recognizable sound-categories) in respect of previous tape sequences. The evolution of vertical pitch-references is however, in its comparative slowness, also here finalized to emphasize timbral component: in addition to usual variations of bow-manners and -positions, it's exhalted by the employment of different "a vuoto" strings for the realization of the over-mentioned bichords and trichords, each of them so can have an own timbral physiognomy.

The active task, even if hierarchic, of differents parameters in this last section, can be showed also by rhytmus. After the suspensions, rhythmis category tends to remain blocked still later, as the instruments realize generally a persistent mensurate-tremolo (demisemiquaver, then six semiquaver in a querter), perforated by sforzato events. These creates on the whole a net of rhithmic saliences, perceptible as well as it is, but provided of a segnaletic role towards timbral mutations. Furthermore, changes of tremolo fastness, together with connected variations of bow-manners ("martellato" or "detachè"), have a relevant timbral effect. Magnetic tape amplifies this component by some lines of beating-again sounds, amid bow-beated and plucked sounds, furthermore not perfectly regular in repetition, but a little fluctuating and asynchronous.

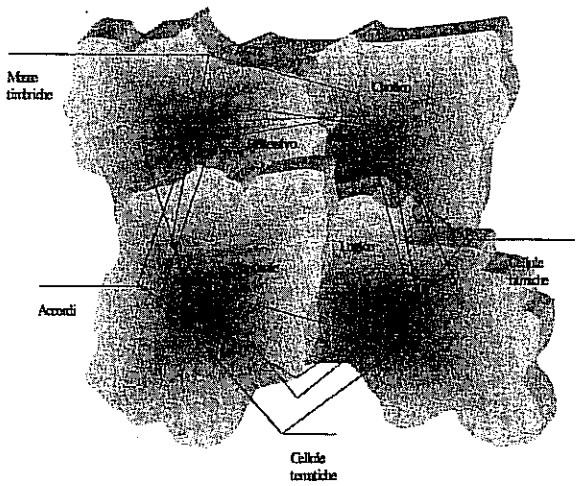


Fig. 1

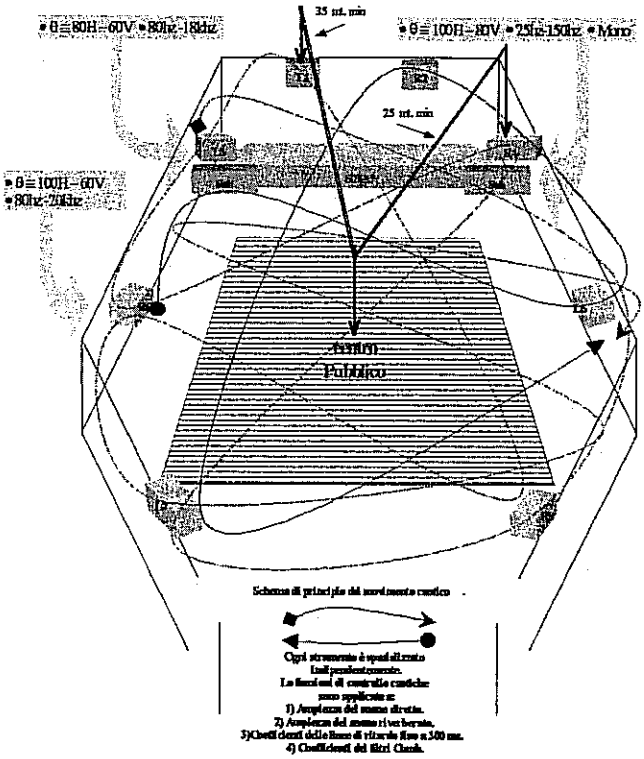
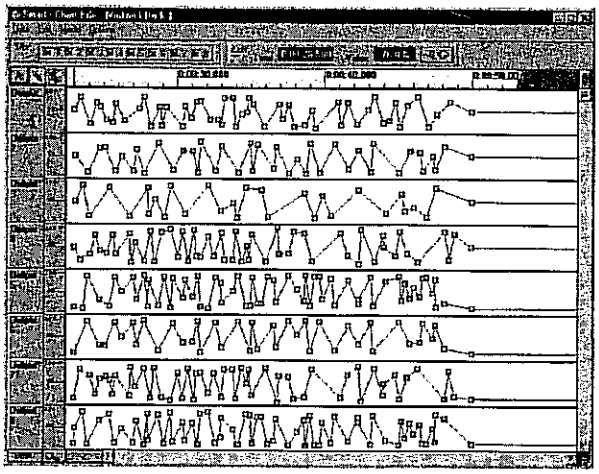


Fig. 2

Fig. 3



[1] M. Seno - L. Palumbi, *Corda di Metallo. Un nuovo modello di simulazione per modelli fisici di strumenti ad arco*, unpublished, 1998, 10 pp..

[2] S. Cappelletto, notes on *Corda di metallo* in the hall-booklet for its wordl premiere, Roma, 29-5-1997, Accademia Filarmonica Romana.

[3] J. Blauert, *Spatial hearing*, Cambridge (Mass.), 1997, The MIT Press.

See, in addition:

M. Lupone, *Corda di metallo*, score and explicative materials, 1997.

Friday 25th

DEMOS

AESTHETIC QUALITY OF STATISTIC AVERAGE MUSIC PERFORMANCE IN DIFFERENT EXPRESSIVE INTENTIONS

Giovanni Umberto Battel
Conservatorio B. Marcello di Venezia
San Marco 2810 - 30124 Venezia
Tel. +39 (41) 5225604 - E-mail gubattel@adria.it

Riccardo Fimbianti
CSC-DEI University of Padova,
Via S. Francesco, 11 – 35131 Padova
phone: +39 -49-8273757 E-mail rf@csc.unipd.it

Abstract

Statistic average of phenomenon or the prototype of a class is often perceived as more representative than a member belonging the same class. It is acknowledged that a musician can play the same piece in different way, all equally musically correct. But all these interpretation converged to the same idea. All different interpretations set in a hypothetical multidimensional space should find the ideal performance at the centre. The average performance seems to be the closer interpretation at the centre.

Starting from this hypothesis we make some recording of the same piece. Our aim was to make performances converge to different musical ideas to prove our hypothesis. By a test we found that average synthesis was judged more attractive then the original ones.

1. Introduction

Statistic average of phenomenon or the prototype of a class is often perceived as more representative than a member belonging to the same class. Already Langlois and Roggman presented a study about some human faces that are obtained aligning and digitalising images of single faces and averaging the grey shade degree for every pixel. In a test, subjects judged average faces more attractive than the original. Such average faces seem to be closer to our patterns of thought than real ones. It seems that we construct a pattern of thought out of all faces we meet and this seems to correspond to an average face.

Musical language, as any other means of communication, remains a quantification of free musical creativeness: Beecham tells that during his first public concert: " ... for some reason the musical sound was strangely different from the idea I had in mind ...". The interpreter tries to interpolate the message written in the score and makes his own idea of the musical piece. His idea is filtered by his personal taste as well as by his musical culture which depends on the historical and musical background he has been training in. Therefore, on the one hand, there is a general idea given by the musical environment, shared by the other musicians, and on the other hand, personal taste.

We could say that the musician achieves his performance starting from the performance praxis and from a general idea shared with all the other musicians. There should be a reference musical idea shared by both the musicians and the audience, connected to the score and the historical period and therefore objective and different from any subjective performance. The passage from idea to music and sound presents some imperfections due to the interaction musician/instrument. The musician cannot always materialise his idea. Moreover, shifting from idea to sound, to some degrees the interpreter moves away from the general shared vision and consciously or not he introduces personal interpretative choices that make the originality of his performance recognisable when we listen to it. As already stated in literature, the more the performance is original, the more it is discussed about and can be really appreciated or totally despised, while a performance that fulfils the audience expectations is appreciated by a wider audience. The *average performance*, artificially made, is closer to the audience expectations because it is not the result of the most subjective choices of an individual musician but the ideal compromise of individual choices that are related, as it is in the different human faces, to an ideal "human-like" average.

Through the *average performance*, subjective deviations are filtered so as to come closer to the ideal performance, which, as we already said, can be considered as objective. The average is, in fact, impersonal as it does not belong to any interpreter.

To make the research easier, we limited the field to some precise musical ideas or expressive intentions. Such demand came also from previous studies on expressive intention showing the necessity to synthesize an expressive intention not connected to a specific performer.

2. Analysis

It is universally known that that there are different ways to play the same piece, all equally good. But every performance, we believe, may converge towards the same musical idea. Different performances arranged in an hypothetical multidimensional space would find in the centre the ideal performance. The *average performance* should be the closest to such ideal one. Starting from this hypothesis, we made some recordings of the same piece trying to make these performances converge towards different and well characterised interpretative ideas.

Five senior piano students at the Conservatorio di Venezia, where asked to play the same piece, Mozart's *Sonata K545*, 2° Tempo *Andante*, in eight different ways trying to correlate them to different interpretative choices we proposed: bright (crystalline), dark (gloomy), hard (strict), soft (tender), heavy (massive), light (gentle) passionate and flat. We chose students as it is reasonable to think that, not having yet developed a personal style, they can play closer to the standard performance. We also avoided a possible school uniformity choosing students tutored by different teachers.

We first made a listening test with some researchers of the C.S.C. of Padua. For each different musical idea we chose the three best performances (among those of the five pianists), that is to say those that better reflected the proposed musical idea. Then the average of the three pieces was made. First we made a temporal normalization so that each piece had the same *tempo*, in this specific case corresponding to the average of the three performances, so that each note would have the same weight in the timing average. We then made the note IOI and relative duration average. The same procedure was adopted for dynamics, first normalizing to the intensity average and subsequently making the average of the intensity of each note. For each value average a synthesized version was made from the average values. Various musicians, the co-author of this paper included, regarded them better then the single recordings of the five pianists. Particularly, for the "passionate" adjective, the average version was definitely judged the best. In this case there is not doubt that the *average performance* is musically correct and at the same time it does not belong to any performer in particular. From these results, it seems reasonable to try to understand which are the rules each pianist applies to musical performance starting precisely from the *average performance* that does not directly belong to any individual.

Moreover, the inevitable little performance imperfections due to the not perfect interaction musician/instrument, are filtered and practically eliminated. Supposing the noise superimposed to the performance to be white noise, i.e. with null average, it is possible, after the average, to analyse the pianist's deviations as significant.

There is nonetheless a difficulty when analysing the performance of a particular pianist. It is sometime difficult to define if a little anticipation/delay is due to the pianist's will or to a sort of mistake in the performance. These doubts should be overcome by studying the *average performance*. For example the downbeat/upbeat pattern is definitely a performance characteristic belonging to the correct performance of the analysed musical piece -

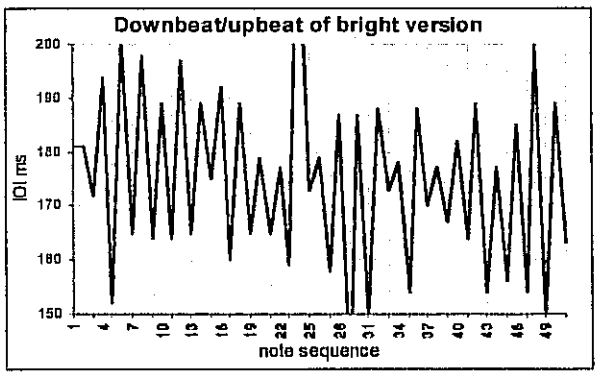


Fig.2.1: downbeat/upbeat of the first four bars of k545 Mozart Sonata. In the ordinate there are IOIs of accompaniment.

The *average performance* is therefore the ideal tool for musical analysis.

To obtain the ideal performance we used the mathematical average on normalized values, as above described. There can be different methods and averages to achieve the same result. Since the importance of the very concept of average in a so highly subjective activity as performance, the authors tried other methods such as the average without normalization and geometrical average, and obtained the same results with a variation of a few ms with respect to the method here described. Data referring to each pianist were quite homogeneous. Of course this is partly because the piece we chose is structurally rather simple and the performers were therefore driven to converge towards a precise musical idea with different expressive intentions. The results obtained in this preliminary phase led us to carry on the study on other recordings. Four pianists of the Conservatorio di Venezia were asked to play Franz Liszt's *La Marquise de Blocqueville, portrait en musique* without any indication. The score had been given them in advance so that they could study it and avoid technical flurries during the performance. Once the recordings were made, Schenker's analysis of the piece was explained to the pianists and we repeated the recordings five days later.

The aim of the experiment was to understand if the musical analysis *suggested* had an influence, in any way, on the pianist's interpretative choices. In this case it is useful to test the soundness of our hypothesis when the piece is structurally more complex and when the performances are not driven by interpretative hints, such as the adjectives proposed in the experiment above described. The analysis is now under study and we shall show the results in the future.

3. Average performance

The listening test on averages demonstrated that there is a good performance independently from the pianist, related to the musical idea fixed on the score. This

allows us to study the close relations between score and musical structure, and also to investigate with new instruments the margin of freedom the score gives to the musician.
In the first case, we found correlations between the legato/staccato degree and harmonic structure, corroborating the results obtained in the expressive intention analysis the authors of this paper previously made [3]. There is a significant correlation between the DRO (Duration Offset as ratio IOI/DR) of each note and the harmonic tension obtained with the Lerdhal method. This confirms the correlation between performance parameters and harmonic structure. To find out the margin of freedom, we found the maximum difference (IOI max – IOI min) for each note within the average expressive intentions, normalised with the method described above to the same temporal length. Then we analysed the score with different methods used in literature so to find the accents or the rhythmic/melodic content of each note.
The notes in the score do not possess all the same structural importance. According to our hypothesis, in conformity with some previous studies [5], the performer modifies the notes in relation to their structural importance.
The difference between the maximum and the minimum, which should represent the performer's margin of freedom, lies in good correlation with some used methods [4, 6, 10]. Tab. 1 shows the methods applied with the relative correlation to the melody notes (113 events) and the coefficient of significance.

	LBDM	D&P	H tens	Att mel
r	0.4086	0.2252	0.2728	-0.2178
P <	0.0001	0.013	0.003	0.016

Tab.1: Correlation table. In the first column there are correlation coefficient (r) and significant coefficient (p). In the first row there are LBDM [4], Drake and Palmer accents[6], harmonic tension method and melodic attraction method [10].

The hypothesis could be the followings: the bigger the harmonic charge on a note, the more the pianist can modify it to characterise an expressive intention from the other; or, the pianist uses mainly notes with a high tension charge to convey expression.
Both the interpretation and the margin of freedom can be connected to the musical structure of the piece and so, according to it, it would make sense to orient the research also in this field. In fig. 3.2 we can notice that the difference between maximum and minimum is not at random but it follows the musical structure.

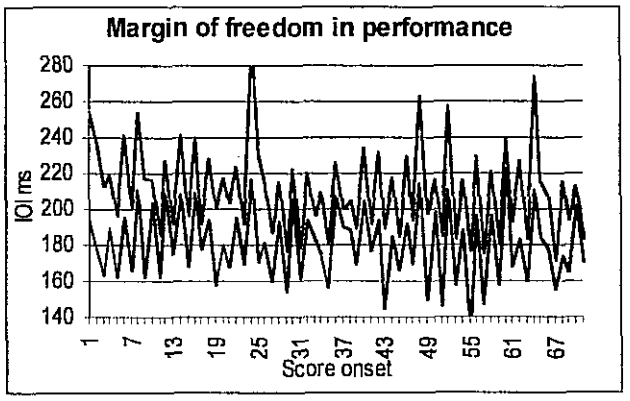


Fig. 3.1: Difference between max and min accompaniment IOI in different average excerpts.

Conclusions

The present work represents a first attempt to assess the aesthetic and musical qualities of the quantitative average in musical performance. The results allow us to confirm the hypothesis that the *average performance* is a good performance and it represents a prototype or an aesthetic idea which is closer to the musical idea the listeners recognise to be the one written in the score. Listening tests, in fact, prove that the *average performance* is appreciated more than the individual pianists' ones.
Moreover, the *average performance*, not being dependent from a particular musician, is a good example for eventual future musical analysis to understand, for instance, the musician's margin of freedom or the correlation between musical structure and performance.
This study is far from reaching a conclusion, and it leads to further hypothesis and researches.
The analysis of the results permitted to correlate statistical analysis to musical performance. No doubt the *average performances* presented are the result of numerical analysis and at the same time they represent correct musical performances. The present work may narrow the gap between two far away worlds such as musical interpretation and numerical analysis. The latter,
consequence of new technologies and research methodologies, can be a useful tool for performers and for musical analysis purposes.

References

[1] Battel G.U., Bresin R. (1993). Analysis by synthesis in piano performance: a study on the theme of the Brahms' Variations on a theme of Paganini op. 35. Proceedings of the Stockholm music acoustic conference. Royal Swedish Academy of Music n°79, Stockholm, pp. 69-73

[2] Battel G. U. (1995). Un progetto per l'analisi dell'esecuzione pianistica. Rivista Italiana di Musicologia XXX n. 2 pp.419-452

[3] Battel G. U., Fimbianti R. (1997). Analisys of expressive intentions in piano performances. Proceeding of AIMI International Workshop on Kansei - The technology of Emotion. pp. 128-133.

[4] Cambouropoulus E. (1996), Musical rhythm: inferring accentuation and metrical structure from grouping structure. Proceedings of JIC96-Brugge, 15-22.

[5] Canazza S., De Poli G., Roda' A., and Vidolin A. 1997. Analysis and synthesis of expressive intentions in musical performance. Proc. ICMC '97, Thessaloniki

[6] Clarke E. (1993), "Imitating and evaluating real and transformed musical performances". Music Peception, 10, n°3, 317-341.

[7] Eco U. (1990): "I limiti dell'interpretazione", Bompiani, Milano 1990.

[8] Friberg A. 1991. Generative rules for musical performance: a formal description of a rule system. Computer Music Journal, 15(2): 56-71.

[9] Gabrielson A.: "Music Performance". The psychology of music, D. Deutch (Ed.).

[10] Lerdahl F. 1996. Calculating Tonal Tension. Music Perception, 13(3): 319-363

[11] Repp B. H. (1995), Acoustic, perception, and production of legato articolation on the piano. Journal of Acoustical Society of America, 97, 3862-3874.

[12] Repp B. H. (1997), The Aesthetic Quality of a Quantitatively Average Music Performance: Two Preliminary Experiments. Music Perception, summer 1997, Vol 14, No. 4, 419-444.

THE STUDIO ÉLECTRO-ACOUSTIQUE OF THE ACADÉMIE DE FRANCE À ROME: A STUDIO REPORT

Nicola Bernardini
Studio Eléctro-Acoustique
ACADÉMIE DE FRANCE À ROME
Viale Trinità dei Monti, 1
00187 Roma
e-mail: nicb@axnet.it

Thierry Coduys
IRCAM
1, place Igor Stravinsky
75004 Paris
e-mail: thierry.coduys@ircam.fr

Abstract

Mainly an analogue electroacoustic studio with equipment going back to the 70's, the STUDIO ÉLECTRO-ACOUSTIQUE DE L'ACADÉMIE DE FRANCE À ROME was recently overhauled and fitted with professional digital equipment, thus becoming a valuable and uncommon resource in the gorgeous environment of the Villa Medici in Rome. The upgrade coincided with the desire of the direction of the ACADÉMIE DE FRANCE À ROME to open the studio to external production work, thus making the resource available to contemporary music productions.

The studio sports some classical digital production and post-production equipment, such as a Pro-Tools 888 workstation along with a quadraphonic listening environment, and fairly new and unusual equipment as the french digital matrix/bus Muxi-paire which could prove to be an extremely interesting general application hardware in electro-acoustic works.

This report describes the studio and discusses the availability of these resources.

1 A bit of history

The ACADÉMIE DE FRANCE is one of most prominent centers of cultural activity of Rome. Created in 1666 by desire of King Louis XIV (who was trying to educate the french artists at the italian school of painters and sculptors) to host the recipients of the *Prix de Rome*[1] (a three-year long scholarship in Italy) it was initially located in a small apartment near the Vatican. It was then moved to Palazzo Salviati on the Corso (built by the Duc of Nevers) and in 1803, Napoléon exchanged the Palazzo Salviati with the final location of the Académie, the Villa Medici over Trinità dei Monti (just above the spanish steps), built in 1544 by Annibale Lippi and Bartolomeo Ammannati. Villa Medici is indeed one of the architectural masterpieces of Rome, with his monumental building abundantly decorated by works of art of all times and a magnificent park which extends down to the antique Mura Aureliane and to Villa Borghese (the biggest park in Rome). Many of the towers in the Mura Aureliane and the small buildings surrounding the Villa have been adapted to host artists of all disciplines (now ranging from classic visual arts to *gourmet cuisine*, from film-making to composition to architecture, etc.). Nowadays, the *Prix de Rome* does not exist anymore under the same name, but neverthless these artists (not necessarily french) are all invited by the french government to spend varying periods of time at the Académie where they get the chance of working in this fertile environment. All artists are encouraged to submit projects and activities to the direction and, where/when possible, these get funded in efficient and professional ways: here's where and how the history of the STUDIO ÉLECTRO-ACOUSTIQUE DE L'ACADÉMIE DE FRANCE À ROME begins.

