procedings

melbourne - july 6 - 8

australasian computer music conference

form space time

:music architecture & design ACMC 2002

Proceedings of the Australasian Computer Music Conference 2002

6th – 8th July 2002

Organising Committee:

Roger Alsop
Paul Doornbusch
Lawrence Harvey
Tim Kreger
Peter McIlwain

Paper Reading Committee:

Roger Alsop
Paul Doornbusch
Lawrence Harvey
David Hirst
Tim Kreger
Peter McIlwain
Gordon Monroe

Published by the Australasian Computer Music Association, P.O. Box 284, Fitzroy, Victoria, VIC 3056, Australia. May 2002.

Welcome from the Organising Committee

We are extremely happy and honoured to welcome this delegation to Melbourne. It is with great anticipation that the ACMC 2002 Organising Committee expects to share the exciting work of the ACMA membership, and, indeed, practitioners of computer music worldwide. It is your efforts and creativity that makes the ACMC what it is: the region's largest and most important showcase of state-of-the-art work, both artistic and scientific, related to music and technology.

From time to time, advancements in technology make possible new achievements in our area of endeavour, or make some things more easily possible. We are at such a time with sound spatialisation. Music has always had a function in space, indeed there is no such thing as non-spatial hearing. While electronic and computer music has been concerned with space and spatialisation since its inception, the proliferation of consumer-format, multi-channel sound, and thus the ready availability of tools and systems to work on, has brought new emphasis on this area in recent times. This recent emphasis has again led composers who use technology to look at form in terms of space. The conference theme of "Form, Space, Time: Music Architecture and Design" makes this connection, and another important one with the discipline of architecture and design, which has always approached the concept of form in terms of space. Many papers and discussions in the conference will take place around this theme, including some significant international work, and we hope you find this stimulating and enlightening.

We wish to thank the authors for sharing their exciting work and thoughts, and also all delegates. It is because of you that we have this community, and it is through the efforts of the community that we are able to support each other, learn, and grow this field we all love.

This conference would not have been possible without the close cooperation and support of RMIT University and the Victorian College of the Arts. We wish to thank them for their assistance, and at times indulgence, to allow this conference to proceed. Coordinating the process was only possible through the enthusiastic and untiring efforts of all concerned and I would like to personally thank everyone involved.

We wish you a wonderful time at ACMC 2002, and we hope you also enjoy your stay in Melbourne.

Welcome!

Paul Doornbusch, for the ACMC 2002 Organising Committee
Roger Alsop, Lawrence Harvey, Tim Kreger, Peter McIlwain

Table of Contents

Fully-refereed papers

Author	Title	
Steve Adam	Control and Mapping Strategies for *HybriD* ……………………	1
Ros Bandt	Spatial Counterpoint as a Design Principle in the Australia Gallery, The Melbourne Museum ………………………………	7
Jim Barbour	Applying Aural Research: the Aesthetics of 5.1 Surround …...	17
Andrew Brown	Opportunities for Evolutionary Music Composition …………..	27
Paul Doornbusch	The Application of Mapping in Composition and Design …….	35
David Hirst	Developing Analysis Criteria Based on Denis Smalley's Timbre Theories ……………………………………………….	43
Ian Kaminskyj	Multi-feature Musical Instrument Sound Classifier with user determined generalisation performance …..……………………	53
Ian Kaminskyj & Mark Williams	Web Based Automatic classification of Musical Instrument Sounds …………………………………..…………….….	63
Andrew Lyons	Abstractly Related and Spatially Simultaneous Auditory-Visual Objects …………………………..……………………...	71
Peter McIlwain	A Survey of Software Designs from the Sonic Art Group …….	81
Johnathan Mustard	Correlating Movement in Space to the Parameters of Sound	91
Timothy Opie	Granular Synthesis: Experiments in Live Performance ……...	97
Ian Stevenson	Spatialisation, Method and Madness Learning from Commercial Systems …………………………………….………	103
Lindsay R. Vickery	The RoboSax Project (1991-2001): forms of performer/machine interaction in works by Johnathan Mustard and Lindsay Vickery …………………………………..	113

Table of Contents

Non-refereed papers, reports and artist-talks

Author	Title	
Ros Bandt & Ian Mott	Report on the Australian Sound Design Project ..………….	121
Warren Burt	Developing and Composing with Scales based on Recurrent Sequences ………………………………………..	123
Hannah Clemen	Enhancing the Experience of Music-Ritual through Gesturally Controlled Interactive Technology ………………..	133
Robert Coburn	Composing Space: The Integration of Music, Time, and Space in Multi-Dimensional Sound Installations …………….	143
Angelo Fraietta	Smart Controller – Artist Talk ………………………………….	149
Elsa Justel	Space as a Structural Function in electroacoustic Music …..	153
Lissa Meridan	Elastic Horizon: Mapping Collaborations …………………….	165
Robert G. Morgan	Metamorphosis as a Musical Algorithm ……………………...	167
Andrei Bajurnow & Alexandra L. Uitdenbogerd	Towards Audio-Based Music Information Retrieval in the RMIT University MIRT Project ……………………………….…..	175
Lindsay R. Vickery	The Yamaha MIBURI MIDI jump suite as a controller for STEIM's Interactive Video software Imag/ine ...……………..	181
Ian Whalley	The University of Waikato: Studio Report 1998-2001 ………	189

Control and Mapping Strategies for *HybriD*

Steve Adam
School of Music – Conservatorium, Monash University
email: spectro@ozemail.com.au

Abstract

HybriD is a collaborative work in progress between contemporary music group re-sound and the author for multiple acoustic instruments and live computer processing and synthesis. This article outlines some of the strategies explored and methods employed in the development of software to explore the creation of "hybrid" sounds based on the combined elements of multiple performers' musical gestures.

1. Introduction

HybriD is a collaborative work in development between contemporary music group *re-sound* and the author, for multiple acoustic instruments (2 saxophones, 2 clarinets, violin, electric bass), MIDI Percussion instrument and live computer processing and synthesis. In many ways, this piece builds on the ideas initially formed during the composition of *Nexus,* which was commissioned by the contemporary performance group *Australysis* in the mid 90s. In that piece, the interpretation of MIDI-based performance data from three different performers was gradually merged to create a flavour of 'Hyperinstrument' (Machover and Chung 1989). With the given advances in computer processing power and software design, new methods of merging and mapping real-time performance data become available and some approaches are explored in this piece. Additionally, the resources offered by a relatively large ensemble offer multi-dimensional control of synthesis and signal processing parameters to achieve an electro-acoustic sound world with the potential of great depth and complexity, much of which is offered by the detail and nuance provided by an instrumental performer.

Development of the software and writing and refinement of the piece has been an ongoing affair. Individual workshops with ensemble performers were held during March and April and workshops with the entire ensemble commenced as of May 2002. At the time of writing, the ensemble has had the opportunity to 'play with the piece' in a few of sessions. These sessions have already provided valuable feedback on the efficacy of various mapping strategies and have served to prompt further development of many aspects of the software.

In its current form, the piece is realized with Max/MSP object-oriented programming software running on a 550MHz Apple Titanium Powerbook. Audio I/O is provided via a Mark of the Unicorn 828. Though a contemporary system, its performance benchmarks are relatively modest and the issue of maximizing CPU efficiency is an ever-present consideration in the design of the software.

2. Audio Analysis and Derived Control Sources.

Given the reliance that this work has on instrumental performance as a means of control, methods of tracking and representing performance parameters through an instruments audio signal are an important factor. Though a detailed investigation into this area is beyond the immediate scope of this project, a brief description of the means employed is appropriate.

The audio signal of each acoustic instrument, captured by instrument microphones or 'bugs', is fed to an analysis stage in the *HybriD* program (a Max/MSP patch) as a sampled audio signal. The MSP object *fiddle~* (Puckette et al.1998) is employed to track real-time pitch and amplitude characteristics of each signal. *fiddle~* is freely available, fairly efficient, and above all, capable of providing generally useful results for the purposes of this work. Each analysis stage generates an output in the form of discrete notes (in MIDI note format) and continuous pitch and amplitude outputs at the Max scheduler (or control) rate. Audio signals from each of the instruments are also available as sources for routing to real-time signal processing algorithms. Coupled with the MIDI performance data provided by the ensemble's percussionist via pads and data from the computer operator via a fader based MIDI controller and computer keyboard, the outputs of the analysis stages make up the available complement of externally derived control sources.

Drawing on software previously developed with Max for the composition *Nexus,* some feature detection in the form of performance data (MIDI note information, duration and inter-onset time stored in cyclic buffers of arbitrary dimensions) is also undertaken. This data is used to track tempo in a rudimentary manner and generate continuous control sources intended in part to facilitate experimentation with controller mapping.

Finally, some control sources are generated internally, such as Low Frequency Oscillators (LFOs) and breakpoint envelopes. Numerous parameters of these sources can be controlled by the performance control sources described above. For example, the rate and amplitude parameters of an LFO can be modified by the pitch and amplitude control values generated by a performer.

3. Mapping Strategies

Within the context of this work, emphasis is placed on utilizing the resources of the ensemble through the exploration of interactive networks that can be formed between performers and the resultant sound(s) they may produce. The ability to define a path of influence for each performer's input to the system independently is necessary, and consequently, a practical method of coupling performance features (control sources) to the processes that determine the sonic results (synthesis parameters) is one of the first considerations.

Using Max/MSP, the routing of control and audio sources to a variety of destinations is most readily achieved by using a number of switching matrices similar in concept to the patching panels popular in analogue synthesizers. An arbitrary number of states of each matrix can be stored and recalled for later use. In its current guise, *HybriD* uses two matrices, one to route control sources and the other audio.

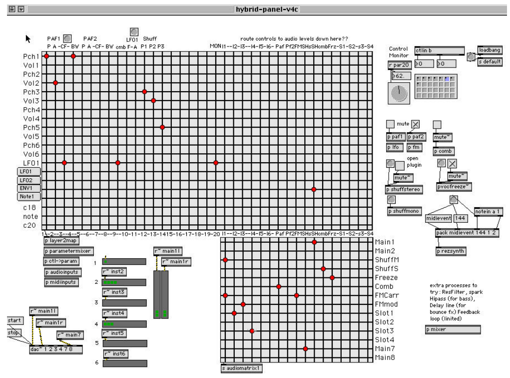


Figure 1. A view of the *HybriD* Matrix Panel showing the larger upper control routing matrix and the lower smaller audio routing matrix.

The two matrices are positioned in the way that they are so that control sources can be routed to effect audio channel parameters, such as volume and panning. Aside from the fact that legible representations of these matrices take up a lot of screen space and become difficult to navigate as larger numbers of control sources and destinations are accommodated, their use is straightforward. Those functions that are less frequently called upon in the development of the software can be assigned to the matrix via a menu when required. This enables a larger number of potential control sources to be available than the dimensions of the matrix can accommodate, albeit not all simultaneously.

The issue of interconnection becomes one of greater significance when considering the suitability of the control source to its intended destination. As an example, *fiddle~* provides pitch data in the form of a floating point MIDI Note value where each equal-tempered semitone is represented by a value of 1.00 (for 100 cents). In this form, the typically encountered pitch range of a violin is conservatively around 36 (3 octaves) commencing from 55 (G below middle C). If this value range were used to control the gain of an audio buffer which would typically operate in floating point values between the range of 0 to 1, we would have a massively distorted signal as a result.

From this example, questions of how a particular pitch range as extracted from a performance should affect the amplitude of a sound (or even why this should be done at all) arise. Removed from the context in which these pitch values are derived and placed in a context where they have little if any meaning, there can be no single correct response. Nevertheless, the facility to arbitrarily map a source to a destination is an implicit requirement of the structure.

Setting aside the consideration of linear and logarithmic characteristics inherent in various control sources, an initial solution to the problem of managing extremely large (or small) values is to scale all control sources and destinations so they operate within a standardized range. Doing so enables any control source to be coupled to any synthesis or processing parameter with confidence that the results will fall into an acceptable (if somewhat arbitrary) range.

Within the domain of research into performance-based synthesis control there is a general consensus that simple one-to-one mapping of control sources to synthesis parameters rarely yields the most useful or interesting results (Hunt et al. 2000) (Garnett and Goudeseune 1999). (One-to-one mapping is implicitly illustrated in our earlier example where the detected pitch of one instrumental source is used to control the amplitude of a - possibly unrelated - sound source but imposes no other influence in the way the resulting sound behaves.) Comparing the dynamic interplay of sonic qualities of acoustic instruments or, more appropriately, aesthetically appealing synthesis algorithms, to synthesis algorithms that employ only one-to-one mapping quickly bears this observation out.

Enabling the capacity to merge control sources (many-to-one) of variable weights or alternately, having numerous synthesis parameters each affected by differing amounts in response to the same control source (one-to-many) raises numerous questions regarding how such a system might operate and what implications this has for the architecture of the system.

The use of an intermediate mapping layer as proposed elsewhere (Wanderley et al. 1998) proceeds by dividing the mapping scheme into independent layers that more readily enable the expressive use of different input devices (acoustic instruments or MIDI in this case) to control a variety synthesis or processing algorithms. This proposal serves as a model for implementation within HybriD.

A model for the interconnection of control sources to parameter destinations grows in complexity and inefficiency as new, combinatorial possibilities, transformative processes and special cases are added. On the other hand, a model that can deal with the most frequently required mappings enables it to be used often and the occasion of conditions for which unique treatments need to be devised are encountered less frequently if at all.

As it currently stands, the scheme is fairly primitive, however one-to-one, one- to-many, many-to-one and many-to-many controller mappings are, to a limited extent, available. This is achieved by way of routing sources to a secondary mapping layer or control mixer. The number of control sources in any control mixer is currently limited to four, though an additional mixer can be used to increase the number of control sources that can be combined if they are required. A similar arrangement is available for distribution of one control source to multiple synthesis parameters.

One of a number of problems that arise with the current implementation is that the control mixing connections made are not reflected in the visual state of the connection matrix. In order to maintain consistency with this representation, a more elaborate system of routing control sources is planned, but not yet implemented. This would most probably require the ‘tagging’ of control sources so that they could be merged into a single data stream (i.e. the matrix buss to which they are assigned) and identified by their tag at the receiving end.

There are a number of possibilities by which two or more different control sources may be combined. Varied weighting between different control sources by setting a ‘tap point’ along a linear or logarithmic curve between two or more states similar to the way in which a panning control in an audio mixer operates is one method. This is equivalent to the addition of two weighted values. The use of one control source to determine the range of another by multiplication is also possible. Both methods are currently in use in HyBriD, however further refinement and generalization of the treatment and combination of control sources is left as an area for future work.

4. Signal Processing network

The audio system provides initial sources in the form of the captured instrument audio signals and a number of interchangeable synthesis algorithms. Signal processing elements, busses and output channels, can accept inputs from all sources (except their own outputs in most cases) via the matrix. Initially, the algorithms used were all rather complex but after some experimentation, it has been found that using a larger complement of simpler processing elements enables finer structural control of the algorithms, better integration between the control sources to the sound production process and reduces the need for interpolation of multiple parameters simultaneously (as is the case with more complex algorithms). The only real disadvantage of this approach is the requirement for a greater number of interconnections in the audio matrix arises which leads to inefficiency and conceptual difficulties.

Some of the synthesis algorithms used include simple oscillators, additive (oscillator bank) and formant synthesis. Signal processing algorithms in use include various types of filtering including filter banks, frequency modulation, amplitude modulation and spatialization, delay lines and comb filters, granulation and looping among others. As with the control matrix, some of the less frequently used elements are relegated to channels from which they can be selected from a pool of processes and enabled when a signal is routed to them. This serves to keep the audio matrix to a manageable size and reduce processor overhead by switching off unused components that would otherwise consume valuable CPU resources.

The process of determining successful mappings in this work is largely as the result of experimentation during workshop sessions with the ensemble or, more frequently at this point in time, by simulation of a performance via playback of pre-recorded fragments of the performers in action.

While the potential exists to map a large range of sources to destinations, resulting in a huge range of permutations, it is likely that only a relatively small subset of all possible mappings will be effective for the stated aims of the work. The general pattern observed is that better results are achieved by maintaining a balanced distribution of the influence of the performers on the synthesis process. This balance becomes more difficult to manage as the number of contributing performers increases toward the full complement of the ensemble.

To illustrate the kinds of mappings possible, a practical example is provided: A cluster of sine tones is generated using an additive synthesis algorithm. The pitches of two instruments can be used to independently determine the lower and upper boundaries of the oscillator frequencies while the overall volume of the cluster may be controlled by a third player. A fourth performer's input could be used to influence spatial disposition of the resultant sound across a multi-channel sound system. In this example, many performers are involved and simple one-to-one mapping of control source to synthesis parameter is predominant. By reducing the number of performers (and consequently, control sources) the scenario changes: If the pitch of one performer were used to control the lower frequency boundary, while a function of pitch and amplitude (pitch, scaled by amplitude) from the same performer were used to determine the upper frequency boundary and yet another variant of these two control sources (a weighted sum of amplitude and pitch) was assigned to control the volume of the resultant sound, both one-to-many and many-to-one mappings arise.

Aural results that reflect a high degree of expressive reaction to the performers' sonic gestures are the primary goal, though there is no real reason to reject mappings that don't strictly fulfill this requirement. In a few cases at least, subtle processing of instrumental audio signals makes up the entire output of the system. Processing parameters may not react in any way to the performers' gestures but simply process the audio signals in a fixed manner over a given period of time.

It is often the case that a particular mapping yields promising aural results from a composer's (or listener's) perspective, while the performers' sense of their affect on the resulting sound may be counter-intuitive or even difficult to determine and thus prove less engaging for them. As the mapping is likely to mutate frequently in a developed performance of the work, each performer's relationship to the sound will change with each mutation, offering performers little opportunity to get comfortable with any particular mapping.

5. Future Work

Ongoing development of the *Hybrid* software offers new possibilities and refinements, but doing so between rehearsal sessions with the ensemble does have negative implications for development of the work overall. It is expected that the further development of the software will be suspended in the near future so that a greater sense of stability ensues and mappings that yield results that are consistent from one rehearsal to another are possible. It should be pointed out however, that even when mappings are totally reproducible from one session to another, there is no guarantee (or expectation) that the aural results will be identical due to variations in performance, bug placement and any number of other variables.

At this time, no tracking or estimation of the timbral or spectral qualities of the captured sounds is used as control sources. This limitation is in part the result of CPU efficiency considerations as tracking and processing six instruments in real-time (among other DSP tasks) places fairly heavy burdens on the processor. Future work with a single performer or a small group will provide greater scope for exploration in this area.

In moving between different mappings the structure of the program actually changes at both the control and audio level. While gradual alteration of states in the audio matrix is possible, control mappings are switched from one state to another more or less instantaneously. It is not anticipated that this limitation will be easily resolved, though possible methods of interpolation between different states need to be investigated.

In its current form, complex mapping is a tedious affair. It is anticipated that further development of practical and effective mapping strategies of performance-control sources to synthesis parameters as outlined in this article will be enhanced by the experience gained from the work on this project.

6. Acknowledgments

Thanks to Dr Thomas Reiner and the performers of *re-sound* for their generous support in the realization of this work. This project has been assisted by the Commonwealth Government through the Australia Council, its arts funding and advisory body.

References

Garnett, G. E. and C. Goudeseune. 1999. “Performance Factors in Control of High-Dimensional Spaces.” *In Proceedings of the 1999 International Computer Music Conference*, San Francisco: International Computer Music Association. pp. 268-271

Hunt, A., M. M. Wanderly and R. Kirk. 2000. “Towards a Model for Instrumental Mapping in Expert Musical Interaction.” In *Proceedings of the International Computer Music Conference*, San Francisco: ICMA.

Machover, T., and J. Chung. 1989. “Hyperinstruments: Musically Intelligent and Interactive Performance and Creativity Systems.” In *Proceedings of the 1998 International Computer Music Conference*. San Francisco: ICMA

Puckette, M. S., T. Apel, D. Zicarreli. 1998. “Real-Time audio analysis tools for Pd and MSP.” In *Proceedings of the 1998 International Computer Music Conference*. San Francisco: ICMA. pp. 109-112

Rowe, R. 1993. *Interactive Music Systems*. Cambridge, Mass: MIT Press.

Wanderley, M. M., N. Schnell, et al. (1998). “ESCHER - Modelling and Performing Composed Instruments in real-time.” In *Proceedings of the 1998 IEEE International Conference on Systems, Man and Cybernetics* (SMC). Pp 1080-1084.

Weinberg, G. and S. L. Gan. 2001. "The Squeezables: Toward an Expressive and Interdependent Multi-player Musical Instrument." *Computer Music Journal* 25(2): 37-45.

Spatial Counterpoint as a design principle in the Australia Gallery, the Melbourne Museum.

Dr Ros Bandt

The Australian Centre, The University of Melbourne

email: r.bandt@unimelb.edu.au

Abstract

The Australia Gallery located on the north east of the first floor represents the shifting intersection of sound environments significant Melbourne's urban life. The diverse sound elements pertaining to the playground, the streetscape, the news stand, the Pharlap pavilion, the Markets and the MCG will be analysed and their interdependent relationships examined as a composed spatial polyphony. Particular attention will be given to the spatial design of the speaker placement as well as durational aspects of the design. The paper will be illustrated with original design information provided by John Keane, curator, and Garth Paine, sound designer and composer and Peter Jago of System Sound.

1 Introduction

Spatial Counterpoint involves the complex design of sound in many layers through an acoustic space.

The counterpoint, the line against line, layer against layer is dispersed through an entire physical space. Spatial geometries involving sound spreading, have been part of the history of sound and architecture for centuries. One is reminded of the powerful antiphonal canzonas of the Gabrielis and other spatially designed double and triple contrapuntal polychoral works of late sixteenth and early seventeenth century Italy.

Contemporary buildings and museums rarely explore the potential of their acoustic properties in clever or creative ways. All spaces are acoustic spaces whether recognized as such or not, whether acoustically designed or not. The Sydney Museum and parts of the Melbourne Museum are exceptions in that they have consciously used spatial sound as the powerful design tool it is. Both museums have considered the on effect complex layered sound designs can have for the museum visitor. This paper focuses on the way in which the Melbourne Museum has approached sound with particular reference to the Australia Gallery in the upper east galleries of the museum..

1.1 Acoustic spaces

The Australia Gallery (**Figure 1**), contains a variety of acoustic spaces in itself as a result of the architectural structure of the museum coupled with the interior partitioning from the upper floor. There is audio spill from the lower floor of the East Superspace where the sound of water lapping under the suspended Polynesian canoes can be heard from the adjoining exhibition. This gives an interesting contextual orientation to the Australia Gallery which is quite poetic. The presence of this sound encourages the posing of a question, Can one float into this space from another entry point? Indeed there is entry from the lower gallery by a staircase.

By viewing **Figure 1**, all the chief areas of the Australia Gallery can be identified. The central entrance opens into a concourse identified as the Piazza which serves as the heart of the gallery. The sound design for this space as a soundscape set the space for all the other exhibits whether isolated, (the Playground, Neighbours, the recessed Pharlap), the semi- partitioned, (the streetscapes, the archeological display) or centrally prominent, (the MCG, the sewer tour, a kiosk with news stands and a delivery van.)

From the outset, sound design was an intrinsic guiding force in the total design of the Australia Gallery in the new Melbourne Museum. Although the initial tender included multimedia, film and lighting control."the sound led the way " according to chief curator John Keane. Garth Paine, the main sound designer of the audio content of the museum.decribes his role "I provided completed stereo mixes of the sound environments and the lighting design was done in sync to defined sections of the sounds. Those sections of the sound design in the Piazza actually come up on a semi-random order however the lighting and the temporal structure of those sections are predefined."

In my interviews with Garth Paine and John Keane it is clear that the audio was the dominant strain in integrating the various exhibits in order to give them a feeling of seamless openness, similar to being out on the street. A subtle interplay of diverse scenarios would effect this for the museum visitor.

The historian Andrew Brown -May describes the new museum as "a limitless place. With rooms never before entered.. an open book. Here there is nowhere to hide. The space of the Australia Gallery gives you the impression you're standing in the Bourke St Mall." (Brown May 2001 pp94-95). The Piazza soundscape design guidelines set criteria to ensure this was so.

2. Overall Design Guidelines

Some general principles were arrived at for the upper east piazza soundscape. These were:

1. The soundscape will be a valuable addition to the Upper East visitor experience. It provides an important way of broadening visitors 'Melbourne' sensibilities through a variety of ambient and 'key point' sounds within the Piazza space.

2. The key curatorial concern is for the soundscape not to dominate or distract visitors from their experiences within the exhibition segments themselves.

3. High points should be well spaced apart.

4. There should be room for 'quiet time' with care taken to avoid aural visitor fatigue.

5. Where possible, sounds could be related to what is in the exhibitions, especially particular objects

6. Sounds should relate directly to Melbourne street/outdoor spaces. This should provide useful parameters for focussing the soundscape and avoid trying to do too much.

7. Sounds can be recreated unless it is a voice which requires historical authenticity eg voice declaring Olympics open or Menzies declaring war.

8. These focuses were to be applied in specific and generalized ways. For example the Flinders St station would also allow for a generalised focus of transportation (i.e.. air planes - transportation - cars - trucks etc The Vic Market would allow for a generalised focus on multiculturalism in Melbourne (ethnic languages/music etc)

9. The Weather was a constant element of the soundscape. It can be perceived as subliminal most of the time, but at times it takes the foreground in rain squalls, thunder or lightning)

10. There was an attempt to historicise the sound sources of the specified sites for as many historical periods as possible i.e.1920, 1940, 1960, 1980, 2000. The soundscape progresses through these as the main thread of the soundscape. It is composed in such a way that the sound sources are perceivable individually but also as one layer in the many layers, evoking the generalised experience of the site over time.

The temporal layers of history are here being treated as spatial counterpoint.in a thoughtful way. The content of the sound design as the weaving of aggregates is the craft of the composer and sound designer. How it is dispersed spatially yields another perceptual layer of this sensitivity so that the planned complexity is elegant rather than dense to the ear. The spatial counterpoint is a principle at all stages of the design, its content and its implementation. Gradually the design process became ever more explicit in the ways this was to be achieved as it was expressed in the following document provided by the design team.

2.1 Specific Sound Design Criteria

The following Piazza Soundscape Design Guidelines were the specific application of the general guidelines and directions .

1. The focus of the sound palette is to be out-Door public spaces
2. The soundscape is not to actively deliver historical events (i.e.. speeches)
3. The Piazza sound level should approach natural outdoor public space levels occasionally
4. The Piazza should be a space in which people can sit and contemplate
5. The Piazza soundscape should not be intrusive or demanding
6. The boundaries of the locations will be bound by the CBD (i.e. not the beach or docks)
7. Sounds will be sourced from within living memory (not before 1920)

8. The soundscape is not to act as a vehicle for Sound Ecology

The soundscape will be structured into five focuses.

1. Train Stations - focusing on Flinders St and Spencer St stations
2. Markets - Vic market
3. Sporting venues - i.e. MCG
4. Parks and Gardens - Treasury gardens
5. Industry

2.2 The Sound Source Archive

Specific ideas and thousands of sound sources were discussed, recorded and collated in an archive under the general headings of sport, transport, festivals, media personalities, commerce, cultural icons, big news, past cultural events, establishment, immigration/multiculturalism and other. Other included the calls of a Herald Boy, the milkcart, a military gun. The final selection of sound sources is itemised by Garth Paine as follows:

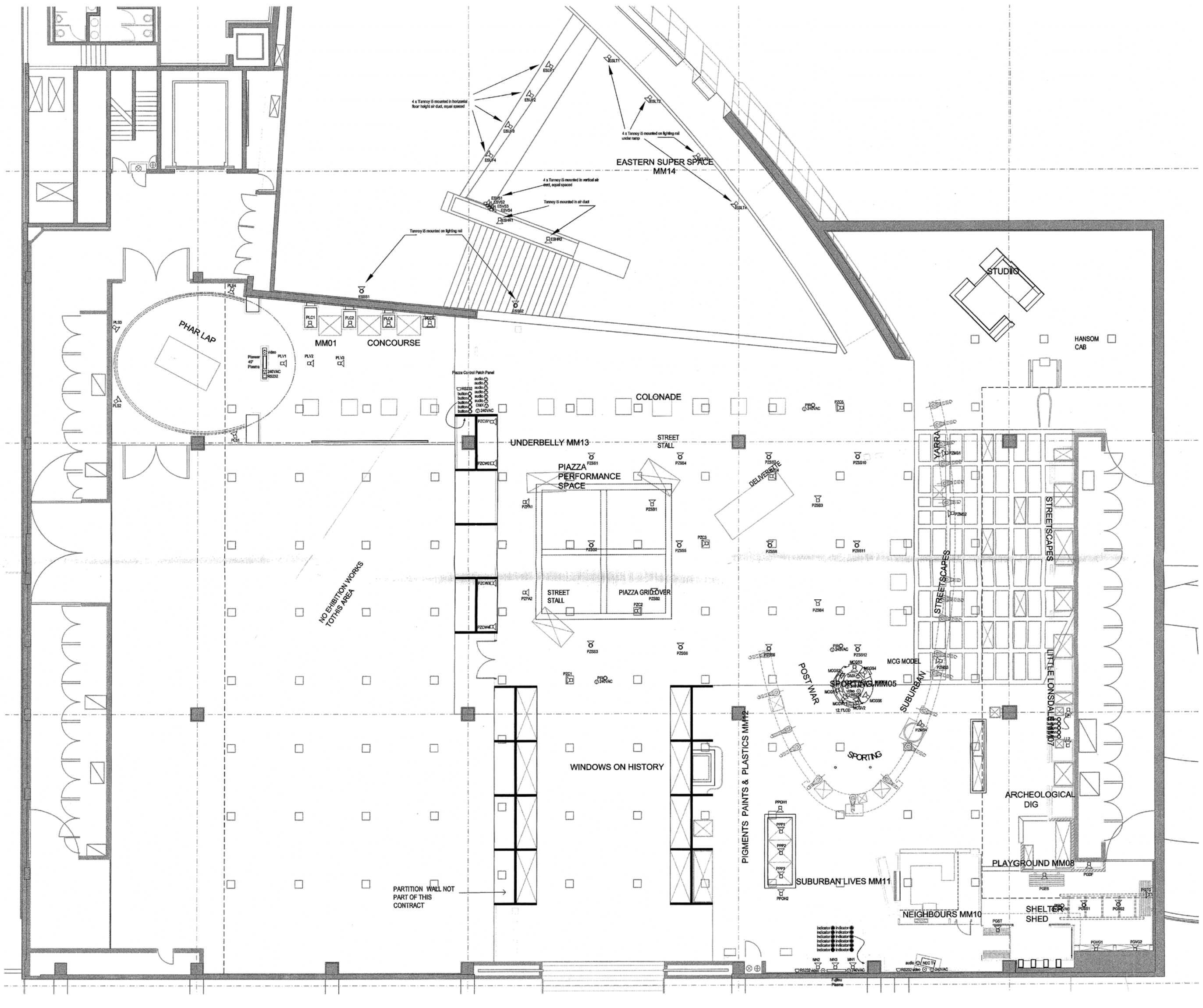


Figure 1 Floor Plan, Upper East Galleries

' Copyiright System Sound. Credits to Johns Scandrett, Garth Paine, Peter Jag and Ian Stevenson

The categories listed below were thought to best represent the experience of public space, and to present sufficient breadth of potential sound sources to make the Piazza soundscape interesting and engaging whilst presenting a quality that will allow all visitors to sit and contemplate within the exhibition space.

Location Focuses

1.Flinders Street
2.Treasury Gardens
3.Industry - Printing
4.Victoria Market
5.City Square

1. Finders Street -station and intersection

Natural Sounds
Small birds
Seagulls
Wind - clunking tins,
papers blowing,
rustling leaves
Rain - on pavement,
in gutters,
tyres through puddles,
footsteps in rain
Tyre's/footsteps in dry weather
Thunder
Man-made sounds
Trucks
Trams
Trains
Horse and cart
Bike bell - bike tyres
Boats - steam whistle
Taxis
Motorbikes
Vehicle horns
Vehicles parking, arriving, idling
Station announcements
Chit chat - various ages
varied abuse (distant yells)
Asking for and giving directions
Police whistles
St Paul's bells
Ambulance - Police - fire engine sounds
Herald boy

2. Transition to gardens

Natural sounds
Footsteps on paths
Picnics
Birds of all types
Rain on footpath
Rain on grass
Wind in trees
Dawn and dusk chorus
Dogs
Thunder and lightning

Man made sounds
Marching - parades - protest - demonstrations
Aeroplanes - flyovers Trams going past
Fountains
Children playing
Ball games (thwack thwack)
Hide and seek
Trams - cars: in groups then silence
Chit chat
Transition of accents
Shop sounds - cafes
Waiters
Sprinklers
Lawnmowers (old and new)
Rubbish collection - bin emptying
Industry (printing)

3. Industry - public space sounds

Man made sounds
Construction site
Building workers
Jack hammer - digging - saws
Timber crashing
Compressors
Cranes - whistles - dog men
Yelling - male - macho
Horses and carts
Loading of trucks
Calling out - yelling instructions -
- heads up

Fork lifts
Chunka Chunka - machinery sounds
Hisses and clunks
Tea chests - cases -
Delivery of beer
Blocks and tackles
Horse and cart harmless sound
Taxis
Air on bus and truck breaks
Reversing warning sounds

4. Victoria market

Natural sounds
Rain
Wind
Thunder

Man made Sounds
Babble
Languages
Wooden barrow wheels
Competition
Deliveries
Bartering
Arguing
Spruking

5. City square

Natural Sounds

Pigeons
Sea gull s
Footsteps on blue stone
Rain and wind

Man-made Sounds

Busking
Musical performance
Clowns - strikers
Soapbox - Speaker's corner
Passing chatter
Kids play
Skateboards
Distant traffic
Distant city sounds
Fountain
Steps - walking upanddown

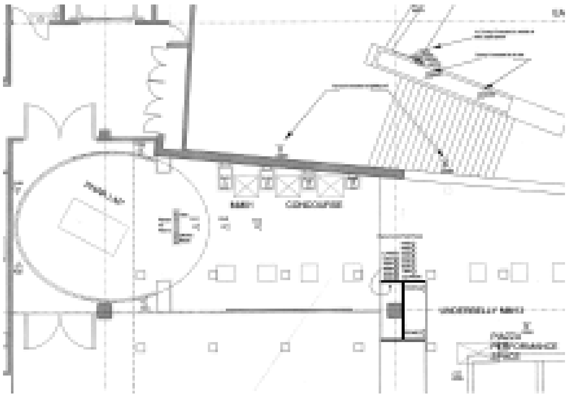
At this stage it can be seen that many decisions have been made regarding the assignment of sound to location.

3 Sound Placement

The task of refining and locating the material through the space continued as a collaborative process. Personnel involved in the architectural planning, speaker design and systems control provided consultation and input. Sound placement had to integrate the acoustic limitations of the areas of the gallery, the building and the speaker placement and be compatible with the overall systems network of the museum.

Apart from the Piazza soundscape, two other spaces impacting upon the concourse had to be worked around as they involved separately designed sound as part of their exhibits. One was the partitioned and separate Pharlap room, an independent shrine/mausoleum composed as a separate audio work, yet having a video with sound at its entrance. Once inside the soft velvet curtained environment, one is acoustically partitioned from the rest of the outdoor piazza, due to the subdued absorbent material providing the regal splendour of the race track, establishing a more independent listening field. It is a place where one can focus for longer on the sonic material composed into the set as an independent theatrical composition of racing history. Here one can ponder the poetic significance of Pharlap as a national icon through the subtlety of the soundspace. The video in relation to it comes and goes providing a reference point of the actual rather than the revered world.

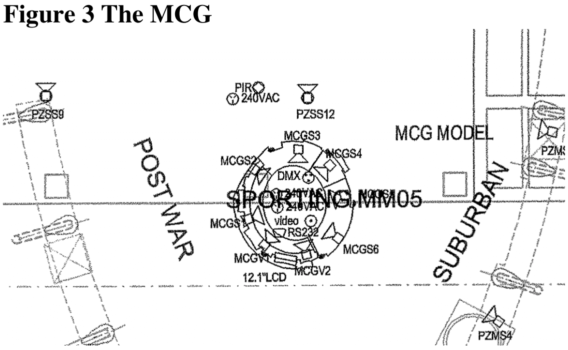
Figure 2 Pharlap



The second space, the MCG impacts more directly on the soundscape as it has a more central position as an object exhibit, the MCG arena Its bursts of distant cheering pierce the soundscape and had to be integrated to achieve the desired perspective and temporal interruptions.

The principal sounds of sport being specific to the sport, or generalised crowd noises would not lend any useful qualities to the overall ambience of the Piazza. It was felt that the sounds of sport are more related to individual activity than the experience of public space, and so became focused at the MCG arena. The design of the timing and volume of these two intrusive sound fields, The MCG and the Pharlap video had to be skillfully integrated with the overall subtler soundscape design., so their overlapping was not destructive to other material.

Figure 3 The MCG



One of the most successful areas of the Australia Gallery is the children's playground and shelter shed. Made of the materials and to the scale familiar from shelter sheds of most state schools, the area already boasts its own graffiti, proof of the public's acceptance of it. The sound, the materials, the furnishings and location in the furthest corner of the gallery, all contribute to its charm and make possible the nostalgic effect enjoyed by so many identifying the contents of their own school lunchboxes as historical cuisine. The intimacy of the acoustic and the playground stories coming and going are designed with successful randomness and unpredictability of a real schoolyard.

3.1 Speaker Placement

As can be seen from the overall floor plan **Figure 1** every attention has been given to concealed and diffuse speaker positioning to ensure the seamless spread of sound throughout the space to give the most realistic and poetic results.

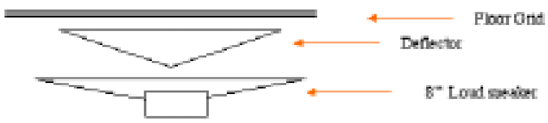
The placement reveals a thorough consideration of the height, depth and width of the acoustic space and a knowledge of the changing relationships possible between the speakers.

The following description provided by Garth Paine is the result of the continuing collaborative process of the larger team and it reflects the thought and detailed analysis of the physical space, a sensitivity essential to the success of this installation.

"In the Piazza for the Melbourne Museum, the spatialisation is generated using speakers in the floor, at about 100mm above the floor, at about 1600mm, and in an array through the roof.

Thirty-six loud-speakers were placed through the space. Six loudspeakers were mounted in the floor, forming a diagonal across the space. These speakers were placed about 200mm below floor level and were covered with a stainless steel cone to deflect the sound sideways, and then by a heavy grid to protect them. The floor loudspeakers were designed to carry sounds that would ground the environment. These included footsteps on pavements and gravel paths and the sound of bicycle tyres on gravel. These sounds were also spread slightly on higher loudspeakers to make the sound occupy the space even though it were 'grounded'. The floor loudspeakers were also used to pull the apparent height of some sound images down. For instance, the sound of a sprinkler running in a public park was placed in the low wall speakers and the ceiling speakers, but was primarily allocated to the floor speakers. The use of the floor speakers grounded the sound image, and the addition of small amounts of the signal in the low wall speakers lifted the sound image slightly. The addition of a very small amount of the sound image in the 1600 mm height speakers creates the sense of space experienced when hearing one of these sprinklers in an outdoor setting where the reverberation is minimal. The sprinkler sound being sent to the 100mm height speakers was also spatialised, so that the sound moved around the space in the manner of a large public sprinkler. This created a dynamic sense of movement, the kind that makes you look to see if you are about to get wet!!

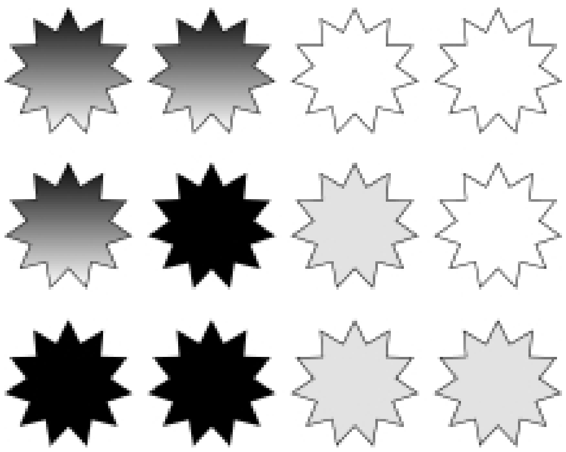
Figure 4 Speaker Mountings



A further 6 loudspeakers were placed in the kick plates of the exhibition stands approximately 100mm above floor level. These loudspeakers were mounted to face directly out into the space, and were designed to create a sense of space in the sounds emanating primarily from the floor speakers, and assist in the adjustment of the vertical position of the sound image from the ceiling and 1600mm speakers. These speakers were placed, two at either end of the space and one along each of the sides.

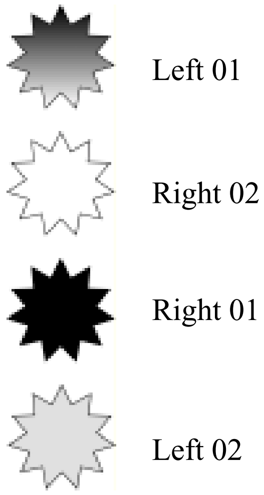
The 1600mm speakers were placed at the top of the exhibition cases. A total of four of these speakers were placed, two on each side of the exhibition space. They were used to help lift the lower sounds when I didn't want to use the overhead speakers. They assisted in the creation of *space* in the sound image. The overhead loudspeakers formed an array of four groups of three, placed in a triangle formation as follows:

Figure 5 Speaker Groupings



This arrangement meant that it was possible to move the sound image either side from the corner. Each group took one side of the stereo image as follows:

Figure 6 Speaker Assignment



This arrangement allowed for a passing of the signal in a circular pattern, and allowed the sound to moved down the side of the array and into the centre."

From the perceiver's point of view, this information is hidden but its presence creates a spatial theatre of coming and going, or being close up and far away, so that sound is felt as proximity, dimensionality and moving matter. For example, the concealed speaker in the wall of the basket- ball ring adjoining the playground makes you look up to the real source of the sound as though the ball just missed you. One is unaware of the presence of artificial sound allowing a sense of immersion to perceive the subtle fluctuations of sound in the space.

An elegant extension of innovative speaker placement is also present on the way in to the east gallery via the escalators where one hears the windmills, water and birds as if entering a large actual , not simulated agricultural space. Speakers are carefully placed at the back of the windmill structures facing the window panes, causing an extra dimension of sound reflection. The spatial listening distance is considerably enhanced.

4 Temporal and compositional design

Garth Paine's prowess as a composer of elaborate computer music systems was fully engaged to implement the results of this consultation in a creative and artistic way..He encoded hundreds of sound files, shaped them, morphed them and organised them into temporal aggregates which would move through the space according to the assigned areas and designed speaker placement. Although no audience interactive triggers or systems were employed, the design has inbuilt fluidity so that the various units come together at different times in different ways with varying degrees of change. All hardware systems, the spatial features of the audio tool box, the computer encoded temporal systems and the general museum networks are seamlessly integrated to effect the intention. The Australia Gallery is a fine example of sound design where many layers of subtleties unfold from different areas at different times. The result is that a space is experienced as if wandering through a genuine piazza, into an arcade, into a theatrical racing retreat, or into the school yard. The richness of the experience is sourced from the contrapuntal layers, each carefully designed as a continuum, but each having many possible aggregate lives.

It is a sound design which emulates the natural world, a sound world which comes and goes, engulfs, confronts, disrupts, floats, disappears, sustains with the ebb and flow of natural activity. The ear pulls the strands together so that each visitor has his/her own experience as they define their listening pathways and the length of stay in each location. Walking out of the Australia Gallery one is aware of having been somewhere, not quite real; but somehow familiar, engaging and immersive, integrating the past and present.

The spatial counterpoint works at many levels, the physical integrations of many parts to the whole, the layering within each particular sounding area, the intersection of each part against the other, the temporal shifts from past to present and the dramatic interplay of the elements as changeable as the weather. John Keane's commitment to sound as the focus and vehicle to set this theatricality in motion was entirely successful.. Time based systems such as these are all too rare but their power to move is readily apparent.

5 Conclusion

The Australia Gallery perhaps represents one of the most powerful and dedicated sound designs in the state of Victoria at the present time. The subtlety of the experience is achieved through the integrity of the planning, at every stage engaging spatial counterpoint as the first design principle. Layer intersects layer in four dimensions, height width, depth and in temporal shifts of small and large, modern and nostalgic, moving at different speeds, creating shifting universes. Its power creates the physical, temporal, sculptural and emotional elements of this theatrical world, entered by the player. It is a credit to the entire team that this process achieved what it has.

It is only rivalled by the Museum of Sydney, an entirely sound designed building where every gallery relates to every other on a temporal and sonic level from the outside in and the inside out, (Bandt, pp1-23.)

From understanding and experiencing clever and sensitive sound designs such as these, the public will come to demand greater input of spatial sound design in public spaces in the future.

6 Acknowledgments

Thanks to Garth Paine, composer, John Keane, Curator, Peter Jago of System Sound who provided information, interviews and discussion about their involvement in the Australia Gallery. Thanks also to Iain Mott for scanning and video editing.

7 References

Bandt, Ros. 2001. *Sound Sculpture: Intersections in Sound and Sculpture in Australian Artworks*. pp1-23

Brown-May, Andrew. 2001 "The Surreal Thing" *Meanjin On Museums*., Vol 60, No4, pp 93-98

Gillespie, Richard. 2001 "Making and Exhibition" *Meanjin On Museums*., Vol 60, No4, pp111-121

Rasmussen, Carole. 2002 A Museum for the People, a History of Museum Victoria and its predecessors.

Applying Aural Research: the aesthetics of 5.1 surround

Jim Barbour

School of Social and Behavioural Sciences, Swinburne University
email: jbarbour@swin.edu.au

Abstract

Multi-channel surround sound audio offers a richer perceptual experience than traditional stereo reproduction, recreating three-dimensional aural soundfields. Understanding the limitations of human auditory perception and surround sound technology may assist the exploration of acoustic space as an expressive dimension.

1 Introduction

Research and development into new digital audio software and hardware has resulted in multi-channel digital audio recording technology that allows the creation of immersive aural environments reproduced by multiple loudspeakers arranged around and above the listening position. The audio delivery platform offered by the Digital Versatile Disc has provided the facility for musicians, composers and sound designers to produce three-dimensional audio works in studio environments and to have them recreated in distant listening rooms. The technical quality of digital audio allows the recording, delivery and reproduction to be transparent to a listener, removing the aural artefacts of analogue technology. No longer are we aware of the sound of the medium itself, the hiss and scratches of mechanical sculpture.

However, digital audio technology has presented new challenges to aural artists. Recording engineers, producers and musicians who are familiar with the paradigm of stereo are experimenting with surround sound technology, unsure of how to best use the new medium to deliver a three-dimensional immersive soundfield to consumers. This paper will explore some of the audio research that lies behind the current debates among sound professionals about the aesthetics of localization, spatial impression and envelopment, and will present some of the solutions that have been achieved by audio producers. References will be made to music releases and sonic art to illustrate some of the aural possibilities. Despite significant limitations inherent in the 5.1 surround sound specification, it is the position of this paper that it represents an opportunity to deliver an exciting and immersive aural experience to a consumer today.

2 Consumer Surround

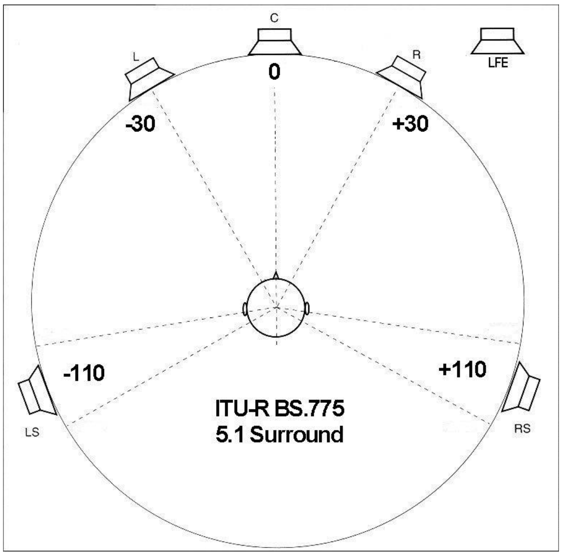
For many decades, it has been a goal of many audio practitioners to deliver the immersive, three-dimensional soundfield of the real world, or of their creative imagination, to listeners in their home environment.

2.1 Quad: 1970s

The development of multi-track tape recorders in the 1970's led to successful experiments with four channel quadraphonic recording, which succeeded in capturing and reproducing a high quality three-dimensional soundfield in a controlled studio environment. However, all efforts to deliver it to listeners in their homes failed. In reality, the delivery medium of vinyl was not aurally transparent, since we could hear disc noise and scratches, which destroyed any illusion of an immersive soundfield. Consumers who had purchased the quadraphonic reproduction equipment, including expensive turntable cartridges, amplifiers and extra loudspeakers, were so disillusioned by the poor quality that the concept died a marketplace death that still resonates today.

2.2 DVD:1990s

This immersive soundfield could not be delivered into consumer's homes with any significant success until the digitisation of video and audio permitted the introduction of the Digital Versatile Disc (DVD) format in the mid 1990's. The technical specifications for DVD-video include the audio standard ITU-R BS.775 (ITU 1993), known as 5.1 surround sound. This utilises five full range loudspeakers distributed around a listener and a low frequency sub-woofer loudspeaker, as shown in the accompanying diagram.



ITU-R BS.775 5.1 Surround specification

With the extraordinary acceptance and purchase by consumers of the DVD-video format, which now numbers more than 50 million players worldwide, there is a platform for the delivery of high quality multi-channel audio into any listening environment. However, audio researcher Francis Rumsey, author of the book *Spatial Audio* and chairperson of the committee that developed the ITU 5.1 standard, identified a problem for the audio community:

> Although 'purist' sound engineers find it hard to accept that they must use a layout intended for movie reproduction, …, most pragmatists realise that they are unlikely to succeed in getting a separate approach adopted for audio-only purposes and that they are best advised to compromise on what appears to be the best chance for a generation of enhancing the spatial listening experience for a large number of people. (Rumsey, 2001, 18)

3 Psychoacoustic Phenomena

Aural research has identified the parameters to be used for a qualitative assessment of the 5.1 surround sound process. According to Rumsey, the 'psychoacoustic phenomena that appear most relevant to the design and implementation of audio systems' are source perception (identity and location), spatial impression and acoustic envelopment. (Rumsey, 2001, 21) Of these, source localization appears to be the most difficult to accurately deliver with a sound reproduction system, but equally, it is the most interesting for many sound practitioners. Therefore, let us consider localization in detail first, since many of the findings for localization affect our perception of spatial impression and acoustic envelopment.

Acoustic space is generally considered to be a sphere with our head at its centre. (Carpenter and McLuhan, 1970, 67) Our perception of the identity, location and distance of a sound source can be aided by visual cues, but is built on binaural discrimination and our experience of the variations in the loudness and timbral qualities of a sound source at different distances. The perception of the environment in which we hear the sound source is based on our perceptual memory of different acoustic spaces, built from the experiences of hearing the reflections of sound waves from surfaces within spaces, and the direction, timbre and loudness of those reflections.

4 Two Channel Stereo

Our experience of high quality audio has been mediated by the stereo paradigm for more than thirty years, based on the implementation of two-channel recording and reproduction through two loudspeakers. It has been stated by the results of considerable research into loudspeaker placement that the ideal listening position is at the apex of an equilateral triangle where the distance between the two speakers is equal to the distance from each speaker to the listener.[1] This position, approximately two to three metres from the speakers with an angle between the speakers of sixty degrees, (±30º from the centre front), is often referred to as the reference listening position or 'sweet spot'. At this reference position, it is possible to perceive with relative accuracy the location of a reproduced sound source within the front arc of 60º.

> Two-channel stereo … is essentially limited to reproducing both sources and reverberation from an angle of about 60º. This is adequate for many purposes as the majority of listeners' attention is likely to be focused in front of them when listening to music or watching television. (Rumsey, 2001, 64)

4.1 The Failings of Stereo

Stereo reproduction has delivered accurate localization within the front arc, but only limited envelopment. Directional perception between the two loudspeakers is only an artificial or phantom image, since there is no true loudspeaker source at that position. If the position of the listener changed to be closer to one loudspeaker, the proximity to the closer

[1] For more detailed information concerning stereo loudspeaker positioning, see Holman's *5.1 Surround Sound, Up and Running*, (2000), Boston: Focal Press, or Eargle's *Stereophonic Technique* (1986) AES.

speaker would make the sound from that loudspeaker arrive sooner and be louder than the sound from the more distant loudspeaker. Due to this change in precedence and relative loudness from the two loudspeakers, the phantom image of the source would appear to shift toward the closer loudspeaker, destroying accurate localization.

There is a timbral change for phantom images due to acoustic crosstalk, with the sound from the left loudspeaker arriving at the right ear slightly later than the sound from the right loudspeaker, and vice versa. Mid to high frequency colouration results with the effect most pronounced for central images. Some research and experimentation has focussed on extending the directional cues to beyond the 60º arc using anti-phase signals mixed into the stereo channels, sometimes referred to as 'trans-aural stereo'. (Rumsey, 2001, 74) This research has analysed the perceptual effects of acoustic crosstalk from, for example, the left loudspeaker to the right ear. Rumsey outlines the possibilities for extended azimuth localization:

> Crosstalk cancelling systems perform this task by feeding an anti-phase version of the left channel's signal into the right channel and vice versa, filtered and delayed according to the HRTF characteristic representing the crosstalk path. The effect of this technique can be quite striking, and in the best implementations enables fully three-dimensional virtual sources to be perceived, including behind the listener, from only two loudspeakers located at the front. (Rumsey, 2001, 75)

4.2 Stereo Envelopment

Another component of the sound source reproduced through the loudspeakers is the enveloping reflections of the original sound source in its acoustic space. For a listener in the reference position, the two loudspeakers will reproduce the enveloping soundfield of the original recording with reasonable accuracy, allowing the listener to perceive the location of the source and some illusion of spaciousness as soundfield depth within the front arc. However, as the listening position changes to be closer to one speaker, the perception of the acoustic space will collapse due to precedence and loudness changes.

5 ITU 5.1 Surround Sound

The development of 5.1 surround sound provides two separate possibilities to extend the soundfield beyond the front arc of 60º, and to provide enhanced envelopment and more accurate and stable localization of sound sources. The front centre channel is equidistant between the main left and right front loudspeakers and impacts upon imaging in the front arc of 60º. The left and right surround loudspeakers provide the possibility to extend localization beyond the front arc into the full horizontal 360º.

> Surround sound provides an opportunity to create something that works over a much wider range of listening positions than two-channel stereo, does not collapse rapidly into the nearest loudspeaker when one moves, and enhances the spatial listening experience. (Rumsey, 190)

For each of these additional loudspeakers, there are some advantages and disadvantages to their use.

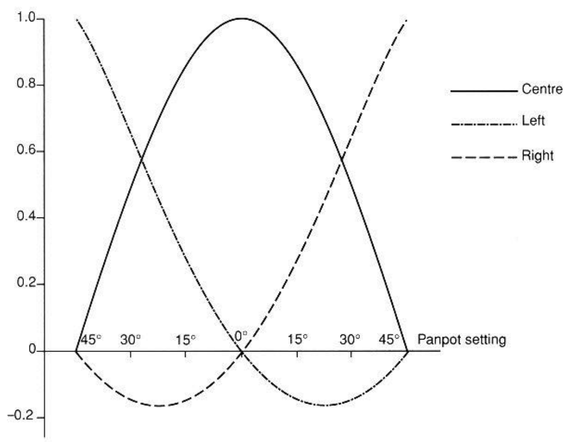
5.1 Centre Channel Advantages

The centre channel loudspeaker provides a solid, real central image to a soundfield created across the front left-right arc. When sounds in the centre of the soundfield are reproduced through the centre loudspeaker, there will be a reduction in the impact of the precedence effect, allowing a broader reference listening position. This is valuable for music only listening, but is considered crucial when the audio is tied to images, for example in film or television reproduction. The total soundfield image is therefore more stable across the front arc. Also, timbral modification is reduced with the introduction of a centre loudspeaker, since there is a real sound source producing one set of sound waves travelling to the ears, rather than the two sets of sound waves for stereo creating the phantom centre image.

5.2 Centre Channel Disadvantages

For some practitioners, a sound that is reproduced through the centre loudspeaker only can be a nuisance, with sounds too focussed at that position, confined to the loudspeaker. To overcome the problem of central focus, it has been suggested (Holman, 2000, 86) that some of the centre channel sound should 'bleed' into front left and right to de-focus the source, or possibly reverberation should be added into front left and right. While this may overcome some of the problems of focus, it does partially destroy the accurate localization that is possible with a centre loudspeaker. The issue of centre loudspeaker position is also relevant when it is difficult to place the loudspeaker in the true centre position due to screen limitations, for example when a television occupies that spot, and the depth or height of the resulting centre loudspeaker position may compromise the quality of the sound.

While many audio practitioners are attempting to recreate an accurate soundfield, there are also many creating an illusion of localization using mono sound sources that are positioned or panned according to amplitude or time differences between loudspeakers. When using two-channel stereo only, the amplitude panpot law that has been found to be most effective psychoacoustically is a –3db reduction in each channel for the centre position, and greater than –40db at the extremes of the left and right loudspeakers. (Rumsey, 2001, 177) This amplitude law has been easy to implement in recording consoles and very simple to operate. However, there is significant difficulty in panning left to right when there is a centre channel. According to Michael Gerzon, his research concluded that true psychoacoustic panning across the left-centre-right sound stage should include frequency and amplitude variations, with out-of-phase amplitude components for the extreme ends. (Gerzon in Rumsey, 2001, 178).

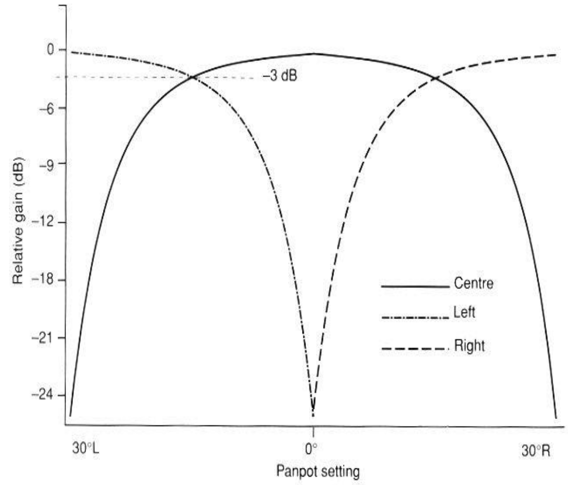


Gerzon’s psychoacoustic panning laws

While this may be optimal, it is very expensive to implement in recording consoles, except in some of the latest generation of digital audio consoles. Instead, the most common implementation is pair-wise amplitude panning, as Rumsey described:

> Typical three-channel (or more) panpots are rather crude devices, tending to work on simple positive-valued gain relationships between pairs of channels at a time, treating each pair of speakers (left-centre or centre-right) as a straight amplitude-panned pair as in two channel stereo. (Rumsey, 2001, 180)

Rumsey’s pair-wise –3db panning laws:



5.3 Surround Channels

Let us now consider the possible uses for two surround loudspeakers, positioned according to the ITU standard at ±110º from front centre. As Rumsey states:

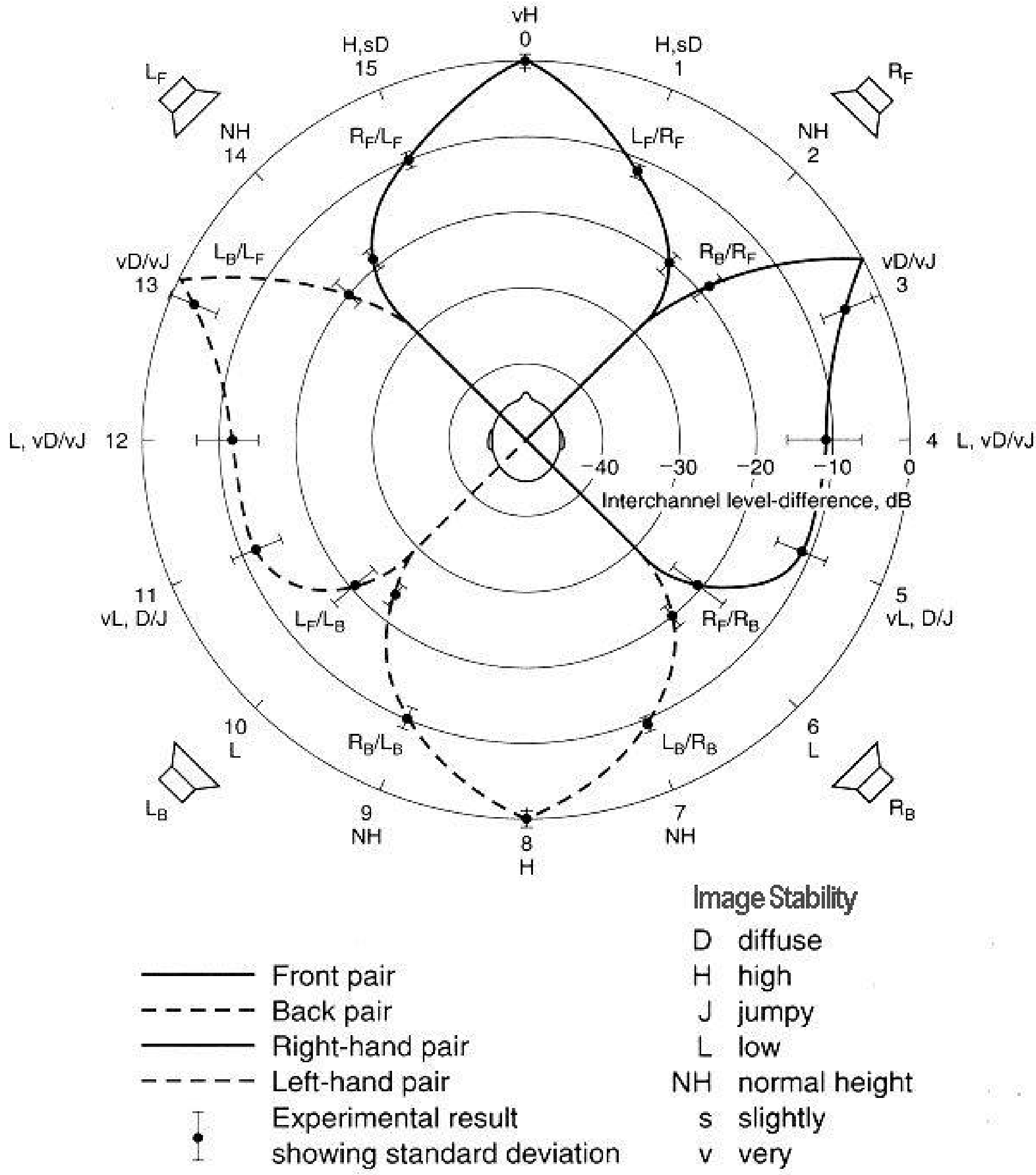
> ...the primary intention for these (surround) channels is for non-localisable ambience and effects information that adds to spatial impression. In the words of the ITU standard: ‘it is not required that the side/rear loudspeakers should be capable of prescribed image locations outside the range of the front loudspeakers’. (Rumsey, 127)

However, for most users of surround sound systems, it is the possibility of specific localization at the sides of and behind a listener that most excites them, offering the potential to deliver an aural experience that goes far beyond the limitations of the stereo frontal image. As classical recording engineer Richard King said in a recent edition of *Mix* magazine:

> When you work in surround a lot, your ear gets used to it after a while and you might even think it's not really working. Then you flip it back to stereo and see the dimension you're missing. It's like night and day. (King, 2000, 48)

5.3.1 Accurate Side and Rear Locations

The greatest practical difficulty in using surround loudspeakers lies in this desire for accurate localization at the sides and behind the listener. However, it is difficult to create images where there are no loudspeakers due to a lack of significant interaural differences, leading to poor phantom images and poor image stability. In his book *Spatial Audio*, Francis Rumsey cited research undertaken by Paul Ratcliffe at the BBC in the 1970s to identify localization and image stability for a square, quadraphonic loudspeaker arrangement, which produced this diagram:



Ratliffe, 1974, BBC Research and Development

There are several important points to consider in this diagram. Firstly, the soundfield across the front arc, in this case ±45º, has strong image stability and clear localization with a precise phantom centre and distinct left or right positions at –25db between channels. However, as soon as the location moves beyond the front arc to incorporate the rear loudspeakers on either side, the image stability becomes poor and the perceived position is very different to the panned position. For example, on the left side, when the front and rear loudspeakers have equal amplitude, the perceived position is approximately 67º from front centre, rather than the expected 90º. Interestingly, for the rear arc, there is clear localization and good stability, very similar to the front arc.

> (Ratliffe) concluded that phantom images based on amplitude differences between side pairs were poorly localised and that sources appeared to jump rapidly from front to back rather than panning smoothly down the sides. This is attributed to the difficulty of creating interaural differences of any significance from differences between loudspeaker pairs to the same sides of the head. (Rumsey, 2001, 33)

5.3.2 Side Panning Inaccuracies

While these measurements were derived from a quadraphonic square arrangement, the findings can be applied within reasonable limits to the ITU 5.1 surround layout. Despite difficulties in smoothly panning along the side arcs, where the large angles make precise localization virtually impossible, what is the practical requirement for such precision in side panning? For many rock recordings, the individual sounds are often mono recordings that are amplitude panned to specific locations in the final mix. Remixing albums into a surround sound format that have previously been released in stereo represents a potential economic bonus for artists and record companies, but does present many challenges. For the producer and engineer, they must decide where instruments and voices will be located, and what acoustic space they will create for those sounds to exist in. This is not without artistic challenges, as engineer Elliot Scheiner explained in an article in *Surround Professional* in July 2000:

> I've got to admit that sometimes I'm a little frightened about where I'm going to pan things and how that's going to alter the mix for the band and the fans. I've been blasted for the work I did on [*Steely Dan* album] *Gaucho* - one critic said that I destroyed that record. (Scheiner, 2000, 30)

In this case, Scheiner placed saxophones and background vocals into the surround loudspeakers only, with some artificial reverberation added to rear and front channels. A listener is very aware that these sounds are coming from the surround loudspeakers, not from positions half way between the front and rear or the rear centre, that is, they are clearly localised at the loudspeaker position. This source perception is probably part of the reason why many fans of Steely Dan were unhappy with the surround mix, as it is very different to the original stereo mix, which had all sources in the front arc.

5.3.3 Loudspeaker Identity and Masking

Other problems that may arise with surround localization include precedence effects collapsing images to the nearest loudspeaker, and the sounds from the front loudspeakers masking rear sounds due to our visual focus reinforcing our forward facing perceptual preference. Also, as we considered for the centre loudspeaker location, if the surround loudspeakers are not properly sited in the home environment, any attempts at specific localization are further doomed to failure. It is for these reasons that many proponents of surround sound consider it essential that there is no attempt at distinct localization beyond a broadening of the front arc to perhaps 120º.

5.3.4 Surround Envelopment

In principle, film surround mixing aims to achieve aural spaciousness by placing ambience information in stereo into the surround loudspeakers with de-correlation between the channels. (Rumsey, 2001, 85) Only sounds that will not draw attention away from the screen will be placed clearly in the surround channels, and only for significant special effects. Many music-only releases are now also using this as a general principle, avoiding specific localization in the rear channels, and using reverberation and ambience to create a spatial impression and acoustic envelopment.

6 Acoustic Envelopment

Let us now move away from specific localization in a surround sound environment to consider how we may create the other two characteristics of an immersive three-dimensional soundfield, namely spatial impression and acoustic envelopment.

A sound rarely exists in isolation, even if it is the only source. There are likely to be surfaces within the space from which the sound will reflect. At each and every reflection, the timbre and loudness of the sound will change, as some energy is absorbed by the surface or partially passes through the surface. Depending on the distances between the source, the listener and any reflective surfaces, we will perceive an aural identity for the room, with characteristics of size, shape and surface materials. Rumsey defined our sense of spatial impression:

> Spaciousness is used most often to describe the sense of open space or ‘room’ in which the subject is located, usually as a result of some sound sources such as musical instruments playing in that space. (Rumsey, 2001, 38)

The complex patterns of reflections will surround and envelop us in a natural soundfield. The aural envelopment we perceive is aided by our ability to localize the original sound source and our ability to perceive the aural spaciousness with no particular direction.

> Envelopment is a similar term (to spatial impression) and is used to describe the sense of immersivity and involvement in a (reverberant) soundfield, with that sound appearing to come from all around. (Rumsey, 2001, 38)

Human perception of acoustic environments is learned through a lifetime of experience of moving through different spaces and acoustically mapping the auditory characteristics into memory. Plenge identified our ability to perceive acoustic environments, noting that:

> Only a few seconds of listening was sufficient for calibration of a room's acoustic properties, which were stored for as long as the listener stayed in the room, and then cleared immediately upon leaving, so that the listener could recalibrate at once for a new acoustic environment. (Plenge, 1974, 44)

7 Applying the Research

The issue for sound practitioners then becomes how do they apply the principles of localization, spatial impression and envelopment to the task of producing satisfying immersive soundfields? Also, what is the capacity of the 5.1 surround system to deliver these soundfields to a listener in their home environment? There are perhaps three clear schools of thought concerning the extent to which the surround channels should produce clear localization as well as spaciousness and envelopment: classical music reproduction, rock and pop music releases, and sonic art and virtual reality. While it is acknowledged that these areas may overlap or further sub-divide, they will provide an opportunity to analyse three approaches to surround sound recording and reproduction. Rumsey argues that:

> The primary aim of most commercial media production is not true spatial fidelity to some notional original soundfield, although one might wish to create cues that are consistent with those experienced in natural environments. (Rumsey, 19)

7.1 Classical Surround

If there is one characteristic of classical music that remains almost religiously rigid, it is the strict protocol of performing classical music exactly as it was written. Performers spend many years perfecting their craft to play only the notes on the page at the tempo defined, with the differences between good and great performers judged by their skill at interpretation by subtle variations and expressive gestures. Similarly, listeners judge the quality of recordings based on similar subtle differences in acoustic clarity, space and definition. The Western paradigm of performance is well established with performers seated in front of an audience. We do not expect to hear a Beethoven symphony from a chair placed in the centre of the orchestra, and it is unlikely that the majority of classical music listeners would accept a surround sound recording creating this aural perspective. As classical producer Steven Epstein suggested in an article in *Mix* magazine, he doesn't 'want to be in the centre of the orchestra with the brass in the back and the fabric of the ensemble torn apart' (Epstein, 2000, 48). So the conventions of classical music are applied to performers and recordings alike, with most producers only using ambience in the surround channels.

7.2 Rock Surround

The initial interest in releasing music in the 5.1 surround sound format follows McLuhan's proposition that a new medium is used to distribute old content first, before new content emerges. The back catalogue of many music artists is being remixed into the surround format with some consumer success. It is the subject of much debate in music industry magazines concerning the techniques to be applied to the placement of sounds into the three dimensional space and the degree of envelopment that is both appropriate and desirable. Remix engineer Jake Nicely summed up one view on remixing well known albums in a recent article in *Mix* magazine:

> You still have to create a coherent acoustic space, and in the case of a record that a lot of people know and love, you have to be faithful to the original mix to a large degree or it won't sound right to people. You have to be respectful; you can't just have everything all over the place. (Nicely, 2000, 40)

7.3 Sonic Arts and Virtual Reality

Sonic artists have also been at the forefront of experimentation with digital audio technologies, using computer audio software to manipulate sounds in the realisation of their ideas. The advantages of digital audio recorders make them ideal to capture and process quiet natural sounds and aural environments with exceptional transparency and depth. Digital technology allows sonic artists to explore sound in ways that could not be achieved with analogue technology, including dissecting sound into smaller sections and layering sounds together to create unique new sounds. Sonic artists can also design their works using digital technology so that the audience can experiment with sound, creating interactive installations that allow the audience to hear their own unique aural environment.

7.4 Experimentation

The possibilities opened by the immersive qualities of digital surround sound are exactly the type of tools that many composers, musicians and sonic artists are seeking to influence their production process and to expand the sounds and spaces that listeners can perceive. For many years, composers and performers have attempted to create new and interesting environments in a live performance by surrounding an audience with performers or loudspeakers and creating unusual and unique acoustic spaces. Producer Steven Epstein knows of composers who are planning to use digital surround formats to realise spatially on disc what they're achieving in a live situation. (Epstein, 2000, 48) Engineer Jake Nicely is working with musician Bela Fleck to produce an album that sounds 'as if you were standing in the centre of a blue-grass jam session and all the players are around you'. (Nicely, 2000, 38) Producer Chris Steinmetz identifies exactly the different approach taken to digital surround sound by popular music compared to classical music, when he describes 'building the mix from the centre - being inside the soundfield, as opposed to the live situation where they have ambience in the back and are more conservative'. (Steinmetz, 2000, 44)

All these different approaches are valid in the context of exploring new ideas and techniques available with 5.1 surround sound delivery. Classical music production may well represent one end of a spectrum of views on digital surround sound, where the surround format is used to create more clearly the feeling of being immersed in the acoustic environment in which the performance has taken place. Mixing engineer Mick Guzauski recently completed the surround remix for Michael Jackson's Thriller album, well known to many millions of listeners worldwide for hits including Billy Jean and the title track. This was Guzauski's first experience of mixing for 5.1 and when asked for his first impressions, he replied:

> I actually found it easier to do than a stereo mix because you don't have the clutter that can build up just from having to put all the elements into two channels. Also, you could make the individual sounds bigger; you don’t have to filter or EQ little portions of the range up or down to make it fit into a smaller soundfield. (Gazauski, 2001, 30)

7.5 Consumer Demand

Many listeners are purchasing the home theatre systems required for surround sound reproduction, are discovering that the new soundfield is exciting to listen to and are demanding more material be available. They may currently be prepared to pay a premium above normal stereo CD prices for the new format, but they are now more discerning about the quality of the technology. They can hear the envelopment of the three dimensional soundfield more clearly, and can hear every little nuance of every instrument. Producer Chris Steinmetz has a focus on where the technology is leading:

> But when the new generation gets in tune with this new format, they're so technically savvy they're not going to be worried about it, [new mixes of old favourites]. … Part of what's happening now is a generational switchover: some people who are used to having things in stereo don't want to hear 5.1. But there are a lot of young people who love the home theatre experience and they're dying to hear more in surround. (Steinmetz, 2000, 46)

7.6 Set-up Issues

Over reliance on the centre channel can cause reproduction problems when a home surround set-up uses poor quality loudspeakers or incorrect positions. If the lead vocal of a rock recording is mixed to the centre channel only, replay could become a ‘karaoke’ version with no lead vocal, or the vocal could be too loud or severely coloured through a low quality loudspeaker. Current consensus among audio producers from the pages of various industry magazines[2] tends to favour traditional left-right phantom imaging with reduced amplitude centre reinforcement, usually only for sources that are centre panned like lead vocals, bass guitar and snare and bass drum.