Memories of an electro-acoustic studio at the ACADÉMIE DE FRANCE À ROME date back to the beginning of the '70s[2] under the direction of Jean Mathieu. Considering the characteristics and the functions of the ACADÉMIE DE FRANCE À ROME, a fairly fast turnaround of people and projects has always been the usual practice. The building of an electro-acoustic studio however has established a fairly solid permanent activity which, though having physiological ups and downs, has evolved continously over time.

In the beginning the studio was located in one of the *ateliers* (n.13) of the *pensionnaires*. Originally, the studio was fitted with analogue tape recorders and oscillators. At the end of the '70s, under the supervision of composer Marc Monnet the studio was moved to its actual location in two remodeled garages along the Viale Trinità dei Monti. It established a solid connection with the **Centre de Recherches et de Formation Musicales de Wallonie** which provided the studio with quadraphonic listening, computer-controlled analogue ring-modulators, oscillators and a connection matrix. Another change happened at the end of 1992 with composer Philippe Mion who involved Thierry Coduys in the re-design of the studio with digital equipment. The re-design was completed at the end of 1997 and it became clear that such a precious resource was more and more needing a stable maintenance and development. Nicola Bernardini was appointed to this end by the direction of the Académie at the beginning of 1998. Under his supervision, a complete internal re-modeling of the studio was performed and the

1. the so-called *pensionnaires de l'Académie*

2. As surprising as it can be (especially because the Académie keeps historical archives of all activities back to the times of Napoléon and beyond), it has been fairly difficult to trace the existence of the studio at the beginning

last technological details were set into place.

2 Current status of the studio

The current disposition of the studio is as follows:

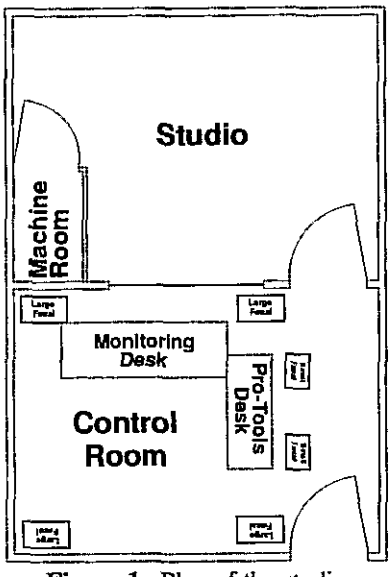


Figure 1. Plan of the studio

Among its equipment, the studio lists:

- a computer-controlled fully digital 56x48 french matrix *Muxi-Paire*
- an 8 input/8 output Pro-Tools TDM system run by a PowerMac 8100/100 which features several SCSI disks (up to 6.5 GBytes of memory), a 1 GByte magneto-optical removable disk and a CD mastering unit
- an Eventide DSP4000 processor
- several MIDI devices (a Yamaha SY99, three Yamaha TX802, a Yamaha TX816, a Yamaha KX88, a Proteus, an Akai S1000)
- a Studio 5 MIDI matrix interface
- a quadraphonic listening environment (large *Focal* loudspeakers)
- a close listening environment (small active *Focal* loudspeakers)
- a 24x12 TAC Scorpion monitoring console
- two DAT machines
- a professional microphone pool (some Neumann U87s, AKG 451s, Sennheiser 441s, etc.)

The characteristics of this studio make it particularly well suited for post-production work and tape pieces. The digital matrix software allows the studio to be always fully connected: the users can create their own patches of the whole studio, save them, and recall them at wish at a later time without touching a single cable. This is very convenient since there is a very quick weekly turnaround of projects here (every user is allotted ca. 48 hours at a time). Another advantage is that every signal entered in the matrix can be completely processed in the digital domain without any other passage into the analog one. Furthermore, since the matrix is splitted in two racks between input and output stages in some special cases/occasions (like concerts etc.) the input stage can be moved up to the production location while the output can continue to stay in the studio (the two stages are connected together by two 'thin wire' cables).

The software running on the Mac features the usual professional suites like Digidesign's *Pro-Tools*, *Finale*, *Studio-Vision*, *MAX-MSP*, *Soundhack*, etc. and more experimental and sophisticated tools like csound and the forum-IRCAM suite (i.e. *Audiosculpt*, *Patchwork*, *Modalys*, etc.). An installation of Linux for PowerMac as alternative operating system is on the way.

All noisy machines (including the computer's main body) are located in a sound-proof closet to keep the control room perfectly quiet.

Here's how the studio works:

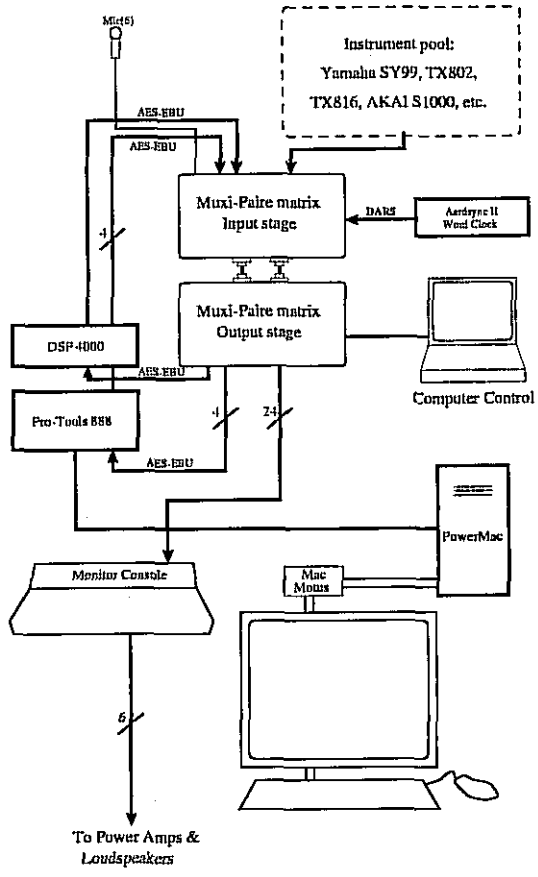


Figure 2. Connection schema of the studio

3 Access policies

Up until now, access to the studio has been granted only to the composers in residence at the ACADÉMIE DE FRANCE. In consideration of the quality and of the value this resource has acquired the current director of the Académie, M.Bruno Racine, is now thinking to open up the studio and has asked the *pensionnaires* to establish a plan for inviting selected projects by young promising composers to be produced in the STUDIO ÉLECTRO-ACOUSTIQUE DE L'ACADÉMIE DE FRANCE À ROME. At the time of this writing this project is in the works.

Acknowledgments

The making of a studio is indeed a combination of many people's efforts, goodwill and especially political and cultural (long-sighted) vision. Furthermore, the current STUDIO ÉLECTRO-

ACOUSTIQUE DE L'ACADÉMIE DE FRANCE À ROME is the collection of all these efforts over a time period that spans well over twenty five years. Therefore, it would be very hard to acknowledge all who have contributed to the studio: at the risk of being unfair, we will refrain from mistakes quoting scarcely documented activities and we will stick to the last year of activity which has seen a complete change in the studio's technology and in its internal modeling.

All these changes are due to the direction of the ACADÉMIE DE FRANCE — Bruno Racine, Director, Gérard Fontaine, *Secretaire Général* and Michel Ferdinand, *Intendant*: without their precise intervention the studio would still be a collection of analogue tape recorders with an old quadraphonic listening environment. The current *pensionnaires* composers (Jean-Louis Agobet, Daniel D'Adamo, Thierry Machuel, Yi Xu) ought to be thanked for providing the direction of the Académie with convicing cultural arguments to support the studio, and to endure with patience all the difficulties the recent changes have implied (as in all changing enviroments). Many thanks are due to Sandro Guarneri, responsible for the works at the Académie, for providing efficient solutions to the many problems posed by a working studio and to Mario Tornesi and Maria-Teresa De Bellis for providing much of the historical background for both the ACADÉMIE DE FRANCE and the studio that is included in this paper.

PASSACAGLIA, BY ALDO CLEMENTI: WRITING DISPOSABLE ALGORITHMIC COMPOSITION PROGRAMS

Nicola Bernardini
Conservatorio "C.Pollini"
35100 Padova, Italy
e-mail: nicb@axnet.it

Alvise Vidolin
Conservatorio "B.Marcello"
Venezia, Italy
e-mail: vidolin@dei.unipd.it

Abstract

The realization of the tape part of Passacaglia, *by Aldo Clementi, is taken as an example to illustrate the following approach: write few simple programs in powerful text filtering languages (such as* `awk`, `perl` *and* `tcl`*) to the composition requirements of a specific piece. Besides the speed of development of such programs and their fast prototyping and debugging, their size and the effort required in writing them makes their inherent disposable characteristics an advantage rather than a drawback.*

This approach was preferred over defining and/or adopting some general compositional tool. Usually, these tools require very extended programming efforts and are hardly general enough for musical composition, a task which encompasses a tremendous variety of actions over extremely diversified data (pitches, tempo, articulation, rhythms, dynamics, position, fragments, etc.).

This paper covers the technical details of such an approach, along with an overview of its advantages and of the problems encountered.

1 Introduction

Computer-assisted composition is indeed one of the first (if not the very first) applications of computing machinery to music. Since the pioneering times of Hiller and Isaacson some 45 years ago (cf.[7]), the similarities between compositional thinking and applied logic have often struck the mind of engineers and composers alike. Therefore, it is no wonder that many efforts towards compositional programs and towards what is termed "Computer-assisted composition" (CAC) have been and are currently spent (cf.[4, 5, 9, 10]). Many of these programs are fully integrated suites that assist composers in common compositional problems and provide results in a number of output formats (graphical, sonic textual, etc.).

There are several problems encountered with such programs. They can be summarized under the following categories:

1. they very often imply a specific vision and/or language of musical composition; it is very hard, indeed, to generalize a 'mainstream fashion' of composition — but since programming must start from some deterministic point of view, CAC programs tend to imply a specific musical language;
2. the integration of graphics, sound and text tends to make these programs inherently heavy and non-portable from platform to platform;
3. the specific file formats they use for holding their data is not very portable too (unless of course, the data is not specific as in the case of common sound and graphics file formats.

These categories of problems produce the following patterns of behaviour among composers:

a. they tend to pick up a specific program which is 'as close as possible' to her/his necessities rather than practicing the art of building specifications that suit in the fullest sense their compositional needs;
b. the choice of a computer platform leads to a number of other choices which are apparently unrelated to it (studio, home logistics, etc.) — indeed, the computer is being viewed more as a musical instrument with its own limitations and idiosyncrasies rather than a general tool;
c. communication among composers and developers tend to be restricted to the user's pool of specific applications rather than being freely interchangeable.

Of course, not all CAC programs suffer of all these problems. However, CAC programs in which none of these problems exists are hard to find.

2 A different approach

Instead of adopting such programs, another approach can be picked up in CAC applications. The basic principles of this approach are:

1. stick as closely as possible to the compositional ideas and to the way of thinking of the composer
2. write small *ad hoc* programs using small language interpreters to
3. transform the compositional ideas in whatever other languages are needed to the purpose

This approach works particularly well adopting pure ASCII text as a message passing standard through different applications and the pipe mechanism which is present in most advanced operating systems (actually, pipes are a fast and easy substitutions for temporary files — which can instead be used in diminished environments). Thus, the programs mentioned in are, effectively, text filtering programs: they accept text in some arbitrary format and can transform/produce text in some other arbitrary format. There are several such programs available in the public domain and functional on all most diffused platforms (a lot of documentation is available on the web, but for a quick reference on the most widely used "small" languages cf.[1] for `awk`,[13] for `perl` and[8, 14] for `tcl/tk` — many others are available).

The ASCII standard allows easy porting the data from one platform to another and many specialized applications (both for graphics and

sound) accept text as input, thus completing the production circle. Reusability

Furthermore, *ad hoc* writing does not imply that written code cannot be reused: it simply means that reusability is not the first and most prominent topic in the programmer's agenda.

3 A Case Study

To demonstrate the validity of this approach, we will illustrate the elaboration of the tape part of *Passacaglia*, a piece for flute and tape by italian composer Aldo Clementi. The tape part is composed out of the combination of twelve flute fragments

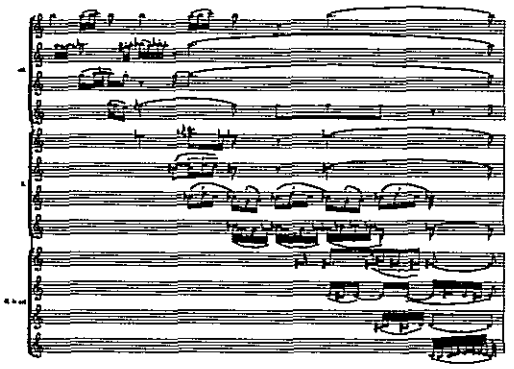


Figure 1. The twelve original flute fragments

which get transposed/permutated circularly for twelve transpositions in this way:

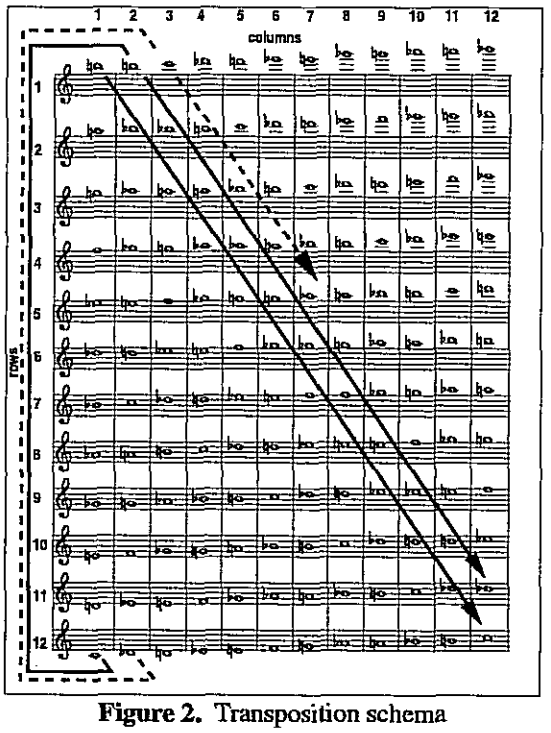


Figure 2. Transposition schema

The central note of each fragment corresponds to the notes in the first column. The fragments get transposed diagonally according following the arrows. The metronomic tempo changes by the same ratio of the transposition interval (this may reveal a 'tape-recorder' oriented mind behind this composition).

The complex compound of the fragments and their permutations sounding together compose a basic block of the composition. As in serial-oriented compositions, there are four basic blocks which include the original one (mentioned above), its inversion, its retrograde and its retrograde inversion. The transformations involve both the fragments in themselves (although they are not literal transformations) and the transposition schemas (which are instead literal ones). Differently from other compositions, the verticality of superpositions of fragments is maintained according to some disposition rules. To make a long story short, here's how the four compound forms can be built together[1]:

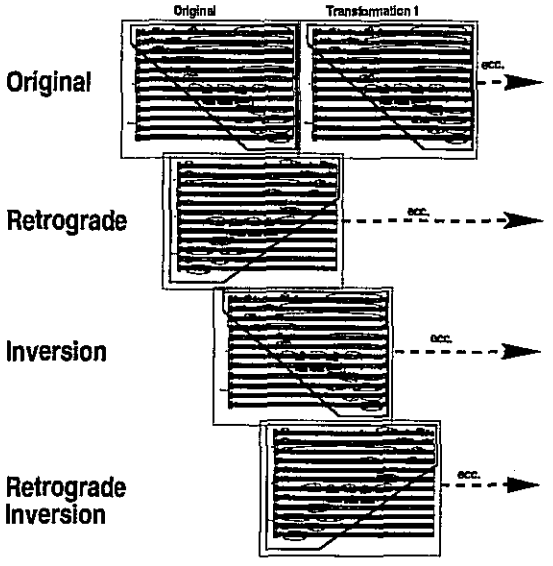


Figure 3. *Passacaglia*: formal construction

Furthermore, the amplitude (and as we will see, space location) of each fragment follow some other specific rule to perceptually alternate foreground/background planes.

Even though the compositional idea is fairly simple to begin with, several practical problems arise:

1. since the fragments had to sound real, we decided that we would have recorded them and we chose to use `csound`[3, 12] to play the fragments in the correct order
2. the total number of fragments is $12 \times 12 \times 4 = 576$: a bit too much to ask to a flute player and a bit too much for the time we had available in the recording studio: we decided that we would record only a fraction of them (180, which were already a bit much for the flute player and for us[2]) and that we would have had `csound` do the necessary interpolations for all the fragments in between (this was helped by the tight tempo/pitch relation

1. Actually, Clementi proposed only one of the possible construction schemes and implied that anybody could build a version of the composition using a different scheme

2. that is, for each transformation: the 12 fragments of the first column plus the 11 fragments of the first lines of each instrument (ott., fl. and alto fl.), thus: $[12 + (11 \times 3)] \times 4 = 180$

required by the composer)

3. though exactly played at the correct tempo by the flute player, each fragment had a specific offset to adjust to its position in the compound, and that offset would change along with the transposition; furthermore, considering the composition not all fragments had a sync point in the same position, so we would need to sync them to each other according to their own characteristics;
4. the permutations of amplitudes, spectrum filtering, positioning etc. were to be calculated during the construction of the tape parts; thus they were to be incorporated in the calculations to produce the `csound` score file
5. after some failed attempts, we realized that if we were to use the `loscil` opcode in `csound` (an opcode which was used to read samples from files into core memory) we needed to do some extra effort in cleaning up the core memory by unloading sample tables when they were not used anymore, otherwise `csound` would just blow up out of exhausted memory after a while (currently, with the recently added `diskin` opcode this is not a requirement anymore — but it was at the time we did the piece...)

Thus, we decided to try and write a database of the recorded fragments which would hold the following data:

- the file name
- the duration
- the offset time where the sync was to be found
- the offset in beats wher the sync was to be found
- the reference metronome at which the fragment was played
- the reference pitch class which served as central note for that particular fragment
- the instrument with which the fragment was played

As an example, a typical record of the database would look like:

```
D-B1-B1.aiff|4.34|0.0635|6|110|9.11|ott
```

with pipe (`|`) separated data fields. Then we wrote the software to perform all the necessary operations to write a `csound` score out of single pipe-separated record lines like this:

```
0.0573|A|full|-2.2|0
```

where the field

1. represents an offset in time positioning of the absolute action time of each compound (if the number is preceded by a + or a −, the number is taken to be an offset from the end of the last compound — that is, the previous one of these lines)
2. indicates the fragment set to be picked up (`A` = original, `B` = inversion, `C` = retrograde, `D` = retrograde inversion)
3. can be a specific column or a full scheme when this field is specified "full". In this case,
4. is used to set the delay between one column and the next

The software is composed of 570-line `awk` script which performs the following operations:

a. parse the input record lines
b. for every compound, and then for every fragment inside each compound:

- finds the original fragment to be played
- selects and performs the appropriate permutation, transposition or retrogradation
- calculates the appropriate transposition factor,
- the relative amplitude,
- the appropriate duration of the fragment
- and the relative inter-fragment distance

This first pass produces, for every input line, the 144 lines of `csound` score plus the needed lines for reverberation, function table selection and garbage collection. A second 57-line `awk` script pipelined to the first pass performs the function table sorting, compacting and cleanup. In the first version of the piece, each data file produces a stereo track to be recorded on an eight-track tape. Four of these tracks get set up together and then projected into space with a real-time spatializer. A second version is in the works which should be running out of a single stereo tape and spatialized with 3-d techniques explained in another paper in these proceedings[2].

Judging from the revision tags, the writing of the first pass required three two-days sessions while the second pass about one hour of work. The performance is acceptable: running both passes in a pipeline takes approximately 32 seconds for each line on an old Pentium 75 with 14 MBytes of RAM and a fairly recent version of the Linux operating system. Machine time is evenly spread between application time and system time, and it grows linearly with the number of input lines.

4 Conclusions

Admittedly, the example at hand is a fairly simple one. Also, it certainly takes the resources of a typeless language like `awk` to its upper limit: the absence of data structures can be faked in awk but it makes the programmer's life fairly complicated beyond simple problems. With the example at hand, however, we wanted to show that:

a. when data structures are simple and the composition is made out of sheer multiplication of elements (which is sometimes the case in computer-controlled compositions), small typeless languages like `awk` fare perfectly well and allow quick and precise definition of compositional algorithms (whose logic complexity is only a matter of taste);
b. furthermore, other small typeless languages (like `perl`) sport a wide palette of system calls to perform over a network or deal with otherwise inaccessible system resources when needed;
c. and finally, still other small (almost) typeless languages like `tcl` allow the definition of graphic interfaces, for those of us who cannot do without (there's even an addition to tcl which performs MIDI, for those of us who can't do without...)
d. if more complex data structures are required (as in some kind of expert systems), some bigger interpreter languages like `prolog` or `lisp` can be used; care should be taken, however, to use the language that exactly fits the problem at hand, because using the wrong tool and

language can result, in the best of cases, in a big waste of time — sometimes, writing a small `lex`—`yacc` `C` language filter program can be quicker and better (from the development point of view) than hacking up some twisted code in some interpreter language

e. a complex task should be subdivided in small self-contained filter programs which pipeline an evolving ASCII text down to a final code which should be used by the specialized application;

f. if the quantity of ASCII text poses a problem (as in the case of sounds or spectrum analysis), inline compression/decompression (which does a great job on ASCII characters) can be interspersed into the pipeline

5 Software References

All of the languages presented in this paper have at least one version available in the public domain (the main being `ftp://prep.ai.mit.edu/pub/gnu`). `tcl` and its suite of accessories can be found at `http://www.tcl-consortium.org`. Public domain `lisps` come in many flavours[3] Many public-domain versions of `prolog` are available too[4].

Acknowledgements

First of all, many thanks are due to Aldo Clementi without whose work the *Passacaglia* would not exist. Roberto Fabbriciani, the first performer of the piece and the unique companion of a 10-hour long marathon to record the 180 fragments is to be thanked for its patience and its supportive and friendly attitude (on top of being among the top-class contemporary music players in the world). The first version of the tape part of the *Passacaglia* was built at the **Centro Tempo Reale** which has been indeed the unique place, from 1994 to 1997, were this work could develop.

References

[1] Aho, Alfred V., Kernighan, Brian W., and Weinberger, Peter J., *The AWK Programming Language,* Addison-Wesley Publishing Company, Reading, MA. (1988).

[2] Bernardini, Nicola and Vidolin, Alvise, "Recording "Orfeo Cantando... Tolse" by Adriano Guarnieri: Sound Motion and Space Parameters on a Stereo CD" in *Proceedings of the XII Colloquio di Informatica Musicale - Gorizia 1998,* Gorizia (September 24-26, 1998). in press.

[3] Boulanger, Richard (ed.), *The Csound book: Tutorials in Software Synthesis and Sound Design,* MIT Press, Cambridge, MA (1998). in press.

[4] Buxton, William, *Design Issues in the Foundations of a Computer-Based Tool for Music Composition,* Computer Systems Research Group, Toronto (1978).

[5] Cope, David, "Computer Modeling of Musical Intelligence in EMI", *Computer Music Journal,* 16, 2, pp. 62-83 (1992).

[6] Dannenberg, Roger B., "Machine Tongues XIX: Nyquist, a Language for Composition and Sound Synthesis", *Computer Music Journal,* 21, 3, pp. 50-60, MIT Press Journals, Cambridge, MA (Fall 1997).

[7] Hiller, L.A. Jr and Isaacson, L.M., "Musical Composition with a High-Speed Digital Computer", *Journal of the Audio Engineering Society,* 6, 2, pp. 154-160 (July 1958).

[8] Osterhout, John K., *Tcl and the Tk Toolkit,* Addison-Wesley Professional Computing Series, Addison-Wesley, Reading, MA (1994). ISBN 0-201-63337-X.

[9] Rodet, Xavier and Cointe, Pierre, "FORMES: Composition and Scheduling of Processes", *Computer Music Journal,* 8, 3, pp. 32-50, MIT Press Journals, Cambridge (1984).

[10] Rowe, Robert, *Interactive Music Systems: Machine Listening and Composing,* MIT Press, Cambridge (1992).

[11] Schottstaedt, William, "Machine Tongues XVII: CLM: Music V Meets Common Lisp", *Computer Music Journal,* 18, 2, pp. 30-37, MIT Press Journals, Cambridge, MA (Spring 1994).

[12] Vercoe, Barry, *CSOUND: A Manual for the Audio Processing System and Supporting Programs,* MIT Media Lab, Cambridge, MA (1986). Program Documentation.

[13] Wall, Larry and Schwartz, Randal, *Programming Perl,* O'Reilly and Associates, Inc. (1991). ISBN 0-937175-64-1.

[14] Welch, Brent, *Practical Programming in Tcl and Tk,* Prentice Hall (1995). ISBN 0-13-182007-9.

---

3. worthwile documenting here are `scheme` (available from `ftp://ftp.cs.cmu.edu/user/ai/lang/scheme/`), `common lisp` (available from `ftp://ma2s2.mathematik.uni-karlsruhe.de/pub/lisp/clisp/`), and in particular Bill Schottstaedt's `common music`[11] (available from `ftp://ccrma-ftp.stanford.edu/pub/Lisp/`) and Roger Dannenberg's `nyquist`[6] (available from `http://www.cs.cmu.edu/afs/cs/project/music/web/nyquist/`) which are music and synthesis oriented variants which combine the algorithmic power of `lisp` with sound processing capabilities.

4. a fairly complete free implementation (`SWI-Prolog`) can be found at `ftp://swi.psy.uva.nl/pub/SWI-Prolog/`.

RECORDING *ORFEO CANTANDO... TOLSE* BY ADRIANO GUARNIERI: SOUND MOTION AND SPACE PARAMETERS ON A STEREO CD

Nicola Bernardini
Conservatorio "C.Pollini"
35100 Padova, Italy
e-mail: nicb@axnet.it

Alvise Vidolin
Conservatorio "B.Marcello"
Venezia, Italy
e-mail: vidolin@dei.unipd.it

Abstract

*This presentation describes the realization of a transposition for CD of a work (*Orfeo cantando... tolse *by Adriano Guarnieri) which calls in its score for sound spatialization and motion, giving a detailed account of the encountered problems and their practical solutions.*

The simulation model is based on the design of a virtual listening environment which depends on heuristic considerations of typical domestic listening rooms (size, wall absorption, speaker placement). As in general spatial modeling approaches, the design of the virtual positioning of instruments and their paths through space produce all the necessary data for early reflection delays, dry/wet reverberation balances and amplitude/filtering levels.

1 Introduction

Spatialization and motion simulation of sounds are currently considered "hot" topics in ongoing music research, and a large literature has been produced on the many aspects that these techniques imply (only a brief excerpt of it can be referenced here:[1, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13]).

However, scientific and fully controlled practical applications of these techniques in published CDs have hardly been documented to date, perhaps in consideration of the fact that such applications would have to overcome a number of practical difficulties which can be categorized as follows:

1. financial/contingent difficulties
2. performance difficulties
3. simulation difficulties

When these difficulties sum up, many compromises have to be put up with and the results can hardly have any scientific value to be worth documenting. Thus, rather than a presentation of brand new techniques the main focus of this paper is centered around the solutions we adopted to the above mentioned difficulties (as we shall see, a privileged environment did not make difficulties disappear: it simply gave us the possibility of finding interesting solutions).

2 *Orfeo cantando... tolse*

Orfeo cantando... tolse, composed by Adriano Guarnieri in 1994, is a 30 minutes long work for two sopranos, a small female choir (6 voices) two electric guitars and small instrumental ensemble.

The score calls for sound motion in space for the two sopranos (two separate paths), the female choir (two separate paths), the two guitars (two separate paths). In some sections of the piece, a flute solo and the double-bass too get projected in a three-dimensional space which surrounds the public.

In concert, the sound motion was performed using *MiniTrails*, an 8x8 spatializer matrix built at the **Centro Tempo Reale** and capable of moving continously 8 sound sources through 8 loudspeaker groups distributed in the hall (for details on the *Trails* and *MiniTrails* projects, cf.[2]).

MiniTrails allows to run a stream of well-defined sound motion scores so that the live implementation of the spatialization cues in the score was pretty straightforward.

Furthermore, after the first performance of the piece it was obvious that the motions the score required were not additional options: they were essential to the intelligibility of the subtle and thick contrapuntal textures written by Guarnieri.

3 The recording project

When Ricordi came up with a recording project for *Orfeo cantando... tolse* in 1996, it was clear to Guarnieri and us that we would have to find a way to convey the sense of the spatialization cues written in that score. Furthermore, the problems that we were facing were:

1. motions would be recorded on a stereo CD: we needed to transform the multiple-output *MiniTrails* scores to a stereo output while mantaining the 3-D motion
2. we could not count on precise speaker locations for reproduction, since the CD would be played in many different situations (as many as there are domestic listening environments)

On the other hand, given the nature of recorded material we could very well work off-line (in non-real-time) with any specialized software we desired.

Of course, the usual two-microphones DAT recording configuration that is so often used in contemporary music recording projects would not allow us to perform any movement separation here, so we required the recording to be done on a multi-track digital machine (actually three stacked DA-88 tape recorders were used). The recordings were to be done with the maximum possible feed-through separation between tracks (this was actually one of the biggest problems: on one hand, bad instrument separation would smear movement precision and on the other, orchestra

players and singers are not very happy playing very difficult music without hearing each other directely; as for the other problems, we had to select an in-between compromise which turned out to be less-than-satisfactory, at least for us).

This was already a big financial effort for our producers, and in order to convince them to pursue this goal we proposed to pursue a fully experimental path in collaboration with the **Centro Tempo Reale** — basically, we were allowed to experiment for quite a long time (about a year) and they would get the results for free. The production was complicated by many other factors so time was not an issue... and we got the deal.

4 Basic Motion Patterns

Orfeo cantando... tolse calls for the following spatialization patterns:

circular	soprano I
	soprano II
	choir I
	choir II
front-back	guitar I
	guitar II
random	double-bass
	flute

Furthermore:

1. the rest of the orchestra has no spatialization features: it is to be aurally placed in its conventional frontal position;
2. the motion parameters of all movements change from section to section;
3. the above mentioned instruments and voices have some static sections in which they do not move
4. no motion implies sound elevation (the vertical plane — what would be some sort of z dimension — is not taken into consideration)

5 Spatialization algorithm

The basic idea was, as usual, to rebuild the delay patterns of the virtual walls of an hypothetical concert hall for each instrument/voice modifying each delay time, amplitude and reverberation mix as it moved through space. Stepping from the basic idea down to practice, however, we had to face several practical considerations:

a. we had to pick up a standard listening setup as a model; this setup would be our 'typical domestic environment' and all delay and amplitude computations would necessarily subtract the contributions of this setup;
b. experimenting with the standard setup and changing listening positions, rooms, etc. we found out that very detailed simulations did not fare very well with different listening environments; as a matter of fact, a middle point was to be found between motion perception and listening versatility
c. the implementation would need to be able to deal with changing patterns with a scoring similar to the one used for *MiniTrails*; for every monophonic source we would rebuild a stereophonic output which would retain all the virtual space information

After some experimenting, we defined the following solutions:

a. the standard listening setup would be configured as follows:

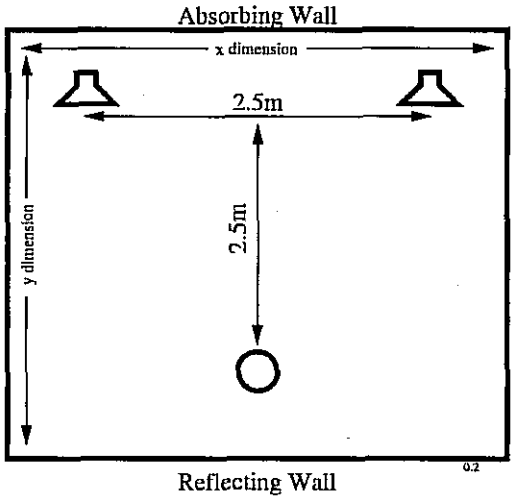


Figure 1. Standard Listening Setup

No other distances (such as loudspeaker-to-front wall, listener-to-back wall, etc.) were taken into consideration;

b. the only computed variables would be:
   - first order delay times for the four walls, along with their amplitudes (no floor, no ceiling)
   - amplitude of the source
   - amplitude of the send auxiliary to the reverberator

   the amplitude were calculated with the usual $\frac{1}{d}$ (d = distance) for the direct signals and $\frac{1}{\sqrt{d}}$ for the reveberation send auxiliary[6]
c. only very heuristic spectral modifications were performed, placing a first-order low pass filter controlled by the y position with the following algorithm:

$$frq_{cutoff} = \left(\frac{frq_{sr}}{2}\right) + \left(\frac{k_{const}\, y_{instr}}{y_{room}}\right)$$

and an emphasis second order bandpass filter with the following transfer function:

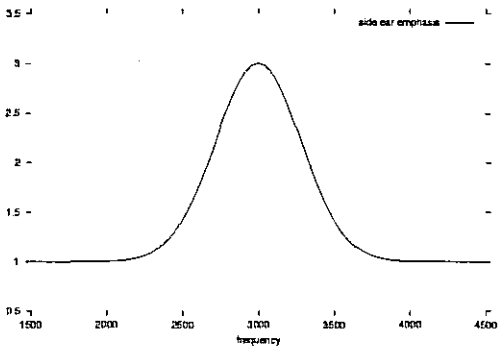


Figure 2. Side emphasis transfer function

applied to the signal virtually placed

perpendicularly to one of the two ears (i.e. at $y = 0$)

d. we had to take extreme care that the signal would never "pass through the head" of the listener (virtually speaking, of course): because of the amplitude algorithms we had chosen, the center of the head of the listener is an asyntotical locus of infinite amplitude — even though this can never happen in reality, it is easy to 'blow-up' signal processing units while testing motion patterns (especially traversals and random patterns; cf. below);

Given these limitations, and figuring out a way of carrying out these coherent variable transformations gave us some reasonable results during experimentations, so we decided to build our model for final processing of the master tracks. We decided to use csound[5, 14] to carry over the processing of the 150+ Mbyte mono files of every single track that needed spatialization, and every full run took approximately 8 to 12 hours on a Pentium 120 machine running the Linux operating system.

6 Csound implementation

The csound orchestra implementation was broken down into multiple instruments feeding each other in a more-or-less top-down fashion:

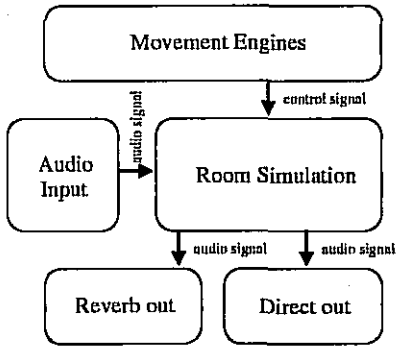


Figure 3. Csound orchestra structure

The 'motion engine' part was by itself composed of four different instruments:

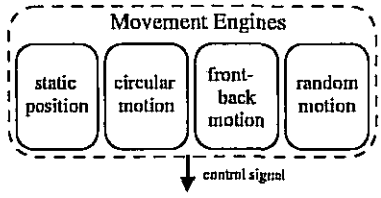


Figure 4. Motion engines

Each motion engine incorporated an interpolation envelope to be able to switch from one motion to the other without jumps, and the random engine had to be carefully redesigned to avoid getting 'too close' to the listener (or even to 'go through his head'). This was achieved calculating random patterns in terms of modulus and phase and allowing only a restricted range on the randomness of the modulus. The following plots are an actual output of the csound instrument:

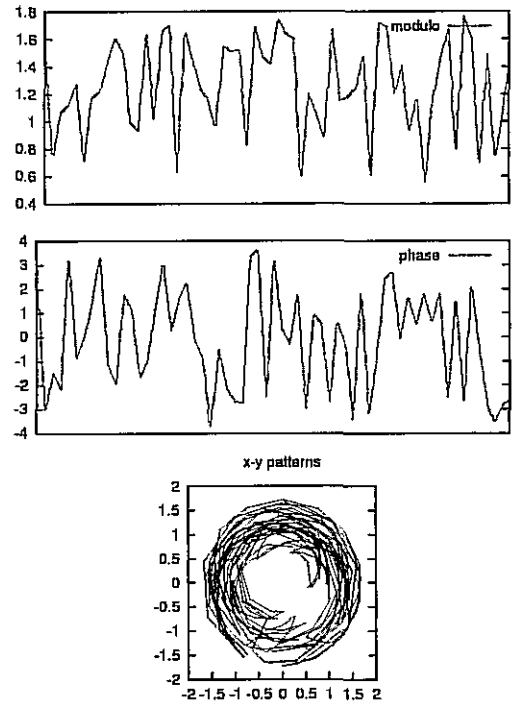


Figure 5. Polar random engine output

Another source of problems was the heavy doppler shift effect caused by fast tangential moving sounds. In order to avoid disturbing out-of-tune and glissando effects, we remapped tangential movement speed through the following function:

$$s = x_{pos} tg(\alpha_0 y + \phi_0)$$

where

$$\alpha_0 = \frac{atan\left(\frac{y_{\min}}{x_{pos}} - \phi\right)}{y_{\max}}$$

and

$$\phi = atan\left(\frac{y_{\min}}{x_{pos}}\right)$$

This functions generates mappings like:

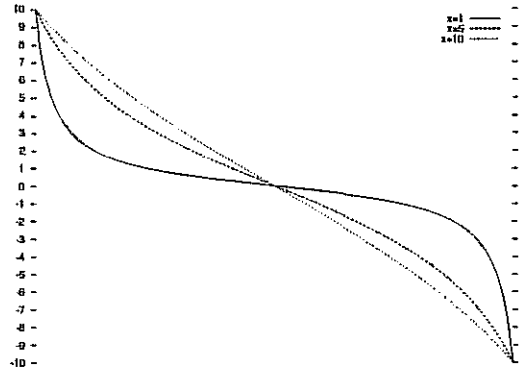


Figure 6. Tangential speed maps

In the above figure, three mapping curves are presented for a front-back passage with three different values of x (respectively 1, 5 and 10 meters

to the right of the listener). As it can be seen, a 10-meter distant sound source does hardly have any variation while a 1-meter distant one slows down considerably in proximity of the listener.

7 Considerations on results

During the preparation of the tracks and after mixing, we have made several listening sessions in many different environments, both conforming and non-conforming to the model we had adopted for calculations. While the conforming listening environments responded as expected the nice surprise was given by the response of non-conforming environments and listening positions: even if sound motions do not behave exactly as written, a three-dimensional characteristic is always clearly present even in thick textural passages. Thus, the musical function of sound spatialization is preserved.

Acknowledgements

This work could have not been achieved without the collaboration of several institutions and many individuals. Many thanks go, first of all of course, to Adriano Guarnieri without whose work and visions nothing would have existed. Many thanks are due also to the Cantiere Internazionale d'Arte of Montepulciano which commissioned *Orfeo cantando... tolse* in the first place, and to the Orchestra della Toscana which played in the recording enduring all the added difficulties of the necessary recording conditions. Both (difficult) situations were fully understood and mediated by Giorgio Battistelli, artistic director at different times of both institutions. Ricordi, the publisher and producer of the recording, has been great in providing and defending the "optimal" research environment: Morena Cambri, the executive producer, Pietro Borgonovo, the conductor, and Michael Seberich, the recording engineer were essential in obtaining good results. Riccardo Dapelo wrote the first prototype csound orchestra for the room simulation and much of our work is due to his precision in debugging it. Last but not least, the **Centro TEMPO REALE** has been indeed the unique place, from 1994 to 1997, were this work could develop; the quality of the work and life there was secured by several individuals who deserve recognition here: Fabio Fassone, Francesca Tortelli, Mariapia Redditi, Detlev Schumacher and Damiano Meacci.

References

[1] Begault, Durand R., *3-D Sound for virtual reality and multimedia,* Academic Press, Chestnut Hill, MA 02167 (1994).

[2] Bernardini, Nicola and Otto, Peter, "TRAILS: An interactive system for sound location" in *Proceedings of the International Computer Music Conference 1989,* CMA, San Francisco (September 1989).

[3] Bernardini, Nicola and Vidolin, Alvise, "La localizzazione spaziale dei suoni" in *Musica e Fisica,* ed. Frova, Andrea (1997). (in press).

[4] Blauert, Jens, *Spatial Hearing: the Psychophysics of Human Sound Localization,* MIT Press, Cambridge, MA (1983).

[5] Boulanger, Richard (ed.), *The Csound book: Tutorials in Software Synthesis and Sound Design,* MIT Press, Cambridge, MA (1998). in press.

[6] Chowning, J., "The dimensions of loudness and auditory perspective", *I Quaderni della Civica Scuola di Musica,* 21-22, pp. 99-105 (December 1992). (italian translation).

[7] Chowning, John, "The simulation of moving sound sources", *Journal of the Audio Engineering Society,,* 19, 19, pp. 2-6 (1971).

[8] Rocchesso Davide, "The Ball within the Box: a sound-processing metaphor", *Computer Music Journal,* 19, 4, pp. 47-53 (1995).

[9] Kendall, G.S. and Martens, W.L., "Simulating the cues of spatial hearing in natural environments" in *Proceedings of the 1984 International Computer Music Conference,* International Computer Music Association, San Francisco (1984).

[10] Kendall, G.S., "A 3-D sound primer: directional hearing and stereo reproduction", *Computer Music Journal,* 19, 4, pp. 23-46 (winter 1995).

[11] Kendall, G.S., "The decorrelation of audio signals and its impact on spatial imagery", *Computer Music Journal,* 19, 4, pp. 71-87 (winter 1995).

[12] Moore, F.R., "A General Model for Spatial Processing of Sound", *Computer Music Journal,* 7, 3, pp. 6-15 (1982).

[13] Rocchesso, Davide and Vidolin, Alvise, "Sintesi del movimento e dello spazio nella musica elettroacustica" in *La Terra Fertile. Atti del Convegno 1996,* ed. De Amicis, Maria Cristina and Prignano, Ignazio, pp. 27-31, L'Aquila (1996).

[14] Vercoe, Barry, *CSOUND: A Manual for the Audio Processing System and Supporting Programs,* MIT Media Lab, Cambridge, MA (1986). Program Documentation.

A REAL-TIME PHYSICAL MODEL OF THE PIANO

Gianpaolo Borin
Centro di Sonologia Comp.
Università di Padova
via S. Francesco, 11
35131 Padova, Italy
borin@dei.unipd.it

Davide Rocchesso
Dip. Scient. e Tecnol.
Università di Verona
Strada Le Grazie
37134 Verona, Italy
rocchesso@sci.univr.it

Francesco Scalcon
Generalmusic S.p.A.
via delle Rose, 12
47048 S. Giovanni in M., Italy
francescos@generalmusic.com

Abstract

The implementation of a complete physical model of the piano is demonstrated. It includes 88 lossy and dispersive strings (with full polyphony), hammers and dampers, and a soundboard load/radiation.