7.7 Adding Width and Height

Recording company Chesky has been experimenting with a completely different approach. They believe that the centre and Low Frequency Effects channels are redundant in an ITU 5.1 system used for music only reproduction, and have instead repositioned the centre and LFE loudspeakers at ±55º increasing the real width of the soundstage.[3] In their recordings, they have captured acoustic reflections from these directions and reproduced it in the listening environment, with critical acclaim. This represents a departure from the ITU specification that may be difficult for many home listeners to implement physically, since it requires more amplifier channels and more loudspeakers. Chesky’s discs must also be clearly labelled so that consumers are not misled with incompatible product. Holman has also proposed delivering extra width and also height information, as the next development in aural immersion, but his 10.2

---

[2] Magazines include *Mix*, *Audio Technology, Surround Professional* and *EQ*, see bibliography for details.

[3] For more information on Chesky recordings see www.chesky.com

system involves using ten channels for surround and height loudspeakers with two LFE sub-woofers, which is undeliverable using current technology. (Holman, 2000, 232) However, the idea of adding height information to enhance the horizontal dimension of 5.1 surround could be very exciting to sonic artists in particular, who are always seeking new opportunities to explore.

8 Conclusion

While deviations from the ITU specification for loudspeaker locations may be rich avenues for exploration, the focus of this paper has been on the possibilities offered by the original format. It is clear that there are many reasons why the chosen loudspeaker layout is not ideal for true aural immersion, with particular problems in the side and rear listening arcs. However, it is also very clear that there is now an audio standard that has been adopted by equipment manufacturers and embraced by consumers that has enormous potential to deliver an immersive and exciting aural experience. New delivery platforms including the Super Audio CD (SACD) and DVD-Audio offer improved audio quality for multi-channel reproduction. Software 5.1 encoders for DTS and Dolby Digital AC-3 are available for many music recording programs and as stand alone products. While there may be limitations, this should not deter audio practitioners from exploring the opportunities available today to expand listener's appreciation of the aural environment. And as composer David Worrall remarked recently on an email list:

> For example, once one has achieved sounds whizzing around in 3D, what else is there to do with surround, compositionally speaking? The answer is: many things, beautiful, profound and delicate. (Worrall, 2001)

References

Alten, Stanley, R, 1999, *Audio in Media*, Belmont: Wadsworth Publishing Company

Audio Technology Magazine, Sydney: Alchemedia Publishing Pty. Ltd., bi-monthly.

Broadcast Engineering News Magazine, Sydney: Reed Business Information, monthly.

Carpenter, E and McLuhan, M, 1970, "Acoustic Space", in *Explorations in Communications,* London: Jonathon Cape Ltd.

Chesworth, David and Leber Sonia, 2000, *5,000 Calls in the Urban Forest,* Sydney: Olympic Co-ordination Authority.

Eargle, J (ed), 1986, *Stereophonic Technique: An anthology of reprinted articles.* Audio Engineering Society.

Epstein, Steven, 2000, "Orchestra Remains Up Front', *Mix Magazine*, August 2000, California: Primedia Inc.

EQ Magazine, New York: Miller Freeman PSN, Ltd., monthly.

Holman, Tomlinson, 2000, *5.1 Surround Sound, Up and Running*, Boston: Focal Press.

———, 1997, *Sound for Film and Television*, Boston: Focal Press.

ITU-R, 1993, *Recommendation BS.775: Multi-channel stereophonic sound system with or without accompanying picture.* International Telecommunications Union.

Jackson, Blair, 2000, "Six Views on 5.1", in *Mix Professional Recording* Magazine, August 2000, California: Intertec / Primedia.

Jourdain, Robert, 1997, *Music, the Brain and Ecstasy*, New York: William Morrow and Co.

King, Richard, 2000, "Orchestra Remains Up Front', *Mix Magazine*, August 2000, California: Primedia Inc.

Malham, D. G, 1998, "Homogeneous and Nonhomogeneous Surround Sound Systems", University of York, available at http://www.york.ac.uk/inst/mustech/3d_audio/homogeneous.htm. [accessed 3 August 2000],

———, 2000, "Spatial Hearing Mechanisms and Sound Reproduction", University of York, at http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm, [accessed 3 August 2000].

Mix Professional Recording Magazine, California: Primedia Inc., monthly.

Nicely, Jake, 2000, "Inside the Band", *Mix Magazine*, August 2000, California: Primedia Inc.

Osborn, Ed, 1996, "Creaking Grounds: Tectonics and SoundCulture 96", in *Essays in Sound 3, Diffractions,* ed. S Davies, A Jonson, A Cavallaro, Sydney: Contemporary Sound Arts.

Plenge, G, 1974, "On the Differences between Localisation and Lateralisation", in Journal of the Acoustical Society of America, Volume 56

Rumsey, Francis, 2001, *Spatial Audio*, Boston: Focal Press.

Steinmetz, Chris, "Percussion Power and Mystique", *Mix Magazine*, August 2000, California: Primedia Inc.

Surround Sound Professional, September 2000, New York: United Entertainment Media, Inc., monthly.

Warren, Richard, 1999, *Auditory Perception, A New Analysis and Synthesis*, Cambridge: Cambridge University Press.

Worrall, D, 6 Sep 2001 13:24:39, email to Suround list sursound@music.vt.edu

Opportunities for Evolutionary Music Composition

Andrew R. Brown

Queensland University of Technology
Email: a.brown@qut.edu.au

Abstract

Traditional compositional techniques and many cmputer-assisted composition systems have been focused on the production of linear musical products. In an age where non-linear media are increasingly prominent there is a need to reassess these technologies in the light of new opportunities for making music with non-linear outcomes. This paper examines the current state of music making for non-linear media with a particular emphasis on evolutionary music and how and where it might be applied. In addition, some of the implications for computer-based tool design will be outlined.

1 Introduction

Music making has long involved interpretation and improvisation, but with the advent of sound recording technologies musical outcomes could be frozen and reproduced unchanged. This development had many ramifications for the music community including the extension of commodification from scores and instruments to include recorded performances, and the establishment of musical styles and aesthetics where recorded products were the intended outcome. The freezing of sound as recordings is central to electroacoustic music that dominates the computer music community. However, despite the advent of programmable digital music systems where static reproduction was no longer required, the traditions of tape music persisted. Responding to the non-linear character of digital media, computer musicians reclaimed interpretive and improvisational practices as hybrid combinations of recorded and live elements in human-computer partnerships, often referring to this as interactive composition or interactive performance. The programming of systems that created music autonomously, generative music, was also pursued. Evolutionary music relates to these performative and generative traditions in a hybrid computer music form that could be labelled, generative improvisation. This paper explores the motivations for and techniques of evolutionary music and suggests that there is significant potential in this form as a vehicle to invigorate computer music making.

Evolutionary music is music that changes over time or in response to external variations, such as interaction with a user or performer. Evolutionary music involves a feedback mechanism such that the current state is based upon previous states and circumstantial conditions, as distinct from traditional musical structures that are often deterministic. For example, traditional compositional techniques focus on the musical function of particular attributes, such as harmony, but research into artificial life suggests that behavioural emergence replaces traditional functional hierarchy (Hendriks-Jansen 1996). In evolutionary music systems, structural outcomes emerge from an iterative process in real-time, rather than being specifically designed from start to finish. Emergence, in this sense, involves the historical development of musical complexities from simple traits and forms of organization (Nagel 1961). The composer or designer of evolutionary music requires new techniques that focus on creating classes of musical potential, as opposed to existing techniques that describe predefined linear outcomes.

The rise of evolutionary music marks a shift in computer music research from the study of musical structures to an examination of the dynamic interaction between aspects of musical systems. There have been some significant advances in this area by researchers examining interactive performance, notably the work of Robert Rowe (1993, 2001), and evolutionary systems can build on this work and extend it into semi or fully autonomous systems. This shift requires the development of new compositional theories and techniques which are often influenced by biological evolutionary systems. These techniques support designers of both interactive and autonomous music systems. The need for a new class of compositional techniques is mainly due to the dynamic nature of most evolutionary music systems. In such systems musical structure becomes undeterminable the use of structuralist compositional paradigms is ineffective. The development of a new class of algorithmic compositional techniques will assist music making with non-linear digital media, and add to the traditional musical forms and processes which assume a predefined linear musical narrative.

An example of the non-linear musical forms are those that have indefinite durations and unpredictable interruptions due to changing contexts or user interactions, in such cases evolutionary music systems can play a significant role. It is important to note that musical narrative is not lost in such systems, it is just not entirely predetermined, rather, narrative emerges as the music evolves. This emergent feature of evolutionary systems presents a new challenge to composers, system designers, and music theorists. This challenge is an opportunity for the developers of new techniques and tools to aid composers.

Activities in evolutionary artistic practice have a significant presence in Australia, as recently seen during the series of "Iteration" conferences held in Melbourne. Notable Australians working in this area include composers Rod Berry and Alan Dorin, visual artist Paul Brown and digital art social theorist Mitchell Whitelaw. Australian work builds upon international activity in evolutionary music in Europe, particularly by Eduardo Miranda at Sony CSL in Paris, by a individual researchers in America, in particular David Tudor, John Biles, Bruno Degazio, Robert Rowe, Gary Lee Nelson and David Cope. Evolutionary music is a field developing in the wake of the expanding interactive media explosion which has recently seen the computer game market exceed the film industry in gross sales, the recognised need for cultural content for broadband internet, and the current roll out of second and third generation mobile phone networks worldwide. It seems that it is now time for evolutionary music to flourish.

2 Evolutionary processes in the creative arts

The use of evolutionary metaphors for artistic purposes has been documented by Mitchell Whitelaw (2002) who traces its origins back to Richard Dawkins' work on *Biomorphs* (Dawkins 1987). *Biomorphs* is a program that generates and evolves graphical stick figures. This program inspired other evolutionary visual artists, including William Latham and Stephen Todd who worked on evolving two-dimensional images in the 1980s and 1990s (Todd and Latham 1992), and Karl Sims' work with three dimensional images in the 1990s (Sims 1994). Work in evolutionary art continues to be dominated by visual artists, with only a few researchers applying the techniques to music or sound. There are great opportunities to transfer the techniques from the visual to the sonic realm and move beyond the quite literal adherence to biological metaphors. In the visual arts, there has been recent deviation from the strict biological metaphor where features of an entity are changed, to a more abstract evolutionary model where rules and structural constraints are evolved. An example of this work is the computational art works produced by Erwin Driessens and Maria Verstappen in the late 1990s, and presented at *Iterations 2* in Melbourne (Driessens and Verstappen 2001). Research into evolutionary music has begun to appear in the past decade, but was proceeded by research into generative and interactive composition which addressed similar issues.

Despite the dominance of biological metaphors in evolutional computation research outside of artistic fields, music and other evolutionary digital art are best considered as a cultural phenomena adhering to the characteristics of cultural change, such as being relevant to the short time spans to human perception and expectation, but acknowledging the longer time spans of changes in social preferences and cultural values. The differences between biological and cultural evolution are clearly articulated by Yri Lotman from the perspective of semiotic hermeneutics.

> The evolution of culture is quite different from biological evolution; the word 'evolution' can be quite misleading. Biological evolution involves species dying out and natural selection. The researcher finds only living creatures contemporary with him. Something similar happens in the history of technology: when an instrument is made obsolete by technical progress it finds a resting place in a museum, as a dead exhibit. In the history of art, however, works which come down to us from remote cultural periods continue to play a part in cultural development as living factors. A work of art may 'die' and come to life again; once thought to be out of date, it may become modern and even prophetic for what it tells of the future. What 'works' is not the most recent temporal selection but the whole packed history of cultural texts (Lotman 1990:127).

In a cultural evolutionary model, musical material as conceived as *memes* (idea cells) rather than genes (Dawkins 1976). A notion of evolutionary music as stylistic morphing depicts compositions that change in musically meaningful ways as they adapt to different situational contexts. Techniques of evolutionary music have been strongly influenced by biological evolution as a metaphor and therefore computational processes from the fields of genetic algorithms, artificial intelligence, and artificial life (Langton 1989) have often been inspirational. The computational processes used include genetic algorithms, Markov transitions, genetic programming, fuzzy logic, and cellular automata, however, the primary goal of the evolutionary music system designer and researcher is the development of new musical understanding and compositional techniques rather than the sonification of computational formalisms. New musical knowledge derived from the development of evolutionary music systems demonstrates that digital media can expand upon, rather than replace, the conventions of musical composition.

3 Identifying stylistic features

In order to generate appropriate musical material using evolutionary processes it is first necessary to establish features of the desired musical style and to set targets and bounds to assist the composer and algorithm to navigate musical space. Well-established theories of diatonic music and compositional texts based on composer experience have previously been used because they have provided useful guidelines in previous research (Towsey, Brown, Wright, and Diederich, 2000). Some research in this area has focused on superficial features, that Rowe in his book *Interactive Music Systems* calls "Level-1 analysis," including pitch, loudness, duration, contour, and harmony (Rowe 1993:122). Successful music evolution over several minutes or hours requires examination beyond this level of analysis, in particular consideration of the temporal nature of musical structure; as has been identified as crucial to musical development in studies of music theory (Lerdahl and Jackendoff 1983) and in computer modelling of musical style (Cope 1991, 2000, 2001). At an even larger scale, attributes of musical changes from one style to another could be examined, however, most research to date has focused on compositional evolution within one style and have payed attention to event organization (Miranda 2001), timbral evolution via changes in synthesis, or adjustments in performance interpretation, or interactive improvisation (Rowe 1993, 2001, Biles 1994).

Evolutionary music provides an effective means to tackle one of the aesthetic goals that all machine systems must resolve when simulating human action; the trade-off between *novelty* and *structure*. In short, evolutionary music systems can build on the structure of human compositional practice like rule-based systems, while providing techniques that offer emergent variations that add surprise and novelty, hopefully to an appropriate degree. Of course, measures of aesthetic appropriateness are at the discretion of the composers who create evolutionary music systems, and their audiences. As with all musical systems, evolutionary music is stylistically constrained by a set of compositional heuristics dependent upon the composer's preferences and intention. The success of evolutionary music systems should be judged against the aesthetic appropriateness of the music they produce and by their value to composers in creating music for non-linear media. A challenge for the computer music community is to create evolutionary music systems that are successful against these criteria. The development of computer-assisted techniques will provide composers with a broader pallet of creative opportunities and reposition compositional practice so that it is ready to confront the non-linear media expansion sure to characterise the early part of this century. The introduction of evolutionary compositional processes will transform composition for digital systems to a degree comparable to the advent of audio recording which enabled music to be kept indefinitely static.

A significant challenge for designers of evolutionary music systems is to adequately represent a musical world in which to evolve. This is a problem common to artificial intelligence and artificial life researchers, which they commonly call the "frame" problem, describing how much of the context to represent such that the computational world is sufficiently comprehensive to enable relevant decisions to be made, but not so complex as to be unworkable. In the case of evolutionary systems, this manifests itself when system designers try to place appropriate constraints on pitch and rhythm values, make decisions about how much memory of prior musical events is maintained for future reference, decide how the events in one musical part should influence other parts, and so on. It is at times like these that computer musicians appreciate the enormous accumulation of knowledge and skill that an instrumental improviser performing in a live ensemble maintains.

A secondary challenge for designers is to appropriately map evolutionary processes to musical parameters. Having decided upon the musical representation and evolutionary model, for example using the jMusic data structure and emulating Darwinian selection, the task is to determine which aspects of notes, phrases or timbres will be important in judging fittness and by what criteria. In particular, the complexity of this task is increased because there is rarely a clear goal for musical outcomes. That is, there is not just one *good* musical solution to a given situation, and evolutionary processes often require such ranking of possible next steps. According to Margaret Boden (2002), an eminent writer on the philosophy of artificial life, evolutionary algorithms and processes have two characteristics. First, they have a way of changing or adapting their own rules and, second, a way of selecting from the array of possibilities available through change. Boden underlines the difficulty of establishing evaluative criteria for making selections from all possible changes, and automating this process in generative algorithms. She suggests that because humans design the algorithms, finally the key issue is human preference and epistemology. In the field of music this relates to musical understanding and aesthetics, which is why evolutionary music system development is fundamentally a musical project, despite having interesting and necessary computational aspects.

These challenges notwithstanding, the field of evolutionary music has great potential that will only increase as non-linear modes of music delivery become more prevalent. In his summary of the future of music systems, the computer-assisted music specialist Paul Berg suggests that potential trends in composition include a "radical change in the non-linearity of current media" leading to a redefinition of the role of the composer. To meet this challenge he maintains that "useful musical generators are needed. They should reflect usable and general concepts that can be applied to create musical expressions" (Berg 1996:26). Evolutionary music system designers have an opportunity to meet this challenge head on by developing generative improvisational processes that can contribute to musical expression in non-linear musical circumstances.

4 Compositional techniques for evolutionary music

Techniques previously used in computer-generated music have predominantly been re-combinatorial or knowledge-based and, as such, have been limited by their inability to introduce novelty and variation. The use of artificial intelligence techniques has provided some success in generating novelty, for example, the use of neural networks (Mozer 1991, Tudor 1995), augmented transition networks (Cope 1997), and genetic algorithms (Degazio 1996, Towsey et al. 2000, Biles 2002). There are opportunities to extend this work to include techniques of artificial life (Boden 1996, Resnick 1994, Miranda 2000) and add complexity by looking at emergent musical behaviour and dynamic environments where musical goals are linked to unpredictable situational changes.

The use of heuristic principles for automated music composition is well established in computer music (Moorer 1972, Laske 1992) but the use of heuristics based on cultural-evolution is less well established. In previous studies, the stylistic objectives have been based on a fixed historical style enabling Darwinian evolutionary fitness objectives to be employed, in particular in the use of genetic algorithms. Effective evolutionary music systems will need to move beyond imitative musical processes (however complex) to establish new techniques of composition that focus on generative music making (Miranda 2002), and move beyond simplistic notions of optimised fitness attainment to embrace broader directions for progression as evident in cultural systems. Examples of evolutionary music systems can be described as falling under particular categories.

Behavioral Recombination. This style of system draws on aspects of previous musical examples, as for example in Ames and Domino's *Cybernetic Composer,* David Zicarelli's *M*, and David Cope's *EMI* software. These systems are generative and can become evolutionary if provided with a feedback mechanism, such as the readjustment of probability weightings in a Markov transition matrix at each iteration by adding the data from the just-generated score.

Cellular Automata. These systems have a number of elements (cells) that change state according to rules related to the state of neighbouring cells. These rules are applied at each iteration creating an ongoing variety of state situations across the system. Cellular automata processes were one of the earliest artificial life computation models. The cell and their state can relate to any musical parameter, but a feature of cellular automata is that change is progressive rather than revolutionary (unlike random changes), and at times oscillating patterns can occur that provide some stability. Examples of cellular automata musical systems include Eduardo Miranda's *ChaOs* and Kenny McAlpin's *CAMUS* system, and the closely related *Boolean Sequencer* of Alan Dorin.

Genetic Algorithms. Directly related to biological evolution, genetic algorithms process data as a string of 'genes' in a virtual genome. Changes to which are traditionally done by mutating the state of a gene at random, and applying crossover techniques where sections of one genome are recombined with a section from another. A population of genomes are created and some survival-of-the-fittest selection criteria ranks members of the population and keeps those that are fittest and discards those that are weakest. In a simplistic musical example the pitch of notes in a melody can be used as a genome (each pitch is a gene), and pitches mutated by transposition and fitness judged by melodic coherence to rules of voice leading. Examples of cellular automata music systems include John Biles' *GenJam* and the *Vox Populi* software of Artemis Moroni and his colleagues.

Evolvable Hardware. As well as the more prevalent software systems for evolutionary music there are also a few hardware-based systems. Evolvable hardware uses special chips which are reconfigurable and a software program, often based on generative algorithms, that iterates through various recofigurations. The hardware is tested at each iteration to see if it performs the desired task; for example, playing or recognising a musical tone. Interesting results come from the fact that the system tries circuit design patterns that are unlikely to be concieved by human designers, and interesting results (and sounds) can occur. An example of this work is *The Sound Gallery* by Woolf and Thompson.

5 Composing evolutionary music with a computer

Computer-assisted composition is an active area of musical, technical, and humanistic research. Evolutionary music systems do not need to be autonomous systems, but can also extend the role of computer-assisted composition to include semi-active participation through the automated-evolution of new and varied musical material. The computer has traditionally acted as a compositional assistant in numerous ways which can be differentiated as modes of compositional engagement (Brown 2000). Computer systems are generally used for musical presentation, usually in the form of notated scores, or audible rendering to MIDI files or audio recordings. They have been less commonly used for compositional support via algorithmic design, and rarely for the design of evolutionary algorithms.

Computer-assisted composition processes are used to some degree in almost all commercial, artistic, and academic music making. Typically, this involves the externalisation in some computer model which simulates either a printed score or an audio recording device. The composer can then manipulate the model in some way to produce a final composition. All current commercial systems are of this type. Commercial solutions that seek to address the non-linear nature of digital media are beginning to enter the computer game market; these include *Direct Music* (Microsoft), *MusyX* (Factor 5, Nintendo) and *Koan* (SSEYO/TAO). These commercial entrants highlight the growing interest in music for non-linear media. However, the current implementations employ recombinatorial and stochastic processes, rather than evolutionary processes. Active research into evolutionary music is underway at Sony CSL, in Paris (Miranda 2002) primarily focused on the use of cellular automata, and by isolated researchers in other locations (Biles 1994, Degazio 1996, Dorin 2000, Rowe 2001, Cope 2001, Todd & Werner 1998).

Musical systems that focus on evolutionary music have so far been limited to tightly constrained environments or they have dealt with limited musical material or concepts. Examples of the former include Rod Berry's works "Feeping Creatures," "Gakki-mon Planet," and "Listening Sky." Each of these works in an intentionally limited musical domain with no intention to provide a general compositional pallet. There is clearly an opportunity for development of more generalised evolutionary computer music tools for composers.

6 Conclusion

The field of evolutionary music shows significant promise as a branch of computer music for use in computer-assisted composition and generative music. The use of evolutionary music in non-linear delivery media is particularly pertinent given the general expansion of this style of music through platforms such as DVD and computer games. This paper has explored the history and current practice in evolutionary music and identified important issues and areas for future exploration in the field. Evolutionary music, and evolutionary art in general, can advance both artistic practice and contribute to artificial life research by introducing models of evolution that take into account cultural development that enhance existing models based on biological evolution. Evolutionary music adds computational improvisation to the expressive opportunities of the computer musician.

References

Ames, C and Domino, Michael. 1992. "Cybernetic Composer: An overview," in *Understanding Music With AI: Perspectives on Music Cognition*. Balban, M., Ebcioglu, K. and Laske, O. eds. Menlo Park: The AAAI Press, pp.186-205.

Berg, P. 1996. "Abstracting the Future: The search for musical constructs," *Computer Music Journal,* 20(3):24-27.

Biles, J. A. 1994. "GenJam: A genetic algorithm for generating jazz solos," in the proceedings of the *International Computer Music Conference*. San Francisco: ICMA,

—. 2002. "GenJam: Evolution of a jazz improviser," *Creative Evolutionary Systems*, P. J. Bently and D. W. Corne, eds. San Francisco: Morgan Kaufmann, pp. 165-188.

Boden, M., ed. 1996. *The Philosophy of Artificial Life*. Oxford: Oxford University Press.

—. 2002. "Foreword," in *Creative Evolutionary Systems*, P. J. Bently and D. W. Corne, eds. San Francisco: Morgan Kaufmann.

Brown, A. R. 2000. "Modes of Compositional Engagement," *MikroPolyphony,* 6 http://farben.latrobe.edu.au/mikropol/volume6/brown_a/brown_a.html

Cope, D. 1991 *Computers and Music Style*. Oxford: A-R Editions.

—. 1997. "The Composers Underscoring Environment: CUE," *Computer Music Journal,* 21(3): 20-37.

—. 2000 *The Algorithmic Composer*. Madison: A-R Editions.

—. 2001 *Virtual Music: Computer synthesis of musical style*. Cambridge, MA: MIT Press.

Dawkins, R. 1976 *The Selfish Gene*. Oxford: Oxford University Press.

—. 1987 *The Blind Watchmaker*. Essex: Longman Scientific and Technical.

Dorin, A. 2000. "Boolean Networks for the Generation of Rhythmic Structure." in proceedings of *Interfaces: The Australasian Computer Music Conference*, A. R.

Brown and R. Wilding, eds. Brisbane: Australasian Computer Music Association, pp. 38-45
Degazio, B. 1996. "Evolution of Musical Organisms." Proceedings of the International Computer Music Conference. San Francisco: ICMA.
Hendriks-Jansen, H. 1996. “In Praise of Interactive Emergence, Or Why Explanations Don't Have to Wait for Implementations,” in *The Philosophy of Artificial Life*, M. Boden, ed. Oxford: Oxford University Press, pp. 282-302.
Langton, C. G. 1989. “Artificial Life,” in *Artificial Life: Proceedings on an interdisciplinary workshop on the synthesis and simulation of living systems*, C. G. Langton, ed. Redwood City: Addison-Wesley.
Laske, O. 1992. “Artificial Intelligence and Music: A cornerstone of cognitive musicology,” in *Understanding Music with AI: Perspectives on music cognition*, M. Balban, K. Ebcioglu and O. Laske, eds. Meno-Park: AAAI Press, pp. 3-29.
Lerdahl, F. and R. Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
Lotman, Yuri M. 1990. *Universe of the Mind: A semiotic theory of Culture*. Bloomington: Indiana University Press.
Miranda, E. R. 2001. Composing Music with Computers. Oxford: Focal Press
—. 2002. “On the Origins and Evolution of Music in Virtual Worlds,” in *Creative Evolutionary Systems*, P. J. Bently and D. W. Corne, eds. San Diego: Academic Press, pp. 189-204.
Moorer, J. A. 1993. (1972) “Music and Computer Composition,” in *Machine Models of Music*, S. Schwanauer and D. Levitt, eds. Cambridge, MA: MIT Press, pp. 167-188.
Moroni, A., Manzolli, J., Van Zuben, F. and Gudwin, R. 1990. "Vox Populi: An interactive Evolutionary System for Algorithmic Music Composition." *Leonardo Music Journal*, 10: 49-54.
Mozer, M. C. 1991. “Connectionist music composition based on melodic, stylistic, and psychophysical constraints,” *Music and Connectionism*, P. Todd and G. Loy, eds. Cambridge, MA: MIT Press, pp. 195-211.
Nagel, E. 1961 *The Structure of Science: Problems in the logic of scientific explanation*. London: Routledge & Kegan Paul.
Resnick, M. 1994 *Turtles, Termites, and Traffic Jams*. Cambridge, MA: The MIT Press.
Reynolds, C. W. 1987. “Flocks, Herds, and Schools: A distributed behavioral model,” *Computer Graphics,* 21(4): 25-34.
Rowe, R. 1993. *Interactive Music Systems: Machine listening and composing*. Cambridge, MA: MIT Press.
—. 2001 *Machine Musicianship*. Cambridge, MA: MIT Press.
Sims, K. 1994. “Evolving 3D Morphology and Behaviour by Competition,” in *Artificial Life IV*, R. A. Brooks and P. Maes, eds. Cambridge, MA: The MIT Press,
Todd, P. and G. M. Werner. 1998. “Frankensteinian Methods for Evolutionary Music Composition,” *Musical Networks: Parallel distributed perception and performance*, N. Griffith and P. Todd, eds. Cambridge, MA: MIT Press,
Todd, S. and W. Latham. 1992. *Evolutionary Art and Computers*. London: Academic Press.
Towsey, M., A. R. Brown, S. Wright and J. Diederich. 2000. “Towards Melodic Extension Using Genetic Algorithms,” in proceedings of *Interfaces: The Australasian Computer Music Conference*, A. R. Brown and R. Wilding, eds. Brisbane: Australasian Computer Music Association, pp. 85-91.
Tudor, D. 1995. *Neural Synthesis Nos.6-9*. CD, Electronic Music Foundation.
Whitelaw, M. 2002. “Breeding Aesthetic Objects: Art and artificial evolution,” in *Creative Evolutionary Systems*, P. J. Bently and D. W. Corne, eds. San Diego: Academic Press, pp. 129-146.

The Application of Mapping in Composition and Design

Paul Doornbusch
Interactive Information Institute, RMIT University
email: pauld@iii.rmit.edu.au

Abstract

Mapping concerns the connection between structures, or gestures and audible results in a musical performance or composition. This is important to computer music composition as most of it is algorithmic in some way or another and it involves mapping. Algorithmic composition is sometimes the process of imagining a gesture or structure - perhaps physical or visual - and then applying a mapping process to turn that structure of the conceptual domain into sound which may display the original conception in some way. This article looks at mapping from the point of view of Australasian algorithmic composition practice, particularly where persistence is an issue, such that the structure (conceptual domain) is embodied and perceptible in the musical result. An attempt is then made to draw some parallel to the practice of (visual) design that uses mapping and examine the similarities and differences with algorithmic music composition.

1 Introduction

Mapping as part of algorithmic composition has not been as thoroughly investigated as mapping in instrument design. This is possibly because it is usually an integrated part of the composition process and not seen as a discrete stage of practice. Mapping in algorithmic composition is different from mapping in instrument design because composition is a process of planning and instruments are for real time music production. The use of mapping which is integral to an instrument has both differences and similarities to the mapping that is an inherent part of algorithmic composition. This paper investigates how some Australasian composers use mapping in their practice and moves towards a taxonomy of the process.

Another aspect of this investigation is how mapping is used by architects and designers. Some designers say that they explicitly use mapping in their practice. There is an attempt to analyse this in a small way and see if there are parallels with composition practice.

2 Defining The Terms

Compositional structures and ideas can take many forms, but they are often abstract in some way, to a greater or lesser degree, from the music that is composed. Composers sometimes use visual ideas of shapes, mathematical functions, physical processes, or phenomena and so on as ideas for creating music. Mapping is the process of taking the (possibly abstract) compositional structures and generating musical parameters. Gesture is an ill-defined term, which for the purposes of this paper means a musical concept; it is not a physical movement.

A musical gesture is a planned change (randomness can be planned) in musical parameters as part of a piece of music. The parameters could be, for example; timbre, density, intensity, timing, pitch, and so on. A compositional gesture is the underlying conception, structure and planning of the musical gesture.

"Conceptual domain" is a term that will be used to cover the entire conceptual area of compositional practice, which includes other organisation strategies as well as compositional gesture.

3 Mapping In Composition

Algorithmic composition can be the practice of composing music through manipulating a number of compositional concepts, structures or gestures to produce a musical piece. The manipulation is usually part of the compositional process, where a multitude of gestures or structures are arranged and then through another process, possibly very direct, musical parameters are extracted. This is most usually done with the aid of a computer, although it should be noted that this is usually a matter of efficiency and expediency only. It is possible to do this from the micro-composition scale of sound synthesis (where it can only be done with a computer) through to the macro scale of the structural model of the piece. Mapping is the process of extracting the musical parameters from the compositional gestures or structure, such that one set of data is *mapped* onto another. It should also be noted that algorithmic composition might use specific organisations of data that do not necessarily require “gestures”; for example, Markov chains, finite-state machines, fractals and space grammars. These constructs may also require significant mapping processes to output musical data.

There may appear to be some similarities between the methods of mapping in algorithmic composition and the mappings used in the general practice of the sonification of data. This can sometimes be a valid observation with some pieces of music, and there is certainly some overlap between the two areas. However, as the intent of sonification is to illustrate a process or some data, it is primarily instructional. This type of mapping is not the focus of this paper, which is primarily about *composition* and mapping, which is somewhat more ephemeral. The closest sonification comes to composition, what Larry Polansky calls “manifestation” (Polansky 2002), is where the intent is to creatively use a type of sonification of a formal or mathematical process to create a musical idea. For the purposes of this paper, that process will be included as a part of composition.

4 Historical Examples

Mapping has been used by some composers for many years, although it may not have been identified as a separate element of practice. These historical examples will help to explain the function of mapping, showing how it has been used by composers.

A famous example of the mapping process is the part of *Pithoprakta* (1955-56) as reproduced in Formalized Music pp17-21 (Xenakis 1991). Xenakis used the Brownian motion of gas particles, combined with Bernoulli’s Law of Large Numbers, as his basic model for the cloud pizzicato-glissandi section. After calculating, statistically, over 1000 velocities of gas particles at given instants of time (the measurement of this was impossible), he then graphed them on an XY plane and directly mapped the straight lines of the velocities to glissandi for 46 string instruments. Xenakis divided his graph vertically into 15 pitched sections, each corresponding to a major third. This was then mapped to the ranges of the string instruments. The mapping was directly of pitch in the vertical direction. This is a particularly direct and concrete example of mapping. All intensities and durations are the same, but to ensure the sensation of a cloud of particles, Xenakis used a complex temporal arrangement of overlapping timing subdivisions that are factorially-unrelated. This is a complex mapping of linear time, designed to represent the instantaneous nature of the movement of the gas particles. Along with many other algorithmic composers, the mapping phase is implicit in much of Xenakis’ work. He often used direct mapping as a result of the deliberate organisation of one set of data in such a way that it maps directly to musical parameters.

Another famous example of algorithmic composition and mapping is Charles Dodge’s *Earth’s Magnetic Field* (1970). Here, Dodge used data from the effects of the radiation of the sun on the magnetic field of Earth. A Bartels diagram showed fluctuations in the Earth’s magnetic field for 1961 and this data formed the basis for the piece. Dodge mapped this data, the Kp index (a measure of the average magnetic activity of the earth) to pitches and rhythms. From the program notes of the recording of *Earth’s Magnetic Field* (Dodge, 1970), we may glean an insight into the mapping used:

"The succession of notes in the music corresponds to the natural succession of the Kp indices for the year 1961. The musical interpretation consists of setting up a correlation between the level of the Kp reading and the pitch of the note (in a diatonic collection over four octaves), and compressing the 2,920 readings for the year into just over eight minutes of musical time." (Dodge, 1970)

While the pitches appear to be a direct mapping from the Kp index, some elements of the composition such as the timbres were chosen purely for aesthetic effect. An arrangement of the data that plotted the length of sections of the data against the maximum amplitude in the section was used to determine the speed and direction of the sound spatialisation and the rhythms. The data was also sometimes read multiple times to generate the musical parameters. That, combined with the similarity of the fluctuations in the Kp index to 1/f noise data, contributes to the aspects of self-similarity in the piece.

The two previous examples contrast different approaches to mapping. It may be linear and direct, but it may also be nonlinear and more complex. Both examples use the data as a structural component and the music achieves some structural unity for that. Also, both of these examples come from a time when computer music practice was in its infancy. With current tools and thinking, composers can use mapping in much more creative and indirect ways.

5 How Mapping Is Used

There has not been the same formalisation of algorithmic composition and mapping as there has been for other musical composition. Additionally, there is little formal analysis of music in terms of mapping or compositional gesture. As there is no set method for defining the mappings or structures, each composer tends to use their own methods for their own reasons.

Compositional structures and mappings are also used differently by different composers, thus making this a problematic area for analysis and study: For example, the structural model and mapping for a piece of music will have different applications for a composer who is focused on, say, spectral composition in contrast to another focused on some other form of algorithmic composition.

It is clear from descriptions of how composers use mapping between the conceptual or abstract domain and the musical domain, that it is sometimes used in a similar combination of ways to how it is used in instrument design. That is; mapping one compositional parameter to many musical parameters (one to many), mapping many compositional parameters to one musical parameter (many to one) and mapping many compositional parameters to many musical parameters (many to many). Additionally, there is another parallel between instrument designers and composers, the mappings themselves may be linear mappings or nonlinear mappings (Hunt and Wanderley 2000, Hunt, Wanderley and Kirk 2000). However, in compositional mapping there is the additional possibility of repetitive nonlinear mappings and most importantly, each composer has their own combination of mapping techniques.

From the comments made by the architectural designers questioned, it seems that there is some overlap in the use of mapping to move from the abstract to the concrete. This is an area that is only touched upon here, to demonstrate that this is (yet another) intersection between music and architectural design.

6 Questions Of Mapping

To understand how Australasian composers use mapping, a number of them were asked questions concerning their practice. The broad answers are summarised below. The composers were; Roger Alsop, Rodney Berry, Chris Cree Brown, Phil Brownlee, Warren Burt, Densil Cabrera, Tim Kreger, Peter McIlwain, Gordon Monroe, Garth Paine, Greg Schiemer, and Paul Doornbusch. The five designers questioned were Pia Ednie-Brown, Tom Kovak, Paul Minifie, Vivian Mitsogianni and Julian Raxworthy. For the designers the questions were modified to exchange “architectural design” for “composition”, “viewer” for “listener” and “visual output” for “musical output”, and so on. The questions, asked were:

1. Is mapping something that you are conscious of when you are composing?

2. Do you have a consistent approach to mapping and (algorithmic) composition or does it vary and why?

3. When implementing a mapping strategy for (part of) a composition, do you organise this in a particular ‘analytical’ way (decomposing the problem in a technical manner), or in a more creative and holistic way for a purely aesthetic result?

4. Is the mapping component of your compositions something that might be perceptible by a listener, or of interest to them and why?

5. Is the mapping component of algorithmic composition something that is pre-determined for you or is it part of a process of exploration?

6. Do you use individual mapping strategies for individual parameters or is there reuse of mapping strategies or a global system? (ie are they monoparametric or multiparametric?)

7. What elements do you control algorithmically in a composition and can you comment on the function and importance of these?

8. Is the mapping consistent within these elements or not?

9. Are the mapping schemas you use mostly linear mappings or non-linear in some fashion, and why?

10. Can mapping be considered a composition technique in itself?

There was deliberate overlap in the questions to elicit as much information as possible and the respondents were also asked to make a yes/partly/no, multiple-choice response to each question.

7 Responses

The results of the composers' responses are summarised below, making it possible to note similarities and divergences. The answers supplied by the composers will be provided in an appendix to another, larger, paper that can be downloaded or will be sent to anyone requesting it. In this summary, it is possible to achieve the an initiall overview of this area.

7.1 General remarks by respondents:

While composers may use mapping to move from the conceptual domain to musical parameters, mapping can work backwards for listeners, providing access to structures and concepts from the music. Composers tend to use complex ideas and structures for music. Focusing on mapping assumes that something pre-exists before the act of composition and not all composers may think or work like that. If mapping is a means of moving from the conceptual domain to the concrete domain, then it may be important for all composition, but it is central to algorithmic composition.

7.2 Collated Responses:

The tables below show the general response versus the number of composers (C) or designers (D) who had that response. While the number of people questioned means that this is a statistically minuscule sample, it does point to trends and variations.

Question 1.	C	D
General Response	#	#
Yes, I am conscious of using mapping	8	3
Sometimes I am conscious of using mapping	3	2
No, I never consciously use mapping	0	0

Question 2.	C	D
General Response	#	#
I have a consistent approach to mapping	0	0
I sometimes have a consistent approach to mapping	5	3
I do not have a consistent approach to mapping	6	2

Question 3.	C	D
General Response	#	#
I mostly use mapping analytically	0	1
Mapping is sometimes used analytically	3	2
I mostly use mapping creatively	6	2

Question 4.	C	D
General Response	#	#
Mapping is of great interesting for listeners	2	1
Sometimes mapping is interesting to listeners	9	4
Mapping is not interesting to listeners	0	0

Question 5.	C	D
General Response	#	#
Mapping is predetermined in composition	1	0
Mapping is sometimes predetermined, sometimes part of exploration	4	2
Mapping is part of the exploration of composition	6	3

Question 6.	C	D
General Response	#	#
I use individual mapping strategies for individual parameters (monoparametric)	2	0
Both individual mappings strategies are used as well as reusing some	7	2
I reuse mapping strategies for multiple parameters (multiparametric)	1	3

Question 7.	C	D
General Response	#	#
I usually control as many elements as possible	2	1
I control most elements sometimes, fewer at other times	4	0
I usually control some elements only	5	2

Question 8.	C	D
General Response	#	#
The mapping is mostly consistent for each element	9	3
The mapping is sometimes consistent for each element	1	0
The mapping is mostly not consistent for each element	1	2

Question 9.	C	D
General Response	#	#
Mostly linear mappings are used (very complex structures)	1	1
A mixture of linear and nonlinear mappings are used (depends on structure)	7	1
Mostly nonlinear mappings are used (simpler structures)	3	3

Question 10.	C	D
General Response	#	#
Mapping can be a compositional technique in itself	6	3
Mapping is part of compositional technique	5	2
Mapping cannot be considered a compositional technique at all	0	0

8 Types Of Mapping

From the composers' descriptions of how they use mapping, it seems that there is no set method for mapping data from the domain of the conceptual, gestural or structural to the musical domain. Linear and simple mappings are *sometimes* used by composers, but it is used significantly less often than mappings that are more complex.

Simple mappings have been most clearly described as ratiometric (Polansky 2002). In this type of mapping, a doubling of the data to be mapped results in a "doubling" of the musical parameter. This could be, for example, an octave pitch displacement or a doubling of the loudness. Note that this second linear mapping might be perceptually linear to a particular person under particular conditions, but it would be mathematically nonlinear. Because ratiometric mappings are the easiest to perceive they are particularly useful when the composer has some data, gesture or concept that should be translated as directly as possible for the listener. The previous example of the pitch mappings from Xenakis' Pithoprakta is an illustration of this. This particular example is musically successful because the data to be mapped is very complex from the outset. Simple mappings can be less musically successful if the data to be mapped is very simple.

Complex mappings may be regularly nonlinear, such as an exponential law. For example, a square law will produce a fourfold increase in the musical parameter from a doubling of the data to be mapped. If this is perceptually based it can be perceived as a more extreme or less extreme mapping with changes in the data to be mapped. Other types of complex mappings are as varied as can be imagined. They can sometimes be related to a complex arrangement of the data to be mapped, or other potentially chaotic functions can be involved. This may obscure the original concept, compositional gesture, or data, or it may embellish it and give it another dimension. However, there is clearly a limit to how far such complexity in mapping can be taken before all sense of the original data is lost in the mapping. Polansky convincingly suggests that the cognitive weight of complex mappings degenerates rapidly and nonlinearly such that beyond a certain point everything is just "complex".

It has been a repeated outcome from instrument design research that humans prefer complex mappings (Hunt and Kirk 2000). This would appear to translate somewhat into the domain of composition, given the previous situation that linear mappings may be most appropriate for very complex data.

In one way, the variety of sophisticated, creative and exploratory approaches to mapping, as embodied in the works of algorithmic composers, is a wonderful outcome and it ensures the variety of music from algorithmic composition. In another way, it means that mapping strategies are reinvented by every composer and young composers do not have easily accessible models to work from. If there was created a categorisation of compositional mapping strategies and a catalogue was built of mapping techniques, then other composers could refer to this and build on it in their own practice. This would be of particular use to student composers (Polansky 2002). In this way, mapping in algorithmic composition could be demystified and more complex, varied and musically appropriate practices could be developed by building on the work of others. The counter argument to this is that there is something valuable in the effort of developing sophisticated mappings for oneself. In instrument design there is certainly interest in the effort required to play an instrument and how mapping relates to this (Ryan 1992).

The answers from the designers are interesting for a number of reasons. There appears to be some degree of overlap in the way that composers use mapping and how architectural designers use it. This is an additional parallel than those that are usually drawn between architecture and music; that they both communicate through, and are deeply concerned with, the concept of form, and that one divides space while the other divides time. The similarities in the ways both designers and composers use mapping, and also how they do not use mapping (the 0 scores in the tables above), points to some similarity in the creative process between the disciplines. However, I think there is both too much semantic variation in the understanding of the questions presented, and too few people involved, to make anything more than a general claim of similarity.

Something which has not been explored here, but which would be of interest to some composers, are cultural associations with mapping. For example, are there culturally invariant mappings with some parameters such as pitch or intensity with height? One example from instrument design that has caused problems is the original Theremin mapping of intensity with proximity to a horizontal antenna such that to make a louder sound the performer would bring their hand *down* to the volume antenna. This was done for practical reasons so that without a performer there is no sound, but modern theremins allow this to be swapped because of the cognitive dissonance caused by the playing action.

9 Conclusions

One clear outcome of this is that the questions need improving if this is going to be attempted again. This is indicated clearly in the detailed responses where not all of the questions seemed sensible to some people, but it is also evident in the spread of answers to some of the questions, although some spread is to be expected of course. In addition, the semantic understanding of many of the terms surrounding mapping and composition (and design) seems to be ill defined and this leads to a varied understanding of the questions and therefore variable answers.

Not withstanding the problems with the questions and semantics, some valuable information can be extracted from the answers. The answers indicate that composers who use algorithmic techniques certainly use mapping from one domain to another, often in very sophisticated ways. Also indicated is that there is no ideal solution or single method, but that it is largely a process of exploration. Very simple or obvious mappings are sometimes problematic in composition probably because of the oversimplification of the resultant music, but they may be appropriate under some circumstances such as with very complex structures or data. More complex mapping strategies are more common and more musically useful, but very numerous and their application and musical usefulness appears to be an aesthetic judgement of the composer. Composers appear to be aware of complexity that does not scale linearly for listeners such that there is some limit to what makes sense to a listener's perceptions and beyond this it is so complex that there is no sense of the original concept or data. There seems to be almost as many approaches to mapping as there are composers or designers. It is the character of composition, being an expansive and creative activity, that if ever a theory provided a "standard", "optimised" or "ideal" mapping technique or repertoire, that composers would ignore it and go beyond that. Having said that, there would appear to be *something* to gain by creating a catalogue of historical and current mapping practices, even if only in a pedagogical sense (Polansky 2002) and this could be started by expanding on the questions presented above.

The nature of composition and composers means that there will never be "solution" to the mapping problem in algorithmic composition, that it will remain a part of the technique and exploration of composers.

Not being a designer, I am unqualified to comment on the area of design, but the survey showed significant areas of similarity in the responses. Significant enough, I feel, that there is some similarity of practice and theoretical approach such that further investigation and cross-discipline discourse is warranted to inform both disciplines.

10 Acknowledgments

I would like to thank and gratefully acknowledge the contributions of the composers and designers who made much of this work possible. The composers are; Roger Alsop, Rodney Berry, Chris Cree Brown, Phil Brownlee, Warren Burt, Densil Cabrera, Tim Kreger, Peter McIlwain, Gordon Monroe, Garth Paine, Greg Schiemer. The designers are Pia Ednie-Brown, Tom Kovak, Paul Minifie, Vivian Mitsogianni and Julian Raxworthy.

Also, much of the material in this paper is the result of stimulating conversations with Paul Berg, Larry Polansky and Marcelo Wanderly.

References

Wanderley, M. M. and Battier, M. eds. 2000. *Trends in Gestural Control of Music*. Paris: IRCAM

Cadoz, C. and Wanderley, M. M. 1999. *Gesture-Music*. In Trends in Gestural Control of Music. Paris: IRCAM CDROM.

Dodge, C. 1970. *Earth's Magnetic Field program notes*. In Nonesuch LP H 71250 EARTH'S MAGNETIC FIELD: Realizations in Electronic Sound.

Dodge, C. and Jerse, T. 1997. *Computer Music second edition*. New York: Schirmer Books

Hunt, A., Wanderley, M. M. and Kirk, R. 2000. *Towards a Model for Instrumental Mapping in Expert Musical Interaction*. In Proceedings of the 2000 International Computer Music Conference. San Francisco: International Computer Music Association, pp. 209 - 212.

Hunt, A. and Wanderley, M. M. (eds.) 2000. *Mapping of Control Variables to Musical Variables*. Interactive Systems and Instrument Design in Music Working Group. Website: http://www.notam02.no/icma/interactivesystems/mapping.html

Hunt, A. and Kirk, R. 2000. *Trends in Gestural Control of Music*. Paris: IRCAM

Kramer, G. 1996. *Mapping a single data stream to multiple auditory variables: A subjective approach to creating a compelling design*. In Proceedings of International Conference on Auditory Display - ICAD'96.

Matossian, N. 1986. *Xenakis*. London: Kahn & Averill

Polansky, L. 2002. *Manifestation and Sonification*. Unpublished paper.

Roads, C. 1996. *The Computer Music Tutorial*. Cambridge, Massachusettes: The MIT Press

Ryan, J. 1992. *Effort and Expression*. In Proceedings of the 1992 International Computer Music Conference. San Francisco: International Computer Music Association, pp. 414 - 416.

Thieberger, E. M. 1995. *An Interview with Charles Dodge*. Computer Music Journal 19(1): 11-25

Varga, B. A. 1996. *Conversations with Iannis Xenakis*. London: Faber and Faber

Xenakis, I. 1991. *Formalized Music Pendragon Revised Edition*. New York: Pendragon Press

Appendix 1: Detailed Responses

This is a one page version of the detailed responses to the first question. The rest of the responses are on the ACMC 2002 website where there is a downloadable version of the whole paper. See http://www.iii.rmit..edu.au/sonology/ACMC2002

The various approaches to mapping in algorithmic composition, and design, can be easily seen in the responses. Additionally, the comments by the designers show the overlap and differences with composers in thinking about mapping. The composers questioned were; Roger Alsop (**RA**), Rodney Berry (**RB**), Chris Cree Brown (**CCB**), Phil Brownlee (**PB**), Warren Burt (**WB**), Densil Cabrera (**DC**), Tim Kreger (**TK**), Peter McIlwain (**PM**), Gordon Monroe (**GM**), Garth Paine (**GP**), Greg Schiemer (**GS**), and Paul Doornbusch (**PD**). The designers questioned were Pia Ednie-Brown (**PEB**), Tom Kovak (**TXK**), Paul Minifie (**PXM**), Vivian Mitsogianni (**VM**) and Julian Raxworthy (**JR**).

1. Is mapping something that you are conscious of when you are composing (designing)?

PM: In the sense that mapping is some process that transforms material from state to another then all standard compositional procedures (such as transposition, inversion etc) are a form of mapping. I am not always conscious of these procedures as they have become more automatic over time. However, the term mapping is often used in a specific way to refer to transformations of musical materials in correlation with extra-music materials. When I have done this I am usually very aware of the process chiefly because I feel that must make judgments about the effectiveness in musical terms (than is, my own subjective musical definitions) of this, or that, mapping procedure.

WB: If the piece I'm doing involves the application of some sort of numerical data, whether a simple use of a random number generator, or something more complex, I'm concerned with mapping. Numerical sources only assume meaning when they're applied to materials that have a kind of "graspability" to them. That is, to hear a number pattern for its own sake, it might be better at first to apply it to a set of 5 pitches rather than all 12 (if we're talking 12tone equal temperament for the moment), because with fewer pitches we might be able to "forget" the traditional musical implications of the fewer pitches (ie, they're so bland we can forget them for the moment) and concentrate on the qualities of our numerical structure. When it comes time to put those structures into a larger "piece", then a much more conscious selection of materials can take place.

GM: Most, but not all, of my compositions involve mapping in some way.

DC: It is not a term that I normally apply. However, I'm conscious of the ideas of 'transcription' or 'sonification', which seem to be closely related. In some cases, my works have realised the acoustic implications of a spatial journey, and I suppose the term 'mapping' is especially apt there.

GP: When composing a piece of music, or a soundscape, I view the mapping of the sound material to the intended emotional context of the work to be central to the craft of composition. I am usually contemplating at least two layers of material at any one time - an environment/broader context, and individual lines of activity/dialogue/narrative.

PD: Yes, I'm conscious of mapping when I'm working, but it's not the first thing I think of. I'm very conscious of it when I'm working on transcribing a structure or gesture into music or sound. However, I am not sure if I think that it is important, in my work, that the compositional structures or gestures are somehow transmitted to the audience. In a way it is, but mostly I'm trying to give the audience a powerful musical experience, so as far as transmitting a gesture or structure is important for that, then yes it is important. I try to map from compositional gestures and structures to musical parameters in such a way that the gesture or structure is evident when *I* listen to the result, but this may not translate to these being perceived by the audience. Having said that, most of the compositional gestures and structures I use are abstract concepts and I usually have movement in three or four aspects of them simultaneously, otherwise I find the result boring musically.

PEB: Mapping is always an integral part of architectural design. All aspects of design involve representational devices The act of moving between various modes of representation involves many types or modes of mapping. As such, mapping is integral and inevitable; some aspects of which are more conscious than others.

The question gathers a more elaborate and more precise answer for me when I think of the sculptural work I've done. In this case the outcome is the object itself, there is not an imagined future event of actualisation (building), making what I see as the primary nature of mapping very directly accessible. I made a series of casts that emerged through experiments with latex and plaster. Each one of the series is the product of a different deformation of the same mould. They were made through pouring liquid plaster into the latex mould of a shower tap. The act of making these involved a kind of puppeteering act in which the stretchy, elastic moulds were suspended with string, positioned with wire and held down with tape under chairs, off coat hangers - whatever was at hand. The latex stretched and deformed with the weight of the fluid plaster, particularly as the makeshift supports fell away or shifted. Each casting was enacted with a different placement or 'context' of the mould. Although the process could be guided toward expected results, the final cast was always a surprise, being very contingent to the moment. The resulting series of different casts arose from the coalescing involutions of a myriad of variable relations (ie the outcome is a mapping of the complexity that is involved in the situation - it is a qualitative mapping of an event) all struck into an accord: form extracted from dynamic interactions, falling into one another in a collaborative agreement with a responsively overarching skin. In this example, it is a *situation* that is explicitly mapped more than anything else. I would argue that this is *always* the case in a creative act, no matter what the process, even if the dynamic variability of the moment or situation is embedded (and rendered opaque) within a rigid, linear, direct mapping process. Often, where extremely non-linear mapping positions are taken, there is an assumption that the material being mapped is the subjectivity of the author. What I like about the cast example is that, while being manifestly non-linear, like many programmed acts of mapping, there is an implied shift in the role of the author/creator in that the outcomes are, to a relatively explicit extent, a product of the process (where designing the process itself becomes the primary role of the composer).

PXM: In some projects mapping is of crucial interest - it is a key way of capturing relationships and structures into a design work. It helps because it is an antitectonic strategy, disrupting a normal tectonic set of relations.

VM: My own work deals with a 'machine based' design methodology, that is the process is designed rather than the object as a starting point. The 'process' or 'machine' will often include mapping, diagramming in part. This is a generalisation, but the work generally involves the use of techniques that contain abstract information (such as diagrams, scientific systems of behaviour) and are used as part of the larger design process. This is more akin to a 'machine' based design process in the sense of Deleuze's use of the term 'machine' in that the 'mapping' is a part of the machine that (along with other ideas and techniques) are 'applied' to something, they are not the thing itself. That is, mapping is not the main tool used.

Developing Analysis Criteria Based on Denis Smalley's Timbre Theories

David Hirst

Teaching, Learning and Research Support Dept, University of Melbourne
email: d.hirst@unimelb.edu.au

Abstract

This paper represents part one of a two-part process. The aim in part one is to outline a number of Denis Smalley's theories and to create a framework that could provide the criteria to analyse his musical works. The second part of the process would be to actually analyse the one or more works according to the identified criteria. A substantial part of the current paper is an attempt to summarise and make comment on a Smalley article. Having identified the key concepts, the paper establishes relationships between the concepts and proposes a methodology for analysis.

Keywords

Denis Smalley, timbre, electroacoustic music, analysis.

1 Introduction

Denis Smalley has written and composed extensively within the field of electroacoustic music. In his writings he has developed new concepts and theories to describe acousmatic music. Probably the best known concept is his spectromorphology - the temporal unfolding and shaping of sound spectra (Smalley 1986).

This paper represents part one of a two-part process. The aim in part one is to outline a number of Smalley's theories and to create a framework that could provide the criteria to analyse his musical works. The second part of the process would be to actually analyse the one or more works according to the identified criteria.

Rather than survey all of Smalley's written works, this paper concentrates on one of his seminal articles called *Defining Timbre – Refining Timbre* (Smalley 1994). A substantial part of the current paper is an attempt to summarise and make comment on the Smalley article. It is also an attempt to demystify many of Smalley's concepts since they are shrouded in, sometimes difficult, terminology which is much of his own invention. Although the jargon is difficult, it does offer the possibility of facilitating a certain dialogue concerning some very hard-to-grasp electroacoustic music concepts. Having identified Smalley's salient theoretic concepts, the paper aims to codify the relationships between the concepts in a way that may prove useful in analysing Smalley's music.

The processes of identifying entities and establishing functional relationships are just as relevant to musical analysis as they are to musical theoretic concepts, as we shall see in the paper below.

We begin with a precautionary note. Organising timbre can be a primary driving force in electroacoustic music, but timbre has had an infamous reputation for being very ephemeral. Smalley begins his article with the following caution:

> (Timbre) is one of those subjects where the more you read and the more you have hands on compositional experience the more you know, but in the process you become less able to grasp its essence. (Smalley 1994:35)

2 Defining Timbre

The difficulties in defining timbre are even highlighted by the contradictions in naming timbre. While the French term *timbre* identifies the object that actually creates the sound, the German term *klangfarben*, or sound colour, is more abstract and detaches the sound from any source.

Like Grey(1975), Erickson(1975) highlights the multi-dimensional nature of timbre, and its subjective and objective variations. He talks of the subjective constancy of timbre and its use as a carrier of other musical information (e.g. melody) and of timbres as objects. These come together in *Klangfarbenmelodie* as contrast and continuity.

Denis Smalley presents four definitions from four different perspectives:

The American National Standards Institute:

> ...that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. (Smalley 1994:36)

The instrumental composer:

> timbre is an extension of harmony, or vice versa. The composer uses spectral analysis as a basis for conceptualising the relationship between pitch and sound qualities, and attempts to negotiate fluent border crossings between the two. (Smalley 1994:36)

The researcher:

> Through research and through electroacoustic compositional experience we have become very aware of the multiple variables which determine timbral identity. And we have also become concerned to differentiate what is acoustically present in sounds from what is psychoacoustically pertinent.

Everyone:

> The everyday language of qualitative description is accessible to everyone. It is closely allied to the "matter" of sound. Terms like bright/dull, compact/spread, hollow, dense... (Smalley 1994:36)

3 Smalley's Theories of Timbral Organisation

3.1 Timbre and Source

Smalley highlights the importance of the sound source when discussing timbre. He provides a preliminary definition, adapted from Michel Chion, that takes the sound source into account:

> Timbre is a general, sonic physiognomy through which we identify sounds as emanating from a source, whether the source be actual, inferred or imagined. (Smalley 1994:36)

Smalley defines the term *source bonding* to encapsulate:

> the natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins. (Smalley 1994:37)

Smalley then acknowledges the socio-cultural contribution the listener makes in the "adventure of bonding play" when listening to electroacoustic music, which he regards as a perceptual activity.

Smalley explains that source bonding is extrinsic in that it refers to **sounding** experiences outside the work itself. He then defines the *extrinsic matrix* which, in addition to external sounding experiences, includes links to a wide range of real and imagined **non-sounding** phenomena too. (Smalley 1994:37)

Smalley doesn't expand on the notion. Instead he turns to the notion that timbre is concerned with the temporal unfolding and shaping of sound spectra or *spectromorphology*. Motion, growth and energy are associated with spectormorphology and therefore have a sonic reality but they also can be interpreted metaphorically and symbolically.

We can see that in addition to the physical unfolding of the temporal shape of the sound (amplitude, spectrum) Smalley has added the listener, metaphor, symbology, and perceived source. This quite complex combination Smalley attempts to define in an abstract and succinct way. His definition of timbre has become:

> a general, sonic physiognomy whose spectromorphological ensemble permits the attribution of an identity. (Smalley 1994:38)

3.2 Source-cause Texture

In a section called "source-cause texture", Smalley invents a lot of his own jargon which can be difficult to interpret. It is worth persisting, however, as many of his ideas strike at the heart of the function of timbre in electroacoustic music.

He begins this section by looking at the identity of instruments, exposed through our experience of what he calls *source-cause levels*.

Imminent level - the ongoing, intrinsic musical context where we encounter the instrument (eg. the violin). Associated with the imminent level is *registration* - the articulation of note objects and their chaining in phrases over a continuum of registers.

Cumulative level - which includes our previous experiences of violin sources in the hands of other violinist-causes who articulate the same music and other genres and styles.

Extended level - extends the source-cause base to include the immediate family of stringed sources. (eg. viola, 'cello, double bass)

Dispersed level - spreads over the widest possible range of source-causes to include all bowed and plucked (string) instruments and the (string) instruments of other cultures.

The four levels together he calls *source-cause texture*.

We may accept this scheme for instrumental music, but how can we apply his notion of source-cause texture to electroacoustic music? Smalley answers by noting that in electroacoustic music we do not find such a definable hierarchical basis for establishing the source-cause aspect of timbral identity.

The discussion then turns to the concrete example of a "water" source-cause. "In a musical work, water, like any sounding source, can exist on both the imminent and cumulative levels." (Smalley 1994:39) That is, the water sounds can function within this musical work and other musical works, but it is much harder to identify families of water sounds or extended families of water sounds (as they function within musical works).

So although the extended and dispersed levels cannot be identified in electroacoustic music, Smalley goes outside the confines of the music.

> ...it can be extended beyond the cumulative level by referring outside the musical works to the extrinsic matrix. (Smalley 1994:39)

That is, to both the sounding and non-sounding inferred phenomena.

> ... the sounding area of the extrinsic matrix can provide a substitute for the extended and dispersed levels of source-cause texture, ... (Smalley 1994:39)

So in our example of water, there are many experiences of water **sounds** beyond the cumulative level, in nature and culture. But what about non-sounding inferences?

> When the non-sounding area (of the extrinsic matrix) is entered we are no-longer on secure common ground because source bonding is no longer operable: we cannot identify real sources and causes. (Smalley 1994:39)

Smalley does, however, go on to highlight the importance of such non-sounding inferences and contends that it is the spectromorphological attributes and ideas that evoke non-sounding substitutes for (instrumental) extended and dispersed levels.

While this seems very ephemeral, it may be the very essence of being for electroacoustic music.

> I believe that such ideas, intangible as they might be, are a means of articulating that necessary, shared, higher-level cultural basis for a music with non-existent source-causes. (Smalley 1994:39)

Smalley begins to expand on the idea that spectromorphological attributes can evoke non-sounding real and imagined phenomena by first considering instrumental gesture.

> ... behind the causality of instrumental gesture lies both a broader experience of the physical gesture and its proprioceptive tensions, and a deeper, psychological experience of gesture. In instrumental music human-bonded source-cause texture *represents* these primal levels of gesture found in the extrinsic matrix. In electroacoustic music where source-cause links are severed, access to any deeper, primal, tensile level is not mediated by source-cause texture. That is what makes such types of acousmatic music difficult for many to grasp. In a certain physical sense there is nothing to grasp - source-cause texture has evaporated. (Smalley 1994:39)

Smalley is desperately searching for a "Something to hold on to factor" (Landy 1994) when considering the most abstract aspects of electroacousticmusic.

Finally he turns to the spectromorphology of the musical work in the search for the identity of timbre.

3.3 Registration

Smalley states that timbral identity in electroacoustic music is heavily reliant on the immediate experience of the work itself and, moreover, the ready identification of the functional elements in the work.

> If timbral identity in electroacoustic music is so heavily reliant on the single imminent level then registration in the work is key. (Smalley 1994:40)

I interpret his concept of registration in music as being the recognition and connection of events or phenomena in the music. While this can be quite systematic in instrumental music, it is problematic in electroacoustic music where there is no consistent low-level note object.

Registration then becomes concerned with variable spectromorphological attributes. Smalley notes that this could become so general we could lose our way, so we need to discriminate the **incidental** from the **functional**, i.e. to perform a functional analysis of electroacoustic music.

We have to decide which spectromorphological attributes matter, and we have to discover this anew in the imminent level of each electroacoustic work. A pair of related variables underpins our attempts at determining identities:

> 1. the coherence and strength of spectromorphological identity;
> 2. the duration needed to establish existence and expose registration. (Smalley 1994:41)

3.4 Existence

There is no standard duration for establishing existence.

> There are certain types of spectromorphology whose existence can only be established after a certain evolution time because the completion or partial completion of a pattern is integral to identity. (Smalley 1994:41)

Continuous transformation can become what Smalley calls a *registration generator*. Change becomes fundamental to existence and we apprehend identity as a consequence of change.

3.5 Coherence

Smalley introduces coherence by stating:

> The notions of existence and registration imply a certain coherence, otherwise identity would not be feasible. (Smalley 1994:41)

To my mind, the process of searching for coherence is the same as the process of creating connections within a work. Smalley, however, considers coherence in a much more subtle way, on both the micro and macro levels. He uses it to establish identity. Along the way he creates the notions of *integration* and *disintegration*.

> Coherence in the case of instruments is usually associated with spectral fusion, which in turn is associated with harmonicity. ...Since fusion is so often closely aligned with harmonicity, and since I find it too rigid a word, I prefer the term "integration". That also allows us to talk of an integration-disintegration continuum. ...Integration means that within a sonic physiognomy the distribution of spectral components in spectral space, and their behaviour over time should not be such that a component or sub-group of components can be perceived as an independent entity. Certainly a fairly high degree of integration is necessary if something called timbral identity is to be established. (Smalley 1994:41-42)

Smalley also highlights the problem of discriminating between integration and disintegration.

> ... but where are the borders between tight fusion, a loose interdependence, and a dissolution towards independence? When is a timbre a timbre and when is it a collection of timbres? When does the physiognomy crack?
> The course of even a relatively short spectromorphology can travel between integration and disintegration. (Smalley 1994:42)

His solution to creating connections between identities in electroacoustic music is to introduce the concept of discourse.

> Playing with integration and disintegration is at the very heart of electroacoustic musical discourse, a discourse which becomes spectromorphology itself once the timbre complex is spilt open, a discourse where the notion of timbre can at one moment perhaps be grasped, but at the next it evaporates. (Smalley 1994:42)

3.6 Discourse Stability and Variability

In electroacoustic music, a timbre may be a discrete object, easily separated from its context or it may be a continuity intertwined with other continuities and not so easily separated. To help identify this phenomenon, Smalley introduces the concept of *timbral level*; which concerns the relationship between two continua.

> 1. *duration*: short-term entity .... longer term evolution
> 2. *separability*: discrete object .... continuing context
>
> ... Timbral level in traditional note-based music is quite simple. The note is the lowest level and is articulated by an instrumental source. Form develops from note articulations. In electroacoustic music continuing contexts resist and deny low-level segmentation. Thus once timbral level ceases to be clearcut we cannot separate timbre and discourse: timbral attributes become woven into the spectromorphological fabric. (Smalley 1994:43)

Now that Smalley has established some bases for determining identity he discusses discourse in more detail. He asserts that discourse is about maintaining and developing some of the established identities.

3.7 Transformational discourse

The first approach to discourse is the *transformational discourse*.

> where an identity is transformed while retaining significant vestiges of its roots (i.e. an identity-base in the imminent level, or a strong extrinsic identity). To achieve this, certain attributes must remain stable while others vary. (Smalley 1994:43)

He gives some examples of prevalent techniques used for digital transformations:

> 1. stretching or contracting in time which does not threaten too much the aural, evolutionary integration of spectral components. Compress too much and you destroy identity; stretch too much and you destroy, for example, attack identity.
> 2. changes in spectral density by thickening, or by spreading in spectral space, which are perceived as addition — a growth process.
> 3. changes in spectral weighting (e.g. brighter, or emphasizing internal intervallic pitch) without altering the essential content (i.e. frequency spacing) of the spectral envelope.
> 4. variation in or reshaping of a morphology (for example, the attack impact) in such a way that it does not effect the identity of the continuing body of spectromorphology. This implies that
>   a. the sound is a strong spectromorphological type well known through extrinsic experience (e.g. a bell, the voice); or
>   b. that the base-spectromorphology has been previously announced and that if this was some time previously then it was striking enough in context for relevant features to be memorable; if the base- spectromorphology is less memorable it would need to be relatively contiguous to the present event.
>
> (Smalley 1994:43-44)

Smalley makes some important points in passing regarding memorability and contiguity.

> It is undoubtedly true that pitch relations and source-causes are the most easily memorable sonic phenomena, and that the problem with multiple spectrological attributes is that we do not know which ones are to be relevant in the discourse. This is why electroacoustic music which avoids pitch phenomena and strong source-cause references will be most frequently concerned with contiguous relationships. Non-contiguous discourse can only be highly developed within the precise and detailed memorability of culturally imbedded pitch and rhythm systems. (Smalley 1994:44)

The above quotes contain important implications for the analysis of Smalley works, e.g. where pitch, rhythm and strong source-cause references are not used then we should find contiguous relationships and continuous transformations.

3.8 Typological discourse

Smalley's second type of discourse is typological discourse.

> Identities are recognised as sharing timbral qualities but are not regarded as being descendants of the same imminent identity — they do not possess a common identity base. Typological discourse is associative. (Smalley 1994:44)

Smalley invents the term *generic timbres* to describe larger groupings of timbres. He gives two examples of generic timbres: *Noise* and *Inharmonicity*.

To transformational and typological discourses Smalley adds a third type: *source-cause discourse* – concerned with the bonding play of specific or inferred sounding identities.

Compare Smalley's notion of discourse with Emmerson's mimetic and aural discourse plotted against an abstract and abstracted syntax continuum on his "Language Grid". (Emmerson 1986:24)

3.9 Discourse Summary

Smalley points out that these various discourses are not mutually exclusive, and "tipping the scales either way can be a question of listening choice and listening experience – listening experience is variable and subjective." (Smalley 1994:45) Thus he ends up with six interactive types of electroacoustic discourse. The first three are those mentioned already:

> *Source-cause discourse*
> *Transformational discourse*
> *Typological discourse*

The second three discourses are concerned with relations among identities. They are:

> *Behavioural discourse* — the changing states of identities cohabitation/conflict and dominance/subordination;
> *Motion discourse* — the relations of types of motion and growth and their directional tendencies;
> *Tensile discourse* — how the previous five discourses together create formal tensions.
> (Smalley 1994:46)

3.10 The Timbre of Pitch – the Pitch of Timbre

Smalley also writes of the pitch of timbre and the timbre of pitch. This section is a very important and very revealing discussion in regard to Smalley's own works. One always has the vague sense that pitch is important in Smalley's works but its manipulation is not overt and its use is hard to grasp.

He introduces the notion that pitch and timbre can cohabit the same space.

> Once tonality and intervallic pitch are no longer regarded as the predominant carriers of musical messages, pitch and timbre can cohabit in a spectromorphological music where the ear has opportunities for shifting in and out of pitch values. (Smalley 1994:40)

Surprisingly Smalley postulates that: "Pitch is even present when not perceived." (Smalley 1994:40). This statement may provide a telling insight when his own works are examined.

How can this non-perceived pitch operate? Smalley elaborates:

> Perhaps it is resting, hidden deep in a spectromorphology, awaiting possible attention, a moment when, for example, the context might change so that perceptual focus becomes directed towards what was a sleeping attribute. (Smalley 1994:40)

Pitch becomes contextual in electroacoustic music, and:

> Timbre rather than being that part of the sound which is not pitch, encompasses the inherent qualities of the whole sound. Perceived pitch is one of these inherent qualities, but the term timbre cannot be expected to shoulder the burden for all others. (Smalley 1994:40)

For the first time Smalley poses the question:

> Under such conditions does not the concept of timbre become so general as to be meaningless? (Smalley 1994:40)

Smalley attempts an answer in his concluding remarks.

3.11 Smalley's Concluding Remarks

In his concluding remarks Smalley rightly confesses that he has not been concerned with the acoustic nature of timbre, nor with its psychoacoustic properties, but with its apprehension, identity and functions in musical contexts from the listener's point of view. Timbre has become the *timbre complex* which has been spilt open to allow its essence to seep through musical discourse.

Smalley seems to provide an answer for Monro's question regarding acousmatic composers' attitude to note-based (pitch-based) composition. (Monro 2000)

> Composing with timbre, composing within timbre, means confronting and enjoying its dissolution.
> This can only be really pursued in an *acousmatic* electroacoustic music. In contrast, adventurous

> contemporary instrumental music, and works which mix instruments with acousmatic element, are rooted in the umbilical security of instrumental source-cause coherence and directly apprehended sound-making gesture. This equates not with a burning desire to explore timbre, but with a hesitant reserve about cutting loose in order to pursue a freer exploration. (Smalley 1994:47)

Re-assuring us that traditional modes of working and listening are still valid, Smalley concludes:

> There is no reason why the traditional notion of timbre should fade away. A notion of musical timbre will always exist alongside its dissolved attributes. In keeping with this ambivalence I can summarise this discourse in six words:
>
> Timbre is dead. Long live timbre. (Smalley 1994:47)

4 Developing an Analytical Framework for the Music of Denis Smalley

4.1 Identifying the theoretic concepts

First of all let us examine the important Smalley Concepts, and then we will go on to attempt to establish relationships between the concepts.

Source Bonding

Sounding experiences outside the work.

Extrinsic Matrix

Both sounding experiences and non-sounding phenomena (real and imagined).

Spectromorphology

The temporal unfolding and shaping of sound spectra.

Definition of Timbre

A general, sonic physiognomy whose spectromorphological ensemble permits the attribution of an identity.

Source-Cause Levels (Instrumental)

Imminent level - the ongoing, intrinsic musical context where we encounter the instrument (eg. the violin). Associated with the imminent level is *registration* - the articulation of note objects and their chaining in phrases over a continuum of registers.

Cumulative level - our previous experiences of violin source-causes for the same music and other genres and styles.

Extended level - extends the source-cause base to include the immediate family of stringed sources. (eg. viola, 'cello, double bass)

Dispersed level - spreads over the widest possible range of source-causes to include all bowed and plucked string instruments for all cultures.

Source-Cause Texture

All levels combined to contain the total long-term experience of imminent registrations, in many musics, across space and time.

Source-Cause Levels (Electroacoustic Music)

Imminent and Cumulative levels – as for instrumental source-causes. (For electroacoustic music *registration* - see below.)

Extended and Dispersed levels – much more difficult to identify in electroacoustic music. The sounding area of the extrinsic matrix can provide a substitute for the extended and dispersed levels of source-cause texture (source bonding).

The non-sounding area of the extrinsic matrix is more problematic as source bonding does not apply. However non-sounding experiences and phenomena are evoked, so how is this accomplished? Smalley contends it is through spectromorpholocial attributes operating in a similar manner to instrumental gesture.

Registration (Electroacoustic Music)

Is concerned with variable spectromorpholocial attributes. I interpret his concept of registration in music as being the recognition and connection of events or phenomena in the music. Smalley says we must discriminate the incidental from the functional.

Establishing Identities (Recognition)

Existence

Duration of the phenomena may vary. Continuous transformation can become what Smalley calls a *registration generator*. Change becomes fundamental to existence and we apprehend identity as a consequence of change.

The notion of existence must be examined in combination with coherence.

Coherence

Coherence is associated with the integration-disintegration continuum.

Integration means that within a sonic physiognomy the distribution of spectral components and their behaviour over time should not be such that a sub-group of components can be perceived as an independent entity. A high degree of integration is required for timbral identity.

The distinction between coherence, with its integration-disintegration continuum, and timbral level is not clear by Smalley.

Timbral Level

Concerns the relationship between *duration* (short-term entity to longer term evolution) and *separability* (discrete object to continuing context). It would appear to concern the degree to which one can unravel an entity from its context. *Timbral Level* overlaps with *discourse*.

Discourse

Transformational Discourse

An identity is transformed while retaining significant vestiges of its roots.

It is important to note that electroacoustic music which avoids pitch phenomena and strong source-cause references will be most frequently concerned with contiguous relationships. Non-contiguous discourse can only be highly developed within the precise and detailed memorability of culturally imbedded pitch and rhythm systems.

Typological Discourse

Identities are recognised as sharing timbral qualities but are not regarded as being descendants of the same imminent identity – they do not possess a common identity base. Typological discourse is associative. Smalley invents the term *generic timbres* to describe larger groupings of timbres, e.g. noise, inharmonicity.

Source-Cause Discourse

Is concerned with the bonding play of specific or inferred sounding identities.

The relationship between Smalley's source-cause discourse and the non-sounding area of the extrinsic matrix is not clear. It would appear there is no relationship. In fact the relationship is between the non-sounding extrinsic matrix and some parts of the registration process. A 'something to hold on to factor' is required here!

Relationships between identities is defined by another three discourses.

Behavioural Discourse

The changing states of identities' cohabitation/conflict and dominance/subordination;

Motion Discourse

The relations of types of motion and growth and their directional tendencies;

Tensile Discourse

How the previous five discourses together create formal tensions.

4.2 Establishing Relationships Between the Concepts

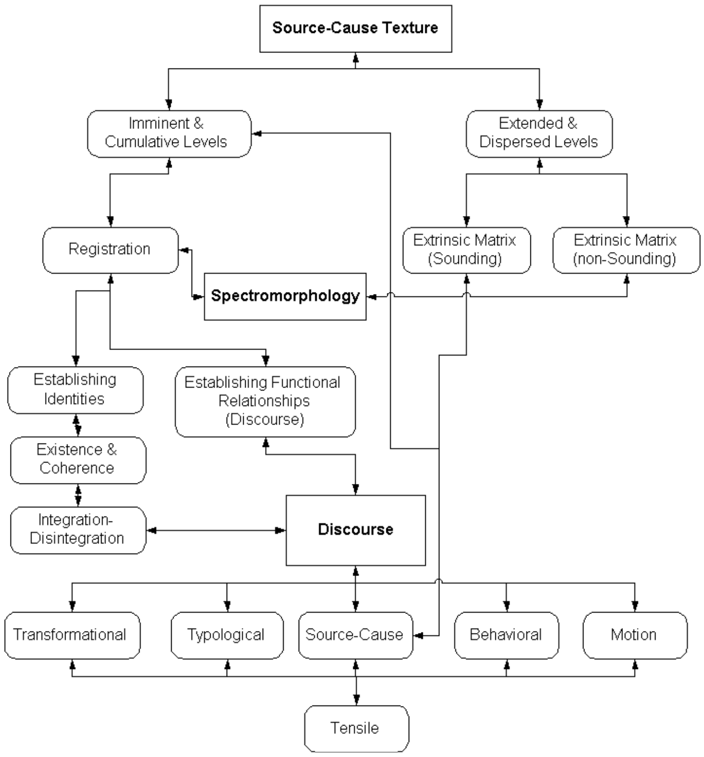


Figure 1 Concept map showing Smalley's concepts and the relationships between them. The connectors show arrows at each end denoting there is some sort of relationship but without saying what type of relationship. All the relationships are quite tenuous, and some overlap. The most tenuous relationship is between the non-sounding area of the extrinsic matrix and the registration process (via spectromorphology).

5 Developing an Analysis Methodology

We have reached a point where we can develop a protocol for the analysis of Smalley's music based on his own theoretical constructs.

An inside-out approach would look at the detailed unfolding of the moment-to-moment details of the music itself, and could begin by examining the imminent/cumulative levels via the registration of spectromorphologies. The steps may involve:

- Establishing the identities by looking at coherence, integration and disintegration.
- Establishing functional relationships between the identities.
- Identifying transformational, typological, and source-cause discourses.
- Identifying behavioral and motion discourses.
- Discussing the tensile ebb and flow.

An examination of the extended and dispersed levels would reveal more of the extrinsic nature of the work. The process would involve connecting the internal sound world of the work to external sonic manifestations and to the more ephemeral, non-sounding, conceptual implications.

The form of reporting this analytical work would then constitute a part of a future study.

References

Emmerson, S. 1986. "Relation of Language to Materials." in *The Language of Electroacoustic Music*, (S. Emmerson Ed.) Basingstoke, Macmillan Press, pp. 17-39.

Erickson, R. 1975. *Sound Structure in Music*. Berkeley, University of California Press.

Grey, J. 1975. *An Exploration of Musical Timbre*. Doctoral Dissertation, Stanford University.

Landy, L. 1994. "The 'Something to Hold on to Factor' in Timbral Composition." in *Contemporary Music Review*, Vol. 10 Part 2, Switzerland, Harwood Academic Publishers, pp. 49-60.

Monro, G. 2000. "An Acousmatic Experience." *Chroma* 26-27, Newsletter of the Australasian Computer Music Association, Melbourne, Australia.

Smalley, D. 1986. "Spectromorphology and Structuring Processes." in *The Language of Electroacoustic Music*, (S. Emmerson Ed.) Basingstoke, Macmillan Press, pp. 61-93.

Smalley, D. 1994. "Defining Timbre - Refining Timbre." in *Contemporary Music Review*, Vol. 10 Part 2, Switzerland, Harwood Academic Publishers, pp. 35-48.

Multi-feature Musical Instrument Sound Classifier w/user determined generalisation performance

Ian Kaminskyj

Electrical & Computer Systems Engineering, Monash University
email: ian.kaminskyj@eng.monash.edu.au

Abstract

A previously developed musical instrument sound classification system has been extended in the following ways: (a) three additional features have been extracted (cepstral coefficients, spectral centroid and presence of vibrato), (b) automatic instrument elimination, and (c) hierarchic and hybrid architectures introduced. Testing the system using the leave-one-out strategy with recordings taken from the same instruments used during system development has yielded a classification accuracy of 89-92% and reliability of 89-92% depending on the classification architecture used. Instrument families were identified with an accuracy of 96%, while sustain versus impulsive instruments were identified with 100% accuracy. Extensive generalisation testing using a single stage classifier over three different collections of instrument recordings showed that the system classification accuracy varied from 33% up to 61%. Applying a number of enhancements to improve the generalisation performance of the system improved the classification results by anywhere from 13 to 50 percentage points. Across four different collections of recordings, the single stage classifier consistently achieved a classification accuracy of 87-90% when the leave-one-out test strategy was used. Using instead the hierarchic classifier, a further 3% improvement would be possible. This shows that when allowed to do so, the system homes in remarkably well to the defining characteristics of any given set of test recordings. The user of the system is given direct control over its performance, which is dictated by (a) which classifier architecture is used, and (b) how much information the user is willing to provide it prior to classification being attempted.

1 Introduction

In the last three years, there has been an explosion of interest in developing a classification system that reliably identifies monophonic sounds produced by musical instruments.

Fujinaga and MacMillan (2000) designed a computer-based classifier to recognise simple continuous tones from the MUMs CDs of musical instrument sounds. 39 different timbres from 23 orchestral instruments played at different pitches were used as samples for an exemplar-based learning system that incorporated a *k*NNC with genetic algorithm.

The weighted features used during classification included moments calculated from both the steady-state and dynamically changing spectral envelopes as well as spectral irregularity and tristimulus features describing the spectral envelopes. While the recognition accuracy varied greatly between the instruments, the average overall accuracy was 68%.

Eronen (2000) developed a pitch independent isolated tone musical instrument recognition system that was tested using the full pitch range of thirty orchestral instruments from the string, brass and woodwind families, played with different articulations. Thirty-two spectral and temporal features were employed during classification. Both hierarchic and direct forms of classification were evaluated using 1498 test tones obtained from the McGill University Masters Samples (MUMs) CDs as well as from amateur musicians.

Best results were obtained with direct classification, where instrument families were classified with 95% accuracy and individual instruments with 81% accuracy.

Eronen (2001) extended his earlier work to develop a classification system that allowed him to compare several features with regard to musical instrument recognition performance. In addition to the features described earlier, the following features were also evaluated: (i) mel-frequency cepstral coefficients, (ii) linear prediction cepstral coefficients calculated with both uniform and warped frequency scales, (iii) delta cepstral coefficients and (iv) linear prediction analysis reflection coefficients.

The features were evaluated using a database of 5286 acoustic and synthetic solo tones from 29 different orchestral instruments, out of which sixteen were used for testing purposes. Five different sources of recordings were used: (i) MUMs, (ii) acoustic guitar recordings made at Tampere University of Technology, (iii) University of Iowa, (iv) IRCAM Studio online (SOL), and (v) a Roland XP-30 synthesizer.

The best performing combination of features included mel-frequency cepstral coefficients calculated over both the onset and the steady state and a subset of the earlier spectral and temporal features. These features allowed 35% of individual instruments and 77% of instrument families to be identified correctly. The results however were found to vary across the different sources of recordings, ranging from 21% for the synthetic tones upto 87% for the University of Iowa recordings.

Agostini et al (2001) evaluated a number of classification techniques to determine which provided the lowest error rate when classifying monophonic sounds from 27 different musical instruments. The classification schemes tested were canonical discriminant analysis, quadratic discriminant analysis, support vector machines and nearest neighbours.

For each instrument sound analysed, eighteen temporal and spectral features were extracted. The test results showed that quadratic discriminant analysis provided the best results, with an accuracy of 93% for all 27 instruments. Instrument families were classified with an accuracy of 97% while pizzicato versus sustain instruments were classified correctly 97%.

In their most recent work, Kostek and Czyzewski (2001) compared the effectiveness of 23 wavelet based features derived from the attack portion of musical tones to fourteen temporal and spectral features used in previous musical instrument sound classification experiments. Experiments were carried out using a two-layer artificial neural network (ANN) having many similar characteristics to that used in their earlier work. Eighteen different instruments were tested, but taken four at a time.

In all the experiments carried out, the recognition accuracy was found to be always higher using the original fourteen features than when using the wavelet-based features. Significant variation in classification accuracy was also found to occur between instrument groups. For example, using the original fourteen features, classification accuracies ranged from 80% to 99% depending on which group of four instruments was used.

Wieczorkowska (1999a and 1999b) explored the use of both decision trees and rough sets for automatically classifying sounds from eleven orchestral (string, woodwind and brass) musical instruments played with different articulations. 62 features were used during classification. These were extracted from each sound at the beginning, middle and end of the attack as well as during the maximum and minimum amplitude portions of the steady state.

Wieczorkowska found that the rough sets outperformed the decision tree. Using a validation strategy employing 10% testing data with the rough sets achieved 90% classification accuracy.

In later work, Wieczorkowska (2000) used the same two classifiers to evaluate the use of the wavelet transform as a feature for automatically identifying musical instrument sounds. 162 features were used during classification, representing various wavelet coefficient values for different filter types taken from both the attack and steady of each sound. Out of a total of 1358 recordings, 70% were used for classifier training and 30% for testing.

In contrast to her earlier work, the experimental results showed that in general, the decision tree classifier outperformed the rough set classifier. In terms of individual instrument classification, taking into account articulation style, the rough set classifier yielded a classification accuracy of 43% compared to the decision tree classifier's 58% classification accuracy.

These results show that great progress has been made in developing a reliable musical instrument sound classifier. Some deficiencies, however, still remain. The results obtained have been generally limited to orchestral instruments. None of the studies for example has incorporated percussive instruments, nor both non-vibrato and vibrato recordings of the same instrument (cello, flute and violin). In addition, a very limited amount of work has been carried out on the generalisation capabilities of the systems developed using more than one set of instrument recordings. Finally, no work has been done to the knowledge of the author in adapting the classification systems developed to the particular set of instrument recordings being used during testing, in an attempt to improve the generalisation performance of the system. The classification system to be described addresses these shortcomings.

2 Data Collection

Sounds from nineteen musical instruments of definite pitch were chosen for system development and testing. 604 recordings at a single dynamic level were taken directly from the MUMS CDs, of which 517 were non-vibrato and the remainder vibrato recordings. The instruments included: guitar (plucked string), violin, cello and double bass (bowed string), piano (struck string), flute (air reed wind), accordion, clarinet, saxophone (single mechanical reed wind), oboe and bassoon (double mechanical reed wind), organ (air/mechanical reed wind), trumpet, trombone, French horn, and tuba (lip reed wind) and xylophone, glockenspiel and marimba (percussive definite pitch). Both vibrato and non-vibrato recordings were used for cello, flute and violin.

The note range used was C3-C6 of the equally tempered musical scale. A three-octave note range was chosen to limit the system complexity, to provide practical limits on the number of recordings required and to simplify instrument selection. Not all instruments covered the complete note range, but each had at least some notes falling within it.

For generalisation testing, three additional sets of recordings were used. The first set of recordings were obtained from the Creative Essentials (CE) set of sound collection CDs, codeveloped by the companies Zero G and Time+Space (http://www.timespace.com/index1.asp). Recordings were provided at a single dynamic level and available for only certain notes for instruments: bassoon, cello, clarinet, flute, guitar, horn, oboe, piano, trombone, trumpet and violin. In total, 46 non-vibrato and 23 vibrato recordings were available.

The second set of recordings was obtained from Ircam, the French institute research centre into Acoustics and Music, via the Studio Online (SOL) project (http://www.ircam.fr/produits/technologies/sol/index-e.html). For consistency with the MUMs and CE collections of instrument recordings, only one dynamic level, i.e. *forte,* was used from this set. Recordings were used for the following instruments: bassoon, cello, clarinet, double bass, flute, guitar, horn, oboe, saxophone, trombone, trumpet, tuba and violin. In total, 333 non-vibrato and 61 vibrato recordings were available across the complete chromatic scale for each instrument.

The final set of recordings was kindly supplied by Prof. Kostek from the Sound and Vision Engineering Department, Faculty of Electronics, Telecommunication and Informatics, Technical University of Gdansk (TUG), Poland. The whole chromatic scale characteristics of the following instruments were used: bassoon, cello, clarinet, French horn, oboe, saxophone, trombone, trumpet and tuba. In total, 312 non-vibrato and 54 vibrato *forte* recordings were available.

3 Feature Extraction

The original classification system developed utilised only three features extracted from each musical recording (Kaminskyj 2000). These were (a) Root Mean Square (RMS) amplitude envelope, (b) Constant Q transform (CQT) frequency spectrum, and (c) Multidimensional Analysis Scaling (MSA) trajectories (see (Kaminskyj 2000) for more details). In order to improve the performance of the system, three additional features were added. These included (d) cepstral coefficients, (e) spectral centroid and (f) presence of vibrato. They were chosen either because they had been used successfully in previous research in solving similar problems (Brown 1999) or had proven important in allowing humans to discern between musical instrument sounds (Handel 1995) or simply looked intuitively promising.

3.1 Cepstral Coefficients

The cepstrum is the Fourier transform of the log magnitude spectrum of the musical sound waveform. There exist many different ways of calculating the cepstral coefficients, principally determined by the spectral measure used. The CQT spectrum with quartertone spaced spectral bins mentioned above was used to calculate the cepstral coefficients (CC_n):

$$CC_n = \sum_{k=0}^{176} \log(X[k])\cos[n(k+0.5)\frac{\pi}{177}]$$

where: $X[k]$ is the CQT amplitude value for

spectral bin k, and
$n = 1..176$ since CC_{177} is always zero.

Although this equation produces 176 cepstral coefficients (ignoring the final zero coefficient), it was determined empirically that for instrument classification purposes, using only the first eleven coefficients produced the best classification results. Figure 1 shows the resultant waveform formed by the first eleven coefficients for piano playing C4.

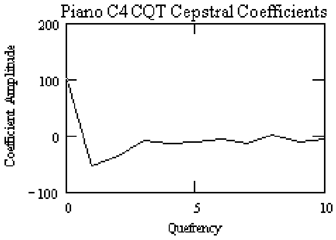


Figure 1 CQT spectrum cepstral coefficients waveforms for piano note C4

3.2 Spectral Centroid

The spectral centroid (S_k), also called brightness, is essentially a measure of the 1st moment of the spectral energy distribution. It is calculated over the duration of musical instrument recording using:

$$S_k = \frac{\sum_{j=0}^{176} A_{j,k} f_j}{\sum_{j=0}^{176} A_{j,k}}$$

where: S_k is the spectral centroid at time interval k,
$A_{j,k}$ is the amplitude of CQT spectral bin j at time interval k, and
f_j is the frequency of CQT spectral bin j

The time varying spectral centroid for four piano notes C3, C4, C5 and C6 is shown in Figure 2.

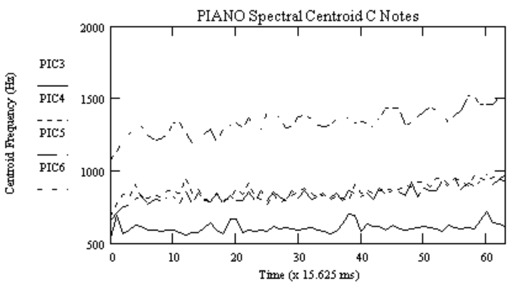


Figure 2 Time-varying spectral centroid musical for various C notes of the Piano.

3.3 Vibrato Detection

A vibrato detection system was developed to automatically detect the presence/absence of vibrato in a musical instrument recording (Kaminskyj 1998). A number of enhancements were however made in order to improve its reliability. As a result, the following conditions were required to be satisfied to qualify a recording as have been played with vibrato:

- the recording was produced by a sustain instrument,
- the peak-to-peak amplitude of the frequency variation of the largest amplitude harmonic exceeded 20 cents, and
- the amplitude fluctuation of the tremolo waveform or the frequency variation of the largest amplitude harmonic following two point moving average filtering, was in the range 5 – 8 Hz.

These conditions were based on the assumption that only cello, flute and violin were capable of being played with vibrato. If necessary, this number and range of instruments can be easily modified if required by the user.

4 Classifier Operation

The system was developed using the MUMs recordings and its structure is shown in Figure 3. The vibrato detector firstly determines the presence of any vibrato within the input test sample. If vibrato is deemed to be present, the vibrato instrument classifier is invoked, which classifies the input as having been produced by one of three possible vibrato instruments: cello, flute and violin. Alternatively, if no vibrato exists, the non-vibrato instrument classifier is executed. It classifies the input as having been produced by one of nineteen possible non-vibrato instruments.

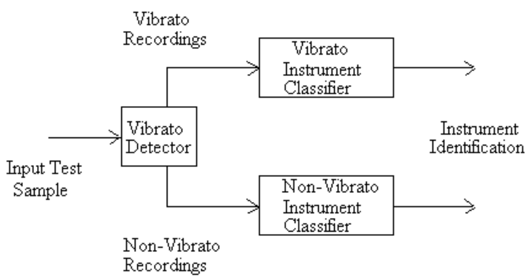


Figure 3 Classification system structure

Given the increased complexity of the non-vibrato instrument classifier, in terms of the number of instruments it needs to classify, three different classifier architectures were evaluated: (a) single stage, (b) hierarchic and (c) hybrid. It is intended that the user of the system will be able to decide which s/he wishes to use, balancing the performance achieved with the computational effort expended.

The simplest, called the single stage classifier, in shown in Figure 4 and uses five single feature classifiers to independently determine the most likely instrument to have produced the input test sample. Each classifier employs the *k* nearest neighbour classification (*k*NNC) algorithm (Duda & Hart 1973) using one of the remaining five features listed above. The results of all five classifiers are then combined using a confusion matrix weighting scheme which makes maximum use of the known strengths and weaknesses of each classifier. The resultant combined classification results provide the best estimate of the instrument that produced the input test sample.

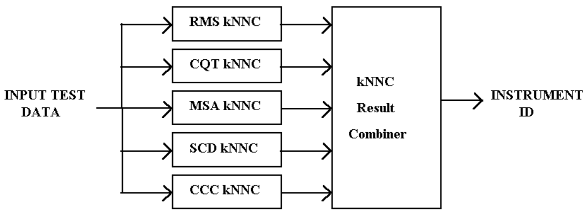


Figure 4 Single Stage Classifier

In an attempt to improve on the classification results obtained using the single stage classifier, both the hierarchic and hybrid architectures were evaluated. These used the same classifier structure as the single stage as a building block to form a taxonomic hierarchic classification tree. Each level of this tree essentially comprised a single stage classifier, but classifying a different collection of instruments. At the top of the tree of the hierarchic architecture shown in Figure 5, instruments were simply classified as being sustain or impulsive. As classification progressed down the tree, the test sample was classified into smaller and smaller groups of instruments, until finally a leaf was reached. This leaf indicated directly the most likely instrument to have produced the test input.

The hybrid classifier shown in Figure 6 comprised some of the features of both the single stage and hierarchic classifier, yielding essentially a compromise between speed of classification and performance.

Given that only three instruments were capable of producing vibrato recordings, the vibrato instrument classifier simply comprised a single stage classifier, with the same structure as described above for the non-vibrato instrument classifier.

5 Results

For the non-vibrato instrument classifier, the classification accuracy of the three classifiers evaluated was single stage 89%, hybrid 89% and hierarchic 92%. Given that the computational complexity increased dramatically when moving from the single stage classifier to the hybrid and then to the hierarchic classifier, the single stage classifier provided the best compromise in terms of classification accuracy versus computational effort. In terms of instrument family classification accuracy, both the single stage and hierarchic classifiers achieved 96%. Finally, the hierarchic classifier achieved a 100% classification accuracy for sustain versus impulsive instruments.

For the vibrato instrument classifier, 100% classification accuracy was achieved using the single stage classifier. Therefore, the overall MUMs classification accuracy taken across both the non-vibrato and vibrato recordings using the leave-one-out test strategy ranged from 90 – 93%, depending on the non-vibrato instrument classifier used.

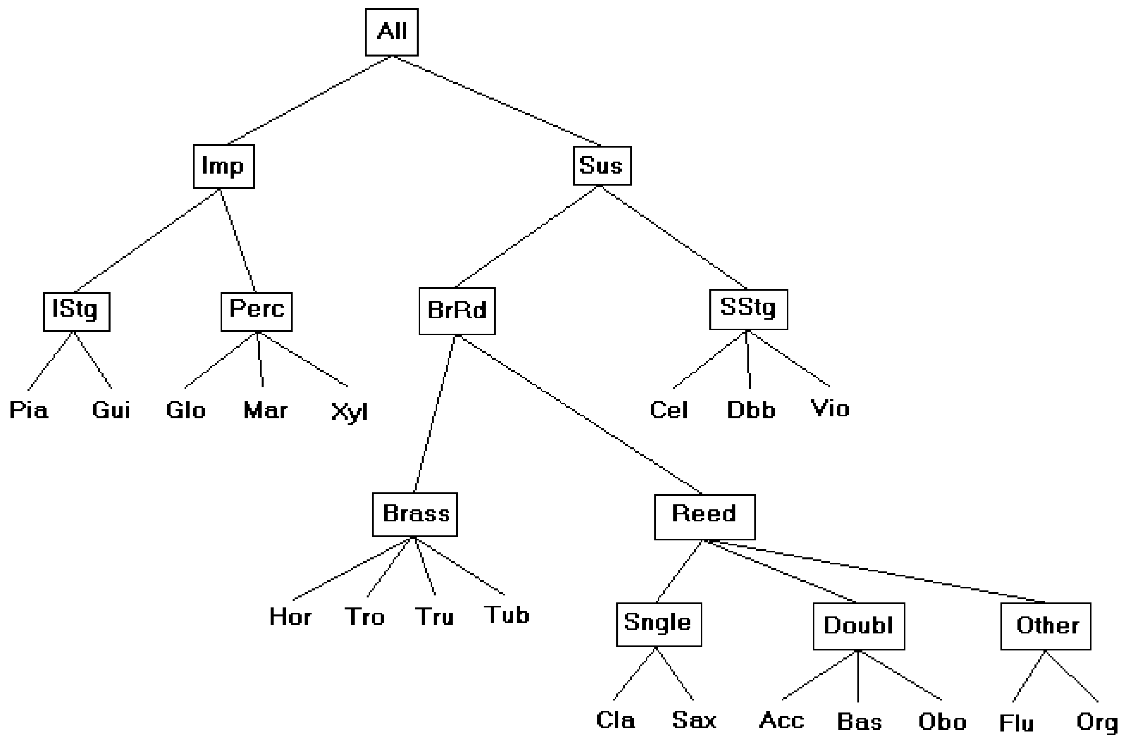


Figure 5 Hierarchic Classifier architecture

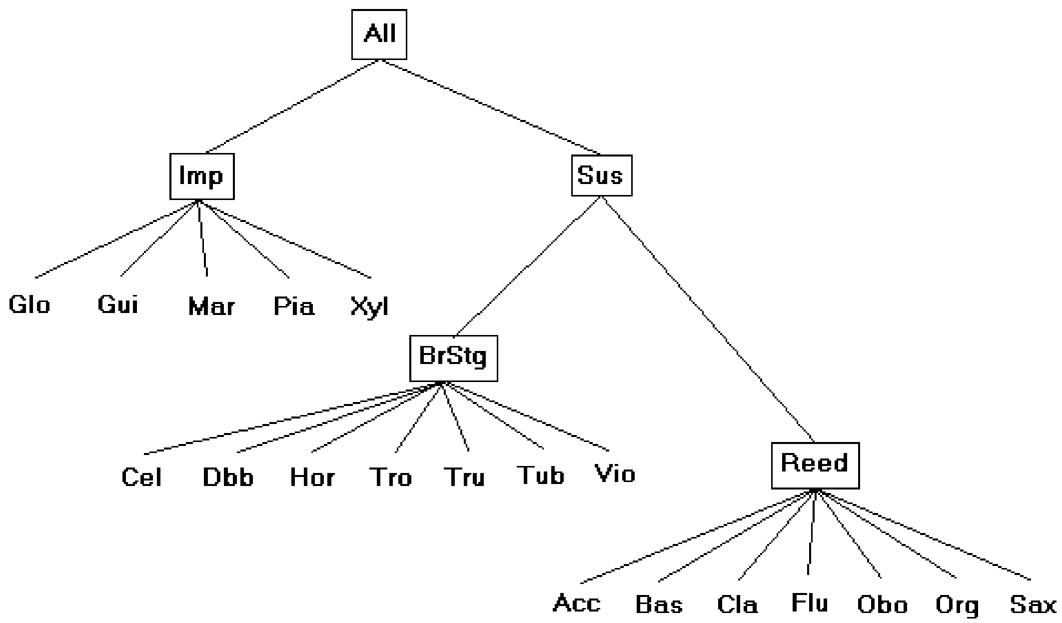


Figure 6 Hybrid Classifier architecture

The generalisation performance of the system was measured using the single stage classifier for both the vibrato and non-vibrato instrument classifiers for the three sets of recordings described above: (a) CE, (b) SOL and (c) TUG. These recording sets comprised of different combinations of vibrato and non-vibrato instrument tones. Direct classification of these sets of recordings, where the MUMs recordings were used as references by both the non-vibrato and vibrato instrument classifiers, provided a wide variety of results, ranging from 33-61%. These differences were attributed to uncontrollable factors during recording production such as different recording environments, players, playing styles, articulations, volume levels and instrument types (e.g. acoustic vs. electric guitar).

In order to make the classification system adaptive to the defining characteristics of any individual set of recordings, and thereby improve its generalisation performance, four enhancements were explored. Firstly, any instruments known to be absent from the test recording sets were excluded from the MUMs reference set during classification. In most cases, the user of the system would know in advance what instruments s/he was working with and attempting to classify. Secondly, a small collection of test recordings, spaced at third octave intervals, was added to the MUMs reference set. In practice, the user would simply need to provide the system with a small number of labeled recordings from the instrument s/he was attempting to classify. Thirdly, the first and second enhancements were combined. As a fourth enhancement, each of the test recording sets was classified using the leave-one out test strategy, with the reference set comprising the test set recordings. Therefore, to apply this enhancement, the user would be required to supply the system with a complete collection of labeled recordings for all instruments that were being classified.

Table 1 summarises the results of applying the first three enhancements to (a) the vibrato and non-vibrato classifiers separately as well as (b) when their results were combined. Due to the small number of vibrato recordings available, the fourth enhancement was only applied to the non-vibrato recordings.

The first enhancement provided anywhere from 13 – 25 percentage point improvement in classification accuracy. For example, for the CE test set, the overall classification accuracy improved from 61% to 74% The second enhancement provided improvements in the range 36 – 39 percentage points. Given the small number of CE recordings (69), it was not feasible to apply this enhancement to these recordings. Combining these two enhancements provided further improvements in the range of 3-12 percentage points. By this stage, the classification accuracy obtained across all sets of test set recordings ranged from 74-83%. Finally, using the fourth enhancement provided classification accuracies for the non-vibrato recordings in the range 87-90%. This compared very well with the 89% classification result obtained with the non-vibrato MUMs recordings using the same leave-one-out testing strategy. As these generalisation results were obtained using the single stage classifier, a further 3% improvement could be reasonably expected when using the hierarchic classifier instead.

Table 1 Overall Generalisation performance for NVIB & VIB recordings for each recording set

Recording Set	Enhancement	Classification Performance		
		NVIB (N (%))	VIB (N (%))	Overall (N (%))
CE (Nt=69) Originally Nnv=46/Nv=23 After Vib Det Nnv=63/Nv=6	Standard	38 (60.3)	4 (66.7)	42 (60.9)
	RIE	47 (68.3)	4 (66.7)	51 (73.9)
	TstRef	N/A	N/A	N/A
	Combined	N/A	N/A	N/A
SOL (Nt=394) Originally Nnv=333/Nv=61 After Vib Det Nnv=352/Nv=42	Standard	134 (38.1)	33 (85.0)	167 (42.4)
	RIE	186 (52.8)	33 (85.0)	219 (55.6)
	TstRef (Nt=292) Nnv=261/Nv=31	203 (77.8)	27 (90.0)	230 (78.8)
	Combined (assume RIE N/A)	213 (81.6)	27 (90.0)	240 (82.2)
TUG (Nt=366) Originally Nnv=312/Nv=54 After Vib Det *Couldn't be run*	Standard	101 (32.4)	19 (35.2)	120 (32.8)
	RIE	156 (50.0)	54 (100.0)	210 (57.4)
	TstRef (Nt=320) Nnv=272/Nv=48	185 (68.0)	33 (68.8)	228 (71.3)
	Combined	211 (77.6)	54 (100.0)	265 (82.8)

Note:
1. For TUG, since only cello vibrato recordings were available, applying RIE always provided 100%

classification accuracy. 2. For TUG, due to short duration recordings available, vibrato detector could not be run. 3. For SOL combined, assumption for vibrato recordings is that since RIE provided no improvement in results on its own, then the same will occur when it is combined with the TstRef enhancement.
LEGEND NVIB: Non-vibrato; VIB: vibrato; N: number of recordings; % : percentage correctly identified Ntot: total number of vibrato and non-vibrato recordings; Nnv: total number of non-vibrato recordings Nv: total number of vibrato recordings; Vib Det: vibrato detection; RIE: reference instrument elimination enhancement; TstRef: enhancement incorporating test tones at third octave spacings into the reference set

6 Discussion

The vibrato detection results summarised in the first column of Table 1 show that the vibrato detector is very good at detecting correctly non-vibrato recordings. Across 896 non-vibrato recordings from three sets of recordings (MUMs, CE and SOL), only two were incorrectly deemed to have been played with vibrato. This corresponds to an accuracy of 99.8%.

Where the vibrato detector had some difficulty related to the correct detection of recordings that were deemed to have been played with vibrato. Across 171 vibrato recordings from the same three sets of recordings, only 119 were correctly identified as having been played with vibrato. This corresponds to an accuracy of 69.6%. The overall detection accuracy across all 1067 vibrato and non-vibrato recordings was 94.9%.

The difficulty the vibrato detector had in detecting vibrato recordings was not based on any fundamental shortcoming of the detector. It brings to light instead the lack of agreement amongst musicians as to what constitutes a vibrato recording. Many of the vibrato recordings simply did not satisfy *all* the detection threshold criteria of the vibrato detector.

Many of the vibrato recordings that were incorrectly classified satisfied some of the "vibrato present" criteria and not others. These could therefore be thought of as being "weak vibrato" recordings. For example, the following SOL recordings contained frequency modulation of the correct frequency (between 5-8 Hz), were deemed to be sustain instrument recordings, but the pk-pk vibrato amplitude was well below 20 cents: cello notes A3#, E4, C3#, and violin notes F5#, G5, D5 and F4. Having the parameters of the vibrato detector user-adjustable, however seems to be most appropriate solution to cope with such a situation.

The classification results of the non-vibrato instrument classifier with the MUMs recordings show that the more effort expended during the classification process, the better the results obtained. A minor improvement in performance was however obtained for a significant increase in computational effort when moving from the single stage to the hybrid to the hierarchic classifiers. The single stage classifier provided the best compromise in terms of performance achieved with computational effort expended. These results comparing the relative performance of direct versus hierarchic classification are consistent with those obtained by Martin and Kim (1998) but contrary to those obtained by Eronen (2000).

The direct generalisation results obtained are consistent with those obtained by Eronen (2001). Eronen is the only researcher who studied the generalisation performance of his system across many sets of recordings. As mentioned in the introduction, he found that on average, only 35% of instruments were correctly identified when many sets of recordings were used. This compares well with the direct generalisation results obtained (33-61%) using a totally different classification system. This further supports the notion that most of the challenge in developing classification system is in achieving good generalisation performance, across many sets of recordings, not just the one set used during system development.

The generalisation results obtained with the system enhancements indicate, not suprisingly, that the more information the user is able and willing to provide the system with regard to the test set being used, the better the generalisation performance that is obtained. This information not only includes any knowledge of the instruments present within the test set, but also any example tones of the various instruments present. The more of these example tones included within the reference set, the better the generalisation results become. The system has shown itself to be remarkably sensitive to the defining characteristics of the specific instruments present within the test set. This is demonstrated by the consistent 87-90% classification results obtained with the single stage classifier across four sets of recordings using the leave-one-out test strategy.

The phenomenon of classification systems being made adaptive to improve their performance concurs with the everyday experiences of speech recognition software users. These systems only begin to offer reasonable performance when trained on the peculiar characteristics of the speaker using it.

7 Conclusion

The results obtained using the classification system developed are consistent with those obtained by a variety of researchers working with a significant number of musical instruments (more than fifteen). It appears that given enough features, it is possible to achieve classification accuracies of the order of 90% or higher with a variety of classification algorithms. These results however have only been obtained using a single collection of instrument recordings for both system development and testing. The main challenge in classification system development, therefore lies in achieving good generalisation performance, across multiple collections of recording collections. This is particularly so given that musicians are most creative in finding new ways of producing different sounds from their instruments. One could argue that this is one of the reasons why music is so interesting to listen to.

In achieving this goal, two possible directions are available. The first approach would be to continue adding more features to the classification system in the hope that the new features would become increasingly less sensitive to any peculiar characteristics of any individual musical instruments. It is the belief of the author that this direction is futile. There exists too much variability in the sounds produced by musical instruments to make such a system practical.

The second approach borrows heavily from the research carried out in developing speech recognition systems. It relies instead on developing a system that is adaptive to the peculiar characteristics of the instruments being classified. The user is given the choice in defining what level of performance s/he would be happy with, based on the amount of information s/he is willing to provide the system. This is the approach currently favoured by the author.

To this end, a system architecture has been developed that works well over a significant number of general instruments. For example, it is not restricted to orchestral non-percussive instruments. Instruments that can be played both with and without vibrato can be classified whether they are played with vibrato or not. If required, the system can be easily expandable to incorporate more features during classification, simply by adjusting the number of individual feature classifiers and the classifier combiner. If necessary, it can also be easily adapted to recognise more instruments playing across a wider range of notes, volume levels, articulation types, simply by including more instruments recordings into the reference set. Most importantly, the system has scalable generalisation performance, where an improvement in results can be achieved based on the amount of effort put in by the user.

The system developed thereby represents but one step in the development of an adaptive classification system, which when properly trained, provides acceptable levels of performance.

8 Bibliography

Agostini, G., Longari, M., and Pollastri, E., 2001, "Musical instrument timbres classification with spectral features , *2001 IEEE Fourth Workshop on Multimedia Signal Processing,* pp. 97 —102.

Brown, J. C., 1999, "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features", *Journal of the Acoustical Society of America,* 105:1933-41

Duda, R. O., and Hart, P. E., 1973, "Pattern Classification and Scene Analysis", Wiley-Interscience.

Eronen, A., and Klapuri, A., 2000, "Musical instrument recognition using cepstral coefficients and temporal features", *Proceedings 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '00,* 2:753-6.

Eronen, A., 2001, "Comparison of Features for Musical instrument recognition", *2001 IEEE Workshop on the Applicationis of Signal Processing to Audio and Acoustics*, pp 19-22.

Fujinaga, I., and MacMillan, K., 2000, "Realtime recognition of orchestral instruments", *Proc. Int. Comp. Music Conf. 2000*, pp. 141-3.

Handel, S., 1995 , "Timbre perception and auditory object identification", in *Hearing*, Moore, B. C. J., New York: Academic Press.

Kaminskyj, I., 1998, "Automatic Musical Vibrato Detection System", *Proceedings ACMA Conference,* pp. 9-15.

Kaminskyj, I., 2000, "Multi-feature Musical Instrument Sound Classifier", *Proceedings ACMA Conference,* pp. 46-54.

Kostek, B., and Czyzewski, A, 2001, "Representing Musical Instrument Sounds for Their Automatic Classification", *Journal Audio Engineering Society,* 49(9):768-85

Martin, K. D. and Kim, Y. E., 1998, "Musical instrument identification: A pattern-recognition approach", *136th Meeting, Journal of the Acoustical Society of America,* 103(3 pt 2):1768.

Wieczorkowska, A., 1999a, "Classification of Musical Instrument Sounds Using Decision Trees", *Proc. Int. Symp. on Sound Engineering and Mastering ISSEM'99,* pp 225-230.

Wieczorkowska, A., 1999b, "Rough Sets as a Tool for Audio Signal Classification", *Foundations of intelligent systems, 11th International Symposium on Methodologies for Intelligent Systems, Proc. ISMIS'99,* 1609:367-375.

Wieczorkowska, A., 2000, "Towards Musical Data Classification via Wavelet Analysis", *Foundations of intelligent systems, Proc. ISMIS'00,* 1932:292-300.

WEB Based Automatic Classification of Musical Instrument Sounds

Mark Williams, Ian Kaminskyj

Faculty of Engineering, Monash University

email: markwil@yoyo.cc.monash.edu.au

email: ian.kaminskyj@eng.monash.edu.au

Abstract

A Web based automatic musical instrument classifier has been developed. It is intended primarily to assist researchers working on the problem of automatic identification of musical instrument sounds. It is a web-based implementation of the classification system developed by Ian Kaminskyj. The system operates on unknown sound files, detecting the presence or absence of vibrato and extracting five features. The system uses the features to identify the most likely instrument to have produced the test input. Java and Java 3D has been used to provide a system that is platform independent. This web based system produces graphical representations of the features extracted from the unknown recording. JAVA3D allows for rotation, zooming and translation of three dimensional feature waveforms. The WWW site provides interactivity, a graphical user interface and ease of use.

1 Introduction

There are many software tools available to researchers for processing sound files, e.g. Pitch Detection SDK (www.mb-usa.com/english/pdsdk.shtml) for automatically detecting and correcting pitch. There are a number of useful web sites for researchers working on the problem of analysis of musical instrument sounds. These include: The Studio Online (SOL) Project (www.ircam.fr/produits/technologies/sol/index-e.html), the Sandell Harmonic Archive (SHARC) Database (http://sparky.parmly.luc.edu/sharc/) and Tae Hong Parks' feature extraction software (http://silvertone.princeton.edu/~park/thesis/).

The SOL project not only provides a source for musical sounds but also furnishes an interactive feature that allows for online transformation (stretching, transposition, filtering etc.) of reference sounds together with the users own sounds. The McGill University Master Samples (MUMS) provides a number of reference CD's with musical instrument recordings. They have been utilised on the SHARC website, wherein static plots of predetermined spectral analysis results are carried out on a number of recordings. The feature extraction software described by Park (2000) is a Java application that provides useful tools including root mean square (RMS) amplitude envelope/spectral centroid calculation and pitch detection (currently, the software is not readily available on the WWW).

A World Wide Web site to automatically classify musical instruments and dynamically calculate features has been developed. It is intended that the site will further assist researchers in both the analysis and automatic recognition of musical instrument sounds.

2 Theory

The classifiers system structure is shown in Figure 1. The vibrato detector firstly determines the presence of any vibrato within the input test sample. If vibrato is deemed to be present, the vibrato instrument classifier is invoked, which classifies the input as having been produced by one of three possible vibrato instruments: cello, flute and violin. Alternatively, if no vibrato exists, the non-vibrato instrument classifier is executed. It classifies the input as having been produced by one of nineteen possible non-vibrato instruments.

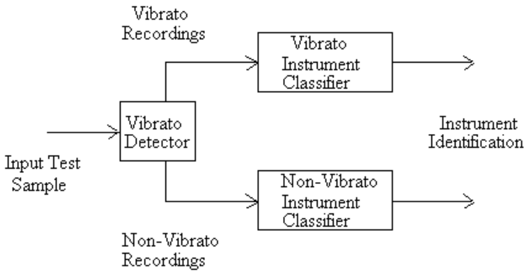


Figure 1 Classification System Structure

The vibrato/non-vibrato instrument classifier is shown in Figure 2 and uses five single feature classifiers to independently determine the most likely instrument to have produced the input test sample. The features extracted by the system and thereby used by the classifier include the Root Mean Squared (RMS) amplitude envelope, cepstral coefficients (CCC), Multidimensional Scaling Analysis trajectories (MSA), Constant Q Transform (CQT) frequency spectrum and the spectral centroid (SCD). Kaminskyj (2000) describes the features in detail. Five classifiers (vibrato or non-vibrato depending on the result of the detector) use these five features to independently identify the test sample. The results of the individual classifiers are combined together via a confusion matrix accuracy and reliability-weighting scheme which produces an estimate of the most likely instrument to have produced the test input. Kaminskyj (2002) provides further details regarding the classification scheme.

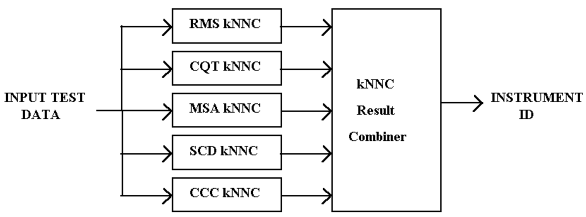


Figure 2 Classifier Details

3 Implementation Method

HTML, JavaScript and Java and were used to develop the web-based system. HTML coupled with JavaScript provides more functionality than solely using HTML. JavaScript was, for example, used to validate the user's input (parameters) before being passed to the musical instrument classifier program. Java provides many classes for dealing with file I/O and drawing objects and is a platform independent language. It can also be used to create either standalone applications or java applets that are executed over the web. Java's Abstract Window Toolkit (AWT) was chosen to implement the output feature plots rather than the more recently developed Java Swing classes. The rationale behind this was that the AWT classes provide all of the functionality required and do not require additional WWW browser plugins. Unlike Java Swing, it is compatible with older versions of Java. The integrated development environment used to write the software was Sun Microsystems' Forte for Java.

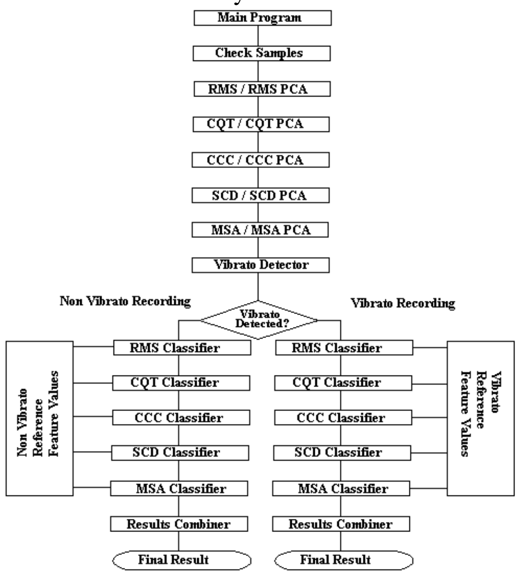


Figure 3 Web classifier software flow chart

Figure 3 shows a software flow diagram, illustrating the order in which the programs are executed and the way the system classifies the instruments.

The main controlling program is a web page form that accepts the various parameters (name of the WAV file to be uploaded, pitch information, which feature classifiers to bypass, classifier search limits etc) from the user and validates them.

The primary purpose of the main program is to execute the sundry programs in sequence and to pass user parameters. The sequence of the programs is significant as results of some programs are used as input to subsequent programs (for example the output of the feature extraction programs is used as input to the principal component analysis programs). After the user-defined parameters have been verified, the length of the input WAV file is examined to determine if there are enough samples to perform the feature extraction (if there are not enough samples the applet will stop and generate an error message). Assuming that there are enough samples, the header information of the WAV file is removed, leaving the raw data in a text file format, suitable for use as the input to the feature extraction programs.

Five features are then extracted from the unknown WAV file directly. These feature waveforms are then used as input to the principal component analysis (PCA) programs, which reduces the amount of information processed by the classifier (Jollife 1986); PCA feature waveforms are thereby obtained.

The file is then checked for the presence of vibrato. The vibrato check is a fundamental feature of the system (as it reduces the number of reference recordings that the unknown recording needs to be compared to during classification). The main program runs a set of vibrato/non-vibrato classification and combiner programs (as illustrated in Figure 1) based on the vibrato detection result obtained. The parameters of the vibrato detector can be changed to allow for different thresholds of vibrato: the parameters comprising the minimum amplitude, and upper and lower frequency limits.

Following the vibrato detection, each feature (either direct or PCA) is used by an individual feature classifier to identify the input sound.

The RMS, CCC, CQT, MSA and SCD classifiers compare the direct or PCA features of the unknown input recording to a set of reference recordings and select the closest match via the *k* Nearest Neighbour Algorithm (*k*-NNC) (Duda & Hart 1976).

3 Graphical User Interface

The parameter settings are obtained by a HTML form, which has been broken into two parts. Figure 4 shows the first component of the input form; here the user specifies the unknown input WAV file to be uploaded together with its pitch. (The system requires that the pitch is known *a priori* and there are a number of programs available to perform this function (e.g. Pitch Detection SDK (www.mb-usa.com/english/pdsdk.shtml), TUNE!IT (www.zeta.org.au/~dvolkmer/tuneit.html) and RBC Voice Tweaker (www.thedirectxfiles.com/features/rbcvt/rbc1.htm)). The user can turn vibrato detection on/off, alter the threshold values of the vibrato detector and enable/disable any of the five individual feature classifiers. Figure 4 shows the default settings of the classifier that were determined to be optimum for the system. For vibrato recordings, CQT, SCD and RMS feature classifiers are excluded from classification. For non-vibrato recordings however, the RMS feature has been included. The user can modify these settings.

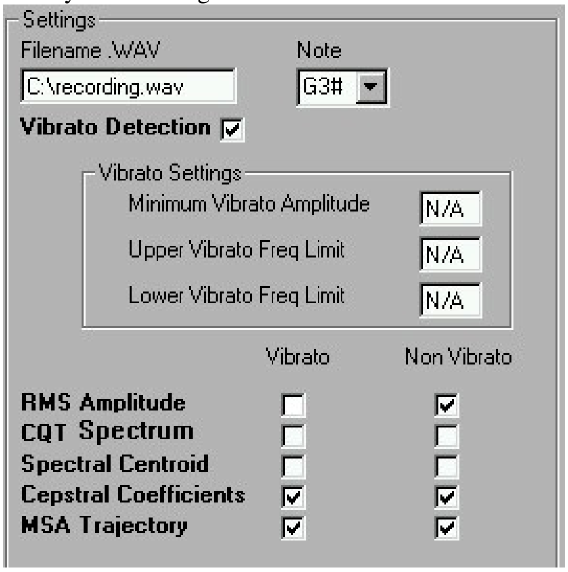


Figure 4 Input form

The second part of the input form is the classifier and combiner settings, shown in Figure 5. These settings allow the user to change the classifier and combiner settings to improve the accuracy of the result. The instrument elimination check box allows the classification program to exclude instruments when comparing the unknown input WAV file to the reference recordings. For example, when the unknown instrument sound has a pitch that is outside the playing range of an instrument, the classifier will exclude this instrument from classification (e.g. this would ensure that a marimba C3 test tone was never misclassified as a glockenspiel, an instrument whose note range does not extend below G5). The confusion matrix accuracy and reliability settings of the combiner can also be adjusted by selecting from the drop down box. By default, the combiner is based on an empirically determined reliability confusion matrix weighing scheme. The settings however, can be changed to have additional schemes such as (a) based on accuracy and reliability, (b)

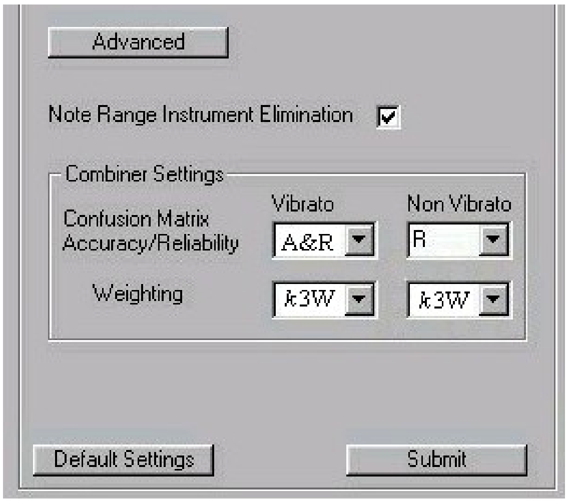


Figure 5 Classifier and Combiner parameters

solely based on accuracy only, or (c) a scheme independent of both accuracy and reliability.

The combiner NNC majority weighting scheme can also be adjusted to $k1$, $k3$ unweighted/weighted and $k5$ unweighted/weighted NNC majority weighting scheme. By default, the weighting for the combiner is set to the $k3$ weighted majority voting scheme. The individual feature classifier settings can be adjusted by pressing the advanced button, opening a window of options shown in Figure 6.

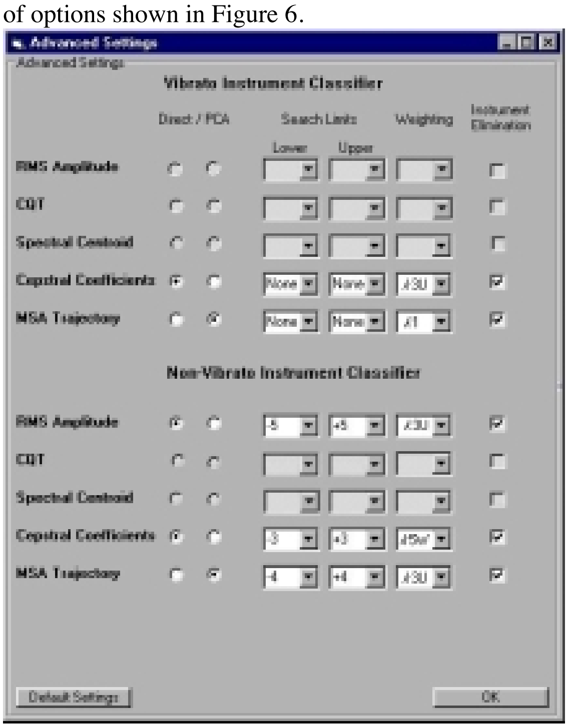


Figure 6 Advanced Settings

As can be seen the classifier settings can be changed to limit the classification to a smaller number of references occupying a restricted range of notes. The search limits indicate what note range to restrict the comparison of references with the test input. The direct or PCA features can be selected for use by the individual classifiers. Instrument elimination for each feature classifier can be individually enabled/disabled by selecting the appropriate checkbox. This however, is only possible if the instrument elimination checkbox on the main parameter page was checked. The weighting of each feature classifier can be set in the same way as the combiner settings of Figure 5. Pressing the OK button will submit the user chosen parameters to the classifier and return the user to the main page. Note that the options for RMS, CQT and SCD are not available under the vibrato recording settings because they were not been selected in Figure 4. Similarly, for the non-vibrato classifier, the CQT and SCD options were not available.

The graphical results of the feature calculations are displayed in the right hand side of the web page as illustrated in Figure 7. After pressing the submit button, the user is presented with plots of the enabled feature waveforms for the unknown recording. For comparison purposes, the corresponding waveforms of the closest reference that was found for each enabled individual feature classifier is also shown. Again note that the RMS, CQT and SCD feature plots and closest references are not shown because they were disabled for the vibrato instrument classifier in the settings of Figure 4 (the unknown sound file was found to contain vibrato).

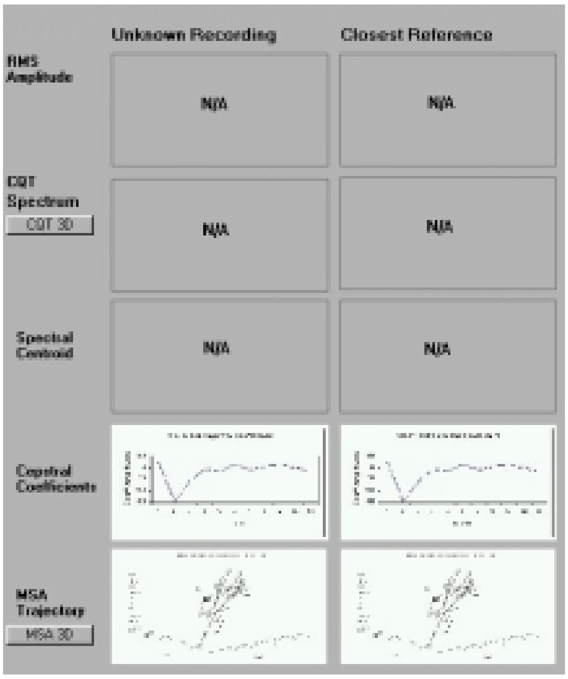


Figure 7 Graphical output of features

Finally, the results of the automatic classification system are displayed at the bottom of the page as shown in Figure 8. The vibrato detection result is displayed, together with the individual feature results. The overall result based on the combiner settings used is also presented.

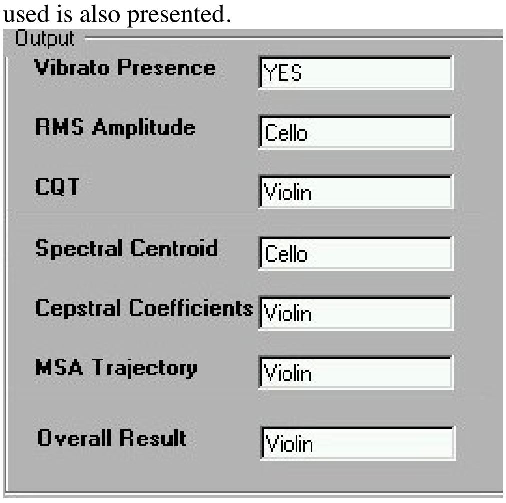


Figure 8 Results pane

4 An Example Session

In order to clarify how the WWW page would typically be used, a hypothetical example session will be detailed. In the sample web page shown in Figure 9, the user uploaded the unknown source file HORD3SC.WAV (a French horn playing the note D3 sharp). The pitch of the unknown input WAV file has been selected from the drop down box (pitches C3 to C6 are available for selection). The vibrato detection feature was enabled and its settings were not changed. Therefore the threshold default values were used (minimum vibrato amplitude = 20 cents, upper vibrato frequency limit = 8 Hz and lower vibrato frequency limit = 5 Hz). The user elected to include all feature classifiers by selecting all five checkboxes. If, for example, the user chose to deselect the RMS feature, the RMS feature classifier would not run and hence not be used in the combination of the individual feature classifier results.

Figure 10 shows the advanced settings for this session. As the user elected to use all features, the advanced feature settings were available for all individual feature classifiers. Only the direct feature values were used during classification and instrument elimination was enabled for all feature classifiers. A number of different search limits and weightings were used on each feature classifier. For example, the non-vibrato RMS feature classifier search limits were set to +/- six notes (an octave). This restricted the RMS classifier to six notes above the unknown input WAV file pitch and six notes below, when comparing it against the reference instruments. The user then set the combiner settings on Figure 9 to use an accuracy and reliability confusion matrix weighting scheme for both the vibrato and non-vibrato instrument classifiers. The accuracy and reliability confusion matrix weighting scheme uses empirical accuracy and reliability results to weight the results of the individual feature classifiers.

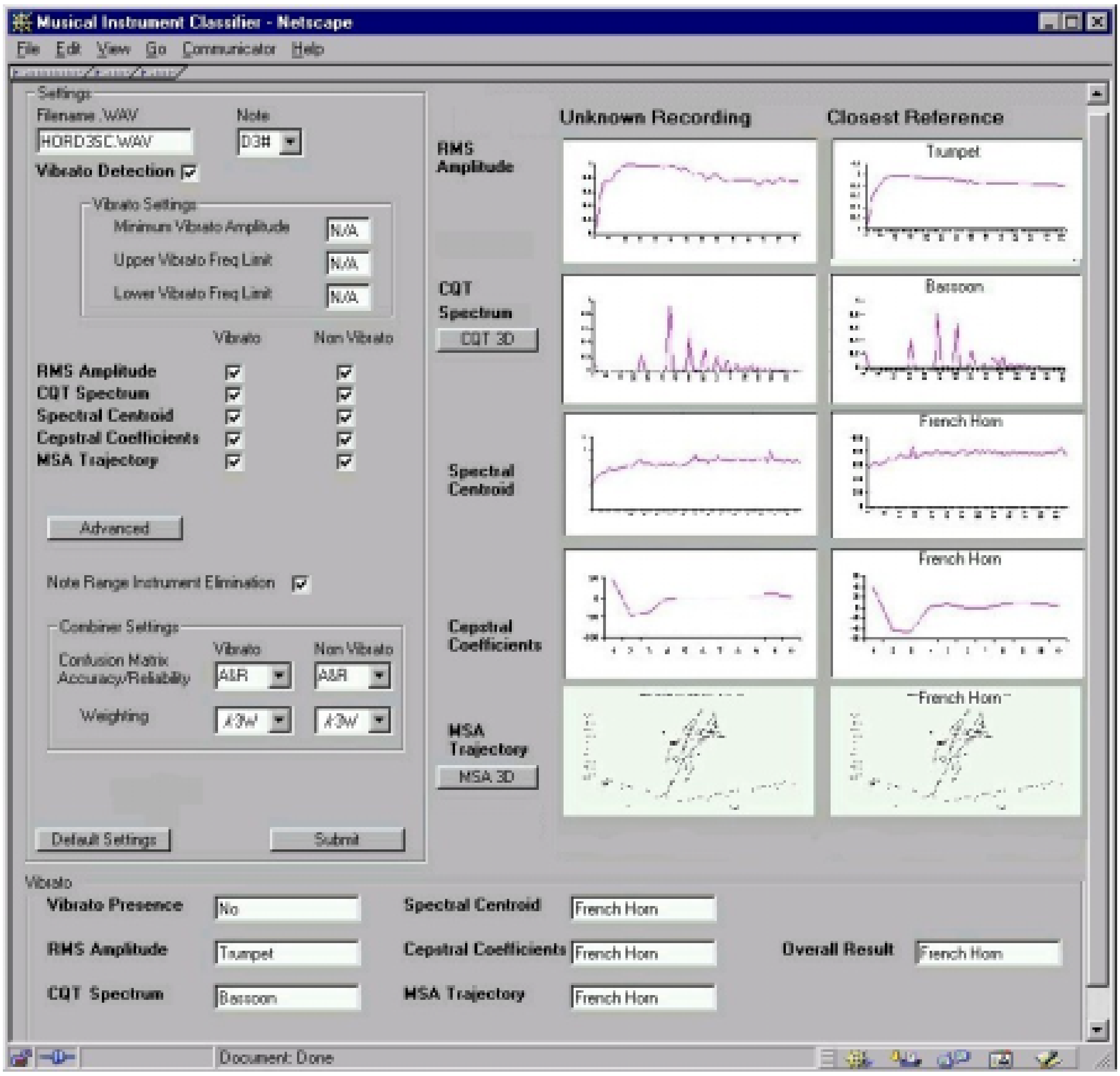


Figure 9 Sample Session

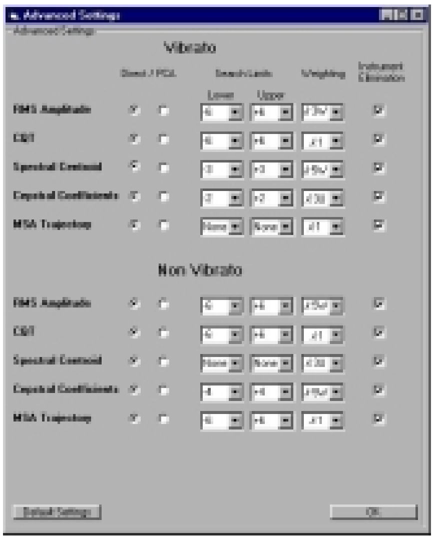


Figure 10 Advanced Settings: Sample Session

The submit button was used to start the program, after which the feature graphs and results appeared. (The default button restores the parameter settings on the form to their default values.)

The graphs plotted in Figure 9 show the RMS, CQT, SCD, CCC, and MSA extracted feature waveforms. To the right of these graphs is the feature data for the closest reference plotted. This allows for a visual comparison of the calculated feature and the closest reference found. In the Figure 9 example, the output on the right hand side shows an RMS Amplitude plot of a trumpet which was deemed by the RMS classifier to be the closest reference to the RMS amplitude envelope of the unknown input WAV file. Below this is a CQT plot of a bassoon, deemed to be the closest reference to the unknown input WAV file by the

CQT classifier. The CQT 3D button plots a 3D spectral waterfall plot of the unknown input recording as shown in Figure 11. The Java 3D graph allows the user to zoom in/out rotate and translate the graph. This feature is also provided to plot the MSA trajectory by pressing the MSA 3D button.

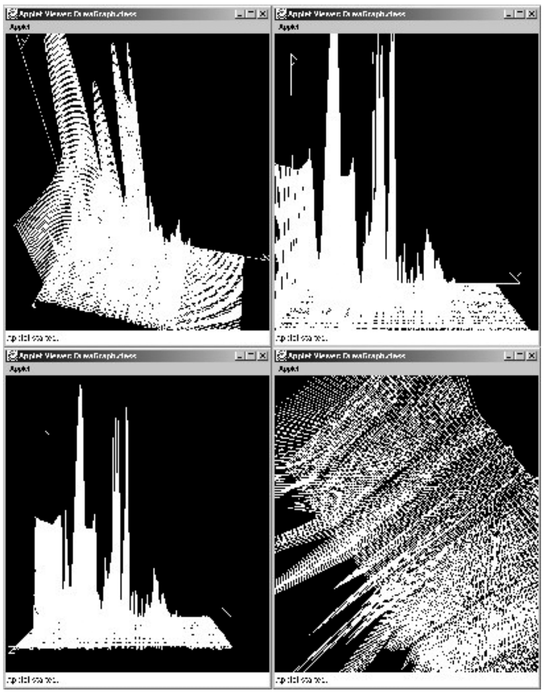


Figure 11 Java 3D: CQT Spectral Waterfall

The spectral centroid, cepstral coefficient and MSA trajectory plots of the closest reference are of a French horn, which is, as we know the same instrument as the input file. While three out of the five feature classifiers were correct, the combiner weighted these results based on the chosen accuracy and reliability weighting scheme. As can be seen at the bottom of Figure 9, the overall result was that the most likely instrument to have played the test input sound was a French horn.

4 Conclusion

The web based automatic musical instrument classifier presented provides remote access to sound processing and analysis via an interactive, platform independent application. The system can be readily distributed to the user via web applet (located at http://www-personal.monash.edu.au/~kaminski/) or as a standalone Java application that can be run offline. There is much scope for advancement including: supporting different sound file formats, providing automatic pitch detection, the ability to update the reference files to include test tones (which will enhance the general performance of the system Kaminskyj 2002), reference instrument elimination (i.e. if it is known beforehand that some instruments were absent from the test set recordings), expanding program parameter options, the ability to classify multiple recordings and extract new features to help identify instruments.

References

Kaminskyj, I., 2000, "Multi-feature Musical Instrument Sound Classifier", *Proceedings ACMA Conference*, pp. 46-54.

Kaminskyj, I. 2002, "Multi-feature Musical Instrument Sound Classifier w/user determined generalisation performance", Proc. Aust. Comp. Music Assoc. Conf. 2002 Melbourne, Australia, July 2002, (accepted for publication)

McGill University Master Samples, http://www.music.mcgill.ca/resources/mums/html/mums.html [Accessed 27/04/02]

The Studio Online Project, www.ircam.fr/produits/technologies/sol/index-e.html [Accessed 27/04/02]

Sandell, G.J, 1995, "SHARC Timbre Database", http://sparky.parmly.luc.edu/sharc/ [Accessed 27/04/2002]

Park T.H, 2000, "Salient Feature Extraction of Musical Signals", http://silvertone.princeton.edu/~park/thesis/ [Accessed 29/04/2002]

Jollife, I. T., "Principal Component Analysis", Springer-Verlag, 1986.

Duda, R. O., and Hart, P. E., "Pattern Classification and Scene Analysis", Wiley-Interscience, 1973

Pitch Detection SDK, 2001 www.mb-usa.com/english/pdsdk.shtml [Accessed 27/04/02]

TUNE!IT - Musical Instrument Tuning Software for Windows, 2001, www.zeta.org.au/~dvolkmer/tuneit.html [Accessed 27/04/02]

RBC Voice Tweaker, 2001, www.thedirectxfiles.com/features/rbcvt/rbc1.htm [Accessed 27/04/02]

Abstractly Related and Spatially Simultaneous Auditory-Visual Objects

Andrew D. Lyons

Composition Unit, Sydney Conservatorium of Music, University of Sydney
email: tstex@tstex.com . *web:* http://www.tstex.com

Abstract

This paper discusses various design issues related to the integrated synthesis of 3D sound and 3D graphics. Issues of particular concern are those related to a style of audio-visual integration in which the perceived attributes of sonic and visual phenomenon are mapped to each other. The relationship between auditory and visual perception can draw on synesthesia, mental imagery and creativity. Unique problems result from the combination of these cross-modal design concerns and concerns for dynamic and realistic spatialisation. These problems are discussed within the context of design for reproduction involving traditional screen and multiple channel audio theatre systems. Works by the author are used as examples including a recently completed 3D audio-visual work for DVD performance called "Heisenberg". The paper concludes with a call for greater treatment of visual artifacts involved in music cognition within music education.

1 Introduction

Works of the sort discussed here can be described using a schema of four design criteria. Issues related to these design criteria and their combination form the basis of all discussion in this paper. They will be referred to in this paper numerically as denoted in table 1.1 below:

1. Spatialised Sound - The realistic and dynamic spatialisation of sonic objects for performance using multiple speakers or headphones.
2. Spatially Simultaneous Sonic and Visual Objects - The visual representation of spatialised sound source locations using 3D computer graphics and a single screen display.
3. Abstractly Related Sonic and Visual Objects The representation of mental imagery resulting from music and the abstraction into the visual domain of other cognitive artifacts attributable to sonic objects.
4. Effective Design - The development of audiovisual works which satisfy the fundamental design tenets of both sonic and visual arts.

Table 1.1 The four design criteria applied by the author.

2 Primary Problems

In the authors research, each of the first three criteria in table 1.1 have been responsible for what may be considered simple or primary problems. Complex or compound problems begin to arise when attempts are made to integrate the first three criteria in a way that satisfies the demands of the fourth criteria. Because the compound problems often result from combinations of primary problems each shall be discussed in turn.

2.1 Primary Problems Related to Sound Spatialisation

2.1.1 Design Criteria One

3D sound spatialisation is defined here as any system which is based on, or extends the original John Chowning (1971) spatial synthesis system. 3D sound systems as defined here do more than just pan sounds around an n-speaker array. A 3D sound system should also synthesise distance cues, doppler shifts, the six early reflections of sound off walls, local and global reverberation and perhaps include other features in order to generate an aural virtual reality. (Begault 1994) For the creation of *Heisenberg* an automated 3D sound spatialisation system of this sort dubbed "*Pan*" was designed.

2.1.2 Spatial Resolution in Sound Spatialisation

It is well known that the number of speakers available in a specific speaker array has a direct relationship to the degree of spatial precision achievable with that array. (Begault 1994) As more speakers are added to an array, each individual sound source can be differentiated more easily from others, and its location determined more accurately. (Begault 1994)

The greatest spatial resolution in the ITU-R BS.755-1 5.1 speaker array corresponds to the sixty degree arc immediately in front of the audience - where three speakers and the visual display are normally located. The author observed that the movement of spatially coincident sounds and images across this frontal 60 degree arc precipitated a generalised perceptual plausibility which in turn enhanced the evocative power of the acousmatic space beyond the area of the screen

Besides the problematic limited spatial resolution inherent in five speaker arrays, the absence of speakers in the vertical plane might also be considered a limitation. The absence of speakers above and below the horizontal plane removes the ability to create sonic images above and below the audience. *Heisenberg* was subsequently animated so that all motion takes place not too far from a horizontal plane level with the camera.

2.1.3 Scale Issues and the Inverse Square Law

One problem associated with all spatial synthesis systems involves ambiguity surrounding the correct function with which to create distance attenuation effects. The inverse square law is suggested widely to be the correct function with which to attenuate the amplitude and spectral content of a sound source to create the illusion of distance. (Dodge 1997) (Roads 1997) (Begault 1994). In Moore (1990) however, the inverse cubed law is stated to provide a more perceptually tenable relationship between distance and amplitude. It would seem that the exponent in this function holds the key to creating different relationships between distance and attenuation.

It is well known that the volume of an extremely loud sound decays over greater distances than that of a quieter one. (Bregman 1990) When working with digitised sound files however it is not always good practice to digitally store such scales of amplitude due to problems with either clipping or narrow bit resolution. By making the exponent in the distance inversion function a parameter called "scale", it was possible to work with an optimally sampled sound file and create either the perception of a nearby insect or a distant Jumbo Jet. This level of control makes itself most apparent in *Heisenberg* in the dramatic yet slowly shifting doppler shifts created by fast moving loud distant sound sources. Without the scale parameter distant sound sources and the dramatic doppler shifts created by large relative velocities would be inaudible. This is a feature not documented in some classic texts on spatialisation nor implemented in some proprietary 3D sound systems.

2.2 Problems Related to Spatially Simultaneous Sonic and Visual Objects.

2.2.1 Design Criteria Two

In our day to day life we are able to attribute sounds causally to an event or object which is more often than not visible in some way. In this way our vision and hearing provide complimentary information about the location of events around us. It has been shown that each modality influences the spatial localisation of a stimuli source established by the other. (Lewis, Beauchamp and DeYoe. 2000) The synthesis of such spatial audio-visual relationships lends considerable authenticity to virtual environments. (Begault 1990) While there is a large amount of visual bias in situations when discrepancies between visual and auditory spatial occur, (Welch and Warren 1980) the correlation of cues from each modality may serve to encourage and reinforce the perception of intended abstract relationships between auditory and visual objects.

2.2.2 Temporal resolution

The main strength of *Pan* for the author is its ability to provide perfect temporal synchronisation between sounds, their trajectories and their animated representations. No external devices are needed to achieve accuracy within the range of one audio sample. This kind of accuracy does however seem a little like overkill when one considers the temporal resolution of the television screens for which the animation is prepared. With such displays there is no satisfactory means to visually represent a sound which has a dynamic spectral envelope that endures no longer than the minimum duration for which a frame can be displayed. This is one twenty fifth of a second in the PAL format used in Australia and Britain and roughly one thirtieth of a second for the NTSC format used in the Unites States.

This problem becomes more insidious when percussive passages take place at a regular rate that is slightly out of phase with the redraw rate of whatever television standard is being used. For this reason all percussive passages in *Heisenberg* are synchronised with the 25 frames per second redraw rate of the PAL television standard. In most cases this means tempos of 93.75 bpm are used. While it remains impossible to accurately represent percussive sounds with any temporal detail, at least it is possible to match their onset and offset with the onset and offset of an image. Creating passages that synchronise in this way, and at this rate creates an effect commonly referred to as strobing. Strobing and flicker effects are regarded favourably in some circles and not so favourably in others, and ideally screens and projection systems with much faster redraw rates will be available in the long term.

2.2.3 Spatial Dimensions

When creating spatial works for cinema style performances, one becomes aware of a great disparity between the spatial range available for representing objects sonically and the spatial range available for representing objects visually. Irrespective of spatial resolution, a sonic object can be located anywhere around the audience on a horizontal plane when using five surround speakers. The representation of all visual objects however must vie for space within whatever dimensions are afforded by the available screen.

2.2.4 Calibrating for Spatial Simultaneity.

In a 3D animation in which the camera is both moving around a scene and panning from side to side, it is important that sound sources and their visual representation move together relative to the camera position and angle of rotation. If the camera pans to the left, then related sonic and visual objects should move off towards the right together relative to the central axis of view. It is one of the functions of *Pan* to ensure that all sounds are spatialised relative to the position and rotation of the camera being used to shoot the animation. It is essential that the audiovisual field remains spatially coincident during camera transformations, because ultimately the position and rotation of the camera represents the spatial position and alignment of the audience during theatre performances.

An extension of this problem is the need to spatially calibrate the visual field to match the rendered sound field. To calibrate the sound field spatially it should only be necessary to centre the circle of speakers around the supposed central audience position and arrange them using the angles indicated in the ITU 5.1 theatre sound specification. To calibrate the visual field however one should consider - before the animation is created - the angular relationship that exists between the central audience position and the side edges of the screen in the target theatre situation. This angle must then be used as the field of vision in the camera that shoots the 3D animation. In this way sounds and associated objects will appear to be spatially coincident. Figure 2.1 below may help explain this problem.

Figure 2.1 - Field of Vision and Spatial coincidence.