1 Introduction

In the last four years we have been developing physical models of the piano, gradually including more details and optimizing the algorithms for design and simulation [1]. In the meanwhile, the speed of affordable computers has increased to the point that we can now demonstrate an 88-notes polyphonic piano model running in real time on a Pentium-based workstation. So far, commercial digital pianos have been influenced by physical modeling only marginally. For example, the resonance effect given by the damper pedal was simulated by means of a network of simplified string models and included in a PCM digital piano as a post-processing effect [2]. Since the dynamic behavior of physical models is considered more realistic than sampling techniques, we expect that digital piano makers will replace more and more components of their sound generators with physical models.

2 Model Architecture

Our piano model is based on the decomposition of the actual instrument into functional blocks. Each block is either simulated by a physical model or replaced by an "equivalent" signal processing module. For each note, an explicit physical model of the hammer-string interaction is implemented. The string is simulated by a feedback delay loop [1] having loop filters accounting for dispersion and losses. Starting from measured distributions of partials and decay rates, the coefficients of these filters are identified by means of filter design techniques. As far as the losses are concerned, only a smooth lowpass component is ascribed to the string, the remaining ripples being due to the non-resistive load of the soundboard and to string coupling [10, 11]. Therefore, we connect the strings at a loaded junction [8], where the load is implemented as an irregularly-rippled digital filter. The resulting scheme is depicted in fig. 1, where $S = (\sum_i Z_i + Z_L)^{-1}$, being Z_i the wave impedance for the i-th string, Z_L the impedance of the load, v_i^+ and v_i^- the velocity waves respectively incoming and outgoing from the bridge.

The hammer-string interaction mechanism has been described elsewhere [3, 4]. It is worth noting that we use a method for eliminating the non-computable loops which result from discretization of the non-linear equations of the hammer [5]. This implementation avoids artificial instabilities and reproduces a reliable force signal, thus producing a more natural sound.

Piano strings exhibit frequency-dependent losses and dispersion, which have to be simulated in order to attain realistic sounds. According to a well-established tradition brought by the literature of digital waveguides [9], losses and dispersion are lumped for the whole string, and simulated, respectively, by lowpass and allpass filters. In our experience, the problem of simulating string dispersion is the most demanding in terms of computations. We developed a method for designing dispersive (allpass) filters where it is possible to set a frequency-dependent weight, in such a way that the partials in low frequency are more accurately put on their exact (inharmonic) positions [6]. A psycho-acoustic investigation on the perception of inharmonicity has also been performed in order to come up with better criteria for guiding the allpass-filter design [7].

3 System Considerations

The implementation that we are currently demonstrating runs on a workstation based on two processors Intel Pentium II working at a clock rate of 300 MHz. The simulation has been written in C language and compiled by means of the Portland Group parallelizing C compiler. The operating system supporting multiprocessing in our application is Linux Red Hat 5.0 (kernel 2.0.32). The full-fledged version of our simulation program achieves speedup 1.73 due

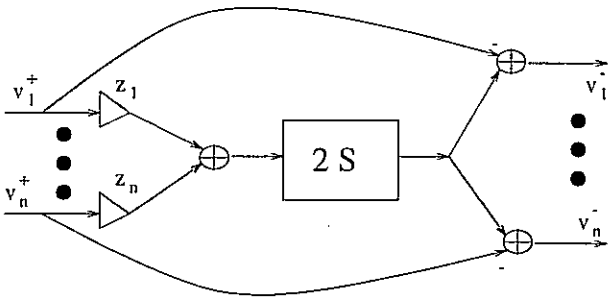


Figure 1: Connecting n strings at the soundboard

to multiprocessing.

Audio and MIDI devices are controlled by means of the OSS drivers (4Front Technologies). By introducing a small amount of buffering, we can limit the overhead due to audio/MIDI input/output to only 1% of total running time without introducing large latency times.

The running time largely depends on the options that we activate in the program call. In other words, by increasing the number of details being modeled we also increase the running time. At the moment of writing this paper, we achieve a performance less than 1.5 times slower than real time when all the options are active. A real-time performance is achieved either by slightly relaxing the accuracy of the model or by reducing its (88-notes) polyphony. In any case, accurate physical modeling of the piano in real time is just around the corner.

4 Acknowledgments

This work has been developed at C.S.C.-D.E.I., Università di Padova, under a Research Contract with Generalmusic S.p.A.

References

[1] G. Borin and D. Rocchesso and F. Scalcon, "A Physical Piano Model for Music Performance," in *Proc. Int. Comp. Music Conf.*, Thessaloniki, Greece, pp. 350–353, 1997.

[2] M. Ambrosini and F. Campetella and F. Scalcon and G. Borin, "Simulazione dell'Effetto del Pedale di Risonanza nei Pianoforti Digitali," in *Proc. Int. Conf. on Acoustics and Musical Research,* Ferrara, Italy, pp. 101–106, 1995.

[3] G. Borin and G. De Poli, "A Hammer-String Interaction Model for Physical Model Synthesis," in *Proc. XI Colloquium on Musical Informatics,* Bologna, Italy, pp. 89–92, 1995.

[4] G. Borin and G. De Poli, "A Hysteretic Hammer-String Interaction Model for Physical Model Synthesis," in *Proc. Nordic Acoustical Meeting,* Helsinki, Finland, pp. 399–406, 1996.

[5] G. Borin and G. De Poli and D. Rocchesso, "Elimination of Delay-free Loops in Discrete-Time Models of Nonlinear Acoustic Systems," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY, pp. 12.1.1–4, 1997.

[6] D. Rocchesso and F. Scalcon, "Accurate Dispersion Simulation for Piano Strings," in *Proc. Nordic Acoustical Meeting,* Helsinki, Finland, pp. 407–4414, 1996.

[7] D. Rocchesso and F. Scalcon and G. Borin, "Subjective Evaluation of the Inharmonicity of Synthetic Piano Tones," in *Proc. NATO Advanced Study Institute on Computational Hearing,* Il Ciocco (Tuscany), Italy, pp. 251–255, 1998.

[8] J. O. Smith, "Music Applications of Digital Waveguides," report stan-m-39, CCRMA - Stanford University, Stanford, California, 1987.

[9] J. O. Smith, "Physical Modeling Using Digital Waveguides," *Computer Music J.*, vol. 16, num. 4, pp. 74–91, Winter 1992.

[10] G. Weinreich, "Coupled Piano Strings," *J. Acoust. Soc. Am.*, vol. 62, num. 6, pp. 1474–1484, 1977.

[11] K. Wogram, "The Strings and the Soundboard," in *Five Lectures on the Acoustics of the Piano*, Royal Swedish Academy of Music, num. 64, Stockholm, pp. 83–98,1990.

A new real-time sound synthesis system intended for live performances

Giorgio Nottoli*, Mario Salerno**, Giovanni Costantini**

* Conservatorio di Musica "L. Refice"
via Roma, 25
03100 Frosinone, Italy
mc5980@mclink.it

** Department of Electronic Engineering
"Tor Vergata" University of Rome
via di Tor Vergata, 110
00133 Rome, Italy
giovanni.costantini@uniroma2.it

Abstract

We have developed a new electronic system for real-time sound synthesis. It is called Betel Orionis and it is based on the dedicated DSP Orion. This system is intended for live performances in concert halls and it is interfaced with personal computer and MIDI controls that allow the composer or performer to interact with it. Interesting features are its complete programmability from the lowest level (DSP) up to the highest level (control algorithms and interactive interface for the user) and its processing power that allows complex real-time synthesis.

Introduction

A system has been designed which allows synthesis, processing and spatialization of sound in real-time (it is intended to be used live in a concert hall): this system is SAIPH.

SAIPH is designed as a modular system: it is made of modules which are dedicated to perform specific musical signal processing. The Betel Orionis system, presented in this paper, is the module of the SAIPH system allotted for sound synthesis. It has been utilised for a live performance at the MACBA (Museum of Contemporary Art of Barcelona) on 31 January 1998.

In that performance, Betel Orionis synthesized eight independent sound structures in parallel using a kind of granular synthesis with both sine-waves and PCM grain events. The 16 independent Betel Orionis outputs have been connected to two 8x8 digital mixers/spatializers Mixtral [3] driving 16 speakers positioned in the MACBA museum. Before describing the Betel Orionis system, a brief description of the Orion DSP architecture will be given.

The DSP Orion has been designed in 1990 by one of the authors [1]. It is a DSP dedicated to the musical signal synthesis and it is the core of this system.

The Orion microcircuit

Orion is a Digital Signal Processor specifically designed to synthesize sound events according to all the principal synthesis algorithms utilised in computer music. Orion architecture is based on a set of top-level units interconnected by an unidirectional multi-bus structure. In particular, the interconnection architecture has been properly designed to support the implementation of the synthesis primitives. So, the basic primitives are implemented by special hardware units designed for maximum execution speed. Each main unit is able to execute elementary sequence of operations under the supervision of the main chip microcontroller.

Orion is constituted of four specialized Arithmetic Logic Units, three data RAMs and three I/O units.

Executing music oriented digital signal processing algorithms, the computing work is distributed among the four ALUs to maximize parallelism while the data transfer work takes advantage of the three data rams structure.

Additive, phase and frequency modulation synthesis algorithms can be implemented using the internal sine-wave generator. The Orion DSP can address up to 16 Msamples of dynamic memory in witch you can store waveforms to make more complex sounds with PCM looping algorithms.

The set of microinstructions enable a complete support to carry out hardware envelopes for high quality sound generation and reduction of the control microprocessor computing charge.

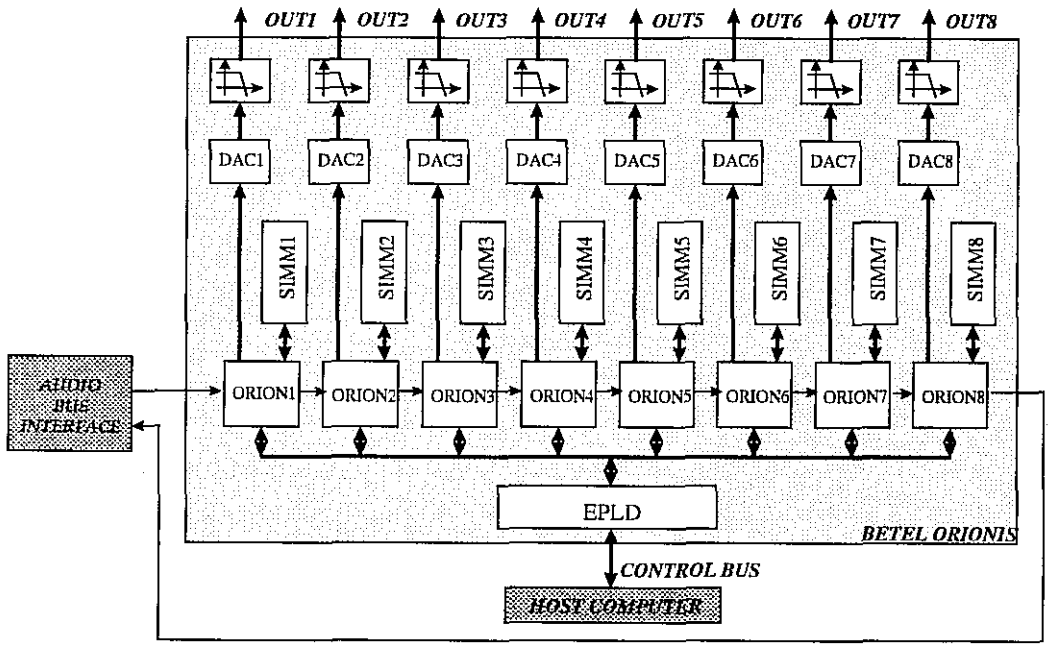
The Betel Orionis system

Betel Orionis is a multiprocessor system for real-time sound synthesis based on the dedicated DSP Orion and it is a hardware module of the SAIPH system, described previously.

The project is based on a multiprocessor architecture and implements all the most important algorithms for real-time sound synthesis with a capacity of eight stereo output audio channels. In particular, each channel is controlled by an Orion DSP driven in real-time by a host personal computer.

Thanks to the Orion DSP architecture dedicated to the implementation of methods for sound synthesis, the system reaches a considerably high processing power.

In order to get an idea of the efficiency of the system we



Betel Orionis block scheme

may refer to two very common methods of synthesis:

- additive synthesis:
  in this case, the system offers a capacity of 1024 virtual oscillators
- PCM synthesis:

the whole system disposes of a capacity of 240 PCM channels with a memory of 128 Msamples.

The system core is made of eight Orion DSPs synchronized by an external 32 MHz clock generator. Every Orion is intefaced with:

a) the PC ISA bus by the host interface
b) another Orion DSP or a stereo DAC device by the synchronous serial interface (the choice is made by the microprogram); the eight DACs, together with anti-aliasing filters, constitute the system audio interface
c) a SIMM module up to 32 Mbytes (16 Msamples) by the dynamic memory interface.

The Betel Orionis system block scheme is shown in Figure.

Betel Orionis automatically generates complex sound textures and allows the composer or the performer to control them in real-time. The system is completely programmable from the lowest level (DSP) to the higher level (control algorithms and interactive interface towards the user), thus making it extremely versatile for the most varied user requests.

The system software architecture is set on three levels:

1) user interface level, designed in such a way to allow the control of a large number of events via a few high level parameters which are activated via gestual controls (mouse, PC keyboard, sliders and other MIDI controls)
2) control level, with control algorithms based on deterministic and random methods that are located within the host computer (PC)
3) sound generation level, with sound synthesis algorithms that are located within the microprograms for the Orion DSP.

The user interface is the most innovative part of the system and it is designed to meet the expressive needs of the musical composition and execution. It allows the control of a large number of complex events via a few high level parameters driven by gestual controls.

References

[1] Giorgio Nottoli: ORION: a single chip digital sound processor/synthesizer, Proc. of IX Colloquium on Musical Informatics, 1991, Genova, Italy

[2] Mario Salerno, Fausto Sargeni, Giorgio Nottoli, Giovanni Costantini: Tecnologia e musica all'Università "Tor Vergata" di Roma, Proc. of La terra fertile, Ottobre 1996, L'Aquila, Italy

[3] Giorgio Nottoli, Carlo Alberto Paterlini, Attila Baldini: Software Aided Mixing on a Low-Cost Digital Console, Proc. of AES /103rd Convention, 1997

[4] Giorgio Nottoli, Giovanni Costantini: Betel Orionis: a real time, multiprocessing sound synthesis system, Proc. of Journées d'Informatique Musicale, May 1998, Agelonde, France

[5] Giorgio Nottoli, Mario Salerno, Giovanni Costantini: Betel Orionis: un sistema multiprocessore orientato alla sintesi del segnale musicale in tempo reale, La terra fertile, September 4-6, 1998, L'Aquila, Italy

ORPHEUS: Software for Interactive Sound Synthesis and Computer-Assisted Composition

Michael Hamman
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
705 W. Nevada #4
Urbana, IL 61801
m-hamman@uiuc.edu

Abstract

In this paper, I give a "flyover" description of the aesthetic, musical, philosophical motivation behind the creation of ORPHEUS, as well as brief overview of its functional behavior.

1. Introduction

In modeling a task domain, we are effectively modeling a human performance. Human performance embodies various kinds of interactions. At one extreme, those interactions provide feedback in real-time -- that is, one can observe the effects of one's actions as they are executed. At another extreme, the effects of those actions are observable only at a latter moment in time. Colloquially, these are referred to respectively as "real-time" interaction and "non real-time" interaction.

Real-time interaction benefits the investigation of systems whose behaviors are observable in what Otto Laske has termed "conscious-time"[1]. In conscious-time, one is interested in being able to observe features which evolve over a duration of a few seconds and, most particularly, in response to gestures and actions taken. The kinds of events which one may observe are, by necessity, temporally linear and non-reversable. Direct manipulation displays facilitate this mode of interaction.

Non real-time interaction benefits the investigation of systems whose outputs are not temporally linear and in which events occur over what Laske has termed "interpretive time." As Laske notes, "the illusion of lasting time is created by memory through interpretations of events on a high level of abstraction" [1]. Programming displays facilitate these kinds of interactions since they allow a composer to hypothesize musical structure as outside-time.

As a software system, *Orpheus* is an experiment in designing a possible task environment for the composition of acoustical and musical structures. Toward this end, it attempts to join real-time interactive tools (for the design of sound morphologies) with non real-time tools (for the organization and deployment of sound morphologies across different levels of time). Moreover, it attempts to join direct manipulation interfaces with programming interfaces in an effort to capture different moments of compositional design.

This paper describes some of the features of the system, as well as detailing what the author views as some of its shortcomings while emphasizing the authors strong feeling that composers must be involved in the design of computer-based task environments for musical production and research, not so much as a way of guarenteeing the preservation of historical musical paradigms, but as a means toward retarding the rate at which music software systems surrender to those historical musical paradigms.

2. Overview of the System

Orpheus is a software system for investigating and designing acoustical and musical structures. It is built on top of a sound synthesis library and real-time audio scheduler [2]. Currently it employs a single synthesis algorithm (described elsewhere in the this Proceedings volume).

While the system is designed primarily to investigate control interfaces to synthesis algorithms it can, for precisely this reason, be a useful tool in the development of synthesis algorithms, at the source code level.

Orpheus is built on a Model/View architecture. It provides displays for direct manipulation and real-time interaction with synthesis algorithms. It also provides text-based editors and a command line utility for interactively sketching and instantiating acoustical and musical data structures.

3. Direct Manipulation Displays

Orpheus currently provides two direct manipulation displays for any given synthesis algorithm. The first display presents a bank of sliders, one slider per synthesis parameter. With this display, the composer can investigate the synthesis algorithm through independent control of various parameters. When studying the behavior of a system, this is often a good place to start.

The second display presents an "integrated control display" (figure 1). The integrated control display allows simultaneous manipulation of multiple parameters. This notion of interaction is predicated on

the idea that when investigating the behavior of a system one is often interested in the effect that various *groupings* of parameters will have on the behavior of the system.

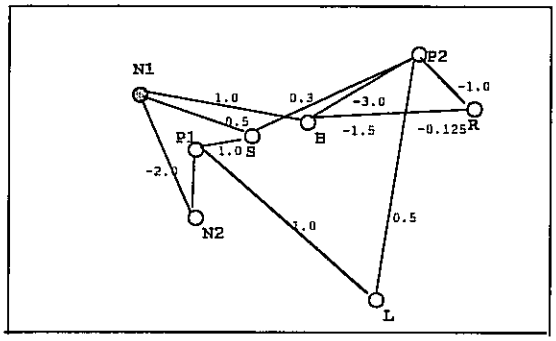


Figure 1: Integrated Control Display

Using the integrated control display tool, one is empowered to create different interactive "views" into the underlying system. A "view" constitutes a selection and grouping of parameters. Figure 1 depicts such a view. Here parameters N1, N2, P1, P2, B, R, S, and L are selected into the view. Groupings of parameters are formed through the linking of one parameter to another.

Parameters "N1", "N2", "S", and "B" constitute one such grouping. To engage this particular grouping, one might click on the node labelled "N1" and drag it around, observing acoustical feedback to one's movements. Movement of node N1 engenders movement of all nodes to which that node is connected (depicted by lines in the diagram): N2, S, and B. Each connection is defined by a "weight" according to which the movement of one node will effect the movement of its attached node. So, for instance, movement of the node labeled "N1" would cause movement of nodes "N2", "S", and "B" by weight factors of -2.0, .5, and 1.0 respectively.

Within the integrated control display, one can add connections, remove them, or change their weight. Any such display can be made to be persistent: they are saved as "sound configurations" into files with the file extension .cfg.

Temporally-bound sonic structures unfold as a consequence of dragging different nodes around. Any one such structure traces a "path" followed by the movement of a node. The integrated control display allows the composer to save paths for future retrieval. In addition, if when investigating the behavior of a particular grouping, one finds a particularly interesting region within the parameter space being tracked, one can use a zoom tool in order to enlarge the view and zoom in on that area.

Through the combined use of the sliders display and the integrated control display, the composer investigates aspects of the behavior of the underlying synthesis algorithm through direct manipulation. Moreover, s/he can create different views and save different paths within any such view for future use and reference.

4. Acoustical and Music Data Structures

Currently *Orpheus* defines three primary data structures. These are (1) the *SoundObject*; (2) the *EventSpawner*; and (3) the *StreamBuilder*.

4.1 Sound Objects

Saved paths can form the basis for the definition of a particular acoustical 'prototype' or 'template', from which events are generated and projected in time. Within the current version of *Orpheus*, such a template is called a "SoundObject." The term is purposefully reminiscent of Schaeffer's term: it refers not to a specific acoustical artifact *per se*, but to a model according to which many artifacts might be generated . A SoundObject forms the basis upon which actuated *variants* are produced.

A SoundObject includes another data structure called a *SpawnFilter*. A SpawnFilter determines the constraints according to which acuated acoustical events are spawned from a SoundObject.