2.3 Problems Related to Abstractly Related Sonic and Visual Objects

2.3.1 Design Criteria Three

It has been a major objective of the author's research to create animations that explore cross-modal exchanges between audition and vision, and which explore mental imagery associated with the experience of listening to acousmatic and computer music. The author appreciates that mental imagery takes highly individual forms, and the imagery of one does not often resemble the imagery of another when responding to the same piece of music. It is hoped however that research describing universal styles and recurring themes in mental imagery, synesthesia and other cross-modal categorisation systems will help offset this subjectivity. A system involving this research has been utilised in the development of *Schwarzchild*, *Loucid* and *Heisenberg*.

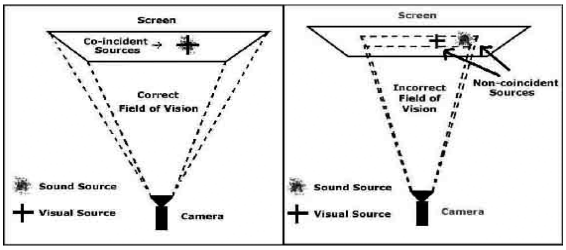
2.3.2 Mental Imagery

When discussing mental imagery it is important to differentiate imagery specific to each sensory modality. Auditory, visual, kinesthetic, tactile, and olfactory imagery are all separate but related fields of cognitive psychology. (Richardson 1999) Such is the dominance of occularity that in its broader usage "mental imagery" may be taken to mean the visual variety of mental imagery. The auditory kind of mental imagery is generally referred to as "auditory imagery" in this and other literature. (Reisberg 1992)

It has been suggested that fifty percent of people experience mental imagery of some sort when listening to music. (Huron 1999) Recent PET and EEG scans of the human brain have confirmed the employment of neural areas usually associated with vision in the exercise of musical tasks. (Nakamura et al. 1999) (Platel et al. 1997) Psychologists regard music as being a powerful source of mental imagery. (Quittner 1980) The use of mental imagery in music therapy techniques such as the Bonny Method of Guided Imagery (Goldberg 1995) has created a body of research describing the features of such imagery. (Cho 2002) (Lem 1998) Some recurring themes in mental imagery in response to musical stimuli include "'Nature scenes' i.e. sun, sky, ocean, plant and animal, etc., 'Journey', 'I' and 'Emotion' i.e. happiness, sadness, depression, etc." (Cho 2001)

2.3.3 Synesthesia

While many experience mental imagery as a result of musical audition, only one in twenty five thousand people experience synesthetic perception, in which coloured forms appear involuntarily in the field of vision in response to sonic stimuli. (Cytowic 1992)



Auditory-visual synesthetic imagery may be described as the superimposition of geometric figures known as "photisms" over the normal field of vision. While being largely flat and figurative, synesthetic photisms do have a sense of extruded depth.

Following a query by the author regarding dimensional features in synesthetic photisms, a synesthete named Sarah provided the following description of her synesthetic perception: "I have sound>colour synesthesia; to put it simply I see colours, and sometimes patterns, when hearing sound. To me these colours and patterns seem two-dimensional. When I hear words and simple sounds the colours are also simple; flat or graduating colours depending on what I'm hearing. When the sounds have some sort of rhythm, as with music or even poetry, these colours form moving patterns… To me it looks like a sort of filter or overlay. I'm not even sure sometimes, if the colours are projected into space at all, or if I'm just seeing them "in my mind's eye" so to speak. I still see these colours with my eyes closed, for instance. In fact I see them more strongly with my eyes closed."

A synesthete named Lisa then offered a description of her perception: "My sound->sight synesthesia is sometimes projected as flat, like Sarah's, but mine can just as easily be 3d with depth. For instance, I was riding in a friend's van the other day, and the side door wasn't all the way shut. It kept making this noise that to me looked like slightly asymmetrical orangey-yellow cylinders coming from the top left of my vision to somewhat near me on my bottom right. (It drove me nuts for almost an hour, 'til I was able to re-close it.)"

There is a notable difference between these description of synesthetic perception and the three dimensional landscapes and journeys common in non-pathological mental imagery. (Cho 2001) For the development of spatially dynamic audio-visual works, other systems must be drawn on to compliment any understanding of cross-modal exchange based on synesthetic perception.

2.3.4 A Systematic Approach

The author has devised a systematic approach to the visual representation of aural attributes which may assist others in developing such works. The first stage in this process involves reducing imagery to four main levels.

This reduction involves an illustrative consideration of the neural dispersion of cognitive activity responsible for each of: synesthesia, spatial cognition, associative mental imagery, and causal analysis. The reduction of auditory-visual imagery using this schema permits the further analysis of each of the four main levels using more specialised modal translation systems.

Level	Location	Features
Level One	Sub – Cortical	Figurative
Level Two	Parietal	Spatial
Level Three	Right Frontal	Associative
Level Four	Left Frontal	Causal

Table 2.1 - The author's four part reduction of auditory-visual imagery. (It should be noted that the neural locations are very generalised.)

2.3.5 Level One - Synesthesia

The first level of the four part scheme in table 2.1 describes relationships between sound and visual imagery purported to take place at a sub-cortical level. Many neurologists involved in research into Synesthesia believe that this is the area of the brain responsible for Synesthetic perception. (Cytowic 1989) (Baron-Cohen and Harrison 1996).

Research into Synesthesia has suggested constant relationships between certain aural qualities and resultant visual percepts. (Marks 1978) In many of the exchanges between the modalities of audition and vision in synesthetic perception, intensity plays a significant role. (Marks 1978) In the context of aural perception, intensity is a product of both high frequency spectra and loudness. In visual perception intensity is a function of brightness. The relationship of aural intensity to visual intensity experienced by most synesthetes is similar to that of non-synesthetes. (Marks 1978) Some cross modal relationships are set out below in Table 2.2:

Aural Feature	**Visual Counterpart**
High Pitch	Small (bright) photism
Low Pitch	Large (dull) photism
High Loudness	Bright Photism
Low Loudness	Dull Photism
"Course" sonic texture.	Rough/Sawtooth shapes
"Smooth" sonic texture.	Smooth flowing shapes.

Table 2.2 Mapping aural features to visual figurative features in synesthetic photisms.

While it is not explicitly described in any cross-modal systems, intensity due to loudness can also be associated with the size of an object. This is perhaps related to the way that objects become larger and louder as they come closer to an observer. It may be noted that a direct relationship between loudness and size would at times conflict with the idea that objects with high primarily frequency spectra are small in visible size. There is also a potential for conflict in situations involving loud sounds with a low pitch - they need to be both bright and dull at the same time. The similar applies to quiet, high frequency sounds.

Besides these conflicts of attribution, the relationship between aural texture and visual texture would seem to be intuitive. Sounds with saw-tooth like amplitude envelopes precipitate saw tooth shaped photisms. Also, sounds with smooth envelopes produce smooth photisms. Not in reference to synesthetic photisms this sonic-visual topological equivalence is also described as "physiognomic perception". (Davies 1978)

Colour is considered to be a highly subjective and widely varying visual attribute of sound by both synesthetes and non-synesthetes.

2.3.6 Level Two — Spatial Aspects

In a positional context, the relationship between space and music is the co-existence of sonic and visual objects. While there is a large amount of visual bias in situations when discrepancies between visual and auditory spatial occur, (Welch and Warren 1980) the careful correlation of cues from each modality may serve to reinforce the perception of intended abstract relationships between auditory and visual objects.

In a design context however, the relationship between space and sound can become more complex. In classical literature, architecture is often described as having a relationship to music in the use of proportion. (Yi 1991) (Bragdon 1939) (Ong 1994) (Treib 1996) Comparisons based on the broader principle of gestalt isomorphism are also useful in formulating relationships between music and architecture. (Wertheimer 1938) (Lyons 2000).

A dynamic variant of isomorphism may however be more useful in temporal works. Such an idea has been expounded by Candace Brower (1998) in which Density 21.5 by Edgar Varese is described in terms of pathways, containment and blockage. This idea of containment and release was a central visual design feature in the author's 2000 animation: "Loucid" and then again in *Heisenberg*.

In some contexts the relationship between human gesture and music (Battey 1998) can be more useful in describing spatial-auditory relationships. This is for three reasons. The first is the powerful existential relationship between kinesthetic, auditory and visual cognition. (Priest 1998) The second is related to the fact that human gesture is at once spatial and temporal whereas architecture is spatial, but frozen in time. The third is the ability of human gesture to be performed within a confined visual field - such as that afforded by a visual display.

2.3.7 Level Three — Associative Imagery

The relationship of mental imagery to sound stimuli is highly subjective in nature. Generally the cognition of such imagery involves exchanges between encoded properties and associative memory as described in Kosslyn's (1994) protomodel of visual perception. For this reason such imagery is delimited by one's visual experiences and mediated by the cognition of similarity. (Sloman 1999)

Level three concerns mental imagery created by a low level, primary process style of association. (Dailey 1995) It is imagery that is associated in a fuzzy way with general auditory features during passive listening. Associative imagery as described here is that utilised in the Bonny Method of Guided Imagery. (Goldberg 1995) Level three imagery is usually derived from the overall gestalt of a sonic passage. No effort is made to isolate individual sounds or to determine their sources. The associated question in this stage is: "What does that sound remind me of?"

2.3.8 Level Four - Causal Attribution

This final level describes an analytical approach to attributing a sound source of initially unknown origin to an object / event. As opposed to the previous stage - which is passive, involuntary and not consciously directed - this stage is active, conscious, constructive and analytic. It depends as much on memory as the associative imagery described above, however the way in which memory is accessed is more directed. In this stage each individual sound source is addressed in series, and the question asked: "What physical object could be making that sound?" The overall nature of the scene is then constructed from the composite of each individual object.

2.3.9 Resolving the Four Levels.

It may be apparent at this point that each level of the four part system would suggest different types of imagery. In the authors work different levels of the system are given priority in each scene. In some scenes - such as the green room and the last cloudfall scene in *Heisenberg* - associative imagery is the dominant guide to visual design. In other scenes the causal approach is taken. Generally once this decision has been made the detail of a scene is developed using the spatial and figurative levels. Imagery suggested by these two low levels is only implemented where it doesn't conflict with any imagery pre-determined by the causal or associative levels.

2.4 Problems Related to Effective Design.

2.4.1 Limiting the Scope.

Design aesthetics will only be discussed briefly in reference to aspects of the authors own approach to design where it pertains to the sort of works and problems being discussed here.

2.4.2 Designing Sound Spatialisation

When animating sound sources, dynamic and spectral qualities are dramatically modified by the proximity of a sound source to the camera/microphone position. The application of compression, dynamic loudness attenuation, or dynamic spectral filtering may compromise the illusion of moving sound sources. The relative loudness of sounds in a piece and certain rhythmic effects are dependent on sound source proximity.

Doppler pitch shifts create a situation where a sound's pitch content is bent upward while approaching the camera and bent downward whilst travelling away. The idea that this might be useful in any melodic sense is defeated by the fact that a sound can only move towards the camera for a certain amount of time before it collides with it. Similarly a sound can only travel away from the camera for a certain amount of time until it is inaudible.

2.4.3 Timbral Composition

In the development of sonic material for *Heisenberg*, primary sonic constituents were initially chosen on the basis of desirable timbral properties. In combining sounds the author's concerns are divided between various approaches. In a more analytic approach, consideration is given to the spectro-morphology of initially selected sounds. (Smalley 1986, 1997)

An advantage of the elemental system of categorisation (Stewart 1987) is that it can be applied equally well to both sonic and visual objects. Throughout the authors animated works there are literal implementations of this system. Fire, water, clouds and landscapes are all visual representational components in various pieces. Also, elemental systems have the rare attribute of near universality in category systems across all human cultures.

With elemental and spectro-morphological considerations in mind, sounds of differing natures are combined to produce balance, juxtaposition and other structural features where desired. By necessity this must be done with a mind to spatial trajectories and visual composition. In *Heisenberg* at least, fractal noise and other random procedures are then used to make decisions and generate detail at the microcosmic level.

2.4.4 Visual Concerns

A primary visual concern for the author is the composition of objects within the frame of the screen. In most scenes in *Heisenberg* for example, attempts are made to ensure that there is a balanced distribution of visible objects within the screen area. At no time for instance is there nothing to be seen on-screen, and in most cases at least one instance of the majority of auditory objects can be seen. Ideally the composition of objects will be balanced at all times.

The choice of hue for a scene is generally developed using the elemental or associative cross-modal systems described above. In *Heisenberg*, the use of high levels of saturation and the tendency towards chromatic homogeneity within a section are based purely on aesthetic choices.

3 Complex Problems

3.1 Overview

With the concerns and problems native to each of the four separate design criteria established, it is now possible to consider conflicts arising from certain combinatorial permutations of these criteria.

3.2 Occlusion

3.2.1 Occlusion

The most significant complex problem for the author in the creation of works such as *Heisenberg* is that of visual occlusion. In visual perception, when opaque objects are superimposed, the foreground object always occludes those behind it. The closer an object is to the viewer, the larger is it's perceived size and its capacity to conceal objects behind it. In a scene in which objects are moving and changing size and shape, occlusion problems arise constantly. Some examples are discussed below.

3.2.2 Occlusion and Spatial Arrangement

Occlusion sets up limitations on the way in which objects can be arranged spatially. The position and size of an object must be taken into account when planning the animation of a scene. When combined with a concern for cross-modal mapping and physiognomic perceptual styles, a tendency arises for scenes to be full of layers of objects at increasing depth. In *Schwarzchild*, *Loucid* and *Heisenberg* large dull objects – which resemble wall planes, ground planes or entire rooms – are used to represent low frequency sounds. As sounds become higher in frequency they tend to become smaller and move into the foreground. Objects within the room must be animated spatially in such a way that they do not pass through the geometry associated with the low frequency sound object.

3.2.3 Occlusion and Audio-Visual Proximity

Occlusion problems effect the proximity of sonic objects to the camera during spatial animation. In the visual domain, if an object is too close to the camera it can not only conceal the rest of the scene, it can also intersect or overlap the position of the camera. In 3D animation, a geometry-camera intersection will create unwanted artifacts. In the visual domain proximity also creates loudness levels which may distort audio signals and even damage audio amplification equipment.

To avoid such effects of proximity it is necessary to hand animate objects carefully so that no camera collisions occur. In systems that are animated using algorithmic techniques, this can be achieved by converting cartesian position coordinates to polar coordinates and setting a minimum distance limit between objects and the camera location. To avoid sudden collision type motion when the minimum distance is achieved, a gaussian envelope filter can be applied to distance data to smooth out sudden changes in motion.

3.3 Other Criteria Conflicts

3.3.1 Aural Immersion - Field of View Conflicts

One conflict between the four design criteria in Table 1.1 results from the aforementioned desire to frame within a screen of limited size, an object associated with every sound audible within that scene. This of course creates a conflict of interest when one wishes to immerse the audience in sound and yet have the sound sources visible at the same time.

In *Heisenberg,* this is solved by arranging multiple instances of certain objects around the camera. In this way the audience is immersed in sound, and at least one instance of each sound object is visible whichever way the camera points. The audience should even be able to infer the appearance of invisible instances from one visible instance of that object type. In *Heisenberg* objects which exist in only one instance tend to be spatialised so that for at least fifty percent of the time they are clearly visible in the field of vision. Instances are usually different waveforms of a similar type, or a de-correlated and delayed version of the original instance. (Kendall 1995)

3.4 Feedback Loops in Audio-Visual Design

3.4.1 Overview

Problems like occlusion, which are due to simple conflicts of interest, can usually be solved with a little compromise. Difficulties arise when the manipulation of an attribute in either the visual or aural domain adversely effects related attributes in the other domain. Solutions for problems involving multiple conflicts of interest compounded by a potential for feedback between spatial, auditory and visual attributes of a scene can be a little trickier to conceptualise solutions for. Their solution usually involves tweaking spatial audio-visual relationships until the problematic system settles down and becomes acceptable in both audio and visual design domains.

3.4.2 Integrated Spatial Audio Visual Composition

When spatial simultaneity and cross modal exchanges are a creative concern, feedback between sonic, visual and spatial elements has the potential to enter infinite feedback loops. When adding a new sound to a sonic passage, the new sound may suggest a certain object, and while the new sound may blend nicely in amongst pre-existing audio material the new object may not. Finding a spatial location for that object where occlusion isn’t a problem will then becomes an issue. If the object is moving, altering it’s position will effect it’s doppler shifts and other dynamic audio information. This in turn may effect the way that sound would look in the first place. Changing the object’s look may fix this, so long as animation and other data doesn’t need to be changed with the new visual features. Still, with the object added and altered thus, the overall sonic dynamic of the entire passage may have altered. Other objects in the scene may need to be altered and added to represent these changes faithfully. The entire visual strata of the scene may in fact need to be re-designed. Besides any re-animation this may require, this may also in turn effect the perception of the new object relative to the new nature of the scene. The altered object type may not now be appropriate.

This kind of thing can go on indefinitely and the mental pre-visualisation of a solution which will permit the system to work on all levels will save a lot of time in trial and error. In the authors experience pre-visualised scenes are also generally the most satisfactory visually anyway.

3.4.3 Low Frequency and Reverberation

As described in section 3.2.2, low frequency sounds are often mapped into ground planes, wall planes or entire rooms. The animation of low frequency objects therefore has the potential to feed back into the reverberation in the scene. In turn, reverberation has the capacity to feedback into general visual qualities within the scene – especially if wall reflection doppler shifting is enabled.

4 Conclusion – Problems in Music Education

The initial problem for many attempting to develop interdisciplinary works such as those described here might be the lack of a conceptual starting point. This may be due in part to the traditional absence of discussion about visual artifacts involved in music cognition in music scholarship and education. While the visual arts have a considerable tradition regarding the depiction of auditory phenomena visually, (Kandinsky 1912) visual-auditory relationships are not an integral part of traditional music scholarship.

This may be related to the fact that mental imagery is experienced less often by trained musicians. (Huron 1999) This in turn has been suggested to be due to a shift in music cognition towards linguistic areas of the brain as linguistic abstractions of music are assimilated during traditional music education. (Crowder and Pitt 1992) Bio-musicologists have suggested that the emphasis in western music on reading notated music from left to right as one would read a spoken language, has had a major influence on the cognition of music and the evolution of the art as a whole. (Wallin 1991) This is perhaps reflected in the popularity of linguistic and grammatical models of music, such as the generative theory of tonal music (GTTM) of Lerdahl and Jackendoff. (1985) As Sir Thomas Beecham once quipped, “A musicologist is a man who can read music but can't hear it.”

While the system of Common Music Notation (CMT) underlying western music scholarship is still useful for musicians and composers working with traditional western music and instruments, it is not always applicable to the music of other cultures, or to many new types of electronic music.

For some time musicologists have considered the possible shortcomings of describing music purely in terms derived from CMT. (Nattiez 1978) (Padham 1996) CMT and many models of music based on it are now increasingly regarded as being outmoded and irrelevant by modern sonic artists working with new technology and a broad timbral palette. (Wishart 1986) More recently it has been shown that cognitive musicology relying on CMT fails to meet basic adequacy criteria. (Leman 1999).

Many have argued that musical experience has ineffable qualities – that is qualia that cannot be expressed in any other way. (Raffman, 1993) There is in fact no a-priori reason why the analysis and discussion of music should be limited to linguistic models based on CMT, when it has been shown repeatedly that a significant degree of musical experience does not reduce well to CMT or related linguistic models.

Of the new music models developed in the 20th century, many make extensive use of visual analogies that do not involve CMT. (Mattis 1992), (Palombini 1992), (Smalley 1986, 1997) A visual approach to sonic art based on mental imagery is much more compatible with dominant dual coding theories of cognition, in which linguistic modes of cognition are complimented by mental imagery modes, and vice versa. (Paivio 1978) Music analysis based on such descriptive techniques might be described as *Musicography*. (Palombini 1992) (Lyons 1999)

The development of Musicography as a field within music research and education would do much to stimulate creativity in music culture. (Dailey 1994) While there can be no doubt that linguistic approaches to music discussion and education are valuable, there is increasingly no reason why they should not be complimented by analysis of the integral visual aspects of musical experience. Such a change to music scholarship would certainly do much to propagate musical concerns in future interdisciplinary artworks. The absence of such teaching will however perpetuate the developmental retardation of musicians seeking to work with new media and interdisciplinary art.

5 References

Baron-Cohen, S., and J. Harrison. eds. 1996. *Synesthesia: Classic and Contemporary Readings*. Oxford: Blackwells.

Battey Bret. 1998. *An investigation into the Relationship between Language, Gesture and Music*. http://staff.washington.edu/bbattey/Ideas/lang-gest-mus.html

Begault, Durand R. 1994. *3D Sound for Virtual Reality and Multimedia*. Cambridge, MA. Academic Press.

Brower, Candace. 1997. "Pathway, Blockage, and Containment in 'Density 21.5'," *Theory and Practice 22*. New York: Columbia University Press.

Davies, John Booth. 1978. *The Psychology of Music*. Stanford University Press.

Bragdon, Claude. 1978, (c1939). *The Beautiful Necessity - Architecture as frozen Music.* Wheaton, Ill.: Theosophical Pub. House.

Bregman, Albert S. 1990. *Auditory Scene Analysis: The Perceptual Organisation of Sound*. Cambridge, Mass.: Bradford Books, MIT Press.

Cho, Young-soo. 2001 "The influence of music on individual mental imagery." *Korean Journal of Music Therapy*. Vol. 3, No. 1, pp. 31.49 http://www.ouimoi.com/mt/KAMT.htm

Cho, Young-soo. 2002. *A Bibliography of UMI Dissertations Dealing with Mental Imagery*. Referenced 3-5-2002. http://www.ouimoi.com/mt/UMI1.htm

Chowning, J. 1971. "The simulation of moving sound sources." *Journal of the Audio Engineering Society*. Vol 19, No.1.

Crowder, R.G; Pitt, M.A.; 1992 "Research on Memory/Imagery for Musical Timbre." In. Reisberg, D. (Ed.) *Auditory Imagery*. Hillsdale, New Jersey: Lawrence Erlbaum.

Dodge, Charles; Jerse, Thomas A. 1997 *Computer music : synthesis, composition, and performance. 2nd ed*. New York : Schirmer Books ; London : Prentice Hall International.

Cytowic, Richard E.1989. *Synesthesia: a Union of the Senses*, New York: Springer Verlag.

Dailey, Audrey, R. 1995. *Creativity, Primary Process Thinking, Synesthesia, and Physiognomic Perception*. Unpublished doctoral dissertation. University of Maine.

Goldberg, Frances Smith. 1995. "The Bonny Method of Guided Imagery." In, Wigram, T; Saperston, B; West, R. (Eds.) *The Art and Science of Music Therapy*. Harwood Academic Publishers.

Huron D. 1999 "Music838 Exam Questions and Answers." Ohio State University. http://www.music-cog.ohio-state.edu/Music838/exam_questions_answers.html#Cognition

Kandinsky, W. 1982 (1912) "On the Spiritual in Art" in *Kandinsky: Complete Writings on Art*, edited and

translated by K.C. Lindsay and P. Vergo. London: Faber and Faber.
Kendall, Gary S. 1995 “The Decorrelation of Audio Signals and Its Impact on Spatial Imagery.” *Computer Music Journal*. 19(4) P.71
Kosslyn, S.M. 1994. *Image and Brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
Lem, A. 1998 “EEG reveals potential connections between selected categories of imagery and the psycho-acoustic profile of music.” *The Australian Journal of Music Therapy*. Volume 9. P.3.
Leman, Marc. 1999. “Adequacy Criteria for Models of Musical Cognition.” In J.N. Tabor (Ed.*) Navigating New Musical Horizons*. Westport CT: Greenwood Publishing Company.
Lerdahl, Fred and Jackendoff, Ray. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
Lewis JW; Beauchamp MS; DeYoe EA. 2000. "A comparison of visual and auditory motion processing in human cerebral cortex." *Cerebral Cortex* Sep;10(9):873-88
Lyons, Andrew D. 1999. *A Course in Applied Musicography*.Unpublished report. www.vislab.usyd.edu.au/user/alyons/Musicography.htm l
Lyons, Andrew D. 2000. *Gestalt Approaches to the Gesamtkunstwerk*. Unpublished paper. www.vislab.usyd.edu.au/user/alyons/gestalt.htm
Lyons. Andrew D. 2001. “Synaesthesia: A Cognitive Model of Cross Modal Association." *Consciousness, Literature and the Arts*. Spring 2001. www.vislab.usyd.edu.au/user/alyons/synaesthesia.htm
Marks, L. E. 1978. *The Unity of the Senses: Interrelations among the Modalities*. New York: Academic Press.
Mattis, Olivia. 1992. *Edgard Varese and The Visual Arts*. Unpublished Dissertation. Stanford University.
Moore, Richard. F. 1990. *Elements of computer music*. Englewood Cliffs, N.J : Prentice Hall.
Nakamura, S., et al. 1999 Analysis of music-brain interaction with simultaneous measurement of regional cerebral blood flow and electroencephalogram beta rhythm in human subjects. *Neuroscience Letters*, 275 (3), 222 –226.
Nattiez, Jean-Jacques. 1990. *Music and Discourse: Toward a Semiology of Music* Translated from French by Carolyn Abbate. Princeton University Press.
Ong, Tze-Boon. 1994. *Music as a generative process in Architectural form and space composition*. Unpublished Doctoral Dissertation. Rice university: Houston, Texas.
Priest, Stephen. 1998. *Mearleau Ponty*. London: Routledge.
Raffman, Diana. (1993) *Language Music and Mind*. Cambridge, M.A.: MIT Press. A Bradford Book.
Reisberg, D. (Ed.) 1992. *Auditory Imagery*. Hillsdale, New Jersey: Lawrence Erlbaum.
Richardson, J.T.E. 1999. *Imagery*. East Sussex, UK: Psychology Press Ltd.
Padham, Max. 1996 *Adorno Modernism & Mass Culture – essays on critical theory and music*. London: Kahn and Averill.
Paivio, A. 1978. “The relationship between verbal and perceptual codes.” In E.C. Carterette and M.P. Friedman (Ed.) *Handbook of Perception: Vol VIII*.New York: Academic Press.
Palombini, Carlos. 1993. *Pierre Schaeffer’s Typo Morphology of Sonic Objects*. University of Durham. Unpublished Doctoral Dissertation.
Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S. J., Lechevalier, B., & Eustache, F. 1997. The structural components of music perception - A functional anatomical study. Brain, 120, 229-243.
Quittner, A.L. 1980. *The facilitative effects of music on mental imagery:* A multiple measures approach. Unpublished masters thesis. Florida State University, Tallahassee, Florida.
Roads, Curtis. et al. (eds.) 1997. *Musical signal processing*. Exton, PA: Swets & Zeitlinger.
Sloman, Steven A. and Rips, Lance J. (eds.) *Similarity and symbols in human thinking*. Cambridge, Mass. : MIT, Press, 1999.
Smalley, Denis. 1986. “Spectro-morphology and Structuring Processes.” In S.Emmerson, ed., *The Language of Electroacoustic Music*. London: Macmillan.pp.61-93.
Smalley, Denis. 1997. "Spectromorphology : Explaining Sound-Shapes." *Organised Sound*, 2(2). Cambridge University Press.
Stewart, R.J. 1987. *Music and the elemental psyche*. Wellingborough: Aquarian Press.
Summer, Lisa. 1985. “Imagery and Music.*” Journal of Mental Imagery*. 9(4). New York: Brandon House.
Treib, Marc. 1996. *Space Calculated in Seconds*. Princeton, N.J. : Princeton University Press.
Wertheimer, Max. 1938 "Laws of Organization in Perceptual Forms". in Ellis, W. *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.
Wallin, Nils Lennart. 1991. *Biomusicology: Neurophysiological, Neuropsychological and evolutionary perspectives on the origins and purposes of music*.Stuyvesant, NY: Pendragon Press.
Welch, Robert R and Warren, David H. (1980) “Immediate Perceptual Response to Intersensory Discrepancy.” *Psychological Bulletin*. 8(3) 638-667.
Wishart, Trevor. 1985 *On Sonic Art*. York: Imagineering Press.
Yi, Dae-Am. 1991 *Musical Analogy in Gothic and Renaissance Architecture*. Unpublished Doctoral Dissertation. University of Sydney: Sydney, Australia.

A Survey of Software Designs from the Sonic Art Group

Peter McIlwain

School of Music - Conservatorium, Monash University
email: Peter.McIlwain@arts.monash.edu.au

Abstract

A survey is presented of a range of approaches to software design that have been developed in the Sonic Art Group, an electro-acoustic music ensemble based in the School of Music - Conservatorium at Monash University. The software is focused on real-time application and therefore interface design is an important element in the discussion of this work. This will be discussed in the context of the Max/MSP programming environment, which was used to create the software. Techniques covered in the survey include: methods for real-time additive and subtractive synthesis, spatialisation (with an 8 channel sound system), and automated musique concrète techniques.

1 Introduction

The Sonic Art Group is an ensemble that is based in the School of Music – Conservatorium at Monash University. One of its functions is to provide a research base for the development of software that can be used for the creation of electro-acoustic and computer music. The ensemble mainly works with real-time sound generation using the *Max/MSP* programming environment (which is especially suited to these applications).

A number of software applications have been developed during 2001 and 2002. These include; *Matrixsynth* – a software synthesizer, *Additive Gendèr* – an additive synthesis instrument, *Fractured Delay* – a gate based delay application, and *Cubist Audio* – automated musique concrète software that incorporates *haas* panning. Screen shots of the software interfaces are included in the appendix.

The function and operation of these applications will be described and discussed. As all of these applications are intended for real-time use, efficient and useable interface design is critical. Comments will therefore be made regarding the various user interface designs that have been adopted.

2 Matrixsynth

Matrixsynth is a software synthesiser based on the synthesis architecture of the *VCS3* synthesiser (Hilton, 1998). Like other synthesisers of its time, the VCS3 incorporates a modular design in which almost all possible signal paths between its various synthesis components can be configured. The VCS3 is unique in that module configuration is done using a pin matrix system whereas other systems employ lead and socket connections. Many software synthesis environments, such as: *Max/MSP, Pd, Reason, Audio Mulch and Turbosynth,* have used the latter paradigm in the design of the user interface. This design paradigm affords a high level of flexibility in that a large number of modules and configurations are possible, however it is difficult to implement and, for the user, somewhat time consuming to configure.

Matrixsynth on the other hand, offers a limited range of synthesis modules, but its advantage lies in the implementation of the *matrix* object (a standard Max/MSP object) that offers a succinct overview of the configuration of the synthesiser (which is not possible with patch cords) and it enables rapid, real-time patch configuration. Figure 1 below shows a section of the matrix interface.

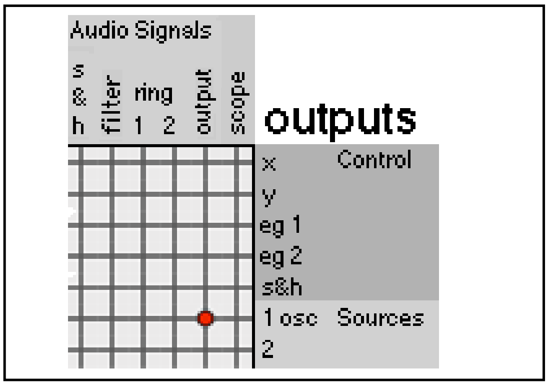


Figure 1. Part of the matrix control interface. Connections are made by clicking on the intersection between inputs, arranged on the horizontal axis and outputs, arranged on the vertical axis. In the example above the output of oscillator 1 is sent to the main output.

2.1 Matrixsynth Components

Matrixsynth comes in two versions Matrixsynth-1 and Matrixsynth-2 (the second version is a simplified design for training applications). The components of Matrixsynth-1 include:

Sound Generators

- Oscillators (with selectable waveforms, sine, triangle and square waveforms)
- Noise Generators
- Audio file and external analogue input signals

Modifiers

- Filter (with selectable highpass, lowpass, bandpass and notch functions)
- Ring Modulator
- Stereo Delay

Controllers

- Envelope Generators (with attack, decay, sustain and release stages)
- Sample and Hold
- X/Y control square (both outputs are scalable)

The oscillator design allows for control of the pulsewidth of the three waveform types. Thus a sawtooth wave can be obtained from extreme settings of the pulsewidth parameter when the triangle wave is selected. Other significant features include a preset facility and selectable triggering sources for the envelope generators. These are: spacebar, tab key, cycle (where the envelope generator operates as an oscillator), mouse, and MIDI (yet to be implemented).

Any output signal can be used as a modulation signal. A wide range of parameters can be modulated in Matrixsynth. These are:

- Oscillator frequency
- Oscillator amplitude
- Oscillator pulsewidth
- Noise amplitude
- Audio file and external analogue input signal amplitude
- Filter cutoff frequency
- Delay time

2.2 Evaluation

While the synthesis modules and control sources are relatively simple, the design affords a high level of predictable control. These design characteristics, together with the flexible patching that the matrix interface offers, enables the creation of very mobile sonic textures. This makes Matrixsynth especially suited to improvisation, a process that is a prominent feature of the music created in the Sonic Art Group.

In comparison with the original VCS3, Matrixsynth offers a greater range of possibilities and a greater level of predictability in the sonic result. One disadvantage that Matrixsynth has, in its current design, is that it only allows for the control of a maximum of two simultaneous parameters with the mouse via the x/y control space. In the VCS3 three parameters can be controlled at once (one hand controlling two parameters with the joystick and the other controlling the one of the various knobs and dials on the synthesizer). This can be addressed by the use of MIDI slider controllers however it has to said that the knobs and dials on the original instrument are more satisfying to use.

3 Additive Gendèr

The *gendèr* is a metallophone found in the Javanese gamelan orchestra[1]. It is pronounced "gndair". The Additive Gendèr is a software instrument - its name referring to the synthesis method used and the characteristic sounds that it produces.

The software uses an additive synthesis approach that utilizes an oscillator bank consisting of 256 oscillators. The amplitude of each oscillator is controlled by the standard Max/MSP *multislider* object (configured with 256 sliders). The frequency of each oscillator is allocated according to the harmonic series based on 20 Hz.

The multislider object functions in three ways: 1) as a trigger that causes a specific oscillator to sound, 2) as a table in which the amplitude values of all the oscillators are stored, and 3) as a visual representation of the power spectrum of the synthesized sound. The multislider object is illustrated in Figure 2 below.

[1] For more information about this instrument see Gend r (Kartomi and Ornstein, 1984).

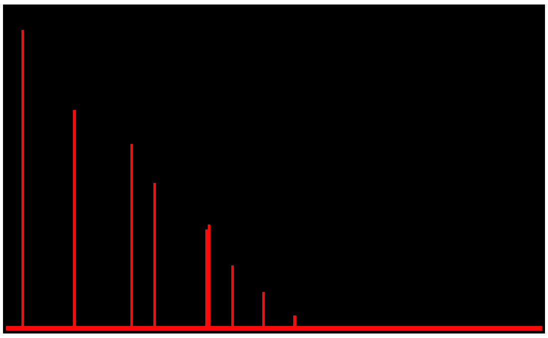


Figure 2. An illustration of the multislider object showing bars (which are independent slider controls). Each bar represents the amplitude of the corresponding oscillator in the oscillator bank.

3.1 Oscillator Triggering and Decay

Each oscillator is triggered when the corresponding slider is set in the multislider object. This is done by clicking in the multislider edit space. Clicking and dragging across the multislider will set and trigger multiple oscillators. The amplitude value that is set is then continuously altered by multiplication with a decay/growth factor. So that the oscillator bank is kept up to date with the current amplitude values, all the values in the multislider object are sent to the oscillator bank at 100 ms intervals.

If the decay/growth factor is below 1 the function will progressively decay the amplitude values in the multislider and oscillator bank. Alternatively values above 1 will cause the amplitude values to rise. The rate of decay or growth is determined by the distance from 1 of the decay/growth factor. Boundaries are set on the function so that no amplitude value can be greater than 1 or less than 0.

The function also includes a randomising feature that alters the decay/growth value. With no randomisation set, the decay curve will be a smooth exponential shape. A high randomising value causes a staggered or erratic decay/growth curve. The randomisation occurs separately for each of the 256 oscillator envelopes. When multiple oscillators are triggered the randomisation results in a shimmering or ringing effect reminiscent of the timbre of a metallophone.

3.2 Stereo Treatment

A stereo effect is created by the use of two oscillator banks (consisting of 256 oscillators each) that output to left and right channels. As mentioned above, the frequency of each oscillator is set according to the harmonic series however these values are slightly randomized according to a user controlled value. This causes a complex detuning effect between the two oscillator banks and accentuates the stereo space.

Another method of spatial articulation used is the implementation of a multi-tap stereo delay line. An interesting feature of the delay line design is the incorporation of a ring modulator that is positioned after the output stage of the delay line and before the feedback function. A simplified signal flow diagram of this design is shown below in Figure 3.

The effect of a ring modulator inserted in a delay line, as shown in Figure 3, is easiest to describe if we look at a simple scenario in which the input signal is an impulse consisting of a sinusoidal wave at 400Hz and the oscillator is producing a sinusoidal wave at 300Hz. Here the component frequencies in the output waveform will be as follows:

1st pass) 700 (sum), 100 (difference)
2nd pass) 1000, 400, -200
3rd pass) 1300, 700, 100, -500
4th pass) 1600, 1000, 400, -200, -800,
5th pass) 1900 1300 700 100 -500 –1100
6th pass) 2200 1600 1000 400 -200 -800 -1400

Here the terms *first pass, second pass* etc. refer to the number of times the same signal has passed through the delay as a result of the feedback function. As can be seen in the frequency list above, the ring modulator produces two effects, 1) alternating frequencies, and 2) an upwards cascade of additional higher frequencies.

The implementation of the ring modulator in Additive Gendèr is more complex than is described above. For each of the two stereo processing chains there are three tapped delay lines. This produces more complex patterns of modulation; however the upwards cascade effect is still present. An additional complication is that the frequency of the oscillator is slowly shifted in response to a random number generator. This produces a rich metallic sound.

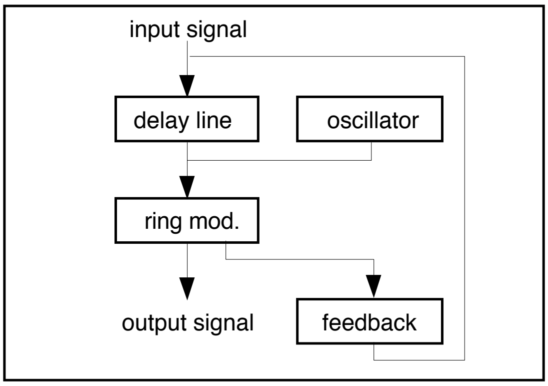


Figure 3. Delay line with a ring modulator and feedback function.

3.3 Additive Gendèr Performance

When the software is set so that amplitudes will decay after triggering, the sounds that are produced are similar to that of a metallophone. The initial amplitude is determined by the level set in the multislider so, in the case of decaying envelopes, triggering in the higher regions works well. For a growth envelope setting, triggering works best when it occurs in the lower regions. This produces a kind of bowed sound.

If single oscillators are triggered in succession they will be heard as pitches. Multiple oscillators triggered at once causes a singular sound with a complex harmonic structure.

As the triggering occurs in response to mouse clicks, the rate of successive triggering is slow. This is not necessarily a problem as it can be seen as a characteristic of the instrument, just as it is not possible to quickly trigger a succession of gongs. Music performed on gongs is by necessity slow, however this characteristic works well because there is time between gong strikes to appreciate the rich and complex evolution of the sound. The same can be said for the Additive Gendèr however it may be worth exploring other forms of interface that could allow for faster triggering rates. A MIDI keyboard has been tried as an alternative triggering source. This of limited use however as the MIDI specification has a 128 as the maximum number of pitches (or frequency units) whereas the Additive Gendèr requires twice that number. One interesting possibility here might be the use of a proximity sensor to set frequency and a pressure pad to set amplitude.

3.4 Performance Gestures

The most interesting, and arguably successful, aspect of the Additive Gendèr software is the integration of control interface with sonic display via the multislider object. Here the object that triggers the sound, also to displays the current state of the sound (as a power spectrum). This makes performance gestures feel natural because the control interface gives feedback to the performer about the current state of the sound generator. The same kind of process occurs when one is performing on a physical, acoustic instrument.

This approach has a secondary benefit in that the multislider object can be displayed via video projection so that an audience can see the performance gestures and possibly divine the sound production process.

4 Fractured Delay

Fractured Delay is a software program that distributes a monophonic signal to eight separate delay lines. The output of these delays are then panned across a stereo field or, in a alternative design, sent to 8 discreet outputs for use in multi-speaker arrays. There are two versions of Fractured Delay; *Fractured Delay 2* for stereo output and *Fractured Delay 8* for eight-channel output.

Each of the eight delay lines has an input gate that is designed to open and close at varying times, and at varying intervals. This causes successive fragments of the input sound to be distributed to the eight delay lines.

The on and off interval values for each gate (called *duration* and *rest* in the program) are generated by two random number generator modules that incorporate the *Gauss* object by Peter Castine (Castine, 1997). Output from gauss is distributed in a bell curve around a center value that is determined by the user. The width of the distribution can also be set by the user.

Fractured Delay is comprised of several components organized in the signal path as shown in Figure 4.

4.1 Input, Filter and Gates

The input signal first passes through an FFT based filter that allows for graphic equaliser style filtering of 253 frequency bands. It then passes to the eight input gates. At the start command for the program, the gates are successively turned on at random intervals ranging from 0 to 6 sec. They are also assigned duration and rest values. Upon completion of the rest stage (in each delay module) a message is sent from the gate to the two random number generator modules. This initiates the generation of new duration and rest values and the gate is opened again. The gates will cycle continuously through the duration and rest stages until the program is stopped.

An x/y control surface is used to set the centre point for the generation of the duration and rest values. The range of possible values is from 14ms to 22 sec. The use of an X/Y control interface enables rapid control of these two values. This is desirable as the relationship between duration and rest times greatly influences the texture that is generated.

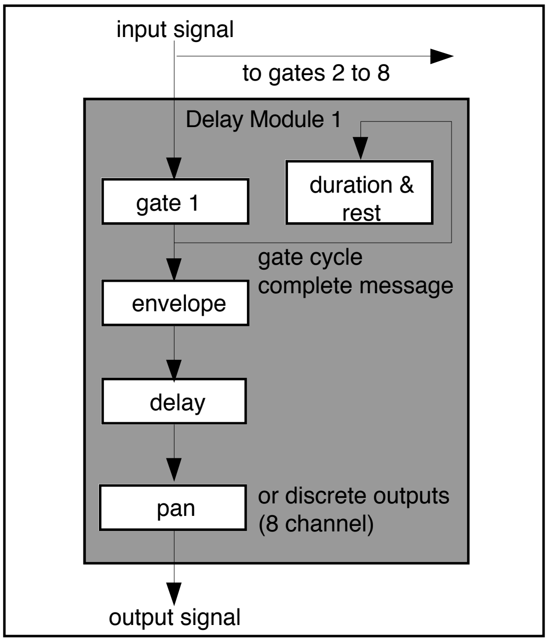


Figure 4. The signal path of the various components of Fractured Delay.

4.2 Envelope

An envelope generator is included in each delay module. The function of the envelope generator is to alter the amplitude envelope of the input signal. When a gate is opened, a message is sent to trigger the start of the envelope generator. The total duration of the envelope is equal to the duration time (the on time of the gate). Control of the envelope shape is done using the standard Max/MSP *function* object that allows for the creation of multiple breakpoint envelopes.

4.3 Delay

Each delay line has a maximum delay of 6 sec. The design of the delay line incorporates a feedback function and delay time modulation using a sine wave oscillator. These functions are controlled by four parameters; 1) delay time, 2) feedback, 2) modulation frequency, and 4) modulation depth.

As there are 8 separate delay lines, there is a total of 32 parameters that must be set. An interface is provided for this using simple number box entry. Due to the large number of parameters this interface is not practical for real-time control. Therefore an alternative interface has also been provided that sets the four parameters for all eight delay lines from four main number controls (one for each parameter). This approach utilises a random number generator that distributes varied values around the value set in the main control. The width of the distribution can also be set. This enables rapid control over all 32 parameters, which is very useful in live performance situations. The interface is illustrated in Figure 5 below.

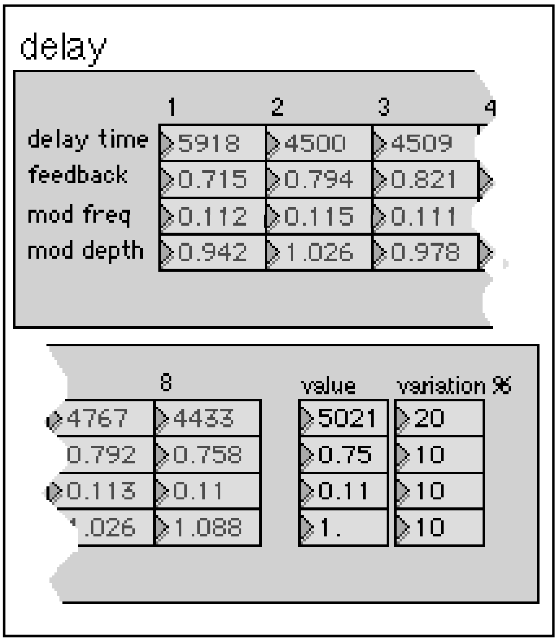


Figure 5. The control interface for the eight delay lines. The value number boxes set the centre value for the distribution of values to the corresponding parameter for each delay line. The variation number boxes set the width of the distribution.

4.4 Output

In the stereo version, Fractured Delay 2, the output of the delay lines are panned to left and right outputs. In the 8 channel version, Fractured Delay 8, the output of each delay is sent to one of the 8 output channels. Due to the action of the input gates, as described above, the input sound is distributed across the virtual stereo space or through an 8 channel speaker array. In order to ensure an even spatial distribution, the generation of duration and rest values is done using the random number module (as discussed above).

The sonic textures generated by Fractured Delay are various however they can be characterised as sounding similar to those created by delay loops. The difference however is that the textures generated by Fractured Delay are more complex and less repetitious. When very short intervals, in the 40 to 100 ms range are used for the duration and rest parameters, the process is similar to granular synthesis.

5 CubistAudio

One of the most significant techniques developed in the *Cubist* painting movement (as practiced by Picasso, Braques and Gris), was the presentation of multiple views of an object, or scene, in the one image (Hughes, 1980). This technique is a manifestation of the idea that our concept of reality is as much a function of memory as it is of instantaneous perception. For example our concept of the reality of an object is informed by the memory of multiple perspectives of that object and other similar objects. It is through our memories that we assemble and assimilate the multiple perspectives of our experience.

Unlike painting however, the cubist approach applied to sound, is much more difficult to achieve. This is because it is possible to visually represent simultaneous and discrete views of an object or scene whereas a sound is an object that must exist in time. Simultaneous and complete representations of the same sound are not possible. However parallel presentations of a sound object can be made and this approach is relatively common. The basic compositional technique of the *canon* is an example of this. Minimalist techniques can present multiple aspects of the sound or sound series. This is achieved by selectively presenting portions of a sound series against other portions, as is found in techniques such as phasing and additive repetition.

CubistAudio explores some aspects of the cubist technique by simultaneously presenting different sections of the same sound and rendering them in different locations in a virtual space. As there are several similarities between the techniques used in musique concrète and cubist painting, the program employs some of these techniques as well. These include: retrogradation, pitch shifting, the isolation and reordering of events in a sound or series of sound objects, and the alteration of the amplitude envelope. As the same sonic material is presented in various forms, and positioned in a virtual space, the musical textures that are created have continuity as a sonic landscape[2].

5.1 Segment Selection

The program includes eight modules that select and play segments from an audio file that is loaded into a buffer. This enables a maximum of eight segments to be played simultaneously. A segment is defined by the:

- Start position within the audio file,
- Duration (of the segment),
- Pitch, or transposition (this influences the actual duration),
- Orientation – whether the segment is played forwards or backwards.

The start position is generated in two processes:

1) A scan factor is applied. When the program is run, the start position of each segment generator is at the beginning of the audio file. When the next segment is generated the start position is advanced along the timeline of the audio file. The distance that the start position is advanced by is determined by the scan factor (a user controlled variable that is measured as a percentage of the total length of the audio file).

2) The position determined by the scan factor is randomized by the use of the *wprob10* object (see 5.3 below).

The maximum length of a segment is equal to the total length of the audio file. Segment durations are

2 For further discussion on the issue of sonic landscapes, see Spatialised Sound: the listener s perspective (Mcilwain, 2001) and On Sonic Art (Wishart, 1996)

generated stochastically (see section 5.3) with an upper limit that is set by the user and is measured as a percentage of the total length of the audio file. As used by the author, segment lengths tend to be in the range of 10 – 15% of audio files that are around 20 seconds long.

While the process of segmenting a sound object has similarities with that used in granular synthesis the effect is different due to the longer intervals involved. In granular approaches, segment durations tend to be around the threshold at which a listener can discern successive separate events, therefore combinations of grains aggregate into a continuous timbre. With longer segments the identity of the segment, as a sound object, is retained so that the aggregations tend to be heard as textures consisting of several streams.

5.2 Segment Processing and Output

After the segment is selected, it is played and altered according to several processes. Firstly, the amplitude envelope is altered according to a user-controlled interface (featuring the Max/MSP function object). Secondly the segments are processed for positioning in a virtual space using the *haaspan* object (see 5.4 below). Lastly the output from the program can either be recorded or sent to monitoring equipment only.

5.3 wprob10

Most of the parameters for segment selection and spatialisation are generated using the wprob10 object (which is created by the author). This object generates random values in 10 bands. Values within a band have an equal probability of being generated, however the bands are selected according to probability weights. Therefore the generation of values by wprob10 is a two step process:

1) a band is selected based on weightings assigned by the user,
2) a value within the range of the selected band is generated using a simple random number generator.

Figure 6. illustrates the user interface for wprob10 together with a graph illustrating the frequency distribution for all possible values that are generated from various settings of the weightings. The height of the bars (which can be altered by the user) determines the relative weightings for the various bands. The use of weights in the generation process described above enables a wide range of distribution patterns to be generated.

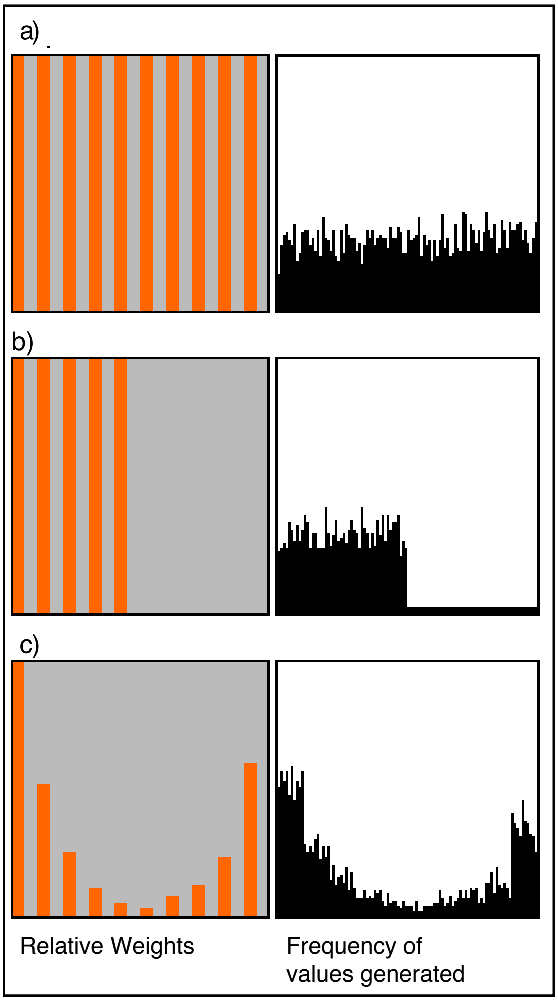


Figure 6. Various distributions patterns for wprob10. a) even distribution, b) limited distribution, and c) weighted distribution.

5.4 haaspan

Each of the eight segment generators uses an instance of the haaspan[3] object to position the playback of the segment in a stereo virtual space. The haaspan object is created by the author according to the design outlined in "Spatialisation of Sounds over Loudspeakers" (Moore, 1989). It creates a panning effect via interaural time difference and through changes in amplitude. There are two user controlled parameters, x and y, that determine the position of the

3 The name of this object is a reference to the *haas effect* after Haas who is credited with discovering the spatial effect described in this paper (Haas. 1951).

sound in a cartesian space. A third variable, *speaker width*, controls the distance that the loudspeakers are apart. Figure 7 shows the relationship of the virtual space created by haaspan to the position of the loudspeakers.

The object also includes a reverb send that is scaled according to distance. The scale adopted is inversely proportional so that the balance between direct and reverberated sound is altered in favor of reverberation the further away a sound is located from the speakers. Reverb is created using a *VST* plugin, the one currently used in CubistAudio is Freeverb (Jezar, 2000).

Apart from dynamic variations in the original audio file, the amplitude scaling inherent in the spatialisation process used in haaspan is primary way in which dynamic variation occurs.

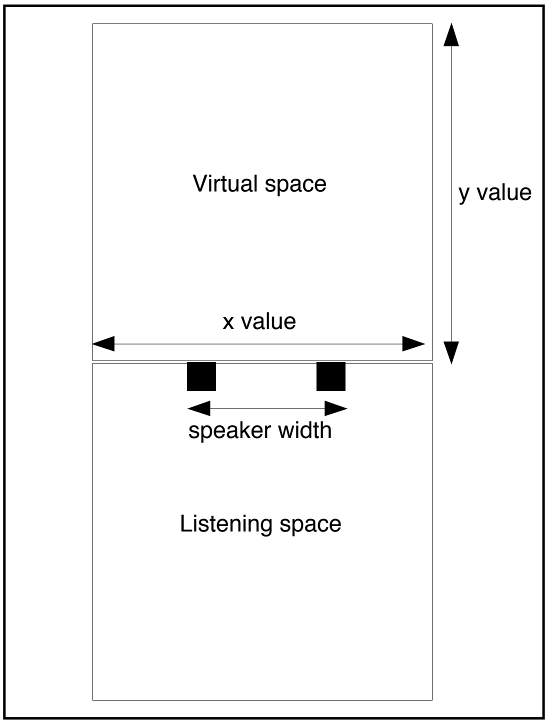


Figure 7. The relationship between the three user variables in haaspan and virtual space.

5.5 Landscapes

The textures that are generated by Cubist Audio have similarities with those that occur as a result of Minimalist techniques. Here textures are characterised by recurring gestures that are combined in a variety of ways by processes such as the addition, subtraction, re-ordering, and displacement of notes, and the scaling of temporal elements. The textures generated by the software however are significantly different in two respects; 1) metrical structures are not adhered to so that they have more in common with the *micropolyphonic* textures of Ligeti or Lutoslawski's *aleatoric counterpoint*, and 2) the elements in the texture are spatialised. The resultant textures from the process tend to create a unified sonic landscape inhabited by aggregations of like gestures (as they are all sourced from the same audio file). For example, if the source audio file is a recording of a dog barking, the resultant texture (if segments are in the 1 to 2 second range) will sound like a field of dogs distributed around the virtual space.

6 Acknowledgments

Thanks to Paul Doornbusch for knowledge and assistance in a number of ways and to the members of the Sonic Art Group; Matt Baxter, Tom Dunstan, Eden Krumins, Kim Lajoie, Andrew Ross, Clint Small and Simon.

References

Castine, P. 1997 "The Litter Package" *Peter Castine's Software* http:://www.bek.no/~pcastine/Music/programs.html Accessed 29/04/02

Haas, H. 1951. "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache" Acustica 1:49-58.

Hilton, G. *Electronic Studios Home Page*, http://www.ems-synthi.demon.co.uk/ Accessed 6/5/02.

Hughes, R. 1980 *The Shock of the New*. London, British Broadcasting Corporation. 20

Jezar, 2000. *Freeverb*, http://www.dreampoint.co.uk Accessed 6/5/02.

Kartomi, M.J., R. Ornstein. 1984. "Gendèr" In S. Sadie (Ed.) *The New Grove Dictionary of Musical Instruments*, London: Macmillan Press. Vol. 2: 35.

Mcilwain, P. A. 2001 "Spatialised Sound:the listener's perspective" *Proceedings of the Australasian Computer Music Conference 2001* ACMA Melbourne. 57-65

Moore, F. R. 1989. "Spatialization of Sounds over Loudspeakers" in M. V. Mathews & J. R. Pierce (Eds) *Current Directions in Computer Research,* Cambridge, Massachusetts: MIT Press. 89-103.

Wishart, T. 1996 *On Sonic Art*, Harwood, Amsterdam

Appendix: Screen shots of the software

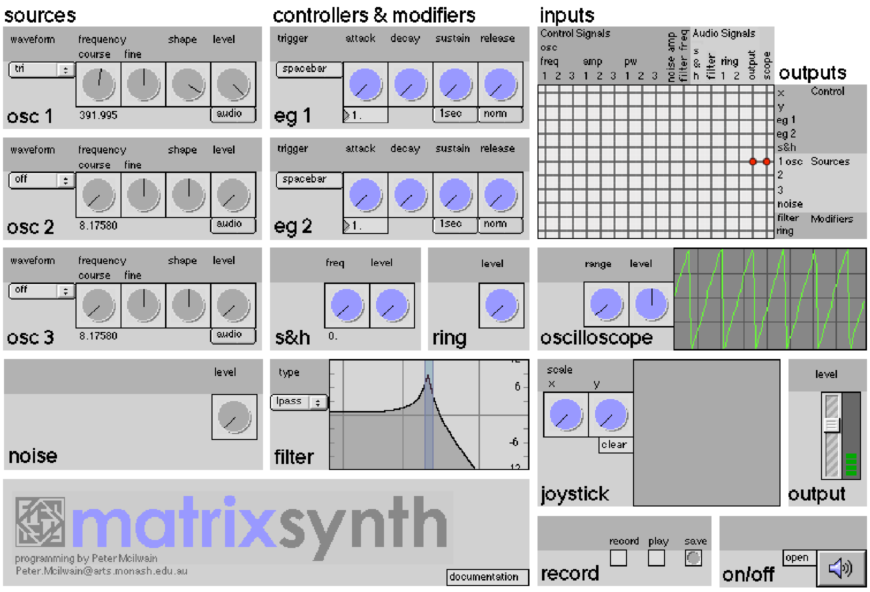


Figure 9. The interface for Matrixsynth-2.

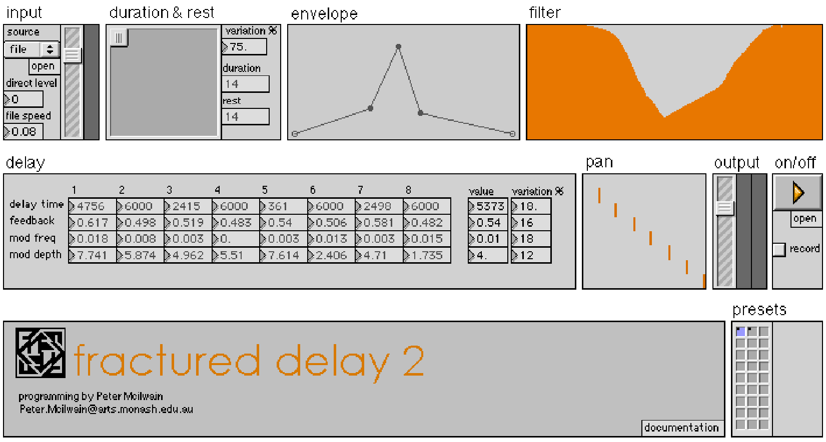


Figure 10. The interface for Fractured Delay 2.

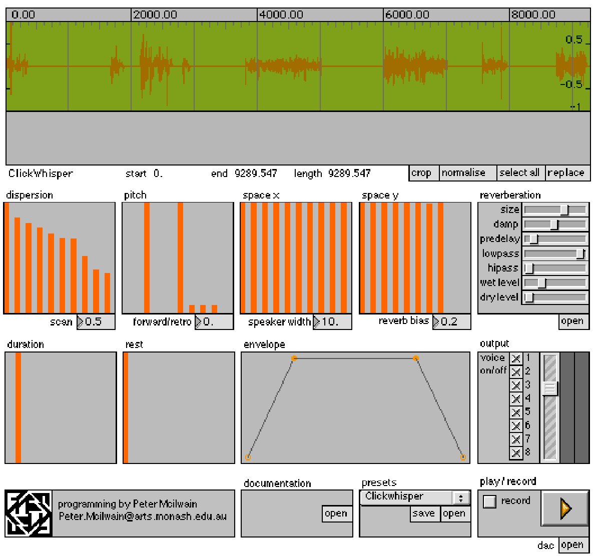


Figure 11. The interface for CubistAudio.

Correlating movement in space to the parameters of sound

Jonathan Mustard

Classical Music Department: Western Australian Academy of Performing Arts, Edith Cowan University
email jmustard@student.ecu.edu.au

Abstract

This paper addresses some of the issues associated with integrating sound and body movement in space and the mapping of movement to the various parameters of sound using computers and sensors as the mediating devices.
Historically composers have made the decision as to how particular movements or trajectories through space will affect sound. This has sometimes meant that the dancers' palette is arbitrarily limited by the composer's choices or visa versa – that a choreographer's imperatives impose arbitrary limitations on the sound/music.
The question this paper addresses is whether there are any general rules or assumptions that can be made about correlating aspects of sound to aspects of movement that are at once flexible and practical and maintain integrity in both disciplines.
The paper analyses a number of examples of interactive, sensing environments and poses some possible directions from the questions above. The directions presented are supported by descriptions of objects moving in virtual space and the sound that results using a variety of mapping procedures.

Introduction

With the increase in sophistication and reliability of motion sensing equipment, more attention can now be given to compositional and choreographic concerns over the technical concerns of the instruments (Winkler, 1995). Affordability of some of the technology is still some way off. I am thinking particularly of motion capture here. But there are cheaper technical means that offer a high degree of useful motion data. Of particular interest are systems that use video to capture motion in an environment. They are relatively inexpensive and integrate with well-known programs such as MAX/MSP. Eric Singer's "VideoIn" and associated objects for MAX, "Big Eye" and David Rokeby's "Very Nervous System" (VNS) are three such systems. All these systems measure the variation in light that is produced when objects are moving in space.

There is also a wide variety of sensing devices and systems that enable the capture of data from purely gestural movement as opposed to topographical positioning. MidiDancer, created by Mark Coniglio, captures the flexion of joints (wrists, elbows, hips and knees). The Miburi bodysuit, developed by Yamha similarly captures flexion in a number of joints and has pressure sensitive pads that attach to the soles of the feet and a keyboard-like device that fits on each hand that can send both on/off information and thumb controls that can send a range of voltages to be interpreted as pitch-bend or other MIDI controller type information. Infusion Systems makes ICube, an analogue to digital converter that accepts a variety of transducers and actuators like light, touch, flexion, thermal and proximity sensors, switches, L.E.D's and motors. SensorLab, created at STEIM is another analogue to MIDI interface for the prototyping of musical instruments and interactive control systems. And BodySynth, developed by Chris van Raalte uses electromyographic sensors to generate MIDI information from the contraction and relaxation of muscles.

There is also another category of gestural devices that were initially intended as MIDI controller instruments for musicians rather than for their potential as controllers for dancers, but may still have useful application in this area. Pat Downs' Airdrums, the GoldBrick Interface for the Nintendo PowerGlove, and well before MIDI there were Theramins, to name a few.

- ## Some examples of interactive works

2.1 "Dark Around the Edges" and "Songs for the Body Electric"

Two works by Todd Winkler,Walter Ferrero and Gerry Girouard, *Dark Around the Edges* and *Songs for the Body Electric* use Rokeby's Very Nervous System as the medium for interaction between dancers, lighting, video and music. In these works the video input and therefore the performance space is divided into a number of regions.

"Each region has the ability to trigger a particular MIDI note or a series of notes, report continuous changes in movement, reconfigure the software, and start or stop a musical process." (Winkler, 1998).

Winkler is able to reconfigure response behaviours in the system by using a number of presets embedded in the software Max. So a single movement gesture in the same position within the performance space can produce a variety of outcomes at different times, though in the works mentioned they are used for different styles of dancing. On the one hand for movements of a precise robotic nature where the musical response is machine-like percussive sounds and on the other characterised by a more free-flowing style of dance that produces continuously evolving sound.

"In several sections of the works, physical gestures have an obvious and immediate impact on the quality of timbre by mapping VNS values to control DSP functions, such as filters, low frequency oscillators, and distortion algorithms. Further processing under MIDI control is available from the Ensoniq DP/4 processor. Several techniques may be combined, such as in "Gazelles" from *Songs for the Body Electric*, where location triggers play chords at a maximum rate of two per second, while speed effects a low pass filter and pitch bend." (Winkler, 1998).

In these works as well as others by Winkler, highly complex and variable mapping procedures have been successfully deployed to create remarkable works that integrate sound, movement and visuals in a way that only a few years ago were impossible. The success of the works is not just due to the emerging technology, but very much because of the lengthy collaborative process that lies behind them. A lot of energy was spent in research and experimentation with different set-ups to determine the optimum for both composer and choreographer.

2.2 Your Sky is Filled with Billboards of the Sky

In Lindsay Vickery's *Your Sky is Filled with Billboards of the Sky* for dancer, video and computer generated sound, the mediating device for movement is the body suit hooked up to a Macintosh PowerBook and a Mac' G4. The PowerBook runs STEIM's *Image/ine* software while the G4 runs IRCAM/Cycling 74's MAX/MSP.

MAX/MSP takes note number, velocity, pitchbend, modulation and elapsed time between events from the MIBURI. Note numbers are mapped to samples, velocity to sample playback speed (pitch), controller 1 to panning, pitchbend to sample volume and if the time between events is greater than zero a loop of the previous few seconds is turned on[1].

Similar procedures are followed to control the video images that include a QuickTime video of clouds passing over the moon, a live video feed of the dancer, text and a still image of a graffitied wall.

For example, note numbers control the angle of displacement of the still image, velocity determines the 'x' coordinate of the still, pitchbend determines the 'y' coordinate, controller 1 alters the area of the image that is displayed, its hue, saturation, colour-shift and the level of "keyer" effect[2].

Various sensors allow the dancer to distort her image with 'PhotoShop'-like effects.

Your Sky is Filled with Billboards of the Sky was developed for the REV festival at the Brisbane Powerhouse in April 02 with Katherine Duhiggt performing. I was fortunate to see two performances during the Totally Huge New Music Festival in Perth this year with visiting artist Zhang Ping in the suit. Vickery states that Zhang Ping's dance, which is strongly influenced by Chinese classical, folk and modern dance, was particularly well suited to the piece. She structured her movements around expansion and contraction of the kinesphere (Laban & Ullman, 1966) with particular emphasis on the arms and torso as the prime articulators of transition from and expression within one extreme to the other.

[1] From conversations and information supplied by the composer.
[2] The amount of transparency applied to whiteness that allows a background image to show through.

A kind of slow, abstracted bird-like movement of the arms was a strong motif in these performances and was very effective in managing the panning of sounds and the coordinates of the image buffer.

So it was a fortunate collaboration that was not made easier by the language barrier. But even so Vickery laments the limited number of rehearsals that were available for the dancers to familiarise themselves with the controls.

2.3 MAP1

MAP1 by Garth Payne is an audience interactive installation that again uses VNS to collect movement data in the environment.

"This installation focuses on sound using the immersive, fluid and emotive qualities of the medium to generate a rich, enveloping and ever evolving environment. The sounds are made more fluid by the use of a system capable of moving the apparent source of the sound through the physical environment. This creates a dynamic relationship between the presence and position of a body and the position and movement trajectory of the sounds. The ability to move sound through space affords the sound a physicality and in so doing the sense that the sound becomes another physical character or presence within the installation.

Those within the installation will sense a physical interaction with the sound. A wide range of different aural qualities are mapped in qualitative groupings to different regions within the installation space, generating a plethora of aural textures and densities, chosen on the basis of the quality of movement of the body within that region of the exhibition" (Payne, 1998)

In the above examples, a number of paradigms are apparent in terms of performer control of the various media. In the first example, it is the extensive collaborative process that allows for the level of complexity exhibited. The choreographers at least if not all of the dancers spent a considerable time familiarising themselves with the technology and potential sonic responses.

Your Sky is Filled with Billboards of the Sky I think takes a pragmatic approach so that a successful outcome is achieved in a relatively short rehearsal period due to the compositional processes inherent in the piece and by limiting the parameters of control over the visual. Having said that though, the performer in this case still had a fearsome array of controls to think about while making a satisfactory choreography.

In the last example a more pedagogical approach is taken that necessarily involves a certain amount of trial and error. Participant interactions become more fluent over time with intent gradually matching outcome.

In many senses we reinvent the wheel each time we make a new movement-interactive piece, drawing up new maps that make various correspondences between (human) movement and sound. We look at movement and make decisions about how it might translate into sound. We make a sound and try to figure out a way to move in a sensitised space that might recreate that sound object. We look at a space and start chopping it up into little bits that can be managed by the technology. We ask performers or participants to take responsibility for completing the work in an area in which they have little expertise. In saying this I place no particular value or judgement on these decisions or methods, but it does raise the question of whether there might be some generalisations that we can make about correlating sound, space and movement and whether certain solutions might be more transportable from one situation to another. And not just physical situations, but social, technological, aesthetic and temporal.

One of the main objections to simple one-to-one relationships where for example a certain gesture will always result in the same sonic response is that it quickly becomes boring and predictable.

Virgil Thompson, composer of many works for dance, age 90 in an interview with Katherine Teck. "Music which illustrates the dance is just as boring as dance which illustrates the music." . . .

"Music walks down the street beside the dance". (Teck, 1989.p 43)

Norma Reynolds Dalby - "Furthermore she pointed out how very exciting it can be to find out 'what can happen in terms of enhancing both media if you do not follow the dancer precisely'". (Teck, 1989.p 53)

And Donald York - "But if the music and dance breathe independently and not exactly hand in glove, it is exciting." (Teck, 1989.p 71)

The music should never 'Mickey Mouse' the movement so we are told and yet we strive for an ever-greater intimacy between the two in the systems that we build. How to reconcile these seemingly conflicting ideas. My current research includes a search for a generalised map that is readily adaptable to a range of spacial situations and dance styles.

Firstly a paradigm that might be useful is one in which the dancer is free to go about making a structure as freely as possible. Free from the responsibility of having to look after a massive number of musical interrelations and controls. Free from the necessity for extensive collaboration. A paradigm in which the sound/music has both elements of unison and elements of parallelism to the dance.

Some hunches I am following include making broad generalisations about space, sound and movement. Then building a system where these generalisations can be linked to see which might yield the best results across the widest range of possibilities.

3 One Map?

Before we are able to correlat e movement to sound we must have some generalised concept of space and a generalised relationship to sound. Ignoring the social, cultural or political sphere that any space might inhabit for a moment, let us assume that it is a completely empty space. A space devoid of objects or meaning except that it has boundaries like a floor, ceiling and walls or just open ground. A space in which any object placed in it will affect how we see that space and through our sensing system affect what we hear.

So what is the most obvious correlation between an empty space and sound? Silence surely. And if I were to fill that space entirely the result in sound would be white noise. If I was particularly perverse I could reverse the situation and say that the correlate of empty space is white noise and the full space silence. Either way a partial fullness of space, i.e. an object in the space is the equivalent of a band stop or band pass filter.

Of course we could substitute some other sub-set of white noise to represent fullness or the material with which a space is filled. All the audible tones of a chromatic scale for example or some other aggregate of tones or all the samples in our sound sample library.

If one were to build a single map that might serve for more than one dance, movement or installation work what parameters of sound would be mapped to the spacial dimensions of height, depth and width? If I used frequency, amplitude and disposition as the three primary elements of sound that a person moving in space controls, which axis applies to each of these elements?

The possibilities are:

1. X=frequency, Y=amplitude, intersection of XYZ=disposition
2. X=frequency, Z=amplitude, intersection of XYZ=disposition
3. Y=frequency, Z=amplitude, intersection of XYZ=disposition
4. X=amplitude, Y=frequency, intersection of XYZ=disposition
5. Y=amplitude, Z=frequency, intersection of XYZ=disposition
6. X=amplitude, Z=frequency, intersection of XYZ=disposition.

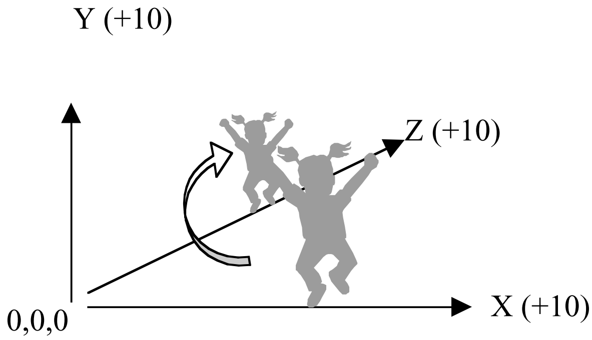


Figure 1. X = width, Y = height, Z = depth.

So what are we likely to hear in each of these possibilities?

1. In 1: As an object moves from co-ordinates X5, Y5, Z0 to X0, Y5, Z10 or in staging terms: down stage centre to up stage left (maintaining the same height all the way) we hear a fall in pitch and a steady volume while the sound travels from centre rear to left front. (Assuming here that there is a four-speaker placement with the observer in the centre of the four speakers.)
2. Using the same movement as above 2. gives us an increase in volume and the same fall in pitch as in 1. The same spacial distribution of sound applies to all the possibilities.
3. gives us a steady pitch and an increase in volume.
4. a decrease in volume from medium to very soft over a steady pitch.
5. a steady volume and a rise in pitch from one extreme to the other.

6. a decrease in volume from medium to very soft over a rise in pitch from one extreme to the other.

Now when we are dealing with an object in relation to a camera position the above results are modified. If the camera is positioned at, for example, a performer's eye level and in the first row of the audience, the performer fills more of the lens the closer she gets. So even when the height of the performance is static, but there is movement in the Z or depth axis the Y-axis will read an increase or decrease. There may also be some variation in the X axis as a receding object appears to our camera lens to get slightly thinner. (We'll ignore this as a marginal variation for the time being. Dancers generally don't need to look any thinner!) The above movement from down stage centre to up stage left gives us:

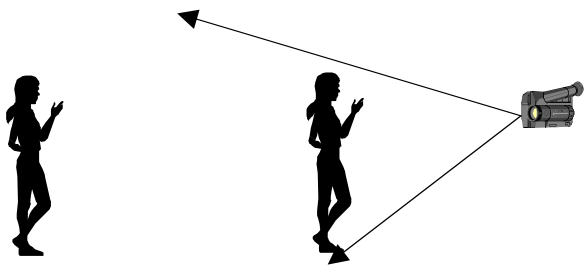


Figure 2. Objects closer to the lens fills more of the screen.