A SoundObject also defines aspects of the auditory environment in which spawned events are to be produced. This is accomplished through an *Environment* data member object. Depending on the particular circumstance at any given moment, the Environment member of a SoundObject can constrain the unfolding of other simultaneously occuring SoundObject events. In this regard, the SoundObject views acoustical events not as isolated occassions, but rather as parts within a larger whole, in which the larger whole can influence the structure of otherwise homegenous parts.

The SoundObject data structure is depicted in figure 2.

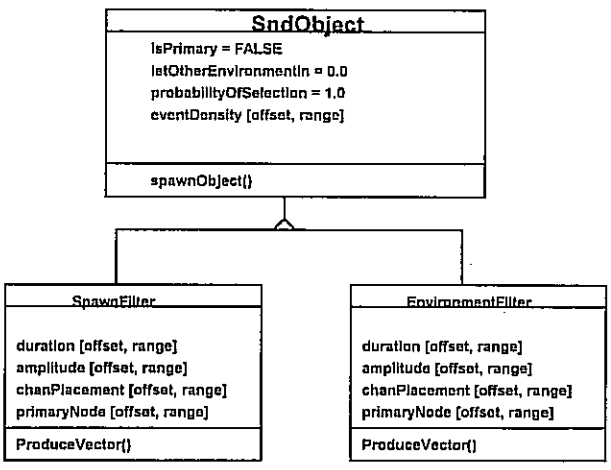


Figure 2: SndObject data structure

4.2 Event Spawners

Within *Orpheus*, a composer can generate more-or-less traditional "playlists" of SoundObjects. A more compelling tool however is the *EventSpawner*. By defining an EventSpawner, a composer articulates the

conditions for the unfolding of events in time by specifying the unfolding of data. An EventSpawner links three data structures: the SoundObject, a SequenceProducer, and a ProcessObject (figure 3).

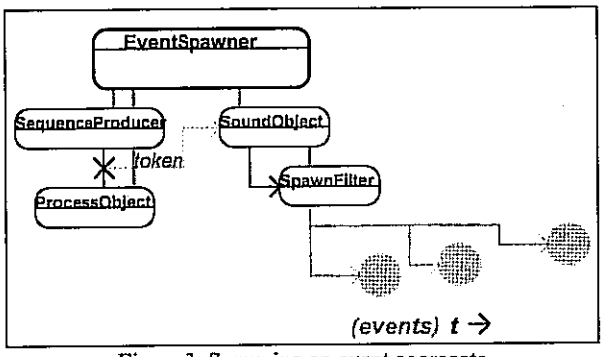


Figure 3: Spawning an event aggregate

A SequenceProducer generates tokens on the basis of which events are generated. SequenceProducers can be of many different types: they can effect the output of particular grammars; of stochastic systems; or of iterative systems, such as non-linear chaotic systems.

A SequenceProducer generates structures that are of local scope: they apply to currently computing aggregate of events. ProcessObjects propagate higher level structures in that the output of a Sequence Producer is combined with iterations generated within a ProcessObject. For example, a ProcessObject might engender the gradual decay of particular acoustical features. This could result in event aggregates in which each successive aggregate is quieter, more sparse, or more disparate than the previous one.

EventSpawners spawn events whose morphology derives from a SoundObject. First, it determines the number of events for a particular event aggregate. Then, it tells the SequenceProducer to generate that number of tokens. These tokens are combined with output from the ProcessObject to produce an input to the SoundObjects spawn filter. The result is the generation of a sound event which is placed on the system scheduler for production at some specific point in time.

When that time arises, that event is computed *at that time* (it is **not** computed in advance). This allows for greater flexibility than would be the case if all events were computed ahead of time. Its cost however, is greater taxation of computing power and possible audio dropouts.

4.3 StreamBuilders

StreamBuilders are data structures for aggregating streams of events. A StreamBuilder contains from one to three EventSpawners. It schedules the execution of event aggregates in much the same manner that EventSpawners schedule the appearance of individual events. Moreover, StreamBuilders allow for interaction between constituent EventSpawners and, as such, of unfolding streams of events.

Within a particular StreamBuilder, simultaneous streams are mutually determinative. One way in which this mutual determinateness is manifested is through activation of a SoundObject's *Environment* object. When an EventSpawner is launched, it is given a value which determines which of the EventSpawners is "primary". Which ever EventSpawner is primary is given a pointer to the other EventSpawner's SoundObject. The Environment of the primary SoundObject is used, in this instance, to constrain the synthesis parameters of any other simultaneous unfolding SoundObjects.

This situation is depicted in figure 4. Here a single StreamBuilder has launched two EventSpawners, ES1 and ES2. Since ES2 has been given primacy, the Environment of its member SndObject (SndObject__2) acts as a secondary filter to the SpawnFilter belonging to SndObject__1. As a consequence, the events of which SndObject__1 is a prototype are constrained according to the Environment specified for SndObject__2.

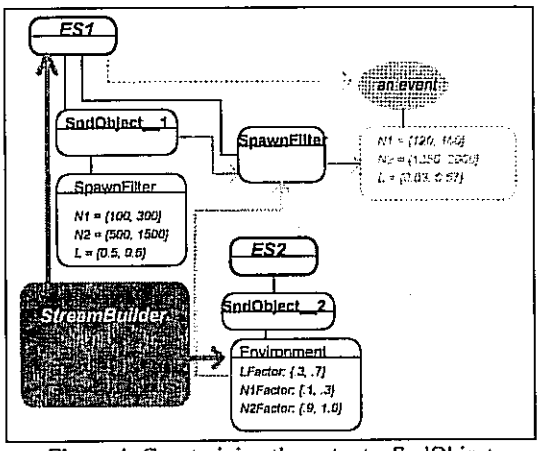


Figure 4: Constraining the output a SndObject

5. Musical Form and Data Structure

One of the principles I have sought to demonstrate in developing *Orpheus*, is the principle of the permeability of musical form [2]. By this I mean that generative structure is not top-down. Rather, while aspects of the global framework may influence the evolution of events in time, the reverse is also true: that the particularity of the unfolding of events determines, in part, the eventual unfolding of a higher level framework.

This is currently accomplished through the ProcessObject. Each time it is used, the ProcessObject is left with some, oftentimes minimal, trace of the data object that used it, be it an EventSpawner, or a SoundObject. The ProcessObject can be initialized with a minimal kernel structure. Then, as it is used, that structure is gradually fleshed out during the process of its use.

6. Using *Orpheus*

In addition to the direct manipulation displays described above, Orpheus presents a text editor for

editing various data structures, and a command line interpreter for initializing and launching data structures and processes.

Figure 5 shows an example of the command line interpreter with several commands. The first command instantiates a SndObject called "so" from a sound configuration displayed within an integrated control window. It will be remembered that within an integrated control display, particular "views" can be defined and saved to disk. Moreover, any such view can have any number of "paths." Here "so" is instantiated from path #4 of the sound configuration called "saw1."

```
>SndObject so = sndConfig saw1 path 4
>so play
>so play amp=1.3
>so play dur=2.0
>so play dur=2.0 prim=1.2
>so spawn so2 amp=.5 dur=2.0 prim=1.2
```

Figure 5: *Orpheus* Command line Interpreter

The second command plays the SndObject. Even though a SndObject has been described as a sound "prototype", we can still "play" it in its original form. The second command--"so play amp1.3"--plays the sound object so, with its amplitude multipled by 1.3. The next command plays the sound object, but this time with the duration doubled. Doubling the duration effectively time-stretches the unfolding of the object.

The next command causes the Sound Object to play itself at twice the duration. In addition, it multiplies the value of its pimary node by 1.2. Remember that the primary node is the node which you dragged with your mouse, and whose movement occasioned the movement of any parameter nodes attached to it.

6. Conclusions and Future Developments

This document serves as very sparse introduction to a fairly complicated software project: many details have been left out for the sake of brevity. Part of the reason is also that the project is in a state of constant flux.

The software is better tailered to the composer who is interested in investigating non-standard approaches to music composition and sound design. It is, as such, not particularly usefull as a "production" tool.

In the future, I would like to add support for *VSS*, an authoring tool developed by the Audio Development Group at NCSA[ ] or for Perry Cooks STK synthesis library. Currently, it is built around a single synthesis algorithm which, though rich in its behaviors, some may find lacking in variety. I would also like to add support for handling network messages so that it might perform in a live performances environment. I would like to continue enriching the data structure layer in order to incorporate a variety of procedural scenarios. Finally, I would like to build a strong language interpreter into it, such as ELK.

7. Availability

Source code, executables, and examples are available on-line at *http://duracef.shout.net/~mhamman.*

8. Acknowledgements

My thoughts in the development of *Orpheus* have benefited from many other similar projects. These include *Modalys* [4], *Foo* [5], *Bol Processor* [6], and *Kyma* [7]. Other influences include the work of Roger Dannenberg, research within the ACROE group, and the research and tools within the Audio Development Group at NCSA. Acknowledgement is also made to Camille Goudeseuene who provided software development advice when it was most needed, and who was a partner in the development of AREAL, the real-time scheduling library on top of which *Orpheus* is built.

9. References

[1] Laske, O. E. "Considering Human Memory in Designing User Interfaces for Computer Music." *Computer Music Journal* 2(4).

[2] Goudeseune, C., and Hamman, M. "A Real-Time Audio Scheduler for Pentium PCs." *Proceedings of the 1998 ICMC.* San Francisco: Computer Music Association. 1998.

[3] Koenig, G.-M. "Composition Processes." Lecture delivered at the UNESCO Workshop on Computer Music. Aarhus, Denmark. 1978.

[4] Morrison, J. D., and Adrien, J.-M. "MOSAIC: A Framework for Modal Synthesis." *Computer Music Journal* 17(1), pp. 45-56. 1993.

[5] Eckel, G., and Gonzales-Arroyo, R. "Musically Salient Control Abstractions for Sound Synthesis." *Proceedings of the 1994 ICMC.* San Francisco: Computer Music Association, pp. 256-259. 1994.

[6] Bel, B. "Migrating Musical Concepts: An Overview of the Bol Processor." *Computer Music Journal* 22(2). pp. 56-64. 1998.

[7] Scaletti, C. "The Kyma/Platypus Computer Workstation." in *The Well-Tempered Object: Musical Applications of Object-Oriented Software Technology*, ed. S. T. Pope. Cambridge, MA: The MIT Press. 1991.

"M.P.S."
Multimedia Performance System

Alberto Rapetti
L.I.M. Laboratorio di Informatica Musicale
Dipartimento di Scienze dell'Informazione
Università degli studi di Milano
Via Comelico, 39
I-20135 Milano (Italia)
Fax +39 2 55006373
e-mail : alberto@lim.dsi.unimi.it

The contemporary social universe has developed in relation with a basic event: the generalized proliferation of the images and the affirmation of their constitutive model in every field of the human experience.

Confusedly assemble in the metropolitan view, individually proposed in newspapers or in promotional posters, offered in rapid, disorderly, overwhelming assemblages in cinematographic sequences, images are able to propose their system of reality production as universal model: system in which the "social actors" use images to orientate themselves in the world; they live, desire, act through them.

Multimedia performance is a show in which a multiplicity of media are involved.

In this show, music and images are unitedly the object of the performance; this is the way to make concrete the expectations of telematic public.

MPS is the result of my degree thesis work realized during 1997. [1]

MPS is born following the steps of projects realized in L.I.M. (Laboratorio di Informatica musicale) in the field of Multimedia Performance such as the applications Temper [2] and MMP (Multimedia performer) [3].

MPS is an application written in Think C in Apple Macintosh programming environment.

It supplies musicians an integrated environment for the registration and synthesis of sequences of images in QuickDraw picture format.

MPS provides tools for the visualization in Real Time of images according to the events generated playing musical controllers with MIDI interface such as musical keyboards, percussion pads, MIDI guitars.

MPS consists of three main phases: the Registration phase, the Synthesis phase, the Visualization phase.

Registration phase

In the Registration phase, MPS provides mechanisms allowing the musician to define and organize his personal archive of visual information.

With MPS, the musicians can define and organize categories of images in QuickDraw picture format.

Musicians can import PICT files into categories.

Images forming the archive are shown in scrollable windows called "PICTwindows"; each document contains six pictures.

Categories are identified by boxes with a caption inside; categories are shown in windows called "Categorywindows".

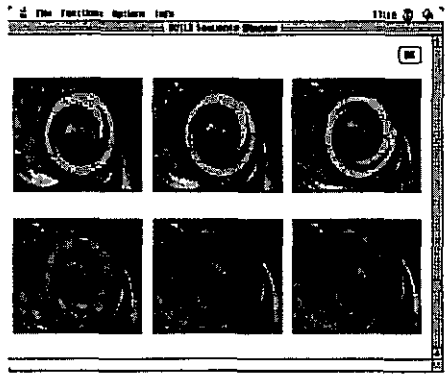


Figure 1: Pictwindow

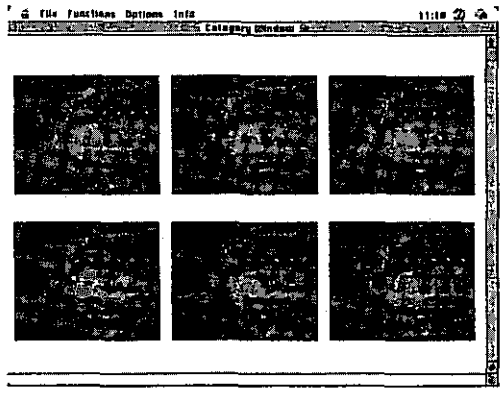


Figure 2: Categorywindow

Clicking the mouse inside a box representing a category, the PICTwindow containing the category's pictures will be shown.

A lot ot PICTwindows can be managed at the same time.

Infact a special memory management based on spool from disk reduce the memory request.

Synthesis phase

In the Synthesis phase the musician creates animation frame by frame taking picture from movies or from the images archive.

Musicians can build movies juxtaposing pictures taken from MPS' archive.

It is possible to import frames from Quicktime movies, to choose inside them the more interesting frames, to build a new movie frame by frame taking and transforming frames from other movies.

Musicians, using the mouse, can select images contained in PICTwindows and specify their order inside the movie.

Frames choiced are shown in windows called "Sequencewindows".

All sequences can be saved in files and loaded with a special menu call.

With MPS it is possible to build and organize sets of animations taken frame by frame.

These animations are the object of the Activation phase.

Activation phase

In the Activation phase the visualization of sequences' frames is performed.

The video process timeline can coincide with the one of the musical execution.

In the resulting multimedia score, each musical object, distinct on temporal plain, is composed by both a sound and an image.

If the video is composed by correlated frames, the result is the animation that follows the musical execution's metrics.

MPS implementation has required the use of techniques based on buffers offscreen [4] to increase the visualization images' speed in the Activation phase.

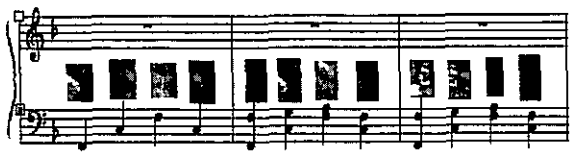


Figure 3: example of multimedia score

With MPS, it is possible to define the video process temporal scansion as a subset of the musical execution's one, according to the following parameters:

- **MIDI velocity.**

Only musical events, with MIDI velocity's value higher than the value of this parameter, activate images.

It is possible a dynamic control of the musical execution.

The musicians can emphasize, using images, parts of the musical execution according to the dynamics of the music played.

- **Position of the notes within the piano keyboard.**

With this parameter it is possible to define the part of the musical keyboard to which associate the visualization of the images.

For example the musicians could want images to be activated only by the bassline, to emphasize the skeleton of the musical execution.

In this case it is possible to set a low range on the keyboard.

A graphic interface showing a piano keyboard makes easy this operation.

Figure 4: KeyRange dialog window

- **Temporal interval between two consecutive music events.**

Musicians can specify the temporal interval between two consecutive musical events below whom stop the visualization.

This temporal interval is set indicating the speed of the musical execution and a rhythmic figure of reference in the following dialog window.

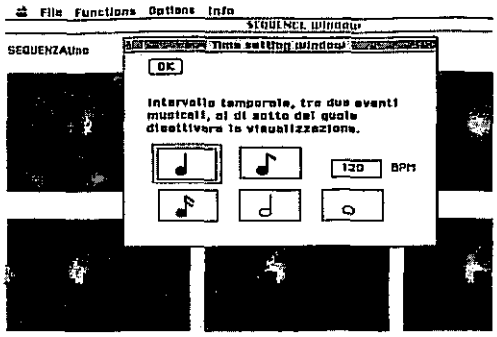


Figure 5: Timewindow

This parameter is useful to avoid that uncorrelated frames could be activated by a speedy sequence of musical events, resulting in this case not enjoyable.

- **Position inside computer's screen of images in the Activation phase.**

For example to control the position of images according to note's hight.

- **Color of the screen's background during the performance.**

With MPS the musician can organize the Multimedia performance direction defining tree-structures in which nodes contain sequences of images, and arcs contain musical events such as NOTE ON, Program change, chords.

When these events occur MPS cause the automatic crossing of the images_trees.

The following picture shows the dialog that MPS predisposes to build images-tree.

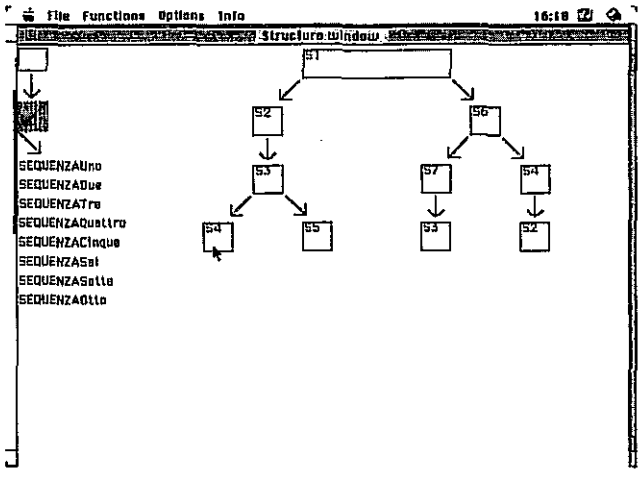


Figure 6: Structurewindow

In the left side of the window images' sequences loaded are indicated.

Using the mouse it is possible to select tools from the palette (such as a rectangle or a row) and build the images-tree.

The musicians can organize the sequences, assembled in the Synthesis phase, in order to their function inside the performance.

They can choose the sequences'order and the events that cause their activation.

The musicians can make choices in Real time.

They are able to preset visual paths related to musical paths.

The musician with MPS are both actor and director of the multimedia performance

References

[1] Rapetti, A. Tesi di Laurea in Scienze dell'Informazione *"Metodi, strumenti e architetture per la sintesi di processi video guidata da performance musicale. Realizzazione di un prototipo sperimentale"*, sessione 1997-1998.

[2] Morini, P., Tanzi, D., 1992. *Guida operativa del modulo Temper.*

[3] Conca, L. Tesi di Laurea in Scienze dell'Informazione *"Un Sistema per la Performance Multimediale basato su protocollo MIDI"*, sessione 1992-1993.

[4] Knaster, S., Rollin, K., *Macintosh Programming Secrets,* Addison-Wesley, 1992.

3D MUSICAL NOTATION - PROVIDING MULTIPLE VISUAL CUES FOR MUSICAL ANALYSIS

Dimitrij Hmeljak

Università di Trieste
DEEI, Via Valerio 10
34127, Trieste, Italy

Indiana University
Computer Science Dept., Lindley Hall
47405, Bloomington, Indiana, USA

Abstract

Visualization of musical information provides help and support for many musical analysis activities. This paper presents a system where standard music notation is augmented by multiple cue information display, using 3D visualization of spatially positioned auditory stimuli.

Virtual reality systems are visualization media for interpreting interactive dynamic simulations of physical systems. We propose a mapping that provides a visualization system whose goal is to improve the visual and auditory information available for music analysis.

The various parts of our system span different research domains: music cognition, visual perception and data visualization; in this paper, we focus mainly on visualization issues.

1 Introduction

This paper presents a Virtual Reality (VR) environment where standard music notation is augmented by multiple, 3D-positioned visual and auditory information.

The proposed VR system is based on the idea that standard music notation can be enhanced on the basis of perception-related results in the areas of music cognition, visual perception and data visualization.

Potential uses of this system include music analysis, computer music composition, and music education.

The final target platform for the system is the CAVE, a VR environment developed by the Electronic Visualization Laboratory (EVL) at the University of Illinois in Chicago [1].

Standard musical notation is the only indisputable standard way to visually represent the vast majority of music material.

It can be enhanced in two ways: augmenting the number of ways information is displayed (multiple cues) and augmenting the amount of displayed information (adding extra visualization paradigms). Our system approaches both problems.

Adding multiple visual cues

In a traditional computer-based musical notation system, music is played and simultaneously displayed using a scrolling notation; The user reads chords and melodic lines directly from the staves as they are played. To this standard feature, we add more visual cues:

Chord quality, key and function are presented on a separate table; played notes are highlighted on the staves. Melodic lines are color-coded, scale information is displayed on a different table.

Augmenting the amount of displayed information

These cues can be all presented on a single 2D image representation. While a 2D visualization could be overwhelmed by this amount of data, using the 3rd dimension allows some extra relevant information to be represented. Examples of 3D display possibilities include the following:

Melodic lines and chords can be extruded from the main staff and displayed on a parallel staff plane in 3D notation space; color coding establishes identity with the original source staves.