1. A decrease in volume and a fall in pitch while the sound travels from **high-to-low** centre rear to left front **medium-to-low**.
2. Is the same except for the alteration in distribution that applies to all possibilities.
3. gives us a fall in higher pitches and a rise in lower frequencies and an increase in volume.
4. a decrease in volume from medium to very soft over a fall in higher pitches and a rise in lower frequencies.
5. a fall in volume and a rise in pitch from one extreme to the other.
6. a decrease in volume from medium to very soft over a rise in pitch from one extreme to the other.

The above must be further qualified because we are dealing with a 3D object. So when I am talking about a rise or fall in pitch I am not talking about a single pitch but rather a band of frequencies and it is the whole band that rises or falls. And in relation to frequencies mapped to the Y axis (height) the band will be narrower or broader depending on where the object is on the Z axis.

So far I have been talking about just 3 sound parameters and the choice of frequency, amplitude and disposition might seem arbitrary. I think there are advantages in this choice. In particular the advantage of clarity and transparency, but other choices are available if we want to be less obvious. I could substitute resonance, or some other signal processing, for disposition resulting in:

H	X	A	A	F	F	R	R
H	Y	F	R	R	A	A	F
H	Z	R	F	A	R	F	A

Figure 3. Where A= amplitude, F = frequency and R = resonance

And so far I have not said much about the temporal domain. Speed of movement or tempo is another obvious element that should be built into the above permutations. I am currently playing with tempo as a means of introducing elements that go beyond the immediate one to one connections to movement. At higher rates of change in movement a slowed-down sonic "after-image" is triggered. Basically this is an isorhythmic canon and the idea behind it is essentially that the ear has another chance to catch the sound at a slower pace while also introducing some polyphony where there is just a solo dancer.

4 Conclusion

At this stage, without having completed and applied a fuller set of maps, it is not possible to say whether some of these generalisations will bear fruit for wider application even in just my own work. It is still a work in progress, but it is my belief though that more research into "obvious" correlations needs to be carried out so that more manifestations of artists' imagination have greater chance of making meaningful connections between collaborators, participants and audiences.

With the few example maps that have been experimented with, the results are encouraging. Even continuous unison of sound with movement has not proven to be so boring. (Certainly for the minimally inclined it can already be said to have been fruitful).

5 Acknowledgments

Thanks to Lindsay Vickery for further information on his work *Your Sky is Filled with Billboards of the Sky*. Thanks to David Rokeby for a copy of the SoftVNS and getting me started with it.

References

Coniglio, M. (1999) “MidiDancer”. See Troika Ranch website http://www.troikaranch.org/mididancer.html. April 7th 2002

Laban, R. (1966). “The Language of Movement: A Guidebook to Choreutics.” Edited by Lisa Ullman, *Boston, Plays, Inc*

Puckett, M & Ziccarelli, D. (1990-2002) “MAX/MSP”. MIDI, Audio and signal processing program developed at IRCAM and in association with Opcode and now Cycling ’74.

Rokeby, D. (2002) “SoftVNS” Beta version 7.1 see personal website http://www.interlog.com/~drokeby, May 23rd 2002.

Singer, E. (1996) “VideoIn.readme” Release notes for the MAX object VideoIn, p1.

Teck, K. (1989). “Music for the Dance: Reflections on a Collaborative Art.” *Westport, Greenwood Press*.

Winkler, T. (1995). “Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music”. *International Computer Music Conference*, p1.

Winkler, T. (1998). “Motion-Sensing Music: Artistic and Technical Challenges in Two Works for Dance”. *International Computer Music Conference*,p1-4

Granular Synthesis: Experiments In Live Performance

Timothy Opie

School of Music, Queensland University of Technology
email: tim_opie@yahoo.com
http://www.granularsynthesis.com

Photo 1: Timothy Opie & Fish

Abstract

This paper describes current research into Granular synthesis. It outlines the design and performance of a musical instrument based upon this Synthesis method. The instrument, design, process and products are used as a research tool for examining granular synthesis in real-time performance on a low budget.

1 Introduction

Granular synthesis has been used in composition for many years now, but is still not widely used as a performance tool. The intention of this paper is to give an overview of the making of a real-time granular synthesis instrument that can be used for live performances and improvisation with other instruments including traditional acoustic instruments. This paper gives a description of some of the goals and specifications, how these were met, and how the instrument performed in a live setting. It also addresses some future development for the instrument. The paper also outlines the building of a low budget granular synthesis instrument.

2 Background

Granular synthesis is perceived as a relatively recent development in sound synthesis, but it can also be seen as a reflection of long-standing ideas about the nature of sound. Quantum physics has shown that sound can be atomically reduced to physical particles (Wiener 1964). This physical form of sound was first envisioned by the Dutch scientist Isaac Beeckman (Cohen 1984). He explained that sound travels through the air as globules of sonic data. Later works including those by Gabor (Gabor 1946) and more recently Xenakis (Xenakis 1971), Roads (Roads 1988), and Truax (Truax 1990) has evolved the particle theory of sound into a synthesis method whereby the natural sound particle is imitated and magnified. The particle is then layered with other imitation particles, either cloned or extracted through a similar process as the original to create different sounds.

3 Design Specification

My study of granular synthesis is focused on making an instrument for both research and live performance. The research element will examine and catalogue a large range of sounds produced using pure granular synthesis techniques. It will also look at ways that this can be notated for musical scores. The performance element focused on creating an easy to use instrument that can be played live on stage and interact with other instruments.

As a musical instrument, I wanted to create something that could be used alongside traditional acoustic instruments for both improvisation and scored compositions. It was also important to have a control interface that was easy to learn, but which allowed the musician virtuosic control over the instrument. A major design consideration of the instrument was budget. It needed to be inexpensive to create and use.

Another major goal of the instrument was that it had to be the focal point for performance. It had to be visually interesting and give the audience a clear association between the sound and the instrument. The computer that actually generated the sounds as dictated by the instrument interface was to be invisible to the audience.

The actual granule producing protocol was a major concern. I wanted it to be portable and open source. Portable in the sense that people could download it to any kind of computer and just run it without having to recompile it.

The first place of exhibition for the instrument was to be at the REV festival in the Brisbane Powerhouse. This was a public event for the whole family to come and try out different ways of producing sound. With this in mind, I had to create an instrument shape that would be appealing to an entire family, young and old, which would encourage them to hold the instrument and try playing it.

4 Methods

The two key elements in working with granular synthesis are the method chosen to extract the sound grains and the method by which to layer them. Bowcott (1989), Keller & Truax (1998) and Hamman (1991) have researched various mathematical and scientific algorithms that have been used for the layering process. Many of these methods were not useful for this instrument design, as they are automated. Initially I examined and produced ways of incorporating mathematical algorithms in order to reduce the number of controllers needed, but eventually decided to incorporate a direct parameter control to the instrument interface. This was done to assist in the research element of the instrument design. I wanted to work from basics and then build from that. For the basic design, I mapped 10 sliders to 10 parameters that I thought were most sonically interesting. These parameters were:

Grain duration & random offset
Grain density & random offset
Grain frequency & random offset
Grain amplitude & random offset
Grain panning & random offset

During implementation this evolved slightly, and I am working on grain envelope shape control, sound source selection, and a more refined grain frequency control with a brass instrument based fingering style.

I decided that in order to make the instrument as versatile as possible I would get it to generate the parameters in the MIDI protocol. I used a CV to MIDI converter designed by Angelo Fraietta. I just bought a basic unit consisting of a circuit board and a pre-programmed PIC chip. I soldered all the sliders and other control devices to the circuit board and then placed the circuitry inside a large 1kg Nestle Quik tin, with the controllers protruding through holes drilled in it. Wooden panels were added around the controllers to ensure there were no sharp edges. This was an important feature considering it would be open for public use.

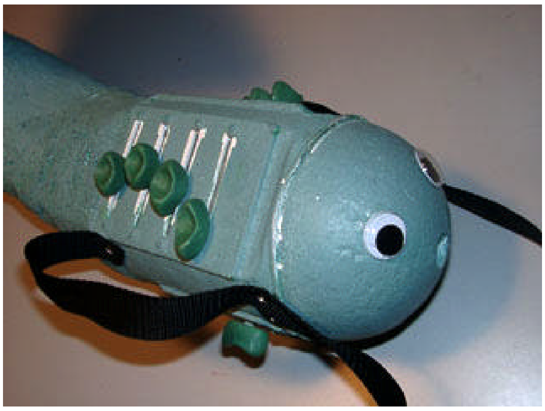


Photo 2: Close up of controls

Photo 3: Close up of MIDI, power and MIC sockets

I was concerned about the feel of the controllers so I made the finger controls out of hard setting plasticine and shaped them to each finger.

Once all of the components were installed in the tin, I then had to make a shape for the instrument. The shape was chosen in relation to a saxophone, but I knew that many children would be playing it at the exhibition, so I added more character to it. I made it a saxophone shaped fish complete with googly eyes. I made the shape out of Bostik gap filling foam by creating a mould around the tin and then spraying the gap filler into the mould. I chose the gap filler because I knew it would set quite hard but be extremely light. Lastly, I painted the instrument a marble green/blue colour.

Photo 4: The full instrument interface

The granular engine running inside the computer was done in two stages. First a prototype was made using Csound. This was primarily to check the interface as it was being made to ensure it worked correctly, and to set a benchmark. The final product was written in jMusic. jMusic is a java based sound synthesis and compositional tool written by Andrew Brown and Andrew Sorenson. Being Java-based allowed me to develop the instrument on a number of different computer platforms. jMusic has adopted the GNU Public License which my instrument also inherited making it open source, so that others may develop it further and modify it as they like.

The software development in jMusic went through a number of stages. I started by looking at different ways of creating grains of sound in jMusic. One very simple method was to create notes with lengths of only 20-30ms. This gave a good result, but the computation involved was too intense and was not suitable for a real-time instrument. I then started looking at creating a Java audio object to integrate into jMusic. Andrew Sorenson helped by creating a base, which I modified to suite the granular synthesis process and the instrument interface. The final part of the project involved getting the whole process running in real-time without any lag or unpredictable events.

The jMusic component is still being refined to include more features, as mentioned previously, envelope shape control, sound source control, and more refined frequency control which I am integrating through the use of a brass instrument based fingering style. This will include adding four buttons to the instrument the first three will follow the standard three valve brass instrument fingering style from C4 to B4, whilst the forth will be used as a substitute for the lack of breath control the musician has over the pitch. For example the note E4 and A4 have the same fingering, but the fourth button will differentiate between the two notes. There will also be a slider to control the octave. It will also retain the current frequency controlling slider, which can be used to offset the current pitch. Another feature that is being worked on is microphone input. I specifically want to incorporate microphone input because it adds a large array of sound sources and control that can be changed as quickly as the performer can create a vocal sound. In previous experiments, the microphone was found to lag when the whole process started working in real-time. This latency issue is something I wish to resolve in the near future.

Another main component to the instrument was the actual computer it ran on. Despite the computer being kept hidden away, it still would have been impractical to use a large powerful computer, or network of computers. Itagaki, Manning and Purvis suggested a live real-time granular synthesis process using 53-networked computers that could control nine voices at once [Itagaki, Manning, Purvis 1996]. Whilst this opens up a fantastic range of options, it poses many problems that would negate some primary objectives of this instrument. These include primarily cost, and computer visibility. The instrument was designed with only one voice, so in order to have multiple voices you would need more performers, each playing an instrument in a granular synthesis orchestra. The program ran on just a single computer of reasonable speed, which reduced costs a lot. Due to the nature of Java, I was able to test and run the instrument on a number of different machines and operating systems. For the REV performance I chose to use Windows 2000 on a Pentium 4 based computer. This gave the fastest Java results, with the least hassles, and at a less expensive cost than a Macintosh or other non-i386 based computer. Using BSD or Linux as the main operating system would decrease the cost even more, but the MIDI-IN support required for the sound card was not adequate. I am currently in the process of improving the MIDI-IN support for FreeBSD as this will add to the stability of the instrument.

5 Performance

The first public exhibition of this instrument was held at the REV festival in Brisbane in April 2002. For the most part the instrument was located in the elevator, with an accompanying performer. The location of the instrument had some positive and negative results. On the positive side, it gave the performer and the listeners an intimate space where they could absorb many of the sounds without being distracted by other exhibitions at the festival. On the negative side many people missed out on seeing it because they weren't expecting anything to be in the elevator, and most of the displays were on the split ground level, for which an elevator was not required. It was also cramped in the elevator meaning that many people didn't want to take the time to stay and learn more about the instrument. The elevator broke down on the final day of the exhibition. The worst part was traveling on the elevator so long gave me motion sickness, although the other two performers handled it much better.

As expected, children loved it. They were curious about the shape of the instrument and of what the different controllers did. The two students who helped me with performances, Rafe Sholer and Joel Joslin, both picked up the instrument playing techniques very quickly. Within half an hour, they were already picking out certain finger combinations that worked together and gave pleasing aural results. They could both see a lot of potential with the instrument and enjoyed working with it. Joel said "the instrument definitely has virtuosic potential, but at the same time it is very easy to create interesting sounds without much instruction".

One highlight of the exhibition was to do a live performance with Benn Woods, Andrew Kettle and Greg Jenkins. We improvised music for a silent experimental Russian film. This performance really allowed me to try out the different sounds I could create in a large space and a large sound system. The effect was vastly different from that of the elevator performances. We will be doing some future performances together as a result.

The main reason I am now working on ways to modify the frequency control of the instrument is because I found it too restrictive being limited to sliders. I could jump quickly from one frequency to another with very little slide, but it was inaccurate. With a slider and preset frequencies, I could play more accurately. This would enhance performance especially when being played as part of an ensemble. For the generation of soundscapes the slider method proved to be adequate.

I found that using a physical interface shaped like a more conventional instrument helped the audience to become more involved with the music, and I was likewise more able to interact with the audience. I could watch the audience at all times and not have to look at the instrument. If I was to perform the exact same music but with the computer sitting on the desk, and myself in front of it playing with a mouse, or even a mixing panel, the audience participation would be decreased, as I would be interacting with the computer rather than with the intended audience. This restriction applies to all computer-based music. The shape and lightness of the instrument helped with the performances. As seen in Photo 1, the performer can hold the instrument quite comfortably and play, without even needing to watch the instrument. It also includes a strap to go around the neck so the performer is freer with their actions and body movements. The shape suited my hand very well. For a small child it was too large, but as I was the main performer, it was more suitable to shape it to my hand size. Joel and Rafe also had similar sized hands, so they also found it easy to hold and maneuver.

Another finding I made was that the audience did not expect such a wide range of sounds to come from one instrument. I was able to create sounds from growly rumbles that vibrated the floor, right up to bursts of high pitched flutters.

I found the granular synthesis real-time instrument to be very useful and very flexible in performance. Even though it was created on a small budget it provides the performer with a wide range of instantly variable sounds to explore.

One last performance issue of exceptional worth was the fact that the program did not crash once in the 3 days that it was running.

6 Conclusion

This paper has discussed my method of creating the instrument, and some of the issues I addressed whilst making it. I have shown that a small budget can be used to create an instrument that is visually as well as gesturally and aurally stimulating. I have briefly explored the experiences of other performers and the audience in relation to the instrument.

Granular synthesis has been used as an effects mechanism in many recent live performances, but this instrument shows that it can be used as a pure sound generation and music-performing instrument. Whilst this has been done before, it has not been used as just an instrument with the expectation of performing with other instruments.

The instrument met most of the design specifications. It still has a large area that can be worked on and refined to increase the versatility and playability of the instrument, as have been noted.

This instrument allows quick observations to be made and documented which will help with the planned research into notational conventions.

Most importantly, this instrument entertained audiences with its visual characteristics and sound qualities.

7 Acknowledgments

These people have helped me with my project in some way: Andrew Brown, Andrew Sorenson, Gabriel Maldonado, Angelo Fraietta, Craig Fisher, Stuart Favilla, Aaron Veryard, Steve Langton, Steve Adams, Steve Dillon, Paul Cohen, Adam Kirby, Rene Wooller, Joel Joslin, and Rafe Sholer.

References

Bowcott, P. 1989 'Cellular Automation As A Means Of High Level Compositional Control Of Granular Synthesis' ICMC Proceedings 1989. ICMA 55-57

Brown, A & Sorensen, A. (2001) jMusic http://jmusic.ci.qut.edu.au/

Cohen, H. 1984 Quantifying Music Dordrecht: D. Reidel Publishing Company.

Gabor, D. 1946 'Theory of Communication' The Journal of the Institution of Electrical Engineers. London: Unwin Brothers. 93(3): 429-457.

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

Hamman, M. 1991 'Mapping Complex Systems using Granular Synthesis' ICMC Proceedings 1991. ICMA 75-478.

Itagaki, T. & Manning, P. & Purvis, A. 'Real-time Granular Synthesis on a Distributed Multi-processor Platform' ICMC Proceedings 1996, ICMA 287-288.

Keller,D.& Truax, B. 1998 'Ecologically-Based Granular Synthesis' ICMC Proceedings 1998. ICMA

Roads, C. 1988 'Introduction to Granular Synthesis' Computer Mus. J. MIT Press 12(2):11-13.

Roads, C. et al. (ed) 1996 The Computer Music Tutorial. Cambridge:MIT Press.

Truax, B. 1990 'Composing with Real-time Granular Sound' Perspectives of New Music. USA: Hamilton Printing Company. 120-134.

Xenakis, I. 1971 Formalized Music. Bloomington: Indiana University Press.

Wiener, N. 1964 'Spatio-Temporal Continuity, Quantum Theory and Music' The Concepts of Space and Time. Dordrecht: D. Reidel Publishing Company. 539-546.

There will be some sound samples of this instrument available at: http://www.granularsynthesis.com

Spatialisation, Method and Madness Learning from Commercial Systems

Ian Stevenson

Fairlight ESP Pty. Ltd, Sydney Australia, *email:* audile@bigpond.com

Abstract

Computer music systems have supported dynamic speaker map models for sometime. With the advent of multiple surround delivery formats, commercial film and TV mixing systems now also support automated down-mixing, albeit for limited range of speaker formats. This paper compares the methods and approaches to producing multi-channel material for multi-speaker replay in the various contexts of music, film and TV, theatre, and installation. The author's experiments in automated spatialisation developed in the PD composition environment are also introduced. In addition to this technical review, the paper also considers aspects of aesthetic and psychological significance in the presentation of sound material in multi-speaker replay systems.

1 Introduction (Method and Madness)

This paper provides an overview of the areas of interest that have influenced my experiments in developing an interface for multi-speaker diffusion. In some ways the process is more interesting than the outcomes, as my experiments are somewhat idiosyncratic and not generally applicable.

Recently, I have been working in a commercial environment as part of a team developing a new mixing system for post-production. This experience has opened my eyes to the priorities that influence the development of such systems and the demands of those professionals working with this style of equipment. These priorities represent a strange amalgam of method and madness, excellence and compromise. However, I have become aware that the technical infrastructure behind much contemporary music is subject to very similar forces.

1.1 Philosophy of Design

There are as many approaches to the design of multi-speaker sound reproduction systems as there are applications for presenting sound in space. It is also obvious that the creative applications to which technology is applied often subvert the original intentions of the designers.

Reproducing Reality Reproducing acoustic reality or synthesising a reasonable representation or reproduction of an acoustic event is the principle behind many systems. The most ubiquitous example of this approach is the theoretical foundation of the stereo reproduction system (Rumsey 1998). Other systems, which apply this approach, are: ambisonics (Malham 1990, 2000; Gerzon 1984, 1985, 1992); virtual reality and industrial simulation systems employing various approaches including very large speaker arrays and other methods of wavefront synthesis (Boone, et al 1996).

These types of systems attempt to recreate a sound field, or at least a part of the sound field.

Pragmatic Approaches Many, more pragmatic approaches attempt merely to present enough auditory information to give the listener the impression of sounds in spatial relation to each other. These systems often rely on intensity panning and the resultant limited time domain cues available within a limited listening area. Examples of these approaches included cinema surround sound systems and their associated encoding and panning methods (Stevenson 2001, 2002); vector based amplitude panning systems and their employment in musical systems and various entertainment installations (Pulkki 1997; Drost and Rolfe 1999)[1]; and triangulated panning methods (LCS 2002).

---

[1] Note: Chris Rolfe of Thirdmonk Software has indicated that the methods employed in ABControl

Spatial Encoding Spatial encoding methods also show significant diversity. Whereas A and B format ambisonic signals (Gerzon 1985) encode an entire sound field in a finite number of matrixed audio channels, other systems store automation or metadata along with a number single channel audio signals or pre-mixed mutichannel stems.

In many cases, these methods of encoding provide, to a greater or lesser degree, a level of speaker system independence. This allows a system that records positional information along side raw audio signals, to be decoded for playback on a speaker system of different proportions to the one on which the material was encoded. This feature usually requires the speaker system to conform to a basic set of design criteria specific to each encoding method. For example, a system may only contain dimensional information for a single plane, or may require the loudspeakers only to be positioned only on the perimeter of the listening area.

Some encoding systems allow spatial transformations to be performed on the encoded material. These processes include divergence, rotation, and soundfield contraction or expansion (Wishart 1996). Most commercial systems offer hybrid approaches, allowing the use of discrete, encoded and pre-mixed elements to be combined.

2 Overview of Systems

There are a wide range of acoustic and psycho-acoustic properties that may be employed and exploited, or ignored in the design of a multi-speaker replay system. In this paper I will merely list the significant parameters to give an idea of what has and hasn't been included in the various approaches. Please refer to the relevant literature for more detailed information.

The auditory system can detect a number cues which assist it in evaluating the properties of individual sound sources, their size, location and motion; properties of the environment, its material composition and disposition; and properties of the composite soundfield which is the result of the temporal, spatial, and timbral or spectro-morphological relations between individual sound sources, and the spatial relations between the sources and elements of the environment. Nearly every acoustic and psycho-acoustic parameter is interdependent and the designer must select a subset, or group the combination of effects to arrive at a set of parameters which can be controlled to produced a reasonable simulation of the spatial location of static and moving sound sources.

Of course, we do not only perceive sound as the result of bodies behaving according to the laws of physics. Sounds come to us in moments of reverie and confusion, in dreams, and from electronic music and the ever-present electronic media, which create synthetic sound images for which evolutionary psychology has not prepared us.

2.1 Potential System Parameters

Acoustic/Physical Properties The acoustical properties that may be considered in the specification of a diffusion system include:

- Source position – this is the primary physical property we are aiming to control. The perceived source position is to be determined relative to the listener and relative to the loudspeakers. We may need to consider that there is more than one listener and that the source and or the listener may be moving.
- Listener position – for large audiences this may be difficult to specify. However a listening area is usually assumed as part of the system specification. Lack of a single listener influences the range of parameters which will have effect. The listener position maybe used as a higher order control parameter in spatial transformations performed on an entire sound field.
- Relative intensity – although sound intensity diminishes in proportion to distance from the listener, intensity must be used in combination with other parameters to achieve an effective distance cue. The relative intensity of signals produced by each loudspeaker is the primary tool used to simulate the location of sound sources.

software should be distinguished from the methods described by V.Pulkki. Informal tests have revealed significant differences in the performance of the two systems.

- Relative loudness – the relative loudness of two sources gives the listener information on the physical properties of sound sources. This includes the size and kinetic energy of the sources.
- Fixed propagation velocity of sound – the fixed propagation velocity of sound in air or any other stable medium results in a temporal offset between sounds from different source locations. Control of relative arrival time can be used in some situations but a range of listener positions may reduce its effectiveness. The fixed propagation velocity of sound also affects the perception of moving sources. This results in the Doppler effect where the pitch of moving sources changes with their angular velocity relative to the listener.
- Frequency dependent transmission properties of air and enclosed spaces - high frequency attenuation with distance in free air, influences the perception of relative spatial locations of sources, and the trajectory of moving sources. The frequency dependent transmission properties of enclosed spaces produces similar effects though not necessarily on high frequencies alone.
- Spectral balance – the overall spectral balance is related to various other physical properties.
- Reverberation time – the overall reverberation time is determined by the volume and material properties of an enclosed space.
- Direct/reverberant ratio – the ratio of direct energy to reverberant energy provides a strong cue to the distance of the source and the overall properties of the space.
- Direct/early reflection arrival time difference – the time delay between the arrival of the initial direct propagation signal and the first cluster of early reflections from nearby surfaces gives the listener cues to relative locations of sources and surfaces and the size and composition of space.
- Diffusion – the acoustical properties of diffusion relate to both the spatial distribution of reverberant energy and the coherence or correlation of that energy. These parameters effect both individual sources and the entire soundfield.
- Source frequency dependent directivity – the source directivity will effect the perception of position, distance, trajectory and orientation of the source.
- Interaction of interconnected spaces and objects within the environment – complex spaces will cause a range of acoustic interactions that may be modeled by a diffusion system.
- Occlusion and Exclusion – the transmission properties of the interface between spaces will effect the perception of spatial relations with the listening environment.
- Obstruction – transmission path obstructions will effect the perception of both static and moving sources. Reflection, refraction and diffraction effects all give clues as to the position and the physical properties of sources. These effects are often responsible for mistakes and confusion in the analysis of real acoustic environments.
- Acoustic properties of environmental materials – absorptivity and specularity affect the overall sound field.

Psycho-acoustic/Perceptual Properties These psycho-acoustic properties are responsible for creating the perceptual cues that our auditory system responds to. The psycho-acoustic properties that may be considered in the specification of a diffusion system include:

- Inter-aural time difference (ITD) – the difference in arrival time between the two ears is a strong cue for the position of a source in the horizontal plane. The auditory system has a higher resolution for sources lying in the front ninety degrees. ITD cues may be manipulated when two or more loudspeakers are used to reproduce a sound in space. The effectiveness of this approach is reduced by inter-aural cross-talk and non-fixed listener positions.

- Inter-aural level difference (ILD) – otherwise known as inter-aural intensity difference (IID) is the primary means of creating the illusion of spatial source location. ILD and ITD interact through vector summation when continuous signals are panned between loudspeakers. However, the auditory system is not very effective in localizing continuous signals and continuous signal are not very common in music or the environment. Continuous signals are most likely to cause confusion in the perception of source location.
- Reverberation – the reverberation properties of moving sources as described above are strong cues for distance perception.
- Head related transfer function (HRTF) – the perception of source location in both horizontal and median plane is affected by the filtering properties of pinna, shoulders and head. Luckily this complex tool is built in to our hearing but its effects may be exaggerated in the replay system.
- Head movement – head movement is probably the most effective tool the auditory system has for discriminating source position. Head movement relies on all the other cues simultaneously. It also destroys or seriously modifies the most subtly created multi-speaker soundfield (however, this effect is reduced as the number of loudspeakers employed increases).
- Inter-aural correlation or diffusion – the similarity of signals presented to the two ears effects the evaluation of the acoustic properties of the environment.

I will briefly mention some of the systems and methods employed in various fields of sound production.

2.2 Computer Music Systems

Vector Based Amplitde Panning (VBAP) as described by Pulkki (1997) and as employed in Csound (http://www.csounds.com), Max/MSP (http://www.cycling74.com) and PD (http://www-crca.ucsd.edu/~msp/software.html) signal processing objects, describes the loudspeaker and source locations in terms of vectors. Signals are distributed between any three loudspeakers in a three-dimensional loudspeaker system. The signal level fed to each loudspeaker is modified by a gain coefficient determined by the length of orthogonal vector components. This method allows source locations and trajectories to be specified in terms of vectors in a unit sphere. Loudspeaker locations may be specified independently but must lie on the surface of the unit sphere. Total intensity is directly proportional to vector length therefore the vector representation does not reflect the actual source location and source trajectories may not pass through the centre of the space with contiguous vectors. As has been pointed out, the centre of the listening space is difficult to define for more than one listener.

Csound signal modifiers (Vercoe 1992) in addition to “VBAP” include “Pan”, “Locsig”, and “Space”. These signal processing routines use intensity panning with the addition of sends to global reverberators. They allow the use of user definable pan laws but are limited to quad type loudspeaker configurations. Source trajectories may be read from a predefined table. Trajectories lie within the boundaries of the loudspeakers.

Symbolic Sound Kyma (Symbolic Sound Corp. 2001) includes a spatialisation system where source locations are specified in terms of polar coordinates relative to a unit circle. Source locations within the unit circle have a vector length of 0 – 1, outside the unit circle source locations approach the acoustic horizon at a vector length of 2. This system allows an intuitive relationship between source position and loudspeaker position. Trajectories may be entered on a graphical timeline alongside other signal processing functions including reverberation parameters. Source locations and speaker locations may be specified independently allowing for re-rendering spacialisations on different loudspeaker systems.

Gmebaphone/Acousmonium The Gmebaphone (Clozier 2001) and GRM Acousmonium are complex and unique systems employing intensity panning, reverberation, equalisation and loudspeaker design or "registers" to modify the perception of individual sources and pre-mixed tracks. Loudspeakers are distributed throughout the auditorium. The unique acoustic properties of these systems are exploited in much the same way as a musical instrumental ensemble. It is my impression that adapting a piece for these ensembles would be considered similar to the process of re-orchestration of instrumental music

2.3 Virtual Reality, Auditory Display, Simulation and Research

This is a category of convenience that I am using to introduce spatialisation systems where spatial resolution is related to the number of loudspeakers employed. This is a general principle that may be employed in most systems. In auditory display systems, the human auditory system's increased ability to discriminate multiple signals through spatial separation is often exploited through the application of a loudspeaker-per-source approach. Wave field synthesis (Boone, et al 1996) and very large speaker arrays offer approaches to synthesising facsimiles of real sound fields.

2.4 Commercial Installation and Entertainment Systems

Level Control Systems SuperNova and CueStation (LCS 2002) products employ a method of defining a listening space or speaker map as a set of triangles or tri-sets of loudspeaker locations including virtual source locations. Virtual locations may distribute signals amongst other loudspeakers or may feed external signal processing systems for adding reverberation and smoothing trajectories. Virtual locations may also be placed beyond the perimeter of the listening area to provide an acoustic horizon. Signal intensity is modified in proportion to the distance from the source to any of the three vertices of the surrounding tri-set. This may result in the signal being sent to any number of loudspeakers. Source trajectories can be plotted on a graphical interface and the speaker map may be modified independently of the trajectories.

2.5 Other/Cross-over

This category includes systems that are employed in any of the above areas.

Richmond Sound Design AudioBox This hardware platform provides user defined pan law tables, a cross-point gain matrix, audio playback and a cue sequencing system. Any software capable of producing MIDI system exclusive messages can control the system. ABEdit (Clampett 2002) offers a rudimentary event based interface. ABControl (Drost and Rolfe 1999) offers a vector based panning system including algorithmic trajectory utilities. This vector based system offers loudspeaker independence but similar to other VBAP systems, the vector locations do not map directly to source locations within the unit circle.

Ambisonics (Gerzon 1984,1985,1992; Malham 1990,2001) as mentioned earlier, is a method of encoding an entire sound field. B Format ambisonic signals include three orthogonal velocity and one omni-directional pressure component. From these signals, discrete loudspeaker components can be derived for any position on a scaled unit sphere. Ambisonics provides a convenient means of specifying spatial information for discrete sources or recording acoustic events with specially design microphone arrays. In addition to its general applicability, ambisonic encoding allows high order spatial transformations, including rotation, and scaling, to be applied to the entire soundfield.

Ambisonics has been employed in multi-channel broadcasting, music recording, and computer music applications.

2.6 Film and TV Mixing

Multichannel audio formats have been a part of cinema since the 1930s. Dolby Surround, introduced in the 1970s made four-channel audio a standard in cinemas around the world. Dolby's mono and stereo compatibility enabled it to be introduced as a no additional cost added feature in stereo television broadcasting, requiring the consumer to invest in a consumer pro-logic decoder.

By the end of the 1980s digital audio had made its way into the cinemas and Dolby digital 5.1 become the standard. During the 1990s Sony's attempt at vertical market integration, following the purchase of CBS, saw the introduction of SDDS 7.1. More recently, the Dolby empire struck back with Dolby EX 6.1 for the most recent George Lucas "Star Wars" extravaganza. With the advent of DVD as a medium for domestic movie viewing and innovative music-in-surround plus audio-visual content, media producers have had to contend with multiple delivery formats.

Luckily all these surround sound formats are two-dimensional and controlled by well adhered to standards. Cinema dubbing mix engineers have been very conservative in their use of surround effects. Due to extensive testing and trial showings mixers have learned what will and will not translate to different listening environments. Their priority has tended to be supporting the narrative and this translates to dialogue. Some practitioners such as the famed Walter Murch have managed to utilise spatialisation as a narrative tool. Interestingly, Murch cites Orson Wells mono radiophonic and cinema soundtracks as being his greatest influence (School of Sound Lecture, London 1997). Wells proved that not only could space be used as a narrative tool but that all the space you needed could be encoded onto a single low dynamic range, low bandwidth audio channel.

To briefly summarise the speaker layouts of the various different formats: mono is still the predominant format used by most televisions and radios; most formats must provide stereo compatibility for radio and television; Dolby Surround includes left, centre, right and a band limited surround channel. The surround systems in cinemas usually comprise an array of loudspeakers down each side of the cinema. Dolby Pro-logic 2 includes two surround channels, the domestic standard allows for two discrete surround loudspeakers; Dolby Digital 5.1 and DTS digital both provide five full range loudspeakers plus one sub-bass channel; Sony Dynamic Digital Sound (SDDS) provides left, left-centre, centre, right-centre, right, left-surround, right-surround, plus sub-bass (this arrangement of five loudspeakers across the screen was first introduced in 1954 for the 70mm wide-screen releases of the musicals “Oklahoma” and “South Pacific”); Dolby EX introduces a centre surround channel allowing the rear soundfield to be more consistent and integrated by filling the hole-in-the-middle where a phantom image may otherwise have existed.

Any mixing system intended for use with such a range of target formats must be inherently flexible in its mapping of spatial position to speaker system. Many current post-production systems allow material to be mixed in more than one format simultaneously with monitoring systems that can be switched between formats. Two such systems are the DSP Media Postation 2 (Stevenson 2001) and the Fairlight DREAM Console (Stevenson 2002). These products are integrated disk recorders editors and mixers. Both systems allow channel panning independent of speaker format. The Postation offers a touch-screen panning interface which may be superimposed on top of the video image. Pan position is recorded along side other mix automation data as a series of X,Y co-ordinates. When played back these coordinates are translated into individual bus element gain coefficients determined by their position relative to the loudspeakers in a virtual map of the soundfield.

Fairlight’s DREAM system offers several additional features. The monitoring system provides a user modifiable matrix of bus element down-mixing coefficients to support differences between mixing and monitoring requirements for domestic and theatrical target formats. These differences are primarily in the area of bass management. The DREAM system includes support for pre-mixed surround stems. Stem mixes are often produced separately for dialogue music and effects and may include many separately spatialised tracks. Multi-channel stems may be controlled like any other signal path with provision for level, equalisation, dynamic range control, grouping and routing. In addition to these features, higher order spatial transformations performed on an entire soundfield, such as rotation, contraction, expansion and divergence can be performed on a premixed stem. All these controls may be automated, trimmed and edited during the mixing process.

Panning control in the DREAM system is provided with motorised joysticks. This brings up the interesting problem that the spatial mapping system must satisfy the conflicting requirements to be both a useable physical control interface for specifying perceived source location in real-time plus an accurate space map for up or down-mixing to other formats. Computer music practitioners who have used joystick during diffusion performance will know that the equal power pan curves used in stereo panning systems do not translate well to the psychological expectation of joystick response. In fact a linear relationship of joystick position to source location is often more acceptable. Similarly the curves suitable for left to right panning are not necessarily appropriate for front to back panning. Not only is the auditory system less discriminating in this axis but the speaker systems on cinema mixing stages have radical differences in type and topology between the screen and surround system. The earlier Fairlight "Fame 2" system attempted to overcome this by providing user configurable pan tables that could be specified independently for left to right and front to back panning operations. This also fulfilled the requirement for –3dB equal power panning for music applications and other laws such as –4.5 dB at centre for broadcast applications (Rumsey 1998).

3 Pan-O-Rama (not another half-baked spatialisation system)

Having taken into consideration all of the possible approaches outlined above. I have attempted to develop an idiosyncratic system which fulfils the requirements which are of most importance to me.

3.1 Compositional Tool

This system is a compositional tool for sonic art as opposed to a hybrid performance tool such as the Gmebaphone. Although simple trajectories can be recorded and controlled in real-time, more general human computer interaction problems have not been dealt with. Trajectories can be defined and edited on-screen with a pointing device.

3.2 Arbitrary Speaker Placement

Some argue, with reasonable force, that the seating arrangements in an auditorium have as much impact on the reception of the music as some of the more subtle spatial relations of which the music is composed. In addition to this consideration, loudspeakers should not be restricted to lying on the perimeter of the listening area. From these points you should be aware that I am not concerned with creating accurate representations of real soundfields. Although the orientation of the soundfield with respect to the listener may well be critical in conveying some of the meaning bearing gestures of spatial music. Simulation and approximations and effects are the concepts I am working with.

3.2 Pan Through Centre

Although the centre of the listening area may not be well defined in any loudspeaker arrangement, the fly over is one of the most effective gestures in cinematic spatial language. It can also be used with some subtlety, if sparingly, in more refined works of acousmatic music. The control interface should relate source position directly to perceived source location as accurately as possible given the deficiencies of multi-speaker intensity panning noted earlier. This requires the source target to be moved through the listening area and into the distance beyond the loudspeakers.

3.3 High Order Spatial Transforms

The user interface is not quite there yet but the system must support effects such as rotation and contraction.

3.4 Speaker System Independent

The panning system must support the recreation of spatial effects on systems with different numbers and positions of speakers. The speaker layout can be edited as simply as the trajectories.

3.5 Front End GUI and Scripting Output Independent of DSP

Software such as PD, running on the latest general-purpose personal computers are now providing signal-processing resources adequate for simple spatialisation. Sequencing data can be read from simple text files with interpolating look-up tables providing gain coefficients in realtime.

3.6 Flexible Pan Laws

Pan laws can be generated arbitrarily and saved as text files to be loaded into interpolating look-up tables. Floating-point indexes allow pan tables to be of any size.

3.7 Support for Position Dependent Reverberation and Decorrelation

The system produces a distance parameter and allows the control of reverberation through the use of position information and scripted control of reverb time and direct/reverberant energy ratios.

3.8 Triangulated Pan Map

Energy is distributed between any three loudspeakers. Virtual loudspeaker positions allow for smooth panning in complicated loudspeaker layouts. These ideas are borrowed from the Level Control Systems SuperNova and CueStation products. However, the geometry model may be unique. Figure 1 demonstrates the method of generating indices for the pan table look-ups. Although this looks complicated, it results in equal loudness for any source position within the triangle.

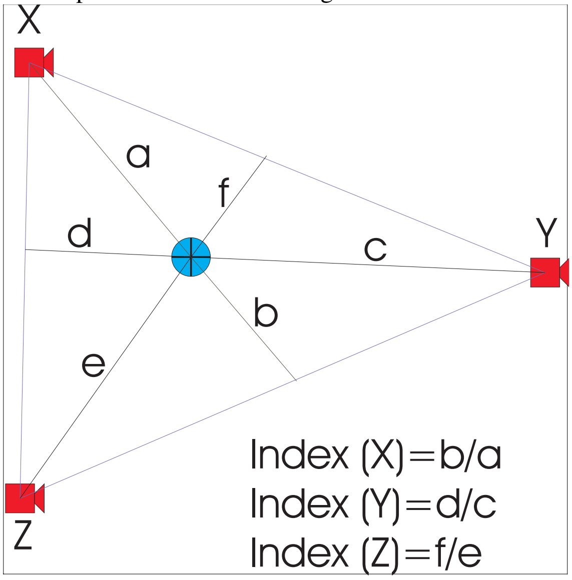


Figure 1. Triangulation Method

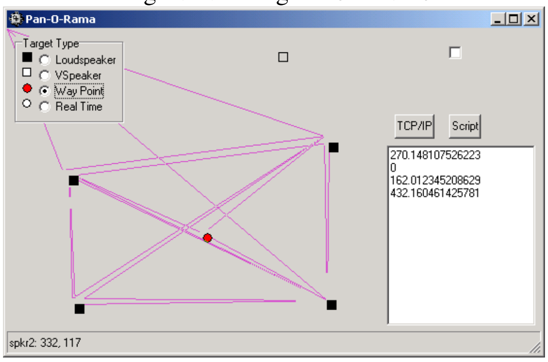


Figure 2. – The Pan-O-Rama Interface

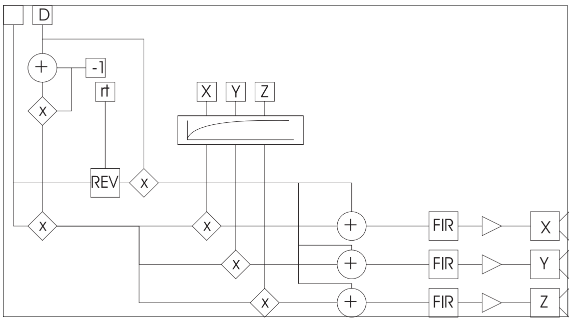


Figure 3. – Typical Signal Processing Arrangement

4 Aesthetic Considerations

Unfortunately, for a non-programmer, the work involved in developing projects such as this, tend to distract from the objective of making interesting work with sound. What, after all is the point of being able to whiz sounds around at will?

4.1 Validity of Space as Musical Property

Much debate surrounds the degree of importance to be attached to spatial manipulation as a significant element of musical language. The arena of musical practice is as broad today as it has ever been and there is undoubtedly continued scope for interesting work in this area. I would merely like to add some observations from areas not strictly aligned with contemporary music debate.

Elementary psychoacoustics tells us that the auditory system's capacity to distinguish discrete sonic events or streams is aided by spatial separation of sources. This explains part of the pleasure of listening to chamber music in the recital hall rather than at home on the record player. The other part of the explanation is the drinks at interval.

Cinematic language as I have mentioned earlier has occasionally managed to overcome the temptation for more meaningless special effects. Spatial effects in the soundtrack have been put to good use in supporting and augmenting the narrative path of drama. Additionally, surround sound can offer support to the less tangible aspects of developing non-verbal relationships between characters, locations, and spaces. This is the work that is otherwise done in the descriptive passages of literature which cinema must replace with sounds and images. Interestingly, the language of spatial cinema sound was developed primarily during an era when a monophonic audio track was the standard.

Illusion and Confusion in auditory perception is the area of most interest to me. Although I have done little research in this area I am convinced that there is something significant about the psychological effect of perceptually confusing sounds. In my experience, when the spatial origin of a sound is difficult to pin down, its cause or acoustic source is also difficult to determine. And the reverse is often true. This effect can be a strong stimulus to the imagination. You may call me a dreamer, but these states of imaginative reverie are exceptionally rewarding and often lead to a more "musical" perception of the natural soundscape. This process can often be triggered by the confusion caused by conflicting auditory cues when sounds a played back over multiple loudspeakers. In my case this either leads to intense frustration or enhanced absorption in the spectro-morphological content. I have to admit that too much objective listening or reduced listening in the past has caused me to overlook these experiences until recent years when I have become less involved in other peoples music.

5 Conclusions

This wide ranging discussion has touched on the broad issues concerning the development of spatialisation systems for a diversity of applications. The somewhat idiosyncratic approach has been steered by my particular interests which I hope will serve as an introduction to the field for some and a stimulus for further debate for others.

6 References

Blauert, J. 1983. Spatial Hearing: The Psychophysics of Human Sound Localisation. Cambridge, Massachusetts: MIT Press.

Blesser, B. 2001 "An Interdisciplinary Synthesis of Reverberation Viewpoints." *Journal of the Audio Engineering Society* 49:10; 867-903.

Boone, M.M., Verheijen, E.N.G., and Jansen, G. 1996. "Virtual Reality by Sound Reproduction based on Wave Field Synthesis." Proceedings of the 100th Convention of the Audio Engineering Society 1996. Copenhagen, Denmark.

Bregman, A. 1999. *Auditory Scene Analysis – The Perceptual Organisation of Sound*. Cambridge, Massachusetts: MIT Press.

Chamberlin, H. 1980 *Musical Applications of Microprocessors*. USA: Hayden

Chion, M. 1994 *Audio-Vision - Sound on Screen.* New York, USA: Columbia University Press.

Chowning, J. 1971. "The Simulation of Moving Sound Sources" *Computer Music Journal* 1(3):48-52.

Clampett, S. 2002. *ABEdit The User Manual* Richmond, Canada: Richmond Sound Design Ltd.

Clozier, C. 2001. "The Gmebaphone Concept and the Cybernëphone" *Computer Music Journal* 25(4):81-90.

Drost, J. and Rolfe, C. 1999 *ABControl User Manua.l* Vancouver,Canada:Third Monk Software, Inc.

Gerzon, M.A. and Barton, G.J. 1984. *Ambisonic Surround-Sound Mixing for Mutitrack Studios* Conference Paper C1009 of the 2nd Audio Engineering Society International Conference, Anaheim, USA.

Gerzon, M.A. 1985. "Ambisonics in Multichannel Broadcasting and Video" *Journal of the Audio Engineering Society* 33:11; 859-871.

Gerzon, M.A. 1992 "General Metatheory of Auditory Localistation" Preprint 3306 of the 92nd Audio Engineering Society Convention, Vienna 1992.

Harrison, J. 1999. "Diffusion: theories and practices, with particular reference to the BEAST system " eContact 2.4 http://cec.concordia.ca/econtact/Diffusion/Beast.htm

Kinsler,L.E., Frey, A.R., Coppens, A.B., Sanders, J.V. 1982 *Fundamentals of Acoustics* New York, USA: John Wiley and Sons.

Level Control Systems 2002. *CueStation v3.1 Users Guide* Sierra Madre,CA: Level Control Systems Inc.

Malham, D.G. 2001 "Towards Reality Equivalence in Spatial Sound Diffusion" *Computer Music Journal* 25(4):31-38.

Malham, D.G. 1990 *Ambisonics – A Technique for Low Cost, High precision, Three Dimensional Sound Diffusion*. Proceedings of the International Computer Music Conference, Glasgow 1990.

Moore, F.R. 1991. "Spatialization of Sounds over Loudspeakers" in Mathews, M.V. and Pierce, J.R. 1991. *Current Directions in Computer Music Research*. Cambridge, Massachusetts: MIT Press.

Moore, F.R. 1989. "A General Model for Spatial Processing of Sounds" in Roads, C. Ed. 1989 *The Music Machine – Selected Readings from Computer Music Journal*. Cambridge, Massachusetts: MIT Press.

Moorer, J.A. 1977. "Signal Processing Aspects of Computer Music – A Survey" *Computer Music Journal* February,1977:4-33.

Moorer, J.A. 1979. "About This Reverberation Business" *Computer Music Journal* 3(2):13-27.

Pulkki, V. 1997. "Virtual Sound Source Positioning Using Vector Based Amplitude Panning." *Journal of the Audio Engineering Society* 45:6; 456-466.

Pulkki, V. and Karjalainen, M. 2001 "Localization of Amplitude-Panned Virtual Sources, Part 1: Stereophonic Panning." *Journal of the Audio Engineering Society* 49:9; 739-752.

Pulkki, V. 2001 "Lacalisation of Amplitude-Panned Virtual Sources, Part 2: Two- and Three-Dimensional Panning." *Journal of the Audio Engineering Society* 49:9; 753-767.

Roads, C. Ed. 1996 *The Computer Music Tutorial*. Cambridge, Massachusetts: MIT Press.

Rumsey, F. 1998 "Stereo" in Talbot-Smith, M. Ed. 1998. *Audio Engineer's Reference Book*. Oxford, U.K: Focal Press.

Stevenson, I. 2002 *DREAM Console User Manual*. Sydney, Australia: Fairlight ESP Pty Ltd.

Stevenson, I. 2001. *VCS4 Virtual Control System User Manual*. Sydney, Australia: DSP Media Pty Ltd.

Symbolic Sounds 2001. *Kyma 5.0 Walkthrough* Illinois, USA: Symbolic Sound Corporation

Vercoe,B. 1992. *The Public CSound Reference Manual Canonical Version V4.01*. Cambridge, Massachusetts: Media Lab MIT.

Wishart, *T*. 1996. *On Sonic Art*. Amsterdam, Netherlands: Harwood Academic Publishers.

Mathews, M. V. 1969. *The Technology of Computer Music*. Cambridge, Massachusetts: MIT Press.

The RoboSax Project (1991-2001): forms of performer/machine interaction in works by Jonathan Mustard and Lindsay Vickery

Lindsay R. Vickery

Music Department, Western Australian Academy of Performing Arts,
Edith Cowan University
email: l.vickery@cowan.edu.au

Abstract

Since 1990 Jonathan Mustard and I have been creating a body of work for solo wind instrument (acoustic or electronic) and interactive electronics. This paper gives an overview of these works and their exploration of interaction and control issues between the performer and machine. It discusses the evolution of the continuing project through the use of software and hardware from the Yamaha TX81z, through MAX to MAX/MSP.

1 Introduction

The 'works for solo instrument and interactive electronics' genre probably begins with Gordon Mumma's work for horn and electronics, *Hornpipe* (1967) which used 'microphones for "listening" to sounds made by the horn as well as for analyzing the acoustical resonance of the space' (Winkler 1998 p.12). It was an ambitious start for a genre that has had to overcome many challenges in order to develop a diverse, expressive and meaningful repertoire.

The attractions of the genre for composers are extremely varied, but some principal motives include: the desire to extend or transform traditional instruments; the ability to create textural and compositional complexity from a single live source; the non-linear possibilities of interactive work; and the possibility of creating a 'custom' improvising partner

A number of different systems have sprung up to deal generically with the types of interactivity composers may wish to employ. The best known of these began to be commercially available in the early nineties: Interactor (Subotnik/Buchla/Coniglio), Hyperinstruments (Machover), Cypher (Rowe), Kyma (Scaletti) and Max (Puckett/Zicarelli).

The RoboSax Project began in 1991, and has continued over the last ten years as a forum for composer Jonathan Mustard and myself to explore the possibilities available for solo wind player and electronics. The work has in general been sustained through personal resources and has necessarily developed a pragmatic approach to materials and technology. This has resulted in a growing concentration on software rather than hardware solutions to musical problems.

The RoboSax Series of works encompass a variety of combinations. It is interesting to note the increasing tendency toward extremely straightforward set-ups:

The RoboSax Project:

1. Mustard: *RoboSax I* (1991) Yamaha WX Series Windcontroller, 2 Yamaha TX81z Tone Generators and Sequencer

2. Vickery: *DiceGame* (1991) Clarinet, Microphone and Ensoniq DP/4 Effects Unit

3. Vickery: *27Matrix* (1995) Yamaha WX Series Windcontroller, Yamaha TG500 Tone Generator and Macintosh [MAX]

4. Mustard: *Robosax III* † (1996) Soprano Saxophone, MIDI Footcontroller, General MIDI Synthesizer and Macintosh [MAX]

5. Mustard: *RoboSax IV 'The Arsonist'* (2000) Alto Saxophone, Microphone and Power Macintosh [MAX/MSP]

6. Vickery: *<as viewed from above>* (2000/01) Soprano Saxophone, Microphone and Power Macintosh [MAX/MSP]

† RoboSax II (1992) Yamaha WX Series Windcontroller, 2 Yamaha TX81z Tone Generators and Sequencer is currently withdrawn

In common to the six pieces is a desire to create sonic environments of textural and compositional complexity that can be generated in real-time through the actions of a single performer. The instrumentation of the successive works shows a continuous progression through available technologies from programmable tone generators controlled via MIDI to digital audio controlled by analogue audio.

It is possible to detect differing principal preoccupations underlying the works. In a broad sense these preoccupations embody three contrasting paradigms of musical organization Dramatic; Structural; and Symbolic. Mustard's contributions, under the generic designation *RoboSax*, explore elements of interaction and control between the performer and machine. Their effectiveness is due principally to intrinsic dramatic tensions set up by the composer between the live performer and the interactive machine components.

My works *DiceGame* and *27Matrix* both focus on 'real-time' composition possibilities, in particular the generation of algorithmic structures via improvisation. Their success resides in their ability to mediate between fixed 'pre-compositional' processes and free improvised performance. My most recent work *<as viewed from above>* in contrast seeks to establish a new structural model based on formalizing certain characteristics of psychological obsession in a non-linear interactive format. Its relies on the symbolic quality of music to mirror certain patterns of thought in a way that is, if not consciously recognizable, at least resonant with its audience.

2 Mustard's RoboSax works

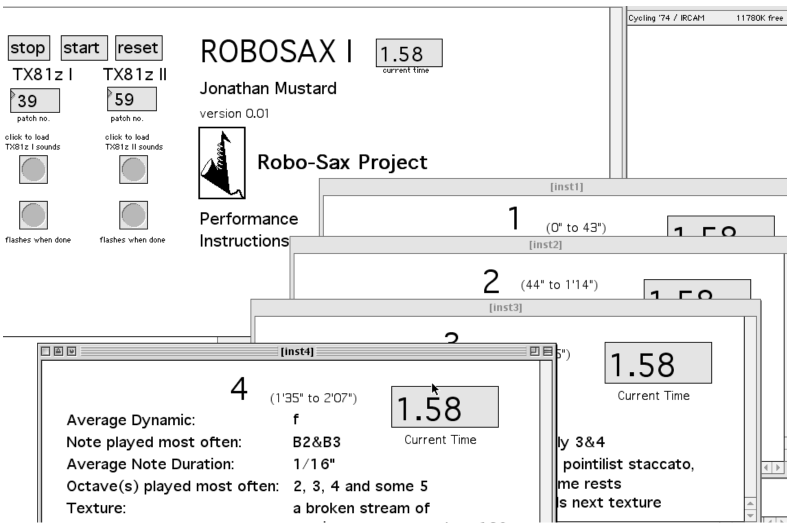


Figure 1:Jonathan Mustard's *RoboSax I* (Max Version) (Mustard and Vickery 1996)

In his programme notes for *RoboSax IV* Mustard states his credo for the series as a whole:

They are the protagonists in the drama or perhaps the dilemma of 20th / 21st Century existence where the interaction of humans and machines (computers in particular) is ubiquitous and the question of which element in the equation has the control at any one time becomes ambiguous. (Mustard 2000)

His choice of title is typical of his oeuvre as a whole, reflecting in equal measure a strong engagement with social issues and self-deprecating humour. "RoboSax" alludes to Paul Verhoven's 1987 film RoboCop and the spin-off series that followed it. Mustard both references and satirizes his own interest in the film's central theme - Human's increasing integration with machinery. At the same time he draws attention to the parallels between composers' and film makers' tendencies to create series of related works under the same title, with its undercurrent of cynical exploitation of success and implication of mechanical mass-production.

2.1 RoboSax I

Mustard's first work in the series places the computer firmly in control. The piece took up the opportunity of exploring the possibilities of the Yamaha WX series of Windcontrollers (WX7 [1988] and WX11 [1989]) that had recently emerged on the market. The performer plays a WX11 MIDI windcontroller while a computer changes patches on a synthesizer that can between them generate up to 32 notes for each note triggered by the performer. In the original version the WX performer with a score and a stopwatch attempted to carry out 'meta-compositional' instructions within particular timeframes.

The catch in carrying out these instructions is that the stream of MIDI information generated by the WX instrument is realized through the filter of a sequencer's continuous alteration of synthesizer patches - over which the performer has no control. Mustard's ingeniously programmed patches on the TX81z add to the computer's ascendancy, confounding the performer's expectations by reordering and/or harmonizing the pitch of each key or changing sounds with each new re-articulation.

The restraint of dividing control of various musical parameters between performers and machine was being independently explored in the same year by David Jaffe and Andrew Schloss in their work *Wildlife* (1991) (Rowe 1993 p.85). Although more complex in its capabitities, *Wildlife* required a NeXtstation, a Macintosh IIci, a RadioDrum(baton), Zeta Violin, a Yamaha TG77 and two sound cards.

The original version of *RoboSax I* called for an Atari computer to send patch changes to the tone generators. In 1996 I created a MAX version of the piece combining the score, stopwatch and Atari's functions together. In this version the performer is presented with the instructions at the appropriate time via MAX patcher windows that open synchronously with the sequence.

But *RoboSax I* 's subtitle (The Strathfield Massacre) reflects a more profound level in which the work attempts to form some kind of response to the arbitrariness and horror of this infamous incident in which a gunman went on the rampage in one of Sydney’s outer Western suburbs killing 13 people including himself and injuring many others.

At the core of the work's formal structure then is the knot of our incomprehension of such events: the performer placed in the position of a powerless observer who is required to play from a list of verbal instructions without really knowing what is going to be the end result.

2.2 RoboSax III

In his *Robosax III* Mustard turned this model on its head, placing the performer (this time on an acoustic instrument) directly in control of most of the musical parameters via MIDI Foot-pedal Controller. In *RoboSax I* the performer is continuously required to make performance decisions based on (often incorrect) expectations of what the aural result might be. In contrast *RoboSax III* puts the performer in the taxing position of controlling not just their own instrument, but up to seven other virtual instruments via the foot-pedal simultaneously.

The Yamaha MFC-10 foot-pedal used for this piece is capable of sending 128 program changes (via a bank of ten buttons at a time), as well as continuous controller information from up to 5 "volume" style (CV) pedals. Program changes sent from the MIDI foot-pedal control harmonic, timbral, tempo and several other MIDI controller parameters.

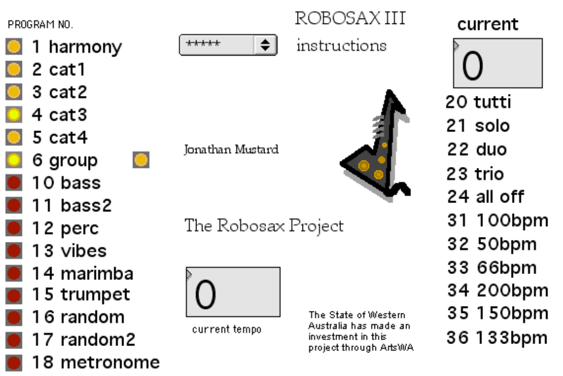


Figure 2. Mustard's RoboSax III

The first six program changes send changes to the harmonic and textural composition of the piece; program changes 10 through 17 make timbral changes (i.e. choose different instrumentations); 18 turns on and off the `metro` object - effectively the engine that generates MIDI events in RoboSax III; 20-26 create various instrumental (timbral) combinations; and 31-36 control the tempo of the metronome and therefore of the whole piece. Additional parameters not visible on this front panel allow continuous tempo, pitch-bend and panning control via additional foot-pedals.

A problem central to all electronic music is the issue of obsolete/superseded equipment rendering works impossible or difficult to reproduce in the future. *RoboSax III* attempts to partially redress this situation by standardizing the equipment required to perform the work: a MIDI foot-pedal, a Macintosh computer capable of running MAXplay and a general MIDI synthesizer. Other versions of the work are of course possible using any MIDI compatible tone generator or sampler, however Mustard's General MIDI format establishes a minimum technical requirement that is readily realizable.

Compositionally *RoboSax III* attempts to create something akin to a virtual 'jazz combo' to accompany the soloist. This is reflected in the choice of typical jazz instruments as the principal timbres (ie Muted Trumpet, Vibraphone, Double Bass, Kit...) and also in more general characteristics such as the texture and syncopated dynamics.

RoboSax III belongs to the 'improvising partner' paradigm - creating a virtual combo under control of the soloist. Mustard's aims mirror those of improviser Bruno Spoerri. '*The important thing for me was to have partner who threw balls at me, who gave me a reason to react in a certain way, but would react in a logical way.* ' *(Bruno Spoerri in Chadabe 1997 p.322)*

Below the 'pseudo-combo' surface however *RoboSax III* employs many techniques typical of Mustard's notated linear works like *Nine Realizations of a Brass Guerilla* (1988) or the *Automaton* Series (1989-90). In the MAX software a central `metro` object sends 'clock' information in the form of a continuous stream of pulses. Mustard utilizes a form of restricted random choice to disregard certain pulses in order creating gaps in the stream and therefore rhythmic variety. Additionally, in some cases rather than deleting the pulses, he reroutes them to different instruments creating an interplay of interlocking rhythms (hoketus) typical of his music. Pitch aggregates used to create melodic and harmonic material are chosen from tables along similar principles. One of the most interesting features of *RoboSax III* is two 'Random' buttons (Program change 16 and 17), both capable of sending all of the patch's variable parameters into overdrive – rhythms, voices and instruments proliferate exponentially in a disturbing crescendo of activity.

2.3 RoboSax IV

In *RoboSax IV 'The Arsonist'* Mustard establishes a middle ground in the struggle between performer and computer control. Again the soloist performs on an acoustic instrument, the Alto Saxophone, but this time reading a relatively conventional notated score. The computer (running MAX/MSP) also performs its samples according to a linear script. What differentiates this work from the traditional 'soloist plus tape' format is that Mustard sets up a feedback loop whereby the live performance and the computer performance modulate one another. The audio via microphone from the live performance modulates both the characteristics of the computer's samples and its own input via filtering to create a third stream. The resultant modulations of the computer's audio in turn influence the live performance.

The same motif of modulation occurs in the score for *RoboSax IV* where Mustard has the soloist modulating their own performance via a complex system of cross-fingerings that create pitch and timbral variation. Elements of chance resulting from the use of some graphic notation, relatively free timing in the score and use of random number generators in the script enliven the work, elevating it into a region somewhere between freedom and control.

3 Vickery's RoboSax works

3.1 DiceGame

My initial interest in this field began with the concept of live processing of an analogue instrument. *DiceGame,* like Mustard's *RoboSax IV,* is a (mostly) traditionally notated work for clarinet with an added electronic component, in this case the Ensoniq DP/4 Effects Processor. *DiceGame* belongs to a series of my works dating from the early 90s, each using a short series of numbers (cypher) as a means for generating all musical parameters from the micro- to macro-scale. The work's fractal nature is augmented, and to some extent its definition is enhanced, by the addition of electronic processing.

The interactivity in *DiceGame* is limited by the available interfaces between the machine and the performer. By far the most straightforward control is achieved through the use of a foot-pedal to step through processor settings, and CV pedal to make continuous changes in certain parameters. Despite the fact that in 1995 interactive signal processing was still in relatively early stages, some greater level of interactivity is achieved through simple use of signal amplitude measurement. The most obvious example is the use of Gate processors to limit, depending on their volume, what signals are passed on to other effect patches. This approach is somewhat analogous to that used by Todd Winkler in his work *Snake Charmer* (1991) where the modulation speed of a chorus effect is controlled by the changing dynamic level of the clarinet soloist.(Winkler 1998 p.249)

The DP/4 part consists of 14 separate patches that are ordered into a 24 patch sequence. The overall form of the work is of nine main sections, consisting of five contrasting types of material. The sections are in the order 532214451: and are relatively easy to hear as they have gradually diminishing pitch resources: section 5 uses a 15-pitch group, 4 uses 12, 3 - 9, 2 - 3 and Section 1 uses 1 pitch. During the repeated sections (there are adjacent 2 and 4 sections) the electronic processes help to define the structure through processing changes. For example the second '2' section is marked by a change in harmonization.

The processes involved are used to reinforce the work's structure - emphasizing the changes between sections and, to some degree, the sub-sections within them. The effects can essentially be broken down into combinations of pitch and time distortions of the live signal.

Some examples will suffice to illustrate:

• *Velocity Octaver* is the principal effect during the opening section '5'. It processes signals above a certain volume (db) adding four delayed semi-quavers each harmonized at the octave below the original note in tempo MM.= 90. Signals below that volume pass through with no effect. The aural result is that accented notes are doubled with a lower octave, which is repeated, as four delayed semi-quavers.

• *Rainbody* delays and pitch-shifts notes downwards by small amounts (in inharmonic intervals) with the aural result being a proliferation of asynchronous descending micro-tonal lines similar to Ligeti Micro-Polyphony.

• The *Tempo Digital Delay* delays the audio signal to four different degrees. Each is then panned hard right, hard left, middle right and middle left in the stereo spectrum. The length of the delay is determined by the position of the control voltage (volume) pedal notated in the part. The aural result is a five-part canon (including the live signal) panned across the stereo spectrum.

• '*Ascending Whole-tones'* delays notes by a semi-quaver (within the tempo) with four repeats. Each delayed note is a whole-tone higher than the last, creating a four note ascending whole-tone scale.

3.2 27matrix

My second work for solo wind and electronics was *27matrix*. It is an interactive improvisation environment written in MAX. An improvising soloist (again a WX series Windcontroller) provides the raw MIDI data that is transformed by MAX in real time into various structures.

27matrix combined live improvisation with the sort of formal processes that I had been using in my music between 1990 and 1995. These structures are rigorously governed by a principle of self-similarity - all material created by the computer draws on a nine-digit cypher (the same one used in *DiceGame*) as its generating kernel.

This type of formal structure originating in the works of Webern has found particular resonance in the field of electronic music where simultaneous control is applicable to an ever-increasing range of parameters – even to the level of the shapes of soundwaves themselves (Yadegari 1991). In interactive computer music one of the best early examples is Gary Lee Nelson's *Fractal Mountains* (1988-89) which maps the performer's notes via MIDI into a series of graphs that control a wide variety of musical parameters in real-time (Nelson 1994). In *27matrix* the fractal algorithm is predetermined, but give rise to a diverse range of musical textures.

Like *DiceGame, 27matrix* comprises five basic environments that employ a number of different strategies to transform the live MIDI input from the Windcontroller. Each reiterates the cypher structure in the pitch and/or rhythm domain. They are discussed in the order in which they are heard.

Environment Five: This environment is an inversion of typical jazz improvisatory practice: pitches from the performer's improvisation are fed into the accompaniment and successive notes are gradually spread out over five octaves. The rhythmic structure of the accompaniment itself is constructed of canonic versions of the cypher rhythm.

Environment Three: The performer's incoming notes are repeated between two and five times depending on their velocity: the higher the velocity the more repetitions. The frequency of the repetitions is dependant on the octave in which they are played: the higher the register the faster the repetitions.

Environment Two: The first note played in this patch is captured and played in the original and retrograde version of the cypher rhythm. The performer’s improvisation is harmonized by between one and nine pitches (dependant on the note’s velocity). Every 27th note the soloist plays triggers a rapidly repeated note to travel across the stereo field.

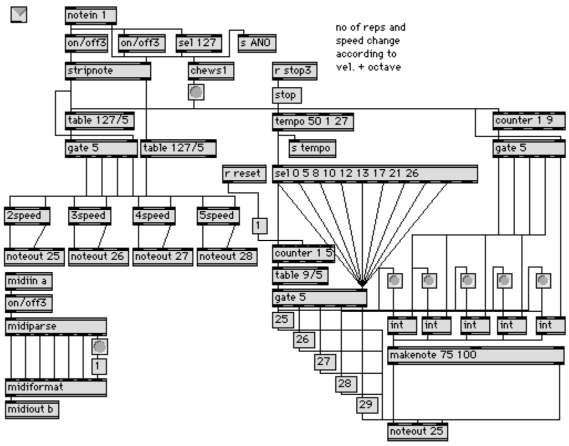


Figure 4. *27matrix* environment 3

Environment One: The first note the soloist plays is captured harmonized and repeated in the cypher rhythm. On the cypher's completion another rapidly repeated chord travels across the stereo spectrum.

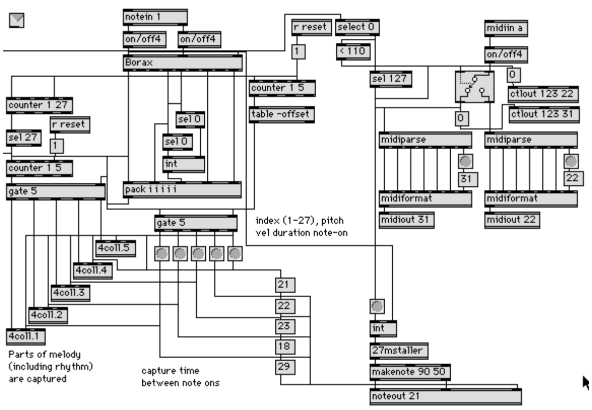


Figure 5. *27matrix* environment 4

Environment Four: The soloist's improvisations are captured, and played back on different instruments, in different registers and at different speeds surrounding the performer with different versions of their material. The solo instrument is a flute-like sound if above a certain velocity and a gamelan type sound if below. Changes of instrument create 'stuck notes' that can only be turned off by playing a full velocity note. A full velocity note also triggers the cypher rhythm to be played.

3.3 <as viewed from above>

<as viewed from above> is written in MAX/MSP and like Mustard's *RoboSax IV* 'listens' to the live performance of an acoustic instrument as a trigger for proceeding through the samples of a text. An increasing involvement with texts in my work since 1997 has led to a desire to expand the techniques developed in *27matrix* and to develop models of interaction that are based more psychological considerations.

Whereas Mustard's *RoboSax* works focus primarily on the drama inherent in performing interactive music, and *DiceGame* and *27matrix* are both concerned with the real-time generation of fractal structures, *<as viewed>* springs from a desire to re-evaluate the implications and possibilities of computer interaction itself. It was evident that there was a degree of tension between the highly deterministic and linear nature of my earlier works' fractal structure and the inherently non-linear potentialities of real-time interactivity. Traditional musical formal models had been developed, in the main, for a linear environment where complex structures could be coordinated only through the synchronization of multiple performers to a single linear temporal framework. The advent of interactive computer technology capable of creating complex non-linear structures has brought about a need both to re-assess these kinds of formal structures and consider the potential for new ones.

<as viewed> is an initial attempt at developing such a new formal structure. It draws partly on developments in literature such as the ‘hypertext rhizome’. Defined by Slovenian philosopher Slavoj Zizek as being a structure that ‘*does not privilege any*

order of reading or interpretation; there is no ultimate overview of "cognitive mapping", no possibility to unify the dispersed fragments in a coherent encompassing narrative framework. (Zizek, 2000 pp.37)

The hypertext rhizome has been developed into a powerful dramatic paradigm by MIT professor Janet Murray. She coined the term 'Violence Hub' to designate hypertextural works in which a central event is examined from different perspectives.

The proliferation of interconnected files is an attempt to answer the perennial and ultimately unanswerable question of why this incident happened. These violence hub stories do not have a single solution like the adventure maze or a refusal of solution like post-modern stories; instead they combine a clear sense of story structure with a multiplicity of meaningful plots. The navigation of the labyrinth is like pacing the floor; a physical manifestation of trying to come to terms with the trauma; it represents the mind's repeated efforts to keep returning to a shocking event in an effort to absorb it and finally, get past it. (Murray 1997 p. 135-6)

Zizek identifies the potency of this novel formal structure in Lacanian terms as referring to the '*trauma of some impossible Real which forever resists its symbolization (all these narratives are ultimately just so many failures to cope with this trauma*) (Zizek 2000 p.38). <*as viewed*> attempts to sonically reproduce a formal structure of this type.

At the heart of <*as viewed*> is a short text. The text is structured in a similar way to Murray's 'Violence Hub': it circles a central irreducible problem. I would, however, like to suggest a more generalized term for this kind of formal structure - 'Event Hub' – in recognition of the wide variety of circumstances, not merely violent ones, that one is cable of obsessing over.

Each line of text is recorded as a separate sound file. The computer can choose to replay and manipulate any previously chosen sound file of text, but is constantly narrowing its own number of text choices. In effect the patch left to its own devices will choose to 'obsess' over - in this case repeating and deforming - an ever diminishing group of samples. The live performance 'distracts' this process and forces it to act upon new material until all of the samples have been exhausted.

The MAX/MSP pitch mapping object `fiddle~` (Puckett et al 1998) forms the bridge between the live performer and the computer. This object is used to approximately map the current frequency and amplitude of sounds from the performer and also makes a guess as to the beginnings of phrases based on amplitude changes. Computer pitch-tracking/score following was one of the tasks MAX was created in 1985 to solve (Chadabe p.183). The effectiveness of this technology still remains an issue for interactive computer music. The accuracy of the CPU intensive object `fiddle~` for example is a trade-off with the number of other objects/processes that can be included in the patch. For this reason <*as viewed*> uses `fiddle~` only to obtain generalized contours of the live performance.

This information is processed by <*as viewed*>in three distinct layers. Layer one cues text samples based on the beginnings of the live performer's phrases. It also manages the samples so that the texts do not play simultaneously and have appropriate pauses between groups of lines of text. Layer two manipulates the samples that have been played up until that point. It uses frequency and amplitude information as well as information pertaining to the amount of activity in the live part to change playback speed, assign loops and loop lengths and pan the samples. The final layer creates an overall mix between the live performance, the expanding text and the manipulated text and processes the result using comb filters and reverbs.

<*as viewed*> responds to a need for new structural models to better take advantage of developments in interactive technology and in particular their non-linear possibilities. Its response is lyrical/abstracted rather than based on any formal linking of 'real' information about the construction of human memory (such as it is understood).

3 Conclusion

Despite their surface similarities the six works in the *RoboSax Project* make use of a great variety of different paradigms of interaction between soloist and electronic sound. In these works the correlation between the form, content and technical means appear to be closely linked. The *RoboSax Project*, developed in a relatively isolated non-institutional environment, are good examples of the potential for the creation of engaging work with minimal resources.

References

Chadabe, J., 1997. *Electric Sound: The Past and Promise of Electronic Music*. Upper Saddle River, New Jersey. Prentice Hall

Murray, J.H., 1997. *Hamlet on the Holodeck: The Future of Narative in Cyberspace*. Cambridge Massachusetts, MIT Press.

Mustard, J.A., 2000. *Program notes for RoboSax IV*
Personal correspondence with the author

Nelson, G. L., 1994. *Real Time Transformation of Musical Material with Fractal Algorithms.* http://timara.con.oberlin.edu/~glnelson/gnelson.htm

Rowe, R.,1993. Interactive Music Systems: Machine Listening and Composing. Cambridge Massachusetts, MIT Press.

Puckett, M., T. Apel, and D. Zicarelli, 1998. Realitme audio analysis tools for Pd and MSP. *Proceedings of the International Computer Music Conference.* International Computer Music Association, *1998* pp. 109-112

Winkler, T., 1998. *Composing Interactive Music: techniques and ideas using Max.* Cambridge Massachusetts, MIT Press.

Yadegari, S.D., 1991. "Using Self-Similarity for Sound/Music Synthesis." *Proceedings of the International Computer Music Conference.* International Computer Music Association, pp. 423-424.

Zizek, S., 2000. *The Art of the Ridiculous Sublime*. Seattle Washington, Walter Chapin Centre for the Humanities: Occaisional Papers 1, University of Washington Press.

Report on the Australian Sound Design Project

Dr Ros Bandt with Iain Mott

The Australian Sound Design Project is a nationwide initiative to document on the web the works of Australian sound designers who work in public space. It was intitiated by Dr Ros Bandt through an ARC large grant and is hosted by the interdisciplinary Australian Centre at the University of Melbourne where she is honorary senior fellow. The research assistants, originally Garth Paine, and now Iain Mott facilitate the project, managing the database and web site programming and design.