Staves can be grouped and moved in 3D space from plane to plane or within a staff plane, either manually or using predefined groupings (by instrument, pitch range, spatial positioning in the input data etc.).

Moving staves in different regions of available 3D space causes stereo left-right repositioning in the simultaneously played sound output.

Automatic 3D camera positioning can be set to follow the execution of the musical piece, or blocked to fixed points in the score while staff planes keep scrolling.

2 Potential uses of the system

2.1 Computer-aided music analysis

In the realm of music analysis, there is a multitude of theories that can be reinterpreted as algorithms or other rule-based systems, and thus at least partially implemented in software. One advantage of a

software realization is the possibility of visualizing the outcome of any performed analysis.

In tonal music Form Analysis, it is beneficial to add and visualize extra structures that are only implied by standard musical notation.

As an example of cognition-oriented studies, Lerdahl and Jackendoff's Generative Theory of Tonal Music provides several different *reductions* to "represent hierarchical relationships among pitches in a piece" [3]. These reductions are notated by means of trees. This might be one of the most obvious cases where the outcome of an analysis, obtained by applying a music theory model, is naturally a graphical object - a graph tree structure superimposed on a music stave.

The reductions presented in Lerdahl and Jackendoff's model, have the goal of researching principles of music cognition. In this effort, a flexible and powerful visualization tool is very likely to provide new insight.

Both analysis examples need methods for displaying their outcomes as additional information to the standard musical notation. Drawing a parallel with scientific visualization tools used in fields such as quantum physics and algorithm parallelization, when a massive amount of information is analyzed, sophisticated visualization techniques are beneficial.

2.2 A tool for computer music composition

Another field, where there is promising use for a VR system capable of real-time musical notation visualization and manipulation, is the realm of computer music composition. Since the first distribution of the Music-V language for direct synthesis, a concept pioneered by Max Mathews at Bell Labs in the 1950s (leading to the Music-V in the 60s), powerful algorithmic paradigms have been available to computer music composers. Only recently have the descendents of such systems been equipped with adequate visual interfaces, in the attempt to remove inevitable human-computer interface barriers, and to allow a real-time control. One such example is the SuperCollider software environment, which provides a graphical interface to the underlying sound synthesis language [4].

2.3 A tool for music education

A third field that can gain benefit from an interactive VR musical notation system is *music education*, starting from the introductory level, in theory and analysis classes.

Musical Education exploiting Musical Intelligence

Gardner's Theory of Multiple Intelligences [2] provides a theoretical foundation for recognizing abilities in such different areas as music, spatial relations and bodily control. It is mainly a psychological theory - not a direct educational strategy - but understanding its principles should allow a wider range of students to successfully profit from classroom learning.

An implication of this theory is that different approaches are needed to efficiently teach students whose intelligences are (in Gardner's classification) not musical, but rather visual, logical, verbal or kinesthetic.

An enriched musical notation stimulates mental models about harmony, exploiting visual abilities instead of kinesthetic ones, thus stimulating the visual/spatial intelligence, according to Gardner's model. Methods for displaying multiple visual cues are needed again.

3 Rendering Musical Notation

Mapping musical data to a geometric representation can assume different specific meanings in the presented system. For example, the structure of a Bach's Canon can be mapped in space using the structure obtained from a form analysis. This method does not use standard music notation, but it's been extensively used in interactive explanations of form analysis bases [5].

Another approach starts by adding color to standard musical notation. Coloring musical phrases can help in recognizing similar horizontal structures in the standard musical notation.

When fine-grain analysis tools are used, such as the GTTM method [3], spatial representation can be shifted from the traditional flat 2D view to the 3D space.

The outcome of an analysis such as the GTTM method, being a tree, consists of several levels of detailed information about the same music material. These levels can be layered in 3D to denote the fact that they represent the same input data.

Additional experiments are needed to define the best possible visualization of naturally linear and directed data, when each layer of information hides the previous ones and simple transparency can not be used easily - legibility has to be preserved.

3.1 Multi-Modal Perception and Color-Sound Associations

A system using VR and 3D visualization techniques for Music Analysis should aim at the most intuitive representation of musical data. That includes not only an appropriate use of symbology and geometry, but the use of colors and color mapping as well. Directions about mapping decisions can be searched in multi-modal perception studies, which explore the interactions between auditory and other sensory processing in the human brain.

3.2 System Prototypes

To display standard musical notation, the building blocks consist of canonically defined symbols for notes, staves, embellishment, expressivity notations etc. Few musical notation software packages exist in publicly available source code form; examples are the Vivace and Rosengarden systems - neither one is easily transportable to a 3D library such as OpenGL. Therefore, the first prototypes for the presented system are built using CosmoWorld - an authoring tool for VRML scenes. VRML allows to combine animation, and audio/video information for fast prototyping.

Labels are in the form of 3D-positioned table objects, small rectangles containing a visual presentation of classes in a categorization of music data.

Colors can be used used in pitch-class representations. The choice of a color scale is not unequivocal. For example, when analyzing functional harmony in tonal music, different chord functions denote their *distance* from the tonal center of a piece. A dominant chord is in this case often considered the closest to the root, being the one which most strongly tends to resolve to the tonal center. It could be therefore visualized with a *warm* color, to denote its proximity to the center. However, it can also be considered the most *distant* chord, because of its tensions.

Similar considerations can be made when choosing colors for chord quality classifications. Major and minor chords are often simplistically described as *happy* and *sad*-sounding chords, and in some cultures these two emotions could be depicted using red and purple colors - such a choice is very delicate.

When the two classifications are combined - quality and function of chords displayed simultaneously - two color scales have to be used and even more care is needed to avoid potential perceptual clashes.

4 Conclusion

Previously presented similar systems for 3D music visualization represent musical information using an ad-hoc, nonstandard notation, using 3D objects such as spheres and ellipsoids to represent notes, their resizing to identify played notes in real time, and a fixed space positioning to represent category grouping.

Our system relies on standard music notation as the main visualization of music material. We add multiple visualization cues to enhance this notation instead of sacrificing the user's familiarity. Played notes are displayed in real time using color-coding and highlighting; separate 3D objects are used as information sources to emphasize temporally related material - an undistorted view of music notation is always available. Finally, the spatial positioning system is interactively controlled by the user.

References

[1] Cruz-Neira,C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., Hart, J.C., "The CAVE: Audio Visual Experience Automatic Virtual Environment," *Communications of the ACM*, Vol. 35, No. 6, pp. 65-72, June 1992.

[2] Gardner, H., *Frames of Mind,* Basic Book Inc, 1983.

[3] Lerdahl, F., and Jackendoff, R., *A Generative Theory of Tonal Music,* MIT Press, 1983.

[4] Pope, S.T., *Sound and Music Processing in SuperCollider,* Center For Research in Electronic Art Technology, University of California, Santa Barbara, 1997.

[5] Smith, T., *Bach, the Baroque and Beyond: an Asynchronous Course in Music Theory Via the World-Wide Web,* Society for Music Theory, 1997.

CCARH'S MUSICAL DATABASES ON THE WEB: A GOLD MINE FOR MUSICOLOGISTS

Andreas Kornstädt
Center for Computer Assisted Research in the Humanities
Braun Music Center
Stanford University
Stanford, CA 94305-3076

Abstract

Musicology has entered the computer age. But despite the substantial number of educational applications with pre-fabricated analyses, there are hardly any programs for customized analyses by individual musicologists. The major reason for this is the almost complete lack of computer-readable, high-quality, musicologically valuable musical resources. CCARH was founded to alleviate this situation and has encoded well over 600 pieces of classical music using the Center's MuseData format, complete with high-resolution graphical renditions and conversions to MIDI. It is complemented by the Themefinder database of musical themes. Now, that the data has been meticulously scrutinized over and over again, CCARH has made its databases freely available on the WorldWide Web. Data from both sources can be searched by various criteria that cater to the needs of scholars and laypersons alike.

1 Introduction

CCARH's goal is the development of data resources and software for applications in musical history, theory, analysis, performance, perception and cognition, and in related areas of study in other disciplines. Although several musical resources are available on the Internet, most of them are limited to sound (MIDI, etc.) and printing (PostScript, GIF images, etc.) and none of them combines free availability with extremely high quality and analytical value. On CCARH's newly designed Web pages, two databases now give musicologists what they have long been waiting for.

2 MuseData

The *MuseData* project aims at providing high-quality scores in electronic formats to scholars and music lovers. Since 1984, 640 complete electronic renditions of musical scores from the baroque, classical, and romantic period were created, edited and proof-read by CCARH data entry specialists to meet the highest possible standards. Composers comprise Bach, Beethoven, Corelli, Händel, Haydn, Mozart and Vivaldi. Works include cantatas, chorales, concertos, fugues, operas, sonatas and symphonies. Over the years, *MuseData* scores were used for performances and recordings (among others at *l'Opéra de Marseille* and the *Göttingen Händel Festival*) and for musical analyses by David Cope and Max Matthews.

	Encoded musical works			Derivative output formats	
	MuseData Stage1 (no layout)	MuseData Stage2 (with layout)	*total*	MIDI	GIF
J.S. Bach	150	300	*450*	249	300
Corelli	48	24	*72*	24	24
Händel	48	10	*58*	10	10
Vivaldi	0	24	*24*	0	24
Beethoven	3	3	*6*	2	3
Haydn	71	10	*81*	1	10
Mozart	28	21	*49*	6	21
Telemann	96	2	*98*	96	2
total	*444*	*394*	*838*	*292*	*380*

Fig.1 Musical works in the MuseData database

Although the data has been freely available for years, it was only due to the WorldWide Web interface that it became easily accessible for scholars and performers.

2.1 List of Features

In order to make the data retrievable without contacting CCARH staff members, an extensive body of reference information was attached to every part of each piece of music. This information includes the composer's name and dates, the country of composition, the work's title in the original language and in English, its parent work (if applicable, e.g. *"Das Wohltemperierte Clavier"*), popular title, opus number, scholarly catalogue number, and - if it is a movement - its movement number. Also encoded are the work's genre, stylistic period, the original document from which the electronic score was prepared, the electronic editor and the encoder.

By specifying one or more of these criteria, users can search the database with any Web browser at ***http://musedata.stanford.edu/databases/musedata***.

Matching records are displayed 10 at a time. In addition to a short description, the data formats[1] in which the work is available are indicated:

- *MuseData Stage 1* files contain pitches, durations and bar lines for one vocal or instrumental part of each movement of each work.
- *MuseData Stage 2* files extend Stage 1 and add information about enharmonic spelling, articulation, ornamentation, lyrics, texts and layout (stem directions and lengths, beaming, etc.).
- *high-resolution GIF images* can be printed right away.
- *MIDI files* can be played back on MIDI instruments and personal computers.

After acknowledging CCARH's copyright on the materials, files can be downloaded to the user's computer.

2.2 Technical Background

The database is held in two separate parts: The reference information and the musical data itself. User requests are processed by a search engine (written in C for highest speed) that examines the reference information and produces a list of unique keys and links to the musical data.

2.3 Future Developments

Although *MuseData* is highly expressive and valuable for researchers, there are hardly any *MuseData* application programs outside CCARH that would facilitate analytical work. Therefore a project is under way to translate all musical data into David Huron's *Humdrum Kern* and into Leland Smith's *SCORE* format.

3 Classical Themefinder

While the *MuseData* database contains complete works, the *Classical Themefinder* database holds over 2000 melodic representations of themes that cover the full range of works from the classical period, from Adolphe Adam Efrem Zimbalist and from chamber music to orchestral works. Search results are displayed graphically and with full reference information.

3.1 List of Features

In order to illustrate the numerous ways in which users can specify musical themes or parts thereof at ***http://musedata.stanford.edu/databases/themefinder***, all examples pertain to this theme:

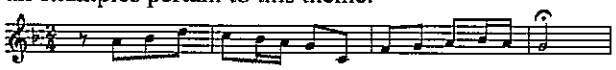


Fig. 2 The first theme from Beethoven's Symphony No. 6

Searches can be done in the following ways:
By pitch ("`A Bb D C Bb A G C F G A Bb A G`", if exact pitch sequence, key and enharmonic spelling are known), by pitch-class ("`9A20A970579A97`" - if exact pitch sequence and key are known but not its enharmonic spelling), by interval ("`+m2+M3-M2-M2-m2-M2-P5+P4+M2+M2+m2-m2-M2`"- if exact interval and key are known) and semitone interval ("`+1+4-2-2-1-2-7+5+2+2+1-1-2`" - if exact interval is known but not the key), by scale degree ("`34654325123432`" - if diatonic pitch sequence and key are known but no chromatic information), by gross contour ("`//\\\\\\////\\`"- if neither exact nor diatonic pitch sequence nor key are known but a rough sequence of upward and downward steps) or by refined contour ("`^/vvvv\/^^^vv`" - as "gross contour" but upward and downward steps can be subdivided into small ("`^v `") and big ("`/\`") steps).

Optionally, the search string can be anchored to the beginning of the theme. It is also possible to treat several consecutive unisons as one item (e.g. the first theme from Beethoven's *Symphony No. 3* ("`5553`" in scale-degree-format) will also show up when searching for "`53`").
The repertoire can be restricted by composer (Adam to Zimbalist), genre (orchestral, concerto, piano concerto, violin concerto, chamber, string quartet, and solo piano), key (key note and mode can be specified separately or set to "`all`" respectively) and meter (numerical or literal: simple, compound, etc.). All of the above elements can be arbitrarily combined.
On the results page, the number of matching themes and the first 10 matches are displayed with complete reference information (composer, work's name, opus number (if available) and genre, movement number, and theme identification) and graphical rendition.

3.3 Technical background

The database contains themes that were compiled from various sources which are all free from copyright. All themes were originally encoded in *Humdrum Kern* format but were pre-compiled into a list of representations that is identical to the syntax of the 7 different ways of specifying themes in *Classical Themefinder* (pitch, pitch-class, etc.). Thus, all queries from the Web interface can swiftly be matched against those representations without time-consuming invocation of several Humdrum commands.

3.4 Future Developments

Due to its simple architecture, the repertoire of *Classical Themefinder* can easily be extended to include an almost unlimited number of themes, provided that they are free from copyright. In addition to increasing the quantity of themes, it is also conceivable to include searches for rhythmic patterns, to provide links to the *MuseData* database to download the whole piece (if available) and to offer MIDI renditions of the themes for acoustic verification.

4 Conclusion

CCARH's two databases on the Web are high-quality resources for the musicological and musically interested community alike. Whether you want to perform an opera, wake up at night with a classical theme springing to your mind or want to analyze the properties of a substantial corpus of baroque music: All you need is at ***http://musedata.stanford.edu***.

References

[1] Eleanor Selfridge-Field, *Beyond MIDI: The Handbook of Musical Codes*, MIT Press, 1997.

The CyberWhistle - An Instrument For Live Performance

Dylan Menzies, David Howard

Department of Electronics, University of York UK
rdmg101@york.ac.uk dmh8@york.ac.uk

Abstract. The CyberWhistle is an integrated approach to a new electronic instrument, with emphasis given to aesthetic and practical considerations. It consists of a *penny whistle*[1] fitted with sensors and electronics connected by cable to a desktop computer, a Silicon Graphics Indy. In contrast to many windcontrollers, the continuous position of the fingers is sensed. Audio is produced on multiple channels using various software synthesis methods, including waveguide modeling. The design of software and hardware have proceeded in parallel, which has helped in the musical unification of the instrument. The word *cyber* is chosen to reflect the welding of a 'natural' form, the whistle, with modern technology, and also, hopefully, the fusing of the player with the instrument.

1 Introduction

Aesthetic Background The development of the CyberWhistle was strongly influenced by the traditional whistle music of Ireland. The effect, on the whistle, of gradually withdrawing the lowest covering finger is to raise the pitch smoothly to the note above. Irish music exploits this technique to the full by combining such shifting *shadings* in rapid succession. The resulting intricate patterns of sound help to compensate for the simplicity of tuning and tone. In the contemporary electroacoustic style, tuning, or at least the ability to create complex hierarchies with pitch, is frequently not very important. More important is the ability to shape the sound in detail. The whistle thus suggests itself as a possible means of performance in electroacoustic music: The movements of the fingers simultaneously provide a rich method for continuous control of musical parameters. Complementary to this is the technique of *tonguing*, in which the player rapidly blocks and unblocks the mouthpiece opening with the tongue tip, affording a way to generate discontinuities.

Note Transitions Even for standard, pitch based, windcontrollers there are good reasons for tracking the position of the fingers, rather than just using switches. The depression of a switch can only effect the sound when it becomes closed, and is unaffected by the speed of closure. In real instruments the internote transitions depend on the closure time profile. This is not only of interest in emulating real instruments, but of general use in creating instruments with rich response characteristics.

The Justified Whistle Many interfaces are possible for measuring continuous finger movements, for instance there exist several *dataglove* types. It is therefore important to identify reasons why the whistle should be retained and not developed into some other form unaffected by the restrictions of acoustic design. The whistle has several excellent ergonomic qualities. The finger holes are wide and can be accurately sensed at the rims, giving feedback on the position of the fingers [2]. Instrument controllers which track movements of free body parts limit force-feedback to that which occurs inside the players body. Conversely, force feedback which requires sustained application of pressure by the player is tiring. The fingers are in view, which helps visual feedback. The mouthpiece is easily used. The blowing feel, or force feedback, of a real whistle is substantially unaltered in the CyberWhistle: Many windcontrollers attempt to emulate reed instruments for which it is not possible to retain the more complex mouth and throat interactions. The visual appearance is uncluttered and pleasing. A more abstract aesthetic comes from the whistle, one of the earliest instruments, being combined with modern technology. The form of the penny whistle is familiar to a great many penny whistle players, many of whom began playing the whistle before moving to more elaborate wind instruments. In the author's experience such players are attracted by the CyberWhistle.

Practical Considerations The whistle provides a very convenient outer shell. The electronics can be mounted on cylindrical former which slides into the whistle bore.(In fact the whistle can be used as normal by withdrawing the the former.) Penny whistles are inexpensive. Six channels of continuous finger movement turns out to be technically reasonable. Certainly a more elaborate finger scheme would need a more elaborate approach than described in the following section.

[1] Also known as a *tin whistle*. The metal bore is cylindrical and contains six finger holes on the up side.

[2] Other wind instruments with open holes such as the clarinet, have much smaller holes.

2 The Electronic Development

2.1 Finger Sensing

The penny whistle player bends a note by rocking the finger to the side of the hole. Once the finger rises more than a few mm it ceases to have any effect. For the CyberWhistle we should like to have good resolution in this lower region, possibly with some resolution at wider separations, to extend the technique. Electric field sensing has a long history of use in electronic instruments dating back to the Theremin [2]. Of the various operation modes described clearly in [5] the *shunt mode* proves to be workable. The finger is earthed to the player, and effectively reduces the conduction across the capacitor-sensor. The metal bore helps to isolate the finger sensors. Circuitry for accurately measuring the change in conduction is described in [5]. For six channels the total circuitry becomes a little awkward.

Light Sensing with LEDs One possibility is to measure the strength of infra-red light reflected from the finger back into the fingerhole. The main drawback is dependence on skin reflectance. Another approach is to measure the occluded ambient light intensity relative to the ambient level detected by a free sensor, **Figure 1**.

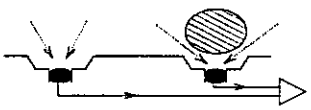


Fig. 1. Compensated ambient light sensor.

The ratio of these quantities is constant to ambient level change, and hence a measure of finger position. Combining a photo diode or transistor with a transconductance amp gives a voltage nearly proportional to light intensity. So, dividing the occluded voltage by the free voltage gives finger position. Unfortunately dividing voltages is not easy. The best option would be to divide with a suitable ADC by supplying the ambient level voltage as the reference voltage, but even this would add significantly to the total circuit complexity.

Light Sensing with LDRs LDRs are can be used as switching devices, for a musical example see [3], but can also be used for measuring light intensity. Miniature LDRs are available from several manufacturers which suit the dimensions of finger holes very well. They are approximated by a pure resistance which under steady lighting conditions satisfies

$$R = AI^B$$

for constants A and B, intensity I, resistance R. If the ambient intensity is I_a and the shaded intensity on the finger hole is I_f, then

$$\frac{R_f}{R_a} = \left(\frac{I_f}{I_a}\right)^B$$

The intensity ratio is constant to ambient light level change, as for the LEDs. So the resistance ratio will also be constant, and so can be used to measure finger position. Producing a voltage related to the resistance ratio is simple using a voltage divider. However, we should like to compensate six finger hole LDRs using one free LDR. The circuit used for achieving this is shown in **Figure 2**.

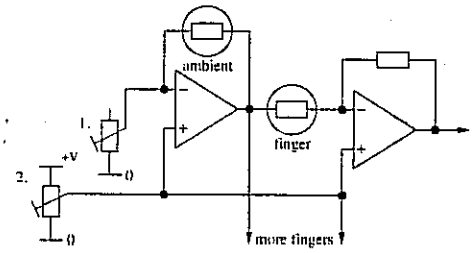


Fig. 2. LDR compensation circuit.

Preset 1 is used to adjust the working distance range. Preset 2 controls the voltage output range. Some variation is encountered in the behaviour of a batch of LDRs. Usually the variations are small, and it is more convenient to correct these in software than in hardware, using a simple calibration test.