The aims and objectives of the Australian Sound Design Project are published on the website, http://www.sounddesign.unimelb.edu.au on the ABOUT page and more fully in the paper given at the Waveform Conference at the launch of this project last year. Designing public acoustic space: Australian sound designs, a database and website, for a more considered acoustic environment is published on the website under papers.

I reiterate an invitation to all sound designers to be part of this project, visit the site and complete the online form and become part of the first Australia wide network in sound design. It is also a free multi-media publishing opportunity. The questionnaire and copyright deed must be filled out and there is an electronic form on line. Hard copies can also be sent.

The ASDP has become a serious resource as well as a source of documentation. The powerful database of over 500 people has a search facility by designer and by sound designs or site or date. It includes all those practitioners who work in the field of designing sound in public space, artists, curators, electronic programmers, software and audio engineers, and all forms of acoustic designers. Two works from each artist have been requested and these are being designed for the web at the present time. Over 50 artists have responded with original materials so far. As well as publishing the sound designs in multi media formats of images, video, sound, text and graphics, it also contains a powerful list of research tools, including important unpublished papers, (Carter, Belfrage, Atwell), a list of Australian and international links, links about sound, links to the sites of sound artists who have their own sites, a list of software and on-line tools, and a comprehensive bibliography. Since its inception the site has received 35,690 hits.

On a technical level, the project is currently adding design and designer information to a database that can generate web pages with sophisticated cross-referencing and browsing capacity. The database is called the Online Heritage Resource Manager developed by the Australian Science and Technology Heritage Centre (www.austehc.unimelb.edu.au). Browsing will be performed by alphabetical listings of entries as well as by category, for example searching by designer, curator, host or funding body. In addition, users will be able to access information through two separate search mechanisms. The first utilises software called ht://Dig . This indexes web files and enables simple keyword searches. The second makes use of PHP scripting and a server-side database containing information duplicated from the primary database maintained at the Australian Centre. This system allows for advanced searches utilising individual fields within the database. Individuals will be able to search for entries using a selection of criteria.

The ASDP has established the first comprehensive resource about sound in Australia and is a powerful research, publishing and networking facility. Be part of it by filling out the on line form:

http://www.sounddesign.unimelb.edu.au/site
/InvitationLetter.html

Biographies

Dr Ros Bandt, ARC director of the Australian Sound Design Project, a nation wide data base and website, author of 3 books on spatial sound design in Australian artworks and internationally acclaimed sound artist, composer and performer.

Iain Mott, is a Melbourne based sound artist.He has exhibited nationally in Australia and internationally at exhibitions including the Ars Electronica. Iain was the Artist in Residence at the CSIRO Mathematical and Informations Sciences.

Developing and Composing With Scales based on Recurrent Sequences

Warren Burt
freelance composer
PO Box 2154, St. Kilda West, VIC 3182
email: waburt@melbourne.dialix.com.au

Abstract:
Following up on suggestions by Ervin Wilson, who showed that an infinity of recurrent sequences can be obtained from taking the sums of the diagonals of Pingala s Meru Prastala, known in the West as Pascal s Triangle, the author has generated a number of scales using these sequences. Among the sequences used are familiar ones such as the Fibonacci series, and other more exotic sequences. Just intonation scales made directly from the proportions of the number series are made, as well as Pythagorean scales made from taking the Means the series tend to, and treating them as generating intervals for scales. The resulting scales are then divided into pseudo -White Key and -Black Key subsets using the Moments of Symmetry technique of Wilson and Chalmers. Examples of these scales, and the harmonies implied by them are heard as musical examples. Compositional work with these scales is then discussed, including using the intervals of the scales not only as generators for musical harmony, but also as generators for rhythm, timbre and spatialisation. A discussion of the spatial nature of a sound as an emergent property of its harmonic content concludes the paper

Keywords: Sonification of Non-Musical Data, Mapping Strategies, Harmonic Design, Microtonal Tunings as Generators of Sonic Space

For some time now, Ervin Wilson has been suggesting that I would be interested in investigating the properties of scales made with recurrent sequences. His 1993 article, The Scales of Mt. Meru (Wilson, 1993a), following up on work by Thomas M. Green (Green, 1968) and A. N. Singh (Singh, 1936), shows how an infinite series of recurrent sequences can be derived by taking the sums of different diagonals of Pingala s Meru Prastara, otherwise known as Pascal s triangle. This diagram was first described by the Hindu mathematician Pingala as early as 200 BCE, and other triangular number diagrams existed in Hindu mathematics as early as 2000 BCE, according to Ernest McClain. (McClain, 1976). Ervin Wilson has suggested that the recurrent sequences found in the Meru Prastara could be useful as a source for musical scales.

A recurrent sequence is any sequence of numbers generated by following a rule for adding previous elements of the sequence to get the next one. For example, the fibonacci series, one of the most famous of these, follows the formula An = An-2 + An-1. That is, each new number An is the sum of the two previous elements in the series, the number just before it (An-1) and the number two elements before it (An-2). This yields the famous sequence 1 1 2 3 5 8 13 21 34 55 89 etc. In his 1993 article, Wilson enumerated the first 192 recurrent sequences that could be yielded by summing the elements found on different diagonals on Pascals triangle. A few examples will suffice - the first five sequences are:

An = An-2 + An-1 (An/An-1 converges on 1.618033989...)
Bn = Bn-3 + Bn-1 (Bn/Bn-1 converges on 1.465571232...)
Cn = Cn-3 + Cn-2 (Cn/Cn-1 converges on 1.324717957...)
Dn = Dn-4 + Dn-1 (Dn/Dn-1 converges on 1.380277569...)
En = En-4 + En-3 (En/En-1 converges on 1.220744085...)

In each of these sequences, the farther one goes out them, the more the ratio between any two elements comes closer to a mean. The means that each of the above five series approach are given in the parentheses after the formula for that series.

While each of these series is interesting in its own right, the problem of how to turn them into scales puzzled me for some time. At first I thought I might just take the interval between each successive two elements of the series and treat it as a pitch in a just intonation scale. So, for example, for the fibonacci series, that would mean a scale consisting of the ratios 2/1, 3/2, 5/3, 8/5, 13/8, etc. Putting these ratios and their inversions (4/3, 6/5, 5/4, 16/13, etc.) into Emanuel Op de Coul s Scala program yielded the following scale, which though it might be charming, was not too useful. The preponderance of intervals around 833 cents and 366 cents can be explained by the fact that as you go further and further out the series, the ratio between each elements tends towards 1.618etc....which, expressed as the ratio of two musical frequencies, is an interval of around 833 cents. The inversion of this interval gives the interval of around 366 cents. So, this clearly wasn t the way to go.

Scale made by taking each fibonacci pair as a ratio expressing a scale degree.

0:	1/1	0.000
1:	6/5	315.641
2:	16/13	359.472
3:	21/17	365.825
4:	110/89	366.752
5:	288/233	366.887
6:	89/72	366.970
7:	68/55	367.324
8:	26/21	369.747
9:	5/4	386.314
10:	4/3	498.045
11:	3/2	701.955
12:	8/5	813.686
13:	21/13	830.253
14:	55/34	832.676
15:	144/89	833.030
16:	233/144	833.113
17:	89/55	833.248
18:	34/21	834.175
19:	13/8	840.528
20:	5/3	884.359
21:	2/1	1200.000

Then I thought, What about using the ratios of successive integers as intervals and pile them up, like a chain of 3/2s, but here each element in the chain is different? So here you would have 3/2 + 5/3 + 8/5 + 13/21 etc. as intervals and then you could reduce your chain of intervals into an octave. I did this with the first 12 elements of the fibonacci series, and here is what I got:

Degree	Ratio	Cents	Element of series
0:	1/1	0.000	(1)(2)(8)
1:	17/16	104.955	(34)
2:	9/8	203.910	(144)
3:	305/256	303.199	(610)
4:	5/4	386.314	(5)
5:	21/16	470.781	(21)
6:	89/64	570.880	(89)
7:	377/256	670.105	(377)
8:	3/2	701.955	(3)
9:	13/8	840.528	(13)
10:	55/32	937.632	(55)
11:	233/128	1037.023	(233)
12:	2/1	1200.000 octave	

This looked more promising. I also noticed that, due to the nature of interval addition, where ratios multiply against each other, that the numerators and denominators cancelled out, making the scale, in effect, a scale of harmonics numbered according to each of the elements of the fibonacci series.

For example: The first 15 elements of the fibonacci series are 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610. Looking at our scale, we see 1/1 (for the ones) 2/1 (for the octave) 3/2 (3rd harmonic), 5/4 (5th harmonic expressed as a ratio within an octave), 8/1 = 2/1 = 1/1 (8th harmonic), 13/8 (13th harmonic), 21/16, 34/32=17/16 (34th harmonic expressed as a ratio within an octave) 55/32, 89/64, 144/128 = 72/64 = 36/32 = 9/8 (144th harmonic is an octave of the 9th harmonic...), etc.

So I then decided to take this scale out to 23 degrees and this is what I got. The last column is the number of the number series that the pitch is expressing. Another way of thinking about this is that the number of the series N becomes the Nth harmonic above the fundamental. (For reference, the first 23 unique elements of the fibonacci series are contained within: 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987 1597 2584 4181 6765 10946 17711 28657 46368 75025 121393)

SCALE A (Just) :

0:	1/1	0.000	(1,2,8)
1:	4181/4096	35.559	(4181)
2:	17/16	104.955	(34)
3:	17711/16384	134.830	(17711)
4:	9/8	203.910	(144)
5:	75025/65536	234.101	(75025)
6:	305/256	303.199	(610)
7:	5/4	386.314	(5)
8:	323/256	402.468	(2584)
9:	21/16	470.781	(21)
10:	5473/4096	501.739	(10946)
11:	89/64	570.880	(89)
12:	1449/1024	601.010	(46368)
13:	377/256	670.105	(377)
14:	3/2	701.955	(3)
15:	1597/1024	769.378	(1597)
16:	13/8	840.528	(13)
17:	6765/4096	868.649	(6765)
18:	55/32	937.632	(55)
19:	28657/16384	967.920	(28657)
20:	233/128	1037.023	(233)
21:	121393/65536	1067.191	(121393)
22:	987/512	1136.288	(987)
23:	2/1	1200.000	

This seemed like a scale with many useable properties. In fact, Los Angeles composer Stephen James Taylor has been using it for many years. I wondered what the sequence of intervals made by the process of generating this scale would look like. I knew that it would start with a sequence of pure small-integer just ratios (3/2, 5/3, 8/5, etc), and that the intervals would get closer and closer to 833 cents, and that it would generate all the pitches in my scale, and that it would lead on and on to an infinite chain of intervals, but I was curious to see how the details of this sequence worked out:

Fibonacci Scale in Interval Order (first 10 intervals)

Int:		3/2	5/3	8/5	13/8	21/13
Pitch:	1/1	3/2	5/4	2/1	13/8	21/16
Cents:	0	702	386	1200	841	471
Scale Nr.	0	14	7	0 (23)	16	9

Int:	34/21	55/34	89/55	144/89	233/144
Pitch:	17/16	55/32	89/64	9/8	233/128
Cents:	105	938	571	204	1037
SclNr:.	2	18	11	4	20

Extending this out through the entire 23 note scale (listing only newly generated non-redundant pitches like the second 2/1 (1/1) in the above example) gave the following note order:

0 14 7 16 9 2 18 11 4 20 13 6 22 15 8 1 17 10 3 19 12 5 21

Looking at the intervals of the fibonacci series themselves, we can see how they do, indeed rapidly converge on the mean of 1.618033989...., which, expressed as a ratio over one, gives a value of 833.090297...cents:

Ratio:	3/2	5/3	8/5	13/8	21/13
Cents:	701.06	884.36	813.69	840.53	830.253

Ratio:	34/21	55/34	89/55	144/89	233/144
Cents:	834.18	832.60	833.23	833.11	833.08

This led me to speculate on what an equal tempered, or at least an equal-interval scale made up of intervals of 1.618...(approximately 833.090297c) would be. That is, if you added up, in a Pythagorean manner, 833.090297c intervals, what sort of a scale would you get? You would, in fact, get a scale that would be, for all practical purposes, identical to one made from the higher elements of the fibonacci series. The charm of the Just scale generated above, however, is that it DOES contain intervals which are more or less distant from the mean. The charm of the scale made from the upper elements of the series would be that it DOESN T contain intervals which deviate from the mean. You pays your money, and you takes your choice. Taking 23 intervals, to get a 23 note scale, like our fibonacci scale, yielded the following:

SCALE A (PYTHAGOREAN):

a chain of 23 intervals of 1.618..../1.000 made into cents

0:	1/1	0.000
1:	30.174 cents	30.174
2:	99.271 cents	99.271
3:	129.445 cents	129.445
4:	198.542 cents	198.542
5:	228.716 cents	228.716
6:	297.813 cents	297.813
7:	327.987 cents	327.987
8:	397.084 cents	397.084
9:	466.181 cents	466.181
10:	496.354 cents	496.354
11:	565.451 cents	565.451
12:	595.625 cents	595.625
13:	664.722 cents	664.722
14:	694.896 cents	694.896
15:	763.993 cent	763.993
16:	833.090 cents	833.090
17:	863.264 cents	863.264
18:	932.361 cents	932.361
19:	962.535 cents	962.535
20:	1031.632 cents	1031.632
21:	1061.806 cents	1061.806
22:	1130.903 cents	1130.903
23:	2/1	1200.000

The elements of the scale, taken in the order they were generated, were

0 16 9 2 18 11 4 20 13 6 22 15 8 1 17 10 3 19 12 5 21 14 7

So, taking the notes in the order they were generated gave a scale of 833.090 cent intervals.

I then began to realize, that as these were infinite series, they could indeed be extended infinitely. Was there any way, I wondered, to know when to stop , in order to get useable scales that had some sort of property of recognizability or coherence. I put this question to Erv Wilson, and he sent me a diagram of a metallophone he designed using the 23 note just scale given above, grouped into interlocking sets of 13 + 10 notes, in an analogous manner to the way the 7 note white key and 5 note black key scales are interlocked on the piano keyboard. He told me that the property known as Moments of Symmetry might provide some way of grouping these new scales into useable subsets.

A Moment of Symmetry (MOS) occurs when, in piling up a chain of equal intervals, a scale of two and only two melodic scale degree sizes results. There have been a number of articles on the workings of Moments of Symmetry (Wilson, 1975a, 1975b, Chalmers, 1975), but briefly, if we were making a chain of 3/2 perfect 5ths (701.955 cents), we would generate MOS scales with 2, 3, 5, 7, 12, 17, etc degrees in them. All the rest of the scales (for example 4, 6, 8, 9 scale degrees) will have 3 kinds of melodic scale degree intervals in them. Ignoring the trivial examples of scales of 2 and 3 scale degrees, the lowest example in the chain of 5ths scale where two MOS scales add up to a 3rd MOS scale with NO intervening examples of a MOS scale is the sequence of 5 scale degrees + 7 scale degrees = 12 scale degrees.

Examining the scales generated by piling up 833.090297 cent intervals, we find Moments of Symmetry with scales of 2, 3, 4, 7, 10, 13 23, etc. scale degrees. The lowest example of two scales that add up to a third with no intervening scales is 10 scale degrees + 13 scale degrees = 23 scale degrees. So if we divide our 23 tone just and Pythagorean scales into groups of 13 and 10 scale degrees we would get a scales interlocked in the following manners:

Just Fibonnacci scale divided up into 13 + 10 note groups:

Black Keys 1 3 5 8 10 12 15
White Keys 0 2 4 6 7 9 11 13 14 16
Black Keys 17 19 21
White Keys 18 20 22 0(23)

Pythagorean Fibonacci scale made from piling up 833.090297c intervals divided into 13 + 10 groups:

Black Keys 1 3 5 7 10 12 14
White Keys 0 2 4 6 8 9 11 13 15

Black Keys 17 19 21
White Keys 16 18 20 22 0 (23)

Although the breaks of 2 successive scale degrees occur at different places in the two scales, it should be obvious that the two structures simply transpositions of each other.

Intrigued by this, I generated Pythagorean and Just series for each of the succeeding 4 recurrent series, using the phenomena of the lowest two interlocking MOS scales that added up to a 3rd MOS scale as my guide for how many pitches my scales should have, and what structure they should have. For the next four scales the number of MOS degrees was as follows:

An = An-2 + An-1 (An/An-1 converges on 1.618033989...) MOS: 13 + 10 = 23
Bn = Bn-3 + Bn-1 (Bn/Bn-1 converges on 1.465571232...) MOS: 11 + 9 = 20
Cn = Cn-3 + Cn-2 (Cn/Cn-1 converges on 1.324717957...) MOS: 7 + 5 = 12
Dn = Dn-4 + Dn-1 (Dn/Dn-1 converges on 1.380277569...) MOS: 15 + 13 = 28
En = En-4 + En-3 (En/En-1 converges on 1.220744085...) MOS: 10 + 7 = 17.

Here are the scales generated from these sequences, with their divisions into the MOS groupings. In each case, the Just scale, made from the actual intervals of the number series are given, then the scale made from piling up the limit interval in a Pythagorean manner is given. In the Just Scales, the element of the number series a particular pitch is expressing is given in parentheses after the interval name.

SCALE B (JUST):

Just scale made from number series Bn=Bn-3+Bn-1.
MOS division 11 + 9 = 20

(The series: 1 1 1 2 3 4 6 9 13 19 28 41 60 88 129 189 277 406 595 872 1278 1873 2745 4023 5896)

0:	1/1	0.000	(1)(2)(4)
1:	129/128	13.473	(129)
2:	277/256	136.491	(277)
3:	9/8	203.910	(9)
4:	595/512	260.095	(595)
5:	19/16	297.513	(19)
6:	639/512	383.607	(1278)
7:	41/32	429.062	(41)
8:	2745/2048	507.109	(2745)
9:	11/8	551.318	(88)
10:	737/512	630.625	(5896)
11:	189/128	674.691	(189)
12:	3/2	701.955	(3)
13:	203/128	798.403	(406)
14:	13/8	840.528	(13)
15:	109/64	921.821	(872)
16:	7/4	968.826	(28)
17:	1873/1024	1045.362	(1873)
18:	15/8	1088.269	(60)
19:	4023/2048	1168.867	(4023)
20:	2/1	1200.000	

Scale degrees generated in order:
0 12 3 14 5 16 7 18 9 1 11 2 13 4 15 6 17 8 19 10

Division into MOS 11 + 9 = 20:
Black Keys 2 4 6 8 10 13 15
White Keys 0 1 3 5 7 9 11 12 14 16
Black Keys 17 19
White Keys 18 0(20)

SCALE B (PYTHAGOREAN):

Pythagorean Scale Made of 1.465571232/1.0000, an interval of 661.755708c. MOS 11+9=20.

0:	1/1	0.000
1:	79.313 cents	79.313
2:	123.511 cents	123.511
3:	202.824 cents	202.824
4:	247.023 cents	247.023
5:	326.336 cents	326.336
6:	370.534 cents	370.534
7:	449.847 cents	449.847
8:	494.046 cents	494.046
9:	573.358 cents	573.358
10:	617.557 cents	617.557
11:	661.756 cents	661.756
12:	741.069 cents	741.069
13:	785.267 cents	785.267
14:	864.580 cents	864.580
15:	908.779 cents	908.779

16:	988.091 cents	988.091
17:	1032.290 cents	1032.290
18:	1111.603 cents	1111.603
19:	1155.801 cents	1155.801
20:	2/1	1200.000

Scale degrees generated in order:
0 11 2 13 4 15 6 17 8 19 10 1 12 3 14 5 16 7 18 9

Divsion into MOS 11 + 9 = 20
Black Keys 1 3 5 7 9 12 14 16
White Keys 0 2 4 6 8 10 11 13 15
Black Keys 18
White Keys 17 19 0(20)

SCALE C (JUST):

Just scale made from the number series Cn = Cn-3+Cn-2. MOS division 7 + 5 = 12

(The series: 1 0 1 1 1 2 2 3 4 5 7 9 12 16 21 28 37 49 65 86 114 151)

0:	1/1	0.000	(1)(2)(4)
1:	65/64	26.841	(65)
2:	9/8	203.910	(9)
3:	37/32	251.344	(37)
4:	151/128	286.086	(151)
5:	5/4	386.314	(5)
6:	21/16	470.781	(21)
7:	43/32	511.518	(86)
8:	3/2	701.955	(3)
9:	49/32	737.652	(49)
10:	7/4	968.826	(7)
11:	57/32	999.468	(114)
12:	2/1	1200.000	

Scale degrees generated in order:
0 8 5 10 2 6 3 9 1 7 11 4

Division into MOS: 7 + 5 = 12
Black Keys 1 4 7 9 11
White Keys 0 2 3 5 6 8 10 0(12)

SCALE C (PYTHAGOREAN):

Pythagorean scale made from ratios of 1.324717975/1.000 - 486.822277c MOS 7+5=12

0:	1/1	0.000
1:	34.111 cents	34.111
2:	68.223 cents	68.223
3:	260.467 cents	260.467
4:	294.578 cents	294.578
5:	486.822 cents	486.822
6:	520.934 cents	520.934
7:	555.045 cents	555.045
8:	747.289 cents	747.289
9:	781.400 cents	781.400
10:	973.645 cents	973.645
11:	1007.756 cents	1007.756
12:	2/1	1200.000

Scale degrees generated in order:
0 5 10 3 8 1 6 11 4 9 2 7

Division into MOS: 7+5 =12
Black Keys 2 4 7 9 11
White Keys 0 1 3 5 6 8 10 0(12)
(note that this is a different division than above)

SCALE D (JUST):

Just scale made from the number series Dn = Dn-4+Dn-1. MOS division 15 + 13 = 28
(The series:1 1 1 2 3 4 5 7 10 14 19 26 36 50 69 95 131 181 250 345 476 657 907 1252 1728 2385 3292 4544 6272 8657 11949 16493 22765 31422)

0:	1/1	0.000	(1)(2)(4)
1:	16493/16384	11.479	(16493)
2:	131/128	40.108	(131)
3:	8657/8192	95.582	(8657)
4:	69/64	130.229	(69)
5:	71/64	179.697	(4544)
6:	9/8	203.910	(36)
7:	2385/2048	263.728	(2385)
8:	19/16	297.513	(19)
9:	313/256	348.023	(1252)
10:	5/4	386.314	(5)
11:	657/512	431.699	(657)
12:	345/256	516.543	(345)
13:	22765/16384	569.436	(22765)
14:	181/128	599.815	(181)
15:	11949/8192	653.523	(11949)
16:	95/64	683.827	(95)
17:	3/2	701.955	(3)
18:	49/32	737.652	(6272)
19:	25/16	772.627	(50)
20:	823/512	821.698	(3292)
21:	13/8	840.528	(26)
22:	27/16	905.865	(1728)
23:	7/4	968.826	(7)(14)
24:	907/512	989.950	(907)
25:	119/64	1073.781	(476)
26:	15711/8192	1127.385	(31422)
27:	125/64	1158.941	(250)
28:	2/1	1200.000	

Scale degrees generated in order:
0 17 10 23 8 21 6 19 4 16 2 14 27 12 25 11 24 9 22 7 20 5 18 3 15 1 13 26

Division into MOS: 15+13=28
Black Keys 1 3 5 7 9 11 13 15
White Keys 0 2 4 6 8 10 12 14 16
Black Keys 18 20 22 24 26
White Keys 17 19 21 23 25 27 0

SCALE D (PYTHAGOREAN):

Pythagorean scale made from ratios of 1.380277569/1.000 - 557.950101c MOS 15+13=28

0:	1/1	0.000
1:	53.351 cents	53.351
2:	106.703 cents	106.703
3:	137.451 cents	137.451
4:	190.802 cents	190.802
5:	221.551 cents	221.551
6:	274.902 cents	274.902
7:	305.651 cents	305.651
8:	359.002 cents	359.002
9:	389.750 cents	389.750
10:	443.102 cents	443.102
11:	473.850 cents	473.850
12:	527.202 cents	527.202
13:	557.950 cents	557.950
14:	611.301 cents	611.301
15:	664.653 cents	664.653
16:	695.401 cents	695.401
17:	748.753 cents	748.753
18:	779.501 cents	779.501
19:	832.852 cents	832.852
20:	863.601 cents	863.601
21:	916.952 cents	916.952
22:	947.701 cents	947.701
23:	1001.052 cents	1001.052
24:	1031.800 cents	1031.800
25:	1085.152 cents	1085.152
26:	1115.900 cents	1115.900
27:	1169.252 cents	1169.252
28:	2/1	1200.000

Scale degrees generated in order:
0 13 26 11 24 9 22 7 20 5 18 3 16 1 14 27 12 25 10 23 8 21 6 19 4 17 2 15

Division into MOS: 15+13=28
Black Keys 2 4 6 8 10 12 15
White Keys 0 1 3 5 7 9 11 13 14 16
Black Keys 17 19 21 23 25 27
White Keys 18 20 22 24 26 0
(note again that this MOS division is different than the MOS division of the Just scale)

SCALE E (JUST):

Just scale made from the number series En = En-4+En-3. MOS division 10 + 7 = 17

(The series:1 0 0 1 1 0 1 2 1 1 3 3 2 4 6 5 6 10 11 11 16 21 22 27 37 43 49 64 80 92 113 144 172 205 257 316 377 462)

0:	1/1	0.000	(1,2,4, 16, 64)
1:	257/256	6.749	(257)
2:	9/8	203.910	(144)
3:	37/32	251.344	(37)
4:	79/64	364.537	(316)
5:	5/4	386.314	(5, 10)
6:	21/16	470.781	(21)
7:	43/32	511.518	(43,172)
8:	11/8	551.318	(11, 22)
9:	23/16	628.274	(92)
10:	377/256	670.105	(377)
11:	3/2	701.955	(3, 6)
12:	49/32	737.652	(49)
13:	205/128	815.376	(205)
14:	27/16	905.865	(27)
15:	113/64	984.215	(113)
16:	231/128	1022.099	(462)
17:	2/1	1200.000	

Scale degrees generated in order:
0 11 5 8 6 14 3 7 12 9 15 2 13 1 4 10 16

Division into MOS: 10+7=17
Black Keys 1 2 4 10 13 15 16
White Keys 0 3 5 6 7 8 9 11 12 14 0
(note the incredibly strange division of the scale here - maybe this is due to the very redundant nature of this series. The MOS division of the Pythagorean scale, below, is much more typical of MOS divisions.)

SCALE E (PYTHAGOREAN):

Pythagorean scale made from ratios of 1.220744085/1.000 - 345.312945c MOS 10+7=17

0:	1/1	0.000
1:	17.191 cents	17.191
2:	34.381 cents	34.381
3:	181.252 cents	181.252
4:	198.442 cents	198.442
5:	345.313 cents	345.313
6:	362.504 cents	362.504
7:	379.694 cents	379.694
8:	526.565 cents	526.565
9:	543.755 cents	543.755
10:	690.626 cents	690.626
11:	707.817 cents	707.817
12:	725.007 cents	725.007
13:	871.878 cents	871.878
14:	889.068 cents	889.068
15:	1035.939 cents	1035.939
16:	1053.129 cents	1053.129
17:	2/1	1200.000

Scale degrees generated in order:
0 5 10 15 3 8 13 1 6 11 16 4 9 14 2 7 12

Division into MOS: 10 + 7 = 17
Black Keys 2 4 7 9 12 14 16
White Keys 0 1 3 5 6 8 10 11 13 15 0(17)

I noticed that in generating each of the Just scales, that at a certain point, an interval was generated that was much less than a semitone away from one of the other tones of the scale. In the 12 note Pythagorean scale, made by stacking 702 cent fifths, after 12 fifths, we get a scale degree of 24 cents, which is very close to our starting point. This is called the Pythagorean comma , and one reason for stopping at 12 tones in this series is that that small interval was thought to be musically not useful. So for each of my series, I thought that I would see where the first narrow, or commatic interval occurred. For the first scale (Fibonacci series - Scale A), this yielded a 10 tone scale (the first comma was between 3/2 -702 c and 377/256 - 670 c)

Note:	1/1	17/16	9/8	5/4	21/16
Cents:	0	105	204	386	471
Note:	89/64	3/2	13/8	55/32	233/128
Cents:	571	702	841	938	1037

For Scale B a 9 tone scale resulted (comma between 1/1 - 0c and 129/128 - 13.5c):

Note:	1/1	9/8	19/16	41/32	11/8
Cents:	0	204	298	429	551
Note:	3/2	13/8	7/4	15/8	
Cents:	702	841	969	1088	

For Scale C a 7 tone scale resulted (comma between 3/2 - 702c and 49/32 - 737.7c):

Note:	1/1	9/8	37/32	5/4	21/16
Cents:	0	204	251	386	471
Note:	3/2	7/4			
Cents:	702	969			

For Scale D a 9 tone scale resulted (comma between 3/2 - 702c and 95/64 - 683.8c):

Note:	1/1	69/64	9/8	19/16	5/4
Cents:	0	130	204	298	386
Note:	3/2	25/16	13/8	7/4	
Cents:	702	773	841	969	

For Scale E a 7 (or 8) tone scale resulted (comma almost exactly halfway between 11/8 - 551.3c and 21/16 - 470.7c - which is 43/32 - 511.5c. Or, if a 40c comma is too large for you, the next interval in the series 49/43 - 737.6c is only 30 cents away from 3/2 - 702c. For practical purposes, though, the series of three tiny intervals in the scale below made by admitting the 43/32 might be thought redundant.)

Note:	1/1	37/32	5/4	21/16	(43/32)
Cents:	0	251	386	471	512
Note:	11/8	3/2	27/16		
Cents:	551	702	906		

Note that with the exception of scale D, all the scales have the smaller number of elements in their MOS division partners. That is for A, MOS13 + MOS 10 = MOS 23, and the commatically determined scale has 10 degrees. And for the exception, scale D, in the progression of MOS scales that leads eventually to the additive series MOS 15 + MOS 13 = MOS 28, we find that there is, indeed, a MOS with a scale of 9 scale degrees. These subsets of the Just scales might be very useful, especially for constructing modes based on these series.

In fact, Wilson, in his 2001 article, Pingala s Meru Prastara, and Sums of the Diagonals (Wilson, 2001) refers to scales resulting from the Second (Bn=Bn-3+Bn-1) and Third Recurrent Sequences respectively, as Meta-Pelog and Meta-Slendro . His Meta-Pelog is indeed a 9 note scale, encompassing elements 6 9 13 19 28 41 60 88 129 of the series. That is, his Meta-Pelog and my 9 tone scale above have only one difference - mine includes a 1/1 at 0 cents, and his scale eliminates this, substituting a 129/128 at 13.5 cents. Wilson is, of course, famous for his scales with no tonal centre (Rapoport, 1994), so his choice of series elements here is not surprising. His Meta-Slendro is, as he says, an unusual but compelling 12 tone scale made from elements of the 3rd recurrent sequence (Cn=Cn-3+Cn-2) as follows: 9 12 16 21 28 37 49 65 86 114 151 200. Here, he goes one element further in the series than I do for his 12 tones, stopping at element 200. Further, because he starts at element 9, he eliminates element 5. So his 12 tone scale based on this series is identical to mine with the sole exception that there is no 5/4 at 386c, but there is, instead, a 25/16 at 773 cents.

Moreover, it is to be noted that these Just scales resulted from taking the elements of each series as members of the harmonic series. It would be just as feasible to take them as members of the subharmonic series, and get the inversions of the scales, and further, to then combine the harmonically generated and subharmonically generated scales to get symmetrical scales with very many melodic and harmonic possibilities. However, that might be a topic of exploration for another time. Already, with just the material explored, I ve generated between 10 and 35 new scales (depending on what you consider an independent scale, and what you consider just a subset of a larger scale). Wilson, has, to date, enumerated the first 192 of the Meru-Pascal recurrent sequences, of which I ve looked at only the first five. For the numerologically inclined, there would be an endless set of resources here to explore.

However, despite my fascination with the patterns in all the preceding, I m actually not one of the numerologically inclined. I m principally a composer, and my interest in all these things is in how they SOUND, what uses can be made of them, and how hearing these new relationships affects us, both emotionally and physically. I first wanted to hear how the scales sounded, in the order of generating their pitches. I thought that if a sequence of pitches as generated were heard canonically, we could also hear the harmonies produced by stacking the generating intervals of the scales. Using Emanuel op de Coul s Scala, and John Dunn s Softstep and Microtone, I proceeded to generate the following examples:

Example 1 is Scale A from the 1st recurrent sequence (the fibonacci sequence) played using the order of scale degrees as generated, with a piano timbre. First we hear the melody solo, on a pulse, and then with 2 voices, canonically, then 3 voices canonically, with a delay of one note between each of the voices. Note how at the end of the example, triadic type harmonies emerge from the result of the stackings and octave reductions of the 833 cent intervals.

Example 2 is the Pythagorean scale equivalent of Scale A (833 cents per generating step) treated in the same way. Here, the small-integer just intervals which characterize the beginning of Scale 1 are missing. Everything has the same quality, as it will in a Pythagorean scale. Notice that the harmonic quality of this scale is almost identical to the harmonic quality of the previous scale near the end of its sequence. That s because this scale is made exclusively of the interval the Just scale tends to as it progresses further and further through the series.

Example 3 is Just Scale B treated the same way. The mean interval here is 662 cents, a very sharp tritone, or an extremely flat fifth. Note the atonal type chords (sorry for the imprecise terminology!) that result from the canonic playing of this sequence as a canon.

Example 4 is the Pythagorean scale equivalent of Scale B (662 cents per generating step) treated in the same way. Again, note that the harmonic quality of this scale is almost identical with the harmonic quality of the just scale near the end of its sequence.

Example 5 is Just Scale C treated the same way. The mean interval here is 487 cents, a slightly flattened fourth, and indeed, at the end of the example, a three voice canonic playing of the sequence yields a series of fourth-like chords.

Example 6 is the Pythagorean scale equivalent of Scale D (487 cents per generating step) treated in the same manner.

Example 7 is Just Scale D treated the same way. The mean interval here is 558 cents, and the chords resulting from canonic playing of the sequence do indeed have a tritone quality, but to my ear sound less dissonant than the chords resulting from the canonic playing of Just Scale B.

Example 8 is the Pythagorean scale equivalent of Scale D (558 cents per generating step).

Example 9 is Just Scale E treated the same way. The mean interval here is 345 cents, which is a neutral third. Chords made from stacking these have a neutral triad quality. However, with the larger Just intervals at the start of the sequence, and the fact that this series takes a little longer to settle down into intervals around its mean, this harmonic quality is only heard very briefly at the end of the series.

Example 10 is the Pythagorean scale equivalent of Scale E (345 cents per generating step). Here, the neutral triad quality implied by Just Scale E is heard much more clearly.

So after all this work on scales, and hearing some examples of the quality of harmonies implied by some of their aspect, it s now time to begin to compose with these materials. Questions about pitch organization naturally occur, but additional questions come to mind at this point about timbre (could these series generate interesting spectra, or if we use simple sine waves to play these scales, will the inherent harmonies of the scales and the pure sine waves combine to make additive synthesis timbres on their own?), rhythm (without getting bogged down in serialism or fibonacciana, can interesting rhythmic sequences be made with these proportions?), and spatiality (can these proportions be used to structure space as well as time? Or, again, if we simply use sine wave timbres, will the proportions of the scale then naturally interact with the acoustics of the space the sounds are played in to create an emergent spatial architecture determined by the wavelengths of the pitches in conjunction with the dimensions of the performing space?). Clearly, after all this work on number series and proportion, we are actually only at the beginning of a sonic adventure exploring these materials.

A first exploration of these scales was made in the compositions grouped under the title of The Mossy Slopes of Mt. Meru. In the first of this series, a series of imaginary heterophonic duets for sampled flute and two sampled harp lines was made. Each piece was in one of the complete Just Scales, A - E, described earlier. In each piece, the three melodies (one flute and two harp) were assembled by using the sequence of generating intervals for that scale as a kind of row. The program moved back and forth through the row by steps of 1, 2, 3 or other number of steps, based on what numbers were in the corresponding number series. Durations for each of the three simultaneous melodies were also drawn from numbers available in the number series (a random choice from a small set of durations was used here), and dynamics were chosen in the same way. The character of each melody then, was determined by proportions inherent in the number series that generated the scale. Over the course of this short piece (five movements of 90 seconds each) different characters of harmony, tempo and gesture can clearly be heard. These characters are an emergent property of the different characters of each number series. As an example, here are the beginnings of three of the pieces:

Examples 11, 12, 13: The first 20" each of Just Scale A , Just Scale B and Just Scale C from The Mossy Slopes of Mt. Meru - II: Imaginary Flute and Harp in a Virtual Garage.

A more thorough exploration of the harmonies available in the MOS divisions of these scales was made in the hour long composition The Mossy Slopes of Mt. Meru - I: Architectural Chords. In this piece, each of the 30 possible scales outlined in this paper is used in a two minute section. They are played with pure sine waves, over a 3, 4 or 5 octave range, with chords that can have between 2 and 10 elements. Melodic motion through the scales was step-wise, in numbers of scales steps determined by low numbers of the relevant number series. For example, there could be steps of 1, 2, 3, 5, or 8 scale degrees in the scales derived from the Fibonacci series. The exact number of steps was chosen by a shift register feedback algorithm I first used in 1972, now embodied in software. (Burt, 1975) Similarly, dynamics and durations were also chosen in a similar manner from lists of numbers taken from the relevant number series. The results of using pure sine waves were what I had hoped for. The sine waves at times did indeed add up to make composite timbres, especially when a wide range of dynamics was chosen for a chord, at times sounding like a simple interval or chord, but at other times fusing into some quite delicious and complex timbres. When I used timbres with more harmonics, this fusing effect was nowhere near as pronounced as when I used pure sines, so I chose to stay with them. Further, because sine waves carve any performance space up into areas of high and low pressure, each chord created quite remarkable spatial effects. These are heard most clearly by moving one s head slightly while listening to these chords. Single pitches, and sometimes whole ranges of the chords seem to appear and disappear as one s head moves. This clearly demonstrates that each chord is indeed dividing the performing space up into a sonic architecture proportioned by the ratios of the pitches in each chord. Sine waves are notoriously non-directional, but just for a little extra dimensionality, I also panned each sine wave slightly in space. In most spaces this seems to make no difference, but when heard over headphones, the effect of this panning is quite marked. Again, when I tried this with more complex timbres, the effect was not as pronounced as when I used pure sines. Since I wanted this piece to be as pure an exposition of the qualities of these scales as I could get, I decided to stay with the sine wave timbres. In live performance, I control the tempo of the chords, how many notes make a chord, and which scale is being used. Pitch, durations and dynamics are chosen algorithmically. I find this piece very beautiful, and I think I can safely say that a large measure of the beauty is the result of emergent properties of these scales and number series being allowed to reveal themselves as a result of my processes and algorithms.

Examples 14 & 15: 30 second excerpts from 1: Just Scale A: 13/23 MOS and 23: Pythagorean Scale D: 28/28 MOS from The Mossy Steps of Mt. Meru - I: Architectural Chords showing the different kinds of harmonies, timbres, and spatialities available from these scales.

REFERENCES:

Burt, Warren Aardvarks IV: A Real-Time Composing Machine Unpublished Masters Thesis, University of California, San Diego, 1975

Chalmers, John, 1975 Cyclical Scales Xenharmonikon 4, Vol. 2, No. 2, Fall 1975. Frog Peak Music, Lebanon, NH.

Green, Thomas M., 1968 Recurrent Sequences and Pascal s Triangle , Mathematics Magazine, Vol. 41, 1968

McClain, Ernest, 1976, The Myth of Invariance New York, Nicholas Hays

Rapoport, Paul, 1994, Just Shape, Nothing Central Musicworks 60, Fall 1994, pp42-49

Singh, A. N., 1936, On the use of series in Hindu Mathematics Osiris, Vol I, pp. 623-624, 1936; http://www.anaphoria.com/hindu.PDF

Wilson, Ervin, 1975a, Letter to John Chalmers pertaining to Moments of Symmetry/Tanabe Cycle, http://www.anaphoria.com/mos.PDF

Wilson, Ervin, 1975b, On the Development of Intonational Systems by Extended Linear Mapping , Xenharmonikon 3, Vol 2, # 1, Spring 1975. Frog Peak Music, Lebanon, NH.

Wilson, Ervin, 1993a The Scales of Mt. Meru http://www.anaphoria.com/meruone.PDF

Wilson, Ervin, 1993b Some Exploratory Triangles for Scales of Mt. Meru http://www.anaphoria.com/merutwo.PDF

Wilson, Ervin, 2001 Pingala s Meru Prastara and the Sums of the Diagonals http://www.anaphoria.com/MERU.PDF

Enhancing the Experience Of Music-Ritual through Gesturally-Controlled Interactive Technology

Hannah Clemen

Western Australian Academy of Performing Arts, Edith Cowan University
email: hclemen@iinet.net.au

Abstract

This paper presents the possibility of applying modern gesturally-controlled interactive music technology to the aesthetic framework of "music-ritual", with the aim of enhancing the experience for the audience and participants by more directly mapping gesture with sound production. The potential of this kind of application could be to increase the communicability of symbolic meaning through concept, sound, and gesture.

1 Introduction

The concept of "music-ritual" has been explored and utilised extensively by many Western "art music" composers during the 20th Century, particularly during the 1960s and 1970s, and was to a large extent motivated by the social and artistic "disillusionment" of intellectual left-wing at the time. My own interest in these ideas has been particularly focussed on how different physically expressive forms (such as movement in space and gesture) can be used in conjunction with sound to enhance the experience for both the performer and the audience. This stemmed from my activities as an instrumental composer, which for some time have been focussed on how meaning can be communicated by sound. These works made use of various methods, applying concepts derived from psychoacoustics, symbolism and ritual, and were often inspired by composers such as Karlheinz Stockhausen and Pauline Oliveros.

In 1999 I began composing electroacoustic music (primarily using Pro Tools to treat and mix found sounds), which I believe has given me much more creative freedom than in purely notated music, as it has allowed me to manipulate the spatial and textural qualities of the sound itself. This kind of work can induce a feeling of complete immersion and a sense of physical "disembodiment", thus allowing both the composer and listener to explore a much deeper, internalised level of musical communication.

My most recent works have begun to make use of both ritualistic forms and electronically modified sound, as it is my contention that by using currently available gestural and movement technology as the medium of control, these two forms of creative expression could be combined, and the realization of ritual music forms could be further enhanced.

The ability of this kind of technology to accurately map movement to sound or sound manipulation makes it ideal for directly representing sound in movement. This means that a meaningful connection needs to be made between the two forms of expression, so that they can be experienced as two parts of the one whole.

This research is still in its formative stages, and has yet to draw any definite conclusions. This paper presents an outline of my background research, my ideas, and provides a model for the actual realization of these ideas. There are 3 stages to this discussion:

Background Research. I will begin by briefly discussing the concept of sound as a communicator of extra-musical ideas, with particular reference to Tibetan Buddhist traditions. These concepts will also be examined within the context of 20th Century art music, and will examine some of the relevant work of Karlheinz Stockhausen and Pauline Oliveros.

New Media and Methods. I will then examine some of the new aesthetic concepts brought forward by new gestural and movement controller technology, with particular reference to the work of Atau Tanaka and Tod Machover.

Conclusion. This section consolidates the various points presented throughout the discussion.

2 Background Research

2.1 Theory, Physics and Metaphysics

Essentially, my interest in the connection between gesture and sound is based on the belief that music – the organised relationships and interactions in time between different pitches and rhythms – reflects the organised relationships and interactions between the different elements of other systems within the universe, and hence the universe itself. Through the numerical symbolism of music, the interplay between forces that would normally lie beyond the realm of the human senses becomes something experiential. "In musical experience we come into direct contact with these principles [of numerical symbolism]; the connection between physical reality and metaphysical principles can be felt in music as nowhere else" (Daniélou 1943, p. 1).

Some ancient texts (e.g. the Indian *Vedas*) suggest that sound is in fact a *direct* – not merely symbolic – representation of these universal interplays, implying that each object, be it an atom or a star, has a vibrational frequency and thus a sound. Although these sounds are normally too subtle to be heard, audible sounds can be made that correspond directly to these more subtle vibrations. This belief is strongly held by Stockhausen, who has referred to it many times in his writings and conversations. "Every object in the world, down to the smallest atom, produces waves which can be transformed into acoustic waves. [...] Everything makes sounds" (Stockhausen 1971, cited in Cott 1974, p. 76). In China, music was used as a tool for the diagnosis and rectification of social or political disturbances (Campbell 1959, p. 453).

The belief that sound is both a physically and psychologically active force is one that has found wide application in electronic and computer-generated music. For many years, electronic music artists such as Steven Halpern have explored these theories, and believe that by correctly applying audible sounds that correspond to more subtle vibrations, a healing and self-transformative effect can take place. Other artists such as the theorist/composer Dr. Jeffrey D. Thompson of the Center for Neuroacoustic Research have been using computer-modelling and various synthesis/spectral filtering techniques to make audible such phenomena as the magnetic fields of planets and various chaotic processes. Many of the artists within the "stochastic music" movement of the 1950s, of whom Iannis Xenakis was a major exponent, were attempting to use sound to make the complex and esoteric aesthetics of mathematics more accessible on an *experiential* level.

2.2 The Nature of Ritual

Before we continue, it seems pertinent to clarify the definition of the word "ritual" as it will be used in this paper. Unfortunately, the word does carry different meanings for different people, and can suggest rigid, anachronistic religious and social role-plays (for this reason, Pauline Oliveros uses the term "ceremony" rather than "ritual" in her work). Within this discussion, I am using the word "ritual" to describe an activity that incorporates the following features. Firstly, time and space is symbolically defined by the participants and observers. Secondly, an underlying concept is being communicated, usually considered to be important or relevant to the lives of the participants and observers. Thirdly, physical, verbal or sonic metaphors are used to translate the meaning of the concept into a more easily understood form (i.e. to "physically manifest" an esoteric concept that cannot normally be understood). The amount of metaphorical translation can be varied, depending on how important it is for "lay-people" to understand the meaning behind the concepts.

Christopher Small (1998) suggests a slightly different terminology, emphasising the importance of ritual as an affirmation of different levels of *relatedness*:

> *Ritual is a form of organized behaviour in which humans use the language of gesture, or paralanguage, to affirm, to explore and to celebrate their ideas of how the relationships of the cosmos (or of part of it), operate, and thus of how they themselves should relate to it and to one another.* (p. 95)

Small (p. 99) goes on to isolate the term "myth" as being the concept being expressed by a ritual, and "metaphor" (the articulation of the elements and relationships within a myth) being the method of communication of that concept within a ritual. He emphasizes the importance of ritual as an experiential tool for understanding; that the participation in or observation of the ritual process allows the individual to become part of the symbolic representation of relationships, and thus the experience becomes more intimate than if the concepts were just "explained".

Because of their personal, experiential, and universally shared nature, the physical senses are an effective medium for the metaphorical translation of conceptual meaning. This belief is particularly evident in the complex ritualistic traditions of Tibetan Buddhism, in which meaning is manifested through meditation and ritual using three primary technical aids: "…*yantra, mantra*, and *mudrã*[1]: the parallelism of the visible, the audible and the tangible (i.e., what can be felt). They are the exponents of mind (*citta*) [e.g., mandala symbolism], speech (*vãc, vãcã*), and body (*kãya*)" (Govinda 1960, p. 92).

Yantra is a term that refers to a visual meditational aid for focus and inner clarity. The arrangement of sacred symbols represent a universal set of relationships, and in meditation are either studied or created. *Mantra* is an aural aid. It exists in the form of various sacred seed-syllables that are combined and uttered by the student. These sounds are believed to initiate subtle inner vibrations within the student that resonate with more universal cosmic vibrations. *Mudrã* is a series of bodily (usually hand) movements that are performed during rituals and meditation, and often accompany the chanting of *mantra*. *Mudrã* serves to emphasise the concepts being visualised and vocalised – it reinforces, through association, the power of the symbolism, and makes the meditational process more complete by including the body as well as the voice and the mind in the process.

Govinda asserts that each concept needs to be understood for the action to be effective, and that sound is the vehicle for understanding, not the understanding itself (pp. 93). However, modern artists like Jeffrey Thompson believe that these subtle vibrations are universal and can cause physical and psychological affects that are entirely subliminal. Either way, it is clear that sound can be used as a powerful tool for the realisation of mentally visualized concepts, and that gesture is an excellent method for symbolically "physicalising" these concepts into one's own body, making them more directly experienced. The importance of metaphor and ritual is paramount for these concepts to be effectively communicated.

All of these concepts are directly applicable to interactive music technology, with the added implication that by simply *using* this technology, there can be the implication of certain symbolic associations. When choosing a particular piece of technology, a composer must decide how overtly present (or subtle) the visual and sonic aspects of the "machine presence" should be, and whether this adds or detracts from the audience's musical experience.

> *Audience members, untrained in the intricacies of computers or composition, may not be interested in the complexities of software design, the sophistication of musical processes, or the amount of memory in a computer. They come to concerts for a variety of reasons, not the least being to enjoy the drama of a live performance.* (Winkler 1999, p. 9)

This kind of judgment needs to be equally as sensitive when choosing sounds and instruments. For example, to the Marind-anim people of South-East Asia, a bull-roarer carries a different set of sonic and cultural associations to a drum. Similarly, to a modern Western listener, a synthetic sound will carry a different aesthetic to a raw sample, as much as one of Machover's hyperinstruments would suggest a different kind of visual concept to a totally "novel" controller such as Waisvisz's The Web.

2.3 20th Century Explorations

The 1960s and 1970s was a time of social and political turbulence in much of the Western world. Many European traditions were being challenged, including those artistic and musical. Many artists and composers began to search outside their own cultures to find answers to their questions, with many incorporating philosophies and aesthetics derived from non-Western traditions into their works. Composers such as John Cage, Cornelius Cardew and LaMonte Young began to create works in which the structures, sounds and actions became deliberately charged with symbolic meaning. Indeterminacy and non-linear structures might reflect a Zen-derived concept of time. In Cage's "infamous" 4'33" (1952), ambient sounds become symbols for representing his philosophy regarding the nature of music; in many of the works of LaMonte Young, the role of the audience is altered (sometimes without their

[1] In Microsoft Word, the Anglicised spelling of the Sanskrit words *mudra, vac, vaca* and *kaya* make use of the '~' symbol above the 'a' rather than the correct '-'.

knowledge) in order to challenge their ideas regarding audience/performer relationships. Sometimes, even a musical instrument is used as a symbolic object in a ritual act (e.g. *Piano Piece for David Tudor No. 2* (Nyman 1974, p. 72)).

Although the concepts, methods and motivations of 20th Century avant-garde composers are very different to those of Tibetan monks, the essential tools are very similar. A myth (concept or philosophy) is presented in the form of metaphor (structural, aural and physical representation) to communicate to – and personally involve – the participants (the line between audience/performer start to become blurred). Of course one does not need to be aligned with any particular "non-Western" form of thought to experiment with ritual forms in music. These methods can be used to serve any kind of aesthetic philosophy, and can be either explicit or implicit.

Pauline Oliveros began experimenting with her so-called "ceremonial" works in the late 1970s. She had abandoned standard notation in 1964 in favour of a more theatrical approach (e.g. *Pieces of Eight* (1964), *Aeolian Partitions* (1969)), through which "…she found that her penchant for imagery allowed her to make statements about music that were impossible to say in a totally abstract medium, … and theater pieces provided the opportunity to use material objects to augment her message" (von Gunden 1983, p. 70).

In the early 1970s Oliveros began to study some of the "deeper" aspects of listening and music-making, and became interested in creativity on both metaphysical (dreams, myths, meditation) and physical (Tai-Chi, karate) levels. This co-exploration of the senses culminated in the *Sonic Meditations* (1971), a collection of 12 text works carefully designed to guide the participant into new levels of interplay between two models of information processing – focal attention and global awareness. "Attention is narrow, pointed and selective. Awareness is broad, diffuse and inclusive" (Oliveros 1984, p. 139).

These meditations led to her so-called "ceremonial works" of the later 1970s such as *Crow II* (1976) and Rose *Moon* (1977), in which the fundamental concepts of awareness outlined in the *Sonic Meditations* are used, with the added elements of physical symbolism and ritual action.

The ritual of *Rose Moon* begins with the mandala-style spatial layout, one of the most important visual/conceptual symbols in Oliveros' work. The circular performance space (to be surrounded by the audience) is marked with the points of the compass, and two runners – who alternate every half-hour – circumnavigate the mandala while making a continuous, repetitive percussive sound with bells attached to belts around their waists. 8 musicians sit just within the circumference, and sound their instruments each time the runner passes directly behind their spine. These musicians recite at will the word "moon" in various languages, as well as names of people they feel should be remembered. Audiences may also participate in this activity if they wish.

Inside this circle, 7 other musicians move in procession in the opposite direction to the runner, sounding their ceremonial percussion instruments each time they circle the mandala while following the sound of a "moon rattle", which is passed from player to player and must always be shaken in time with the runners' bells. Like most mandalas, there are segmentations within the circle, and in *Rose Moon* these are divided (using coloured sheets) into 4 triangular quadrants representing each of the 4 elements. These quadrants are occupied by 12 "cloth people" who meditate and sing chords within their 4 groups, while manipulating the sheets in various symbolically choreographed ways. The centre of the mandala is marked with a black and white tent, in which two "moon figures" – one male and one female – enter and remove their clothing.

Rose Moon does not function like a "regular" performance piece, it has no linear narrative. Time is cyclical and multidimensional; there is action and reaction, and the presentation of opposites. Oliveros makes it clear that the work should be performed as an act of highly focussed meditation for both the audience and performers. Much of the significance of this gruelling 2-hour ritual lies in the co-operation of the performers, and in the active involvement of the audience (this is certainly not a work designed for passive entertainment). If the activities are not performed in harmony, the sound is discontinuous and the ritual disintegrates.

The tools of ritual are also strongly represented in *Rose Moon*: the mandala concept is clearly articulated visually (*yantra*) by the placement of the various characters; each set of characters utter different kinds of sounds (*mantra*), giving further definition each group; the sounds and identity of each group or character are further emphasised by their symbolic actions (*mudrã*).

This model of spatially, aurally and gesturally symbolic performance art could present a vast number of creative possibilities to the modern interactive computer music artist. For example, the role of the audience could be greatly increased by using a video monitoring system such as Eric Singer's Max object Cyclops or STEIM's BigEye. The symbolic sounds generated by the characters within the mandala structure could be changed depending on their (or someone else's) position within the space (e.g. by using movement and proximity sensors attached to an Infusion Systems I-CubeX Digitizer), or by a change of their gesture or its relationship with someone else's. Depending on the software programming, the process could be made as complex or simple as the composer desires, and the sounds much more flexible. This is not to suggest that *Rose Moon* could (or should) be improved, but that the kind of ritualistic structure used in this piece could be a potent and flexible creative foundation for the application of interactive technology.

Karlheinz Stockhausen is another composer who adopted ritualistic multimedia techniques in some of his works during the 1970s. Throughout the 1950s and 1960s, his explorations of musical symbolism were generally confined to "sound in space" as the symbolic medium. Earlier works such as *Gruppen* (1955-57) and *Carré* (1959-60) dealt with the relationships between spatially separated sounds, and those relationships carried much of the symbolic meaning. In the 1960s he began dealing with space in a metaphorical sense, by exploring the relationships between frequency and timbre, pitch and rhythm, or electronic and natural sound. Many of his works during the 1970s extend these ideas further into the realm of fully realised music-ritual works by the addition of visual and physical symbolism. This new development was possibly due to his short period of experimentation with so-called "intuitive" music, in which notation was abandoned in favour of text instructions. The series *Aus den Sieben Tagen* (1968) deals with various aspects of intuitive creativity, meditation and sonic awareness in what Robin Maconie (1976) describes as "…training the mind to prolong the moment of intuition indefinitely" (pp. 173).

In 1970, Stockhausen returned to a more structured composition style in his piano work *Mantra*, but the new importance placed on personal experience, meditation and consciousness in music remained, and he developed these ideas in many ritualistic multimedia works, eventually leading to the monumental *LICHT* opera cycle. Some works, such as *Inori* (1973-4) use gesture as a way of visually symbolizing the sonic and conceptual elements of the music, while others such as *Sternklang* (1971) and *Musik im Bauch* (1975) are fully realised pieces of ritualistic music-theatre.

Musik im Bauch resembles a rite of passage in which 3 "brothers" explore their own melodies, eventually discovering that the same melodies are being played inside the belly of a mannequin, dressed in Stockhausen's clothes, adorned with bells and with the head of an eagle. They open the belly and extract three music boxes, one for each "brother". First they play along with their musical boxes using glockenspiels, but the piece comes to a close with the boxes winding down naturally by themselves, the stage having been abandoned. During the course of the whole piece, 3 other percussionists play stretched out versions of the same musical material, as well as provide cues for sudden actions from the three "actor/percussionists". Through the augmented melodies, two alternate time scales are suggested that create a multidimensional quality to the perceived passage of time within the work. These performers also eventually leave the stage, moving like tin soldiers as the music boxes slow down.

Much of the charm of this work lies in the fact that the meaning is interpretative, giving it an element of "audience interactivity" by allowing each member of the audience to make it "his own" ritual. Unlike Oliveros' *Rose Moon*, *Musik im Bauch* is more dramatic and narrative in that it tells a story, rather than create a meditative cyclical time structure. Maconie's own interpretation is as follows: "The moral seems to be that what we take to be humanly inspired is in fact entirely preordained in the divine melodies of which the composer, in his guise as bird-man, is the mysterious messenger. Music is the ultimate meaning of the visible drama" (p. 246).

Once again, the three-fold method of concept transmission is clear: the mentally visualised concept (*yantra*), demonstrated by the stage layout, multilayered treatment of time and the significance of the visual objects used in the performance (the music boxes and mannequin); the sonic element (*mantra)*, which is generated by "human" means (in this case by only playing instruments, not singing) or by "automatic" means (music boxes); and the bodily element (*mudrã*), represented by the character of the different acts performed by both the actors and percussionists. Again we can see that the overall meaning is conveyed not by the visual elements, sounds or actions themselves, but by the entire ritualistic process of how the elements interact with each other, and how the audience relates to them.

There are many aesthetic and practical differences between the work of Oliveros and Stockhausen, but they have both made very effective and personally creative use of similar ritualistic elements. In addition, both these composers have and do use interactive technology and live computer applications in their music, and both were composing electronic music as early as the 1950s.

For many years Oliveros has used complex user-controlled delay techniques, both in her own solo performances and in ensemble projects such as The Deep Listening Band (primarily using the "Expanded Instrument System": Lexicon PCM 42 digital delays, Lexicon PCM 70 reverb units, MIDI controlled amplifiers and a Reson8 DSP Processor), and has collaborated with artists such as Doug Van Noort, whose installation using various controllers and Opcode's Max/MSP formed part of her recent work *The Library of Maps: An Opera in Many Parts*.

Stockhausen is also taking advantage of new technology, such as his use of highly complex sound diffusion systems in the more recent developments of the *LICHT* opera cycle.

It is difficult to say, if today's technology was available 25 years ago, whether they would have employed such technology in their ritualistic works of the 1970s, and if so, how different the works would be. Needless to say, the temporal and spatial structures presented by these works, as well as the use of certain "universal" ritual tools and methods, provide us with some fascinating models for application in modern interactive computer music.

3 New Media and Methods

One of the features of modern gesturally and spatially controlled instruments is that they do not follow the same kind of idiomatic "rules" as acoustic instruments: either in the playing technique or non-musical associations (which are usually as a result of the instrument's socio-cultural history).

For traditional instruments, new developments or improvements not only resulted in a complete change to the style of playing and composing, but were also a symbolic representation of the social and cultural climate of the time. It would be impossible to play a Franz Liszt etude on a harpsichord, not only because the instrument would be physically incapable of this feat (due to its narrower range, lighter frame, different sound, etc.), but it would also be *aesthetically* inappropriate. Thus each instrument's identity is defined by its physical capabilities, as well as by its many non-musical implications. "The invention of the modern piano, for example, ... fuelled the imagination and technique of nineteenth-century Romantic composers…The aesthetic and the instrument are wedded" (Winkler 1999, p. 313).

Interactive computer music is much less imbued with social and historical associations, simply because the history is much shorter. Although computer music as a general field has carried various associations over the years, it has followed the rapidly changing attitudes towards computers in general. "In a great deal of music since the advent of electronic sound production, the electronic elements have been used to imply the non-human. […] Part of my contention here is that this view of technology is now no longer relevant. Technology is beginning to empower people" (Garnett 2001, p. 21). Today there is much more choice in how the technology will communicate symbolically – its visual and sonic impact can be highly apparent (even intrusive) or so subtle it can be barely noticeable. Generally, these instruments are also more physically and aurally adaptable. Because there is no *direct* correlation between the gesture and sound of the instrument, the composer must "invent" these correlations – choices must be made carefully in order for these correlations to communicate successfully to an audience.

These two general features of identity can vary greatly with different controller instruments, either physically or mechanically. It may or may not have a connection with a traditional instrument (e.g., by its physical appearance, playing style or sound output). Indeed, an instrument may not even have to be associated with a single object, such as the "…new paradigms…" (Tanaka 2000, p. 391) of public multimedia installations and network music.

Because of the "newness" of so many of these instruments, there is still much exploration being done into the essential aesthetic qualities of this kind of music making. In many ways, a full circle has been traversed, beginning with the advent of electronic recording media. "When recording technologies became socially effective, they brought about … a shift from the prominence of music production processes (composition and interpretation) to the prominence of listening activities as cultural experience" (Iazzetta 2000, p. 259). Electronically-generated sound has deepened these tendencies, but as an interesting recent development, new gesturally-controlled technology has once again brought us to music as something physically *experienced*, with the added excitement of new uncharted aesthetic territory to explore.

Atau Tanaka and Tod Machover are two artists who are making many new practical and theoretical innovations in the field of gesturally-controlled music and its application. Although their methods and ideologies are very different, they do represent two sides of the same aesthetic coin.

One of the main differences between these two artists is that Tanaka is a practicing performer specializing generally in one instrument (the BioMuse, by BioControl Systems, Inc.), whereas Machover realizes his art through composition and multiple instrument design.

Atau Tanaka is involved in many projects that are interesting both from the point of view of new instrument aesthetics, as well as within the context of music ritual. I will briefly discuss one of his major preoccupations – his performances using the BioMuse interface, and his performances with the ensemble Sensorband.

Because of his instrumental specialization, Tanaka aims to communicate to the audience through the way he and his instrument communicate with each other. He describes this combination of the "…physical, tactile or sonic and musical…" as a kind of "feedback" between the performer and his instrument. "It is a confidence in his instrument that helps the musician to create a flowing musical dynamic that is conveyed to the audience" (Tanaka 2000, p. 400).

In many ways this idea could be seen in a similar way to the idea of the mind, voice and body as tools in the process of conveying meaning through ritual. However in Tanaka's case, the ritual is experienced by both the performer and the audience. The BioMuse is a highly expressive instrument, sensing muscle tension in the performers arms via two armbands, and thus is very responsive to physical gesture. Non-essential gestures (e.g. head or leg movements) can also be incorporated into the performance to enhance the level of expressivity, but care must be taken that "… the thin line between musical artistry and vain theatre…" (Tanaka 2000, p. 401) is not crossed.

It is in Atau Tanaka's performances with the group Sensorband that the elements of ritual are fully realized. With Edwin van der Heide and Zbigniew Karkowski (who play their own instruments, the ultrasound MIDI-Conductor and an infared sensor "cage", respectively), the ensemble performs highly engaging group-improvisations and compositions in a way that would be impossible on traditional instruments.

There is a mass of computer-generated electronic sounds coming from three musicians on stage. The audience must distinguish who is playing what. At some moments it is clear, and there are other moments where it is unclear. We can play with this ambiguity. It is a kind of natural reaction on the part of the audience to try to make a connection between the physical gesture they see and what they hear. However, to do so is difficult, because these sounds are unknown. (Tanaka, cited in Bongers 1998, p. 18)

Thus a ritual is created by the very "newness" of the media, combined with the communicative relationship between the receptive, actively listening audience and the sensitive, knowledgeable performer.

I think that what we are doing is really without tradition. Or if there is tradition, it would be right at the roots – music as ritual, trance, and pure energy. […] We want to communicate. Our concerts exploit energy, we want the audience to feel like they have just gotten their batteries reloaded. We want them to feel stronger and like better human beings. (Karkowski/Bongers, p. 23)

Tod Machover is also renowned for his work in refining and expanding the subtle expressivity of musical instruments (e.g. his collaborations with performers such as Yo-Yo Ma (1991) in his design of the so-called "hyperinstruments" – gestural controllers designed to work in conjunction with existing instrument such as the 'cello), and he is also a highly creative designer of completely new instruments and compositional models.

Much of his work involves the creation of instruments designed to be played by non-musicians. As opposed to the hyperinstruments, which are designed to enable professional performers to reach new levels of expressivity, instruments such as the Melody Easel, Gesture Wall and Drum-Boy are designed for ease of use, allowing amateurs to create and shape their own sounds and musical structures without being intimidated by instruments that are difficult to play. "If we could find a way to allow people to spend the same amount of concentration and effort on listening and thinking and evaluating the difference between things and thinking about how to communicate musical ideas to somebody else, how to make music with somebody else, it would be a great advantage" (Machover 1999, cited in Oteri).

Machover has put a lot of thought into how gesture and sound are intuitively joined, and has developed instruments that are designed to be played in such a way. With colleagues Maggie Orth and Gili Weinberg, he has designed musical toys designed for children such as the "Squeezables" and "Stretchables" – instruments made of soft, pliable materials that take full advantage of the malleability of synthesized sound, and thus simulate this flexibility in a fully experiential, tactile way. In 1996 Machover launched the first stage of a large-scale project called the *Brain Opera*, in which hyperinstruments and instruments for amateurs are used in a combination of live performance and real-time Internet activity. In this project, Machover is attempting to find a perfect balance between predetermined structure and performer/participant free will. "My interest lies in understanding the balance between central organization and anarchy – in our minds and in our lives. The Brain Opera is intended to encourage audiences to reflect on this process" (Machover 1996, unattributed interview).

Whereas the realization of Tanaka's artistic vision lies in his sharing a ritualistic experience with his audience, Machover is exploring the possibilities in creating a set of tools and circumstances that provide the audience with the opportunity to create their own musical ritual. "To create precisely the situation where somebody can do something really personal and special and contribute and feel like something wonderful has happened, that's I think a major goal for a certain kind of work that should be done now, and it's very hard to do" (Machover/Oteri 1999).

4 Conclusion

Although there is no one conceptual model for the communication of meaning, there is still the essential quality of *communication*. As human beings, we do not derive meaning from isolated elements, but from the relationships between elements (as applied in the comparative nature of spoken language). It is for this reason that, no matter what technology is being used or what conceptual/aesthetic framework is being communicated, the foundation of that communication lies in our ability to relate to that framework. As physical beings, we experience and relate to the world by sensation, so a metaphorical language built from the senses serves as an excellent format for the contextualization of meaning. This active, sensory format consists of the concept (mind), the sound (voice), and the gesture (body).

Ritual is essentially the communication of a concept through acted-out metaphors, and how those metaphors relate to each other and to the individual. Modern movement-controlled interactive technology is an ideal medium for the realization of these metaphors, in that it is not bound by any particular set of gesture-to-sound relationships. The ritual can be shaped and manipulated by the composer or by the participants to suit the relationships being explored, thus making it a much more personally relevant and participatory experience.

5 References

Bongers, Bert. 1998. "An Interview with Sensorband. *Computer Music Journal*. 22(1): pp. 13-24

Campbell, Joseph. 1959. *The Masks of God: Primitive Mythology*. New York (USA): Viking Penguin Inc.

Cott, Jonathan. 1974. *Stockhausen: Conversations with the Composer*. London (UK): Picador, Pan Books Ltd.

Daniélou, Alain. 1943 (rev. 1995). *Music and the Power of Sound: the influence of tuning and interval on consciousness*. Rochester, Vermont (USA): Inner Traditions International.

Garnett, Guy E. 2001. "The Aesthetics of Interactive Computer Music." *Computer Music Journal* 25(1): pp. 21-33.

Govinda, Lama Anagarika. 1960. *Foundations of Tibetan Mysticism*. London (UK): Rider and Company.

Iazzetta, Fernando. 2000. "Meaning in Musical Gesture." *Trends in Gestural Control of Music*. CD-ROM released by Ircam – Centre Pompidou: pp. 259-268

Maconie, Robin. 1976 (2nd ed. 1990). *The Works of Karlheinz Stockhausen: 2nd Edition*. Oxford (UK): Clarendon Press.

MIT Media Lab Website. http://web.media.mit.edu/~tod/

Nyman, Michael. 1974. *Experimental Music: Cage and Beyond*. New York (USA): Schirmer Books.

Oliveros, Pauline. 1976. "On Sonic Meditation." *Painted Bride Quarterly*. 3(1). Reprinted in 1984. *Software for People*. Baltimore, MD (USA): Smith Publications.

Oteri, Frank J. 1999. "Technology and the Future of Music." *NewMusicBox*. E-Journal http://www.newmusicbox.org.

Scientific American interview Website: Interview with Tod Machover. 1996. http://www.sciam.com/interview

Tanaka, Atau. 2000. "Musical Performance Practice on Sensor-based Instruments." *Trends in Gestural Control of Music*. CD-ROM released by Ircam – Centre Pompidou: pp. 389-405.

Von Gunden, Heidi. 1983. *The Music of Pauline Oliveros*. Metuchen, NJ (USA) and London (UK): The Scarecrow Press, Inc.

Winkler, Tod. 1999. Composing Interactive Music: Techniques and Ideas Using Max. Cambridge, MA (USA): The MIT Press.

Composing Space: The Integration of Music, Time, and Space in Multi-Dimensional Sound Installations

Robert Coburn

Conservatory of Music, University of the Pacific
email: rcoburn@uop.edu

Abstract

The creation of art to be situated within an architectural setting, landscape, or other three dimensional space offers an artist the opportunity to consider relationships beyond those normally associated with any individual art form. The art work will be experienced within the context of the total environment and its success will depend upon how it interacts with or integrates into its surroundings. When the art work utilizes sound as its primary material, the possibility for integration is extended beyond the consideration of architectural and visual space to include both temporal and pitch space. The need for developing a creative approach which, within its design, integrates all aspects of the experience, becomes fundamental to the creation of the work and to the work's integration into the space. In my own work, 39 Bells, physical space, temporal space, pitch space, and musical structure are integrated through a process which unifies all aspects of the experience around the symmetrical nature of the original environment. This integration will be discussed at length. The experience of a mile long soundwork constructed from 39 bells hung along a public street offers interesting perceptual challenges. The role the body plays in parsing, measuring, sensing, and relating to the reality of the full scope of the constructed site will be discussed with the intent of understanding how a listener can find meeting points between his or her position within the work and the physical presence of the work itself. In conclusion, Between...Beyond, my recent temporary gallery installation, will be described with reference to how these larger principles effect a more local experience through the use of sound, image, and object to manipulate space, time, memory, and imagination.

Keywords

soundworks, perception, public art, symmetrical structure, musical structure, installation, site-specific

1 The Experience of Listening

Traditionally concert music is performed in a setting optimized for listening, the underlying assumption being that music is primarily an aural experience, but for centuries composers have shaped their music using the structural language of spatial art forms applied to temporal relationships. The music theory literature documents many examples of composers, from Dufay (Powell 1979) to Debussy and Bartok (Howat 1986) drawing on the golden section as a means of structuring durational form and pitch space on all levels from the smallest scale to the entire work.

The 20th century has seen an increase in the application of spatial relationships to musical structure through the influx of technology. Earlier experiments with the spatial placement of performers throughout the concert hall have given way to the multiple speaker electroacoustic projection of works created with and by computers, and the "home theater" 5.1 surround-sound system. In all cases the music is distributed spatially but the architectural construct still takes place at the perceptual level as it always has in traditional music. Music is perhaps inherently architectural in its structural manifestation but through the process of listening alone the construct is inevitably a mental one. In this sense there is little difference between experiencing a Beethoven string quartet, a spatially distributed acoustic composition by Henry Brant, or an abstract multi-channel computer music work. As long as music is defined by the abstractions of pitch/sound structure and confined to the space in which it is presented, its architectural character is destined to represent only a musically "simple" space and to occur only at the perceptual level. Until a listener moves through a work physically and experiences it from different spatial and locational perspectives or, through the experience of referential sounds, mentally perceives other spaces beyond the confines of the physical space itself, music remains simply music.

2 Perception as Spatiality

In The Phenomenology of Perception (1962/92), French philosopher Maurice Merleau-Ponty puts forth the treatise that

". . . all senses are spatial if they are to give us access to some form or other of being . . ."

This clear and in some ways simple statement reinforces our understanding of the relationship of music experience to spatial structure. Merleau-Ponty posits a philosophy of perception where "all (senses) open on the same space", a space experienced as spatiality. Each sense provides its own window on this space and its own particular embodiment of spatiality. He says, "sensation as it is brought to use by experience is no longer some inert substance or abstract moment, but one of our surfaces of contact with being, a structure of consciousness," and ". . . as the universal condition of all qualities, we have with each (sensation), a particular manner of being in space and, in a sense, of making space." (Merleau-Ponty 1962/92)

Maurice Merleau-Ponty sees the experience of music, as he does all sense experience, as spatial. "Music is not in visible space, but it besieges, undermines, and displaces that space . . .". He sees the interaction of all perception within the spatial domain as a "primary organization" of experience. Perception brings all senses together in the meeting point of spatiality and at that point creates a unique, shared space of experience (Merleau-Ponty 1962/92).

3 Composing Space – 39 Bells

In 1993 I received a commission from the Public Art Program of the city of Philadelphia to create a permanent soundwork for the Avenue of the Arts, a mile long section of south Broad Street that extends north and south from the City Hall to Washington Avenue. I had previously completed a large-scale soundwork for the Oregon Convention Center (Bell Circles II, 1987-91) and was actively seeking opportunities to expand my creative interest in site-specific, spatially distributed soundworks.

The city of Philadelphia's intention was to centralize arts activity along the Avenue of the Arts and create a focus for the downtown area that would enliven the city. Arts organizations along Broad Street include the Academy of Music, the Philadelphia Orchestra, the Opera Company of Philadelphia, the Pennsylvania Academy of Fine Arts, the Clef Club (a prominent venue for jazz performance), the Brandywine Workshop and gallery, the Merriam Theater and many others. The city was seeking a site-specific artwork that would identify the unique nature of this mile long arts area while unifying the diverse elements along the street.