Testing An LDR Fingerhole A simple test rig for the LDR fingerhole reveals more strengths and weaknesses. The old term for covering a finger hole, *shading*, is particularly apt because the player can see the shadows cast by the fingers directly. For small source lights the shadows are harder, but the finger control is still continuous because the shadows move across the receptive area of the LDRs. Photo semiconductors have much smaller receptive areas. Most lighting conditions give a suitable variation profile. The greatest voltage variation is near the finger down position, ensuring better resolution in this region when sampled. The circuit can be made to work over a wide ambient light range, but in the lower region the response rapidly deteriorates due to a *cooling* effect in the LDRs. The response to a rapid finger depression is a slowing up near saturation. To work around this, the soft ware can generate a finger down state before the voltage saturates. The finger release doesn't suffer the same delay because the light level is immediately increased.

2.2 Breath Sensing

Breath Noise Detection Two schemes have been implemented for breath control. The first measures the noise generated by the breath passing over a rough plastic surface, using an inexpensive sub-miniature electret microphone of the kind found on computer desktops (See **Figure 3**).

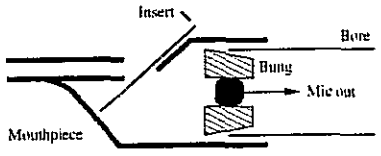


Fig. 3. Using an electret mic in the mouthpiece.

The audio signal is fed directly to the computer audio input and processed digitally as outlined in **Figure 4.**

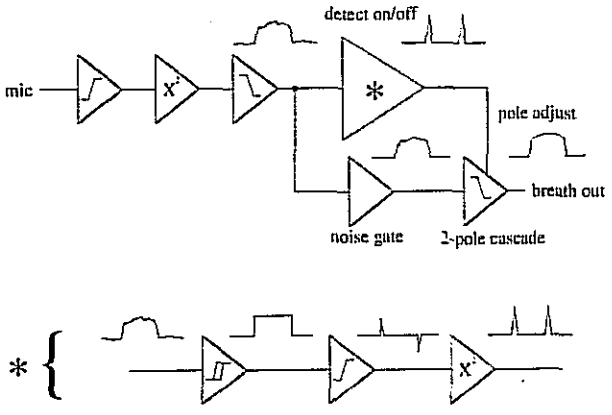


Fig. 4. Mic processing overview.

Tonguing can be followed quickly by detecting the changes to and from low level signals, when the noise generated by air turbulence ceases. In-note signals are filtered to reject noise, while still offering enough response so that vibrato technique can be used. The computational load is moderate but not inappropriate in relation to the high load demanded by the synthesis software. The system suits a low pressure, relaxed blowing style. Generally, windcontrollers operate at higher pressures. The plastic insert helps keep moisture from the mic, and also shields the mic from audio feedback. Breath moisture is carried directly into the air and doesn't build up in the bore. The mouthpiece is easily cleaned by removing.

Breath Pressure Detection The second scheme uses a miniature breathsensor of the micro-silicon bridge kind that have recently become available. This is much more expensive than the condenser mic, but offers wider pressure ranges and direct pressure measurement. The sensor is conveniently mounted in a bung like the mic. The back pressure can be changed by adjusting a block fitted to the mouthpiece exhaust. By blocking the air exhaust completely the player can play without expending breath. The breath pressure is sampled and conveyed by serial link to computer. While it is convenient to combine finger and breath together on one serial line, it is found that lower latency can be achieved for breath control by using audio generated according to the noise detection scheme. A DC block and a 2-pole low pass filter with a cutoff of 400 Hz applied to the pressure sensor signal are effective in isolating a reasonably clean audio signal which can be used, for example, to implement *growling* style effects by modulating the synthesis process with the player's own voice.

2.3 Sensor Sampling

The sensor voltages are sampled and converted to midi controller messages using a low cost microcontroller microcontroller. Additional analog inputs are read via a multiplexer. MIDI provides ample bandwidth for the six finger holes. A more serious problem lies in the desktop software. The processor disruption when receiving the serial information is very high, despite the low bandwidth, and reduces the potential for software synthesis. Various compromises can be made. The microcontroller has dip switches for setting the control message blanking time and the resolution of the sample: The average bandwidth can be reduced, maintaining fast response to sudden changes at the expense of lower resolution. This reduces the ability for subtle expression, for example vibrato. Reducing the message rate to 1 per 4 ms with a resolution of 64 (6 bits) made a useful compromise for most of the software used. The interface electronics and microcontroller have been integrated into a single circuit board which fits inside a Bb penny whistle. An earlier prototype of the CyberWhistle has an external control box, which requires many more connecting wires, and introduces parasitic feedback problems in the high impedance lines.

3 Software Synthesis

3.1 The Programming Environment

Initially, real time Csound, [7], was used to prototype ideas rapidly. However the need for greater flexibility and efficiency prompted a move to C++.[3] The

[3] For example, the audio block size in Csound, *ksmps*, must be set to the order of 100 samples to achieve good efficiency, but this limits the minimum global feedback time and minimum audio latency. Correcting the feedback time is awkward.

object programming style is particularly appropriate for physical models which reflect the hierarchical nature of real objects. It is also useful for the more general abstract structures frequently found in music, possibly because the human mind is biased towards natural structures. The inlining of small audio functions proves effective in improving speed, especially when fully optimized. The core audio code is necessarily compact, so inlining does not produce an excessively large executable; around 100 Kbyte.

I/O, Latency The best response time for incoming midi data is achieved by making a shared process sleep with a Unix *select* command until woken by a midi event. As noted earlier, it is important to limit the midi rate so as not to disrupt the main audio process too much. Using the interrupt method the best latency from sensor voltage to audio output is 10 ms. The audio buffers can be set to a minimum of 256 samples[4]. In normal free running operation with the output buffer full, the audio latency from input to output is then 6 ms.

3.2 Software Synthesis

Physical Models A range of simple physical models were constructed consisting of woodwind and string elements, [6], [1], [4], combined with some greatly simplified models for the finger holes (**Figure 5**). Even with such a minimal arrangement very interesting results can be produced by careful control of the finger holes.

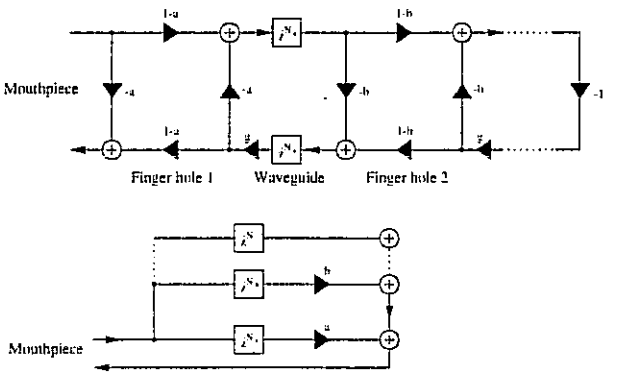


Fig. 5. Simplified bore models.

Relaxing the bore filter and extending the instruments to very low registers is effective in generating a rich spectrum. Different bore tunings lead to contrasting multiphonic effects.

Control Filter Models *Control Filtering* is a broad technique for trying to bridge the gap between traditional live synthesis systems and physical modeling. In the older systems the controls typically map directly into the audio rate synthesis engine. For example, a key triggers an envelope, or breath pressure maps onto depth of frequency modulation. The result is that the output at a given time depends only on the very recent input, and possibly some isolated previous events such as a pedal sustain. In a physical modeling instrument the combination of long delays, feedback and non-linearity ensure that the instantaneous output is dependent on a much broader *history*. The idea of control filtering is to enrich the dynamics of the response to the controls without creating an audio rate physical model. The control signals are filtered, possibly nonlinearly and nonindependently, then upsampled to drive the audio rate synthesis section. For example, take a simple linear filter consisting of DC pass with some resonance at 15 Hz. Apply a finger control through this filter to the index of an FM oscillator. The output ripples when the input is displaced.

4 Summary

The CyberWhistle demonstrates that additional effort expended in controller design may be well rewarded, even when using simple synthesis processes and inexpensive hardware. Likewise, care taken in combining the synthesis process with the continuous control parameters is worthwhile. Frequently too much emphasis is given to the complexity of the synthesis and not the overall instrument. The serial communication method seriously compromises performance, and this is a general feature of pre-emptive operating systems. Using the audio bus as an indirect communication channel is more efficient and offers lower latency.

References

1. Perry R. Cook. A meta wind instrument physical model and a meta controller for real time performance. In *ICMC proceedings*, pages 273–276, 1992.
2. Robert L Doerschuk. The life and legacy of leon theremin. *Keyboard*, pages 49–68, February 1994.
3. Stuart Favilla. The ldr controller. In *ICMC Proceedings*, pages 177–180, 1994.
4. M Karjalainen, UK Laine, T Laasko, and V Valimaki. Transmission-line modeling and real-time synthesis of string and wind instruments. In *ICMC Proceedings*, pages 293–296, 1991.
5. JA Paradiso and N Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, 21(2):69–89, 1997.
6. Julius O. Smith. Efficient simulation of the reed bore and bow string mechanisms. In *ICMC proceedings*, pages 275–280, 1986.
7. Barry Vercoe. *CSOUND: a manual for the audio processing system*, 1992.

[4] Only in 4 channel mode on the Silicon Graphics Indy

Thursday 24th **h. 17.00**

MUSICAL COMPOSITIONS

Listening Session I

Listening Session

Lawrence Fritts

Thought-Forms

Thought-Forms is an acousmatic computer composition created in the University of Iowa Electronic Music Studios. The work is inspired by the early twentieth-century spiritualist movement whose proponents believed that images and matter were physical forms of thought. The musical gestures of Thought-Forms were created from a wide variety of sounds that originated in the physical world. Treated musically, however, these sounds lose their material identity as their physical continuity is transformed into musical thought-forms.

Gianantonio Patella

Canevon

For this composition I developed a plan of work that is based on construction of three sound-layers. The material of final layer is therefore product of elaboration of the two previous layers. In practice after having synthetized the first layer, that is a piece of about 76 second, I prepared an other layer of the same length using as basic material the first layer, and so I did for final layer which expands the duration of the piece until about thirteen minutes, besides in this layer I also used the same file that I was synthetizing as basic material. Synthesis of the first layer has been obtained with a very simple instrument that utilizes algorithm for synthesis of plucked string (Karplus-Strong).

I built the second layer of piece handling, breaking up, distorting the first layer, I wanted in fact to supply some material enough various to modulate the next layer. I synthetized the last layer with 1 only instrument that was fed by a score that I generated by a chaotic algorithm, I used the same methodology also for score of the first layer. Timbre of this instrument can be modified during the time by a sequence of instructions. These can modify parameters from the outside of the instrument. A single note of this instrument can produce 2 sounds that weave on the 2 channels by functions which control the movement of sounds for the whole duration of piece. Algorithm that I used for the generation of score is built by a chaotic function that can generate many values per time, these values are connected between them. I think existence of a strong degree of link between values is truly important, when we utilize them as generators of score. On the contrary we would find us in front of a kind of generator of random numbers. Utilizing this type of function let me obtain 2 advantages, the first one is just interdependence of values, the second one is a kind of structural memory, as a matter of fact this algorithm maintains a kind of memory of values generated before. For a few of these starting values we obtain a memory that we can consider structural, we can actually observe some recurrent microstructures. Besides a few structures present themselves more times interlaced together with other structures. Generator creates the following parameters: starting instant, duration, amplitude of signal and frequency. These parameters are enough to synthetize the first layer because algorithm of synthesis is very simple and static. These same parameters lose importance constructing the final layer, even if they remain basic substance for synthesis.

Characteristic imprint of the piece is formed by a series of instructions modifying: timbre, region of frequency and sonorous intensity. And they are supplied by outer controls, but they can even modify many parameters of instrument. One of these instructions can therefore modify parameters intervening during the starting phase on notes until a new instruction modifies the same parameter. Almost always these instructions modify a parameter changing 2 values because each note produces 2 distinguished sounds, but it is possible intervening in a few cases to modify a single parameter like for example selection of modulating file that can be either the second layer or the third. These instructions can intervene on other parameters, for example: depth of modulation, amplitude of signal, filtering of

modulating signal, multiplier of carrier frequency, temporal region in which is possible to select modulating signal, presence or absence of distortion.

A very important factor for synthesis is the choice of a modulating signal. This is put in relationship to temporal region of modulating file (always introduced by outside instructions) and is besides put in relationship to the frequency of note that we are generating in that instant. For this reason choice of modulating signal is proportional to carrier frequency of note. In this way temporal references of modulating sound-layers are strictly put in relation with frequency parameter and one of the principal parameters that characterize timbre.

Pete Stollery

Onset/Offset

My previous tape piece *Alterated Images* was concerned with the dual interpretation of the word "image" on both aesthetic and sonic levels, Onset/Offset is concerned, even more than before, with exploiting the interplay between the original "meaning" of sound objects and their spectro-morphological characteristics. Thus, there are many recognisable sounds in this piece which can, and should, be perceived on both levels – the sound of a key in a lock is on one level refers to the action of unlocking a door, but on another, is also interesting as a pure sound in itself.

Marco Biscarini

Percorsi

Percorsi is a composition in which the electronic processing of the sound, used in a different ways in the six sections making up the piece, creates a strong musical idea. This musical idea is made choerent by a harmonic area that resounds in the whole piece.

Some chords that reminds the clime from Secon Wien School are inserted in an absolutely original timbral texture.

Two paths, one harmonic and the other timbral, develop side by side towards a unique sound object.

The piece was realized at the Studio di Musica Elettronica of the Conservatorio G.B. Martini, Bologna and in the composer's private studio.

Eric Chasalow

Left To His Own Devices

I was a graduate student at Columbia from 1977 to 1985. I spent much of this time working on tape and instrument pieces at the Columbia-Princeton Electronic Music Center and would occasional take a break from the splicing block to sit in the reception area and catch up with the other the people working at the center, usually composer Pril Smiley, technician Virgil Decarvalo, Vladimir Ussachevsky and, Milton Babbitt. The RCA synthesizer housed at the center was Miltons' instrument. On one occasion, Milton told me that he intended to write only one more piece for tape and instrument – a violin and tape piece that he had promised long ago. It was to have been titled Left To His Own Devices – a fabulous title I thought. Sadly, not long after that discussion, the studio was broken into and the RCA vandalized, rendering it inoperable. With the RCA gone, Milton's piece could not be written.

I have been thinking for a long time about using the computer to precisely control the manipulation of preexisting material – especially material that has very strong associations. It is a challenge to do more than simply quote well-known sources and take a "free ride" on their fame. A composer needs to digest the material and come up with a piece

that reveals its sources, yet through their manipulation and recontextualization has something to say of its own. In Left To His Own Devices, I have combined archival interviews with Milton Babbitt that go back as far as the 1960's with a virtual RCA synthesizer of my own creation. This has allowed me to write music that drows on quotations from Babbitt's instrumental music but to have it "performed" by the RCA. The text is my own composite of phrases that some of us have heard Milton speak many times over the years. In the best tradition of text-setting, I have tried to intensify these phrases by building a dramatic, musical structure both *from* them and aroun them.

Silvia Lanzalone

Intersezioni

Intersezioni, for magnetic tape and performer (1997), is based on the novel of Edoardo Sanguineti "Capriccio italiano" (1963) and takes the expressive contents, the structure, part of the text. The different narrative planes of the novel are made musically like sonorous layers that superimpose or separate, intersect or penetrate.

The text of the novel was subdivided in order to reconstruct the different courses of narration; each of such courses was subsequently subjected to decomposition and to a selection of its more significant elements. I didn't want to tell with music that is already definite and autonomous in the literary text; out of the narrative plot remain the words that are for me more meaningful, the phonemes more pregnant of rhythm and timbre.

The sound matter was build in two stages:

- I played and interpreted the text emphasising its acoustic and expressive peculiarity;
- I subjected the text and its parts to various levels of transformation using digital filters and processes of granulation (System Fly30).

Biagio Putignano

Oscillum

The OSCILLUM is an earthen vase, usually put outside the doorside by the ancient Messapians for protection against negative influences.

Its form is round and flat, and it is empty inside; at the frontside the vase is decorated with anthropomorphous drawings. Oscilla of different dimensions, lumped together on trees or other suitable places, probably gained a magic meaning when waved by the wind. Their sound could contain divine messages of manifold meanings and could bode well to anyone who came across. Some of these arcane objects are being conserved at the Provincial Museum S.Castromediano of Lecce, Italy.

Finally, the peculiar assonance between oscillum and oscillator, the symbolic instrument of electronic music, goes right to the heart of the matter. The composition drew its inspiration from this object. The diffusion of the piece is thought to be in picture galleries or avant-garde art galleries, expositions of contemporary art, etc., as background music, welcoming and accompanying visitors.

The piece starts from a basic concept: the gradual filtering of a white noise in order to obtain pitches (creating euphonic agglomerates). For this reason the whole composition uses only a subtractive synthesis realised with Csound. The piece originates from silence, discreetly, with an exploration of the sonorous field by means of easy flows; the white noise immediately gets coloured and polarises on hinge-frequencies, whose glissando's design simple trajectories. The element of the initial figure is characterised by a group of percussive sounds always obtained by filtering. Around the frequency of 466 Hz other frequencies start to be polarised, forming three harmonies of the ninth of dominant. The continuous narrowing of the pass-band filter, determining the maximum possible filtering , is reducing the noise to a sinusoid. Simultaneously, also the profiles of the envelopes start to diminish until the whole energy of the amplitude of the starting point is stored. This determines the appearance of other percussive sounds on adjacent pitches. These percussive sounds differ from the initial figure by the more pronounced decay in such a way that the same sounds seem to resound in a different architectonic surrounding. The acoustic illusion determines a dichotomy between the two figures. Although the initial figure remains inalterated, the other undergoes a process of

adjustment of the internal components to every single tone. Nevertheless, the different choice of the decay profiles keeps the two groups separated by reason of the described acoustic illusion. The bands that ridge the final acoustic space simulate the trajectories of the initial glissando's, though reduced remarkably in amplitude. The piece finishes with the hinge-frequency of 466 Hz, that fades into white noise. This dissipates all the energetic accumulation caused by the harmonies in the final part of the composition.

Luca Pavan

The Impossible Planet

The idea of this work comes from the homonymous short story The Impossible Planet (1953) by philip K. Dick (1928-1982). The story is a representation of the Earth in a future time after a catastrophic event. The piece wants to recall, in a metaphorical sense, the description of the sounds in the landscape. The whole work is realized with the soun synthesis program Csound. I used only an original Csound instrument and a recorded sound of a female singing voice. This sound has been transformed, filtered and put in different locations in the stereophonic environment. I used the voice after some experiments with different kinds of sound: the voice sounds, because of their characteristics, gave the effect I was searching for producing a wind simulation. To produce the wind effect the voice is stretched many times and the pitch is controlled by a random process. The spatialization has been obtained with the creation of a circular movement of sounds: if the higher frequencies of a sound are progressively eliminated one perceives the sound behind him. I used this technique to obtain a simulation of a wind coming from different directions. In my Csound instrument I have included the control of the number of sound rotations in the stereophonic environment: when I have changed the number of sound rotations it has been possible ti realize a dynamic sound movement. The structure of the piece is divided in two parts: the first, which is only one minute of lenght, is composed with no filtered sounds. The second part is realized with filtered sounds and a large number of sound rotations. The filtered sounds come from a bank of resonant filters in series. The second part is recognizable especially by the presence of a higher pitch of sounds, that is the result of using the bank of filters. Two different reverberation units have been used, one to transform sounds, the other for reverberation of the whole piece. The first used reverberation unit has a control parameter of high frequencies decay in the reverberated sounds. This parameter is controlled by a different function per channel. The second reverberation unit has been used for simple reverberation. The goal of the piece were to obtain dark sounds, evoking a dead landscape: this was obtained especially with time stretching of the voice sound, with a particular Csound unit (Soundwarp, R. Karpern, 1992-97): it can stretches a sound and changes its pitch at the same time, like a *phase vocoder*. So the piece is realized with only a Csound orchestra: this is possible with the Csound powerful features. The piece was realized in the electronic music classroom of the Music Conservatorio "O. Respighi" (Latina, Italy).

Michelangelo Lupone

Controfiato (Counterbreath)
performance on a one dancer's breath

Breath is common to the conceptions of many cultures about the appearance of life.

The first and last breath usually represent the beginning and the end of contact with physical and sensory realty, but also tha beginning and the end of the spiritual quest, the creative act, of conscious participation in the cosmic order. Rhythmical articulation and the various types of inspiration and expiration, emerge from the dense network of emotions and meditation techniques furnishing a constant reminder of the way nature regulates and inscribes the human body.

Controfiato uses these considerations as a starting point for narrating the transformations of a breath through the gestures and movement of a body as it dances.

Starting from a static and inert condition of the body, in which breath represents nothing more than the interior echo life, the exploration of movement and space lead the dancer to self possession and expression; the different positions (lying down, on all fours, upright) modulate the sounds of different breaths and the clashing rhythms

distinguish the spaces occupied. The sound of every breath is added on and interacts with the preceding ones, tracing a musical form which contracts and expands (like a breath which accumulates and cancels itself).

The work was created at the CRM Music Research Center of Rome. A Fly30 system was used for the sound elaboration.

James Dashow

Media Survival Kit
a lyric satire in three parts for radio
text by Bruno Ballardini

Our lives are more and more being determined by interactions with some sort of screen: first there was the cinema, then the tv, now the computer. The latter captures us more than its predecessors by inviting us to take part in an experience much less passive than that based on merely the one way presentation of video images. But this participation is only an illusion: the information and modes of interaction (the "instructions for use") are easily manipulated by others, producing a cultural conformism far more insidious than before... we become digital lotus eaters.

1. NICO' Our hero, a certain Nicò, is recounting a few early memories, he is already captured, hypnotized by his computer screen. The voices are those of the world of informatics which become more and more hallucinatory as Nicò falls deeper and deeper into the digital world. Reality continues to call him, to pull him out of his vortex, but in vain; the further he gets inside, the more reality seems mere light-play.
2. CREMA. (CREAM) As persistent as a virus, the publicity spot appears wherever there's a screen.
3. TUTTI COLLEGATI. (EVERYBODY CONNECTED) The ultimate triumph of the screen: its universe is the Net, the electron flux is all: But as for us, are we really there?

Media Survival Kit was produced for Audiobox, a program of National Italian Radio (RAI), Radio 3:

The voices, in order of their materialization: Alfredo Lombardozzi, Bruno Ballardini, Claudio Bianchi, Lucia Bova, Pinotto Fava.

The musicians: Lucia Bova, harp; Corrado Canonici, contrabass; Paul Goldfield, percussion; Barbara Lazotti, soprano; Paola Buccian, cello; with the special participation of whistler Nicholas Anagnostis.

Elaboration of recorded sounds and electronic sounds synthesis was done using the MUSIC30 system of digital sound synthesis, on the Spirit30 accelerator board for personal computers by Sonitech International.

Digital editing by Giancarlo Grevi, Paolo Antonini, Antonio Giordano.

With thanks to Roberto Carapellucci and especially to Pinotto Fava, producer of Audio Box.

P.S. The composer and the author of the words worked together almost exclusively via Internet; nevertheless, towards the end of the project each verified that the other really existed.

Luigi Ceccarelli

Tupac Amaru (La deconquista, il Pachacuti)
musical opera based on a Gianni Toti's text

This piece was born from a poetic text by Gianni Toti about the epos of Inca prince Tupac Amaru.

The action starts at the rising of indios against the spanish conquistadores, nine years before the French Revolution. The conclusion of the poem takes place, though, in our time, and talks about the cruel ending of the kidnapping at the Japanese Embassy, caused by the present President of Peru. In these epic vicissitudes past and present melt, mingle and get united within the legend of Latin American People's liberation from the conquistadores of all times.

The aim of the composer was that of transforming this fascinating and complex text into a charming and amazing postmodern adventure.

In Tupac Amaru Luigi Ceccarelli tries to convey the expressive force of present music into moulding the words and giving them a strong emotional power, expanding the original meaning of the text. This work was realized at Edison Studio - Rome in september 1997. The sound processing is mainly based on analysis and resynthesis but also on conventional studio techniques used in a radical way. All the vocal parts of the work were realized with Giovanna Mori's voice.

Tupac Amaru was premiered at "Europa Festival" in Ferentino in september 1997: the actress was performing live, with a real-time use of a video camera, and slides projection. This version instead is for solo tape. It is available in two versions: for two or eight tracks tape.

Thursday 24th **h. 21.00**

MUSICAL COMPOSITIONS
Listening Session II

Listening Session

Fabio Cifariello Ciardi

Giochi di fondo

Sounds might be linked with the net of our autobiographical and shared memories in different ways. Many of "Giochi di fondo" sounds and events apply for shared long-term memory traces: rock timbres, rock and reggae rhythms and "licks", concrete sounds that are commonly considered as a part of a recognizable sonic knowledge. Looking at sound events in terms of mental images might be useful to grasp a part of the listener knowledge that might stand out clearly during the listening process, even without being directly related with musical parameters (e.g. a knowledge related with abstract concepts or with emotions). In "Giochi di fondo" I used sounds to explore the abstract concept of game. Within the infinite correlates of this concept I decided to stress the links between the idea of "games" and concepts as "interaction", "confrontation", "conflict".

"Giochi di fondo" is based on "contacts", "fugues", "conquests" and "defeats" among sonic organisms that dwell different virtual spaces. Listener is intended as the main player of a metaphorical and surrealistic "audio game". "Giochi di fondo" derives from the electroacoustic materials of "Games" for double bass, live electronics and quadraphonic tape that has been realized during the summer-autumn of 1995 at EMS (Stockholm). At that time spatialization was realized on the workstation MARS-IRIS controlled with the software written by Davide Rocchesso, Luca Mozzoni e Oscar Ballan of Centro di Sonologia Computazionale (University of Padua). The definitive elaboration of "Giochi di fondo" sonic materials has been realized at Edison Studio (Roma) in winter 1997.

Giovanni Cospito

Le stelle intorno...

The principal idea of this composition, is to use archaic musical material with new informatic tecnolgies. The first material is the speech: I used a poetry of Saffo in ancient greek language. The poetry is seen as a complex sound world: phonems, words, rhythms of metric, melodic lines of prosody, pre-verbarl expression, micro-transition between phonems and so on. The second one is an ambient of several synthetic sounds which are related to some aspects of the first material.

The composition is also structurated about archaic compositional elements: the rhythmic and formal elements are taken from the metric structure of the poetry and the pitch organizzation use a particolar kind of contrapunctum based on the ancient greek harmonies, genders and "nomoi". It is not a simple reconstraction but a particolar start-point of musical elements to construct complex structure possible to immagine today, with that kind of material and with modern tecnological procedures.

Francesco Scagliola

..e organizzar

My work ...e Organizzar arises from a need: to find an original mark which generates the whole immanently. The organization of the Urline, the exploration and deformation of the given mark is surely my most enchanting adventure of these years. To make this I formalized a mathematical model by computer. Therefore, I composed this piece, aided by a software I wrote in MSBasic. The input data to software are: parametric controls as pitch, density, duration and fractal functions tables I use, above all, as probability distributions for composing those several parameters.

I composed ...e Organizzar using the three following models of sound synthesis :

- one instrument FM with double modulating,
- one instrument AM with a very low modulating frequency, approximately from .1 to 3 Hz,
- one granular synthesis instrument.

Francesco Giomi

Agnaby (1997)

Realized at the Sheffield University Sound Studio (UK), Agnaby is an electroacoustic drama freely inspired by the novel "L'Etranger" by Albert Camus. It is also an attempt to get in touch, from a musical point of view, with Arab culture and language.

The material is derived from 1) environmental sounds of the Arab urban life and culture; 2) fragments from Arab traditional music and texts; 3) excerpts from the French text of the book read by Daniel Arfib. The objects, having pre-existing functions and carrying intrinsic associations and meanings, have been processed with different algorithms in order to create different degrees of transformation and recognizability.

The formal structure of the piece is divided into two parts; the first follows the idea of a story taking place in an open and free space, with its noises and its sonic life; the second moves along the idea of a close and narrow environment, the courthouse of the novel, with its people, its atmospheres and its more introverted and predestined character.

Agostino Di Scipio

Studio 97-98
(in crini, legni, fili d'acciaio, voci fuori campo, e risonanze d'ambiente)

Studio 97-98 was conceived as a preliminary study in the process of composing a larger scale work, INSTALL QRTT. The latter is a piece for string quartet, tape and computer processing. Ultimately, Studio 97-98 can be tought of as one of three components in the overall INSTALL QRTT project – the other two being 5 pezzi brevi in crini, legni e fili d'acciaio (5 short pieces in hairs, woods and steel wires) for string quartet, and Interazioni cicliche, alle differenze sensibili (Difference-sensitive circular interactions), for string quartett and interactive computer processing.

Studio 97-98 itself features two musical components: the "out of context" one, i.e. a tape with string quartet sounds and voices, and the "in Context" one, i.e. the live granular computer processing of the tape sounds. The latter is such that it continually changes upon recognition of slight differences between the taped sounds and the sounds come out from the loudspeakers and heard in the performance room or hall. Therefore, the resulting textural sonic fabric is shaped by machine/ambiance interaction, i.e. upon contact with the particular physical place (including the audience) where the performance takes place. The musical pace and density of granular material is then specific to the particular

ambiance, including its own room acoustics, not to mention number and displacement of listeners, quality and displacement of space, as is on the concretization and sensible experience of a real space.

Voices are heard from the tape which whisper short fragments from works by several authors: Goethe's *Metamorphosis of plants*, Edgar Morin's *La methode*, Paul Feyerabend's *Against the method*, and Italo Calvino's *Palomar*.

In the making of this piece, I benefited from the help and skill of the members of the *Prometeo String Quartet*. Thanks are also due to Istituto Gramma, Toni Balthasar, Maria Di Giulio, Lucia Di Giulio and Silvia Schiavoni, for their cooperation. First performance took place at Museo Laboratorio d'Arte Contemporanea of La Sapienza University, Rome, March 1998.

Friday 25th **h. 17.00**

MUSICAL COMPOSITIONS
Listening Session III

Listening Session

Elsa Justel

Au loin... bleu

Clouds of whispers fade away into the blue space, as fugitive wings hastening to a distant horizon.

The verb becomes the material of an unknown language that tries to communicate the sense of something lived, perfumed, something unreal and sensible.

The words, the phonemes, the vocal gestures are transformed into a musical flux enriched by the own colours of each language. The speech is disguised into atmosphere, it expresses its essence fusing with music.

Thousands samples of voices speaking in five different languages were regrouped to model the musical "paste". In its expressive run the language curve transports the articulation of musical objects. As in the oral expression, the music is inhabited by sounds more or less harmonics interwoven with another of non-harmonic character. The speech is underlined and and broken by the respiration and by multiple noises of articulation. We have expoited that natural noises to create our musical material. By means of the spectral analysis of the material we have established the intimate substance of the voice sound. Then we have realized transformations and re-synthesis such as cross-synthesis, convolution, interpolation, etc. In this way we have developed the different phonetic cells into a variety of surfaces and textures.

Antonio Augusto Caminhoto Neto

Paisagens Londrinenses 1

Paisagens Londrinenses 1 tries to represent oniric soundscapes where real and imaginary co-exist, using sounds whose source (natural or syntetic) remains nebulous.

This piece was composed using recorded sounscapes from Londrina city (Brazil): cathedral bells, children playing, cicadas and other insects, birds, storms. The recorded sounds were processed with Csound, wich served also to synthesize other sounds. Processing techniques used were mainly granulation of sampled sounds, linear predictive coding, phase vocoder, subtractive synthesis.

Massimiliano Messieri

Aquilae
for tape

Aquilae is a composition for magnetic tape produced using concrete material only (percussion and friction onto a metal book stand) and synthetic material (saw tooth wave). After changing the enveloping of those two materials they have been divided into two groups according to the sound source. The first group having tonic sounds with continous and iteractive facture with different kinds of envelopes, the second group being made up of nodal sounds with impulsive or continous facture. The diversity of those sound gestures further elaborated with signal amplification in a frequency band, filtering (passa alto, passa basso), enlargenings, compressions, overlapping, trasposition, etc. (made

with IBM compatible programmes, Sound Toolkit, Cool, Goldwave and phase Vocoder), has created between them a cause/effect relationship which has suggested the composition.

Guido Facchini

Invettiva di Aiace

This paper proposes the way in which the author composed this "Invettiva" and his attempt to keep a close relationship between the greek mith of Aiax and the author's way of writing music.

1.The title and the mith

I think that such a title needs more explications. Aiace (Aiax from Locri, not to be confused with Aiax Thelamonious) was one of the greek heroes that burned down Troia. He was going back home with his own fleet, but he sank down with his boat during a storm. He hardly saved himself,swimming to safety.He climbed upon a rock, and from there he shouted abuse at Juppiter and he said that he didn't need any help from the gods to save himself. Of course, Juppiter immediately fulminated Aiax and he falled down into the sea.

2.The material: outer and inner structure

This ancient greek mith has always made me wonder how could it be the last cry of a man shouting abuse to the gods.So the most important material that I employed to build this music is a shout; it was this shout that I worked upon with the computer, until it has grown into an articulated "invective".

Such a vocal material is of course a sounding object without a constant pitch, and the glissando is the most important allure.I tried to articulate this material with different kinds of stretching. I made use of SOUND TOOLKIT, COOL EDIT and PHASE VOCODER, and I combined the results to obtain a kind of continuos rallentato.

3.Acknowledgments

I must thank Lelio Camilleri, my teacher in Bologna, who gave me the wai to keep my musical ideas in touch with the technology I could made use of, and Trevor Wishart, whose music is and will remain a font of inspiration for me.

Riccardo Dapelo

Sul Cuore Della Terra
(On earth 's hearth)

Ognuno sta solo sul cuor della terra
trafitto da un raggio di sole...
(Everyone stays alone on the hearth of the earth
pierced by a ray of sunlight)

These verses by Italian poet Salvatore Quasimodo can describe completely the sensations felt in the incredible and wild countries of Barbagia, Baronia and Supramonte (Sardinia-Italy). Often the most interesting phenomena of human experiences occur in borderlands, lands only superficially scratched by process of History. The isle of Sardinia is one of this lands, an extreme land, in which flows something ancient and misterious, primitive and solar. Is sufficient to go away a little from the coasts full of tourists, to find oneself, unexpectedly, in a "natural harmony", "On earth's hearth".

And in the same natural harmony, rough and wild, are born also the voices of "Tenore de Orosei" an ethnic poly-vocal group, whose voices are the starting point of this piece. Every sound of the entire work is derived from the manipulation (re-synthesis, cross-synthesis, convolutions, granular time-shifting and so on) of the above mentioned voices plus several samples of water and Launeddas (a typical bagpipe of Sardinia).

The aim of the piece is the reconstruction and the interpretation of a sensorial experience of the Sardinia, rejecting however any musical description but rather searching for an ancient feeling *archetype (Ognuno sta solo... Everyone stays alone...*)

Laura Bianchini

Aura, for electronic sounds (1996) (stereo playback version)

The piece is based on a text written by Sandro Cappelletto, inspired to the myth of AURA.

The piece, interely created with processed vocal sounds, evolves around a theatrical action of metamorphosis. The sound interlacing derives from the dialectics among the sound elements of the language which mantain their sense, and the same elements transformed in a way to become sonds with their "own" sense.

Breath, whispering, stammering, gradually transformed, are considered complex sounds materials as well as dramaturgical components.

The sounds are moved trough the acoustic space defined by the loudspeakers generating a perception of movement or localization.

For vocal sound processing computation algorithms designed by the author for the Fly30 system have been utilized. The piece was produced and realized at CRM – Centro Ricerche Musicali of Rome.

Friday 25th **h. 21.00**

MUSICAL COMPOSITIONS
Listening Session IV

Listening Session

Diego Garro

Voci dall'Aldiquà

The title 'Voci dall'Aldiqua' is a play on words that means approximately 'Voices from the Material World'. 'Aldiqua', as a matter of fact, usually means 'on-this-side' but in the context of the title the it is meant to contrast the word 'Aldila', which is one among the thousands of different ways italians refer to heaven or, broadly speaking, to life-after-death. Since the title of the composition may be misleading it is probably worth an initial remark.

Many composers and performers in the past, distant or recent, have already explored the possibilities of various utterances achievable by the human vocal apparatus. This body of research has been put into music-composition in various artistic context. Moreover, during the last twenty years the vast world of computer processing applied onto vocal sounds has also been widely investigated and experimented on. In 'Voci dall'Aldiqua' my intention was to use some particular transformations onto a limited set of vocal sounds as a triggering tool to develop a musical discourse. Thus my first concern was the creation of a coherent flow of musical information in which sometimes the vocal component is incidentally predominant and sometimes it is merely marginal. The word 'voices' mentioned in the title, therefore, has to be read in a broad context. Along the piece, sometimes the 'voice' is human voice, clearly recognizable, either speaking or singing. Sometimes it is human voice deeply transformed, almost unrecognizable. Some other times the 'voice' is, methaphorically speaking, the echo of the material world responding to the human call. The finale features the triumph of human voice singing an angelic chant which, nevertheless, dissolves in a sort of self-decomposition: does transience deprive human attributes from beauty and spirituality?

Roberto Doati

IV Felix Regula

Felix Regula is a work commissioned by and realized at the Centre de Recherches et Formation Musicales de Wallonie in Liège. When I received the invitation to realize a new piece with instruments and electronics it has been natural for me, living in Padova, to think to Johannes Ciconia (1340-1411). Not only because the great composer and theorician from liège lived his last years in Padova, but also for the deep interaction between science and music there is in his life and work. As a composer working with computer since long time, I developed a musical thought shaped on this new technology. As technology I do not simply mean here the machine. I aam referring to the technology as an ensemble of new scientific procedures to investigate and transform the nature (tekhné=Arts and Crafts).

The "nature" to be transformed is a virelai by Ciconia (Sus une fontayne) which represents for me an archetype of the interest many composers had and still have on mirror games. So in the five different versions of the piece I realized, I broke and rebuilt the form of the Ciconia virelai with musical instruments (violin, flute-flute in G, clarinet-bass clarinet-double bass clarinet) mirroring not only each other, as in the music of the past, but also in my preferred mirror: the computer technology.

The computer transformation of the instrumental sounds are therefore conceived as a sort of double of each instrument, but differently disposed in time according the esprit de géométrie peculiar of Ciconia's work. The instruments are also acoustically treated, as the original pitches of the Ciconia's song are changed as concern the modalities of their emission using instrumental contemporary techniques (slap, tongue ram, multiphonics, etc.).

Michele Brugnaro

Epigenetic Landscape n.1

Epigenetic Landscape n.1 finds its originating stimulus in the desire to join together different and seemingly independent operative conceptual fields in a coherent whole provided with homogeneous structural properties. The morphogenesis of the piece seems to be characterized by a series of contrasting forces which interact dialectically and eventually coalesce to define an unitary, comprehensive overall picture. The composition project has taken place from two primary needs which form a first basic opposition:

i) "Abstract" predetermination of the generative algorithm of the formal-syntactical units (intermediate-superior levels);

ii) "Concrete" raw material used as a generative nucleus in the microlevel.

The work has followed two directed paths that are to be considered at the same time parallel and complementary. The activity inherent in the preliminary drafting of the project has concerned mainly the definition of the treatment of the concrete material according to the new conceptual dyad:

i) "Denaturation" and consequent unidentification of the original source, together with the possibility both of altering interactively the timbral and textural behaviour of the sounds and of creating grouping variations in compound sonic objects (density, prevailing colorations, etc.);

ii) Preservation of some pertinent features of the original sound objects, such as, for example, their microrhythmic organization. The samples that have been chosen as the original raw material are essentially recordings of African and Asian percussive solos. Their particular (and interesting) internal structuration can be considered as a sort of matrix from which complex durational patterns are derived, which show both periodic and aperiodic aspects.

a) PERIODICITY:

i) Structural internal attributes of the sound objects that are based on periodicity, such as rhythmical patterns already present in the samples or arising from the automated modifications of the single sounds (accelerando/rallentando effects caused by transposition);

ii) In the interactive treatment of more complex sonic textures some other periodic features are discernible through the particular use of delay lines ["Pseudo-Imitation" generated from delay lines with long time values or sound "coloration" (choral thickening, "comb" effects) obtained by very short delay times].

b) APERIODICITY:

i) Aleatoric fragmentation of the pattern with random combinations between contiguous fragments. But conversely some coherent, cohesive images may be extracted from few basic nuclei;

ii) Casual variations in the temporal distance between fragments, only tendentially determinate, together with the control of the "density" of events and of the offset time within the sample.

The definition of a casual-reading algorithm of the sonic samples with all the potential timbral transformations offered by the different types of treatment [ring modulation, delay lines with different time values, from very short (filtering/ "comb" effect) to very long (imitative, rhythmical effects)] has been coupled with the decision of creating a control panel with real-time modifications of the parameters. Another conceptual pair is needed to fix the strategies involved when defining complex formal objects:

i) Ever-changing automated creation of the sonic organisms of the microlevel;

ii) Option of intervening, with relative and immediate feedback, by modifying in real-time the parameters which regulate the global behaviour of the algorithm, interacting stochastically with the computerized procedure.

The problem is to achieve a good compromise between the "predetermination" of the algorithm and the capacity to "sculpt" the sound directly in real time, imposing over the generative automatism the composer's subjective will, in some way regaining the expressive freedom already present in the original percussive sounds.

The piece, in its final form, is the result of a mixing of four textures generated from a performative improvisation guided by the feedback traces of the internal development of the materials. In substance this textural growth oscillates between

i) a maximum density state (condensation);

ii) a minimum density state (rarefaction);

iii) dynamic variations;

iv) relative expansion within the diastematic space;

v) different locations in the stereophonic field;

vi) changes in the perceived distance.

Zack Settel

Punjar

Punjar is a work for solo soprano saxophone and live electronics. The electronics are used to: (1) expand the timbral range of the instrument, (2) allow for the possibility of self accompaniment, providing additional "ensemble voices" in the musical structure, based on material played by the soloist. Almost all of the electronically produced sounds are initiated and/or modified according to the material played by performer. Finally, an important underlying idea for this piece, "an ensemble controlled by one player", is inspired by John Cage's work in his Sonatas and Interludes for Prepared Piano (1946-48).

Contents

Musical Informatics: Expression and Performance Analysis II

Musical Systems and Computer Assisted Composition

Music Analysis and Cognition

Music Archives

Restoration of Audio Documents I (Special Session)

Restoration of Audio Documents II (Special Session)

Poster Session I

Poster Session II

Demos

Musical Compositions

Listening Session I

Listening Session II

Listening Session III

Listening Session IV