When working on an expanded physical scale sound offers many unique aspects not found in other art forms. Sound objects can be located to define specific points in the environment while the sound disperses in all directions to fill a much larger space than the object itself occupies. Depending upon the needs of the site, sound objects can either provide significant visual elements in the environment or practically invisible ones with only their sound identifying their presence. Sound objects can create a presence that activates a space and/or defines a place of sanctuary. And, most importantly, sounds themselves distributed throughout a larger environment make possible a link between physical/spatial relationships, temporal relationships, pitch relationships, and musical structure.

3.1 Site-Specific Considerations

Any site-specific project must develop directly from the needs and opportunities of the site itself and the community that inhabits it. In the case of the Avenue of the Arts project the area covered one mile running north to south. The northern half, starting at the city hall, was predominately well established businesses, restaurants, and arts organizations (orchestra, opera, theater), and actively represented the heart of an historically old but modern city. The southern half was less affluent with apartments, gas stations, fast food restaurants, some deserted buildings, and a surrounding residential neighborhood. One of my first decisions was to incorporate this north/south division in a way that would equally represent the two halves of the environment while also delineating the difference in current use. This led me to the decision to use symmetrical structures and mirrored relationships as a means for representing and unifying these two spatially equal but environmentally different areas as one.

3.2 Symmetrical Relationships

Mirrored, symmetrical relationships are a common structure in traditional music often represented as an A B C B A arch form. This structure is symmetrical across the center section and can be mirrored at the same point. It can reflect this relationship at various levels of structure often including similar mirrored symmetries within smaller musical sections and, at times, within the choice of pitch and rhythmic material as well. For this project I chose to hang 39 bells from lampposts installed along the length of the street and to represent both the differences between the use of the north and south halves and their unification as one community by symmetrically mirroring the pitches of the bells across the axis dividing north and south. Bell pitches begin at the north end of the avenue with g4, ascend chromatically through 19 bells to d6 at the center of the area and then reverse, descending back down to g4 at the southern end. The choice of G as the fundamental pitch served to relate the street to the site because of the presence of an existing bell (low G) at the top of the PNB building which anchors the north end of the street. The pitches along the street ascend over one and a half octaves from the northern end of the street to the mid-point and then reverse order and descend back down to the starting pitch at the southern end creating a symmetrical mirror of pitched material and sounding objects. This initial site-specific decision, growing out of the nature of the environment itself, became a determining factor for all subsequent structural choices made in the design of the physical, musical, and temporal aspects of 39 Bells. The existing environment provided the key to the integration of all elements within the site.

With the physical/spatial aspect of the work determined, the temporal and musical structures developed in close integration to the physical structure. 39 Bells rings musical patterns on a daily basis. The musical activity of these patterns reflects the usage of the street at the time of ringing. The temporal structure of these patterns across the day is based on a symmetrically mirrored arrangement of ringing times. The work day naturally assumes a symmetrical pattern of activity. Ringing times are shown in figure 1.

8 a.m.	opening
8:15	Ia
8:30	Ia+b
8:45	Ia+b+c
9:00	Ia+b+c+d
10:00	II
11:00	III
12:00	IV
12:15	Va
12:30	Va+b
12:45	Va+b+c
1:00	Va+b+c+d+c+b+a
1:15	Vc+b+a
1:30	Vb+a
1:45	Va
2:00	IV (retrograde)
3:00	III (retrograde)
4:00	II (retrograde)
5:00	Id+c+b+a
5:15	Ic+b+a

Figure 1. Schedule of daily ringing times.

The ringing schedule is organized to mirror times across 1 p.m., the center of the business day. Each ringing between 8 a.m. and 1 p.m. has a corresponding symmetrically mirrored time for ringing between 1 and 6 p.m.. Ringings reflect the use of the street with more ringings occurring during the hours of 8 to 9 a.m. when people are on their way to work, 12 to 2 p.m. when they are out of their offices for lunch, and 5 to 6 p.m. when they are leaving work for home.

The musical structures of these ringings are also organized by symmetrically mirrored relationships. The pattern of each ringing is composed of one or more modular section. These modular sections accumulate between 8:15 and 9 a.m. and 12:15 and 1 p.m., and disperse during their corresponding mirrored ringings between 1 and 1:45 p.m. and 5:00 and 5:45 p.m. reflecting the higher level of street activity during these times. At times of less activity the ringings are made up of single sections. The musical structures are shown in figure 2.

8 a.m.		1:15
8:15		1:30
8:30		1:45
8:45		2
9		3
10	1 p. m.	4
11		5
12 noon		5:15
12:15		5:30

Figure 2. The musical structure of ringing patterns across the day in 39 Bells.

All ringings that occur after 1:00 p.m. are reversed ringings of the patterns prior to 1:00. In some instances both pitch and rhythm are mirrored while in others pitch is mirrored but rhythm is not. The central ringing at 1:00 contains within itself a retrograde of its first half. The closing at 6:00 p.m. is a retrograde variation of the opening at 8:00 a.m. allowing for the different experiential nature of the time of day each occurs.

The temporal structure of the ringings across the day and the corresponding musical structures represent an integration of the mirrored symmetricality that occurs in the physical structure of the sounding objects and the pitch system on the street. This becomes evident when, due to the reversal of the physical positioning of the bells along the south and north ends of the street, patterns which sound on their corresponding bells at the north and south ends simultaneously play as mirrored aural images across the axis dividing north and south. At its most fundamental, all aspects of 39 Bells are related to the initial observation of symmetricality evident in the existing environment.

4 Experiencing 39 Bells

Experiencing a mile long soundwork constructed from 39 bells hung along a public street offers some interesting perceptual challenges. Certainly the entire work cannot be heard from any one position. Each time a ringing pattern is played a single listener hears only a fragment of the work's greater self. This is one characteristic of 39 Bells that sets it and most other environmental soundworks apart from other forms of musical expression. The physical expanse of the work makes it necessary for an individual to assemble the experience through repeated listenings from a variety of physical locations. This can occur through an act of deliberate attention or through the slow accumulation of experience, living with and hearing the work on a daily basis. In both cases the demands of experience more closely resemble the manner in which individuals perceive and understand architectural structures. On an on-going basis the listener finds meeting points between his or her position within the work and the physical and aural presence of the work itself. The relationship of the body to the physicality of the work is fundamental. The dimensions of the work are measured and experienced only by physical interaction with and movement through the space itself.

This is often the missing element in the correlation of architectural systems and musical systems: the physical experience accumulated through time of the unseen and unheard expanses of the work. Through a slow accumulation of information the viewer/listener assembles a larger sense of the whole which can never be perceived simultaneously. Although both music and architecture do in a sense unfold over time, the physical unfolding of architecture is seldom matched in the musical experience except through large scale installations such as 39 Bells. The vital presence of the body in experience and the role it plays in parsing, measuring, sensing, and relating to the reality of the full scope of the constructed site can only be matched in an equally expanded musical experience that requires the same active role for the body. Through the body architectural scale and structure are measured and understood. Through the body musical scale and structure may also be sensed and measured to bring an eventual experiential understanding even to the most expanded of works as music, sound, space, and environment become one.

5 Between… Beyond

How then does this relate back to the integration of all aspects of the experience of site-specific installations designed for and confined to a physically limited space? Between… Beyond is an installation piece I created for the Richard and Marjorie Reynolds Gallery. It was exhibited from March 26 to April 25, 2002. Designed for a space of approximately 24 x 24 feet square it appears to place no unique demands on the listener/viewer. All aspects of the work are simultaneously heard and seen within a confined space. But in both design and experience it shares much in common with 39 Bells and other extended works.

Between… Beyond consists of 6 boxes of varying heights placed at specific points within the space to create a balance between themselves as objects and the emptiness surrounding them. They define the spatial aspect of the room without creating any obvious division of the space. Each box is illuminated from within and houses two superimposed glass lantern slides whose images demonstrate early 20th century acoustical experiments to make invisible waveforms visible. The slides are seen on the upper surface of the boxes located and oriented on each in a unique position. Sound plays into the space from four speakers mounted high on each wall and located along each wall in a similar relationship to that used to locate the boxes within the space. Two stereo audio cds play simultaneously into the space, their lengths, c. 27 and 30 minutes respectively, allowing them to loop back and cycle against each other to create a constantly changing layering of sound across the day. The sounds used in this installation were collected as site recordings during an extended stay in Japan. They include the sounds of temple bells, Buddhist rituals, nature soundscapes, city sounds, and some excerpts from my own computer music performance piece, PatternsLuminous. For the most part the sounds are presented without modification. Integration of time and space occurs through the use of a similar temporal proportion to that used to locate the speakers along the walls and the boxes within the space to distribute the sounds in time on the cds. Silence divides the extended sound passages to create a balance between the sounds themselves and the emptiness of the silence around them.

All aspects of this work serve to draw the listener into a relationship with the spaces between objects and beyond direct experience. Katherine Norman (1996) refers to real-world music, music which draws on sounds collected from the natural soundscape, as not being "concerned with realism because it seeks, instead, to initiate a journey, which takes us away from our preconceptions, so that we might arrive at a changed, perhaps expanded, appreciation of reality." It is this sense of journey, of physical and experiential transport and expanded reality, that links Between…Beyond to 39 Bells. In the confined gallery space it is not the distribution of sounds throughout the space by a four-channel sound system that creates an expanded spatial experience but rather the nature of the sounds themselves that draws the listener into an evocative, expanded space, shaped by memory and beyond time. In the expanded physical space of 39 Bells it is not the sounds of bells alone but the physical interaction of the listener/viewer with the sounding environment that unifies the experience of time, space, and sound to ultimately create a "changed, perhaps expanded appreciation of reality". (Norman 1996) In both works it is the eventual renewed understanding of our place within that reality that rewards our efforts.

References

Coburn, R. 2001. “Composing Space.” *Proceedings of the 3rd Annual Symposium on Systems Research in the Arts*. International Association for Informatics, Cybernetics, and Systems Research, pp. 8-12.

Howat, R. 1986. *Debussy in Proportion: A Musical Analysis*. Cambridge University Press.

Merleau-Ponty, M. 1962/1992. *Phenomenologie de la perception* (Colin Smith, trans.). Routledge Press.

Norman, K. 1996. “Real-World Music as Composed Listening.” *Contemporary Music Review* 15(1-2): 1-27.

Powell, N. W. 1979. “Fibonacci and the Gold Mean: Rabbits, Rumbas, and Rondeaux.” *Journal of Music Theory* 23(2): 227-273.

Smart Controller —Artist Talk

Angelo Fraietta

PO Box 859, Hamilton NSW, 2303
email: angelo_f@bigpond.com

Abstract

The Smart Controller is a portable hardware device that allows performers to create music using Programmable Logic Control. The device can be programmed remotely through the use of a patch editor or Workbench, which is an independent computer application that simulates and communicates with the hardware. The Smart Controller responds to input control voltage and MIDI messages, producing output control voltage and MIDI messages (depending upon the patch). The Smart Controller is a stand alone device -- a powerful, reliable, and compact instrument -- capable of reducing the number of electronic modules required, particularly the requirement of a laptop computer, in a live performance. This talk will detail the progress of the project.

1 Introduction

I discussed the Smart Controller at Waveform 2001 (Fraietta 2001), explaining the methodology whereby I was able to develop the software for the Smart Controller using a desktop computer as a hardware simulator. The research, however, had to eventually lead me to the development of a physical hardware device (otherwise the Smart Controller would always remain a theoretical device), which in turn spawned two other products that have commercial potential, enabling the research to partly fund itself before its completion. Brief details of these products will be given later in the document.

The last year has been quite successful in that I have been able to develop a prototype of the Smart Controller in a stand-alone box. The other side of the coin, however, is that this research has been extremely taxing – mentally, emotionally, and spiritually -- in that there were many problems that occurred that made it look as if the project was doomed to failure; however, the people acknowledged at the end of this paper helped to ensure this was not the case.

2 The Resource Model

I had originally intended using a PC-104 embedded PC to run the Smart Controller under RTEMS. I found on the RTEMS newsgroup that a particular PC-104 system had successfully run RTEMS (Wasierski, 2001), and obtained the supplier details. I found the URL describing the device online, however when I jumped to the online catalog, it led me to the DIMMPC. The DIMMPC evaluation board had a PC-104 interface and so I assumed this was the device I required. I bought the evaluation board and DIMMPC (costing approximately $1300 after shipping, customs and GST), and upon opening the box found that I did not get a PC-104 single board computer at all -- the DIMMPC was a whole 386 PC motherboard on a single chip. I had originally intended returning the items, however, after connecting the evaluation board to a monitor and power supply, I found that I was able to run the Smart Controller software on it with no problems at all. This effectively gave me a smaller device, however, I had to design a circuit board and obtain the 144-pin socket to mount the chip. This has now turned out to be a very cost effective and space efficient alternative to the PC-104 system that I had originally intended. Additionally, an advantage of this system is that someone can upgrade to a faster CPU simply by replacing the chip – a viable alternative if I choose to make the Smart Controller perform DSP.

The next stage in the development was the implementation of the MIDI and control voltage I/O. I implemented this using the 16F877 PIC microcontroller. NKA (formally Neil Kilgour and

Associates) provided me with parts, development tools, and engineering advice at no cost, thus making this stage of development relatively painless. I designed the circuit board as a separate module to the Smart Controller CPU, thus enabling the board to be used as a standalone CV to MIDI / MIDI to CV controller, without the Smart Controller. I originally offered these on the ACMA post at cost, however, I received a lot of negative feedback, particularly as I did not offer an option to configure the device. I found that I was able to store configuration data within the EEPROM of the device, and as such was able to configure the device using software through the MIDI ports. What has eventuated from this is that I now have a fully configurable CV (control voltage) to MIDI / MIDI to CV controller available for sale at a very competitive price. These devices are now sold internationally via the Internet. This is the first example of the Smart Controller research obtaining a source of funding generated from technology developed before the project's completion. The devices have been designed so they can be easily upgraded to a Smart Controller later.

A multi-object file stream developed for communication with the external Patch Editor became another source of funding. Quikscribe adopted this steaming methodology for their Digital Transcribing, whereby they have created an Intelligent Audio File (IAF).

> The .iaf (Intelligent Audio File) provides the Quikscribe Transcription System with a powerful "unique advantage". Rather than just being able to record, edit and transcribe audio files, the .iaf (Intelligent Audio File) provides Quikscribe with the ability to offer a lot of advanced features, not currently available in any other dictation/transcription product. For example, Quikscribe is able to insert Text Attachments, capture Screen Shots and insert File Attachments. It also has a powerful Built-in Database for management purposes. Lastly it can record, edit (Undo or Redo) and compress audio in real-time. (Quikscribe, 2002)

Apart from the financial benefits, this has led to the satisfaction of providing solutions for industries outside of Creative Arts.

The next stage in development was the intercommunication between the PIC I/O card and the 386 Smart Controller card. Communication between the two boards was achieved by using a PLA (programmable logic array), which communicated with the 386 in a parallel data stream, while communicating with PIC using a serial data stream. In developing this area, I found that there were many areas that errors could and did occur, which had to be identified and corrected. The biggest problem, however, was the speed of the data interchange between the two boards. The maximum acceptable interchange between the two boards must be no greater than 320 microseconds as this was the maximum MIDI transfer rate. I was unable, however, to get the rate below 450 microseconds on the DIMMPC, which in turn caused the device to lose MIDI data bytes. I created this situation by sending continuous sysex blocks of 1024 bytes into the device from my PC MIDI output – this caused a MIDI overflow to occur. I ultimately had to reduce the amount of time in the interrupt and the data exchange. Joel Sherrill asked “Aren't you down to the point of counting instructions?” (Sherrill, 2002) I thought that this was some sort of programmer’s figure of speech, however, I found out that this was exactly what was required – I had to actually count the number of CPU instructions that were taking place in the exchange. I performed this by stepping through the code in the MPLAB simulator and was able to reduce the time by implementing some methods in the PIC that are normally considered poor programming practice. The first methods I used were implementing global variables instead of passing function parameters, and using “USE_FAST_IO” directives to prevent unnecessary changes to the data direction registers. This proved effective; however, the exchange rate using the DIMMPC was still 370 microseconds – 50 microseconds too slow. The next method I used was actually counting the instructions using the MPLAB simulator and comparing my “C” code with the compiler generated assembly code. Consider the following code fragment, which writes the most significant bit of a variable to a pin of the PIC.

```
output_bit (SPI_PLA_DATA, pla_out_data.flags & 0x80);
```

This fragment took five machine cycles to execute. The following fragment performs the same function, however, only taking three cycles.

```
if (bit_test (pla_out_data.flags, 7))
```

```
{
  output_bit (SPI_PLA_DATA, 1);
}
else
{
output_bit (SPI_PLA_DATA, 0);
}
```

This saving of two instructions actually becomes a saving of sixty-four instructions as the code is executed thirty-two times per exchange. The biggest saving, however, was in the omission of "for" loops in the exchange. The following statement causes the code within the loop to be executed eight times.

```
for (byte_num = 0; byte_num < 8; byte_num++)
```

The problem with the code, however, is that it takes ten cycles every time to perform that line of code, which becomes an eighty machine cycle overhead to the loop. This type of iteration is performed four times per exchange, becoming 320 machine cycles of overhead, which translates to eighty microseconds per exchange. I overcame this by placing the code from the block within the "for" loop into inline functions, and literally called them each eight times. These changes enabled the exchange between the two processors to take place in 250 microseconds – well within the required time. This, however, produced another problem – the speed was now too fast in that sometimes the 386 did not sense the interrupt, which in turn caused a lock up when there was no MIDI or CV input at the PIC. This problem was overcome without too much difficulty.

The next problem encountered caused me the greatest distress in the entire project to date. I had the simulator on the Windows machine running well for over a year, however, attempting to run some patches on the RTEMS machine would create access violations that caused the machine to crash if I clicked madly on the Patch Editor. I searched for days, unable to find where the problem in the code could be. The whole point of the simulator was that the majority of the code was identical, and so it would be easy to find the problems by debugging the Windows machine. The problem, however, was that it would not crash on the Windows machine. After three days of intense debugging, I was physically ill from the stress. That night, I cried out in anguish "Lord, I can't find it! It is beyond me. You have to show me where the problem is." The next day I sat down at the computer and started clicking madly on the Windows Patch Editor. Almost immediately, I received a Code Guard error message. Code Guard generated a report to a text file that actually showed me lines of code that had the error. This was miraculous! The error had been in my code for more than a year; however, it did not show itself until that moment. The reason that it occurred so regularly on the RTEMS machine was because the code runs faster in RTEMS on a 40MHz 386 than it does on a Windows 2000 machine running at 1.133GHz. This supports the concept that a machine designed specifically for this purpose would probably be more effective as an instrument than a laptop or desktop computer running software -- such as Max, PD, or Algorithmic Composer -- because the specific machine does not have to waste time performing unnecessary operations such as updating displays, servicing the many tasks that the operating system starts up, etc…, but rather allocates CPU resources only where they are required. I have been able to make the Smart Controller generate a 50 HZ square wave by toggling the digital output every ten milliseconds using a metronome object, while at the same time generating 100 MIDI messages per second out of each of its MIDI output ports. I have measured the digital waveform with an oscilloscope and was surprised to find that the pulses were exactly 10 milliseconds wide – I would not expect to see this level of accuracy using a non-dedicated device.

3 Conclusion

The Smart Controller is now at a stage where it could actually be used in an installation or performance. There are, however, still more features that I would like to add to the hardware device such as non-volatile storage and retrieval of last loaded patch in the case of a power outage at an installation. Additionally, I must now commence work on the patch editor in order to make it run on multiple platforms. I hope to have prototypes of the Smart Controller available for testing by the beginning of 2003, and hopefully, presenting the device at the International Computer Music Conference (ICMC) next year.

4 Acknowledgments

I would like to firstly give thanks to God in that His grace and mercy "delivered me from the paw of the lion"(Holy Bible 1978, I Samuel 17:37) in that it was often not until I broke down and specifically asked for His help did the solutions to the problems become revealed; to Neil Kilgour for generously providing parts, equipment, and technical advice; and to all the people who participate on the RTEMS users group.

References

Fraietta, A. 2001. "Smart Budgeting for a Smart Controller." *Proceedings of Waveform 2001 -- Australasian Computer Music* Association *Conference*. Australasian Computer Music Association, pp. 43-47.

Holy Bible. New International Version. 1978. New York: International Bible Society.

Quikscribe. 2002 "iaf (Intelligent Audio File." http://www.quikscribe.com.au/PopUps/iaf.htm.

Sherrill, J. 2002. "Re: Interrupts occurrence." RTEMS User Newsgroup Archives, http://www.oarcorp.com/rtems/maillistArchives/rtems-users/2002/march/msg00210.html

Wasierski, R. 2001. "Re: PC-104." RTEMS User Newsgroup Archives, http://www.oarcorp.com/rtems/maillistArchives/rtems-users/2001/august/msg00129.html

Space As Structural Function In Electroacoustic Music

Elsa Justel
Composer, Doctor in Aesthetics, Sciences and Technology of Arts, University of Paris 8
email: justelsa@wanadoo.fr

Abstract

This paper tries to demonstrate some of the strategies used by different composers to create their compositional spatial models. Throughout our lecture, we will hear some examples of space organization and demonstrate their physical parameters by means of Fourier analysis. We will see different methods of spatialization by temporal pitch changes, amplitude panning changes, spectral transformations, re-synthesis by convolution and transposition, expansion, granulation etc) on works by the composers J.C.Risset, H.Vaggione, B.Truax and others.

Introduction

The use of space as an element of organization is one of the new preoccupations for composers of electroacoustic music. This spatial dialectic is the most relevant topic in electroacoustic music. Space has acquired a structural role that was absent in instrumental music. However, the morphological analysis of this aspect is difficult, not only because we have not the appropriate tools to measure it, but also because we are confronted by multiple spaces.

This new alternative of considering space as a musical and structural phenomenon requires a different kind of hearing, more demanding from a physiological and psychological point of view.

It is useful to look at the works themselves and to try to extract descriptions of the different behavior of materials in relation to space in electroacoustic music. We can perhaps determine a kind of syntax of space.

1. Internal and external spaces

To begin with, we must consider two principal aspects: the internal and the external spaces. Both phases of space operate on our perception in a way that is at the same time both centrifugal and centripetal. There is a diversity of criteria about the importance of space and its role of articulation in electroacoustic music, but in any case the interaction between time, space and timbre constitutes the crux of a new conception of musical structure. This most ancient tendency is intended to give different types of expressive relief to music by manipulating the sound parameters. They were inspired on the instrumental music concept that space amplitude is a result of the combination of frequency, intensity and harmonic spectra. In fact, the differences in spectral field between low and high sounds (that is the consequence of absorption rate of partials), as well as relationships between duration and registers, are determining agents of space in instrumental music.

Most acousmatic composers talk to us about virtual and real space, that is: the space created during the composition, and the space of performance. In the first case, the internal space will be incorporated to the material and stay fixed in the support. The external space will enter into action at the moment of projection, during the concert. This space is then a composition act, as well as an interpretation one. In this way both spaces become complementary. The audience will have the sensation of an imaginary space emerging from the environment all around. The composer, however, is conscious of the physical reality of space, in both forms, even if he applies his own flow of imagination to the composition. In fact, it is by means of perception that space acquires a symbolic sense, because brain mechanisms award it subjacent signification.

1.1 Models of internal space

There are numerous resources, but even by means of spectral analysis, we cannot determine which techniques were employed to obtain these effects. Nevertheless our experience with computer treatments can help us, by simple hearing, to know how composers achieved such results.

We have several very simple procedures, applied in the mixing of tracks. For example to slide elements in both stereo tracks, or to give different amplitude panning to each track in order to get larger or narrow perceptive spaces.

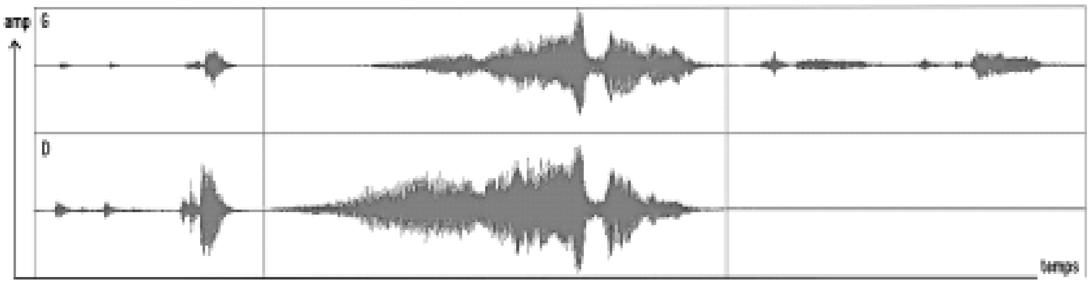


Figure 1 - Savouret : “Scène d'intérieur” [1]

The design of both stereo tracks makes evident the manipulations with the mixing table. We can observe that the evolution of amplitude curves represents objects and voice displacements in the space. The first two segments show a movement from right to left and the third segment represents a vocal sound fixed clearly at left.

We detected a similar mixing situation in “Petit Poucet Magazine” de B. Ferreyra[2].

In figure 2 we can observe a shifting at the same time temporal and of amplitude, between both stereo tracks.

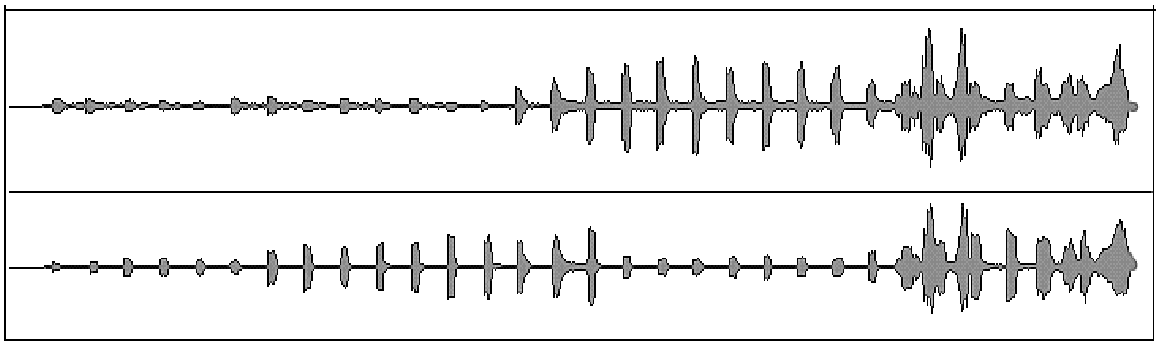


Figure 2 - Ferreyra : “Petit Poucet Magazine”

Besides, in spite of the similarity between both voices, we can see that there is still another additional shifting: this of pitches. I mean, both melodic lines have the same common elements but they are not identical. Melodic lines have an oblique trajectory: they begin at the same pitch and then they cross to go in opposite directions.

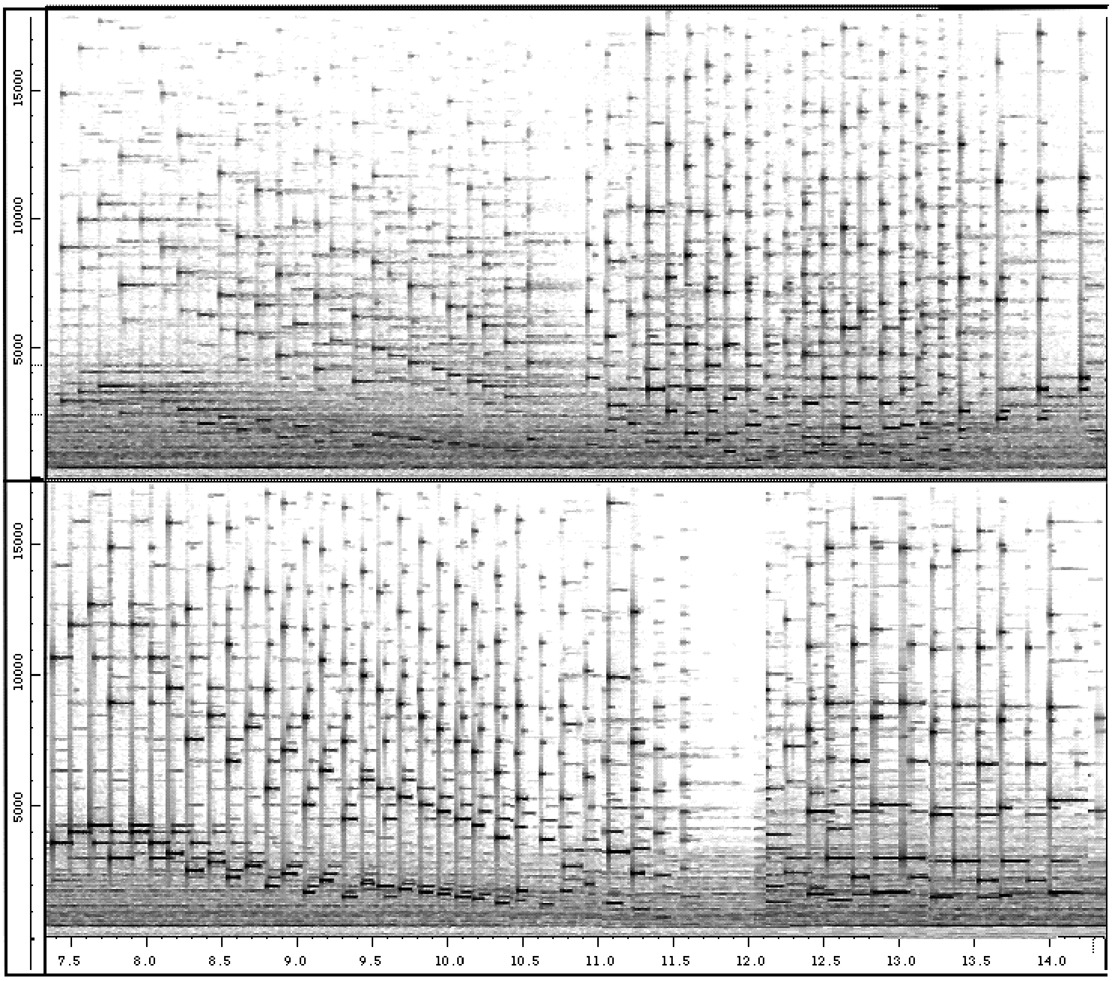


Figure 3 [3]

Another way to produce sensations of virtual space is by using temporal pitch changes. The classical examples are Little Boy and Mutations of J. C. Risset [4-5], based on Shepard's [6] experience with the illusion of indefinite glissando. Using an additive synthesis program, Risset produced an ascendant one octave scale in which each partial is doubled one octave higher. So that in a loop we have a sensation of infinitely descending pitch because when one component reaches the end of the curve a new component appears at the beginning.

………………

Other more complex procedures are those which apply transformations into the spectral components. We know that perceptual magnitude of sound depends on its spectral richness, its duration and non-synchronized temporal partials.

We will now see an example of composer H.Vaggione [7]. In this figure we can see three instrumental sound spectra. Sounds 1 and 2 belong to Double bass and sound 3 is a pizzicato of cello.

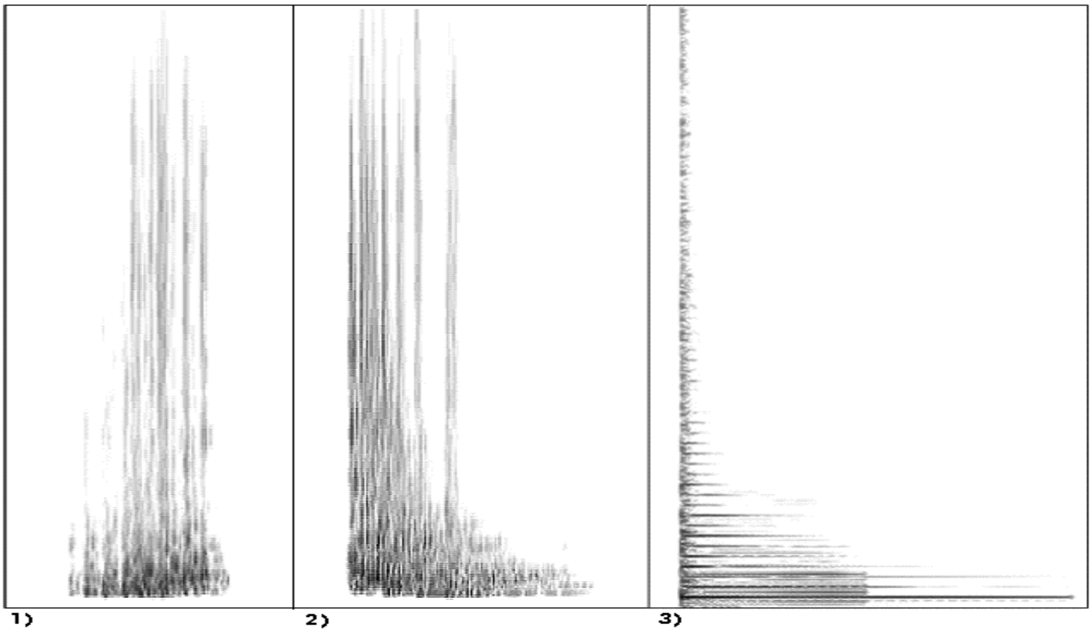


Figure 4

Let's see how the composer transforms sound 2. By means of re-synthesis by convolution the composer obtains a kind of filter that raises higher partials of spectra. (5a)

We can see the doubled vertical lines after the attack, these are the higher harmonics. As high frequencies are more piercing as far as perception is concerned, we have the sensation that the Double bass sound is nearer in this example, even if it has still the same qualities of timbre and duration, as the original.

The example that shows fig. 5b corresponds to the same sound, in this case with a slight panoramic shifting. We can see a kind of “out of phase” spectra in the attack transient and in the first partials. On the other hand we also observe that superior harmonics fall down faster than in example a.). Besides, the design seems to be flattened and the whole ensemble of the object seems larger. However its duration is still the same (1s750ms). The perceptual sensation is of a trajectory going from right to left.

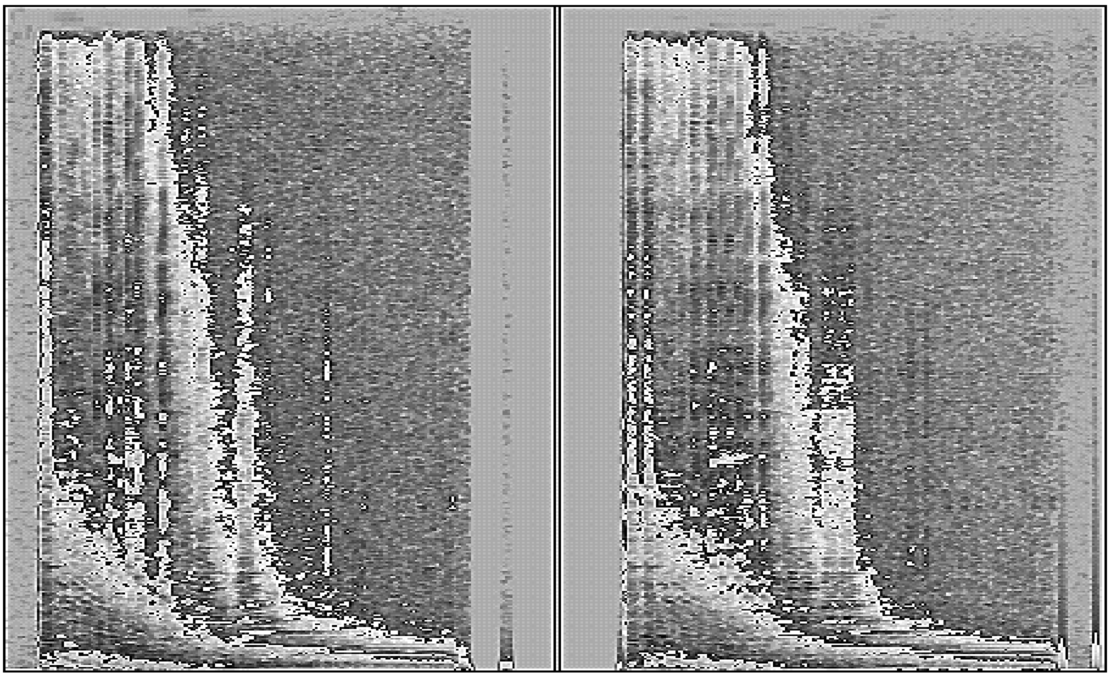


Figure 5.a Figure 5.b

The expansion process is also interesting, because in special conditions it can produce a sensation of being far away. Figure 6 shows sound 1) transformed by expansion. This kind of process sometimes produces little deformations in signal. Comparing this spectra with its original, we can see that both have the same morphodynamic aspect. In fact, frequencies are the same but the object format is larger and we perceive it as being far away.

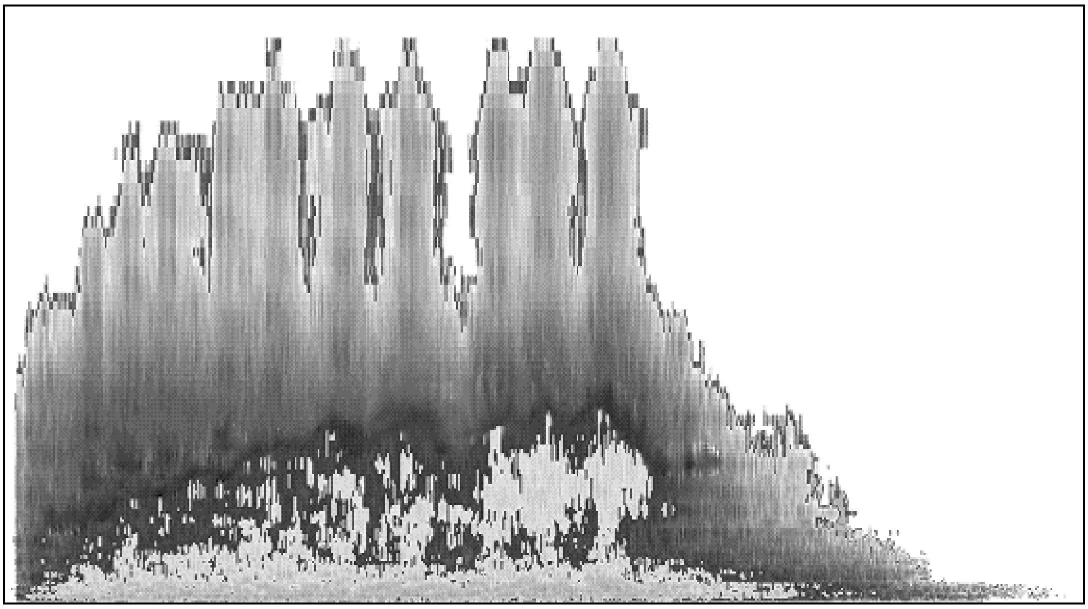


Figure 6

By granular methods we can obtain effects of polyphonic space. Spectra in Figure 7, correspond to a counter-tenor voice sound (belonging to the piece of B. Truax [8] "Powers of two"). Over those spectra, the composer used effects of granulation, expansion and transposition. (Figure 8) As a result of these treatments Truax obtained a thick mass that evolves very slowly. This mass shows a remarkable contrast with the original sound whose spectral quality is mostly pure. In that way, the composer got a perceptual illusion that embraces a large spatial field. We can verify that enlargement of spatial field in the extension of signal as well as in thickness of spectral lines. That means that we perceive a larger space both in field depth and in density of material.

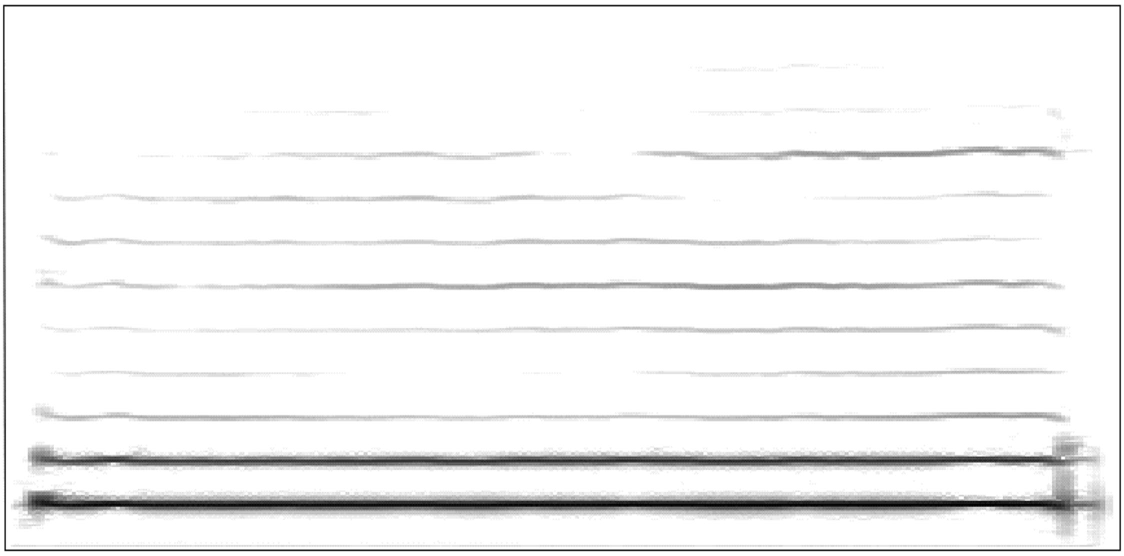


Figure 7

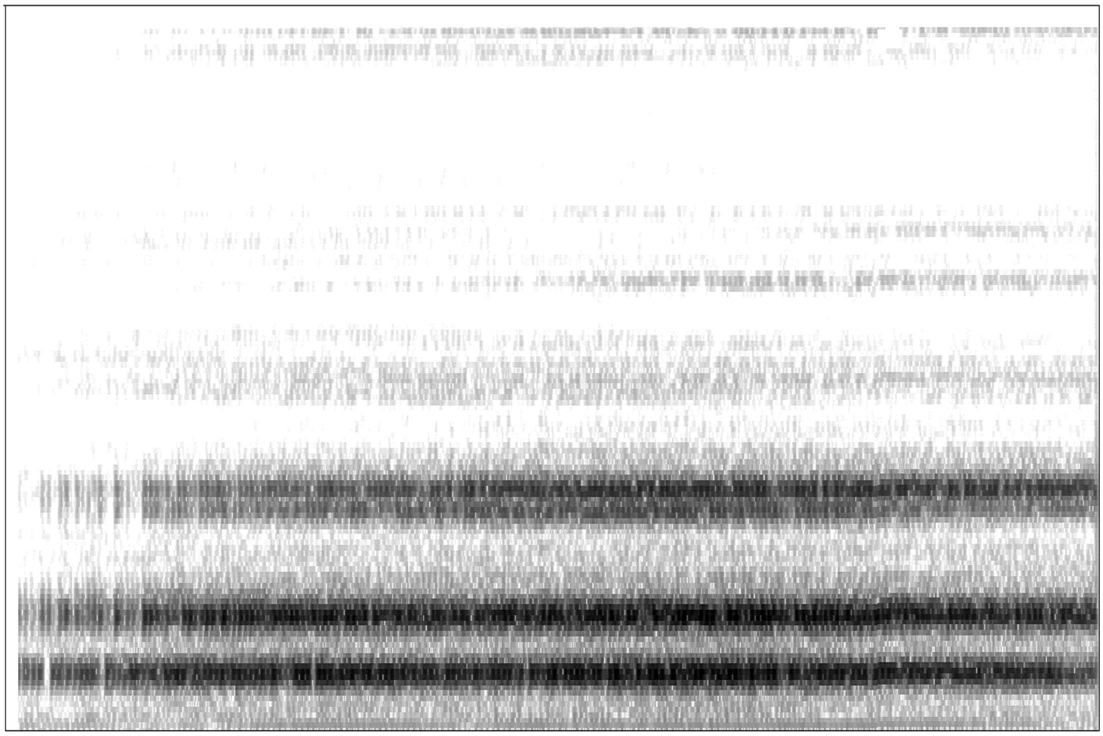


Figure 8

These and many other models of sound space can be integrated with music in a functional way. We can hear the example of C. Zanési in Arkheion [9], who uses a characteristic spatial scheme as a kind of “leit-motiv”. This scheme is constituted by lateral trajectories and fixed points in different plans of space. The composer always uses the same scheme to delimitate sequences and to create a good balance between different articulations. Thus it is evident that the composer thinks in terms of a structural element. (Figure 9)

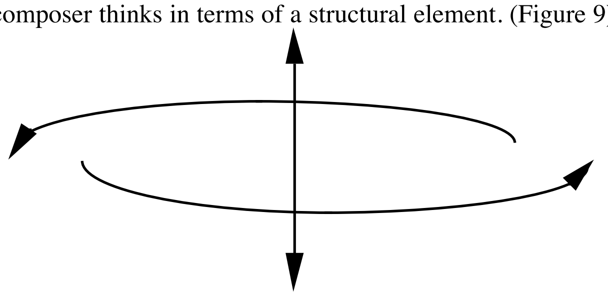


Figure 9

The gradual temporal transformation is one of the preferred resources of electroacoustic music composers. This aspect, as polyphonic structure, is closely connected with spatial sensations. We will analyze two examples of the piece “Elementa” of J. C. Risset[10], in which the composer uses instrumental and natural sounds to create a material that mutates permanently. In this way, he creates a floating atmosphere in which material evolves in an infinite space.

....

In the third movement of Aer, the composer develops a complex structure, using a great number of different materials: flute sounds (slaps, aeolian sounds, melodies etc), wind sound, insects and little animal voices. The whole movement is a long series of metamorphosis.

This movement can be compared to Mauritius Escher [11] “Metamorphosis” (Figure 10). This Dutch graphic artist was also preoccupied by notions of space and time emerging from bi-dimensional figures; in this case, engravings. We can observe how bees go out of hives towards the birds that become fishes and then horses, and so on. In ìAerî we find similar transformations, when wind sound changes the register and then becomes the melody of a flute which gradually disintegrates to become crickets songs... then, all is again transformed to become a complex mass.

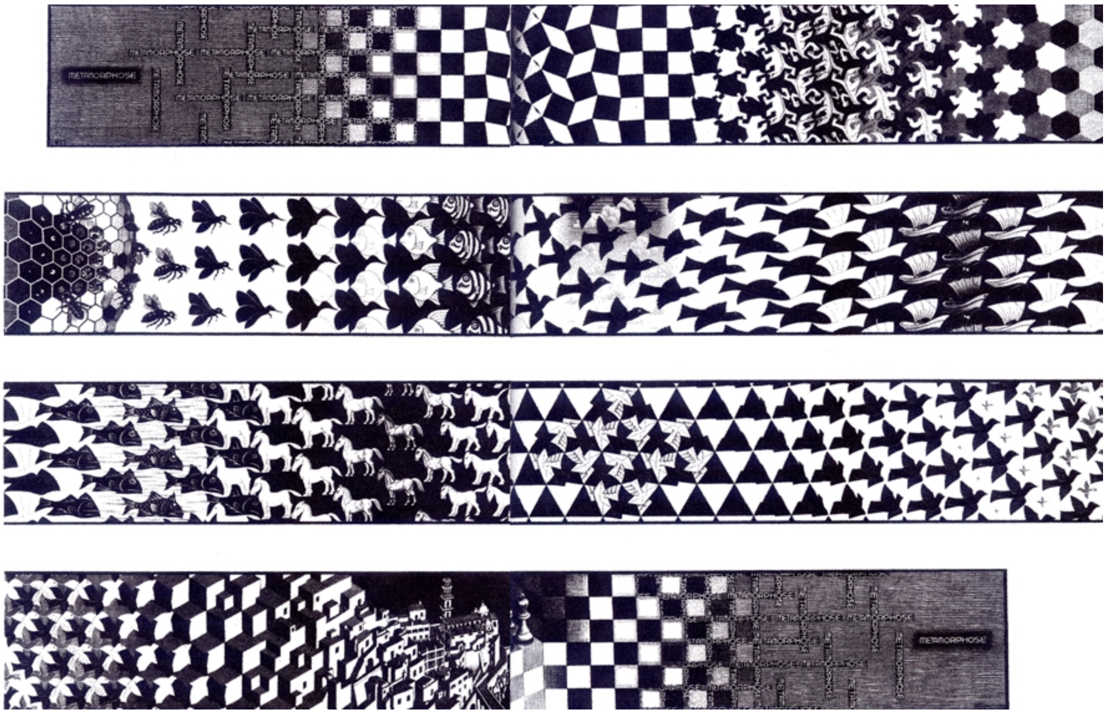


Figure 10

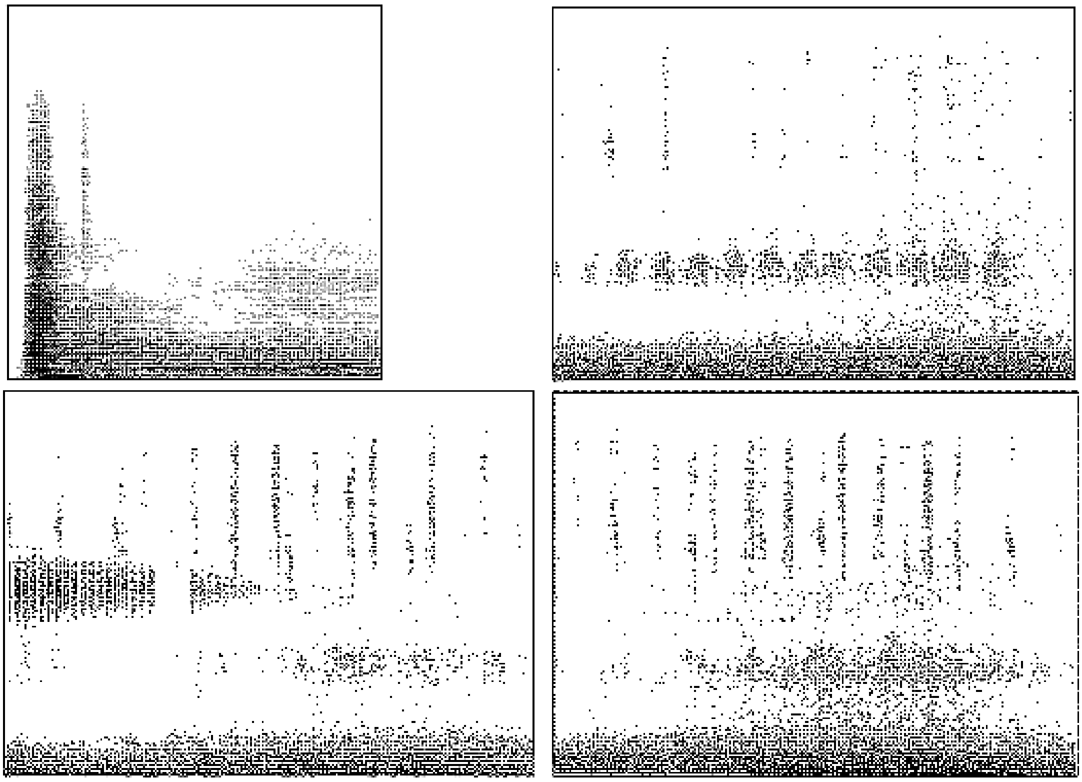


Figure 11

In Figure 11 we can observe: in the top left square, the characteristic spectra of a flute slap.

The figure on the right shows us two layers that represent the sounds of wind and flute respectively. In the lower squares we have: on the left: the same elements superposed to a new object that represents the cricket sound. In the final square on the right we can observe an heterogeneous group of elements of different classes.

The most important characteristic of this movement is the fusion of surfaces, particularly those of wind and flute. This complex surface is present during the entire fragment describing a continuum of streamer figures. This is, in fact, another good example of the phenomenon linking material and space.

As in Escher's "Moebius band" (Figure 12a) and "Magic cube with ribbons" (Fig.12b) the external and internal surfaces of the band, are blurred. So our perception catches alternatively the harmonic sound and the white noise, without realizing when they are muted.

Figure 12a

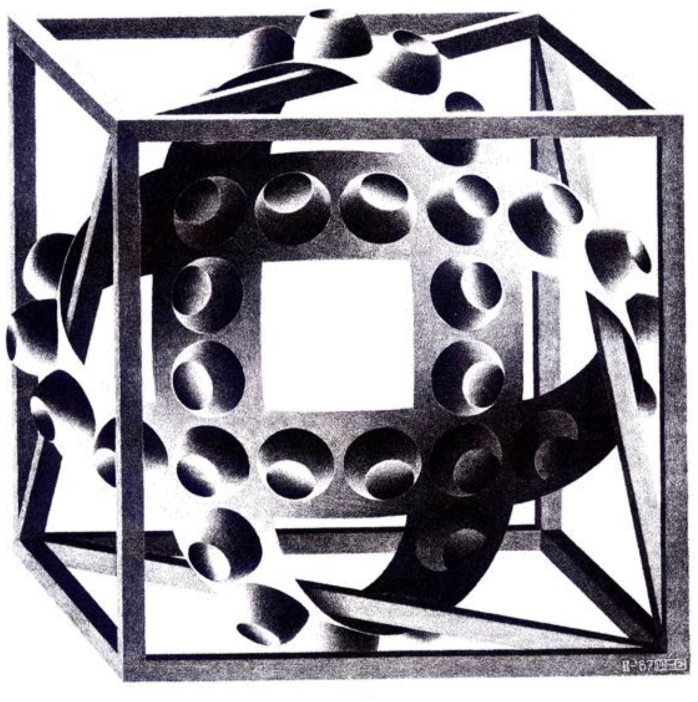


Figure 12b

In these visual labyrinths the observer cannot establish whether he/she is seeing the internal or external face of the figure. In Rissetís piece this notion of virtual space also leads to a reversing of perceptive phenomenon and at the same time it suggests a reflection on the symbolic sense of music.

This notion of spatial architecture is present all throughout the piece Elementa. The whole piece is recorded in four tracks. This aspect, evidently contributes to enhance polyphonic structures and consequently also spatial aspect.

We will analyze two fragments of the first movement: Aqua. In the following scheme we can observe the first segment of 1'30, in which there are four levels of spatial depth, that we numbered from forward to backwards from 1 to 4. Events of different density are superposed and interwoven. Some of these events are in fix plans of space, others move into right/left and forward to backward trajectories, simultaneously.

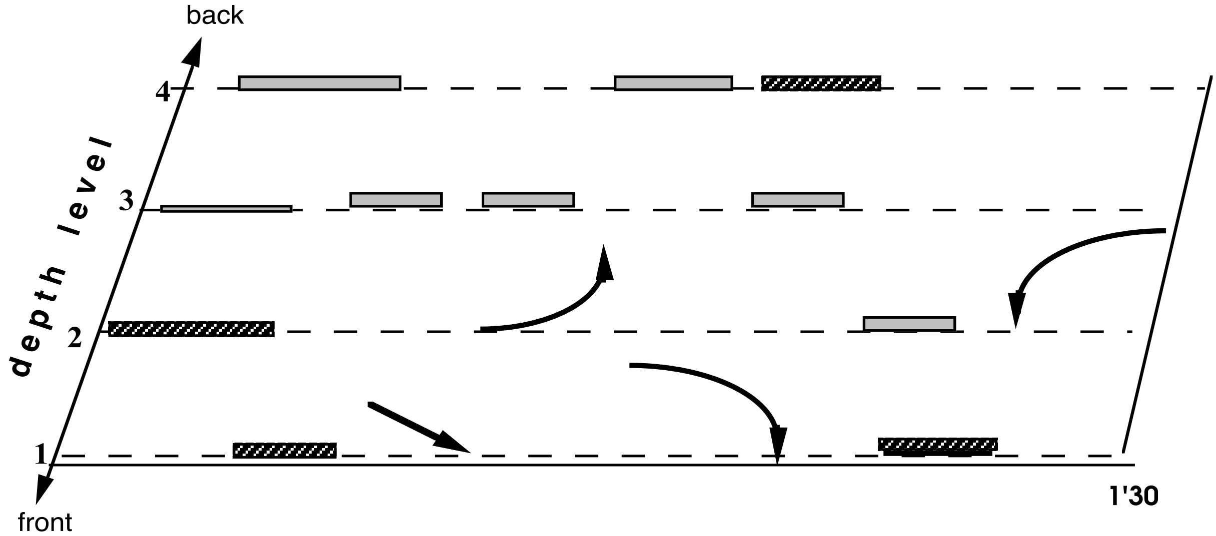


Figure 13

After one region of great activity there is a release and then a new moment of shaking (between 2' and 3'30) - Figure 14

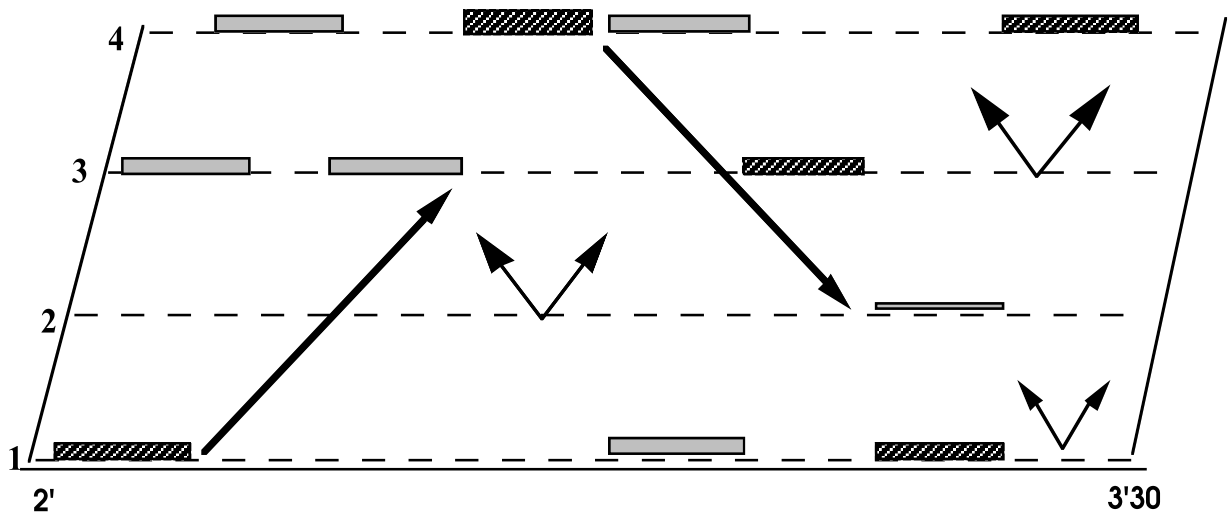
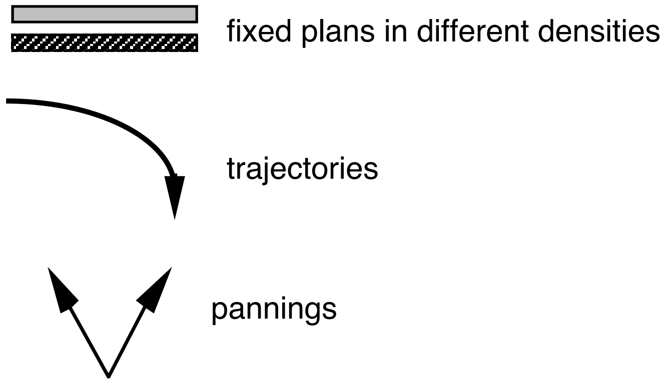


Figure 14

Graphic symbols:



The diagram shows the complexity of the fragment where long trajectories cross events at different levels. At the same time, events on fixed plans placed in the center of the scene, are juxtaposed to textures embracing a larger panoramic field. All these elements added to material diversity of the fragment, let us imagine the richness of texture and movement of the piece. The moving character of both space and material, interwoven, can be compared again with functions of figures and backgrounds in Escher's engravings. The ear, as the eye, has a tendency to fix on a precise object, leaving in the background all of the surrounding.

Figure 15 can be an example of graphic representation for the movement Aqua.

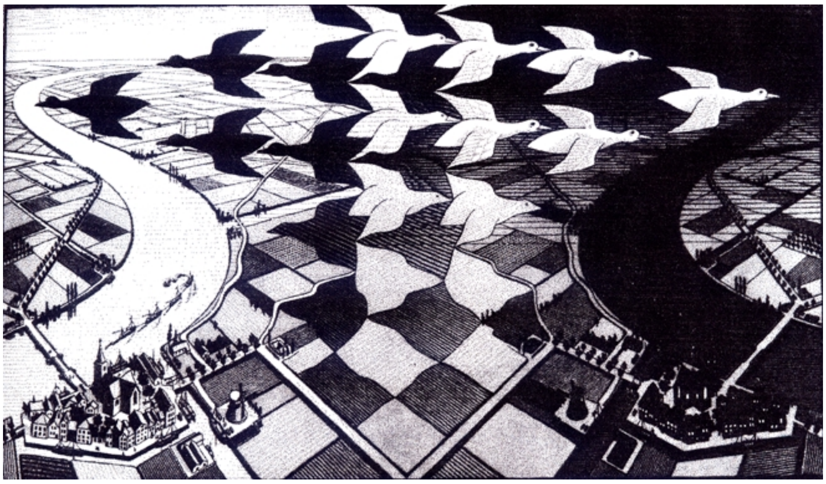


Figure 15

Day and Night: (1938) "At the top of the image, the gray rectangular fields become figures of black and white birds. Black birds fly to the left and white ones to the right in two opposed rows. On the left of the image, white birds fusion to become the light of the day and at the same time a landscape. On the right, the black birds become the night. The clear and dark landscapes constitutes the inverse of the other and they are linked by the gray field that become again birds."

Space of performance

The French school of acousmatic music has given one of the first models in real time spatialization, with the systems Cybernophone and Acousmonium. This type of system, composed of multiple loudspeakers, allows the composer/performer to "color" and spatialize the diffusion of electroacoustic music. In this way, the disposition contributes not only to ornament to the music but also to go deeply in the sense of the work. Nevertheless, this type of system is not convenient for all kinds of music. For example, in the piece "Turenas" de J. Chowning [12] needs four identical sound sources, with flat phase response and a minimum of intermodulation distortion.

In any case it is necessary to consider several conditions, such as; the quality of loud speakers, the form and size of the hall, the absorptive surfaces and the regulation of diffusion mechanisms. If all these conditions are adequate, the diffusion can enable the realization of spatial figures, and produce important changes into the musical texture.

It is important to note that the reverberation is a very important element in the structure of electroacoustic music. So, the composer must take care of the influence of hall reverberation at the moment of using it. The combination of both reverberations can be a disconcerting element for perception. It would be interesting if the composer could study the hall characteristics before beginning the composition.

Readingness of space

All these reflections give us an idea of the interrelation between internal and external spaces. If we can say that the projection contributes to a better "reading" of musical structure, it is important to know that this "reading" is related to the perception phenomenon and to the acoustic conditions.

As an example of these interrelations we can mention the studies on timbre by Risset and Wessel. The phase difference between two spectrally identical sounds, for example, is one of the problems for the "reading" of projected sound. It is very well known that two sounds of the same timbre and harmonic amplitude, may have a phase difference from the acoustic point of view, that is not perceivable by the ear.

The research by Mathews, MacAdams and Deutsch [13-14], among others, on separation of auditive flux, had allowed us to establish models of perception behavior, which enables us to clarify some aspects of musical prosody.

The composer should be able to govern this type of phenomenon such that a discourse is perceptible. The tools of analysis and synthesis can help, even if they are not absolutely perfect.

We believe that a collaborative and interdisciplinary work between scientists and artists will be important to resolve these conflicts.

Aesthetics of space

The last point, but not the less important, is the esthetic of space. All throughout our paper we have mentioned the word structure. It means that we have penetrated the esthetic domain. In fact, in the spatial conception there are two dimensions: the acoustical and the musical. Between both of them it must exist a transmutation able to create a consolidation of musical discourse. It means that we must consider two correlation factors:

a) the interaction between the components of emission, that is source parameters (such as timbre, duration, amplitude), and the propagation conditions (the hall characteristics such as reflection, distance, reverberation etc).

b) the implication of these factors for the structural organization of music. That is; the awareness of these external interactions in musical project.

On the other hand, spatialization can be used as:

1) An esthetical ornament; we can mention the examples of circular trajectories and revolving effects in Little Boy of Risset or in Gesang der Jungling of Stockhausen.

2) Or as a construction element; in this sense the knowledge of physical reality of sound material is important to obtain the most interesting spatial effects.

3) Or still as metaphor in order to create sound images, as in anecdotal music tendency.

We can add to those conceptions the criteria of those who attribute to diffusion an aesthetic function. They consider that it is during the performance that the work takes its real sense because it is re-composed and loaded of sense by means of interpretation.

Conclusion

We have shown, by all these assertions, that composers, in spite of their conceptual divergences, stay loyal to the final objective: the musical work. The different points of view in relation to space or to its operational form or even to its aesthetic function, are only theoretical deviations over one single plan.

Acknowledgments

Thanks to all composers mentioned in this paper who authorised us to publish the analysis of their pieces.

References

[1] Savouret, Alain: "*Sonate Baroque*" - Chrysopée électronique-Bourges, Harmonia mundi , LDC 278 1101 - 1993

[2] Ferreyra, Beatriz : "*Petit Poucet Magazine*" - Chrysopée électronique - Bourges, LDC 278 1109-1998

[3] We have used mostly FFT analysis by sliding windows, in 2048 window size and 256.000 steps.

[4] Risset, Jean-Claude : "*Computer suite from Little Boy*" - "Music with computers" Wergo 2013-50 - 1988

[5] Risset, Jean-Claude : "*Extraits de Little Boy et Mutations*" - CD in Actes de l'Académie de Bourges,Editions Mnemosyne - 1996

[6] Shepard, Roger : *"Le son musical"* Ed. Pour la Science, Paris, 1984, et in Richard Moore " Elements of computer music " Prentice Hall,New Jersey, pp.221-226, 1990.

[7] Vaggione, Horacio : *"Transformations morphologiques para analyse/resynthèse"* in *Actes II de l'Académie de Bourges,* 1996.

[8] Truax, Barry *: "Le symbolisme électroacoustique dans Powers of Two :The artist*", in Actes de l'Académie de Bourges, 1996

[9] Zanési, Christian : "*Arkheion,* INA-GRM (INA e 5001), 1996

[10] Risset, Jean-Claude : *"Elementa*", recording given by the composer himself, 1998

[11] Escher, Mauritius : "*L'œuvre graphique*",Benedikt Taschen Verlag Berlin GmbH, 1990

[12] Chowning, John (turenas)

[13] MacAdams et ..: "*La musique et les sciences cognitives*" Pierre Mardaga éditeur-Bruxelles, 1989

[14] Deutsch, Diana: "*Grouping Mechanisms in Music*" , The Psychology of music, Academy Press Inc.,1980

Elastic Horizon: Mapping Collaboration

Lissa Meridan

School of Music, Victoria University of Wellington,
PO Box 600, Wellington, NZ
*email:*lissa.meridan@vuw.ac.nz

Keywords: collaboration, mapping, installation, interactive, gesture.

Abstract

The artist talks about aims and outcomes of her recent research trip to CNMAT, UC Berkeley, with specific reference to the experience of collaborative arts, and the contrasts in research methods and objectives between the university environment and the underground studios she was working in.

Artist Statement

Shifting between the familiar and the abstract, *Elastic Horizon* extends the boundaries of the known world into a fantastic 'other' – a labyrinth of visual and aural experience.

Elastic Horizon is an interactive audiovisual installation, which invites the visitor to manipulate and transform images and sounds from the natural environment. Participant's actions within the space are interpreted, in real time, into a sequence of visual and aural effects. The number of people, their rate of movement and their location in the space all contribute to changes within the installation.

This project is the result of the collaborative work of Lissa Meridan (New Zealand) and Antonio Funiciello (Venezuela) and was supported by the Adam Art Gallery, Victoria University of Wellington, MagnumMac and B&H Visual Communication & Presentation Solutions.

Lissa Meridan (b. 1972) is currently Director of the Electroacoustic Music Studios at Victoria University of Wellington, where she also teaches instrumental composition, orchestration, counterpoint and acoustics. She is a committee member and webmaster for the Composers Assoc. of NZ and Vice President of the Australasian Computer Music Association.

Her primary artistic concern is with the recontextualisation of new music outside of the concert hall performance model, and this has manifested itself in multi-media collaborations with video and dance, gallery installations and interactive performances involving acoustic instruments and electronic media. Her research into performance practice in electroacoustic music has also led her into experimental turntablism and live video art – she has perfomed as *fierce angel* at events both in NZ, Venezuela and the USA. Recent recordings are available on the Atoll release Fanfares for the New Millienium, Sonic 2000: New Zealand Electronic Music and NZ Sonic Art Vol 2.

Metamorphosis as a Musical Algorithm

Robert Griffin Morgan

Music Department, Edith Cowan University
email: r.morgan@ecu.edu.au

Abstract

This paper presents a personal inquiry into processes of musical transformation open to the electro-acoustic composer of today. The case study is the writer's own compositional formulae drawing from acoustic and electro-acoustic examples. A transformative process of metamorphosis is raised as the thesis and it's various subdivisions described. The elements of electronic composition are discussed in relation to the concept of metamorphosis – by filtering techniques such as wave synthesis, additive synthesis, granular synthesis, as well as frequency modulation, in the attempt of 'differentiating' and 'integrating' sound. The ratio of the Golden Section used in formal musical structure as well as in sound formation is also discussed. The genres of Musique Concrete and SoundScape provide preliminary topics for electronic formal structure before the main antithesis of this paper is raised; the electronic interactive genre that is available to the post-modern composer. Examples from the composer's Second Symphony are used as a summary for both thesis and antithesis.

Key Words

Metamorphosis, Golden Mean, Interactive Electronic Music, MAX, Electronic Sound

Introduction

By inquiry into the compositional processes open to the electro-acoustic composer of today, I will attempt to provide a critique of my own adopted methods of metamorphosis and display its relevance to musical investigation. To elucidate the meaning of this word, I will propose an aesthetic that preoccupies my mind; As all life is concerned with character growth and development, I like to represent this experience as musical material distorts and reforms through metamorphosis. As with life, a musical organism must, to my mind, be in constant evolution of being.

Jointly, I will discuss the marriage of acoustic and electro-acoustic fields, which has reached a new significance for today's tone poet.

1. A Preliminary Electronic Technique

In order to facilitate discussion into the essence of an electronic sound technique, a preliminary inquiry into the nature of its constituent ingredients needs to be conducted. The fundamental building block for sound is waveform vibration. Analysis of this basic ingredient will pave the way to electro-acoustic techniques.

Via processes of 'differentiating' and 'integrating' sound by filtering techniques, a capacity for producing a wide range of timbral sound evolution can take place. These two processes become the new tools for compositional structure. Wave synthesis, additive synthesis, granular synthesis as well as frequency modulation are all techniques that manipulate the frequency, amplitude and/or velocity of a soundwave. The process of reducing or adding harmonics to a tone is the preliminary idea behind my attempts to create a timbral metamorphosis. The range of wave manipulation from sinusoidal sound (sine wave), square waves, sawtooth waves through to white noise provides the building blocks for the

electro-acoustic musician. It seems in this sphere of electronic music that there is a return to nature and to sound in its unmediated or primordial state. The essence of sound becomes the thesis. Rather than predetermined hierarchies of scales and harmonies, the composer discovered the nature of sound- in its pure state.

I wish to include comment on the ratio of the Golden Mean, since it is not only widely used in electronic and acoustic formats but is also a key component in my methods of composing music. Besides its use in areas of macrostructure, microstructure and cellular/motific structure, the Golden Mean ratio can also control the harmonic content of tones. John Chowning in his piece Stria written in 1977, used the Golden Section as a means of relating partials within sounds and from this time, others such as Debussy, Bartok, Stockhousen to name a few have also applied the device. Certainly the proportion can be used other than in terms of formal structure- in the manufacture of sound itself, thus potentially creating a virtual space whereby all aspects of a musical work are dictated by this proportion.

Golden Section as MAX patch.

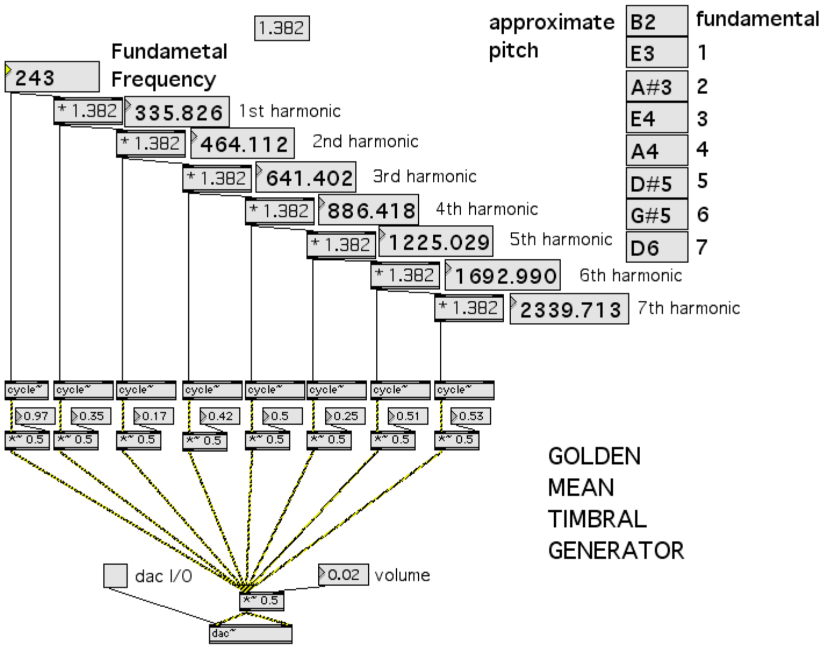


Figure 1. Max patch showing formation of tone based upon the Golden Mean.

2. Musique Concrete and the Post-modern Soundscape

A parent form of my metamorphic process is the sound art developments of the 1940's, that of Musique Concrete. In an attempt to contextualise my ideas of Metamorphosis, an account of these early methods as well as modern applications of Musique Concrete should be addressed.

Following a process of collecting 'sound objects' (as in the visual art image), the early attempts at Musique Concrete combined sampled resources to produce a collage-like structure with the aim of creating a pastiche of everyday life. After manipulating these samples through processes of filtering, distortion and modification, the result would often be unrecognisable. By the very nature of these samples, there is often evident an adopted symbolic representation of the original source, even when distorted from the original context. Certainly if common sound samples or other musical structures have been commonly ascertained to a culture, style or a work, the use of these components draws the ear to the cultural framework that is implied. Naturally the link between affective stimuli and cultural connotations needs to be taken into account if an intended meaning or affect is to be truly conveyed.

Of course hybrid forms employing electronically produced sounds and natural sounds are explored today. The interaction of the Musique Concrete principal and pure electronic sound sources is a valid form of compositional hybridity open to the post-modern composer. The term for this hybrid form is the SoundScape, and as the name implies, a virtual space is constructed by a localisation of sound.

3. Metamorphic Processes

Before describing my metamorphic processes and their various applications to electronic and/or acoustic genres, it is pertinent to provide an introduction exploring the nature of transformation. An elementary definition of the word metamorphosis is a logical beginning. Hence the following two descriptions; firstly, a complete change of form or substance occurs and, secondly, a structural principle which allows a small cell to expand into an organic, self-perpetuating whole. Two approaches are immediately apparent through these two definitions. Firstly, something already formed gradually changes into another form. Secondly, a sense of growth is witnessed by the duplication of cells (cell division) from a single cellular source.

Two paradigms of contrasting origins have influenced my concept. The first belongs to the natural organic growth and mutation of a cell from the natural sciences, and the second derives from the techniques of development within the Sonata Form of western music. In order to gain full understanding of my musical research, preliminary discussions concerning this second paradigm need to be undertaken before examples of my processes exploring both can be given.

To validate the inclusion of such a topic as Sonata Form in this paper, I will provide a brief account of the validity of historical inquiry, regardless of the discipline. It is of paramount importance to criticise any new philosophical law or artistic endeavour in light of its historical antecedents, since all generations of subsequent achievement are direct or indirect results of past developments. Of course the genesis of Sonata Form really grew from the 18th century aesthetic of rational thought and the logical exploration of an idea. The proposition of hypothesis and the subsequent examination towards supporting this hypothesis was central to the idea of debate so influential to the 18th century psyche. Naturally there were musical works bearing the title of the Sonata before this time and these earlier formal developments naturally influenced subsequent formats, but the evolution of the sonata form technique can be seen as owing its genesis to 18th century science and philosophy.

The central idea influencing this philosophy was the thesis and antithesis- the proposition of an idea that is compared and contrasted to another idea whereby an ordered argument is set up before reaching a conclusion. The philosophy can relate directly to the structure of sonata form in the following way; a statement of the thesis in the tonic key is followed by the contrasting antithesis (second subject) in the dominant key. The process of debate though justification and comparison can be seen in the subsequent development section whereby fragments of the thesis and antithesis are extended through a variety of interpretations. The recapitulation, thus, can be seen as a restatement of the original conceit and the argument explained and/or opinions validated- in musical terms, the subject and second subject are both stated in the tonic key. In all, the third section can be construed as a summary of the progression. Though perhaps simplistic to the modern psyche, one can witness the sonata form’s relevance to the development or growth of a musical idea.

I return to my metamorphic principle. Preliminary musical developments in metamorphosis first began for me in an acoustical context. Starting with a musical cell, motif, or texture, I formulated procedures through which the musical structure witnesses growth resulting in a complete change in musical information and thus giving rise to a completely new theme, motif or texture. The process is gradual and occurs in a step by step fashion whereby each stage of development moves logically on to the next. Below is a simplistic numeral representation showing five stages of this process;

A ABA AB BAB B.

With B representing completely new material, the whole progression reflects the gradual incorporation of B, beginning in the second stage, the equal interaction of the two forms in the third (AB), the prominence of B as A loses its character in the fourth (BAB) and the final establishment of the new material in the fifth (B).

Taking this concept as a departure point for musical structure, I have formulated four methods of transforming musical material. A further three methods specifically applicable to electronic metamorphosis will be discussed later. The first is concerned with the simple metamorphosis of a theme/motif or acoustic texture. Utilising the intervallic and rhythmic characteristics of musical stimuli, the material gradually adopts new characteristics, almost imperceptibly, until the new foundations rise in prominence and dictate the new musical theme/motif or texture. Below this is exemplified.

Musical Example 1

Extract from Morgan s Symphony No 1 — Genesis

Figure 2. Metamorphosis of a Theme.

Stages of Metamorphosis for Figure 2;

1. Theme
2. Theme is simplified.
3. Characteristic cell is isolated with a cell from the new theme.
4. The theme starts to be modified.
5. Material is reduced with the inclusion of the new cell.
6. The new cell dominates the musical line.
7. 2nd cell of new theme is introduced. Traces of original material are few.
8. First section of new theme is linked to second via a bridge from the old material.
9. Presentation of new theme.

The second method of transformation involves corresponding two thematic germs to a musical evolution, whereby the two gradually adopt similar characteristics and finally merge to create one final whole containing influences from both the originals. Furthermore, these two musical thumbprints can be stated separately and then developed side by side as they influence each other. The resolution of this method will logically be the fusion of both thematic germs to create a single new theme containing related parts from both originals. The process can be seen as one of unification- the interaction of the two, how they influence one another and finally end in a union. Below is an example from my own work whereby this system is in place.

Musical Example 2

Metamorphic union of two musical lines.

Figure 3. Excerpt from Morgan s first ballet Orlando .

The third method could be labelled Micro-metamorphosis, whereby the actual process of musical growth begins from pointillistic random stimuli. Key cellular components are established out of this milieu that gradually rises in prominence. The music then leads to the formation of a musical cell, motif or a much longer musical line. The following musical analysis is derived from my second Symphony subtitled '<u>Dialogues for Chamber Orchestra and MAX'</u>, whereby this technique is utilised using a live orchestra and a computer using the MAX environment. The start of the work is unformed and thus the first section is concerned with the growth of a musical cell or group of cells arising from out of a random musical chaos. MAX's random capabilities are utilised here until an organic conception is established. As a work in progress, the symphony aims to explore the culmination of individual voices by the interaction of acoustic orchestra and computer manipulated sound samples derivative of the instruments themselves.

Musical Example 3

Stage 2 of Micro-metamorphosis

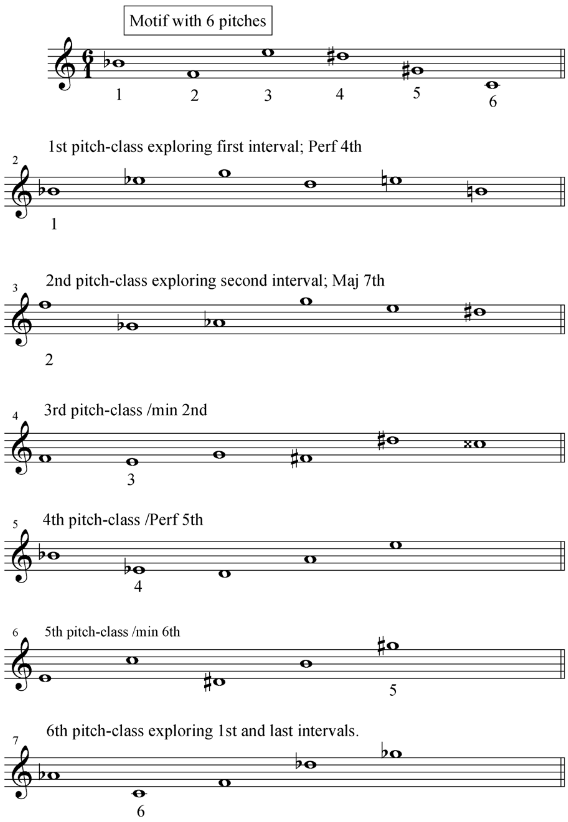


Figure 4. Excerpt from Morgan's Second Symphony- stage 2 of micro-metamorphosis.

Stage 1 of the process is concerned with the random incorporation of all 12 pitch-classes and all intervallic permutations. Stage 2 starts the filtering process whereby the six pitches and five intervals of the theme dominate. Stage 2 witnesses the reduction of pitch-classes to the six of the theme. As these occur more frequently, other pitch-classes become infrequent. Stage 3 is the presentation of the theme's intervallic ordering. Stage 4 applies rhythmic details to the pitches, leading to the final product- before this stage, no rhythmic characteristics are established.

The fourth method is concerned with the transformation of a texture. The permutations of this method are obvious. Further inquiry into textural metamorphosis will take place when electronic treatments are discussed.
To summarise, it is important to contextualise these procedures in light of musical form. Naturally these methods or techniques can be applied to a moment in the musical work, a section or they can be used as a means to construct a complete macrostructure.

Musical language and gesture become of paramount concern in this unstable and evolutionary progression. The level of harmonic dissonance can be used as a tool for creating tension as the thematic development progresses. The result, I hope, gives one the sensation of witnessing a growing, writing, thirsting musical organism constantly flowing through time and its enveloping energies.

4. Electronic Metamorphosis

So far, issues of musical transformation raised have been confined to the manipulation of pitch sets and the tonalities they imply. The whole realm concerning the manipulation of soundwaves lends itself to further discussions in relation to the application of the above principles.

The parameters in which the electro-acoustic musician can manipulate sound stimuli are summarised in the ideas of Stan Tempelaars. Tempelaars differentiates between ‘global’ modulatory sound (whereby the entire sound is affected) and micro-modulation (which causes change instant to instant). Furthermore, the developments of the 1990’s gave rise to the term ‘frequency modulation’. The following quotation from Electric Sound by Chadabe is a valuable concept for the manipulation of a waveform frequency via ‘peak deviation’;

In *frequency modulation*, the instantaneous frequency of one waveform, called the *carrier*, is varied by another waveform, called the *modulator*. The extent to which the carrier frequency varies, called *peak deviation*, is determined by the amplitude of the modulator. The rate at which the carrier frequency varies is determined by the frequency of the modulator.

When carrier and modulator are at audio frequencies, audible extra partials, called *sidebands*, appear in the carrier spectrum. These sidebands are located symmetrically above or below the carrier at intervals equal to the modulator frequency….The number of sidebands with significant amplitude is determined by the *modulation index*, which is the ratio between the peak deviation and the modulator frequency (index = peak deviation/modulatory frequency) (Chadabe 1997 p251).

I return to my ideas of metamorphosis. Trevor Wisthart used processes of sampling birdsong, machine sounds and other various sounds taken from the real world. Symbolic representation was then associated to the sounds just as in any other Concrete work. More interesting in this example is the juxtaposition of samples and gradual transformations between them; “A flock of birds, for example, transforms into a rustling of a book’s pages, then into the sound of closing a book, then into a slammed door, into a creaking door…”(Chadabe 1997 p137). Here is the epitome of the electronic transformative process.

The use of vibrato to imply timbral change is another method. By applying a vibrato of increasing range and depth to a tone, the illusion of timbral fluctuation is evident, giving the implication of an evolving spectrum of timbre. The application of vibrato is a unique way of creating a depth of sound and by controlling the speed of the oscillation and amplitude manipulation, the technique can be utilised as another approach to metamorphosis.

That master of sound-art, Jonathan Harvey summarises these ideas of timbral transformation used as a structural device; “The timbral combinations were very interesting because in each melody the timbre changed- the shakuhachi turned into an oboe, for instance- so mixing melodies was also mixing timbres”(Chadabe 1997 p130). The implications of this statement are obvious – the range of timbral experimentation is infinite, especially considering the developments in hardware and software since 1990, the time of Harvey’s observation.

To elucidate the effect of the above upon my own creative work, I again return to the word metamorphosis. I would like to offer further definitions to the word, summarising the individual processes I utilise in the act of musical composition.

.

Metamorphosis of Tone. Manipulating a clearly defined pitch. The transformation to a more complex array of partials can finally result in noise. Or another example, a whole combination of tones create a texture that is slowly raising (or lowering) in intonation

Metamorphosis of Timbre.
Transformation of a specific timbral pitch to another timbre occurs by manipulating the harmonic content of the pitch.

Textural Metamorphosis.
Single tones grow into chords. This process was used in much of my earlier acoustic works. The expansion from a unison pitch to a thick instrumental texture explores various shades of density that shift and fluctuate. A further density transformation can be found in the shifting of a soundscape texture where several timbral voices all shift to one timbre. Or a polyphonic texture utilising only one interval, say a major second, widens until all voices reach another interval, say an augmented 4th. Certainly Per Norgard's Infinity Hierarchies relate to this approach, or the textural growth

that is evident in Ligeti's works. This process also lends itself to a wide range of variation – one being the reduction of a texture incorporating all the 12 semitones to five defined pitch classes. This particular example also shows the golden ratio in operation- 12/5 = 0.618/0.382.

Musical Example No 4
Textural Metamorphosis

Figure 5. The expansion of a texture over four bars from a minor 2rd to an Augmented 4th.

Rhythmic Metamorphosis.This last form leaves the treatment of pitch, timbre and harmony. All permutations of rhythmic transformation are available. Techniques I have adopted are various rhythmic accelerations and retardations and layered rhythmic metamorphosis, whereby one layer of musical fabric starts to travel at a different speed and then all other layers speed up from that departure point. Further more, this process can be used to set up a new tempo with a new rhythmic pulse and beat subdivision. The new tempo in relation to the old can be a predetermined ratio; for example a two to three ratio, or the golden mean ratio. In musical terms, an example would be the shift of pulse from a quaver to a quaver triplet.

All of the above processes exploring the theory of metamorphosis can be utilised as passing compositional gestures and are seen in many existing works. My approach is to use them as a means of travelling through the musical structure - from one section to the next. I am also preoccupied with the rate of change; emphasising the journey before arriving at a new destination gives the new section more meaning and definition. In an attempt to summarise all my categories of the word metamorphosis, I propose the following hypothesis;

> While a musical section is established, the introduction of a new element takes root (causing a sense of conflict) and is perpetuated as it grows in importance before dominating and finally causing all material to follow and adapt until the new section based on the new element is secured.

5. A last note on Interactive Music- MAX and the Acoustic Orchestra

A fashionable artistic association of today is the performance-orientated collaboration of the musician with a computer- one that certainly lends itself successfully to my adopted techniques. The combination of sampled art, electronic synthesis, acoustic performance and live interaction produces a completely new aesthetic. Certainly with the rise of the software application MAX and its MSP enhancement, infinite compositional and multi-media possibilities now exist in an interactive context. It is also important to acknowledge John Chowning and Todd Winkler and their contribution to computer interactive music. To conclude, I will contextualise the issues raised in this paper by relating them to this interactive environment.

My proposed composition Symphony No 2 'Dialogues for Chamber Orchestra and MAX' will explore the duality of highly detailed construction (various processes of metamorphosis permeated by the Golden ratio) and sections of *Maxian* Improvisation. As the ensemble is pitted against MAX, samples of the section will be utilised in the form of a computer generated improvisation using MAX's random related objects (guided by the operator who is following certain text instructions written into the patch).

Thus, as the thematic material progresses, the computer manipulates this data to produce an electronically produced synthetic response, until the compositional algorithms take over once more. As the computer cognises the order, timing and timbre of the pitches played by the instrumentalists, the principle of metamorphosis follows two independent streams of musical activity guided by the one set of musical material. As stated earlier, the cognitive modelling of the musical stimuli will be confined to certain parameters dictated by the operator and the messages incorporated into the patch -the result is somewhat like the cadenza within a concerto. Episodic in form, the work will move from sections of organised developments to those of guided improvisatory exploration.

One can see the relevance of interactive music to my ideas of Sonata Form and the 18th century aesthetic. The interaction of two sentient beings that share and develop thoughts before leading to conclusive hypotheses is the essence behind both paradigms. In the context of my second symphony, the aesthetic produced should be one whereby an isolated individual (MAX) explores a relationship with a group of individuals (the instrumental ensemble) that are all connected in some sense of censorship. How they influence one another as they converge and diverge is the thesis of this compositional style.

References

Chadabe, J. (1997). Electronic Sound.
New Jersey, Prentice Hall, Upper Saddle River.

Winkler, T. (1999). Composing Interactive Music- Techniques and Ideas Using Max.
Massachusetts, The MIT Press.

Towards Audio-Based Music Information Retrieval in the RMIT MIRT Project

Andrei Bajurnow and Alexandra L. Uitdenbogerd*
School of Computer Science and IT, RMIT University
email:{abajurno,alu}@cs.rmit.edu.au

Abstract

Musical Information Retrieval is the field of research that explores ways of retrieving music, mainly using non-text techniques, such as through melody queries or rhythm. At RMIT Computer Science we have worked on melody-based information retrieval of MIDI files and are now exploring the use of audio features, such as rhythm, tempo, tonality, noisiness, and timbre to answer a range of music queries. In this paper we summarise the achievements of the MIRT project, and describe our current projects. In addition we show some preliminary work that explores the application of thresholding on standard deviation of frequency spectrum values to the analysis of cymbal hits.

1 Introduction

Traditional textual databases of musical information don't allow certain types of music query that are commonly asked, such as, "where does this fragment of melody come from?", "who wrote the Alleluia that goes like this?", "how similar is my composition to existing works?", "I like the song 'I've seen that face before'; what else is there that is similar?", "I want a piece of music for my film, that expresses defiance". The field of Music Information Retrieval aims to address these types of content-based queries.

RMIT's MIRT project commenced in January 1997 as work for a PhD thesis (Uitdenbogerd 2002). The initial aim was to develop techniques that allow a melody query to be presented to a system, which would then provide a ranked list of matches to the query based on melodic similarity. This work led to the development of a three stage approach to music matching, consisting of melody extraction (TRIMming), melody standardisation (TIRMing), and similarity measurement (MITRing). These three stages work together to ensure that meaningful matches can be found to melody queries, given the processing constraints and varied musical content associated with matching against a large collection. The work focused on musical data in the MIDI file format and searching polyphonic music for matches to melody queries. Techniques that worked well for this problem were the extraction of melodies based on the highest pitch note commencing at each instant in each MIDI track, representing these melodies as strings of pitch differences, and using a matching method that counts all common substrings of length 5 (n-grams), ignoring duplicates (For more details see Uitdenbogerd and Zobel (2002)).

While we are continuing to answer the research questions associated with the melody matching problem, we have expanded our project to include other music retrieval applications. Methods of retrieving music that are likely to be useful include those based on style, mood, and musical taste. In order to address these types of queries, we are exploring methods of extracting the features from music in audio format that may indicate style or mood. We will be combining features that are already published, such as those used by (Welsh et al. 1999), and Tzanetakis, Essl, and Cook (2001), with others that try to capture aspects that have been shown to be important through studies of human perception of music. Early work in music psychology suggests that for determining the mood of a piece, tempo is the most important, followed by tonality and then whether the pitch is high or low (discussed in Farnsworth (1958) Chapter 5). For musical style, human-classified genres have been based on genealogical, historical, geographical, functional, instrumental, as well as other enumerated types (Pachet and Cazaly 2000).

For musical taste, we expect to answer queries based on a user's existing music preferences, that is, we are developing techniques for music recommender systems. An approach that has been successfully applied to this type of problem is known as collaborative filtering. We are building a test platform for evaluating different methods of recommending music. These will include user ratings and behaviour in addition to other forms of evidence, such as the answers provided to questionnaires presented to the user, and user-provided links between pieces.

Another problem that has yet to be solved is how the lay person can reliably find music for which they know a part of the tune. Parsons (1975) used a melody "contour" index of themes, that is, an alphabetically sorted list of themes represented by a sequence of symbols representing *up*, *down*, and *same* note transitions.

*Contact author for correspondence

This made it possible to look up a theme in the index without being able to read music. Some prototypes have been developed that allow users to sing a melody query, a melody contour is then extracted and used to look up answers in a database (Blackburn and Roure 1998) (Ghias et al. 1995; McNab et al. 1996). It has been found, however, that the use of contour information is insufficient for larger collections of musical data (Uitdenbogerd and Zobel 1999).

Beyond the question of music representation, it is still unclear how to best present the answers to a music query so that the user can determine whether they are indeed relevant. We concluded from an experiment in collecting music relevance judgements that it is important to provide match location information when the user wants to listen to the potential matches. There are further difficulties for users when listening to unfamiliar music (Uitdenbogerd, Chattaraj, and Zobel). In work that extends our earlier melody-based retrieval research, we are exploring different types of music retrieval user interface, and evaluating them from a usability perspective. Our initial experiment will compare the usability of different textual representations of melody for novice users.

In other work, we hope to examine the playing style of musicians using various audio analysis techniques. Initially we are focusing on percussion style, which should also be a useful indicator of genre. Percussion playing style can be characterised by timing, loudness, and timbral variation of hits. Using existing techniques for locating beats, we can explore timing variations in pieces of music. However, these techniques do not isolate percussion from other strong beats occurring in the music. As an indicator of percussion rhythm it may be interesting to isolate the high-hat and other cymbals that are used for fast beats. As a preliminary step we are analysing audio samples of high-hat and cymbal hits using Fourier Transforms and statistical approaches, to determine the nature of a cymbal attack. This preliminary work is reported in this paper. Other researchers have published on related topics, such as instrument detection (Fujinaga and MacMillan 2000; Martin and Kim 1998), beat detection (Cemgil et al. 2000) (Dixon 2001), transcription of music (Garcia 2001; Hainsworth and Macleod 2001; Kashino and Tanaka 1993; Martin 1996a; Martin 1996b), and the analysis of playing styles (Cambouropoulos 2001; Trilsbeek, Desain, and Honing 2001). To our knowledge, however there hasn't been much work specifically on percussion analysis.

2 Background

In the preliminary work we report here we used Fourier transforms of a stored wave-form. In this section we provide a brief review of the relevant concepts.

When audio is stored on a computer, the sound signal is converted from a continuous one to a discrete one via analogue to digital conversion. Each sample represents an amplitude at a specific time unit. While it is possible to glean some useful information from audio stored in this manner, more meaningful details are able to be extracted if the data is transformed into another representation. The Fourier transform is one such transformation.

The Fourier transform takes a signal such as one from an audio source, and converts it from a function of time to one on frequency. It is based on the Fourier series, which can be used to represent a periodic signal in terms of an infinite weighted sum of sines and cosines. A finite signal that is not periodic can be treated as periodic by assuming that the period of the signal is equal to the duration of the entire signal. In practice, when applying these techniques to discrete audio signals, the transform usually provides frequency bins of a particular range and the magnitudes associated with them. For detailed analysis of changes in the frequency spectrum over time, short sequences of audio are transformed, allowing a sequence of frequency snapshots to be compared. These can be analysed statistically.

When examining the timbre of an instrument, the frequency range, or *bandwidth*, is an important indication. Much of the frequency spectrum for a particular timbre remains the same regardless of the pitch of the base note (Dodge and Jerse 1997). Conversely, if the frequencies shift with the base note, the timbre appears to change. For example, consider the effect of changing the pitch and speed of a voice sample.

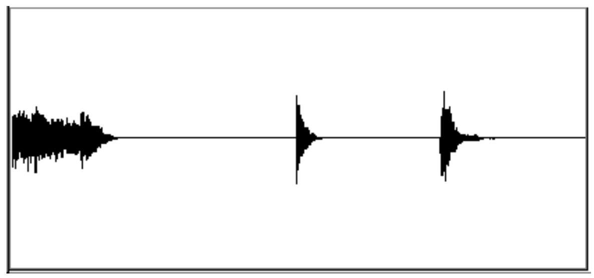


Figure 1: The wave-form used for our preliminary work. It consists of a high-hat being closed, and two crash cymbal hits.

3 Preliminary Work

The goal of this initial project was to establish whether a line of investigation into the possibilities of profiling timbres according to levels of volatility or variation was promising enough to pursue. The idea came about while viewing the three-dimensional spectrograms of many instruments and their unique timbres. It was a view of the profile of percussion instruments, especially various crash cymbals, that highlighted the magnitude of noise propagated across the frequency spectrum for the duration of the cymbal's decay. The

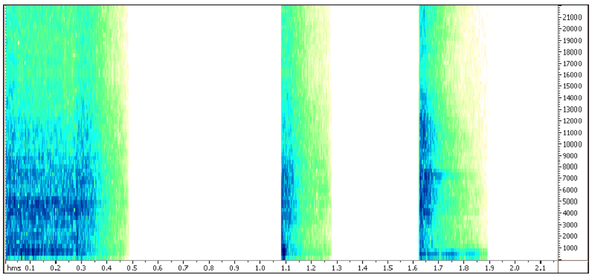


Figure 2: The power spectrum of the wave-form shown in Figure 1. Observe the spread of frequencies (bandwidth).

term noise is being used in the sense of not having the clear fundamental and regular harmonic properties produced when playing most pitched instruments. A clarinet, for example, will have a strong fundamental frequency and some overtones (harmonics) which characterise the timbre as being that of the clarinet. The degree of noise produced when sounding particular instruments could well disrupt an audio analysis model. If it were possible to do a probabilistic model, assigning a particular noise interference profile to a particular event within the sample, it may be possible to compensate for this in an analysis, such as focusing on the area in question and using statistical techniques to remove or reduce the effect of the increased variation, thereby enabling a finer analysis of what may be masked by such noise. The scope of such techniques was well beyond the limits of this initial investigation, which was limited to determining the effectiveness of such an approach to timbres in a non-polyphonic setting.

3.1 Method

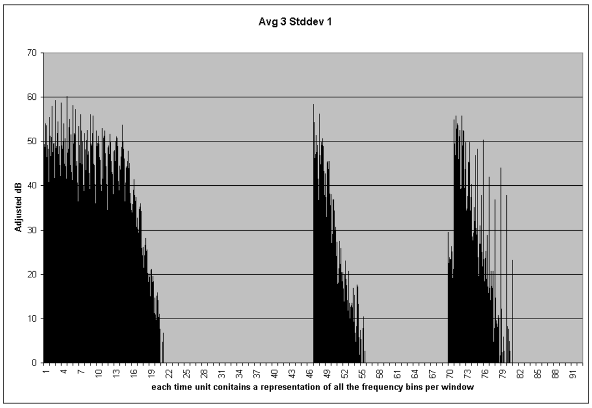
The aim was to determine what kind of timbral profiles would be returned when assessing a sound in relation to an average measure of amplitude variation, in a particular frequency spectrum. The two main variables were the length of the time average and the level of amplitude extremity (or distributions falling within certain deviations away from the average variation). Variation in the time domain was taken to mean "Is the current amplitude of this frequency bin, greater or lesser than the average of an arbitrary amount of consecutive previous frequency bins in the same spectrum?" The level of extremity variable means "Is the current amplitude of the frequency bin greater than N standard deviations away from the average of the amplitudes of the same frequency over an arbitrary distance?" Combinations of these results had some interesting effects.

Since the idea came from viewing crash cymbals, we commenced the investigation by creating a small library of different types of cymbals with differing timbral profiles and sequencing them into a wave file (see Figures 1 and 2). This enabled the extraction of the frequency profiles by using the Fast Fourier Transform (FFT). After some experimenting with various lengths, the use of a window length of 1024 and FFT length of 1024 was selected, as it seemed the most appropriate for this current analysis, in that percussion was the focus and therefore accuracy in the time domain was more important than accuracy in the pitch domain. Higher window values were tested but the resolution in the time domain was reduced such that the data returned was not accurate enough to provide the detail required for this style of analysis.

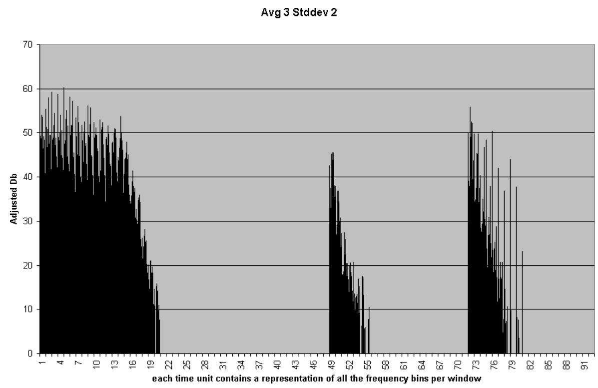
3.2 Observations

Figures 3, 4, and 5 show the values for different averaging windows and thresholds. The horizontal axis shows two dimensions: each unit represents a time unit, and within these are vertical lines representing the amplitude for each specific frequency band.

In the cases of the average of 3 time units, there is not a great deal of change between the different deviations (See Figure 3). When the average is increased to 6, the change is more immediate. The first plot (6,1) (See Figure 4) is not much different from all the previous ones but the next, (6,2) displays a severe curtailing of the decay. This is to be expected as the effects of the initial amplitude spike has "washed" through the calculations. The final plot of the average of 6 (6,3) shows the other effect of having such an amplitude spike; the deviation from the mean is so great, that when using a factor of 3 in the calculations, very little else exceeds it until there is great calm in the music, a situation not obtained in this short sample. The second and third cymbal hits do not even register and are absent from the plot entirely. As the average is increased to 9 units (not shown), the effects of altering the sound profiles are more predictable. With a steady increase of the deviation factor, the decays reduce proportionately. As the time factor here is three times greater than in the first instance, the averages will be less reactive to sudden spikes but the effects will remain in the calculations for a longer period. The final set, based around an average of 12 (For example, see Figure 5), display similar properties to 9 but have slight differences in which frequencies are more persistent. It is interesting to note that these differences are only those that could be picked up by the naked eye and it is possible that with more advanced comparison functions, the program might quantify other significant differences, especially within each window, which has been compressed by the nature of the graphical pot. Further development is needed to extract the differences between the models and distil them into meaningful patterns. It is obvious that a profile as generated by such a routine is a degree more descriptive than one that is just looking for a spike above a given level. This also provides data relating key features of the timbre of the sound. The effect is to attempt to create a data set that could be stored for correlation or comparison routines at a a)



b)



c)

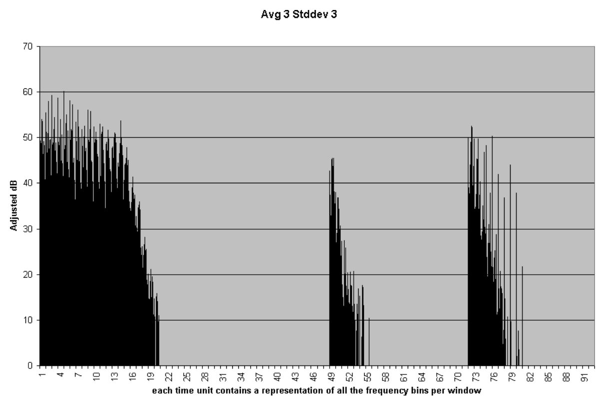
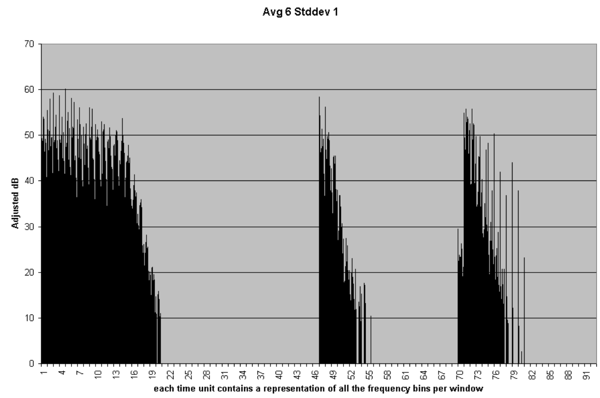


Figure 3: Graphs of the change in frequencies over time. The X axis is divided into bins containing the full frequency domain for a particular FFT window. For these graphs, the current frequency amplitude was compared to the mean of the previous 3 amplitudes. At a), amplitudes that were within one standard deviation were mapped to zero, at b) those within two standard deviations, and at c), within three standard deviations. Values that were less than an arbitrarily selected number, representing low level noise, were also removed.

a)



b)

c)

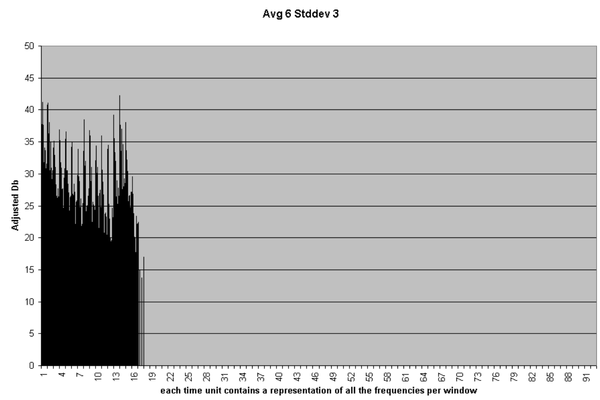


Figure 4: Graphs of the change in frequencies over time. The X axis is divided into bins containing the full frequency domain for a particular FFT window. For these graphs the current frequency amplitude was compared to the mean of the previous 6 amplitudes. At a), amplitudes that were within one standard deviation were mapped to zero, at b) those within two standard deviations, and at c), within three standard deviations.

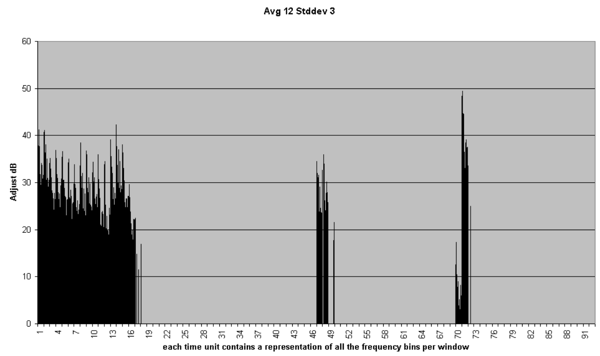


Figure 5: A graph of the change in frequencies over time. The X axis is divided into bins containing the full frequency domain for a particular FFT window. For this graph the current frequency amplitude was compared to the mean of the previous 12 amplitudes. Those that were within three standard deviations were mapped to zero.

later stage.

3.3 Discussion

There are many potential issues surrounding the use of such a technique: how is an optimum average and variation to be arrived at? What relation does this have to the tempo of the music? When would accuracy of frequency outweigh the need for accuracy of pitch within the FFT? It is possible that a number of test runs might be necessary to determine the probability of success of the employment of such an algorithm. It also may have questionable effects when the music is played by pitched instruments which display less variation within a pitch and move across many pitches within a phrase. More testing would be needed to determine a strategy to deal with this.

To conclude, we think that this investigation is worth pursuing. We suspect that any comparison routine based on variation could become quite complex due to the amount of data that would be collected and the number of comparisons to be made. There might exist a trade-off between using a less descriptive average to accommodate more comparisons, and only employing a search using this method of comparison if all others had failed.

4 Summary and Conclusion

There are many interesting research questions related to music information retrieval. Practical systems for melody-based music retrieval from MIDI or note-based data can now be built, but are limited by the data that is available. Most music is available as audio recordings, but extracting meaningful data from audio data is quite difficult. Our MIRT project hopes to address some of the problems that have not yet been solved for audio-based music retrieval. Our focus is on feature extraction of indicators of style or mood, recommendations based on a combination of evidence, and detecting instruments. We are interested in collaborating with other researchers on these ideas.

The preliminary work reported here revealed that a combination of Fourier Transform and the detection of changes in deviation from the mean are reasonably effective in locating the onset, and detecting certain characteristics of a percussion instrument, however, there are many questions that remain unanswered, such as whether the approach will work for more complex audio samples, and whether other types of transform, such as wavelets, will provide this information more readily. We intend to develop these ideas further in the near future.

Acknowledgements

We thank James Thom, Justin Zobel, Lin Padgham, Ron van Schyndel, and Santha Sumanasekara for their involvement in some stages of this project. This work has been partly supported by an RMIT School of Computer Science and Information Technology department grant.

References

Blackburn, S. and D. D. Roure (1998, September). A tool for content-based navigation of music. In *Proc. ACM International Multimedia Conference*. ACM.

Cambouropoulos, E. (2001, August). The local boundary detection model (LBDM) and its application in the study of expressive timing. See ICMA (2001).

Cemgil, A. T., B. Kappen, P. Desain, and H. Honing (2000, August). On tempo tracking: Tempogram representation and kalman filtering. See ICMA (2000).

Dixon, S. (2001, August). An interactive beat tracking and visualisation system. See ICMA (2001).

Dodge, C. and T. A. Jerse (1997). *Computer Music* (2nd ed.). New York: Schirmer Books.

Farnsworth, P. R. (1958). *The Social Psychology of Music*. New York: Holt, Rinehart and Winston.

Fujinaga, I. and K. MacMillan (2000, August). Real-time recognition of orchestral instruments. See ICMA (2000), pp. 141–143.

Garcia, G. (2001, August). A genetic search technique for polyphonic pitch detection. See ICMA (2001).

Ghias, A., J. Logan, D. Chamberlin, and B. Smith (1995). Query by humming — musical information retrieval in an audio database. In *Proc. ACM International Multimedia Conference*.

Hainsworth, S. W. and M. D. Macleod (2001, August). Automatic bass line transcription from polyphonic music. See ICMA (2001).

ICMA (2000, August). *Proc. International Computer Music Conference*, Berlin, Germany. ICMA.

ICMA (2001, August). *Proc. International Computer Music Conference*, Havana, Cuba. ICMA.

Kashino, K. and H. Tanaka (1993). A sound source separation system with the ability of automatic tone modeling. In *Proc. International Computer Music Conference*, pp. 248–255.

Martin, K. and Y. Kim (1998). Musical instrument identification: A pattern-recognition approach. Paper read at the 136th meeting of the Acoustical Society of America.

Martin, K. D. (1996a). Automatic transcription of simple polyphonic music: robust front end processing. Technical Report 399, MIT.

Martin, K. D. (1996b). A blackboard system for automatic transcription of simple polyphonic music. Technical Report 385, MIT.

McNab, R. J., L. A. Smith, I. H. Witten, C. L. Henderson, and S. J. Cunningham (1996). Towards the digital music library: Tune retrieval from acoustic input. In *Proc. ACM Digital Libraries*.

Pachet, F. and D. Cazaly (2000, April). A taxonomy of musical genres. In *Content-Based Multimedia Information Access Conference (RIAO)*, Paris.

Parsons (1975). *The Directory of Tunes*. Cambridge, England: Spencer Brown and Co.

Trilsbeek, P., P. Desain, and H. Honing (2001, August). Spectral analysis of timing profiles of piano performances. See ICMA (2001).

Tzanetakis, G., G. Essl, and P. Cook (2001, October). Automatic musical genre classification of audio signals. In J. S. Downie and D. Bainbridge (Eds.), *International Symposium on Music Information Retrieval*, Volume 2, Bloomington, Indiana, USA.

Uitdenbogerd, A. L. (2002). *Music Information Retrieval Technology*. Ph. D. thesis, School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria, Australia. to appear.

Uitdenbogerd, A. L., A. Chattaraj, and J. Zobel. Music information retrieval: Past, present and future. In D. Byrd, J. S. Downie, and T. Crawford (Eds.), *Current Research in Music Information Retrieval: Searching Audio, MIDI, and Notation*. Kluwer. to appear.

Uitdenbogerd, A. L. and J. Zobel (1999, November). Melodic matching techniques for large music databases. In D. Bulterman, K. Jeffay, and H. J. Zhang (Eds.), *Proc. ACM International Multimedia Conference*, Orlando Florida, USA, pp. 57–66. ACM: ACM Press.

Uitdenbogerd, A. L. and J. Zobel (2002, January). Music ranking techniques evaluated. In M. Oudshoorn (Ed.), *Proc. Australasian Computer Science Conference*, Melbourne, Australia.

Welsh, M., N. Borisov, J. Hill, R. von Behren, and A. Woo (1999, November). Querying large collections of music for similarity.

The Yamaha MIBURI MIDI jump suit as a controller for STEIM's Interactive Video software Image/ine

Lindsay R. Vickery

Classical Music Department, Western Australian Academy of Performing Arts,
Edith Cowan University
email: l.vickery@cowan.edu.au

Abstract

The desire on the part of composers to extend the scope of interactive electronics into the visual domain has resulted in the development of a number of systems that use gestural information as a means for controlling video. This paper will discuss the combination of two of these technologies - the Yamaha MIBURI MIDI jump suit and STEIM's Interactive Video software Image/ine - to provide real-time control of images via a performer.

The paper includes an introduction to the Yamaha MIBURI MIDI jump suit and an overview of STEIM's Interactive Video software Image/ine. A number of different MIBURI/Image/ine interface paradigms developed by the author at STEIM are examined in light of the advantages and limitations of the two systems.

1. Introduction

> *In the mid-1990s digital cameras became more widely available to the average user, a step some regard as the 'cinematization' of video. Accompanied by more sophisticated editing equipment, especially digital non-linear systems..video production attained closer links to cinema. (Rush 1999 p.164)*

These advances have opened the doors to a profusion of different systems involving video, sound and movement. Many of these developments – Eric Singer: *Cyclops*, *Palindrome*: *EyeCon*, David Rokeby*: Very Nervous System* etc - have focussed on the use of video as a motion sensing input device for controlling sound. The development of the opposite paradigm – using movement to control video – has been driven by Techno/Dance/VJ-oriented systems such as *Arkaos* and *U&I software*'s *Videodelic*. The 'art' music/dance world has also responded to these developments with software such as *NATO (0f0003.MASHIN3KUNST), Isadora (Troika Ranch)* and *Image/ine (STEIM)*.

The other side of the movement/video equation is the method in which information is gathered from dancer(s)' movements to control interactive video. Gestural controller innovator Axel Mulder in his 1994 Technical report *Human Movement Tracking Technology* defines three types of systems for motion sensing:

- inside-in - sensor(s) and source(s) that are both on the body;
- inside-out - on-body sensors that sense artificial external sources;
- outside-in - external sensors that sense artificial sources on the body (Mulder 1994).

Cyclops, *EyeCon* and *Very Nervous System* are clearly outside-in type systems. My current work with interactive video follows the 'inside-in' paradigm, where the dancer's gestures are transduced via Yamaha's MIBURI MIDI jump suit – a product of the company's Tokyo-based experimental division, and the Video is processed by STEIM's Image/ine software.

2. Yahama's MIBURI MIDI jump suit

The MIBURI was released commercially by the Yahama Company's Tokyo-based experimental division in 1994. It conforms to what Todd Winkler refers to as the 'body sensor' group of controllers (the other are spatial sensors, acoustic models and 'new instruments') (Winkler p. 315-8).

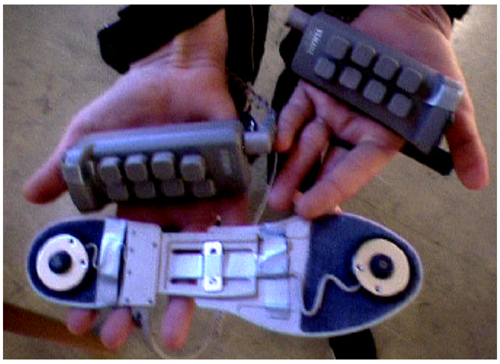


Figure 1. MIBURI handgrips and foot-sensor

The MIBURI system comprises a vest with embedded flex-sensors, two hand-grips, shoe inserts with pressure sensors, and a belt-worn signal distribution unit joined by a cable to a small synthesizer/MIDI converter. A wireless version, conforming to Japanese wireless frequency regulations was available within Japan only.

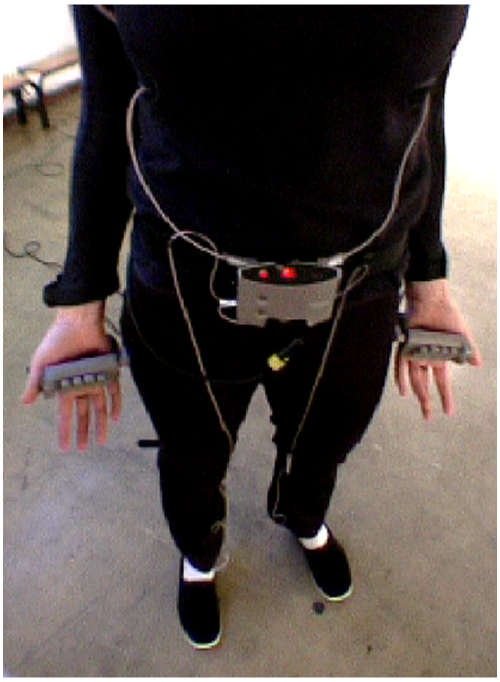


Figure 2. The MIBURI

The MIBURI's belt unit processes data from the sensors into MIDI pitch and velocity information. The unit can be programmed to interpret the data using three 'trigger' modes: 'Cross-point' mode; 'Stop' mode and 'All' a combination of both modes. 'Cross-point' mode measures the speed of the tranducer's flexion as it traverses its zero point (when the flex sensor is straight). 'Stop' mode sends note and maximum velocity values at the conclusion of a gesture. 'All' interprets sensor data in both modes simultaneously (Yamaha MIBURI Manual p.41).

Sensor type	**No.**	**trigger**	**MIDI**
pressure sensors	8	finger tips	Notes (8) / Velocity
pressure sensors	4	heel and toe	Notes (4) / Velocity
'flex' sensors	6	arm joints	Notes (12†)/ Velocity
'mod-wheel' style benders	2	thumbs	Pitchbend/ Controller
buttons	2	thumbs	Program change +/-

Figure 3. The MIBURI 's sensor array

† The six 'flex' sensors send 12 notes – this is because they measure inward and outward movement of each joint as separate notes.

The mapping of each sensor is highly programmable. Each sensor can be mapped on the synthesizer unit to any MIDI note, interpreted in any of the three modes outlined above according to 48 different response modes. The response modes (preset by Yamaha) define the manner in which the sensor's output is graphed to velocity. All the above definitions are components of a single Map 'Preset', there are 32 programmable preset positions available.

These features make the MIBURI extremely effective as a controller. However the MIBURI's synthesizer unit is limited in its possibilities as a sound source and more importantly is only able to process gestures in a direct relationship to the sounds they produce.

Teresa Marrin in her 1996 MIT Master's thesis *Toward an Understanding of Musical Gesture* assesses these shortcomings:

'One place where the Miburi might be strengthened, however, is in its reliance on triggers; the synthesizer unit which plays the music has no "intelligence" other than to provide

a one-to-one mapping between actions and sounds. In fact, the Miburi's responses are as direct as the Theremin's. But because it has so many more controls available for the user than the Theremin it might be advantageous to make use of those controls to create or shape more complex, autonomous musical processes

(Marrin 1996).

For this reason, I have chosen to combine the MIBURI with more sophisticated sound sources and software-based interactive mapping in MAX/MSP and Image/ine. (Note: MAX/MSP mapping will not be discussed in this paper.) The need to 'tether' the MIBURI to its synthesizer box is also clearly a drawback for movement detection and a restriction for the dancer. However, the MIBURI has the robust design, and very predictable sensor output that might be expected from one of the principal electronic instrument manufacturers.

The Miburi represents a rare excursion by a large, commercial organization into the uncharted waters of interface design.

(Marrin 1996)

3. Image/ine

Image/ine was developed at STEIM in the mid-1990s by Tom Demeyer. It is designed to allow the interactive manipulation and processing of live video, QuickTime movies, still images and text in real-time. Image/ine allows control of a number of parameters via a range of controllers including MIDI, the computer ASCII keyboard, mouse position (also the output of a WACOM table) and audio-in.

Image/ine allows three layers or channels of images to be separately manipulated. The layers are called *Foreground*, *Background* and *Displace Source* and each is capable of supporting all of the image sources listed above. The image input sources are viewed via the formats Video Signal, QuickTime Movie, Buffer, Text, Frame or Drawing.

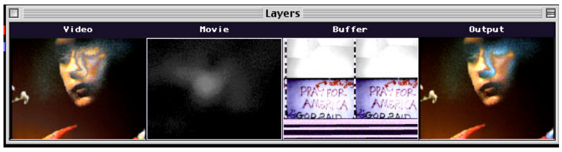


Figure 4. Image/ine's Layer Window from the author's *your sky is filled with billboards of the sky*

Still images are stored in an image buffer. The dimensions of the buffer are user-assignable. It can be navigated in two ways:

• **Frame Mode** divides the buffer into distinct frames (the example below is a 3x3 frame buffer) and allows the user to jump to any particular frame

• **Buffer Mode** in which a Frame with assignable dimensions is navigated around the buffer using the x/y co-ordinates of the Frame's centre.

The strength of *Image/ine,* however, lies in its ability to manipulate and blend these layers in real-time. Real-time processing is determined in the Mapping Window which defines more than 60 editable parameters and what is controlling them. These parameters include the input sources for the three layers, frame selection and buffer frame size and position as described above, as well as parameters controlling the mix between the layers, QuickTime movie playback, Drawing, Text and a number of Photoshop-like processes.

Unsurprisingly, considering its development at STEIM and developer Tom Demeyer, Image/ine is highly programmable. Each Mapping Window parameter has the option of 14 bit mapping of the control source to the parameter source via editable tables. Multiple arrangements of performance parameters can be saved via *Presets* and *Display States* that can create a 'snapshot' of the current state of parameter assignments values.

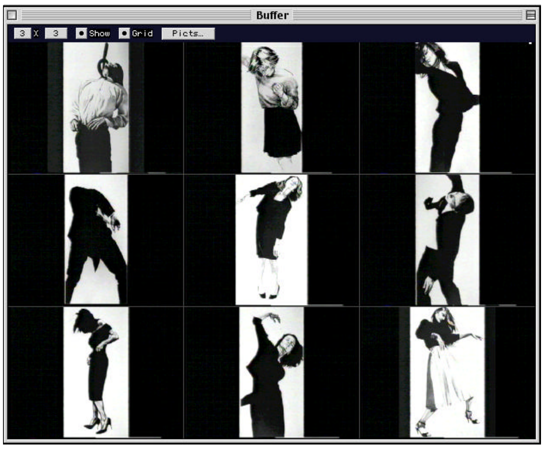


Figure 5. Image/ine's Buffer Window from the author's Image/ine interface for Amy Knoles' *Men in the Cities* 1990/2002 (images by Robert Longo)

The development of *Image/ine* came to an abrupt halt in 2000 when Demeyer resigned from STEIM, and has only recently begun to proceed with external development on a voluntary basis. This hiatus has resulted in the software being burdened with some difficulties. The most important of these are that the screen resolution has not kept up with rapid developments in digital video - *Image/ine* still operates best at 480x320 pixel resolution and therefore suffers from a pixelated 'jaggy' appearance – it is also not optimized for multi-tasking and is intolerant of other software running simultaneously.

4. Examplar Mappings

In December 2001 I had the opportunity to be in-residence at STEIM, working with Daniel Schorno on Image/ine interfaces. The pieces I have made since then focus on three of the software's strengths: the rapid access to still images available in the *Buffer* and *Frame* Modes; the use of QuickTime video in a similar way and the ability to process and blend live video.

4.1 Men in the Cities and Scan

My first foray into *Image/ine* programming was for Californian percussionist Amy Knoles' *Men in the Cities* (1990), a work she wrote in response to an invitation to perform a computer/electronic realization of the work of artist Robert Longo at the LA County Museum of Art.

The interactive video component for *Men in the Cities* used a relatively straightforward method in which nine drawings by Longo were arranged in the buffer (Figure 5). Two *Presets* corresponding to the principal sections in the work displayed the images in different ways. In *Preset* 1, MIDI note numbers simply selected a particular frame in *the Foreground Layer*, and in *Preset* 2 this same process is blended with a *Background Layer* comprising the *Buffer Frame*. The *Buffer Frame*'s position (both x and y) was determined by note's velocity and its size by pitchbend.

The resultant images (Figure 6 is an example), dynamically combine the two *Layers*, 'keying' them so that the *Background Layer* appears 'through' the white regions of the *Foreground Layer*. These two modes corresponded to sections of Knoles' composition, in which the musical material was sparse and clear (*Preset* 1) or overlaid and dense evoking the textural complexity of city life (*Preset* 2).

Figure 6. *Men in the Cities: Preset* 1 blended *Foreground* and *Background* Layers

Use of the rapid transitions possible from the Buffer also featured in my most recent Image/ine interface for the *skadada* Dance company's dancework *Scan* (2002). Like many 'New Media" works *Scan* had a number of authors: the images were provided by *Tissue Culture and Art* (Oron Catts, Ionat Zurr and Guy Ben-Ary), audio samples by John Patterson and choreography by *skadada* (Jon Burtt and Katie Lavers).

The interactive component of scan comprised only the first six minutes or so of the performance. The 25 microscopic images were accessed using a method roughly opposite to that of *Men in the Cities*. This time MIDI note number determined the *Buffer Frame* size with a result that cueing consecutive note numbers zoomed-in or out from a particular point in the *Buffer* a situation that was factored into the choreography. The zoom point was defined by the MIBURI's thumb operated toggles: left thumb gave 'x' coordinates and right thumb gave 'y' coordinates.

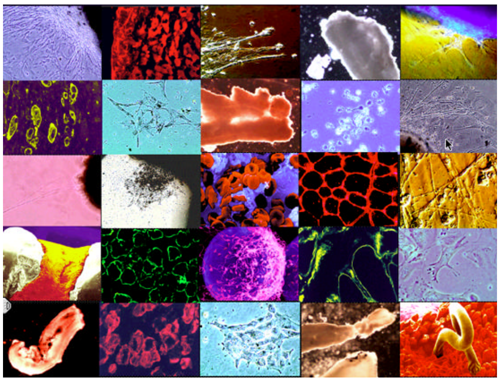


Figure 8. *Scan* Buffer
(Images by Tissue Culture and Art)

The extreme speed of the response time of the MIBURI/Image/ine combination gave the allusion of a hyper-accelerated examination of slides by an electron-microscope. The theme of surveillance/ forensic examination was also a feature of similar sections in the *Image/ine* components of two other works of mine. *Delicious Ironies (Noir)* where a down-and-out private eye dancer manipulates images and sounds from my opera *Rendez-vous* (1995) and dance piece Noir (2000). (*Noir* itself used the MIBURI to interactively control sound and a MIDI lighting desk) The other was Perth digital artist Vikki Wilson's dark meditation on the Claremont Serial Killer *a throw of the dice can never abolish chance* with images by Vikki Wilson and text by Vikki Wilson and Erin Hefferon.

4.3 a throw of the dice

Vikki Wilson initiated this project concerning the serial killer 'still at large' in the urban family paradise of Perth. The work highlights the contrast between the city's naïve sun-bathed exterior and the dark reality of continued disappearance and murder of young women. It also seeks to highlight the ambiguities of modern society's simultaneous repulsion and fascination with serial killing (as evidenced by 100s of Hollywood movies).

The final version of the work is planned to be a non-linear video installation using Finale Cut Pro as the engine to randomly juxtapose text, images and sound. Vikki had allowed me to experiment at STEIM with an interactive version of the piece where the juxtaposition of images is made by a live performer using the MIBURI as a kind of 'all-body' VJ console.

In the Image/ine component of *a throw of the dice* QuickTime videos provided by Vikki Wilson were manipulated by the dancer. The MIBURI's right hand buttons defined the current clip and the pitchbender became a fast-forward/reverse toggle causing the frame-rate to advance when pushed up and decrement when pushed back. Releasing the toggle to its centre position caused the video to stop on a particular frame.

The dancer's other hand defines the arrangement of *Layers* – principally whether the video is keyed against images in the *Buffer*. Note number determined *Buffer Frame* position, and velocity determined *Buffer Frame* size as well as the degree of keying between the Video and Buffer. In essence the dancer transitions between victim and voyeur through her control of the medium and non-control on its content.

We plan to expand the work considerably in the future, creating a hypertext-like non-linear maze of pathways in which the dancer can access only material according to her previous choices.

4.4 your sky is filled with billboards of the sky

your sky [2002] was developed for the REV festival at the Brisbane Powerhouse in April 02. Thematically, it is a sibling work to my song cycle (for soprano and DVD) **songs of virtual love+war** [1998] – exploring in a more abstract format similar issues of defining identity in the context of a world increasingly comprised of simulated experiences. In **your sky** both the sound

(MAX/MSP) and video elements are manipulated by the dancer, giving the impression that she exists in a loop in which she is called upon to respond authentically to an environment entirely under her own control.

Figure 9. Live Video Image displacement in *your sky*

The interactive nature of the technology used in **your sky** opens the possibility for an open-ended non-linear formal structure, but in this version a mix of improvisation and predetermined structure was applied. The determined formal structure rested on three principal sections: Breath, Text and Exterior each a different *Image/ine Preset*. **your sky** used a live video feed for the bulk of the performance. A brief explanation of the *Breath* section will serve to demonstrate the degree of image control achieved.

In *Breath*, the *Layers* were arranged: Foreground: video, Background: QuickTime Movie and Displace Source: buffer. A large number of parameters were assigned to the MIBURI: Note number: *Buffer* displace angle (the maximum amount the image can be twisted on a central axis); Note Velocity: Buffer Frame pan 'x'; Controller 1 : Buffer Frame Size as well as image hue/saturation and the degree of Keying between Layers; Pitchbend: Buffer Frame pan 'y' and its displace amount (how much the image can is twisted on a central axis within the 'displace angle'). The bender also determined the luminousity of the Video Input.

Figure 10. Live Video Image displacement, hue/saturation and keying against a QuickTime video in *your sky*

5. Conclusion

Interactive video is still in its infancy. There are currently a number of competing systems and it is likely that as technologies like Cinema, Television and the Internet converge the desire for reliable, high resolution software will continue. Large, well-resourced companies such as Avid and Adobe are also likely to extend their products further in this direction. There are still a number of limitations inherent in the combination of the MIBURI and Image/ine, most importantly question marks about the continuing development of both products. However together, and especially in combination with interactive audio capabilities (which I have not discussed here) they create an effective and responsive system for the development of work in this field

References

Demeyer, T. 1997. Image/ine Manual. Amsterdam, Holland: STEIM Foundation

Marrin, T.A., 1996. *Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton*, Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology. (Available at *web.media.mit.edu /~marrin/Thesis.htm*)

Mulder, A., 1994. *Human movement tracking technology,* Hand Centered Studies of Human Movement Project, Technical Report 94-1 *http://www.cs.sfu.ca/~amulder/personal/vmi/ HMTT.pub.html*

Rush, M., 1999. New Media in late 20th-Century Art. London, United Kingdom: Thames and Hudson

Yamaha 1996. MIBURI R3 Manual. Tokyo, Japan: Yamaha Corporation. (In Japanese, no author listed)

Winkler, T., 1998. Composing Interactive Music: techniques and ideas using Max. Cambridge Massachusetts, MIT Press.

Other Internet References

0f0003.MASHIN3KUNST: Nato
www.bootsquad.com/nato/nato02.html
Arkaos
www.arkaos.net/site/en/index.html
Palindrome: EyeCon
www.palindrome.de/
U&I Software: Videodelic
www.uisoftware.com/videodelic/index.html
David Rokeby: Very Nervous System
www.interlog.com/~drokeby/home.html
Eric Singer: Cyclops
www.ericsinger.com/
STEIM: Image/ine
www.image-ine.org
Troika Ranch: Isadora
www.troikaranch.org/isadora.html

The University of Waikato: Studio Report 1998-2001

Ian Whalley (Studio Director)

Music Department, The University of Waikato
email: musik@waikato.ac.nz

Abstract: *This studio report from Waikato University updates a previous one presented at ACMA97 (Whalley 1997). Covering the period 1998-2001, it outlines current developments and directions; research, composition, and collaborations; CD production; international administrative contributions; and some current speculations.*

1. Introduction: Studios & Campus

The University of Waikato first launched its major in Music in 1991, and now offers a professional BMus/MMus programme that extends to doctoral level.

The computer music studios were housed in two purpose built rooms in 1996. The two main rooms now include six sub-studios. In 2001, a multi-venue performing arts complex was opened on campus with a concert chamber wired to accommodate electroacoustic music interaction and broadcast. In 2002, a further studio room was added in the department with a suite of entry-level machines focusing on notation/composition and aural training.

The three-year undergraduate digital music programme (including Music and Computers, Composition for the Screen, and Electroacoustic Music) has now run for over ten years, supporting a graduate programme (studio based electroacoustic music, and live/interactive work) . Technically, the undergraduate programme covers sequencing and DA based composition, sample manipulation and synthesis generation, and work with MAX/MSP as well as generative systems. The undergraduate courses also contribute to a conjoint multimedia programme run by the Department of Computer Science. Graduate work is focused on the IRCAM suite of programs.

Ian Whalley established and directs the studio teaching and research programme. In 2000, Senior Tutor Lisa Meridan-Skipp assisted him, later appointed as lecturer and Studio Director at Victoria University of Wellington beginning in 2001. Composer Michael Williams took up the position from 2002. The current studio technician is Kim Walker, an accomplish percussionist with both music and computer science qualifications.

As the technology stream of the music programme has become a defining characteristic of the Department, we are fortunate to be well resourced with regular software and hardware upgrades, and can afford to be experimentally focused. The studios are designed to focus on digital music generation and composition, rather than recording.

The current studio configurations are on the website (www.waikato.ac.nz/music). They are supported by a 1Gig network connected to the Department’s server. An advantage is that a studio can be individually configured for each user as they log on, and the 'studio' is in reality a network of computers that can be accessed both on and off campus.

An emerging issue facing us, as with many academic studios, is that hardware and technological literacy is becoming less a defining difference between professional and home studios. The differentiation is now largely made through the range of software used, the exploration of non-commercial aesthetics, and the development of skills to explore new composition strategies the technology affords. Artistic outcomes, enquiry and discussion then remain the primary focus.

Waikato University was the original gateway for the internet into New Zealand, and we are fortunate to maintain a first rate digital infrastructure.

A further noted international strength of the campus is its Computer Science Department, involved in a wide range of government funded and commercial research projects, and responsible for various spin-off companies. Computer music remains a strong interest in the Department and it makes ongoing international contributions. Much of this work centred on the output of Ian Witten (machine learning, digital libraries [Witten 1999]), David Bainbridge (optical musical recognition music information retrieval [Bainbridge 1997],) and the now departed Lloyd Smith (artificial intelligence applications and web based melody indexing services).

2. Studio Research, Contributions& Collaborations

2.1 Research and Composition

Studio research has centred on the interests of the director, tutors, and collaborators. Output covers a wide gambit, and has been extensively published internationally.

Digital music aesthetics (Whalley 1998b, 1999c, 2000m, 2001e) as an aspect of the problems faced by computer musicians is a continuing interest, as are issues of technology facing contemporary composers, such as the application of physical modelling synthesis or web-based music distribution (Whalley 1998c, 2000b, 2001c).

An interest has been in generative systems in music (Whalley 1999a&b) particularly in the application of system dynamics modelling as a way of simulating and generating new works.

This work formed the basis of an interest in building system dynamic models of emotional/cognitive response in computer music (Whalley 2000a&e, 2001g) and my being invited to lead the cognition workshop at ICMC2000 in Berlin (Whalley 2000d).

The 0-0 approach to cognition using system dynamics software afforded the development of a number of models that could be used as the basis for a closed system non-linear composition system, embedded in interactive artworks. Further work in this area is continuing, and the University has now registered a provisional patent over the prototype developments (Whalley 2001f).

Composition remains at the heart of the studio, and the large number of students passing through the studios contribute to a growing body of work contributed to local and international events. Increasingly, students are now producing works for mix media, and for live interactive performances. Briefly, recent staff contributions include *Ga no mita mono* (Whalley 2000c) written at Kunitachi's Centre for Computer Music and Music Technology; IZUII (Whalley 1998) premiered at ACMA; and Lissa Meridan's *twitter tourniquet* was released on CD in 2001 (Meridan 2001).

2.2 Reviews & Professional Publications

Reviews and professional documentation is part of the ongoing work associated with the studio. Over the period I wrote extensive book, software and CD reviews for a range of publications such as *Array Live* (ICMA), *The Computer Music Journal* (MIT), *Music In New Zealand*, *Contexts* (AUST). A sample of these from 2000 is given in the reference section (Whalley 2000 h-l). Although time consuming (for example a recent CD review of 'Southern Cones' from *Leonardo Music Journal* took four days), it provides a way of keeping abreast of the range of material being released, and integrating our activities back into the international circuit.

Documenting the work of other New Zealand composers involved in computer music is also ongoing. Recent contributions include articles on Kit Powell, now resident in Switzerland, and Matthew Suttor, now resident in the United States (Whalley 2000f, 2001b).

2.3 Visitors and Collaborations

Recent studio visitors who gave lectures and interviews included Warren Burt in 2001 (Meridan-Skipp 2001).

Outward visits independent of conference and festival contributions, included my being Composer-in-residence at Kunitachi's International Centre of Computer Music and Music Technology (Tokyo 1998), Visiting Fellow in the Computer Science Department at Meiji University (Tokyo 2001), and awarded a York/Waikato Visitors in 2001 (yet to be completed).

The visit to Meiji allowed the continuation of work in generative systems through applying artificial intelligence programmes to non-linear models of music cognition.

The Waikato Campus recently launched the award winning graphics and multi-media degree programme under the direction of Ian Gwilt. This is based on the conjoint programme run for some time in conjunction with Wanganui Polytechnic. The cluster in Digital Arts combining our studio, computer science, and computer graphics affords the ongoing development of a teaching and research programme where there is an alignment of international expertise (Whalley 2001o). This recently allowed us to bid for a Centre of Excellence grant as part of the Waikato (host) Auckland/Auckland University of Technology Information and Communication Technology proposal. The development of the cluster is ongoing, and more staff with an interdisciplinary approach and skill have recently been hired to support the development.

3. CD Production

With the major studio upgrades that took place in 1999, the main studio was ideally suited to CD mastering.

Through Waikato University acting as publisher, I led the New Zealand Sonic Art CD series initiative as editor and producer. The first call for works for the first CD went out in 1999, resulting in the *New Zealand Sonic Art 2000* (Whalley 2000d). This included nine composers, and aimed at updating the series of electroacoustic works last released by Ode Record Company in 1994.

There were far more submissions than space allowed, and the CD attempted to represent some of the diversity of work being undertaken, and balance new with established composers.

New Zealand Sonic Art Vol.II (Whalley 2001d) included six composers and was released the following year, dedicated to the memory of the country's electroacoustic music pioneer Douglas Lilburn who recently passed away. The dedication restricted the outcomes somewhat to a 'tape pieces' approach, with many of the works being more reflective than the first CD.

Both discs have been widely internationally distributed, and received many positive reviews from commercial sources such as *The Wire* UK, and academic sources such as SAN UK, and EMF. Both CDs also made the New Zealand Concert FM listening charts, finding a wider audience for the content outside the traditional computer music conference circuit.

A call for works for the third CD was released early in 2002, with the aim of completing the master by the end of the year. The focus of this third CD is slightly different than the first two, so as to encourage a wider variety of artists to submit material, use more local source material, and move beyond the approach dominant in the first two CDs. I hope to include works from Maori composers and include live instruments with soundscape processing more typical of the upper North Island aesthetic.

4. Administrative Contributions

In 1999 the studios took charge of administering the web site for ACMA. This involved registering acma.asn.au in Australia, redesigning the site, and mounting it on local servers. Electronic versions of *Chroma* were also added. The site averages around 180 hits a month, with most interest being in downloading *Chroma*. Hits increase greatly when the annual conference is announced.

In the same year we formed a new listserv group (acma-l) as a service to the Australasian Computer Music Community. This incorporated the oz-computermusic and New Zealand Sonic Art lists. The list currently has around 140 members, quite in excess of current ACMA membership. Traffic is not heavy. Perhaps current state of government support for computer music in Australia is partly responsible for this ?

Other regional activities involve acting on the reading panel for *Micropolyphonie*, and acting as the regional editor for *Organised Sound* (Cambridge University Press).

Making and administrative contribution to computer music internationally has been an ongoing process. For example, in 2000 I sat on the final paper selection committee for the International Computer Music Association Conference in Berlin. The panel for took more than a month of work, selecting 133 out of 375 papers from international field. The process was a difficult one because of the high level of most work.

In 2001 I co-ordinated an issue of *Organised Sound* focusing on Computer Music in Australasia/South East Asia (Whalley 2001a). The request for submissions began by noting that along with increasing globalisation, each world region has its distinct characteristics, and that Australasia/South East Asia represents an exciting crossroads of established, traditional and new approaches. The questions authors were asked to address included: Are there distinct national or regional aesthetics in electroacoustic and computer music in this corner of the globe? Which developments in relevant areas of music technology are taking place or are about to take place in Australasia/South East Asia? Are they related to distinct aesthetic ideas, compositional procedures, technological applications, and/or technological developments? Are there environmental or cultural circumstances that influence computer music in this region?

Authors from different areas submitted summaries of developments, cultural/economic backgrounds, events and approaches, written either individually or collectively. They also dealt with wider thematic and national issues as they say fit.

Nine papers were selected: two from New Zealand, one from SEA, and six from Australia. The material chosen attempted to cover most of the region and the themes that arose, although some geographical areas of Australia are neglected through a lack of material or it being insufficiently addressed to the central themes. The responses were as diverse and the region itself, and the collection captures some of the similarities, contradictions and differences that one would expect to emerge in a complex, evolving and diverse region. (Whalley 2001a). For the Auckland/Waikato summary, see (Elmsly, Dart, Whalley 2001).

5. Speculations

The New Zealand Government recently released a new *Strategy for Tertiary Education 2002-2007*, signalling its intention to focus university activity on research outcomes with links to industry to underpin its move towards a 'knowledge economy'. This will require greater networking and collaboration between universities, recognising each other's recent unique 'profile' and expertise: a change from the competitive model based on funding undergraduate student numbers. Within this new context, computer music/electroacoustic music can find a good niche nationally, integrated into the international computer music community.

Given the current developments on Campus, and the increasing integration of digital arts, our programme is likely to become increasingly web based, non-linear, and multimedia based, in tandem with increasing specialisation in computer music composition, interaction, and research.

Increasing government support for the creative industries also is likely to see the use of more local source material in products, a path more of our ex-students are discovering with increasing success outside academic music. This is also a part and reflection New Zealand beginning to finding an international artistic niche in a variety of new artistic media.

Selected References: 1997-2001

Bainbridge, D., N. Carter, N. 1997. Automatic Recognition of Music Notation, *Handbook of Optical Character Recognition and Document Image Analysis*, H. Bunke and P. Wang (eds), World Scientific, Singapore: 557-603.

Meridan-Skipp, L. 2001. Imaginary Space/Warren Burt Interview. *Canzona 99,* New Zealand Composers Association Yearbook, 2000.

Meridan, L. 2001. twitter tourniquet (2000) *New Zealand Sonic Art II*. UWMD. University of Waikato.

Elmsly, J. Dart, W. Whalley, I. 2001. A view of Computer Music from New Zealand: Auckland, Waikato, and the Asia/Pacific connection. *Organised Sound* 6 (1): 11-20

Whalley, I. 1997. Studio Report. Digital music studios, Music Department, The University of Waikato. *The Fifth Annual Conference of the Australian Computer Music Association*, University of Auckland. 11 - 13 July, 4 p.

--1998a. Izu II. (7'04) Premiered at ACMA98. *The Australasian Computer Music Association Sixth Annual Conference*. Australasian Centre for Arts and Technology. Australian National University, Canberra, Australia, 10-12 July.

--1998b. Digital Music Technology and Artistic Response Since 1980. Music Sessions. *Music Sessions, Crossroads in Cultural Studies: Second International Conference*. Tampere, Finland. June 28-July 1: 1-18.

--1998c. Web Sites for Music : Intention and Uses. Convergence. *The Journal of Research into New Media Technologies*. University of Luton Press. Volume 3, Number 4, Winter 1997. p. 21-26.

--1999a. Beyond Pitch/Duration Scoring: Towards A System Dynamics Model of Electroacoustic Music. *Proceedings of 17th International Conference of the*

System Dynamics Society. Wellington, 20-23 July: 73-83.

--1999b. Thematic Juxtaposition and Resolution in Computer Music: A System Dynamics Model. *Proceedings of 1st Iteration Conference in Generative Systems in the Electronic Arts*. Monash University, Australia, 1-3 December: 158-174.

--1999c. Towards a New Aesthetic in Electroacoustic Music Practice in New Zealand Universities. *Enculturation* 2(2): 1-15.

--2000a. Modelling Emotional Narrative through System Dynamics: Applications to Computer Music. *Proceedings of 13th Colloquium on Musical Informatics*. L'Aquila, Italy. 2-5 September: 213-216.

--2000b. Real time To Hyper-Real Gestures: Expanding Instrumental Expression Through Physical Modelling Synthesis. *Proceedings of The 10th International IASPM Conference*. University of Technology, Sydney, Australia: 395-401

--2000c. Broadcast of Ga no mita mono (What the moths saw) (14'30). *KAJX, Aspen*, USA, 17 January. Original work commissioned by Kunitachi University of Music as part of International meet the composer series, premiered 3rd Dec. 1998.

--2000d. Cognition and Perception Issues in Computer Music. Workshop at ICMC2002. *Proceedings of the International Computer Music Conference* 26-28 August, Berlin: 81.

--2000e. Emotion, Theme and Structure: Enhancing Computer Music through System Dynamics Modelling. *Proceedings of the International Computer Music Conference*. Berlin: 49-52.

--2000f. Kit Powell: Changing Approaches and New Zealand Sounds. *Music in New Zealand,* 36, 30-31.

--2000g. (Ed.) New Zealand Sonic Art 2000. University of Waikato. UWMD1200

--2000h. Review of CD: Darren Copeland Rendu visible. empreintes DIGITALes CD: IMED 9841, 1998, in, *Array*, International Computer Music Association, November.

--2000i. Review of CD: Yves Daoust Musiques Naives Musiques Naives. empreintes DIGITALes IMED 9843, 1988, in, *Array*, International Computer Music Association Journal, November.

--2000j. Review of Music and Signs: Semiotic and Cognitive Studies, in, *Music in Contexts Journal of New Music*, Melbourne, 18: 99-102.

--2000k. Review of Musical Visions, Wakefield Press. *Media International Australia*. Boustien, G. (ed.) No. 95, May 2000: 260-261.

--200ll. Review of CD: Ned Bouhalassa Aérosol. empreintes DIGITALes CD: IMED 8940, 1998, in, *Array*, International Computer Music Association , November.

--2000m. Tales, Tellers, and Taking Part: The Computer Music/audience dialectic. *Interfaces, Proceedings of the Australasian Computer Music Association Conference*. 5-8 July, Queensland University of Technology. Brisbane, Australia: 92 -98.

--2000o. Digital Arts in the Knowledge Economy: Towards a New Zealand Centre for Arts and Technology. Submission to the Tertiary Education A.C. http://www.teac.govt.nz/submissions/sec%2049.pdf

--2001a. Editorial. Computer Music in Australasia and South East Asia. *Organised Sound* 6 (1): 1-2

--2001b. Matthew Suttor: Visualising Music and Things Theatrical. *Music in New Zealand*. 38:42-4.

--2001c. Music and Recent Internet Developments: Opportunities and Perils for New Zealand Composers *Canzona 99,* New Zealand Composers Association Yearbook, 2000.

--2001d. (Ed.) *New Zealand Sonic Art Vol. II*. University of Waikato. UWMD1201

--2001e. Redundancy, Innovation and Integration: Digital Music Techniques and Neo-World Music . B. Enders /J. Stange-Elbe (eds.): *Global Village - Global Brain - Global Music, KlangArt 99* Osnabrueck University Press 2001: 190-203.

--2001f. *ENACS* Emotion Non-linear Automated Composition System. Provisional Patent Lodged by Waikato University Unilink Office.

--2001g. Systems Dynamics Modelling: Applications to Computer Music. *Organised Sound*, 5(3) 149-157.

Witten, I. Manzara, L., Conklin, D. 1999 Computational Models of Music Prediction, IEEE *Computer Society Press*, Los Alamitos, CA.